Probabilités
et
inférence statistique

Notes à l'usage des étudiants de 2ème année de bachelier
en sciences économiques, sciences de gestion et ingénieur de gestion

Année académique 2005 – 2006

Préambule

Ce syllabus doit beaucoup à des notes de cours originales rédigées par Laurence Broze, Professeur à l'Université de Lille III, dont il reprend, en les aménageant et en les complétant, la structure générale, des extraits du texte et la plupart des exercices. Je voudrais ici la remercier de m'avoir gracieusement autorisé ces emprunts.

En complément de ce syllabus, l'étudiant peut utilement consulter :

- Wannacott T.H.et Wannacott R.J.(1991), *Statistique*, 4e édition, Economica.

L'étudiant désireux d'en savoir plus peut compulser (par ordre croissant de difficulté) :

- Droesbeke J-J. (1997), *Eléments de Statistique*, 3e édition, Editions de l'Université de Bruxelles / Editions Ellipses.
- Lecoutre J-P. (1998), *Statistique et Probabilités*, Dunod.
- Saporta G. (1990), *Probabilités, Analyse des Données et Statistique*, Editions Technip.

Bernard Lejeune

Table des matières

Partie I

Probabilités

Chapitre 1

L'aléatoire et les probabilités

1.1. Epreuve aléatoire

La théorie des probabilités offre un cadre d'analyse permettant d'appréhender ce qu'elle appelle dans son jargon une *épreuve* (ou *expérience*) *aléatoire*.

Définition 1.1 *On entend par épreuve aléatoire une situation, une action ou encore un processus qui peut donner lieu à différents résultats et dont on ne peut dire à l'avance lequel de ces résultats va se produire.*

Le sexe des enfants à naître d'un couple qui envisage d'avoir 3 enfants, le lancer de 2 dés, la durée de vie d'une ampoule, sont des exemples d'épreuves aléatoires.

Plus précisément, la théorie des probabilités permet d'une part de décrire une épreuve aléatoire, et d'autre part, sur cette base, de calculer les chances de réalisation — les *probabilités* — d'*événements* relatifs à cette épreuve aléatoire. Les chances d'avoir au moins une fille, que la somme des résultats des dés soit égale à 4, que la durée de vie de l'ampoule soit comprise entre 800 et 900 heures, sont des exemples d'événements relatifs aux épreuves aléatoires évoquées ci-dessus.

Dans ce chapitre, on décrit les éléments de base du cadre d'analyse que constitue la théorie des probabilités.

1.2. Ensemble des résultats possibles

On suppose que l'on peut décrire l'ensemble de tous les résultats possibles de l'épreuve aléatoire. On note généralement cet ensemble Ω.

Le lancer d'un dé[1] est une épreuve aléatoire dont l'ensemble des résultats

[1] Notons que le mot "aléatoire" vient du latin *alea* qui signifie "jeu de dé".

possibles est :

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Si on lance deux dés, on a 36 $(= 6^2)$ couples (D_1, D_2) — où D_1 est le résultat du premier dé et D_2 celui du second — de résultats possibles :

$$\Omega = \{(1, 1), (1, 2), ..., (2, 1), (2, 2), ..., (6, 6)\}$$

De même, le sexe des enfants à naître d'un couple qui envisage d'avoir 2 enfants est une épreuve aléatoire qui a 4 $(= 2^2)$ couples (S_1, S_2) — où S_1 est le sexe du premier enfant et S_2 celui du second (chaque fois Fille ou Garçon) — de résultats possibles :

$$\Omega = \{(F, F), (G, F), (F, G), (G, G)\}$$

Le nombre de résultats possibles d'une épreuve aléatoire peut être *fini*, comme dans le cas du lancer d'un dé, ou *infini*, comme par exemple dans le cas de la durée de vie d'une ampoule qui a priori peut prendre toutes les valeurs de l'intervalle $[0, +\infty)$. Dans la suite, nous supposerons que ce nombre est fini. Nous ne relâcherons cette hypothèse que dans le Chapitre 4, lorsque nous étudierons les variables aléatoires.

1.3. Distribution de probabilité

Lorsqu'on lance une pièce de monnaie, il y a deux résultats possibles : $\Omega = \{\text{pile,face}\}$. Si la pièce est n'est pas truquée (est équilibrée), il n'y a pas de raison de s'attendre à ce que l'un se réalise plutôt que l'autre. On les considère donc comme *également probables* ou *équiprobables*. On dit communément qu'on a une chance sur deux d'obtenir pile et une chance sur deux d'obtenir face. En termes de probabilités, on dira que la probabilité d'obtenir pile est de $\frac{1}{2}$ et que la probabilité d'obtenir face est aussi de $\frac{1}{2}$:

Résultats possibles	pile	face
Probabilités	$\frac{1}{2}$	$\frac{1}{2}$

De même, lorsqu'on jette un dé, il y a six résultats possibles que l'on considère comme équiprobables si le dé n'est pas truqué. On dit alors que chacun de ces résultats a une chance sur six de se réaliser, que la probabilité d'obtenir chacun de ces résultats est de $\frac{1}{6}$:

Résultats possibles	1	2	3	4	5	6
Probabilités	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Quid si la pièce et le dé sont truqués ? Cette conjecture peut s'exprimer en termes de probabilités. Dans le cas du lancer de la pièce, si on suspecte que la pièce n'est pas bien équilibrée, on dira que les chances d'obtenir pile sont pas égales aux chances d'obtenir face, par exemple qu'on a 6 chances sur 10 d'obtenir pile et 4 chances sur 10 d'obtenir face. En termes de probabilités, on dira que la probabilité d'obtenir pile est de $0,6$ $(=\frac{6}{10})$ et que la probabilité d'obtenir face est de $0,4$ $(=\frac{4}{10})$:

Résultats possibles	pile	face
Probabilités	$0,6$	$0,4$

De même dans le cas d'un dé pipé, on dira que les probabilités d'obtenir chacun des résultats possibles ne sont plus égales à $0,1666...$ $(=\frac{1}{6})$, mais par exemple (en supposant que la face correspondant au 1 a été lestée et que le 6 est à l'opposé du 1), qu'on a respectivement $6, 10, 10, 10, 10$ et 14 chances sur 60 d'obtenir $1, 2, 3, 4, 5$ et 6. En termes de probabilités :

Résultats possibles	1	2	3	4	5	6
Probabilités	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{7}{30}$

Dans tous ces exemples, on a décrit les chances de réalisation des différents résultats possibles de l'épreuve aléatoire en attribuant aux divers résultats possibles des nombres compris entre 0 et 1, dont la somme est égale 1.

La généralisation à partir de ces exemples est évidente.

Définition 1.2 *Soit une épreuve aléatoire ayant N résultats possibles désignés par i $(i = 1, 2, ..., N)$. On appelle distribution de probabilité (sur Ω) pour cette épreuve aléatoire un ensemble de N couples (i, p_i), où les p_i sont des nombres réels tels que :*

$$\begin{cases} 0 \leq p_i \leq 1, \quad \forall\, i = 1, ..., N \\ p_1 + p_2 + ... + p_N = \sum\limits_i p_i = 1 \end{cases}$$

Le couple "ensemble des résultats possibles – distribution de probabilité" constitue un *modèle de probabilité* d'une épreuve aléatoire.

1.4. Evénement et probabilité d'un événement

Un modèle de probabilité pour une épreuve aléatoire étant défini, l'objet du calcul des probabilités est de déterminer la probabilité d'*événements* relatifs à cette épreuve aléatoire.

Définition 1.3 *On entend par événement une proposition logique relative à l'issue d'une épreuve aléatoire, qui peut ou non se réaliser selon le résultat de l'épreuve aléatoire.*

On dit qu'un événement est réalisé ou non selon que, au vu du résultat de l'épreuve aléatoire — la réalisation d'un des résultats possibles —, la proposition est vraie ou fausse.

Par exemple, lorsqu'on prend l'épreuve aléatoire du jet d'un dé, les événements suivants peuvent être considérés : "obtenir un 6", "obtenir un nombre pair", "obtenir un nombre positif ou nul", ...

L'événement "obtenir un 6" se réalise lorsque le résultat de l'épreuve aléatoire est un 6. Il ne se réalise pas lorsque le résultat de l'épreuve aléatoire est $1, 2, 3, 4$ ou 5. L'événement "obtenir un nombre pair" se réalise lorsque le résultat de l'épreuve aléatoire est $2, 4$ ou 6. Il ne se réalise pas lorsque le résultat est $1, 3$ ou 5. L'événement "obtenir un nombre positif ou nul" se réalise dans tous les cas.

On constate que certains événements ont un statut un peu particulier : ils s'identifient exactement avec un des résultats possibles de l'épreuve aléatoire. C'est le cas de "obtenir un 6". De tels événements sont appelés *événements élémentaires* (ou *simples*). Dans le cas de l'épreuve aléatoire du lancer du dé, il y a six événements élémentaires qui sont : "obtenir un 1", "obtenir un 2", ..., "obtenir un 6".

Comment calculer la probabilité $\mathbb{P}(A)$ d'un événement quelconque A relatif à une épreuve aléatoire ?

Examinons tout d'abord le cas le plus simple où les différents résultats possibles de l'épreuve aléatoire sont équiprobables.

Supposons qu'on jette un dé non truqué. Par définition, la probabilité des différents résultats possibles — les événements élémentaires — est égale à $p_i = \frac{1}{6}$: chaque résultat possibles a 1 chance sur 6 de se réaliser. Considérons maintenant l'événement $A =$ "obtenir un nombre pair". Cet événement se réalise si le résultat de l'épreuve aléatoire est $2, 4$ ou 6, autrement dit, pour 3 résultats sur 6. Cet événement a donc 3 chances $(1 + 1 + 1)$ sur 6 de se réaliser ou, ce qui revient au même, une probabilité de réalisation $\mathbb{P}(A) = 0,5$.

La généralisation à partir de cet exemple est claire.

Définition 1.4 *Soit une épreuve aléatoire ayant N résultats possibles, tous équiprobables, et un événement quelconque A relatif à cette épreuve aléatoire. On appelle probabilité de l'événement A :*

$$\mathbb{P}(A) = \frac{N_A}{N}$$

où N_A est le nombre des résultats possibles de l'épreuve aléatoire pour lesquels A se réalise.

Ainsi, la probabilité d'obtenir pile en lançant une pièce équilibrée est $\frac{1}{2} = 0,5$. La probabilité d'obtenir 6 en lançant un dé non truqué est $\frac{1}{6} = 0,166...$; la probabilité d'obtenir un résultat supérieur ou égal à 4 est $\frac{3}{6} = 0,5$.

Comme autre exemple d'une épreuve aléatoire de ce type , considérons le cas du tirage au sort d'un élève parmi les 40 élèves d'une classe composée de 16 garçons et de 24 filles. Ce tirage est réalisé en mettant les noms des élèves sur 40 bulletins mis dans un sac et en tirant un bulletin au hasard dans le sac. On considère l'événement $F =$

"l'élève tiré est une fille". Le tirage est une épreuve aléatoire ayant 40 résultats possibles ($N = 40$), tous équiprobables (il n'y a pas de raison qu'un nom sorte plutôt qu'un autre) et l'événement F se réalise pour 24 de ces résultats ($N_A = 24$). L'événement F a donc 24 chances sur 40 de se réaliser. Sa probabilité est :

$$I\!P(F) = \frac{24}{40} = \frac{6}{10} = 0,6$$

Passons au cas où les différents résultats possibles de l'épreuve aléatoire ne sont pas équiprobables.

Considérons à nouveau l'épreuve aléatoire du jet d'un dé et l'événement $A =$ "obtenir un nombre pair", mais supposons cette fois que le dé est pipé et que la distribution de probabilité des différents résultats possibles est :

Résultats possibles	1	2	3	4	5	6
Probabilités	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{7}{30}$

On l'a vu, cette distribution indique qu'on a respectivement $6, 10, 10, 10, 10$ et 14 chances sur 60 d'obtenir $1, 2, 3, 4, 5$ et 6. Notez qu'en disant cela, on s'est en fait ramener au cas d'une épreuve aléatoire dont les résultats possibles sont équiprobables et qui consiste à tirer au sort une boule dans une urne contenant 6 boules marquées du numéro 1, 10 boules marquées du numéro 2, du numéro 3, du numéro 4 et du numéro 5, et 14 boules marquées du numéro 6. Cette opération mentale est possible chaque fois que l'ensemble des p_i de la distribution sont des nombres rationnels (c'est-à-dire peuvent s'écrire comme le rapport de deux nombres entiers.)

L'événement $A =$ "obtenir un nombre pair" se réalise si le résultat de l'épreuve aléatoire est $2, 4$ ou 6. Cet l'événement a donc 34 chances $(10 + 10 + 14)$ sur 60 de se réaliser ou, ce qui revient au même, une probabilité de réalisation $I\!P(A) = \frac{17}{30} = 0,566....$ Cette probabilité n'est autre que la somme des probabilités p_i des résultats $2, 4$ et 6 : $\frac{1}{6} + \frac{1}{6} + \frac{7}{30} = \frac{17}{30}$.

A nouveau, la généralisation est évidente.

Définition 1.5 *Soit une épreuve aléatoire ayant N résultats possibles désignés par i $(i = 1, 2, ..., N)$ dont la distribution de probabilité est donnée par un ensemble de N couples (i, p_i) et soit un événement quelconque A relatif à cette épreuve aléatoire. On appelle probabilité de l'événement A :*

$$I\!P(A) = \sum_{i \in A} p_i$$

où $\sum\limits_{i \in A}$ signifie "somme pour tous les résultats possibles i pour lesquels A se réalise".

Ainsi, pour le dé truqué dont la distribution est donnée ci-dessus, la probabilité d'obtenir 6 est $\frac{7}{30} = 0,233...$; la probabilité d'obtenir un nombre supérieur ou égal à 4 est $\frac{1}{6} + \frac{1}{6} + \frac{7}{30} = \frac{17}{30} = 0,566...$; la probabilité d'obtenir un multiple de 3 est $\frac{1}{6} + \frac{7}{30} = \frac{12}{30} = 0,4$; etc...

Il va sans dire la Définition 1.4 est simplement un cas particulier de la Définition 1.5 : il suffit de prendre $p_i = \frac{1}{N}$.

La probabilité de tout événement A est la somme de tout ou partie des p_i. Par définition, on a $0 \leq p_i \leq 1, \forall\, i = 1, ..., N$ et $\sum\limits_i p_i = 1$. On en déduit la propriété suivante.

Propriété 1.1 *La probabilité d'un événement A quelconque est un nombre compris entre 0 et 1 :*

$$0 \leq I\!P(A) \leq 1$$

Remarque : Dans une copie d'examen, si vous trouvez une probabilité plus grande que 1 (ou négative), c'est que vous avez fait une erreur, et vous devez le dire. Sinon, vous serez sévèrement pénalisé. On peut être indulgent pour une erreur de calcul, mais pas lorsqu'on écrit sans sourciller quelque chose d'absurde.

1.5. Exercices

1- Une boule est tirée au hasard d'une boîte contenant 4 boules rouges, 6 boules blanches, 2 boules vertes et 8 boules noires. Déterminez la probabilité qu'elle soit :

   a- rouge.
   b- verte.
   c- non blanche.
   d- verte ou noire.
   e- noire ou blanche.

2- On considère un jeu de 52 cartes dans lequel on tire au hasard une carte. Calculez la probabilité des événements suivants :

   a- obtenir la reine de coeur.
   b- obtenir un roi.
   c- obtenir une carte qui ne soit pas un trèfle.
   d- obtenir un trèfle ou un carreau.
   e- ne pas obtenir un valet.

3- Un groupe d'étudiants est formé de 20 étudiants de première année (10 filles et 10 garçons) et de 30 étudiants de deuxième année (18 filles et 12 garçons). On choisit une personne au hasard dans ce groupe. Déterminez la probabilité qu'elle soit :

   a- de première année.
   b- un garçon.
   c- une fille de deuxième année.

4- Une classe de 100 étudiants se compose des quatre groupes suivants :

	Garçons	Filles
Inscrit en gestion	17%	38%
Inscrit en économie	23%	22%

On choisit un étudiant au hasard comme délégué de classe.

a- Quelle est la probabilité que ce soit :

i- un garçon ?
ii- une fille ?
iii- un(e) étudiant(e) de gestion ?
iv- un garçon inscrit en économie ?
v- un(e) étudiant(e) de gestion ou un garçon ?
vi- une fille non inscrite en gestion ?
vii- ni un(e) étudiant(e) d'économie, ni une fille ?

b- Que deviennent ces résultats si on ne précise pas l'effectif de la classe ?

5- Sur la base de l'inventaire communal réalisé par l'INSEE (Institut National de la Statistique et des Etudes Economiques en France) en 1979, on peut établir le tableau suivant qui croise, pour les 36 414 communes métropolitaines, la fréquence de ramassage des ordures ménagères, durant l'hiver 1979, avec la taille de la commune.

Fréquence de ramassage	Taille de la commune (nb. hab.) moins de 500	500 à 2 000	2 000 à 100 000	plus de 100 000	Total
Quotidienne	260	403	807	26	1 496
2, 3 ou 4 fois par semaine	2 622	3 221	2 357	12	8 212
1 fois par semaine	9 352	5 377	335	0	15 064
1 fois par quinzaine ou moins	4 217	704	7	0	4 928
Jamais	6 479	235	0	0	6 714
Total	22 930	9 940	3 506	38	36 414

a- On tire une commune au hasard. Quelle est la probabilité qu'elle compte moins de 500 habitants et qu'elle bénéficie d'une collecte quotidienne ?

b- On tire au hasard une commune de plus de 100 000 habitants. Quelle est la probabilité qu'elle bénéficie d'une collecte quotidienne ?

c- On tire au hasard une commune qui bénéficie d'une collecte par semaine au moins. Quelle est la probabilité qu'elle compte moins de 500 habitants ou plus de 100 000 habitants ?

6- L'enquête sur l'emploi réalisée par l'INSEE en 1981 permet de construire le tableau suivant donnant la répartition des adultes élevant seuls leurs enfants selon le sexe et la classe socio-professionnelle (en milliers).

Classe Socio-Professionnelle (CSP)	Sexe F	H
Personnel de service	99	6
Ouvriers	145	78
Employés	168	17
Cadres moyens	93	21
Prof. lib., cadres sup.	28	19
Patrons, ind., com.	30	16
Salariés agricoles	2	2
Agriculteurs	11	9
Autres actifs	2	4
Inactifs	156	22

a- On choisit au hasard un adulte élevant seul ses enfants. Quelle est la probabilité que ce soit :

i- une femme ?
ii- un homme ?
iii- une personne active ?

b- On choisit au hasard une femme élevant seule ses enfants. Déterminez la probabilité de chacune des CSP. Faites les mêmes calculs pour un homme choisi au hasard. Représentez toutes ces probabilités à l'aide d'un tableau.

c- On choisit au hasard un employé élevant seul ses enfants. Quelle est la probabilité que ce soit une femme ? Construisez un tableau donnant pour chaque CSP la probabilité d'obtenir une femme, un homme.

7- On lance deux fois une pièce de monnaie (équilibrée) et on représente le résultat de cette épreuve aléatoire par un couple (X, Y), où X est le résultat du premier lancer et Y celui du second lancer.

a- Combien y a-t-il de résultats possibles ?

b- Quelle est la probabilité :

i- d'obtenir au moins une fois face ?
ii- de ne pas obtenir un pile ?

8- On lance deux dés (non truqués) et on représente le résultat de cette épreuve aléatoire par un couple (X, Y) où X est le résultat du lancer du premier dé et Y celui du second dé.

a- Combien y a-t-il de résultats possibles ?

b- Quelle est la probabilité :

i- d'obtenir un double 6 ?
ii- de ne pas obtenir un double ?
iii- d'obtenir un résultat dont la somme vaut 11 ?
iv- d'obtenir un résultat dont la somme vaut 8 ?
v- d'obtenir un résultat dont la somme vaut au moins 8 ?
vi- d'obtenir un résultat dont la somme vaut au plus 8 ?
vii- d'obtenir un résultat dont la somme vaut plus de 8 ?
viii- d'obtenir un résultat dont la somme vaut au moins 9 ?

ix- d'obtenir une somme paire ?

c- Sur quelle valeur de la somme est-il le plus intéressant de parier ? Pourquoi ?

9- A chaque naissance, la probabilité d'avoir un garçon est égale à celle d'avoir une fille, soit $0,5$. On considère la composition d'une famille de trois enfants et on représente le résultat de cette épreuve aléatoire par le triplet (S_1, S_2, S_3) où S_1 est le sexe du premier enfant, S_2 celui du second enfant et S_3 celui du troisième enfant.

a- Combien y a-t-il de résultats possibles ?

b- Quelle est la probabilité :

i- que le deuxième enfant soit une fille et le troisième un garçon ?

ii- d'avoir au moins deux filles ?

iii- que tous les enfants soient du même sexe ?

iv- de n'avoir aucune fille ?

v- d'avoir exactement une fille ?

vi- d'avoir exactement deux filles ?

vii- d'avoir exactement trois filles ?

Chapitre 2

Le calcul des probabilités – I

Dans ce chapitre, on répond à la question suivante : étant donnés deux ou plusieurs événements ayant entre eux certaines relations logiques, quelles relations en résulte-t-il pour leurs probabilités ? La réponse à cette question fournit des règles qui permettent de calculer les probabilités de certains de ces événements, connaissant celles des autres. Jointes à celles développées dans le chapitre suivant, elles forment les règles fondamentales du calcul des probabilités.

Nous établirons ces règles en supposant que les conditions de la Définition 1.5 sont satisfaites, c'est-à-dire qu'on a affaire à une épreuve aléatoire ayant N résultats possibles désignés par i ($i = 1, 2, ..., N$), dont la distribution de probabilité est donnée par un ensemble de N couples (i, p_i), et où la probabilité d'un événement A quelconque est donc donnée par la formule :

$$\mathbb{P}(A) = \sum_{i \in A} p_i$$

où $\sum\limits_{i \in A}$ signifie "somme pour tous les résultats possibles i pour lesquels A se réalise".

On peut cependant montrer que les règles que nous aurons établies sont en fait d'application générale, valables que les résultats possibles de l'épreuve aléatoire sous-jacente soient ou non en nombre fini (comme par exemple dans le cas de la durée de vie d'une ampoule qui a priori peut prendre toutes les valeurs possibles — un continuum de valeurs — dans l'intervalle $[0, +\infty)$).

2.1. Représentation ensembliste d'un événement

On peut représenter les événements relatifs à une épreuve aléatoire en termes ensemblistes, comme des parties (sous-ensembles) de l'ensemble Ω de tous les résultats possibles. Cette représentation est très commode car elle permet de visualiser de façon précise d'une part les événements, et surtout, d'autre part, les relations logiques entre événements, qui apparaissent alors comme des relations ensemblistes entre parties de l'ensemble Ω.

Ainsi, à tout événement A, on peut associer la partie de Ω formée de tous les résultats possibles de l'épreuve aléatoire pour lesquels A est réalisé.

Par exemple, dans le cas du lancer d'un dé, à l'événement A = "obtenir un nombre pair", on peut associer la partie de Ω :

$$\mathcal{A} = \{2, 4, 6\}$$

De même, à l'événement B = "obtenir un multiple 3", on peut associer la partie de Ω :

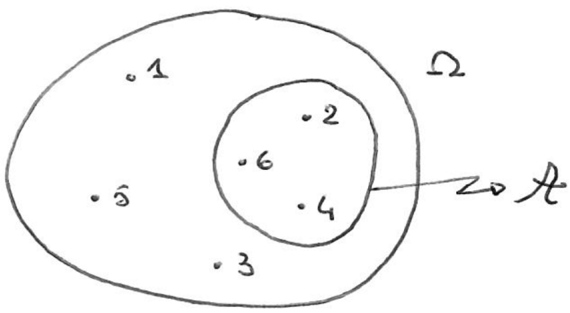
$$\mathcal{B} = \{3, 6\}$$

à l'événement C = "obtenir un nombre premier", on peut associer la partie de Ω :

$$\mathcal{C} = \{1, 2, 3, 5\}$$

et ainsi de suite.

Cette façon d'associer à tout événement un sous-ensemble de Ω permet un représentation graphique commode de cet événement sous la forme d'un diagramme :



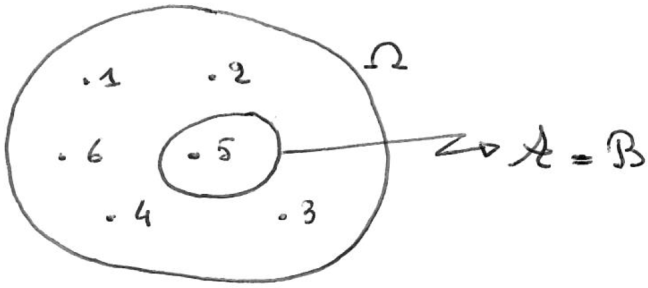
Graphique 1.1 : L'événement A

2.2. Evénements équivalents

Définition 2.1 *Deux événements A, B sont dits équivalents si chaque fois que l'un des deux se réalise, l'autre se réalise aussi : autrement dit, si les résultats possibles de l'épreuve aléatoire pour lesquels ils se réalisent sont les mêmes.*

En termes ensemblistes, deux événements équivalents correspondent au même sous-ensemble de Ω.

Par exemple, dans le cas du lancer d'un dé, les événements A = "obtenir un 5" et B = "obtenir un nombre premier supérieur à 3" sont équivalents : chaque fois que B est réalisé, A l'est aussi, et vice versa ; ils se réalisent pour le même résultat de l'épreuve aléatoire et correspondent donc en termes ensemblistes à la même partie de Ω, qui ici contient un seul élément, à savoir le nombre 5.



Graphique 1.2 : Les événements A et B sont équivalents

Comme autre exemple, supposons qu'on lance cinq fois une pièce, et considérons les deux événements suivants :

$$\begin{array}{lcl} A & = & \text{“obtenir plus de 2 fois pile”} \\ B & = & \text{“obtenir au moins 3 fois pile”} \end{array}$$

Les résultats de l'épreuve aléatoire pour lesquels A et B se réalisent sont les mêmes : ce sont ceux où pile est obtenu $3, 4$ ou 5 fois. Ces deux événements sont donc équivalents.

En fait, comme le montre clairement ces exemples, deux événements équivalents ne sont qu'un seul et même événement décrit par deux énoncés différents. Pour cette raison, on exprime l'équivalence de deux événements A et B en écrivant tout simplement : $A = B$. Une conséquence importante est que deux événements équivalents ont la même probabilité, puisqu'en fait il s'agit du même événement.

Propriété 2.1 *Deux événements A, B équivalents ont la même probabilité.*

2.3. Evénement certain et événement impossible

On lance trois fois une pièce. Considérons l'événement "obtenir au plus 3 fois pile". Cet événement se réalise toujours. On dit alors qu'il est certain. Au contraire, l'événement "obtenir au moins 4 fois pile" ne se réalise jamais. On dit qu'il est impossible.

Définition 2.2 *On appelle événement certain tout événement qui se réalise quel que soit le résultat de l'épreuve aléatoire. On le note Ω. On appelle événement impossible tout événement qui ne se réalise pour aucun résultat de l'épreuve aléatoire. On le note $\emptyset$.*

Les notations $\emptyset$ et Ω découle de la représentation ensembliste des événements : à l'événement certain correspond en effet tous les résultats possibles de l'épreuve aléatoire, soit l'ensemble Ω, et à l'événement impossible, aucun des résultats possibles de l'épreuve aléatoire, soit l'ensemble vide $\emptyset$.

Propriété 2.2 *La probabilité de l'événement certain est* 1 *et celle de l'événement impossible est* 0 *:*

$$\mathbb{P}(\Omega) = 1, \quad \mathbb{P}(\emptyset) = 0$$

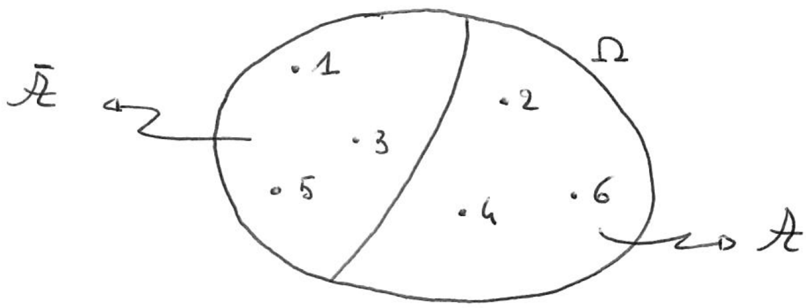
On pourrait être tenté de compléter cette propriété en disant que réciproquement, si la probabilité d'un événement est 1 (est 0), alors cet événement est certain (est impossible). Si nous ne le faisons pas, c'est parce que cette réciproque, qui est vraie dans la situation particulière sur laquelle nous raisonnons (Ω fini), ne l'est pas dans des situations plus générales (quand il y a un continuum de résultats possibles, cf. Chapitre 4).

2.4. Evénement contraire

Définition 2.3 *L'événement contraire (ou complémentaire) d'un événement A est celui qui se réalise si et seulement si A ne se réalise pas ; autrement dit, qui se réalise pour tous les résultats possibles de l'épreuve aléatoire pour lesquels A ne se réalise pas, et seulement ceux-là. On le note $\overline{A}$.*

En termes ensemblistes, l'événement contraire $\overline{A}$ correspond une partie $\overline{\mathcal{A}}$ de Ω composée de l'ensemble des éléments de Ω qui n'appartiennent pas à la partie $\mathcal{A}$ de Ω correspondant à A: $\overline{A}$ correspond donc à la partie complémentaire de $\mathcal{A}$ (par rapport à Ω).

Par exemple, dans le cas du lancer d'un dé, l'événement complémentaire de l'événement $A =$ "obtenir un nombre pair" est $\overline{A} =$ "obtenir un nombre impair" : chaque fois que A n'est pas réalisé, $\overline{A}$ l'est, et vice versa ; A correspond à la partie $\mathcal{A} = \{2, 4, 6\}$ et $\overline{A}$ à la partie $\overline{\mathcal{A}} = \{1, 3, 5\}$, qui est bien la partie complémentaire de $\mathcal{A}$ (par rapport à Ω).



Graphique 1.3 : L'événement A et son événement contraire $\overline{A}$

Comme autre exemple, considérons le lancer d'une pièce. Pour un seul lancer, l'événement contraire de "obtenir pile" est "obtenir face". Pour n lancers, l'événement contraire de "obtenir n fois pile" est "obtenir au moins une fois face".

Relativement à une épreuve aléatoire quelconque, l'événement contraire de l'événement certain Ω est l'événement impossible $\emptyset$, et vice versa. Par ailleurs, il est clair que pour tout événement A, l'événement contraire de $\overline{A}$ est A lui-même. Ainsi, les événements se groupent par paires d'événements contraires l'un de l'autre.

Propriété 2.3 *Pour tout événement A, on a :*

$$I\!\!P(\overline{A}) = 1 - I\!\!P(A)$$

En effet, on a :

$$1 = \sum_i p_i = \sum_{i \in A} p_i + \sum_{i \notin A} p_i = I\!\!P(A) + I\!\!P(\overline{A})$$

où $\sum_{i \notin A}$ signifie "somme pour tous les résultats possibles i pour lesquels A ne se réalise pas".

2.5. Evénement "A ou B", événement "A et B" et loi d'addition

Définition 2.4 *L'événement "A ou B" est celui qui se réalise si et seulement si l'un au moins des deux événements A, B se réalise ; autrement dit, qui se réalise pour tout les résultats possibles de l'épreuve aléatoire pour lesquels l'un au moins des deux événements A, B se réalise, et seulement ceux-là.*

Définition 2.5 *L'événement "A et B" est celui qui se réalise si et seulement si l'événement A et l'événement B se réalisent tous les deux ; autrement dit, qui se réalise pour tout les résultats possibles de l'épreuve aléatoire pour lesquels l'événement A et l'événement B se réalisent tous les deux, et seulement ceux-là.*

Il est très important de remarquer que l'événement "A ou B" peut a priori être réalisé de trois manières différentes :

- lorsque A est réalisé et que B ne l'est pas,
- lorsque B est réalisé et que A ne l'est pas,
- lorsque A et B sont réalisés tous les deux,

tandis que l'événement "A et B" n'est réalisé que d'une seule manière (la troisième).

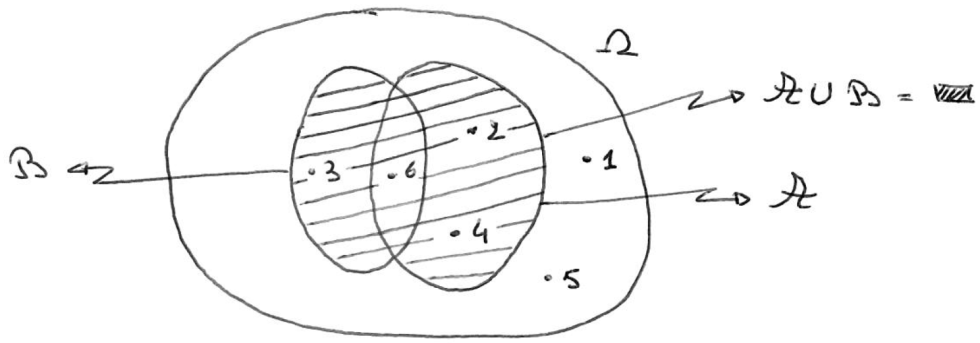
En termes ensemblistes, l'événement "A ou B" correspond à la réunion (ou union) de la partie $\mathcal{A}$ de Ω correspondant à A et de la partie $\mathcal{B}$ de Ω correspondant à B : $\mathcal{A} \cup \mathcal{B}$ Pour cette raison, l'événement "A ou B" est aussi appelé la *réunion* (ou *union*) des deux événements A, B, et est alors noté $A \cup B$. L'événement "A et B" correspond pour sa part à l'intersection des parties $\mathcal{A}$ et $\mathcal{B}$ de Ω : $\mathcal{A} \cap \mathcal{B}$. A nouveau pour cette raison, l'événement "A et B" est aussi appelé l'*intersection* (ou *conjonction*) des deux événements A, B, et est alors noté $A \cap B$.

Supposons par exemple qu'on lance un dé et considérons les deux événements suivants :

$A \quad = \quad$ "obtenir un nombre pair"
$B \quad = \quad$ "obtenir un multiple de 3"

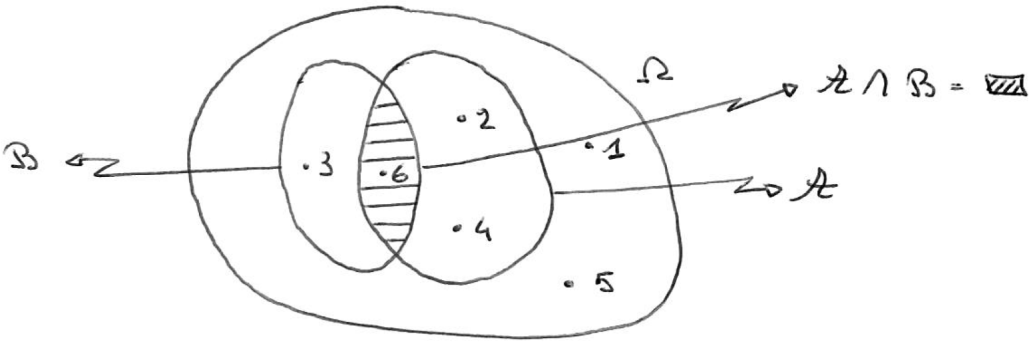
L'événement "A ou B" est "obtenir $2, 3, 4$ ou 6" : chaque fois que l'un au moins des deux événements A, B est réalisé, "A ou B" l'est aussi, et vice versa ; A correspond à la partie $\mathcal{A} = \{2, 4, 6\}$, B à la partie $\mathcal{B} = \{3, 6\}$, et "A ou B" à la partie $\mathcal{A} \cup \mathcal{B} = \{2, 3, 4, 6\}$, qui est bien la réunion de $\mathcal{A}$ et $\mathcal{B}$. Par ailleurs, il peut bien a priori être réalisé de trois manières différentes :

1- en obtenant un nombre pair autre qu'un multiple de 3 (c-à-d. 2 ou 4) : alors A est réalisé et B ne l'est pas.
2- en obtenant un multiple de trois autre qu'un nombre pair (c-à-d. 3) : alors B est réalisé et A ne l'est pas.
3- en obtenant un nombre qui est à la fois pair et multiple de 3 (c-à-d. 6) : alors A et B sont réalisés tous les deux.



Graphique 1.4 : L'événement "A ou B"

L'événement "A et B" est "obtenir un 6" : chaque fois que les événements A, B sont tout les deux réalisés, "A et B" l'est aussi, et vice versa ; "A et B" correspond à la partie $\mathcal{A} \cap \mathcal{B} = \{6\}$, qui est bien l'intersection de $\mathcal{A}$ et $\mathcal{B}$. Par ailleurs,"A et B" n'est bien réalisé que d'une seule manière (la troisième).



Graphique 1.5 : L'événement "A et B"

Les notions d'événements "A ou B" et "A et B" s'étendent de façon naturelle au cas de plus de deux événements. Ainsi, pour trois événements A, B, C, on peut définir l'événement "A ou B ou C" : c'est celui qui est réalisé si et seulement si l'un au moins des trois événements est réalisé. L'événement "A et B et C" est celui qui se réalise si et seulement si A, B et C se réalisent tous les trois. En termes ensemblistes, ces deux événements correspondent respectivement à la réunion et à l'intersection des trois parties $\mathcal{A}, \mathcal{B}$ et $\mathcal{C}$ de Ω correspondant à A, B et C, et pour cette raison sont aussi notés respectivement $A \cup B \cup C$ et $A \cap B \cap C$.

Considérons par exemple l'épreuve aléatoire qui consiste à prélever un couteau dans un lot sortant d'un atelier de coutellerie. On suppose que les couteaux peuvent être affectés par trois types de défectuosité : le manche peut être défectueux, la lame peut être défectueuse et la lame peut être mal fixée au manche. Un couteau peut donc être observé à la sortie de l'atelier dans 8 ($= 2^3$) états différents. Considérons les événements suivants :

$$\begin{array}{lcl} A & = & \text{“le manche est défectueux”} \\ B & = & \text{“la lame est défectueuse”} \\ C & = & \text{“la lame est mal fixée au manche”} \\ D & = & \text{“le couteau est défectueux”} \end{array}$$

et leurs contraires :

$$\begin{array}{lcl} \overline{A} & = & \text{“le manche est bon”} \\ \overline{B} & = & \text{“la lame est bonne”} \\ \overline{C} & = & \text{“la lame est bien fixée au manche”} \\ \overline{D} & = & \text{“le couteau est bon”} \end{array}$$

Clairement, on a "A ou B ou C" $= D$. Tout aussi clairement, on a "$\overline{A}$ et $\overline{B}$ et $\overline{C}$" $= \overline{D}$.

Cet exemple illustre une propriété générale qu'on appelle Loi de De Morgan[2] : les conjonctions "ou" et "et" s'échangent par passage aux événements contraires.

Propriété 2.4 (Loi de De Morgan) *Quels que soient les événements A, B, on a :*

$$\begin{array}{rcl} \overline{A \text{ ou } B} & = & \overline{A} \text{ et } \overline{B} \\ \overline{A \text{ et } B} & = & \overline{A} \text{ ou } \overline{B} \end{array}$$

On a également les propriétés suivantes.

Propriété 2.5 *Quels que soient les événements A, B, C, on a :*

$$\begin{array}{rcl} A \text{ ou } (B \text{ et } C) & = & (A \text{ ou } B) \text{ et } (A \text{ ou } C) \\ A \text{ et } (B \text{ ou } C) & = & (A \text{ et } B) \text{ ou } (A \text{ et } C) \end{array}$$

Ainsi, si on considère les trois événements suivants, relatifs au résultat du lancer

[2] A. DE MORGAN, mathématicien anglais (1806–1878).

d'un dé :

$$\begin{array}{lcl} A & = & \text{“obtenir un nombre pair”} \\ B & = & \text{“obtenir un multiple de 3”} \\ C & = & \text{“obtenir un nombre inférieur ou égal 4”} \end{array}$$

on a bien :

$$\begin{array}{lcll} B \text{ et } C & = & \text{“obtenir 3”} & \\ A \text{ ou } (B \text{ et } C) & = & \text{“obtenir } 2,3,4 \text{ ou } 6\text{”} & \\ A \text{ ou } B & = & \text{“obtenir } 2,3,4 \text{ ou } 6\text{”} & \\ A \text{ ou } C & = & \text{“obtenir } 1,2,3,4 \text{ ou } 6\text{”} & \\ (A \text{ ou } B) \text{ et } (A \text{ ou } C) & = & \text{“obtenir } 2,3,4 \text{ ou } 6\text{”} & = \quad A \text{ ou } (B \text{ et } C) \end{array}$$

et

$$\begin{array}{lcll} B \text{ ou } C & = & \text{“obtenir } 1,2,3,4 \text{ ou } 6\text{”} & \\ A \text{ et } (B \text{ ou } C) & = & \text{“obtenir } 2,4 \text{ ou } 6\text{”} & \\ A \text{ et } B & = & \text{“obtenir 6”} & \\ A \text{ et } C & = & \text{“obtenir 2 ou 4”} & \\ (A \text{ et } B) \text{ ou } (A \text{ et } C) & = & \text{“obtenir } 2,4 \text{ ou } 6\text{”} & = \quad A \text{ et } (B \text{ ou } C) \end{array}$$

Voyons à présent les effets des relations logiques “et”, “ou” sur le calcul des probabilités. Supposons qu'on tire au hasard une carte d'un jeu de 52, et considérons les événements :

$$\begin{array}{lcl} A & = & \text{“obtenir un as”} \\ B & = & \text{“obtenir un trèfle”} \end{array}$$

On a :

$$I\!P(A) = \sum_{i \in A} p_i = \frac{4}{52} \quad (= \frac{N_A}{N})$$

et

$$I\!P(B) = \sum_{i \in B} p_i = \frac{13}{52} \quad (= \frac{N_B}{N})$$

puisque il y a dans le jeu 4 as et 13 trèfles, et que chaque carte a la même probabilité $p_i = \frac{1}{N} = \frac{1}{52}$ d'être tirée (résultats possibles équiprobables).

L'événement “A ou B” est “obtenir un as ou un trèfle”. Il peut être réalisé de trois manières différentes :

1- En obtenant un as qui n'est pas l'as de trèfle. C'est là un événement que nous appellerons A'. On a :

$$I\!P(A') = \sum_{i \in A'} p_i = \frac{3}{52} \quad (= \frac{N_{A'}}{N})$$

2- En obtenant un trèfle qui n'est pas l'as de trèfle. C'est là un événement que nous appellerons B'. On a :

$$I\!\!P(B') = \sum_{i \in B'} p_i = \frac{12}{52} \quad (= \frac{N_{B'}}{N})$$

3- En obtenant l'as de trèfle. C'est l'événement "A et B". On a :

$$I\!\!P(A \text{ et } B) = \sum_{i \in (A \text{ et } B)} p_i = \frac{1}{52} \quad (= \frac{N_{A \text{ et } B}}{N})$$

Au total, on a donc :

$$\begin{aligned}
I\!\!P(A \text{ ou } B) &= \sum_{i \in (A \text{ ou } B)} p_i \quad (= \frac{N_{A \text{ ou } B}}{N}) \\
&= \sum_{i \in A'} p_i + \sum_{i \in B'} p_i + \sum_{i \in (A \text{ et } B)} p_i \quad (= \frac{N_{A'}}{N} + \frac{N_{B'}}{N} + \frac{N_{A \text{ et } B}}{N}) \\
&= I\!\!P(A') + I\!\!P(B') + I\!\!P(A \text{ et } B) \\
&= \frac{3}{52} + \frac{12}{52} + \frac{1}{52} = \frac{16}{52}
\end{aligned}$$

De fait, l'événement "A ou B" = "obtenir un as ou un trèfle" est réalisé si on obtient une des 16 cartes suivantes :

$$1\heartsuit, 1\diamondsuit, 1\spadesuit, 1\clubsuit, 2\clubsuit, 3\clubsuit, 4\clubsuit, 5\clubsuit, 6\clubsuit, 7\clubsuit, 8\clubsuit, 9\clubsuit, 10\clubsuit, V\clubsuit, D\clubsuit, R\clubsuit$$

Si on ne prend pas garde au fait que les événements A et B peuvent être réalisés en même temps — ne sont pas mutuellement exclusifs, cf. infra — et que, pour obtenir $I\!\!P(A \text{ ou } B)$, on additionne tout simplement $I\!\!P(A)$ et $I\!\!P(B)$, on obtient :

$$I\!\!P(A) + I\!\!P(B) = \frac{4}{52} + \frac{13}{52} = \frac{17}{52}$$

En fait, si on additionne $I\!\!P(A)$ et $I\!\!P(B)$ dans le but d'obtenir $I\!\!P(A \text{ ou } B)$, le résultat de l'épreuve aléatoire qui réalise à la fois A et B ($= 1\clubsuit$) est compté deux fois, donc une fois de trop. De façon générale, si on additionne $I\!\!P(A)$ et $I\!\!P(B)$ dans le but d'obtenir $I\!\!P(A \text{ ou } B)$, les résultats de l'épreuve aléatoire qui réalisent à la fois A et B sont comptés une fois de trop, c'est-à-dire qu'on a :

$$I\!\!P(A) + I\!\!P(B) = I\!\!P(A \text{ ou } B) + I\!\!P(A \text{ et } B)$$

d'où :

$$I\!\!P(A \text{ ou } B) = I\!\!P(A) + I\!\!P(B) - I\!\!P(A \text{ et } B)$$

On a ainsi établi la propriété suivante.

Propriété 2.6 *Quels que soient les événements* A, B*, on a :*

$$\mathbb{P}(A \text{ ou } B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \text{ et } B)$$

Cette propriété est connue sous le nom de *loi* (ou *règle*) *d'addition*. Elle peut être étendue au cas où il y a plus de deux événements. Ainsi, un raisonnement semblable à celui développé ci-dessus montre que pour trois événements A, B, C, on a :

$$\begin{aligned}\mathbb{P}(A \text{ ou } B \text{ ou } C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \text{ et } B) - \mathbb{P}(A \text{ et } C) \\ &\quad - \mathbb{P}(B \text{ et } C) + \mathbb{P}(A \text{ et } B \text{ et } C)\end{aligned}$$

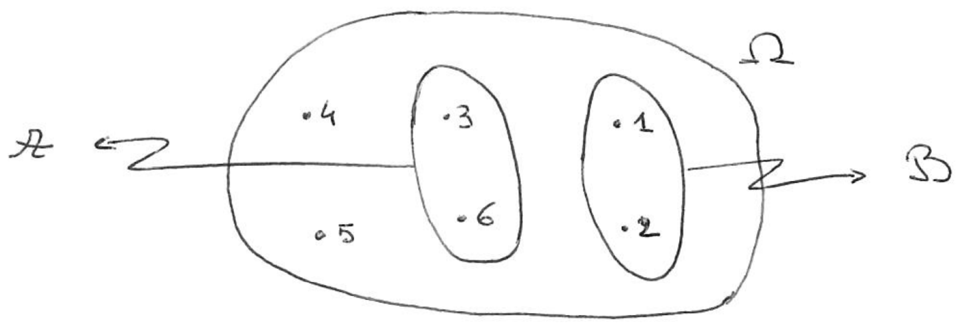
Pour plus de trois événements, cette formule devient vite très complexe.

2.6. Evénements mutuellement exclusifs

Définition 2.6 *Deux événements* A, B *sont dits mutuellement exclusifs (ou incompatibles) s'ils ne peuvent pas être réalisés en même temps : autrement dit, si les résultats possibles de l'épreuve aléatoire pour lesquels ils se réalisent sont différents.*

En termes ensemblistes, deux événements A, B mutuellement exclusifs sont des événements correspondants à des parties $\mathcal{A}$ et $\mathcal{B}$ de Ω disjointes (dont l'intersection est l'ensemble vide $\emptyset$) : $\mathcal{A} \cap \mathcal{B} = \emptyset$.

Par exemple, dans le cas du lancer d'un dé, les événements A = "obtenir un multiple de 3" et B = "obtenir un nombre inférieur ou égal à 2" sont mutuellement exclusifs : A et B ne peuvent pas être réalisés en même temps ; A correspond à la partie $\mathcal{A} = \{3, 6\}$, B à la partie $\mathcal{B} = \{1, 2\}$, et on a $\mathcal{A} \cap \mathcal{B} = \emptyset$.



Graphique 1.6 : Les événements A et B sont mutuellements exclusifs

Deux événements contraires l'un de l'autre sont par définition mutuellement exclusifs, mais l'inverse n'est pas vrai. Ainsi, si l'on tire une boule d'un sac qui contient des boules blanches, noires et rouges, les événements "obtenir une boule blanche" et "obtenir une boule rouge" sont mutuellement exclusifs sans être contraires l'un de l'autre (l'événement contraire de "obtenir une boule blanche" est "obtenir une boule noire ou rouge").

Propriété 2.7 *Si A, B sont deux événements mutuellement exclusifs, alors "A et B" est un événement impossible et $I\!P(A \text{ et } B) = 0$, ce qui implique que $I\!P(A \text{ ou } B) = I\!P(A) + I\!P(B)$.*

Le dernier élément de cette propriété résulte de la Propriété 2.6. Il constitue la *loi d'addition pour deux événements mutuellement exclusifs.*

La notion d'événements mutuellement exclusifs s'étend de façon naturelle au cas de plus de deux événements. On dit que les événements $A_1, A_2, ..., A_r$ sont mutuellement exclusifs (ou deux à deux incompatibles) si pour toute paire A_i, A_j de ces événements, A_i et A_j sont deux événements mutuellement exclusifs. En termes ensemblistes, le fait que $A_1, A_2, ..., A_r$ sont mutuellement exclusifs s'identifie à la situation où les parties $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_r$ de Ω correspondant à $A_1, A_2, ..., A_r$ sont deux à deux disjointes : $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$, $\forall\, i \neq j$. Dans ce cas, les événements "A_i et A_j", "A_i et A_j et A_k", ..., "A_1et A_2 et ... et A_r" sont tous des événements impossibles — donc de probabilité nulle — et la loi d'addition se réduit à :

$$I\!P(A_1 \text{ ou } A_2 \text{ ou } ...\text{ou } A_r) = I\!P(A_1) + I\!P(A_2) + ... + I\!P(A_r)$$

C'est la *loi d'addition pour r événements mutuellement exclusifs.*

2.7. Système complet d'événements mutuellement exclusifs

On rencontre souvent la situation que voici. On a des événements $A_1, A_2, ..., A_r$ qui remplissent les deux conditions suivantes :

1- l'événement "A_1 ou A_2 ou ... ou A_r" est certain.
2- les événements $A_1, A_2, ..., A_r$ sont mutuellement exclusifs (deux à deux incompatibles).

Dans cette situation, quel que soit le résultat de l'épreuve aléatoire, un des événements $A_1, A_2, ..., A_r$ est réalisé (à cause de la condition 1), et un seul (à cause de la condition 2).

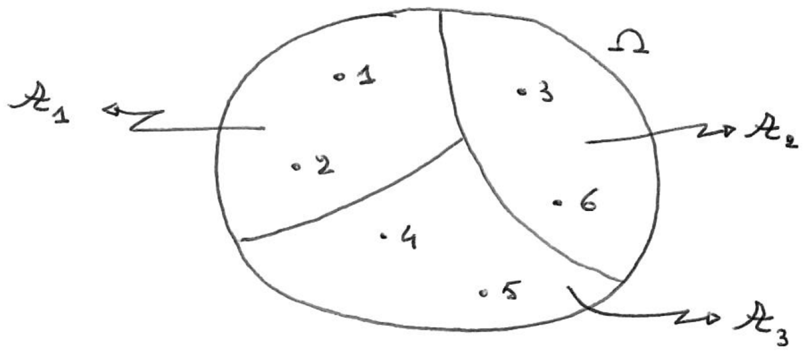
Définition 2.7 *Quand des événements $A_1, A_2, ..., A_r$ satisfont les deux conditions précédentes, on dit qu'ils forment un système complet d'événements mutuellement exclusifs (ou système complet d'événements deux à deux incompatibles) ou encore une partition.*

En termes ensemblistes, le fait que $A_1, A_2, ..., A_r$ forment un système complet d'événements mutuellement exclusifs s'identifie à la situation où les parties $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_r$ de Ω correspondant à $A_1, A_2, ..., A_r$ remplissent les deux conditions suivantes :

1- $\mathcal{A}_1 \cup \mathcal{A}_2 \cup ... \cup \mathcal{A}_r = \Omega$.
2- les parties $\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_r$ sont deux à deux disjointes : $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset, \forall i \neq j$.

Ces deux conditions définissent une *partition*, d'où l'utilisation de ce terme ensembliste.

Par exemple, dans le cas du lancer d'un dé, les événements A_1 = "obtenir un nombre inférieur ou égal à 2", A_2 = "obtenir un multiple de 3" et A_3 = "obtenir 4 ou 5" forme une partition : "A_1 ou A_2 ou A_3" est certain et A_1, A_2, A_3 sont mutuellement exclusifs ; A_1 correspond à la partie $\mathcal{A}_1 = \{1, 2\}$, A_2 à la partie $\mathcal{A}_2 = \{3, 6\}$ et A_3 à la partie $\mathcal{A}_3 = \{4, 5\}$, de sorte qu'on a bien $\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 = \Omega$ et $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset, \forall i \neq j$.



Graphique 1.7 : Les événements A_1, A_2, A_3 forment une partition

Comme autre exemple de partition, considérons un atelier qui comprend quatre machines numérotées $1, 2, 3$ et 4, et qui fabriquent les mêmes pièces. On prélève une pièce dans un lot sortant de l'atelier. On peut définir quatre événements A_1, A_2, A_3, A_4 par :

$$A_i = \text{"la pièce a été fabriquée par la machine numéro } i\text{"}$$

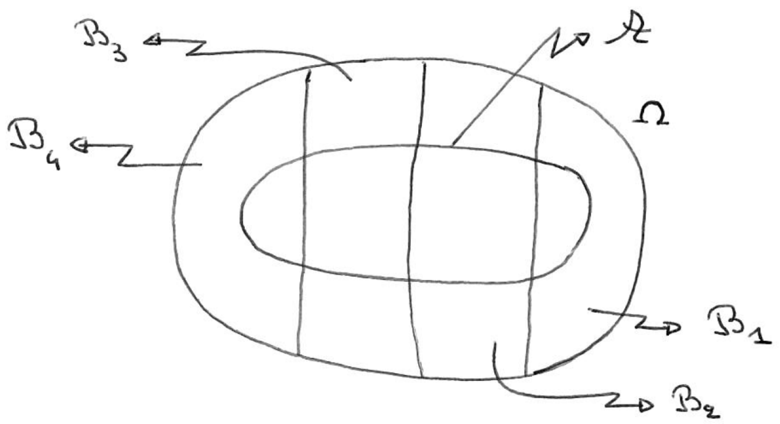
Il est clair que ces quatre événements forment une partition.

Propriété 2.8 *Soit $A_1, A_2, ..., A_r$ une partition. On a :*

$$I\!P(A_1) + I\!P(A_2) + ... + I\!P(A_r) = 1$$

En effet, on a : "A_1 ou A_2 ou ... ou A_r" $= \Omega$, de sorte que $I\!P(A_1$ ou A_2 ou ... ou $A_r) = I\!P(\Omega) = 1$ et, les r événements étant mutuellement exclusifs, $I\!P(A_1$ ou A_2 ou ... ou $A_r) = I\!P(A_1) + I\!P(A_2) + ... + I\!P(A_r)$.

Voici une propriété très importante des partitions. Soit $B_1, B_2, ..., B_r$ une partition, et soit A un événement quelconque. Graphiquement, en termes ensemblistes, on a par exemple :



Graphique 1.8 : Une partition B_1, B_2, B_3, B_4 et un événement A quelconque

Comme $B_1, B_2, ..., B_r$ forme une partition, l'événement A et l'événement "(A et B_1) ou (A et B_2) ou ... ou (A et B_r)" sont équivalents :

$$A = (A \text{ et } B_1) \text{ ou } (A \text{ et } B_2) \text{ ou } (\mathrm{A} \text{ et } B_2) \text{ ou ... ou } (A \text{ et } B_r)$$

Chaque fois que A est réalisé, "(A et B_1) ou (A et B_2) ou ... ou (A et B_r)" l'est aussi, et vice versa. Pour la même raison, chaque fois que A est réalisé, un et un seul des événements $B_1, B_2, ..., B_r$ est réalisé en même temps que lui. Les événements "A et B_1", "A et B_2", ..., "A et B_r" sont donc mutuellement exclusifs (deux à deux incompatibles). On en déduit la propriété suivante.

Propriété 2.9 *Si $B_1, B_2, ..., B_r$ est une partition et A un événement quelconque, on a :*

$$\mathbb{P}(A) = \mathbb{P}(A \text{ et } B_1) + \mathbb{P}(A \text{ et } B_2) + ... + \mathbb{P}(A \text{ et } B_r)$$

En effet, on a : $A =$ "(A et B_1) ou (A et B_2) ou ... ou (A et B_r)", de sorte que $\mathbb{P}(A) = \mathbb{P}[(A \text{ et } B_1) \text{ ou } (A \text{ et } B_2) \text{ ou ... ou } (A \text{ et } B_r)] = \mathbb{P}(A \text{ et } B_1) + \mathbb{P}(A \text{ et } B_2) + ... + \mathbb{P}(A \text{ et } B_r)$ puisque les r événements "A et B_1", "A et B_2", ..., "A et B_r" sont mutuellement exclusifs.

Une partition particulièrement simple est donnée par une paire d'événements contraires l'un de l'autre (c'est le cas où $r = 2$). Ainsi, B et $\overline{B}$ désignant deux événements contraires l'un de l'autre, on a pour tout événement A :

$$A = (A \text{ et } B) \text{ ou } (A \text{ et } \overline{B})$$

et les deux événements "A et B", "A et $\overline{B}$" sont mutuellement exclusifs. Par application de la Propriété 2.9, on a donc :

Propriété 2.10 *Quels que soient les événements A, B, on a :*

$$\mathbb{P}(A) = \mathbb{P}(A \text{ et } B) + \mathbb{P}(A \text{ et } \overline{B})$$

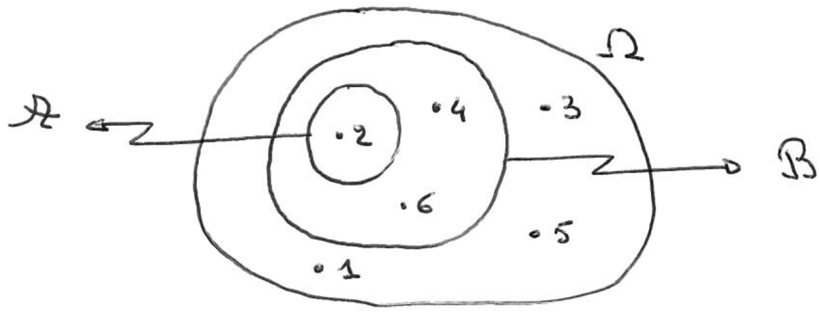
2.8. Implication

Définition 2.8 *On dit qu'un événement A implique un événement B si, chaque fois que A est réalisé, B l'est aussi ; autrement dit si B se réalise pour tous les résultats possibles de l'épreuve aléatoire pour lesquels A se réalise, plus éventuellement d'autres.*

Pour exprimer que A implique B, on écrit $A \Rightarrow B$.

En termes ensemblistes, le fait qu'un événement A implique un événement B s'identifie à la situation où la partie $\mathcal{A}$ de Ω correspondant à A est incluse dans la partie $\mathcal{B}$ de Ω correspondant à B : $\mathcal{A} \subset \mathcal{B}$ Pour cette raison, $A \Rightarrow B$ s'écrit aussi $A \subset B$.

Par exemple, dans le cas du lancer d'un dé, l'événements $A =$ "obtenir un 2" implique l'événement $B =$ "obtenir un nombre pair" : chaque fois que A est réalisé, B l'est aussi (mais pas vice versa) ; A correspond à la partie $\mathcal{A} = \{2\}$, B à la partie $\mathcal{B} = \{2, 4, 6\}$, et on a bien $\mathcal{A} \subset \mathcal{B}$.



Graphique 1.9 : L'événement A implique l'événement B

Comme autre exemple, supposons qu'on tire trois cartes d'un jeu de 52 cartes, et qu'on considère les deux événements :

$$\begin{aligned} A &= \text{"obtenir deux piques et un coeur"} \\ B &= \text{"obtenir au moins un pique"} \end{aligned}$$

Il est clair que A implique B : si A est réalisé, alors B l'est aussi automatiquement. En revanche l'événement :

$$C = \text{"obtenir au moins un pique ou un coeur"}$$

n'implique pas B, car il peut être réalisé en obtenant au moins un coeur, mais aucun pique. Par contre B implique C.

Soit deux événements A, B tels que A implique B. De la Propriété 2.10, on a :

$$I\!P(B) = I\!P(B \text{ et } A) + I\!P(B \text{ et } \overline{A})$$

Comme $A \Rightarrow B$, on a aussi "B et A"$= A$. En effet, d'une part, chaque fois que "B et A" est réalisé, A l'est aussi. D'autre part, chaque fois que A est réalisé, puisque A implique B, "B et A" l'est aussi. Ainsi, l'événement "B et A" et l'événement A sont équivalents. On en déduit que

$$I\!P(B) = I\!P(A) + I\!P(B \text{ et } \overline{A})$$

d'où $I\!P(B) \geq I\!P(A)$, puisque $I\!P(B \text{ et } \overline{A}) \geq 0$.

On a ainsi établi la propriété suivante.

Propriété 2.11 *Si un événement A implique un événement B, on a :*

$$I\!P(A) \leq I\!P(B)$$

2.9. Remarque

On a d'un côté des événements, de l'autre des probabilités, qui sont des nombres. Ce sont des objets de nature différente.

Il y a des *opérations logiques* sur les *événements*, comme les opérations "ou" et "et", et il y a d'autre part des *opérations arithmétiques* sur les *nombres*, comme l'addition et la multiplication. On ne peut pas appliquer aux événements des opérations qui portent sur les nombres, ni aux probabilités, qui sont des nombres, des opérations qui portent sur les événements.

Ainsi pour deux événements A, B il est parfaitement correct d'écrire des expressions comme :

$$I\!P(A \text{ ou } B) \qquad\qquad I\!P(A \text{ et } B)$$
$$I\!P(A \cup B) \qquad\qquad I\!P(A \cap B)$$

mais il est incorrect d'écrire des expressions comme :

$$I\!P(A) \text{ ou } I\!P(B) \qquad\qquad I\!P(A) \text{ et } I\!P(B)$$
$$I\!P(A) \cup I\!P(B) \qquad\qquad I\!P(A) \cap I\!P(B)$$

alors qu'il est correct d'écrire des expressions comme :

$$I\!P(A) + I\!P(B) \qquad\qquad I\!P(A) \times I\!P(B)$$

2.10. Exercice résolu

On lance un dé truqué. On sait que ce dé est tel que les probabilités des événements :

$$\begin{array}{rcl} A & = & \text{“obtenir un nombre pair”} \\ B & = & \text{“obtenir un nombre premier”} \\ C & = & \text{“obtenir 4”} \end{array}$$

sont $\mathbb{P}(A) = 0,6$, $\mathbb{P}(B) = 0,5$ et $\mathbb{P}(C) = 0,1$.

1- Quel est la probabilité de l'événement $D =$ “obtenir un nombre impair” ?

   On a $D = \overline{A}$. Dès lors :

$$\mathbb{P}(D) = \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A) = 1 - 0,6 = 0,4$$

2- Quel est la probabilité de l'événement $E =$ “obtenir un nombre pair ou un nombre premier” ?

   On a $E =$ “A ou B”. Or, “A ou B” est un événement certain. Dès lors :

$$\mathbb{P}(E) = \mathbb{P}(A \text{ ou } B) = \mathbb{P}(\Omega) = 1$$

3- Quel est la probabilité de l'événement $F =$ “obtenir 2” ?

   On a $F =$ “A et B”. On sait que $\mathbb{P}(A \text{ ou } B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \text{ et } B)$ et que ici $\mathbb{P}(A \text{ ou } B) = 1$. Dès lors :

$$\begin{array}{rcl} \mathbb{P}(F) & = & \mathbb{P}(A \text{ et } B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \text{ ou } B) \\ & = & 0,6 + 0,5 - 1 = 0,1 \end{array}$$

4- Quel est la probabilité de l'événement $G =$ “obtenir 6” ?

   Les événements B, C et G forment une partition. On a donc $\mathbb{P}(B) + \mathbb{P}(C) + \mathbb{P}(G) = 1$. Dès lors,

$$\mathbb{P}(G) = 1 - \mathbb{P}(B) - \mathbb{P}(C) = 1 - 0,5 - 0,1 = 0,4$$

5- Quel est la probabilité de l'événement $H =$ “obtenir 2 ou 4” ?

   On a $H =$ “F ou C”. Les événements F, C sont mutuellement exclusifs. Dès lors :

$$\mathbb{P}(H) = \mathbb{P}(F) + \mathbb{P}(C) = 0,1 + 0,1 = 0,2$$

6- Soit l'événement $I =$ “obtenir 2 ou 5”. Est-il possible qu'on ait $\mathbb{P}(I) = 0,6$?

   On a $I \Rightarrow B$. Dès lors :

$$\mathbb{P}(I) \leq \mathbb{P}(B) = 0,5$$

   C'est donc impossible.

2.11. Exercices

1- Considérons le tirage au sort d'une carte d'un jeu de 52 cartes et définissons les événements suivants :

A	$=$	"obtenir un as"	D	$=$	"obtenir un coeur ou un pique"
B	$=$	"obtenir un coeur"	E	$=$	"obtenir un coeur et un pique"
C	$=$	"obtenir l'as de coeur"	F	$=$	"obtenir un pique"

a- Quel est le lien entre les événements A, B et C ?
b- Quel est le lien entre les événements B, D et F ?
c- Quel est le lien entre les événements B, E et F ?
d- Que peut-on dire de l'événement E ?
e- Que représentent les événements suivants ?
   i- $\overline{A}$, $\overline{F}$, $\overline{D}$.
   ii- "$\overline{A}$ et B", "$\overline{C}$ ou B".
f- Les événements A et B sont-ils mutuellement exclusifs ?
g- Donnez l'événement contraire de chacun des événements considérés.
h- Donnez deux exemples de systèmes complets d'événements mutuellement exclusifs relatifs à cette épreuve aléatoire.
i- Donnez les relations d'implication existant entre les événements considérés.

2- Une urne contient 10 boules numérotées de 1 à 10. Les boules portant les numéros $1, 4, 5$ sont rouges, les autres sont blanches. On tire au hasard une boule de l'urne et on considère les événements suivants :

B	$=$	"la boule est blanche"
R	$=$	"la boule est rouge"
$N1$	$=$	"la boule porte le numéro 1"
P	$=$	"la boule porte un numéro pair"
I	$=$	"la boule porte un numéro impair"

a- Déterminez la probabilité de chacun de ces événements.
b- Quel est le lien entre les événements P et I, $N1$ et I ?
c- Quel est l'événement contraire de $N1$?
d- Les événements B et R sont-ils mutuellement exclusifs ?
e- Que représentent les événements suivants : $\overline{P}$, $\overline{I}$, "P ou $\overline{P}$", "P et R", "I et $N1$", "P ou R", "I ou $N1$", "$\overline{B \text{ et } P}$"; "$\overline{B \text{ ou } P}$".
f- Donnez quelques exemples de partition.

3- On tire au sort une boule parmi 37 numérotées de 1 à 37.
a- Calculez la probabilité que la boule tirée porte un numéro multiple de 3 (événement T).
b- Calculez la probabilité que la boule tirée porte un numéro multiple de 4 (événement Q).
c- Calculez la probabilité que la boule tirée porte un numéro multiple de 6 (événement S).
d- Que représente chacun des événements suivants : "T et Q", "T ou Q", "Q et S", "Q ou S", "T et Q et S", "T ou Q ou S" ? Calculez leurs probabilités.

4- On tire au hasard trois cartes d'un jeu de 52 cartes. On considère les événements suivants :

A = "obtenir deux coeurs et un carreau"
B = "obtenir au moins deux coeurs"
C = "obtenir trois as"
D = "ne pas obtenir d'as"

Répondez aux questions suivantes sans calculer de probabilité.

a- Quel lien y a t-il entre les événements A et B ?
b- Quel est le contraire de l'événement C ?
c- Quel est le contraire de l'événement B ?
d- Les événements A et C sont-ils mutuellement exclusifs ? Forment-ils une partition ?
e- Que représente l'événement "B et D" ?

5- On considère la population adulte d'une ville. Celle-ci est composée de femmes et d'hommes, de Français et d'immigrés. On choisit une personne au hasard et on considère les événements suivants :

A = "la personne choisie est une femme"
B = "la personne choisie est un homme"
C = "la personne choisie est de nationalité française"
D = "la personne choisie est immigrée"
E = "la personne choisie est une femme française"
F = "la personne choisie est une femme immigrée"
G = "la personne choisie est un homme français"
H = "la personne choisie est un homme immigré"

Indiquez votre réponse par une croix dans le tableau suivant :

	Vrai	Faux
Les événements A et B sont mutuellement exclusifs		
A est le contraire de B		
E est le contraire de G		
Les événements B, E, F forment une partition		
L' événement "C et D" est certain		
L'événement "C ou D" est certain		
L'événement "C ou D" n'est pas certain		
A implique E		
Le contraire de F est "E ou B"		

6- La probabilité qu'un individu pris au hasard apprécie à la fois le théâtre et le cinéma est de $0,15$. D'autre part, la probabilité qu'il apprécie le théâtre est de $0,3$ et celle qu'il apprécie le cinéma ou le théâtre est de $0,6$. Déterminez la probabilité qu'il n'aime pas le cinéma.

7- Dans une entreprise, 40% des vendeurs réalisent au moins 3 ventes par jour, 10% ont suivi un stage de négociation, 7% ont suivi le stage et réalisent au moins 3 ventes par jour. Quelle est la probabilité qu'un vendeur pris au hasard :

a- ait suivi le stage ou réalise au moins trois ventes par jour ?

b- réalise au moins 3 ventes par jour et n'ait pas suivi le stage ?
c- réalise moins de 3 ventes par jour et ait suivi le stage ?
d- réalise au plus 2 ventes par jour et n'ait pas suivi le stage ?

8- Un service météorologique a établi qu'au cours d'une journée d'avril il pleut avec une probabilité $0,35$ et il neige avec une probabilité de $0,08$. D'autre part, il y a une température supérieure ou égale à 20 degrés avec une probabilité de $0,52$. Il y a de la pluie et une température inférieure à 20 degrés avec une probabilité $0,2$. La pluie et la neige sont des événements incompatibles, de même que la neige et une température supérieure à 20 degrés. Quelle est la probabilité que lors d'une journée d'avril :

a- il pleuve ou il neige ?
b- il n'y ait ni pluie ni neige ?
c- il y ait de la neige et une température inférieure à 20 degrés ?
d- il pleuve ou la température soit inférieure à 20 degrés ?

9- Les probabilités qu'un ménage pris au hasard possède certains objets sont données dans le tableau suivant :

Objet			Probabilité
A	=	une télévision couleur	$0,45$
B	=	une automobile	$0,56$
C	=	deux automobiles ou plus	$0,24$
D	=	une télévision couleur et une automobile	$0,18$
E	=	une télévision couleur et deux autos ou plus	$0,21$

Quelle est la probabilité qu'un ménage pris au hasard :

a- possède au moins une automobile ?
b- ne possède pas d'automobile ?
c- possède une télévision couleur et au moins une automobile ?
d- possède une télévision couleur ou au moins une automobile ?

10- Le tableau suivant est calculé à partir des données de l'article :"Du domicile à l'établissement scolaire : les trajets quotidiens des jeunes en 1991-1992", Brigitte Baccaïni, *Economie et statistique*, 293, 1996. Il donne la répartition d'un échantillon de 1 000 élèves selon la durée du trajet entre le domicile et l'établissement et selon le type d'établissement.

Type d'établissement	Moins de 5 minutes	5 à 14 minutes	15 à 29 minutes	30 à 44 minutes	45 à 59 minutes	60 minutes et plus	Total
Public	76	438	182	67	27	29	819
Privé	8	80	51	23	11	8	181
Total	84	518	233	90	38	37	1000

a- On tire un élève au hasard parmi les 1 000 élèves considérés.
   i- Quelle est la probabilité qu'il fréquente l'école publique ? Quelle est la probabilité qu'il fréquente l'école privée ?
   ii- Quelle est la probabilité que son trajet dure moins de 5 minutes et qu'il fréquente l'école privée ?
   iii- Quelle est la probabilité qu'il fréquente l'école publique ou que son

trajet scolaire dure moins de 5 minutes ?

iv- Pour quelle proportion d'élèves le trajet dure-t-il moins de 45 minutes ?

b- On tire au hasard un élève parmi ceux qui fréquentent l'école publique. Quelle est la probabilité que son trajet scolaire soit compris entre 5 et 44 minutes ?

c- Parmi ceux dont le trajet scolaire est supérieur à 60 minutes, quel est le pourcentage d'élèves qui fréquentent l'école privée ? l'école publique ?

Chapitre 3

Le calcul des probabilités – II Probabilité conditionnelle, dépendance et indépendance

3.1. Probabilité conditionnelle

3.1.1. Définition

Supposons qu'on lance un dé équilibré sans vous montrer le résultat. On vous demande de deviner, au hasard, le résultat obtenu. Vous dites, par exemple, 5. Quelle est la probabilité que vous ayez deviné juste ?

Puisque vous ne savez rien du résultat, votre situation est la même que si le dé n'avait pas encore été lancé : les six résultats $1, 2, 3, 4, 5$ et 6 sont tous possibles et également probables. La probabilité d'avoir deviné juste est de $1/6$.

Supposons maintenant qu'on vous donne une information sur le résultat. Par exemple, on vous dit qu'il est impair. Dès lors, les résultats $2, 4, 6$ deviennent impossibles, seuls restent les résultats $1, 3, 5$, tous également probables pour vous. Donc en disant au hasard que le résultat est 5, vous évaluerez votre probabilité d'avoir deviné juste non plus à $1/6$, mais à $1/3$. Vous avez réévalué la probabilité du résultat 5 en tenant compte d'une information que vous possédiez sur l'issue de l'épreuve aléatoire.

Que représente cette probabilité réévaluée de $1/3$? C'est tout simplement, dans l'épreuve aléatoire du lancer d'un dé équilibré, la probabilité que l'événement $A =$ "obtenir un 5" soit réalisé lorsque l'événement $B =$ "obtenir un nombre impair" est réalisé.

La probabilité qu'un événement quelconque A soit réalisé lorsqu'on sait qu'un certain événement B est réalisé est appelée la *probabilité conditionnelle* de A *sachant que B est réalisé.* On la note $\mathbb{P}(A|B)$, ce qui se lit "probabilité de A sachant B"

ou "probabilité de A si B". Dans notre exemple introductif, $I\!P(A|B) = 1/3$.

Par opposition à la notion de probabilité conditionnelle, la probabilité d'un événement évaluée en l'absence de toute information est parfois appelée *probabilité a priori* ou encore *probabilité non conditionnelle.* C'est tout simplement $I\!P(A)$. Dans notre exemple introductif, $I\!P(A) = 1/6$.

Il faut être attentif aux rôles fondamentalement dissymétriques que jouent les événements A et B. La probabilité conditionnelle $I\!P(A|B)$ est une probabilité relative à l'événement A. L'événement B apporte simplement une information quant à l'issue de l'épreuve aléatoire.

Définition 3.1 *Soient deux événements A, B relatifs à une épreuve aléatoire quelconque. On appelle probabilité conditionnelle de l'événement A sachant que l'événement B est réalisé :*

$$I\!P(A|B) = \frac{I\!P(A \ et \ B)}{I\!P(B)} \qquad (I\!P(B) \neq 0)$$

Notons que lorsque l'épreuve aléatoire possède N résultats possibles, tous équiprobables, c'est-à-dire lorsque $p_i = \frac{1}{N}(i = 1, ..., N)$, la formule de $I\!P(A|B)$ se réduit à :

$$I\!P(A|B) = \frac{I\!P(A \text{ et } B)}{I\!P(B)} = \frac{\sum\limits_{i \in (A \text{ et } B)} p_i}{\sum\limits_{i \in B} p_i} = \frac{\frac{N_{A \text{ et } B}}{N}}{\frac{N_B}{N}} = \frac{N_{A \text{ et } B}}{N_B} \qquad (N_B \neq 0)$$

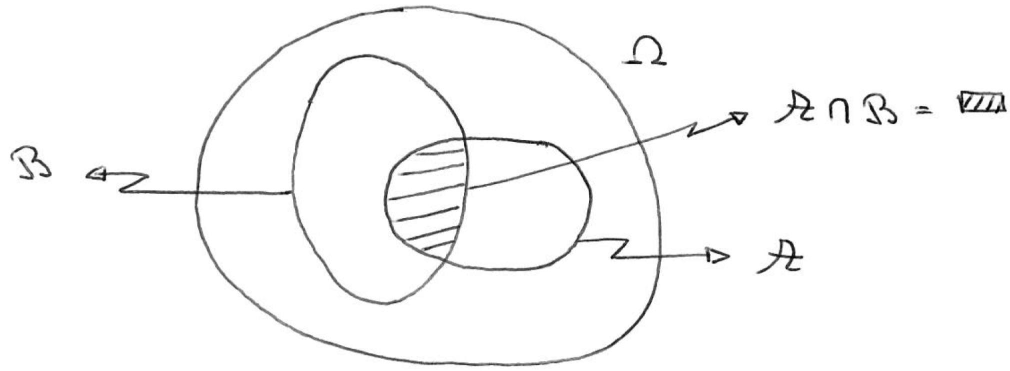
où N_X désigne le nombre des résultats possibles i pour lesquels l'événement X se réalise.

Ainsi dans notre exemple introductif, l'événement "A et B" étant "obtenir un 5", on a bien :

$$I\!P(A|B) = \frac{I\!P(A \text{ et } B)}{I\!P(B)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3} = \frac{N_{A \text{ et } B}}{N_B}$$

La formule donnée par la Définition 3.1 se comprend plus aisément si on note que la probabilité $I\!P(A|B)$ qu'un événement A soit réalisé sachant qu'un événement B est réalisé est tout simplement la probabilité que le résultat possible qui réalise B soit un résultat qui réalise aussi A ; autrement dit, que parmi les résultats qui réalisent l'événement B, ce soit un résultat qui réalise à la fois A et B, c'est-à-dire l'événement "A et B".

En termes ensemblistes, $I\!P(A|B)$ est donc la probabilité que, lorsque l'épreuve aléatoire est réalisée pour un des résultats possibles appartenant à la partie $\mathcal{B}$ de Ω correspondant à B, il appartienne également à la partie $\mathcal{A} \cap \mathcal{B}$ de Ω correspondant à l'intersection des parties $\mathcal{A}$ et $\mathcal{B}$ de Ω, la partie $\mathcal{A}$ de Ω étant bien entendu la partie correspondant à l'événement A. Graphiquement :



Graphique 3.1 : $I\!\!P(A|B)$ est la probabilité que le résultat possible qui réalise B soit un résultat qui réalise aussi "A et B"

L'exemple suivant devrait finir de clarifier les choses. Supposons qu'on lance un dé pipé dont la distribution de probabilité des différents résultats possibles est :

Résultats possibles i	1	2	3	4	5	6
Probabilités p_i	$0,2$	$0,1$	$0,1$	$0,1$	$0,1$	$0,4$

et considérons les deux événements :

A = "obtenir un nombre premier"
B = "obtenir un nombre supérieur à 2"

L'événement A est réalisé pour les résultats possibles $1, 2, 3$ et 5, et l'événement B pour les résultats possibles $3, 4, 5$ et 6. On a donc :

$$I\!\!P(A) = \sum_{i \in A} p_i = 0,2 + 0,1 + 0,1 + 0,1 = 0,5$$

et

$$I\!\!P(B) = \sum_{i \in B} p_i = 0,1 + 0,1 + 0,1 + 0,4 = 0,7$$

Quand on sait que l'événement B est réalisé, la probabilité conditionnelle de chacun des résultats possibles — la probabilité que lorsque B est réalisé, il le soit par l'un ou l'autre de ces résultats — est donnée par :

$$P(E_i|B) = \frac{I\!\!P(E_i \text{ et } B)}{I\!\!P(B)} = \frac{\sum\limits_{i \in (E_i \text{ et } B)} p_i}{\sum\limits_{i \in B} p_i}$$

où les E_i sont les événements élémentaires "obtenir i". Dans notre exemple, ces probabilités sont égales à :

E_i	1	2	3	4	5	6
$P(E_i\|B)$	$\frac{0}{0,7} = 0$	$\frac{0}{0,7} = 0$	$\frac{0,1}{0,7} = \frac{1}{7}$	$\frac{0,1}{0,7} = \frac{1}{7}$	$\frac{0,1}{0,7} = \frac{1}{7}$	$\frac{0,4}{0,7} = \frac{4}{7}$

Notez que $P(E_1|B) = P(E_2|B) = 0$. L'événement B est uniquement réalisé pour les résultats $3, 4, 5$ et 6. Lorsque B est réalisé, la probabilité qu'il le soit par l'un des deux résultats 1 et 2 est donc nulle. En termes plus techniques, les événements "E_1 et B" et "E_2 et B" sont des événements impossibles, et donc $P(E_1 \text{ et } B) = P(E_2 \text{ et } B) = 0$, ce qui implique $P(E_1|B) = P(E_2|B) = 0$.

Notez également que les probabilités conditionnelles $P(E_3|B)$, $P(E_4|B)$, $P(E_5|B)$ et $P(E_6|B)$ sont strictement proportionnelles à leur probabilité a priori $p_i = P(E_i)$ — on a en effet $P(E_i|B) = p_i/\mathbb{P}(B)$, $\forall\, i \in B$, soit $i = 3, 4, 5$ et 6 — et que leur somme est égale à 1.

On le voit, lorsqu'on sait que B est réalisé, tout se passe en fait comme si on avait affaire à une "nouvelle" épreuve aléatoire dont les résultats possibles sont uniquement les événements élémentaires E_i qui réalisent B, et les probabilités de ces résultats possibles sont données par $P(E_i|B) = p_i/\mathbb{P}(B)$. Dans notre exemple, les résultats possibles E_i de cette "nouvelle" épreuve aléatoire sont $3, 4, 5$ et 6, et les probabilités de ces résultats possibles sont respectivement $1/7, 1/7, 1/7$ et $4/7$. Remarquons au passage que si on avait supposé que le dé était équilibré, ces probabilités auraient toutes les quatre été égales à $0,25$.

Dans cette perspective, la probabilité conditionnelle $P(A|B)$ est tout simplement la somme des probabilités $P(E_i|B) = p_i/\mathbb{P}(B)$ des résultats possibles E_i de cette "nouvelle" épreuve aléatoire — uniquement ceux qui réalisent B donc — qui réalisent A, c'est-à-dire de tous les résultats possibles de l'épreuve aléatoire "initiale‘" qui réalise A, parmi ceux qui réalisent B. Ces résultats possibles ne sont autres que les résultats possibles qui réalisent l'événement "A et B". On a en effet bien :

$$\begin{aligned}
\mathbb{P}(A|B) &= \sum_{i \in (A \text{ et } B)} P(E_i|B) = \sum_{i \in (A \text{ et } B)} \frac{p_i}{\mathbb{P}(B)} \\
&= \frac{1}{\mathbb{P}(B)} \sum_{i \in (A \text{ et } B)} p_i = \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)}
\end{aligned}$$

Ainsi, dans notre exemple, où l'événement "A et B" est réalisé pour les résultats 3 et 5, on a :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)} = \frac{\sum\limits_{i \in (A \text{ et } B)} p_i}{\sum\limits_{i \in B} p_i} = \frac{0,1 + 0,1}{0,1 + 0,1 + 0,1 + 0,4} = \frac{0,2}{0,7} = \frac{2}{7}$$

ce qui est bien égal à la somme des probabilités $P(E_i|B) = p_i/\mathbb{P}(B)$ des deux résultats possibles 3 et 5 qui réalisent A dans la "nouvelle" épreuve aléatoire.

3.1.2. Propriétés

On notera que la définition de la probabilité conditionnelle que nous avons donnée est d'application générale, valable que les résultats possibles de l'épreuve aléatoire sous-jacente soient ou non en nombre fini.

Une restriction toutefois. Pour que le quotient $\mathbb{P}(A \text{ et } B)/\mathbb{P}(B)$ ait un sens, le dénominateur $\mathbb{P}(B)$ doit être différent de 0, d'où la condition $\mathbb{P}(B) \neq 0$ contenue dans la Définition 3.1.

De la définition de la probabilité conditionnelle $P(A|B)$, on peut déduire une série de propriétés, semblables à celles que satisfont le calcul des probabilités a priori. Etant donnée l'interprétation possible de la probabilité conditionnelle d'un événement comme la "probabilité standard" de cet événement dans une "nouvelle" épreuve aléatoire, cela ne devrait guère vous étonner.

Dans les propriétés ci-dessous, il est toujours implicitement supposé que $\mathbb{P}(B) \neq 0$.

Propriété 3.1 *La probabilité conditionnelle de tout événement A sachant qu'un événement B est réalisé est un nombre compris entre* 0 *et* 1 *:*

$$0 \leq \mathbb{P}(A|B) \leq 1$$

En effet, d'une part, on a $\mathbb{P}(A \text{ et } B) \geq 0$ et $\mathbb{P}(B) > 0$, ce qui assure que $\mathbb{P}(A|B) \geq 0$. D'autre part, comme "A et B" $\Rightarrow B$, de la Propriété 2.11, on a $\mathbb{P}(A \text{ et } B) \leq \mathbb{P}(B)$, ce qui assure que $\mathbb{P}(A|B) \leq 1$.

Propriété 3.2 *Si deux événements A_1, A_2 sont équivalents, sachant un même événement B, ils ont la même probabilité conditionnelle.*

En effet, si deux événements A_1, A_2 sont équivalents, on a $A_1 = A_2$, et donc "A_1 et B" = "A_2 et B", ce qui implique $\mathbb{P}(A_1 \text{ et } B) = \mathbb{P}(A_2 \text{ et } B)$, de sorte que $\mathbb{P}(A_1|B) = \mathbb{P}(A_1 \text{ et } B)/\mathbb{P}(B) = \mathbb{P}(A_2 \text{ et } B)/\mathbb{P}(B) = \mathbb{P}(A_2|B)$.

Propriété 3.3 *Pour tout événement A, B tel que $B \Rightarrow A$, on a :*

$$\mathbb{P}(A|B) = 1$$

ce qui implique en particulier que $\mathbb{P}(B|B) = \mathbb{P}(\Omega|B) = 1$.

En effet, lorsque $B \Rightarrow A$, on "A et B" = B, de sorte que $\mathbb{P}(A|B) = \mathbb{P}(A \text{ et } B)/\mathbb{P}(B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$.

Ainsi, conditionnellement à B, tout événement impliqué par B est un événement certain.

Propriété 3.4 *Si deux événements A, B sont mutuellement exclusifs, on a :*

$$\mathbb{P}(A|B) = 0$$

ce qui implique en particulier que $\mathbb{P}(\emptyset|B) = 0$.

En effet, si A, B sont mutuellement exclusifs, on a $\mathbb{P}(A \text{ et } B) = 0$ et donc $\mathbb{P}(A|B) = \mathbb{P}(A \text{ et } B)/\mathbb{P}(B) = 0/\mathbb{P}(B) = 0$.

Donc, conditionnellement à B, tout événement disjoint de (incompatible avec) B est un événement impossible.

Propriété 3.5 *Pour tout événement A, B tel que $A \Rightarrow B$, on a :*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A)}{\mathbb{P}(B)}$$

ce qui implique en particulier que $\mathbb{P}(A|\Omega) = \mathbb{P}(A)$.

En effet, lorsque $A \Rightarrow B$, on a "A et B" $= A$, de sorte que $\mathbb{P}(A|B) = \mathbb{P}(A \text{ et } B)/\mathbb{P}(B) = \mathbb{P}(A)/\mathbb{P}(B)$.

Propriété 3.6 *Pour tout événement A, B, on a :*

$$\mathbb{P}(\overline{A}|B) = 1 - \mathbb{P}(A|B)$$

En effet, de la Propriété 2.10, on a $\mathbb{P}(B) = \mathbb{P}(B \text{ et } A) + \mathbb{P}(B \text{ et } \overline{A})$, ou de façon équivalente, $\mathbb{P}(\overline{A} \text{ et } B) = \mathbb{P}(B) - \mathbb{P}(A \text{ et } B)$. On en déduit que :

$$\begin{aligned} \mathbb{P}(\overline{A}|B) &= \frac{\mathbb{P}(\overline{A} \text{ et } B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B) - \mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)} \\ &= 1 - \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)} = 1 - \mathbb{P}(A|B) \end{aligned}$$

Propriété 3.7 *Pour tout événement A_1, A_2, B, on a :*

$$\mathbb{P}[(A_1 \text{ ou } A_2)|B] = \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) - \mathbb{P}[(A_1 \text{ et } A_2)|B]$$

En effet, des Propriétés 2.5 et 2.6, on a :

$$\begin{aligned} \mathbb{P}[(A_1 \text{ ou } A_2) \text{ et } B] &= \mathbb{P}[(A_1 \text{ et } B) \text{ ou } (A_2 \text{ et } B)] \\ &= \mathbb{P}(A_1 \text{ et } B) + \mathbb{P}(A_2 \text{ et } B) \\ &\quad -\mathbb{P}[(A_1 \text{ et } B) \text{ et } (A_2 \text{ et } B)] \end{aligned}$$

Comme $\mathbb{P}[(A_1 \text{ et } B) \text{ et } (A_2 \text{ et } B)] = \mathbb{P}[(A_1 \text{ et } A_2) \text{ et } B)]$, on en déduit que :

$$\begin{aligned} \mathbb{P}[(A_1 \text{ ou } A_2)|B] &= \frac{\mathbb{P}[(A_1 \text{ ou } A_2) \text{ et } B]}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(A_1 \text{ et } B) + \mathbb{P}(A_2 \text{ et } B) - \mathbb{P}[(A_1 \text{ et } A_2) \text{ et } B]}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(A_1 \text{ et } B)}{\mathbb{P}(B)} + \frac{\mathbb{P}(A_2 \text{ et } B)}{\mathbb{P}(B)} - \frac{\mathbb{P}[(A_1 \text{ et } A_2) \text{ et } B]}{\mathbb{P}(B)} \\ &= \mathbb{P}(A_1|B) + \mathbb{P}(A_2|B) - \mathbb{P}[(A_1 \text{ et } A_2)|B] \end{aligned}$$

On le voit, les probabilités conditionnelles satisfont une même loi d'addition que les probabilités a priori. Comme dans le cas des probabilités a priori, elle peut être étendue au cas où il y a plus de deux événements. Ainsi, un raisonnement semblable à celui développé ci-dessus montre que pour trois événements A_1, A_2, A_3, on a :

$$\begin{aligned} I\!P[(A_1 \text{ ou } A_2 \text{ ou } A_3)|B] &= I\!P(A_1|B) + I\!P(A_2|B) + I\!P(A_3|B) - I\!P[(A_1 \text{ et } A_2)|B] \\ &\quad - I\!P[(A_1 \text{ et } A_3)|B] - I\!P[(A_2 \text{ et } A_3)|B] \\ &\quad + I\!P[(A_1 \text{ et } A_2 \text{ et } A_3)|B] \end{aligned}$$

Pour plus de trois événements, cette formule devient vite très complexe.

Propriété 3.8 *Pour tout événement B et deux événements A_1, A_2 mutuellement exclusifs, on a :*

$$I\!P[(A_1 \textit{ et } A_2)|B] = 0$$

ce qui implique que :

$$I\!P[(A_1 \textit{ ou } A_2)|B] = I\!P(A_1|B) + I\!P(A_2|B)$$

En effet, si les événements A_1, A_2 sont mutuellement exclusifs, $I\!P[(A_1 \text{ et } A_2) \text{ et } B)] = 0$, et donc $I\!P[(A_1 \text{ et } A_2)|B] = I\!P[(A_1 \text{ et } A_2) \text{ et } B)] / I\!P(B) = 0/I\!P(B) = 0$. La seconde partie de la propriété découle de la Propriété 3.7.

Les probabilités conditionnelles satisfont donc une même loi d'addition pour des événements mutuellement exclusifs que les probabilités à priori. Comme pour ces dernières, la Propriété 3.8 peut être étendue aux cas de plus de deux événements. Ainsi, pour r événements $A_1, A_2, ..., A_r$ mutuellement exclusifs, on a :

$$I\!P[(A_1 \text{ ou } A_2 \text{ ou ...ou } A_r)|B] = I\!P(A_1|B) + I\!P(A_2|B) + ... + I\!P(A_r|B)$$

Propriété 3.9 *Pour tout événement B et r événements $A_1, A_2, ..., A_r$ qui forment une partition, on a :*

$$I\!P(A_1|B) + I\!P(A_2|B) + ... + I\!P(A_r|B) = 1$$

En effet, de la Propriété 2.9, on a :

$$I\!P(B) = I\!P(B \text{ et } A_1) + I\!P(B \text{ et } A_2) + ... + I\!P(B \text{ et } A_r)$$

et donc :

$$\frac{I\!P(B)}{I\!P(B)} = \frac{I\!P(B \text{ et } A_1) + I\!P(B \text{ et } A_2) + ... + I\!P(B \text{ et } A_r)}{I\!P(B)} = 1$$

soit, comme $I\!P(B \text{ et } A_i)/I\!P(B) = I\!P(A_i|B)$:

$$I\!P(A_1|B) + I\!P(A_2|B) + ... + I\!P(A_r|B) = 1$$

Pour terminer, une dernière propriété dont nous laisserons, à titre d'exercice, la preuve au lecteur.

Propriété 3.10 *Pour tout événement B et deux événements A_1, A_2 tel que $A_1 \Rightarrow A_2$, on a :*

$$\mathbb{P}(A_1|B) \leq \mathbb{P}(A_2|B)$$

3.1.3. Exercices résolus

1- On tire au hasard une carte d'un jeu de 52 cartes et on considère les événements :

A = "la carte est un trèfle"
B = "la carte n'est pas l'as de trèfle"

a- Quel est la probabilité de l'événement A ?

Les résultats possibles (chacune des 52 cartes) sont équiprobables. On a $N_A = 13$ et $N = 52$. Donc :

$$\mathbb{P}(A) = \frac{N_A}{N} = \frac{13}{52} = 0,25$$

b- Quel est la probabilité conditionnelle de l'événement A sachant que B est réalisé ?

L'événement "A et B" est "la carte est un trèfle autre que l'as de trèfle". On a donc $N_{A \text{ et } B} = 12$. Comme $N_B = 51$, on obtient :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)} = \frac{N_{A \text{ et } B}}{N_B} = \frac{12}{51} = 0,235...$$

2- Une classe de 100 étudiants se compose des quatre groupes suivants :

	Garçons	Filles
Inscrit en gestion	17%	38%
Inscrit en économie	23%	22%

On choisit un étudiant au hasard comme délégué de classe.

a- Quelle est la probabilité que ce soit un étudiant inscrit en gestion ?

Posons :

A = "le délégué de classe est inscrit en gestion"

On cherche $\mathbb{P}(A)$. Les résultats possibles (chacun des 100 étudiants) sont équiprobables. On a $N_A = 17 + 38 = 55$ et $N = 100$. Donc :

$$\mathbb{P}(A) = \frac{N_A}{N} = \frac{55}{100} = 0,55$$

b- Le délégué de classe choisi est une fille. Quelle est la probabilité qu'elle soit inscrite en gestion ?

Posons :

$$B \quad = \quad \text{“le délégué de classe est une fille”}$$

On cherche $I\!P(A|B)$. L'événement "A et B" est "le délégué est une fille inscrite en gestion". On a donc $N_{A \text{ et } B} = 38$. Comme $N_B = 38 + 22 = 60$, on obtient :

$$I\!P(A|B) = \frac{I\!P(A \text{ et } B)}{I\!P(B)} = \frac{N_{A \text{ et } B}}{N_B} = \frac{38}{60} = 0,633...$$

c- Que deviennent ces résultats si, les proportions d'étudiants dans les différents groupes restant les mêmes, l'effectif de la classe passe de 100 à 500 ?

Ils restent inchangés. $I\!P(A)$ dépend uniquement de la proportion d'étudiants inscrits en gestion parmi tous les étudiants et $I\!P(A|B)$ de la proportion d'étudiantes inscrites en gestion parmi les filles. Ces deux probabilités restent donc inchangées tant que ces proportions sont elle-mêmes inchangées, ce qui est le cas lorsque les proportions d'étudiants dans les différents groupes restent les mêmes.

3- On tire au hasard successivement et sans remise trois boules d'une urne contenant 6 boules rouges et 4 boules noires.

a- Quelle est, sachant que la première boule tirée est noire, la probabilité que la seconde boule tirée soit rouge ?

Posons :

$$\begin{aligned} A &= \text{“la première boule tirée est noire”} \\ B &= \text{“la seconde boule tirée est rouge”} \end{aligned}$$

On cherche $I\!P(B|A)$. Sachant qu'une boule noire a été tirée lors du premier tirage (l'événement A est réalisé), il reste lors du deuxième tirage 9 résultats possibles, tous équiprobables : les dix boules initiales moins la boule noire tirée au premier tirage, soit 6 boules rouges et 3 boules noires. On a donc :

$$I\!P(B|A) = \frac{6}{9} = 0,333...$$

b- Quelle est, sachant que la première boule tirée est noire et la deuxième boule tirée est rouge, la probabilité que la troisième boule tirée soit à nouveau rouge ?

Posons :

$$C \quad = \quad \text{“la troisième boule tirée est rouge”}$$

On cherche maintenant $I\!P[C|(A \text{ et } B)]$. Sachant qu'une boule noire et une boule rouge ont été tirées lors des deux premiers tirages (l'événement "A et B" est réalisé), il reste lors du troisième tirage 8 résultats possibles, tous équiprobables : les dix boules initiales moins la boule noire et la boule rouge tirées lors des deux premiers tirages, soit 5 boules rouges et 3 boules noires. On a donc :

$$I\!P[C|(A \text{ et } B)] = \frac{5}{8} = 0,625$$

4- Un examen comporte une épreuve théorique et une épreuve pratique. On réussit à l'examen si on réussit à la fois à l'épreuve théorique et à l'épreuve pratique. La probabilité de réussir à l'épreuve théorique est de $0,6$, la probabilité de réussir à l'épreuve pratique est de $0,7$ et la probabilité de réussir à l'examen est de $0,5$. Quelle est la probabilité de réussir à l'épreuve théorique sachant qu'on réussit à l'épreuve pratique et vice-versa ?

Posons :

$$\begin{array}{lcl} A & = & \text{“réussir l'épreuve théorique”} \\ B & = & \text{“réussir l'épreuve pratique”} \end{array}$$

On a "A et B" = "réussir à l'examen". Des données, on a :

$$\mathbb{P}(A) = 0,6 \quad \mathbb{P}(B) = 0,7 \quad \text{et} \quad \mathbb{P}(A \text{ et } B) = 0,5$$

On doit calculer les probabilités conditionnelles $\mathbb{P}(A|B)$ et $\mathbb{P}(B|A)$. On a :

$$\begin{array}{lcl} \mathbb{P}(A|B) & = & \dfrac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)} = \dfrac{0,5}{0,7} = 0,714... \\ \mathbb{P}(B|A) & = & \dfrac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(A)} = \dfrac{0,5}{0,6} = 0,833... \end{array}$$

5- On lance trois fois de suite une pièce de monnaie équilibrée. Quelle est, sachant qu'on n'a pas obtenu pile au premier lancer, la probabilité :

a- de n'obtenir aucun pile ?

Posons :

$$\begin{array}{lcl} A_1 & = & \text{“n'obtenir aucun pile en trois lancers”} \\ B & = & \text{“ne pas obtenir pile au premier lancer”} \end{array}$$

et désignons pile et face par respectivement P et F.
L'épreuve aléatoire possède 2^3 résultats possibles, tous équiprobables :

$$\Omega = \left\{ \begin{array}{c} (P,P,P),(P,P,F),(P,F,P),(F,P,P), \\ (P,F,F),(F,P,F),(F,F,P),(F,F,F) \end{array} \right\}$$

On cherche $\mathbb{P}(A_1|B)$. On a $N_B = 4$ et $N_{A_1 \text{ et } B} = 1$. Donc :

$$\mathbb{P}(A_1|B) = \frac{\mathbb{P}(A_1 \text{ et } B)}{\mathbb{P}(B)} = \frac{N_{A_1 \text{ et } B}}{N_B} = \frac{1}{4} = 0,25$$

b- d'obtenir au moins une fois pile ?

Posons :

$$A_2 \quad = \quad \text{“obtenir au moins une fois pile en trois lancers”}$$

On cherche $\mathbb{P}(A_2|B)$. On a $A_2 = \overline{A_1}$. Donc :

$$\mathbb{P}(A_2|B) = \mathbb{P}(\overline{A_1}|B) = 1 - \mathbb{P}(A_1|B) = 1 - 0,25 = 0,75$$

c- d'obtenir au plus une fois pile ?

Posons :

A_3 = "obtenir au plus une fois pile en trois lancers"
A_4 = "obtenir exactement une fois pile en trois lancers"

On cherche $I\!P(A_3|B)$. On a A_3 = "A_1 ou A_4", les événements A_1, A_4 étant mutuellement exclusifs. On a donc :

$$I\!P(A_3|B) = I\!P[(A_1 \text{ ou } A_4)|B] = I\!P(A_1|B) + I\!P(A_4|B)$$

On a déjà trouvé $I\!P(A_1|B) = 0,25$. On a pour $I\!P(A_4|B)$:

$$I\!P(A_4|B) = \frac{I\!P(A_4 \text{ et } B)}{I\!P(B)} = \frac{N_{A_4 \text{ et } B}}{N_B} = \frac{2}{4} = 0,5$$

Au total, donc :

$$I\!P(A_3|B) = 0,25 + 0,50 = 0,75$$

3.1.4. Probabilité conditionnelle et loi de multiplication

De la définition de la probabilité conditionnelle $I\!P(A|B)$ découle directement la propriété suivante.

Propriété 3.11 *Pour tout événement A, B, on a :*

$$I\!P(A \text{ et } B) = I\!P(A|B) I\!P(B)$$

Ainsi donc, on peut obtenir la probabilité de l'intersection de deux événements A, B, c'est-à-dire la probabilité de l'événement "A et B", à partir de la probabilité conditionnelle de A sachant B et de la probabilité a priori de B (ou de façon équivalente, de la probabilité conditionnelle de B sachant A et de la probabilité a priori de A : il suffit de permuter les rôles de A et de B) : connaissant $I\!P(A|B)$ et $I\!P(B)$ (ou de façon équivalente $I\!P(B|A)$ et $I\!P(A)$), on peut par simple multiplication en déduire $I\!P(A \text{ et } B)$. Cette propriété est connue sous le nom de *loi* (ou *règle*) *de multiplication*.

Cette loi s'étend de façon naturelle au cas de plus de deux événements. Ainsi pour trois événements A, B, C, on a la propriété suivante.

Propriété 3.12 *Pour tout événement A, B, C, on a :*

$$I\!P(A \text{ et } B \text{ et } C) = I\!P[A|(B \text{ et } C)]\, I\!P(B|C) I\!P(C)$$

Cette propriété découle directement de la Propriété 3.11. On a en effet :

$$\begin{aligned} I\!P(A \text{ et } B \text{ et } C) &= I\!P[A \text{ et } (B \text{ et } C)] \\ &= I\!P[A|(B \text{ et } C)]\, I\!P(B \text{ et } C) \\ &= I\!P[A|(B \text{ et } C)]\, I\!P(B|C) I\!P(C) \end{aligned}$$

De la même façon, pour quatre événements A, B, C, D, on aura :

$$I\!P(A \text{ et } B \text{ et } C \text{ et } D) = I\!P[A|(B \text{ et } C \text{ et } D)]\, I\!P[B|(C \text{ et } D)]\, I\!P(C|D) I\!P(D)$$

et ainsi de suite.

Les probabilités conditionnelles satisfont également une loi de multiplication. La probabilité d'un événement "A_1 et A_2" sachant qu'un événement B est réalisé peut ainsi être obtenue de la probabilité conditionnelle de A_1 sachant "A_2 et B" et de la probabilité conditionnelle de A_2 sachant B (ou de façon équivalente, de la probabilité conditionnelle de A_2 sachant "A_1 et B" et de la probabilité conditionnelle de A_1 sachant B : il suffit de permuter les rôles de A_1 et A_2).

Propriété 3.13 *Pour tout événement A_1, A_2, B, on a :*

$$I\!P[(A_1 \textit{ et } A_2)|B] = I\!P[A_1|(A_2 \textit{ et } B)]\, I\!P(A_2|B)$$

On a en effet :

$$I\!P[(A_1 \text{ et } A_2)|B] = \frac{I\!P[(A_1 \text{ et } A_2) \text{ et } B]}{I\!P(B)} = \frac{I\!P(A_1 \text{ et } A_2 \text{ et } B)}{I\!P(B)}$$

Or, de la Propriété 3.12, on a $I\!P(A_1 \text{ et } A_2 \text{ et } B) = I\!P[A_1|(A_2 \text{ et } B)]\, I\!P(A_2|B) I\!P(B)$. Dès lors :

$$\begin{aligned} I\!P[(A_1 \text{ et } A_2)|B] &= \frac{I\!P[A_1|(A_2 \text{ et } B)]\, I\!P(A_2|B) I\!P(B)}{I\!P(B)} \\ &= I\!P[A_1|(A_2 \text{ et } B)]\, I\!P(A_2|B) \end{aligned}$$

Comme la règle de multiplication pour les probabilité a priori, cette règle de multiplication s'étend de façon naturelle aux cas de plus de deux événements. Pour trois événements, on aura :

$$I\!P[(A_1 \text{ et } A_2 \text{ et } A_3)|B)] = I\!P[A_1|(A_2 \text{ et } A_3 \text{ et } B)]\, I\!P[A_2|(A_3 \text{ et } B)]\, I\!P(A_3|B)$$

et ainsi de suite.

Notons pour conclure que la validité de toutes les formules ci-dessus requiert que les probabilités conditionnelles qui entrent dans leur composition soient toutes bien définies. Dans la cas de la Propriété 3.12 par exemple, cela signifie qu'on ait bien $I\!P(B \text{ et } C) \neq 0$ et $I\!P(C) \neq 0$.

3.1.5. Exercices résolus

1- Une association est composée de 43% d'hommes et de 57% de femmes. 23% des femmes lisent régulièrement le périodique de l'association pour seulement 5% des hommes. On considère un membre de l'association pris au hasard. Quelle est la probabilité que ce membre :

a- soit une femme qui lit régulièrement le périodique de l'association ?

Posons :

$$\begin{array}{lcl} A & = & \text{“le membre est une femme”} \\ B & = & \text{“le membre lit régulièrement le périodique de l'association”} \end{array}$$

On cherche $I\!\!P(A \text{ et } B)$. Des données, on a :

$$I\!\!P(A) = 0,57 \quad \text{et} \quad I\!\!P(B|A) = 0,23$$

On en déduit :

$$I\!\!P(A \,\text{et}\, B) = I\!\!P(A)I\!\!P(B|A) = 0,57 \times 0,23 = 0,1311$$

b- soit un homme qui ne lit pas régulièrement le périodique de l'association ?

On cherche maintenant $I\!\!P(\overline{A} \text{ et } \overline{B})$. Des données, on a :

$$\begin{array}{rcl} I\!\!P(\overline{A}) & = & 1 - I\!\!P(A) = 0,43 \\ I\!\!P(\overline{B}|\overline{A}) & = & 1 - I\!\!P(B|\overline{A}) = 1 - 0,05 = 0,95 \end{array}$$

Donc :

$$I\!\!P(\overline{A} \,\text{et}\, \overline{B}) = I\!\!P(\overline{A})I\!\!P(\overline{B}|\overline{A}) = 0,43 \times 0,95 = 0,4085$$

2- On tire au hasard successivement et sans remise quatre boules d'une urne contenant 3 boules blanches et 4 boules noires et 2 boules rouges. Quel est la probabilité :

a- d'obtenir une boule noire puis une boule rouge lors des deux premiers tirages ?

Posons :

$$\begin{array}{lcl} A & = & \text{“la première boule tirée est noire”} \\ B & = & \text{“la seconde boule tirée est rouge”} \end{array}$$

On cherche $I\!\!P(A \text{ et } B)$. Des données, on a :

$$I\!\!P(A) = \frac{4}{9} = 0,444... \quad \text{et} \quad I\!\!P(B|A) = \frac{2}{8} = 0,25$$

On en déduit :

$$I\!\!P(A \,\text{et}\, B) = I\!\!P(A)I\!\!P(B|A) = \frac{4}{9} \times \frac{2}{8} = \frac{1}{9} = 0,111...$$

b- d'obtenir une boule rouge, puis deux boules noires, puis une boule blanche ?

Posons :

$$\begin{array}{lcl} A_1 & = & \text{“obtenir une boule rouge au 1}^{\text{er}}\text{ tirage”} \\ A_2 & = & \text{“obtenir une boule noire au 2}^{\text{ème}}\text{ tirage”} \\ A_3 & = & \text{“obtenir une boule noire au 3}^{\text{ème}}\text{ tirage”} \\ A_4 & = & \text{“obtenir une boule blanche au 4}^{\text{ème}}\text{ tirage”} \end{array}$$

On cherche $I\!\!P(A_1 \text{ et } A_2 \text{ et } A_3 \text{ et } A_4)$. De façon semblable à ci-dessus, on trouve :

$$\begin{aligned} & I\!\!P(A_1 \text{ et } A_2 \text{ et } A_3 \text{ et } A_4\ ) \\ = \ & I\!\!P(A_1)I\!\!P(A_2|A_1)I\!\!P[A_3|(A_1 \text{ et } A_2)]\,I\!\!P[A_4|(A_1 \text{ et } A_2 \text{ et } A_3)] \\ = \ & \frac{2}{9}\times\frac{4}{8}\times\frac{3}{7}\times\frac{3}{6} = \frac{72}{3\,024} = 0,0238... \end{aligned}$$

c- d'obtenir une boule rouge puis une boule noire aux deux derniers tirages, sachant qu'on a obtenu une boule blanche et une boule noire aux deux premiers tirages ?

Posons :

$$\begin{array}{lcl} A_1 & = & \text{“obtenir une boule rouge au 3}^{\text{ème}}\text{ tirage”} \\ A_2 & = & \text{“obtenir une boule noire au 4}^{\text{ème}}\text{ tirage”} \end{array}$$

et

$$\begin{array}{lcl} B & = & \text{“obtenir une boule blanche et une boule noire lors des deux} \\ & & \quad\text{premiers tirages”} \end{array}$$

On cherche $I\!\!P[(A_1 \text{ et } A_2)|B]$. A nouveau de façon semblable à ci-dessus, on trouve :

$$\begin{aligned} I\!\!P[(A_1 \text{ et } A_2)|B] &= I\!\!P(A_1|B)I\!\!P[A_2|(A_1 \text{ et } B)] \\ &= \frac{2}{7}\times\frac{3}{6} = \frac{1}{7} = 0,142... \end{aligned}$$

3.1.6. Théorème de Bayes

Ce théorème est dénommé d'après le statisticien anglais Thomas BAYES (1702–1761). On appelle habituellement “théorème de BAYES” la formule (3.4) ci-dessous, mais par commodité nous mettrons aussi sous ce nom la formule (3.1), que nous allons présenter d'abord, et dont la formule (3.4) est une conséquence.

La première formule relie $I\!\!P(A|B)$ et $I\!\!P(B|A)$. On suppose qu'on a bien $I\!\!P(A) \neq 0$ et $I\!\!P(B) \neq 0$.

De la Propriété 3.11, on a :

$$I\!\!P(A \text{ et } B) = I\!\!P(A|B)I\!\!P(B)$$

En permutant les rôles de A et de B, on a de même :

$$\mathbb{P}(A \text{ et } B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

puisque "B et A" est la même chose que "A et B". On a donc :

$$\mathbb{P}(B|A)\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B)$$

On en déduit la première formule de BAYES.

Propriété 3.14 (Formule de BAYES I) *Pour tout événement A, B, on a :*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \tag{3.1}$$

Passons maintenant à la deuxième formule de BAYES, c'est-à-dire au théorème de BAYES proprement dit.

On considère d'une part un événement A, et d'autre part des événements $B_1, B_2, ..., B_r$ qui forment une partition. On suppose uniquement connues les probabilités a priori $\mathbb{P}(B_i)$, qui doivent toutes être non nulles, ainsi que les probabilités conditionnelles $\mathbb{P}(A|B_i)$. On se propose d'en déduire les probabilités conditionnelles $\mathbb{P}(B_i|A)$. Celles-ci sont données par la formule (3.1) :

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A)} \tag{3.2}$$

Il reste à exprimer $\mathbb{P}(A)$ à partir des données. De la Propriété 2.9 des probabilités a priori, on a :

$$\mathbb{P}(A) = \mathbb{P}(A \text{ et } B_1) + \mathbb{P}(A \text{ et } B_2) + ... + \mathbb{P}(A \text{ et } B_r)$$

Or, $\mathbb{P}(A \text{ et } B_j) = \mathbb{P}(A|B_j)\mathbb{P}(B_j)$. On a donc aussi :

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \cdots + \mathbb{P}(A|B_r)\mathbb{P}(B_r) \\ &= \sum_{j=1}^{r} \mathbb{P}(A|B_j)\mathbb{P}(B_j) \end{aligned} \tag{3.3}$$

En substituant (3.3) dans (3.2), on obtient la deuxième formule de BAYES.

Propriété 3.15 (Formule de BAYES II) *Pour tout événement A et r événements $B_1, B_2, ..., B_r$ qui forment une partition, on a :*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^{r} \mathbb{P}(A|B_j)\mathbb{P}(B_j)} \tag{3.4}$$

C'est la forme classique du théorème de BAYES. Dans la pratique, pour l'appliquer, on calcule souvent le dénominateur, qui est $\mathbb{P}(A)$, et on divise $\mathbb{P}(A|B_i)\mathbb{P}(B_i)$ par ce dénominateur, ce qui revient à appliquer la formule (3.1).

Un cas particulier important du théorème BAYES est celui où la partition est simplement composée d'un événement B et de son contraire $\overline{B}$.

Propriété 3.16 *Pour tout événement A, B, on a :*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|\overline{B})\mathbb{P}(\overline{B})}$$

Remarquons pour terminer que nous avons au passage dans cette section établi une nouvelle formule permettant de calculer la probabilité d'un événement à partir d'une partition.

Propriété 3.17 *Pour tout événement A et r événements $B_1, B_2, ..., B_r$ qui forment une partition, on a :*

$$\mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \cdots + \mathbb{P}(A|B_r)\mathbb{P}(B_r)$$

3.1.7. Exercices résolus

1- Une usine comporte trois ateliers numérotés $1, 2$ et 3 qui fabriquent la même pièce. L'atelier 1 fabrique 30% de la production, l'atelier 2, 50%, et l'atelier 3, 20%. 2% des pièces sortant de l'atelier 1 sont défectueuses, cette proportion étant de 3% pour les pièces sortant de l'atelier 2 et de 5% pour les pièces sortant de l'atelier 3. On prélève au hasard une pièce à la sortie de l'usine, et on constate qu'elle est défectueuse. Quelle est la probabilité qu'elle provienne de l'atelier 1 ?

Posons :

D = "la pièce est défectueuse"
A_i = "La pièce provient de l'atelier i" $(i = 1, 2, 3)$

Il est clair que les A_i $(i = 1, 2, 3)$ forment une partition.
Les données nous fournissent d'une part les probabilités a priori des A_i :

$$\mathbb{P}(A_1) = 0,3 \quad \mathbb{P}(A_2) = 0,5 \quad \text{et} \quad \mathbb{P}(A_3) = 0,2$$

et d'autre part les probabilités conditionnelles $\mathbb{P}(D|A_i)$:

$$\mathbb{P}(D|A_1) = 0,02 \quad \mathbb{P}(D|A_2) = 0,03 \quad \text{et} \quad \mathbb{P}(D|A_3) = 0,05$$

On cherche la probabilité conditionnelle $\mathbb{P}(A_1|D)$. Du théorème de BAYES, on a :

$$\mathbb{P}(A_1|D) = \frac{\mathbb{P}(D|A_1)\mathbb{P}(A_1)}{\mathbb{P}(D)}$$

où :

$$\begin{aligned} I\!P(D) &= I\!P(D|A_1)I\!P(A_1) + I\!P(D|A_2)I\!P(A_2) + I\!P(D|A_3)I\!P(A_3) \\ &= 0,02 \times 0,3 + 0,03 \times 0,5 + 0,05 \times 0,2 = 0,031 \end{aligned}$$

$I\!P(D)$ est la probabilité qu'une pièce sortant de l'usine (sans qu'on sache de quel atelier) soit défectueuse.
La probabilité demandée est :

$$I\!P(A_1|D) = \frac{0,02 \times 0,3}{0,031} = 0,193...$$

On peut calculer de même :

$$\begin{aligned} I\!P(A_2|D) &= \frac{0,03 \times 0,5}{0,031} = 0,484... \\ I\!P(A_3|D) &= \frac{0,05 \times 0,2}{0,031} = 0,323... \end{aligned}$$

On vérifie par ailleurs qu'on a :

$$I\!P(A_1|D) + I\!P(A_2|D) + I\!P(A_3|D) = 1$$

comme il se doit (cf. Propriété 3.9) puisque les A_i forment une partition.

2- Un premier panier contient 6 pommes et 5 pamplemousses. Un deuxième panier contient 6 pommes et 8 pamplemousses.

  a- Supposons tout d'abord qu'on tire au hasard un fruit dans le premier panier. Quelle est la probabilité qu'on obtienne une pomme ?

  Posons :

$$\begin{aligned} P &= \text{“le fruit tiré est une pomme”} \\ P_1 &= \text{“le fruit est tiré dans le panier 1”} \\ P_2 &= \text{“le fruit est tiré dans le panier 2”} \end{aligned}$$

  La probabilité demandée est :

$$I\!P(P|P_1) = \frac{6}{11} = 0,545...$$

  b- Même question si on tire dans le second panier.

  La probabilité demandée est cette fois :

$$I\!P(P|P_2) = \frac{6}{14} = 0,428...$$

  c- Choisissons à présent un panier au hasard et prélevons-y un fruit. Quelle est la probabilité que l'on obtienne une pomme ?

  On cherche $I\!P(P)$. On sait que $I\!P(P_1) = I\!P(P_2) = 0,5$.

Comme P_1 et P_2 forment une partition, de la Propriété 3.17, on a :

$$\begin{aligned} I\!P(P) &= I\!P(P|P_1)I\!P(P_1) + I\!P(P|P_2)I\!P(P_2) \\ &= \frac{6}{11} \times \frac{1}{2} + \frac{6}{14} \times \frac{1}{2} = 0,486... \end{aligned}$$

d- Comparez la probabilité ci-dessus avec celle que l'on obtient si tous les fruits sont réunis dans un seul panier.

Dans ce cas, la probabilité est :

$$I\!P(P) = \frac{6+6}{11+14} = \frac{12}{25} = 0,480$$

3.2. Dépendance et indépendance

3.2.1. Définitions

Considérons deux événements A, B relatifs à une même épreuve aléatoire. On suppose dans un premier temps que $I\!P(B) \neq 0$.

Définition 3.2 *On dit que l'événement A est indépendant de l'événement B si on a :*

$$I\!P(A|B) = I\!P(A) \tag{3.5}$$

Si $I\!P(A|B) \neq I\!P(A)$, on dit que A est dépendant de B.

$I\!P(A|B) = I\!P(A)$ signifie que le fait de savoir que B est réalisé ne rend ni plus ni moins probable le fait que A soit réalisé.

En multipliant les deux membres de la relation (3.5) par $I\!P(B)$, on obtient :

$$I\!P(A|B)I\!P(B) = I\!P(A)I\!P(B)$$

c'est-à-dire :

$$I\!P(A \text{ et } B) = I\!P(A)I\!P(B) \tag{3.6}$$

Si on suppose aussi que $I\!P(A) \neq 0$, en divisant les deux membres de la relation (3.6) par $I\!P(A)$, on obtient :

$$\frac{I\!P(A \text{ et } B)}{I\!P(A)} = I\!P(B)$$

c'est-à-dire :

$$I\!P(B|A) = I\!P(B) \tag{3.7}$$

en d'autres termes, que B est aussi indépendant de A.

En résumé, lorsque $I\!\!P(A) \neq 0$ et $I\!\!P(B) \neq 0$, si A est indépendant de B, B est aussi indépendant de A, et vice-versa.

La relation (3.6) fait apparaître de façon évidente le caractère symétrique de la relation d'indépendance. Par ailleurs, contrairement aux relations (3.5) et (3.7), elle a un sens quelles que soient les probabilités $I\!\!P(A)$ et $I\!\!P(B)$, nulles ou non. C'est donc elle qu'on retient pour définir de façon générale l'indépendance de deux événements.

Définition 3.3 *On dit que les événements A, B sont indépendants si on a :*

$$I\!\!P(A \text{ et } B) = I\!\!P(A) I\!\!P(B) \tag{3.8}$$

Si $I\!\!P(A \text{ et } B) \neq I\!\!P(A) I\!\!P(B)$, on dit que les événements A, B sont dépendants.

En dehors du cas où l'un des événements est de probabilité nulle, la Définition 3.3 est équivalente à la Définition 3.2 (et à sa "variante" symétrique où les rôles de A et de B sont permutés), et dans la pratique on utilise tantôt l'une tantôt l'autre.

Illustrons l'utilisation de ces définitions par l'exemple d'un tirage au hasard d'une carte dans un jeu de 52 cartes. Considérons les événements :

$$\begin{array}{rcl} A & = & \text{"la carte n'est pas un trèfle"} \\ B & = & \text{"la carte n'est pas un as"} \end{array}$$

On a :

$$\begin{array}{rcll} N & = & 52 & \text{(il y a 52 résultats possibles, tous équiprobables)} \\ N_A & = & 39 & \text{(il y a 39 cartes qui ne sont pas des trèfles)} \\ N_B & = & 48 & \text{(il y a 48 cartes qui ne sont pas des as)} \\ N_{A \text{ et } B} & = & 36 & \text{(il y a 36 cartes qui ne sont ni des trèfles, ni des as)} \end{array}$$

et donc :

$$\begin{array}{rcl} I\!\!P(A) & = & \dfrac{N_A}{N} = \dfrac{39}{52} = 0,75 \\ I\!\!P(A|B) & = & \dfrac{N_{A \text{ et } B}}{N_B} = \dfrac{36}{48} = 0,75 \end{array}$$

On le voit $P(A|B) = P(A)$. Les événements A et B sont donc indépendants : le fait de savoir qu'on a pas un as n'apporte aucune information sur les chances de ne pas avoir un trèfle.

On aurait pu procéder en comparant $I\!\!P(A \text{ et } B)$ et $I\!\!P(A) I\!\!P(B)$. On a :

$$\begin{array}{rcl} I\!\!P(A \text{ et } B) & = & \dfrac{N_{A \text{ et } B}}{N} = \dfrac{36}{52} = 0,692... \\ I\!\!P(A) I\!\!P(B) & = & \dfrac{N_A}{N} \times \dfrac{N_B}{N} = \dfrac{39}{52} \times \dfrac{48}{52} = \dfrac{1872}{2704} = 0,692... \end{array}$$

et la conclusion est évidemment la même.

On aurait pu encore procéder en comparant $I\!P(B)$ et $I\!P(B|A)$. On a :

$$\begin{aligned} I\!P(B) &= \frac{N_B}{N} = \frac{48}{52} = \frac{12}{13} \\ I\!P(B|A) &= \frac{N_{A \text{ et } B}}{N_A} = \frac{36}{39} = \frac{12}{13} \end{aligned}$$

et donc encore la même conclusion : de façon symétrique, le fait de savoir qu'on a pas de trèfle n'apporte aucune information sur les chances de ne pas avoir un as.

3.2.2. Propriété de la relation d'indépendance

Propriété 3.18 *Si les événements A et B sont indépendants, il y a aussi indépendance entre les événements A et $\overline{B}$, entre les événements $\overline{A}$ et B, et entre les événements $\overline{A}$ et $\overline{B}$.*

Il suffit de démontrer l'indépendance entre les événements A et $\overline{B}$, les deux autres cas s'ensuivent. De la Propriété 2.10, on a $I\!P(A) = I\!P(A \text{ et } B) + I\!P(A \text{ et } \overline{B})$, et donc aussi :

$$I\!P(A \text{ et } \overline{B}) = I\!P(A) - I\!P(A \text{ et } B)$$

Comme les événements A, B sont indépendants, $I\!P(A \text{ et } B) = I\!P(A)I\!P(B)$. On en déduit que :

$$\begin{aligned} I\!P(A \text{ et } \overline{B}) &= I\!P(A) - I\!P(A)I\!P(B) \\ &= I\!P(A)\,[1 - I\!P(B)] \\ &= I\!P(A)I\!P(\overline{B}) \end{aligned}$$

Les événements $A, \overline{B}$ sont donc bien indépendants.

On peut illustrer cette propriété en continuant l'exemple de la section précédente (tirage au hasard d'une carte dans un jeu de 52 cartes). On avait :

$$\begin{aligned} A &= \text{“la carte n'est pas un trèfle”} \\ B &= \text{“la carte n'est pas un as”} \end{aligned}$$

et

$$I\!P(A) = \frac{N_A}{N} = \frac{39}{52} = 0,75 \quad I\!P(A|B) = \frac{N_{A \text{ et } B}}{N_B} = \frac{36}{48} = 0,75$$

de sorte que les événements A, B sont indépendants.

L'événement $\overline{B}$ est “la carte est un as” et l'événement “A et $\overline{B}$” est “la carte est un as qui n'est pas un trèfle”. On a $N_{\overline{B}} = 4$ et $N_{A \text{ et } \overline{B}} = 3$. Donc :

$$\mathbb{P}(A|\overline{B}) = \frac{N_{A \text{ et } \overline{B}}}{N_{\overline{B}}} = \frac{3}{4} = 0,75$$

On a $P(A|\overline{B}) = P(A)$. Les événements $A, \overline{B}$ sont donc indépendants. On peut de la même façon vérifier que les événements $\overline{A}, B$ et $\overline{A}, \overline{B}$ sont bien indépendants.

3.2.3. Evénements mutuellement exclusifs et événements indépendants

Une erreur fréquente est de confondre "événements mutuellement exclusifs" (ou "événements incompatibles") avec "événements indépendants". En fait, deux événements mutuellement exclusifs sont généralement dépendants.

Propriété 3.19 *Si A, B sont deux événements mutuellement exclusifs, alors ils sont dépendants, sauf si l'un d'eux est de probabilité nulle.*

En effet, si A, B sont deux événements mutuellement exclusifs, et de probabilités non nulles, on a $\mathbb{P}(A \text{ et } B) = 0$ (puisque "A et B" $= \emptyset$). D'autre part, on a $\mathbb{P}(A)\mathbb{P}(B) \neq 0$. Donc, en dehors du cas où l'un d'eux a une probabilité nulle, deux événements mutuellement exclusifs sont toujours dépendants.

3.2.4. Evénements indépendants dans leur ensemble

Considérons trois événements A, B, C.

Définition 3.4 *On dit que A, B, C sont indépendants deux à deux s'il y a indépendance entre A et B, entre B et C, et entre A et C, autrement dit si on a les trois relations :*

$$\begin{aligned} \mathbb{P}(A \text{ et } B) &= \mathbb{P}(A)\mathbb{P}(B) \\ \mathbb{P}(B \text{ et } C) &= \mathbb{P}(B)\mathbb{P}(C) \\ \mathbb{P}(A \text{ et } C) &= \mathbb{P}(A)\mathbb{P}(C) \end{aligned} \tag{3.9}$$

Définition 3.5 *Trois événements indépendants deux à deux qui satisfont de plus la relation :*

$$\mathbb{P}(A \text{ et } B \text{ et } C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \tag{3.10}$$

sont dits indépendants dans leur ensemble.

On peut montrer que les trois relations (3.9) n'implique pas la relation (3.10). Réciproquement, la relation (3.10) n'implique aucune des trois relations (3.9). C'est pourquoi la notion d'événements indépendants dans leur ensemble requiert que toutes ces relations soient simultanément vérifiées.

Définition 3.6 *Des événements $A_1, A_2, ..., A_r$ sont dits indépendants dans leur ensemble si, pour toute sélection de k événements $A_{i_1}, ..., A_{i_k}$ de ces événements, on a :*

$$\mathbb{P}(A_{i_1} \text{ et } ... \text{ et } A_{i_k}) = \mathbb{P}(A_{i_1}) \times ... \times \mathbb{P}(A_{i_k}), \quad 2 \leq k \leq r \tag{3.11}$$

Dans la suite, on omettra les mots “dans leur ensemble”, et lorsque nous parlerons d'événements indépendants sans autre précision, il sera entendu que cela veut dire “indépendants dans leur ensemble”.

3.2.5. Indépendance et loi de multiplication

L'indépendance de deux ou plusieurs événements relatifs à une épreuve aléatoire donnée n'est pas toujours une propriété à vérifier, à démontrer. Il arrive souvent que l'indépendance de deux ou plusieurs événements soit admise comme hypothèse théorique étant donné la nature même des événements et de l'épreuve aléatoire considérés. Ainsi, lorsqu'on tire au hasard avec remise n boules dans une urne, il est naturel de considérer les événements relatifs à des tirages différents comme indépendants.

De façon générale, on considérera comme raisonnable l'hypothèse d'indépendance de deux événements A, B si on ne voit aucun mécanisme, direct ou indirect, par lequel la réalisation de l'événement A pourrait influer sur les chances — la probabilité — que l'événement B se réalise. Le même raisonnement vaut pour l'hypothèse d'indépendance de plusieurs événements.

Lorsque deux ou plusieurs événements peuvent, par hypothèse, de par la nature même de l'épreuve aléatoire, être considéré comme indépendants, par application des Définitions 3.3 et 3.6, la probabilité de l'intersection — opérateur logique “et” — de ces événements peut être obtenue par simple multiplication de leurs probabilités a priori. Ainsi, pour deux événements A, B indépendants, on aura :

$$\mathbb{P}(A \,\text{et}\, B) = \mathbb{P}(A)\mathbb{P}(B)$$

Pour trois événements A, B, C indépendants :

$$\mathbb{P}(A \,\text{et}\, B \,\text{et}\, C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

et ainsi de suite.

Cette règle de calcul étant un cas particulier de la loi de multiplication donnée à la Propriété 3.11 et des généralisations aux cas d'un nombre quelconque d'événements qui en découlent, elle est appelée *loi de multiplication pour des événements indépendants.*

3.2.6. Exercice résolu

1- On tire au hasard successivement et avec remise trois boules d'une urne contenant 6 boules rouges et 4 boules noires. Quelle est la probabilité :

a- d'obtenir une boule noire puis deux boules rouges ?

Posons :

$$\begin{array}{rcl} A & = & \text{“la première boule tirée est noire”} \\ B & = & \text{“la seconde boule tirée est rouge”} \\ C & = & \text{“la troisième boule tirée est rouge”} \end{array}$$

On cherche $I\!P(A \text{ et } B \text{ et } C)$. Les tirages étant effectués avec remise, les événements relatifs à des tirages différents, donc les événements A, B, C, peuvent être considérés comme indépendants. A chaque tirage, les résultats possibles — les 6 boules rouges et les 4 boules noires — sont équiprobables, de sorte qu'on a :

$$I\!P(A) = \frac{4}{10} = 0,4 \quad \text{et} \quad I\!P(B) = I\!P(C) = \frac{6}{10} = 0,6$$

De l'indépendance de A, B, C on déduit :

$$\begin{array}{rcl} I\!P(A \,\text{et}\, B \,\text{et}\, C) & = & I\!P(A)I\!P(B)I\!P(C) \\ & = & \dfrac{4}{10} \times \dfrac{6}{10} \times \dfrac{6}{10} = 0,144 \end{array}$$

b- d'obtenir deux boules rouges lors des deux derniers tirages ?

On cherche maintenant $I\!P(B \text{ et } C)$. De la même façon que ci-dessus, on trouve :

$$I\!P(B \,\text{et}\, C) = I\!P(B)I\!P(C) = \frac{6}{10} \times \frac{6}{10} = 0,36$$

2- On suppose qu'à chaque naissance, la probabilité d'avoir un fille est de $0,48$, et celle d'avoir un garçon de $0,52$. On considère la composition d'une famille de trois enfants et on représente le résultat de cette épreuve aléatoire par le triplet (S_1, S_2, S_3) où S_1 est le sexe du premier enfant, S_2 celui du second enfant et S_3 celui du troisième enfant.

a- Combien y a-t-il de résultats possibles ?

Désignons fille et garçon par respectivement F et G.
L'épreuve aléatoire possède 2^3 résultats possibles :

$$\Omega = \left\{ \begin{array}{c} (F,F,F),(G,F,F),(F,G,F),(F,F,G), \\ (G,G,F),(G,F,G),(F,G,G),(G,G,G) \end{array} \right\}$$

b- Quelle est la probabilité de chacun des résultats possibles ?

Posons :

$$\begin{array}{lcl|lcl} F_1 & = & \text{“le } 1^{\text{er}} \text{ enfant est une fille”} & \overline{F_1} & = & \text{“le } 1^{\text{er}} \text{ enfant est un garçon”} \\ F_2 & = & \text{“le } 2^{\text{ème}} \text{ enfant est une fille”} & \overline{F_2} & = & \text{“le } 2^{\text{ème}} \text{ enfant est un garçon”} \\ F_3 & = & \text{“le } 3^{\text{ème}} \text{ enfant est une fille”} & \overline{F_3} & = & \text{“le } 3^{\text{ème}} \text{ enfant est un garçon”} \end{array}$$

On a :

$$I\!P(F_1) = I\!P(F_2) = I\!P(F_3) = 0,48 \quad \text{et} \quad I\!P(\overline{F_1}) = I\!P(\overline{F_2}) = I\!P(\overline{F_3}) = 0,52$$

Les sexes des enfants obtenus lors des différentes naissances peuvent être considérés comme indépendants. On a donc :

$$\begin{array}{rcl} I\!P(\text{“obtenir } (F,F,F)\text{”}) & = & I\!P(F_1 \text{ et } F_2 \text{ et } F_3) = 0,48 \times 0,48 \times 0,48 \approx 0,11 \\ I\!P(\text{obtenir “}(G,F,F)\text{”}) & = & I\!P(\overline{F_1} \text{ et } F_2 \text{ et } F_3) = 0,52 \times 0,48 \times 0,48 \approx 0,12 \\ I\!P(\text{obtenir “}(F,G,F)\text{”}) & = & I\!P(F_1 \text{ et } \overline{F_2} \text{ et } F_3) = 0,48 \times 0,52 \times 0,48 \approx 0,12 \\ I\!P(\text{obtenir “}(F,F,G)\text{”}) & = & I\!P(F_1 \text{ et } F_2 \text{ et } \overline{F_3}) = 0,48 \times 0,48 \times 0,52 \approx 0,12 \\ I\!P(\text{obtenir “}(G,G,F)\text{”}) & = & I\!P(\overline{F_1} \text{ et } \overline{F_2} \text{ et } F_3) = 0,52 \times 0,52 \times 0,48 \approx 0,13 \\ I\!P(\text{obtenir “}(G,F,G)\text{”}) & = & I\!P(\overline{F_1} \text{ et } F_2 \text{ et } \overline{F_3}) = 0,52 \times 0,48 \times 0,52 \approx 0,13 \\ I\!P(\text{obtenir “}(F,G,G)\text{”}) & = & I\!P(F_1 \text{ et } \overline{F_2} \text{ et } \overline{F_3}) = 0,48 \times 0,52 \times 0,52 \approx 0,13 \\ I\!P(\text{obtenir “}(G,G,G)\text{”}) & = & I\!P(\overline{F_1} \text{ et } \overline{F_2} \text{ et } \overline{F_3}) = 0,52 \times 0,52 \times 0,52 \approx 0,14 \end{array}$$

c- Quelle est, sachant que la famille comporte au moins deux garçons, la probabilité que le troisième enfant soit une fille ?

Posons :

A = “la famille comporte au moins deux garçons”

On cherche $I\!P(F_3|A)$. L'événement A est réalisé pour les résultats possibles : “(G,G,F)”, “(G,F,G)”, “(F,G,G)” et “(G,G,G)”. Dès lors :

$$I\!P(A) = \sum_{i \in A} p_i \approx 0,13 + 0,13 + 0,13 + 0,14 = 0,53$$

D'autre part, l'événement “F_3 et A” est réalisé pour le seul résultat possible “(G,G,F)”. Donc :

$$I\!P(F_3 \text{ et } A) = \sum_{i \in (A \text{ et } B)} p_i \approx 0,13$$

On en déduit :

$$I\!P(F_3|A) = \frac{I\!P(F_3 \text{ et } A)}{I\!P(A)} \approx \frac{0,13}{0,53} = 0,245...$$

d- Quelle serait la probabilité de chacun des résultats possibles de cette épreuve aléatoire si on supposait qu'à chaque naissance, la probabilité d'avoir un fille est égale à celle d'avoir un garçon ?

Dans ce cas, tous les résultats possibles seraient équiprobables et leur probabilité égale à $1/8$. On a ici la justification de ce qu'on avait jusqu'à présent accepté intuitivement. La justification des résultats possibles équiprobables est la même pour le cas de lancers successifs d'une pièce équilibrée ou d'un dé non truqué.

3.3. Exercice résolu

On considère une population où 45% des ménages ont un revenu annuel inférieur à 100 000 F et 8% des ménages ont un revenu annuel supérieur à 300 000 F. La proportion des ménages qui possèdent une voiture est de 80%. Cette proportion est de 85% lorsqu'on considère uniquement les ménages ayant un revenu compris entre 100 000 F et 300 000 F. Elle est de 90% pour les ménages ayant un revenu supérieur à 300 000 F. On tire un ménage au hasard.

1- Calculez la probabilité que le ménage tiré dispose d'un revenu supérieur à 300 000 F et d'une voiture.

Posons :

$$\begin{array}{rcl} A_1 & = & \text{“le ménage tiré a un revenu inférieur à 100 000 F”} \\ A_2 & = & \text{“le ménage tiré a un revenu compris entre 100 000 F et 300 000 F”} \\ A_3 & = & \text{“le ménage tiré a un revenu supérieur à 300 000 F”} \\ B & = & \text{“le ménage tiré possède une voiture”} \end{array}$$

Les événements A_1, A_2 et A_3 forment une partition. Des données, on a :

$$I\!P(A_1) = 0,45 \qquad I\!P(A_3) = 0,08$$

et donc :

$$I\!P(A_2) = 1 - 0,45 - 0,08 = 0,47$$

Des données, on a encore :

$$I\!P(B) = 0,8 \qquad I\!P(B|A_2) = 0,85 \qquad I\!P(B|A_3) = 0,9$$

La probabilité demandée dans cette question est :

$$I\!P(B \text{ et } A_3) = I\!P(B|A_3)I\!P(A_3) = 0,9 \times 0,08 = 0,072$$

2- Calculez la probabilité que le ménage tiré dispose d'un revenu supérieur à 300 000 F ou d'une voiture.

La probabilité demandée est :

$$\begin{aligned} I\!P(B \text{ ou } A_3) &= I\!P(B) + I\!P(A_3) - I\!P(B \text{ et } A_3) \\ &= 0,8 + 0,08 - 0,072 = 0,808 \end{aligned}$$

3- On constate que le ménage tiré possède une voiture.

a- Quelle est la probabilité qu'il dispose d'un revenu compris entre 100 000 F et 300 000 F ?

La probabilité demandée est $I\!P(A_2|B)$. Elle est donnée par :

$$\begin{aligned} I\!P(A_2|B) &= \frac{I\!P(B|A_2)I\!P(A_2)}{I\!P(B)} \\ &= \frac{0,85 \times 0,47}{0,8} = 0,499... \end{aligned}$$

b- Quelle est la probabilité qu'il dispose d'un revenu supérieur à 100 000 F ?

La probabilité demandée est $I\!P\left[(A_2 \text{ ou } A_3)|B\right]$. Comme A_2 et A_3 sont mutuellement exclusifs, on a :

$$I\!P\left[(A_2 \text{ ou } A_3)|B\right] = I\!P(A_2|B) + I\!P(A_3|B)$$

On vient de trouver $I\!P(A_2|B) = 0,499...$. On a de même :

$$\begin{aligned} I\!P(A_3|B) &= \frac{I\!P(B|A_3)I\!P(A_3)}{I\!P(B)} \\ &= \frac{0,9 \times 0,08}{0,8} = 0,09 \end{aligned}$$

Donc :

$$I\!P\left[(A_2 \text{ ou } A_3)|B\right] = 0,09 + 0,499... = 0,589...$$

4- Un revenu élevé (supérieur à 300 000 F) et le fait de posséder une voiture sont-ils des événements indépendants ? Pourquoi ?

On a $I\!P(B|A_3) = 0,9$ et $I\!P(B) = 0,8$. Ces deux probabilités sont différentes. Donc les événements A_3 et B ne sont pas indépendants.

3.4. Exercices

1- Dans un jeu de roulette russe, une personne choisit au hasard une arme parmi trois revolvers disponibles et tire une fois. Les revolvers contiennent tous six chambres. Le nombre de chambres vides dans chaque arme est respectivement égal à $4, 3$ et 2. Trouvez la probabilité pour que la personne survive à ce jeu.

2- Trois quotidiens régionaux A, B, C sont diffusés dans une ville. Une enquête montre que parmi la population :

20%	lisent A exclusivement	5%	lisent A et C exclusivement
16%	lisent B exclusivement	4%	lisent B et C exclusivement
14%	lisent C exclusivement	2%	lisent A, B et C
8%	lisent A et B exclusivement		

a- On tire un individu au hasard.

i- Quelle est la probabilité qu'il lise au moins un journal ?
ii- Quelle est la probabilité qu'il en lise exactement deux ?

b- On tire un individu au hasard parmi les personnes qui lisent au moins un journal. Quelle est la probabilité qu'il lise exactement les deux journaux A et B ?

3- La probabilité pour un candidat quelconque d'être recruté dans une entreprise est de 3/20. La probabilité que Monsieur et Madame Dupont soient recrutés tous les deux dans cette même entreprise est de $0,02$.

a- Quelle est la probabilité qu'au moins un des deux époux soit recruté ?
b- Quelle est la probabilité que Monsieur Dupont soit recruté sachant que Madame Dupont l'est ?
c- Quelle est la probabilité que Monsieur Dupont soit recruté sachant que Madame Dupont ne l'est pas ?

4- Au cours d'une soirée, trois amis, Albert, Bernard et Charles, conviennent d'aller voir le film “Le Distrait” le lendemain. Ce jour-là, il y a 2 séances de projection et ils ont oublié de s'accorder sur l'heure de la séance. Supposez qu'il est également probable qu'Albert assiste à la première ou à la seconde séance, qu'il y a 7 chances sur 10 que Bernard assiste à la première et 8 chances sur 10 que Charles assiste à la seconde.

a- Quelle est la probabilité que tous aillent à la seconde séance ?
b- Quelle est la probabilité que tous aillent à la même séance ?
c- Quelle est la probabilité que 2 au moins des 3 amis assistent à la première séance ?

5- L'urne U_1 contient cinq billes rouges et cinq billes noires. L'urne U_2 contient quatre billes rouges et huit billes noires. L'urne U_3 contient trois billes rouges et six billes noires. On prélève au hasard une bille dans U_1 et, sans regarder sa couleur, on la place dans U_2. Ensuite on prélève au hasard une bille dans U_2 et, sans la regarder, on la place dans U_3. On tire au hasard une bille dans U_3. Quelle est la probabilité qu'elle soit rouge ?

6- Une manufacture emploie 100 ouvriers, dont les 3/4 ont plus de 30 ans, les 2/5 sont célibataires et les 4/5 sont syndiqués. On sait en outre que :

– parmi les ouvriers de moins de 30 ans, 40% sont célibataires.
– parmi les ouvriers célibataires, la proportion de syndiqués est de 3/4.

On demande :

a- La probabilité qu'un ouvrier pris au hasard :

i- soit célibataire ou syndiqué.
ii- soit célibataire sachant qu'il est syndiqué.
iii- soit célibataire et âgé de plus de 30 ans.
iv- soit célibataire sachant qu'il a plus de 30 ans.

b- S'il y a indépendance entre :

i- le fait d'être syndiqué et celui d'être célibataire. Pourquoi ?
ii- le fait d'être célibataire et celui d'avoir plus de 30 ans. Pourquoi ?

7- On considère un examen qu'on a le droit de présenter trois fois. La probabilité qu'on a de réussir du premier coup est de $0,4$, celle de réussir quand on a échoué une fois est de $0,5$ et celle de réussir quand on a échoué deux fois est $0,3$.

- a- Quelle est la probabilité de réussir la deuxième fois ? de réussir la troisième fois ?
- b- Quelle est la probabilité de réussir ?
- c- Sachant qu'une personne a réussi, quelles sont les probabilités qu'elle ait réussi à la première, à la deuxième, à la troisième tentative ?

8- On considère le jeu suivant. Un joueur A et un banquier B lancent chacun, simultanément, un dé correctement équilibré. Le joueur A est déclaré vainqueur si le banquier obtient un 4 ou si le résultat du banquier est inférieur ou égal au résultat du joueur A. Quelle est la probabilité que A gagne, étant donné que le résultat du banquier est différent de 4 ? Quelle est la probabilité que A gagne ?

9- Monsieur Legris est un grand distrait. Le matin, il oublie d'emporter son cartable avec une probabilité de $0,1$. Lorsqu'il oublie son cartable, il oublie aussi son parapluie avec une probabilité de $1/3$. La probabilité qu'il n'oublie ni son parapluie ni son cartable est de $11/15$.

- a- Quelle est la probabilité que Monsieur Legris oublie son cartable et son parapluie ?
- b- Quelle est la probabilité que Monsieur Legris n'oublie que son cartable ?
- c- Quelle est la probabilité que Monsieur Legris n'oublie pas son parapluie ?
- d- Sachant qu'il a oublié son parapluie, quelle est la probabilité que Monsieur Legris n'oublie pas son cartable ?
- e- Quelle est la probabilité que Monsieur Legris oublie son cartable ou son parapluie ?

10- Mme Dumur a un chien nommé Artaban et une chienne nommée Popsy. On considère le comportement des deux animaux quand le facteur passe. Artaban aboie 6 fois sur 10. Quand Artaban aboie, Popsy aboie aussi 7 fois sur 10. Quand Popsy aboie, Artaban aboie aussi 8 fois sur 10. Le facteur arrive.

- a- Quelle est la probabilité que les deux animaux aboient ?
- b- Quelle est la probabilité que Popsy aboie ?
- c- Quelles sont les probabilités qu'Artaban aboie seul, que Popsy aboie seule ?
- d- Quelle est la probabilité qu'aucun des deux n'aboie ?

11- Parmi les abonnés d'un quotidien, 50% résident en région parisienne, 40% en province et 10% à l'étranger. Ce quotidien publie un supplément mensuel. Parmi les abonnés de la région parisienne, 20% sont aussi abonnés au supplément. Parmi les abonnés de province, 90% ne sont pas abonnés au supplément. Enfin, 17% de tous les abonnés au quotidien sont aussi abonnés au supplément. Pour un abonné au quotidien pris au hasard, on demande de calculer les probabilités suivantes :

- a- qu'il réside à l'étranger et soit abonné au supplément.
- b- qu'il réside à l'étranger ou soit abonné au supplément.
- c- qu'il soit abonné au supplément sachant qu'il réside à l'étranger.
- d- qu'il réside à l'étranger sachant qu'il est abonné au supplément.
- e- qu'il réside en région parisienne sachant qu'il n'est pas abonné au supplé-

ment.

12- La probabilité que Monsieur Legris ait la grippe cet hiver est de $0,4$. Madame Legris est beaucoup plus résistante : la probabilité qu'elle ait la grippe n'est que de $0,15$. Lorsque Monsieur Legris a la grippe, Madame Legris l'a aussi avec une probabilité de $0,2$.

   a- Quelle est la probabilité que les époux Legris aient la grippe tous les deux ?
   b- Quelle est la probabilité que Monsieur Legris ait la grippe sachant que sa femme l'a ?
   c- Quelle est la probabilité qu'au moins un des deux époux ait la grippe ?
   d- Quelle est la probabilité que Monsieur Legris ait la grippe et pas Madame Legris ?
   e- Quelle est la probabilité que Monsieur Legris ait la grippe sachant que sa femme ne l'a pas ?
   f- Quelle est la probabilité qu'un seul des époux Legris ait la grippe ?

13- On dispose de deux dés équilibrés A et B. Le dé A comporte 4 faces rouges et 2 blanches, le dé B comporte 3 faces rouges et 3 blanches.

   a- Avant de lancer les dés, on joue à pile ou face (pièce équilibrée). Si on obtient pile, on joue uniquement avec le dé A, pour toute la suite du jeu. Si on obtient face, on joue uniquement avec le dé B.
      i- Quelle est la probabilité d'obtenir une face rouge après un premier lancer de dé ?
      ii- Quelle est la probabilité en trois lancers d'obtenir 3 fois une face rouge ?
   b- On n'utilise plus la pièce de monnaie et on lance ensemble les deux dés. Quelle est la probabilité d'obtenir deux faces de la même couleur ?

14- Pour tirer une boule d'une urne, on procède de la façon suivante. On lance d'abord un dé équilibré.

   – Si le résultat est pair, la boule est tirée au hasard dans une première urne qui contient 4 boules blanches et 6 boules rouges.
   – Si le résultat est 3 ou 5, la boule est tirée au hasard dans une deuxième urne qui contient 3 boules blanches, 3 boules rouges et 4 boules noires.
   – Si le résultat est 1, la boule est tirée au hasard dans une troisième urne qui contient 5 boules blanches, 2 boules rouges et 3 boules noires.

   Si on tire une boule rouge, on a gagné et le jeu s'arrête. Si c'est une boule noire, on a perdu et le jeu s'arrête également. Si c'est une boule blanche, on tire à nouveau dans la même urne sans remettre la boule blanche tirée dans l'urne.

   a- Quelle est la probabilité que le joueur gagne au premier tirage ?
   b- Quelle est la probabilité que le joueur perde au deuxième tirage ?
   c- Quelle est la probabilité que le jeu se termine en 2 tirages au plus ?

15- Une urne contient 3 boules blanches et 2 boules noires. Une autre urne contient 2 boules blanches et 5 boules noires. Une troisième urne contient 1 boule blanche et 4 boules noires. On tire au hasard une urne dans laquelle on tire au hasard une boule. On constate que la boule est blanche. Quelle est la probabilité qu'elle provienne de la deuxième urne ?

16- Supposons que 5 hommes sur 100 et 25 femmes sur 10 000 soient daltoniens. Choisissons un daltonien au hasard. Quelle est la probabilité que cette personne soit un homme, si l'on suppose que les hommes et les femmes sont en nombre égal ?

17- On considère 100 boîtes de 5 fusibles. 20 boîtes proviennent de l'entreprise A, 30 boîtes proviennent de l'entreprise B et 50 boîtes proviennent de l'entreprise C. La probabilité qu'un fusible venant de A soit défectueux est de 5%, venant de B, la probabilité est de 4%, et de C, la probabilité est de 2%. On tire une boîte au hasard et on teste un fusible. Il est défectueux. Quelle est la probabilité qu'il provienne de B ?

18- Un concessionnaire automobile vend deux modèles de voitures, la VX9 et la Sultane. Il a vendu 250 voitures dont 150 VX9 et 100 Sultanes. La probabilité qu'une VX9 nécessite une réparation dans l'année est de $0,21$, et cette probabilité est de $0,13$ pour une Sultane. Un client vient faire réparer sa voiture achetée depuis moins d'un an. Quelle est la probabilité que ce soit une Sultane ?

19- Un chercheur de pétrole évalue la probabilité de la présence de pétrole (P) à $0,2$. Une étude spéciale d'échantillons de roches est réalisée. Sur la base des résultats obtenus par le passé, une telle étude se conclut par un verdict positif (VP) dans 70% des cas où effectivement il y a du pétrole et elle se conclut négativement (VN) dans 95% des cas où il n'y a effectivement pas de pétrole. Déterminez la probabilité de la présence de pétrole conditionnellement à :

   a- un verdict positif.
   b- un verdict négatif.

20- Une pochette contient une paire de dés truqués (T) : un dé marque la valeur 3 sur chaque face tandis que l'autre marque la valeur 4 sur chaque face. Une deuxième pochette contient une paire de dés normaux (N). On choisit une pochette au hasard et on lance les dés qu'elle contient.

   a- Quelle est la probabilité que les dés truqués soient choisis ?
   b- La somme obtenue s'élève à 7. Déterminez la probabilité que les dés ainsi lancés soient les dés truqués.

21- Un livre de 100 pages numérotées de 1 à 100 comporte 3 chapitres. Le premier chapitre comporte 20 pages, le second 30 pages et le troisième 50 pages. Il y a 70 pages de texte et 30 pages de tableaux réparties comme suit :

   – Chapitre I : 20 pages de texte et 0 page de tableaux.
   – Chapitre II : 20 pages de texte et 10 pages de tableaux.
   – Chapitre II : 30 pages de texte et 20 pages de tableaux.

   On tire au hasard une page.

   a- On constate que c'est une page de texte. Quelle est la probabilité qu'elle provienne du chapitre III ?
   b- Quelle est la probabilité de tirer une page paire ?
   c- Quelle est la probabilité d'obtenir la page 39 ?
   d- Quelle est la probabilité de tirer une page paire du Chapitre II ?
   e- Quelle est la probabilité de tirer une page de texte ?

f- Quelle est la probabilité de tirer un tableau du Chapitre I ?

22- La probabilité que Marius annonce qu'il a attrapé un poisson est de 1 s'il en a réellement attrapé un. Cette probabilité est de $0,7$ s'il n'a rien pêché du tout. En supposant que Marius attrape un poisson un jour sur quatre, quelle est la probabilité qu'il dise la vérité s'il annonce ce soir qu'il en a pêché un ?

23- M. Dupont joue tous les dimanches avec son ami Dupond. Par expérience, on sait qu'ils vont décider à pile ou face s'ils vont jouer aux échecs ou au tennis. S'ils jouent aux échecs, la probabilité que Dupond gagne est de $0,8$ mais s'ils jouent au tennis, la probabilité que Dupond gagne est de $0,1$. Lundi dernier, Dupont rencontre son ami Tryphon et lui annonce triomphalement que la veille, il a battu son adversaire. Quelle est la probabilité que ce soit au tennis ?

24- La Belgique comporte 3 régions linguistiques : la Wallonie, la Flandre et Bruxelles, couvrant respectivement 50%, 45% et 5% du territoire. Un martien est tombé quelque part en Belgique mais on ne sait où. On sait que la première personne à qui il s'est adressé lui a répondu en français. Etant donné qu'il y a 10% de francophones en Flandre, 80% à Bruxelles et 100% en Wallonie, quelle est la probabilité que ce martien soit tombé en région bruxelloise ?

25- En Belgique, il pleut 4 jours sur 10 et il fait beau les 6 autres jours. Malgré les instruments sophistiqués dont il dispose, M. Météo se trompe souvent : il prévoit de la pluie alors qu'il va faire beau dans 30% des cas et il prévoit du soleil alors qu'il va pleuvoir dans 10% des cas.

   a- Quelle est la probabilité qu'il pleuve demain ?
   b- Si M. Météo prévoit de la pluie pour demain, quelle est la probabilité qu'il pleuve ?
   c- Si M. Météo prévoit de la pluie pour demain, quelle est la probabilité qu'il fasse beau ?

26- Un terrain carré de 200 m de côté recouvre trois temples grecs, l'un d'Héra qui mesure 40 m sur 20, l'autre d'Apollon qui mesure 35 m sur 18, et le troisième de Déméter qui mesure 30 m sur 15. Un archéologue fait un sondage en un point pris au hasard dans ce terrain.

   a- Quelle est la probabilité qu'il tombe sur un des trois temples ?
   b- Sachant qu'il tombe sur un des temples, quelle est la probabilité que ce soit celui de Déméter ?
   c- L'archéologue a un texte qui lui apprend que le site renferme ces trois temples et qui en donne les dimensions, mais qui ne dit rien sur leur situation précise. Au fond de son sondage, il trouve une statuette votive. Il sait que dans un temple d'Héra on a une probabilité de $0,5$ de trouver ce type de statuette, dans un temple d'Apollon une probabilité de $0,3$, et dans un temple de Déméter, une probabilité de $0,6$.

      i- Quelle est l'identification la plus probable du temple dans lequel il est tombé ?
      ii- Quelle est l'identification la moins probable ?

27- Une entreprise gère deux camps de vacances, l'un à la mer, l'autre à la montagne. Ces camps sont ouverts en juillet et en août. 60% de la clientèle passe ses vacances à la mer. D'autre part, il y a autant de clients qui prennent leurs vacances en juillet qu'en août. Enfin, 30% de la clientèle passe ses vacances à la montagne pendant le mois d'août.

a- Quelle est la probabilité qu'un client pris au hasard passe ses vacances à la montagne au mois de juillet ?

b- Quelle est la probabilité qu'un client pris au hasard passe ses vacances à la mer au mois de juillet ?

c- Sachant qu'un client passe ses vacances à la mer :

  i- quelle est la probabilité que ce soit en juillet ?

  ii- quelle est la probabilité que ce soit en août ?

d- La probabilité qu'un client ait un accident pendant son séjour est de $0,03$ pour un client en vacances à la mer, et de $0,02$ pour un client en vacances à la montagne. Sachant qu'un client a eu un accident, quelle est la probabilité que ce soit à la mer ?

28- Les employés d'une entreprise manufacturière sont regroupés par service. Chaque employé est affecté à un seul service. Ainsi 20% des employés sont affectés au service de la comptabilité, 28% au service du marketing, et les autres sont affectés au service de la production. Parmi les employés affectés au service de la comptabilité, 65% travaillent depuis plus de 10 ans. Au service du marketing, 43% des employés travaillent depuis plus de 10 ans, alors qu'il y en a 40% au service de la production. Le pourcentage d'employés gagnant plus de 200 000 F par année est indépendant du service : il est de 20%. Quelle est la probabilité qu'un employé de cette entreprise choisi au hasard :

a- soit affecté au service de la comptabilité depuis plus de 10 ans dans cette entreprise ?

b- travaille depuis plus de 10 ans dans cette entreprise ?

c- soit affecté au service de la production s'il a 10 ans ou moins d'ancienneté ?

d- soit affecté au service du marketing ou de la production, et gagne plus de 200 000 F par année ?

e- gagne plus de 200 000 F par année s'il est affecté au service du marketing ou de la production ?

29- Trois secrétaires tapent des adresses sur un lot d'enveloppes. La secrétaire A remplit 50% des enveloppes, la secrétaire B en remplit 30%, la secrétaire C, 20%. La probabilité qu'il y ait une faute dans une adresse est de $0,03$ si elle est tapée par la secrétaire A, de $0,02$ si elle est tapée par la secrétaire B et de $0,01$ si elle est tapée par la secrétaire C.

a- On tire une lettre au hasard. L'adresse est fausse. Quelle est la probabilité que cette adresse ait été tapée par la secrétaire A ?

b- On tire une lettre au hasard. L'adresse est exacte. Quelle est la probabilité que cette adresse ait été tapée par la secrétaire A ?

30- Une enquête est faite auprès d'une population étudiante d'un campus universitaire. On note F la population féminine et I l'ensemble des étudiants, garçons et filles, sachant jouer d'un instrument de musique. L'enquête révèle que F représente 48% de la population étudiante, I représente 40% de la population étudiante et que, chez les étudiants du groupe I, 45% sont des filles.

a- Quelle est la probabilité qu'un étudiant choisi au hasard soit musicien ?

b- On interroge un étudiant au hasard.

  i- Quelle est la probabilité que ce soit une fille sachant jouer d'un instrument de musique ?

  ii- Sachant qu'il est musicien, quel est la probabilité que ce soit un garçon ?

c- On interroge une fille au hasard sur le campus. Quelle est la probabilité que cette fille sache jouer d'un instrument de musique ?

d- Parmi les musiciens, trois garçons sur cinq jouent de la guitare et deux filles sur cinq jouent de la guitare. On interroge un étudiant au hasard sur le campus. Sachant qu'il est guitariste, quelle est la probabilité que ce soit un garçon ?

31- Les professeurs Nimbus et Tournesol partagent leur bureau avec une collègue qu'ils ennuient beaucoup lorsqu'ils parlent de leur passion commune, le football. Lorsque Nimbus parle, la probabilité que ce soit de football est de $0,6$. Tournesol est plus raisonnable car, lorsqu'il parle, la probabilité que ce soit de football n'est que de $0,4$. Hélas lorsque Nimbus parle football, la probabilité que Tournesol entame alors avec lui une conversation footballistique est de $0,6$. La collègue est excédée et vous demande :

a- Quelle est la probabilité que Nimbus et Tournesol lorsqu'ils sont ensemble parlent de football ?

b- Le mardi où nous sommes tous les trois dans le bureau, quelle est la probabilité que je n'entende pas parler de football ?

c- Quelle est la probabilité que Nimbus ne parle pas de football quand Tournesol n'en parle pas ?

32- L'hypermarché Aupré est desservi par deux parkings A et B.

a- La probabilité que le parking A soit saturé le samedi matin à 11 h est de $0,3$, celle que le parking B soit saturé le même jour à la même heure est de $0,4$. La probabilité que les deux parkings soient saturés est de $0,18$.

  i- Y a-t-il indépendance entre les saturations des deux parkings ?

  ii- Quelle est la probabilité que le parking A soit saturé sachant que le parking B l'est ?

  iii- Quelle est la probabilité qu'aucun des deux parkings ne soit saturé ?

  iv- Quelle est la probabilité que le parking A soit saturé sachant que B ne l'est pas ?

b- M. Legris va à l'hypermarché chaque samedi matin à 11 h. Trois samedis sur quatre, venant de chez lui, il se présente d'abord au parking A. Un samedi sur quatre, venant de son bureau, il sa présente d'abord au parking B.

  i- Quelle est la probabilité, un samedi quelconque, que M. Legris ne trouve pas de place au premier parking où il va ?

  ii- Samedi dernier, M. Legris n'a pas trouvé de place au premier parking où il est allé. Quelle est la probabilité que ç'ait été au parking A ?

33- Un laboratoire vient de mettre au point un nouveau test pour dépister la maladie de Badhealth, qui atteint 2% de la population. Les essais montrent que :

- 97 fois sur 100, le test donne un résultat positif alors que la personne est vraiment malade.
- 96 fois sur 100, le test donne un résultat négatif alors que la personne n'est pas malade.

On pratique le test sur une personne prise au hasard et on constate que le test est positif. Quelle est la probabilité que la personne ne soit pas malade ?

34- Le tableau suivant est issu de l'article : "Du domicile à l'établissement scolaire : les trajets quotidiens des jeunes en 1991-1992", Brigitte Baccaïni, *Economie et statistique*, 293, 1996. Il donne la distribution de la durée du trajet entre le domicile et l'établissement selon le cycle scolaire (en %).

	Moins de 5 minutes	5 à 14 minutes	15 à 29 minutes	30 à 44 minutes	45 à 59 minutes	60 minutes et plus
Maternelle	16,5	70,8	10,7	1,8	0,2	0,0
Primaire	12,1	69,8	13,9	3,0	0,6	0,7
Secondaire	1,6	35,4	36,6	15,5	6,2	4,8
Supérieur	0,0	7,5	33,4	23,3	13,0	22,7
Ensemble	8,4	51,8	23,3	9,0	3,8	3,7

a- Que représentent exactement les chiffres donnés dans ce tableau ? Interprétez-les en termes probabilistes.

b- On tire au hasard un élève de maternelle.

i- Quelle est la probabilité que son trajet scolaire dure moins de 5 minutes ?

ii- Quelle est la probabilité qu'il dure moins de 30 minutes ?

c- On tire un élève au hasard.

i- Quelle est la probabilité que son trajet scolaire dure moins de 5 minutes ?

ii- Quelle est la probabilité qu'il dure 30 minutes ou plus ?

35- Un enquêteur est chargé d'étudier diverses caractéristiques de la population d'un quartier situé dans la banlieue d'une grande ville. Après avoir consulté les données de l'INS (Institut National de la Statistique), il sait que 34% des familles qui habitent le quartier sont d'origine immigrée et que 70% de celles-ci comportent au moins deux enfants. D'autre part, il sait que, globalement, 55% des familles du quartier comportent au moins deux enfants.

a- Quelle est la probabilité qu'une famille d'origine immigrée du quartier prise au hasard comporte au plus un enfant ?

b- Quelle est la probabilité qu'une famille du quartier prise au hasard soit d'origine immigrée et comporte au plus un enfant ?

c- Quelle est la probabilité qu'une famille du quartier prise au hasard soit d'origine immigrée et comporte au moins deux enfants ?

d- Quelle est la probabilité qu'une famille du quartier prise au hasard ne soit pas d'origine immigrée ou comporte au plus un enfant ?

e- On considère une famille choisie au hasard dans le quartier. On constate que la famille comporte au moins deux enfants. Quelle est la probabilité que cette famille soit d'origine immigrée ?

36- On étudie certaines pratiques culturelles d'un échantillon de cadres composé pour 35% de cadres supérieurs et 65% de cadres moyens. On a les données suivantes : 51% des cadres supérieurs et 45% des cadres moyens lisent plus de 20 livres par an, 28% des cadres supérieurs et 18% des cadres moyens assistent à un concert de musique classique au moins une fois par an, 27% des cadres supérieurs lisent plus de 20 livres par an et assistent à un concert de musique classique au moins une fois par an.

a- On tire un individu au hasard dans l'échantillon.

i- Quelle est la probabilité qu'il lise plus de 20 livres par an ?

ii- Quelle est la probabilité qu'il assiste à un concert de musique classique plus d'une fois par an ?

b- On tire un cadre moyen au hasard dans l'échantillon. Quelle est la probabilité qu'il ne lise pas plus de 20 livres par an ?

c- On tire au hasard un individu qui lit plus de 20 livres par an au hasard dans l'échantillon. Quelle est la probabilité qu'il s'agisse d'un cadre supérieur ?

d- On tire un cadre supérieur au hasard dans l'échantillon. Quelle est la probabilité qu'il lise moins de 20 livres par an et assiste à un concert moins d'une fois par an ?

37- Le journal *Libération* publiait le 18 janvier 1994 les résultats d'un sondage réalisé par l'Ifop durant la manifestation du 16 janvier en faveur de l'école publique. Près de 2 400 personnes ont été interrogées pour comprendre leur motivation et cerner leur personnalité. *Libération* publie les tableaux suivants.

Profil des manifestants

Sexe	
Homme	60%
Femme	40%
Total	100%

Age	
15 - 24 ans	13%
25 - 34 ans	20%
35 - 49 ans	45%
50 - 64 ans	19%
65 et plus	3%
Total	100%

Niveau d'études	
Primaire	5%
Secondaire	20%
Professionnel, technique ou commercial	12%
Supérieur	63%
Total	100%

Scolarisation des enfants	
Dans un établissement public	59%
Dans un établissement privé	3%
N'a pas d'enfants scolarisés	11%
N'a pas d'enfants du tout	30%
Total supérieur à 100% en raison des réponses multiples	

Statut professionnel	
Travaillent dans le secteur public	58%
Travaillent dans le secteur privé	19%
Etudiant ou lycéen n'ayant jamais travaillé	11%
Femme au foyer ou sans profession	1%
Chômeur	3%
Retraité	8%
Total	100%

Profession	
Etudiant	11%
Ouvrier	3%
Employé	11%
Cadre moyen, technicien	18%
Instituteur	18%
Enseignant dans le secondaire	20%
Enseignant dans le supérieur	4%
Commerçant, artisan	1%
Cadre supérieur	8%
Profession libérale	3%
Agriculteur	1%
Inactif	2%
Total	100%

Profil politique des manifestants

Sympathie partisane déclarée	
Parti des Travailleurs, Lutte Ouvrière, et Ligue Communiste Révolutionnaire	9%
Alternative Démocratie Socialiste et Parti Communiste	15%
Mouvement des Citoyens, Parti Socialiste et Mouvement des Radicaux de Gauche	52%
Génération Ecologie et Les Verts	9%
UDF et RPR	2%
Front National	0%
Ne se prononcent pas	13%
Total	100%

Affiliation	
A un parti politique	18%
A un syndicat	43%
A une association de parents d'élèves	25%
A une autre association	29%
Total supérieur à 100% en raison des réponses multiples	

Les raisons des manifestants

Quelle est la raison la plus importante pour laquelle vous êtes venus manifester aujourd'hui ?

	Ensemble	Selon la sympathie partisane déclarée				
		Extrême gauche	PC ADS	PS, MRG MDC	Eco-logistes	Droite
Pour défendre l'école publique	88%	79%	89%	90%	88%	89%
Pour répondre l'appel d'un syndicat ou d'une organisation	3%	4%	6%	2%	2%	0%
Pour manifester votre opposition au gouvernement	11%	19%	11%	9%	13%	5%
Pour aider la gauche	3%	4%	3%	3%	1%	3%
Ne se prononcent pas	0%	0%	0%	0%	0%	3%
Total supérieur à 100% en raison des réponses multiples						

Dans l'idéal, en ce qui concerne les établissements privés, que souhaitez-vous ?

	Ensemble	Selon la sympathie partisane déclarée				
		Extrême gauche	PC ADS	PS, MRG MDC	Eco-logistes	Droite
Maintenir le mode actuel de financement sous conditions	29%	14%	18%	36%	31%	49%
Créer un service public unifié d'éducation	27%	22%	33%	26%	31%	11%
Ne pas financer du tout les établissements privés	42%	62%	48%	37%	37%	35%
Ne se prononcent pas	2%	2%	1%	1%	1%	5%
total	100%	100%	100%	100%	100%	100%

Aujourd'hui, pour vous, que représente la laïcité ?

	Ensemble	Selon l'âge				
		15 à 24 ans	25 à 34 ans	35 à 49 ans	50 à 64 ans	65 et plus
L'égalité des chances	28%	25%	29%	28%	26%	32%
la séparation de l'Eglise et de l'Etat	13%	15%	14%	9%	16%	14%
La tolérance	16%	10%	11%	16%	21%	32%
L'école pour tous	43%	47%	46%	45%	34%	37%
Un moyen d'intégration	9%	6%	8%	10%	11%	12%
Ne se prononcent pas	0%	0%	0%	0%	0%	0%
Total supérieur à 100% en raison des réponses multiples						

Répondez aux questions suivantes ou expliquez pourquoi il est impossible d'y répondre, en précisant les données qui manquent pour le faire.

a- Proposez une représentation graphique pour le niveau d'études des manifestants.
b- Tracez l'histogramme des fréquences des âges des manifestants.
c- Quelle est la probabilité qu'un manifestant choisi au hasard :
    - i- soit âgé de plus de 65 ans ?
    - ii- soit un homme de plus de 65 ans ?
    - iii- soit enseignant ?
    - iv- ne soit affilié ni à un parti, ni à un syndicat, ni à une association de parents, ni à une autre association ?
    - v- soit affilié à une association de parents d'élèves ?
    - vi- soit affilié à un parti politique et à un syndicat ?
    - vii- soit affilié à un parti ou à un syndicat ?
    - viii- soit d'extrême gauche ?

ix- soit motivé par la défense de l'école publique sachant qu'il est d'extrême gauche ?
x- soit motivé par la défense de l'école publique et soit d'extrême gauche ?

d- Quelle est la probabilité qu'un manifestant écologiste choisi au hasard :
i- soit présent pour répondre à l'appel d'un syndicat ou d'une organisation ?
ii- pense qu'il ne faut pas du tout financer l'école privée ?

e- Quelle est la probabilité qu'un manifestant choisi au hasard soit écologiste et pense qu'il ne faut pas du tout financer l'école privée ?

f- Sachant qu'un manifestant choisi au hasard a des enfants scolarisés dans l'enseignement public, quelle est la probabilité qu'il ait aussi des enfants scolarisés dans l'enseignement privé, et vice versa ?

g- Dans le tableau reprenant les raisons de manifester, on voit apparaître 3% dans la colonne "sympathisants de droite" et 0% dans la colonne "ensemble", pour les manifestants qui ne se prononcent pas. Ne pensez-vous pas qu'il y a une erreur ?

Chapitre 4

Variables aléatoires

Dans ce chapitre et le suivant, on développe de nouveaux outils de description d'une épreuve aléatoire, basé sur la notion de *variable aléatoire*.

Définition 4.1 *On appelle variable aléatoire une variable qui prend différentes valeurs suivant une certaine distribution de probabilité.*

Une variable aléatoire X décrit une épreuve aléatoire dont les résultats possibles sont les différentes valeurs que X peut prendre et dont les chances de réalisation des résultats possibles sont dépeintes par la distribution de probabilité de X.

De nombreuses épreuves aléatoires peuvent être représentées, directement ou par l'intermédiaire d'un *codage*, par une variable aléatoire.

Considérons par exemple le jeu de hasard suivant. On jette un dé équilibré :

– si on obtient un 4, on gagne 100 F.
– si on obtient un 6, on gagne 20 F.
– si on obtient un 2 ou un 3, on perd 50 F.
– si on obtient un autre résultat, on ne perd ni ne gagne rien.

On peut représenter cette épreuve aléatoire par une variable aléatoire X donnant le gain (positif ou négatif) du jeu. L'ensemble des *valeurs possibles* de X, que l'on note $\mathcal{X}$, est pour ce jeu :

$$\mathcal{X} = \{-50, 0, 20, 100\}$$

Voyons deux autres exemples :

1- On lance une pièce de monnaie. On peut encore représenter cette épreuve aléatoire par une variable aléatoire X qui, par convention (il s'agit tout simplement d'un codage), prend la valeur 0 lorsqu'on obtient face et la valeur 1 lorsqu'on obtient pile. On a ici :

$$\mathcal{X} = \{0, 1\}$$

2- On prend au hasard une ampoule d'une marque donnée et on teste sa durée de vie. On peut à nouveau représenter cette épreuve aléatoire par une variable aléatoire X mesurant le temps qui s'écoule avant que l'ampoule testée claque. On a cette fois (a priori) :

$$\mathcal{X} = [0, +\infty)$$

Parmi les variables aléatoires, on distingue les variables aléatoires *discrètes* et les variables aléatoires *continues*. Les variables aléatoires discrètes sont celles qui ne prennent que des valeurs isolées : un nombre fini de valeurs (comme dans notre exemple introductif, $\mathcal{X} = \{-50, 0, 20, 100\}$) ou, tout au plus, un nombre infini mais dénombrable de valeurs (par exemple $\mathcal{X} = \{1, 2, 3, 4, ...\}$). Les variables aléatoires continues prennent un continuum de valeurs dans un (ou plusieurs) intervalle(s) (comme dans notre exemple de la durée de vie d'une ampoule).

Il est important de bien distinguer une variable aléatoire des valeurs qu'elle prend. C'est pourquoi nous adoptons la convention suivante : les *variables aléatoires* seront notées en *lettres capitales* tandis que les *valeurs prises* par ces variables seront notées en *lettres minuscules*. Ainsi, lorsque nous écrirons "$X = x$", "$X < x$" ou encore "$x_1 \leq X \leq x_2$", il faut lire respectivement "la variable aléatoire X prend la valeur x", "la variable aléatoire X prend une valeur inférieure à x" et "la variable aléatoire X prend une valeur comprise entre x_1 et x_2 (bornes incluses)". Notez que "$X = x$", "$X < x$" ou encore "$x_1 \leq X \leq x_2$" expriment des événements relatifs l'épreuve aléatoire représentée par la variable aléatoire X.

4.1. Variables aléatoires discrètes

4.1.1. Loi d'une variable aléatoire discrète

Définition 4.2 *Une variable aléatoire discrète X est une variable aléatoire qui prend un nombre fini, ou infini mais dénombrable, de valeurs x_i.*

Définition 4.3 *On appelle loi ou distribution de probabilité d'une variable aléatoire discrète X un ensemble de couples (x_i, p_{x_i}), où les $p_{x_i} = \mathbb{P}(X = x_i)$ sont des nombres réels tels que :*

$$\begin{cases} 0 \leq p_{x_i} \leq 1, \quad \forall\, x_i \in \mathcal{X} \\ \displaystyle\sum_{x_i} p_{x_i} = 1 \end{cases} \tag{4.1}$$

$\mathcal{X}$ désignant l'ensemble des valeurs possibles x_i de X et $\sum_{x_i}$ signifiant "somme pour toutes les valeurs possibles x_i de X".

Tout ensemble de couples (x_i, p_{x_i}) satisfaisant les conditions (4.1) de la définition ci-dessus définit la loi d'une variable aléatoire discrète.

On le voit, la notion de loi d'une variable aléatoire discrète n'est guère différente de la notion de distribution de probabilité définie au Chapitre 1. Simplement, les résultats possibles de l'épreuve aléatoire qu'elle représente sont ici des *nombres réels*. Comme on le verra dans la suite, on peut abondamment tirer parti de cette caractéristique.

Reprenons notre exemple de jeu de hasard du début. Les différentes valeurs possibles de X (les différents gains possibles) sont :

$$\mathcal{X} = \{-50, 0, 20, 100\}$$

On perd 50 F lorsqu'on obtient un 2 ou un 3 en jetant le dé. Le dé étant supposé équilibré, on a :

$$I\!P(X = -50) = I\!P(\text{“obtenir un 2 ou un 3”}) = \frac{1}{3}$$

De la même façon, on a :

$$I\!P(X = 0) = I\!P(\text{“obtenir un 1 ou un 5”}) = \frac{1}{3}$$
$$I\!P(X = 20) = I\!P(\text{“obtenir un 6”}) = \frac{1}{6}$$
$$I\!P(X = 100) = I\!P(\text{“obtenir un 4”}) = \frac{1}{6}$$

La loi de la variable aléatoire X décrivant notre jeu est donc :

Valeurs possibles x_i	-50	0	20	100
$p_{x_i} = I\!P(X = x_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

On vérifie que la somme des probabilités est bien égale à 1.

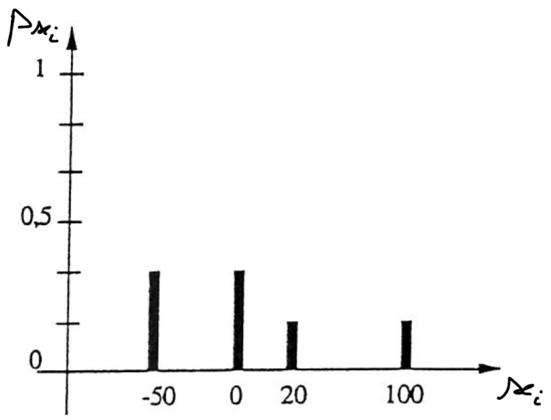
Une variable aléatoire discrète X et sa loi associée étant définie, le calcul de la probabilité d'événements relatifs à X tels que “$X = x$”, “$X < x$” ou encore “$x_1 \leq X \leq x_2$” s'effectue de la façon habituelle, de manière générique en faisant la somme des probabilités des différentes valeurs possibles x_i de X pour lesquelles l'événement considéré se réalise. Ainsi, on a par exemple :

$$\begin{aligned} I\!P(-10 \leq X < 100) &= I\!P(X = 0 \text{ ou } X = 20) \\ &= I\!P(X = 0) + I\!P(X = 20) \\ &= \frac{1}{3} + \frac{1}{6} = \frac{1}{2} \end{aligned}$$

ou encore :

$$\begin{aligned} I\!P(X \geq 0) &= 1 - I\!P(X < 0) = 1 - I\!P(X = -50) \\ &= 1 - \frac{1}{3} = \frac{2}{3} \end{aligned}$$

Outre un tableau donnant, pour les différentes valeurs possibles x_i de X, la valeur des probabilités $p_{x_i} = I\!P(X = x_i)$, on représente souvent la loi d'une variable aléatoire discrète à l'aide d'un diagramme en bâtons. Le Graphique 4.1 illustre cette représentation pour notre exemple.



Graphique 4.1 : La loi de la variable aléatoire X

Cette représentation n'est pas sans rappeler le diagramme en bâtons rencontré en statistique descriptive pour représenter la *distribution de fréquence* d'une variable x prenant un petit nombre de valeurs distinctes x_i dans une population. Le lien entre distribution de fréquence et distribution de probabilité est simple : on passe de la distribution de fréquence dans une population pour une variable x donnée à une distribution de probabilité identique pour cette même variable au travers de l'épreuve aléatoire "tirer au hasard un individu dans la population et observer la valeur de la variable x pour l'individu tiré".

Ainsi par exemple, supposons que la variable "nombre de véhicules possédés par un ménage" dans une population ait une distribution de fréquence donnée par (on suppose pour simplifier qu'un ménage ne possède jamais plus de 4 véhicules) :

Nbr. de véhicules possédés par un ménage	0	1	2	3	4
Fréquence dans la population	18%	62%	15%	4%	1%

On tire au hasard un ménage dans cette population et on appelle X la variable aléatoire donnant le nombre de véhicules possédés par le ménage tiré. La loi de X est de façon évidente donnée par :

Valeurs possibles x_i	0	1	2	3	4
Probabilités p_{x_i}	$0,18$	$0,62$	$0,15$	$0,04$	$0,01$

c'est-à-dire par une distribution identique à la distribution de fréquence de la variable d'intérêt x dans la population.

4.1.2. Fonction de répartition d'une variable aléatoire discrète

Définition 4.4 *On appelle fonction de répartition d'une variable aléatoire discrète X la fonction $F(x)$ définie par :*

$$F(x) = I\!P(X \leq x) = \sum_{x_i \leq x} p_{x_i}$$

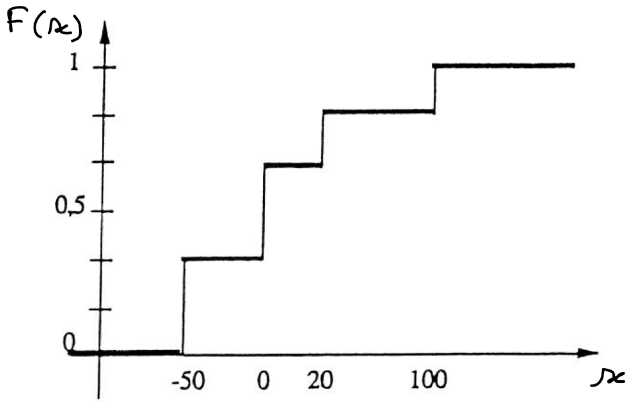
Il s'agit, en quelque sorte, de *probabilités cumulées.* Illustrons ce concept à l'aide de notre exemple du jeu de hasard, dont la loi de X (= gains associés au jeu) est donnée par :

x_i	-50	0	20	100
p_{x_i}	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

On a :

$$F(x) = I\!P(X \leq x) = \begin{cases} 0 & \text{pour } x < -50 \\ \dfrac{1}{3} & \text{pour } -50 \leq x < 0 \\ \dfrac{1}{3} + \dfrac{1}{3} = \dfrac{2}{3} & \text{pour } 0 \leq x < 20 \\ \dfrac{2}{3} + \dfrac{1}{6} = \dfrac{5}{6} & \text{pour } 20 \leq x < 100 \\ \dfrac{5}{6} + \dfrac{1}{6} = 1 & \text{pour } x \geq 100 \end{cases}$$

Graphiquement, dans le cas de variables aléatoires discrètes, la fonction de répartition est une fonction en escalier. Elle a un point de discontinuité en chaque valeur $x_i \in \mathcal{X}$. Elle est croissante et varie de 0 à 1.



Graphique 4.2 : La fonction de répartition de la variable aléatoire X

La notion de fonction de répartition d'une variable aléatoire discrète X est évidemment à rapprocher de la notion analogue, rencontrée en statistique descriptive, de *fréquence cumulée* d'une variable x prenant un petit nombre de valeurs distinctes x_i dans une population.

4.1.3. Espérance et variance d'une variable aléatoire discrète

Tout comme on peut résumer la distribution de fréquence d'une variable dans une population par différents indicateurs synthétiques, on peut résumer la loi d'une variable aléatoire par différentes quantités destinées à en mesurer la tendance centrale, la dispersion, l'asymétrie, l'aplatissement, etc... On se concentrera ici sur les deux indicateurs les plus utilisés que sont l'*espérance* (ou moyenne) et la *variance* de X, qu'on qualifie aussi de *moments d'ordre* 1 *et* 2 de la loi de X.

Définition 4.5 *On appelle espérance (ou moyenne) d'une variable aléatoire discrète X la quantité :*

$$E(X) = \sum_{x_i} x_i p_{x_i}$$

L'espérance de X est la moyenne pondérée par leur probabilité p_{x_i} des différentes valeurs possibles x_i de X. Par définition, elle est toujours comprise entre la plus petite et la plus grande des valeurs possibles de X. Dans notre exemple du jeu de hasard, où la loi de X (= gains associés au jeu) est donnée par :

x_i	-50	0	20	100
p_{x_i}	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

on a :

$$\begin{aligned} E(X) &= -50 \times \frac{1}{3} + 0 \times \frac{1}{3} + 20 \times \frac{1}{6} + 100 \times \frac{1}{6} \\ &= \frac{-100 + 20 + 100}{6} = \frac{10}{3} = 3,333... \end{aligned}$$

On dit que l'espérance du gain, ou encore le gain espéré, vaut $3,333$ F.

L'espérance de X est un indicateur synthétique de la *tendance centrale* (ou *localisation*) de la loi de X, de la tendance centrale (ou localisation) des valeurs susceptibles d'être prises par X. $E(X)$ est souvent utilisée pour caractériser la valeur de X à laquelle on peut s'attendre, comme *prévision* de la valeur qu'est susceptible de prendre X.

Dans l'optique de l'analogie entre distribution de fréquence d'une variable x prenant un petit nombre de valeurs distinctes x_i dans une population et distribution de probabilité d'une variable aléatoire discrète X, l'espérance de X correspond tout simplement à la *moyenne* $\overline{x}$ de la variable x en question dans la population, moyenne qui peut être calculée par la formule :

$$\overline{x} = \sum_{x_i} \frac{n_{x_i}}{n} x_i$$

où $\sum_{x_i}$ signifie ici “somme pour toutes les valeurs distinctes x_i prises par la variable x dans le population”, n_{x_i} désigne le nombre d’individus de la population pour lesquels la variable x prend la valeur x_i, n le nombre total d’individus dans la population, et donc $\frac{n_{x_i}}{n}$ la fréquence de la valeur x_i dans la population, qui est l’analogue de la probabilité p_{x_i}.

Définition 4.6 *On appelle variance d’une variable aléatoire discrète X la quantité :*

$$V(X) = \sum_{x_i} (x_i - \mu)^2 p_{x_i}, \quad \textit{où } \mu = E(X)$$

et écart-type de X la quantité :

$$\sigma(X) = \sqrt{V(X)}$$

La variance de X est la moyenne pondérée par les probabilités p_{x_i} du carré des écarts à sa valeur espérée des différentes valeurs possibles x_i de X. L’écart-type de X est une simple transformation (la racine carrée) de la variance. Par définition, $V(X)$ et $\sigma(X)$ sont toujours positifs ou nuls. Pour notre exemple du jeu de hasard, on a (aux arrondis près) :

$$\begin{aligned}
V(X) &= (-50 - 3,333)^2 \times \frac{1}{3} + (0 - 3,333)^2 \times \frac{1}{3} \\
&\quad +(20 - 3,333)^2 \times \frac{1}{6} + (100 - 3,333)^2 \times \frac{1}{6} \\
&= 948,136 + 3,703 + 46,298 + 1557,418 \\
&= 2555,555 \\
\sigma(X) &= \sqrt{2555,555} = 50,552
\end{aligned}$$

La variance de X, et par conséquent son écart-type, est un indicateur synthétique de la *dispersion* de la loi de X, de la dispersion des valeurs susceptibles d’être prises par X. Elle est d’autant plus grande que la plage des valeurs possibles de X (autour de son espérance) est étendue et que les probabilités de ses valeurs possibles extrêmes (éloignées de son espérance) sont importantes.

L’intérêt de l’écart-type comme indicateur de dispersion est qu’il donne une grandeur dont l’unité de mesure est la même que l’unité de mesure de X, ce qui n’est pas le cas de la variance.

Dans l’optique de l’analogie entre distribution de fréquence d’une variable x prenant un petit nombre de valeurs distinctes x_i dans une population et distribution de probabilité d’une variable aléatoire discrète X, $V(X)$ et $\sigma(X)$ correspondent tout simplement à la *variance* σ^2 et l’*écart-type* σ de la variable x en question dans la population, quantités qui peuvent être calculées par les formules :

$$\sigma^2 = \sum_{x_i} \frac{n_{x_i}}{n} (x_i - \overline{x})^2 \quad \text{et} \quad \sigma = \sqrt{\sigma^2}$$

On peut établir une autre formule pour la variance $V(X)$, souvent plus commode pour les calculs. On a :

$$\begin{aligned}
V(X) &= \sum_{x_i}(x_i-\mu)^2 p_{x_i} \qquad\qquad (\text{où } \mu = E(X)) \\
&= \sum_{x_i}(x_i^2 - 2\mu x_i + \mu^2)p_{x_i} \\
&= \sum_{x_i} x_i^2 p_{x_i} - 2\mu \underbrace{\sum_{x_i} x_i p_{x_i}}_{=\mu} + \mu^2 \underbrace{\sum_{x_i} p_{x_i}}_{=1} \\
&= \sum_{x_i} x_i^2 p_{x_i} - 2\mu^2 + \mu^2 \\
&= \sum_{x_i} x_i^2 p_{x_i} - \mu^2
\end{aligned}$$

Propriété 4.1 *Soit X une variable aléatoire discrète. On a :*

$$V(X) = \sum_{x_i} x_i^2 p_{x_i} - \mu^2, \quad \textit{où } \mu = E(X)$$

4.1.4. Exercice résolu

Les élèves d'une classe de sixième primaire viennent de passer une interrogation de mathématique. Sur les 20 élèves de la classe, 11 ont obtenu le maximum, 6 ont obtenu 9/10, 2 ont obtenu 8/10 et 1 seul a obtenu 7/10. On tire au hasard un élève de la classe et on désigne par X la variable aléatoire donnant sa note (sur 10).

1- Quelle est la loi de X ?

L'ensemble des valeurs possibles de X est :

$$\mathcal{X} = \{7, 8, 9, 10\}$$

De façon évidente, les probabilités de ces valeurs possibles sont :

$$\mathbb{P}(X=7) = \frac{1}{20} = 0,05 \qquad \mathbb{P}(X=8) = \frac{2}{20} = 0,1$$
$$\mathbb{P}(X=9) = \frac{6}{20} = 0,3 \qquad \mathbb{P}(X=10) = \frac{11}{20} = 0,55$$

La loi de X est donc :

x_i	7	8	9	10
p_{x_i}	0,05	0,1	0,3	0,55

On vérifie que la somme des probabilités est bien égale à 1.

2- Calculez la probabilité que la note de l'élève tiré soit :

a- égale à $8,5$.

La probabilité demandée est :

$$\mathbb{P}(X = 8,5) = 0$$

puisque "$X = 8,5$" est un événement impossible.

b- inférieure à $9,5$.

La probabilité demandée est :

$$\mathbb{P}(X < 9,5) = 1 - \mathbb{P}(X \geq 9,5) = 1 - \mathbb{P}(X = 10) = 0,45$$

c- égale à 9 sachant qu'elle est comprise entre $7,5$ et $9,5$.

La probabilité demandée est :

$$\begin{aligned}
\mathbb{P}(X = 9 | 7,5 \leq X \leq 9,5) &= \frac{\mathbb{P}(X = 9 \text{ et } 7,5 \leq X \leq 9,5)}{\mathbb{P}(7,5 \leq X \leq 9,5)} \\
&= \frac{\mathbb{P}(X = 9)}{\mathbb{P}(X = 8) + \mathbb{P}(X = 9)} \\
&= \frac{0,3}{0,1 + 0,3} = \frac{3}{4} = 0,75
\end{aligned}$$

3- Calculez l'espérance de la note de l'élève tiré.

On trouve :

$$E(X) = \sum_{x_i} x_i p_{x_i} = 7 \times 0,05 + 8 \times 0,1 + 9 \times 0,3 + 10 \times 0,55 = 9,35$$

4- Calculez l'écart-type de X.

On a :

$$\begin{aligned}
V(X) &= \sum_{x_i} x_i^2 p_{x_i} - \mu^2, \quad \text{où } \mu = E(X) \\
&= 7^2 \times 0,05 + 8^2 \times 0,1 + 9^2 \times 0,3 + 10^2 \times 0,55 - (9,35)^2 = 0,7275
\end{aligned}$$

D'où :

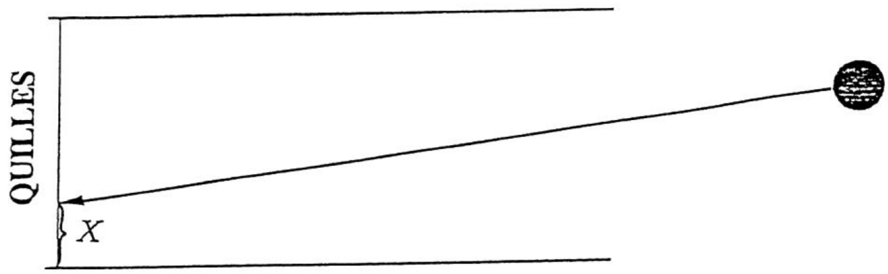
$$\sigma(X) = \sqrt{V(X)} = \sqrt{0,7275} = 0,852...$$

4.2. Variables aléatoires continues

4.2.1. Loi d'une variable aléatoire continue

Définition 4.7 *Une variable aléatoire continue X est une variable aléatoire qui prend un continuum de valeurs x dans un (ou plusieurs) intervalle(s).*

Considérons par exemple un individu qui joue au bowling et supposons qu'on s'intéresse au point d'impact de la boule lancée. On peut représenter cette épreuve aléatoire par une variable aléatoire X mesurant la distance entre le bord gauche de l'espace où se situent les quilles et le point d'impact de la boule.



Graphique 4.3 : La variable aléatoire X

Si on suppose que la largeur de l'espace où se situent les quilles est de 2 m, l'ensemble $\mathcal{X}$ des valeurs possibles de X est :

$$\mathcal{X} = [0, 2]$$

On est dans la situation d'une épreuve aléatoire représentée par une variable aléatoire X continue, qui peut prendre un continuum de valeurs x dans un intervalle donné.

Lorsqu'une variable aléatoire X peut ainsi prendre un continuum de valeurs, il n'est plus possible de définir la loi ou distribution de probabilité de X par un ensemble de couples (x_i, p_{x_i}) : il est impossible de faire la liste d'un continuum de valeurs.

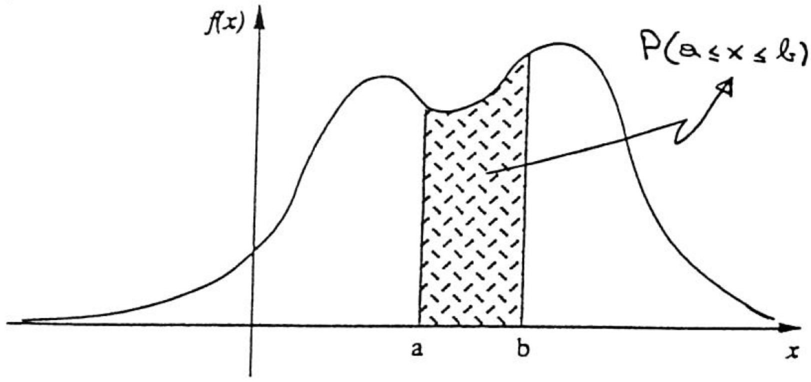
La loi ou distribution de probabilité d'une variable aléatoire continue X est définie par l'intermédiaire d'un concept particulier, le concept de *densité de probabilité*.

Définition 4.8 *On appelle densité de probabilité d'une variable aléatoire continue X une fonction $f(x)$ telle que :*

$$\begin{cases} f(x) \geq 0, \quad \forall\, x \in \mathbb{R} \\ \displaystyle\int_{-\infty}^{+\infty} f(x)dx = 1 \end{cases} \tag{4.2}$$

et

$$I\!P(a \leq X \leq b) = \int_a^b f(x)dx$$



Graphique 4.4 : Une fonction de densité $f(x)$

Une fonction de densité $f(x)$ est une fonction positive, strictement positive pour l'ensemble des valeurs possibles x de X et nulle pour toutes les autres valeurs, dont l'aire totale sous la courbe qu'elle représente est égale à 1, et dont la surface sous la courbe dans un intervalle $[a, b]$ — notez qu'on peut prendre $a = -\infty$ et/ou $b = +\infty$ — définit la probabilité que X prenne sa valeur dans cet intervalle.

Toute fonction $f(x)$ satisfaisant les conditions (4.2) de la Définition 4.8 définit la loi ou distribution de probabilité d'une variable aléatoire continue. Jointe à la condition $f(x) \geq 0, \forall x \in \mathbb{R}$, la condition $\int_{-\infty}^{+\infty} f(x)dx = 1$ assure que, comme il se doit, la probabilité de tout événement relatif à X est toujours un nombre compris entre 0 et 1.

La condition $\int_{-\infty}^{+\infty} f(x)dx = 1$ est à mettre en parallèle avec la condition $\sum_{x_i} p_{x_i} = 1$ du cas discret. De même, la formule générique de la probabilité $I\!P(a \leq X \leq b) = \int_a^b f(x)dx$ est à mettre en parallèle avec la formule générique $I\!P(a \leq X \leq b) = \sum_{a \leq x_i \leq b} p_{x_i}$ valable dans le cas discret. On le voit, le passage du cas discret au cas continu revient pour l'essentiel à remplacer une liste de probabilités ponctuelles (les couples (x_i, p_{x_i})) par une fonction de densité et des sommes par des intégrales[3].

Par définition, la probabilité qu'une variable aléatoire continue X prenne une valeur ponctuelle "$X = a$" est donnée par :

$$I\!P(X = a) = I\!P(a \leq X \leq a) = \int_a^a f(x)dx = 0$$

La probabilité de toute valeur a prise isolément, qu'elle soit ou non possible — lorsqu'elle est impossible, on a en plus $f(a) = 0$ —, est donc nulle.

Propriété 4.2 *Si X est une variable aléatoire continue, quelle que soit la valeur*

[3] Rappelons qu'une intégrale n'est rien d'autre que la limite d'une somme.

a considérée, on a :

$$\mathbb{P}(X = a) = 0$$

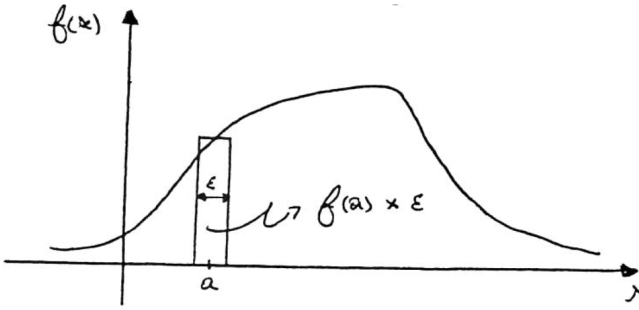
On en déduit immédiatement la propriété suivante.

Propriété 4.3 *Si X est une variable aléatoire continue, quelles que soient les valeurs a et b considérées, on a :*

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b) = \int_a^b f(x)dx$$

Comment intuitivement interpréter la notion de densité de probabilité d'une variable aléatoire continue ? Considérons un *petit intervalle* ε centré sur une valeur quelconque a. On a :

$$\mathbb{P}(a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f(x)dx \approx f(a) \times \varepsilon$$

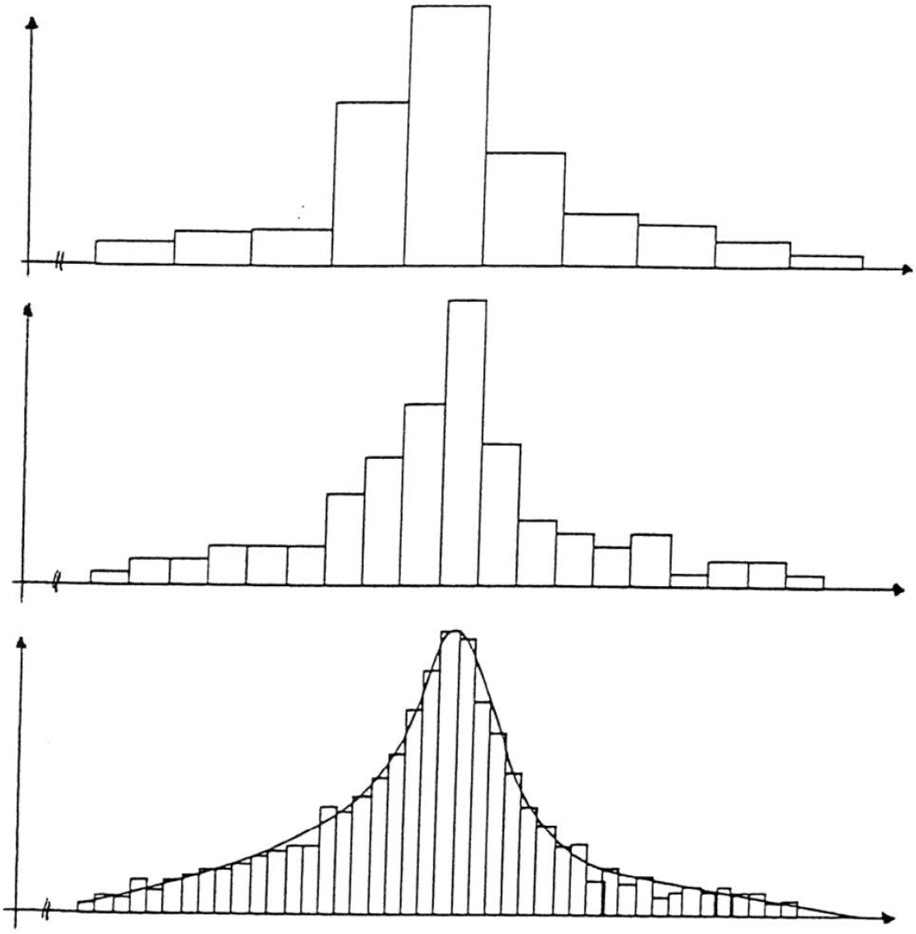


Graphique 4.5 : Interprétation d'une fonction de densité $f(x)$

En d'autres termes, la densité $f(x)$ est approximativement proportionnelle à la probabilité que la variable aléatoire X prenne sa valeur dans une petit intervalle ε centré sur x. La fonction de densité $f(x)$ nous renseigne donc sur les probabilités *relatives* d'observer X dans un *même* petit intervalle ε autour des différentes valeurs de x.

La notion densité de probabilité d'une variable aléatoire continue est par ailleurs à rapprocher de la notion d'*histogramme de fréquence* utilisée en statistique descriptive pour décrire, dans une population, une variable x prenant un grand nombre de valeurs distinctes très proches les unes des autres.

Supposons qu'on mesure la taille de tous les individus d'une population. Si la population est grande et que la mesure de la taille est très précise (disons au millimètre près), on aura un grand nombre de valeurs distinctes très proches les unes des autres. Si on regroupe ces valeurs en *classes*, disons dans un premier temps de 10 cm, et qu'on en trace l'histogramme de fréquence — rappelons que dans un histogramme de fréquence, la fréquence des individus dans les différentes classes est représentée par la *surface* des rectangles correspondant à ces classes, et donc que la surface totale des rectangles est égale à 1 — on obtiendra sans doute quelque chose de très semblable à la figure du haut du Graphique 4.6.



Graphique 4.6 : Histogramme de fréquence et densité de probabilité

En répétant la même opération pour des classes de plus en plus petites — 5 cm, $2,5$ cm, 1 cm, etc..., on obtiendra des histogrammes de fréquence dont les rectangles seront de plus en plus étroits tout en préservant une surface totale égale à l'unité. A la limite, comme le suggère la figure du bas du Graphique 4.6, si la population était infinie et que la mesure de la taille était exacte, on obtiendrait une courbe continue, avec une surface sous la courbe égale à 1. Un densité de probabilité est tout simplement l'analogue — comme dans le cas discret, au travers de l'épreuve aléatoire "tirer au hasard un individu dans la population et observer la valeur de la variable x pour l'individu tiré" — en termes de probabilité de cette courbe limite "définie" en termes de fréquence.

Dans la réalité, bien entendu, on a ni population infinie et ni mesure exacte. Dès lors, dire qu'une variable aléatoire est continue constitue au mieux, une idéalisation, une approximation de la réalité. Il se fait cependant qu'en pratique, c'est une approximation très commode. C'est dans cette perspective qu'il convient d'appréhender la notion de variable aléatoire continue, de même que certaines de ses caractéristiques

— en particulier la Propriété 4.2 établissant que toute probabilité ponctuelle est nulle — qui peuvent paraître assez peu intuitive : il ne faut pas chercher à trouver quelque chose de “naturel” dans le concept de continuité d'une variable aléatoire.

4.2.2. Fonction de répartition d'une variable aléatoire continue

La définition de la fonction de répartition d'une variable aléatoire continue est semblable à celle du cas discret, sauf qu'elle prend ici la forme d'une intégrale.

Définition 4.9 *On appelle fonction de répartition d'une variable aléatoire continue X la fonction $F(x)$ définie par :*

$$F(x) = I\!P(X \leq x) = I\!P(-\infty \leq X \leq x) = \int_{-\infty}^{x} f(t)dt$$

La fonction de répartition $F(x)$ est la primitive de la densité $f(x)$. Autrement dit, $f(x)$ est la dérivée de $F(x)$.

Comme dans le cas discret, la fonction de répartition $F(x)$ d'une variable aléatoire continue est une fonction croissante et varie de 0 à 1. Au contraire du cas discret, elle est ici continue.

La notion de fonction de répartition peut sembler quelque peu difficile et inutile. En fait, on l'utilise assez peu dans le cadre de problèmes faisant intervenir des lois discrètes. Par contre, la propriété suivante montre l'intérêt de la fonction de répartition lorsqu'on se trouve dans un contexte de loi continue : son utilisation permet d'éviter le calcul d'intégrales.

Propriété 4.4 *Soit X une variable aléatoire continue de fonction de répartition $F(x)$. On a :*

$$I\!P(a \leq X \leq b) = F(b) - F(a)$$

En effet, on a :

$$I\!P(a \leq X \leq b) = I\!P(a < X \leq b) = I\!P(X \leq b) - I\!P(X \leq a) = F(b) - F(a)$$

Pour les lois les plus courantes, on dispose de tables donnant des valeurs de la fonction de répartition. On peut lire dans les tables les valeurs $F(a)$, $F(b)$ et obtenir, par simple soustraction, la probabilité recherchée $I\!P(a \leq X \leq b)$.

Illustrons ce qui précède en revenant à notre exemple du bowling. Supposons que la personne soit suffisamment maladroite pour qu'il soit raisonnable de considérer qu'elle a autant de chances d'obtenir n'importe quelle valeur de X dans l'ensemble $\mathcal{X} = [0, 2]$.

Pour rendre compte de l'aspect équiprobable du continuum de valeurs possibles, on utilise une fonction de densité *constante* :

$$f(x) = \begin{cases} k & \text{pour } 0 \leq x \leq 2 \\ 0 & \text{partout ailleurs} \end{cases}$$

La valeur de la constante k est déterminée par la contrainte $\int_{-\infty}^{+\infty} f(x)dx = 1$ que doit satisfaire toute fonction de densité. On doit avoir :

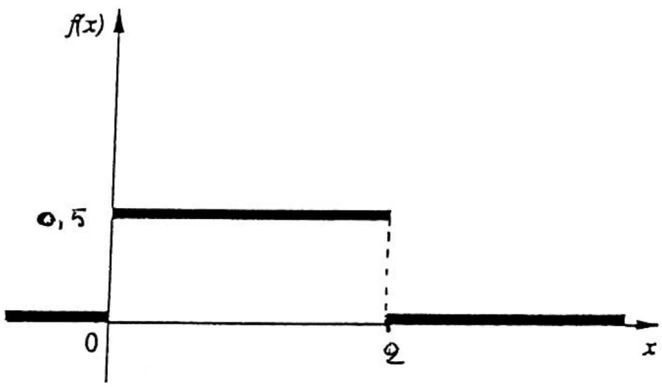
$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^2 kdx = [kx]_0^2 = 2k = 1$$

soit :

$$k = \frac{1}{2} = 0,5$$

La densité $f(x)$ est donc égale à :

$$f(x) = \begin{cases} 0,5 & \text{pour } 0 \leq x \leq 2 \\ 0 & \text{partout ailleurs} \end{cases}$$



Graphique 4.7 : La fonction de densité $f(x)$

La fonction de répartition est donnée par :

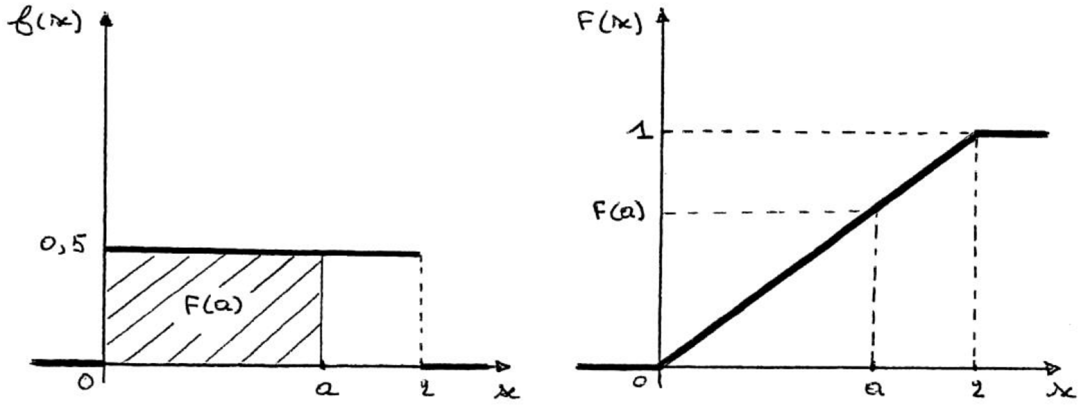
$$F(x) = I\!P(X \leq x) = \begin{cases} 0 & \text{pour } x < 0 \\ \int_0^x f(t)dt & \text{pour } 0 \leq x \leq 2 \\ 1 & \text{pour } x > 2 \end{cases}$$

On a :

$$\int_0^x f(t)dt = \int_0^x 0,5dt = [0,5t]_0^x = 0,5x$$

Donc :

$$F(x) = \begin{cases} 0 & \text{pour } x < 0 \\ 0,5x & \text{pour } 0 \leq x \leq 2 \\ 1 & \text{pour } x > 2 \end{cases}$$



Graphique 4.8 : La fonction de répartition $F(x)$

On peut utiliser la fonction de répartition ainsi trouvée pour calculer la probabilité que X prenne sa valeur dans l'un ou l'autre intervalle. Ainsi par exemple, on a :

$$I\!P(0,2 \leq X \leq 1,3) = F(1,3) - F(0,2) = 0,5 \times 1,3 - 0,5 \times 0,2 = 0,55$$

ou encore :

$$I\!P(X \geq 0,7) = 1 - I\!P(X \leq 0,7) = 1 - F(0,7) = 1 - (0,5 \times 0,7) = 1 - 0,35 = 0,65$$

4.2.3. Espérance et variance d'une variable aléatoire continue

Définition 4.10 *On appelle espérance d'une variable aléatoire continue X la quantité :*

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Cette formule de l'espérance est la simple transposition au cas continu de la formule qui s'applique aux variables aléatoires discrètes :

$$E(X) = \sum_{x_i} x_i p_{x_i}$$

La somme est simplement remplacée par une intégrale et les valeurs prises par la variable sont pondérées par la densité au lieu des probabilités ponctuelles.

La définition de la variance d'une variable aléatoire continue correspond à une même transposition au cas continu de la formule qui s'applique aux variables aléatoires discrètes.

Définition 4.11 *On appelle variance d'une variable aléatoire continue X la quantité :*

$$V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx, \quad \text{où } \mu = E(X)$$

et écart-type de X la quantité :

$$\sigma(X) = \sqrt{V(X)}$$

L'interprétation de $E(X)$, $V(X)$ et $\sigma(X)$ d'une variable aléatoire continue X est la même — à la transposition discret / continu près — que dans la cas discret.

Comme dans le cas discret, la variance peut également s'exprimer sous une forme équivalente, plus simple pour les calculs. On a en effet :

$$\begin{aligned}
V(X) &= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx \qquad\qquad (\text{où } \mu = E(X)) \\
&= \int_{-\infty}^{+\infty} (x^2 - 2\mu x + \mu^2) f(x)dx \\
&= \int_{-\infty}^{+\infty} x^2 f(x)dx - 2\mu \underbrace{\int_{-\infty}^{+\infty} x f(x)dx}_{=\mu} + \mu^2 \underbrace{\int_{-\infty}^{+\infty} f(x)dx}_{=1} \\
&= \int_{-\infty}^{+\infty} x^2 f(x)dx - 2\mu^2 + \mu^2 \\
&= \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2
\end{aligned}$$

Propriété 4.5 *Soit X une variable aléatoire continue. On a :*

$$V(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2, \quad \text{où } \mu = E(X)$$

Illustrons le calcul de l'espérance, de la variance et de l'écart-type d'une variable aléatoire continue dans notre exemple du bowling où la fonction de densité est donnée par :

$$f(x) = \begin{cases} 0,5 & \text{pour } 0 \le x \le 2 \\ 0 & \text{partout ailleurs} \end{cases}$$

On trouve :

$$\begin{aligned}E(X) &= \int_{-\infty}^{+\infty} xf(x)dx = \int_0^2 0,5xdx = \left[0,5\frac{x^2}{2}\right]_0^2 \\ &= 0,5\times\frac{2^2}{2} - 0,5\times 0 = 1\end{aligned}$$

On a d'autre part (aux arrondis près) :

$$\begin{aligned}V(X) &= \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2, \quad \text{où } \mu = E(X) \\ &= \int_0^2 0,5x^2dx - 1^2 = \left[0,5\frac{x^3}{3}\right]_0^2 - 1 \\ &= \left[0,5\times\frac{2^3}{3} - 0,5\times 0\right] - 1 = 0,333\end{aligned}$$

et donc :

$$\sigma(X) = \sqrt{V(X)} = \sqrt{0,333} = 0,577$$

4.3. Fonction d'une variable aléatoire

Si X est une variable aléatoire (discrète ou continue) et $g(X)$ une fonction de X — une application $g(.)$ qui a toute valeur x associe *une et au plus une* valeur $y = g(x)$ —, alors $Y = g(X)$ est encore une variable aléatoire. Par exemple, $Y = a + bX$, $Y = X^2$, $Y = \ln X$, ... sont des variables aléatoires.

4.3.1. Loi d'une fonction d'une variable aléatoire

Connaissant la loi de X, il est en général aisé de déterminer la loi de la variable aléatoire Y définie par $Y = g(X)$, a tout le moins dans le cas discret.

4.3.1.1. Cas discret

Comme la loi de toute variable aléatoire discrète, la loi de la variable aléatoire $Y = g(X)$ est donnée par un ensemble de couples $\left(y_j, p_{y_j}\right)$, où les y_j désignent les valeurs possibles de Y — notons l'ensemble de ces valeurs possibles $\mathcal{Y}$ — et les $p_{y_j} = \mathbb{P}(Y = y_j)$ les probabilités de ces valeurs possibles.

Etant donné la loi de X et la fonction $g(X)$, on peut aisément trouver les valeurs possibles y_j de Y : il s'agit de toutes les valeurs distinctes $y_j = g(x_i)$ obtenues en passant en revue les différentes valeurs possibles x_i de X.

Les probabilités $p_{y_j} = I\!P(Y = y_j)$ s'obtiennent tout aussi aisément : elles sont tout simplement égales aux probabilités des événements "$Y = y_j$" dans l'épreuve aléatoire décrite par la variable aléatoire X, c'est-à-dire pour chaque valeur possible y_j de Y égale à la somme des probabilités des valeurs possibles x_i de X pour lesquelles l'événement "$Y = y_j$" se réalise :

$$p_{y_j} = I\!P(Y = y_j) = \sum_{x_i : g(x_i) = y_j} p_{x_i} \tag{4.3}$$

où $\sum_{x_i : g(x_i) = y_j}$ signifie "somme pour tous les valeurs possibles x_i de X telles que $g(x_i) = y_j$".

Comme par définition l'application $g(.)$ associe a tout x_i une et au plus une valeur y_j, les différents événements "$Y = y_j$" forment bien une partition dans l'épreuve aléatoire décrite par X, ce qui assure, comme il se doit, que $\sum_{y_j} p_{y_j} = 1$.

Reprenons notre jeu du début dont la loi de X (= gains associés au jeu) est donnée par :

x_i	-50	0	20	100
p_{x_i}	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

Supposons qu'on modifie les règles du jeu et que la personne souhaitant participer soit tenue de payer un droit d'entrée de 3 F. Supposons également que toutes les transactions financières ont été exprimées en francs français et qu'on veut les réexprimer en francs belges (1 FF = 6 FB). Le gain net du joueur est alors :

$$Y = (X - 3) \times 6 = -18 + 6X$$

Etablissons la loi de la variable aléatoire Y. Ses valeurs possibles sont :

$$\begin{array}{ll} -18 + 6 \times (-50) = -318 & \quad -18 + 6 \times 0 = -18 \\ -18 + 6 \times 20 = 102 & \quad -18 + 6 \times 100 = 582 \end{array}$$

c'est-à-dire :

$$\mathcal{Y} = \{-318, -18, 102, 582\}$$

Calculons à présent les probabilités. On a :

$$\begin{array}{ll} I\!P(Y = -318) = I\!P(X = -50) = \frac{1}{3} & \quad I\!P(Y = -18) = I\!P(X = 0) = \frac{1}{3} \\ I\!P(Y = 102) = I\!P(X = 20) = \frac{1}{6} & \quad I\!P(Y = 582) = I\!P(X = 100) = \frac{1}{6} \end{array}$$

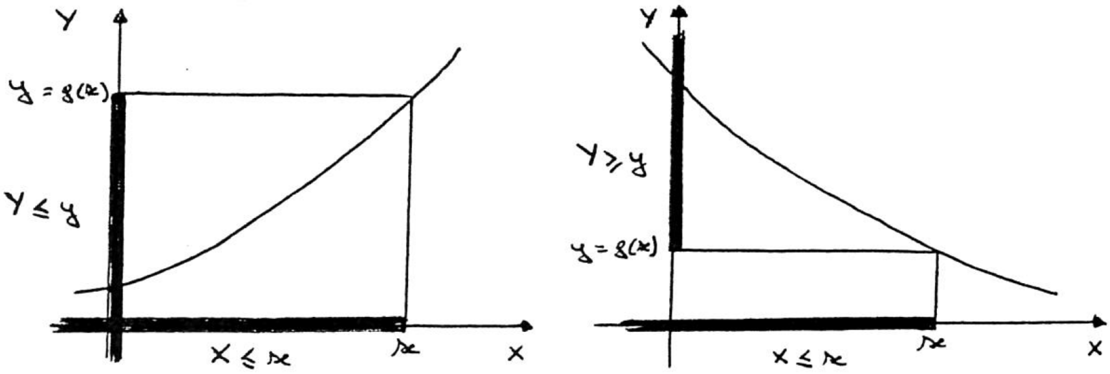
La loi de Y est donc :

y_j	-318	-18	102	582
p_{y_j}	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

On vérifie que la somme des probabilités est bien égale à 1.

4.3.1.2. Cas continu

Le cas continu est plus difficile. Nous n'envisagerons ici que le cas où la fonction $g(X)$ est strictement croissante ou décroissante, et différentiable. Dans ce cas, la fonction $g(X)$ est inversible. On notera cette fonction inverse $g^{-1}(Y)$.



$g(X)$ est strictement croissante — $g(X)$ est strictement décroissante

Graphique 4.9 : Loi d'une fonction d'une variable aléatoire continue

Désignons les fonctions de densité et de répartition de X par respectivement $f(x)$ et $F(x)$, celles de $Y = g(X)$ par $h(y)$ et $H(y)$, et supposons tout d'abord que la fonction $g(X)$ est strictement croissante. Comme le montre le Graphique 4.9, dans ce cas, on a :

$$F(x) = I\!P(X \leq x) = I\!P(g(X) \leq g(x)) = I\!P(Y \leq y) = H\,(y)$$

soit :

$$F(g^{-1}(y)) = H\,(y)$$

En dérivant par rapport à y, on obtient :

$$\frac{dF(g^{-1}(y))}{dy} = \frac{dF\,(x)}{dx}\frac{dg^{-1}(y)}{dy} = f(g^{-1}(y))\frac{dg^{-1}(y)}{dy} = \frac{dH(y)}{dy} = h(y)$$

soit :

$$h(y) = \frac{dg^{-1}(y)}{dy} f(g^{-1}(y)) \tag{4.4}$$

Supposons maintenant que la fonction $g(X)$ est strictement décroissante. Comme le montre également le Graphique 4.9, dans ce cas, on a :

$$F(x) = I\!P(X \leq x) = I\!P(g(X) \geq g(x)) = 1 - I\!P(Y \leq y) = 1 - H\,(y)$$

soit :

$$F(g^{-1}(y)) = 1 - H\,(y)$$

De la même façon que ci-dessus, en dérivant par rapport à y, on obtient :

$$f(g^{-1}(y))\frac{dg^{-1}(y)}{dy} = -h(y)$$

soit :

$$h(y) = -\frac{dg^{-1}(y)}{dy}f(g^{-1}(y)) = \left|\frac{dg^{-1}(y)}{dy}\right| f(g^{-1}(y)) \tag{4.5}$$

puisque lorsque la fonction $g(X)$ est strictement décroissante, $\frac{dg^{-1}(y)}{dy} < 0$.

La formule (4.5) peut être appliquée tant au cas où la fonction $g(X)$ est décroissante que croissante, puisque dans ce dernier cas, on a de toute façon $\frac{dg^{-1}(y)}{dy} > 0$. On en déduit la propriété suivante.

Propriété 4.6 *Soit X une variable aléatoire continue de densité de probabilité $f(x)$. Si $Y = g(X)$, où $g(X)$ est une fonction strictement croissante ou décroissante et différentiable de X, la densité de probabilité $h(y)$ de Y est donnée par :*

$$h(y) = \left|\frac{dg^{-1}(y)}{dy}\right| f(g^{-1}(y))$$

Illustrons cette propriété dans notre exemple du bowling où la fonction de densité de X est donnée par :

$$f(x) = \begin{cases} 0,5 & \text{pour } 0 \leq x \leq 2 \\ 0 & \text{partout ailleurs} \end{cases}$$

Considérons la variable aléatoire Y définie par $Y = g(X) = X^3$. La fonction $g(X) = X^3$ est strictement croissante (et différentiable) et sa fonction inverse est $g^{-1}(Y) = \sqrt[3]{Y} = Y^{\frac{1}{3}}$. L'intervalle des valeurs possibles de X étant $\mathcal{X} = [0,2]$, l'intervalle des valeurs possibles de Y est $\mathcal{Y} = [0, 2^3] = [0,8]$. On a par ailleurs :

$$\frac{dg^{-1}(y)}{dy} = \frac{1}{3}y^{-\frac{2}{3}} = \frac{1}{3\sqrt[3]{y^2}}$$

La fonction de densité $h(y)$ de la variable aléatoire Y est donc donnée par :

$$h(y) = \begin{cases} \frac{0{,}5}{3\sqrt[3]{y^2}} & \text{pour } 0 \leq y \leq 8 \\ 0 & \text{partout ailleurs} \end{cases}$$

On vérifie que $h(y)$ définit bien une fonction de densité, puisqu'on a bien $h(y) \geq 0$, $\forall\, y \in \mathbb{R}$, et :

$$\begin{aligned}\int_{-\infty}^{+\infty} h(y)dy &= \int_0^8 \frac{1}{6}y^{-\frac{2}{3}}dy = \left[\frac{1}{6}\frac{y^{\frac{1}{3}}}{\frac{1}{3}}\right]_0^8 \\ &= \left[\frac{1}{2}\sqrt[3]{y}\right]_0^8 = \frac{1}{2}\times\sqrt[3]{8} - 0 = 1\end{aligned}$$

4.3.2. Espérance et variance d'une fonction d'une variable aléatoire

La loi d'une variable aléatoire Y, discrète ou continue, définie par $Y = g(X)$ étant établie, on peut calculer son espérance et sa variance (et par conséquent son écart-type) de la façon habituelle, comme pour toute variable aléatoire, c'est-à-dire, dans le cas discret, par (cf. Définitions 4.5 et 4.6, et Propriété 4.1) :

$$\begin{aligned}E(Y) &= \mu_Y = \sum_{y_j} y_j p_{y_j} \\ V(Y) &= \sum_{y_j}(y_j - \mu_Y)^2 p_{y_j} = \sum_{y_j} y_j^2 p_{y_j} - \mu_Y^2\end{aligned}$$

et dans le cas continu, par (cf. Définitions 4.10 et 4.11, et Propriété 4.5) :

$$\begin{aligned}E(Y) &= \mu_Y = \int_{-\infty}^{+\infty} yh(y)dy \\ V(Y) &= \int_{-\infty}^{+\infty} (y-\mu_Y)^2 h(y)dy = \int_{-\infty}^{+\infty} y^2 h(y)dy - \mu_Y^2\end{aligned}$$

On peut également calculer ces quantités directement à partir de loi de X, sans avoir à établir la loi de Y, ce qui est souvent plus commode. Cette possibilité découle de la propriété générale suivante — que nous admettrons —, qui donne la formule permettant d'obtenir l'espérance de $Y = g(X)$ et dont, comme nous allons le voir, on peut déduire celle valable pour la variance de $Y = g(X)$.

Propriété 4.7 *Soit X une variable aléatoire. Si $Y = g(X)$, où $g(X)$ est une fonction quelconque de X, on a dans le cas discret :*

$$E(Y) = E\left[g(X)\right] = \sum_{x_i} g(x_i)p_{x_i} \tag{4.6}$$

et dans le cas continu :

$$E(Y) = E\left[g(X)\right] = \int_{-\infty}^{+\infty} g(x)f(x)dx \tag{4.7}$$

On note que pour la fonction $g(X) = (X-\mu)^2$, où $\mu = E(X)$, les formules (4.6) et (4.7) donnent respectivement :

$$E(Y) = E\left[(X-\mu)^2\right] = \sum_{x_i}(x_i-\mu)^2 p_{x_i}$$

et

$$E(Y) = E\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty}(x-\mu)^2 f(x)dx$$

c'est-à-dire les définitions de la variance de X dans, respectivement, le cas discret et le cas continu.

On voit ainsi que la variance $V(X)$ d'une variable aléatoire X, discrète ou continue, n'est rien d'autre que l'espérance d'une fonction particulière — en termes plus précis, d'une variable aléatoire définie par une fonction particulière, en l'occurrence $(X-E(X))^2$ — de X.

Propriété 4.8 *Quelle que soit la variable aléatoire X, on a :*

$$V(X) = E\left[(X-E(X))^2\right] = E(X^2) - (E(X))^2$$

On vérifie aisément la deuxième égalité de la propriété ci-dessus : elle découle des Propriétés 4.1 et 4.5, et des formules (4.6) et (4.7) pour la fonction $g(X) = X^2$.

De la Propriété 4.8, on a pour la variance de la variable aléatoire $Y = g(X)$:

$$\begin{aligned} V(Y) = V\left[g(X)\right] &= E\left[(Y-E(Y))^2\right] = E\left[(g(X) - E\left[g(X)\right])^2\right] \\ &= E(Y^2) - (E(Y))^2 = E\left[(g(X))^2\right] - (E\left[g(X)\right])^2 \end{aligned}$$

Par application de la Propriété 4.7, on en déduit des formules pour calculer la variance de $Y = g(X)$ sur base de la loi de X.

Propriété 4.9 *Soit X une variable aléatoire. Si $Y = g(X)$, où $g(X)$ est une fonction quelconque de X, on a dans le cas discret :*

$$V(Y) = V\left[g(X)\right] = \sum_{x_i}\left(g(x_i) - \mu_g\right)^2 p_{x_i} = \sum_{x_i}(g(x_i))^2 p_{x_i} - \mu_g^2$$

et dans le cas continu :

$$V(Y) = V\left[g(X)\right] = \int_{-\infty}^{+\infty}\left(g(x) - \mu_g\right)^2 f(x)dx = \int_{-\infty}^{+\infty}(g(x))^2 f(x)dx - \mu_g^2$$

où $\mu_g = E(Y) = E\left[g(X)\right]$.

L'espérance d'une fonction linéaire d'une variable aléatoire X, discrète ou continue, est particulièrement simple à calculer.

Propriété 4.10 *Soit X une variable aléatoire quelconque. Si $Y = a + bX$, on a :*

$$E(Y) = E(a + bX) = a + bE(X)$$

En particulier :

$$E(a) = a$$

En effet, de la Propriété 4.7, dans le cas discret, on a :

$$E(a + bX) = \sum_{x_i} (a + bx_i) p_{x_i} = a \underbrace{\sum_{x_i} p_{x_i}}_{=1} + b \underbrace{\sum_{x_i} x_i p_{x_i}}_{=E(X)} = a + bE(X)$$

De même, dans le cas continu, on a :

$$E(a + bX) = \int_{-\infty}^{+\infty} (a + bx) f(x)dx = a \underbrace{\int_{-\infty}^{+\infty} f(x)dx}_{=1} + b \underbrace{\int_{-\infty}^{+\infty} xf(x)dx}_{=E(X)} = a + bE(X)$$

Dans le cas d'une fonction linéaire, on a donc $E[g(X)] = g[E(X)]$. Attention : ceci n'est vrai que dans le cas d'une fonction linéaire. Pour une fonction quelconque, non linéaire, on a généralement $E[g(X)] \neq g[E(X)]$: par exemple, $E(X^2) \neq (E(X))^2$.

Comme l'espérance, la variance (et par conséquent l'écart-type) d'une fonction linéaire d'une variable aléatoire X, discrète ou continue, peut être obtenue de façon assez simple.

Propriété 4.11 *Soit X une variable aléatoire quelconque. Si $Y = a + bX$, on a :*

$$V(Y) = V(a + bX) = b^2 V(X)$$

et

$$\sigma(a + bX) = \sqrt{V(a + bX)} = |b|\sigma(X)$$

En particulier :

$$V(a) = \sigma(a) = 0$$

En effet, de la Propriété 4.8, on a $V(a + bX) = E\left[(a + bX - E[a + bX])^2\right]$. Posons $Z = a + bX - E(a + bX)$. De la Propriété 4.10, on a :

$$Z = a + bX - E(a + bX) = a + bX - (a + bE(X)) = b(X - E(X))$$

de sorte que :

$$V(a + bX) = E\left[(a + bX - E(a + bX))^2\right] = E\left(Z^2\right) = E\left[b^2 (X - E(X))^2\right]$$

soit, en utilisant encore les Propriétés 4.10 et 4.8 :

$$V(a+bX) = b^2 E\left[(X - E(X))^2\right] = b^2 V(X)$$

dont on déduit :

$$\sigma(a+bX) = \sqrt{V(a+bX)} = \sqrt{b^2 V(X)} = |b|\sigma(X)$$

Illustrons ces propriétés à l'aide de notre exemple de jeu de dé. On a déjà calculé que, pour la variable aléatoire X, on a (aux arrondis près) $E(X) = 3,333$, $V(X) = 2555,555$ et $\sigma(X) = 50,552$.

Pour la variable aléatoire $Y = -18 + 6X$, on a donc (aux arrondis près) :

$$\begin{aligned}
E(Y) &= E(-18+6X) = -18 + 6E(X) = -18 + 20 = 2 \\
V(Y) &= V(-18+6X) = 6^2 V(X) = 36 \times 2555,555 = 92\,000 \\
\sigma(Y) &= \sigma(-18+6X) = 6\sigma(X) = 6 \times 50,552 = 303,32
\end{aligned}$$

On vérifie qu'en utilisant la loi de Y qui est donnée par :

y_j	-318	-18	102	582
p_{y_j}	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

on trouve de même :

$$\begin{aligned}
E(Y) &= \sum_{y_j} y_j p_{y_j} = -318 \times \frac{1}{3} - 18 \times \frac{1}{3} + 102 \times \frac{1}{6} + 582 \times \frac{1}{6} = 2 \\
V(Y) &= E(Y^2) - (E(Y))^2 = \sum_{y_j} y_j^2 p_{y_j} - (E(Y))^2 \\
&= (-318)^2 \times \frac{1}{3} + (-18)^2 \times \frac{1}{3} + 102^2 \times \frac{1}{6} + 582^2 \times \frac{1}{6} - 2^2 \\
&= 33\,708 + 108 + 1\,734 + 56\,454 - 4 = 92\,000 \\
\sigma(Y) &= \sqrt{V(Y)} = \sqrt{92\,000} = 303,32
\end{aligned}$$

4.3.3. Standardisation d'une variable aléatoire

On peut toujours s'arranger pour définir une variable aléatoire d'espérance 0 et de variance 1 à partir de n'importe quelle variable aléatoire X. Il suffit de *centrer* la variable, c'est-à-dire de lui soustraire son espérance, et de la *réduire*, c'est-à-dire de la diviser par son écart-type. Autrement dit, il suffit de considérer la variable aléatoire Y définie par :

$$Y = \frac{X - E(X)}{\sigma(X)}$$

En effet, en utilisant les Propriétés 4.10 et 4.11, on vérifie qu'on a bien :

$$E(Y) = E\left(\frac{X - E(X)}{\sigma(X)}\right) = \frac{1}{\sigma(X)} E(X - E(X)) = \frac{1}{\sigma(X)}\left(E(X) - E(X)\right) = 0$$

et

$$V(Y) = V\left(\frac{X - E(X)}{\sigma(X)}\right) = \frac{1}{(\sigma(X))^2} V(X - E(X)) = \frac{1}{V(X)} V(X) = 1$$

On dit que Y est une variable aléatoire *centrée réduite*, ou encore *standardisée*.

4.3.4. Exercice résolu

On considère une variable aléatoire X dont la loi est donnée par :

x_i	-10	-5	0	5	10
p_{x_i}	$0,125$	$0,25$	$0,125$	$0,375$	$0,125$

1- Etablissez la loi de $Y = X^2$.

Les valeurs possibles y_j de Y sont :

$$\mathcal{Y} = \{0, 25, 100\}$$

On trouve :

$$\begin{aligned}
I\!P(Y = 0) &= I\!P(X^2 = 0) = I\!P(X = 0) = 0,125 \\
I\!P(Y = 25) &= I\!P(X^2 = 25) = I\!P(X = 5 \text{ ou } X = -5) \\
&= I\!P(X = 5) + I\!P(X = -5) = 0,375 + 0,25 = 0,625 \\
I\!P(Y = 100) &= I\!P(X^2 = 100) = I\!P(X = 10 \text{ ou } X = -10) \\
&= I\!P(X = 10) + I\!P(X = -10) = 0,125 + 0,125 = 0,25
\end{aligned}$$

Au total, on a donc :

y_j	0	25	100
p_{y_j}	$0,125$	$0,625$	$0,25$

2- Calculez $\sigma(1 + 2X^2)$.

On a :

$$V(1 + 2X^2) = V(1 + 2Y) = 2^2 V(Y) = 4V(Y)$$

De la loi de Y, on obtient :

$$\begin{aligned}
E(Y) &= \sum_{y_j} y_j p_{y_j} = 0 \times 0,125 + 25 \times 0,625 + 100 \times 0,25 \\
&= 15,625 + 25 = 40,625
\end{aligned}$$

et

$$\begin{aligned} E(Y^2) &= \sum_{y_j} y_j^2 p_{y_j} = 0^2 \times 0,125 + 25^2 \times 0,625 + 100^2 \times 0,25 \\ &= 390,625 + 2500 = 2\,890,625 \end{aligned}$$

de sorte que :

$$V(Y) = E(Y^2) - (E(Y))^2 = 2\,890,625 - 40,625 = 2\,850$$

Au total, on a donc :

$$\sigma(1 + 2X^2) = \sqrt{V(1 + 2X^2)} = 2\sqrt{V(Y)} = 2 \times \sqrt{2\,850} = 106,770...$$

3- Déterminez la loi de la version standardisée de Y.

La version standardisée de Y est la variable aléatoire :

$$Z = \frac{Y - E(Y)}{\sigma(Y)}$$

On a $E(Y) = 40,625$ et $\sigma(Y) = \sqrt{V(Y)} = \sqrt{2\,850} = 53,385$ (à l'arrondi près). Dès lors :

$$Z = \frac{Y - 40,625}{53,385}$$

Les valeurs possibles de Z sont (aux arrondis près) :

$$\mathcal{Z} = \{-0,760, -0,292, 1,112\}$$

On vérifie facilement que la loi de Z est donnée par :

z_k	$-0,760$	$-0,292$	$1,112$
p_{z_k}	$0,125$	$0,625$	$0,25$

4.4. Exercices

1- Soit une variable aléatoire X dont la loi est :

x_i	0	1	2	3	4	5
p_{x_i}	0,10	0,30	0,40	0,10	0,05	0,05

a- Représentez graphiquement la distribution de probabilité et la fonction de répartition de X.

b- Calculez :

i- $I\!P(X = 2,5)$.
ii- $I\!P(X < 4,5)$.
iii- $I\!P(X > 2)$.
iv- $I\!P(2 < X < 4,5)$.
v- $I\!P(2 \leq X < 4,5)$.

vi- $I\!P(X < 4,5|X > 1,5)$.
vii- $I\!P(X \leq 3|2 \leq X \leq 4)$.
viii- $I\!P(|X - 2,5| > 1)$.
ix- $E(X)$, $V(X)$ et $\sigma(X)$.

2- Deux jeux de hasard sont caractérisés par les distributions de gains suivantes :

Gains du jeu 1	20	50	80	120
Probabilités	$0,2$	$0,4$	$0,3$	$0,1$

Gains du jeu 2	28	70	125
Probabilités	$0,5$	$0,3$	$0,2$

a- Représentez graphiquement la distribution de probabilité et la fonction de répartition des gains de chacun de ces deux jeux.
b- On vous propose de participer (gratuitement) à un des deux jeux. Lequel choisirez-vous ? Pourquoi ?
c- Accepterez-vous de participer si pour cela il vous faut payer un droit d'entrée de 68 F ? Pourquoi ?

3- Un agriculteur produit des pommes de terre. Il en obtient de 3 calibres différents A, B et C selon la distribution de probabilité suivante :

Calibres	A	B	C
Probabilités	$0,3$	$0,5$	$0,2$

Il peut vendre les pommes de terre du calibre A au prix de $2,2$ F le kilo, celles du calibre B au prix de 2 F le kilo et celles du calibre C au prix de $1,8$ F le kilo.
a- Quelle est la loi du revenu du producteur par kilo produit.
b- Quelle est l'espérance du revenu pour 40 000 kilos de pommes de terre produites ?
c- Supposons que les conditions climatiques changent et que la loi du calibre devient :

Calibres	A	B	C
Probabilités	$0,4$	$0,6$	0

Supposons également que cela affecte la production qui passe à 38 000 kilos. Que devient l'espérance du revenu ?

4- Dans un restaurant, un aquarium contient 25 grosses truites, 30 moyennes et 50 petites. Les prix en sont respectivement de 60 F, 40 F et 30 F. Le patron vous propose de payer 45 F pour une truite à prendre au hasard dans l'aquarium. Ce prix paraît-il raisonnable ? Sinon, de quel autre prix pourriez-vous convenir ?

5- Le richissime M. Mockseller vient de mourir. Il laisse pour tout héritage trois collections. Sa collection de timbres comporte 50 000 lots équivalents d'une valeur totale de 5 millions de dollars, celle de tableaux est de 200 lots équivalents d'une valeur totale de 60 millions de dollars et celle de porcelaines comporte 3 000 lots équivalents pour un total de 6 millions de dollars. Vous avez droit, de par le testament, à un lot tiré au sort dans l'ensemble de l'héritage. Un héritier du défunt voudrait vous acheter ce droit avant le tirage. De quel prix pourriez-vous convenir ? Votre prix serait-il différent si le défunt avait spécifié qu'on tirerait d'abord au sort une des collections et ensuite un lot dans cette collection ?

6- Le producteur d'un très grand concert en plein air prévu pour le 16 juin, estime que l'affluence à ce concert est fonction du temps. Il a obtenu du service météo local des statistiques concernant le temps au mois de juin durant les 10 années précédentes. Ces chiffres sont donnés dans le tableau ci-dessous :

Temps	Fréquence	Affluence prévue
Humide, froid	$0,20$	5 000
Humide, chaud	$0,20$	20 000
Sec, froid	$0,10$	30 000
Sec, chaud	$0,50$	50 000

a- Quelle est l'affluence espérée ?

b- Les tickets seront vendus 9 \$ pièce. Les coûts sont de 2 \$ par personne pour le nettoyage et le contrôle des billets, plus 150 000 \$ pour l'orchestre, plus 60 000 \$ pour l'administration (y compris les installations). Le producteur a-t-il ou non intérêt à organiser ce concert ? Pourquoi ?

7- Le producteur du problème précédent a lancé son projet de concert et il a obtenu le 10 juin des prévisions météorologiques relativement plus pessimistes. Les 4 situations atmosphériques ont, suivant ces prévisions, maintenant des probabilités respectivement égales à $0,30$, $0,20$, $0,20$ et $0,30$. S'il annule son concert, il lui faudra de toute façon payer 60 000 \$ de frais administratifs, plus 15 000 \$ de dédit aux musiciens. A votre avis, devrait-il ou non annuler le concert ?

8- Monsieur Legris veut regagner son domicile après une soirée animée chez des amis. Il a garé sa Sultane dans une rue si sombre qu'il ne parvient pas à reconnaître la bonne clé parmi son trousseau. Celui-ci comporte quatre clés différentes et Monsieur Legris décide de les essayer méthodiquement, tour à tour, jusqu'au moment où il parvient à ouvrir la voiture. On appelle X la variable aléatoire donnant le nombre d'essais réalisés par Monsieur Legris pour parvenir à ouvrir la portière de sa voiture.

a- Quelle est la loi de X ?

b- Quelle est la probabilité qu'il parvienne à ouvrir la portière en deux essais maximum ?

c- Donnez une prévision du nombre d'essais nécessaires pour ouvrir la portière.

d- On suppose à présent que Monsieur Legris n'a pas ses clés sur un trousseau mais que celles-ci sont dans sa poche, non reliées les unes aux autres. On suppose également que la soirée fut tellement animée que Monsieur Legris essaye ses clés totalement au hasard : il en prend une au hasard dans sa poche et, si elle ne convient pas, il la remet en poche, en prend à nouveau une au hasard dans sa poche, et ainsi de suite. Que deviennent dans ces conditions :

i- la loi de X (donnez une formule générale permettant de calculer $\mathbb{P}[X = x]$, où x est un nombre entier quelconque) ?

ii- la probabilité qu'il parvienne à ouvrir la portière en deux essais maximum ?

9- Une compagnie d'assurance propose une police qui couvre un sinistre dont la gravité est classée selon une échelle à 4 niveaux A, B, C et D. Lorsque le sinistre est de gravité A, il est remboursé 40 000 F. lorsqu'il est de gravité B, C et D, il est remboursé respectivement 60 000 F, 80 000 F et 100 000 F. Un client pris au hasard a, au cours d'une année, une probabilité de $0,007$ de subir un sinistre de gravité A, une probabilité de $0,004$ de subir un sinistre de gravité B, une probabilité de $0,002$ de subir un sinistre de gravité C et une probabilité de $0,001$ de subir un sinistre de gravité D. La compagnie rembourse au plus un sinistre par an et par client.

a- Quelle est la loi du montant que la compagnie aura à rembourser annuellement à un client pris au hasard ?

b- La compagnie propose sa police moyennant une prime annuelle de 530 F. L'organisme de contrôle des assurances lui suggère de revoir son tarif. Pourquoi ?

10- On considère le jeu de hasard suivant. Une urne contient 5 boules blanches et 3 boules noires. On tire au hasard 2 boules sans remise. Si les deux boules sont blanches, le joueur gagne 10 F. Si une boule seulement est blanche, le joueur gagne 3 F. Si les deux boules sont noires, le joueur ne gagne rien. On appelle X la variable aléatoire donnant le gain du joueur.

a- Quelle est la loi de X ?

b- Donnez son espérance, sa variance et son écart-type.

c- Déterminez et représentez graphiquement sa fonction de répartition.

11- Madame Gamble joue au jeu suivant. Elle tire une carte au hasard dans un jeu de 52 cartes. Si la carte choisie est un coeur, elle gagne 10 F. Si la carte est un trêfle, elle perd 10 F. Si la carte est un carreau, elle joue à pile ou face (la pièce est supposée équilibrée). Si le résultat est pile elle gagne 10 F, sinon elle perd 10 F. Dans tous les autres cas, elle ne prend ni ne gagne rien.

a- Quelle est la loi du gain de ce jeu ?

b- Quelle est la somme que Madame Gamble peut s'attendre à gagner (ou à perdre) ?

c- On définit un nouveau jeu dont les gains sont égaux au carré des gains du jeu initial. Quel est la loi du gain dans ce nouveau jeu ?

12- La loi du nombre de clients ($= X$) attendant d'être servis au rayon boucherie d'un supermarché est donnée par :

x_i	0	1	2	3	4	5	6	7
p_{x_i}	$0,1$	$0,2$	$0,25$	$0,15$	$0,1$	$0,08$	$0,06$	$0,06$

Monsieur Legris s'occupe du service à ce rayon, mais lorsque le nombre de clients est supérieur ou égal à 3, il fait appel à Madame Dubois pour l'aider, et lorsque ce nombre est supérieur ou égal à 6, il appelle en plus Madame Dumur.

a- Représentez la loi de X par un diagramme en bâtons.

b- Calculez l'espérance, la variance et l'écart-type de X.

c- Quelle est la probabilité que seulement Madame Dubois intervienne ?

d- Madame Dubois et Madame Dumur ont été appelées toutes les deux au rayon. Quelle est la probabilité que le nombre de clients soit égal à 7 ?

e- On appelle Y le nombre de personnes servant au rayon boucherie.

i- Quelle est la loi de Y ?

ii- Calculez son espérance, sa variance et son écart-type.

13- La Syldavie est un pays à forte activité sismique. La loi de l'intensité des séismes (mesurée sur l'échelle de Richter) qui s'y produisent est donnée par le tableau suivant :

Intensité du séisme	1	2	3	4	5	6	7	8	9
Probabilité	$0,25$	$0,19$	$0,17$	$0,14$	$0,11$	$0,07$	$0,04$	$0,02$	$0,01$

a- Représentez graphiquement la fonction de répartition de l'intensité des séismes.

b- Donnez une prévision de l'intensité du prochain séisme.

c- Un plan catastrophe est déclenché lorsque l'intensité du séisme est supérieure ou égale à 5. Le plan catastrophe a été déclenché. Quelle est la probabilité que l'intensité du séisme soit supérieure à 7 ?

d- On évalue que le montant Y des dommages causés par un séisme d'intensité X est donné par $Y = X^2 - 1$. Déterminez la loi de Y.

e- Un fonctionnaire syldave a calculé l'espérance du montant des dommages causés par le prochain séisme par la formule $[E(X)]^2 - 1$. Son chef veut l'envoyer suivre un cours de statistique. Pourquoi ?

14- On considère une variable aléatoire continue X dont la fonction de densité est donnée par :

$$f(x) = \begin{cases} 2x & \text{pour } 0 \leq x \leq 1 \\ 0 & \text{partout ailleurs} \end{cases}$$

a- Vérifiez que $f(x)$ définit bien une densité de probabilité et représentez-la graphiquement.

b- Déterminez la fonction de répartition de X et représentez-la graphiquement.

c- Calculez :

i- $I\!P(X = 0,6)$.
ii- $I\!P(X > 1,2)$.
iii- $I\!P(0,1 < X < 1)$.
iv- $I\!P(0 \leq X < 0,9)$.
v- $I\!P(0,2 \leq X \leq 0,7)$.
vi- $I\!P(X \geq 0,5)$.
vii- $I\!P(X < 0,5 | X \geq 0,2)$.
viii- $I\!P(X \leq 0,6 | 0,1 < X \leq 0,7)$.
ix- $I\!P(|X - 0,4| > 0,2)$.
x- $E(X)$, $V(X)$ et $\sigma(X)$.

d- On considère la variable aléatoire Y définie par la fonction $Y = 1 - \frac{X}{2}$.

i- Déterminez la fonction de densité de Y, vérifiez qu'il s'agit bien d'une densité de probabilité et représentez-la graphiquement.
ii- Déterminez la fonction de répartition de Y et représentez-la graphiquement.
iii- Calculez de deux façons différentes (en utilisant vos réponses au point c.x. ci-dessus et par intégration) l'espérance, la variance et

l'écart-type de Y.

15- On considère une variable aléatoire continue X dont la fonction de densité est donnée par :

$$f(x) = \begin{cases} e^{-x} & \text{pour } x \geq 0 \\ 0 & \text{partout ailleurs} \end{cases}$$

a- Vérifiez que $f(x)$ définit bien une densité de probabilité et représentez-la graphiquement.
b- Déterminez la fonction de répartition de X et représentez-la graphiquement.
c- Calculez :
   i- $I\!P(X > 1)$.
   ii- $I\!P(0,4 \leq X \leq 2,1)$.
   iii- $I\!P(1 \leq X < 2|X > 0,5)$.
   iv- $I\!P(|X - 3| \leq 2)$.
d- On considère la variable aléatoire Y définie par la fonction $Y = e^X$.
   i- Déterminez la fonction de densité de Y, vérifiez qu'il s'agit bien d'une densité de probabilité et représentez-la graphiquement.
   ii- Déterminez la fonction de répartition de Y et représentez-la graphiquement.
   iii- Calculez l'espérance, la variance et l'écart-type de Y.

16- On considère une variable aléatoire continue X dont la fonction de densité est donnée par :

$$f(x) = \begin{cases} 1 + x & \text{pour } -1 \leq x < 0 \\ 1 - x & \text{pour } 0 \leq x \leq 1 \\ 0 & \text{partout ailleurs} \end{cases}$$

a- Vérifiez que $f(x)$ définit bien une densité de probabilité et représentez-la graphiquement.
b- Déterminez la fonction de répartition de X et représentez-la graphiquement.
c- Calculez :
   i- $I\!P(X \geq 0)$.
   ii- $I\!P(-1 < X < 1)$.
   iii- $I\!P(X \geq 0,52|X < -0,5)$.
   iv- $I\!P(-0,7 \leq X \leq 0,9|0,3 < X < 0,8)$.
   v- $E(X)$, $V(X)$ et $\sigma(X)$.

Chapitre 5

Couples et vecteurs de variables aléatoires

Définition 5.1 *On appelle couple de variables aléatoires (resp. vecteur de n variables aléatoires) un couple de variables (resp. un vecteur de n variables) qui prend différentes valeurs suivant une certaine distribution de probabilité.*

Un couple de variables aléatoires (X, Y) (resp. un vecteur de n variables aléatoires $(X_1, X_2, ..., X_n)$) décrit une épreuve aléatoire dont les résultats possibles sont les différentes valeurs que le couple de variables (X, Y) (resp. le vecteur de n variables $(X_1, X_2, ..., X_n)$) peut prendre et dont les chances de réalisation des résultats possibles sont dépeintes par la distribution de probabilité du couple (X, Y) (resp. du vecteur $(X_1, X_2, ..., X_n)$).

De nombreuses épreuves aléatoires peuvent — directement ou par l'intermédiaire d'un codage — être représentées à l'aide d'un couple ou plus généralement d'un vecteur de variables aléatoires. C'est en particulier le cas des épreuves aléatoires "composites", dont les différents aspects sont alors décrits par les variables aléatoires qui composent le couple ou le vecteur. Ainsi par exemple, on peut représenter le lancer de deux dés par un couple de variables aléatoires, chacune des deux variables du couple décrivant le résultat du lancer d'un des dés. De la même façon, on peut représenter n lancers successifs d'un dé par un vecteur de n variables aléatoires.

On concentrera notre étude sur l'examen du cas bivarié, c'est-à-dire d'un couple de variables aléatoires, en distinguant toujours les cas discret et continu. Nous n'évoquerons que brièvement, en fin de chapitre, le cas multivarié, c'est-à-dire d'un vecteur de n $(n > 2)$ variables aléatoires.

5.1. Couples de variables aléatoires discrètes

Définition 5.2 *Un couple de variables aléatoires discrètes (X, Y) est un couple de variables aléatoires qui prend un nombre fini, ou infini mais dénombrable, de couples de valeurs (x_i, y_j).*

Supposons par exemple qu'on s'intéresse à la fréquentation et aux ventes d'un magasin de téléviseurs au cours d'un après-midi quelconque. On peut représenter cette épreuve aléatoire par un couple de variables aléatoires discrètes (X,Y), dont les variables X et Y sont définies par :

$$\begin{aligned} X &= \text{le nombre de clients entrant dans le magasin durant un après-midi} \\ Y &= \text{le nombre de téléviseurs vendus pendant la même période} \end{aligned}$$

5.1.1. Loi jointe d'un couple de variables aléatoires discrètes

Définition 5.3 *On appelle loi ou distribution de probabilité jointe d'un couple de variables aléatoires discrètes* (X,Y) *un ensemble de couples* $\big((x_i,y_j),p_{x_iy_j}\big)$*, où les* $p_{x_iy_j} = I\!P(X = x_i \text{ et } Y = y_j)$ *sont des nombres réels tels que :*

$$\begin{cases} 0 \leq p_{x_iy_j} \leq 1, \quad \forall\,(x_i,y_j) \in \mathcal{X} \times \mathcal{Y} \\ \displaystyle\sum_{x_i}\sum_{y_j} p_{x_iy_j} = 1 \end{cases} \tag{5.1}$$

$\mathcal{X}$ *désignant l'ensemble des valeurs possibles* x_i *de* X*,* $\mathcal{Y}$ *l'ensemble des valeurs possibles* y_j *de* Y*,* $\sum_{x_i}$ *signifiant "somme pour toutes les valeurs possibles* x_i *de* X*" et* $\sum_{y_j}$ *"somme pour toutes les valeurs possibles* y_j *de* Y*".*

Tout ensemble de couples $\big((x_i,y_j),p_{x_iy_j}\big)$ satisfaisant les conditions (5.1) de la définition ci-dessus définit la loi jointe d'un couple de variables aléatoires discrètes (X,Y).

Le concept de loi jointe d'un couple de variables aléatoires discrètes n'est guère différent de la notion de distribution de probabilité définie jusqu'ici, que ce soit au Chapitre 1 ou au Chapitre 4 (pour le cas discret, évidemment). Simplement, les résultats possibles de l'épreuve aléatoire qu'elle décrit sont ici des *couples* (x_i,y_j) de *nombres réels*.

On représente généralement la loi jointe d'un couple de variables aléatoires discrètes (X,Y) à l'aide d'un tableau à double entrée reprenant en ligne les différentes valeurs possibles x_i de X, en colonne les différentes valeurs possibles y_j de Y, et dans le corps du tableau les probabilités $p_{x_iy_j}$ des couples de valeurs (x_i,y_j). Une représentation graphique par un diagramme en bâtons est toujours possible mais peu pratique puisqu'elle implique une figure à trois dimensions.

Reprenons notre exemple de magasin de téléviseurs. Si on suppose pour simplifier que $\mathcal{X} = \{0,1,2,3\}$ et $\mathcal{Y} = \{0,1,2\}$, la loi jointe de (X,Y) pourrait par exemple être donnée par le tableau suivant :

$x_i \backslash y_j$	0	1	2
0	0,62	0,00	0,00
1	0,12	0,07	0,01
2	0,04	0,07	0,02
3	0,01	0,02	0,02

De la lecture de ce tableau, on voit par exemple que $\mathbb{P}(X = 2 \text{ et } Y = 1) = 0,07$: la probabilité d'avoir 2 clients et de vendre 1 téléviseur durant un après-midi quelconque est de $0,07$. On note que la somme de toutes les probabilités $p_{x_i y_j}$ est bien égale à 1.

Un couple de variable aléatoire discrète (X, Y) et sa loi jointe associée étant définie, le calcul de la probabilité d'événements relatifs au couple (X, Y) s'effectue de la façon habituelle, de manière générique en faisant la somme des probabilités des différents couples de valeurs possibles (x_i, y_j) de (X, Y) pour lesquelles l'événement considéré se réalise. Ainsi, on a par exemple pour le cas de notre magasin de téléviseurs :

$$\begin{aligned}
\mathbb{P}(X \geq 2 \text{ et } Y < 1) &= \mathbb{P}[(X = 2 \text{ et } Y = 0) \text{ ou } (X = 3 \text{ et } Y = 0)] \\
&= \mathbb{P}(X = 2 \text{ et } Y = 0) + \mathbb{P}(X = 3 \text{ et } Y = 0) \\
&= 0,04 + 0,01 = 0,05
\end{aligned}$$

De façon semblable au cas univarié, la notion de distribution de probabilité jointe d'un couple de variables aléatoires discrètes (X, Y) est à rapprocher de la notion de *distribution de fréquence jointe* utilisée en statistique descriptive pour décrire, dans une population, les valeurs prises par un couple de variables (x, y) prenant un petit nombre de valeurs distinctes (x_i, y_j). Le lien entre les deux est toujours le même : il s'agit de l'épreuve aléatoire “tirer au hasard un individu dans la population et observer la valeur du couple de variables (x, y) pour l'individu tiré”.

5.1.2. Lois marginales d'un couple de variables aléatoires discrètes

De la loi jointe du couple (X,Y), on peut déduire les *lois marginales* de X et de Y.

La loi marginale de X (resp. de Y) est tout simplement la loi — un ensemble de couples valeur possible / probabilité — de la variable aléatoire X (resp. Y) considérée séparément, telle qu'elle découle de la loi jointe. On a (en désignant par r le nombre de valeurs possibles de Y) :

$$\begin{aligned}
& \mathbb{P}(X = x_i) \\
= \;& \mathbb{P}(X = x_i \text{ et “} Y \text{ est quelconque”}) \\
= \;& \mathbb{P}[X = x_i \text{ et } (Y = y_1 \text{ ou } Y = y_2 \text{ ou ... ou } Y = y_r)] \\
= \;& \mathbb{P}[(X = x_i \text{ et } Y = y_1) \text{ ou } (X = x_i \text{ et } Y = y_2) \text{ ou ... ou } (X = x_i \text{ et } Y = y_r)] \\
= \;& \mathbb{P}(X = x_i \text{ et } Y = y_1) + \mathbb{P}(X = x_i \text{ et } Y = y_2) + ... + \mathbb{P}(X = x_i \text{ et } Y = y_r) \\
= \;& \sum_{y_j} \mathbb{P}(X = x_i \text{ et } Y = y_j)
\end{aligned}$$

Un raisonnement symétrique s'applique bien entendu au calcul de $\mathbb{P}(Y = y_j)$.

On a ainsi la définition suivante.

Définition 5.4 *Soit un couple de variables aléatoires discrètes* (X, Y) *et sa loi ou distribution de probabilité jointe donnée par un ensemble de couples* $\big((x_i, y_j), p_{x_i y_j}\big)$*. On appelle loi ou distribution de probabilité marginale de* X *l'ensemble des couples* $(x_i, p_{x_i.})$ *où :*

$$p_{x_i.} = I\!P(X = x_i) = \sum_{y_j} I\!P(X = x_i \text{ et } Y = y_j) = \sum_{y_j} p_{x_i y_j}$$

De même, on appelle loi ou distribution de probabilité marginale de Y *l'ensemble des couples* $\big(y_j, p_{.y_j}\big)$ *où :*

$$p_{.y_j} = I\!P(Y = y_j) = \sum_{x_i} I\!P(X = x_i \text{ et } Y = y_j) = \sum_{x_i} p_{x_i y_j}$$

On vérifie aisément que l'ensemble des couples $(x_i, p_{x_i.})$ et l'ensemble des couples $\big(y_j, p_{.y_j}\big)$ définissent bien des distributions de probabilité. Par définition, on a en effet d'une part :

$$0 \leq p_{x_i.} \leq 1, \ \ \forall\, x_i \in \mathcal{X} \quad \text{et} \quad 0 \leq p_{.y_j} \leq 1, \ \ \forall\, y_j \in \mathcal{Y}$$

et d'autre part :

$$\begin{aligned} \sum_{x_i} p_{x_i.} &= \sum_{x_i} \left(\sum_{y_j} p_{x_i y_j} \right) = \sum_{x_i} \sum_{y_j} p_{x_i y_j} = 1 \\ \sum_{y_j} p_{.y_j} &= \sum_{y_j} \left(\sum_{x_i} p_{x_i y_j} \right) = \sum_{x_i} \sum_{y_j} p_{x_i y_j} = 1 \end{aligned}$$

En pratique, partant du tableau décrivant la loi jointe de (X, Y), les probabilités $p_{x_i.}$ sont obtenues en additionnant toutes les probabilités de la ligne correspondant à la valeur x_i et les probabilités $p_{.y_j}$ sont de même obtenues en additionnant toutes les probabilités de la colonne correspondant à la valeur y_j.

Voyons cela sur notre exemple du magasin de téléviseurs. On a :

$x_i \backslash y_j$	0	1	2	$p_{x_i.}$
0	0,62	0,00	0,00	0,62
1	0,12	0,07	0,01	0,20
2	0,04	0,07	0,02	0,13
3	0,01	0,02	0,02	0,05
$p_{.y_j}$	0,79	0,16	0,05	1

Les lois marginales de X et de Y sont donc :

x_i	0	1	2	3
$p_{x_i.}$	0,62	0,20	0,13	0,05

y_j	0	1	2
$p_{.y_j}$	0,79	0,16	0,05

Comme il se doit, la somme des probabilités $p_{x_i.}$ et la somme des probabilités $p_{.y_j}$ vaut bien 1.

Les lois marginales de X et de Y étant comme toute loi d'une variable aléatoire discrète, il va sans dire que toutes les notions (fonction de répartition, espérance, variance, etc...) discutées au Chapitre 4 s'appliquent sans modification à chacune d'elles. On peut en particulier en calculer de la façon habituelle l'espérance et la variance, et ainsi obtenir l'espérance et la variance des variables aléatoires X et Y considérées séparément, telles qu'elles découlent de leur loi jointe.

5.1.3. Lois conditionnelles d'un couple de variables aléatoires discrètes

De la loi jointe du couple (X,Y), on peut également déduire les *lois conditionnelles* de X sachant $Y = y_j$ et de Y sachant $X = x_i$.

La loi conditionnelle de X sachant $Y = y_j$ (resp. de Y sachant $X = x_i$) est tout simplement la loi de la variable aléatoire X (resp. Y) lorsque la variable aléatoire Y (resp. X) est égale à y_j (resp. x_i), telle qu'elle découle de la loi jointe.

Définition 5.5 *Soit un couple de variables aléatoires discrètes (X, Y) et sa loi ou distribution de probabilité jointe donnée par un ensemble de couples $\big((x_i, y_j), p_{x_i y_j}\big)$. On appelle loi ou distribution de probabilité conditionnelle de X sachant $Y = y_j$ l'ensemble des couples $\big(x_i, p_{x_i|y_j}\big)$ où :*

$$p_{x_i|y_j} = I\!P(X = x_i|Y = y_j) = \frac{I\!P(X = x_i \text{ et } Y = y_j)}{I\!P(Y = y_j)} = \frac{p_{x_i y_j}}{\sum\limits_{x_i} p_{x_i y_j}} = \frac{p_{x_i y_j}}{p_{.y_j}} \quad (\forall\, y_j \in \mathcal{Y})$$

De même, on appelle loi ou distribution de probabilité conditionnelle de Y sachant $X = x_i$ l'ensemble des couples $\big(y_j, p_{y_j|x_i}\big)$ où :

$$p_{y_j|x_i} = I\!P(Y = y_j|X = x_i) = \frac{I\!P(X = x_i \text{ et } Y = y_j)}{I\!P(X = x_i)} = \frac{p_{x_i y_j}}{\sum\limits_{y_j} p_{x_i y_j}} = \frac{p_{x_i y_j}}{p_{x_i.}} \quad (\forall\, x_i \in \mathcal{X})$$

Il y a autant de lois conditionnelles de X sachant $Y = y_j$ qu'il n'y a de valeurs possibles y_j de Y, et de même autant de lois conditionnelles de Y sachant $X = x_i$ qu'il n'y a de valeurs possibles x_i de X.

On vérifie facilement que l'ensemble des couples $\big(x_i, p_{x_i|y_j}\big)$ (quelle que soit la valeur possible y_j de Y considérée) et l'ensemble des couples $\big(y_j, p_{y_j|x_i}\big)$ (quelle que soit la valeur possible x_i de X considérée) définissent bien des distributions de probabilité. Par définition, on a effet d'une part :

$$0 \leq p_{x_i|y_j} \leq 1, \;\; \forall\, x_i \in \mathcal{X} \qquad \text{et} \qquad 0 \leq p_{y_j|x_i} \leq 1, \;\; \forall\, y_j \in \mathcal{Y}$$

et d'autre part :

$$\begin{aligned}\sum_{x_i} p_{x_i|y_j} &= \sum_{x_i}\left(\frac{p_{x_iy_j}}{p_{.y_j}}\right) = \frac{1}{p_{.y_j}}\sum_{x_i} p_{x_iy_j} = \frac{p_{.y_j}}{p_{.y_j}} = 1 \\ \sum_{y_j} p_{y_j|x_i} &= \sum_{y_j}\left(\frac{p_{x_iy_j}}{p_{x_i.}}\right) = \frac{1}{p_{x_i.}}\sum_{y_j} p_{x_iy_j} = \frac{p_{x_i.}}{p_{x_i.}} = 1\end{aligned}$$

On voit que les lois conditionnelles de X sachant $Y = y_j$ et de Y sachant $X = x_i$ sont comme toute loi — un ensemble de couples valeur possible / probabilité — d'une variable aléatoire discrète. La seule différence est qu'elles font référence à des probabilités conditionnelles plutôt qu'à des probabilités à priori.

Poursuivons notre exemple du magasin de téléviseurs et déterminons les lois conditionnelles du nombre de téléviseurs vendus sachant le nombre de clients qui se présentent au magasin. On a pour $X = 0$ (aucun client) :

y_j	0	1	2
$p_{y_j\|x_i=0}$	$\frac{0{,}62}{0{,}62} = 1$	$\frac{0}{0{,}62} = 0$	$\frac{0}{0{,}62} = 0$

pour $X = 1$ (1 client) :

y_j	0	1	2
$p_{y_j\|x_i=1}$	$\frac{0{,}12}{0{,}20} = 0,6$	$\frac{0{,}07}{0{,}20} = 0,35$	$\frac{0{,}01}{0{,}20} = 0,05$

pour $X = 2$ (deux clients) :

y_j	0	1	2
$p_{y_j\|x_i=2}$	$\frac{0{,}04}{0{,}13} = 0,31$	$\frac{0{,}07}{0{,}13} = 0,54$	$\frac{0{,}02}{0{,}13} = 0,15$

et $X = 3$ (trois clients) :

y_j	0	1	2
$p_{y_j\|x_i=3}$	$\frac{0{,}01}{0{,}05} = 0,2$	$\frac{0{,}02}{0{,}05} = 0,4$	$\frac{0{,}02}{0{,}05} = 0,4$

On note qu'à chaque fois la somme des probabilités est bien égale à 1.

Comme on peut le constater, les lois conditionnelles de Y sachant $X = x_i$, considérées pour les diverses valeurs possibles x_i de X, montrent de manière très parlante la façon dont les valeurs prises par Y sont liées à celles prises par X, c'est-à-dire la façon dont Y dépend de X. Pour voir la façon dont les valeurs prises par X sont liées à celles prises par Y, c'est-à-dire la façon dont X dépend de Y, c'est bien entendu les lois conditionnelles de X sachant $Y = y_j$, pour les diverses valeurs possibles y_j de Y, qu'il convient d'examiner.

Notons finalement que, par définition, les lois conditionnelles font un lien entre la loi jointe et les lois marginales.

Propriété 5.1 *Quelles que soient les variables aléatoires discrètes X et Y, on a* *pour toutes les valeurs possibles x_i de X et y_j de Y :*

$$p_{x_i y_j} = p_{x_i|y_j} p_{.y_j} \qquad et \qquad p_{x_i y_j} = p_{y_j|x_i} p_{x_i.}$$

En utilisant ces relations, on peut en particulier construire des lois jointes à partir de lois conditionnelles et de lois marginales.

5.1.3.1. Espérance conditionnelle et variance conditionnelle

On l'a vu, pour chaque valeur possible x_i de X, la loi conditionnelle de Y sachant $X = x_i$ (et, de façon symétrique, pour valeur possible y_j de Y, la loi conditionnelle de X sachant $Y = y_j$) est comme toute loi d'une variable aléatoire discrète. On peut donc, comme pour toute loi de ce type, "résumer" sa distribution par diverses quantités, en particulier son espérance et sa variance.

Définition 5.6 *On appelle espérance conditionnelle de Y sachant $X = x_i$ d'un couple de variables aléatoires discrètes (X, Y) la quantité :*

$$E(Y|X = x_i) = \sum_{y_j} y_j p_{y_j|x_i}$$

variance conditionnelle de Y sachant $X = x_i$ la quantité :

$$\begin{aligned} V(Y|X = x_i) &= \sum_{y_j} (y_j - \mu_{Y|x_i})^2 p_{y_j|x_i} = E\left[(Y - E(Y|X = x_i))^2 \,|X = x_i\right] \\ &= \sum_{y_j} y_j^2 p_{y_j|x_i} - \mu_{Y|x_i}^2 = E(Y^2|X = x_i) - (E(Y|X = x_i))^2 \end{aligned}$$

où $\mu_{Y|x_i} = E(Y|X = x_i)$, et écart-type conditionnel de Y sachant $X = x_i$ la quantité :

$$\sigma(Y|X = x_i) = \sqrt{V(Y|X = x_i)}$$

L'interprétation ces diverses quantités est la même que dans le cas non conditionnel (cf. Chapitre 4) : elles font simplement référence à une loi conditionnelle plutôt qu'à une loi non conditionnelle. Il en est de même de la preuve de l'équivalence des différentes expressions de $V(Y|X = x_i)$.

Illustrons le calcul de $E(Y|X = x_i)$, $V(Y|X = x_i)$ et $\sigma(Y|X = x_i)$ en poursuivant notre exemple du magasin de téléviseurs. Pour $X = 0$, on trouve :

$$\begin{aligned} E(Y|X = 0) &= \sum_{y_j} y_j p_{y_j|x_i=0} = 0 \times 1 + 1 \times 0 + 2 \times 0 = 0 \\ E(Y^2|X = 0) &= \sum_{y_j} y_j^2 p_{y_j|x_i=0} = 0^2 \times 1 + 1^2 \times 0 + 2^0 \times 0 = 0 \end{aligned}$$

de sorte que :

$$\begin{aligned} V(Y|X=0) &= E(Y^2|X=0) - (E(Y|X=0))^2 = 0 - 0^2 = 0 \\ \sigma(Y|X=0) &= \sqrt{V(Y|X=0)} = 0 \end{aligned}$$

ce qui reflète le fait que, lorsque $X = 0$, la valeur de Y est certaine et égale à 0.

De même, pour $X = 1$, on obtient :

$$\begin{aligned} E(Y|X=1) &= 0 \times 0,6 + 1 \times 0,35 + 2 \times 0,05 = 0,45 \\ E(Y^2|X=1) &= 0^2 \times 0,6 + 1^2 \times 0,35 + 2^2 \times 0,05 = 0,55 \end{aligned}$$

et donc :

$$\begin{aligned} V(Y|X=1) &= 0,55 - (0,45)^2 = 0,3475 \\ \sigma(Y|X=1) &= \sqrt{0,3475} = 0,5894... \end{aligned}$$

Pour $X = 2$, on obtient :

$$\begin{aligned} E(Y|X=2) &= 0 \times 0,31 + 1 \times 0,54 + 2 \times 0,15 = 0,84 \\ E(Y^2|X=2) &= 0^2 \times 0,31 + 1^2 \times 0,54 + 2^2 \times 0,15 = 1,14 \end{aligned}$$

et donc :

$$\begin{aligned} V(Y|X=2) &= 1,14 - (0,84)^2 = 0,4344 \\ \sigma(Y|X=2) &= \sqrt{0,4344} = 0,6590... \end{aligned}$$

Finalement, pour $X = 3$, on a :

$$\begin{aligned} E(Y|X=3) &= 0 \times 0,2 + 1 \times 0,4 + 2 \times 0,4 = 1,2 \\ E(Y^2|X=3) &= 0^2 \times 0,2 + 1^2 \times 0,4 + 2^2 \times 0,4 = 2 \end{aligned}$$

et donc :

$$\begin{aligned} V(Y|X=3) &= 2 - (1,2)^2 = 0,56 \\ \sigma(Y|X=3) &= \sqrt{0,56} = 0,7483... \end{aligned}$$

En résumé, on a donc :

x_i	0	1	2	3
$E(Y\|X=x_i)$	0	0,45	0,84	1,2
$V(Y\|X=x_i)$	0	0,34	0,43	0,56
$\sigma(Y\|X=x_i)$	0	0,58	0,65	0,74

On notera l'augmentation de $E(Y|X = x_i)$ et $V(Y|X = x_i)$ (ou $\sigma(Y|X = x_i)$) en fonction de x_i.

$E(Y|X=x_i)$ et $V(Y|X=x_i)$ sont des quantités très commodes pour résumer la façon dont Y dépend de X. Regardée comme une fonction de x_i, l'espérance conditionnelle de Y sachant $X=x_i$ est appelée *fonction d'espérance conditionnelle de Y sachant X* ou *courbe de régression de Y en X*, et est notée $E(Y|X)$. De façon semblable, regardée comme une fonction de x_i, la variance conditionnelle de Y sachant $X=x_i$ est appelée *fonction de variance conditionnelle de Y sachant X* et est notée $V(Y|X)$.

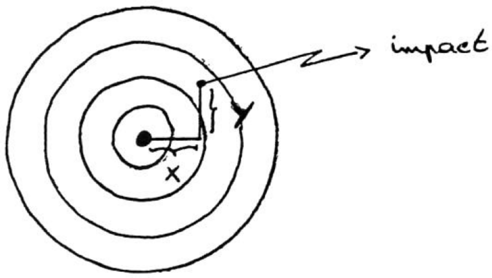
Il va de soi que toutes les considérations ci-dessus s'appliquent, de façon symétrique, aux lois conditionnelles de X sachant $Y=y_j$.

5.2. Couples de variables aléatoires continues

Définition 5.7 *Un couple de variables aléatoires continues (X,Y) est un couple de variables aléatoires qui prend un continuum de couples de valeurs (x,y) dans un (ou plusieurs) intervalle(s).*

Considérons par exemple un tireur à l'arc et supposons qu'on s'intéresse au point d'impact de la flèche. On peut représenter cette épreuve aléatoire par un couple de variables aléatoires (X,Y), où X mesure la distance horizontale et Y la distance verticale entre le point d'impact de la flèche et le centre de la cible.

On est dans le même type de situation que le cas du bowling évoqué au Chapitre 4 : le couple de variables aléatoires (X,Y) peut prendre un continuum de couples de valeurs (x,y) dans un intervalle donné. Dans notre cas, si l'archer ne manque jamais sa cible et que celle-ci mesure par exemple 100 cm de côté, on a $\mathcal{X}=[-50,+50]$ et $\mathcal{Y}=[-50,+50]$, de sorte que le couple (X,Y) peut prendre un continuum de couples de valeurs (x,y) dans $\mathcal{X}\times\mathcal{Y}$.



Graphique 5.1 : Le couple de variables aléatoires (X,Y)

5.2.1. Loi jointe d'un couple de variables aléatoires continues

De façon semblable au cas univarié, lorsqu'un couple de variables aléatoires

(X, Y) peut ainsi prendre un continuum de couples de valeurs, il n'est plus possible de définir la loi ou distribution de probabilité jointe de (X, Y) par un ensemble de couples $\big((x_i, y_j), p_{x_i y_j}\big)$: il est impossible de faire la liste d'un continuum de valeurs.

Pour définir la loi ou distribution de probabilité jointe d'un couple de variables aléatoires continues (X, Y), on recourt à nouveau au concept de *densité de probabilité*, qui cette fois sera bivariée.

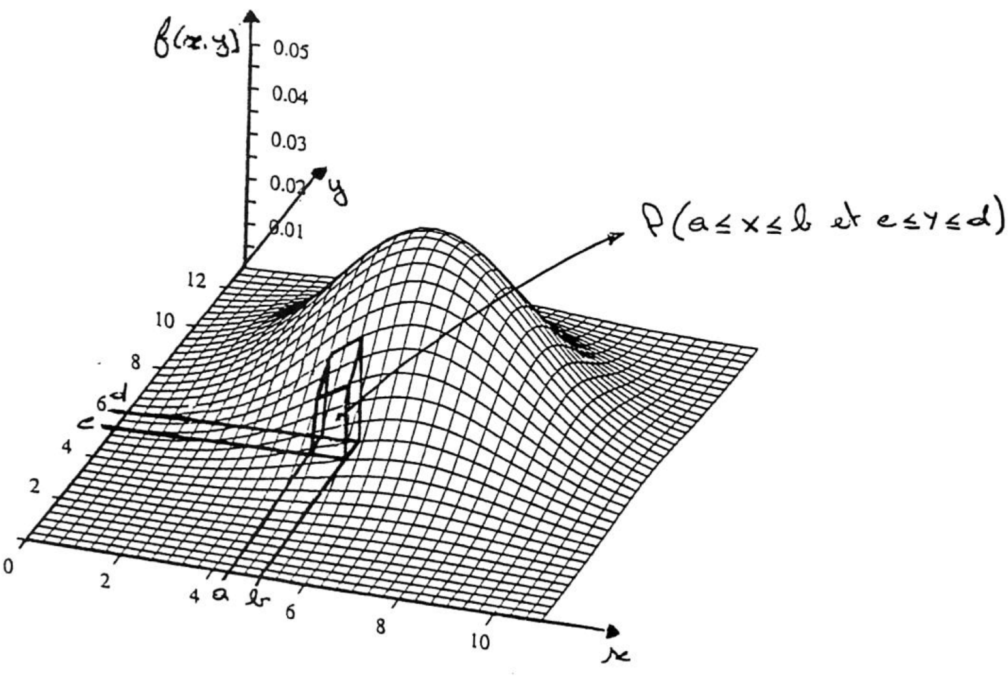
Définition 5.8 *On appelle densité de probabilité jointe d'un couple de variables aléatoires continues* (X, Y) *une fonction* $f(x, y)$ *telle que :*

$$\begin{cases} f(x,y) \geq 0, \quad \forall\,(x,y) \in \mathbb{R}^2 \\ \displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1 \end{cases} \tag{5.2}$$

et

$$I\!\!P(a \leq X \leq b \;\, et \; c \leq Y \leq d) = \int_a^b \int_c^d f(x,y)dxdy$$

Une fonction de densité jointe $f(x, y)$ est une fonction positive, strictement positive pour l'ensemble de couples de valeurs possibles (x, y) de (X, Y) et nulle pour tous les autres couples de valeurs, dont le volume total sous la surface qu'elle représente est égal à 1, et dont le volume sous la surface délimité par le rectangle $([a, b], [c, d])$ — notez qu'on peut prendre $a = -\infty$ et/ou $b = +\infty$ et/ou $c = -\infty$ et/ou $d = +\infty$ — définit la probabilité que le couple (X, Y) prenne sa valeur dans ce rectangle.



Graphique 5.2 : Une fonction de densité $f(x, y)$

Toute fonction $f(x,y)$ satisfaisant les conditions (5.2) de la Définition 5.8 définit la loi ou distribution de probabilité jointe d'un couple de variables aléatoires continues. Jointe à la condition $f(x,y) \geq 0$, $\forall\,(x,y) \in \mathbb{R}^2$, la condition $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$ assure que, comme il se doit, la probabilité de tout événement relatif au couple (X,Y) est toujours un nombre compris entre 0 et 1.

Comme dans le cas univarié, la condition $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$ est à mettre en parallèle avec la condition $\sum_{x_i}\sum_{y_j} p_{x_iy_j} = 1$ du cas discret. De même, la formule générique de la probabilité $I\!P(a \leq X \leq b \text{ et } c \leq Y \leq d) = \int_a^b\int_c^d f(x,y)dxdy$ est à mettre en parallèle avec la formule générique $I\!P(a \leq X \leq b \text{ et } c \leq Y \leq d) = \sum_{a\leq x_i\leq b}\sum_{c\leq y_j\leq d} p_{x_iy_j}$ valable dans le cas discret. On constate encore que le passage du cas discret au cas continu revient pour l'essentiel à remplacer une liste de probabilités ponctuelles (les couples $\big((x_i,y_j),p_{x_iy_j}\big)$) par une fonction de densité (bivariée) et des sommes (doubles) par des intégrales (doubles)[4].

Pareillement au cas univarié, la probabilité que (X,Y) prenne une valeur ponctuelle "$X = a$ et $Y = c$" (qu'elle soit ou non possible, lorsqu'elle est impossible, on a en plus $f(a,c) = 0$) est nulle puisque :

$$I\!P(X = a \text{ et } Y = c) = I\!P(a \leq X \leq a \text{ et } c \leq Y \leq c) = \int_a^a\int_c^c f(x,y)dxdy = 0$$

Par conséquent, comme dans le cas univarié, le fait que les bornes des intervalles sur X et Y soient définies par des inégalités strictes ($<$) ou non ($\leq$) est sans effet sur le calcul des probabilités. Ainsi par exemple :

$$I\!P(a \leq X \leq b \text{ et } c \leq Y \leq d) = I\!P(a < X < b \text{ et } c < Y < d)$$

A nouveau de façon semblable au cas univarié, la densité jointe $f(x,y)$ est approximativement proportionnelle à la probabilité que le couple de variables aléatoires (X,Y) prenne sa valeur dans un petit carré de côté ε centré sur (x,y). On a en effet :

$$\begin{aligned} &I\!P(a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2} \text{ et } c - \frac{\varepsilon}{2} \leq Y \leq c + \frac{\varepsilon}{2}) \\ &= \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}}\int_{c-\frac{\varepsilon}{2}}^{c+\frac{\varepsilon}{2}} f(x,y)dxdy \approx f(a,c) \times \varepsilon^2 \end{aligned}$$

En d'autres termes, la fonction de densité jointe $f(x,y)$ nous renseigne sur les probabilités *relatives* d'observer le couple de variables aléatoires (X,Y) dans un *même* petit carré de côté ε autour des différentes valeurs du couple (x,y).

La notion de densité de probabilité jointe est encore à rapprocher de la notion d'*histogramme de fréquence*, mais cette fois *bivarié*, utilisé en statistique descriptive pour décrire, dans une population, un couple de variables (x,y) — par exemple le

[4] Rappelons que, de façon semblable au cas univarié, une intégrale double n'est rien d'autre que la limite d'une double somme.

couple taille – poids d'un individu — prenant dont chacune prend un grand nombre de valeurs distinctes très proches les unes de autres. Dans cette perspective, comme dans le cas univarié, la fonction de densité jointe est tout simplement l'analogue en termes de probabilité de la surface limite qui apparaîtrait si la population était infinie, les mesures exactes, et que l'on considérait pour les deux variables des regroupements en classes de plus en plus fines. On l'a déjà souligné, l'hypothèse de continuité doit essentiellement être comprise comme une approximation commode.

5.2.2. Lois marginales d'un couple de variables aléatoires continues

Comme dans le cas discret, de la loi jointe du couple (X,Y), on peut déduire les *lois marginales* de X et de Y. Plus précisément, de la fonction de densité jointe, on peut déduire les fonctions de densité de X et de Y, qu'on appelle *densité de probabilité marginale* de X et *densité de probabilité marginale* de Y, densités marginales qui définissent les *lois marginales* de X et de Y, c'est-à-dire les lois de X et de Y considérées séparément, telles qu'elles découlent de la loi jointe.

Définition 5.9 *Soit un couple de variables aléatoires continues* (X,Y) *et sa densité de probabilité jointe* $f(x,y)$*. On appelle densité de probabilité marginale de* X *la fonction* $f_X(x)$ *donnée par :*

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy$$

De même, on appelle densité de probabilité marginale de Y *la fonction* $f_Y(y)$ *donnée par :*

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx$$

Ces formules sont très semblables à celles du cas discret :

$$p_{x_i\cdot} = \sum_{y_j} p_{x_iy_j} \quad \text{et} \quad p_{\cdot y_j} = \sum_{x_i} p_{x_iy_j}$$

et peuvent être, au moins intuitivement, interprétées de la même façon : rappelez-vous simplement qu'une intégrale est la limite d'une somme et que la valeur d'une densité en un point (resp. un couple de points) est approximativement proportionnelle à la probabilité de se trouver dans un petit intervalle (resp. un petit carré) centré sur ce point (resp. ce couple de points).

On vérifie facilement que les fonctions $f_X(x)$ et $f_Y(y)$ définissent bien des densités de probabilité. Par définition, on a en effet d'une part :

$$f_X(x) \geq 0,\ \forall\, x \in \mathbb{R} \qquad \text{et} \qquad f_Y(y) \geq 0,\ \forall\, y \in \mathbb{R}$$

et d'autre part :

$$\begin{aligned}
\int_{-\infty}^{+\infty} f_X(x)dx &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y)dy \right) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dxdy = 1 \\
\int_{-\infty}^{+\infty} f_Y(y)dy &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x,y)dx \right) dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dxdy = 1
\end{aligned}$$

On vérifie également facilement que les densités marginales $f_X(x)$ et $f_Y(x)$ définissent bien les lois ou distributions de probabilité marginales de X et de Y, c'est-à-dire les lois de X et de Y considérées séparément, telles qu'elles découlent de la loi jointe du couple (X,Y). On a en effet :

$$\begin{aligned}
I\!P(a \leq X \leq b) &= I\!P(a \leq X \leq b \text{ et "}Y \text{ est quelconque"}) \\
&= I\!P(a \leq X \leq b \text{ et } -\infty \leq Y \leq +\infty) \\
&= \int_a^b \int_{-\infty}^{+\infty} f(x,y)dxdy = \int_a^b \left(\int_{-\infty}^{+\infty} f(x,y)dy \right) dx \\
&= \int_a^b f_X(x)dx
\end{aligned}$$

et de même :

$$\begin{aligned}
I\!P(c \leq Y \leq d) &= I\!P(c \leq Y \leq d \text{ et "}X \text{ est quelconque"}) \\
&= I\!P(-\infty \leq X \leq +\infty \text{ et } c \leq Y \leq d) \\
&= \int_{-\infty}^{+\infty} \int_c^d f(x,y)dxdy = \int_c^d \left(\int_{-\infty}^{+\infty} f(x,y)dx \right) dy \\
&= \int_c^d f_Y(y)dy
\end{aligned}$$

Les lois marginales de X et de Y étant comme toute loi d'une variable aléatoire continue, il va de soi que, comme pour le cas discret, toutes les notions (fonction de répartition, espérance, variance, etc...) discutées au Chapitre 4 s'appliquent sans modification à chacune d'elles. On peut en particulier en calculer de la façon habituelle l'espérance et la variance, et ainsi obtenir l'espérance et la variance des variables aléatoires X et Y considérées séparément, telles qu'elles découlent de leur loi jointe.

5.2.3. Lois conditionnelles d'un couple de variables aléatoires continues

A nouveau comme dans le cas discret, de la loi jointe du couple (X,Y), on peut également déduire les *lois conditionnelles* de X sachant $Y = y$ et de Y sachant $X = x$. Plus précisément, de la densité jointe du couple (X,Y), on peut déduire les *densités conditionnelles* de X sachant $Y = y$ et de Y sachant $X = x$, densités conditionnelles qui définissent les *lois conditionnelles* de X sachant $Y = y$ et de Y sachant $X = x$, c'est-à-dire respectivement la loi de X lorsque Y est égale à y et de façon symétrique la loi de Y lorsque X est égale à x, telles qu'elles découlent de la loi jointe.

Définition 5.10 *Soit un couple de variables aléatoires continues (X, Y) et sa densité de probabilité jointe $f(x, y)$. On appelle densité de probabilité conditionnelle de X sachant que $Y = y$ la fonction $f_{X|Y}(x|y)$ donnée par :*

$$f_{X|Y}(x|y) = \frac{f(x,y)}{\int_{-\infty}^{+\infty} f(x,y)dx} = \frac{f(x,y)}{f_Y(y)} \qquad (\forall\, y \text{ tel que } f_Y(y) \neq 0)$$

De même, on appelle densité de probabilité conditionnelle de Y sachant que $X = x$ la fonction $f_{Y|X}(y|x)$ donnée par :

$$f_{Y|X}(y|x) = \frac{f(x,y)}{\int_{-\infty}^{+\infty} f(x,y)dy} = \frac{f(x,y)}{f_X(x)} \qquad (\forall\, x \text{ tel que } f_X(x) \neq 0)$$

Il y a un continuum de densités conditionnelles de X sachant $Y = y$, une pour chacune des valeurs du continuum de valeurs possibles y de Y (c'est-à-dire $\forall\, y$ tel que $f_Y(y) \neq 0$), et pareillement un continuum de densités conditionnelles de Y sachant $X = x$, une pour chacune des valeurs du continuum de valeurs possibles x de X (c'est-à-dire $\forall\, x$ tel que $f_X(x) \neq 0$).

Les formules données à la Définition 5.10 sont à nouveau très semblables à celles du cas discret :

$$p_{x_i|y_j} = \frac{p_{x_iy_j}}{\sum\limits_{x_i} p_{x_iy_j}} = \frac{p_{x_iy_j}}{p_{.y_j}} \quad \text{et} \quad p_{y_j|x_i} = \frac{p_{x_iy_j}}{\sum\limits_{y_j} p_{x_iy_j}} = \frac{p_{x_iy_j}}{p_{x_i.}}$$

peuvent encore être, au moins intuitivement, interprétées de la même façon.

On vérifie aisément que les fonctions $f_{X|Y}(x|y)$ (quelle que soit la valeur possible y de Y considérée) et $f_{Y|X}(y|x)$ (quelle que soit la valeur possible x de X considérée) définissent bien des densités de probabilité, densités qui sont comme toute densité d'une variable aléatoire continue. Par définition, on a en effet d'une part :

$$f_{X|Y}(x|y) \geq 0, \;\; \forall\, x \in \mathbb{R} \qquad \text{et} \qquad f_{Y|X}(y|x) \geq 0, \;\; \forall\, y \in \mathbb{R}$$

et d'autre part :

$$\begin{aligned}
\int_{-\infty}^{+\infty} f_{X|Y}(x|y)dx &= \int_{-\infty}^{+\infty} \frac{f(x,y)}{f_Y(y)}dx = \frac{1}{f_Y(y)} \int_{-\infty}^{+\infty} f(x,y)dx = \frac{f_Y(y)}{f_Y(y)} = 1 \\
\int_{-\infty}^{+\infty} f_{Y|X}(y|x)dy &= \int_{-\infty}^{+\infty} \frac{f(x,y)}{f_X(x)}dy = \frac{1}{f_X(x)} \int_{-\infty}^{+\infty} f(x,y)dy = \frac{f_X(x)}{f_X(x)} = 1
\end{aligned}$$

On vérifie également que les densités conditionnelles $f_{X|Y}(x|y)$ (pour chaque valeur possible y de Y) et $f_{Y|X}(y|x)$ (pour chaque valeur possible x de X) définissent bien les lois ou distributions de probabilité conditionnelles de respectivement X

sachant $Y = y$ et Y sachant $X = x$, c'est-à-dire respectivement la loi de X lorsque Y prend une valeur fixe et égale à y et la loi de Y lorsque X prend une valeur fixe et égale à x, telles qu'elles découlent de la loi jointe. On a en effet :

$$\begin{aligned}
I\!\!P(a \leq X \leq b|Y = c) &= \lim_{\varepsilon \to 0} I\!\!P(a \leq X \leq b|c - \tfrac{\varepsilon}{2} \leq Y \leq c + \tfrac{\varepsilon}{2}) \\
&= \lim_{\varepsilon \to 0} \frac{I\!\!P(a \leq X \leq b \text{ et } c - \frac{\varepsilon}{2} \leq Y \leq c + \frac{\varepsilon}{2})}{I\!\!P(c - \frac{\varepsilon}{2} \leq Y \leq c + \frac{\varepsilon}{2})} \\
&= \lim_{\varepsilon \to 0} \frac{\int_a^b \int_{c-\frac{\varepsilon}{2}}^{c+\frac{\varepsilon}{2}} f(x,y)dxdy}{\int_{-\infty}^{+\infty} \int_{c-\frac{\varepsilon}{2}}^{c+\frac{\varepsilon}{2}} f(x,y)dxdy} = \frac{\int_a^b f(x,c)dx}{\int_{-\infty}^{+\infty} f(x,c)dx} \\
&= \frac{\int_a^b f(x,c)dx}{f_Y(c)} = \int_a^b \frac{f(x,c)}{f_Y(c)}dx \\
&= \int_a^b f_{X|Y}(x|c)dx
\end{aligned}$$

et de façon symétrique :

$$\begin{aligned}
I\!\!P(c \leq Y \leq d|X = a) &= \lim_{\varepsilon \to 0} I\!\!P(c \leq Y \leq d|a - \tfrac{\varepsilon}{2} \leq X \leq a + \tfrac{\varepsilon}{2}) \\
&= \lim_{\varepsilon \to 0} \frac{I\!\!P(a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2} \text{ et } c \leq Y \leq d)}{I\!\!P(a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2})} \\
&= \lim_{\varepsilon \to 0} \frac{\int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} \int_c^d f(x,y)dxdy}{\int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} \int_{-\infty}^{+\infty} f(x,y)dxdy} = \frac{\int_c^d f(a,y)dy}{\int_{-\infty}^{+\infty} f(a,y)dy} \\
&= \frac{\int_c^d f(a,y)dy}{f_X(a)} = \int_c^d \frac{f(a,y)}{f_X(a)}dy \\
&= \int_c^d f_{Y|X}(y|a)dy
\end{aligned}$$

On découvre au passage une nouvelle particularité des lois continues : il est possible d'y définir des probabilités conditionnelles par rapport à des événements de probabilité nulle (les valeurs ponctuelles $Y = y$ et $X = x$, où y et x sont des valeurs possibles de respectivement Y et X ; cela ne marche évidemment pas pour des valeurs impossibles de Y et X, pour lesquelles on a $f_Y(y) = 0$ et $f_X(x) = 0$). Rappelons ce que nous avons déjà dit : il ne faut pas chercher à trouver dans le concept de continuité et les propriétés qui en découlent quelque chose de "naturel". Il s'agit avant tout d'un outil commode pour traiter les situations où considérer qu'une (plusieurs) variable(s) aléatoire(s) peut (peuvent) prendre un continuum de valeurs possibles constitue une approximation raisonnable de la réalité.

Comme dans le cas discret, les lois conditionnelles de X sachant $Y = y$ et Y sachant $X = x$ définies par les densités conditionnelles $f_{Y|X}(y|x)$ et $f_{X|Y}(x|y)$ sont les lois / densités qu'il convient de considérer lorsqu'on s'intéresse à la façon dont, respectivement, X dépend de Y et Y dépend de X.

Notons encore finalement que, par définition, les densités conditionnelles font un lien entre la densité jointe et les densités marginales.

Propriété 5.2 *Quelles que soient les variables aléatoires continues X et Y, on a pour toutes les valeurs possibles x de X et y de Y :*

$$f(x,y) = f_{X|Y}(x|y)f_Y(y) \qquad \textit{et} \qquad f(x,y) = f_{Y|X}(y|x)f_X(x)$$

Comme dans le cas discret, on peut en particulier utiliser ces relations pour construire des densités jointes à partir de densités conditionnelles et de densités marginales (en d'autres termes, des lois jointes à partir de lois conditionnelles et de lois marginales).

5.2.3.1. Espérance conditionnelle et variance conditionnelle

On l'a vu, pour chaque valeur possible x de X, $f_{Y|X}(y|x)$ (et, de façon symétrique, pour chaque valeur possible y de Y, $f_{X|Y}(x|y)$) définit un densité de probabilité pour Y (de façon symétrique, pour X) semblable à toute densité d'une variable aléatoire continue. On peut donc, comme pour toute loi d'une variable aléatoire continue, "résumer" sa loi / densité par diverses quantités, en particulier son espérance et sa variance.

Définition 5.11 *On appelle espérance conditionnelle de Y sachant $X = x$ d'un couple de variables aléatoires continues (X, Y) la quantité :*

$$E(Y|X=x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x)dy$$

variance conditionnelle de Y sachant $X = x$ la quantité :

$$\begin{aligned} V(Y|X=x) &= \int_{-\infty}^{+\infty} \left(y-\mu_{Y|x}\right)^2 f_{Y|X}(y|x)dy = E\left[(Y-E(Y|X=x))^2 \,|X=x\right] \\ &= \int_{-\infty}^{+\infty} y^2 f_{Y|X}(y|x)dy - \mu_{Y|x}^2 = E(Y^2|X=x) - (E(Y|X=x))^2 \end{aligned}$$

où $\mu_{Y|x} = E(Y|X=x)$, et écart-type conditionnel de Y sachant $X = x$ la quantité :

$$\sigma(Y|X=x) = \sqrt{V(Y|X=x)}$$

Ces formules sont la simple transposition au cas continu des formules données pour le cas discret. Elle s'interprètent — à la transposition discret / continu près — de la même façon que dans le cas discret.

Comme dans le cas discret, vues comme des fonctions de x, $E(Y|X=x)$ et $V(Y|X=x)$ sont respectivement appelées *fonction d'espérance conditionnelle de Y sachant X*, ou *courbe de régression de Y en X*, et *fonction de variance conditionnelle de Y sachant X*, et sont respectivement notées $E(Y|X)$ et $V(Y|X)$.

Il va sans dire que toutes les considérations ci-dessus s'appliquent, de façon symétrique, au lois conditionnelles de X sachant $Y = y$.

5.3. Indépendance de deux variables aléatoires

L'indépendance de deux variables aléatoires est définie sur la base de la définition de l'indépendance de deux événements. Deux variables aléatoires X et Y sont dites indépendantes si tous les événements relatifs à X, c'est-à-dire tous les événements décrits à l'aide de X uniquement, sont indépendants de tous les événements relatifs à Y, c'est-à-dire de tous les événements décrits à l'aide de Y uniquement. De façon formelle, on a la définition suivante.

Définition 5.12 *Deux variables aléatoires X et Y sont indépendantes si et seulement si, dans le cas discret, on a pour toutes les valeurs possibles x_i de X et y_j de Y :*

$$p_{x_i y_j} = p_{x_i .} p_{. y_j} \tag{5.3}$$

ou de façon équivalente :

$$p_{x_i | y_j} = p_{x_i .}$$

ou encore de façon équivalente :

$$p_{y_j | x_i} = p_{. y_j}$$

et dans le cas continu, on a pour toutes les valeurs possibles x de X et y de Y :

$$f(x, y) = f_X(x) f_Y(y) \tag{5.4}$$

ou de façon équivalente :

$$f_{X|Y}(x|y) = f_X(x)$$

ou encore de façon équivalente :

$$f_{Y|X}(y|x) = f_Y(y)$$

Autrement dit, deux variables aléatoires X et Y (discrètes ou continues) sont indépendantes si et seulement si leur loi jointe est égale au produit de leur loi marginale respective, ou façon équivalente, si leurs lois conditionnelles respectives sont égales à leur loi marginale respective.

Intuitivement, lorsque deux variables aléatoires sont indépendantes, la connaissance de la valeur prise par l'une d'entre elle n'apporte aucune information quant à la valeur que pourrait prendre l'autre.

Dans notre exemple du magasin de téléviseurs, les variables aléatoires X et Y ne sont pas indépendantes. On a effet vu que les distributions conditionnelles de Y sachant $X = x_i$ variaient selon x_i et étaient donc différentes de la distribution marginale de Y : on n'a pas $p_{y_j|x_i} = p_{.y_j}$ pour toutes les valeurs possibles x_i et y_j. De façon équivalente, on vérifie que :

$$0,07 = I\!P(X = 1 \text{ et } Y = 1) \neq I\!P(X = 1)I\!P(Y = 1) = 0,20 \times 0,16 = 0,032$$

autrement dit que l'on n'a pas $p_{x_i y_j} = p_{x_i.}p_{.y_j}$ pour toutes les valeurs possibles x_i et y_j.

Comme l'indépendance de deux événements, l'indépendance de deux variables aléatoires (discrètes ou continues) n'est pas toujours une propriété à vérifier, à démontrer. Il arrive souvent qu'elle soit *admise* comme *hypothèse théorique* étant donné la nature même de l'épreuve aléatoire considérée. Ainsi, lorsqu'on représente le lancer de deux dés par un couple de variables aléatoires (X, Y), X décrivant le résultat du lancer d'un des dés et Y de l'autre, il est naturel de considérer les deux variables aléatoires X et Y comme indépendantes.

De façon générale, on peut considérer comme raisonnable l'hypothèse d'indépendance de deux variables aléatoires X et Y chaque fois que l'on ne voit aucun mécanisme direct ou indirect par lequel les valeurs prises par l'une pourrait influer les valeurs prises — plus précisément la probabilité des valeurs prises — par l'autre.

Lorsque deux variables aléatoires X et Y peuvent, par hypothèse, de par la nature même de l'épreuve aléatoire, être considéré comme indépendantes, par application des formules (5.3) et (5.4), on peut déduire leur loi jointe de leur loi marginale respective. Ainsi, dans notre exemple de lancer de deux dés, si les dés sont équilibrés, X et Y ont comme lois marginales :

x_i	1	2	3	4	5	6
$p_{x_i.}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

et

y_j	1	2	3	4	5	6
$p_{.y_j}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

lois marginales dont on peut déduire sous l'hypothèse d'indépendance par application de (5.3) la loi jointe du couple (X, Y) :

$x_i \backslash y_j$	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

5.4. Fonction de deux variables aléatoires

Considérons un couple de variables aléatoires (X, Y) (discrètes ou continues) et une fonction $g(X, Y)$ de X et Y, c'est-à-dire une application $g(.,.)$ qui à tout couple de valeurs (x, y) associe *une et au plus une* valeur $z = g(x, y)$. $Z = g(X, Y)$ est une nouvelle variable aléatoire.

5.4.1. Loi d'une fonction de deux variables aléatoires

Connaissant la loi jointe du couple (X, Y), il est en général aisé de déterminer la loi d'une variable aléatoire Z définie par $Z = g(X, Y)$, a tout le moins dans le cas discret. Dans le cas continu, le problème est nettement plus délicat. Nous ne l'aborderons pas ici.

Dans le cas discret, la mécanique est semblable à celle développée à la Section 4.3.1 pour le cas d'une fonction d'une seule variable aléatoire.

Comme la loi de toute variable aléatoire discrète, la loi d'une variable aléatoire $Z = g(X, Y)$ est donnée par un ensemble de couples (z_k, p_{z_k}), où les z_k désignent les valeurs possibles de Z — notons l'ensemble de ces valeurs possibles $\mathcal{Z}$ — et les $p_{z_k} = \mathbb{P}(Z = z_k)$ les probabilités de ces valeurs possibles.

Etant donné la loi jointe de (X, Y) et la fonction $g(X, Y)$, on peut aisément trouver les valeurs possibles z_k de Z : il s'agit de toutes les valeurs distinctes $z_k = g(x_i, y_j)$ obtenues en passant en revue les différents couples de valeurs possibles (x_i, y_j) de (X, Y).

Les probabilités $p_{z_k} = \mathbb{P}(Z = z_k)$ s'obtiennent tout aussi aisément : elles sont tout simplement égales aux probabilités des événements "$Z = z_k$" dans l'épreuve aléatoire décrite par le couple de variables aléatoires (X, Y), c'est-à-dire pour chaque valeur possible z_k de Z égale à la somme des probabilités des couples de valeurs possibles (x_i, y_j) de (X, Y) pour lesquels l'événement "$Z = z_k$" se réalise :

$$p_{z_k} = \mathbb{P}(Z = z_k) = \sum_{(x_i, y_j): g(x_i, y_j) = z_k} p_{x_i y_j} \tag{5.5}$$

où $\sum_{(x_i, y_j): g(x_i, y_j) = z_k}$ signifie "somme pour tous les couples de valeurs possibles (x_i, y_j) de (X, Y) tels que $g(x_i, y_j) = z_k$".

Comme par définition l'application $g(.,.)$ associe a tout (x_i, y_j) une et au plus une valeur z_k, les différents événements "$Z = z_k$" forment bien une partition dans l'épreuve aléatoire décrite par le couple (X, Y), ce qui assure, comme il se doit, que $\sum_{z_k} p_{z_k} = 1$.

Illustrons l'application de cette règle de calcul en reprenant notre exemple de magasin de téléviseurs. Supposons que notre magasin vende aussi des magnétoscopes et redéfinissons les variables aléatoires X et Y par :

X = le nombre de téléviseurs vendus pendant un après-midi quelconque
Y = le nombre de magnétoscopes vendus durant la même période

Supposons que la loi jointe de (X, Y) et les lois marginales correspondantes de X et de Y soient données par :

$x_i \backslash y_j$	0	1	2	$p_{x_i.}$
0	0,59	0,14	0,06	0,79
1	0,09	0,05	0,02	0,16
2	0,02	0,02	0,01	0,05
$p_{.y_j}$	0,70	0,21	0,09	1

Notez que la loi marginale du nombre de téléviseurs vendus (appelé Y dans notre exemple du début) est inchangée.

Supposons encore que la recette (en milliers de F) réalisée lors de la vente d'un téléviseur est de 40, et de 20 lors de la vente d'un magnétoscope. La recette totale au cours d'un après-midi quelconque est donc : $Z = 40X + 20Y$.

Etablissons la loi de la variable aléatoire Z. On vérifie aisément que ses valeurs possibles sont :

$$\mathcal{Z} = \{0, 20, 40, 60, 80, 100, 120\}$$

On trouve comme probabilités de ces valeurs possibles :

$$\begin{aligned}
I\!\!P(Z=0) &= I\!\!P(X=0 \text{ et } Y=0) = 0,59 \\
I\!\!P(Z=20) &= I\!\!P(X=0 \text{ et } Y=1) = 0,14 \\
I\!\!P(Z=40) &= I\!\!P([(X=1 \text{ et } Y=0) \text{ ou } (X=0 \text{ et } Y=2)] \\
&= I\!\!P(X=1 \text{ et } Y=0) + I\!\!P(X=0 \text{ et } Y=2) \\
&= 0,09 + 0,06 = 0,15 \\
I\!\!P(Z=60) &= I\!\!P(X=1 \text{ et } Y=1) = 0,05 \\
I\!\!P(Z=80) &= I\!\!P([(X=2 \text{ et } Y=0) \text{ ou } (X=1 \text{ et } Y=2)] \\
&= I\!\!P(X=2 \text{ et } Y=0) + I\!\!P(X=1 \text{ et } Y=2) \\
&= 0,02 + 0,02 = 0,04 \\
I\!\!P(Z=100) &= I\!\!P(X=2 \text{ et } Y=1) = 0,02 \\
I\!\!P(Z=120) &= I\!\!P(X=2 \text{ et } Y=2) = 0,01
\end{aligned}$$

La loi de Z est donc :

z_k	0	20	40	60	80	100	120
p_{z_k}	0,59	0,14	0,15	0,05	0,04	0,02	0,01

On vérifie que la somme des probabilités est bien égale à 1.

5.4.2. Espérance et variance d'une fonction de deux variables aléatoires

La loi d'une variable aléatoire Z, discrète ou continue, définie par $Z = g(X,Y)$ étant établie, on peut calculer son espérance et sa variance (et par conséquent son écart-type) de la façon habituelle, comme pour toute variable aléatoire.

Comme dans le cas d'une fonction d'une seule variable aléatoire, on peut également calculer ces quantités directement à partir de la loi jointe de (X,Y), sans avoir à établir la loi de Z, ce qui est souvent plus commode. Cette possibilité découle de la propriété générale suivante — que nous admettrons —, qui donne la formule permettant d'obtenir l'espérance de $Z = g(X,Y)$, et dont on peut déduire celle valable pour la variance de $Z = g(X,Y)$.

Propriété 5.3 *Soit (X,Y) un couple de variables aléatoires. Si $Z = g(X,Y)$, où $g(X,Y)$ est une fonction quelconque de X et Y, on a dans le cas discret :*

$$E(Z) = E\left[g(X,Y)\right] = \sum_{x_i}\sum_{y_j} g(x_i,y_j)p_{x_iy_j} \tag{5.6}$$

et dans le cas continu :

$$E(Z) = E\left[g(X,Y)\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy \tag{5.7}$$

On note que pour la fonction $g(X,Y) = X$, les formules (5.6) et (5.7) donnent l'espérance de X, telle qu'elle peut être calculée sur base de la loi marginale de X. En effet, on a dans le cas discret :

$$E(X) = \sum_{x_i}\sum_{y_j} x_i p_{x_iy_j} = \sum_{x_i} x_i\left(\sum_{y_j} p_{x_iy_j}\right) = \sum_{x_i} x_i p_{x_i\cdot}$$

et dans le cas continu :

$$E(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy = \int_{-\infty}^{+\infty} x\left(\int_{-\infty}^{+\infty} f(x,y)dy\right)dx = \int_{-\infty}^{+\infty} xf_X(x)dx$$

Pour $g(X,Y) = (X - E(X))^2$, on obtient pareillement la variance $V(X) = E\left[(X - E(X))^2\right]$ de la variable aléatoire X, qu'elle peut être calculée sur base de la loi marginale de X. Le même raisonnement vaut bien entendu pour Y (considérez $g(X,Y) = Y$ et $g(X,Y) = (Y - E(Y))^2$).

De la Propriété 4.8, on a pour la variance de la variable aléatoire $Z = g(X,Y)$:

$$\begin{aligned} V(Z) = V\left[g(X,Y)\right] &= E\left[(Z - E(Z))^2\right] = E\left[(g(X,Y) - E\left[g(X,Y)\right])^2\right] \\ &= E(Z^2) - (E(Z))^2 = E\left[(g(X,Y))^2\right] - (E\left[g(X,Y)\right])^2 \end{aligned}$$

Par application de la Propriété 5.3, on en déduit des formules pour calculer la variance de $Z = g(X, Y)$ sur base de la loi jointe de (X, Y).

Propriété 5.4 *Soit (X, Y) un couple de variables aléatoires. Si $Z = g(X, Y)$, où $g(X, Y)$ est une fonction quelconque de X et Y, on a dans le cas discret :*

$$\begin{aligned} V(Z) = V\left[g(X, Y)\right] &= \sum_{x_i}\sum_{y_j}\left(g(x_i, y_j) - \mu_g\right)^2 p_{x_i y_j} \\ &= \sum_{x_i}\sum_{y_j} g(x_i, y_j))^2 p_{x_i y_j} - \mu_g^2 \end{aligned}$$

et dans le cas continu :

$$\begin{aligned} V(Z) = V\left[g(X, Y)\right] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(g(x, y) - \mu_g\right)^2 f(x, y)dxdy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(g(x, y)\right)^2 f(x, y)dxdy - \mu_g^2 \end{aligned}$$

où $\mu_g = E(Z) = E\left[g(X, Y)\right]$.

Comme celle d'une fonction linéaire d'une seule variable aléatoire, l'espérance d'une fonction linéaire d'un couple de variables aléatoires (X, Y), discrètes ou continues, est particulièrement simple à calculer.

Propriété 5.5 *Soient X et Y deux variables aléatoires quelconques. Si $Z = a + bX + cY$, on a :*

$$E(Z) = E\left(a + bX + cY\right) = a + bE(X) + cE(Y)$$

et donc en particulier :

$$E(X + Y) = E(X) + E(Y) \quad \text{et} \quad E(X - Y) = E(X) - E(Y)$$

En effet, de la Propriété 5.3, on a dans le cas discret :

$$\begin{aligned} E\left(a + bX + cY\right) &= \sum_{x_i}\sum_{y_j}\left(a + bx_i + cy_j\right) p_{x_i y_j} \\ &= a\underbrace{\sum_{x_i}\sum_{y_j} p_{x_i y_j}}_{=1} + b\underbrace{\sum_{x_i}\sum_{y_j} x_i p_{x_i y_j}}_{=E(X)} + c\underbrace{\sum_{x_i}\sum_{y_j} y_j p_{x_i y_j}}_{=E(Y)} \end{aligned}$$

De même, dans le cas continu, on a :

$$E\left(a + bX + cY\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(a + bx + cy\right) f(x, y)dxdy$$

$$= a\underbrace{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy}_{=1} + b\underbrace{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)dxdy}_{=E(X)}$$

$$+c\underbrace{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dxdy}_{=E(Y)}$$

Illustrons l'application de cette règle dans le cas discret en poursuivant notre exemple du magasin de téléviseurs, dont pour rappel la loi jointe du couple (X, Y) et les lois marginales correspondantes de X et de Y (X = nombre de téléviseurs vendus et Y = nombre de magnétoscopes vendus) sont données par :

$x_i \backslash\ y_j$	0	1	2	$p_{x_i.}$
0	0,59	0,14	0,06	0,79
1	0,09	0,05	0,02	0,16
2	0,02	0,02	0,01	0,05
$p_{.y_j}$	0,70	0,21	0,09	1

Supposons que le vendeur souhaite savoir à quelle recette totale il peut s'attendre. Sa recette totale étant $Z = 40X + 20Y$, sa recette totale espérée est donnée par :

$$E(Z) = E(40X + 20Y) = 40E(X) + 20E(Y)$$

On trouve :

$$\begin{aligned} E(X) &= \sum_{x_i} x_i p_{x_i.} = 0 \times 0,79 + 1 \times 0,16 + 2 \times 0,05 = 0,26 \\ E(Y) &= \sum_{y_j} y_j p_{.y_j} = 0 \times 0,70 + 1 \times 0,21 + 2 \times 0,09 = 0,39 \end{aligned}$$

et donc :

$$E(Z) = E(40X + 20Y) = 40 \times 0,26 + 20 \times 0,39 = 10,4 + 7,8 = 18,2$$

On vérifie qu'en utilisant la loi de Z qui est donnée par :

z_k	0	20	40	60	80	100	120
p_{z_k}	0,59	0,14	0,15	0,05	0,04	0,02	0,01

on trouve de même :

$$\begin{aligned} E(Z) &= \sum_{z_k} z_k p_{z_k} = 0 \times 0,59 + 20 \times 0,14 + 40 \times 0,15 \\ & +60 \times 0,05 + 80 \times 0,04 + 100 \times 0,02 + 120 \times 0,01 = 18,2 \end{aligned}$$

Nous verrons ci-dessous que comme l'espérance, la variance (et par conséquent l'écart-type) d'une fonction linéaire d'un couple de variables aléatoires (X, Y), discrètes ou continues, peut aussi être obtenue de façon assez simple.

5.4.3. Covariance et corrélation entre deux variables aléatoires

5.4.3.1. Covariance

Considérons la fonction $g(X,Y) = (X - E(X))\,(Y - E(Y))$ d'un couple de variables aléatoires (X, Y), discrètes ou continues. L'espérance de la variable aléatoire définie par cette fonction est appelée la covariance entre X et Y.

Définition 5.13 *On appelle covariance entre deux variables aléatoires X et Y quelconques la quantité :*

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = Cov(Y,X)$$

c'est-à-dire, dans le cas discret, la quantité donnée par :

$$\sum_{x_i}\sum_{y_j}(x_i - \mu_X)(y_j - \mu_Y)p_{x_i y_j} \tag{5.8}$$

et dans le cas continu, par :

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - \mu_X)(y - \mu_Y)f(x,y)dxdy \tag{5.9}$$

où $\mu_X = E(X)$ et $\mu_Y = E(Y)$.

Les formules (5.8) et (5.9) découlent bien entendu directement de la Propriété 5.3. Notez le caractère symétrique de la définition de la covariance. Remarquez également que la covariance entre une variable aléatoire X quelconque et elle-même est simplement égale à la variance de X : $Cov(X,X) = V(X)$. La covariance est évidemment à rapprocher de son analogue rencontré en statistique descriptive.

La covariance, ou plus précisément son *signe*, caractérise le *type de dépendance* — une relation positive ou négative — existant entre deux variables aléatoires. Si les valeurs de X supérieures à son espérance $(X - E(X) > 0)$ ont tendance à survenir — entendez ont une probabilité élevée de survenir — en même temps que les valeurs de Y supérieures à son espérance $(Y - E(Y) > 0)$ et que les valeurs de X inférieures à son espérance $(X - E(X) < 0)$ ont tendance à survenir en même temps que les valeurs de Y inférieures à son espérance $(Y - E(Y) < 0)$, on voit aisément de sa définition que la covariance $Cov(X,Y)$ prendra une valeur *positive*, indiquant une relation positive (croissante) entre X et Y. On peut par exemple s'attendre à une covariance positive entre le poids ($= X$) et la taille ($= Y$) d'un homme pris au hasard. A l'opposé, si les valeurs de X supérieures à son espérance $(X - E(X) > 0)$ ont tendance à survenir en même temps que les valeurs de Y inférieures à son espérance $(Y - E(Y) < 0)$ et que les valeurs de X inférieures à son espérance ($X - E(X) < 0$) ont tendance à survenir en même temps que les valeurs de Y supérieures à son espérance ($Y - E(Y) > 0$), on voit que la covariance $Cov(X, Y)$ prendra une valeur *négative*, indiquant une relation *négative* (décroissante) entre X et Y. On peut par exemple s'attendre à une covariance négative entre la durée de vie ($= X$) et la consommation hebdomadaire de tabac ($= Y$) d'un homme pris au hasard.

La covariance entre X et Y peut être exprimée sous une forme souvent plus commode pour les calculs. En utilisant la Propriété 5.5, on obtient :

$$\begin{aligned} Cov(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= E[XY - E(X)Y - XE(Y) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Propriété 5.6 *Soient X et Y deux variables aléatoires quelconques. On a :*

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

c'est-à-dire, dans le cas discret :

$$Cov(X, Y) = \sum_{x_i} \sum_{y_j} x_i y_j p_{x_i y_j} - \mu_X \mu_Y \qquad (5.10)$$

et dans le cas continu :

$$Cov(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x, y)dxdy - \mu_X \mu_Y \qquad (5.11)$$

où $\mu_X = E(X)$ et $\mu_Y = E(Y)$.

Les formules (5.10) et (5.11) découlent à nouveau directement de la Propriété 5.3.

Lorsqu'on connaît les variances et la covariance de deux variables aléatoires X et Y, discrètes ou continues, on peut facilement obtenir la variance (et par conséquent l'écart-type) d'une fonction linéaire de ces variables.

Propriété 5.7 *Soient X et Y deux variables aléatoires quelconques. Si $Z = a + bX + cY$, on a :*

$$V(Z) = V(a + bX + cY) = b^2 V(X) + c^2 V(Y) + 2bcCov(X, Y)$$

et

$$\sigma(a + bX + cY) = \sqrt{V(a + bX + cY)}$$

et donc en particulier :

$$\begin{aligned} V(X + Y) &= V(X) + V(Y) + 2Cov(X, Y) \\ V(X - Y) &= V(X) + V(Y) - 2Cov(X, Y) \end{aligned}$$

En effet, de la Propriété 4.8, on a :

$$V(a+bX+cY) = E\left[(a+bX+cY-E[a+bX+cY])^2\right]$$

Posons $W = a+bX+cY-E(a+bX+cY)$. De la Propriété 5.5, on a :

$$\begin{aligned} W &= a+bX+cY-(a+bE(X)+cE(Y)) \\ &= b(X-E(X))+c(Y-E(Y)) \end{aligned}$$

de sorte que :

$$\begin{aligned} & V(a+bX+cY) \\ = \; & E\left(W^2\right) = E\left[(b(X-E(X))+c(Y-E(Y)))^2\right] \\ = \; & E\left[b^2(X-E(X))^2+c^2(Y-E(Y))^2+2bc(X-E(X))(Y-E(Y))\right] \end{aligned}$$

soit, en utilisant encore les Propriétés 5.5 et 4.8, et la Définition 5.13 :

$$\begin{aligned} & V(a+bX+cY) \\ = \; & b^2E\left[(X-E(X))^2\right]+c^2E\left[(Y-E(Y))^2\right]+2bcE\left[(X-E(X))(Y-E(Y))\right] \\ = \; & b^2V(X)+c^2V(Y)+2bcCov(X,Y) \end{aligned}$$

On ne manquera pas de noter que, contrairement au cas de l'espérance, la variance d'une somme de deux variables aléatoires n'est généralement pas égale à la somme de leur variance. Il n'y a égalité que dans le cas où $Cov(X,Y) = 0$.

Illustrons ce qui précède à l'aide de notre exemple de magasin de téléviseurs. Pour rappel, la loi jointe de (X,Y) et les lois marginales correspondantes de X et de Y (X = nombre de téléviseurs vendus et Y = nombre de magnétoscopes vendus) sont données par :

$x_i \backslash y_j$	0	1	2	$p_{x_i.}$
0	0,59	0,14	0,06	0,79
1	0,09	0,05	0,02	0,16
2	0,02	0,02	0,01	0,05
$p_{.y_j}$	0,70	0,21	0,09	1

Calculons la covariance entre X et Y. On a déjà obtenu $E(X) = 0,26$ et $E(Y) = 0,39$. Il faut à présent déterminer $E(XY)$. On a :

$$\begin{aligned} E(XY) &= \sum_{x_i}\sum_{y_j} x_i y_j p_{x_i y_j} \\ &= 0\times 0\times 0,59+0\times 1\times 0,14+0\times 2\times 0,06 \\ &\quad +1\times 0\times 0,09+1\times 1\times 0,05+1\times 2\times 0,02 \\ &\quad +2\times 0\times 0,02+2\times 1\times 0,02+2\times 2\times 0,01 = 0,17 \end{aligned}$$

La covariance entre X et Y vaut donc :

$$Cov(X,Y) = E(XY)-E(X)E(Y) = 0,17-0,26\times 0,39 = 0,0686$$

Evaluons maintenant la variance et l'écart-type de la recette totale $Z = 40X + 20Y$, dont on a déjà calculé que l'espérance $E(Z) = E\left[40X + 20Y\right]$ vaut $18,2$. On a besoin de $V(X)$ et de $V(Y)$. On trouve :

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 = \sum_{x_i} x_i^2 p_{x_i.} - (E(X))^2 \\ &= 0^2 \times 0,79 + 1^2 \times 0,16 + 2^2 \times 0,05 - (0,26)^2 = 0,2924 \end{aligned}$$

$$\begin{aligned} V(Y) &= E(Y^2) - (E(Y))^2 = \sum_{y_j} y_j^2 p_{.y_j} - (E(Y))^2 \\ &= 0^2 \times 0,70 + 1^2 \times 0,21 + 2^2 \times 0,09 - (0,39)^2 = 0,4179 \end{aligned}$$

On a donc finalement :

$$\begin{aligned} V(Z) &= V(40X + 20Y) = 40^2 V(X) + 20^2 V(Y) + (2 \times 40 \times 20) Cov(X,Y) \\ &= 1600 \times 0,2924 + 400 \times 0,4179 + 1600 \times 0,0686 \\ &= 467,84 + 167,16 + 109,76 = 744,76 \\ \sigma(Z) &= \sigma(40X + 20Y) = \sqrt{V(40X + 20Y)} = \sqrt{744,76} = 27,29... \end{aligned}$$

On vérifie qu'en utilisant la loi de Z qui est donnée par :

z_k	0	20	40	60	80	100	120
p_{z_k}	0,59	0,14	0,15	0,05	0,04	0,02	0,01

on a :

$$\begin{aligned} E(Z^2) &= \sum_{z_k} z_k^2 p_{z_k} = 0^2 \times 0,59 + 20^2 \times 0,14 + 40^2 \times 0,15 \\ &\quad +60^2 \times 0,05 + 80^2 \times 0,04 + 100^2 \times 0,02 + 120^2 \times 0,01 \\ &= 1076 \end{aligned}$$

de sorte qu'on trouve de même :

$$\begin{aligned} V(Z) &= E(Z^2) - (E(Z))^2 = 1076 - 18,2^2 = 744,76 \\ \sigma(Z) &= \sqrt{V(Z)} = \sqrt{744,76} = 27,29 \end{aligned}$$

5.4.3.2. Corrélation

On l'a dit, la covariance, ou plus précisément son *signe*, permet de caractériser le *type de dépendance* — une relation positive ou négative — existant entre deux variables aléatoires. On pourrait être tenté d'utiliser la *valeur* de la covariance pour mesurer le *degré de dépendance* entre deux variables aléatoires X et Y. Ce n'est pas une bonne idée car, comme le montre la propriété ci-dessous, la covariance entre X et Y dépend des unités de mesure de X et de Y.

Propriété 5.8 *Soient X et Y deux variables aléatoires quelconques. On a :*

$$Cov(a + bX, c + dY) = bdCov(X, Y)$$

En effet, de la Définition 5.13, on a :

$$Cov(a + bX, c + dY) = E\left[(a + bX - E\left[a + bX\right])\left(c + dY - E\left[c + dY\right)\right]\right)]$$

Posons $W_1 = a + bX - E(a + bX)$ et $W_2 = c + dY - E(c + dY)$. De la Propriété 4.10, on a :

$$\begin{aligned} W_1 &= a + bX - (a + bE(X)) = b(X - E(X)) \\ W_2 &= c + dY - (c + dE(Y)) = d(Y - E(Y)) \end{aligned}$$

de sorte que :

$$Cov(a + bX, c + dY) = E(W_1 W_2) = E\left[bd(X - E(X))(Y - E(Y))\right]$$

soit, en utilisant encore la Propriété 4.10 et la Définition 5.13 :

$$Cov(a + bX, c + dY) = bdE\left[(X - E(X))(Y - E(Y))\right] = bdCov(X, Y)$$

Cette dépendance aux unités de mesure rend difficilement interprétable la valeur de $Cov(X, Y)$: une grande (petite) valeur pour $Cov(X, Y)$ n'est pas un indicateur fiable d'un lien important (peu important) entre X et Y.

Pour contourner ce problème, on utilise une autre mesure, appelée coefficient de corrélation linéaire.

Définition 5.14 *On appelle coefficient de corrélation linéaire ou, par abus de langage, corrélation entre deux variables aléatoires quelconques X et Y la quantité :*

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)}$$

Notons que la corrélation entre X et Y est égale à la covariance entre les versions standardisées de X et de Y. En effet, en utilisant la Propriété 5.8, on obtient :

$$\begin{aligned} Cov\left(\frac{X - E(X)}{\sigma(X)}, \frac{Y - E(Y)}{\sigma(Y)}\right) &= Cov\left(-\frac{E(X)}{\sigma(X)} + \frac{1}{\sigma(X)}X, -\frac{E(Y)}{\sigma(Y)} + \frac{1}{\sigma(Y)}Y\right) \\ &= \frac{1}{\sigma(X)}\frac{1}{\sigma(Y)}Cov(X, Y) = \rho(X, Y) \end{aligned}$$

Remarquons également que $\rho(X, Y) = \rho(Y, X)$, que $\rho(X, X) = 1$ et que $\rho(X, Y)$ a nécessairement le même signe — et donc de ce point de vue la même interprétation — que $Cov(X, Y)$. Le coefficient de corrélation est évidemment à rapprocher de son analogue rencontré en statistique descriptive.

On vérifie aisément que $\rho(X, Y)$ ne dépend pas des unités de mesure choisies.

Propriété 5.9 *Soient X et Y deux variables aléatoires quelconques. On a :*

$$\rho(a+bX, c+dY) = \begin{cases} \rho(X,Y) & si\ bd > 0 \\ -\rho(X,Y) & si\ bd < 0 \end{cases}$$

En effet, de la Définition 5.14 et des Propriétés 5.8 et 4.11, on a :

$$\rho(a+bX, c+dY) = \frac{Cov(a+bX, c+dY)}{\sigma(a+bX)\sigma(c+dY)} = \frac{bdCov(X,Y)}{|b|\sigma(X)|d|\sigma(Y)} = \frac{bd}{|bd|}\rho(X,Y)$$

Une propriété remarquable du coefficient de corrélation linéaire est qu'il est borné : il est toujours compris entre -1 et 1.

Propriété 5.10 *Soient X et Y deux variables aléatoires quelconques. On a :*

$$-1 \leq \rho(X,Y) \leq 1$$

En effet, considérons les versions standardisées de X et Y :

$$X^* = \frac{X - E(X)}{\sigma(X)} \quad \text{et} \quad Y^* = \frac{Y - E(Y)}{\sigma(Y)}$$

On a vu que :

$$\rho(X,Y) = Cov(X^*, Y^*)$$

De la Propriété 5.7, on a :

$$\begin{aligned} V(X^* + Y^*) &= V(X^*) + V(Y^*) + 2Cov(X^*, Y^*) \\ V(X^* - Y^*) &= V(X^*) + V(Y^*) - 2Cov(X^*, Y^*) \end{aligned}$$

soit, comme par définition on a $V(X^*) = V(Y^*) = 1$ et qu'une variance est toujours positive ou nulle :

$$V(X^* + Y^*) = 2\left(1 + \rho(X,Y)\right) \geq 0 \tag{5.12}$$

$$V(X^* - Y^*) = 2\left(1 - \rho(X,Y)\right) \geq 0 \tag{5.13}$$

De la relation (5.12), on tire :

$$1 + \rho(X,Y) \geq 0 \Leftrightarrow \rho(X,Y) \geq -1$$

De même, de la relation (5.13), on a :

$$1 - \rho(X,Y) \geq 0 \Leftrightarrow \rho(X,Y) \leq 1$$

Au total, on a donc :

$$-1 \leq \rho(X,Y) \leq 1$$

Notons au passage que le contrainte $-1 \leq \rho(X,Y) \leq 1$ implique qu'on a toujours :

$$(Cov(X,Y))^2 \leq V(X)V(Y) \Leftrightarrow |Cov(X,Y)| \leq \sigma(X)\sigma(Y)$$

Cette inégalité est connue sous le nom d'*inégalité de* SCHWARZ.

Les deux bornes $\rho(X,Y) = 1$ et $\rho(X,Y) = -1$ du coefficient de corrélation linéaire correspondent à des situations particulières où il existe entre X et Y une relation linéaire exacte.

Propriété 5.11 *Soient X et Y deux variables aléatoires quelconques. On a :*

$$\rho(X,Y) = \begin{cases} 1 & \text{si et seulement si } Y = a + bX \text{ avec } b > 0 \\ -1 & \text{si et seulement si } Y = a + bX \text{ avec } b < 0 \end{cases}$$

En effet, si $Y = a + bX$, de la Définition 5.14, on a :

$$\rho(X,Y) = \rho(X, a+bX) = \frac{Cov(X, a+bX)}{\sigma(X)\sigma(a+bX)}$$

soit, en utilisant les Propriétés 5.8 et 4.11 :

$$\rho(X,Y) = \frac{bCov(X,X)}{\sigma(X)\,|b|\,\sigma(X)} = \frac{bV(X)}{|b|\,V(X)} = \frac{b}{|b|}$$

ce qui est égal à 1 si b est positif et -1 si b est négatif.

Pour montrer la réciproque, considérons à nouveau les versions standardisées de X et Y. On a vu ci-dessus qu'on avait :

$$V(X^* + Y^*) = 2\,(1 + \rho(X,Y)) \qquad (5.14)$$

$$V(X^* - Y^*) = 2\,(1 - \rho(X,Y)) \qquad (5.15)$$

Supposons tout d'abord que $\rho(X,Y) = 1$. Dans ce cas, de la relation (5.15), on a $V(X^* - Y^*) = 0$, ce qui implique que $X^* - Y^*$ soit égal à une constante : $X^* - Y^* = c$. Comme $c = E(c) = E(X^* - Y^*)$ et que $E(X^* - Y^*) = 0$, cette constante est égale à zéro. On a donc :

$$X^* - Y^* = 0 \quad \Leftrightarrow \quad \frac{X - E(X)}{\sigma(X)} - \frac{Y - E(Y)}{\sigma(Y)} = 0$$

soit :

$$Y = \underbrace{E(Y) - \tfrac{\sigma(Y)}{\sigma(X)}E(X)}_{=a} + \underbrace{\tfrac{\sigma(Y)}{\sigma(X)}}_{=b>0} X$$

Supposons maintenant que $\rho(X,Y) = -1$. De façon semblable, de la relation (5.14), on a $V(X^* + Y^*) = 0$, ce qui implique que $X^* + Y^*$ soit égal à une constante : $X^* + Y^* = c'$. Comme $c' = E(c') = E(X^* + Y^*)$ et que $E(X^* + Y^*) = 0$, cette constante est égale à zéro. On a donc :

$$X^* + Y^* = 0 \quad \Leftrightarrow \quad \frac{X - E(X)}{\sigma(X)} + \frac{Y - E(Y)}{\sigma(Y)} = 0$$

soit :

$$Y = \underbrace{E(Y) + \tfrac{\sigma(Y)}{\sigma(X)} E(X)}_{=a} + \underbrace{\left(-\tfrac{\sigma(Y)}{\sigma(X)}\right)}_{=b<0} X$$

Donc, lorsque $\rho(X,Y) = 1$, il y a bien une relation linéaire croissante entre X et Y, et lorsque $\rho(X,Y) = -1$, une relation linéaire décroissante. Notez que les paramètres a et b de ces relations sont entièrement déterminés par $E(X)$, $E(Y)$, $\sigma(X)$ et $\sigma(Y)$.

Comme le suggère la propriété ci-dessus, le coefficient de corrélation mesure le degré de dépendance *linéaire* entre les variables aléatoires X et Y, d'où le nom de coefficient de corrélation *linéaire*. Lorsque $\rho(X,Y)$ est proche de 1, cela signifie que X et Y varient dans le même sens et sont proches de satisfaire une relation linéaire positive exacte. Lorsque $\rho(X,Y)$ est proche de -1, cela signifie que X et Y varient en sens opposés et sont proches de satisfaire une relation linéaire négative exacte. De façon générale, plus $|\rho(X,Y)|$ est proche de 1, plus X et Y sont linéairement dépendants.

Calculons la corrélation entre X et Y pour notre exemple du magasin de téléviseurs. On a déjà calculé que $Cov(X,Y) = 0,0686$, $V(X) = 0,2924$ et $V(Y) = 0,4179$. On en déduit :

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{0,0686}{\sqrt{0,2924}\sqrt{0,4179}} = 0,1962...$$

On dit que X et Y sont positivement corrélées. Lorsque $\rho(X,Y) < 0$, on dit que X et Y sont négativement corrélées.

5.4.3.3. Indépendance et non-corrélation

Deux variables aléatoires X et Y quelconques sont dites non-corrélées lorsque $\rho(X,Y) = Cov(X,Y) = 0$. Remarquez bien que $\rho(X,Y) = 0$ *si et seulement si* $Cov(X,Y) = 0$.

Rappelons au passage que lorsque $Cov(X,Y) = 0$, la variance d'une combinaison linéaire de X et Y prend une forme très simple.

Propriété 5.12 *Si X et Y sont deux variables aléatoires non-corrélées, alors :*

$$V(a + bX + cY) = b^2 V(X) + c^2 V(Y)$$

et donc en particulier :

$$V(X + Y) = V(X) + V(Y) = V(X - Y)$$

Cette propriété découle directement de la Propriété 5.7.

Deux variables aléatoires indépendantes ont une corrélation nulle, sont non-corrélées. Ce résultat se déduit aisément de la propriété suivante.

Propriété 5.13 *Si X et Y sont deux variables aléatoires indépendantes, alors :*

$$E(XY) = E(X)E(Y)$$

En effet, de la Définition 5.3, on a dans le cas discret :

$$E(XY) = \sum_{x_i} \sum_{y_j} x_i y_j p_{x_i y_j}$$

Par définition, lorsque X et Y sont indépendantes, $p_{x_i y_j} = p_{x_i.} p_{.y_j}$, de sorte que :

$$E(XY) = \sum_{x_i} \sum_{y_j} x_i y_j p_{x_i.} p_{.y_j} = \left(\sum_{x_i} x_i p_{x_i.} \right) \left(\sum_{y_j} y_j p_{.y_j} \right) = E(X)E(Y)$$

De même, dans le cas continu, on a :

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x,y)dxdy$$

Or, lorsque X et Y sont indépendantes, $f(x,y) = f_X(x)f_Y(y)$, de sorte qu'à nouveau :

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf_X(x)f_Y(y)dxdy = \left(\int_{-\infty}^{+\infty} xf_X(x)dx \right) \left(\int_{-\infty}^{+\infty} yf_Y(y)dy \right) \\ &= E(X)E(Y) \end{aligned}$$

Propriété 5.14 *Si X et Y sont deux variables aléatoires indépendantes, alors :*

$$\rho(X,Y) = Cov(X,Y) = 0$$

En effet, si X et Y sont indépendantes, de la Propriété 5.13, on a $E(XY) = E(X)E(Y)$, de sorte que de la Propriété 5.6 et la Définition 5.14, on voit que $Cov(X,Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$, et donc aussi $\rho(XY) = 0$.

Attention : *la réciproque de cette propriété est fausse.* Par conséquent, il ne suffit pas de montrer que deux variables aléatoires sont non-corrélées pour démontrer leur indépendance : deux variables aléatoires non-corrélés peuvent très bien être dépendantes. Ceci provient du fait que le coefficient de corrélation $\rho(X,Y)$ ne mesure que la dépendance *linéaire*, et pas d'autres types de dépendance.

Pour s'en convaincre, considérons par exemple le couple de variables aléatoires (X,Y) dont la loi jointe et les lois marginales correspondantes sont données par :

$x_i\backslash\, y_j$	0	25	$p_{x_i.}$
-5	0	$0,25$	$0,25$
0	$0,50$	0	$0,50$
5	0	$0,25$	$0,25$
$p_{.y_j}$	$0,50$	$0,50$	1

Clairement, X et Y ne sont pas indépendantes. On a par exemple :

$$0 = \mathbb{P}(X = -5 \text{ et } Y = 0) \neq \mathbb{P}(X = -5)\mathbb{P}(Y = 0) = 0,25 \times 0,5 = 0,125$$

En fait, on voit facilement qu'il existe une relation *non-linéaire exacte* entre X et Y : $Y = X^2$. Lorsqu'on connaît X, la valeur que prend Y est connue avec certitude et égale à X^2.

X et Y sont cependant non-corrélées. En effet, on a :

$$\begin{aligned} E(XY) &= \sum_{x_i}\sum_{y_j} x_i y_j p_{x_i y_j} = (-5) \times 0 \times 0 + (-5) \times 25 \times 0,25 \\ &\quad +0 \times 0 \times 0,50 + 0 \times 25 \times 0 + 5 \times 0 \times 0 + 5 \times 25 \times 0,25 \\ &= 0 \end{aligned}$$

$$\begin{aligned} E(X) &= \sum_{x_i} x_i p_{x_i.} = (-5) \times 0,25 + 0 \times 0,50 + 5 \times 0,25 = 0 \\ E(Y) &= \sum_{y_j} y_j p_{.y_j} = 0 \times 0,50 + 25 \times 0,50 = 12,5 \end{aligned}$$

de sorte que :

$$Cov(X,Y) = E(XY) - E(X)E(Y) = 0 - 0 = 0 = \rho(X,Y)$$

5.5. Vecteurs de variables aléatoires

On peut aisément généraliser ce que nous venons de faire et passer de 2 variables à n variables aléatoires. On parle alors de vecteurs de variables aléatoires, discrètes ou continues.

Définition 5.15 *Un vecteur de n variables aléatoires discrètes $(X_1, X_2, ..., X_n)$ est un vecteur de n variables aléatoires qui prend un nombre fini, ou infini mais dénombrable, de n-uples de valeurs $(x_{1k}, x_{2l}, ..., x_{nm})$.*

Définition 5.16 *Un vecteur de n variables aléatoires continues $(X_1, X_2, ..., X_n)$ est un vecteur de n variables aléatoires qui prend un continuum de n-uples de valeurs $(x_1, x_2, ..., x_n)$ dans un (ou plusieurs) intervalle(s).*

Ces définitions sont une généralisation directe de celles du cas discret bivarié.

Comme dans le cas bivarié, la loi ou distribution de probabilité jointe d'un vecteur de n variables aléatoires discrètes est définie par un ensemble de couples valeur possible / probabilité. De la même façon, la loi ou distribution de probabilité jointe d'un vecteur de n variables aléatoires continues est définie par l'intermédiaire d'une fonction de densité jointe multivariée. Bref, rien de vraiment nouveau, il s'agit toujours d'une généralisation directe du cas bivarié.

Définition 5.17 *On appelle loi ou distribution de probabilité jointe d'un vecteur de n variables aléatoires discrètes $(X_1, X_2, ..., X_n)$ un ensemble de couples $((x_{1k}, x_{2l}, ..., x_{nm}), p_{x_{1k}x_{2l}...x_{nm}})$, où les $p_{x_{1k}x_{2l}...x_{nm}} = I\!P(X_1 = x_{1k}$ et $X_2 = x_{2l}$ et ... et $X_n = x_{nm})$ sont des nombres réels tels que :*

$$\begin{cases} 0 \leq p_{x_{1k}x_{2l}...x_{nm}} \leq 1, \quad \forall\,(x_{1k}, x_{2l}, ..., x_{nm}) \in \mathcal{X}_1 \times \mathcal{X}_2 \times ... \times \mathcal{X}_n \\ \displaystyle\sum_{x_{1k}} \sum_{x_{2l}} ... \sum_{x_{nm}} p_{x_{1k}x_{2l}...x_{nm}} = 1 \end{cases}$$

$\mathcal{X}_r$ désignant l'ensemble des valeurs possibles $x_{r.}$ de la variable aléatoire X_r et $\sum_{x_{r.}}$ signifiant "somme pour toutes les valeurs possibles $x_{r.}$ de la variable aléatoire X_r" $(r = 1, ..., n)$.

Définition 5.18 *On appelle densité de probabilité jointe d'un vecteur de n variables aléatoires continues $(X_1, X_2, ..., X_n)$ une fonction $f(x_1, x_2, ..., x_n)$ telle que :*

$$\begin{cases} f(x_1, x_2, ..., x_n) \geq 0, \quad \forall\,(x_1, x_2, ..., x_n) \in \mathbb{R}^n \\ \displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} f(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n = 1 \end{cases}$$

et

$$\begin{aligned} &I\!P(a_1 \leq X_1 \leq b_1 \text{ et } a_2 \leq X_2 \leq b_2 \text{ et ... et } a_n \leq X_n \leq b_n) \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} ... \int_{a_n}^{b_n} f(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n \end{aligned}$$

De façon semblable au cas bivarié (discret ou continu), à partir d'une loi jointe multivariée, on peut retrouver les lois marginales de chaque variable aléatoire, déterminer les lois conditionnelles de chaque variable aléatoire sachant une ou plusieurs autres variables, calculer l'espérance et la variance de ces lois, déterminer la loi et calculer l'espérance d'une fonction des n variables aléatoires, calculer la covariance et la corrélation entre ces variables aléatoires prises deux à deux, etc... On peut de même obtenir la loi jointe marginale d'un sous-ensemble des n variables aléatoires, déterminer la loi jointe conditionnelle d'un sous-ensemble des n variables sachant un autre sous-ensemble de ces variables, calculer l'espérance de fonctions des variables intervenant dans ces lois, etc...

Nous n'entrerons pas ici dans ces détails. On se contentera de donner une généralisation de la définition de l'indépendance et des formules permettant de calculer l'espérance et la variance de fonctions linéaires de n variables aléatoires.

Définition 5.19 *n variables aléatoires $X_1, X_2, ..., X_n$ sont indépendantes si et seulement si, dans le cas discret, on a pour toutes les valeurs possibles x_{1k} de X_1, x_{2l} de X_2, ... et x_{nm} de X_n :*

$$p_{x_{1k}x_{2l}...x_{nm}} = p_{x_{1k}} p_{x_{2l}} \times ... \times p_{x_{nm}}$$

où les $p_{x_{r.}} = \mathbb{P}(X_r = x_{r.})$ sont les probabilités définissant les lois marginales de chacune des variables aléatoires X_r $(r = 1, ..., n)$, et dans le cas continu, on a pour toutes les valeurs possibles x_1 de X_1, x_2 de X_2, ... et x_n de X_n :

$$f(x_1, x_2, ..., x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \times ... \times f_{X_n}(x_n)$$

où les $f_{X_r}(x_r)$ sont les densités de probabilité marginales chacune des variables aléatoires X_r $(r = 1, ..., n)$.

Propriété 5.15 *Soient n variables aléatoires $X_1, X_2, ..., X_n$ quelconques. Si $Y = a_0 + \sum_{i=1}^n a_i X_i$, on a :*

$$E(Y) = E\left(a_0 + \sum_{i=1}^n a_i X_i\right) = a_0 + \sum_{i=1}^n a_i E(X_i)$$

et

$$V(Y) = V\left(a_0 + \sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i) + \sum_{\substack{i=1 \\ i \neq j}}^n \sum_{j=1}^n a_i a_j Cov(X_i, X_j)$$

Cette propriété est une généralisation directe des Propriétés 5.5 et 5.7.

Le calcul de la variance d'une fonction linéaire de n variables aléatoires se simplifie évidemment grandement lorsque les covariances sont nulles. C'est en particulier le cas si les variables aléatoires sont indépendantes.

Propriété 5.16 *Si n variables aléatoires $X_1, X_2, ..., X_n$ sont indépendantes, alors :*

$$\rho(X_i, X_j) = Cov(X_i, X_j) = 0, \quad \forall\, i, j = 1, ..., n \ (i \neq j)$$

ce qui implique que :

$$V\left(a_0 + \sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i)$$

5.6. Exercice résolu

On considère une table pouvant accueillir six convives dans un restaurant. On définit les deux variables aléatoires suivantes :

$$\begin{array}{lcl} X & = & \text{le nombre de convives s'installant à cette table} \\ Y & = & \text{le nombre de personnes qui, à cette table, commandent le plat du jour} \end{array}$$

La loi jointe du couple aléatoire (X, Y) est donnée par le tableau suivant.

$x_i \backslash y_j$	0	1	2	3	4	5	6
0	0,01	0	0	0	0	0	0
1	0,01	0,09	0	0	0	0	0
2	0,09	0,02	0,15	0	0	0	0
3	0,06	0,02	0,02	0,02	0	0	0
4	0,08	0,03	0,02	0,03	0,10	0	0
5	0,08	0,02	0	0,02	0,01	0,01	0
6	0,06	0	0,01	0,01	0,02	0	0,01

1- Donnez la loi marginale de X.

En additionnant les probabilités en ligne, on obtient :

x_i	0	1	2	3	4	5	6
$p_{x_i.}$	0,01	0,10	0,26	0,12	0,26	0,14	0,11

2- Calculez l'espérance de X.

On trouve :

$$\begin{aligned} E(X) &= 0 \times 0,01 + 1 \times 0,10 + 2 \times 0,26 + 3 \times 0,12 \\ &\quad +4 \times 0,26 + 5 \times 0,14 + 6 \times 0,11 \\ &= 3,38 \end{aligned}$$

3- Donnez la probabilité d'avoir une commande de deux plats du jour sachant que le nombre de convives est trois.

La probabilité à calculer est :

$$I\!P(Y = 2 | X = 3) = \frac{I\!P(X = 3 \text{ et } Y = 2)}{I\!P(X = 3)} = \frac{0,02}{0,12} = 0,166...$$

4- Calculez la probabilité que tous les convives, quel que soit leur nombre, commandent le plat du jour.

La probabilité demandée est :

$$\begin{aligned} I\!P(X = Y) &= I\!P[(X = 0 \text{ et } Y = 0) \text{ ou } (X = 1 \text{ et } Y = 1) \text{ ou } (X = 2 \text{ et } Y = 2) \\ &\qquad \text{ou } (X = 3 \text{ et } Y = 3) \text{ ou } (X = 4 \text{ et } Y = 4) \\ &\qquad \text{ou } (X = 5 \text{ et } Y = 5) \text{ ou } (X = 6 \text{ et } Y = 6)] \\ &= I\!P(X = 0 \text{ et } Y = 0) + I\!P(X = 1 \text{ et } Y = 1) + I\!P(X = 2 \text{ et } Y = 2) \\ &\quad + I\!P(X = 3 \text{ et } Y = 3) + I\!P(X = 4 \text{ et } Y = 4) + I\!P(X = 5 \text{ et } Y = 5) \\ &\quad + I\!P(X = 6 \text{ et } Y = 6) \\ &= 0,01 + 0,09 + 0,15 + 0,02 + 0,10 + 0,01 + 0,01 = 0,39 \end{aligned}$$

5.7. Exercices

1- Soit deux variables aléatoires X et Y dont la loi jointe est donnée par le tableau suivant :

$x_i \backslash y_j$	0	1	2	3
0	$0,22$	$0,20$	0	$0,02$
1	$0,05$	$0,11$	$0,04$	$0,01$
2	$0,04$	$0,07$	$0,02$	$0,22$

a- Etablissez les lois marginales de X et de Y. Calculez leur espérance et leur variance.
b- Etablissez la loi de X conditionnelle à $Y = 2$. Calculez l'espérance et la variance de cette loi conditionnelle.
c- Calculez la covariance entre X et Y. Calculez le coefficient de corrélation.
d- Les deux variables sont-elles indépendantes ?
e- Etablissez la loi de la variable aléatoire $Z = Y^2 - X$.

2- Le prix (P, en francs) et la quantité vendue (Q, en cartons) d'un bien obéissent à la loi bivariée suivante :

$p_i \backslash q_j$	30	40
6	$0,4$	$0,2$
8	$0,1$	$0,3$

a- Donnez les lois marginales de P et de Q.
b- Donnez la loi de Q conditionnelle à un prix de 8.
c- Calculez $E(P)$ et $E(Q)$.
d- Calculez $Cov(P, Q)$.
e- Calculez le coefficient de corrélation linéaire $\rho(P, Q)$.

3- M. Floréal est un grand amateur de fleurs exotiques. Il possède depuis plusieurs années deux rhododendrons nains de Malaisie. M. Floréal a pu établir un tableau donnant les probabilités pour les nombres de fleurs données par les deux rhododendrons :

		Nbr. de fleurs du premier		
		1	2	3
Nbr. de fleurs	0	$0,1$	$0,2$	$0,3$
du second	1	0	$0,2$	$0,2$

a- Quelle est la probabilité que le second rhododendron ne donne pas de fleurs si le premier en a donné 3 ?
b- Y-a-t-il indépendance entre le nombre de fleurs données par le premier et le nombre de fleurs données par le second ? Justifiez brièvement votre réponse.
c- A quel nombre de fleurs peut-on s'attendre pour le premier rhododendron ?
d- Donnez l'espérance et la variance du nombre total de fleurs données par les deux rhododendrons.

4- Un agriculteur cultive des choux et des pommes de terre. Les choux sont vendus par sacs de 100 à 30 F le sac, les pommes de terre par sacs de 100 kilos à 60 F le sac. L'agriculteur est assuré de pouvoir vendre toute sa récolte à ces prix. Avant la récolte, il évalue les probabilités des quantités vendues comme suit :

Choux (en sacs de 100)	Probabilités
2	$0,2$
3	$0,4$
4	$0,3$
5	$0,1$

Pommes de terres (en sacs de 100 kg)	Probabitités
1	$0,2$
2	$0,5$
3	$0,3$

La récolte des choux et la récolte des pommes de terre sont supposées être des variables aléatoires indépendantes.

a- Calculez l'espérance et la variance de la recette totale que l'agriculteur obtiendra pour la vente de sa récolte.

b- Etablissez la loi jointe des quantités vendues.

5- Le zoo de Washington vient d'acquérir un couple de girafes. Des spécialistes ont pu établir les probabilités suivantes concernant l'effectif et le sexe de l'éventuelle progéniture d'un couple de girafes :

Nbr. de mâles	Nbr. de femelles	Probabilités
0	0	$0,20$
0	1	$0,20$
0	2	$0,10$
1	0	$0,10$
1	1	$0,40$

a- Y-a-t-il indépendance entre le nombre de girafeaux mâles et le nombre de girafeaux femelles donnés par un couple de girafes ?

b- Calculez la covariance entre le nombre de girafeaux mâles et le nombre de girafeaux femelles donnés par un couple de girafes.

c- Quelle est la probabilité qu'un couple ait deux girafeaux femelles, s'il n'a pas de girafeau mâle ?

d- Quelle est la probabilité qu'un couple ait au moins autant de girafeaux mâles que de girafeaux femelles ?

e- Quelles sont l'espérance et la variance du nombre de girafeaux mâles donnés par un couple ?

f- Quelles sont l'espérance et la variance du nombre de girafeaux femelles donnés par un couple ?

g- Quelles sont l'espérance et la variance du nombre total de girafeaux donnés par un couple ?

6- Dans la population des 1 270 diplômés lors du dernier examen cantonal, on a observé la distribution de fréquence jointe suivante pour les notes (/100) obtenues par les élèves à l'épreuve de français et de mathématiques :

		Note en math		
		70	80	90
Note en	70	$0,10$	$0,10$	0
français	80	$0,10$	$0,40$	$0,10$
	90	0	$0,10$	$0,10$

On tire au hasard un élève parmi les diplômés. On appelle X sa note en français et Y sa note en math.

a- Quelle est la loi jointe du couple (X, Y) ?
b- Calculer les lois marginales de X et de Y. Les 2 distributions sont-elles égales ? A quoi correspondent, en termes de fréquence des notes dans la population, ces deux lois ?
c- X et Y sont-elles indépendantes ?
d- Calculer $E(X)$ et $E(Y)$. A quoi correspondent, en termes de fréquence des notes dans la population, ces deux quantités ?
e- Calculer la variance de X et de Y. A quoi correspondent, en termes de fréquence des notes dans la population, ces deux quantités ?
f- Calculez la covariance et le coefficient de corrélation entre les deux notes. A quoi correspondent, en termes de fréquence des notes dans la population, ces deux quantités ?
g- Etablissez la loi de Y sachant $X = 70$. A quoi correspond, en termes de fréquence des notes dans la population, cette loi ?
h- Calculez $E(Y|X = 70)$ et $V(Y|X = 70)$. A quoi correspondent, en termes de fréquence des notes dans la population, ces deux quantités ?
i- On tire au hasard un élève parmi les diplômés ayant obtenu une note de 70 en français et on appelle Y sa note en math.
   i- Quelle est la loi de Y ? A quoi correspond cette loi (cf. point g) ?
   ii- Quelle est l'espérance et la variance de Y ? A quoi correspondent ces quantités (cf. point h) ?

7- La distribution marginale du nombre annuel de sinistres automobiles occasionnés par Madame Chauffard ($= Y$), ainsi que les distributions conditionnelles du nombre annuel de sinistres automobiles occasionnés par Monsieur Chauffard ($= X$) sont données par les tableaux suivants :

y_j	0	1	2
$p_{.y_j}$	$0,25$	$0,50$	$0,25$

x_i	0	1
$p_{x_i\|y_j=0}$	$0,4$	$0,6$
$p_{x_i\|y_j=1}$	$0,5$	$0,5$
$p_{x_i\|y_j=2}$	$0,6$	$0,4$

a- Quelle est la probabilité que M. Chauffard ait cette année un sinistre ?
b- Pensez-vous que le comportement de M. Chauffard au volant ait une influence sur celui de Mme Chauffard ?
c- Si un sinistre occasionné par Mme Chauffard coûte 10 000 F au couple, et qu'un sinistre occasionné par M. Chauffard en coûte 20 000 F, à quels frais de réparation le ménage peut-il s'attendre l'an prochain ?

8- Une urne contient 10 boules, marquées chacune d'une paire de nombres (X, Y). Les paires sont respectivement : $(0, 5)$, $(0, 4)$, $(3, 2)$, $(3, 4)$, $(5, 2)$, $(5, 7)$, $(8, 1)$,

$(10, 4)$, $(10, 4)$, $(10, 5)$. On tire au hasard de cette urne une boule. Soit (X, Y) sa marque, inconnue de l'observateur. Etablissez :

a- la distribution jointe du couple (X, Y).
b- la distribution marginale de la variable X.
c- la distribution de X conditionnellement à $Y = 2$.
d- la distribution de Y conditionnellement à $X = 5$.

9- Sur la base de l'inventaire communal réalisé par l'INSEE en 1979, on peut établir le tableau suivant qui croise, pour les 36 414 communes métropolitaines, la fréquence de ramassage des ordures ménagères, durant l'hiver 1979, avec la taille de la commune.

Fréquence de ramassage	Taille de la commune (nbr. hab.) moins de 500	500 à 2 000	2 000 à 100 000	plus de 100 000	Total
Quotidienne	260	403	807	26	1 496
2, 3 ou 4 fois par semaine	2 622	3 221	2 357	12	8 212
1 fois par semaine	9 352	5 377	335	0	15 064
1 fois par quinzaine ou moins	4 217	704	7	0	4 928
Jamais	6 479	235	0	0	6 714
Total	22 930	9 940	3 506	38	36 414

On appelle X la fréquence (codée 1, 2, 3, 4 et 5, par ordre décroissant de fréquence) de ramassage des ordures et Y la taille (codée 1, 2, 3 et 4, par ordre croissant de taille) d'une commune tirée au sort.

a- Donnez la loi jointe du couple (X, Y).
b- Donnez la loi conditionnelle de X sachant $Y = 3$.

10- Dans un jeu de hasard on utilise une paire de dés (équilibrés) ayant chacun deux faces portant le nombre 1, deux faces portant le nombre 2 et deux faces portant le nombre 3, de sorte que chaque dé ne donne que 3 résultats ayant chacun la même probabilité $2/6 = 1/3$. Les 2 dés sont lancés de façon indépendante. On définit X la variable aléatoire représentant le nombre obtenu avec le premier dé et Y la variable aléatoire représentant le nombre obtenu avec le deuxième dé.

a- Donner la distribution jointe du couple (X, Y).
b- Déterminer la loi de $S = X + Y$.
c- Calculer l'espérance, la variance et l'écart-type de X, de Y et de S.
d- S correspond exactement à la somme de X et Y. L'espérance de S est-elle alors aussi la somme des deux espérances ? Est-ce toujours le cas ? La variance est-elle simplement la somme des deux variances ? Est-ce toujours le cas ? Et l'écart-type est-il simplement la somme des deux écart-types ?

11- Pour essayer de savoir à quoi selon eux tient le bonheur, on a demandé à un groupe de gens de se classer eux-mêmes en fonction de "l'indice de bonheur" suivant : 0 = malheureux, 1 = moyennement heureux ou 2 = très heureux. En même temps, on a relevé pour chaque individu le revenu annuel du ménage (en milliers de \$). Les fréquences observées des différentes combinaisons "indice de bonheur" / revenu annuel du ménage se présentaient en gros de la façon suivante :

		Indice de bonheur		
		0	1	2
	2, 5	0, 03	0, 12	0, 07
Revenu du	7, 5	0, 02	0, 13	0, 11
ménage	12, 5	0, 01	0, 13	0, 14
	17, 5	0, 01	0, 09	0, 14

On considère un individu du groupe pris au hasard. On appelle H "l'indice de bonheur" et X le revenu du ménage de l'individu tiré.

a- Calculer $E(H|X)$ pour les différentes valeurs de X.
b- Calculer la covariance et le coefficient de corrélation entre X et H.
c- les variables aléatoires X et H sont-elles indépendantes ? Répondez sans rien calculer de nouveau.

12- Une entreprise de vente par correspondance analyse l'effet des lettres de relance sur les commandes des clients ayant reçu depuis plus de six semaines le catalogue automne-hiver. Le service d'études marketing divise la clientèle en quatre groupes : le premier groupe reçoit une seule lettre de relance, le second reçoit deux lettres à un mois d'intervalle, le troisième reçoit trois lettres, également espacées d'un mois. Le dernier groupe est un groupe-témoin qui ne reçoit aucune lettre de relance. Les fréquences observées des différentes combinaisons nombre de lettres de relance envoyées au client / nombre de commandes de ce client sont données par le tableau suivant :

		Nbr. de lettres de relance			
		0	1	2	3
	0	0, 10	0, 12	0, 02	0, 01
Nbr. de	1	0, 06	0, 16	0, 02	0, 01
commandes	2	0, 05	0, 15	0, 04	0, 01
	3	0, 04	0, 10	0, 08	0, 03

On considère un client pris hasard. On appelle X le nombre de lettres de relance qu'il a reçu et Y le nombre de commandes qu'il a passé.

a- Etablissez la loi de X et calculez son espérance.
b- Etablissez la loi de Y et calculez son espérance.
c- Les variables aléatoires X et Y sont-elles indépendantes ?
d- Calculer $E(Y|X)$ pour les différentes valeurs possibles de X.
e- Les coûts occasionnées par l'envoi de X lettres de relance à un client sont les suivants : outre les frais postaux ($2,5$ F par lettre), il y a un coût fixe de 15 F lié à la gestion du fichier informatique (mise à jour des adresses, etc...) et un coût variable (cadeaux, tombolas, etc...) qu'on estime proportionnel au carré du nombre de lettres envoyées (avec un coefficient de $1,2$). D'autre part, on constate que le montant des commandes décroît avec leur nombre de telle sorte que la recette R attendue de l'envoi de X lettres à un client est donnée par : $E(R|X) = 120E(Y|X) - 5[E(Y|X)]^2$.
L'entreprise est persuadée de l'efficacité de la politique de lettres de relance mais elle s'interroge sur le nombre de lettres à envoyer par client. Elle hésite entre 1 et 2 lettres. Que lui recommandez-vous ?

13- Soit les deux variables aléatoires suivantes :

X = le nombre de vérifications de l'état des pneus (sur une année)
Y = le nombre de crevaisons (durant la même période)

On considère que, pour un conducteur pris au hasard, la loi jointe du couple (X, Y) est donnée par le tableau suivant :

$x_i \backslash y_j$	0	1	2
1	$0,040$	$0,065$	$0,045$
2	$0,162$	$0,081$	$0,027$
3	$0,252$	$0,072$	$0,036$
4	$0,176$	$0,022$	$0,022$

a- Quelle est la probabilité qu'un conducteur pris au hasard subisse deux crevaisons ?

b- Etablissez la loi marginale de X. Calculez son espérance et sa variance.

c- On constate qu'un conducteur pris au hasard a vérifié trois fois l'état de ses pneus.

i- Quelle est la probabilité qu'il ait subi au moins une crevaison ?

ii- A quel nombre de crevaisons peut-on s'attendre qu'il ait dû faire face ?

d- On suppose qu'un garagiste réclame 80 F pour une vérification de l'état des pneus et 300 F pour réparer un pneu crevé. Donnez l'espérance et la variance du montant des dépenses d'un conducteur pris au hasard pour l'entretien et la réparation des pneus de son véhicule.

14- Les actions composant le portefeuille d'un grand investisseur sont classées en considérant le gain (ou la perte) qu'elles ont occasionné durant les douze derniers mois et l'agent de change qui a recommandé leur achat.

	Agent			
Gain	A	B	C	Total
$> 20\%$	1	2		6
10% à 20%	13	12		45
0% à 10%	20	20		60
-10% à 0%	2	2		20
$< -10\%$				
Total	48	36	59	143

a- Complétez le tableau.

b- On considère un action du portefeuille tirée au sort. On appelle X le gain (ou la perte) de l'action (codé 1, 2, 3, 4 et 5, par ordre croissant de gain) et Y l'agent qui a recommandé son achat (codé 1, 2 et 3 pour respectivement l'agent A, B et C).

i- Convertissez le tableau d'effectifs en tableau donnant la loi jointe du couple (X, Y).

ii- Convertissez le tableau d'effectifs en tableaux donnant les lois conditionnelles de X pour les différentes valeurs possibles de Y.

iii- Donnez la probabilité que ce soit l'agent C qui ait recommandé cette action sachant que l'action a rapporté plus de 20%.

iv- Donnez la probabilité que cette action ait perdu de la valeur.
v- On constate que l'action a été recommandée par l'agent C. Quelle est la probabilité qu'elle ait occasionné un gain ?

15- Monsieur Durand est responsable de l'entretien des machines d'une salle informatique de l'université. Il constate qu'au cours des heures de libre service de la salle, les étudiants ont l'habitude d'y apporter des disquettes de jeux et qu'ils y laissent parfois des virus informatiques. Monsieur Durand décide d'effectuer des inspections de routine des machines de la salle. Faute de temps, il n'en n'inspecte jamais plus de trois lors d'une inspection. On appelle X le nombre de machines vérifiées et Y le nombre de machines sur lesquelles un virus est détecté lors d'une inspection quelconque. La loi jointe du couple (X, Y) est décrite par le tableau ci-dessous.

$x_i \backslash y_j$	0	1	2	3
1	$0,15$	$0,10$	0	0
2	$0,05$	$0,09$	$0,06$	0
3	$0,10$	$0,11$	$0,19$	$0,15$

a- Lors d'une inspection quelconque :
   i- Quelle est la probabilité que Monsieur Durand ne détecte aucun virus ?
   ii- Combien de virus Monsieur Durand peut-il s'attendre à détecter ?
   iii- Quelle est la probabilité que Monsieur Durand détecte un virus sur chacune des machines inspectées ?

b- Aujourd'hui, Monsieur Durand décide d'inspecter trois machines.
   i- Quelle est la probabilité qu'aucune machine ne soit infectée par un virus ?
   ii- A combien de machines infectée par un virus peut-il s'attendre ?

c- Les variables X et Y sont-elles indépendantes ? Justifiez votre réponse.

d- La vérification d'une machine prend 10 minutes. Si un virus y est détecté, 15 minutes supplémentaires sont nécessaires pour "nettoyer" la machine. Monsieur Durand commence toujours son travail d'inspection à 17 heures. A quelle heure peut-on espérer qu'il termine une inspection quelconque ?

16- Un enquêteur est chargé d'interviewer des enseignants en vue d'une étude sociologique. Il se présente le soir à leur domicile et leur propose de répondre à ses questions. Il arrive assez souvent que les gens refusent de répondre de telle sorte que l'enquêteur ne sait jamais précisément combien de questionnaires il va pouvoir remplir. On note X le nombre de personnes à qui, lors d'une soirée quelconque, l'enquêteur propose de participer à l'enquête et Y le nombre de personnes qui acceptent de répondre à ses questions. Par expérience, l'enquêteur a pu établir le tableau suivant qui donne la loi jointe de (X, Y).

$x_i \backslash y_j$	0	1	2	3
0				
1	$0,08$	$0,12$		
2	$0,06$	$0,15$	$0,09$	
3	$0,03$	$0,12$	$0,20$	$0,03$

a- Complétez le tableau ci-dessus.

b- Quelle est la probabilité que, lors d'une soirée quelconque :

i- l'enquêteur se présente chez deux personnes ?

ii- l'enquêteur se présente chez deux personnes et que ces deux personnes acceptent de répondre à ses questions ?

iii- deux personnes acceptent de répondre aux questions de l'enquêteur ?

c- Combien de questionnaires peut-on espérer que l'enquêteur remplisse lors d'une soirée quelconque ?

d- A combien de questionnaires complétés l'enquêteur doit-il s'attendre si, lors d'une soirée donnée, il se présente chez trois personnes ?

e- L'enquêteur a pu remplir un questionnaire sur la soirée. Quelle est la probabilité qu'il se soit présenté chez au moins deux personnes ?

17- Lorsque Madame Dumur prépare des crêpes, elle ne sait jamais exactement quelle quantité de pâte il faut préparer et obtient donc une nombre de crêpes qui peut être considéré comme une variable aléatoire ($= X$). D'autre part, elle est assez distraite et laisse parfois brûler un certain nombre de crêpes ($= Y$). La loi jointe du couple (X, Y) est donnée dans le tableau suivant.

$x_i \backslash y_j$	0	1	2	3
1	$0,08$	$0,02$	0	0
2	$0,14$	$0,09$	$0,03$	0
3	$0,21$	$0,10$	$0,04$	$0,03$
4	$0,14$	$0,08$	$0,03$	$0,01$

Madame Dumur vient de décider qu'elle fera des crêpes demain.

a- Combien de crêpes peut-on espérer qu'elle prépare ?

b- A quel nombre de crêpes brûlées peut-on s'attendre ?

c- Quelle est la probabilité que Madame Dumur brûle toutes les crêpes qu'elle prépare ?

d- Il s'avère que Madame Dumur a finalement préparé 2 crêpes.

i- Quelle est la probabilité qu'elle en ait brûlé une ?

ii- A quel nombre de crêpes brûlées peut-on s'attendre ?

18- Le Président de la Société Speed est un homme très occupé et a du mal à gérer son emploi du temps. Appelant X le nombre de rendez-vous qu'il accorde sur une demi-journée et Y le nombre de ses rendez-vous où il est en retard, sa secrétaire a établi le tableau suivant donnant la loi jointe de (X, Y).

$x_i \backslash y_j$	0	1	2
1	$0,04$	$0,05$	$0,00$
2	$0,06$	$0,15$	$0,08$
3	$0,05$	$0,14$	$0,22$
4	$0,04$	$0,10$	$0,07$

a- Quelle est la probabilité que demain matin le Président soit en retard à deux de ses rendez-vous ?

b- Etablissez la loi de Y. A combien de rendez-vous peut-on s'attendre qu'il soit en retard demain matin ?

c- Le président a 3 rendez-vous cet après-midi. Quelle est la probabilité qu'il n'ait aucun retard ? A quel nombre de retards peut-on s'attendre ?

d- Les variables X et Y sont-elles indépendantes ? Pourquoi ?

Chapitre 6

Lois discrètes usuelles

On peut faire dépendre d'un (ou plusieurs) paramètre(s) la distribution de probabilité d'une variable aléatoire discrète, ou de façon semblable, la fonction de densité d'une variable aléatoire continue. On a alors affaire à une *loi paramétrique*, loi paramétrique qui en fait définit une *famille de lois*, chaque valeur possible du (ou des) paramètre(s) définissant une loi particulière au sein de la famille de loi considérée.

Il existe de nombreuses lois paramétriques "prêtes à l'emploi", représentant des épreuves aléatoires "types", plus ou moins utiles selon le problème qu'on souhaite traiter. Dans le cadre de ce cours, nous nous contenterons d'en présenter les plus importantes, celles dont l'usage est le plus répandu dans les applications pratiques.

Dans ce chapitre nous en étudierons quatre, toutes discrètes et univariées (la loi discrète uniforme, la loi de Bernoulli, la loi binomiale et la loi de Poisson). Dans le chapitre suivant, nous en étudierons sept autres, six continues univariées (la loi continue uniforme, la loi normale, la loi du khi-carré, la loi de Student, la loi de Fisher-Snedecor et la loi exponentielle) et une continue bivariée (la loi normale bivariée).

6.1. La loi discrète uniforme

On considère une urne contenant 10 boules numérotées de 0 à 9. On tire une boule au hasard et on note le numéro. On a donc une variable aléatoire X dont l'ensemble des valeurs possibles x_i est $\mathcal{X} = \{0, 1, 2, ..., 9\}$ et dont la probabilité p_{x_i} de chacune de ces valeurs possibles est égale à $\frac{1}{10} = 0,1$:

x_i	0	1	2	3	4	5	6	7	8	9
p_{x_i}	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$

On dit que la loi de X est *uniforme*.

Définition 6.1 *Une variable aléatoire X suit une loi discrète uniforme si :*

$$\mathcal{X} = \{x_1, x_2, ..., x_n\} \quad (n \text{ fini})$$

et

$$\mathbb{P}(X = x_i) = \frac{1}{n}, \quad \forall\, x_i \in \mathcal{X}$$

On note : $X \rightsquigarrow \mathcal{U}(x_1; x_2; ...; x_n)$.

En clair, une variable aléatoire X suit une loi discrète uniforme si elle prend un ensemble fini quelconque de valeurs $x_1, x_2, ..., x_n$ et que ces différentes valeurs possibles sont équiprobables, donc ont toutes une probabilité égale à $1/n$. On note que seul le nombre n (fini) de valeurs possibles détermine les probabilités qui, comme il se doit, sont toujours comprises entre 0 et 1 et dont la somme est bien égale à 1.

Généralement, on considère des lois discrètes uniformes pour des variables prenant les valeurs entières appartenant à un intervalle donné $[a, b]$, où a et b sont eux-mêmes deux nombres entiers.

6.2. La loi de Bernoulli

Une *épreuve de* Bernoulli est une épreuve aléatoire ayant seulement deux résultats possibles, souvent appelés *succès* et *échec*. En assignant, par convention, la valeur 1 au succès de l'épreuve aléatoire et 0 à l'échec de cette épreuve, on obtient une variable aléatoire X ayant deux valeurs possibles : 0 et 1. Une telle variable aléatoire est appelée *variable aléatoire de* Bernoulli, *variable aléatoire indicatrice* ou encore *variable aléatoire binaire*. Par exemple, on interroge au hasard une personne dans la rue et on lui demande si elle a déjà consommé telle marque de yaourt. On associe la valeur 1 à une réponse positive et 0 à une réponse négative.

Si la variable aléatoire X ainsi définie prend la valeur 1 avec une certaine probabilité conventionnellement notée p, qui est bien entendu comprise entre 0 et 1, comme le somme des probabilités des diverses valeurs possibles doit être égale à 1, elle prendra évidemment la valeur 0 avec une probabilité $q = 1 - p$. Sous la forme d'un tableau :

x_i	0	1
p_{x_i}	$q = 1 - p$	p

De façon plus formelle, on a la définition suivante.

Définition 6.2 *Une variable aléatoire X suit une loi de* Bernoulli *(ou encore loi indicatrice) de paramètre p, où $0 \leq p \leq 1$, si :*

$$\mathcal{X} = \{0, 1\}$$

et

$$\mathbb{P}(X = x_i) = p^{x_i}(1 - p)^{1 - x_i} = p^{x_i} q^{1 - x_i}, \quad \forall\, x_i \in \mathcal{X} \tag{6.1}$$

On note : $X \rightsquigarrow \mathcal{B}(p)$.

On vérifie que pour x_i égal à 0 et 1, la formule (6.1) donne bien :

$$\mathbb{P}(X=0) = p^0(1-p)^{1-0} = 1-p = q \quad \text{et} \quad \mathbb{P}(X=1) = p^1(1-p)^{1-1} = p$$

Remarquons que la loi d'une variable aléatoire de Bernoulli est entièrement déterminée par un seul paramètre : la probabilité p d'obtenir la valeur 1.

On peut facilement calculer l'espérance et la variance d'une variable aléatoire de Bernoulli $X \rightsquigarrow \mathcal{B}(p)$. Par application de la Définition 4.5 et des Propriétés 4.7 et 4.8 (ou de façon équivalente de la Propriété 4.1), on a :

$$\begin{aligned} E(X) &= \sum_{x_i} x_i p_{x_i} = 0 \times (1-p) + 1 \times p = p \\ E(X^2) &= \sum_{x_i} x_i^2 p_{x_i} = 0^2 \times (1-p) + 1^2 \times p = p \end{aligned}$$

et donc :

$$V(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p) = pq$$

Propriété 6.1 *Si $X \rightsquigarrow \mathcal{B}(p)$, alors :*

$$E(X) = p \quad et \quad V(X) = p(1-p) = pq$$

6.3. La loi binomiale

6.3.1. Préliminaire : permutations, arrangements et combinaisons

Avant d'aborder la loi binomiale, il convient de faire un petit rappel d'analyse combinatoire.

6.3.1.1. Permutations

On s'intéresse à la question suivante : à combien de résultats différents peut donner lieu le classement de tous les candidats d'un concours auquel participent n individus ? Exprimé en d'autres termes, combien de *suites ordonnées* différentes de n objets, *sans répétition* du même objet, peut-on former avec n objets ?

On appelle ce nombre le nombre de *permutations* de n objets, et on le note P_n.

Le premier élément de la suite peut être choisi de n façons différentes (chacun des n objets). Pour chacun de ces choix, le deuxième élement de la suite peut être choisi de $n-1$ façons différentes (chacun des $n-1$ objets restants). Pour chacun de ces deux premiers choix, le troisième élement de la suite peut être choisi de $n-2$ façons différentes (chacun des $n-2$ objets restants), etc..., jusqu'au n-ième et dernier élément de la suite, qui ne peut plus être choisi que d'une seule façon (le dernier objet restant). On a donc :

$$P_n = n \times (n-1) \times (n-2) \times ... \times 2 \times 1 = n!$$

Ce produit est représenté par $n!$, qui se lit “factorielle de n”. Il est important de noter que, par convention, factorielle de zéro vaut 1 : $0! = 1$.

6.3.1.2. Arrangements

On s'intéresse maintenant à la question : à combien de résultats différents peut donner lieu le classement des k premiers lauréats d'un concours auquel participent n individus ($n \geq k$) ? Exprimé en d'autres termes, combien de *suites ordonnées* différentes de k objets, *sans répétition* du même objet, peut-on former à partir de n objets ($n \geq k$) ?

On appelle ce nombre le nombre d'*arrangements* de n objets pris k par k, et on le note A_n^k.

Le raisonnement est identique à celui développé pour établir la formule de P_n. Le premier élément de la suite peut être choisi de n façons différentes (chacun des n objets). Pour chacun de ces choix, le deuxième élement de la suite peut être choisi de $n-1$ façons différentes (chacun des $n-1$ objets restants). Pour chacun de ces deux premiers choix, le troisième élement de la suite peut être choisi de $n-2$ façons différentes (chacun des $n-2$ objets restants), etc..., jusqu'au k-ième et dernier élément de la suite, qui peut cette fois encore être choisi de $n-k+1$ façons différentes (chacun des $n-k+1$ objets restants). On a donc :

$$\begin{aligned} A_n^k &= \underbrace{n \times (n-1) \times (n-2) \times ... \times (n-k+1)}_{=k \text{ facteurs}} \\ &= \frac{n \times (n-1) \times (n-2) \times ... \times (n-k+1) \times (n-k) \times ... \times 2 \times 1}{(n-k) \times ... \times 2 \times 1} \\ &= \frac{n!}{(n-k)!} \end{aligned}$$

Notons que les permutations sont un cas particulier des arrangements. On a en effet :

$$A_n^n = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n! = P_n$$

6.3.1.3. Combinaisons

Considérons finalement la question suivante : combien de groupes d'individus différents peuvent former les k premiers lauréats d'un concours auquel participent n individus $(n \geq k)$? Exprimé en d'autres termes, combien de *suites non ordonnées* différentes de k objets, *sans répétition* du même objet, peut-on former à partir de n objets $(n \geq k)$?

On appelle ce nombre le nombre de *combinaisons* de n objets pris k par k, et on le note C_n^k.

On sait déjà que le nombre de suites ordonnées différentes de k objets, sans répétition du même objet, qu'on peut former à partir de n objets est donné par :

$$A_n^k = \frac{n!}{(n-k)!}$$

A chaque suite non ordonnée de k objets correspond $P_k = k!$ suites ordonnées des mêmes k objets. Le nombre de suites ordonnées étant A_n^k, le nombre de suites non ordonnées est donc égal à :

$$C_n^k = \frac{A_n^k}{P_k} = \frac{n!}{(n-k)!k!}$$

On notera que :

$$C_n^{n-k} = \frac{n!}{(n-(n-k))!(n-k)!} = \frac{n!}{k!(n-k)!} = C_n^k$$

et que :

$$C_n^n = \frac{n!}{(n-n)!n!} = \frac{n!}{0!n!} = 1 \quad \text{et} \quad C_n^0 = \frac{n!}{(n-0)!0!} = \frac{n!}{n!0!} = 1$$

6.3.2. Processus de Bernoulli et loi binomiale

Il arrive souvent qu'on ne s'intéresse pas à une seule épreuve de BERNOULLI, mais à une suite, une répétition d'épreuves de BERNOULLI identiques se déroulant indépendamment les unes des autres. On parle alors de *processus de* BERNOULLI. Par exemple :

1- On lance 20 fois un dé. A chacun des lancers, on note la valeur 1 si le résultat de dé est pair et la valeur 0 s'il est impair. On obtient ainsi une suite de 1 et de 0 correspondant à la suite des résultats (pair ou impair) des 20 lancers du dé.
2- On interroge 100 personnes choisies au hasard et de façon indépendante à l'entrée d'un supermarché pour savoir si elles ont déjà consommé telle marque de yaourt. On obtient ainsi une suite de 1 et de 0 correspondant à la suite des 100 réponses

positives ou négatives des personnes interrogées.

Dans chacun de ces deux exemples, la suite d'épreuves de BERNOULLI considérée peut être représentée par une suite de variables aléatoires indépendantes, toutes distribuées selon une même loi de BERNOULLI $\mathcal{B}(p)$.

Définition 6.3 *n variables aléatoires $X_1, X_2, ..., X_n$ suivent un processus de* BERNOULLI *si elles sont indépendantes et si elles suivent toutes une même loi de* BERNOULLI *$\mathcal{B}(p)$.*

Lors d'une répétition d'épreuves de BERNOULLI, on est souvent intéressé par le nombre d'épreuves ayant donné la valeur 1. Ainsi, dans l'exemple des personnes intérrogées à l'entrée du supermarché, on souhaitera savoir le nombre de personnes ayant déjà goûté le yaourt auquel on s'intéresse. Comme la variable aléatoire de BERNOULLI X_i associée à chacune des n épreuves vaut 0 ou 1, il suffit de considérer la somme $X = X_1 + X_2 + ... + X_n$: celle-ci donne le nombre de 1 rencontrés dans l'ensemble du processus de BERNOULLI. On dit que X est une *variable aléatoire binomiale.*

Définition 6.4 *Si $X_1, X_2, ..., X_n$ forment un processus de* BERNOULLI*, c'est-à-dire si $X_1, X_2, ..., X_n$ sont n variables aléatoires indépendantes suivant toutes une même loi de* BERNOULLI *$\mathcal{B}(p)$, alors la somme :*

$$X = X_1 + X_2 + ... + X_n$$

est appelée variable aléatoire binomiale.

On peut également proposer une définition moins formelle.

Définition 6.5 *La variable aléatoire X donnant le nombre de 1 rencontrés lors de n répétitions indépendantes d'une épreuve aléatoire de* BERNOULLI *est appelée variable aléatoire binomiale.*

On peut aisément déterminer la loi de X qu'on appelle alors *loi binomiale* : c'est la loi de la variable aléatoire X définie par la fonction $g(X_1, X_2, ..., X_n) = X_1 + X_2 + ... + X_n$ du n-uple de variables aléatoires $(X_1, X_2, ..., X_n)$.

De façon générale, suivant la même logique que celle développée à la Section 5.4.1 pour le cas d'une fonction de deux variables aléatoires discrètes, la loi d'une variable aléatoire X définie par une fonction $g(X_1, X_2, ..., X_n)$ d'un n-uple de variables aléatoires discrètes $(X_1, X_2, ..., X_n)$ est donnée par un ensemble de couples (x_i, p_{x_i}), où les x_i désignent les valeurs possibles de X et les $p_{x_i} = \mathbb{P}(X = x_i)$ les probabilités de ces valeurs possibles, probabilités qui sont données par :

$$\mathbb{P}(X = x_i) = \sum_{(x_{1k}, x_{2l}, ..., x_{nm}) : g(x_{1k}, x_{2l}, ..., x_{nm}) = x_i} \mathbb{P}(X_1 = x_{1k} \text{ et } X_2 = x_{2l} \text{ et } ... \text{ et } X_n = x_{nm})$$

où $\sum_{(x_{1k}, x_{2l}, ..., x_{nm}) : g(x_{1k}, x_{2l}, ..., x_{nm}) = x_i}$ signifie "somme pour tous les n-uples de valeurs possibles $(x_{1k}, x_{2l}, ..., x_{nm})$ de $(X_1, X_2, ..., X_n)$ tels que $g(x_{1k}, x_{2l}, ..., x_{nm}) = x_i$".

Dans le cas qui nous occupe, l'ensemble des n-uples de valeurs possibles $(x_{1k}, x_{2l}, ..., x_{nm})$ de $(X_1, X_2, ..., X_n)$ est constitué de toutes les suites possibles de n valeurs 0 ou 1. Ces différents n-uples de valeurs possibles sont au nombre de 2^n. Clairement, à ces 2^n n-uples de valeurs possibles correspondent, pour la variable aléatoire $X = X_1 + X_2 + ... + X_n$, l'ensemble des $n+1$ valeurs possibles :

$$\mathcal{X} = \{0, 1, 2, ..., n\}$$

Pour chaque valeur possible x_i de X, la probabilité $I\!P(X = x_i)$ est égale à la somme des probabilités des n-uples de valeurs possibles $(x_{1k}, x_{2l}, ..., x_{nm})$ du processus de Bernoulli pour lesquels $X = x_i$, c'est-à-dire la somme des probabilités de toutes les suites possibles comportant x_i fois la valeur 1 et $n - x_i$ fois la valeur 0.

Calculons la probabilité de la suite particulière : 1 1 ... 1 1 0 0 ... 0 0, commençant par x_i valeurs 1 et se terminant par $n - x_i$ valeurs 0. Comme chacune des variables aléatoires X_i du processus de Bernoulli sont indépendantes, la probabilité de cette suite particulière est $p^{x_i}(1-p)^{n-x_i} = p^{x_i}q^{n-x_i}$. C'est aussi la probabilité d'une suite quelconque comportant x_i fois la valeur 1 et $n - x_i$ fois la valeur 0. La probabilité $I\!P(X = x_i)$ est donc égale à $p^{x_i}q^{n-x_i}$ fois le nombre de ces suites. Ce nombre est égal au nombre de façons de choisir x_i positions pour la valeur 1 parmi les n positions disponibles, ce qui n'est autre que le nombre de combinaisons possibles de n éléments pris x_i par x_i :

$$C_n^{x_i} = \frac{n!}{(n - x_i)!x_i!}$$

Au total, on a donc :

$$I\!P(X = x_i) = C_n^{x_i} p^{x_i} q^{n-x_i} = \frac{n!}{(n - x_i)!x_i!} p^{x_i}(1-p)^{n-x_i}$$

Ces probabilités sont entièrement déterminées par la connaissance de la valeur de deux paramètres : n et p. On dit que X suit une loi binomiale de paramètres n et p, et on note $X \rightsquigarrow \mathcal{B}(n; p)$. Remarquons que n est un nombre entier strictement positif tandis que p est une probabilité, c'est-à-dire un nombre compris entre 0 et 1.

Définition 6.6 *Une variable aléatoire X suit une loi binomiale de paramètres n et p, où $n \in \mathbb{N}^+$ et $0 \leq p \leq 1$, si :*

$$\mathcal{X} = \{0, 1, 2, ..., n\}$$

et

$$I\!P(X = x_i) = C_n^{x_i} p^{x_i} q^{n-x_i} = \frac{n!}{(n - x_i)!x_i!} p^{x_i}(1-p)^{n-x_i}, \quad \forall\, x_i \in \mathcal{X}$$

On note : $X \rightsquigarrow \mathcal{B}(n; p)$.

Si on se rappelle la formule du binôme de Newton :

$$(x+y)^n = \sum_{k=0}^{n} C_n^k x^k y^{n-k}, \quad \forall\, n \in \mathbb{N}^+$$

on vérifie aisément que, comme il se doit, les probabilités $p_{x_i} = \mathbb{P}(X = x_i)$ sont bien, $\forall\, x_i \in \mathcal{X}$, comprises entre 0 et 1, et que leur somme est bien égale à 1 :

$$\sum_{x_i} p_{x_i} = \sum_{x_i=0}^{n} C_n^{x_i} p^{x_i} q^{n-x_i} = (p+q)^n = 1^n = 1$$

Les probabilités $\mathbb{P}(X = x_i) = C_n^{x_i} p^{x_i} q^{n-x_i}$ ne sont pas très difficiles à calculer, à tout le moins lorsqu'on dispose d'une calculatrice performante ou d'un ordinateur. Pour des valeurs de n pas trop grandes, elles sont reprises dans des tables[5], qu'il est parfois commode d'utiliser. Une de ces tables est reproduite à la fin de ce chapitre, à la Section 6.7. Lorsque n est grand (et p suffisamment petit), si on dépasse la capacité de calcul dont on dispose, on utilise souvent des approximations. Nous y reviendrons plus tard.

Comme une variable binomiale est la somme de n variables aléatoires $X_i \rightsquigarrow \mathcal{B}(p)$, il est très facile d'en calculer l'espérance. De la Propriété 5.15 et du fait que pour tout i, $E(X_i) = p$, on a :

$$\begin{aligned} E(X) &= E(X_1 + X_2 + ... + X_n) \\ &= E(X_1) + E(X_2) + ... + E(X_n) \\ &= p + p + ... + p = np \end{aligned}$$

Les n variables aléatoires $X_1, X_2, ..., X_n$ étant par ailleurs aussi indépendantes, il est également très facile de calculer la variance de X. De la Propriété 5.16 et du fait que pour tout i, $V(X_i) = p(1-p) = pq$, on a :

$$\begin{aligned} V(X) &= V(X_1 + X_2 + ... + X_n) \\ &= V(X_1) + V(X_2) + ... + V(X_n) \\ &= pq + pq + \cdots + pq = npq \end{aligned}$$

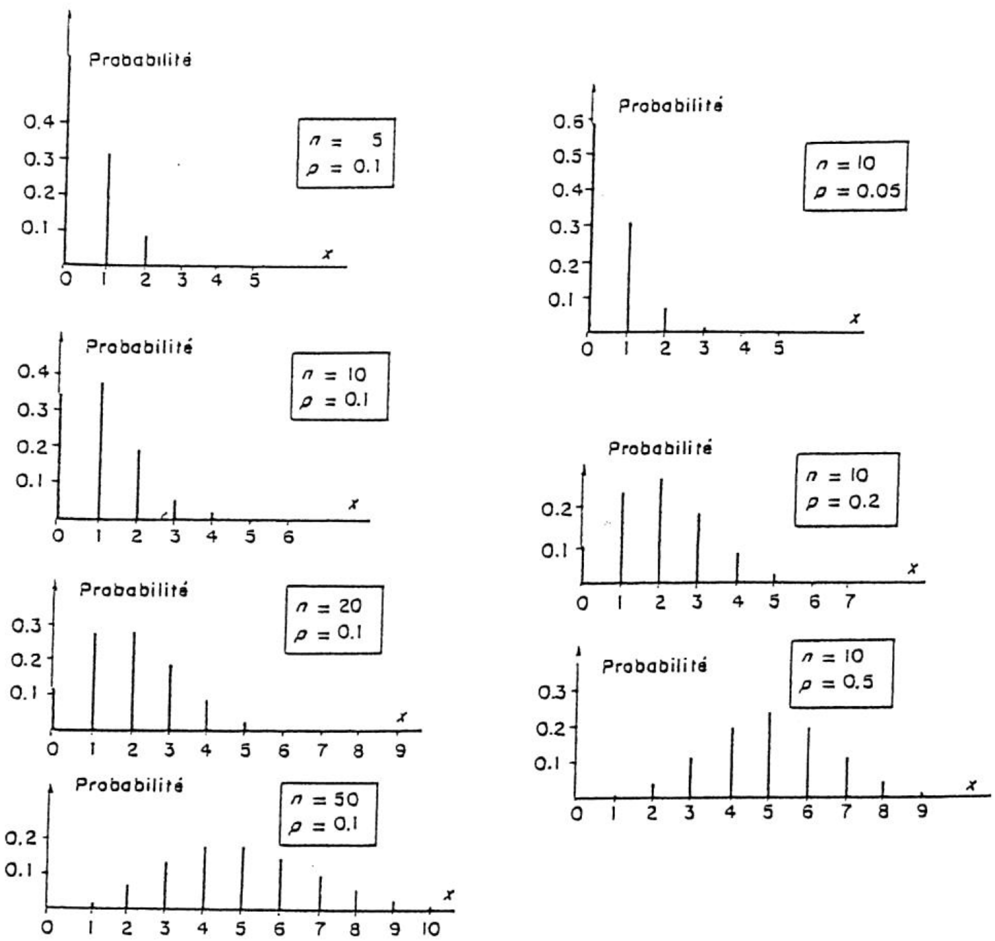
Propriété 6.2 *Si $X \rightsquigarrow \mathcal{B}(n;p)$, alors :*

$$E(X) = np \quad \textit{et} \quad V(X) = npq = np(1-p)$$

Le Graphique 6.1 montre l'allure de la distribution de probabilité binomiale pour

[5] Ces tables donnent les probabilités $\mathbb{P}(X = x_i)$ dans une loi $\mathcal{B}(n;p)$ pour des valeurs de n pas trop grandes (typiquement $n \leq 30$) et un ensemble de valeurs de p $(0,05,\ 0,1,\ 0,15,$ etc...) inférieures ou égales à $0,5$. Pour des valeurs de p supérieures à $0,5$, il suffit d'utiliser le fait que si $X \rightsquigarrow \mathcal{B}(n;p)$, alors $X^* = (n - X) \rightsquigarrow \mathcal{B}(n;p^*)$, où $p^* = 1 - p$: on peut donc obtenir la probabilité $\mathbb{P}(X = x_i)$ dans une loi $\mathcal{B}(n;p)$, où $p > 0,5$, en recherchant dans les tables la probabilité $\mathbb{P}(X^* = x_i^*)$ dans la loi $\mathcal{B}(n;p^*)$, où $x_i^* = n - x_i$ et $p^* = 1 - p$. Nous laisserons au lecteur le soin de vérifier que cette façon de procéder est bien correcte.

diverses valeurs de n et p. On notera que la distribution est symétrique pour $p = 1/2$ et le devient approximativement sinon, dès que n est assez élevé.



Graphique 6.1 : Quelques représentations de la loi binomiale

Pour conclure, notons encore la propriété suivante.

Propriété 6.3 *Soient $Z_1 \rightsquigarrow \mathcal{B}(n_1; p)$ et $Z_2 \rightsquigarrow \mathcal{B}(n_2; p)$. Si Z_1 et Z_2 sont indépendantes, alors :*

$$Z = Z_1 + Z_2 \rightsquigarrow \mathcal{B}(n_1 + n_2; p)$$

En d'autres termes, la somme de deux variables aléatoires binomiales indépendantes et de même paramètre p est encore une variable aléatoire binomiale. Etant donné la définition d'une variable aléatoire binomiale, cela ne devrait guère vous étonner.

Ce résultat est aussi valable pour la somme de r variables aléatoires binomiales indépendantes et de même paramètre p.

Propriété 6.4 *Soient $Z_1 \rightsquigarrow \mathcal{B}(n_1; p), Z_2 \rightsquigarrow \mathcal{B}(n_2; p), ..., Z_r \rightsquigarrow \mathcal{B}(n_r; p)$. Si $Z_1, Z_2,$*

$..., Z_r$ *sont indépendantes, alors :*

$$Z = Z_1 + Z_2 + ... + Z_r \rightsquigarrow \mathcal{B}(n_1 + n_2 + ... + n_r; p)$$

6.3.3. Exercice résolu

On suppose qu'à chaque naissance la probabilité d'avoir une fille est de $0,48$. On considère une famille de 6 enfants.

1- Quelle est la probabilité que tous les enfants soient des filles ?

On peut associer à chacun des enfants une variable aléatoire de Bernoulli, qui prend la valeur 1 si l'enfant est une fille et la valeur 0 si c'est un garçon. On a ainsi $X_1, X_2, X_3, X_4, X_5, X_6$, six variables aléatoires indépendantes et suivant une loi de Bernoulli $\mathcal{B}(0,48)$.
On peut alors définir la variable aléatoire $X = X_1 + X_2 + X_3 + X_4 + X_5 + X_6$ donnant le nombre de filles parmi les 6 enfants. X suit une loi binomiale $\mathcal{B}(6; 0,48)$. On a donc (à l'arrondi près) :

$$I\!P(X = 6) = C_6^6 (0,48)^6 (0,52)^0 = (0,48)^6 = 0,0122$$

2- Quelle est la probabilité que tous les enfants soient des garçons ?

C'est la probabilité qu'aucun des enfants ne soit une fille. Sur base de ce qui précède, on obtient (à l'arrondi près) :

$$I\!P(X = 0) = C_6^0 (0,48)^0 (0,52)^6 = (0,52)^6 = 0,0197$$

3- Quelle est la probabilité que 4 au moins des enfants soient une fille ?

A nouveau sur base de ce qui précède, on obtient (aux arrondis près):

$$\begin{aligned} I\!P(X \geq 4) &= I\!P(X = 4) + I\!P(X = 5) + I\!P(X = 6) \\ &= C_6^4 (0,48)^4 (0,52)^2 + C_6^5 (0,48)^5 (0,52)^1 + C_6^6 (0,48)^6 (0,52)^0 \\ &= 0,2153 + 0,0794 + 0,0122 = 0,3069 \end{aligned}$$

4- Donnez sous forme d'un tableau la loi du nombre de filles parmi les 6 enfants.

La loi demandée est la loi de X définie ci-dessus. On a déjà calculé la probabilité de 4 des 7 valeurs possibles de X. On trouve pour les valeurs restantes (aux arrondis près) :

$$\begin{aligned} I\!P(X = 1) &= C_6^1 (0,48)^1 (0,52)^5 = 0,1094 \\ I\!P(X = 2) &= C_6^2 (0,48)^2 (0,52)^4 = 0,2526 \\ I\!P(X = 3) &= C_6^3 (0,48)^3 (0,52)^3 = 0,3110 \end{aligned}$$

La loi de X est donc au total donnée par (aux arrondis près) :

x_i	0	1	2	3	4	5	6
p_{x_i}	0,0197	0,1094	0,2526	0,3110	0,2153	0,0794	0,0122

On note que la somme des probabilités est bien, comme il se doit , égale à 1 (à l'arrondi près).

6.4. La loi de Poisson

6.4.1. Processus de Poisson et loi de Poisson

La loi de POISSON[6] permet de représenter, au moins approximativement, de nombreux phénomènes aléatoires. Quelques exemples sont le nombre d'accidents sur une autoroute durant une heure de pointe, le nombre d'unités d'un bien vendues au cours d'un après-midi, le nombre d'urgences se présentant dans un hôpital entre 2h et 4h du matin , le nombre d'appels téléphoniques aboutissant à un standard durant cinq minutes, etc...

Tous ces cas peuvent être considérés comme des exemples d'un processus dont les événements surviennent aléatoirement, indépendamment et uniformément au cours du temps. On appelle un tel processus un *processus de* POISSON. La définition suivante en donne une description plus précise.

Définition 6.7 *On appelle processus de POISSON un processus aléatoire qui décrit les arrivées d'un événement au cours du temps et qui est tel que :*

1- la probabilité qu'un événement survienne au cours d'un très petit intervalle de temps Δt est (à peu près) proportionnelle à la durée de cet intervalle et est constante au cours du temps.
2- la probabilité que deux événements ou plus surviennent au cours d'un tel très petit intervalle de temps Δt est négligeable par rapport à celle qu'il ne s'en produise qu'un seul.
3- les nombres d'événements qui surviennent au cours d'intervalles de temps distincts (quelle que soit leur durée) sont indépendants.

Considérons un processus de POISSON et désignons par X le nombre d'événements qui surviennent au cours d'une période de temps quelconque T. On peut montrer que la variable aléatoire X, qui peut prendre n'importe quelle valeur entière supérieure ou égale de 0, suit une loi de POISSON de paramètre λ, où λ est un nombre (entier ou non) strictement positif qui dépend à la fois de l'intensité du processus de POISSON (l'importance de la probabilité qu'un événement survienne au cours d'un très petit intervalle de temps Δt) et de la période de temps T considérée. Cette loi est définie de la manière suivante.

[6] Siméon-Denis POISSON, mathématicien français (1781–1840), auteur de l'ouvrage *Recherche sur la probabilité des jugements en matière criminelle et en matière civile* (1837).

Définition 6.8 *Une variable aléatoire X suit une loi de* POISSON *de paramètre λ, où $\lambda > 0$, si :*

$$\mathcal{X} = \{0, 1, 2, ...\}$$

et

$$I\!P(X = x_i) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}, \ \forall x_i \in \mathcal{X}$$

On note : $X \rightsquigarrow \mathcal{P}(\lambda)$.

Remarquez qu'une variable aléatoire qui suit une loi de POISSON constitue un exemple de variable aléatoire discrète dont le nombre de valeurs possibles est infini, mais dénombrable.

Le paramètre λ a une interprétation simple. On peut en effet vérifier (cf. infra) que λ est égal à l'espérance de X, qui au surplus est égale à sa variance.

Propriété 6.5 *Si $X \rightsquigarrow \mathcal{P}(\lambda)$, alors :*

$$E(X) = \lambda \quad \text{et} \quad V(X) = \lambda$$

En d'autres termes, λ est simplement l'espérance du nombre d'événements qui se produisent par unité de temps T.

Le paramètre λ sera évidemment d'autant plus grand que l'intensité de processus de POISSON est forte. On peut en outre montrer que pour un processus de POISSON d'intensité donnée, λ est directement proportionnel à la durée de la période de temps T considérée. Ainsi, si le processus de POISSON est tel que l'espérance du nombre d'événements qui se produisent par heure est de par exemple 3, l'espérance du nombre d'événements qui se produisent par unité de 10 minutes sera de $\frac{3}{6} = 0,5$.

Le Graphique 6.2 montre l'allure de la distribution de probabilité de POISSON pour diverses valeurs de λ. On voit que cette distribution est plutôt étalée à droite pour des petites valeurs de λ et tend à devenir symétrique à mesure que λ s'accroît.

Le nombre e qui intervient dans la définition de la loi de POISSON est celui qui intervient dans la fonction exponentielle e^x. Ce nombre est égal à :

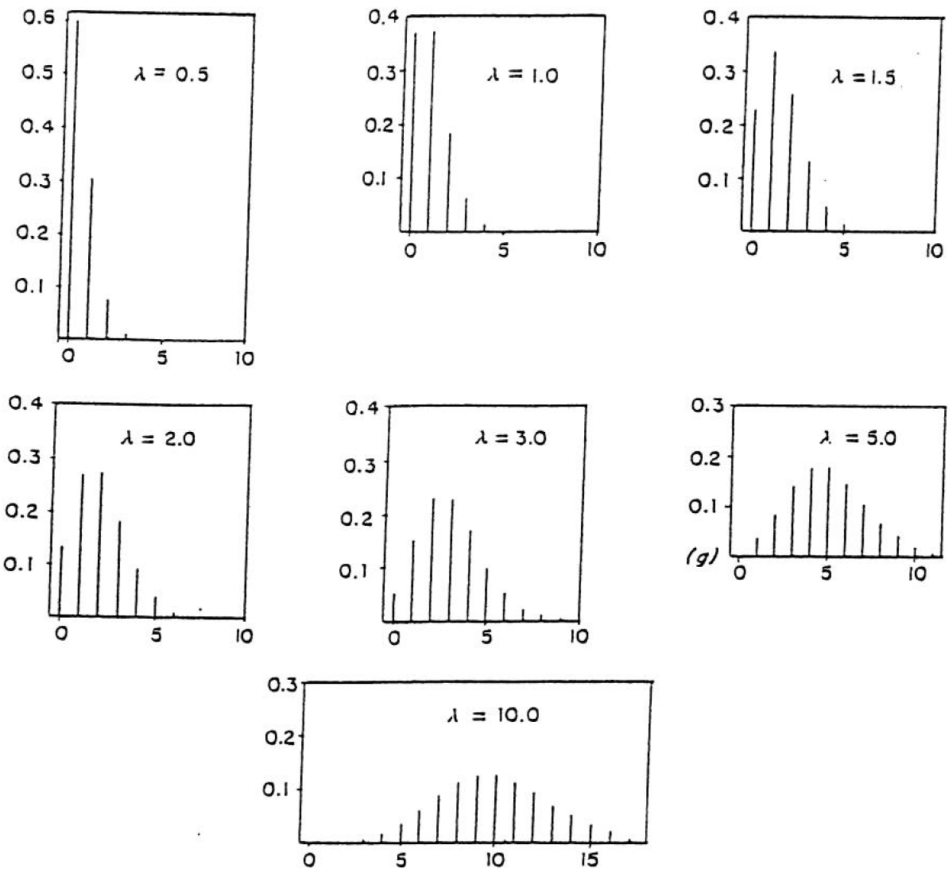
$$e = 2,71828...$$

Si on se rappelle que e^x peut être calculé par le développement de Taylor :

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + ... = \sum_{i=0}^{\infty} \frac{x^i}{i!} \tag{6.2}$$

on vérifie aisément que, comme il se doit, les probabilités $p_{x_i} = I\!P(X = x_i)$ sont bien, $\forall x_i \in \mathcal{X}$, comprises entre 0 et 1, et que leur somme est bien égale à 1 :

$$\sum_{x_i} p_{x_i} = \sum_{x_i=0}^{\infty} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-\lambda} \underbrace{\sum_{x_i=0}^{\infty} \frac{\lambda^{x_i}}{x_i!}}_{=e^{\lambda}} = e^{-\lambda} e^{\lambda} = 1$$



Graphique 6.2 : Quelques représentations de la loi de Poisson

En s'appuyant sur (6.2), on vérifie également aisément que l'espérance et la variance de X sont bien, comme indiqué par la Propriété 6.5, toutes deux égales à λ. Par application de la Définition 4.5 et des Propriétés 4.7, 4.8 et 5.5, on a en effet :

$$\begin{aligned} E(X) &= \sum_{x_i} x_i p_{x_i} = \sum_{x_i=0}^{\infty} x_i e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-\lambda} \sum_{x_i=1}^{\infty} x_i \frac{\lambda^{x_i}}{x_i!} \\ &= e^{-\lambda} \lambda \underbrace{\sum_{x_i=1}^{\infty} \frac{\lambda^{x_i-1}}{(x_i-1)!}}_{=e^{\lambda}} = e^{-\lambda} \lambda e^{\lambda} = \lambda \end{aligned}$$

$$\begin{aligned} E\left[X(X-1)\right] &= \sum_{x_i} x_i(x_i-1)p_{x_i} = \sum_{x_i=0}^{\infty} x_i(x_i-1)e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} \\ &= e^{-\lambda}\sum_{x_i=2}^{\infty} x_i(x_i-1)\frac{\lambda^{x_i}}{x_i!} = e^{-\lambda}\lambda^2 \underbrace{\sum_{x_i=2}^{\infty}\frac{\lambda^{x_i-2}}{(x_i-2)!}}_{=e^{\lambda}} \\ &= e^{-\lambda}\lambda^2 e^{\lambda} = \lambda^2 \end{aligned}$$

et donc :

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 = E\left(X^2) - E(X\right) + E(X) - (E(X))^2 \\ &= \underbrace{E(X^2 - X)}_{=E[X(X-1)]} + E(X) - (E(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

Comme celles de la loi binomiale, les probabilités intervenant dans la loi de Poisson ne sont pas très difficiles à calculer avec les instruments dont on dispose actuellement. Toutefois, il est parfois commode d'utiliser des tables[7] ou certaines approximations dont nous parlerons plus loin. Quelques tables de la loi de Poisson sont reproduites à la fin de ce chapitre, à la Section 6.7.

Pour conclure, remarquons encore qu'on peut établir que la somme de deux variables aléatoires de Poisson indépendantes est encore une variable aléatoire de Poisson.

Propriété 6.6 *Soient $X_1 \rightsquigarrow \mathcal{P}(\lambda_1)$ et $X_2 \rightsquigarrow \mathcal{P}(\lambda_2)$. Si X_1 et X_2 sont indépendantes, alors :*

$$X = X_1 + X_2 \rightsquigarrow \mathcal{P}(\lambda_1 + \lambda_2)$$

Ce résultat est aussi valable pour la somme de r variables aléatoires de Poisson indépendantes.

Propriété 6.7 *Soient $X_1 \rightsquigarrow \mathcal{P}(\lambda_1), X_2 \rightsquigarrow \mathcal{P}(\lambda_2), ..., X_r \rightsquigarrow \mathcal{P}(\lambda_r)$. Si $X_1, X_2, ..., X_r$ sont indépendantes, alors :*

$$X = X_1 + X_2 + ... + X_r \rightsquigarrow \mathcal{P}(\lambda_1 + \lambda_2 + ... + \lambda_r)$$

6.4.2. Exercice résolu

On considère que les ventes de magnétoscopes dans une grande surface suivent un processus de Poisson tel que l'espérance du nombre de magnétoscopes vendus en une journée est de $0,9$.

[7] Ces tables donnent les probabilités $\mathbb{P}(X = x_i)$ dans une loi $\mathcal{P}(\lambda)$ pour un ensemble de valeurs typiques de λ inférieures ou égales à 20.

1- Quelle est la probabilité de vendre 4 magnétoscopes au moins en une journée ?

Notons X le nombre de magnétoscopes vendus durant une journée. On a : $X \rightsquigarrow \mathcal{P}(0,9)$. La probabilité demandée est (aux arrondis près) :

$$\begin{aligned}
I\!P(X \geq 4) &= 1 - I\!P(X \leq 3) \\
&= 1 - I\!P(X=0) - I\!P(X=1) - I\!P(X=2) - I\!P(X=3) \\
&= 1 - e^{-0,9}\frac{0,9^0}{0!} - e^{-0,9}\frac{0,9^1}{1!} - e^{-0,9}\frac{0,9^2}{2!} - e^{-0,9}\frac{0,9^3}{3!} \\
&= 1 - 0,4066 - 0,3659 - 0,1647 - 0,0494 = 1 - 0,9866 \\
&= 0,0134
\end{aligned}$$

2- Quelle est la probabilité de vendre 4 magnétoscopes au moins sur une période de 10 jours ?

Notons X^* le nombre de magnétoscopes vendus sur une période de 10 jours. On a maintenant : $X^* \rightsquigarrow \mathcal{P}(\lambda^*)$ où $\lambda^* = 10\lambda = 9$. La probabilité demandée est (aux arrondis près) :

$$\begin{aligned}
I\!P(X \geq 4) &= 1 - I\!P(X \leq 3) \\
&= 1 - I\!P(X=0) - I\!P(X=1) - I\!P(X=2) - I\!P(X=3) \\
&= 1 - e^{-9}\frac{9^0}{0!} - e^{-9}\frac{9^1}{1!} - e^{-9}\frac{9^2}{2!} - e^{-9}\frac{9^3}{3!} \\
&= 1 - 0,0001 - 0,0011 - 0,0049 - 0,0149 = 1 - 0,0210 \\
&= 0,9790
\end{aligned}$$

6.4.3. Approximation de la loi binomiale par une loi de Poisson

Sous certaines conditions, la loi de Poisson peut être utilisée comme approximation d'une loi binomiale. D'un point de vue formel, cette possibilité repose sur un résultat de convergence en loi dont nous ne parlerons qu'au Chapitre 9. Retenons simplement ici que lorsque n est grand, p petit et np pas trop grand, les probabilités associées à une loi binomiale $\mathcal{B}(n;p)$ sont très proches des probabilités données par une loi de Poisson dont le paramètre λ vaut np :

$$\mathcal{B}(n;p) \approx \mathcal{P}(np)$$

en d'autres termes :

$$I\!P(X = x_i) = C_n^{x_i} p^{x_i} q^{n-x_i} \approx e^{-np}\frac{(np)^{x_i}}{x_i!}, \quad \forall\, x_i \in \{0,1,2,...,n\}$$

Remarquez que les deux lois $\mathcal{B}(n;p)$ et $\mathcal{P}(np)$ ont la même espérance $(= np)$. L'interprétation des qualificatifs "grand", "petit" et "pas trop grand" est relative : tout dépend de l'exigence qu'on a en matière de précision. Dans le cadre de ce cours, nous admettrons que cette approximation est “bonne” lorsque $n \geq 50$, $p \leq 0,1$ et $np \leq 15$.

Examinons un cas pratique : l'approximation d'une loi binomiale $\mathcal{B}(50; 0,05)$ par une loi de Poisson $\mathcal{P}(2,5)$. On a (aux arrondis près) :

Probabilités cherchées	Résultats exacts selon la loi $\mathcal{B}(50; 0,05)$	Résultats approchés selon la loi $\mathcal{P}(2,5)$
$P(X = 1)$	$C_{50}^{1}(0,05)^1(0,95)^{49} = 0,2025$	$e^{-2,5}\ (2,5)^1/1! = 0,2052$
$P(X = 5)$	$C_{50}^{5}\ (0,05)^5(0,95)^{45} = 0,0658$	$e^{-2,5}\ (2,5)^5/5! = 0,0668$
$P(X = 10)$	$C_{50}^{10}\ (0,05)^{10}(0,95)^{40} = 0,0001$	$e^{-2,5}\ (2,5)^{10}/10! = 0,0002$

On le voit, l'approximation est assez précise.

6.5. Exercice résolu

La probabilité qu'un patient ne se présente pas à son rendez-vous chez le dentiste est de $0,13$. Supposons que les patients agissent indépendamment les uns des autres et considérons une consultation où 10 patients ont un rendez-vous. On considère la variable aléatoire X représentant le nombre de patients qui ne se présentent pas à leur rendez-vous durant cette consultation.

1- Quelle est la loi de X ? Justifiez votre réponse.

On peut associer à chacun des 10 patients une variable aléatoire de Bernoulli, qui prend la valeur 1 si le patient ne se présente pas à son rendez-vous et la valeur 0 s'il se présente à son rendez-vous. On a ainsi $X_1, X_2, ..., X_{10}$, dix variables aléatoires suivant une même loi : $X_i \rightsquigarrow \mathcal{B}(0,13)$.
Comme les patients agissent indépendamment les uns des autres, ces 10 variables aléatoires sont indépendantes et on a donc :

$$X = X_1 + X_2 + \cdots + X_{10} \rightsquigarrow \mathcal{B}(10; 0,13)$$

où X représente le nombre de patients qui ne se présentent pas à leur rendez-vous.

2- Quelle est la probabilité que 2 personnes ou plus annulent leur rendez-vous ?

La probabilité demandée est $\mathbb{P}(X \geq 2)$. On a (aux arrondis près) :

$$\begin{aligned}
\mathbb{P}(X \geq 2) &= 1 - \mathbb{P}(X \leq 1) = 1 - \mathbb{P}(X = 0 \text{ ou } X = 1) \\
&= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \\
&= 1 - C_{10}^{0}(0,13)^0(0,87)^{10} - C_{10}^{1}(0,13)^1(0,87)^9 \\
&= 1 - 0,2484 - 0,3712 = 0,3804
\end{aligned}$$

3- On considère à présent que 100 patients ont un rendez-vous et que la probabilité qu'un patient ne se présente pas à son rendez-vous est de $0,07$. Calculez à l'aide d'une approximation la probabilité que 12 patients ne se présentent pas à leur rendez-vous.

La loi de X est maintenant :

$$X \rightsquigarrow \mathcal{B}(100; 0,07)$$

Comme $n = 100 \geq 50$, $p = 0,07 \leq 0,1$ et $np = 7 \leq 15$, on réalise une approximation par une loi de POISSON. On a :

$$X \rightsquigarrow \mathcal{B}(100; 0,07) \approx \mathcal{P}(7)$$

La probabilité demandée est $\mathbb{P}(X = 12)$, que l'on approche donc par :

$$\mathbb{P}(X = 12) \approx e^{-7}\frac{7^{12}}{12!} = 0,0263...$$

6.6. Exercices

1- Calculez $C_3^2, C_5^2, C_5^3, C_{647}^3, C_{45853}^3$.

2- Une petite entreprise de transport possède 6 camions. La probabilité pour chacun d'entre eux de nécessiter l'année prochaine une vérification complète du moteur est de $0,1$.

   a- Décrivez l'épreuve aléatoire à l'aide de variables aléatoires de BERNOULLI et d'une variable aléatoire binomiale. Donnez la loi de ces différentes variables aléatoires.

   b- Quelle est la probabilité :

      i- de ne pas avoir de camions en révision l'an prochain ?

      ii- d'avoir 2 camions en révision complète l'an prochain ?

   c- Calculez l'espérance et la variance du nombre de camions qui iront en révision l'an prochain ?

3- Supposons que le nombre d'appels parvenant à la police le samedi matin entre 6h et 6h30 suit une loi de POISSON dont le paramètre λ vaut $2,5$.

   a- Quelle est la probabilité qu'aucun appel ne parvienne à la police durant cette période ?

   b- Quelle est la probabilité d'avoir 3 appels ou plus ?

   c- Calculez l'espérance, la variance et l'écart-type du nombre d'appels.

4- On suppose que les appels qui arrivent à un central téléphonique suivent un processus de POISSON caractérisé par une espérance du nombre d'appels reçus par heure égale à 30.

   a- Quelle est la probabilité qu'il n'y ait pas d'appel pendant une période de 3 minutes ?

   b- Quelle est la probabilité qu'il y ait plus de 5 appels pendant ces 3 minutes ?

   c- Quelle est la probabilité que sur 6 périodes de 4 minutes, il y en ait plus de deux où il n'y ait aucun appel ?

   d- Quelle est la probabilité que sur 4 périodes de 6 minutes, il y en ait au plus une où il y ait plus de deux appels ?

5- On considère que, durant une heure de pointe, le nombre d'appels qui arrivent au standard téléphonique d'une administration fluctue autour de 600 appels. Le tableau du standard peut faire un maximum de 20 connections par minute. Utilisez la distribution de POISSON pour évaluer la probabilité que le tableau soit surchargé pendant une minute quelconque d'une période de pointe.

6- Soit une variable aléatoire X prenant les valeurs entières entre 0 et 20 et suivant une distribution de probabilité uniforme.

   a- Déterminez $I\!P(2 \leq X \leq 9)$.
   b- Calculez l'espérance et la variance de X.

7- Le mercenaire San Antonio est condamné à être passé par les armes. Le peloton d'exécution est composé de 10 hommes qui sont tous de mauvais tireurs de sorte que chacun d'eux n'a qu'une chance sur deux d'atteindre la cible.

   a- Quelle probabilité San Antonio a-t-il de s'en tirer indemne ?
   b- Quelle est la loi du nombre de tireurs qui réussiront à l'atteindre ?
   c- Quelle est la probabilité qu'il soit touché par au moins deux des tireurs ?
   d- Sur quelle hypothèse implicite vos réponses aux questions ci-dessus s'appuient-elles ?
   e- On suppose à présent que San Antonio est parvenu à acheter la moitié du peloton d'exécution. Tous sont de tellement mauvais tireurs que ceux qui tâcheront de l'atteindre n'ont que 5 chances sur 10 de réussir, tandis que ceux qui tâcheront de l'éviter ont malgré tout 2 chances sur 10 de le toucher. Quelle probabilité San Antonio a-t-il de s'en tirer indemne ?

8- Un magasin emploie deux vendeurs : Marie et Pierre. Le nombre de clients servis par minute suit une loi de POISSON $\mathcal{P}(0,4)$ lorsque Marie est au comptoir et $\mathcal{P}(0,2)$ lorsque c'est Pierre qui est au comptoir.

   a- Un des vendeurs est plus expérimenté que l'autre. A votre avis, duquel s'agit-il et pourquoi ?
   b- Supposons que Marie est seule au comptoir. Quelle est la probabilité qu'elle serve :

      i- 1 ou 2 personnes en une minute ?
      ii- au moins 2 personnes en une minute ?

   c- Supposons que Marie et Pierre sont tous les deux au comptoir et qu'ils servent les clients de manière indépendante. Quelle est la probabilité que le nombre total de clients servis en une minute par Marie et Pierre soit supérieur ou égal à 3 ?

9- Une compagnie d'assurances propose à ses clients un nouveau contrat annuel destiné à les couvrir en cas de vol dans leur résidence secondaire. Sur base d'une étude statistique, la compagnie estime que six résidences secondaires sur 100 sont cambriolées au moins une fois par an et que la valeur des biens volés dépasse 100 000 F une fois sur 50.

   a- Supposons que 10 clients prennent l'assurance proposée.

      i- Quelle est la probabilité que la compagnie ne doive pas intervenir ?
      ii- Quelle est la loi du nombre de clients pour lesquels l'assurance va

intervenir ?

iii- Quelle est la probabilité qu'elle doive intervenir pour deux clients ?

b- Supposons à présent que 100 clients prennent l'assurance proposée.

i- Donnez une prévision du nombre de clients qui bénéficieront de l'intervention de l'assurance.

ii- Quelle est la probabilité que l'assurance doive intervenir pour 3 clients au moins ? (Faites une approximation)

c- On considère seulement 2 clients. On appelle X le nombre de clients qui seront cambriolés et Y le nombre de clients cambriolés pour lesquels la valeur des biens volés dépasse 100 000 F. Calculez la probabilité $I\!P(Y = 1|X = 2)$.

10- La classe de Toto comporte 50 élèves et a aujourd'hui quatre cours . Dans chacun des cours, le professeur va interroger un élève en le tirant au sort. Précisons que les professeurs ne se concertent pas entre eux de telle sorte qu'il est possible pour un même élève d'être interrogé dans plusieurs matières.

a- Quelle est la probabilité que Toto ne soit pas interrogé en mathématiques ?

b- Quelle est la probabilité que Toto ne soit pas interrogé du tout aujourd'hui ?

c- Quelle est la probabilité qu'il soit interrogé en mathématiques et en histoire, et dans ces deux matières seulement ?

d- On appelle X le nombre de fois que Toto sera interrogé aujourd'hui.

i- Quelle est la loi de X ?

ii- Quelle est la probabilité que Toto soit interrogé au moins deux fois ?

iii- Quelles sont l'espérance et le mode (c'est-à-dire la valeur la plus probable) de X ?

11- Une sonde spatiale posée sur Mars observe les impacts de météorites qui se produisent dans un rayon de 10 km autour d'elle. On admet que le nombre d'impacts par jour martien suit une loi de Poisson $\mathcal{P}(\lambda)$.

a- Pendant 50 jours martiens consécutifs, la sonde a détecté en tout 150 impacts. Quelle valeur proposez-vous d'admettre pour λ ? (Dans la suite, on supposera que λ a cette valeur).

b- Quelle est la probabilité qu'il n'y ait aucun impact durant une journée martienne donnée ?

c- Quelle est la probabilité qu'il y ait au moins 5 impacts ?

d- Supposons à présent que sur 100 météorites qui tombent dans la zone d'observation, il y en a une qui pèse plus d'une tonne.

i- Sachant qu'il tombe k météorites un certain jour martien dans la zone d'observation, quelle est la loi du nombre de météorites qui pèsent plus d'une tonne ?

ii- Quelle est la probabilité d'observer deux météorites de plus d'une tonne sachant qu'il est tombé trois météorites durant la journée martienne ?

iii- Quelle est la probabilité d'observer l'arrivée de trois météorites dont deux de plus d'une tonne durant une journée martienne ?

iv- Quelle est la probabilité qu'il ne tombe aucune météorite de plus d'une tonne, quel que soit le nombre de météorites tombées durant la journée martienne considérée ?

12- M et Mme Legris possèdent deux voitures, une VX9 et une Sultane. Ils se servent autant de l'une que de l'autre et chacun conduit indifféremment l'une ou l'autre. On appelle X le nombre de réparations que la VX9 nécessite dans l'année et Y le nombre de réparations que la Sultane nécessite dans l'année. On suppose que X et Y suivent respectivement des lois de POISSON $\mathcal{P}(3)$ et $\mathcal{P}(2)$, et sont des variables aléatoires indépendantes.
   - a- A votre avis, laquelle des deux voitures est de meilleure qualité que l'autre ? Justifiez votre réponse.
   - b- Quelle est la probabilité que la VX9 ne nécessite aucune réparation dans l'année ? Qu'elle en nécessite au moins deux ?
   - c- Quelle est la probabilité que la VX9 et la Sultane nécessite chacune au moins deux réparations dans l'année ?
   - d- Quelle est la loi du nombre total de réparations auxquelles le couple a à faire face dans l'année ?

13- Un texte est donné à dactylographier en deux parties de 1 000 caractères chacune à deux secrétaires différentes. La probabilité que la première secrétaire fasse une erreur de frappe est de $0,002$ par caractère, elle est de $0,001$ pour la seconde secrétaire. Quelle est la probabilité de trouver dans le texte 3 erreurs au moins (Faites une approximation).

14- Une compagnie d'assurance constate que chaque année $0,004\%$ de la population meurt à la suite d'un accident. Quelle est la probabilité pour que la compagnie qui compte 10 000 assurés doive payer plus de 3 fois ?

15- Il y a au Pakistan 97% de Musulmans (le reste de la population étant partagée entre les minorités chrétienne, hindoue et sikh). Un journal pakistanais organise un concours dans lequel 100 personnes gagnent un voyage en Europe. Quelle est la probabilité que parmi les 100 gagnants il y ait moins de 95 Musulmans ?

16- La probabilité qu'une personne quelconque soit allergique est de $0,1\%$.
   - a- Quelle est la loi de probabilité du nombre d'allergiques sur une population de 4 000 personnes?
   - b- Peut-on réaliser une approximation de cette loi par une loi de POISSON ?
   - c- Calculez la probabilité que le nombre d'allergiques ne dépasse pas deux personnes.
   - d- Calculez la probabilité qu'il soit au moins de trois.

17- Le nombre de personnes adressant par jour une réclamation à une administration suit une loi de POISSON $\mathcal{P}(3)$. On considère que les nombres de réclamations enregistrées au cours de journées différentes sont indépendants.
   - a- Quelle est la probabilité que demain le nombre de réclamations soit au plus égal à 1 ?
   - b- A quel nombre de réclamations peut-on s'attendre ? Pourquoi ?
   - c- On considère 4 jours où on examine le nombre de réclamations.
      - i- Quelle est la loi du nombre total de réclamations sur ces 4 jours ?
      - ii- Quelle est la probabilité que le nombre total de réclamations soit inférieur à 4 ?

iii- Quelle est la probabilité que, sur ces 4 jours, il y en ait 2 où le nombre de réclamations soit au plus égal à 1 ?

18- Chaque jour d'été, il pleut avec une probabilité de $0,08$. On considère que les conditions climatiques enregistrées au cours des différentes journées d'été sont indépendantes entre elles.

a- L'hypothèse d'indépendance énoncée ci-dessus est-elle crédible ? Pourquoi ?

b- On maintient l'hypothèse d'indépendance et on considère 7 journées consécutives.

i- Quelle est la loi du nombre de jours de pluie sur les 7 jours considérés ?

ii- A quel nombre de jours de pluie peut-on s'attendre ?

iii- Quelle est la probabilité qu'il ne pleuve pas plus de 2 jours sur les 7 considérés ?

c- On maintient toujours l'hypothèse d'indépendance et on considère à présent 60 journées consécutives.

i- Quelle est la loi du nombre de jours de pluie sur les 60 jours considérés ?

ii- Par quelle loi peut-on approcher cette loi ?

iii- Utilisez cette loi approchée pour calculer la probabilité qu'il pleuve au plus 3 jours.

d- On considère maintenant un pays où un jour de pluie se produit avec une probabilité de $0,93$ et on continue à supposer que les conditions climatiques enregistrées au cours de différentes journées sont indépendantes entre elles. Quelle est la probabilité d'avoir 148 jours de pluie sur 150 jours observés ?

19- On appelle X le nombre de clients se présentant pour acheter une voiture neuve chez le concessionnaire PAGEOT ce mercredi. Parmi les clients, on distingue les femmes ($= X_1$) et les hommes ($= X_2$). On a donc $X = X_1 + X_2$. On suppose que X_1 suit une loi de POISSON $\mathcal{P}(1,2)$, que X_2 suit une loi de POISSON $\mathcal{P}(1,4)$ et que X_1 et X_2 sont des variables aléatoires indépendantes.

a- Quelle est la probabilité que deux femmes se présentent chez PAGEOT ce mercredi ?

b- Quelle est la probabilité que deux femmes et un homme se présentent chez PAGEOT ce mercredi ?

c- Quelle est la probabilité que trois clients se présentent chez PAGEOT ce mercredi ?

d- Donnez une prévision du nombre de clients qui se présenteront chez PAGEOT ce mercredi.

e- On sait que deux clients se sont présentés chez PAGEOT ce mercredi. Établissez la loi du nombre de femmes parmi ces deux clients.

6.7. Tables statistiques

Loi binomiale (source : Saporta G. (1990))

PROBABILITÉS BINOMIALES $C_N^X p^X (1-p)^{N-X}$ POUR $N \leqslant 10$ ET POUR DIVERSES VALEURS DE p

N	X \ p	.01	.05	.10	.15	.20	.25	.30	1/3	.35	.40	.45	.50
2	0	.9801	.9025	.8100	.7225	.6400	.5625	.4900	.4444	.4225	.3600	.3025	.2500
	1	.0198	.0950	.1800	.2550	.3200	.3750	.4200	.4444	.4550	.4800	.4950	.5000
	2	.0001	.0025	.0100	.0225	.0400	.0625	.0900	.1111	.1225	.1600	.2025	.2500
3	0	.9703	.8574	.7290	.6141	.5120	.4219	.3430	.2963	.2746	.2160	.1664	.1250
	1	.0294	.1354	.2430	.3251	.3840	.4219	.4410	.4444	.4436	.4320	.4084	.3750
	2	.0003	.0071	.0270	.0574	.0960	.1406	.1890	.2222	.2389	.2880	.3341	.3750
	3	.0000	.0001	.0010	.0034	.0080	.0156	.0270	.0370	.0429	.0640	.0911	.1250
4	0	.9606	.8145	.6561	.5220	.4096	.3164	.2401	.1975	.1785	.1296	.0915	.0625
	1	.0388	.1715	.2916	.3685	.4096	.4219	.4116	.3951	.3845	.3456	.2995	.2500
	2	.0006	.0135	.0486	.0975	.1536	.2109	.2646	.2963	.3105	.3456	.3675	.3750
	3	.0000	.0005	.0036	.0115	.0256	.0469	.0756	.0988	.1115	.1536	.2005	.2500
	4	.0000	.0000	.0001	.0005	.0016	.0039	.0081	.0123	.0150	.0256	.0410	.0625
5	0	.9510	.7738	.5905	.4437	.3277	.2373	.1681	.1317	.1160	.0778	.0503	.0312
	1	.0480	.2036	.3280	.3915	.4096	.3955	.3602	.3292	.3124	.2592	.2059	.1562
	2	.0010	.0214	.0729	.1382	.2048	.2637	.3087	.3292	.3364	.3456	.3369	.3125
	3	.0000	.0011	.0081	.0244	.0512	.0879	.1323	.1646	.1811	.2304	.2757	.3125
	4	.0000	.0000	.0004	.0022	.0064	.0146	.0284	.0412	.0488	.0768	.1128	.1562
	5	.0000	.0000	.0000	.0001	.0003	.0010	.0024	.0041	.0053	.0102	.0185	.0312
6	0	.9415	.7351	.5314	.3771	.2621	.1780	.1176	.0878	.0754	.0467	.0277	.0156
	1	.0571	.2321	.3543	.3993	.3932	.3560	.3025	.2634	.2437	.1866	.1359	.0938
	2	.0014	.0305	.0984	.1762	.2458	.2966	.3241	.3292	.3280	.3110	.2780	.2344
	3	.0000	.0021	.0146	.0415	.0819	.1318	.1852	.2195	.2355	.2765	.3032	.3125
	4	.0000	.0001	.0012	.0055	.0154	.0330	.0595	.0823	.0951	.1382	.1861	.2344
	5	.0000	.0000	.0001	.0004	.0015	.0044	.0102	.0165	.0205	.0369	.0609	.0938
	6	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0014	.0018	.0041	.0083	.0156
7	0	.9321	.6983	.4783	.3206	.2097	.1335	.0824	.0585	.0490	.0280	.0152	.0078
	1	.0659	.2573	.3720	.3960	.3670	.3115	.2471	.2048	.1848	.1306	.0872	.0547
	2	.0020	.0406	.1240	.2097	.2753	.3115	.3177	.3073	.2985	.2613	.2140	.1641
	3	.0000	.0036	.0230	.0617	.1147	.1730	.2269	.2561	.2679	.2903	.2918	.2734
	4	.0000	.0002	.0026	.0109	.0287	.0577	.0972	.1280	.1442	.1935	.2388	.2734
	5	.0000	.0000	.0002	.0012	.0043	.0115	.0250	.0384	.0466	.0774	.1172	.1641
	6	.0000	.0000	.0000	.0001	.0004	.0013	.0036	.0064	.0084	.0172	.0320	.0547
	7	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0005	.0006	.0016	.0037	.0078
8	0	.9227	.6634	.4305	.2725	.1678	.1001	.0576	.0390	.0319	.0168	.0084	.0039
	1	.0746	.2793	.3826	.3847	.3355	.2670	.1977	.1561	.1373	.0896	.0548	.0312
	2	.0026	.0515	.1488	.2376	.2936	.3115	.2965	.2731	.2587	.2090	.1569	.1094
	3	.0001	.0054	.0331	.0839	.1468	.2076	.2541	.2731	.2786	.2787	.2568	.2188
	4	.0000	.0004	.0046	.0185	.0459	.0865	.1361	.1707	.1875	.2322	.2627	.2734
	5	.0000	.0000	.0004	.0026	.0092	.0231	.0467	.0683	.0808	.1239	.1719	.2188
	6	.0000	.0000	.0000	.0002	.0011	.0038	.0100	.0171	.0217	.0413	.0703	.1094
	7	.0000	.0000	.0000	.0000	.0001	.0004	.0012	.0024	.0033	.0079	.0164	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0002	.0007	.0017	.0039
9	0	.9135	.6302	.3874	.2316	.1342	.0751	.0404	.0260	.0207	.0101	.0046	.0020
	1	.0830	.2985	.3874	.3679	.3020	.2253	.1556	.1171	.1004	.0605	.0339	.0176
	2	.0034	.0629	.1722	.2597	.3020	.3003	.2668	.2341	.2162	.1612	.1110	.0703
	3	.0001	.0077	.0446	.1069	.1762	.2336	.2668	.2731	.2716	.2508	.2119	.1641
	4	.0000	.0006	.0074	.0283	.0661	.1168	.1715	.2048	.2194	.2508	.2600	.2461
	5	.0000	.0000	.0008	.0050	.0165	.0389	.0735	.1024	.1181	.1672	.2128	.2461
	6	.0000	.0000	.0001	.0006	.0028	.0087	.0210	.0341	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0000	.0003	.0012	.0039	.0073	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0009	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0003	.0008	.0020
10	0	.9044	.5987	.3487	.1969	.1074	.0563	.0282	.0173	.0135	.0060	.0025	.0010
	1	.0914	.3151	.3874	.3474	.2684	.1877	.1211	.0867	.0725	.0403	.0207	.0098
	2	.0042	.0746	.1937	.2759	.3020	.2816	.2335	.1951	.1757	.1209	.0763	.0439
	3	.0001	.0105	.0574	.1298	.2013	.2503	.2668	.2601	.2522	.2150	.1665	.1172
	4	.0000	.0010	.0112	.0401	.0881	.1460	.2001	.2276	.2377	.2508	.2384	.2051
	5	.0000	.0001	.0015	.0085	.0264	.0584	.1029	.1366	.1536	.2007	.2340	.2461
	6	.0000	.0000	.0001	.0012	.0055	.0162	.0368	.0569	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0000	.0001	.0008	.0031	.0090	.0163	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0014	.0030	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010

Loi de Poisson (source: Saporta G. (1990))

LOI DE POISSON

k	Probabilités individuelles $Pr(k) = e^{-m} \frac{m^k}{k!}$								
	m = 0,1	m = 0,2	m = 0,3	m = 0,4	m = 0,5	m = 0,6	m = 0,7	m = 0,8	m = 0,9
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4		0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5				0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6							0,0001	0,0002	0,0003

LOI DE POISSON

k	Probabilités individuelles $Pr(k) = e^{-m} \frac{m^k}{k!}$								
	m = 1,0	m = 1,5	m = 2,0	m = 2,5	m = 3,0	m = 3,5	m = 4,0	m = 4,5	m = 5,0
0	0,3679	0,2231	0,1353	0,0821	0,0498	0,0302	0,0183	0,0111	0,0067
1	0,3679	0,3347	0,2707	0,2052	0,1494	0,1057	0,0733	0,0500	0,0337
2	0,1839	0,2510	0,2707	0,2565	0,2240	0,1850	0,1465	0,1125	0,0842
3	0,0613	0,1255	0,1804	0,2138	0,2240	0,2158	0,1954	0,1687	0,1404
4	0,0153	0,0471	0,0902	0,1336	0,1680	0,1888	0,1954	0,1898	0,1755
5	0,0031	0,0141	0,0361	0,0668	0,1008	0,1322	0,1563	0,1708	0,1755
6	0,0005	0,0035	0,0120	0,0278	0,0504	0,0771	0,1042	0,1281	0,1462
7	0,0001	0,0008	0,0034	0,0099	0,0216	0,0385	0,0595	0,0824	0,1044
8		0,0001	0,0009	0,0031	0,0081	0,0169	0,0298	0,0463	0,0653
9			0,0002	0,0009	0,0027	0,0066	0,0132	0,0232	0,0363
10				0,0002	0,0008	0,0023	0,0053	0,0104	0,0181
11					0,0002	0,0007	0,0019	0,0043	0,0082
12					0,0001	0,0002	0,0006	0,0016	0,0034
13						0,0001	0,0002	0,0006	0,0013
14							0,0001	0,0002	0,0005
15								0,0001	0,0002
16									0,0001

Loi de Poisson (suite)

Loi de Poisson

k	Probabilités individuelles $Pr(k) = e^{-m}\dfrac{m^k}{k!}$								
	m = 5,5	m = 6,0	m = 6,5	m = 7,0	m = 7,5	m = 8,0	m = 8,5	m = 9,0	m = 9,5
0	0,0041	0,0025	0,0015	0,0009	0,0006	0,0003	0,0002	0,0001	0,0001
1	0,0225	0,0149	0,0098	0,0064	0,0041	0,0027	0,0017	0,0011	0,0007
2	0,0618	0,0446	0,0318	0,0223	0,0156	0,0107	0,0074	0,0050	0,0034
3	0,1133	0,0892	0,0688	0,0521	0,0389	0,0286	0,0208	0,0150	0,0107
4	0,1558	0,1339	0,1118	0,0912	0,0729	0,0573	0,0443	0,0337	0,0254
5	0,1714	0,1606	0,1454	0,1277	0,1094	0,0916	0,0752	0,0607	0,0483
6	0,1571	0,1606	0,1575	0,1490	0,1367	0,1221	0,1066	0,0911	0,0764
7	0,1234	0,1377	0,1462	0,1490	0,1465	0,1396	0,1294	0,1171	0,1037
8	0,0849	0,1033	0,1188	0,1304	0,1373	0,1396	0,1375	0,1318	0,1232
9	0,0519	0,0688	0,0858	0,1014	0,1144	0,1241	0,1299	0,1318	0,1300
10	0,0285	0,0413	0,0558	0,0710	0,0858	0,0993	0,1104	0,1186	0,1235
11	0,0143	0,0225	0,0330	0,0452	0,0585	0,0722	0,0853	0,0970	0,1067
12	0,0065	0,0113	0,0179	0,0264	0,0366	0,0481	0,0604	0,0728	0,0844
13	0,0028	0,0052	0,0089	0,0142	0,0211	0,0296	0,0395	0,0504	0,0617
14	0,0011	0,0022	0,0041	0,0071	0,0113	0,0169	0,0240	0,0324	0,0419
15	0,0004	0,0009	0,0018	0,0033	0,0057	0,0090	0,0136	0,0194	0,0265
16	0,0001	0,0003	0,0007	0,0014	0,0026	0,0045	0,0072	0,0109	0,0157
17		0,0001	0,0003	0,0006	0,0012	0,0021	0,0036	0,0058	0,0088
18			0,0001	0,0002	0,0005	0,0009	0,0017	0,0029	0,0046
19				0,0001	0,0002	0,0004	0,0008	0,0014	0,0023
20					0,0001	0,0002	0,0003	0,0006	0,0011
21						0,0001	0,0001	0,0003	0,0005
22							0,0001	0,0001	0,0002
23									0,0001
24									

Loi de Poisson (suite et fin)

Loi de Poisson

k	Probabilités individuelles $Pr(k) = e^{-m}\frac{m^k}{k!}$								
	m = 10	m = 11	m = 12	m = 13	m = 14	m = 15	m = 16	m = 17	m = 18
0									
1	0,0005	0,0002	0,0001						
2	0,0023	0,0010	0,0004	0,0002	0,0001				
3	0,0076	0,0037	0,0018	0,0008	0,0004	0,0002	0,0001		
4	0,0189	0,0102	0,0053	0,0027	0,0013	0,0007	0,0003	0,0002	0,0001
5	0,0378	0,0224	0,0127	0,0070	0,0037	0,0019	0,0010	0,0005	0,0002
6	0,0631	0,0411	0,0255	0,0152	0,0087	0,0048	0,0026	0,0014	0,0007
7	0,0901	0,0646	0,0437	0,0281	0,0174	0,0104	0,0060	0,0034	0,0019
8	0,1126	0,0888	0,0655	0,0457	0,0304	0,0194	0,0120	0,0072	0,0042
9	0,1251	0,1085	0,0874	0,0661	0,0473	0,0324	0,0213	0,0135	0,0083
10	0,1251	0,1194	0,1048	0,0859	0,0663	0,0486	0,0341	0,0230	0,0150
11	0,1137	0,1194	0,1144	0,1015	0,0844	0,0663	0,0496	0,0356	0,0245
12	0,0948	0,1094	0,1144	0,1099	0,0984	0,0829	0,0661	0,0504	0,0368
13	0,0729	0,0926	0,1056	0,1099	0,1060	0,0956	0,0814	0,0658	0,0509
14	0,0521	0,0728	0,0905	0,1021	0,1060	0,1024	0,0930	0,0700	0,0655
15	0,0347	0,0534	0,0724	0,0885	0,0989	0,1024	0,0992	0,0906	0,0786
16	0,0217	0,0367	0,0543	0,0719	0,0866	0,0960	0,0992	0,0963	0,0884
17	0,0128	0,0237	0,0383	0,0550	0,0713	0,0847	0,0934	0,0963	0,0936
18	0,0071	0,0145	0,0255	0,0397	0,0554	0,0706	0,0830	0,0909	0,0936
19	0,0037	0,0084	0,0161	0,0272	0,0409	0,0558	0,0699	0,0814	0,0887
20	0,0019	0,0046	0,0097	0,0177	0,0286	0,0418	0,0559	0,0692	0,0798
21	0,0009	0,0024	0,0055	0,0109	0,0191	0,0299	0,0426	0,0560	0,0684
22	0,0004	0,0012	0,0030	0,0065	0,0121	0,0204	0,0310	0,0433	0,0560
23	0,0002	0,0006	0,0016	0,0037	0,0074	0,0133	0,0216	0,0320	0,0438
24	0,0001	0,0003	0,0008	0,0020	0,0043	0,0083	0,0144	0,0227	0,0329
25		0,0001	0,0004	0,0010	0,0024	0,0050	0,0092	0,0154	0,0237
26			0,0002	0,0005	0,0013	0,0029	0,0057	0,0101	0,0164
27			0,0001	0,0002	0,0007	0,0016	0,0034	0,0063	0,0109
28				0,0001	0,0003	0,0009	0,0019	0,0039	0,0070
29					0,0002	0,0004	0,0010	0,0023	0,0044
30					0,0001	0,0002	0,0006	0,0013	0,0026
31						0,0001	0,0003	0,0007	0,0015
32						0,0001	0,0001	0,0004	0,0009
33							0,0001	0,0002	0,0005
34								0,0001	0,0002
35									0,0001
36									0,0001

Chapitre 7

Lois continues usuelles

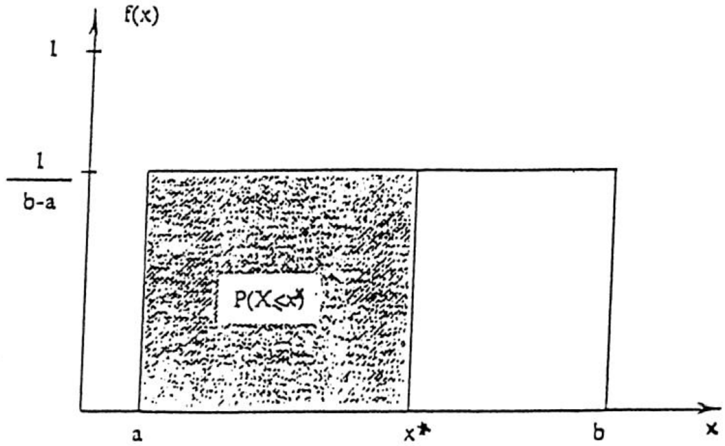
7.1. La loi continue uniforme

La loi discrète uniforme a un équivalent dans le cas continu : il s'agit d'une loi dont la densité est uniforme (constante) sur un intervalle. On a déjà rencontré ce type de loi au Chapitre 4.

Définition 7.1 *Une variable aléatoire X suit une loi continue uniforme sur un intervalle $[a, b]$ si sa densité de probabilité est :*

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \textit{pour } a \leq x \leq b \\ 0 & \textit{partout ailleurs} \end{cases}$$

On note : $X \rightsquigarrow \mathcal{U}(a; b)$.



Graphique 7.1 : La fonction de densité d'une loi continue uniforme $\mathcal{U}(a; b)$

La loi continue uniforme dépend de deux paramètres : a et b. Ces deux valeurs sont respectivement la plus petite valeur et la plus grande valeur de l'ensemble $\mathcal{X} = [a, b]$ des valeurs possibles que peut prendre la variable aléatoire X. Le Graphique 7.1, qui représente une densité uniforme, montre pourquoi on parle parfois de *loi rectangulaire*.

On vérifie aisément qu'on a bien, comme il se doit, $f(x) \geq 0$, $\forall\, x \in \mathbb{R}$, et :

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

En effet, l'aire sous la fonction de densité $f(x)$ est égale à la surface d'un rectangle de largeur $b - a$ et de hauteur $1/(b - a)$:

$$\int_{-\infty}^{+\infty} f(x)dx = (b-a)\frac{1}{b-a} = 1$$

L'espérance et la variance de X sont faciles à calculer. Par application de la Définition 4.10 et des Propriétés 4.7 et 4.8 (ou de façon équivalente de la Propriété 4.5), on trouve :

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b \frac{x}{b-a}dx = \left[\frac{x^2}{2(b-a)}\right]_a^b \\ &= \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(a+b)}{2(b-a)} = \frac{a+b}{2} \end{aligned}$$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_a^b \frac{x^2}{b-a}dx = \left[\frac{x^3}{3(b-a)}\right]_a^b \\ &= \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(a^2+ab+b^2)}{3(b-a)} = \frac{a^2+ab+b^2}{3} \end{aligned}$$

et donc :

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{(b-a)^2}{12} \end{aligned}$$

Propriété 7.1 *Si $X \rightsquigarrow \mathcal{U}(a; b)$, alors :*

$$E(X) = \frac{a+b}{2} \quad \textit{et} \quad V(X) = \frac{(b-a)^2}{12}$$

On note que l'espérance de X est le milieu de l'intervalle $[a, b]$.

Pour le calcul pratique des probabilités, il est utile de déterminer la fonction de répartition de X. Par application de la Définition 4.5, on obtient :

$$F(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{pour } x < a \\ \int_a^x f(t)dt & \text{pour } a \leq x \leq b \\ 1 & \text{pour } x > b \end{cases}$$

où :

$$\int_0^x f(t)dt = \int_0^x \frac{1}{b-a}dt = \left[\frac{t}{(b-a)}\right]_a^x = \frac{x-a}{b-a}$$

Propriété 7.2 *Si $X \rightsquigarrow \mathcal{U}(a;b)$, alors :*

$$F(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & \textit{pour } x < a \\ \dfrac{x-a}{b-a} & \textit{pour } a \leq x \leq b \\ 1 & \textit{pour } x > b \end{cases}$$

7.2. La loi normale

La loi normale est certainement la loi la plus importante pour les applications statistiques. De nombreuses épreuves aléatoires peuvent être représentées par une variable aléatoire suivant une loi normale, ou une loi très proche de celle-ci. Par exemple, la durée de vie d'une machine, la température journalière moyenne en hiver, la taille d'un individu pris au hasard, etc... Malgré son appellation malencontreuse de loi normale, elle est cependant loin de décrire tous les phénomènes aléatoires et on se gardera de considérer comme “anormale” une variable aléatoire ne suivant pas une loi normale. En fait, l'importance de la loi normale en statistique provient pour une large part du rôle central qu'elle joue dans les problèmes d'échantillonnage (cf. Chapitre 9).

La loi normale, qu'on appelle aussi loi de LAPLACE[8] – GAUSS[9] ou loi gaussienne, dépend de deux paramètres notés m et σ^2 (le second paramètre σ^2 est noté comme un carré pour simplifier l'écriture de sa racine carré : $\sqrt{\sigma^2} = \sigma$, qui sera abondamment utilisée dans la suite). m est un nombre réel quelconque tandis que σ^2 est un nombre réel strictement positif.

Définition 7.2 *Une variable aléatoire X suit une loi normale de paramètres m et σ^2, où $m \in \mathbb{R}$ et $\sigma^2 > 0$, si sa densité de probabilité est :*

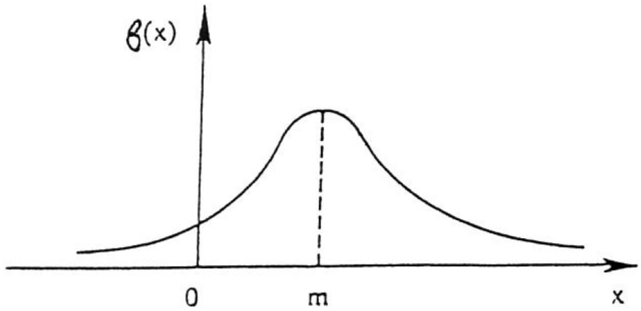
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x-m)^2}{2\sigma^2}}, \quad \forall\, x \in \mathbb{R}$$

On note : $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$.

[8] Pierre Simon de LAPLACE (1749–1827), mathématicien français.
[9] Carl-Friedrich GAUSS (1777–1855), mathématicien allemand.

L'ensemble $\mathcal{X}$ des valeurs possibles d'une variable aléatoire normale $\mathcal{N}(m;\sigma^2)$ est $\mathcal{X} = (-\infty, +\infty)$. Le Graphique 7.2 représente sa fonction de densité $f(x)$, qu'on appelle parfois *courbe de Gauss*. On le voit, la fonction de densité d'une loi normale $\mathcal{N}(m;\sigma^2)$ est une courbe en cloche, unimodale (elle possède un maximum global unique en $x = m$), symétrique par rapport à la verticale en m (on a en effet : $f(m-x) = f(m+x)$) et donc les queues s'approchent de zéro (sans néanmoins jamais s'annuler) lorsque x tend vers $+\infty$ ou $-\infty$:

$$\lim_{x\to-\infty} f(x) = \lim_{x\to+\infty} f(x) = 0$$



Graphique 7.2 : La fonction de densité d'une loi normale $\mathcal{N}(m;\sigma^2)$

On vérifie aisément que, comme il se doit, $f(x) \geq 0, \forall\, x \in \mathbb{R}$. On peut également montrer qu'on a bien :

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 1$$

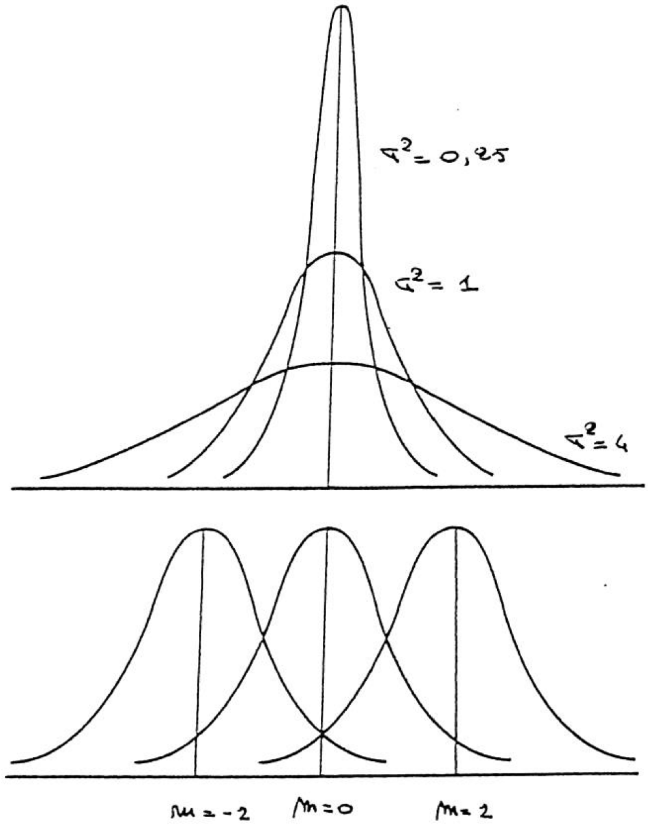
On peut encore prouver le résultat suivant, qui peut être obtenu par des calculs d'intégrales semblables à ceux détaillés aux Sections 4.2.3 et 7.1.

Propriété 7.3 *Si $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$, alors :*

$$E(X) = m \quad \textit{et} \quad V(X) = \sigma^2$$

Ainsi, les paramètres m et σ^2 ont une interprétation simple : m est l'espérance de la variable aléatoire X tandis que σ^2 est sa variance, et donc $\sigma = \sqrt{\sigma^2}$ son écart-type.

Le Graphique 7.3 montre l'influence des paramètres m et σ^2 sur l'allure de la fonction de densité de la loi normale $\mathcal{N}(m;\sigma^2)$. On voit que lorsqu'on modifie m, en gardant σ^2 inchangé, la densité normale est simplement translatée tandis que lorsqu'on modifie σ^2, en gardant m inchangé, c'est la "concentration" de la densité autour de m qui est modifiée.



Graphique 7.3 : Influence de m et σ^2 sur l'allure de la densité de la loi $\mathcal{N}(m;\sigma^2)$

7.2.1. loi normale centrée réduite

Parmi les lois normales, il y en a une qui est particulièrement importante car, comme nous le verrons plus tard, d'une certaine manière les autres s'y ramènent. Il s'agit de la *loi normale centrée réduite* ou encore *loi normale standardisée.*

Définition 7.3 *On dit qu'une variable aléatoire, que l'on désigne alors souvent par Z, suit une loi normale centrée réduite si elle suit une loi normale d'espérance nulle et de variance égale à 1. On note : $Z \rightsquigarrow \mathcal{N}(0;1)$.*

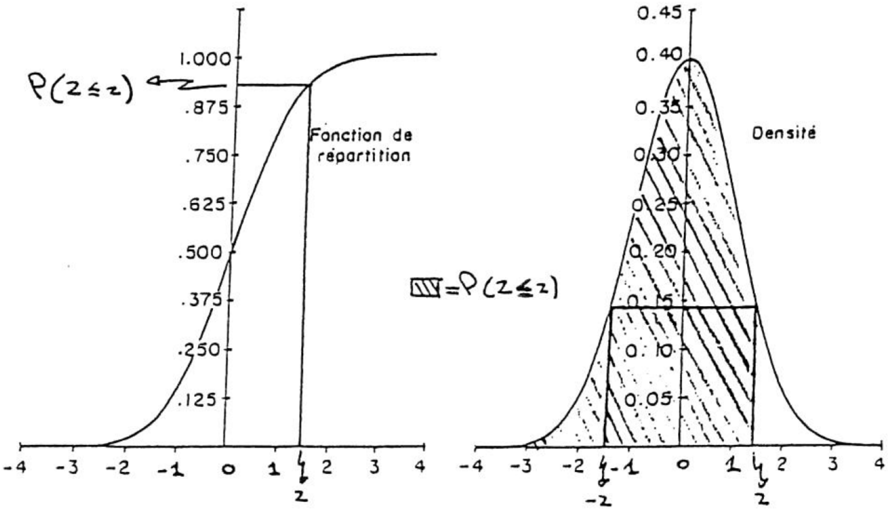
La fonction de densité d'une variable aléatoire Z normale centrée réduite, que l'on note habituellement $\phi(z)$, est obtenue en fixant m et σ^2 à respectivement 0 et 1 dans la densité donnée à la Définition 7.2. On a donc :

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \forall\, z \in \mathbb{R}$$

La fonction de répartition de cette variable Z, qui donne la probabilité $I\!P(Z \leq z)$

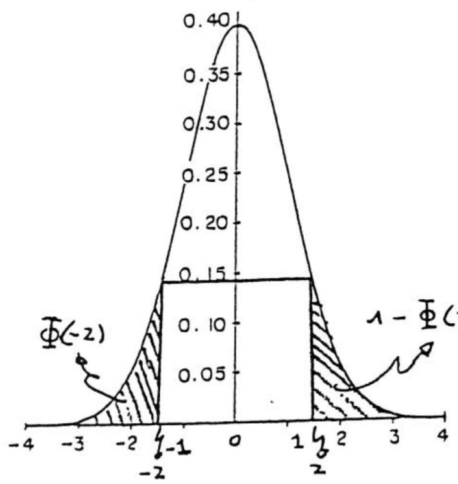
et que l'on note habituellement $\Phi(z)$, est :

$$\Phi(z) = I\!P(Z \leq z) = \int_{-\infty}^{z} \phi(t)dt = \int_{-\infty}^{z} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad \forall\, z \in \mathbb{R} \tag{7.1}$$



Graphique 7.4 : La fonction de répartition $\Phi(z)$ et la densité $\phi(z)$

L'intégrale (7.1) n'a pas de solution analytique : elle ne peut être évaluée que numériquement. En pratique, on peut obtenir les probabilités $\Phi(z) = I\!P(Z \leq z)$ en consultant une table de la fonction de répartition de la loi normale centrée réduite. Une telle table donne, pour une grille assez serrée de valeurs de z, les valeurs de $\Phi(z)$. Nous en avons reproduit un exemplaire à la fin de ce chapitre, à la Section 7.8. En utilisant cette table, on trouve par exemple que la probabilité $\Phi(1,14) = I\!P(Z \leq 1,14)$ est égale à $0,8729$.



Graphique 7.5 : $\phi(z)$ est symétrique, et donc $\Phi(-z) = 1 - \Phi(z)$

Notons que cette table ne donne les valeurs de $\Phi(z)$ que pour des valeurs de

z positives. On peut obtenir les valeurs $\Phi(z)$ pour des valeurs de z négatives en utilisant le fait que, comme la densité $\phi(z)$ est symétrique, on a (comparez les surfaces sur le Graphique 7.5) :

$$\Phi(-z) = I\!P(Z \leq -z) = I\!P(Z \geq z) = 1 - I\!P(Z \leq z) = 1 - \Phi(z)$$

Ainsi par exemple, la probabilité $\Phi(-0,96) = I\!P(Z \leq -0,96)$ est égale à $1-\Phi(0,96)$. En utilisant la table de la Section 7.8, on trouve : $1-\Phi(0,96) = 1-0,8315 = 0,1685$.

Connaissant les valeurs de $\Phi(z)$ pour les diverses valeurs de z, on peut calculer la probabilité que Z prenne sa valeur dans n'importe quel intervalle $[a, b]$ à l'aide de la relation :

$$I\!P(a \leq Z \leq b) = I\!P(Z \leq b) - I\!P(Z \leq a) = \Phi(b) - \Phi(a)$$

Lorsque l'intervalle est centré en 0, on a plus simplement :

$$I\!P(-a \leq Z \leq a) = \Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1$$

Il arrive fréquemment que plutôt que de s'intéresser à la probabilité $\Phi(z) = I\!P(Z \leq z)$ pour une valeur de z donnée, on recherche la valeur z_α qui est telle la probabilité $\Phi(z_\alpha) = I\!P(Z \leq z_\alpha)$ soit égale à une valeur donnée α.

Définition 7.4 *On appelle quantile d'ordre α, où $0 \leq \alpha \leq 1$, de la loi normale centrée réduite $\mathcal{N}(0, 1)$ la quantité z_α qui est telle que :*

$$\Phi(z_\alpha) = I\!P(Z \leq z_\alpha) = \alpha$$

La notion de quantile d'ordre α n'est pas propre à la loi normale centrée réduite : de façon générale, on appelle quantile d'ordre α de la loi d'une variable aléatoire quelconque X la quantité x_α qui est telle que $I\!P(X \leq x_\alpha) = \alpha$. Remarquons que le quantile d'ordre $0,5$ n'est rien d'autre que la *médiane* : c'est la valeur $x_{0,5}$ qui est telle que $I\!P(X \leq x_{0,5}) = I\!P(X \geq x_{0,5}) = 0,5$. Dans le cas de la loi normale centrée réduite, la médiane est égale à 0 : $z_{0,5} = 0$.

Les quantiles d'ordre α de la loi normale centrée réduite $\mathcal{N}(0, 1)$ sont repris dans des tables, qui donnent z_α pour une grille de valeurs de α et dont nous avons à nouveau reproduit un exemplaire à la Section 7.8 ci-dessous. En utilisant cette table, on voit par exemple que le quantile d'ordre $0,95$ est égal à $1,6449$.

7.2.2. Loi normale quelconque et loi normale centrée réduite

On peut aisément passer d'une loi normale quelconque à une loi normale centrée réduite, et vice versa. Cette possibilité repose sur le résultat suivant, qui peut assez facilement être vérifié en utilisant la Propriété 4.6.

Propriété 7.4 *Soit* $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$*. Si* $Y = a + bX$*, alors :*

$$Y \rightsquigarrow \mathcal{N}(a + bm; b^2\sigma^2)$$

En d'autres termes, une fonction linéaire d'une variable aléatoire normale est encore une variable aléatoire normale, dont l'espérance et la variance découlent simplement des Propriétés 4.10 et 4.11. On en déduit immédiatement la propriété suivante.

Propriété 7.5 *Si* $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$*, alors :*

$$Z = \frac{X - m}{\sigma} \rightsquigarrow \mathcal{N}(0; 1)$$

En sens opposé, si $Z \rightsquigarrow \mathcal{N}(0; 1)$*, alors :*

$$X = m + \sigma Z \rightsquigarrow \mathcal{N}(m; \sigma^2)$$

En effet, si $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$ et $Z = (X - m)/\sigma$ (= une fonction linéaire de X), des propriétés 4.10 et 4.11, on a $E(Z) = 0$ et $V(Z) = 1$, et de la Propriété 7.4, $Z \rightsquigarrow \mathcal{N}(0; 1)$. De la même façon, si $Z \rightsquigarrow \mathcal{N}(0; 1)$ et $X = m + \sigma Z$ (= une fonction linéaire de Z), des propriétés 4.10 et 4.11, on a $E(X) = m$ et $V(X) = \sigma^2$, et de la Propriété 7.4, $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$.

La possibilité de passer, au travers d'une transformation simple, d'une loi normale quelconque à une loi normale centrée réduite permet de calculer des probabilités se rapportant à des lois normales quelconques en utilisant uniquement une table se rapportant à la loi normale centrée réduite.

Ainsi, supposons que l'on veuille calculer la probabilité $\mathbb{P}(X \leq a)$ dans une loi normale $\mathcal{N}(m; \sigma^2)$. On a :

$$\begin{aligned} \mathbb{P}(X \leq a) &= \mathbb{P}\left(\frac{X - m}{\sigma} \leq \frac{a - m}{\sigma}\right) = \mathbb{P}\left(Z \leq \frac{a - m}{\sigma}\right) \\ &= \Phi\left(\frac{a - m}{\sigma}\right) \end{aligned}$$

et on obtient cette valeur en consultant une table de la fonction de répartition de la loi $\mathcal{N}(0; 1)$. Par exemple, la probabilité $\mathbb{P}(X \leq 2)$ dans une loi normale $\mathcal{N}(3; 4)$ est égale à $\Phi\left((2 - 3)/\sqrt{4}\right) = \Phi(-0,5) = 1 - \Phi(0,5) = 1 - 0,6915 = 0,3085$.

Si on cherche la probabilité que X prenne sa valeur dans un intervalle, on a pareillement :

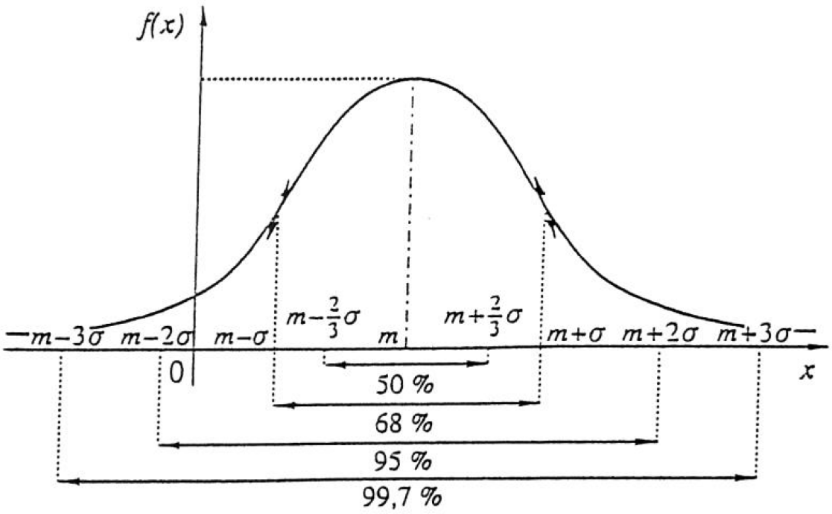
$$\begin{aligned} \mathbb{P}(a \leq X \leq b) &= \mathbb{P}\left(\frac{a - m}{\sigma} \leq \frac{X - m}{\sigma} \leq \frac{b - m}{\sigma}\right) = \mathbb{P}\left(\frac{a - m}{\sigma} \leq Z \leq \frac{b - m}{\sigma}\right) \\ &= \Phi\left(\frac{b - m}{\sigma}\right) - \Phi\left(\frac{a - m}{\sigma}\right) \end{aligned}$$

et on obtient à nouveau cette valeur en consultant la même table. Par exemple, la probabilité $I\!P(-1 \leq X \leq 2)$ dans une loi normale $\mathcal{N}(3;4)$ est égale à $\Phi\left((2-3)/\sqrt{4}\right) - \Phi\left((-1-3)/\sqrt{4}\right) = \Phi(-0,5) - \Phi(-2) = \Phi(2) - \Phi(0,5) = 0,9772 - 0,6915 = 0,2857$.

Supposons que $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$ et considérons les probabilités des événements du type "$m - \lambda\sigma \leq X \leq m + \lambda\sigma$", où $\lambda \geq 0$. On a :

$$\begin{aligned} I\!P(m - \lambda\sigma \leq X \leq m + \lambda\sigma) &= I\!P\left(-\lambda \leq \frac{X-m}{\sigma} \leq \lambda\right) \\ &= I\!P(-\lambda \leq Z \leq \lambda) = 2\Phi(\lambda) - 1 \end{aligned}$$

Pour λ égal à $0,67$, 1, 2 et 3, on trouve respectivement $49,7\%$, $68,3\%$, $95,4\%$ et $99,7\%$. Ces valeurs sont résumées par le Graphique 7.6. Elles permettent de se faire une idée assez précise de la "répartition des probabilités" dans une loi normale. On le voit, la quasi-totalité (95%) de la probabilité est concentrée dans l'intervalle $m \pm 2\sigma$. Bien qu'une variable aléatoire normale soit une variable qui peut prendre toutes les valeurs réelles entre $-\infty$ et $+\infty$, les valeurs hors de l'intervalle $m \pm 3\sigma$ sont extrêmement peu probables. Cette dernière caractéristique permet de comprendre que la loi normale, tout en n'étant pas bornée, est parfois utilisée comme approximation pour décrire des phénomènes bornés.



Graphique 7.6 : Intervalles typiques de la loi $\mathcal{N}(m,\sigma^2)$

7.2.3. Exercice résolu

Une compagnie d'assurance a engagé une nouvelle équipe de secrétaires chargées de dactylographier le courrier. A cette occasion, elle analyse la charge qui incombera à la nouvelle équipe. Le nombre de lettres nécessitant une réponse écrite arrivant chaque semaine à la compagnie est modélisé comme suivant une loi normale d'espérance $m = 2\,500$ et de variance $\sigma^2 = 40\,000$ (et donc d'écart-type $\sigma = 200$).

1- Calculez la probabilité que la compagnie reçoive durant une semaine (a) plus de 2 750 lettres, (b) moins de 2 250 lettres et (c) entre 2 250 et 2 750 lettres.

Notons X la variable aléatoire représentant le nombre de lettres arrivant durant une semaine à la compagnie. On a :

$$X \rightsquigarrow \mathcal{N}(2\,500; 40\,000) \text{ et donc } Z = \frac{X - 2\,500}{200} \rightsquigarrow \mathcal{N}(0;1)$$

En utilisant la table de la fonction de répartition de la loi $\mathcal{N}(0;1)$, on trouve pour les probabilités demandées :

$$\begin{aligned} I\!P(X > 2\,750) &= I\!P(X \geq 2\,750) = 1 - I\!P(X \leq 2\,750) \\ &= 1 - I\!P\left(\frac{X - 2\,500}{200} \leq \frac{2\,750 - 2\,500}{200}\right) \\ &= 1 - I\!P(Z \leq 1,25) = 1 - \Phi(1,25) = 1 - 0,8944 = 0,1056 \end{aligned}$$

$$\begin{aligned} I\!P(X < 2\,250) &= I\!P(X \leq 2\,250) = I\!P\left(\frac{X - 2\,500}{200} \leq \frac{2\,250 - 2\,500}{200}\right) \\ &= I\!P(Z \leq -1,25) = \Phi(-1,25) = 1 - \Phi(1,25) = 0,1056 \end{aligned}$$

$$\begin{aligned} I\!P(2\,250 < X < 2\,750) &= I\!P(X \leq 2\,750) - I\!P(X \leq 2\,250) \\ &= (1 - 0,1056) - 0,1056 = 0,7888 \end{aligned}$$

2- Déterminez le nombre de lettres n tel que la probabilité de recevoir au cours d'une semaine un nombre de lettres inférieur à n est de $0,75$.

On cherche n tel que $I\!P(X < n) = I\!P(X \leq n) = 0,75$, soit n tel que :

$$\begin{aligned} 0,75 &= I\!P(X \leq n) = I\!P\left(\frac{X - 2\,500}{200} \leq \frac{n - 2\,500}{200}\right) \\ &= I\!P(Z \leq \frac{n - 2\,500}{200}) = \Phi\left(\frac{n - 2\,500}{200}\right) \end{aligned}$$

En regardant dans la table des quantiles de la loi $\mathcal{N}(0;1)$, on trouve $\Phi(0,6745) = 0,75$ ($z_{0,75} = 0,6745$: quantile d'ordre $0,75$). On a donc :

$$0,6745 = \frac{n - 2\,500}{200}$$

d'où :

$$n = 0,6745 \times 200 + 2.500 = 2\,634,9 \approx 2\,635$$

3- Dans quel intervalle centré en m a-t-on 9 chances sur 10 de trouver le nombre de lettres reçues au cours d'une semaine ?

On cherche un intervalle du type $[2\,500 - a\ ;\ 2\,500 + a]$ tel que :

$$
\begin{aligned}
0,90 &= I\!P(2\,500 - a \leq X \leq 2\,500 + a) \\
&= I\!P\left(\frac{2\,500 - a - 2\,500}{200} \leq \frac{X - 2\,500}{200} \leq \frac{2\,500 + a - 2\,500}{200}\right) \\
&= I\!P\left(\frac{-a}{200} \leq Z \leq \frac{a}{200}\right)
\end{aligned}
$$

soit la valeur a telle que :

$$2\Phi\left(\frac{a}{200}\right) - 1 = 0,90 \Leftrightarrow \Phi\left(\frac{a}{200}\right) = I\!P\left(Z \leq \frac{a}{200}\right) = 0,95$$

En regardant dans la table des quantiles de la loi $\mathcal{N}(0;1)$, on trouve $\Phi(1,6449) = 0,95$ ($z_{0,95} = 1,6449$: quantile d'ordre $0,95$). On a donc :

$$1,6449 = \frac{a}{200}$$

d'où :

$$a = 1,6449 \times 200 = 328,98$$

L'intervalle demandé est donc :

$$[2\,500 - 328,98\ ;\ 2\,500 + 328,98] \approx [2\,171\ ;\ 2\,829]$$

7.2.4. Autres propriétés de la loi normale

Sous certaines conditions, la loi normale peut être utilisée comme approximation d'une loi de Poisson ou d'une loi binomiale. Comme pour l'approximation d'une loi binomiale par une loi de Poisson, d'un point de vue formel, cette possibilité repose sur un résultat de convergence en loi dont nous parlerons au Chapitre 9.

On se contentera ici de retenir les règles suivantes.

1- Lorsque λ est grand, les probabilités associées à une loi de Poisson $\mathcal{P}(\lambda)$ sont très proches des probabilités données par une loi normale de paramètres $m = \lambda$ et $\sigma^2 = \lambda$:

$$\mathcal{P}(\lambda) \approx \mathcal{N}(\lambda; \lambda)$$

Plus précisément, si $X \rightsquigarrow \mathcal{P}(\lambda)$ et si $Y \rightsquigarrow \mathcal{N}(\lambda; \lambda)$, on a :

$$
\begin{aligned}
I\!P(X = x_i) &= e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}, \quad \forall\, x_i \in \{0, 1, 2, ...\} \\
&\approx I\!P(x_i - 0,5 \leq Y \leq x_i + 0,5) \qquad (7.2)
\end{aligned}
$$

et donc :

$$\mathbb{P}(X = x_i) \approx \Phi\left(\frac{x_i + 0,5 - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{x_i - 0,5 - \lambda}{\sqrt{\lambda}}\right), \ \forall\, x_i \in \{0, 1, 2, ...\}$$

Remarquez la *correction de continuité* (pour "$X = x_i$", on considère "$x_i - 0,5 \leq Y \leq x_i + 0,5$") introduite pour tenir compte du fait que l'on approxime une loi *discrète* par une loi *continue*. Remarquez également que les deux lois $\mathcal{P}(\lambda)$ et $\mathcal{N}(\lambda; \lambda)$ ont même espérance et même variance. En pratique, nous admettrons que cette approximation est "bonne" lorsque $\lambda \geq 15$.

2- Lorsque n est grand et que p n'est ni trop petit ni trop grand, les probabilités associées à une loi binomiale $\mathcal{B}(n; p)$ sont très proches des probabilités données par une loi normale de paramètres $m = np$ et $\sigma^2 = npq$ $(q = 1 - p)$:

$$\mathcal{B}(n; p) \approx \mathcal{N}(np; npq)$$

Plus précisément, si $X \rightsquigarrow \mathcal{B}(n; p)$ et si $Y \rightsquigarrow \mathcal{N}(np; npq)$, on a :

$$\begin{aligned} \mathbb{P}(X = x_i) &= C_n^{x_i} p^{x_i} q^{n-x_i}, \quad \forall\, x_i \in \{0, 1, 2..., n\} \\ &\approx \mathbb{P}(x_i - 0,5 \leq Y \leq x_i + 0,5) \end{aligned} \tag{7.3}$$

et donc :

$$\mathbb{P}(X = x_i) \approx \Phi\left(\frac{x_i + 0,5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{x_i - 0,5 - np}{\sqrt{npq}}\right), \ \forall\, x_i \in \{0, 1, 2..., n\}$$

Remarquez à nouveau la correction de continuité et le fait que les deux lois $\mathcal{B}(n; p)$ et $\mathcal{N}(np; npq)$ ont même espérance et même variance. Dans le cadre de ce cours, nous que admettrons cette approximation est "bonne" lorsque $n \geq 30$, $np \geq 5$ et $nq \geq 5$. Rappelons que l'approximation d'une loi binomiale par une loi de Poisson ne s'applique que lorsque n est grand, p petit et np pas trop grand.

Si la correction de continuité qui apparaît dans les formules (7.2) et (7.3) est importante — et même essentielle puisque sans elle l'approximation donnerait toujours une valeur nulle, ce qui n'est manifestement pas très satisfaisant — lorsqu'on désire évaluer, par une approximation normale, la probabilité qu'une variable aléatoire X qui suit une loi de Poisson ou binomiale prenne une valeur ponctuelle "$X = x_i$", elle l'est nettement moins, voire plus du tout, lorsqu'on désire évaluer par cette même approximation la probabilité que X prenne sa valeur dans un intervalle : $\mathbb{P}(a \leq X \leq b)$. Dans ce cas de figure, la mise en oeuvre de la correction de continuité voudrait que l'on évalue $\mathbb{P}(a \leq X \leq b)$ par $\mathbb{P}(a - 0,5 \leq Y \leq b + 0,5)$, où Y suit une loi normale selon le cas $\mathcal{N}(\lambda; \lambda)$ ou $\mathcal{N}(np; npq)$. Cependant, à moins que la longueur l'intervalle $[a, b]$ ne soit très petite (quelques unités, disons moins de 5), on a $\mathbb{P}(a \leq X \leq b) \approx \mathbb{P}(a - 0,5 \leq Y \leq b + 0,5) \approx \mathbb{P}(a \leq Y \leq b)$. En d'autres termes, dans ce cas de figure, la correction de continuité est négligeable. En pratique, lorsque nous serons amené à évaluer la probabilité que X prenne sa valeur dans un intervalle autre que très petit, nous la négligerons toujours.

Pour conclure, notons encore une propriété très importante des lois normales, dont nous ferons grand usage dans la suite du cours.

Propriété 7.6 *Soient* $X_1 \rightsquigarrow \mathcal{N}(m_1;\sigma_1^2)$ *et* $X_2 \rightsquigarrow \mathcal{N}(m_2;\sigma_2^2)$. *Si* X_1 *et* X_2 *sont indépendantes, alors :*

$$X = a + bX_1 + cX_2 \rightsquigarrow \mathcal{N}(m;\sigma^2)$$

où :

$$m = a + bm_1 + cm_2 \quad et \quad \sigma^2 = b^2\sigma_1^2 + c^2\sigma_1^2$$

En d'autres termes, une combinaison linéaire de variables aléatoires normales indépendantes est encore une variable aléatoire normale, dont l'espérance ($= m$) et la variance ($= \sigma^2$) découlent simplement des Propriétés 5.5 et 5.7 (X_1 et X_2 étant indépendantes, on a $Cov(X_1, X_2) = 0$)

Ce résultat est aussi valable pour une combinaison linéaire de r variables aléatoires normales indépendantes.

Propriété 7.7 *Soient* $X_1 \rightsquigarrow \mathcal{N}(m_1;\sigma_1^2), X_2 \rightsquigarrow \mathcal{N}(m_2;\sigma_2^2), ..., X_r \rightsquigarrow \mathcal{N}(m_r;\sigma_r^2)$. *Si* $X_1, X_2, ..., X_r$ *sont indépendantes, alors :*

$$X = a_0 + \sum_{i=1}^{r} a_i X_i \rightsquigarrow \mathcal{N}(m;\sigma^2)$$

où :

$$m = a_0 + \sum_{i=1}^{r} a_i m_i \quad et \quad \sigma^2 = \sum_{i=1}^{r} a_i^2 \sigma_i^2$$

Les valeurs de m et σ^2 découlent ici simplement des Propriétés 5.15 et 5.16.

7.2.5. Exercice résolu

La proportion de ménages qui possèdent un magnétoscope est de 60%. On sélectionne au hasard et indépendamment 500 ménages.

1- Calculez à l'aide d'une approximation la probabilité que, sur les 500 ménages sélectionnés, 280 ménages possèdent un magnétoscope.

On associe à chacun des 500 ménages une variable aléatoire de Bernoulli, qui prend la valeur 1 si le ménage possède un magnétoscope et la valeur 0 sinon. On a ainsi $X_1, X_2, ..., X_{500}$, 500 variables aléatoires indépendantes et suivant une loi de Bernoulli $\mathcal{B}(0,6)$.
On définit la variable aléatoire $X = \sum_{i=1}^{500} X_i$ représentant le nombre des ménages sélectionnés qui possèdent un magnétoscope. On a :

$$X \rightsquigarrow \mathcal{B}(500; 0,6)$$

Comme $n = 500 \geq 30, np = 300 \geq 5$ et $nq = 200 \geq 5$, on réalise une approximation par une loi normale. On a :

$$\mathcal{B}(500; 0,6) \approx \mathcal{N}(300; 120)$$

La probabilité demandée est $I\!P(X = 280)$. On l'approche de la manière suivante (notez la correction de continuité) :

$$I\!P(X = 280) \approx I\!P(279,5 \leq Y \leq 280,5)$$

où Y désigne une variable aléatoire qui suit la loi $\mathcal{N}(300; 120)$. On a :

$$Z = \frac{Y - 300}{\sqrt{120}} \rightsquigarrow \mathcal{N}(0; 1)$$

et donc (aux arrondis près) :

$$\begin{aligned} I\!P(279,5 \leq Y \leq 280,5) &= I\!P\left(\frac{279,5 - 300}{\sqrt{120}} \leq \frac{Y - 300}{\sqrt{120}} \leq \frac{280,5 - 300}{\sqrt{120}}\right) \\ &= I\!P(-1,87 \leq Z \leq -1,78) \\ &= \Phi(-1,78) - \Phi(-1,87) = \Phi(1,87) - \Phi(1,78) \end{aligned}$$

En utilisant la table de la fonction de répartition de la loi $\mathcal{N}(0; 1)$, on trouve :

$$\Phi(1,87) - \Phi(1,78) = 0,9693 - 0,9625 = 0,0068$$

et donc finalement :

$$I\!P(X = 280) \approx 0,0068$$

2- Calculez à l'aide d'une approximation la probabilité que, sur les 500 ménages sélectionnés, le nombre de ménages possédant un magnétoscope soit supérieur à 260.

La probabilité demandée est $I\!P(X > 260) = I\!P(X \geq 261)$. De façon semblable à ci-dessus, on l'approche de la manière suivante (on peut ici négliger la correction de continuité) :

$$I\!P(X \geq 261) \approx I\!P(Y \geq 261), \quad \text{où } Y \rightsquigarrow \mathcal{N}(300; 120)$$

On a (aux arrondis près) :

$$\begin{aligned} I\!P(Y \geq 261) &= 1 - I\!P(Y \leq 261) = 1 - I\!P\left(\frac{Y - 300}{\sqrt{120}} \leq \frac{261 - 300}{\sqrt{120}}\right) \\ &= 1 - I\!P(Z \leq -3,56) = 1 - \Phi(-3,56) = \Phi(3,56) = 0,9998 \end{aligned}$$

soit finalement :

$$I\!P(X > 260) \approx 0,9998$$

3- Calculez à l'aide d'une approximation la probabilité que, sur les 500 ménages sélectionnés, le nombre de ménages possédant un magnétoscope soit compris (bornes incluses) entre 280 et 320.

La probabilité demandée est $I\!P(280 \leq X \leq 320)$. A nouveau de façon semblable à ci-dessus, on l'approche de la manière suivante (on peut ici aussi négliger la correction de continuité) :

$$I\!P(280 \leq X \leq 320) \approx I\!P(280 \leq Y \leq 320), \quad \text{où } Y \rightsquigarrow \mathcal{N}(300; 120)$$

On a (aux arrondis près) :

$$\begin{aligned} I\!P(280 \leq Y \leq 320) &= I\!P\left(\frac{280-300}{\sqrt{120}} \leq \frac{Y-300}{\sqrt{120}} \leq \frac{320-300}{\sqrt{120}}\right) \\ &= I\!P(-1,82 \leq Z \leq 1,82) = 2\Phi(1,82) - 1 \\ &= (2 \times 0,9656) - 1 = 0,9312 \end{aligned}$$

et donc :

$$I\!P(280 \leq X \leq 320) \approx 0,9312$$

7.3. Lois liées à la loi normale

Plusieurs lois très utilisées peuvent être définies par référence à la loi normale. Nous nous limiterons aux plus importantes : la loi du khi-carré, la loi de Student et la loi de Fisher–Snedecor. Ces lois tirent la majeure partie de leur importance du rôle qu'elles jouent dans les problèmes d'échantillonnage (cf. Chapitre 9). On se bornera ici à les définir et à brièvement décrire leurs principales caractéristiques.

7.3.1. La loi du khi-carré

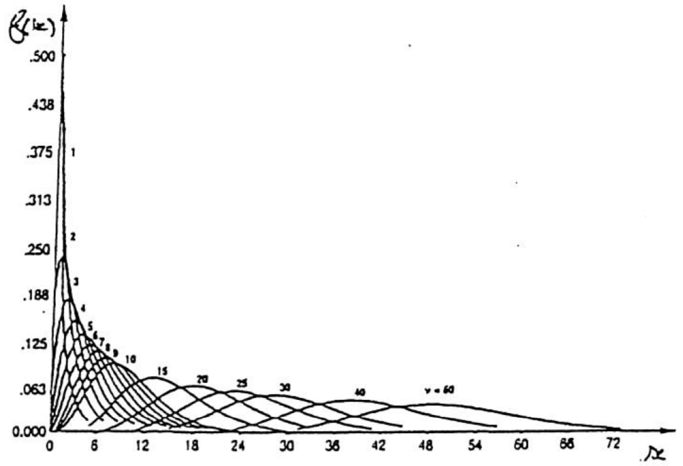
La loi du khi-carré, qu'on appelle aussi loi du khi-deux, est une loi continue dépendant d'un paramètre qu'on appelle *nombre de degrés de liberté*. Ce paramètre, souvent noté ν, est un nombre entier strictement positif. L'ensemble $\mathcal{X}$ des valeurs possibles d'une variable aléatoire distribuée selon cette loi est $\mathcal{X} = [0, +\infty)$. Son lien avec la loi normale apparaît directement dans sa définition.

Définition 7.5 *Si $X_1, X_2, ..., X_\nu$ sont ν variables aléatoires $\mathcal{N}(0;1)$ indépendantes, alors la variable aléatoire X définie par :*

$$X = X_1^2 + X_2^2 + ... + X_\nu^2$$

suit une loi du khi-carré à ν degrés de liberté. On note : $X \rightsquigarrow \chi^2(\nu)$.

Le Graphique 7.7 montre l'allure de la fonction de densité de la loi du khi-carré $\chi^2(\nu)$ pour diverses valeurs du paramètre ν. On voit que la densité est unimodale et étalée à droite, mais que la dissymétrie s'estompe lorsque ν grandit. En fait, lorsque ν augmente, la densité d'une loi $\chi^2(\nu)$ ressemble de plus en plus à la densité d'une loi normale.

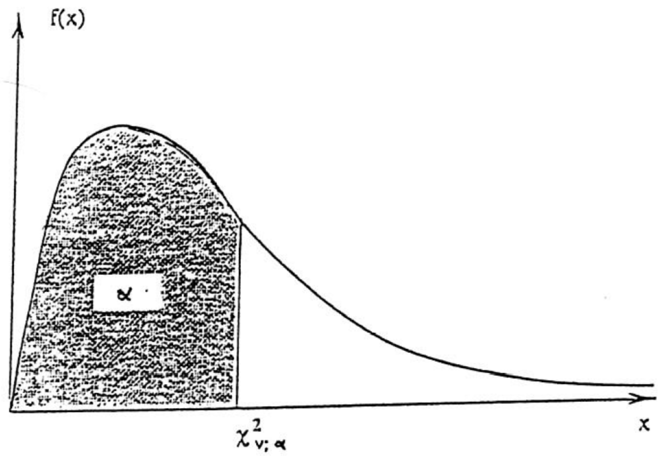


Graphique 7.7 : Quelques représentations de la densité de la loi $\chi^2(\nu)$

Propriété 7.8 *Si $X \rightsquigarrow \chi^2(\nu)$, alors :*

$$E(X) = \nu \quad et \quad V(X) = 2\nu$$

L'espérance et la variance d'une variable aléatoire suivant une loi $\chi^2(\nu)$ sont donc déterminées par le nombre de degrés de liberté. Plus celui-ci augmente, plus l'espérance et la variance augmentent.



Graphique 7.8 : Quantile d'ordre α de la loi $\chi^2(\nu)$

Définition 7.6 *On appelle quantile d'ordre α, où $0 \le \alpha \le 1$, de la loi du khi-carré à ν degrés de liberté la quantité $\chi^2_{\nu;\alpha}$ qui est telle que, si $X \rightsquigarrow \chi^2(\nu)$:*

$$\mathbb{P}(X \le \chi^2_{\nu;\alpha}) = \alpha$$

Les quantiles d'ordre α de la loi $\chi^2(\nu)$ sont repris dans des tables. Elles donnent $\chi^2_{\nu;\alpha}$ pour une grille de valeurs de ν et α. Nous en avons reproduit un exemplaire à la Section 7.8 ci-dessous. En utilisant cette table, on voit par exemple que le quantile d'ordre $0,95$ de la loi $\chi^2(3)$ est égal à $7,81$.

Mentionnons un dernier résultat.

Propriété 7.9 *Soient $X_1 \rightsquigarrow \chi^2(\nu_1)$ et $X_2 \rightsquigarrow \chi^2(\nu_2)$. Si X_1 et X_2 sont indépendantes, alors :*

$$X = X_1 + X_2 \rightsquigarrow \chi^2(\nu_1 + \nu_2)$$

En d'autres termes, la somme de deux variables aléatoires du khi-carré indépendantes est encore une variable aléatoire du khi-carré. Etant donné la définition de la loi du khi-carré, cela ne devrait guère vous étonner.

Ce résultat est aussi valable pour la somme de r variables aléatoires du khi-carré indépendantes.

Propriété 7.10 *Soient $X_1 \rightsquigarrow \chi^2(\nu_1), X_2 \rightsquigarrow \chi^2(\nu_2), ..., X_r \rightsquigarrow \chi^2(\nu_r)$. Si $X_1, X_2, ..., X_r$ sont indépendantes, alors :*

$$X = X_1 + X_2 + ... + X_r \rightsquigarrow \chi^2(\nu_1 + \nu_2 + ... + \nu_r)$$

7.3.2. La loi de Student

La loi de Student[10] est une loi continue qui, comme la loi du khi-carré, dépend d'un seul paramètre qu'on appelle également *nombre de degrés de liberté* et qu'on note ν. Ce paramètre est un nombre entier strictement positif. L'ensemble $\mathcal{X}$ des valeurs possibles d'une variable aléatoire distribuée selon cette loi est $\mathcal{X} = (-\infty, +\infty)$. La loi de Student est directement liée à la loi normale par sa définition.

Définition 7.7 *Si $X \rightsquigarrow \mathcal{N}(0;1)$, si $Y \rightsquigarrow \chi^2(\nu)$ et si X et Y sont indépendantes, alors la variable aléatoire Z définie par :*

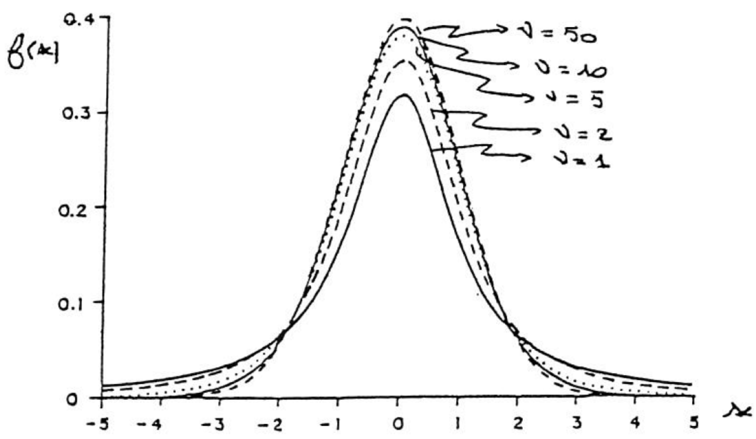
$$Z = \frac{X}{\sqrt{\dfrac{Y}{\nu}}}$$

suit une loi de Student à ν degrés de liberté. On note : $Z \rightsquigarrow t(\nu)$.

Comme le suggère le Graphique 7.9, la loi de Student ressemble beaucoup à la loi normale centrée réduite. Comme la fonction de densité de la loi normale centrée réduite, la fonction de densité de la loi de Student est unimodale, centrée et

[10] STUDENT est le pseudonyme sous lequel le mathématicien anglais William Sealy GOSSET (1876–1937) publia ses travaux.

symétrique. Elle est cependant plus "plate", elle a des queues plus épaisses. Cette différence s'amenuise à mesure le nombre de degrés de liberté grandit : lorsque ν grandit, la loi de STUDENT tend vers la loi normale centrée réduite.

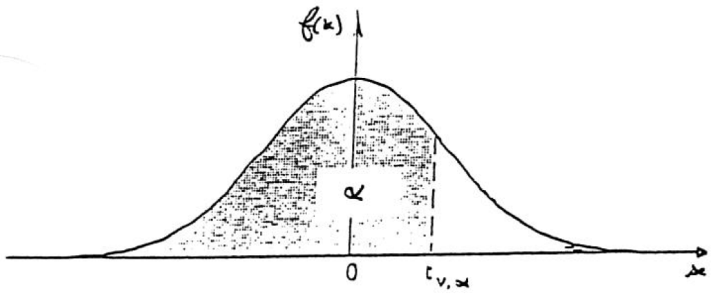


Graphique 7.9 : Quelques représentations de la densité de la loi $t(\nu)$

Propriété 7.11 *Si $X \rightsquigarrow t(\nu)$, alors :*

$$E(X) = 0 \quad \textit{et} \quad V(X) = \frac{\nu}{\nu - 2}, \;\; \textit{pour } \nu > 2$$

Seule la variance d'une variable aléatoire de STUDENT dépend du nombre de degrés de liberté. Notons que pour $\nu = 1$ ou $\nu = 2$, la variance n'est pas déterminée. Lorsque ν grandit, la variance tend vers 1.



Graphique 7.10 : Quantile d'ordre α de la loi $t(\nu)$

Définition 7.8 *On appelle quantile d'ordre α, où $0 \leq \alpha \leq 1$, de la loi de STUDENT à ν degrés de liberté la quantité $t_{\nu;\alpha}$ qui est telle que, si $X \rightsquigarrow t(\nu)$:*

$$\mathbb{P}(X \leq t_{\nu;\alpha}) = \alpha$$

Les quantiles d'ordre α de la loi $t(\nu)$ sont repris dans des tables. Elles donnent $t_{\nu;\alpha}$ pour une grille de valeurs de ν et α. Une de ces tables est reproduite à la Section

7.8 ci-dessous. En utilisant cette table, on voit par exemple que le quantile d'ordre $0,95$ de la loi $t(5)$ est égal à $2,015$.

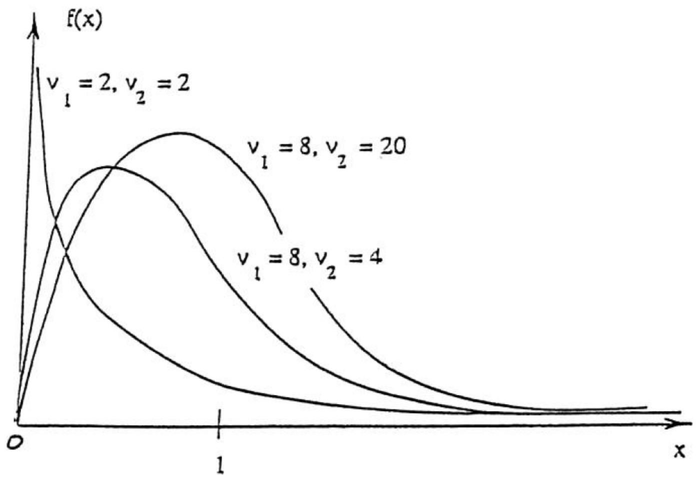
7.3.3. La loi de Fisher-Snedecor

La loi de FISHER[11] – SNEDECOR[12] est une loi continue dépendant de deux paramètres notés ν_1 et ν_2. On dit que ν_1 est le *nombre de degrés de liberté* du numérateur et que ν_2 est le *nombre de degrés de liberté* du dénominateur. L'ensemble $\mathcal{X}$ des valeurs possibles d'une variable aléatoire distribuée selon cette loi est $\mathcal{X} = [0, +\infty)$. La loi de FISHER – SNEDECOR est directement liée à la loi du khi-carré, et donc à la loi normale.

Définition 7.9 *Si $X_1 \rightsquigarrow \chi^2(\nu_1)$, si $X_2 \rightsquigarrow \chi^2(\nu_2)$ et si X_1 et X_2 sont indépendantes, alors la variable aléatoire Y définie par :*

$$Y = \frac{\frac{X_1}{\nu_1}}{\frac{X_2}{\nu_2}}$$

suit une loi de FISHER – SNEDECOR à ν_1 et ν_2 degrés de liberté. On note : $Y \rightsquigarrow F(\nu_1; \nu_2)$.



Graphique 7.11 : Quelques représentations de la densité de la loi $F(\nu_1; \nu_2)$

Le Graphique 7.11 montre l'allure de la fonction de densité de la loi de FISHER – SNEDECOR pour quelques valeurs de ν_1 et ν_2. On le voit, les courbes sont unimodales et étalées à droite. Le mode se rapproche de plus en plus de 1 lorsque les nombres de degrés de liberté augmentent.

[11] Ronald Aylmer FISHER (1890–1962), statisticien anglais.

[12] George Waddel SNEDECOR (1881–1974), statisticien américain, un des premiers utilisateurs de l'informatique en statistique.

Propriété 7.12 *Si $X \rightsquigarrow F(\nu_1, \nu_2)$, alors :*

$$E(X) = \frac{\nu_2}{\nu_2 - 2}, \text{ pour } \nu_2 > 2$$

et

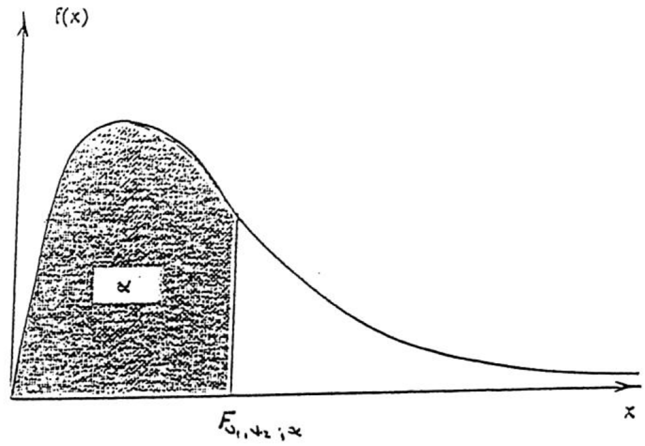
$$V(X) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \text{ pour } \nu_2 > 4$$

Notons que pour $\nu_2 \leq 2$, l'espérance n'est pas déterminée et que pour $\nu_2 \leq 4$, la variance n'est pas déterminée.

Définition 7.10 *On appelle quantile d'ordre α, où $0 \leq \alpha \leq 1$, de la loi de* FISHER – SNEDECOR *à ν_1 et ν_2 degrés de liberté la quantité $F_{\nu_1,\nu_2;\alpha}$ qui est telle que, si $X \rightsquigarrow F(\nu_1; \nu_2)$:*

$$\mathbb{P}(X \leq F_{\nu_1,\nu_2;\alpha}) = \alpha$$

Comme ceux de la loi du khi-carré et de STUDENT, les quantiles d'ordre α de la loi de FISHER – SNEDECOR à ν_1 et ν_2 degrés de liberté sont repris dans des tables. Elles donnent $F_{\nu_1,\nu_2;\alpha}$ pour $\alpha = 0,95$, $\alpha = 0,99$ et une grille de valeurs de ν_1 et ν_2. Nous en avons reproduit un exemplaire à la Section 7.8 ci-dessous. En utilisant cette table, on voit par exemple que le quantile d'ordre $0,95$ de la loi $F(3; 8)$ est égal à $4,07$.



Graphique 7.12 : Quantile d'ordre α de la loi $F(\nu_1; \nu_2)$

7.4. La loi exponentielle

On le sait, un processus de POISSON décrit un processus dont les événements surviennent aléatoirement, indépendamment et uniformément dans le temps. On a vu que pour un tel processus la variable aléatoire donnant le nombre d'événements qui surviennent au cours d'une période de temps quelconque T suit une loi de

POISSON $\mathcal{P}(\lambda)$, où λ est l'espérance du nombre d'événements qui surviennent par unité de temps T.

Pour un processus de POISSON donné, plutôt que de compter le nombre d'événements qui surviennent au cours d'une période de temps quelconque T, on peut s'intéresser au temps qui s'écoule entre deux apparitions successives d'un événement.

Considérons un processus de POISSON tel que l'espérance du nombre d'événements qui surviennent au cours d'une période de temps quelconque T est égale à λ, et désignons par X la durée, mesurée en unités de temps T ($X = 0,5$ indique un intervalle de temps égal à $T/2$, $X = 1$ un intervalle de temps égal à T, etc...), entre deux apparitions successives d'un événement. On peut montrer que la variable aléatoire X suit une loi de exponentielle de paramètre λ. Cette loi est définie de la manière suivante.

Définition 7.11 *Une variable aléatoire X suit une loi exponentielle de paramètre λ, où $\lambda > 0$, si sa densité de probabilité est :*

$$f(x) = \begin{cases} 0 & \textit{pour } x < 0 \\ \lambda e^{-\lambda x} & \textit{pour } x \geq 0 \end{cases}$$

On note : $X \rightsquigarrow \mathcal{E}(\lambda)$.

L'ensemble $\mathcal{X}$ des valeurs possibles d'une variable aléatoire suivant une loi exponentielle $\mathcal{E}(\lambda)$ est $\mathcal{X} = [0, +\infty)$. La Graphique 7.13 montre l'allure de la fonction de densité d'une variable aléatoire distribuée selon une loi exponentielle. On le voit, elle décroît très rapidement (mais ne s'annule jamais).

Graphique 7.13 : Quelques représentations de la densité de la loi $\mathcal{E}(\lambda)$

On vérifie aisément qu'on a bien, comme il se doit, $f(x) \geq 0$, $\forall\, x \in \mathbb{R}$, et :

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{+\infty} \lambda e^{-\lambda x} dx = \left[\lambda \frac{e^{-\lambda x}}{-\lambda}\right]_0^{+\infty} = 0 - (-1) = 1$$

La loi de POISSON et la loi exponentielle offrent deux visions complémentaires d'un processus de POISSON, l'une en termes de comptage, l'autre en termes de durée. On peut aisément passer de l'une à l'autre : pour un processus de POISSON donné, si $Y \rightsquigarrow \mathcal{P}(\lambda)$, où Y représente le nombre d'événements qui surviennent au cours d'une période de temps T et λ est l'espérance du nombre d'événements qui surviennent par unité de temps T, alors $X \rightsquigarrow \mathcal{E}(\lambda)$, où X représente la durée, mesurée en unités de temps T, entre deux apparitions successives d'un événement. Réciproquement, si $X \rightsquigarrow \mathcal{E}(\lambda)$, alors $Y \rightsquigarrow \mathcal{P}(\lambda)$.

Des calculs d'intégrales semblables à ceux détaillés aux Sections 4.2.3 et 7.1 permettent d'établir le résultat suivant.

Propriété 7.13 *Si $X \rightsquigarrow \mathcal{E}(\lambda)$, alors :*

$$E(X) = \frac{1}{\lambda} \quad et \quad V(X) = \frac{1}{\lambda^2}$$

Comme on pouvait s'y attendre étant donné le lien entre loi exponentielle et loi de POISSON, l'espérance de X, c'est-à-dire l'espérance de la durée, mesurée en unités de temps T, entre deux apparitions successives d'un événement, est tout simplement égale à l'inverse du l'espérance du nombre d'événements qui surviennent par unité de temps T. Notons encore l'écart-type de X est égal à son espérance.

On peut facilement obtenir la fonction de répartition de la loi exponentielle. On a :

$$F(x) = I\!P(X \leq x) = \begin{cases} 0 & \text{pour } x < 0 \\ \int_0^x \lambda e^{-\lambda t} dt & \text{pour } x \geq 0 \end{cases}$$

où :

$$\int_0^x \lambda e^{-\lambda t} dt = \left[\lambda \frac{e^{-\lambda t}}{-\lambda}\right]_0^x = -e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}$$

Propriété 7.14 *Si $X \rightsquigarrow \mathcal{E}(\lambda)$, alors :*

$$F(x) = I\!P(X \leq x) \begin{cases} 0 & pour\, x < 0 \\ 1 - e^{-\lambda x} & pour\ x \geq 0 \end{cases}$$

On peut donc aisément calculer des probabilités relatives à cette loi.

7.4.1. Exercice résolu

On considère que les arrivées des clients a un guichet suivent un processus de POISSON tel que l'espérance du nombre de clients qui arrivent par période d'une demi-heure est égale à 6. Un client se présente. Quel est la probabilité qu'il faille attendre au moins 12 minutes avant que le client suivant se présente ?

Notons X la variable aléatoire représentant la durée, en unité de 30 minutes, entre l'arrivée de deux clients successifs. On a :

$$X \rightsquigarrow \mathcal{E}(6)$$

La probabilité demandée est :

$$I\!P(X \geq \frac{12}{30}) = 1 - I\!P(X \leq 0,4) = 1 - \left(1 - e^{-(6\times 0,4)}\right) = e^{-2,4} = 0,0907...$$

On aurait pu procéder en changeant l'échelle de temps. Si l'espérance du nombre de clients qui arrivent par période d'une demi-heure est de 6, l'espérance du nombre de clients qui arrivent par minute est donnée par $6/30 = 0,2$ (cf. Section 6.4.1). On a donc :

$$X^* \rightsquigarrow \mathcal{E}(0,2)$$

où X^* représente la durée, en unités de 1 minute, entre l'arrivée de deux clients successifs. Dans cette nouvelle échelle temporelle, la probabilité demandée est :

$$I\!P(X^* \geq 12) = 1 - I\!P(X^* \leq 12) = 1 - \left(1 - e^{-(0,2\times 12)}\right) = e^{-2,4} = 0,0907...$$

Notons qu'on peut aussi obtenir la probabilité demandée en utilisant uniquement la loi de Poisson. En effet, l'événement "il faut attendre au moins 12 minutes avant que le client suivant ne se présente" est équivalent à l'événement "aucun client ne se présente durant une période de 12 minutes". L'espérance du nombre de clients qui arrivent par période d'une demi-heure étant de 6, l'espérance du nombre de clients qui arrivent par période de 12 minutes est donnée par $(12/30) \times 6 = 2,4$ et on a :

$$X^{**} \rightsquigarrow \mathcal{P}(2,4)$$

où X^{**} représente le nombre d'événements qui se produisent par période de 12 minutes (à nouveau cf. Section 6.4.1). Reformulée de cette façon, la probabilité demandée donnée par :

$$I\!P(X^{**} = 0) = e^{-2,4}\frac{2,4^0}{0!} = e^{-2,4} = 0,0907...$$

7.5. La loi normale bivariée

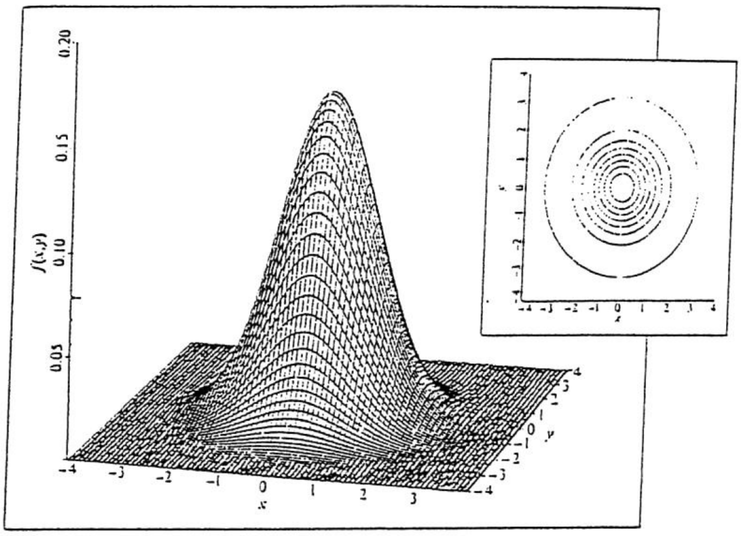
La loi normale bivariée est une généralisation de la loi normale univariée. Comme celle-ci, elle permet de décrire de nombreuses épreuves aléatoires, par exemple le couple de valeurs taille – poids d'un individu pris au hasard, et joue un rôle central les problèmes (bivariés) d'échantillonnage.

Définition 7.12 *Un couple de variable aléatoire (X, Y) suit une loi normale bivariée de paramètres m_x, m_y, σ_x^2, σ_y^2 et ρ, où $m_x \in \mathbb{R}$, $m_y \in \mathbb{R}$, $\sigma_x^2 > 0$, $\sigma_y^2 > 0$ et $-1 \leq \rho \leq 1$, si sa densité de probabilité jointe est :*

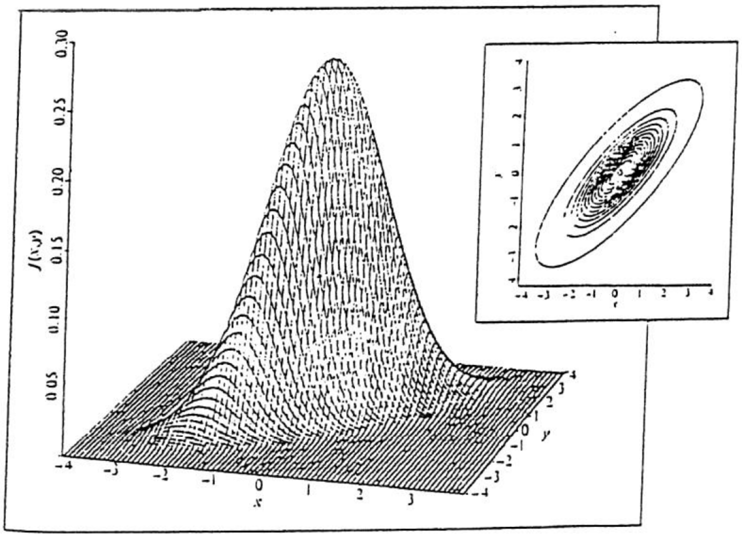
$$f(x,y) = \frac{1}{2\pi\sqrt{\sigma_x^2\sigma_y^2(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]}, \quad \forall\,(x,y) \in \mathbb{R}^2$$

On note : $(X, Y) \rightsquigarrow \mathcal{N}(m_x; m_y; \sigma_x^2; \sigma_y^2; \rho)$.

L'ensemble des couples de valeurs possibles d'un couple (X, Y) de variables aléatoires distribué selon une loi normale bivariée est $\mathcal{X} \times \mathcal{Y}$, où $\mathcal{X} = (-\infty, +\infty)$ et $\mathcal{Y} = (-\infty, +\infty)$. Les Graphiques 7.14.a et 7.14.b illustrent l'allure de la densité jointe $f(x, y)$ et de ses contours pour $m_x = m_y = 0$, $\sigma_x^2 = \sigma_y^2 = 1$ et, respectivement, $\rho = 0$ et $\rho = 0,8$.



Graphique 7.14.a : Fonction de densité de la loi $\mathcal{N}(0; 0; 1; 1; 0)$



Graphique 7.14.b : Fonction de densité de la loi $\mathcal{N}(0; 0; 1; 1; 0,8)$

L'interprétation des paramètres d'une loi normale bivariée est particulièrement simple.

Propriété 7.15 *Si* $(X, Y) \rightsquigarrow \mathcal{N}(m_x; m_y; \sigma_x^2; \sigma_y^2; \rho)$, *alors :*

$$E(X) = m_x, \;\; E(Y) = m_y, \;\; V(X) = \sigma_x^2, \;\; V(Y) = \sigma_y^2 \;\; et \;\; \rho(X, Y) = \rho$$

ce qui implique : $Cov(X, Y) = \rho \sigma_x \sigma_y$.

Bref, les paramètres sont directement liées aux espérances, aux variances et à la corrélation / covariance des variables aléatoires X et Y.

La loi normale bivariée possède encore d'autres propriétés remarquables.

Propriété 7.16 *Si* $(X, Y) \rightsquigarrow \mathcal{N}(m_x; m_y; \sigma_x^2; \sigma_y^2; \rho)$, *alors :*

$$X \rightsquigarrow \mathcal{N}(m_x; \sigma_x^2), \qquad Y \rightsquigarrow \mathcal{N}(m_y; \sigma_y^2)$$

et

$$a + bX + cY \rightsquigarrow \mathcal{N}(m; \sigma^2)$$

où :

$$m = a + bm_x + cm_y \quad et \quad \sigma^2 = b^2\sigma_x^2 + c^2\sigma_y^2 + 2bc\rho\sigma_x\sigma_y$$

Ainsi, si un couple de variables aléatoires suit une loi normale bivariée, chacune des variables de ce couple suit individuellement (loi marginale) une loi normale. Notons que la réciproque est fausse : on peut très bien avoir deux variables aléatoires qui suivent chacune une loi normale mais dont la loi jointe n'est pas normale bivariée. On voit également qu'une combinaison linéaire de deux variables aléatoires dont la loi jointe est normale bivariée suit aussi une loi normale. Les paramètres m et σ^2 de cette loi découlent des Propriétés 5.5 et 5.7. On a déjà vu que c'était le cas si X et Y sont indépendantes (cf. Propriété 7.6). Cela l'est donc aussi si X et Y ne sont pas indépendantes, à condition que leur loi jointe soit normale bivariée.

Propriété 7.17 *Si* $(X, Y) \rightsquigarrow \mathcal{N}(m_x; m_y; \sigma_x^2; \sigma_y^2; \rho)$, *alors :*

$$Y|X = x \rightsquigarrow \mathcal{N}(\alpha + \beta x; \sigma^2),$$

où :

$$\alpha = m_y - \beta m_x, \;\; \beta = \rho\frac{\sigma_y}{\sigma_x} \;\; et \;\; \sigma^2 = \sigma_y^2 - \beta^2\sigma_x^2$$

En d'autres termes, si un couple de variables aléatoires suit une loi normale bivariée, la loi conditionnelle de Y sachant $X = x$ (notée ci-dessus $Y|X = x$) est encore une loi normale, dont l'espérance (conditionnelle) est une fonction *linéaire* de x : $E(Y|X = x) = \alpha + \beta x$, et dont la variance (conditionnelle) est constante (ne dépend pas de x) : $V(Y|X = x) = \sigma^2$, les paramètres de l'une et l'autre étant fonction des paramètres de la loi jointe. Il va de soi que le même résultat vaut pour la loi de X sachant $Y = y$: il suffit de permuter les rôles de X et de Y.

Propriété 7.18 *Soit* $(X, Y) \rightsquigarrow \mathcal{N}(m_x; m_y; \sigma_x^2; \sigma_y^2; \rho)$. *Si X et Y sont non-corrélées,* *c'est-à-dire si $Cov(X,Y) = \rho(X,Y) = 0$, alors elles sont indépendantes.*

On a vu que si deux variables aléatoires quelconques sont indépendantes, alors elles sont non-corrélées, la réciproque étant généralement fausse (cf. Propriété 5.14). On voit que dans le cas normal bivarié, la réciproque est vraie : dans ce cas précis, indépendance et non-corrélation sont équivalents.

Notons pour conclure qu'il existe une généralisation de la loi normale univariée pour des vecteurs de n variables aléatoires. On parle alors de loi normale multivariée. La loi normale bivariée en est évidemment un cas particulier. La loi normale multivariée possède les mêmes types de propriétés que la loi normale bivariée.

7.6. Exercice résolu

On considère que la durée de vie d'une ampoule prise au hasard suit une loi normale $\mathcal{N}(m; 1600)$. Une ampoule est jugée défectueuse lorsque sa durée de vie est inférieure à 90 heures.

1- On suppose que $m = 160$.

a- Calculez la probabilité qu'une ampoule prise au hasard ait une durée de vie supérieure à 190 heures.

En appelant X la durée de vie, on a :

$$X \rightsquigarrow \mathcal{N}(160; 1600) \Rightarrow Z = \frac{X - 160}{\sqrt{1600}} \rightsquigarrow \mathcal{N}(0; 1)$$

Donc :

$$\begin{aligned}
I\!P(X > 190) &= I\!P(X \geq 190) = I\!P\left(\frac{X-160}{40} \geq \frac{190-160}{40}\right) \\
&= I\!P(Z \geq 0,75) = 1 - I\!P(Z \leq 0,75) = 1 - \Phi(0,75) \\
&= 1 - 0,7734 = 0,2266
\end{aligned}$$

b- Calculez la probabilité qu'une ampoule prise au hasard ait une durée de vie comprise entre 140 et 180 heures.

On doit calculer $I\!P(140 < X < 180) = I\!P(140 \leq X \leq 180)$. On a :

$$\begin{aligned}
I\!P(140 \leq X \leq 180) &= I\!P\left(\frac{140-160}{40} \leq \frac{X-160}{40} \leq \frac{180-160}{40}\right) \\
&= I\!P(-0,5 \leq Z \leq 0,5) = \Phi(0,5) - \Phi(-0,5) \\
&= 2\Phi(0,5) - 1 = 2 \times 0,6915 - 1 = 0,3830
\end{aligned}$$

c- Calculez la probabilité qu'une ampoule prise au hasard soit défectueuse.

On doit calculer $I\!P(X < 90) = I\!P(X \leq 90)$. On a :

$$\begin{aligned}
I\!P(X \leq 90) &= I\!P\left(\frac{X-160}{40} \leq \frac{90-160}{40}\right) \\
&= I\!P(Z \leq -1,75) = \Phi(-1,75) = 1-\Phi(1,75) \\
&= 1-0,9599 = 0,0401
\end{aligned}$$

2- Déterminez la plus petite valeur de m telle que la probabilité qu'une ampoule prise au hasard soit défectueuse soit au plus égale à $0,1$.

On cherche la plus petite valeur de m telle que : $I\!P(X \leq 90) \leq 0,1$. On a :

$$I\!P(X \leq 90) = I\!P\left(\frac{X-m}{40} \leq \frac{90-m}{40}\right) = I\!P\left(Z \leq \frac{90-m}{40}\right) = \Phi\left(\frac{90-m}{40}\right)$$

Comme $\Phi(.)$ est croissante, la plus petite valeur de m telle que $\Phi\left(\frac{90-m}{40}\right) \leq 0,1$ est donnée par la valeur de m satisfaisant l'égalité :

$$\Phi\left(\frac{90-m}{40}\right) = 0,1$$

On a $\Phi(-1,2816) = 0,1$ ($z_{0,1} = -1,2816$: quantile d'ordre $0,1$) et donc finalement :

$$\frac{90-m}{40} = -1,2816 \;\Leftrightarrow\; m = 90 + 40 \times 1,2816 = 141,264$$

3- On suppose à nouveau que $m = 160$. On considère un lot de 6 ampoules choisies au hasard.

a- Calculez la probabilité qu'une ampoule au moins soit défectueuse.

On peut associer à chacune des 6 ampoules une variable aléatoire de Bernoulli Y_i qui prend la valeur 1 si l'ampoule est défectueuse et la valeur 0 sinon. On a ainsi $Y_1, ..., Y_6$, six variables aléatoires indépendantes suivant une même loi de Bernoulli $\mathcal{B}(p)$, où d'après la question 1, $p = I\!P(Y_i = 1) = I\!P(X < 90) = 0,0401$. Le nombre $Y = Y_1 + ... + Y_6$ d'ampoules défectueuses parmi les 6 ampoules suit une loi binomiale $\mathcal{B}(6; 0,0401)$. La probabilité demandée est donc (à l'arrondi près) :

$$\begin{aligned}
I\!P(Y \geq 1) &= 1 - I\!P(Y=0) = 1 - C_6^0(0,0401)^0(0,9599)^6 \\
&= 1 - 0,7822 = 0,2178
\end{aligned}$$

b- Calculez la probabilité que le nombre d'ampoules défectueuses soit compris, bornes incluses, entre 2 et 4.

La probabilité demandée est (aux arrondis près) :

$$\begin{aligned}
I\!P(2 \leq Y \leq 4) &= I\!P(Y=2) + I\!P(Y=3) + I\!P(Y=4) \\
&= C_6^2\,(0,0401)^2\,(0,9599)^4 + C_6^3\,(0,0401)^3\,(0,9599)^3 \\
&\quad + C_6^4\,(0,0401)^4\,(0,9599)^2 \\
&= 0,0205 + 0,0011 + 0,0000 = 0,0216
\end{aligned}$$

4- On suppose toujours que $m = 160$ mais on considère cette fois un lot de 100 ampoules choisies au hasard. Calculez la probabilité d'avoir au plus 3 ampoules défectueuses.

On a maintenant : $Y \rightsquigarrow \mathcal{B}(100; 0,0401)$. Comme $n = 100 \geq 50$, $p = 0,0401 \leq 0,1$ et $np = 4,01 \leq 15$, on approche cette loi par une loi de POISSON :

$$\mathcal{B}(100; 0,0401) \approx \mathcal{P}(4,01)$$

La probabilité demandée est :

$$I\!P(Y \leq 3) = I\!P(Y = 0) + P(Y = 1) + I\!P(Y = 2) + I\!P(Y = 3)$$

que l'on approche donc par (aux arrondis près) :

$$\begin{aligned}
I\!P(Y \leq 3) &\approx e^{-4,01}\frac{4,01^0}{0!} + e^{-4,01}\frac{4,01^1}{1!} + e^{-4,01}\frac{4,01^2}{2!} + e^{-4,01}\frac{4,01^3}{3!} \\
&\approx e^{-4,01} \times (1 + 4,01 + \frac{4,01^2}{2} + \frac{4,01^3}{6}) \\
&\approx 0,0181 \times (1 + 4,01 + 8,0401 + 10,7469) = 0,4307
\end{aligned}$$

5- On suppose toujours que $m = 160$ mais on considère à présent un lot de 1 000 ampoules choisies au hasard. Calculez la probabilité d'avoir au plus 30 ampoules défectueuses.

On a maintenant : $Y \rightsquigarrow \mathcal{B}(1000; 0,0401)$. Comme $np = 40,1$, on ne peut plus utiliser l'approximation par la loi de POISSON. Par contre, on a $n = 1000 \geq 30$, $np = 40,1 \geq 5$, $nq = 959,9 \geq 5$. On peut donc approcher cette loi par une loi normale :

$$\mathcal{B}(1000; 0,0401) \approx \mathcal{N}(40,1; 38,49199)$$

La probabilité demandée est $I\!P(Y \leq 30)$. On l'approche de la manière suivante (on peut ici négliger la correction de continuité) :

$$I\!P(Y \leq 30) \approx I\!P(W \leq 30)$$

où W désigne une variable aléatoire qui suit la loi $\mathcal{N}(40,1; 38,49199)$. On a (à l'arrondis près) :

$$\begin{aligned}
I\!P(W \leq 30) &= I\!P\left(\frac{Y - 40,1}{\sqrt{38,49199}} \leq \frac{30 - 40,1}{\sqrt{38,49199}}\right) \\
&= I\!P\left(Z \leq \frac{30 - 40,1}{6,2042}\right) = \Phi(-1,63) \\
&= 1 - \Phi(1,63) = 1 - 0,9484 = 0,0516
\end{aligned}$$

soit finalement :

$$I\!P(Y \leq 30) \approx 0,0516$$

7.7. Exercices

1- On considère une variable aléatoire $X \rightsquigarrow \mathcal{N}(0;1)$. Trouvez les probabilités suivantes :

a- $\mathbb{P}(X \leq 1,47)$.
b- $\mathbb{P}(X < \frac{1}{3})$.
c- $\mathbb{P}(X \geq 0,21)$.
d- $\mathbb{P}(X \leq -2,52)$.
e- $\mathbb{P}(X > -0,74)$.
f- $\mathbb{P}(-1,2 < X \leq 2,47)$.
g- $\mathbb{P}(-2,52 \leq X < 2,52)$.

2- On considère une variable aléatoire $X \rightsquigarrow \mathcal{N}(75;25)$. Calculez les probabilités suivantes :

a- $\mathbb{P}(60 \leq X \leq 80)$.
b- $\mathbb{P}(X \geq 82)$.
c- $\mathbb{P}(X > 95)$.
d- $\mathbb{P}(X < 0)$.
e- $\mathbb{P}(|X-75| \leq 5)$.
f- $\mathbb{P}(|X-75| < 10)$.
g- $\mathbb{P}(|X-70| \leq 10)$.

3- Soit $Z \rightsquigarrow \mathcal{N}(0,1)$. Résolvez les équations suivantes :

a- $\mathbb{P}(Z \leq x) = 0,9 \Rightarrow x = ?$
b- $\mathbb{P}(Z \leq x) = 0,88 \Rightarrow x = ?$
c- $\mathbb{P}(Z \leq x) = 0,5 \Rightarrow x = ?$
d- $\mathbb{P}(Z \leq x) = 0,2 \Rightarrow x = ?$
e- $\mathbb{P}(Z \leq x) = 0,657 \Rightarrow x = ?$
f- $\mathbb{P}(-x \leq Z \leq x) = 0,98 \Rightarrow x = ?$
g- $\mathbb{P}(-x \leq Z \leq x) = 0,2 \Rightarrow x = ?$

4- Soit $Z \rightsquigarrow \mathcal{N}(0;1)$. Déterminez les quartiles de cette loi, c'est-à-dire les valeurs $z_{0,25}$, $z_{0,50}$ et $z_{0,75}$ telles que, respectivement, $\mathbb{P}(Z \leq z_{0,25}) = 0,25$, $\mathbb{P}(Z \leq z_{0,50}) = 0,50$ et $\mathbb{P}(Z \leq z_{0,75}) = 0,75$.

5- Soit $X \rightsquigarrow \mathcal{N}(6;1,69)$. Déterminez les quartiles de cette loi, c'est-à-dire les valeurs $x_{0,25}$, $x_{0,50}$ et $x_{0,75}$ telles que, respectivement, $\mathbb{P}(X \leq x_{0,25}) = 0,25$, $\mathbb{P}(X \leq x_{0,50}) = 0,50$ et $\mathbb{P}(X \leq x_{0,75}) = 0,75$.

6- La taille des soldats de l'armée syldave est en moyenne de 177 cm avec une variance de 29 cm^2. On suppose que la distribution de ces tailles est normale.

a- Quelle est la taille médiane des recrues ?
b- Quelle est la probabilité que la taille d'une recrue prise au hasard soit comprise :

i- entre 170 cm et 190 cm ?
ii- entre 173, 5 cm et 175, 6 cm ?

c- Déterminez un intervalle, centré sur la taille moyenne des recrues, qui soit tel qu'il y ait une probabilité de 0, 6 qu'une recrue prise au hasard ait une taille appartenant à cet intervalle.

d- Quelle est la valeur de la taille qui a une probabilité de $0,7$ de ne pas être dépassée par une recrue prise au hasard ?

7- On considère que le volume de remplissage effectif d'une bouteille par une machine d'embouteillage suit une loi normale d'espérance égale à 252 cm^{3}et d'écart-type égal à 2 cm^3.

a- Quelle est la probabilité que le volume de remplissage d'une bouteille soit inférieur à 250 cm^3 ?

b- Quelle valeur faudrait-il donner à l'espérance pour que cette probabilité soit réduite à 5% ?

8- Trouvez la probabilité d'obtenir 25 fois une somme des résultats égale à sept en 100 jets d'une paire de dés. Que devient cette probabilité si on demande à présent au moins 25 fois une somme des résultats égale à sept ?

9- Un ascenseur peut supporter une charge limite de 5 000 kg. Il a une capacité de 50 personnes. Si la distribution du poids des gens utilisant cet ascenseur est normale avec une moyenne de 95 kg et un écart-type de $12,5$ kg, quelle est la probabilité qu'un groupe de 50 personnes tirées au hasard excède cette charge limite ?

10- L'expérience d'une entreprise amène à considérer que la durée de vie (exprimée en années) des téléviseurs qu'elle fabrique est distribuée selon une loi normale $\mathcal{N}(3; 2,25)$. L'entreprise accorde une garantie totale sur ses téléviseurs durant un an. Pour 100 téléviseurs vendus, à combien de remplacements sous garantie doit-on s'attendre ?

11- On admet que la loi de la durée de vie d'un cocker est approximativement normale d'espérance égale à 7 ans et d'écart-type égal à 2 ans.

a- Donner le mode et la médiane de cette distribution.

b- Supposons que j'achète un cocker qui vient de naître et est en parfaite santé.

i- Quelle est la probabilité qu'il dépasse l'âge de 9 ans ?

ii- Entre quelles limites de durée de vie a-t-il 95 chances sur 100 de se situer ? (on donnera un intervalle centré sur l'espérance).

12- On considère une variable aléatoire X distribuée selon une loi normale de variance égale à 2. On sait que la probabilité d'avoir X supérieur à 28 vaut $0,03$. Trouvez l'espérance de la loi.

13- Soit X une variable aléatoire normale. Sachant que $\mathbb{P}(X \leq 0,22) = 0,04$ et que $\mathbb{P}(X \geq 5,2) = 0,23$, déterminez les valeurs de m et de σ^2.

14- Un discobole, Avau Marc ne réalise plus de progrès. Ses lancers se répartissent à peu près suivant une distribution normale. Dans 95% des cas, la distance qu'il atteint est comprise entre 49 m et 61 m (intervalle centré sur l'espérance). Le record olympique est $59,40$ m. Monsieur Avau participera aux prochains Jeux. Il entrera le premier en poste et aura droit à 10 essais.

a- Quelle est la probabilité qu'il batte le record olympique dès son premier lancer ?

b- Quelle est la probabilité qu'il batte ce record au cours de ses 10 essais ?
c- Sur quelles hypothèses reposent vos réponses aux deux premières questions ?

15- Un processus de fabrication produit des cylindres métalliques. Pour qu'un cylindre soit conforme, il faut que sa longueur (L) soit comprise entre $8,4$ cm et $8,615$ cm, et que son diamètre (D) soit compris entre $1,5404$ cm et $1,5886$ cm. Le processus de fabrication est tel que la longueur des cylindres est distribuée selon une loi normale $\mathcal{N}(8,54; 0,0025)$ et leur diamètre selon une loi normale (indépendante de la précédente) $\mathcal{N}(1,57; 0,0001)$.
   a- Quelle est la probabilité qu'un cylindre prélevé au hasard dans la production soit conforme ?
   b- Quelle est la probabilité qu'un lot de 7 cylindres prélevés au hasard dans la production contienne au plus 1 cylindre non conforme ?
   c- Quelle est la probabilité qu'un lot de 700 cylindres prélevés au hasard dans la production contienne au plus 90 cylindres non conformes ?

16- Un examen comporte trois épreuves E_1, E_2, E_3. Chaque épreuve est notée sur 20. Les coefficients des épreuves sont respectivement 2, 2 et 1, de sorte que la note totale est sur 100. On réussit à l'examen si cette note totale est au moins égale à 50. On suppose que les résultats obtenus aux trois épreuves sont indépendants. On appelle X_1, X_2, X_3 les notes obtenues. On désigne par m_1,m_2, m_3, σ_1^2, σ_2^2, σ_3^2 les espérances et les variances respectives de X_1, X_2, X_3. On appelle X la note totale.
   a- Exprimez X en termes de X_1, X_2, X_3. Exprimez l'espérance m et l'écart-type σ de X en fonction de m_1, m_2, m_3, σ_1^2, σ_2^2, σ_3^2.
   b- Dans toute la suite de l'exercice, on suppose que X_1, X_2, X_3 suivent respectivement les lois normales $\mathcal{N}(11,2; 6,25)$, $\mathcal{N}(10,4; 3,24)$, $\mathcal{N}(9,5; 7,29)$.
      i- Calculez la probabilité d'obtenir à l'épreuve E_1 une note comprise entre 8 et 12.
      ii- Quelle est la loi de X ? Calculez le taux de réussite à l'examen.
   c- Un candidat a fait l'impasse sur la troisième matière, de sorte qu'il obtient 1 à l'épreuve E_3. Quelle est la probabilité qu'il réussisse à l'examen ?

17- On suppose que la durée de vie (en heures) d'une pile d'un certain type suit une loi normale $\mathcal{N}(40; 4,41)$.
   a- Quelle est la probabilité qu'une pile de ce type dure plus de 41 heures et 30 minutes ?
   b- Quelle est la durée minimale d qui a au plus 2% de chances d'être dépassée par une pile de ce type ?

18- Dans une entreprise, un ascenseur peut transporter une seule personne. On suppose que le poids d'une personne est en moyenne de 75 kg avec une variance de $104,04$ kg^2, et qu'il est distribué selon une loi normale. L'ascenseur est supposé en surcharge lorsque le poids excède 92 kg.
   a- Quelle est la probabilité qu'il y ait surcharge ?
   b- A partir de quelle capacité de charge p un ascenseur qui peut transporter une seule personne sera-t-il en surcharge plus d'une fois sur dix ?

19- On considère que le poids d'un paquet de sucre (exprimé en grammes) en morceaux pris au hasard suit une loi normale $\mathcal{N}(1\,000; 144)$. Un paquet est jugé non conforme si son poids est inférieur à 990 g ou supérieur à 1 020 g.

  a- Quelle est la probabilité qu'un paquet pris au hasard soit jugé non conforme ?

  b- On examine 5 paquets de sucre pris au hasard et on appelle X le poids total de ces 5 paquets.

    i- Quelle est la loi de X ? Pourquoi ?

    ii- Quelle est la valeur maximale p qui sera dépassée par le poids total X dans au moins 5% des cas ?

20- Le poids, exprimé en grammes, des pièces produites par une machine est distribué selon une loi normale de moyenne égale à 120 et d'écart-type égal à 7.

  a- On considère une pièce choisie au hasard dans la production. Quelle est la probabilité que son poids ne dépasse pas 118 g ?

  b- On considère 3 pièces prélevées au hasard dans la production.

    i- Quelle est la loi de leur poids total ?

    ii- Dans quel intervalle centré en l'espérance ce poids total se situera-t-il avec une probabilité de 96% ?

  c- On considère deux pièces prélevées au hasard dans la production.

    i- Quelle est la probabilité que leur poids total soit supérieur à 243 g ?

    ii- Quelle est la probabilité que leurs poids diffèrent de plus de 4 g ?

21- Vérifiez, en utilisant la Propriété 4.6 établie au Chapitre 4, que si $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$ et $Y = a+bX$, on a bien, comme l'indique la Propriété 7.4, que $Y \rightsquigarrow \mathcal{N}(m^*; \sigma^{*2})$, où $m^* = a + bm$ et $\sigma^{*2} = b^2\sigma^2$.

7.8. Tables statistiques

Loi normale (source : Droesbeke J-J. (1997))

Valeurs de la fonction de répartition de la distribution normale réduite $Z = \mathcal{N}(0,1)$

$$P(Z \leqslant z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Remarques : $F(-z) = 1 - F(z)$, $P(a \leqslant Z \leqslant b) = F(b) - F(a)$ $(a \leqslant b)$.

Pour $z > 4$, la précision de la table ne permet plus de connaître la valeur exacte de $F(z)$ **Exemples :**.

$F(4) = 0.9999683$; $F(4,5) = 0.9999966$; $F(5) = 0.999999713$, qui est donc arrondi à 1.

Loi normale (suite et fin)

Quantiles z_p de la variable $Z = \mathcal{N}(0,1)$

$P(Z \leqslant z_p) = p \quad (0 < p < 1)$

p	z_p	p	z_p	p	z_p	p	z_p
0.0001	−3.7190	0.15	−1.0364	0.51	0.0251	0.87	1.1264
0.0002	−3.5401	0.16	−0.9945	0.52	0.0502	0.88	1.1750
0.0003	−3.4316	0.17	−0.9542	0.53	0.0753	0.89	1.2265
0.0004	−3.3528	0.18	−0.9154	0.54	0.1004	0.90	1.2816
0.0005	−3.2905	0.19	−0.8779	0.55	0.1257	0.91	1.3408
0.0006	−3.2389	0.20	−0.8416	0.56	0.1510	0.92	1.4051
0.0007	−3.1947	0.21	−0.8064	0.57	0.1764	0.925	1.4395
0.0008	−3.1559	0.22	−0.7722	0.58	0.2019	0.93	1.4758
0.0009	−3.1214	0.23	−0.7389	0.59	0.2275	0.94	1.5548
0.001	−3.0902	0.24	−0.7063	0.60	0.2534	0.95	1.6449
0.002	−2.8782	0.25	−0.6745	0.61	0.2793	0.96	1.7507
0.003	−2.7478	0.26	−0.6434	0.62	0.3055	0.97	1.8808
0.004	−2.6521	0.27	−0.6128	0.63	0.3319	0.975	1.9600
0.005	−2.5758	0.28	−0.5828	0.64	0.3585	0.98	2.0537
0.006	−2.5121	0.29	−0.5534	0.65	0.3853	0.985	2.1701
0.007	−2.4573	0.30	−0.5244	0.66	0.4125	0.99	2.3263
0.008	−2.4089	0.31	−0.4959	0.67	0.4399	0.991	2.3656
0.009	−2.3656	0.32	−0.4677	0.68	0.4677	0.992	2.4089
0.01	−2.3263	0.33	−0.4399	0.69	0.4959	0.993	2.4573
0.015	−2.1701	0.34	−0.4125	0.70	0.5244	0.994	2.5121
0.02	−2.0537	0.35	−0.3853	0.71	0.5534	0.995	2.5758
0.025	−1.9600	0.36	−0.3585	0.72	0.5828	0.996	2.6521
0.03	−1.8808	0.37	−0.3319	0.73	0.6128	0.997	2.7478
0.04	−1.7507	0.38	−0.3055	0.74	0.6434	0.998	2.8782
0.05	−1.6449	0.39	−0.2793	0.75	0.6745	0.999	3.0902
0.06	−1.5548	0.40	−0.2534	0.76	0.7063	0.9991	3.1214
0.07	−1.4758	0.41	−0.2275	0.77	0.7389	0.9992	3.1559
0.075	−1.4395	0.42	−0.2019	0.78	0.7722	0.9993	3.1947
0.08	−1.4051	0.43	−0.1764	0.79	0.8064	0.9994	3.2389
0.09	−1.3408	0.44	−0.1510	0.80	0.8416	0.9995	3.2905
0.10	−1.2816	0.45	−0.1257	0.81	0.8779	0.9996	3.3528
0.11	−1.2265	0.46	−0.1004	0.82	0.9154	0.9997	3.4316
0.12	−1.1750	0.47	−0.0753	0.83	0.9542	0.9998	3.5401
0.13	−1.1264	0.48	−0.0502	0.84	0.9945	0.9999	3.7190
0.14	−1.0803	0.49	−0.0251	0.85	1.0364		
0.15	−1.0364	0.50	0.0000	0.86	1.0803		

Loi du χ^2 (source : Droesbeke J-J. (1997))

Quantiles $\chi^2_{\nu;p}$ *de la variable* $\chi^2(\nu)$

$$P\left(X \leqslant \chi^2_{\nu,p}\right) = p$$

ν	p											
	0.005	0.01	0.025	0.05	0.075	0.10	0.90	0.925	0.95	0.975	0.99	0.995
1	0.00	0.00	0.00	0.00	0.01	0.02	2.71	3.17	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.16	0.21	4.61	5.18	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.47	0.58	6.25	6.90	7.81	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	0.90	1.06	7.78	8.50	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.39	1.61	9.24	10.01	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	1.94	2.20	10.65	11.47	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.53	2.83	12.02	12.88	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.14	3.49	13.36	14.27	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	3.78	4.17	14.68	15.63	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.45	4.87	15.99	16.97	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.12	5.58	17.28	18.29	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	5.82	6.30	18.55	19.60	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	6.52	7.04	19.81	20.90	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.24	7.79	21.06	22.18	23.69	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	7.97	8.55	22.31	23.45	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	8.71	9.31	23.54	24.72	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	9.45	10.09	24.77	25.97	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.21	10.87	25.99	27.22	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	10.97	11.65	27.20	28.46	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	11.73	12.44	28.41	29.69	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	12.50	13.24	29.62	30.92	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	13.28	14.04	30.81	32.14	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.07	14.85	32.01	33.36	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	14.85	15.66	33.20	34.57	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	15.65	16.47	34.38	35.78	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	16.44	17.29	35.56	36.98	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	17.24	18.11	36.74	38.18	40.11	43.20	46.96	49.65
28	12.46	13.57	15.31	16.93	18.05	18.94	37.92	39.38	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	18.85	19.77	39.09	40.57	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	19.66	20.60	40.26	41.76	43.77	46.98	50.89	53.67
31	14.46	15.66	17.54	19.28	20.48	21.43	41.42	42.95	44.99	48.23	52.19	55.00
32	15.13	16.36	18.29	20.07	21.30	22.27	42.59	44.13	46.19	49.48	53.49	56.33
33	15.82	17.07	19.05	20.87	22.12	23.11	43.75	45.31	47.40	50.73	54.78	57.65
34	16.50	17.79	19.81	21.66	22.94	23.95	44.90	46.49	48.60	51.97	56.06	58.96
35	17.19	18.51	20.57	22.47	23.76	24.80	46.06	47.66	49.80	53.20	57.34	60.28
36	17.89	19.23	21.34	23.27	24.59	25.64	47.21	48.84	51.00	54.44	58.62	61.58
37	18.59	19.96	22.11	24.08	25.42	26.49	48.36	50.01	52.19	55.67	59.89	62.88
38	19.29	20.69	22.88	24.88	26.25	27.34	49.51	51.17	53.38	56.90	61.16	64.18
39	20.00	21.43	23.65	25.70	27.09	28.20	50.66	52.34	54.57	58.12	62.43	65.48
40	20.71	22.16	24.43	26.51	27.93	29.05	51.81	53.50	55.76	59.34	63.69	66.77
41	21.42	22.91	25.22	27.33	28.77	29.91	52.95	54.66	56.94	60.56	64.95	68.05
42	22.14	23.65	26.00	28.14	29.61	30.77	54.09	55.82	58.12	61.78	66.21	69.34
43	22.86	24.40	26.79	28.97	30.45	31.63	55.23	56.98	59.30	62.99	67.46	70.62
44	23.58	25.15	27.58	29.79	31.29	32.49	56.37	58.13	60.48	64.20	68.71	71.89
45	24.31	25.90	28.37	30.61	32.14	33.35	57.51	59.29	61.66	65.41	69.96	73.17
46	25.04	26.66	29.16	31.44	32.99	34.22	58.64	60.44	62.83	66.62	71.20	74.44
47	25.78	27.42	29.96	32.27	33.84	35.08	59.77	61.59	64.00	67.82	72.44	75.70
48	26.51	28.18	30.76	33.10	34.69	35.95	60.91	62.74	65.17	69.02	73.68	76.97
49	27.25	28.94	31.56	33.93	35.54	36.82	62.04	63.89	66.34	70.22	74.92	78.23

Loi du χ^2 (suite et fin)

Quantiles $\chi^2_{\nu;p}$ *de la variable* $\chi^2(\nu)$

ν	p											
	0.005	0.01	0.025	0.05	0.075	0.10	0.90	0.925	0.95	0.975	0.99	0.995
50	27.99	29.71	32.36	34.76	36.40	37.69	63.17	65.03	67.51	71.42	76.15	79.49
51	28.74	30.48	33.16	35.60	37.25	38.56	64.30	66.17	68.67	72.52	77.39	80.75
52	29.48	31.25	33.97	36.44	38.11	39.43	65.42	67.32	69.83	73.81	78.62	82.00
53	30.23	32.02	34.78	37.28	38.97	40.31	66.55	68.46	70.99	75.00	79.84	83.25
54	30.98	32.79	35.59	38.12	39.83	41.18	67.67	69.60	72.15	76.19	81.07	84.50
55	31.74	33.57	36.40	38.96	40.69	42.06	68.80	70.74	73.31	77.38	82.29	85.75
56	32.49	34.35	37.21	39.80	41.55	42.94	69.92	71.87	74.47	78.57	83.51	86.99
57	33.25	35.13	38.03	40.65	42.42	43.82	71.04	73.01	75.62	79.75	84.73	88.24
58	34.01	35.91	38.84	41.49	43.28	44.70	72.16	74.15	76.78	80.94	85.95	89.48
59	34.77	36.70	39.66	42.34	44.15	45.58	73.28	75.28	77.93	82.12	87.17	90.72
60	35.53	37.49	40.48	43.19	45.02	46.46	74.40	76.41	79.08	83.30	88.38	91.95
61	36.30	38.27	41.30	44.04	45.89	47.34	75.51	77.54	80.23	84.48	89.59	93.19
62	37.07	39.06	42.13	44.89	46.76	48.23	76.63	78.67	81.38	85.65	90.80	94.42
63	37.84	39.86	42.95	45.74	47.63	49.11	77.75	79.80	82.53	86.83	92.01	95.65
64	38.61	40.65	43.78	46.60	48.50	50.00	78.86	80.93	83.68	88.00	93.22	96.88
65	39.38	41.44	44.60	47.45	49.37	50.88	79.97	82.06	84.82	89.18	94.42	98.11
66	40.16	42.24	45.43	48.31	50.24	51.77	81.09	83.18	85.97	90.35	95.63	99.33
67	40.94	43.04	46.26	49.16	51.12	52.66	82.20	84.31	87.11	91.52	96.83	100.55
68	41.71	43.84	47.09	50.02	51.99	53.55	83.31	85.43	88.25	92.69	98.03	101.78
69	42.49	44.64	47.92	50.88	52.87	54.44	84.42	86.56	89.39	93.86	99.23	103.00
70	43.28	45.44	48.76	51.74	53.75	55.33	85.53	87.68	90.53	95.02	100.43	104.21
71	44.06	46.25	49.59	52.60	54.63	56.22	86.64	88.80	91.67	96.19	101.62	105.43
72	44.84	47.05	50.43	53.46	55.51	57.11	87.74	89.92	92.81	97.35	102.82	106.65
73	45.63	47.86	51.27	54.33	56.39	58.01	88.85	91.04	93.95	98.52	104.01	107.86
74	46.42	48.67	52.10	55.19	57.27	58.90	89.96	92.16	95.08	99.68	105.20	109.07
75	47.21	49.48	52.94	56.05	58.15	59.80	91.06	93.28	96.22	100.84	106.39	110.29
76	48.00	50.29	53.78	56.92	59.03	60.69	92.17	94.40	97.35	102.00	107.58	111.50
77	48.79	51.10	54.62	57.79	59.91	61.59	93.27	95.52	98.48	103.16	108.77	112.70
78	49.58	51.91	55.47	58.65	60.80	62.48	94.37	96.63	99.62	104.32	109.96	113.91
79	50.38	52.73	56.31	59.52	61.68	63.38	95.48	97.75	100.75	105.47	111.14	115.12
80	51.17	53.54	57.15	60.39	62.57	64.28	96.58	98.86	101.88	106.63	112.33	116.32
81	51.97	54.36	58.00	61.26	63.45	65.18	97.68	99.97	103.01	107.78	113.51	117.52
82	52.77	55.17	58.85	62.13	64.34	66.08	98.78	101.09	104.14	108.94	114.69	118.73
83	53.57	55.99	59.69	63.00	65.23	66.98	99.88	102.20	105.27	110.09	115.88	119.93
84	54.37	56.81	60.54	63.88	66.12	67.88	100.98	103.31	106.39	111.24	117.06	121.13
85	55.17	57.63	61.39	64.75	67.01	68.78	102.08	104.42	107.52	112.39	118.24	122.32
86	55.97	58.46	62.24	65.62	67.90	69.68	103.18	105.53	108.65	113.54	119.41	123.52
87	56.78	59.28	63.09	66.50	68.79	70.58	104.28	106.64	109.77	114.69	120.59	124.72
88	57.58	60.10	63.94	67.37	69.68	71.48	105.37	107.75	110.90	115.84	121.77	125.91
89	58.39	60.93	64.79	68.25	70.57	72.39	106.47	108.86	112.02	116.99	122.94	127.11
90	59.20	61.75	65.65	69.13	71.46	73.29	107.57	109.97	113.15	118.14	124.12	128.30
91	60.01	62.58	66.50	70.00	72.35	74.20	108.66	111.08	114.27	119.28	125.29	129.49
92	60.82	63.41	67.36	70.88	73.25	75.10	109.76	112.18	115.39	120.43	126.46	130.68
93	61.63	64.24	68.21	71.76	74.14	76.01	110.85	113.29	116.51	121.57	127.63	131.87
94	62.44	65.07	69.07	72.64	75.03	76.91	111.94	114.39	117.63	122.72	128.80	133.06
95	63.25	65.90	69.93	73.52	75.93	77.82	113.04	115.50	118.75	123.86	129.97	134.25
96	64.06	66.73	70.78	74.40	76.82	78.73	114.13	116.60	119.87	125.00	131.14	135.43
97	64.88	67.56	71.64	75.28	77.72	79.63	115.22	117.71	120.99	126.14	132.31	136.62
98	65.69	68.40	72.50	76.16	78.62	80.54	116.32	118.81	122.11	127.28	133.48	137.80
99	66.51	69.23	73.36	77.05	79.51	81.45	117.41	119.91	123.23	128.42	134.64	138.99
100	67.33	70.07	74.22	77.93	80.41	82.36	118.50	121.02	124.34	129.56	135.81	140.17

Loi de Student (source : Droesbeke J-J. (1997))

Quantiles $t_{\nu;p}$ de la variable de Student $t(\nu)$

$$P(X \leqslant t_{\nu;p}) = p$$

ν	p												
	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	0.289	0.445	0.617	0.817	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.500	4.785	5.408
8	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.897	3.355	4.501	5.041
9	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.813	2.228	2.764	3.169	4.144	4.587
11	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.394	0.538	0.694	0.870	1.080	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.625	2.977	3.787	4.141
15	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.603	2.947	3.733	4.073
16	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.584	2.921	3.686	4.015
17	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.611	3.922
19	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.540	2.861	3.579	3.883
20	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.390	0.532	0.685	0.858	1.060	1.320	1.714	2.069	2.500	2.807	3.485	3.768
24	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
31	0.256	0.389	0.530	0.682	0.853	1.054	1.310	1.696	2.040	2.453	2.744	3.375	3.634
32	0.255	0.389	0.530	0.682	0.853	1.054	1.309	1.694	2.037	2.449	2.739	3.365	3.622
33	0.255	0.389	0.530	0.682	0.853	1.053	1.308	1.692	2.035	2.445	2.733	3.356	3.611
34	0.255	0.389	0.529	0.682	0.852	1.053	1.307	1.691	2.032	2.441	2.728	3.348	3.601
35	0.255	0.389	0.529	0.682	0.852	1.052	1.306	1.690	2.030	2.438	2.724	3.340	3.591
36	0.255	0.388	0.529	0.681	0.852	1.052	1.306	1.688	2.028	2.435	2.720	3.333	3.582
37	0.255	0.388	0.529	0.681	0.851	1.051	1.305	1.687	2.026	2.431	2.715	3.326	3.574
38	0.255	0.388	0.529	0.681	0.851	1.051	1.304	1.686	2.024	2.429	2.712	3.319	3.566
39	0.255	0.388	0.529	0.681	0.851	1.050	1.304	1.685	2.023	2.426	2.708	3.313	3.558
40	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.705	3.307	3.551
41	0.255	0.388	0.529	0.681	0.850	1.050	1.303	1.683	2.020	2.421	2.701	3.301	3.544
42	0.255	0.388	0.528	0.680	0.850	1.049	1.302	1.682	2.018	2.419	2.698	3.296	3.538
43	0.255	0.388	0.528	0.680	0.850	1.049	1.302	1.681	2.017	2.416	2.695	3.291	3.532
44	0.255	0.388	0.528	0.680	0.850	1.049	1.301	1.680	2.015	2.414	2.692	3.286	3.526
45	0.255	0.388	0.528	0.680	0.850	1.049	1.301	1.679	2.014	2.412	2.690	3.282	3.520
46	0.255	0.388	0.528	0.680	0.850	1.048	1.300	1.679	2.013	2.410	2.687	3.277	3.515
47	0.255	0.388	0.528	0.680	0.849	1.048	1.300	1.678	2.012	2.408	2.685	3.273	3.510
48	0.255	0.388	0.528	0.680	0.849	1.048	1.299	1.677	2.011	2.407	2.682	3.269	3.505
49	0.255	0.388	0.528	0.680	0.849	1.048	1.299	1.677	2.010	2.405	2.680	3.265	3.500

Loi de Student (suite et fin)

Quantiles $t_{\nu;p}$ de la variable de Student $t_{(\nu)}$

ν	p												
	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995	0.999	0.9995
50	0.255	0.388	0.528	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.261	3.496
51	0.255	0.388	0.528	0.679	0.849	1.047	1.298	1.675	2.008	2.402	2.676	3.258	3.492
52	0.255	0.387	0.528	0.679	0.849	1.047	1.298	1.675	2.007	2.400	2.674	3.255	3.488
53	0.255	0.387	0.528	0.679	0.848	1.047	1.298	1.674	2.006	2.399	2.672	3.251	3.484
54	0.255	0.387	0.528	0.679	0.848	1.047	1.297	1.674	2.005	2.397	2.670	3.248	3.480
55	0.255	0.387	0.527	0.679	0.848	1.046	1.297	1.673	2.004	2.396	2.668	3.245	3.476
56	0.255	0.387	0.527	0.679	0.848	1.046	1.297	1.673	2.003	2.395	2.667	3.242	3.473
57	0.255	0.387	0.527	0.679	0.848	1.046	1.297	1.672	2.003	2.394	2.665	3.240	3.470
58	0.255	0.387	0.527	0.679	0.848	1.046	1.296	1.672	2.002	2.392	2.663	3.237	3.466
59	0.254	0.387	0.527	0.679	0.848	1.046	1.296	1.671	2.001	2.391	2.662	3.234	3.463
60	0.254	0.387	0.527	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.232	3.460
61	0.254	0.387	0.527	0.679	0.848	1.045	1.296	1.670	2.000	2.389	2.659	3.229	3.457
62	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.670	1.999	2.388	2.658	3.227	3.455
63	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.669	1.998	2.387	2.656	3.225	3.452
64	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.669	1.998	2.386	2.655	3.223	3.449
65	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.669	1.997	2.385	2.654	3.220	3.447
66	0.254	0.387	0.527	0.678	0.847	1.045	1.295	1.668	1.997	2.384	2.652	3.218	3.444
67	0.254	0.387	0.527	0.678	0.847	1.045	1.294	1.668	1.996	2.383	2.651	3.216	3.442
68	0.254	0.387	0.527	0.678	0.847	1.044	1.294	1.668	1.996	2.382	2.650	3.215	3.439
69	0.254	0.387	0.527	0.678	0.847	1.044	1.294	1.667	1.995	2.382	2.649	3.213	3.437
70	0.254	0.387	0.527	0.678	0.847	1.044	1.294	1.667	1.994	2.381	2.648	3.211	3.435
71	0.254	0.387	0.527	0.678	0.847	1.044	1.294	1.667	1.994	2.380	2.647	3.209	3.433
72	0.254	0.387	0.527	0.678	0.847	1.044	1.293	1.666	1.994	2.379	2.646	3.207	3.431
73	0.254	0.387	0.527	0.678	0.847	1.044	1.293	1.666	1.993	2.379	2.645	3.206	3.429
74	0.254	0.387	0.527	0.678	0.847	1.044	1.293	1.666	1.993	2.378	2.644	3.204	3.427
75	0.254	0.387	0.527	0.678	0.846	1.044	1.293	1.665	1.992	2.377	2.643	3.203	3.425
76	0.254	0.387	0.527	0.678	0.846	1.044	1.293	1.665	1.992	2.376	2.642	3.201	3.423
77	0.254	0.387	0.527	0.678	0.846	1.044	1.293	1.665	1.991	2.376	2.641	3.200	3.421
78	0.254	0.387	0.527	0.678	0.846	1.043	1.293	1.665	1.991	2.375	2.640	3.198	3.420
79	0.254	0.387	0.527	0.678	0.846	1.043	1.292	1.664	1.991	2.375	2.640	3.197	3.418
80	0.254	0.387	0.527	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
81	0.254	0.387	0.526	0.678	0.846	1.043	1.292	1.664	1.990	2.373	2.638	3.194	3.415
82	0.254	0.387	0.526	0.677	0.846	1.043	1.292	1.664	1.989	2.373	2.637	3.193	3.413
83	0.254	0.387	0.526	0.677	0.846	1.043	1.292	1.663	1.989	2.372	2.636	3.191	3.412
84	0.254	0.387	0.526	0.677	0.846	1.043	1.292	1.663	1.989	2.372	2.636	3.190	3.410
85	0.254	0.387	0.526	0.677	0.846	1.043	1.292	1.663	1.988	2.371	2.635	3.189	3.409
86	0.254	0.387	0.526	0.677	0.846	1.043	1.292	1.663	1.988	2.371	2.634	3.188	3.407
87	0.254	0.387	0.526	0.677	0.846	1.043	1.291	1.663	1.988	2.370	2.634	3.187	3.406
88	0.254	0.387	0.526	0.677	0.846	1.043	1.291	1.662	1.987	2.370	2.633	3.185	3.405
89	0.254	0.387	0.526	0.677	0.846	1.043	1.291	1.662	1.987	2.369	2.632	3.184	3.403
90	0.254	0.387	0.526	0.677	0.846	1.042	1.291	1.662	1.987	2.369	2.632	3.183	3.402
91	0.254	0.387	0.526	0.677	0.846	1.042	1.291	1.662	1.986	2.368	2.631	3.182	3.401
92	0.254	0.387	0.526	0.677	0.846	1.042	1.291	1.662	1.986	2.368	2.630	3.181	3.399
93	0.254	0.387	0.526	0.677	0.846	1.042	1.291	1.661	1.986	2.367	2.630	3.180	3.398
94	0.254	0.387	0.526	0.677	0.845	1.042	1.291	1.661	1.986	2.367	2.629	3.179	3.397
95	0.254	0.386	0.526	0.677	0.845	1.042	1.291	1.661	1.985	2.366	2.629	3.178	3.396
96	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.661	1.985	2.366	2.628	3.177	3.395
97	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.661	1.985	2.365	2.628	3.176	3.394
98	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.661	1.985	2.365	2.627	3.176	3.393
99	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.660	1.984	2.365	2.626	3.175	3.392
100	0.254	0.386	0.526	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.391
200	0.254	0.386	0.525	0.676	0.843	1.039	1.286	1.653	1.972	2.345	2.601	3.132	3.340
500	0.253	0.386	0.525	0.675	0.842	1.038	1.283	1.648	1.965	2.334	2.586	3.107	3.310
∞	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Loi de Fisher-Snedecor (source : Droesbeke J-J. (1997))

Distribution de Fisher Snédecor $F(\nu_1 ; \nu_2)$

Quantiles d'ordre .95

ν_2 \ ν_1	1	2	3	4	5	6	8	10	12	15	20	30	60	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	238.88	241.88	243.91	245.95	248.01	250.10	252.20	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.40	19.41	19.43	19.45	19.46	19.48	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.85	8.79	8.74	8.70	8.66	8.62	8.57	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.96	5.91	5.86	5.80	5.75	5.69	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.74	4.68	4.62	4.56	4.50	4.43	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.06	4.00	3.94	3.87	3.81	3.74	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.64	3.57	3.51	3.44	3.38	3.30	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.35	3.28	3.22	3.15	3.08	3.01	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.14	3.07	3.01	2.94	2.86	2.79	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.98	2.91	2.85	2.77	2.70	2.62	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.85	2.79	2.72	2.65	2.57	2.49	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.85	2.75	2.69	2.62	2.54	2.47	2.38	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.77	2.67	2.60	2.53	2.46	2.38	2.30	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.60	2.53	2.46	2.39	2.31	2.22	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.54	2.48	2.40	2.33	2.25	2.16	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.49	2.42	2.35	2.28	2.19	2.11	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.45	2.38	2.31	2.23	2.15	2.06	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.41	2.34	2.27	2.19	2.11	2.02	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.38	2.31	2.23	2.16	2.07	1.98	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.35	2.28	2.20	2.12	2.04	1.95	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.32	2.25	2.18	2.10	2.01	1.92	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.30	2.23	2.15	2.07	1.98	1.89	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.37	2.27	2.20	2.13	2.05	1.96	1.86	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.25	2.18	2.11	2.03	1.94	1.84	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.34	2.24	2.16	2.09	2.01	1.92	1.82	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.32	2.22	2.15	2.07	1.99	1.90	1.80	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.31	2.20	2.13	2.06	1.97	1.88	1.79	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.29	2.19	2.12	2.04	1.96	1.87	1.77	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.28	2.18	2.10	2.03	1.94	1.85	1.75	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.16	2.09	2.01	1.93	1.84	1.74	1.62
31	4.16	3.30	2.91	2.68	2.52	2.41	2.25	2.15	2.08	2.00	1.92	1.83	1.73	1.61
32	4.15	3.29	2.90	2.67	2.51	2.40	2.24	2.14	2.07	1.99	1.91	1.82	1.71	1.59
33	4.14	3.28	2.89	2.66	2.50	2.39	2.23	2.13	2.06	1.98	1.90	1.81	1.70	1.58
34	4.13	3.28	2.88	2.65	2.49	2.38	2.23	2.12	2.05	1.97	1.89	1.80	1.69	1.57
35	4.12	3.27	2.87	2.64	2.49	2.37	2.22	2.11	2.04	1.96	1.88	1.79	1.68	1.56
36	4.11	3.26	2.87	2.63	2.48	2.36	2.21	2.11	2.03	1.95	1.87	1.78	1.67	1.55
37	4.11	3.25	2.86	2.63	2.47	2.36	2.20	2.10	2.02	1.95	1.86	1.77	1.66	1.54
38	4.10	3.24	2.85	2.62	2.46	2.35	2.19	2.09	2.02	1.94	1.85	1.76	1.65	1.53
39	4.09	3.24	2.85	2.61	2.46	2.34	2.19	2.08	2.01	1.93	1.85	1.75	1.65	1.52
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.08	2.00	1.92	1.84	1.74	1.64	1.51
50	4.03	3.18	2.79	2.56	2.40	2.29	2.13	2.03	1.95	1.87	1.78	1.69	1.58	1.44
60	4.00	3.15	2.76	2.53	2.37	2.25	2.10	1.99	1.92	1.84	1.75	1.65	1.53	1.39
70	3.98	3.13	2.74	2.50	2.35	2.23	2.07	1.97	1.89	1.81	1.72	1.62	1.50	1.35
80	3.96	3.11	2.72	2.49	2.33	2.21	2.06	1.95	1.88	1.79	1.70	1.60	1.48	1.32
90	3.95	3.10	2.71	2.47	2.32	2.20	2.04	1.94	1.86	1.78	1.69	1.59	1.46	1.30
100	3.94	3.09	2.70	2.46	2.31	2.19	2.03	1.93	1.85	1.77	1.68	1.57	1.45	1.28
110	3.93	3.08	2.69	2.45	2.30	2.18	2.02	1.92	1.84	1.76	1.67	1.56	1.44	1.27
120	3.92	3.07	2.68	2.45	2.29	2.18	2.02	1.91	1.83	1.75	1.66	1.55	1.43	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	1.94	1.83	1.75	1.67	1.57	1.46	1.32	1.00

Loi de Fisher-Snedecor (suite et fin)

Distribution de Fisher Snédecor $F(\nu_1 ; \nu_2)$.

Quantiles d'ordre .99

ν_2 \ ν_1	1	2	3	4	5	6	8	10	12	15	20	30	60	∞
1	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5981.1	6055.8	6106.3	6157.3	6208.7	6260.6	6313.0	6365.9
2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.40	99.42	99.43	99.45	99.47	99.48	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.23	27.05	26.87	26.69	26.51	26.32	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.55	14.37	14.20	14.02	13.84	13.65	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	10.05	9.89	9.72	9.55	9.38	9.20	9.02
6	13.75	10.93	9.78	9.15	8.75	8.47	8.10	7.87	7.72	7.56	7.40	7.23	7.06	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.62	6.47	6.31	6.16	5.99	5.82	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.81	5.67	5.52	5.36	5.20	5.03	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.26	5.11	4.96	4.81	4.65	4.48	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.85	4.71	4.56	4.41	4.25	4.08	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.74	4.54	4.40	4.25	4.10	3.94	3.78	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.30	4.16	4.01	3.86	3.70	3.54	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.30	4.10	3.96	3.82	3.66	3.51	3.34	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.14	3.94	3.80	3.66	3.51	3.35	3.18	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.80	3.67	3.52	3.37	3.21	3.05	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.69	3.55	3.41	3.26	3.10	2.93	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.79	3.59	3.46	3.31	3.16	3.00	2.83	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.71	3.51	3.37	3.23	3.08	2.92	2.75	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.43	3.30	3.15	3.00	2.84	2.67	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.37	3.23	3.09	2.94	2.78	2.61	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.31	3.17	3.03	2.88	2.72	2.55	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.45	3.26	3.12	2.98	2.83	2.67	2.50	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.21	3.07	2.93	2.78	2.62	2.45	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.17	3.03	2.89	2.74	2.58	2.40	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	3.13	2.99	2.85	2.70	2.54	2.36	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	3.09	2.96	2.82	2.66	2.50	2.33	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	3.06	2.93	2.78	2.63	2.47	2.29	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	3.03	2.90	2.75	2.60	2.44	2.26	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	3.00	2.87	2.73	2.57	2.41	2.23	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.98	2.84	2.70	2.55	2.39	2.21	2.01
31	7.53	5.36	4.48	3.99	3.67	3.45	3.15	2.96	2.82	2.68	2.52	2.36	2.18	1.98
32	7.50	5.34	4.46	3.97	3.65	3.43	3.13	2.93	2.80	2.65	2.50	2.34	2.16	1.96
33	7.47	5.31	4.44	3.95	3.63	3.41	3.11	2.91	2.78	2.63	2.48	2.32	2.14	1.93
34	7.44	5.29	4.42	3.93	3.61	3.39	3.09	2.89	2.76	2.62	2.46	2.30	2.12	1.91
35	7.42	5.27	4.40	3.91	3.59	3.37	3.07	2.88	2.74	2.60	2.44	2.28	2.10	1.89
36	7.40	5.25	4.38	3.89	3.57	3.35	3.05	2.86	2.72	2.58	2.43	2.26	2.08	1.87
37	7.37	5.23	4.36	3.87	3.56	3.33	3.04	2.84	2.71	2.56	2.41	2.25	2.06	1.85
38	7.35	5.21	4.34	3.86	3.54	3.32	3.02	2.83	2.69	2.55	2.40	2.23	2.05	1.84
39	7.33	5.19	4.33	3.84	3.53	3.30	3.01	2.81	2.68	2.54	2.38	2.22	2.03	1.82
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.80	2.66	2.52	2.37	2.20	2.02	1.80
50	7.17	5.06	4.20	3.72	3.41	3.19	2.89	2.70	2.56	2.42	2.27	2.10	1.91	1.68
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.63	2.50	2.35	2.20	2.03	1.84	1.60
70	7.01	4.92	4.07	3.60	3.29	3.07	2.78	2.59	2.45	2.31	2.15	1.98	1.78	1.54
80	6.96	4.88	4.04	3.56	3.26	3.04	2.74	2.55	2.42	2.27	2.12	1.94	1.75	1.49
90	6.93	4.85	4.01	3.54	3.23	3.01	2.72	2.52	2.39	2.24	2.09	1.92	1.72	1.46
100	6.90	4.82	3.98	3.51	3.21	2.99	2.69	2.50	2.37	2.22	2.07	1.89	1.69	1.43
110	6.87	4.80	3.96	3.49	3.19	2.97	2.68	2.49	2.35	2.21	2.05	1.88	1.67	1.40
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.47	2.34	2.19	2.03	1.86	1.66	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.51	2.32	2.18	2.04	1.88	1.70	1.47	1.00

Partie II

Inférence statistique

Chapitre 8

Population, échantillon et échantillonnage aléatoire

Dans cette seconde partie du cours, on étudie un ensemble de méthodes statistiques de base, fondées sur les concepts et outils de la théorie des probabilités, permettant de tirer des conclusions fiables concernant les caractéristiques d'une population sur base de l'observation des caractéristiques d'un échantillon issu de cette population.

8.1. Population et échantillon

Commençons par préciser le sens des termes utilisés.

On appelle *population* la totalité des objets de même nature (quelle qu'elle soit) pris en considération dans le cadre d'un problème. Il peut s'agir d'êtres humains, de légumes, d'entreprises, de pièces fabriquées par une machine, etc... On désigne les éléments de la population par le terme général d'*individus* (même s'il s'agit de haricots verts !). Le nombre d'individus composant la population est appelé la *taille* de la population. Ce nombre est généralement noté N.

Par *caractéristique d'une population*, on entend un résumé statistique des valeurs individuelles prises par une variable — qualifiée de *variable d'intérêt* et qui mesure une caractéristique particulière des individus de la population — dans cette population. Par exemple, le revenu moyen des ménages d'une zone géographique donnée, la proportion de pièces défectueuses dans une production, etc...

Un *échantillon* est un ensemble d'individus tirés de la population étudiée. Il peut par exemple s'agir de 2 000 ménages prélevés dans l'ensemble de tous les ménages belges, de 500 pièces prélevées dans une production, etc... On appelle *taille* de l'échantillon le nombre de tirages effectués pour constituer l'échantillon. On note habituellement ce nombre n. De façon analogue au cas d'une population, on entend par *caractéristique d'un échantillon* un résumé statistique des valeurs individuelles prises par une variable dans cet échantillon.

8.2. Recensement et sondage

A priori, lorsqu'on souhaite étudier une caractéristique d'une population, deux démarches sont possibles.

On peut tout d'abord choisir d'observer les valeurs de la variable d'intérêt pour l'ensemble des individus de la population, et d'en déduire la caractéristique de la population à laquelle on s'intéresse. On dit dans ce cas qu'on procède par *recensement*. Ainsi, pour connaître la valeur moyenne du montant hebdomadaire d'argent de poche dont disposent les étudiants de la Faculté, on demandera à chacun des étudiants de la Faculté le montant hebdomadaire d'argent de poche dont il dispose, et on en déduira la valeur moyenne recherchée.

Alternativement, on peut choisir de n'observer les valeurs de la variable d'intérêt que pour un échantillon des individus de la population, et d'en déduire alors une estimation de la caractéristique de la population à laquelle on s'intéresse. On dit dans ce cas qu'on procède par *sondage*. Dans l'exemple ci-dessus, cela signifie tirer un échantillon d'étudiants dans l'ensemble des étudiants de la Faculté, demander à chacun des étudiants tirés le montant hebdomadaire d'argent de poche dont il dispose, et en déduire une estimation de la valeur moyenne recherchée.

Le recensement a pour principal avantage de délivrer (à tout le moins en théorie) des résultats exacts : on peut connaître avec exactitude toutes les caractéristiques de la population étudiée, puisqu'on en observe tous les individus et leurs caractéristiques. Il présente cependant un sérieux inconvénient : lorsque la taille de la population est importante, sa réalisation est coûteuse à la fois en argent (recrutement, formation et rétribution d'un nombre important d'enquêteurs, etc...) et en temps (temps nécessaire au recueil de l'information et au dépouillement des résultats, etc...). Pour cette raison, le choix du recensement est le plus souvent assez peu réaliste : dans nombre de situations, il est impensable d'y recourir.

On procède alors par sondage. Ses avantages sont évidents : coût financier plus limité, plus grande facilité de mise en oeuvre, délai d'obtention des résultats plus court, etc... Ils ont cependant un prix : un sondage ne peut, au contraire du recensement, au mieux que délivrer des résultats approximatifs, puisqu'on n'observe les caractéristiques que d'une partie seulement des individus de la population.

Heureusement, comme nous aurons l'occasion de le développer dans la suite, il est parfaitement possible d'assurer la fiabilité des résultats obtenus par sondage. Pour cela, il est toutefois nécessaire d'adopter des méthodes rigoureuses de prélèvement d'échantillons.

8.3. Echantillonnage aléatoire

La plus simple et aussi la plus importante des méthodes de prélèvement d'échantillons permettant d'assurer la fiabilité des résultats obtenus par sondage est l'*échantillonnage aléatoire simple*, que l'on désigne généralement plus succinctement par l'expression *échantillonnage aléatoire* sans autre précision, ce que nous ferons dans la suite.

Le prélèvement par échantillonnage aléatoire d'un échantillon de taille n dans une population consiste simplement à tirer au hasard et avec remise n individus dans l'ensemble des N individus de la population. Un échantillon obtenu par échantillonnage aléatoire est qualifié d'*échantillon aléatoire*.

On peut physiquement se représenter l'échantillonnage aléatoire de la façon suivante : on place dans une urne N boules, chaque boule représentant un des individus de la population, et on tire au hasard avec remise n boules de l'urne.

Notez que, lorsque la taille n de l'échantillon prélevé est petite par rapport à la taille N de la population, c'est-à-dire lorsque le *taux de sondage* n/N est faible (au plus quelques pour cent, c'est en pratique le plus souvent le cas), la précision tirages au hasard *avec remise* est sans grande importance : comme le suggère l'intuition, dans ce cas de figure, tirer au hasard avec ou sans remise ne fait pas une grande différence. Nous verrons effectivement dans la suite que lorsque la taille de l'échantillon prélevé est petite par rapport à celle de la population, on peut pour l'essentiel traiter les tirages sans remise comme s'il s'agissait de tirages avec remise. Il n'en va évidemment pas de même si la taille de l'échantillon prélevé représente une part substantielle de celle de la population. Dans ce cas, il faut tenir compte des différences. Dans le cadre de ce cours, nous supposerons toujours qu'on a affaire à des tirages avec remise, ou à tout le moins que les conditions sont réunies pour qu'on puisse les considérer comme tels.

En pratique, pour tirer un échantillon aléatoire dans une population, on peut faire appel à des méthodes reposant sur l'utilisation de nombres aléatoires, nombres qui peuvent être obtenus par ordinateur à l'aide de programmes appelés *générateurs de nombres aléatoires*. Ces générateurs de nombres aléatoires fournissent des suites de nombres — appelées *suites de nombres aléatoires* — correspondant à des tirages indépendants dans une loi continue uniforme $\mathcal{U}(0;1)$, c'est-à-dire des suites de réalisations indépendantes de la variable aléatoire $X \rightsquigarrow \mathcal{U}(0;1)$. Voici un exemple d'une telle suite (8 premières décimales de chaque nombre) :

$$0,37005162\,;\ 0,10346736\,;\ 0,35641634\,;\ 0,99716867\,;\ 0,18446503\,;\ 0,78754947\,;\ \ldots$$

Voyons maintenant comment procéder au tirage d'un échantillon aléatoire à l'aide d'une telle suite aléatoire. Supposons qu'on souhaite tirer un échantillon aléatoire de 5 étudiants dans la population des 345 étudiants de candidature de la Faculté. On se procurera au secrétariat de la Faculté une liste donnant le nom et le prénom de tous les étudiants de candidature et on attribuera (arbitrairement) à chaque étudiant de la liste un numéro allant de 1 à 345. On obtiendra ainsi une liste numérotée telle que :

1	Pierre Dupont
2	Jacqueline Durand
3	Annie Martin
⋮	⋮
345	Paul Lebon

Sur base d'une telle liste, tirer un échantillon aléatoire de 5 étudiants parmi les 345 étudiants revient simplement à tirer au hasard avec remise 5 nombres dans l'ensemble des 345 nombres $\{1, 2, ..., 345\}$.

Pour ce faire, on découpe l'intervalle $[0, 1]$ en 345 intervalles d'égale longueur et on attribue à chacune des valeurs x_i de la suite obtenue du générateur de nombres aléatoires les nombres $1, 2, ..., 345$ suivant la règle d'association :

$$\begin{array}{ll}
\text{si } x_i \in [0, \frac{1}{345}[& \rightarrow 1 \\
\text{si } x_i \in [\frac{1}{345}, \frac{2}{345}[& \rightarrow 2 \\
\text{si } x_i \in [\frac{2}{345}, \frac{3}{345}[& \rightarrow 3 \\
\quad\quad \vdots & \quad \vdots \\
\text{si } x_i \in [\frac{344}{345}, 1] & \rightarrow 345
\end{array}$$

Dans le cas de la suite aléatoire reproduite ci-dessus, on obtient :

$$128\,;\ 36\,;\ 123\,;\ 345\,;\ 64\,;\ 272\,;\ ...$$

Par construction, ces nombres forment une suite de réalisations indépendantes de la variable aléatoire discrète uniforme $X \rightsquigarrow \mathcal{U}(1; 2; ...; 345)$. En d'autres termes, la suite ainsi obtenue correspond à une suite de tirages au hasard avec remise dans l'ensemble des 345 nombres $\{1, 2, ..., 345\}$.

Les étudiants dont le numéro d'ordre correspond aux 5 premiers nombres de la suite : 128 ; 36 ; 123 ; 345 ; 64 forment un échantillon aléatoire de 5 étudiants tirés parmi les 345 étudiants de candidature.

Notez que l'utilisation d'une autre suite de nombres aléatoires aurait conduit à un autre échantillon, de même qu'un autre classement des étudiants sur la liste.

La procédure d'échantillonnage aléatoire décrite ci-dessus est assez pratique et relativement simple. On voit toutefois que sa mise en oeuvre requiert la disponibilité d'une liste des individus de la population concernée.

Chapitre 9

Propriétés de l'échantillonnage aléatoire

S'il est possible de tirer des conclusions fiables concernant les caractéristiques d'une population sur base de l'observation des caractéristiques d'un échantillon prélevé par échantillonnage aléatoire dans cette population, c'est parce qu'il existe des liens étroits — qu'on devine de nature probabiliste — entre les caractéristiques d'un échantillon aléatoire et celles de la population dont il est issu.

Le présent chapitre est consacré à l'examen de quelques-uns des plus importants de ces liens, à savoir les liens existant entre les résumés statistiques les plus couramment utilisés pour caractériser une population et un échantillon aléatoire issu de cette population : moyenne et variance des valeurs prises par une variable et fréquence d'une caractéristique individuelle.

9.1. Caractéristiques de la population, échantillon aléatoire et caractéristiques d'un échantillon aléatoire

On commence par décrire en termes probabilistes la population et ses caractéristiques, un échantillon aléatoire issu de cette population et les caractéristiques d'un échantillon aléatoire issu de cette population.

9.1.1. La population et ses caractéristiques

Supposons que nous soyons chargé de réaliser une étude de marché pour une entreprise qui souhaite lancer une nouvelle marque de yaourt. On doit déterminer le marché potentiel pour ce nouveau produit. On commence par analyser la consommation de yaourt en Belgique. Il est évidemment hors de question de procéder à un recensement. On décide donc de réaliser un sondage et de demander aux personnes interrogées le nombre de yaourts qu'elles ont consommés durant la semaine écoulée.

La population est l'ensemble des consommateurs résidant en Belgique et la variable d'intérêt est le nombre de yaourts consommés par semaine. La variable d'intérêt prend une valeur déterminée pour chaque individu de la population. L'ensemble de la population est caractérisée par une certaine distribution de fréquence pour cette variable d'intérêt. A cette distribution de fréquence, on peut associer différents résumés statistiques (moyenne, variance, etc...), qui constituent autant de caractéristiques de la population concernant la variable d'intérêt "nombre de yaourts consommés par semaine".

Supposons qu'on tire au hasard un individu dans la population et qu'on note X sa consommation hebdomadaire de yaourt. La variable aléatoire X a une certaine distribution de probabilité, une loi. Notons cette loi $\ell(x)$, où $\ell(x)$ représente la probabilité que X prenne la valeur x : $\ell(x) = \mathbb{P}(X = x)$. A quoi correspond cette loi ? On le sait (cf. Chapitre 4), tout simplement à la distribution de fréquence de la variable "nombre de yaourts consommés par semaine" dans la population. L'espérance et la variance de la quantité de yaourts consommés sont deux caractéristiques, deux paramètres de la loi de X. Notons l'espérance m et la variance σ^2. A quoi sont égaux m et σ^2 ? On l'a aussi déjà vu (cf. à nouveau Chapitre 4), tout simplement à la moyenne et la variance dans la population du nombre de yaourts consommés par semaine, c'est-à-dire aux caractéristiques "moyenne" et "variance" de la population concernant la variable d'intérêt "nombre de yaourts consommés par semaine". Comme on le conçoit aisément, de la même façon, à chacune des caractéristiques de la population concernant la variable d'intérêt "nombre de yaourts consommés par semaine" correspond ainsi une caractéristique, un paramètre bien précis — tout simplement l'analogue du résumé statistique considéré pour la population — de la loi $\ell(x)$ de la variable aléatoire X.

Le même raisonnement tient si la variable d'intérêt dans la population est une variable (approximativement) continue, comme par exemple la taille des individus. Simplement, dans ce cas, la loi de X n'est plus une loi discrète mais une loi continue et $\ell(x)$ est sa fonction de densité, fonction de densité qui, on le sait (cf. à nouveau Chapitre 4), correspond à la courbe limite, obtenue lorsqu'on considère des classes de plus en plus fines, de l'histogramme de fréquence de la variable en question dans la population.

On déduit de ce qui précède qu'on peut, au travers de l'épreuve aléatoire "tirer au hasard un individu dans la population et observer la valeur de la variable d'intérêt pour l'individu tiré", décrire une population et ses caractéristiques concernant une variable d'intérêt par la loi $\ell(x)$ d'une variable aléatoire X, les paramètres de cette loi $\ell(x)$ identifiant de façon univoque les diverses caractéristiques de la population.

Définition 9.1 *On représente une population et ses caractéristiques concernant une variable d'intérêt par la loi d'une variable aléatoire X et ses paramètres. On appelle loi de la population la loi de la variable aléatoire X.*

9.1.2. Echantillon aléatoire

Le prélèvement par échantillonnage aléatoire d'un échantillon de taille n dans une population en vue d'en étudier les caractéristiques concernant une variable d'intérêt consiste à tirer au hasard et avec remise n individus dans la population et à observer les valeurs de la variable d'intérêt pour les n individus tirés, c'est-à-dire à répéter n fois, en remettant à chaque fois l'individu tiré, l'épreuve aléatoire "tirer au hasard un individu dans la population et observer la valeur de la variable d'intérêt pour l'individu tiré".

Le résultat de chacun de ces tirages, qui forcément varie d'un échantillon à l'autre, peut être représenté par une variable aléatoire X_i $(i = 1, 2, ..., n)$. Les tirages se faisant avec remise, la distribution de la variable d'intérêt dans la population reste inchangée d'un tirage à l'autre (puisque chaque tirage est effectué parmi tous les individus de la population) et le résultat d'un tirage n'a aucun influence sur celui d'un autre tirage. On en déduit que les variables aléatoires $X_1, X_2, ..., X_n$ utilisées pour décrire le résultat de chacun des n tirages sont indépendantes et toutes de même loi que la loi $\ell(x)$ de la variable aléatoire X qui représente la population.

Quid si les tirages sont effectués sans remise plutôt qu'avec remise ? Dans ce cas, la distribution de la variable d'intérêt dans la population change à chaque tirage (puisque chaque tirage est effectué parmi tous les individus de la population, moins ceux déjà tirés au cours des tirages précédents), et donc le résultat de chaque tirage est influencé par les résultats de tous les tirages précédents. On voit toutefois aisément que si la taille N de la population est grande, les changements induits par l'absence de remise sur la distribution de la variable d'intérêt dans la population, et donc l'influence des tirages déjà effectués sur le tirage suivant, est négligeable pour les premiers tirages, et le reste tant que le nombre de tirages effectués, c'est-à-dire la taille n de l'échantillon prélevé, reste petite par rapport la taille N de la population. En d'autres termes, comme suggéré au Chapitre 8, si la taille n de l'échantillon prélevé est petite par rapport à la taille N de la population, c'est-à-dire si le *taux de sondage* n/N est faible, tirer avec ou sans remise, c'est vert-chou et chou-vert : en pratique, on peut négliger la légère différence qu'induit la non-remise et traiter les tirages sans remise comme s'il s'agissait de tirages avec remise.

On voit ainsi qu'on peut décrire le résultat du tirage d'un échantillon obtenu par échantillonnage aléatoire — avec remise ou sans remise si la taille de l'échantillon prélevé est petite par rapport à celle de la population — dans la population par un n-uple de variables aléatoires $(X_1, X_2, ..., X_n)$ indépendantes et toutes de même loi — on dit qu'elles sont *identiquement et indépendamment distribuées* (i.i.d.) — que la variable aléatoire X qui représente la population.

Définition 9.2 *On représente un échantillon aléatoire de taille n prélevé dans une population décrite par la loi d'une variable aléatoire X par un n-uple de variables aléatoires $(X_1, X_2, ..., X_n)$ qui, par construction, sont indépendantes et toutes de même loi que X. On appelle ce n-uple de variables aléatoires $(X_1, X_2, ..., X_n)$ indépendantes et toutes de même loi que X un échantillon aléatoire de taille n relatif à la variable aléatoire X.*

Notons que, pour un échantillon aléatoire $(X_1, X_2, ..., X_n)$, on a par construction :

$$\begin{aligned} E(X_1) &= E(X_2) = ... = E(X_n) = E(X) = m \\ V(X_1) &= V(X_2) = ... = V(X_n) = V(X) = \sigma^2 \end{aligned}$$

Remarquons également que l'indépendance permet d'obtenir une expression simple pour la loi jointe du n-uple de variables aléatoires $(X_1, X_2, ..., X_n)$, c'est-à-dire la loi jointe de l'échantillon aléatoire. Cette loi jointe est appelée *vraisemblance* et notée $\mathcal{L}(x_1, x_2, ..., x_n)$. Elle est donnée par le produit des lois marginales :

$$\mathcal{L}(x_1, x_2, ..., x_n) = \ell(x_1)\ell(x_2) \times ... \times \ell(x_n)$$

Dans le cas discret, $\mathcal{L}(x_1, x_2, ..., x_n) = I\!P(X_1 = x_1$ et $X_2 = x_2$ et ... et $X_n = x_n)$. Dans le cas continu, $\mathcal{L}(x_1, x_2, ..., x_n)$ représente la fonction de densité jointe.

En pratique, le tirage d'un échantillon aléatoire de taille n dans une population donne un ensemble de n valeurs numériques $(x_1, x_2, ..., x_n)$, dans notre exemple le nombre de yaourts consommés par semaine des n individus interrogés. Cet ensemble de n valeurs numériques est appelé un échantillon d'*observations*. Il constitue une réalisation particulière de l'échantillon aléatoire $(X_1, X_2, ..., X_n)$. Il convient de ne pas confondre l'échantillon aléatoire $(X_1, X_2, ..., X_n)$, qui décrit ce qu'on est susceptible d'obtenir lorsqu'on tire un échantillon par échantillonnage aléatoire, et l'échantillon d'observations $(x_1, x_2, ..., x_n)$, qui est un ensemble de nombres (et donc non aléatoire) et qui décrit ce qui est effectivement obtenu lors d'un tirage particulier, c'est-à-dire une réalisation particulière de l'échantillon aléatoire $(X_1, X_2, ..., X_n)$. La notation adoptée jusqu'ici et que nous maintiendrons ne laisse aucun doute : les variables aléatoires sont toujours notées en lettre majuscule et leurs réalisations toujours notées en lettre minuscule.

9.1.3. Caractéristiques d'un échantillon aléatoire

A l'instar de la population, on peut synthétiser les caractéristiques d'un échantillon obtenu par échantillonnage aléatoire par divers résumés statistiques (la moyenne, la variance, etc...) des valeurs prises par la variable d'intérêt dans l'échantillon.

Calculer un résumé statistique, tel que par exemple la moyenne, des valeurs prises par une variable d'intérêt dans un échantillon, cela revient à calculer une certaine fonction, pour la moyenne : $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, des valeurs x_i prises par cette variable dans l'échantillon.

Dans un échantillon obtenu par échantillonnage aléatoire dans une population, les valeurs prises par la variable d'intérêt varient forcément d'un échantillon à l'autre, et, en conséquence, la valeur d'un résumé statistique des valeurs prises par la variable d'intérêt dans un échantillon aléatoire varie elle-même d'un échantillon à l'autre. Les valeurs qu'on est susceptible d'obtenir dans un échantillon aléatoire prélevé dans une population représentée par la loi d'une variable aléatoire X étant décrites par l'échantillon aléatoire $(X_1, X_2, ..., X_n)$, la forme et le résultat du calcul d'un résumé statistique des valeurs prises par la variable d'intérêt dans un tel échantillon aléatoire peuvent être décrits par une variable aléatoire définie comme une fonction — celle qui correspond au résumé statistique étudié — de l'échantillon aléatoire $(X_1, X_2, ..., X_n)$, dans le cas du calcul de la moyenne : $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Une telle variable aléatoire définie comme une fonction d'un échantillon aléatoire $(X_1, X_2, ..., X_n)$ est appelée une *statistique*.

Définition 9.3 *On représente les caractéristiques d'un échantillon aléatoire* $(X_1, X_2, ..., X_n)$ *de taille* n *prélevé dans une population décrite par la loi d'une variable aléatoire* X *par des statistiques. On appelle statistique toute variable aléatoire* G_n *définie comme une fonction* $G_n = g(X_1, X_2, ..., X_n)$ *d'un échantillon aléatoire* $(X_1, X_2, ..., X_n)$.

Ayant formalisé en termes probabilistes, au travers des concepts de loi d'une variable aléatoire X, d'échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à X et de statistiques $G_n = g(X_1, X_2, ..., X_n)$, respectivement la population et ses caractéristiques, un échantillon aléatoire issu de cette population et les caractéristiques d'un échantillon aléatoire issu de cette population, nous sommes maintenant en mesure d'expliciter les liens existant entre les résumés statistiques les plus couramment utilisés pour caractériser une population et un échantillon aléatoire issu de cette population : moyenne et variance des valeurs prises par une variable et fréquence d'une caractéristique individuelle.

9.2. Moyenne dans la population et moyenne-échantillon

On examine tout d'abord les liens existant entre moyenne des valeurs prises par une variable d'intérêt dans une population et moyenne des valeurs prises par cette variable dans un échantillon aléatoire issu de cette population.

9.2.1. La moyenne-échantillon $\overline{X}_n$

Revenons à notre exemple initial. Supposons qu'on s'intéresse à la moyenne dans la population du nombre de yaourts consommés par semaine. Cette moyenne est identifiée par le paramètre $m = E(X)$ de la loi $\ell(x)$ de la variable aléatoire X représentant la population.

Supposons qu'on prélève un échantillon aléatoire de taille n dans la population et qu'on calcule le nombre moyen de yaourts consommés par les individus repris dans l'échantillon. Comme expliqué ci-avant, on peut représenter la forme et le résultat — qui forcément varie d'un échantillon à l'autre — de ce calcul par une statistique que, dans le cas qui nous occupe, on appelle la *moyenne-échantillon*.

Définition 9.4 *Soit* $(X_1, X_2, ..., X_n)$ *un échantillon aléatoire relatif à une variable aléatoire* X. *On appelle moyenne-échantillon la statistique :*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Intuitivement, on s'attend à ce que la moyenne-échantillon $\overline{X}_n$ reflète la moyenne $m = E(X)$ dans la population. Quel lien peut-on établir entre la moyenne $m =$

$E(X)$ dans la population, qui est une valeur fixe, non aléatoire, et la moyenne-échantillon $\overline{X}_n$, qui est une variable aléatoire, dont la valeur varie d'un échantillon à l'autre ? La réponse à cette question est donnée par la loi de $\overline{X}_n$.

9.2.2. Distribution d'échantillonnage de $\overline{X}_n$

Comme toute variable aléatoire, $\overline{X}_n$ a une loi que dans le présent contexte et pour des raisons évidentes on nomme généralement *distribution d'échantillonnage.*

La distribution d'échantillonnage de $\overline{X}_n$ est la loi de la variable aléatoire définie par la fonction $g(X_1, X_2, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ du n-uple de variables aléatoires $(X_1, X_2, ..., X_n)$. Dans le cas discret, suivant la logique développée à la Section 5.4.1 pour le cas d'une fonction de deux variables aléatoires et généralisée au cas d'une fonction de n variables aléatoires à la Section 6.3.2, elle est donnée[13] par un ensemble de couples $(\overline{x}_n, p_{\overline{x}_n})$, où les $\overline{x}_n$ désignent les valeurs possibles de $\overline{X}_n$ et les $p_{\overline{x}_n} = I\!P(\overline{X}_n = \overline{x}_n)$ les probabilités de ces valeurs possibles, probabilités qui sont données par :

$$I\!P(\overline{X}_n = \overline{x}_n) = \sum_{(x_1,x_2,...,x_n):\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n} I\!P(X_1 = x_1 \text{ et } X_2 = x_2 \text{ et ... et } X_n = x_n)$$

où $\sum_{(x_1,x_2,...,x_n):\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n}$ signifie "somme pour tous les n-uples de valeurs possibles $(x_1, x_2, ..., x_n)$ de $(X_1, X_2, ..., X_n)$ tels que $\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n$". Le même type de calcul peut être effectué dans le cas continu.

La distribution d'échantillonnage de $\overline{X}_n$ dépend de deux éléments : d'une part de la loi $\ell(x)$ de X et d'autre part de la taille n de l'échantillon. Le Graphique 9.1 illustre, de façon quelque peu stylisée[14], cette double dépendance.

Hormis quelques cas spécifiques, le calcul de la distribution d'échantillonnage exacte de $\overline{X}_n$ est le plus souvent très malaisé. Cependant, on peut très facilement en calculer ses deux premiers moments : son espérance et sa variance. En effet, de la Propriété 5.15 et du fait que pour tout i, $E(X_i) = m$, on a :

$$E(\overline{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}nm = m$$

Par ailleurs, $X_1, X_2, ..., X_n$ étant indépendantes, de la Propriété 5.16 et du fait que pour tout i, $V(X_i) = \sigma^2$, on a :

$$V(\overline{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

---

[13] Pour alléger la notation, on omet d'indicer les valeurs possibles des variables aléatoires considérées.

[14] Les graphiques de droite correspondent à des lois discrètes. Plutôt que d'être représentées par des petits rectangles, les probabilités devraient être représentées par des bâtonnets. Par ailleurs, ces graphiques sont approximatifs (source : Wannacott T.H.et Wannacott R.J.(1991)).

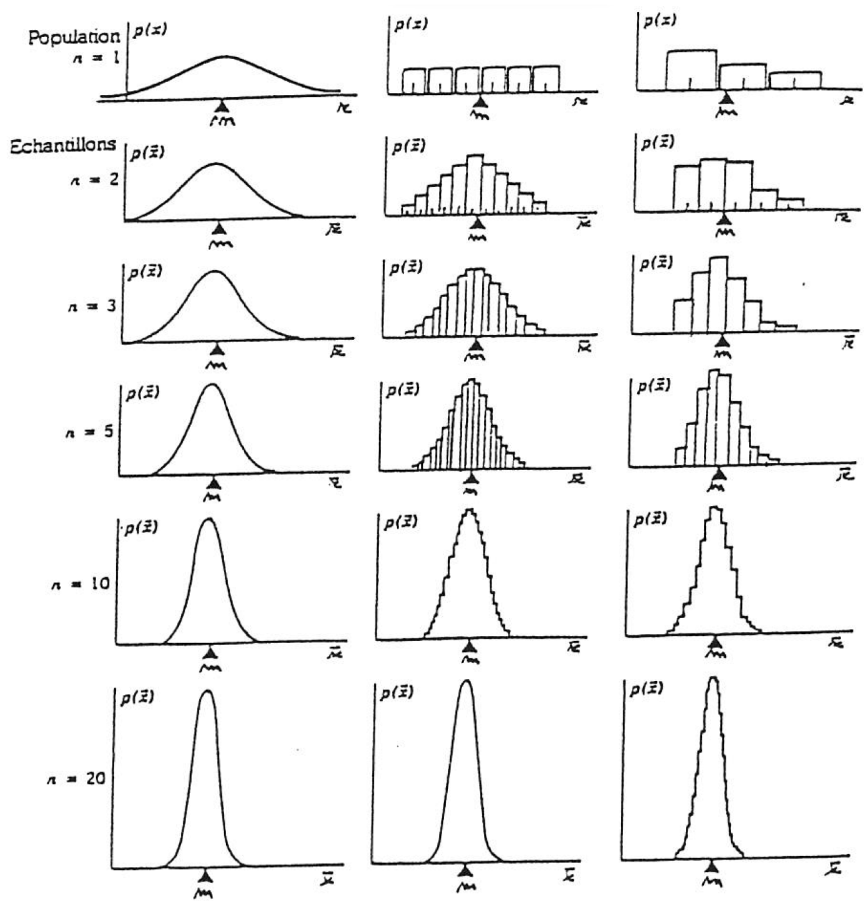
Propriété 9.1 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. On a :*

$$E(\overline{X}_n) = m \quad et \quad V(\overline{X}_n) = \frac{\sigma^2}{n}$$

Notez bien que cette propriété est vraie quels que soient n et la loi de la variable aléatoire X.

On constate que l'espérance de la moyenne-échantillon — la valeur espérée de $\overline{X}_n$, sa tendance centrale — est égale à la moyenne $m = E(X)$ dans la population et que la variance de $\overline{X}_n$, qui mesure la dispersion de $\overline{X}_n$ autour de sa valeur espérée m, est d'autant petite que la taille n de l'échantillon est grande et que la variance $\sigma^2 = V(X)$ dans la population est faible.

Autrement dit, comme le suggère également le Graphique 9.1, la moyenne-échantillon $\overline{X}_n$ prend des valeurs centrées sur la moyenne $m = E(X)$ dans la population, et ces valeurs tendent à être d'autant plus proche de cette moyenne $m = E(X)$ dans la population que la taille n de l'échantillon est grande et que la variance $\sigma^2 = V(X)$ dans la population est faible.



Graphique 9.1 : Quelques exemples de distributions d'échantillonnage de $\overline{X}_n$

9.2.3. Exercice résolu

On suppose que le poids à la naissance d'un nouveau-né est en moyenne de $3,50$ kg, avec un écart-type de 300 gr. On tire trois échantillons aléatoires de respectivement 5, 25 et 100 nouveau-nés et on examine leur poids à la naissance. Quels sont l'espérance et l'écart-type du poids moyen à la naissance des nouveaux-nés dans ces trois échantillons ?

Notons X le poids à la naissance d'un nouveau-né. La loi de X (= la loi de la population) est inconnue. Nous savons toutefois que cette loi est telle que $m = E(X) = 3,5$ (= la moyenne du poids à la naissance dans la population) et que $\sigma = \sigma(X) = 0,3$ (= l'écart-type du poids à la naissance dans la population). Chacun des échantillons aléatoires est représenté par n ($n = 5, 25, 100$) variables aléatoires $X_1, X_2, ..., X_n$ indépendantes et de même loi que X.

De la propriété 9.1, l'espérance, la variance et l'écart-type du poids moyen à la naissance ($= \overline{X}_n$) dans un échantillon aléatoire de n nouveau-nés sont donnés par :

$$E(\overline{X}_n) = m = 3,5 \ , \quad V(\overline{X}_n) = \frac{\sigma^2}{n} = \frac{0,3^2}{n} \quad \text{et} \quad \sigma(\overline{X}_n) = \sqrt{V(\overline{X}_n)} = \sqrt{\frac{0,3^2}{n}}$$

Pour des échantillons aléatoires de 5, 25 et 100 nouveau-nés, on a donc :

n	$E(\overline{X}_n)$	$V(\overline{X}_n)$	$\sigma(\overline{X}_n)$
5	$3,5$	$0,018$	$0,134...$
25	$3,5$	$0,0036$	$0,06$
100	$3,5$	$0,0009$	$0,03$

9.2.4. Comportement asymptotique de $\overline{X}_n$: loi des grands nombres et théorème central limite

La distribution d'échantillonnage exacte — on dit en *échantillon fini* — de $\overline{X}_n$ dépend de la loi de la population et est le plus souvent très malaisée à calculer. Comme on vient de le voir, seules l'espérance et la variance de cette distribution peuvent être aisément obtenues.

En grand échantillon, ou plus précisément lorsque la taille d'échantillon n tend vers l'infini, la distribution d'échantillonnage de la moyenne-échantillon peut toutefois être caractérisée indépendamment de la loi de la population : elle ne dépend plus de la loi $\ell(x)$ de X. Notez que la proposition "la taille d'échantillon tend vers l'infini" est physiquement réalisable : les tirages d'un échantillon aléatoire se faisant avec remise, on peut obtenir des échantillons de taille aussi grande que l'on veut, y compris de taille plus grande que celle de la population Il est clair qu'en pratique on ne tire jamais de tels échantillons. L'analyse de ce qui se passe lorsque "la taille d'échantillon tend vers l'infini" — on parle alors de *comportement asymptotique* — doit simplement être compris comme un moyen commode d'étudier ce qui se passe lorsque la taille d'échantillon grandit.

Revenons un instant au Graphique 9.1. A son examen, on constate que quelle que soit la loi de X, à mesure que n grandit :

1- la distribution d'échantillonnage de $\overline{X}_n$ est de plus en plus concentrée autour de sa valeur espérée $m = E(X)$. Cela est bien entendu lié au fait que $E(\overline{X}_n) = m$ et que la variance $V(\overline{X}_n) = \sigma^2/n$ diminue lorsque n augmente.
2- tout en se concentrant toujours davantage, la distribution d'échantillonnage de $\overline{X}_n$ prend une forme qui se rapproche de plus en plus de la forme en cloche typique de la loi normale.

Ces deux éléments caractérisent ce qu'on appelle la *distribution asymptotique* de la moyenne-échantillon. Ils sont formalisés par deux résultats fondamentaux de la statistique, respectivement la *loi des grands nombres* et le *théorème central limite.*

Pour être en mesure d'aborder ces résultats, il convient d'introduire différentes notions de convergence pour les suites de variables aléatoires.

9.2.4.1. Convergence d'une suite de variables aléatoires

A. Convergence d'une suite de nombres

On commence par rappeler la notion convergence pour une suite numérique, non-aléatoire. Considérons la suite numérique définie de la manière suivante :

$$x_n = 1 + \frac{(-1)^n}{n^2}, \quad n = 1, 2, ...$$

Nous pouvons calculer les premières valeurs :

$x_1 = 0$	$x_2 = 1,25$	$x_3 = 0,8889$	$x_4 = 1,0625$
$x_5 = 0,96$	$x_6 = 1,0278$	$x_7 = 0,9796$	$x_8 = 1,0156$
$x_9 = 0,9877$	$x_{10} = 1,01$	$x_{11} = 0,9917$	$x_{12} = 1,0069$

On constate que ces valeurs numériques se rapprochent de plus en plus de 1. On dit que *la suite converge vers la valeur 1* ou que x_n tend vers 1 lorsque n tend vers l'infini ou encore que *la limite de la suite est 1.* La définition suivante donne une définition plus formelle de la convergence d'une suite numérique.

Définition 9.5 *Une suite numérique $x_1, x_2, ..., x_n, ...$ converge vers la valeur a, appelé limite de la suite, si quel que soit $\varepsilon > 0$, il existe un rang N tel que :*

$$|x_n - a| < \varepsilon, \quad \textit{pour tout } n \geq N$$

On note : $\lim_{n\to\infty} x_n = a$.

Pour avoir $\lim_{n\to\infty} x_n = a$, il faut donc que pour n'importe quelle valeur de $\varepsilon > 0$, il existe un rang N à partir duquel les valeurs de la suite se situent toutes dans l'intervalle $]a-\varepsilon, a+\varepsilon[$. En d'autres termes, $\lim_{n\to\infty} x_n = a$ signifie qu'il existe toujours un certain rang N à partir duquel les valeurs de la suite sont toutes aussi proche que l'on veut de a.

B. Convergence en probabilité

Le concept de convergence d'une suite numérique peut être étendu aux suites de variables aléatoires. Un exemple d'une suite de variables aléatoires $X_1, X_2, ..., X_n, ...$ est donné par la suite des moyennes-échantillons $\overline{X}_1, \overline{X}_2, ..., \overline{X}_n, ...$ c'est-à-dire la suite des moyennes-échantillons obtenues d'échantillons aléatoires de plus en plus grand.

Considérons une suite de variables aléatoires $X_1, X_2, ..., X_n, ...$ De façon semblable au cas des suites numériques, on dit que cette suite *converge en probabilité* vers la valeur a si pour tout $\varepsilon > 0$, la probabilité que la variable aléatoire X_n prenne sa valeur dans l'intervalle $]a-\varepsilon, a+\varepsilon[$ tend vers 1 lorsque n tend vers l'infini.

Définition 9.6 *Une suite aléatoire $X_1, X_2, ..., X_n, ...$ converge en probabilité vers la valeur a, appelée limite de la suite, si quel que soit $\varepsilon > 0$:*

$$\lim_{n\to\infty} I\!P(|X_n - a| < \epsilon) = 1$$

On note : $X_n \xrightarrow{P} a$.

En termes plus simple, $X_n \xrightarrow{P} a$ signifie que la probabilité que X_n prenne une valeur aussi proche que l'on veut de a tend vers 1 lorsque n tend vers l'infini.

Illustrons cette notion de convergence par un exemple. Soit $U_1, U_2, ..., U_n, ...$ une suite de variables aléatoires indépendantes, toutes de même loi continue uniforme sur $[0,1]$: $U_n \rightsquigarrow \mathcal{U}(0;1)$, $n = 1, 2....$ A cette suite aléatoire, on peut associer la suite $X_1, X_2, ..., X_n, ...$ où X_n est défini par $X_n = \min\{U_1, U_2, ..., U_n\}$. Nous allons montrer que :

$$X_n \xrightarrow{P} 0$$

Pour tout $\varepsilon > 0$, la probabilité $I\!P(|X_n - 0| < \epsilon) = I\!P(|X_n| < \epsilon)$ est égale à $I\!P(X_n < \epsilon)$ puisque X_n est toujours positive. Par ailleurs, l'événement "$X_n \geq \epsilon$" est équivalent à l'événement "$U_1 \geq \varepsilon$ et $U_2 \geq \varepsilon$ et ... et $U_n \geq \varepsilon$". Toutes les variables aléatoires $U_1, U_2, ..., U_n$ étant indépendantes et de même loi que U avec $I\!P(U \geq \varepsilon) = 1 - \varepsilon$ pour $0 < \varepsilon \leq 1$ et $I\!P(U \geq \varepsilon) = 0$ pour $\varepsilon > 1$, on a :

$$\begin{aligned}
I\!P(|X_n - 0| < \epsilon) &= I\!P(X_n < \epsilon) = 1 - I\!P(X_n \geq \epsilon) \\
&= 1 - I\!P(U_1 \geq \varepsilon \text{ et } U_2 \geq \varepsilon \text{ et ... et } U_n \geq \varepsilon) \\
&= 1 - I\!P(U_1 \geq \varepsilon) I\!P(U_2 \geq \varepsilon) \times ... \times I\!P(U_n \geq \varepsilon) \\
&= \begin{cases} 1 - (1-\varepsilon)^n & \text{pour } 0 < \varepsilon \leq 1 \\ 1 & \text{pour } \varepsilon > 1 \end{cases}
\end{aligned}$$

On constate que quel que soit $\varepsilon > 0$, on a $\lim_{n\to\infty} I\!P(|X_n - 0| < \epsilon) = 1$. On a donc bien : $X_n \xrightarrow{P} 0$.

On peut établir des conditions suffisantes pour la convergence en probabilité, souvent plus facile à vérifier que la définition elle-même.

Propriété 9.2 *Soit une suite aléatoire $X_1, X_2, ..., X_n, ...$. Si on a :*

$$\lim_{n\to\infty} E(X_n) = a \quad et \quad \lim_{n\to\infty} V(X_n) = 0$$

alors :

$$X_n \xrightarrow{P} a$$

Ce résultat est intuitif. Il indique que X_n converge en probabilité vers a si, lorsque n tend vers l'infini, la loi de la variable aléatoire X_n tend à être centrée en a et de variance nulle, c'est-à-dire tend à se concentrer sur un seul point (une seule valeur possible) : a.

La preuve de ce résultat repose sur une propriété connue sous le nom d'*inégalité de* BIENAYMÉ[15] – TCHEBYCHEV[16].

Propriété 9.3 (Inégalité de BIENAYMÉ – TCHEBYCHEV) *Soit X une variable aléatoire quelconque et un nombre $a > 0$. On a :*

$$I\!P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}$$

On démontrera cette inégalité uniquement dans le cas discret. Le cas continu peut être traité de façon semblable. Posons $Y = (X - E(X))^2$. Par construction, toutes les valeurs possibles y_j de Y sont nécessairement supérieurs ou égales à zéro. On a d'une part :

$$I\!P(|X - E(X)| \geq a) = I\!P\left(\sqrt{Y} \geq a\right) = I\!P(Y \geq a^2) = \sum_{y_j \geq a^2} p_{y_j}$$

où $p_{y_j} = I\!P(Y = y_j)$, et :

$$V(X) = E\left[(X - E(X))^2\right] = E(Y)$$

D'autre part, on a également :

$$\begin{aligned} E(Y) &= \sum_{y_j} y_j p_{y_j} = \underbrace{\sum_{y_j < a^2} y_j p_{y_j}}_{\geq 0} + \sum_{y_j \geq a^2} y_j p_{y_j} \\ &\geq \sum_{y_j \geq a^2} y_j p_{y_j} \geq a^2 \sum_{y_j \geq a^2} p_{y_j} = a^2 I\!P(Y \geq a^2) \end{aligned}$$

[15] Irénée-Jules BIENAYMÉ (1796 –1878), mathématicien français.
[16] Panoufty Lvovitch TCHEBYCHEV (1821–1894), mathématicien russe.

et donc :

$$I\!P(Y \geq a^2) \leq \frac{E(Y)}{a^2}$$

On en déduit que :

$$I\!P\left[(X - E(X))^2 \geq a^2\right] \leq \frac{V(X)}{a^2} \quad \Leftrightarrow \quad I\!P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}$$

Notons au passage que l'inégalité de BIENAYMÉ – TCHEBYCHEV est parfois utilisée pour fournir des bornes dans le calcul de certaines probabilités.

Venons-en maintenant à la preuve de la Propriété 9.2. Appliquée la suite aléatoire $X_1, X_2, ..., X_n, ...$ l'inégalité de BIENAYMÉ – TCHEBYCHEV permet d'affirmer que, pour n quelconque et quelque soit $\varepsilon > 0$, on a :

$$I\!P(|X_n - E(X_n)| \geq \varepsilon) \leq \frac{V(X_n)}{\varepsilon^2}$$

Si $\lim_{n\to\infty} E(X_n) = a$ et $\lim_{n\to\infty} V(X_n) = 0$, quel que soit $\varepsilon > 0$, on a donc :

$$\lim_{n\to\infty} I\!P(|X_n - a| \geq \varepsilon) \leq \frac{0}{\varepsilon^2} \quad \Leftrightarrow \quad \lim_{n\to\infty} I\!P(|X_n - a| \geq \varepsilon) = 0$$

soit, en passant à l'événement contraire :

$$\lim_{n\to\infty} I\!P(|X_n - a| < \varepsilon) = 1$$

En d'autres termes, $X_n \xrightarrow{P} a$.

C. Convergence en loi

On peut définir un autre type de convergence pour les suites de variables aléatoires. Celui-ci est basé sur la fonction de répartition des variables : il s'agit de la *convergence en loi*, qu'on appelle aussi *convergence en distribution*.

Considérons une suite de variables aléatoires $X_1, X_2, ..., X_n, ...$. A chacune des variables aléatoires de cette suite correspond une fonction de répartition que l'on note $F_n(x)$. Considérons également une variable aléatoire X dont la fonction de répartition est notée $F(x)$. On dit que la suite de variables aléatoires $X_1, X_2, ..., X_n, ...$ converge en loi vers la loi de la variable aléatoire X si la suite de fonctions de répartition $F_1(x), F_2(x), ..., F_n(x), ...$ converge vers la fonction de répartition $F(x)$.

Définition 9.7 *Une suite aléatoire $X_1, X_2, ..., X_n, ...$ dont les fonctions de répartitions sont $F_1(x), F_2(x), ..., F_n(x), ...$ converge en loi vers la loi d'une variable aléatoire X dont la fonction de répartition est $F(x)$, appelée loi limite de la suite, si :*

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} I\!P(X_n \leq x) = I\!P(X \leq x) = F(x)$$

en tout point de continuité x de $F(x)$. On note : $X_n \xrightarrow{L} X$, ou encore $X_n \xrightarrow{L} \mathcal{L}oi(.)$, *où "$\mathcal{L}oi(.)$" désigne la loi de X (par exemple $\mathcal{L}oi(.) \equiv \mathcal{N}(0;1)$).*

On tolère que la convergence ne soit pas assurée en un certain nombre de points. Ces points correspondent aux points de discontinuité de la fonction $F(x)$. Il s'agit là d'une condition technique sur laquelle nous ne nous étendrons pas.

En termes plus concret, dans le cas de variables aléatoires continues, $X_n \xrightarrow{L} X$ signifie que, lorsque n tend vers l'infini, la probabilité que X_n prenne sa valeur dans un intervalle $[a,b]$ quelconque tend vers la probabilité que X prenne sa valeur dans le même intervalle :

$$\lim_{n\to\infty} \mathbb{P}(a \leq X_n \leq b) = \mathbb{P}(a \leq X \leq b)$$

De la même façon, dans le cas de variables aléatoires discrètes, $X_n \xrightarrow{L} X$ signifie que, lorsque n tend vers l'infini, la probabilité que X_n prenne une valeur a quelconque tend vers la probabilité que X prenne la même valeur :

$$\lim_{n\to\infty} \mathbb{P}(X_n = a) = \mathbb{P}(X = a)$$

D. Approximations de loi

Une petite digression. Comme suggéré lorsque nous les avons présenté, le concept de convergence en loi est à la base des approximations de lois que nous avons vues aux Chapitres 6 et 7 : approximation d'une loi binomiale par une loi de Poisson, approximation d'une loi de Poisson par une loi normale et approximation d'une loi binomiale par une normale. On énonce ci-dessous les résultats de convergence qui sont à la base de ces approximations, plus encore deux autres.

Propriété 9.4 *Soient une variable aléatoire $X \rightsquigarrow \mathcal{P}(\lambda)$ et une suite $X_1, X_2, ...,$ $X_n, ...$ de variables aléatoires telles que :*

$$X_n \rightsquigarrow \mathcal{B}(n;p_n), \quad \text{où } \lim_{n\to\infty} np_n = \lambda < +\infty$$

La suite des variables binomiales converge en loi vers la loi de la variable aléatoire de Poisson X de paramètre λ :

$$X_n \xrightarrow{L} X \quad \Leftrightarrow \quad X_n \xrightarrow{L} \mathcal{P}(\lambda)$$

Grâce à cette propriété, nous pouvons, comme indiqué au Chapitre 6, approcher une loi binomiale $\mathcal{B}(n;p)$ par une loi de Poisson $\mathcal{P}(np)$:

$$\mathcal{B}(n;p) \approx \mathcal{P}(np)$$

En pratique, on considère généralement que cette approximation est bonne lorsque $n \geq 50$, $p \leq 0,1$ et $np \leq 15$.

Propriété 9.5 *Soient une variable aléatoire $Z \rightsquigarrow \mathcal{N}(0;1)$ et une suite X_1, X_2,..., X_n,... de variables aléatoires telles que :*

$$X_n \rightsquigarrow \mathcal{P}(\lambda_n), \quad \textit{où} \lim_{n\to\infty} \lambda_n = +\infty$$

La suite des variables de Poisson *centrées et réduites converge en loi vers la loi de la variable aléatoire normale centrée réduite Z :*

$$\frac{X_n - \lambda_n}{\sqrt{\lambda_n}} \xrightarrow{L} Z \quad \Leftrightarrow \quad \frac{X_n - \lambda_n}{\sqrt{\lambda_n}} \xrightarrow{L} \mathcal{N}(0;1)$$

On dit qu'une loi de Poisson dont le paramètre tend vers l'infini est asymptotiquement normale. Grâce à cette propriété, comme indiqué au Chapitre 7, pour λ suffisamment grand ($\lambda \geq 15$), nous pouvons faire l'approximation :

$$\mathcal{P}(\lambda) \approx \mathcal{N}(\lambda;\lambda)$$

Propriété 9.6 *Soient une variable aléatoire $Z \rightsquigarrow \mathcal{N}(0;1)$ et une suite X_1, X_2, ..., X_n,... de variables aléatoires telles que :*

$$X_n \rightsquigarrow \mathcal{B}(n;p)$$

La suite des variables binomiales centrées et réduites converge en loi vers la loi de la variable aléatoire normale centrée réduite Z :

$$\frac{X_n - np}{\sqrt{npq}} \xrightarrow{L} Z \quad \Leftrightarrow \quad \frac{X_n - np}{\sqrt{npq}} \xrightarrow{L} \mathcal{N}(0;1) \qquad (q = 1-p)$$

On dit qu'une loi binomiale dont le paramètre n tend vers l'infini est asymptotiquement normale. Grâce à cette propriété, à nouveau comme indiqué au Chapitre 7, nous pouvons approcher une loi binomiale $\mathcal{B}(n;p)$ par une loi normale :

$$\mathcal{B}(n;p) \approx \mathcal{N}(np;npq)$$

Dans la pratique, on considère généralement que cette approximation est bonne lorsque $n \geq 30$, $np \geq 5$ et $nq \geq 5$ (même s'il faut parfois être prudent).

Propriété 9.7 *Soient une variable aléatoire $Z \rightsquigarrow \mathcal{N}(0;1)$ et une suite X_1, X_2, ..., X_n,... de variables aléatoires telles que :*

$$X_n \rightsquigarrow t(n)$$

La suite des variables de Student *converge en loi vers la loi de la variable aléatoire normale centrée réduite Z :*

$$X_n \xrightarrow{L} Z \quad \Leftrightarrow \quad X_n \xrightarrow{L} \mathcal{N}(0;1)$$

On dit qu'une loi de Student dont le nombre de degrés de liberté n tend vers l'infini est asymptotiquement normale. Ceci est à mettre en parallèle avec la ressemblance que nous avions remarquée entre la densité de la loi normale et la densité de la loi de STUDENT.

Un résultat similaire concerne la loi du khi-carré.

Propriété 9.8 *Soient une variable aléatoire* $Z \rightsquigarrow \mathcal{N}(0;1)$ *et une suite* $X_1, X_2, ...,$ $X_n, ...$ *de variables aléatoires telles que :*

$$X_n \rightsquigarrow \chi^2(n)$$

La suite des variables du khi-carré centrées et réduites converge en loi vers la loi de la variable aléatoire normale centrée réduite Z *:*

$$\frac{X_n - n}{\sqrt{2n}} \xrightarrow{L} Z \quad \Leftrightarrow \quad \frac{X_n - n}{\sqrt{2n}} \xrightarrow{L} \mathcal{N}(0;1)$$

On dit qu'une loi du khi-carré dont le nombre de degrés de liberté n tend vers l'infini est asymptotiquement normale.

9.2.4.2. Loi des grands nombres

On revient à l'examen du comportement asymptotique — en grand échantillon — de la moyenne-échantillon $\overline{X}_n$.

La suite des moyennes-échantillons $\overline{X}_1, \overline{X}_2, ..., \overline{X}_n, ...$ obtenues d'échantillons aléatoires de plus en plus grand est une suite de variables aléatoires. On a vu (Propriété 9.1) que, quels que soient n et la loi de la population représentée par la loi de la variable aléatoire X, on a :

$$E(\overline{X}_n) = m \quad \text{et} \quad V(\overline{X}_n) = \frac{\sigma^2}{n}$$

En faisant tendre n vers l'infini, on obtient :

$$\lim_{n \to \infty} E(\overline{X}_n) = m \quad \text{et} \quad \lim_{n \to \infty} V(\overline{X}_n) = 0$$

Par application de la Propriété 9.2, on a donc :

$$\overline{X}_n \xrightarrow{P} m$$

Nous venons d'établir la loi des grands nombres.

Propriété 9.9 (Loi des grands nombres) *Soit* $(X_1, X_2, ..., X_n)$ *un échantillon aléatoire relatif à une variable aléatoire* X *d'espérance* m *et de variance* σ^2*. On a :*

$$\overline{X}_n \xrightarrow{P} m$$

On dit que la moyenne-échantillon $\overline{X}_n$ converge en probabilité vers la moyenne $m = E(X)$ dans la population. Autrement dit, lorsque n tend vers l'infini, la probabilité que la moyenne-échantillon $\overline{X}_n$ prenne une valeur aussi proche que l'on veut de la moyenne $m = E(X)$ dans la population tend vers 1. Notez bien que cette propriété est vraie quelle que soit la loi de la variable aléatoire X.

9.2.4.3. Théorème central limite

La loi des grands nombres formalise l'idée que, quel que soit la loi de la population, lorsque la taille d'échantillon tend vers l'infini, la distribution d'échantillonnage de $\overline{X}_n$ tend à se concentrer sur le seul point $m = E(X)$.

On a vu que, tout en se concentrant, la distribution d'échantillonnage de $\overline{X}_n$ avait tendance à se rapprocher de la forme en cloche typique de la loi normale. Cette idée est formalisée par le théorème central limite.

Considérons la forme standardisée — centrée et réduite — de la moyenne-échantillon $\overline{X}_n$:

$$Y_n = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{V(\overline{X}_n)}} = \frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}}$$

Au contraire de la loi de $\overline{X}_n$, la loi de Y_n ne tend pas à se concentrer sur un seul point lorsque n tend vers l'infini puisque, par construction, quel que soit n, $E(Y_n) = 0$ et $V(Y_n) = 1$.

Le théorème central limite, que nous admettrons sans preuve, assure que, lorsque n tend vers l'infini, la loi de Y_n tend vers une loi normale standardisée, en d'autres termes que la version standardisée de $\overline{X}_n$ converge en loi vers une loi normale standardisée.

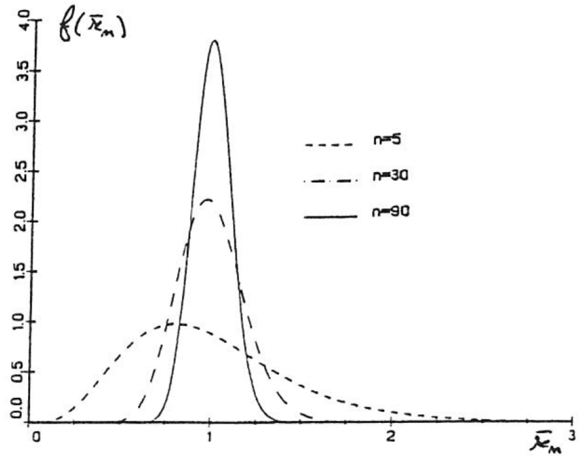
Propriété 9.10 (Théorème central limite) *Soit* $(X_1, X_2, ..., X_n)$ *un échantillon aléatoire relatif à une variable aléatoire* X *d'espérance* m *et de variance* σ^2*. On a :*

$$\frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \xrightarrow{L} \mathcal{N}(0; 1)$$

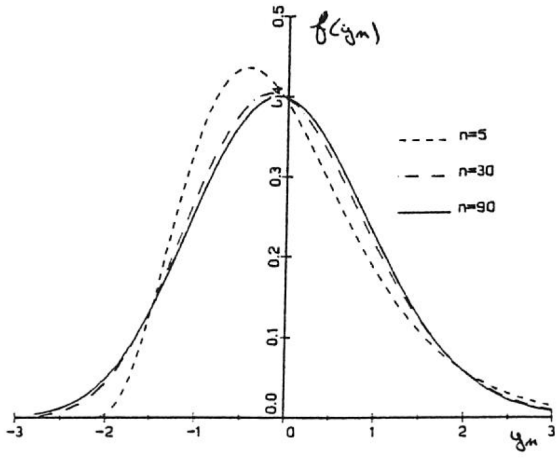
Remarquez à nouveau bien que, comme la loi des grands nombres, ce résultat est vrai quelle que soit la loi de la variable aléatoire X.

Le Graphique 9.2. illustre la loi des grands nombres et le théorème central limite pour une population représentée par une loi exponentielle de paramètre $\lambda = 1$: $X \rightsquigarrow \mathcal{E}(1)$. Pour une telle loi, on a $E(X) = V(X) = 1$, et donc $m = \sigma^2 = 1$.

Le graphique de gauche représente la distribution d'échantillonnage exacte de $\overline{X}_n$ pour n égal à 5, 30 et 90. On voit clairement que, lorsque n s'accroît, la loi de $\overline{X}_n$ se concentre de plus en plus autour de sa valeur espérée 1 et se rapproche de la forme d'une loi normale. Le graphique de droite représente la distribution d'échantillonnage exacte de la version standardisée de $\overline{X}_n$. Il montre clairement la convergence de cette loi vers la loi normale standardisée.



Loi de $\overline{X}_n$



Loi de $Y_n = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{V(\overline{X}_n)}}$

Graphique 9.2 : Lois exactes de $\overline{X}_n$ et $Y_n = \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{V(\overline{X}_n)}}$ pour $X \rightsquigarrow \mathcal{E}(1)$

Le théorème central limite justifie que l'on puisse, lorsque la taille d'échantillon n est suffisamment grande, approcher la distribution d'échantillonnage exacte de $\overline{X}_n$ par l'approximation normale :

$$\overline{X}_n \approx \mathcal{N}(m; \tfrac{\sigma^2}{n})$$

A l'aide de cette approximation, on peut, en grand échantillon et quel que soit la loi de la population, calculer la probabilité que la moyenne-échantillon $\overline{X}_n$ prenne une valeur appartenant à un intervalle quelconque. Il est clair que le même calcul peut, au moins en théorie, être effectué sans approximation sur base de la distribution d'échantillonnage exacte de $\overline{X}_n$. Pour cela, il faut cependant connaître cette distribution. On a vu que son calcul est le plus souvent très malaisé. Il est même impossible lorsqu'on ne connaît pas la loi de la population, mais seulement certaines de ces caractéristiques (sa moyenne et sa variance, cf. l'exercice résolu de la Section 9.2.3 et l'exercice résolu ci-dessous). Le théorème central limite offre un moyen commode de contourner ces problèmes. Il n'est cependant applicable que si la taille d'échantillon n est suffisamment grande.

Que faut-il entendre par "n assez grand" ou "en grand échantillon" ? Comme pour toute approximation, tout dépend du problème traité ainsi que du niveau d'exigence pour la précision des résultats. En pratique, l'approximation ne devrait jamais être faite pour un échantillon dont la taille est inférieure à 30. Au delà de cette taille, elle devient généralement assez rapidement très bonne (pour vous en convaincre, réexaminez les Graphiques 9.1 et 9.2).

9.2.5. Exercice résolu

Des tubes fluorescents fabriqués par une entreprise ont une durée de vie moyenne de 800 heures. L'écart-type de la durée de vie est évalué à 60. On prélève un échantillon aléatoire de 50 tubes dans la production du mois écoulé et on mesure la durée de vie des tubes. Quelle est la probabilité d'obtenir une durée de vie moyenne comprise entre 790 et 810 heures ?

Notons D la durée de vie d'un tube. La loi de D (= la loi de la population) est inconnue. Nous savons toutefois que cette loi est telle que $m = E(D) = 800$ et que $\sigma = \sigma(D) = 60$.

Un échantillon aléatoire de taille $n = 50$ est décrit par le n-uple $(D_1, D_2, ..., D_{50})$. La moyenne-échantillon $\overline{D}_{50}$ vaut :

$$\overline{D}_{50} = \frac{D_1 + D_2 + \cdots + D_{50}}{50}$$

De la propriété 9.1, on a :

$$E(\overline{D}_{50}) = 800 \quad \text{et} \quad V(\overline{D}_{50}) = \frac{60^2}{50} = 72$$

Le théorème central limite assure que :

$$\overline{D}_{50} \approx \mathcal{N}(800; 72) \quad \Leftrightarrow \quad Z = \frac{\overline{D}_{50} - 800}{\sqrt{72}} \approx \mathcal{N}(0,1)$$

Sur base de cette approximation, on trouve pour la probabilité demandée :

$$\begin{aligned}
I\!\!P(790 < \overline{D}_{50} < 810) &= I\!\!P(790 \leq \overline{D}_{50} \leq 810) \\
&= I\!\!P\left(\frac{790-800}{\sqrt{72}} \leq \frac{\overline{D}_{50} - 800}{\sqrt{72}} \leq \frac{810-800}{\sqrt{72}}\right) \\
&= I\!\!P(-1,18 \leq Z \leq 1,18) \\
&= 2\Phi(1,18) - 1 = 2 \times 0,8810 - 1 = 0,762
\end{aligned}$$

9.3. Variance dans la population et variance-échantillon

On examine maintenant les liens existant entre variance des valeurs prises par une variable d'intérêt dans une population et variance des valeurs prises par cette variable dans un échantillon aléatoire issu de cette population.

9.3.1. La variance-échantillon Σ_n^2

Reprenons notre exemple consacré à la consommation de yaourt. La moyenne dans la population du nombre de yaourts consommés par semaine est identifiée par le paramètre $m = E(X)$ de la loi $\ell(x)$ de la variable aléatoire X représentant la population. La moyenne est une mesure de tendance centrale.

Supposons qu'on s'intéresse également à la variabilité dans la population de la consommation de yaourts. Cette variabilité peut être mesurée par la variance dans la population du nombre de yaourts consommés par semaine. Cette variance est identifiée par le paramètre $\sigma^2 = V(X)$ de la loi de X.

Supposons qu'on tire un échantillon aléatoire de taille n dans la population et qu'on calcule la variance du nombre de yaourts consommés par les individus repris dans l'échantillon. De façon semblable au cas de la moyenne, on peut représenter la forme et le résultat — qui forcément varie d'un échantillon à l'autre — de ce calcul par une statistique que, dans le cas présent et de manière assez naturelle, on appelle la *variance-échantillon.*

Définition 9.8 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X. On appelle variance-échantillon la statistique :*

$$\Sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

Intuitivement, on s'attend à ce que la variance-échantillon Σ_n^2 reflète la variance $\sigma^2 = V(X)$ dans la population. Comme dans le cas de la moyenne-échantillon, on peut de façon formelle établir un lien entre la variance σ^2 dans la population, qui est une valeur fixe, non aléatoire, la variance-échantillon Σ_n^2, qui est une variable aléatoire, dont la valeur varie d'un échantillon à l'autre, au travers de la loi — la distribution d'échantillonnage — de Σ_n^2.

9.3.2. Distribution d'échantillonnage de Σ_n^2

Comme celle de $\overline{X}_n$, la distribution d'échantillonnage de Σ_n^2 dépend d'une part de la loi $\ell(x)$ de X et d'autre part de la taille n de l'échantillon. Elle peut être obtenue, comme la loi de toute fonction de variables aléatoires, de façon semblable à celle de $\overline{X}_n$.

De même que dans le cas de la moyenne-échantillon, hormis quelques cas spécifiques, le calcul de la distribution d'échantillonnage exacte de Σ_n^2 est le plus souvent très malaisé. Son espérance et sa variance peuvent cependant être calculées sans trop de difficultés.

On a :

$$\begin{aligned}
\Sigma_n^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}_n\right)^2=\frac{1}{n}\sum_{i=1}^{n}\left[(X_i-m)-\left(\overline{X}_n-m\right)\right]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2-2\left(\overline{X}_n-m\right)\underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i-m)}_{=\overline{X}_n-m}+\left(\overline{X}_n-m\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2-\left(\overline{X}_n-m\right)^2
\end{aligned}$$

et donc, en utilisant la Propriété 5.15 :

$$\begin{aligned}
E\left(\Sigma_n^2\right) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2-\left(\overline{X}_n-m\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}E\left[(X_i-m)^2\right]-E\left[\left(\overline{X}_n-m\right)^2\right]
\end{aligned}$$

Comme $E\left[(X_i-m)^2\right]=E\left[(X_i-E(X_i))^2\right]=V(X_i)=\sigma^2\ (i=1,...,n)$ et $E\left[\left(\overline{X}_n-m\right)^2\right]=E\left[\left(\overline{X}_n-E(\overline{X}_n)\right)^2\right]=V(\overline{X}_n)=\sigma^2/n$, on trouve finalement :

$$E\left(\Sigma_n^2\right)=\frac{1}{n}\sum_{i=1}^{n}V(X_i)-V(\overline{X}_n)=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2$$

Le calcul de la variance de Σ_n^2 est plus difficile. Nous l'admettrons sans preuve.

Propriété 9.11 *Soit $(X_1,X_2,...,X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. On a :*

$$E(\Sigma_n^2)=\frac{n-1}{n}\sigma^2 \quad \textit{et} \quad V(\Sigma_n^2)=\frac{(n-1)}{n^3}\left[(n-1)\mu_4-(n-3)\sigma^4\right]$$

où $\mu_4=E\left[(X-m)^4\right]$.

Notez bien que cette propriété est vraie quels que soient n et la loi de la variable aléatoire X.

On constate que, contrairement au cas de la moyenne-échantillon, l'espérance de la variance-échantillon — la valeur espérée de Σ_n^2, sa tendance centrale — n'est pas égale à la variance $\sigma^2=V(X)$ dans la population : elle est plus petite que σ^2. Pour une taille d'échantillon n grande, elle en est toutefois très proche.

L'expression de la variance de Σ_n^2 est assez complexe. On retiendra simplement que cette variance dépend des moments d'ordre 2 et 4 (respectivement σ^2 et μ_4) de la loi de la population et qu'elle diminue lorsque la taille d'échantillon n s'accroît.

Joint à l'observation que $E(\Sigma_n^2)$ se rapproche de σ^2 lorsque n grandit, ce dernier élément indique que la statistique Σ_n^2 tend à prendre des valeurs d'autant plus proche de la variance $\sigma^2 = V(X)$ dans la population que la taille d'échantillon n est grande.

9.3.3. La variance-échantillon modifiée S_n^2

Nous venons de voir que l'espérance de la variance-échantillon ne correspondait pas parfaitement à la variance dans la population : la distribution de Σ_n^2 n'est pas exactement centrée sur σ^2. C'est pourquoi on considère généralement une version modifiée (ou corrigée) de la statistique Σ_n^2, dont l'espérance est elle exactement égale à σ^2. Cette statistique est appelée la *variance-échantillon modifiée* (ou *corrigée*). Elle représente le résultat de l'application d'une formule quelque peu modifiée pour calculer la variance dans un échantillon.

Définition 9.9 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X. On appelle variance-échantillon modifiée (ou corrigée) la statistique :*

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n\right)^2$$

La statistique S_n^2 est très proche de la statistique Σ_n^2. Elle est liée à Σ_n^2 par la relation :

$$S_n^2 = \frac{n}{n-1} \Sigma_n^2 \tag{9.1}$$

9.3.4. Distribution d'échantillonnage de S_n^2

Comme Σ_n^2, S_n^2 possède une certaine distribution d'échantillonnage qui dépend de la loi $\ell(x)$ de X et de la taille d'échantillon n. Etant donné la relation ci-dessus, on la devine proche — mais malheureusement guère plus facile à calculer de façon exacte — de celle de Σ_n^2.

Les deux premiers moments de la distribution d'échantillonnage exacte de S_n^2 se déduisent aisément de ceux de Σ_n^2. En utilisant la relation (9.1) et les Propriétés 4.10 et 4.11, on trouve :

$$E(S_n^2) = E\left(\frac{n}{n-1}\Sigma_n^2\right) = \frac{n}{n-1}E(\Sigma_n^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$$

et

$$V(S_n^2) = V\left(\frac{n}{n-1}\Sigma_n^2\right) = \left(\frac{n}{n-1}\right)^2 V(\Sigma_n^2)$$

$$= \left(\frac{n}{n-1}\right)^2 \frac{(n-1)}{n^3}\left[(n-1)\mu_4 - (n-3)\sigma^4\right]$$
$$= \frac{1}{n(n-1)}\left[(n-1)\mu_4 - (n-3)\sigma^4\right]$$

Propriété 9.12 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. On a :*

$$E(S_n^2) = \sigma^2 \quad et \quad V(S_n^2) = \frac{1}{n(n-1)}\left[(n-1)\mu_4 - (n-3)\sigma^4\right]$$

où $\mu_4 = E\left[(X-m)^4\right]$.

Comme la Propriété 9.11, la Propriété 9.12 est vraie quels que soient n et la loi de la variable aléatoire X.

On voit que, contrairement à celle de Σ_n^2, l'espérance de la variance-échantillon modifiée — la valeur espérée de S_n^2, sa tendance centrale — est elle exactement égale à la variance $\sigma^2 = V(X)$ dans la population.

L'expression de la variance de S_n^2 est très proche de celle de Σ_n^2. Comme cette dernière, elle dépend des moments d'ordre 2 et 4 (σ^2 et μ_4) de la loi de la population et elle diminue lorsque la taille d'échantillon n s'accroît, ce qui assure que, comme Σ_n^2, la statistique S_n^2 tend à prendre des valeurs d'autant plus proche — et cette fois, quel que soit n, toujours exactement centrée sur σ^2 — de la variance $\sigma^2 = V(X)$ dans la population que la taille d'échantillon n est grande.

Comme $E(S_n^2) = \sigma^2$ et $E(\Sigma_n^2) = \frac{n-1}{n}\sigma^2 < \sigma^2$, on pourrait être tenté de croire que les valeurs prises par la statistique S_n^2 reflète mieux la variance $\sigma^2 = V(X)$ dans la population que celle prise par la statistique Σ_n^2. C'est sans compter que, quel que soit $n = 2, 3...$ fini, on a :

$$V(S_n^2) > V(\Sigma_n^2)$$

Autrement dit, ce que la statistique S_n^2 "gagne" en termes d'espérance, elle le "paye" en termes de variance. On ne peut donc pas dire que les valeurs prises par S_n^2 reflète mieux la variance $\sigma^2 = V(X)$ dans la population que celles prises par Σ_n^2 : elles la reflètent d'une façon différente.

9.3.5. Comportement asymptotique de Σ_n^2 et de S_n^2

Comme dans le cas de $\overline{X}_n$, en grand échantillon, ou plus précisément lorsque la taille d'échantillon n tend vers l'infini, les distributions d'échantillonnage de Σ_n^2 et de S_n^2 peuvent être caractérisées indépendamment de la loi de la population : elles ne dépendent plus de la loi $\ell(x)$ de X.

Les comportements asymptotiques de Σ_n^2 et de S_n^2 sont semblables à celui de la moyenne-échantillon $\overline{X}_n$. Quelle que soit la loi de X, à mesure que n grandit :

1- les distributions d'échantillonnage de Σ_n^2 et de S_n^2 sont de plus en plus concentrées autour de leur analogue dans la population : la variance $\sigma^2 = V(X)$. Sur base des résultats précédemment obtenus, on vérifie en effet aisément qu'on a :

$$\lim_{n\to\infty} E(\Sigma_n^2) = \lim_{n\to\infty} E(S_n^2) = \sigma^2 \quad \text{et} \quad \lim_{n\to\infty} V(\Sigma_n^2) = \lim_{n\to\infty} V(S_n^2) = 0 \tag{9.2}$$

2- tout en se concentrant toujours davantage, les distributions d'échantillonnage de Σ_n^2 de S_n^2 prennent une forme qui se rapproche de plus en plus de la forme en cloche typique de la loi normale.

Le première de ces deux caractéristiques est formalisée par la propriété suivante.

Propriété 9.13 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. On a :*

$$\Sigma_n^2 \xrightarrow{P} \sigma^2 \quad \textit{et} \quad S_n^2 \xrightarrow{P} \sigma^2$$

Ce résultat s'obtient de (9.2) par application de la Propriété 9.2. Il nous indique que, lorsque n tend vers l'infini, la probabilité que tant Σ_n^2 que S_n^2 prenne une valeur aussi proche que l'on veut de la variance $\sigma^2 = V(X)$ dans la population tend vers 1. Notez bien qu'il est vrai quelle que soit la loi de la variable aléatoire X.

La seconde des deux caractéristiques mentionnées est formalisée par la propriété ci-dessous, que nous admettrons sans preuve.

Propriété 9.14 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. On a :*

$$\frac{\Sigma_n^2 - E(\Sigma_n^2)}{\sqrt{V(\Sigma_n^2)}} \xrightarrow{L} \mathcal{N}(0;1) \quad \textit{et} \quad \frac{S_n^2 - E(S_n^2)}{\sqrt{V(S_n^2)}} \xrightarrow{L} \mathcal{N}(0;1)$$

Cette propriété — qui est vraie quelle que soit la loi de X — justifie qu'on puisse, lorsque la taille d'échantillon n est suffisamment grande, approcher les distributions d'échantillonnage exactes de Σ_n^2 et de S_n^2 par :

$$\Sigma_n^2 \approx \mathcal{N}\left(E(\Sigma_n^2); V(\Sigma_n^2)\right) \quad \text{et} \quad S_n^2 \approx \mathcal{N}\left(E(S_n^2); V(S_n^2)\right)$$

On sait que $\lim_{n\to\infty} E(\Sigma_n^2) = E(S_n^2) = \sigma^2$. On peut par ailleurs facilement vérifier que :

$$\lim_{n\to\infty} \frac{V(\Sigma_n^2)}{V(S_n^2)} = 1$$

Autrement dit, les variances de Σ_n^2 et de S_n^2 tendent (assez rapidement) à s'égaliser lorsque n s'accroît. On en déduit que les distributions asymptotiques de Σ_n^2 et de S_n^2 sont identiques.

On voit que, en grand échantillon, contrairement au cas où n est faible, rien ou presque ne distingue les statistiques Σ_n^2 et S_n^2. Cela est bien entendu lié au fait que $S_n^2 = \frac{n}{n-1}\Sigma_n^2$ et que $\lim\limits_{n\to\infty} \frac{n}{n-1} = 1$.

9.4. Fréquence dans la population et fréquence-échantillon

On examine à présent les liens existant entre fréquence d'une caractéristique individuelle dans une population et fréquence de cette caractéristique individuelle dans un échantillon aléatoire issu de cette population.

9.4.1. La fréquence-échantillon F_n

Revenons à nouveau à notre exemple de la consommation de yaourt. Dans la population, un certain nombre d'individus ne consomment jamais de yaourt. Les autres en consomment occasionnellement ou régulièrement. Supposons qu'on s'intéresse à la fréquence — la proportion — dans la population des individus qui ne consomment jamais de yaourt. Notons cette fréquence p.

Supposons qu'on tire un individu au hasard dans la population et qu'on représente le résultat de ce tirage par une variable aléatoire de BERNOULLI X qui prend la valeur 1 si l'individu tiré ne consomme jamais de yaourt et la valeur 0 sinon. Clairement, "$X = 1$" se produira avec une probabilité p et "$X = 0$" se produira avec une probabilité $q = 1 - p$. X suit donc une loi de BERNOULLI de paramètre p: $X \rightsquigarrow \mathcal{B}(p)$. Nous venons, au travers de l'épreuve aléatoire "tirer au hasard un individu dans la population et observer si cet individu consomme ou non du yaourt" de décrire la population concernant la variable qualitative binaire "ne pas consommer / consommer du yaourt" par une variable aléatoire X suivant une loi de BERNOULLI $\mathcal{B}(p)$, où le paramètre p identifie la fréquence des individus qui ne consomment jamais de yaourt dans la population.

De façon générale, toute population composée d'une certaine proportion d'individus possédant une caractéristique donnée — "être un homme", "'peser plus de 80 kg", pour une pièce mécanique, "être défectueuse", etc..., dans notre exemple, "ne pas consommer de yaourt" — peut ainsi être représentée par une variable aléatoire X suivant une loi de BERNOULLI $\mathcal{B}(p)$, où $p = \mathbb{P}(X = 1)$ correspond à la fréquence dans la population des individus qui possèdent la caractéristique en question et $q = 1 - p = \mathbb{P}(X = 0)$ correspond à la fréquence dans la population des individus qui ne possèdent pas la caractéristique en question.

Poursuivons à notre exemple. Supposons qu'on prélève un échantillon aléatoire de taille n dans la population et qu'on calcule la fréquence dans l'échantillon des individus qui ne consomment jamais de yaourt. De façon semblable aux cas de la moyenne et de la variance, on peut représenter la forme et le résultat — qui forcément varie d'un échantillon à l'autre — de ce calcul par une statistique que, dans le cas présent et de manière assez naturelle, on appelle la *fréquence-échantillon.*

Définition 9.10 *Soit* $(X_1, X_2, ..., X_n)$ *un échantillon aléatoire relatif à une variable aléatoire* $X \rightsquigarrow \mathcal{B}(p)$. *On appelle fréquence-échantillon la statistique :*

$$F_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Comme la loi qui représente la population est une loi de Bernoulli $\mathcal{B}(p)$, chacune des variables aléatoires X_i de l'échantillon aléatoire $(X_1, X_2, ..., X_n)$ ne peut prendre que la valeur 0 ou 1 — 1 si le i-ème individu tiré possède la caractéristique étudiée, 0 sinon —, de sorte que F_n donne bien la fréquence — le nombre de 1 parmi les n individus tirés — dans l'échantillon des individus qui possèdent la caractéristique en question.

Intuitivement, on s'attend évidemment à ce que la fréquence-échantillon F_n reflète la fréquence p dans la population. Comme dans le cas de la moyenne-échantillon et de la variance-échantillon, on peut à nouveau de façon formelle établir un lien entre la fréquence p dans la population, qui est une valeur fixe, non aléatoire, la fréquence-échantillon F_n, qui est une variable aléatoire, dont la valeur varie d'un échantillon à l'autre, au travers de la distribution d'échantillonnage — la loi — de F_n.

9.4.2. Distribution d'échantillonnage de F_n

La distribution d'échantillonnage exacte de F_n est facile à obtenir. Par construction, les variables aléatoires X_i $(i = 1, ..., n)$ de l'échantillon aléatoire forment un processus de Bernoulli : elles sont indépendantes et suivent toutes une même loi de Bernoulli $\mathcal{B}(p)$. Par conséquent (cf. Chapitre 6), leur somme — notons-la Y — suit une loi binomiale de paramètre n et p :

$$Y = X_1 + X_2 + \cdots + X_n \rightsquigarrow \mathcal{B}(n; p)$$

La variable aléatoire Y représente le nombre d'individus de l'échantillon aléatoire qui possèdent la caractéristique étudiée, dans notre exemple le nombre d'individus de l'échantillon qui ne consomment jamais de yaourt.

L'ensemble $\mathcal{Y}$ des $n + 1$ valeurs possibles de Y est $\mathcal{Y} = \{0, 1, 2, ..., n - 1, n\}$ et les probabilités de ces valeurs possibles sont données par :

$$I\!P(Y = y) = C_n^y p^y (1 - p)^{n-y}, \quad \forall\, y \in \mathcal{Y}$$

La fréquence-échantillon est égale à Y divisé par la taille de l'échantillon :

$$F_n = \frac{Y}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

On en déduit que l'ensemble $\mathcal{F}$ des $n+1$ valeurs possibles de F_n est $\mathcal{F} = \left\{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\right\}$ et que les probabilités de ces valeurs possibles sont données par :

$$I\!P(F_n = f) = I\!P(\frac{Y}{n} = f) = I\!P(Y = nf) = C_n^{nf} p^{nf}(1-p)^{n-nf}, \quad \forall f \in \mathcal{F}$$

On voit que la distribution d'échantillonnage exacte de F_n est une loi discrète de type binomiale.

Propriété 9.15 *Soit* $(X_1, X_2, ..., X_n)$ *un échantillon aléatoire relatif à une variable aléatoire* $X \rightsquigarrow \mathcal{B}(p)$. *On a :*

$$\mathcal{F} = \left\{0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\right\}$$

et

$$I\!P(F_n = f) = C_n^{nf} p^{nf}(1-p)^{n-nf}, \quad \forall f \in \mathcal{F}$$

Les deux premiers moments de cette distribution sont donnés par la propriété suivante.

Propriété 9.16 *Soit* $(X_1, X_2, ..., X_n)$ *un échantillon aléatoire relatif à une variable aléatoire* $X \rightsquigarrow \mathcal{B}(p)$. *On a :*

$$E(F_n) = p \quad \textit{et} \quad V(F_n) = \frac{pq}{n} = \frac{p(1-p)}{n}$$

En effet, $F_n = Y/n$. Puisque $Y = X_1 + X_2 + \cdots + X_n \rightsquigarrow \mathcal{B}(n;p)$, nous avons $E(Y) = np$ et $V(Y) = npq$ (cf. Propriété 6.2). En utilisant les Propriétés 4.10 et 4.11, on en déduit :

$$E(F_n) = E\left(\frac{Y}{n}\right) = \frac{1}{n}E(Y) = \frac{1}{n}np = p$$

et

$$V(F_n) = V\left(\frac{Y}{n}\right) = \frac{1}{n^2}V(Y) = \frac{1}{n^2}npq = \frac{pq}{n}$$

On constate que l'espérance de la fréquence-échantillon — la valeur espérée de F_n, sa tendance centrale — est égale à la fréquence p dans la population et que sa variance est d'autant plus petite que la taille n de l'échantillon est grande et, de façon plus marginale (à moins que p soit très proche de 0 ou de 1), que la fréquence p dans la population est différente de $0,5$.

Autrement dit, la fréquence-échantillon F_n prend des valeurs centrées sur la fréquence p dans la population, et ces valeurs tendent à être d'autant plus proche de la fréquence p dans la population que la taille n de l'échantillon est grande et, de façon plus marginale, que cette fréquence p est différente de $0,5$.

On aurait pu établir directement, sans passer par sa loi exacte, l'espérance et la variance de F_n. Comme l'aura sans doute noté le lecteur attentif, la fréquence-échantillon est une statistique qui a exactement la même forme que la moyenne-échantillon. Comme la statistique $\overline{X}_n$, la statistique F_n n'est effet rien d'autre que la moyenne d'un ensemble de n variables aléatoires indépendantes et qui suivent toutes une même loi. On se trouve simplement dans le cas particulier où cette loi, la loi de la population, est une loi de BERNOULLI $\mathcal{B}(p)$. La fréquence-échantillon n'étant qu'un cas particulier de la moyenne-échantillon où la loi de X est une loi de BERNOULLI $\mathcal{B}(p)$, on peut lui appliquer toutes les propriétés de la moyenne-échantillon, en particulier la Propriété 9.1. Par application de cette propriété — qui est vraie quels que soient n et la loi de X —, on a :

$$E(F_n) = E(X) \quad \text{et} \quad V(F_n) = \frac{V(X)}{n}, \quad \text{où } X \rightsquigarrow \mathcal{B}(p)$$

On sait (cf. Propriété 6.1) qu'une variable aléatoire X distribuée selon une loi de BERNOULLI $\mathcal{B}(p)$ a une espérance égale à p et une variance égale à $pq = p(1-p)$. Par conséquent :

$$E(F_n) = p \quad \text{et} \quad V(F_n) = \frac{pq}{n} = \frac{p(1-p)}{n}$$

9.4.3. Comportement asymptotique de F_n

La fréquence-échantillon F_n n'étant qu'un cas particulier de la moyenne-échantillon $\overline{X}_n$ où la loi de X est une loi de BERNOULLI $\mathcal{B}(p)$ — loi qui est telle que $E(X) = p$ et $V(X) = pq = p(1-p)$ —, le comportement asymptotique de la statistique F_n est identique à celui de la moyenne-échantillon. Il découle directement de la loi des grands nombres et du théorème central limite, résultats qui, rappelons-le, sont vrais quelle que soit la loi de la population.

Par application de la loi des grands nombres, on a la propriété suivante.

Propriété 9.17 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{B}(p)$. On a :*

$$F_n \xrightarrow{P} p$$

Ce résultat est généralement appelé la *loi des grands nombres de* BERNOULLI. Il nous garantit que, lorsque la taille d'échantillon n tend vers l'infini, la probabilité que la fréquence-échantillon F_n soit aussi proche qu'on veut de la fréquence $p = I\!\!P(X = 1) = E(X)$ dans la population tend vers 1. Il formalise l'idée que la distribution d'échantillonnage de F_n tend à se concentrer sur p lorsque n tend vers l'infini : on vérifie en effet aisément que $\lim\limits_{n\to\infty} E(F_n) = p$ et $\lim\limits_{n\to\infty} V(F_n) = 0$ (ce qui on le sait — cf. Propriété 9.2 — implique $F_n \xrightarrow{P} p$).

Par application du théorème central limite, on a d'autre part la propriété suivante.

Propriété 9.18 *Soit* $(X_1, X_2, ..., X_n)$ *un échantillon aléatoire relatif à une variable aléatoire* $X \rightsquigarrow \mathcal{B}(p)$*. On a :*

$$\frac{F_n - p}{\sqrt{\frac{pq}{n}}} \xrightarrow{L} \mathcal{N}(0;1)$$

Cette résultat formalise l'idée qu'à mesure que n grandit, la distribution d'échantillonnage de F_n prend une forme qui se rapproche de plus en plus de la forme en cloche typique de la loi normale. Il justifie qu'on puisse, lorsque la taille d'échantillon n est suffisamment grande, approcher la distribution d'échantillonnage exacte de F_n par l'approximation normale :

$$F_n \approx \mathcal{N}(p; \tfrac{pq}{n})$$

Pourquoi utiliser une approximation lorsque alors que la distribution d'échantillonnage exacte de F_n est connue ? Tout simplement parce que l'approximation est bien plus facile à manier que la distribution exacte. Dans le cadre de ce cours, nous admettrons que cette approximation est raisonnable si n et p sont tels que : $n \geq 30$, $np \geq 5$ et $nq \geq 5$.

9.4.4. Exercice résolu

Dans une production de plusieurs millions de circuits électroniques intégrés, 2% seulement sont défectueux. Quelle est la probabilité que, sur 1 000 circuits extraits au hasard de la chaîne d'assemblage, plus de 10% soient défectueux ?

Notons X la variable aléatoire de Bernoulli qui prend la valeur 1 si un élément est défectueux et la valeur 0 sinon. La population des circuits électroniques intégrés est composée d'une proportion $p = 0,02$ d'éléments défectueux. La loi de la population est donc : $X \rightsquigarrow \mathcal{B}(0,02)$.

Un échantillon aléatoire de taille $n = 1000$ est décrit par le n-uple $(X_1, X_2, ..., X_{1000})$. La fréquence-échantillon F_{1000} vaut :

$$F_{1000} = \frac{X_1 + X_2 + \cdots + X_{1000}}{1000}$$

On a $n = 1000 \geq 30$, $np = 20 \geq 5$ et $nq = 980 \geq 5$. On est donc dans les conditions d'applicabilité de la Propriété 9.18. Cette propriété nous assure que :

$$F_{1000} \approx \mathcal{N}(0,02; \tfrac{0,02 \times 0,98}{1000}) \quad \Leftrightarrow \quad Z = \frac{F_{1000} - 0,02}{0,00443} \approx \mathcal{N}(0,1)$$

Sur base de cette approximation, on trouve pour la probabilité demandée :

$$\begin{aligned} \mathbb{P}(F_{1000} > 0,1) &= \mathbb{P}(F_{1000} \geq 0,1) \\ &= \mathbb{P}\left(\frac{F_{1000} - 0,02}{0,00443} \geq \frac{0,1 - 0,02}{0,00443}\right) \\ &= \mathbb{P}(Z \geq 18,05) = 1 - \Phi(18,05) \approx 0 \end{aligned}$$

9.5. Distribution d'échantillonnage de $\overline{X}_n$, Σ_n^2 et S_n^2 sous l'hypothèse de normalité

On termine ce chapitre par quelques résultats complémentaires relatifs au cas des populations normales.

Dans le cas où la loi de la population est normale, c'est-à-dire lorsque $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$, les distributions d'échantillonnage exactes de $\overline{X}_n$, Σ_n^2 et S_n^2 sont bien connues. On les détaille ci-dessous.

Propriété 9.19 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$. On a :*

$$\overline{X}_n \rightsquigarrow \mathcal{N}(m; \tfrac{\sigma^2}{n}) \quad \Leftrightarrow \quad \frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \rightsquigarrow \mathcal{N}(0;1)$$

En effet, lorsque $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est une combinaison linéaire de variables aléatoires normales indépendantes. Or, de la Propriété 7.7, une combinaison linéaire de variables aléatoires normales indépendantes est encore une variable aléatoire normale. L'espérance et la variance de cette loi découle de la Propriété 9.1.

Malgré leur ressemblance, il faut éviter toute confusion entre la Propriété 9.19 et le théorème central limite. Le théorème central limite s'applique quelle quoi soit la loi de X. Il donne une propriété asymptotique. La Propriété 9.19 ne s'applique qu'au cas où X est normale et fournit un résultat exact, valable quelle que soit la taille n de l'échantillon considéré.

On peut également prouver les deux propriétés suivantes que nous admettrons sans démonstration.

Propriété 9.20 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$. On a :*

$$\frac{n\Sigma_n^2}{\sigma^2} = \frac{(n-1)S_n^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$$

Il s'agit à nouveau d'un résultat exact, valable quelle que soit la taille d'échantillon n.

Propriété 9.21 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$. Les statistiques :*

$$V_1 = \frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \quad et \quad V_2 = \frac{n\Sigma_n^2}{\sigma^2} = \frac{(n-1)S_n^2}{\sigma^2}$$

sont des variables aléatoires indépendantes.

On sait que $V_1 \rightsquigarrow \mathcal{N}(0,1)$ et que $V_2 \rightsquigarrow \chi^2(n-1)$. V_1 et V_2 étant indépendantes, une simple application de la définition de la loi de STUDENT montre que :

$$\frac{V_1}{\sqrt{\frac{V_2}{n-1}}} = \frac{\frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\frac{n\Sigma_n^2}{\sigma^2}}{n-1}}} = \frac{\frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\frac{(n-1)S_n^2}{\sigma^2}}{n-1}}} = \frac{\overline{X}_n - m}{\sqrt{\frac{\Sigma_n^2}{n-1}}} = \frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \rightsquigarrow t(n-1)$$

Nous avons ainsi établi la propriété suivante.

Propriété 9.22 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$. On a :*

$$\frac{\overline{X}_n - m}{\sqrt{\frac{\Sigma_n^2}{n-1}}} = \frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \rightsquigarrow t(n-1)$$

9.5.1. Exercice résolu

On fabrique des pièces en série à l'aide d'une machine. La machine est réglée de telle sorte que le diamètre des pièces fabriquées est distribué selon une loi normale $\mathcal{N}(32;1)$. On prélève un échantillon aléatoire de 10 pièces dans la production et on calcule le diamètre moyen pour cet échantillon.

1- Quelle est la loi du diamètre moyen ?

Notons X le diamètre d'une pièce. La loi de la population est : $X \rightsquigarrow \mathcal{N}(32;1)$. Le diamètre moyen dans l'échantillon aléatoire est donné par la statistique $\overline{X}_{10}$. Par application de la Propriété 9.19, on obtient :

$$\overline{X}_{10} \rightsquigarrow \mathcal{N}\left(32; \tfrac{1}{10}\right) = \mathcal{N}(32; 0,1)$$

Notez que si on ne connaissait que le diamètre moyen ($m = E(X) = 32$) et la variance du diamètre ($\sigma^2 = V(X) = 1$) du diamètre des pièces fabriquées, on ne pourrait rien dire de la loi de $\overline{X}_{10}$: l'échantillon est en effet trop petit ($n = 10 < 30$) pour pouvoir appliquer le théorème central limite, et considérer que la loi ci-dessus est approximativement correcte.

2- Dans quel intervalle symétrique autour de sa valeur espérée a-t-on 99% de chances d'observer le diamètre moyen ?

La valeur espérée du diamètre moyen est $E(\overline{X}_{10}) = 32$. On cherche un intervalle du type $[32 - a\ ;\ 32 + a]$ tel que :

$$\begin{aligned} 0,99 &= I\!P(32 - a \leq \overline{X}_{10} \leq 32 + a) \\ &= I\!P\left(\frac{32 - a - 32}{\sqrt{0,1}} \leq \frac{\overline{X}_{10} - 32}{\sqrt{0,1}} \leq \frac{32 + a - 32}{\sqrt{0,1}}\right) \\ &= I\!P\left(\frac{-a}{0,3162} \leq Z \leq \frac{a}{0,3162}\right) \end{aligned}$$

où $Z \rightsquigarrow \mathcal{N}(0;1)$, soit la valeur a telle que :

$$2\Phi\left(\frac{a}{0,3162}\right) - 1 = 0,99 \Leftrightarrow \Phi\left(\frac{a}{0,3162}\right) = 0,995$$

Dans les tables, on trouve $\Phi(2,5758) = 0,995$. On a donc :

$$\frac{a}{0,3162} = 2,5758 \Leftrightarrow a = 2,5758 \times 0,3162 = 0,8144$$

L'intervalle demandé est dès lors :

$$[32 - 0,8144\ ;\ 32 + 0,8144] = [31,1856\ ;\ 32,8144]$$

3- On mesure la dispersion du diamètre des pièces dans l'échantillon comme indiqué par la statistique S_n^2. Quelle est ici la loi de cette statistique ? Donnez son espérance et sa variance.

De la Propriété 9.20, on obtient :

$$9S_{10}^2 \rightsquigarrow \chi^2(9)$$

De la Propriété 7.8, on sait que si $X \rightsquigarrow \chi^2(\nu)$, alors $E(X) = \nu$ et $V(X) = 2\nu$. On en déduit :

$$E(9S_{10}^2) = 9 \Leftrightarrow 9E(S_{10}^2) = 9 \Leftrightarrow E(S_{10}^2) = 1$$

et

$$V(9S_{10}^2) = 2 \times 9 = 18 \Leftrightarrow 9^2 V(S_{10}^2) = 18 \Leftrightarrow V(S_{10}^2) = 0,222...$$

9.6. Exercice résolu

Une entreprise d'horlogerie fabrique des montres dont la déviation par rapport à l'heure exacte en 24 heures est en moyenne de 0 seconde et avec un écart-type de 30 secondes. Cette déviation peut être positive ou négative : lorsqu'elle est positive la montre avance, lorsqu'elle est négative la montre retarde.

1- On prélève un échantillon aléatoire de 100 montres. Quelle est la loi de la déviation moyenne dans l'échantillon ?

   On note X la déviation en secondes par rapport à l'heure exacte en 24 heures d'une montre. La loi de la population est inconnue mais on sait qu'elle est telle que $m = E(X) = 0$ et que $\sigma = \sigma(X) = 30$. La déviation moyenne par rapport à l'heure exacte dans un échantillon aléatoire de 100 montres est donnée par la statistique $\overline{X}_{100}$. La loi de X étant inconnue et la taille de l'échantillon grande ($n \geq 30$), on utilise le théorème central limite pour déduire une loi approchée pour $\overline{X}_{100}$:

$$\overline{X}_{100} \approx \mathcal{N}\left(0; \tfrac{30^2}{100}\right) = \mathcal{N}(0; 9)$$

2- On suppose à présent que la déviation par rapport à l'heure exacte en 24 heures est distribuée de façon normale. Calculez la probabilité qu'une montre tirée au hasard avance d'au moins 1 minute par 24 heures.

   Nous avons : $X \rightsquigarrow \mathcal{N}(0; 30^2) \Leftrightarrow Z = (X - 0)/30 \rightsquigarrow \mathcal{N}(0; 1)$. La probabilité demandée est :

$$\begin{aligned} I\!P(X \geq 60) &= 1 - I\!P(X \leq 60) = 1 - I\!P\left(\frac{X-0}{30} \leq \frac{60-0}{30}\right) \\ &= 1 - I\!P(Z \leq 2) = 1 - \Phi(2) = 1 - 0,9772 = 0,0228 \end{aligned}$$

3- On prélève un échantillon aléatoire de 15 montres. Calculez la probabilité que la déviation moyenne dans l'échantillon soit en valeur absolue supérieure à 5 secondes.

   Soit $\overline{X}_{15}$ la déviation moyenne pour cet échantillon aléatoire. Puisque X suit une loi normale, nous déduisons la loi exacte de $\overline{X}_{15}$:

$$\overline{X}_{15} \rightsquigarrow \mathcal{N}\left(0; \tfrac{30^2}{15}\right) = \mathcal{N}(0; 60) \quad \Leftrightarrow \quad Z = \frac{\overline{X}_{15} - 0}{\sqrt{60}} \rightsquigarrow \mathcal{N}(0; 1)$$

   La probabilité demandée est :

$$\begin{aligned} I\!P(\left|\overline{X}_{15}\right| > 5) &= 1 - I\!P(-5 \leq \overline{X}_{15} \leq 5) \\ &= 1 - I\!P\left(\frac{-5-0}{\sqrt{60}} \leq \frac{\overline{X}_{15}-0}{\sqrt{60}} \leq \frac{5-0}{\sqrt{60}}\right) \\ &= 1 - I\!P(-0,65 \leq Z \leq 0,65) \\ &= 1 - (2\Phi(0,65) - 1) = 1 - (2 \times 0,7422) + 1 = 0,5156 \end{aligned}$$

4- On prélève un échantillon aléatoire de n montres. A partir de quelle valeur de

n cette probabilité est-elle au plus égale à $0,1$?

Nous avons :

$$\overline{X}_n \rightsquigarrow \mathcal{N}\left(0; \tfrac{30^2}{n}\right) \quad \Leftrightarrow \quad Z = \frac{\overline{X}_n - 0}{\sqrt{\frac{30^2}{n}}} \rightsquigarrow \mathcal{N}(0;1)$$

On cherche la plus petite valeur de n telle que : $I\!P(|\overline{X}_n| > 5) \leq 0,1$. On a :

$$\begin{aligned} I\!P(|\overline{X}_n| > 5) &= 1 - I\!P(-5 \leq \overline{X}_n \leq 5) \\ &= 1 - I\!P\left(\frac{-5-0}{\sqrt{\frac{30^2}{n}}} \leq \frac{\overline{X}_n - 0}{\sqrt{\frac{30^2}{n}}} \leq \frac{5-0}{\sqrt{\frac{30^2}{n}}}\right) \\ &= 1 - I\!P\left(-\frac{5}{30}\sqrt{n} \leq Z \leq \frac{5}{30}\sqrt{n}\right) \\ &= 1 - \left(2\Phi(\tfrac{5}{30}\sqrt{n}) - 1\right) = 2 - 2\Phi(\tfrac{5}{30}\sqrt{n}) \end{aligned}$$

Comme $\Phi(.)$ est croissante, la plus petite valeur de n telle que $2 - 2\Phi(\frac{5}{30}\sqrt{n}) \leq 0,1$ est donnée par la valeur de n satisfaisant l'égalité :

$$2 - 2\Phi(\tfrac{5}{30}\sqrt{n}) = 0,1 \Leftrightarrow \Phi(\tfrac{5}{30}\sqrt{n}) = \frac{2-0,1}{2} = 0,95$$

Dans les tables, on trouve $\Phi(1,6449) = 0,95$. On a donc :

$$\tfrac{5}{30}\sqrt{n} = 1,6449 \Leftrightarrow n = \left(\tfrac{30}{5} \times 1,6449\right)^2 = 97,405$$

Il faut donc un échantillon d'au moins 98 montres.

5- On prélève un échantillon aléatoire de 15 montres. Quelle est la probabilité que la statistique S_{15}^2 soit :

a- plus petite que 1 680 ?

Nous avons :

$$\frac{(15-1)S_{15}^2}{30^2} = \frac{14}{900}S_{15}^2 \rightsquigarrow \chi^2(14)$$

La probabilité demandée est :

$$\begin{aligned} I\!P(S_{15}^2 < 1\,680) &= I\!P(S_{15}^2 \leq 1\,680) = I\!P\left(\frac{14}{900}S_{15}^2 \leq \frac{14 \times 1\,680}{900}\right) \\ &= I\!P(Y \leq 26,13) \end{aligned}$$

où $Y \rightsquigarrow \chi^2(14)$. Dans les tables, on trouve : $I\!P(Y \leq 26,13) \approx 0,975$.

b- comprise entre 360 et 1 680 ?

La probabilité demandée est :

$$\begin{aligned}
\mathbb{P}(360 < S^2_{15} < 1\,680) &= \mathbb{P}(360 \leq S^2_{15} \leq 1\,680) \\
&= \mathbb{P}\left(\frac{14 \times 360}{900} \leq \frac{14}{900} S^2_{(15)} \leq \frac{14 \times 1\,680}{900}\right) \\
&= \mathbb{P}(5,6 \leq Y \leq 26,13) \quad (\text{où } Y \leadsto \chi^2(14)) \\
&= \mathbb{P}(Y \leq 26,13) - \mathbb{P}(Y \leq 5,6) \\
&\approx 0,975 - 0,025 = 0,95
\end{aligned}$$

6- Toujours en prélevant un échantillon aléatoire de 15 montres, quelle est la probabilité qu'il y en ait au moins deux qui avancent d'au moins 1 minute par 24 heures ?

Notons Y une variable de BERNOULLI qui prend la valeur 1 si une montre tirée au hasard avance d'au moins 1 minute par 24 heures et la valeur 0 sinon. On a : $Y \leadsto \mathcal{B}(p)$, où $p = \mathbb{P}(Y = 1)$ est la fréquence des montres qui avancent d'au moins 1 minute par 24 heures dans la population des montres. Du point 2 ci-dessus, nous savons que $\mathbb{P}(Y = 1) = \mathbb{P}(X \geq 60) = 0,0228$. On a donc : $Y \leadsto \mathcal{B}(0,0228)$ (= la loi de la population concernant la caractéristique individuelle "avancer d'au moins 1 minute par 24 heures / ne pas avancer d'au moins 1 minute par 24 heures"). L'échantillon aléatoire de 15 montres est décrit par un n-uple $(Y_1, Y_2, ..., Y_{15})$, où les variables aléatoires Y_i $(i = 1, ..., 15)$ sont indépendantes et toutes de même loi que celle de Y. Le nombre de montres parmi les 15 avançant d'au moins 1 minute par 24 heures est donné par la variable aléatoire Z :

$$Z = \sum_{i=1}^{15} Y_i$$

Par construction, Z suit une loi binomiale $\mathcal{B}(15; 0,0228)$: $Z \leadsto \mathcal{B}(15; 0,0228)$. La probabilité demandée est donc :

$$\begin{aligned}
\mathbb{P}(Z \geq 2) &= 1 - \mathbb{P}(Z < 2) = 1 - \mathbb{P}(Z = 0) - \mathbb{P}(Z = 1) \\
&= 1 - C^0_{15}\ (0,0228)^0 (0,9772)^{15} - C^1_{15} (0,0228)^1 (0,9772)^{14} \\
&= 1 - 0,7075 - 0,2476 = 0,0449
\end{aligned}$$

9.7. Exercices

1- On suppose que la note en statistique des étudiants d'une grande classe est en moyenne de 58 avec un écart-type de 29. Trouvez la probabilité que la note moyenne dans un échantillon aléatoire de 100 étudiants soit :

a- de moins de 54.

b- de plus de 65.

c- de plus de 50.

2- Des briques sont fabriquées selon une technique qui assure que leur poids est en moyenne de $1,6$ kg avec un écart-type de 30 g. On prélève un échantillon aléatoire dans la fabrication.

a- Quel est l'écart-type de la moyenne-échantillon $\overline{X}_n$:

i- lorsque la taille de l'échantillon vaut 10 ?

ii- lorsque la taille de l'échantillon vaut 50 ?

b- A partir de quelle taille d'échantillon la moyenne-échantillon s'écartera-t-elle, avec une probabilité de 98%, de sa valeur espérée de moins de 5 g ?

c- Supposons qu'on prélève un échantillon aléatoire de 100 briques et, tout à fait indépendamment du premier, un deuxième échantillon aléatoire de 50 briques. On considère les deux moyennes-échantillons $\overline{X}^A_{100}$ et $\overline{X}^B_{50}$. Calculez la probabilité que les moyennes-échantillon diffèrent de plus de 10 g.

3- En Boursouphlie, le poids moyen des hommes est de 90 kg, et celui des femmes de 75 kg. La variance du poids des hommes est de 500 kg^2, et celle des femmes de 400 kg^2. Quelle est la probabilité que :

a- le poids moyen d'un échantillon aléatoire de 625 femmes soit compris entre 74 kg et 77 kg ?

b- le poids moyen d'un échantillon aléatoire de 320 femmes soit au moins égal au 8/10 du poids moyen d'un échantillon aléatoire de 320 hommes ? On suppose que les deux échantillons sont prélevés de manière indépendante.

4- On considère l'ensemble des petites entreprises de la région et on s'intéresse au montant de leur profit net l'an passé, que l'on note X. On représente ce profit X à l'aide d'une loi d'espérance $m = E(X) = 210\,000$ F et d'écart-type $\sigma = \sigma(X) = 52\,000$ F. On tire un échantillon aléatoire de 250 firmes.

a- Quelle est la probabilité que la moyenne-échantillon $\overline{X}_n$ se situe dans un intervalle de $\pm$ 5 000 autour de sa valeur espérée ?

b- Est-ce que cette probabilité serait sensiblement augmentée si la taille de l'échantillon passait à 300 entreprises ?

c- Et si l'écart-type σ passait à 49 000 F ?

d- Compléter le tableau suivant en calculant la probabilité de la première question pour différentes valeurs de σ et de n.

σ	$n = 200$	$n = 250$	$n = 300$
52 000			
49 000			
45 000			

5- On suppose que le nombre de voitures empruntant, par minute, l'autoroute entre Liège et Ans l'après-midi en semaine est distribué selon une loi de Poisson de paramètre $\lambda = 20$. On choisit au hasard 30 périodes d'observation d'une minute durant lesquelles on dénombre les voitures empruntant l'autoroute. On considère la variable aléatoire $\overline{X}_{30}$ mesurant le nombre moyen de véhicules par minute empruntant l'autoroute sur ces 30 périodes.

a- Déduisez du théorème central limite une loi approchée pour $\overline{X}_{30}$.
b- Quel est la probabilité d'avoir un nombre moyen de véhicules par minute sur ces 30 périodes supérieur à 22 ?

6- La durée d'un vol Paris – New York est en moyenne de 5 heures, avec un écart-type de 24 minutes.
   a- On considère la durée moyenne d'un vol calculée sur un échantillon aléatoire de 40 vols.
      i- Donnez l'espérance et la variance de la durée moyenne.
      ii- Quelle est la probabilité que la durée moyenne soit comprise entre $4,9$ et $5,2$ heures ?
      iii- Que devient cette probabilité si le nombre de vols considérés est 80 au lieu de 40 ?
   b- On suppose maintenant que la durée d'un vol Paris – New York est distribuée selon une loi normale $\mathcal{N}(5; 0,16)$ et que la durée d'un vol retour New York – Paris est distribuée selon une loi normale $\mathcal{N}(5,2; 0,1225)$. Quelle est, pour le personnel navigant, la probabilité que la durée totale d'un voyage Paris – New York – Paris pris au hasard soit supérieure à $10,5$ heures :
      i- en supposant qu'il n'y a pas de délai à New York entre l'aller et le retour ?
      ii- en supposant qu'il y a un délai d'une heure ?
   c- On continue à supposer la durée d'un vol Paris – New York est distribuée selon une loi normale $\mathcal{N}(5; 0,16)$. Quelle est la probabilité :
      i- qu'un voyage Paris – New York pris au hasard dure plus de $5,8$ heures ?
      ii- que sur 12 voyages pris au hasard ce phénomène s'observe au moins 1 fois ?

7- Un médecin assure une consultation dans un hôpital tous les mercredis après-midi. Le nombre de patients se présentant à cette consultation est distribué selon une loi de Poisson de paramètre $\lambda = 12$.
   a- On considère une consultation prise au hasard. Quelle est la probabilité que 5 patients au plus se présentent à la consultation ?
   b- On considère 10 consultations tirées au hasard. Quelle est la probabilité d'observer au moins deux consultations où au plus 5 patients se présentent ?
   c- On considère 50 consultations choisies au hasard et on s'intéresse au nombre moyen de patients se présentant à ces consultations. Quelle est la probabilité d'observer un nombre moyen de patients supérieur à 13 ?

8- On suppose que 40% des clients qui fréquentent le magasin Cheapshop en sortent en ayant acheté quelque chose. On tire au hasard 300 clients du magasin et on étudie la fréquence d'achat F_{300}.
   a- Déterminez la loi de F_{300}. Donnez son espérance et sa variance.
   b- Quelle est la probabilité d'avoir une fréquence d'achat inférieure à 35% ?

9- On admet que l'injection d'un cocktail médicamenteux à des souris de laboratoire préalablement infectées par un virus permet à 20% d'entre-elles d'échapper à une mort sinon certaine. On choisit au hasard 81 souris, qu'on infecte par le virus et auxquelles on injecte ensuite le cocktail médicamenteux.

a- Quelle est la loi du nombre de décès parmi ces 81 souris ?
b- Quelle est la loi de la fréquence de décès parmi ces 81 souris ?
c- Quelle est la probabilité d'observer une fréquence de décès parmi ces 81 souris s'écartant de sa valeur espérée de plus de $0,1$?
d- Dans quel intervalle centré autour de $0,8$ peut-on s'attendre à trouver la fréquence de décès parmi ces 81 souris avec une probabilité de 95% ? A quel nombre de décès cela correspond-il ?
e- Et si la probabilité était de 99% ?

10- On considère une élection à un tour où deux candidats sont en compétition : A et B. On cherche à évaluer les intentions de vote des électeurs pour chaque candidat : p_A et p_B. On pratique un sondage par échantillonnage aléatoire de n électeurs (n est supposé grand). On note F_n^A la proportion des personnes interrogées ayant l'intention de voter pour A et F_n^B la proportion des personnes interrogées ayant l'intention de voter pour B. On suppose que $p_B = 1 - p_A$ et que $F_n^B = 1 - F_n^A$.

a- Que signifient concrètement les hypothèses $p_B = 1 - p_A$ et $F_n^B = 1 - F_n^A$?
b- Quelle est la loi de F_n^A ?
c- On cherche à évaluer avec quelle probabilité le sondage donne une image exacte des intentions de vote de la population. Dans cette perspective, compléter le tableau suivant en calculant la probabilité $\mathbb{P}(F_n^A \leq 0,5)$ pour les différentes valeurs de p_A et de n.

p_A	$n = 100$	$n = 400$	$n = 1000$
$0,4$			
$0,45$			
$0,475$			

Comment interprétez-vous les quantités obtenues ?
d- On se propose de limiter à 5% le risque que $F_n^A \geq 0,5$ quand $p_A = 0,48$. Combien faut-il interroger d'électeurs ?

11- Dans une ville, 55% des jurés éligibles sont des femmes. On tire au hasard un jury de 40 membres, quelle est la probabilité qu'il comprenne moins de femmes que d'hommes ?

12- Une maladie est soignée par un traitement classique que l'on sait, par expérience, être efficace dans 60% des cas.

a- On considère 100 malades pris au hasard que l'on traite par ce traitement classique. Quelle est la loi de la fréquence de guérison ? Quelle est la probabilité d'obtenir une fréquence de guérison supérieure à 65% ?
b- Déterminez f de manière telle que, sur un échantillon aléatoire de 100 malades, la proportion de guérisons soit au plus égale à f avec une probabilité de 95% ?

13- Le montant (en FF) que dépensent dans son restaurant les clients d'un restaurateur français est distribué selon une loi normale $\mathcal{N}(180; 400)$. On prélève un échantillon aléatoire de clients.

a- Quelle est la loi de $\overline{X}_n$ lorsque n vaut 5 ?
b- Quelle est la loi de $\overline{X}_n$ lorsque n vaut 100 ?
c- Considérons à présent que $n = 100$. Quelle est la probabilité d'observer une moyenne-échantillon se situant à moins de 5 FF de la valeur espérée ?
d- Dans quel intervalle centré autour de la valeur espérée a-t-on 99% de chances d'observer la moyenne-échantillon ?
e- Quelle est la loi de la statistique S_n^2 ? Quelle est son espérance ? Quelle est son écart-type ?

14- Supposons que le poids des étudiants de l'université est distribué selon une loi normale $\mathcal{N}(68; 9)$. On tire un échantillon aléatoire de 80 étudiants.

a- Quelle est la probabilité d'obtenir un poids moyen pour l'échantillon qui s'écarte de plus de 500 g de la moyenne dans la population ?
b- Quelle taille d'échantillon faut-il considérer pour réduire cette probabilité à moins de 2% ?

15- On suppose que le poids (en grammes) de paquets de farine remplis par une machine en bon état de fonctionnement est distribué suivant une loi normale $\mathcal{N}(501; 1)$. Un inspecteur prend au hasard un échantillon de n paquets de la production pour déterminer si leur poids moyen est au moins de 500 grammes. Dans la négative, la firme risque une amende de 50 000 F. Quelle est la probabilité d'encourir une telle amende si la taille n de l'échantillon est de respectivement 1, 4 et 16 ?

16- La consommation d'essence par jour ouvrable d'un voyageur de commerce est en moyenne de 13 litres, avec un écart-type 2 litres.

a- Quels sont l'espérance et l'écart-type de la consommation totale sur 50 jours ouvrables pris au hasard ?
b- On considère à présent la consommation moyenne par jour sur 50 jours ouvrables choisis au hasard.
    i- Quelle est la probabilité que cette consommation moyenne soit inférieure à $12,5$ litres ?
    ii- Quelle est la probabilité que cette consommation moyenne soit supérieure à $13,5$ litres ?
    iii- Quelle est la probabilité que cette consommation moyenne ne s'écarte pas de plus de $0,1$ litres de son espérance ?
    iv- Pourrait-on calculer ces probabilités si on raisonnait sur 10 jours au lieu de 50 ?
c- On suppose maintenant que la consommation journalière est distribuée selon une loi normale $\mathcal{N}(13; 4)$.
    i- Quelle est la probabilité que sur 5 jours ouvrables pris au hasard la consommation moyenne soit comprise entre $12,5$ et $13,5$ litres ?
    ii- Quelle est la probabilité que sur 50 jours ouvrables choisis au hasard il y en ait au moins 20 dont la consommation d'essence est comprise entre $12,5$ et $13,5$ litres ?

17- La fréquentation journalière, en semaine, d'un supermarché est en moyenne de

4 000 clients, avec un écart-type de 300 clients.

a- Donnez l'espérance et la variance du nombre de clients qui fréquenteront le supermarché mardi prochain.

b- On observe le nombre moyen de clients par jour calculé sur 50 jours pris au hasard.

 i- Quelle est la probabilité que ce nombre moyen soit compris entre 3 900 et 4 050 clients ?

 ii- Et si le nombre de jours considérés était de 12 jours au lieu de 50 ?

c- On suppose maintenant que la fréquentation journalière, en semaine, est distribuée selon une loi normale $\mathcal{N}(4\,000; 300^2)$.

 i- Quelle est la probabilité que le nombre total de clients sur 5 jours choisis au hasard soit plus grand que 24 764 ?

 ii- Quelle est la probabilité que sur ces 5 jours, il y en ait au moins 3 où le nombre de clients dépasse 4 500 ?

d- On suppose que la fréquentation du samedi est différente de celle des jours de semaine. On admet que le nombre de clients fréquentant le supermarché le samedi est distribué selon une loi normale $\mathcal{N}(10\,000; 200^2)$, alors que ce nombre suit une loi normale $\mathcal{N}(4\,000; 300^2)$ les autres jours de la semaine. Quelle est la probabilité que :

 i- le nombre de clients lors d'un samedi choisi au hasard soit plus de deux fois et demi plus grand que le nombre de clients lors d'un jour de semaine pris au hasard ?

 ii- le nombre moyen de clients calculé sur 10 samedis choisis au hasard soit plus de deux fois et demi plus grand que le nombre moyen de clients calculé sur 10 jours de semaine pris au hasard ?

18- Une entreprise fabrique des pâtes alimentaires qui sont vendues en paquets dont le poids est en moyenne de 500 grammes avec un écart-type de 10 grammes.

a- On calcule, sur la base de 50 paquets de pâtes prélevés au hasard dans la production, le poids moyen par paquet.

 i- Donnez l'espérance et l'écart-type du poids moyen d'un paquet.

 ii- Quelle est la probabilité que ce poids moyen soit compris entre 499 et 502 grammes ?

b- On suppose maintenant que le poids des paquets de pâtes est distribué suivant une loi normale $\mathcal{N}(500; 100)$. On emballe 20 paquets de pâtes pris au hasard dans un carton. Quelle est la probabilité que le poids total du carton soit supérieur à 10 050 grammes :

 i- en supposant que le poids du carton est négligeable ?

 ii- en supposant que le poids du carton est 80 grammes ?

 iii- en supposant que le poids du carton, lui-même choisi au hasard, est distribué selon une loi $\mathcal{N}(80; 225)$?

c- On continue à supposer le poids des paquets de pâtes est distribué suivant une loi normale $\mathcal{N}(500; 100)$.

 i- Construisez un intervalle centré en la valeur espérée 500 grammes contenant le poids d'un paquet pris au hasard avec une probabilité de 85%.

ii- On considère un échantillon aléatoire de n paquets. Pour quelle valeur de n la probabilité d'avoir un poids moyen par paquet inférieur à 502 grammes est-elle égale à $0,95$?

19- Un arboriculteur produit des oranges dont le diamètre X est distribué de façon normale avec une moyenne égale à 12 cm et un écart-type égal à $0,8$ cm.

a- Quelle est la probabilité qu'une orange tirée au sort ait un diamètre supérieur à 13 cm ?

b- On a un tas de plusieurs milliers d'oranges issues de la production que l'on trie pour en faire trois tas, l'un formé des oranges de diamètre inférieur à 11 cm, le deuxième des oranges de diamètre compris entre 11 cm et 13 cm, et le troisième des oranges de diamètre supérieur à 13 cm.

i- Dans quelles proportions approximatives les oranges vont-elles se répartir entre les trois tas ? Justifiez votre réponse.

ii- On suppose que le tas à trier contient 5 000 oranges. Quelle est la probabilité qu'après le tri, le deuxième tas contienne moins de 78% des oranges ? et plus 80% des oranges ? Sur quelle hypothèse implicite s'appuie vos réponses ?

c- L'arboriculteur range, au hasard, sans égard à leur taille, les oranges qu'il produit par panier six oranges. On appelle $\overline{X}_6$ le diamètre moyen des oranges d'un panier.

i- Quelle est la loi de $\overline{X}_6$?

ii- Quelle est la probabilité que $\overline{X}_6$ soit supérieur à $11,5$ cm ?

iii- L'arboriculteur livre à un client 80 paniers de six oranges, paniers qu'il a prélevés au hasard dans son stock de paniers. Quelle est la probabilité qu'il y ait au moins dix paniers tels que le diamètre moyen des oranges soit inférieur à $11,5$ cm ?

20- On admet que la durée de vie (en années) d'une femme est en moyenne de 80 ans, avec une variance de $38,44$ ans^2, et qu'elle est distribuée de façon normale.

a- On considère une femme prise au hasard.

i- Quelle est la probabilité qu'elle vive plus de 85 ans ?

ii- Dans quel intervalle centré en 80 ans se situera sa durée de vie avec une probabilité de 90% ?

b- On considère deux femmes choisies au hasard.

i- Quelle est la probabilité qu'elles aient exactement la même durée de vie ?

ii- Quelle est la probabilité que leur durée de vie diffère de 3 mois ou plus ?

c- On considère 100 femmes choisies au hasard.

i- Quelle est la probabilité d'observer une durée de vie moyenne pour ces 100 femmes comprise entre 79 et 80 ans ?

ii- Combien de personnes faut-il considérer pour porter cette probabilité soit d'au moins $48,5\%$?

iii- Combien de femmes peut-on espérer voir atteindre 85 ans parmi les 100 considérées ?

21- Une usine fabrique des tôles galvanisées. Le processus de fabrication est composé de 3 étapes effectuées à la suite les unes des autres et de manière indépendante : décapage de la tôle, laminage à froid et galvanisation. On suppose que la distribution des durées de chaque étape est normale. La moyenne et l'écart-type de la durée (exprimée en minutes) de chaque étape pour une bobine de 25 tonnes d'acier sont donnés dans le tableau suivant :

Etape	Moyenne	Ecart-type
Décapage	12	2
Laminage	18	4
Galvanisation	14	3

a- On considère une bobine choisie au hasard dans la production.

i- Quelle est la probabilité que sa durée de décapage ait été supérieure à 14 minutes ?

ii- Quelle est la probabilité que son traitement complet (décapage, laminage et galvanisation) ait duré moins de 40 minutes ?

iii- Combien de bobines peut-on espérer traiter complètement en 8 heures de travail ?

b- L'étape de fabrication la moins automatisée est le laminage. Celui-ci nécessite beaucoup de manutention. Pour améliorer le rendement de l'atelier de laminage, l'entreprise accorde une prime par bobine dès que son laminage dure moins de 14 minutes.

i- Quelle est la probabilité d'accorder une prime pour une bobine choisie au hasard ?

ii- On considère un échantillon aléatoire de 200 bobines. Quelle est la probabilité d'accorder une prime pour plus de 20% de ces bobines ?

22- La quantité de viande vendue par demi-journée par une boucherie est en moyenne de 33 kg avec un écart-type de $1,2$ kg.

a- On considère la quantité moyenne vendue sur 35 demi-journées choisies au hasard.

i- Quelle est la loi approximative de cette quantité moyenne ?

ii- Calculez la probabilité que cette quantité moyenne s'écarte de sa valeur espérée de moins de $0,4$ kg.

b- On suppose maintenant que la quantité vendue par demi-journée est normale $\mathcal{N}(33; 1,44)$.

i- Quelle est la loi de la quantité moyenne vendue sur 5 demi-journées prises au hasard ?

ii- Quelle est la probabilité que cette quantité moyenne s'écarte de sa valeur espérée de plus de $0,4$ kg ?

iii- Sur 7 groupes de 5 demi-journées prises au hasard, quelle est la probabilité que l'événement envisagé à la question précédente se réalise au moins trois fois ?

23- On suppose que le poids des pièces produites par une machine est distribué de façon normale et en moyenne égal à 120 grammes, avec un écart-type de 7 grammes.

a- On considère une pièce choisie au hasard dans la production. Quelle est la probabilité que son poids ne dépasse pas 118 grammes ?

b- On considère 3 pièces prélevées au hasard dans la production.

  i- Quelle est la loi de leur poids total ?

  ii- Dans quel intervalle centré en l'espérance ce poids total se situera-t-il avec une probabilité de 96% ?

c- On considère 100 pièces prélevées au hasard dans la production. Quelle est la probabilité que leur poids moyen soit supérieur à 121 grammes ?

Chapitre 10

Estimation ponctuelle et intervalle de confiance

En exploitant les liens qui unissent caractéristiques d'une population et caractéristiques d'un échantillon aléatoire issu de cette population, on peut dégager des procédures permettant de tirer des conclusions fiables concernant les caractéristiques d'une population sur base d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ obtenu par échantillonnage aléatoire dans cette population. De telles procédures sont appelées *procédures d'inférence statistique*.

On peut classer les procédures d'inférence statistique en deux grandes catégories : les procédures d'estimation et les procédures de tests. Dans ce chapitre, on se concentre sur l'estimation (estimation ponctuelle et intervalle de confiance). Nous aborderons les tests au chapitre suivant.

10.1. Estimation ponctuelle

10.1.1. Estimateur et estimation

Plantons le décor. On suppose qu'on considère une population, population qu'on représente par la loi d'une variable aléatoire X, et qu'on s'intéresse à une caractéristique bien précise de la population, caractéristique qui est représentée par un paramètre, que nous noterons θ, de la loi de X. Ce paramètre est appelé *paramètre d'intérêt*. C'est un nombre dont la valeur est inconnue et que l'on cherche à estimer.

Dans le cadre de ce cours, on se concentrera sur le paramètre $m = E(X)$ qui représente la moyenne des valeurs prises par une variable d'intérêt dans une population décrite par la loi d'une variable aléatoire X, le paramètre $\sigma^2 = V(X)$ qui représente la variance des valeurs prises par une variable d'intérêt dans une population décrite par la loi de X, et le paramètre $p = I\!P(X = 1)$ d'une loi de Bernoulli $\mathcal{B}(p)$, paramètre qui, on le sait, représente la proportion des individus d'une population — population qui est décrite par la loi $\mathcal{B}(p)$ — qui possèdent une caractéristique donnée.

On suppose que pour estimer θ on dispose d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ obtenu par échantillonnage aléatoire dans la population, et qui donc constitue une réalisation particulière d'un échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à la variable aléatoire X.

Qui dit estimation sur base d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ de la valeur inconnue d'un paramètre d'intérêt θ dit choix d'une règle de décision décrivant la façon dont, quelle que soit la réalisation observée $(x_1, x_2, ..., x_n)$ d'un échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à X, on va évaluer le paramètre inconnu θ. Cette règle de décision, dont on peut décrire la forme et le résultat de l'application — qui forcément varie d'un échantillon à l'autre — par une statistique, est appelée un *estimateur*.

Définition 10.1 *Soient θ un paramètre de la loi d'une variable aléatoire X et $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à cette variable aléatoire. On appelle estimateur du paramètre θ toute statistique $T_n = g(X_1, X_2, ..., X_n)$ représentant une règle de décision utilisée pour estimer la valeur inconnue du paramètre θ.*

Le résultat de l'application à un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier de la règle de décision définie par un estimateur est appelée une *estimation*.

Définition 10.2 *Soient θ un paramètre de la loi d'une variable aléatoire X et $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à cette variable aléatoire. On appelle estimation du paramètre θ une réalisation particulière $\widehat{\theta} = g(x_1, x_2, ..., x_n)$, pour un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier, d'un estimateur $T_n = g(X_1, X_2, ..., X_n)$ du paramètre θ.*

$$\begin{tikzcd}
\boxed{\begin{array}{c}\text{Population :}\\ \text{loi de } X \text{ et paramètre } \theta\end{array}} \arrow[d] & \\
\boxed{\begin{array}{c}\text{Echantillon aléatoire :}\\ (X_1, X_2, ..., X_n)\end{array}} \arrow[d] \arrow[r] & \boxed{\begin{array}{c}\text{Estimateur :}\\ T_n = g(X_1, X_2, ..., X_n)\end{array}} \\
\boxed{\begin{array}{c}\text{Observations :}\\ (x_1, x_2, ..., x_n)\end{array}} \arrow[r] & \boxed{\begin{array}{c}\text{Estimation :}\\ \widehat{\theta} = g(x_1, x_2, ..., x_n)\end{array}}
\end{tikzcd}$$

Graphique 10.1 : Estimateur et estimation

Tout comme il ne faut pas confondre un échantillon aléatoire $(X_1, X_2, ..., X_n)$, qui décrit ce qu'on est susceptible d'obtenir lorsqu'on tire un échantillon par échantillonnage aléatoire, et un échantillon d'observations $(x_1, x_2, ..., x_n)$, qui est un ensemble de nombres (et donc non aléatoire) et qui décrit ce qui est effectivement obtenu lors d'un tirage particulier, on veillera à ne pas confondre un estimateur $T_n = g(X_1, X_2, ..., X_n)$, qui décrit la règle de décision utilisée par estimer le valeur inconnue du paramètre θ et dont le résultat de l'application varie forcément d'un échantillon à l'autre, et une estimation $\widehat{\theta} = g(x_1, x_2, ..., x_n)$ qui est un nombre (et donc non aléatoire) et qui décrit ce qui est effectivement obtenu de l'application de la règle de décision définie par l'estimateur pour un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier.

10.1.2. Estimateurs pour la moyenne, la variance et la fréquence

A l'évidence, les résultats obtenus au Chapitre 9 suggèrent de considérer comme estimateur de la moyenne m, de la variance σ^2 et de la fréquence p dans une population respectivement la moyenne-échantillon $\overline{X}_n$, la variance-échantillon Σ_n^2 ou sa version modifiée S_n^2, et la fréquence-échantillon F_n.

Avant d'examiner la qualité de ces estimateurs, illustrons-en l'utilisation pratique. Supposons qu'on veuille étudier la fréquentation de la bibliothèque par les étudiants de la Faculté. On constitue un échantillon aléatoire de 50 étudiants et on leur demande le nombre de fois qu'ils se sont rendus à la bibliothèque la semaine précédente. On obtient ainsi 50 observations $x_1, x_2, ..., x_{50}$, par exemple :

6	2	5	1	3	2	2	3	1	3
5	4	1	3	3	2	1	2	4	2
2	1	4	2	1	1	2	3	1	1
7	4	2	1	5	1	2	3	1	3
2	4	3	8	2	4	3	3	5	2

Sur base de ces observations, en utilisant l'estimateur $\overline{X}_n$, une estimation — que nous noterons $\widehat{m}$ — de la fréquentation moyenne m de la bibliothèque par les étudiants de la Faculté est :

$$\widehat{m} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{50} \times (6 + 2 + 5 + ... + 5 + 2) = 2,76$$

En utilisant l'estimateur Σ_n^2, une estimation — que nous noterons $\widehat{\sigma}^2$ — de la variance σ^2 de cette fréquentation est :

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{m})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \widehat{m}^2 \\ &= \frac{1}{50} \times \left(6^2 + 2^2 + 5^2 + ... + 5^2 + 2^2\right) - 2,76^2 = 2,6224 \end{aligned}$$

tandis que si on utilise l'estimateur S_n^2, une estimation — que nous noterons $\widehat{s}^2$ — de σ^2 est :

$$\begin{aligned}\widehat{s}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\widehat{m})^2 = \frac{1}{n-1}\sum_{i=1}^{n}x_i^2 - \frac{n}{n-1}\widehat{m}^2 \\ &= \frac{1}{49}\times\left(6^2+2^2+5^2+...+5^2+2^2\right) - \frac{50}{49}\times 2,76^2 = 2,6759...\end{aligned}$$

Finalement, en utilisant l'estimateur F_n, une estimation — que nous noterons $\widehat{f}$ — de la proportion p des étudiants qui fréquentent la bibliothèque au moins cinq fois par semaine est :

$$\widehat{f} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{x_i\geq 5} = \frac{1}{50}\times(1+0+1+...+1+0) = 0,14$$

où $\mathbb{I}_{x_i\geq 5}$ est une variable indicatrice qui est égale à 1 si $x_i \geq 5$ et égale à 0 sinon.

10.1.3. Qualité des estimateurs $\overline{X}_n$, Σ_n^2, S_n^2 et F_n

Que valent les règles de décision que constituent les estimateurs $\overline{X}_n$, Σ_n^2, S_n^2 et F_n ? La réponse à cette question nécessite que l'on définissent des critères de qualité d'un estimateur.

La qualité d'un estimateur $T_n = g(X_1, X_2, ..., X_n)$ se juge sur base des caractéristiques de sa distribution d'échantillonnage : plus cette distribution est "proche" du paramètre d'intérêt θ, meilleur est l'estimateur. En effet, plus cette distribution est "proche" de θ, plus les chances d'obtenir une estimation $\widehat{\theta} = g(x_1, x_2, ..., x_n)$ — qui est une réalisation particulière de T_n — éloignée de θ est faible, et donc plus la procédure d'estimation est fiable.

10.1.3.1. Absence sans biais

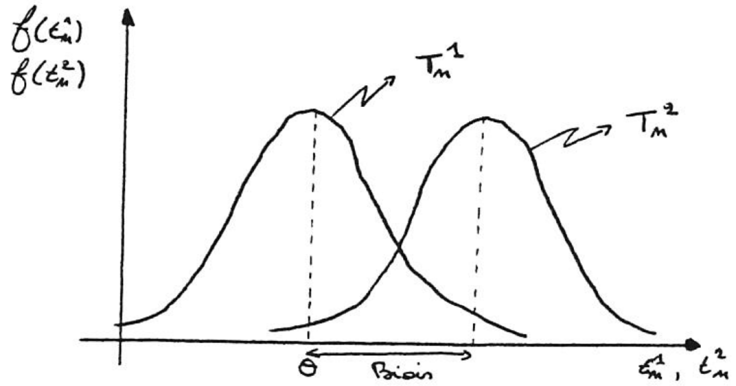
Dans cette perspective, une première propriété souhaitable pour un estimateur est que son espérance soit égale au paramètre d'intérêt θ. Cela revient à demander que la distribution d'échantillonnage de l'estimateur soit centrée en θ, en d'autres termes que les valeurs prises par l'estimateur ne soient pas systématiquement "à côté" de θ. On dit d'un tel estimateur qu'il est *sans biais*, ou encore *non biaisé.*

Définition 10.3 *Soient θ un paramètre de la loi d'une variable aléatoire X et $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à cette variable aléatoire. On appelle estimateur sans biais (ou non biaisé) du paramètre θ un estimateur T_n tel que :*

$$E(T_n) = \theta$$

Lorsque l'estimateur T_n est biaisé, on appelle biais de l'estimateur la quantité :

$$Biais(T_n) = E(T_n) - \theta$$



Graphique 10.2 : Estimateur sans biais (T_n^1) et estimateur biaisé (T_n^2)

On a vu au Chapitre 9 (cf. Propriétés 9.1, 9.12 et 9.16) que $E(\overline{X}_n) = m$, $E(S_n^2) = \sigma^2$ et $E(F_n) = p$. La moyenne-échantillon $\overline{X}_n$, la variance-échantillon modifiée S_n^2 et la fréquence-échantillon F_n sont donc des estimateurs non biaisés de respectivement la moyenne m, la variance σ^2 et la fréquence p dans la population.

Propriété 10.1 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. $\overline{X}_n$ et S_n^2 sont des estimateurs sans biais de respectivement m et σ^2.*

Propriété 10.2 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{B}(p)$. F_n est un estimateur sans biais de p.*

Par contre, la variance-échantillon Σ_n^2 n'est pas un estimateur sans biais de la variance σ^2 dans la population. On a effet vu que (cf. Propriété 9.11) :

$$E(\Sigma_n^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

Le biais de cet estimateur vaut :

$$Biais(\Sigma_n^2) = E(\Sigma_n^2) - \sigma^2 = \left(1 - \frac{1}{n}\right)\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$$

En d'autres termes, l'estimateur Σ_n^2 tend à systématiquement sous-évaluer σ^2. Remarquons cependant que son biais est petit et qu'il tend vers zéro lorsque n tend vers l'infini :

$$\lim_{n\to\infty} Biais(\Sigma_n^2) = 0 \quad \Leftrightarrow \quad \lim_{n\to\infty} E(\Sigma_n^2) = \sigma^2$$

On dit d'un estimateur tel que Σ_n^2 dont le biais tend vers zéro lorsque n tend vers l'infini qu'il est *asymptotiquement sans biais*.

Définition 10.4 *Soient θ un paramètre de la loi d'une variable aléatoire X et $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à cette variable aléatoire. On ap-*

pelle estimateur asymptotiquement sans biais (ou asymptotiquement non biaisé) du paramètre θ un estimateur T_n tel que :

$$\lim_{n\to\infty} E(T_n) = \theta$$

Un estimateur sans biais est évidemment asymptotiquement sans biais.

Propriété 10.3 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. Σ_n^2 est un estimateur asymptotiquement sans biais de σ^2.*

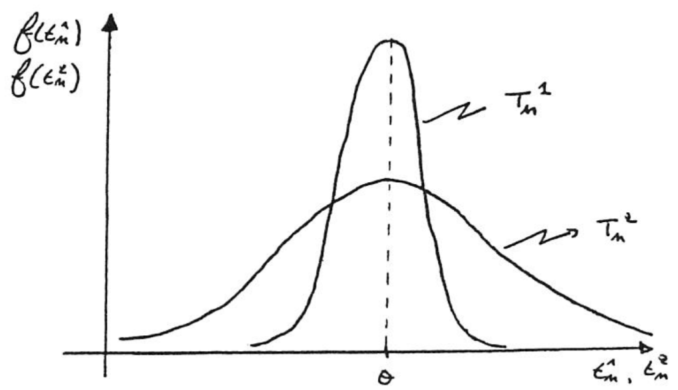
Etant non biaisé, l'estimateur S_n^2 de la variance σ^2 dans la population est généralement préféré à l'estimateur Σ_n^2, qui n'est lui qu'asymptotiquerment non biaisé. Notons cependant qu'en pratique, sauf si n est très petit, l'utilisation de l'un ou l'autre de ces estimateurs ne fait guère beaucoup de différences (cf. Section 9.3.5 et l'exemple numérique ci-dessus).

10.1.3.2. Efficacité

L'absence de biais ne traduit qu'un aspect de la "proximité" entre la distribution d'échantillonnage d'un estimateur et la valeur du paramètre d'intérêt θ. Un autre aspect crucial est la variabilité, la dispersion de l'estimateur. Il est clair qu'entre deux estimateurs non biaisés, on préférera celui dont la variance est la plus faible. On dit de cet estimateur qu'il est *plus efficace.*

Définition 10.5 *Soient θ un paramètre de la loi d'une variable aléatoire X, $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à cette variable aléatoire et T_n^1, T_n^2 deux estimateurs sans biais du paramètre θ. On dit que l'estimateur sans biais T_n^1 est plus efficace que l'estimateur sans biais T_n^2 si :*

$$V(T_n^1) \leq V(T_n^2)$$



Graphique 10.3 : L'estimateur T_n^1 est plus efficace que l'estimateur T_n^2

Illustrons cette définition par un exemple. Considérons comme estimateur de la moyenne m dans la population la statistique :

$$\Gamma_n = \frac{X_1 + X_n}{2}$$

c'est-à-dire un estimateur qui évalue m en se servant uniquement la première et la dernière observation de l'échantillon. En utilisant la Propriété 5.15 et le fait que pour tout i, $E(X_i) = m$, on vérifie aisément que cet estimateur est non biaisé :

$$E(\Gamma_n) = \frac{1}{2}\left(E(X_1) + E(X_n)\right) = \frac{1}{2}(m + m) = m$$

Ne se servant que 2 observations sur n, on s'attend à ce que cet estimateur soit moins efficace que la moyenne-échantillon $\overline{X}_n$, c'est-à-dire à ce que $V(\Gamma_n) \geq V\left(\overline{X}_n\right)$. C'est bien le cas. En effet, $X_1, X_2, ..., X_n$ étant indépendantes, de la Propriété 5.16 et du fait que pour tout i, $V(X_i) = \sigma^2$, on trouve pour la variance de Γ_n :

$$V(\Gamma_n) = \frac{1}{4}\left(V(X_1) + V(X_n)\right) = \frac{1}{4}\left(\sigma^2 + \sigma^2\right) = \frac{\sigma^2}{2}$$

On a donc bien (pour tout $n \geq 2$) :

$$V(\Gamma_n) = \frac{\sigma^2}{2} \geq \frac{\sigma^2}{n} = V\left(\overline{X}_n\right)$$

Il existe généralement de nombreux estimateurs sans biais d'un même paramètre θ. Il est dès lors naturel de se demander si l'un de ces estimateurs n'est pas le plus efficace parmi un sous-ensemble voire la totalité des estimateurs sans biais possibles de ce paramètre.

De ce point de vue, on peut établir que les estimateurs sans biais $\overline{X}_n$, S_n^2 et F_n ont des propriétés remarquables.

On peut ainsi montrer que, quel que soit la loi de X, la moyenne-échantillon $\overline{X}_n$ est un estimateur sans biais de m plus efficace que tout autre estimateur linéaire sans biais de m, c'est-à-dire que tout autre estimateur de la forme :

$$Q_n = \sum_{i=1}^{n} a_i X_i$$

où les a_i $(i = 1, ..., n)$ sont des constantes telles que $\sum_{i=1}^{n} a_i = 1$ (cette dernière condition assurant que $E(Q_n) = m$). Notez $\overline{X}_n$ correspond au cas particulier de Q_n où les $a_i = \frac{1}{n}$ $(i = 1, ..., n)$ et l'estimateur Γ_n définit ci-dessus au cas particulier de Q_n où $a_1 = a_n = \frac{1}{2}$ et $a_2, ..., a_{n-1}$ sont égaux à zéro. On dit que $\overline{X}_n$ est le *meilleur estimateur linéaire sans biais* de la moyenne m dans la population. On peut également montrer que, pour certaines loi de X, en particulier la loi normale et la loi de Poisson (et encore d'autres lois), la moyenne-échantillon $\overline{X}_n$ est un estimateur sans biais de m plus efficace que tout autre estimateur — linéaire au non

— sans biais de m. On dit que pour ces lois $\overline{X}_n$ est le *meilleur estimateur sans biais* de la moyenne m dans la population.

Propriété 10.4 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. $\overline{X}_n$ est le meilleur estimateur linéaire sans biais de m. Pour certaines lois de X (en particulier la loi normale et la loi de* Poisson*), il est le meilleur estimateur sans biais de m.*

La variance-échantillon modifiée S_n^2 possède une propriété du même type, quoique de portée plus limitée. On peut ainsi montrer à son propos que si la loi de la population est normale, S_n^2 est un estimateur sans biais de σ^2 plus efficace que tout autre estimateur — linéaire au non — sans biais de σ^2. On dit que, sous l'hypothèse de normalité, S_n^2 est le *meilleur estimateur sans biais* de la variance σ^2 dans la population.

Propriété 10.5 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. Si la loi de X est normale ($X \rightsquigarrow \mathcal{N}(m; \sigma^2)$), S_n^2 est le meilleur estimateur sans biais de σ^2.*

Enfin, on peut encore montrer que la fréquence-échantillon F_n est un estimateur sans biais de p plus efficace que tout autre estimateur — linéaire ou non — sans biais de p. On dit que F_n est le *meilleur estimateur sans biais* de la fréquence p dans la population.

Propriété 10.6 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{B}(p)$. F_n est le meilleur estimateur sans biais de p.*

Notez bien que toutes ces propriétés sont des propriétés en échantillon fini, valables quelle que soit la taille d'échantillon n.

10.1.3.3. Convergence

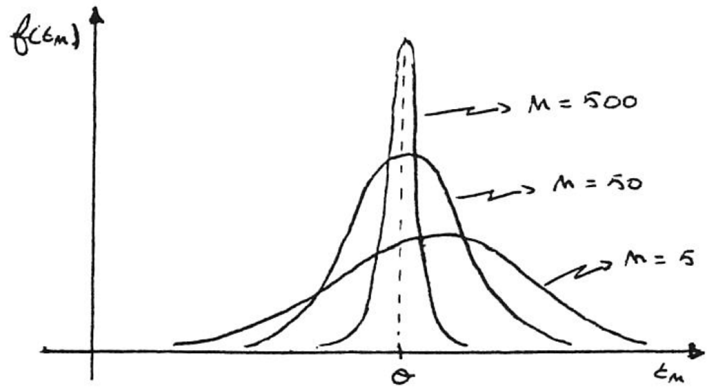
La qualité d'un estimateur se jaugeant à sa capacité à délivrer préférentiellement des valeurs proches du paramètre θ que l'on cherche à estimer, il est naturel d'exiger que la probabilité qu'il en soit bien ainsi s'accroisse avec la taille d'échantillon n. En d'autres termes, il est naturel d'exiger que sa distribution tende à se concentrer de plus en plus sur θ à mesure que n augmente. On dit d'un tel estimateur qu'il est *convergent.*

Définition 10.6 *Soient θ un paramètre de la loi d'une variable aléatoire X et $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à cette variable aléatoire. On appelle estimateur convergent du paramètre θ un estimateur T_n tel que :*

$$T_n \xrightarrow{P} \theta$$

c'est-à-dire tel que :

$$\textit{quel que soit } \epsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}(|T_n - \theta| < \varepsilon) = 1$$



Graphique 10.4 : L'estimateur T_n est convergent

Notons qu'un estimateur convergent peut être biaisé. Néanmoins, la convergence assure que le biais devient de moins en moins important lorsque la taille de l'échantillon grandit.

La convergence est une propriété incontournable d'un estimateur : un estimateur non convergent est un estimateur sans intérêt. Ainsi, bien qu'il soit non biaisé, l'estimateur $\Gamma_n = (X_1 + X_n)/2$ discuté à la section précédente est sans aucun intérêt, certes parce qu'il est moins efficace que $\overline{X}_n$ mais plus fondamentalement parce qu'il n'est pas convergent (sa variance ne diminue pas avec n, elle est constante).

On a vu au Chapitre 9 (cf. Propriétés 9.9, 9.13 et 9.17) que $\overline{X}_n \xrightarrow{P} m$, $\Sigma_n^2 \xrightarrow{P} \sigma^2$, $S_n^2 \xrightarrow{P} \sigma^2$ et $F_n \xrightarrow{P} p$. La moyenne-échantillon $\overline{X}_n$, la variance-échantillon Σ_n^2 et sa version modifiée S_n^2, et la fréquence-échantillon F_n sont donc des estimateurs convergents de respectivement la moyenne m, la variance σ^2 et la fréquence p dans la population.

Propriété 10.7 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2. $\overline{X}_n$, Σ_n^2 et S_n^2 sont des estimateurs convergents de respectivement m et, pour les deux derniers, σ^2.*

Propriété 10.8 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{B}(p)$. F_n est un estimateur convergent de p.*

10.2. Intervalles de confiance

10.2.1. La notion d'intervalle de confiance

Un estimateur T_n délivre une estimation ponctuelle $\widehat{\theta}$ — un nombre — du paramètre d'intérêt inconnu θ.

De façon à rendre compte de la fiabilité, de la précision de l'estimation ponctuelle

$\widehat{\theta}$ du paramètre inconnu θ, on peut lui associer un intervalle de valeurs $[\widehat{c}_1, \widehat{c}_2]$ dont on peut assurer, sur base de la règle de décision suivant laquelle il est calculé, avec un haut niveau de confiance $1 - \alpha$ (traditionnellement 95%) choisi à l'avance qu'il contient la valeur inconnue du paramètre θ. Plus cette intervalle sera petit, plus on pourra considérer comme fiable l'estimation $\widehat{\theta}$ du paramètre inconnu θ.

La règle de décision, que l'on dérive de la distribution d'échantillonnage de l'estimateur T_n, décrivant la façon dont, quelle que soit la réalisation observée $(x_1, x_2, ..., x_n)$ d'un échantillon aléatoire $(X_1, X_2, ..., X_n)$, on calcule cet intervalle est appelée un *intervalle de confiance de niveau* $1 - \alpha$ pour le paramètre θ (ou encore, un *estimateur par intervalle de niveau de confiance* $1 - \alpha$ du paramètre θ). On en représente la forme et le résultat de l'application — qui forcément varie d'un échantillon à l'autre — par un intervalle aléatoire $[C_{1_n}; C_{2_n}]$, où C_{1_n} et C_{2_n} sont deux statistiques.

Définition 10.7 *Soient θ un paramètre de la loi d'une variable aléatoire X et $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à cette variable aléatoire. On appelle intervalle de confiance de niveau $1 - \alpha$ pour le paramètre θ un intervalle aléatoire $[C_{1_n}; C_{2_n}]$ où C_{1_n} et C_{2_n} sont deux statistiques $C_{1_n} = g_1(X_1, X_2, ..., X_n)$ et $C_{2_n} = g_2(X_1, X_2, ..., X_n)$ telles que :*

$$I\!P(C_{1n} \leq \theta \leq C_{2_n}) = 1 - \alpha \tag{10.1}$$

Le résultat de l'application à un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier de la règle de décision définie par l'intervalle de confiance $[C_{1_n}; C_{2_n}]$ donne l'intervalle de valeurs $[\widehat{c}_1, \widehat{c}_2]$, où $\widehat{c}_1 = g_1(x_1, x_2, ..., x_n)$ et $\widehat{c}_2 = g_2(x_1, x_2, ..., x_n)$, qui constitue une réalisation particulière, pour cet échantillon d'observations, de l'intervalle de confiance $[C_{1_n}; C_{2_n}]$.

Notez que dans la pratique, on constate une certaine confusion entre *intervalle de confiance* et *réalisation de l'intervalle de confiance* (qu'on qualifie aussi, d'*intervalle d'estimation* par référence au terme *estimateur par intervalle*) : le même terme "intervalle de confiance" est généralement utilisé pour désigner à la fois la règle de décision décrite par l'intervalle aléatoire $[C_{1_n}; C_{2_n}]$ et une réalisation particulière (donc des nombres) $[\widehat{c}_1, \widehat{c}_2]$ de cet intervalle.

La relation (10.1) nous indique qu'un intervalle de confiance de niveau $1 - \alpha$ est tel qu'il y a une probabilité $1 - \alpha$ que l'intervalle aléatoire $[C_{1_n}; C_{2_n}]$ prenne un couple de valeurs qui recouvre la valeur inconnue (et qui le restera !) du paramètre θ. Autrement dit, il y a a priori une probabilité $1-\alpha$ que l'intervalle $[\widehat{c}_1, \widehat{c}_2]$ calculé pour un échantillon d'observations quelconque, c'est-à-dire une réalisation particulière de $[C_{1_n}; C_{2_n}]$, recouvre la valeur inconnue du paramètre θ. En termes d'échantillonnage répété, cela signifie que si on tirait un grand nombre de d'échantillons et que l'on calculait pour chacun d'eux l'intervalle de valeurs $[\widehat{c}_1, \widehat{c}_2]$ — intervalles qui seront tous différents —, à peu près $100 \times (1 - \alpha)$% des intervalles calculés recouvraient effectivement la valeur inconnue du paramètre de θ. D'où le niveau de confiance $1-\alpha$ que l'on peut avoir quant au fait que l'intervalle $[\widehat{c}_1, \widehat{c}_2]$ calculé pour l'échantillon d'observations dont on dispose recouvre effectivement θ.

Généralement, on considère des intervalles de confiance à risque symétrique, c'est-à dire tels que :

$$\mathbb{P}(C_{1_n} > \theta) = \mathbb{P}(C_{2_n} < \theta) = \frac{\alpha}{2}$$

Dans la suite, on montre comment construire et calculer de tels intervalles de confiance pour la moyenne m, la variance σ^2 et l'écart-type σ (sous l'hypothèse de normalité seulement), et la fréquence p dans une population.

10.2.2. Intervalle de confiance pour la moyenne

10.2.2.1. Préliminaire : cas où σ^2 est connu

Supposons qu'on considère une population représentée par la loi d'une variable aléatoire X d'espérance m inconnue et de variance σ^2 que, dans un premier temps et pour des raisons pédagogiques — nous relâcherons cette hypothèse irréaliste ci-dessous —, l'on présume connue. Disposant d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ — une réalisation particulière d'un échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à X —, on estime la moyenne m dans la population en utilisant l'estimateur sans biais et convergent (et à variance minimale parmi les estimateurs linéaires sans biais, dans certains cas parmi tous les estimateurs sans biais) $\overline{X}_n$ et on obtient une estimation — une réalisation particulière de $\overline{X}_n$ — $\widehat{m}$. On cherche à rendre compte de la fiabilité, de la précision de l'estimation ponctuelle $\widehat{m}$ au travers d'un intervalle de confiance de niveau $1-\alpha$ à risque symétrique pour le paramètre m.

Du théorème central limite (cf. Propriété 9.10) on sait que, quelle que soit la loi de X, la distribution d'échantillonnage de $\overline{X}_n$ est asymptotiquement normale :

$$\frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \xrightarrow{L} \mathcal{N}(0;1)$$

En grand échantillon, on a donc approximativement :

$$\frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \approx \mathcal{N}(0;1) \quad \Leftrightarrow \quad \overline{X}_n \approx \mathcal{N}(m; \tfrac{\sigma^2}{n})$$

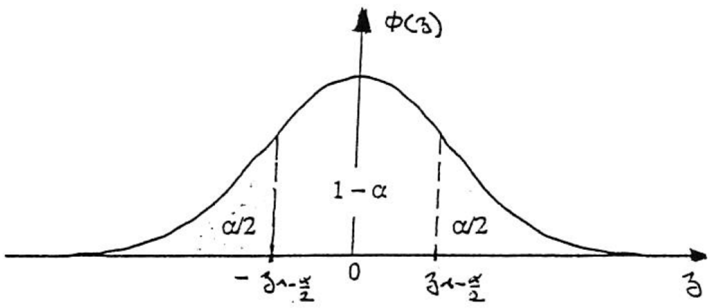
Cette approximation est (généralement) raisonnable si $n \geq 30$.

Sur base de cette distribution d'échantillonnage approximative, on peut déterminer un intervalle de confiance, forcément lui aussi approximatif, de niveau $1-\alpha$ pour le paramètre m.

Comme l'illustre le Graphique 10.5, pour $Z \rightsquigarrow \mathcal{N}(0;1)$, on a :

$$\begin{aligned} & I\!P\big(Z \geq z_{1-\frac{\alpha}{2}}\big) = I\!P\big(Z \leq -z_{1-\frac{\alpha}{2}}\big) = \frac{\alpha}{2} \\ \Leftrightarrow \quad & I\!P\big(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\big) = 1-\alpha \qquad (10.2) \end{aligned}$$

où $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite $\mathcal{N}(0;1)$. Pour $\alpha = 0,05$ et $\alpha = 0,01$, on trouve dans les tables respectivement $z_{0,975} = 1,96$ et $z_{0,995} = 2,5758$.



Graphique 10.5 : Le quantile $z_{1-\frac{\alpha}{2}}$ de la loi $\mathcal{N}(0;1)$

En appliquant, à titre d'approximation, la relation (10.2) à la moyenne-échantillon $\overline{X}_n$ centrée réduite, on obtient :

$$\begin{aligned} & I\!P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \\ \Leftrightarrow \quad & I\!P\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq \overline{X}_n - m \leq z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha \\ \Leftrightarrow \quad & I\!P\left(z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \geq m - \overline{X}_n \geq -z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha \\ \Leftrightarrow \quad & I\!P\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq m - \overline{X}_n \leq z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha \\ \Leftrightarrow \quad & I\!P\left(\overline{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq m \leq \overline{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha \qquad (10.3) \end{aligned}$$

Les bornes de la relation (10.3) donne un intervalle de confiance $[C_{1_n}; C_{2_n}]$ approximatif à risque symétrique de niveau $1-\alpha$ pour le paramètre m :

$$\left[\overline{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}; \overline{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right]$$

Une réalisation particulière de cet intervalle de confiance, obtenue pour un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier, est donnée par l'intervalle $[\widehat{c}_1; \widehat{c}_2]$:

$$\left[\widehat{m} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}; \widehat{m} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right]$$

Propriété 10.9 *Soit $X_1, X_2, ..., X_n$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m inconnue et de variance σ^2 connue. Pour n grand ($n \geq 30$), on a approximativement :*

$$I\!P\left(\overline{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq m \leq \overline{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1 - \alpha \qquad (10.4)$$

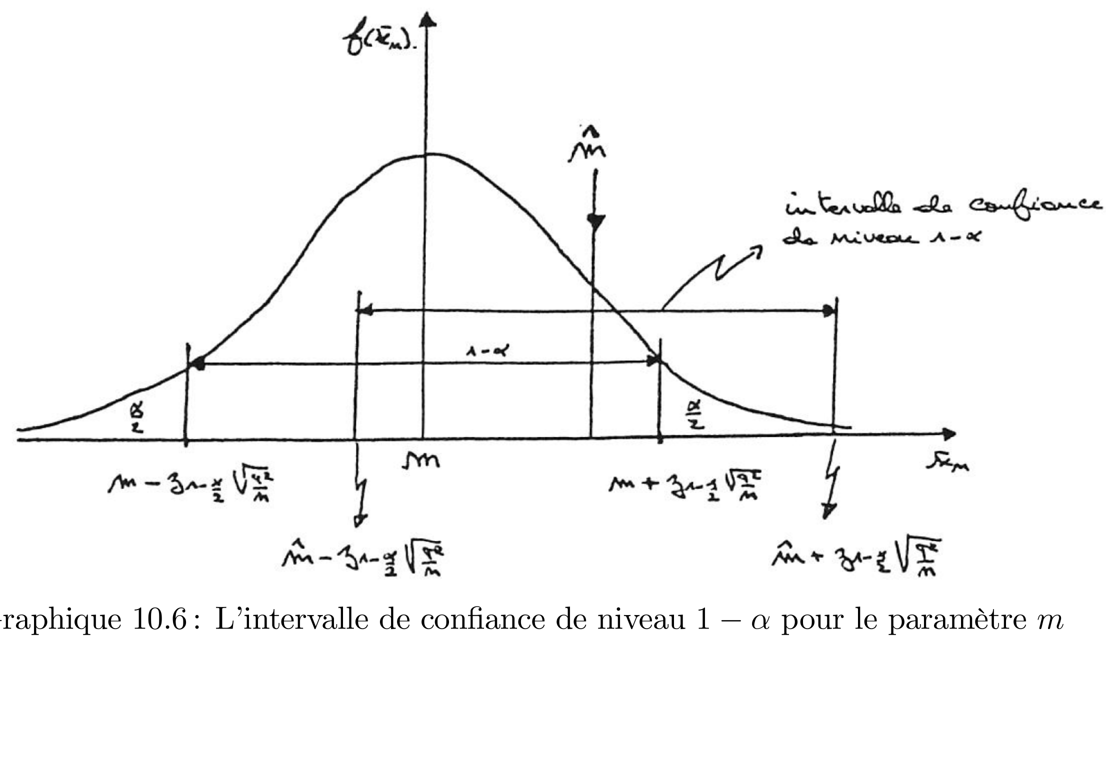
de sorte qu'à une estimation $\widehat{m}$ on peut associer un intervalle de confiance approximatif de niveau $1 - \alpha$ pour le paramètre m donné par :

$$\left[\widehat{m} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}; \widehat{m} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right] \qquad (10.5)$$

Notez que ce résultat est valable quelle que soit la loi de X.

On constate que l'intervalle de confiance est centré en $\widehat{m}$: ses bornes sont obtenues en ajoutant et en retranchant de l'estimation $\widehat{m}$ la quantité $z_{1-\frac{\alpha}{2}}\sqrt{\sigma^2/n}$. Notez que la quantité $\sqrt{\sigma^2/n}$ n'est autre que l'écart-type $\sigma(\overline{X}_n)$ de l'estimateur $\overline{X}_n$.

La relation (10.4) nous garantit qu'il y a a priori (approximativement) une probabilité $1 - \alpha$ que l'intervalle (10.5) calculé pour un échantillon d'observations (de taille $n \geq 30$) quelconque recouvre la valeur inconnue du paramètre m. Le Graphique 10.6 permet de visualiser pourquoi il en est bien ainsi. Il illustre également au passage la "logique d'inversion" de l'information contenue dans la distribution d'échantillonnage qui sous-tend la construction de tout intervalle de confiance.



Graphique 10.6 : L'intervalle de confiance de niveau $1 - \alpha$ pour le paramètre m

La courbe en cloche représente la distribution d'échantillonnage de $\overline{X}_n$ qui, pour n suffisamment grand, est donc approximativement normale $\mathcal{N}(m; \frac{\sigma^2}{n})$. Il y a (approximativement) une probabilité $1-\alpha$ que $\overline{X}_n$ prenne sa valeur — c'est-à-dire qu'on obtienne une estimation $\widehat{m}$ — dans l'intervalle :

$$\left[m - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}; m + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right] \tag{10.6}$$

En effet, sur base de l'approximation normale, on a :

$$\begin{aligned}
& I\!\!P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \\
\Leftrightarrow\quad & I\!\!P\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq \overline{X}_n - m \leq z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha \\
\Leftrightarrow\quad & I\!\!P\left(m - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}} \leq \overline{X}_n \leq m + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right) = 1-\alpha
\end{aligned}$$

Pour une estimation $\widehat{m}$ — une réalisation particulière de $\overline{X}_n$, l'intervalle de confiance de niveau $1-\alpha$ pour m est donné par l'intervalle (10.5). Les intervalles (10.5) et (10.6) ayant la même longueur, on voit que pour toute estimation $\widehat{m}$ qui tombe dans l'intervalle (10.6), l'intervalle (10.5) recouvrira m. Comme il y a a priori (approximativement) $100 \times (1-\alpha)$ chances sur 100 que $\widehat{m}$ tombe dans l'intervalle (10.6), il y a également a priori (approximativement) $100 \times (1-\alpha)$ chances sur 100 que l'intervalle (10.5) recouvre la valeur inconnue du paramètre de m. D'où le niveau de confiance $1-\alpha$ que l'on peut avoir quant au fait que l'intervalle (10.5) calculé pour l'échantillon d'observations dont on dispose recouvre effectivement m.

Pour un niveau de confiance $1-\alpha$ donné, la longueur de l'intervalle de confiance pour m reflète la fiabilité, la précision de l'estimation : il sera d'autant plus court que l'estimation est précise. La précision d'estimation découle directement de la plus ou moins grande variabilité de l'estimateur $\overline{X}_n$ de m : si la taille d'échantillon n est grande et / ou la variance σ^2 dans la population est petite, l'écart-type $\sigma(\overline{X}_n) = \sqrt{\sigma^2/n}$ de l'estimateur $\overline{X}_n$ sera petit, donc l'estimation précise et l'intervalle de confiance court. Au contraire, si la taille d'échantillon n est petite et / ou la variance σ^2 dans la population est grande, l'écart-type $\sigma(\overline{X}_n) = \sqrt{\sigma^2/n}$ de l'estimateur $\overline{X}_n$ sera grand, donc l'estimation peu précise et la longueur de l'intervalle de confiance importante.

Pour un écart-type de $\overline{X}_n$ donné, la valeur du quantile $z_{1-\frac{\alpha}{2}}$ étant une fonction décroissante de α et donc croissante de $1-\alpha$ (cf. Graphique 10.5), l'intervalle de confiance sera d'autant plus long que le niveau de confiance $1-\alpha$ exigé est grand. α mesure le risque que l'intervalle ne recouvre pas la valeur inconnue de m. Plus on choisit ce risque α petit, plus le niveau de confiance $1-\alpha$ que l'on peut avoir dans le fait que l'intervalle recouvre effectivement la valeur inconnue m est grand, et pour assurer ce niveau de confiance, plus l'intervalle de confiance doit être large.

10.2.2.2. Cas où σ^2 est inconnu

Généralement, la variance σ^2 dans la population est inconnue. L'intervalle de confiance de niveau $1-\alpha$ pour le paramètre m :

$$\left[\widehat{m} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}; \widehat{m} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right]$$

ne peut donc en pratique être calculé puisqu'il dépend de la valeur inconnue de σ^2.

Pour obtenir un intervalle de confiance de niveau $1-\alpha$ pour le paramètre m qui est en pratique calculable, on peut procéder de la façon suivante. On sait que, quel que soit la loi de X, on a :

$$\frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \xrightarrow{L} \mathcal{N}(0;1)$$

On sait également que la variance-échantillon modifiée S_n^2 est une estimateur convergent de la variance σ^2 dans la population : $S_n^2 \xrightarrow{P} \sigma^2$. On peut intuitivement comprendre que, comme $S_n^2 \xrightarrow{P} \sigma^2$, on a :

$$\frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} - \frac{\overline{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}} \xrightarrow{P} 0$$

autrement dit que, lorsque n tend vers l'infini, les quantités $\left(\overline{X}_n - m\right)/\sqrt{S_n^2/n}$ et $\left(\overline{X}_n - m\right)/\sqrt{\sigma^2/n}$ deviennent (avec une probabilité qui tend vers 1) indiscernables. Etant indiscernables, elles ont forcément la même distribution limite. On a donc aussi :

$$\frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \xrightarrow{L} \mathcal{N}(0;1)$$

c'est-à-dire, en grand échantillon, approximativement :

$$\frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \approx \mathcal{N}(0;1) \tag{10.7}$$

L'approximation (10.7) est “doublement approximative” : elle repose d'une part sur le théorème central limite (première approximation), et d'autre part sur la convergence vers σ^2 de l'estimateur S_n^2 (deuxième approximation). Elle n'est évidemment raisonnable que si n est suffisamment grand. Nous continuerons à supposer, comme on le fait généralement, que cela est le cas pour $n \geq 30$.

En utilisant comme précédemment, à titre d'approximation, le fait que pour

$Z \rightsquigarrow \mathcal{N}(0;1)$:

$$\begin{aligned} &\mathbb{P}\left(Z \geq z_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(Z \leq -z_{1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \\ \Leftrightarrow \quad &\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \end{aligned}$$

on obtient :

$$\begin{aligned} &\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \\ \Leftrightarrow \quad &\mathbb{P}\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \leq \overline{X}_n - m \leq z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha \\ \Leftrightarrow \quad &\mathbb{P}\left(z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \geq m - \overline{X}_n \geq -z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha \\ \Leftrightarrow \quad &\mathbb{P}\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \leq m - \overline{X}_n \leq z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha \\ \Leftrightarrow \quad &\mathbb{P}\left(\overline{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \leq m \leq \overline{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha \end{aligned}$$

c'est-à-dire un intervalle de confiance $[C_{1_n}; C_{2_n}]$ approximatif de niveau $1-\alpha$ à risque symétrique pour le paramètre m donné par :

$$\left[\overline{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}; \overline{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right]$$

dont une réalisation particulière, obtenue pour un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier, est donnée par l'intervalle $[\widehat{c}_1; \widehat{c}_2]$:

$$\left[\widehat{m} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}}; \widehat{m} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}}\right]$$

Propriété 10.10 *Soit $X_1, X_2, ..., X_n$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2 inconnues. Pour n grand ($n \geq 30$), on a approximativement :*

$$\mathbb{P}\left(\overline{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \leq m \leq \overline{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha \tag{10.8}$$

de sorte qu'à une estimation $\widehat{m}$ on peut associer un intervalle de confiance approximatif de niveau $1-\alpha$ pour le paramètre m donné par :

$$\left[\widehat{m} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}}; \widehat{m} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}}\right] \tag{10.9}$$

Remarquez que, comme la Propriété 10.9, la Propriété 10.10 est valable quelle que soit la loi de X.

La relation (10.8) nous garantit qu'il y a a priori (approximativement) une probabilité $1-\alpha$ que l'intervalle (10.9) calculé pour un échantillon d'observations (de taille $n \geq 30$) quelconque recouvre la valeur inconnue du paramètre m. D'où le niveau de confiance $1-\alpha$ que l'on peut avoir quant au fait que l'intervalle (10.9) calculé pour l'échantillon d'observations dont on dispose recouvre effectivement m.

On voit que l'intervalle de confiance (10.9) n'est guère très différent de l'intervalle de confiance obtenu pour le cas où σ^2 est connu. Comme ce dernier, il est centré en $\widehat{m}$ et, pour un niveau de confiance $1-\alpha$ donné, sa longueur — qui reflète la fiabilité, la précision de l'estimation — est reliée à la plus ou moins grande variabilité de l'estimateur $\overline{X}_n$ de m. Simplement, ici, cette longueur ne dépend plus directement de l'écart-type théorique $\sigma(\overline{X}_n) = \sqrt{\sigma^2/n}$ de l'estimateur $\overline{X}_n$ mais d'une estimation $(= \sqrt{\widehat{s}^2/n})$ de cet écart-type. Pour $\sqrt{\widehat{s}^2/n}$ donné, on a évidemment toujours que l'intervalle de confiance sera d'autant plus long que le niveau de confiance $1-\alpha$ exigé est grand.

Si on a une idée de l'ordre de grandeur de la variance σ^2 dans la population (par exemple sur base des résultats d'une étude antérieure), on peut déterminer à l'avance la taille d'échantillon n approximative qu'il est nécessaire de considérer pour atteindre une précision d'estimation donnée. Ainsi, si l'on désigne l'ordre de grandeur de σ^2 dans la population par σ^{2*} et que l'on souhaite obtenir, avec un niveau de confiance $1-\alpha$, une estimation de m précise à $\pm\ \delta$ près, c'est-à-dire un intervalle de confiance pour m de demi-longueur égale à δ, on choisira n tel que :

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma^{2*}}{n}} = \delta \quad \Leftrightarrow \quad n = \left(z_{1-\frac{\alpha}{2}}\right)^2 \frac{\sigma^{2*}}{\delta^2}$$

Si l'évaluation σ^{2*} de la variance σ^2 est bonne, $\widehat{s}^2$ sera proche de σ^{2*} et la demi-longueur de l'intervalle de confiance pour m proche de la valeur souhaitée δ.

10.2.3. Exercice résolu

Dans un échantillon aléatoire de 50 nouveau-nés, on a observé un poids moyen à la naissance $\widehat{m} = 3,37$ kilos, avec un écart-type modifié[17] $\widehat{s} = 280$ grammes.

1- Donnez un intervalle de confiance de niveau 95% pour le poids moyen m des nouveau-nés.

   Notons X le poids à la naissance d'un nouveau-né. La loi de X (= la loi de la population) est inconnue. Le poids moyen m auquel on s'intéresse correspond à l'espérance de X. Comme $n = 50 \geq 30$, on peut utiliser la Propriété 10.10. De cette propriété, un intervalle de confiance de niveau $1-\alpha$ pour m est donné par :

$$\left[\widehat{m} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}} ; \widehat{m} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}}\right]$$

[17] Dans la suite, lorsque nous parlerons de l'écart-type modifié $\widehat{s} = \sqrt{\widehat{s}^2}$ et de la variance modifiée $\widehat{s}^2$, nous omettrons le qualificatif “modifié” et nous dirons simplement l'écart-type $\widehat{s}$ et la variance $\widehat{s}^2$.

Pour $\alpha = 0,05$, on trouve dans les tables $z_{0,975} = 1,96$. L'intervalle demandé est donc (en kg) :

$$\left[3,37 - 1,96 \times \frac{0,28}{\sqrt{50}}; 3,37 + 1,96 \times \frac{0,28}{\sqrt{50}}\right]$$

$$\approx [3,37 - 0,077; 3,37 + 0,077] = [3,293; 3,447]$$

2- Que devient cet intervalle si on souhaite un niveau de confiance de 99% ?

Pour $\alpha = 0,01$, on trouve dans les tables $z_{0,995} = 2,5758$. L'intervalle devient :

$$\left[3,37 - 2,5758 \times \frac{0,28}{\sqrt{50}}; 3,37 + 2,5758 \times \frac{0,28}{\sqrt{50}}\right]$$

$$\approx [3,37 - 0,102; 3,37 + 0,102] = [3,268; 3,472]$$

3- Sur base de l'information disponible, évaluer la taille d'échantillon n qu'il serait nécessaire de considérer pour obtenir, avec un niveau de confiance de 99%, une estimation de m à précise $\pm$ 50 grammes près.

On cherche la taille d'échantillon n pour laquelle la demi-longueur de l'intervalle de confiance de niveau 99% sera égale à $0,05$ (kg). Sur base l'estimation ci-dessus, une évaluation σ^{2*} de la variance σ^2 dans la population est donnée par $\sigma^{2*} = \widehat{s}^2 = 0,28^2 = 0,0784$. Il conviendrait donc de considérer une taille d'échantillon n donnée par la valeur de n qui est telle que :

$$z_{0,995}\sqrt{\frac{\sigma^{2*}}{n}} = 0,05 \quad \Leftrightarrow \quad n = (z_{0,995})^2 \frac{\sigma^{2*}}{0,05^2} = (2,5758)^2 \times \frac{0,0784}{0,05^2} \approx 208,06$$

Sur base de cette évaluation, pour atteindre la précision souhaitée, il faudrait donc considérer une taille d'échantillon $n \geq 209$.

10.2.4. Intervalle de confiance pour une fréquence

On suppose à présent qu'on considère une population composée d'une certaine proportion inconnue p d'individus possédant une caractéristique donnée, population qu'on représente par une variable aléatoire X suivant une loi de Bernoulli $\mathcal{B}(p)$, où $p = \mathbb{P}(X = 1)$ correspond à la fréquence dans la population des individus qui possèdent la caractéristique en question. Disposant d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ — une réalisation particulière d'un échantillon aléatoire $(X_1, X_2,..., X_n)$ relatif à X —, on estime la fréquence p dans la population en utilisant l'estimateur sans biais, convergent et à variance minimale F_n et on obtient une estimation — une réalisation particulière de F_n — $\widehat{f}$. Comme dans le cas de la moyenne, on cherche à rendre compte de la fiabilité, de la précision de l'estimation ponctuelle $\widehat{f}$ au travers d'un intervalle de confiance de niveau $1 - \alpha$ à risque symétrique pour le paramètre p.

On sait (cf. Propriété 9.18) que la distribution d'échantillonnage de l'estimateur F_n est asymptotiquement normale :

$$\frac{F_n - p}{\sqrt{\frac{pq}{n}}} \xrightarrow{L} \mathcal{N}(0;1)$$

La fréquence-échantillon F_n étant un estimateur convergent de p, on conçoit bien que $\Gamma_n = F_n(1-F_n)$ soit une estimateur convergent de $pq = p(1-p)$ (qui n'est autre que la variance de X puisque pour $X \rightsquigarrow \mathcal{B}(p)$, on a $V(X) = pq$) : $\Gamma_n \xrightarrow{P} pq$. Suivant le même raisonnement que celui développé à la Section 10.2.2.2, du fait que $\Gamma_n = F_n(1-F_n) \xrightarrow{P} pq$, on déduit que :

$$\frac{F_n - p}{\sqrt{\frac{F_n(1-F_n)}{n}}} \xrightarrow{L} \mathcal{N}(0;1)$$

En grand échantillon, on a donc approximativement :

$$\frac{F_n - p}{\sqrt{\frac{F_n(1-F_n)}{n}}} \approx \mathcal{N}(0;1) \qquad (10.10)$$

Comme l'approximation (10.7) de la Section 10.2.2.2, l'approximation (10.10) “doublement approximative” : elle repose d'une part sur le théorème central limite appliqué à la fréquence-échantillon (première approximation) et d'autre part sur la convergence vers pq de l'estimateur $\Gamma_n = F_n(1-F_n)$ (deuxième approximation). Nous considérerons, comme il est d'usage de le faire, que cette approximation est raisonnable si n et p sont tels que : $n \geq 30$, $np \geq 5$ et $nq \geq 5$.

En utilisant à nouveau, à titre d'approximation, le fait que pour $Z \rightsquigarrow \mathcal{N}(0;1)$:

$$\begin{aligned} & I\!P\left(Z \geq z_{1-\frac{\alpha}{2}}\right) = I\!P\left(Z \leq -z_{1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \\ \Leftrightarrow \quad & I\!P\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \end{aligned}$$

on obtient :

$$\begin{aligned} & I\!P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{F_n - p}{\sqrt{\frac{F_n(1-F_n)}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \\ \Leftrightarrow \quad & I\!P\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{F_n(1-F_n)}{n}} \leq F_n - p \leq z_{1-\frac{\alpha}{2}}\sqrt{\frac{F_n(1-F_n)}{n}}\right) = 1-\alpha \\ \Leftrightarrow \quad & I\!P\left(z_{1-\frac{\alpha}{2}}\sqrt{\frac{F_n(1-F_n)}{n}} \geq p - F_n \geq -z_{1-\frac{\alpha}{2}}\sqrt{\frac{F_n(1-F_n)}{n}}\right) = 1-\alpha \end{aligned}$$

$$\Leftrightarrow \quad I\!P\left(-z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{F_n(1-F_n)}{n}} \leq p - F_n \leq z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{F_n(1-F_n)}{n}}\right) = 1-\alpha$$

$$\Leftrightarrow \quad I\!P\left(F_n - z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{F_n(1-F_n)}{n}} \leq p \leq F_n + z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{F_n(1-F_n)}{n}}\right) = 1-\alpha$$

c'est-à-dire un intervalle de confiance $[C_{1_n}; C_{2_n}]$ approximatif de niveau $1-\alpha$ à risque symétrique pour le paramètre p donné par :

$$\left[F_n - z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{F_n(1-F_n)}{n}}; F_n + z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{F_n(1-F_n)}{n}}\right]$$

dont une réalisation particulière, obtenue pour un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier, est donnée par l'intervalle $[\widehat{c}_1; \widehat{c}_2]$:

$$\left[\widehat{f} - z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{f}(1-\widehat{f})}{n}}; \widehat{f} + z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{f}(1-\widehat{f})}{n}}\right]$$

Propriété 10.11 *Soit $X_1, X_2, ..., X_n$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{B}(p)$, où p est inconnu. Pour n grand et p ni trop petit ni trop grand ($n \geq 30$, $np \geq 5$ et $nq \geq 5$), on a approximativement :*

$$I\!P\left(F_n - z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{F_n(1-F_n)}{n}} \leq p \leq F_n + z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{F_n(1-F_n)}{n}}\right) = 1-\alpha \qquad (10.11)$$

de sorte qu'à une estimation $\widehat{f}$ on peut associer un intervalle de confiance approximatif de niveau $1-\alpha$ pour le paramètre p donné par :

$$\left[\widehat{f} - z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{f}(1-\widehat{f})}{n}}; \widehat{f} + z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{f}(1-\widehat{f})}{n}}\right] \qquad (10.12)$$

Les conditions d'applicabilité de ce résultat ($n \geq 30$, $np \geq 5$ et $nq \geq 5$) sont celles de l'approximation (10.10) sur laquelle il est basé. Au sens strict, il est impossible de vérifier si les deux dernières conditions ($np \geq 5$ et $nq = n(1-p) \geq 5$) sont bien remplies puisque la valeur de p est inconnue. Nous pouvons cependant regarder si elles le sont pour l'estimation ponctuelle $\widehat{f}$ de p, et c'est ce qu'on fait dans la pratique.

La relation (10.11) nous garantit qu'il y a a priori (approximativement) une probabilité $1-\alpha$ que l'intervalle (10.12) calculé pour un échantillon d'observations (tel que $n \geq 30$, $np \geq 5$ et $nq \geq 5$) quelconque recouvre la valeur inconnue du paramètre p. D'où le niveau de confiance $1-\alpha$ que l'on peut avoir quant au fait que l'intervalle (10.12) calculé pour l'échantillon d'observations dont on dispose recouvre effectivement p.

On constate que l'intervalle de confiance (10.12) pour le paramètre p est centré en l'estimation ponctuelle $\widehat{f}$ et que sa demi-longueur $z_{1-\frac{\alpha}{2}}\sqrt{\widehat{f}(1-\widehat{f})/n}$ est proportionnelle à $\sqrt{\widehat{f}(1-\widehat{f})/n}$, quantité qui n'est autre qu'une estimation de l'écart-type théorique $\sigma(F_n) = \sqrt{p(1-p)/n}$ de l'estimateur F_n, écart-type théorique qui dépend lui-même de la taille d'échantillon n et de la fréquence p dans la population. On retrouve ici la même structure que celle de l'intervalle de confiance pour la moyenne m (avec σ^2 inconnu) dans la population.

Pour un niveau de confiance $1-\alpha$ donné, la longueur de l'intervalle de confiance pour p sera d'autant plus courte que l'estimation est fiable, précise, la précision de l'estimation étant reflétée par $\sqrt{\widehat{f}(1-\widehat{f})/n}$. Pour $\sqrt{\widehat{f}(1-\widehat{f})/n}$ donné, on obtient un intervalle de confiance d'autant plus long que le niveau de confiance $1-\alpha$ exigé est grand.

Comme dans le cas de la moyenne, si on a une idée de l'ordre de grandeur de la fréquence p dans la population (par exemple sur base des résultats d'une étude antérieure), on peut déterminer à l'avance la taille d'échantillon n approximative qu'il est nécessaire de considérer pour atteindre une précision d'estimation donnée. Ainsi, si l'on désigne l'ordre de grandeur de p dans la population par p^* et que l'on souhaite obtenir, avec un niveau de confiance $1-\alpha$, une estimation de p précise à $\pm\ \delta$ près, c'est-à-dire un intervalle de confiance pour p de demi-longueur égale à δ, on choisira n tel que :

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{p^*(1-p^*)}{n}} = \delta \quad \Leftrightarrow \quad n = \left(z_{1-\frac{\alpha}{2}}\right)^2 \frac{p^*(1-p^*)}{\delta^2}$$

Si l'évaluation p^* de la fréquence p est bonne, $\widehat{f}(1-\widehat{f})$ sera proche de $p^*(1-p^*)$ et la demi-longueur de l'intervalle de confiance pour p proche de la valeur souhaitée δ. Si on n'a aucune idée de la fréquence p dans la population, on prendra pour n la valeur :

$$n = \left(z_{1-\frac{\alpha}{2}}\right)^2 \frac{0,25}{\delta^2} \tag{10.13}$$

ce qui revient à prendre $p^* = 0,5$. Cela est justifié par le fait que la valeur maximum possible de $\widehat{f}(1-\widehat{f})$ est atteinte pour $\widehat{f} = 0,5$. Dès lors, en choisissant n comme indiqué par (10.13), on est assuré que, quel que soit l'estimation $\widehat{f}$ obtenue, la demi-longueur de l'intervalle de confiance pour p sera au maximum égale à δ.

10.2.5. Exercice résolu

Un sondage est réalisé en vue de prévoir le résultat d'une élection présidentielle. On effectue un tirage aléatoire de 250 électeurs dans la population. Parmi ceux-ci, 108 déclarent qu'ils ont l'intention de voter pour le président sortant tandis que les 142 autres électeurs affirment qu'ils vont voter pour l'un des deux autres candidats.

1- Donnez une estimation de la proportion d'électeurs qui vont voter pour le président sortant lors de l'élection.

   On représente la population par une loi de Bernoulli $\mathcal{B}(p)$, où p est la proportion (inconnue) d'électeurs qui vont voter pour le président sortant lors de l'élection :

$X \rightsquigarrow \mathcal{B}(p)$. L'utilisation de l'estimateur F_n nous fournit l'estimation ponctuelle :

$$\widehat{f} = \frac{108}{250} = 0,432$$

2- Donnez des intervalles de confiance de niveau 95% et 99% pour cette proportion.

Comme $n = 250 \geq 30$, $n\widehat{f} = 108 \geq 5$ et $n(1-\widehat{f}) = 142 \geq 5$, on peut utiliser la Propriété 10.11. De cette propriété, un intervalle de confiance de niveau $1-\alpha$ pour p est donné par :

$$\left[\widehat{f} - z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{f}(1-\widehat{f})}{n}}; \widehat{f} + z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{f}(1-\widehat{f})}{n}}\right]$$

Pour $\alpha = 0,05$, on trouve dans les tables $z_{0,975} = 1,96$. L'intervalle à 95% est donc :

$$\left[0,432 - 1,96 \times \sqrt{\tfrac{0,432\times 0,568}{250}}; 0,432 - 1,96 \times \sqrt{\tfrac{0,432\times 0,568}{250}}\right]$$

$$\approx [0,432 - 0,0614; 0,432 + 0,0614] = [0,3706; 0,4934]$$

Notez que cet intervalle prédit, avec niveau de confiance de 95%, que le président sortant ne sera pas réélu.
Pour $\alpha = 0,01$, on trouve dans les tables $z_{0,995} = 2,5758$. L'intervalle à 99% est donc :

$$\left[0,432 - 2,5758 \times \sqrt{\tfrac{0,432\times 0,568}{250}}; 0,432 + 2,5758 \times \sqrt{\tfrac{0,432\times 0,568}{250}}\right]$$

$$\approx [0,432 - 0,0806; 0,432 + 0,0806] = [0,3514; 0,5126]$$

Si on est plus prudent et qu'on exige un niveau de confiance de 99%, on voit que l'intervalle "laisse une chance" au président sortant.

3- Mis au courant des résultats de ce sondage et jugeant l'estimation obtenue trop peu précise, le président sortant commande à une société spécialisée un nouveau sondage. Sur combien d'électeurs devraient porter ce nouveau sondage pour obtenir, avec un niveau de confiance de 95%, une estimation de son score lors de l'élection précise à ± 3% près ?

On cherche la taille d'échantillon n pour laquelle la demi-longueur de l'intervalle de confiance de niveau 95% sera égale à $0,03$. Sur base l'estimation ci-dessus, une évaluation p^* de la fréquence p dans la population (des électeurs) est donnée par $p^* = \widehat{f} = 0,432$. Il conviendrait donc de considérer une taille d'échantillon n donnée par la valeur de n qui est telle que :

$$z_{0,975}\sqrt{\frac{p^*(1-p^*)}{n}} = 0,03 \quad \Leftrightarrow \quad n = (z_{0,975})^2\,\frac{p^*(1-p^*)}{(0,03)^2}$$

$$= (1,96)^2 \times \frac{0,432 \times 0,568}{(0,03)^2} \approx 1047,4$$

Sur base de cette évaluation, pour atteindre la précision souhaitée, il faudrait donc considérer une taille d'échantillon $n \geq 1048$.
Si on veut être sûr d'atteindre la précision de souhaitée, il faut prendre $p^* = 0,5$, ce qui donne pour n :

$$n = (z_{0,975})^2 \frac{0,5 \times (1-0,5)}{(0,03)^2} = (1,96)^2 \times \frac{0,25}{(0,03)^2} \approx 1067,1$$

Pour une taille d'échantillon $n \geq 1068$, on est sûr de toujours atteindre la précision souhaitée.

10.2.6. Cas d'une population normale : intervalle de confiance pour la moyenne, la variance et l'écart-type

On considère pour terminer le cas spécifique où la population peut être représentée par une loi normale d'espérance m et de variance σ^2 inconnues : $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$. Disposant d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ — une réalisation particulière d'un échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à X —, on estime la moyenne m et la variance σ^2 dans la population en utilisant les estimateurs sans biais, convergents et efficaces $\overline{X}_n$ et S_n^2, et on obtient des estimations $\widehat{m}$ et $\widehat{s}^2$ — des réalisations particulières de $\overline{X}_n$ et S_n^2 — des paramètres inconnus m et σ^2.

Comme nous allons le voir ci-dessous, sous l'hypothèse de normalité, on peut établir un intervalle de confiance exact — valable quel que soit la taille d'échantillon n (et non plus seulement pour n grand) — pour la moyenne m dans la population. On peut également déterminer un intervalle de confiance lui aussi exact pour la variance σ^2 dans la population, intervalle dont on peut déduire un intervalle de confiance à nouveau exact pour l'écart-type $\sigma = \sqrt{\sigma^2}$ dans la population.

Notons qu'en s'appuyant sur la distribution asymptotique de S_n^2 donnée au Chapitre 9 (cf. Propriété 9.14), on peut très bien, à l'image de ce que nous avons fait ci-dessus pour la moyenne m, établir un intervalle de confiance approximatif pour σ^2 — intervalle dont on peut déduire un intervalle de confiance approximatif pour σ —, valable en grand échantillon et quel que soit la loi de X. Dans le cadre de ce cours, on se contentera, pour la variance σ^2 et l'écart-type σ dans la population, de traiter le cas normal.

10.2.6.1. Intervalle de confiance pour la moyenne

A la Section 10.2.2.2, nous avons déterminé un intervalle de confiance approximatif pour m en nous appuyant sur le fait que, pour n grand et quel que soit la loi de X, on a approximativement :

$$\frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \approx \mathcal{N}(0; 1) \tag{10.14}$$

Lorsque la loi de la population est normale $\mathcal{N}(m;\sigma^2)$, on dispose d'un résultat plus précis. On a en effet vu (cf. Propriété 9.22) que, sous l'hypothèse de normalité, quel que soit n, on a de façon exacte :

$$\frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \rightsquigarrow t(n-1) \tag{10.15}$$

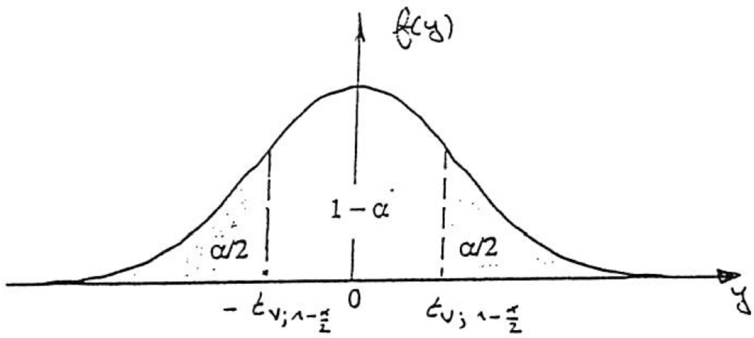
Remarquez que pour n tendant vers l'infini, la loi STUDENT $t(n-1)$ tend vers une loi normale standardisée $\mathcal{N}(0;1)$ (cf. Propriété 9.7), de sorte que l'on retrouve le résultat asymptotique (10.14).

En s'appuyant sur la distribution d'échantillonnage exacte (10.15), on peut établir un intervalle de confiance exact de niveau $1-\alpha$ pour le paramètre m.

Comme l'illustre le Graphique 10.7, pour $Y \rightsquigarrow t(\nu)$, on a :

$$\begin{aligned} &\mathbb{P}\left(Y \geq t_{\nu;1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(Y \leq -t_{\nu;1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \\ \Leftrightarrow \quad &\mathbb{P}\left(-t_{\nu;1-\frac{\alpha}{2}} \leq Y \leq t_{\nu;1-\frac{\alpha}{2}}\right) = 1-\alpha \end{aligned} \tag{10.16}$$

où $t_{\nu;1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi de STUDENT à ν degrés de liberté $t(\nu)$. Pour par exemple $\alpha = 0,05$ et $\nu = 5$, on trouve dans les tables $t_{5;0,975} = 2,571$.



Graphique 10.7 : Le quantile $t_{\nu;1-\frac{\alpha}{2}}$ de la loi $t(\nu)$

Des relations (10.15) et (10.16), on déduit :

$$\mathbb{P}\left(-t_{n-1;1-\frac{\alpha}{2}} \leq \frac{\overline{X}_n - m}{\sqrt{\frac{S_n^2}{n}}} \leq t_{n-1;1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Leftrightarrow \quad \mathbb{P}\left(-t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \leq \overline{X}_n - m \leq t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha$$

$$\Leftrightarrow \quad \mathbb{P}\left(t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \geq m - \overline{X}_n \geq -t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha$$

$$\Leftrightarrow \quad I\!P\left(-t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \leq m - \overline{X}_n \leq t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha$$

$$\Leftrightarrow \quad I\!P\left(\overline{X}_n - t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \leq m \leq \overline{X}_n + t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha$$

c'est-à-dire un intervalle de confiance $[C_{1_n}; C_{2_n}]$ exact de niveau $1-\alpha$ à risque symétrique pour le paramètre m donné par :

$$\left[\overline{X}_n - t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}; \overline{X}_n + t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right]$$

dont une réalisation particulière, obtenue pour un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier, est donnée par l'intervalle $[\widehat{c}_1; \widehat{c}_2]$:

$$\left[\widehat{m} - t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}}; \widehat{m} + t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}}\right]$$

Propriété 10.12 *Soit $X_1, X_2, ..., X_n$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$, où m et σ^2 sont inconnus. On a :*

$$I\!P\left(\overline{X}_n - t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}} \leq m \leq \overline{X}_n + t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{S_n^2}{n}}\right) = 1-\alpha \tag{10.17}$$

de sorte qu'à une estimation $\widehat{m}$ on peut associer un intervalle de confiance exact de niveau $1-\alpha$ pour le paramètre m donné par :

$$\left[\widehat{m} - t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}}; \widehat{m} + t_{n-1;1-\frac{\alpha}{2}}\sqrt{\frac{\widehat{s}^2}{n}}\right] \tag{10.18}$$

Ne s'appuyant sur aucune approximation, ce résultat est exact, valable quelle que soit la taille n de l'échantillon.

La relation (10.17) nous garantit qu'il y a a priori (exactement) une probabilité $1-\alpha$ que l'intervalle (10.18) calculé pour un échantillon d'observations quelconque recouvre la valeur inconnue du paramètre m. D'où le niveau de confiance $1-\alpha$ que l'on peut avoir quant au fait que l'intervalle (10.18) calculé pour l'échantillon d'observations dont on dispose recouvre effectivement m.

L'intervalle de confiance (10.18) a une forme identique à celle n'est celle de l'intervalle de confiance approximatif (10.9) obtenu à la Section 10.2.2.2. Algébriquement, il s'en différentie uniquement par le quantile $t_{n-1;1-\frac{\alpha}{2}}$ qui remplace le quantile $z_{1-\frac{\alpha}{2}}$. Comme son homologue approximatif, il est centré en $\widehat{m}$ et, pour un niveau de confiance $1-\alpha$ donné, sa longueur — qui reflète la fiabilité, la précision de l'estimation — est reliée à la plus ou moins grande variabilité estimée ($= \sqrt{\widehat{s}^2/n}$) de l'estimateur $\overline{X}_n$ de m. Pour $\sqrt{\widehat{s}^2/n}$ donné, la valeur du quantile $t_{n-1;1-\frac{\alpha}{2}}$ étant (comme celle du quantile $z_{1-\frac{\alpha}{2}}$) une fonction décroissante de α (cf. Graphique 10.7), l'intervalle de confiance sera à nouveau — comme tout intervalle de confiance — d'autant plus long que le niveau de confiance $1-\alpha$ exigé est grand.

A l'aide de la Propriété 10.12, on peut calculer un intervalle de confiance pour m dans le cas de petit échantillons ($n < 30$), ce que ne nous permettait pas la Propriété 10.9. Evidemment, elle n'offre cette possibilité que pour le cas où la loi de la population est normale.

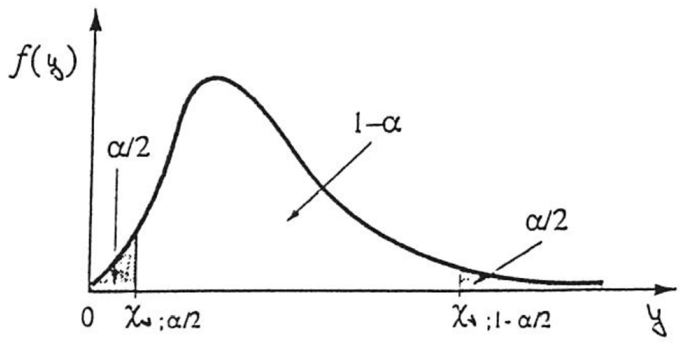
Pour n est grand ($n \geq 30$), en conséquence de la convergence de loi de STUDENT vers la loi normale standardisée, l'intervalle de confiance (10.18) ne se différentie guère de l'intervalle approximatif (10.9) : pour n grand, la loi de STUDENT $t(n-1)$ est très proche de la loi normale standardisée $\mathcal{N}(0;1)$, de sorte que les quantiles $t_{n-1;1-\frac{\alpha}{2}}$ et $z_{1-\frac{\alpha}{2}}$ sont eux-mêmes très proches. Ainsi par exemple, pour $\alpha = 0,05$ et $n = 30$, on a $t_{29;0,975} = 2,042$ et $z_{0,975} = 1,96$. On le voit, la différence entre les quantiles — qui va en s'amenuisant lorsque n augmente — est très faible. Pour n grand, il est donc pour l'essentiel indifférent d'utiliser l'une ou l'autre forme de l'intervalle de confiance pour m. Ainsi, on constate dans la pratique que l'intervalle (10.18) est souvent aussi utilisé pour calculer, lorsque n est grand et quelle que soit la loi de X, un intervalle de confiance pour m.

10.2.6.2. Intervalle de confiance pour la variance

Nous avons vu (cf. Propriété 9.20) que, lorsque la loi de la population est normale $\mathcal{N}(m;\sigma^2)$, quel que soit n, on a de façon exacte :

$$\frac{(n-1)S_n^2}{\sigma^2} \rightsquigarrow \chi^2(n-1) \tag{10.19}$$

De façon semblable au cas de la moyenne ci-dessus, en s'appuyant sur cette distribution d'échantillonnage exacte, on peut obtenir un intervalle de confiance exact de niveau $1-\alpha$ pour le paramètre σ^2.



Graphique 10.8 : Les quantiles $\chi^2_{\nu;\frac{\alpha}{2}}$ et $\chi^2_{\nu;1-\frac{\alpha}{2}}$ de la loi $\chi^2(\nu)$

Comme l'illustre le Graphique 10.8, pour $Y \rightsquigarrow \chi^2(\nu)$, on a :

$$I\!P\left(Y \leq \chi^2_{\nu;\frac{\alpha}{2}}\right) = I\!P\left(Y \geq \chi^2_{\nu;1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

et donc :

$$\mathbb{P}\left(\chi^2_{\nu;\frac{\alpha}{2}} \leq Y \leq \chi^2_{\nu;1-\frac{\alpha}{2}}\right) = 1 - \alpha \tag{10.20}$$

où $\chi^2_{\nu;\frac{\alpha}{2}}$ et $\chi^2_{\nu;1-\frac{\alpha}{2}}$ sont les quantiles d'ordre respectivement $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$ de la loi du khi-carré à ν degrés de liberté $\chi^2(\nu)$. Pour par exemple $\alpha = 0,05$ et $\nu = 10$, on trouve dans les tables $\chi^2_{10;0,025} = 3,25$ et $\chi^2_{10;0,975} = 20,48$.

Des relations (10.19) et (10.20), on obtient :

$$\mathbb{P}\left(\chi^2_{n-1;\frac{\alpha}{2}} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq \chi^2_{n-1;1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\Leftrightarrow \quad \mathbb{P}\left(\frac{1}{\chi^2_{n-1;\frac{\alpha}{2}}} \geq \frac{\sigma^2}{(n-1)S_n^2} \geq \frac{1}{\chi^2_{n-1;1-\frac{\alpha}{2}}}\right) = 1 - \alpha$$

$$\Leftrightarrow \quad \mathbb{P}\left(\frac{1}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \leq \frac{\sigma^2}{(n-1)S_n^2} \leq \frac{1}{\chi^2_{n-1;\frac{\alpha}{2}}}\right) = 1 - \alpha$$

$$\Leftrightarrow \quad \mathbb{P}\left(\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right) = 1 - \alpha$$

c'est-à-dire un intervalle de confiance $[C_{1_n}; C_{2_n}]$ exact de niveau $1 - \alpha$ à risque symétrique pour le paramètre σ^2 donné par :

$$\left[\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}; \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

dont une réalisation particulière, obtenue pour un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier, est donnée par l'intervalle $[\widehat{c}_1; \widehat{c}_2]$:

$$\left[\frac{(n-1)\widehat{s}^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}; \frac{(n-1)\widehat{s}^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$$

Propriété 10.13 *Soit $X_1, X_2, ..., X_n$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$, où m et σ^2 sont inconnus. On a :*

$$\mathbb{P}\left(\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right) = 1 - \alpha \tag{10.21}$$

de sorte qu'à une estimation $\widehat{s}^2$ on peut associer un intervalle de confiance exact de niveau $1 - \alpha$ pour le paramètre σ^2 donné par :

$$\left[\frac{(n-1)\widehat{s}^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}; \frac{(n-1)\widehat{s}^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right] \tag{10.22}$$

Ne s'appuyant, comme dans le cas de la moyenne ci-dessus, sur aucune approximation, ce résultat est exact, valable quelle que soit la taille n de l'échantillon.

Comme à l'accoutumée, la relation (10.21) nous garantit qu'il y a a priori (exactement) une probabilité $1-\alpha$ que l'intervalle (10.22) calculé pour un échantillon d'observations quelconque recouvre la valeur inconnue du paramètre σ^2. D'où le niveau de confiance $1-\alpha$ que l'on peut avoir quant au fait que l'intervalle (10.22) calculé pour l'échantillon d'observations dont on dispose recouvre effectivement σ^2.

L'intervalle de confiance (10.21) a une forme assez différente des intervalles obtenus précédemment. Ainsi, bien qu'à risque symétrique, il n'est pas centré sur l'estimation ponctuelle $\widehat{s}^2$. Cela résulte de la dissymétrie de la loi du khi-carré.

Même si cela n'apparaît pas explicitement dans son expression, pour un niveau de confiance $1-\alpha$ donné, on peut montrer que la longueur de l'intervalle de confiance (10.22) — qui reflète la fiabilité, la précision de l'estimation — est directement reliée à la plus ou moins grande variabilité (estimée) de l'estimateur S_n^2 de σ^2 : plus celle-ci est faible — c'est en particulier le cas plus n est grand —, plus l'intervalle de confiance sera étroit. Pour une précision (estimée) donnée de l'estimateur S_n^2, les quantiles $\chi^2_{n-1;\frac{\alpha}{2}}$ et $\chi^2_{n-1;1-\frac{\alpha}{2}}$ étant des fonctions respectivement croissante et décroissante de α (cf. Graphique 10.8), l'intervalle de confiance sera comme d'habitude d'autant plus long que le niveau de confiance $1-\alpha$ exigé est grand.

Une mise en garde importante : au contraire de l'intervalle de confiance pour la moyenne obtenu à la section précédente qui, pour n grand, est valable — puisque équivalent à l'intervalle approximatif (10.9) — quelle que soit la loi de X, l'intervalle de confiance (10.22) n'est correct que si la loi de la population est normale, et cela même si n est grand.

10.2.6.3. Intervalle de confiance pour l'écart-type

Disposant d'une estimation $\widehat{s}^2$ de la variance σ^2 dans la population, une estimation naturelle de l'écart-type $\sigma = \sqrt{\sigma^2}$ dans la population est $\widehat{s} = \sqrt{\widehat{s}^2}$. Estimer, sur base d'un échantillon d'observations $(x_1, x_2, ..., x_n)$, l'écart-type σ de cette façon signifie utiliser comme estimateur de σ la statistique :

$$S_n = \sqrt{S_n^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}$$

On peut intuitivement comprendre que, comme $S_n^2 \xrightarrow{P} \sigma^2$, on a :

$$S_n \xrightarrow{P} \sigma$$

autrement dit, que S_n est un estimateur convergent l'écart-type σ dans la population. Il n'en n'est cependant pas un estimateur non biaisé (mais bien asymptotiquement non biaisé puisque convergent). En effet, la racine carrée étant une fonction non linéaire, on a :

$$E\left(S_n\right) = E\left(\sqrt{S_n^2}\right) \neq \sqrt{E(S_n^2)} = \sqrt{\sigma^2} = \sigma$$

puisque pour une fonction non linéaire $g(.)$, $E\left[g(X)\right] \neq g\left[E(X)\right]$ (cf. Chapitre 4).

Cette façon d'obtenir un estimateur convergent $\Gamma_n = g(T_n)$ d'une fonction $g(\theta)$ d'un paramètre θ pour lequel on connaît un estimateur convergent T_n est très couramment utilisée.

Disposant d'une intervalle de confiance $[C_{1n}; C_{2_n}]$ de niveau $1-\alpha$ pour le paramètre θ, c'est-à-dire de deux statistiques C_{1n} et C_{2n} telles que :

$$I\!P(C_{1n} \leq \theta \leq C_{2_n}) = 1 - \alpha \tag{10.23}$$

si la fonction $g(.)$ est strictement croissante ou décroissante, on peut aisément déterminer un intervalle de confiance de même niveau pour le paramètre $g(\theta)$. Si $g(.)$ est strictement croissante, un tel intervalle de confiance est donné par l'intervalle $[g(C_{1n}); g(C_{2_n})]$ puisque dans de cas, de (10.23), on a :

$$I\!P\left(g(C_{1n}) \leq g(\theta) \leq g(C_{2_n})\right) = 1 - \alpha$$

Si $g(.)$ est strictement décroissante, il est donné par l'intervalle $[g(C_{2n}); g(C_{1_n})]$ puisque dans de cas, de (10.23), on a :

$$I\!P\left(g(C_{2n}) \leq g(\theta) \leq g(C_{1_n})\right) = 1 - \alpha$$

Appliquons ce raisonnement au cas de l'écart-type σ dans une population distribuée selon une loi normale $\mathcal{N}(m; \sigma^2)$. Nous avons vu ci-dessus que, sous cette hypothèse, on a de façon exacte :

$$I\!P\left(\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right) = 1 - \alpha$$

On en déduit :

$$I\!P\left(S_n\sqrt{\frac{(n-1)}{\chi^2_{n-1;1-\frac{\alpha}{2}}}} \leq \sigma \leq S_n\sqrt{\frac{(n-1)}{\chi^2_{n-1;\frac{\alpha}{2}}}}\right) = 1 - \alpha$$

c'est-à-dire un intervalle de confiance $[C_{1_n}; C_{2_n}]$ exact de niveau $1 - \alpha$ à risque symétrique pour le paramètre σ donné par :

$$\left[S_n\sqrt{\frac{(n-1)}{\chi^2_{n-1;1-\frac{\alpha}{2}}}}; S_n\sqrt{\frac{(n-1)}{\chi^2_{n-1;\frac{\alpha}{2}}}}\right]$$

dont une réalisation particulière, obtenue pour un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier, est donnée par l'intervalle $[\widehat{c}_1; \widehat{c}_2]$:

$$\left[\widehat{s}\sqrt{\frac{(n-1)}{\chi^2_{n-1;1-\frac{\alpha}{2}}}}; \widehat{s}\sqrt{\frac{(n-1)}{\chi^2_{n-1;\frac{\alpha}{2}}}}\right]$$

Propriété 10.14 *Soit $X_1, X_2, ..., X_n$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$, où m et σ^2 sont inconnus. On a :*

$$I\!P\left(S_n\sqrt{\frac{(n-1)}{\chi^2_{n-1;1-\frac{\alpha}{2}}}} \leq \sigma \leq S_n\sqrt{\frac{(n-1)}{\chi^2_{n-1;\frac{\alpha}{2}}}}\right) = 1-\alpha \qquad (10.24)$$

de sorte qu'à une estimation $\widehat{s}$ on peut associer un intervalle de confiance exact de niveau $1-\alpha$ pour le paramètre σ donné par :

$$\left[\widehat{s}\sqrt{\frac{(n-1)}{\chi^2_{n-1;1-\frac{\alpha}{2}}}};\widehat{s}\sqrt{\frac{(n-1)}{\chi^2_{n-1;\frac{\alpha}{2}}}}\right] \qquad (10.25)$$

Ne s'appuyant sur aucune approximation, ce résultat est, comme celui pour σ^2, exact, valable quelle que soit la taille n de l'échantillon.

Comme d'habitude, la relation (10.24) nous garantit qu'il y a a priori (exactement) une probabilité $1-\alpha$ que l'intervalle (10.25) calculé pour un échantillon d'observations quelconque recouvre la valeur inconnue du paramètre σ. D'où toujours le niveau de confiance $1-\alpha$ que l'on peut avoir quant au fait que l'intervalle (10.25) calculé pour l'échantillon d'observations dont on dispose recouvre effectivement σ.

L'intervalle de confiance (10.25), qui à nouveau n'est pas centré en l'estimation ponctuelle $\widehat{s}$, s'interprète de façon semblable à l'intervalle de confiance (10.22) : pour un niveau de confiance $1-\alpha$ donné, il sera d'autant plus court que la variabilité (estimée) de l'estimateur S_n de σ est petite, ce qui est en particulier le cas plus n est grand, et pour une variabilité (estimée) donnée, il sera d'autant plus long que le niveau de confiance $1-\alpha$ exigé est grand.

Pour conclure, insistons sur le fait que, comme l'intervalle de confiance (10.22) dont il est dérivé, l'intervalle de confiance (10.25) n'est valable que sous l'hypothèse de normalité, même pour n grand.

10.2.7. Exercice résolu

On suppose que la concentration en métaux lourds dans le sang des éléphants d'Afrique est distribuée selon une loi normale $\mathcal{N}(m;\sigma^2)$. Dans le cadre d'une étude menée sur un échantillon aléatoire de 8 sujets, on a mesuré une concentration moyenne en métaux lourds ($=\widehat{m}$) égale à 27 microgrammes/litres, avec une variance ($=\widehat{s}^2$) de 16 microgrammes^2/litres^2.

1- Donnez un intervalle de confiance de niveau 90% pour la concentration moyenne en métaux lourds dans le sang des éléphants d'Afrique.

Notons X la concentration en métaux lourds dans le sang d'un éléphant d'Afrique. On a : $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$ (= loi de la population), $\widehat{s}^2 = 16$ et donc $\widehat{s} = 4$. On cherche un intervalle de confiance de niveau 90% pour m. De la Propriété 10.12, un intervalle de confiance (exact) de niveau $1-\alpha$ pour le paramètre m est donné par :

$$\left[\widehat{m} - t_{n-1;1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{s}^2}{n}}; \widehat{m} + t_{n-1;1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{s}^2}{n}}\right]$$

Pour $n = 8$ et $\alpha = 0,1$, on trouve dans les tables $t_{7;0,95} = 1,895$. L'intervalle demandé est donc :

$$\left[27 - 1,895 \times \sqrt{\tfrac{16}{8}}; 106 + 1,895 \times \sqrt{\tfrac{16}{8}}\right]$$

$$\approx [27 - 2,68; 27 + 2,68] = [24,32; 29,68]$$

2- Donnez un intervalle de confiance de niveau 90% pour l'écart-type de la concentration en métaux lourds dans le sang des éléphants d'Afrique.

On cherche maintenant un intervalle de confiance de niveau 90% pour σ. De la Propriété 10.14, un intervalle de confiance (exact) de niveau $1 - \alpha$ pour le paramètre σ est donné par :

$$\left[\widehat{s}\sqrt{\tfrac{(n-1)}{\chi^2_{n-1;1-\frac{\alpha}{2}}}}; \widehat{s}\sqrt{\tfrac{(n-1)}{\chi^2_{n-1;\frac{\alpha}{2}}}}\right]$$

Pour $n = 8$ et $\alpha = 0,1$, on trouve dans les tables $\chi^2_{7;0,05} = 2,17$ et $\chi^2_{7;0,95} = 14,07$. L'intervalle demandé est donc :

$$\left[4 \times \sqrt{\tfrac{7}{14,07}}; 4 \times \sqrt{\tfrac{7}{2,17}}\right] \approx [2,82; 7,18]$$

3- Donnez une estimation de l'intervalle de valeurs centré en m dans lequel se trouve la concentration en métaux lourds dans le sang de 95% des éléphants d'Afrique.

Pour $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$, on a :

$$I\!P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{X - m}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\Leftrightarrow \quad I\!P\left(-z_{1-\frac{\alpha}{2}}\sigma \leq X - m \leq z_{1-\frac{\alpha}{2}}\sigma\right) = 1 - \alpha$$

$$\Leftrightarrow \quad I\!P\left(m - z_{1-\frac{\alpha}{2}}\sigma \leq X \leq m + z_{1-\frac{\alpha}{2}}\sigma\right) = 1 - \alpha$$

Sur base de l'hypothèse de normalité, $100 \times (1-\alpha)$% des éléphants d'Afrique ont donc une concentration de métaux lourds dans le sang ($= X$) dans l'intervalle centré en m :

$$\left[m - z_{1-\frac{\alpha}{2}}\sigma; m + z_{1-\frac{\alpha}{2}}\sigma\right] \tag{10.26}$$

Pour $\alpha = 0,95$, on trouve dans les tables les tables $z_{0,975} = 1,96$. En remplaçant m et σ par leur estimation $\widehat{m}$ et $\widehat{s}$, on obtient l'intervalle de valeurs demandé :

$$[\widehat{m} - 1,96\widehat{s}; \widehat{m} + 1,96\widehat{s}] = [27 - 1,96 \times 4; 27 + 1,96 \times 4] = [19,16; 34,84]$$

Les estimations de m et σ étant relativement peu précises (cf. les intervalles de confiance obtenus ci-dessus), cette estimation de l'intervalle théorique (10.26) est elle-même relativement peu précise.

10.3. Exercice résolu

On considère les notes (sur 10 points) obtenues lors d'un examen cantonal par 50 étudiants prélevés au hasard parmi les milliers d'étudiants qui y ont participé. Ces notes sont reprises dans le tableau suivant, donnant pour chacune des notes x_j $(j = 1, ..., 11)$ le nombre n_j d'étudiants l'ayant obtenue.

x_j	0	1	2	3	4	5	6	7	8	9	10
n_j	5	4	2	4	6	10	8	4	4	2	1

1- Donnez une estimation de la moyenne, de la variance et de l'écart-type des notes obtenues par l'ensemble des étudiants ayant participé à l'examen cantonal.

Sur base de données groupées, on obtient $\widehat{m}$, $\widehat{s}^2$ et $\widehat{s}$ à l'aide des formules :

$$\begin{aligned}
\widehat{m} &= \frac{1}{n}\sum_{j=1}^{J} n_j x_j, \qquad \text{où } n = \sum_{j=1}^{J} n_j \\
\widehat{s}^2 &= \frac{1}{n-1}\sum_{j=1}^{J} n_j\left(x_j - \widehat{m}\right)^2 = \frac{1}{n-1}\sum_{j=1}^{J} n_j x_j^2 - \frac{n}{n-1}\widehat{m}^2
\end{aligned}$$

et évidemment $\widehat{s} = \sqrt{\widehat{s}^2}$. En utilisant les résultats de tableau ci-dessous :

x_j	n_j	$n_j x_j$	$n_j x_j^2$
0	5	0	0
1	4	4	4
2	2	4	8
3	4	12	36
4	6	24	96
5	10	50	250
6	8	48	288
7	4	28	196
8	4	32	256
9	2	18	162
10	1	10	100
Total :	50	230	1396

on trouve :

$$\begin{aligned}
\widehat{m} &= \frac{230}{50} = 4,6 \\
\widehat{s}^2 &= \frac{1396}{49} - \frac{50}{49} \times 4,6^2 \approx 28,489 - 21,591 = 6,898 \\
\widehat{s} &= \sqrt{6,898} \approx 2,6264
\end{aligned}$$

2- <u>Construisez</u> un intervalle de confiance de niveau $0,95$ pour la note moyenne obtenue par l'ensemble des étudiants ayant participé à l'examen cantonal.

Comme la loi de X est inconnue et que $n = 50 \geq 30$, on utilise le résultat approximatif :

$$\frac{\overline{X}_{50} - m}{\sqrt{\frac{S_{50}^2}{50}}} \approx \mathcal{N}(0;1)$$

Pour $Z \rightsquigarrow \mathcal{N}(0;1)$, on trouve en utilisant les tables :

$$I\!P(-1,96 \leq Z \leq 1,96) = 0,95$$

de sorte qu'on a approximativement :

$$I\!P\left(-1,96 \leq \frac{\overline{X}_{50} - m}{\sqrt{\frac{S_{50}^2}{50}}} \leq 1,96\right) = 0,95$$

$$\Leftrightarrow \quad I\!P\left(-1,96\sqrt{\frac{S_{50}^2}{50}} \leq \overline{X}_{50} - m \leq 1,96\sqrt{\frac{S_{50}^2}{50}}\right) = 0,95$$

$$\Leftrightarrow \quad I\!P\left(1,96\sqrt{\frac{S_{50}^2}{50}} \geq m - \overline{X}_{50} \geq -1,96\sqrt{\frac{S_{50}^2}{50}}\right) = 0,95$$

$$\Leftrightarrow \quad I\!P\left(-1,96\sqrt{\frac{S_{50}^2}{50}} \leq m - \overline{X}_{50} \leq 1,96\sqrt{\frac{S_{50}^2}{50}}\right) = 0,95$$

$$\Leftrightarrow \quad I\!P\left(\overline{X}_{50} - 1,96\sqrt{\frac{S_{50}^2}{50}} \leq m \leq \overline{X}_{50} + 1,96\sqrt{\frac{S_{50}^2}{50}}\right) = 0,95$$

Cette relation nous garantit qu'il y a a priori (approximativement) une probabilité de $0,95$ que l'intervalle :

$$\left[\widehat{m} - 1,96\sqrt{\frac{\widehat{s}^2}{50}}; \widehat{m} + 1,96\sqrt{\frac{\widehat{s}^2}{50}}\right]$$

calculé pour un échantillon d'observations (de taille $n \geq 30$) quelconque recouvre la valeur inconnue du paramètre m. Pour notre échantillon, on obtient :

$$\left[4,6 - 1,96 \times \sqrt{\frac{6{,}898}{50}}; 4,6 + 1,96\sqrt{\frac{6{,}898}{50}}\right]$$

$$\approx [4,6 - 0,72; 4,6 + 0,72] = [3,88; 5,32]$$

3- Donnez une estimation de la proportion l'ensemble des étudiants ayant participé à l'examen cantonal qui ont obtenu une note supérieure ou égale à 7.

Sur base de données groupées, on obtient $\widehat{f}$ à l'aide de la formule :

$$\widehat{f} = \frac{1}{n}\sum_{j=1}^{J} n_j \mathbb{I}_{x_j \geq 7}$$

où $n = \sum_{j=1}^{J} n_j$ et $\mathbb{I}_{x_j \geq 7}$ est une variable indicatrice qui est égale à 1 si $x_j \geq 7$ et égale à 0 sinon. On trouve :

$$\widehat{f} = \frac{4+4+2+1}{50} = \frac{11}{50} = 0,22$$

4- Donnez un intervalle de confiance à 99% pour la proportion l'ensemble des étudiants ayant participé à l'examen cantonal qui ont obtenu une note supérieure ou égale à 7.

Comme $n = 50 \geq 30$, $n\widehat{f} = 11 \geq 5$ et $n(1-\widehat{f}) = 39 \geq 5$, on peut utiliser l'intervalle de confiance approximatif de niveau $1-\alpha$ pour p donné par (cf. Propriété 10.11) :

$$\left[\widehat{f} - z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{f}(1-\widehat{f})}{n}}\,;\, \widehat{f} + z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{\widehat{f}(1-\widehat{f})}{n}}\right]$$

Pour $\alpha = 0,01$, on trouve dans les tables $z_{0,995} = 2,5758$. L'intervalle à 99% est donc :

$$\left[0,22 - 2,5758 \times \sqrt{\tfrac{0,22\times 0,78}{50}}\,; 0,22 - 2,5758 \times \sqrt{\tfrac{0,22\times 0,78}{50}}\right]$$

$$\approx [0,22 - 0,151; 0,22 + 0,151] = [0,069; 0,371]$$

10.4. Exercices

1- Supposons que l'on tire un échantillon aléatoire de 100 comptes clients d'une chaîne de grands magasins et que l'on trouve que le solde moyen débiteur de ces 100 comptes est de 74 \$.

a- Sachant que l'écart-type du solde débiteur de l'ensemble des comptes clients est de 86 \$, donnez un intervalle de confiance à 95% pour la moyenne du solde débiteur de l'ensemble des comptes clients.

b- On suppose qu'il y a en tout 243 000 comptes.

i- Donnez une estimation du solde débiteur total de l'ensemble des comptes clients.

ii- Construisez un intervalle de confiance à 95% pour ce solde débiteur total.

Expliquez brièvement au vice-président la signification de vos réponses.

c- Supposons que le vice-président, qui demeure sceptique, entreprenne une vérification complète de tous les soldes débiteurs et qu'il trouve un solde débiteur total de 22 714 000 \$. Que diriez-vous ?

2- Pour déterminer l'âge moyen de ses clients, une grande entreprise de confection pour hommes étudie un échantillon aléatoire de 50 de ses clients et trouve $\widehat{m} = 36$ et $\widehat{s} = 12$.
   - a- Donnez un intervalle de confiance à 95% pour l'âge moyen m de l'ensemble des clients.
   - b- Supposez que pour le même niveau de confiance (95%) on souhaite obtenir une estimation de m à $\pm$ 2 années près. Quelle devrait être approximativement la taille de l'échantillon n ?
   - c- Et si l'on veut une estimation précise à $\pm$ 1 année près ?

3- Le temps de réaction de 30 conducteurs choisis au hasard présente une moyenne $\widehat{m}$ de $0,83$ seconde et une variance $\widehat{s}^2$ de $0,04$ seconde^2.
   - a- Donnez un intervalle de confiance à 99% pour le temps moyen de réaction m de l'ensemble des conducteurs de la population.
   - b- Un conducteur supplémentaire a été choisi au hasard et on constate qu'il a un temps de réaction de $0,98$ seconde. Cette valeur est-elle surprenante ? Expliquez.

4- Le nombre moyen de paquets de cigarettes vendus lundi dernier par un échantillon aléatoire de 120 débits de tabac de la région parisienne a été de $73,21$ paquets, avec un écart-type $9,12$ paquets.
   - a- Construisez un intervalle de confiance de niveau 90% pour le nombre moyen m de paquets de cigarettes vendus par les débits de tabac de la région parisienne lundi dernier.
   - b- On aurait souhaité obtenir avec un niveau de confiance de 99% une estimation de m précise à $\pm$ 1% de m près. Quelle taille d'échantillon approximative aurait-on dû considérer pour atteindre cette précision d'estimation relative ?

5- Lors d'une enquête, on a demandé à un échantillon aléatoire de 500 belges le nombre de fois qu'ils s'étaient rendu au cinéma au cours du dernier mois écoulé. On a obtenu les résultats décrits par le tableau suivant :

x_j	0	1	2	3	4	5
n_j	223	135	69	37	22	14

où n_j $(j = 1, ..., 5)$ désigne le nombre d'individus ayant déclaré s'être rendus x_j fois au cinéma au cours du dernier mois écoulé.
   - a- Donnez une estimation de la moyenne et de l'écart-type du nombre de fois par mois que les belges se rendent au cinéma.
   - b- Donnez un intervalle de confiance à 92% pour la moyenne.
   - c- Donnez une estimation de la proportion des belges qui se rendent plus de deux fois par mois au cinéma et construisez un intervalle de confiance de niveau $0,97$ pour cette proportion.
   - d- On souhaiterait obtenir avec le même niveau de confiance (= 97%) une estimation de cette proportion à $\pm$ $0,01$ près ? Quelle taille d'échantillon

approximative devrait-on considérer ?

6- Lors d'une inspection sanitaire, on découvre 34 bouteilles d'eau minérale impropres à la consommation dans un échantillon aléatoire de 200 issu de la production.
   a- Donnez une estimation de la fréquence des bouteilles impropres à la consommation dans la production totale.
   b- Construisez un intervalle de confiance de niveau 99% pour cette fréquence.

7- On souhaite évaluer la proportion d'électeurs qui approuveront un référendum. Deux sondages aléatoires ont été effectués et ils ont donné les résultats suivants :
   – le 1^{er} octobre, 75 des 130 personnes interrogées approuvaient le référendum.
   – le 15 octobre, 642 des 1056 personnes interrogées approuvaient le référendum.
   a- Pour chaque date :
      i- Donnez une estimation de la proportion des électeurs qui sont en faveur du référendum.
      ii- Donnez un intervalle de confiance à 95% pour cette proportion.
   b- Le scrutin va avoir lieu. On souhaite obtenir, toujours avec le même niveau de confiance de 95%, une nouvelle estimation précise à $\pm$ 1 % près de la proportion d'électeurs qui approuveront le référendum. Quelle taille d'échantillon approximative devrait-on considérer ?

8- Une association de consommateurs a acheté au hasard 500 ampoules électriques d'une certaine marque et a trouvé que 55 d'entre elles étaient défectueuses.
   a- Donnez une estimation de la fréquence des ampoules défectueuses parmi les ampoules de cette marque.
   b- Donnez un intervalle de confiance de niveau 95% pour cette fréquence.
   c- Le fabricant affirme qu'il ne produit pas plus de 10% d'ampoules défectueuses. Qu'en pensez-vous ?

9- La même association a acheté au hasard 400 ampoules d'une marque A et 500 ampoules d'une marque B. Elle a trouvé dans ces lots respectivement 21 et 31 ampoules défectueuses.
   a- En appelant p_A et p_B les proportions respectives d'ampoules défectueuses des deux marques, $F^A_{n_A}$ et $F^B_{n_B}$ les fréquences respectives d'ampoules défectueuses dans des échantillons aléatoires de tailles respectives n_A et n_B, quelle est la loi approximative de la différence $F^A_{n_A} - F^B_{n_B}$?
   b- Donnez une estimation de la différence $p_A - p_B$.
   c- Construisez un intervalle de confiance de niveau 95% pour cette différence.

10- Dans une université américaine, quelques étudiants ont fait un sondage concernant le salaire horaire des étudiants qui travaillent durant l'été. Ils ont tiré au hasard 1000 étudiants et ils ont calculé que leur salaire horaire moyen $\widehat{m}$ l'été dernier était de $6,2$ \$ et que l'écart-type $\widehat{s}$ de leur salaire horaire était de $0,5$ \$. Ils ont également constaté que, parmi les 1000 étudiants interrogés, 560 gagnaient moins de 6 \$ l'heure.

a- Donnez un intervalle de confiance de niveau $0,98$ pour le salaire moyen gagné par les étudiants qui ont travaillé l'été dernier.
b- Quelle taille d'échantillon approximative suffit-il prendre pour s'assurer, toujours avec un niveau de confiance de $0,98$, une estimation de ce salaire moyen à $\pm\ 0,1$ \$ près ?
c- On considère la proportion d'étudiants qui gagnaient moins de 6 \$ l'heure l'été dernier.
   i- Donnez une estimation de cette proportion.
   ii- Donnez un intervalle de confiance de niveau $0,98$ pour cette proportion.
d- On considère la proportion d'étudiants qui gagnaient 6 \$ l'heure et plus l'été dernier.
   i- Donnez une estimation de cette proportion.
   ii- Donnez un intervalle de confiance à 98% pour cette proportion.

11- La teneur en sel (en grammes) d'un produit de régime fabriqué par la société Diet est étudiée par une association de consommateurs. Un échantillon aléatoire de 163 boîtes du produit est analysé et les résultats numériques suivants sont établis :

Teneur observée x_j	Effectif observé n_j	n_jx_j	$n_jx_j^2$
$3,2$	1	$3,2$	$10,24$
$3,3$	4	$13,2$	$43,56$
$3,4$	5	17	$57,8$
$3,5$	6	21	$73,5$
$3,7$	6	$22,2$	$82,14$
$3,8$	4	$15,2$	$57,76$
$3,9$	7	$27,3$	$106,47$
$4,0$	9	36	144
$4,1$	12	$49,2$	$201,72$
$4,2$	17	$71,4$	$299,88$
$4,3$	25	$107,5$	$462,25$
$4,4$	19	$83,6$	$367,84$
$4,5$	13	$58,5$	$263,25$
$4,6$	12	$55,2$	$253,92$
$4,7$	8	$37,6$	$176,72$
$4,8$	7	$33,6$	$161,28$
$4,9$	5	$24,5$	$120,05$
$5,0$	3	15	75
Total :	163	$691,2$	$2957,38$

a- Donnez une estimation de la moyenne m et de l'écart-type σ de la teneur en sel dans les boîtes du produit de régime fabriqué par la société Diet.
b- Donnez un intervalle de confiance à 96% pour m.
c- On s'intéresse au cas où la teneur en sel dépasse strictement $4,5$ grammes parce que cette situation est jugée nocive pour la santé.

i- Donnez une estimation de la proportion de boîtes nocives pour la santé dans la production de la firme.

ii- Construisez un intervalle de confiance de niveau 99% pour cette proportion.

12- Un chef d'une grande entreprise souhaite étudier le trajet que ses ouvriers doivent parcourir pour se rendre de leur domicile à leur travail. Il considère que ce trajet, mesuré en kilomètres, suit une loi normale. Sur la base d'un échantillon aléatoire de 12 ouvriers, il obtient les estimations ponctuelles suivantes : $\widehat{m} = 1,3$ et $\widehat{s} = 0,4$.

a- Construisez un intervalle de confiance de niveau $0,90$ pour la longueur moyenne du trajet des ouvriers de l'entreprise.

b- Construisez un intervalle de confiance de niveau $0,90$ pour la variance et l'écart-type de la longueur de ce trajet.

13- Un syndicat de boulangers fait une enquête sur la consommation mensuelle de pain (en kg) dans un ménage de 4 personnes. Un échantillon aléatoire de 100 ménages est suivi pendant un mois, et on note pour chacun d'eux la quantité totale de pain consommé.

a- On trouve dans l'échantillon une moyenne $\widehat{m} = 21,5$ kg et un écart-type $\widehat{s} = 1,8$ kg. Donnez un intervalle de confiance de niveau 90% pour la consommation mensuelle moyenne m de pain des ménages de 4 personnes.

b- Sur les 100 ménages, 23 ont consommé moins de 20 kg de pain dans le mois.

i- Donnez une estimation de la fréquence des ménages qui consomment mensuellement moins de 20 kg de pain dans la population des ménages de 4 personnes.

ii- Donnez un intervalle de confiance à 90% pour cette fréquence.

c- Un boulanger de quartier fait une enquête analogue après d'un échantillon aléatoire de 11 ménages du quartier. Il trouve également dans son échantillon $\widehat{m} = 21,5$ kg et $\widehat{s} = 1,8$ kg.

i- Au contraire du syndicat de boulangers, pour être en mesure de construire une intervalle de confiance pour m, notre boulanger de quartier doit s'appuyer sur une hypothèse auxiliaire. Pourquoi ? Quelle hypothèse suggérez-vous de faire ?

ii- En supposant cette hypothèse remplie, construisez un intervalle de confiance de niveau 90% pour m.

14- Pour vérifier une machine qui fabrique des bonbons et pour s'assurer que les poids de ceux-ci respectent bien les normes précisées au cahier des charges, on prélève au hasard dans la production 50 bonbons et on les pèse. On suppose que le poids des bonbons produit par cette machine est distribué selon une loi normale. On obtient les poids suivants (en grammes) :

8,1	8,2	8,3	7,9	8,2	7,8	8,1	8,1	8,2	8,4
8,3	8,3	8,0	8,2	8,1	8,1	8,2	8,1	8,3	8,3
8,4	8,4	8,1	8,1	8,1	8,4	8,3	8,1	8,4	8,1
8,1	8,1	8,4	7,9	8,4	7,7	7,9	8,1	8,2	8,4
8,2	8,0	8,5	7,8	7,9	7,8	8,1	8,4	8,2	8,3

a- Estimez le poids moyen, la variance et l'écart-type du poids des bonbons produits par cette machine.
b- Donnez un intervalle de confiance de niveau $0,95$ pour ces trois quantités.
c- Quelle taille d'échantillon approximative devrait-on prendre pour obtenir un intervalle de confiance de niveau $0,95$ pour le poids moyen qui soit deux fois plus court ?

15- Lors d'une étude clinique, on a observé que la durée moyenne de survie d'un échantillon aléatoire de 50 patients atteints d'une grave maladie et traités à l'aide d'une substance A était de $3,8$ années, avec un écart-type de 7 mois. On a par ailleurs observé que durée moyenne de survie d'un autre échantillon aléatoire de 80 patients atteints de la même grave maladie mais traités à l'aide d'une substance B était de $4,3$ années avec un écart-type de 5 mois.

a- En appelant m_A et m_B les durées moyennes de survie des patients atteints de cette grave maladie et traités à l'aide de respectivement la substance A et la substance B, σ_A^2 et σ_B^2 les variances respectives de ces durées de survie, et $\overline{X}_{n_A}^A$ et $\overline{X}_{n_B}^B$ les durées moyennes de survie respectives des patients dans des échantillons aléatoires de tailles respectives n_A et n_B, quelle est la loi approximative de la différence $\overline{X}_{n_A}^A - \overline{X}_{n_B}^B$?
b- Donnez une estimation de la différence $m_A - m_B$.
c- Construisez un intervalle de confiance de niveau 95% pour cette différence.

Chapitre 11

Tests d'hypothèses

Nous avons vu dans le chapitre précédent comment, en exploitant les liens qui unissent caractéristiques d'une population et caractéristiques d'un échantillon aléatoire issu de cette population, on pouvait dégager des procédures permettant, sur base d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ obtenu par échantillonnage aléatoire dans une population, d'estimer la valeur inconnue d'un paramètre θ représentant une certaine caractéristique de la population et de rendre compte, au travers d'un intervalle de confiance, de la précision, de la fiabilité de cette estimation.

Dans ce chapitre, nous allons voir comment, à nouveau en exploitant ces mêmes liens, on peut établir des procédures permettant, toujours sur base d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ obtenu par échantillonnage aléatoire dans une population, de tester si une assertion — une proposition donnée comme vraie — concernant la valeur inconnue d'un paramètre θ représentant une certaine caractéristique de la population est ou non crédible, vraisemblable. On appelle ce type de procédures d'inférence statistique des *tests d'hypothèses*.

Dans le cadre de ce cours, on se concentrera sur l'examen de tests d'hypothèses relatifs au paramètre $m = E(X)$ qui représente la moyenne des valeurs prises par une variable d'intérêt dans une population décrite par la loi d'une variable aléatoire X et au paramètre $p = I\!P(X = 1)$ d'une loi de BERNOULLI $\mathcal{B}(p)$, paramètre qui, on le sait, représente la proportion — la fréquence — des individus d'une population — population qui est décrite par la loi $\mathcal{B}(p)$ — qui possèdent une caractéristique donnée.

On commence par examiner les tests relatifs à une fréquence.

11.1. Tests relatifs à une fréquence

Supposons qu'on considère une population composée d'une certaine proportion inconnue p d'individus possédant une caractéristique donnée, population qu'on représente donc par une variable aléatoire X suivant une loi de BERNOULLI $\mathcal{B}(p)$, où $p = I\!P(X = 1)$ correspond à la fréquence dans la population des individus qui possèdent la caractéristique en question.

Disposant d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ obtenu par échantillonnage aléatoire dans la population, — une réalisation particulière d'un échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à X —, on suppose qu'on souhaite mettre à l'épreuve une assertion concernant la valeur inconnue du paramètre p dans la population, c'est-à-dire tester si cette assertion est ou non compatible avec les observations dont on dispose.

L'assertion qu'on souhaite mettre à l'épreuve des observations est appelée l'*hypothèse nulle* et on la note H_0. L'assertion contraire (complémentaire) à l'hypothèse nulle H_0 — qui sera accréditée si H_0 est jugée incompatible avec les observations — est appelée l'*hypothèse alternative* et est notée H_1. On dit qu'on teste H_0 contre H_1.

Dans les applications pratiques, on est le plus souvent amené à considérer l'une des trois formes suivantes de test :

1- $\mathrm{H}_0 : p = p_0$ contre $\mathrm{H}_1 : p \neq p_0$. On dit d'un test de cette forme qu'il s'agit d'un test *bilatéral.* On l'utilisera par exemple pour tester, sur base des résultats d'un sondage politique, si la proportion inconnue p des électeurs d'un parti dans la population est toujours égale à la proportion p_0 des électeurs qui ont votés pour ce parti lors des dernières élections.

2- $\mathrm{H}_0 : p \leq p_0$ contre $\mathrm{H}_1 : p > p_0$. On qualifie un test de cette forme de test *unilatéral à droite.* C'est par exemple le test que fera, sur base de l'observation d'un échantillon aléatoire de pièces prélevé dans sa production et pour s'assurer qu'il ne s'est pas fait rouler, l'acheteur d'une machine dont le fabriquant affirme que dans des conditions normales d'utilisation le taux (inconnu de l'acheteur) p de pièces défectueuses produites par sa machine n'excède pas un certain taux (faible) p_0.

3- $\mathrm{H}_0 : p \geq p_0$ contre $\mathrm{H}_1 : p < p_0$. Un tel test est appelé un test *unilatéral à gauche.* C'est par exemple le test que mettra en oeuvre une association de consommateurs qui souhaite évaluer, sur base de l'observation d'un échantillon aléatoire de lampes, la crédibilité de l'assertion d'un producteur selon laquelle la proportion (inconnue de l'association) p des lampes qu'il vend ayant une durée de vie supérieure à 5 000 heures n'est pas inférieure à une certaine valeur (élevée) p_0.

Comme dans le cas de l'estimation, qui dit test, sur base d'un échantillon d'observations $(x_1, x_2, ..., x_n)$, d'une hypothèse nulle H_0 contre une hypothèse alternative H_1 dit choix d'une règle de décision décrivant la façon dont, quelle que soit la réalisation observée $(x_1, x_2, ..., x_n)$ d'un échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à X, on va choisir de rejeter ou non H_0.

Evidemment, cette règle de décision, dont l'application à l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose va nous amener à considérer comme ou non crédible l'hypothèse nulle H_0, sera différente selon que l'on considère un test bilatéral, un test unilatéral à droite ou un test unilatéral à gauche.

11.1.1. Test bilatéral

Considérons tout d'abord le cas du test bilatéral :

$$\mathrm{H}_0 : \ p = p_0 \quad \text{contre} \quad \mathrm{H}_1 : \ p \neq p_0$$

Pour évaluer le valeur inconnue du paramètre p dans la population, on dispose de l'estimateur sans biais, convergent et à variance minimale F_n, estimateur dont l'application à un échantillon d'observations $(x_1, x_2, ..., x_n)$ particulier donne une estimation $\widehat{f}$.

Sous l'hypothèse nulle $\mathrm{H}_0 : \ p = p_0$, c'est-à-dire si H_0 est vraie, pour peu que ses conditions de d'applicabilité soient remplies, du théorème central limite appliqué à la fréquence-échantillon F_n (cf. Propriété 9.18), on sait qu'on a approximativement :

$$F_n \approx \mathcal{N}(p_0; \tfrac{p_0 q_0}{n}) \qquad (q_0 = 1 - p_0)$$

ou de façon équivalente :

$$W_n^f = \frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \approx \mathcal{N}(0; 1) \tag{11.1}$$

Sous l'hypothèse alternative $\mathrm{H}_1 : \ p = p_1 \neq p_0$ (que l'on ait $p_1 > p_0$ ou $p_1 < p_0$), c'est-à-dire si H_0 est fausse, on a par contre approximativement :

$$F_n \approx \mathcal{N}(p_1; \tfrac{p_1 q_1}{n}) \qquad (q_1 = 1 - p_1)$$

et donc :

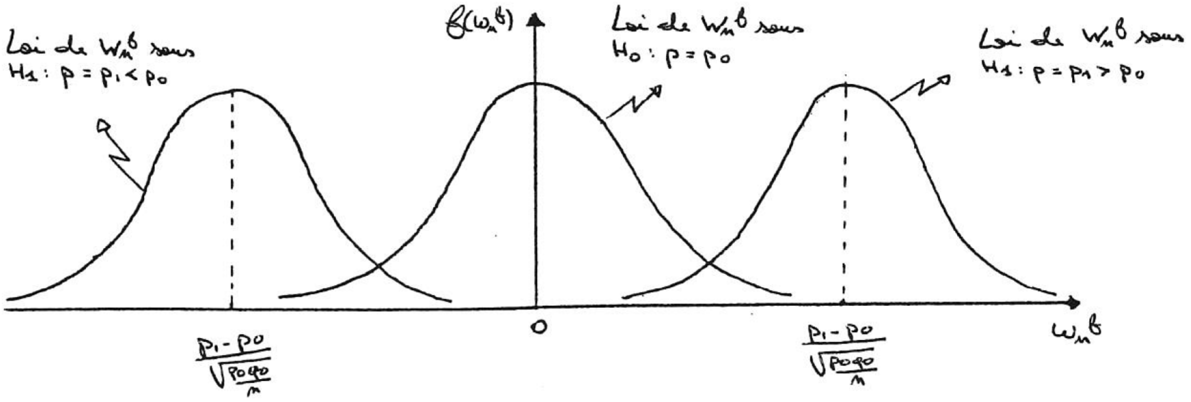
$$F_n - p_0 \approx \mathcal{N}(p_1 - p_0; \tfrac{p_1 q_1}{n})$$

ou encore :

$$W_n^f = \frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \approx \mathcal{N}\left(\frac{p_1 - p_0}{\sqrt{\frac{p_0 q_0}{n}}}; \frac{p_1 q_1}{p_0 q_0} \right) \tag{11.2}$$

On constate que la statistique W_n^f suit sous H_0 (approximativement) une loi normale $\mathcal{N}(0; 1)$ tandis que sous H_1 elle est (approximativement) distribuée selon une loi normale d'espérance non nulle, positive si $p_1 > p_0$ et négative si $p_1 < p_0$, et d'écart-type $\sqrt{\frac{p_1 q_1}{p_0 q_0}}$ différent mais, pour la plupart des couples de valeurs possibles de p_0 et p_1, pas très éloigné de 1.

Comme le montre le Graphique 11.1, sous H_0, on peut s'attendre à ce que la statistique W_n^f prenne des valeurs "proches" de zéro, ce qui reflète simplement le fait que sous H_0 on peut s'attendre à ce que l'estimateur F_n prenne des valeurs — c'est-à-dire délivre des estimations $\widehat{f}$ — "proche" de p_0. Au contraire, sous H_1, la statistique W_n^f aura tendance à prendre des valeurs "éloignées" de zéro, positives (si $p_1 > p_0$) ou négatives (si $p_1 < p_0$), ce qui à nouveau reflète simplement le fait que sous H_1 on peut s'attendre à ce que l'estimateur F_n prenne des valeurs — délivre des estimations $\widehat{f}$ donc — "proches" de p_1. Les termes "proche" et "éloigné" sont évidemment à comprendre en regard des lois des quantités auxquelles ils se réfèrent.



Graphique 11.1 : Distribution d'échantillonnage de W_n^f

Clairement, si W_n^f prend une valeur "proche" de zéro, on n'a aucune raison de remettre en cause H_0 : un tel évémenent est parfaitement compatible avec le fait que H_0 soit vraie. Par contre, si W_n^f prend une valeur "éloignée" de zéro, on peut légitiment douter de la véracité de H_0 : un tel événement est peu compatible avec le fait que H_0 soit vraie tout en étant parfaitement compatible avec le fait que H_0 soit fausse.

Ce raisonnement suggère de retenir, pour choisir de rejeter ou non H_0, la règle de décision :

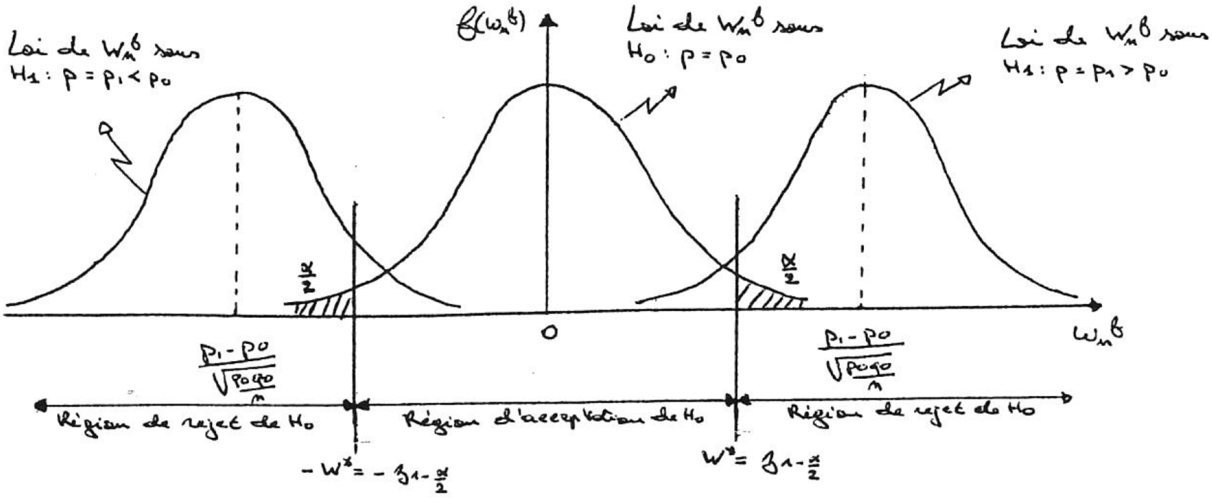
$$\begin{cases} \mathrm{RH}_0 & \text{si } \left|W_n^f\right| = \left|\frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}}\right| > w^* \\ \mathrm{NRH}_0 & \text{sinon} \end{cases} \tag{11.3}$$

où RH_0 signifie "rejet de H_0", NRH_0 "non rejet de H_0" et w^* est une *valeur critique* telle que la probabilité α de rejeter H_0 alors que H_0 est vraie :

$$\alpha = I\!P\left(\left|W_n^f\right| > w^* | \mathrm{H}_0 \text{ est vraie}\right)$$

est (très) petite. Cette règle de décision revient à rejeter H_0 si un indice sérieux, à savoir la réalisation d'un événement (très) peu probable sous H_0 et plus probable sous H_1, suggère qu'elle soit fausse, et à ne pas le rejeter sinon.

La probabilité α est appelée le *seuil* ou encore le *niveau* (*de signification*) du test. Habituellement, on choisit $\alpha = 0,05$ ou $\alpha = 0,01$. La statistique W_n^f sur laquelle est basée la règle de décision est appelée une *statistique de test*, l'ensemble des valeurs de W_n^f pour lesquelles on rejette H_0 est appelée la *région de rejet* de H_0 et l'ensemble des valeurs de W_n^f pour lesquelles on ne rejette pas H_0 la *région d'acceptation* de H_0 (cf. Graphique 11.2).



Graphique 11.2 : La règle de décision du test bilatéral d'une fréquence

La valeur critique w^* correspondant au seuil α du test peut être déterminée en s'appuyant sur le fait que, sous H_0, la statistique W_n^f est approximativement normale $\mathcal{N}(0;1)$. Pour $Z \rightsquigarrow \mathcal{N}(0;1)$, on a (cf. Chapitre 10) :

$$I\!P\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \quad \Leftrightarrow \quad I\!P\left(|Z| > z_{1-\frac{\alpha}{2}}\right) = \alpha \qquad (11.4)$$

où $z_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite $\mathcal{N}(0;1)$. On en déduit que la valeur critique w^* correspondant au seuil α du test est approximativement égale au quantile $z_{1-\frac{\alpha}{2}}$.

Propriété 11.1 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{B}(p)$, où p est inconnu. Pour n grand et p_0 ni trop petit ni trop grand ($n \geq 30$, $np_0 \geq 5$ et $nq_0 \geq 5$), un test approximatif au seuil α de $H_0: p = p_0$ contre $H_1: p \neq p_0$ est donné par la règle de décision :*

$$\begin{cases} RH_0 & si\ \left|W_n^f\right| = \left|\frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}}\right| > z_{1-\frac{\alpha}{2}} \\ NRH_0 & sinon \end{cases} \qquad (11.5)$$

Les conditions d'applicabilité ($n \geq 30$, $np_0 \geq 5$ et $nq_0 \geq 5$) de cette propriété découlent des conditions sous lesquelles on peut, à titre approximatif en échantillon fini, utiliser le résultat donné par la théorème central limite appliqué à la fréquence-échantillon F_n.

L'application de la règle de décision (11.5) à l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose consiste à rejeter H_0 si la réalisation observée $\widehat{w}^f = (\widehat{f} - p_0)/\sqrt{\frac{p_0 q_0}{n}}$ de la statistique de test W_n^f est en valeur absolue supérieure à $z_{1-\frac{\alpha}{2}}$, c'est-à-dire si l'estimation $\widehat{f}$ obtenue est "assez largement différente" de p_0, et à ne pas rejeter H_0 sinon. Le sens de l'expression "assez largement différente" est évidemment à relativiser en fonction des valeurs de n et p_0, c'est-à-dire de la valeur du dénominateur de $\widehat{w}^f$.

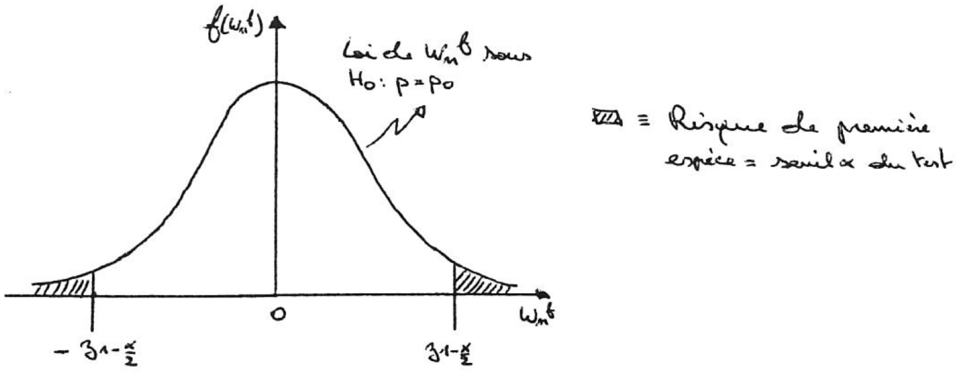
L'utilisation d'une règle de décision telle que la règle de décision (11.5) peut nous faire commettre deux types d'erreurs, synthétisées par le Tableau 11.1.

	H_0 est vraie	H_0 est fausse (H_1 est vraie)
RH_0	erreur de première espèce	pas d'erreur
NRH_0	pas d'erreur	erreur de deuxième espèce

Tableau 11.1 : Erreur de première espèce et erreur de deuxième espèce

La première erreur possible est le rejet de l'hypothèse nulle H_0 alors que H_0 est vraie. Cette erreur est appelée *erreur de première espèce* (ou encore *erreur de type I*) et la probabilité de commettre cette erreur le *risque de première espèce* : c'est la probabilité de rejeter à tort l'hypothèse nulle H_0. Le risque de première espèce est ici la probabilité d'observer une réalisation $\widehat{w}^f$ de W_n^f supérieure en valeur absolue à $z_{1-\frac{\alpha}{2}}$ lorsque H_0 est vraie. Par construction, cette probabilité est (approximativement) égale au seuil α de notre procédure de test (cf. Graphique 11.3) :

$$\mathbb{P}(\mathrm{RH}_0|\mathrm{H}_0 \text{ est vraie}) = \mathbb{P}\left(\left|W_n^f\right| > z_{1-\frac{\alpha}{2}}|\mathrm{H}_0 \text{ est vraie}\right) = \alpha$$



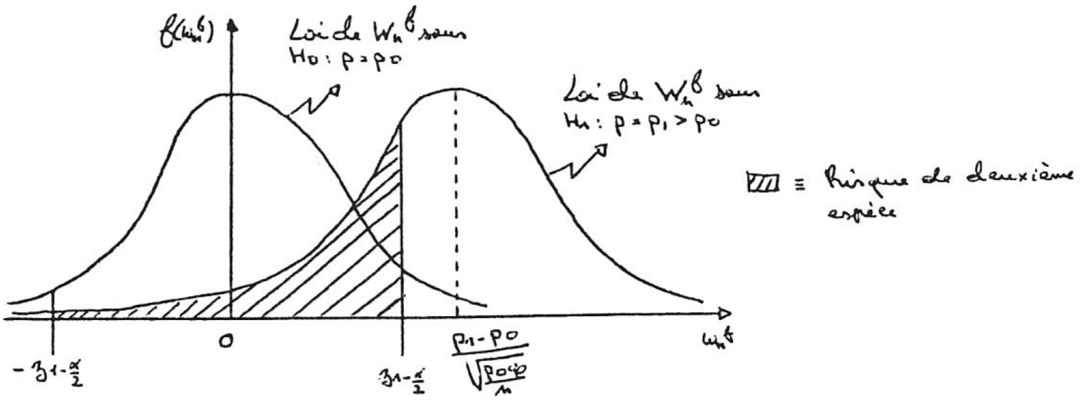
Graphique 11.3 : Le risque de première espèce

Le risque de première espèce est donc entièrement sous notre contrôle. On peut le moduler en choisissant pour α la valeur que l'on souhaite. Clairement, plus α est choisi petit, plus on pourra être confiant dans le fait que lorsque pour l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose on rejette H_0, ce rejet est effectivement dû au fait que H_0 est fausse.

La deuxième erreur possible est le non-rejet de l'hypothèse nulle H_0 alors que H_0 est fausse. Cette erreur est appelée *erreur de deuxième espèce* (ou encore *erreur de type II*) et la probabilité de commettre cette erreur le *risque de deuxième espèce* : c'est la probabilité de ne pas rejeter à tort l'hypothèse nulle H_0. Le risque de deuxième espèce, généralement noté β, est ici la probabilité d'observer une réalisation $\widehat{w}^f$ de W_n^f inférieure ou égale en valeur absolue à $z_{1-\frac{\alpha}{2}}$ lorsque H_0 est fausse, c'est-à-dire H_1 est vraie :

$$I\!P(\mathrm{NRH}_0|\mathrm{H}_0 \text{ est fausse}) = I\!P\left(\left|W_n^f\right| \leq z_{1-\frac{\alpha}{2}}|\mathrm{H}_1 \text{ est vraie}\right) = \beta$$

Comme on peut le vérifier sur base du Graphique 11.4, le risque β de deuxième espèce dépend du seuil α du test choisi : toute autre chose étant égale, plus α est petit, plus β sera grand (on constate que l'on ne peut pas avoir le beurre et l'argent du beurre...). Il dépend aussi de l'importance de la fausseté de l'hypothèse nulle H_0 et de la précision d'estimation : pour α fixé, plus l'hypothèse nulle est fausse ($|p_1 - p_0|$ est grand) et plus la précision d'estimation est grande (l'écart-type $\sigma(F_n)$ de l'estimateur F_n est petit, ce qui est en particulier le cas plus n est grand), plus la loi de W_n^f sous H_1 sera décentrée[18] par rapport à zéro, et donc plus le risque β de deuxième espèce sera faible. Notez que, quels que soient $\alpha > 0$ et $|p_1 - p_0| > 0$, le risque β de deuxième espèce tend vers zéro lorsque la taille d'échantillon n tend vers l'infini.



Graphique 11.4 : Le risque de deuxième espèce pour $p = p_1 > p_0$

La quantité $1 - \beta$ est appelée la *puissance du test.* C'est la probabilité de rejeter l'hypothèse nulle H_0 lorsque H_0 est fausse, c'est-à-dire la probabilité de rejeter à raison l'hypothèse nulle H_0. Elle est ici donnée par :

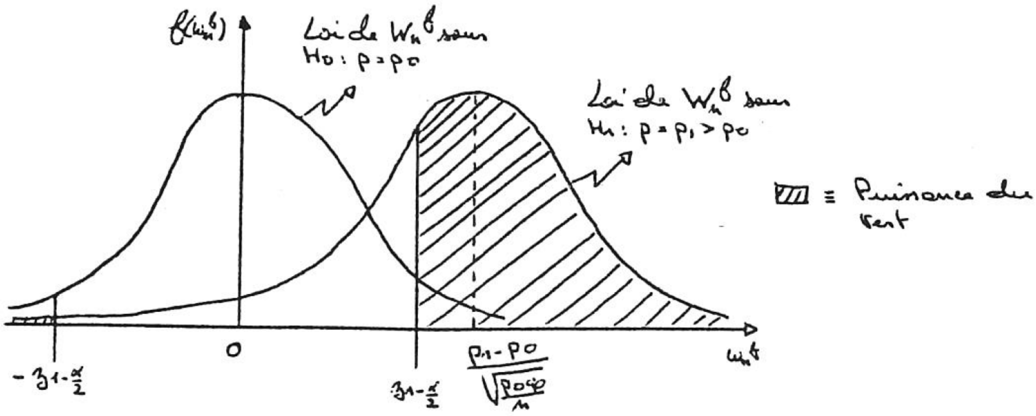
$$I\!P(\mathrm{RH}_0|\mathrm{H}_0 \text{ est fausse}) = I\!P\left(\left|W_n^f\right| > z_{1-\frac{\alpha}{2}}|\mathrm{H}_1 \text{ est vraie}\right) = 1 - \beta$$

c'est-à-dire la probabilité d'observer une réalisation $\widehat{w}^f$ de W_n^f supérieure en valeur absolue à $z_{1-\frac{\alpha}{2}}$ lorsque H_0 est fausse, c'est-à-dire H_1 est vraie (cf. Graphique 11.5).

Evidemment, à un risque β de deuxième espèce faible correspond une puissance $1-\beta$ de test élevée, et vice versa. Ainsi, comme on peut le déduire des déterminants du risque β de deuxième espèce décrits ci-dessus ou le vérifier directement sur base du Graphique 11.5 : (a) toute autre chose étant égale, plus le seuil α est petit, moins la puissance $1 - \beta$ du test sera grande ; (b) pour α fixé, plus l'hypothèse nulle est fausse ($|p_1 - p_0|$ est grand) et plus la précision d'estimation est grande (l'écart-type $\sigma(F_n)$ de l'estimateur F_n est petit, ce qui est en particulier le cas plus n est

[18] Par simplicité, on omet les variations de l'écart-type de la loi de W_n^f sous H_1, dont l'impact est plus marginal.

grand), plus la puissance $1-\beta$ du test sera grande ; et (c) quels que soient $\alpha > 0$ et $|p_1 - p_0| > 0$, la puissance $1-\beta$ du test tend vers 1 lorsque la taille d'échantillon n tend vers l'infini.



Graphique 11.5 : La puissance du test pour $p = p_1 > p_0$

Au contraire du risque de commettre une erreur de première espèce qui est déterminé par le choix du seuil α du test et qui est donc sous notre contrôle, le risque de commettre une erreur de deuxième espèce, ou ce qui revient au même mais est plus parlant la puissance du test, n'est pas sous notre contrôle : α étant fixé, l'un et l'autre dépendent de n, de p_0 et de la valeur inconnue du paramètre $p = p_1$ sous H_1. Pour cette raison, on se gardera, tout spécialement lorsque la précision d'estimation est modérée (l'écart-type $\sigma(F_n)$ de l'estimateur F_n n'est pas très petit, ce qui est en particulier le cas lorsque la taille n de l'échantillon n'est pas très grande) d'interpréter un non rejet de H_0 pour l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose comme un bon indice que H_0 est vraie. On ne peut en effet interpréter un non rejet de H_0 comme un bon indice que H_0 est — au mieux approximativement — vraie que si on a l'assurance que le risque β de deuxième espèce est faible, ou ce qui revient au même que la puissance $1-\beta$ du test est élevée, pour toutes les valeurs de $p = p_1$ qui ne sont pas dans le voisinage immédiat de p_0. Or, cela ne peut être le cas que si la précision d'estimation est très élevée, ce qui assure que, même pour des déviations relativement faibles de $p = p_1$ par rapport à p_0, la loi de W_n^f sous H_1 est substantiellement décentrée[19] par rapport à zéro, et donc le risque β de deuxième espèce est faible, ou ce qui revient au même, la puissance $1-\beta$ du test est élevée[20]. Lorsque précision d'estimation est modérée (ce qui en pratique est le plus souvent le cas), le risque β de deuxième espèce ne sera faible, ou de façon équivalente la puissance $1-\beta$ du test élevée, que pour des déviations relativement importantes de $p = p_1$ par rapport à p_0, d'où le fait qu'il faut se garder, au moins dans ce cas de figure, d'inférer d'un non rejet de H_0 pour l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose que H_0 est vraie. Il s'agit

[19] Par simplicité, on omet à nouveau les variations de l'écart-type de la loi de W_n^f sous H_1, dont l'impact est plus marginal.

[20] Pour fixer les idées, un test au seuil $\alpha = 0,05$ a une puissance de l'ordre de 0,5 (ce qui n'est pas très élevé) lorsque $|p_1 - p_0|/\sqrt{\frac{p_0 q_0}{n}} \approx 2$. Pour obtenir une puissance de test de l'ordre de $0,95$ (et donc un risque de deuxième espèce de l'ordre de $0,05$, c'est-à-dire de niveau comparable au risque de première espèce), il faut avoir $|p_1 - p_0|/\sqrt{\frac{p_0 q_0}{n}} \approx 4$, en d'autres termes que la déviation de $p = p_1$ par rapport à p_0 soit approximativement 4 fois plus importante que la précision d'estimation.

seulement d'une absence de preuve que H_0 est fausse (ce qui n'est déjà pas si mal...).

En résumé, on voit qu'un rejet et un non rejet de H_0 pour l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose ne s'interprète pas de la façon : si un rejet de H_0 peut être vu comme une preuve convaincante que H_0 est fausse, et d'autant plus convaincante que le seuil α choisi est petit, il faut par contre se garder, au moins lorsque la précision d'estimation est modérée, ce qui est en particulier le cas lorsque la taille n de l'échantillon n'est pas très grande, d'interpréter un non rejet de H_0 comme un indice sérieux que H_0 est vraie.

11.1.2. Exercice résolu

On tire un échantillon aléatoire de 100 individus dans une population et on compte le nombre de fois qu'on obtient un homme.

1- Proposez un test d'hypothèse pour évaluer si cette population contient autant d'hommes que de femmes.

On représente la population par une loi de Bernoulli $\mathcal{B}(p)$, où p est la proportion (inconnue) d'hommes dans la population : $X \rightsquigarrow \mathcal{B}(p)$. Si il y a autant d'hommes que de femmes dans la population $p = 0,5$, et $p \neq 0,5$ sinon. Pour évaluer si la population contient autant d'hommes que de femmes, on testera donc :

$$\mathrm{H}_0 : \ p = p_0 \quad \text{contre} \quad \mathrm{H}_1 : \ p \neq p_0 \quad , \text{où } p_0 = 0,5$$

2- Pour quelles valeurs de la proportion observée d'hommes dans l'échantillon dont on dispose ne rejettera-t-on pas, au seuil $\alpha = 5\%$, l'hypothèse que la population contient autant d'hommes que de femmes ?

On a $n = 100 \geq 30$, $np_0 = 50$ et $nq_0 = n(1 - p_0) = 50 \geq 5$. On est donc dans les conditions d'applicabilité de la Propriété 11.1. De cette propriété, un test au seuil α de $\mathrm{H}_0 : \ p = p_0$ contre $\mathrm{H}_1 : \ p \neq p_0$ est donné par la règle de décision :

$$\begin{cases} \mathrm{RH}_0 & \text{si } \left|W_n^f\right| = \left|\frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}}\right| > z_{1-\frac{\alpha}{2}} \\ \mathrm{NRH}_0 & \text{sinon} \end{cases}$$

Pour $\alpha = 0,05$, on trouve dans les tables $z_{0,975} = 1,96$. On ne rejettera donc pas H_0 pour l'échantillon dont on dispose si la réalisation observée $\widehat{w}^f$ de la statistique W_n^f est telle que :

$$\left|\widehat{w}^f\right| = \left|\frac{\widehat{f} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}\right| = \left|\frac{\widehat{f} - 0,5}{\sqrt{\frac{0,5(1-0,5)}{100}}}\right| \leq 1,96$$

c'est-à-dire si :

$$\left|\widehat{f} - 0,5\right| \leq 1,96 \times \sqrt{\tfrac{0,5(1-0,5)}{100}} = 1,96 \times 0,05 = 0,098$$

$$\Leftrightarrow \quad -0,098 \leq \widehat{f} - 0,5 \leq 0,098 \quad \Leftrightarrow \quad 0,402 \leq \widehat{f} \leq 0,598$$

3- Parmi les 100 individus tirés, on observe 57 hommes. Quelle est, sur base de votre réponse au point 2, votre conclusion ?

On a $\widehat{f} = \frac{57}{100} = 0,57$. Au seuil de 5%, on ne rejette donc pas H_0, c'est-à-dire l'hypothèse que la population contient autant d'hommes que de femmes.

4- Si on avait prélevé un échantillon aléatoire de 1 000 individus, pour quelles valeurs de la proportion observée d'hommes dans l'échantillon dont on dispose n'aurait-on pas rejeté, toujours au seuil 5%, l'hypothèse que la population contient autant d'hommes que de femmes ?

Pour $n = 1000$, on ne rejettera pas H_0 pour l'échantillon dont on dispose si la réalisation $\widehat{w}^f$ de W_n^f est telle que :

$$\left|\widehat{w}^f\right| = \left|\frac{\widehat{f} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}\right| = \left|\frac{\widehat{f} - 0,5}{\sqrt{\frac{0,5(1-0,5)}{1000}}}\right| \leq 1,96$$

c'est-à-dire si :

$$\left|\widehat{f} - 0,5\right| \leq 1,96 \times \sqrt{\tfrac{0,5(1-0,5)}{1000}} \approx 1,96 \times 0,0158 \approx 0,031$$

$$\Leftrightarrow \quad -0,469 \leq \widehat{f} - 0,5 \leq 0,031 \quad \Leftrightarrow \quad 0,469 \leq \widehat{f} \leq 0,531$$

5- Calculez la puissance du test proposé au point 1 pour $\alpha = 5\%$ et :

a- $n = 100$ et $p = p_1 = 0,65$.
b- $n = 100$ et $p = p_1 = 0,40$.
c- $n = 100$ et $p = p_1 = 0,55$.
d- $n = 100$ et $p = p_1 = 0,48$.
e- $n = 1000$ et $p = p_1 = 0,40$.
f- $n = 1000$ et $p = p_1 = 0,55$.
g- $n = 1000$ et $p = p_1 = 0,48$.
h- $n = 1000$ et $p = p_1 = 0,51$.
i- $n = 10\,000$ et $p = p_1 = 0,55$.
j- $n = 10\,000$ et $p = p_1 = 0,48$.
k- $n = 10\,000$ et $p = p_1 = 0,51$.

Le risque β de deuxième espèce associé au test au seuil α de $\mathrm{H}_0 : p = p_0$ contre $\mathrm{H}_1 : p \neq p_0$ est donné par :

$$\beta = I\!\!P(\mathrm{NRH}_0|\mathrm{H}_0 \text{ est fausse}) = I\!\!P\left(\left|W_n^f\right| \leq z_{1-\frac{\alpha}{2}}|\mathrm{H}_1 \text{ est vraie avec } p = p_1\right)$$

soit :

$$\beta = I\!P\left(-z_{1-\frac{\alpha}{2}} \leq W_n^f \leq z_{1-\frac{\alpha}{2}}\right) \ \text{, où } W_n^f \approx \mathcal{N}\left(\frac{p_1-p_0}{\sqrt{\frac{p_0q_0}{n}}}; \frac{p_1q_1}{p_0q_0}\right)$$

On a donc :

$$\beta = I\!P\left(\frac{-z_{1-\frac{\alpha}{2}} - \frac{p_1-p_0}{\sqrt{\frac{p_0q_0}{n}}}}{\sqrt{\frac{p_1q_1}{p_0q_0}}} \leq \underbrace{\frac{W_n^f - \frac{p_1-p_0}{\sqrt{\frac{p_0q_0}{n}}}}{\sqrt{\frac{p_1q_1}{p_0q_0}}}}_{=Z\approx\mathcal{N}(0;1)} \leq \frac{z_{1-\frac{\alpha}{2}} - \frac{p_1-p_0}{\sqrt{\frac{p_0q_0}{n}}}}{\sqrt{\frac{p_1q_1}{p_0q_0}}}\right)$$

c'est-à-dire approximativement :

$$\beta = \Phi\left(\frac{z_{1-\frac{\alpha}{2}} - \frac{p_1-p_0}{\sqrt{\frac{p_0q_0}{n}}}}{\sqrt{\frac{p_1q_1}{p_0q_0}}}\right) - \Phi\left(\frac{-z_{1-\frac{\alpha}{2}} - \frac{p_1-p_0}{\sqrt{\frac{p_0q_0}{n}}}}{\sqrt{\frac{p_1q_1}{p_0q_0}}}\right) \tag{11.6}$$

La puissance de test est égale à $1-\beta$. On a $p_0 = 0,5$, et pour $\alpha = 0,05$, $z_{0{,}975} = 1,96$. En utilisant l'expression (11.6) et les tables, on obtient (nous laisserons au lecteur le soin de vérifier les calculs), pour les différents couples de valeurs de n et p_1 demandés, les puissances $1-\beta$ de test données par le tableau suivant :

p_1	$\lvert p_1 - 0,5\rvert$	$n = 100$	$n = 1000$	$n = 10\,000$
0,65	0,15	0,8622	–	–
0,40	0,10	0,5163	≈ 1	–
0,55	0,05	0,1688	0,8865	≈ 1
0,48	0,02	0,0683	0,2439	0,9794
0,51	0,01	–	0,0969	0,5160

6- Etant donné les résultats obtenus au point 5, considérez-vous qu'un non rejet de H_0 pour un échantillon aléatoire de 100 individus est un indice très convaincant que la population contient autant d'hommes que de femmes ?

Pas vraiment, car pour cette taille d'échantillon, la puissance du test reste modeste même pour des déviations relativement importantes (par ex. $\lvert p_1 - 0,5\rvert = 0,10$) par rapport à $p_0 = 0,5$. Clairement, un non rejet de H_0 pour $n = 1000$, ou mieux encore $n = 10\,000$, serait un indice nettement plus convaincant que la population contient (au moins approximativement) autant d'hommes que de femmes.

11.1.3. Tests unilatéraux

11.1.3.1. Test unilatéral à droite

Examinons à présent le cas du test unilatéral à droite :

$$\mathrm{H}_0 : p \leq p_0 \quad \text{contre} \quad \mathrm{H}_1 : p > p_0$$

Des résultats de la Section 11.1.1, on sait que sous H_0 : $p \leq p_0$ avec $p = p_0$, c'est-à-dire si H_0 est vraie avec $p = p_0$, la statistique W_n^f suit (approximativement) une loi normale $\mathcal{N}(0;1)$ (cf. équation (11.1)), que sous H_1 : $p = p_1 > p_0$, c'est-à-dire si H_0 est fausse, elle est (approximativement) distribuée selon une loi normale d'espérance positive et d'écart-type généralement pas très éloigné de 1 (cf. équation (11.2)), et on déduit immédiatement que sous H_0 : $p = p_0' < p_0$, c'est-à-dire si H_0 est vraie avec $p = p_0' < p_0$, on a :

$$W_n^f = \frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \approx \mathcal{N}\left(\frac{p_0' - p_0}{\sqrt{\frac{p_0 q_0}{n}}}; \frac{p_0' q_0'}{p_0 q_0}\right) \qquad (q_0' = 1 - p_0') \tag{11.7}$$

autrement dit que W_n^f suit une loi normale d'espérance négative et d'écart-type également généralement pas très éloigné de 1.

Etant donné ces caractéristiques, comme on peut le voir sur le Graphique 11.6, si W_n^f prend une valeur négative, ou positive mais "proche" de zéro, on n'a aucune raison de remettre en cause H_0 : un tel événement est parfaitement compatible avec le fait que H_0 soit vraie. Par contre, si W_n^f prend une valeur positive et "éloignée" de zéro, on peut légitiment douter de la véracité de H_0 : un tel événement est peu compatible avec le fait que H_0 soit vraie tout en étant parfaitement compatible avec le fait que H_0 soit fausse.

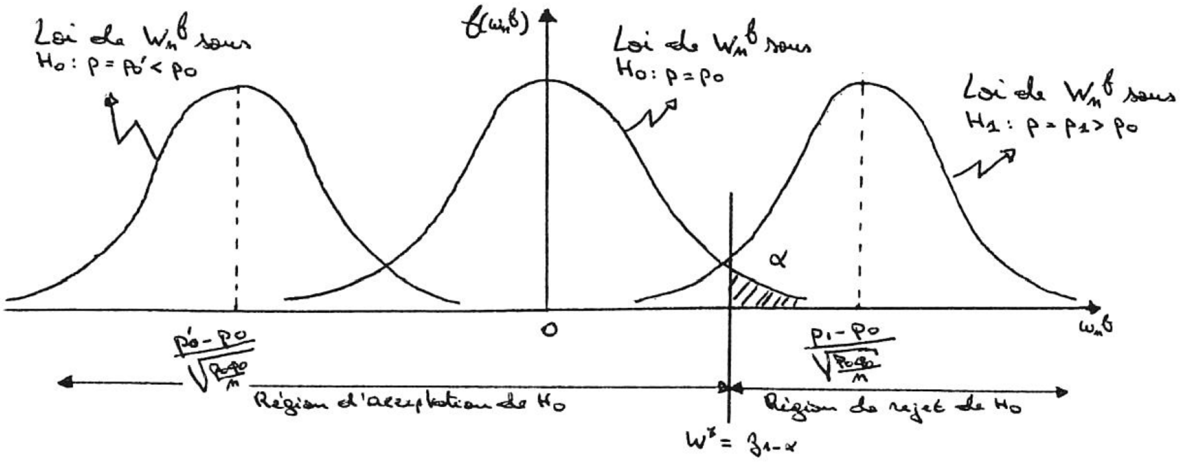
Suivant ce raisonnement, une règle de décision naturelle pour choisir de rejeter ou non H_0 est donc ici :

$$\begin{cases} \mathrm{RH}_0 & \text{si } W_n^f = \frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}} > w^* \\ \mathrm{NRH}_0 & \text{sinon} \end{cases} \tag{11.8}$$

où w^* est une valeur critique telle que la probabilité α — le seuil ou encore le niveau (de signification) du test — de rejeter H_0 lorsque que H_0 est vraie avec $p = p_0$:

$$\alpha = I\!P\left(W_n^f > w^* | \mathrm{H}_0 \text{ est vraie avec } p = p_0\right)$$

est (très) petite. Cette règle de décision revient, comme dans le cas du test bilatéral, à rejeter H_0 si un indice sérieux, à savoir la réalisation d'un événement (très) peu probable sous H_0 et plus probable sous H_1, suggère qu'elle soit fausse, et à ne pas le rejeter sinon.



Graphique 11.6 : La règle de décision du test unilatéral à droite d'une fréquence

On détermine aisément la valeur critique w^* correspondant au seuil α du test en s'appuyant sur le fait que, sous H_0 vraie avec $p = p_0$, la statistique W_n^f est approximativement normale $\mathcal{N}(0;1)$. Pour $Z \rightsquigarrow \mathcal{N}(0;1)$, on a (cf. Chapitre 7) :

$$I\!P(Z \leq z_{1-\alpha}) = 1 - \alpha \quad \Leftrightarrow \quad I\!P(Z > z_{1-\alpha}) = \alpha$$

où $z_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi normale centrée réduite $\mathcal{N}(0;1)$. On en déduit que la valeur critique w^* correspondant au seuil α du test est approximativement égale au quantile $z_{1-\alpha}$.

Propriété 11.2 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{B}(p)$, où p est inconnu. Pour n grand et p_0 ni trop petit ni trop grand ($n \geq 30$, $np_0 \geq 5$ et $nq_0 \geq 5$), un test approximatif au seuil α de $H_0 : p \leq p_0$ contre $H_1 : p > p_0$ est donné par la règle de décision :*

$$\begin{cases} RH_0 & si\ W_n^f = \frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}} > z_{1-\alpha} \\ NRH_0 & sinon \end{cases} \tag{11.9}$$

Comme celles de la Propriété 11.1, les conditions d'applicabilité ($n \geq 30$, $np_0 \geq 5$ et $nq_0 \geq 5$) de cette propriété découlent des conditions sous lesquelles on peut, à titre approximatif en échantillon fini, utiliser le résultat donné par la théorème central limite appliqué à la fréquence-échantillon F_n.

L'application de la règle de décision (11.9) à l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose consiste ici à rejeter H_0 si la réalisation observée $\widehat{w}^f = (\widehat{f} - p_0)/\sqrt{\frac{p_0 q_0}{n}}$ de la statistique de test W_n^f est supérieure à $z_{1-\alpha}$, c'est-à-dire si l'estimation $\widehat{f}$ obtenue est "assez largement supérieure" à p_0, et à ne pas rejeter H_0 sinon. Le sens de l'expression "assez largement supérieure" est évidemment à relativiser en fonction de la valeur du dénominateur de $\widehat{w}^f$.

Comme dans le cas du test bilatéral, l'utilisation cette règle de décision peut nous faire commettre une erreur de première espèce ou une erreur de deuxième espèce.

Le risque de première espèce — la probabilité de rejeter H_0 alors que H_0 est vraie —, qui est ici égal à la probabilité d'observer une réalisation $\widehat{w}^f$ de W_n^f supérieure à $z_{1-\alpha}$ lorsque H_0 est vraie :

$$I\!P(\mathrm{RH}_0|\mathrm{H}_0 \text{ est vraie}) = I\!P\left(W_n^f > z_{1-\alpha}|\mathrm{H}_0 \text{ est vraie}\right)$$

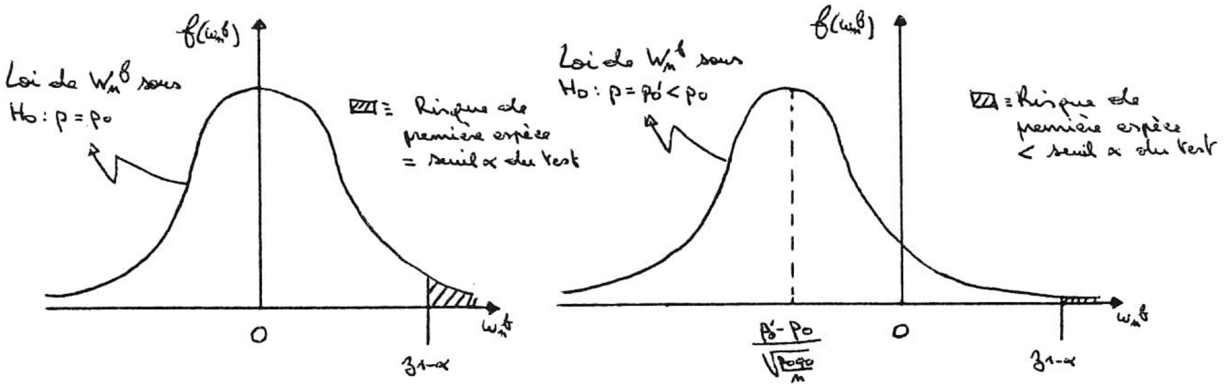
dépend de la valeur inconnue du paramètre p sous H_0. Par construction, lorsque H_0 est vraie avec $p = p_0$, ce risque est (approximativement) égal au seuil α de notre procédure de test :

$$I\!P\left(W_n^f > z_{1-\alpha}|\mathrm{H}_0 \text{ est vraie avec } p = p_0\right) = \alpha$$

et lorsque H_0 est vraie avec $p = p'_0 < p_0$, il est nécessairement inférieur à ce seuil α (cf. Graphique 11.7) :

$$I\!P\left(W_n^f > z_{1-\alpha}|\mathrm{H}_0 \text{ est vraie avec } p = p'_0 < p_0\right) < \alpha$$

On en conclut que, bien qu'il ne le soit plus totalement, le risque de première espèce est toujours bien sous notre contrôle puisqu'il est au maximum (approximativement) égal au seuil α du test. Dès lors, comme dans le cas du test bilatéral, plus α est choisi petit, plus on pourra être confiant dans le fait que lorsque pour l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose on rejette H_0, ce rejet est effectivement dû au fait que H_0 est fausse.

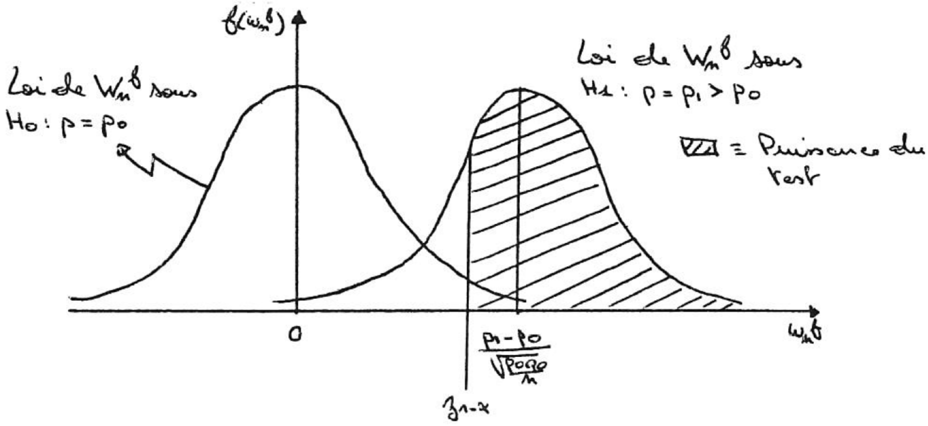


Graphique 11.7 : Le risque de première espèce pour $p = p_0$ et $p = p'_0 < p_0$

De son côté, le risque β de deuxième espèce ou ce qui revient au même mais est plus parlant, la puissance du test — la probabilité $1 - \beta$ de rejeter H_0 alors que H_0 est fausse —, qui est ici égale à la probabilité d'observer une réalisation $\widehat{w}^f$ de W_n^f supérieure à $z_{1-\alpha}$ lorsque H_0 est fausse, c'est-à-dire H_1 est vraie :

$$I\!P(\mathrm{RH}_0|\mathrm{H}_0 \text{ est fausse}) = I\!P\left(W_n^f > z_{1-\alpha}|\mathrm{H}_1 \text{ est vraie}\right) = 1 - \beta$$

dépend, comme dans le cas du test bilatéral, du seuil α du test, de l'importance de la fausseté de l'hypothèse nulle et de la précision d'estimation.



Graphique 11.8 : La puissance du test pour $p = p_1 > p_0$

Comme on peut le vérifier sur base du Graphique 11.8, d'une part, toute autre chose étant égale, plus α est petit et moins la puissance $1 - \beta$ du test sera grande (à nouveau, on constate qu'on ne peut pas avoir le beurre et l'argent du beurre...). D'autre part, pour α fixé, plus l'hypothèse nulle est fausse ($p_1 - p_0 > 0$ est grand) et plus la précision d'estimation est grande (l'écart-type $\sigma(F_n)$ de l'estimateur F_n est petit, ce qui est en particulier le cas plus n est grand), plus la loi de W_n^f sous H_1 sera positivement décentrée[21] par rapport à zéro, et donc plus la puissance $1 - \beta$ du test sera grande. Remarquons qu'à nouveau quels que soient $\alpha > 0$ et $p_1 - p_0 > 0$, la puissance $1 - \beta$ du test tend vers 1 lorsque la taille d'échantillon n tend vers l'infini.

Comme dans le cas du test bilatéral, la puissance $1 - \beta$ du test, ou ce qui revient au même le risque β de deuxième espèce, n'est pas sous notre contrôle : α étant fixé, l'un et l'autre dépendent de n, de p_0 et de la valeur inconnue du paramètre $p = p_1$ sous H_1. Suivant les mêmes arguments que ceux développés pour le test bilatéral, on se gardera donc à nouveau, au moins lorsque la précision d'estimation est modérée, ce qui est en particulier le cas lorsque la taille n de l'échantillon n'est pas très grande, d'interpréter un non rejet de H_0 pour l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose comme un indice convaincant que H_0 est vraie.

11.1.3.2. Test unilatéral à gauche

Le cas du test unilatéral à gauche :

$$\mathrm{H}_0 : p \geq p_0 \quad \text{contre} \quad \mathrm{H}_1 : p < p_0$$

est parfaitement symétrique au cas du test unilatéral à droite.

Le même raisonnement que celui développé pour lui conduit naturellement ici à retenir comme règle de décision pour choisir de rejeter ou non H_0 :

[21] Par simplicité, on omet à nouveau les variations de l'écart-type de la loi de W_n^f sous H_1, dont l'impact est plus marginal.

$$\begin{cases} \mathrm{RH}_0 & \text{si } W_n^f = \frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}} < w^* \\ \mathrm{NRH}_0 & \text{sinon} \end{cases} \tag{11.10}$$

où la valeur critique w^* est telle que la probabilité α — le seuil ou encore le niveau (de signification) du test — de rejeter H_0 lorsque que H_0 est vraie avec $p = p_0$:

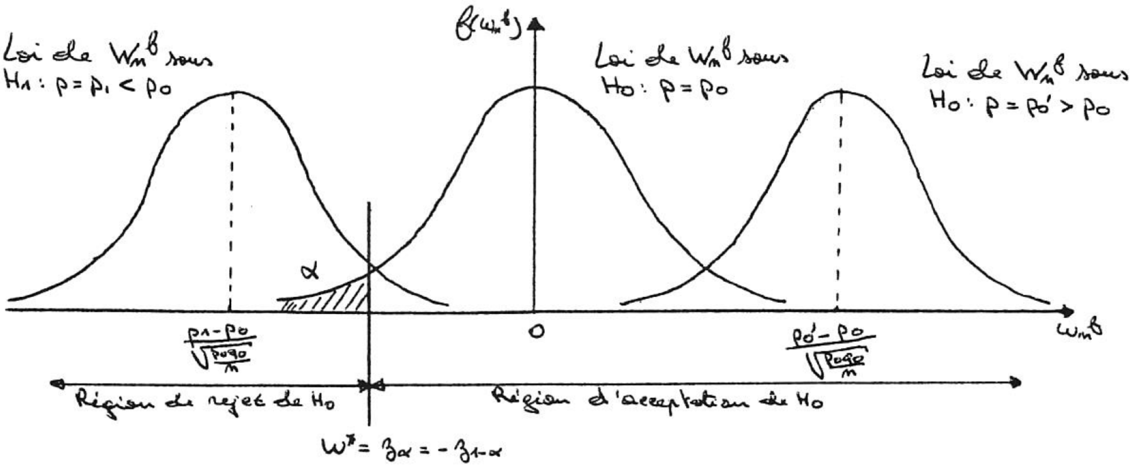
$$\alpha = I\!P\left(W_n^f < w^* | \mathrm{H}_0 \text{ est vraie avec } p = p_0\right)$$

est (très) petite.

La valeur critique w^* correspondant au seuil α du test est obtenue en s'appuyant sur le fait que, sous H_0 vraie avec $p = p_0$, la statistique W_n^f est approximativement normale $\mathcal{N}(0;1)$. Pour $Z \rightsquigarrow \mathcal{N}(0;1)$, on a (cf. Chapitre 7) :

$$I\!P(Z \leq z_\alpha) = \alpha = I\!P(Z \leq -z_{1-\alpha})$$

où z_α et $z_{1-\alpha}$ sont les quantiles d'ordre respectivement α et $1-\alpha$ de la loi normale centrée réduite $\mathcal{N}(0;1)$. On en déduit que la valeur critique w^* correspondant au seuil α du test est approximativement égale au quantile z_α, quantile qui par ailleurs est égal à l'opposé du quantile $z_{1-\alpha}$ $(= -z_{1-\alpha})$.



Graphique 11.9 : La règle de décision du test unilatéral à gauche d'une fréquence

Propriété 11.3 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{B}(p)$, où p est inconnu. Pour n grand et p_0 ni trop petit ni trop grand ($n \geq 30$, $np_0 \geq 5$ et $nq_0 \geq 5$), un test approximatif au seuil α de $H_0 : p \geq p_0$ contre $H_1 : p < p_0$ est donné par la règle de décision :*

$$\begin{cases} RH_0 & si\ W_n^f = \frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}} < z_\alpha \ (= -z_{1-\alpha}) \\ NRH_0 & sinon \end{cases} \tag{11.11}$$

L'origine des conditions d'applicabilité ($n \geq 30$, $np_0 \geq 5$ et $nq_0 \geq 5$) de cette propriété est identique à celle des conditions d'applicabilité des Propriétés 11.1 et

11.2.

L'application de la règle de décision (11.11) à l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose consiste maintenant à rejeter H_0 si la réalisation observée $\widehat{w}^f = (\widehat{f} - p_0)/\sqrt{\frac{p_0 q_0}{n}}$ de la statistique de test W_n^f est inférieure à z_α $(= -z_{1-\alpha})$, c'est-à-dire si l'estimation $\widehat{f}$ obtenue est "assez largement inférieure" à p_0, et à ne pas rejeter H_0 sinon, l'expression "assez largement inférieure" étant évidemment comme précédemment à comprendre en regard de la valeur du dénominateur de $\widehat{w}^f$.

Nous laisserons au lecteur le soin de vérifier que, comme dans le cas du test unilatéral à droite, le risque de première espèce — la probabilité de rejeter H_0 alors que H_0 est vraie — associé à l'utilisation de cette règle de décision est toujours bien sous notre contrôle puisqu'il est à nouveau au maximum (approximativement) égal au seuil α du test.

Nous laisserons également au lecteur le soin de vérifier que la puissance du test — la probabilité $1 - \beta$ de rejeter H_0 alors que H_0 est fausse — (ou ce qui revient au même le risque β de deuxième espèce) dépend, de façon semblable au cas du test unilatéral à droite, du seuil α du test, de l'importance de la fausseté de l'hypothèse nulle (l'écart $p_1 - p_0 < 0$) et de la précision d'estimation (l'écart-type $\sigma(F_n)$ de l'estimateur F_n, qui dépend en particulier de la taille d'échantillon n), que quels que soient $\alpha > 0$ et $p_1 - p_0 < 0$, elle tend vers 1 lorsque n tend vers l'infini, et qu'elle n'est pas sous notre contrôle puisque, α étant fixé, elle dépend de n, de p_0 et de la valeur inconnue du paramètre $p = p_1$ sous H_1.

Du fait que l'erreur de première espèce soit sous contrôle et que l'erreur de deuxième espèce ne le soit pas, on déduit évidemment les mêmes conclusions que précédemment : si un rejet de H_0 pour l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose peut être vu comme une preuve convaincante que H_0 est fausse, et d'autant plus convaincante que le seuil α choisi est petit, il faut par contre se garder, au moins lorsque la précision d'estimation est modérée, ce qui est en particulier le cas lorsque la taille n de l'échantillon n'est pas très grande, d'interpréter un non rejet de H_0 comme un indice sérieux que H_0 est vraie.

11.1.4. Exercice résolu

Le taux de circuits électroniques défectueux qui sortent d'une chaîne d'assemblage est de 7%. Dans le but de réduire ce taux, les ingénieurs de l'entreprise apportent diverses modifications à la chaîne d'assemblage. Pour vérifier que les modifications réalisées diminuent effectivement le taux de circuits défectueux, ils extraient 300 circuits au hasard de la nouvelle chaîne d'assemblage. Ils observent dans leur échantillon un taux de circuits défectueux de 4%.

1- Peut-on en conclure, au seuil de 5%, que les modifications apportées à la chaîne d'assemblage diminuent effectivement le taux de circuits défectueux ?

On représente la population des circuits électroniques issus de la nouvelle chaîne

d'assemblage par une variable aléatoire X suivant une loi de BERNOULLI $\mathcal{B}(p)$, où $p = I\!P(X = 1)$ est le taux inconnu de circuits défectueux.
On pourra affirmer, au seuil de 5%, que les modifications apportées à la chaîne d'assemblage diminuent effectivement le taux de circuits défectueux, si le test au seuil $\alpha = 5\%$ de :

$$\mathrm{H}_0 : \ p \geq p_0 \quad \text{contre} \quad \mathrm{H}_1 : \ p < p_0$$

où $p_0 = 0,07$, rejette H_0.
On a $n = 300 \geq 30$, $np_0 = 21 > 5$ et $nq_0 = n(1 - p_0) = 279 \geq 5$. On est donc dans les conditions d'applicabilité de la Propriété 11.3. De cette propriété, un test au seuil α de $\mathrm{H}_0 : \ p \geq p_0$ contre $\mathrm{H}_1 : \ p < p_0$ est donné par la règle de décision :

$$\begin{cases} \mathrm{RH}_0 & \text{si } W_n^f = \frac{F_n - p_0}{\sqrt{\frac{p_0 q_0}{n}}} < z_\alpha \ (= -z_{1-\alpha}) \\ \mathrm{NRH}_0 & \text{sinon} \end{cases}$$

Pour $\alpha = 0,05$, on trouve dans les tables $z_{0,05} = -z_{0,95} = -1,6449$. La réalisation observée $\widehat{w}^f$ de la statistique W_n^f est :

$$\widehat{w}^f = \frac{\widehat{f} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0,04 - 0,07}{\sqrt{\frac{0,07(1-0,07)}{300}}} \approx \frac{-0,03}{0,01473} \approx -2,0366$$

On a $\widehat{w}^f \approx -2,0366 < -1,6449 = z_{0,05}$. On rejette donc H_0. On en conclut que, au seuil de 5%, les modifications apportées à la chaîne d'assemblage diminuent effectivement le taux de circuits défectueux.

2- Et au seuil de 1% ?

Pour $\alpha = 0,01$, on trouve dans les tables $z_{0,01} = -z_{0,99} = -2,3263$. Pour ce risque de première espèce (maximum) plus faible, on a $\widehat{w}^f \approx -2,0366 > -2,3263 = z_{0,01}$, de sorte qu'on ne rejette pas H_0. A ce seuil de 1%, on ne peut donc pas conclure que les modifications apportées à la chaîne d'assemblage diminuent effectivement le taux de circuits défectueux.

11.2. Tests relatifs à la moyenne

Supposons maintenant qu'on considère une population représentée par la loi d'une variable aléatoire X d'espérance m et de variance σ^2 inconnues. Disposant d'un échantillon d'observations $(x_1, x_2, ..., x_n)$ — une réalisation d'un échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à X —, on suppose qu'on souhaite mettre à l'épreuve une assertion concernant la valeur inconnue de la moyenne m dans la population, c'est-à-dire tester si cette assertion est ou non compatible avec les observations dont on dispose.

Comme pour la fréquence p dans la population, on établit ci-après des règles de décision adaptées aux cas d'un test bilatéral ($\mathrm{H}_0 : \ m = m_0$ contre $\mathrm{H}_1 : \ m \neq m_0$), d'un test unilatéral à droite ($\mathrm{H}_0 : \ m \leq m_0$ contre $\mathrm{H}_1 : \ m > m_0$) et d'un test unilatéral à gauche ($H_0 : m \geq m_0$ contre $H_1 : m < m_0$). Notez qu'aucune hypothèse n'est faite concernant la valeur de la variance σ^2 dans la population.

11.2.1. Test bilatéral

On commence par examiner le cas du test bilatéral :

$$H_0 : m = m_0 \quad \text{contre} \quad H_1 : m \neq m_0$$

Pour évaluer le valeur inconnue de la moyenne m dans la population, on dispose de l'estimateur sans biais et convergent (et à variance minimale parmi les estimateurs linéaires sans biais, dans certains cas parmi tous les estimateurs sans biais) $\overline{X}_n$, estimateur dont l'application à un échantillon d'observations $(x_1, x_2, ..., x_n)$ donne une estimation $\widehat{m}$.

Sous l'hypothèse nulle $H_0 : m = m_0$, du théorème central limite (cf. Propriété 9.10) on sait que, pour n grand ($n \geq 30$) et quel que soit la loi de X, on a approximativement :

$$\overline{X}_n \approx \mathcal{N}(m_0; \tfrac{\sigma^2}{n}) \quad \Leftrightarrow \quad \frac{\overline{X}_n - m_0}{\sqrt{\frac{\sigma^2}{n}}} \approx \mathcal{N}(0; 1)$$

et, comme $S_n^2 \xrightarrow{P} \sigma^2$, également (cf. Chapitre 10, Section 10.2.2.2) :

$$W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} \approx \mathcal{N}(0; 1) \tag{11.12}$$

Sous l'hypothèse alternative $H_1 : m = m_1 \neq m_0$ (que l'on ait $m_1 > m_0$ ou $m_1 < m_0$), on a par contre approximativement :

$$\overline{X}_n \approx \mathcal{N}(m_1; \tfrac{\sigma^2}{n})$$

et donc :

$$\overline{X}_n - m_0 \approx \mathcal{N}(m_1 - m_0; \tfrac{\sigma^2}{n})$$

ou encore :

$$\frac{\overline{X}_n - m_0}{\sqrt{\frac{\sigma^2}{n}}} \approx \mathcal{N}\left(\frac{m_1 - m_0}{\sqrt{\frac{\sigma^2}{n}}}; 1\right)$$

et, comme $S_n^2 \xrightarrow{P} \sigma^2$, suivant le même raisonnement qu'à la Section 10.2.2.2 évoquée ci-dessus, également :

$$W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} \approx \mathcal{N}\left(\frac{m_1 - m_0}{\sqrt{\frac{\sigma^2}{n}}}; 1\right) \tag{11.13}$$

On constate que le comportement de la statistique de test W_n^m sous H_0 et sous H_1 ressemble comme deux gouttes d'eau à celui de la statistique W_n^f discuté à la Section 11.1.1. De façon semblable à W_n^f, W_n^m suit sous H_0 (approximativement) une loi normale $\mathcal{N}(0; 1)$ tandis que sous H_1 elle est (approximativement) distribuée selon une loi normale d'espérance non nulle, positive si $m_1 > m_0$ et négative si $m_1 < m_0$, et d'écart-type (ici exactement) égal à 1 (cf. Graphique 11.10).

Le même raisonnement que celui développé à la Section 11.1.1 conduit naturellement à retenir le même type de règle de décision pour choisir de rejeter ou non H_0 :

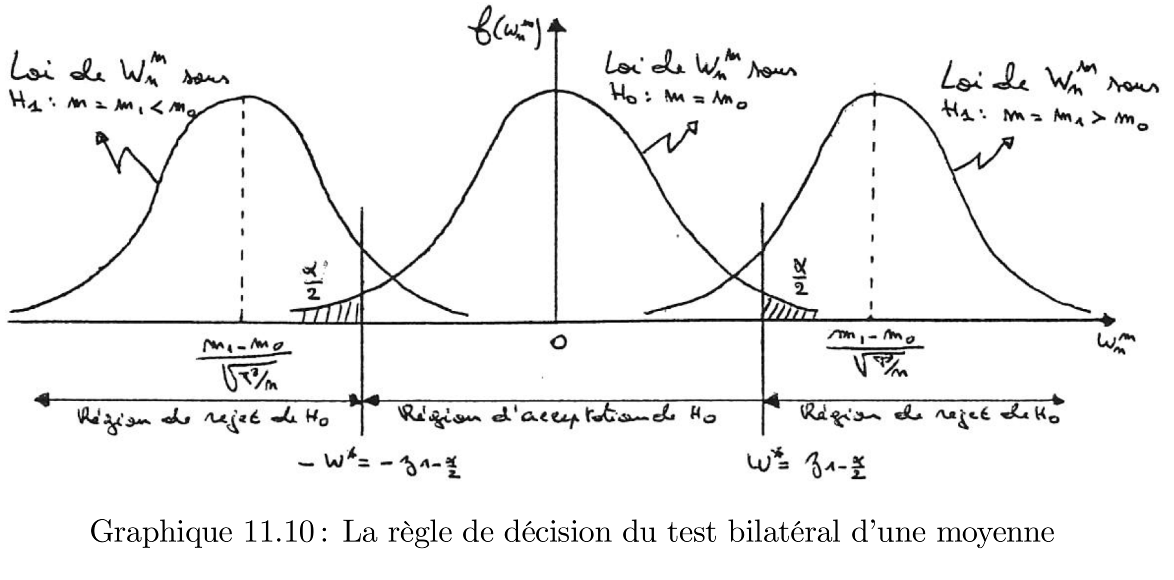
$$\begin{cases} \mathrm{RH}_0 & \text{si } |W_n^m| = \left|\frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}}\right| > w^* \\ \mathrm{NRH}_0 & \text{sinon} \end{cases} \tag{11.14}$$

où la valeur critique w^* est telle que la probabilité α — le seuil ou encore le niveau (de signification) du test — de rejeter H_0 alors que H_0 est vraie :

$$\alpha = I\!P(|W_n^m| > w^* | \mathrm{H}_0 \text{ est vraie})$$

est (très) petite.

La statistique W_n^m étant sous H_0 approximativement normale $\mathcal{N}(0; 1)$, la valeur critique w^* correspondant au seuil α du test est pareillement approximativement égale au quantile $z_{1-\frac{\alpha}{2}}$.



Graphique 11.10 : La règle de décision du test bilatéral d'une moyenne

Propriété 11.4 *Soit* $(X_1, X_2, ..., X_n)$ *un échantillon aléatoire relatif à une variable aléatoire* X *d'espérance* m *et de variance* σ^2 *inconnues. Pour* n *grand* $(n \geq 30)$*, un* *test approximatif au seuil α de H_0 : $m = m_0$ contre H_1 : $m \neq m_0$ est donné par la règle de décision :*

$$\begin{cases} RH_0 & si\ |W_n^m| = \left|\frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}}\right| > z_{1-\frac{\alpha}{2}} \\ NRH_0 & sinon \end{cases} \tag{11.15}$$

Notez bien que cette propriété est valable quelle que soit la loi de X.

De façon semblable au cas du test bilatéral d'une fréquence, l'application de la règle de décision (11.15) à l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose consiste à rejeter H_0 si la réalisation observée $\widehat{w}^m = (\widehat{m} - m_0)/\sqrt{\frac{\widehat{s}^2}{n}}$ de la statistique de test W_n^m est en valeur absolue supérieure à $z_{1-\frac{\alpha}{2}}$, c'est-à-dire si l'estimation $\widehat{m}$ obtenue est "assez largement différente" de m_0, et à ne pas rejeter H_0 sinon. Le sens du terme "assez largement différente" est évidemment à apprécier en fonction de la valeur du dénominateur de $\widehat{w}^m$ qui, notez-le, dépend de l'estimation $\widehat{s}^2$ de la variance σ^2 dans la population.

Par construction, le risque première espèce associé à la règle de décision (11.15) est (approximativement) égal au seuil α du test, et est donc entièrement sous notre contrôle. Nous laisserons au lecteur le soin de vérifier (faites-le en raisonnant graphiquement, comme à la Section 11.1.1) que la puissance $1 - \beta$ du test (ou ce qui revient au même le risque β de deuxième espèce) dépend du seuil α du test (toute autre chose étant égale, plus α est petit, moins la puissance est grande), de l'importance de la fausseté de l'hypothèse nulle (plus l'écart $|m_1 - m_0|$ est grand, plus la puissance est grande) et de la précision d'estimation (plus la précision d'estimation est grande, c'est-à-dire plus l'écart-type $\sigma(\overline{X}_n)$ de l'estimateur $\overline{X}_n$ est petit, ce qui est le cas plus n est grand et σ^2 est petit, plus la puissance est grande), que quels que soient $\alpha > 0$ et $|m_1 - m_0| > 0$, elle tend vers 1 lorsque n tend vers l'infini, et qu'elle n'est pas sous notre contrôle : α étant fixé, elle dépend de n, de m_0, de la valeur inconnue de σ^2 et de la valeur inconnue de $m = m_1$ sous H_1. Il va sans dire que les conclusions à tirer du fait que l'erreur de première espèce soit sous contrôle et que l'erreur de deuxième espèce ne le soit pas sont les mêmes que précédemment (refaites le même raisonnement qu'à la Section 11.1.1). Bref, on le voit, rien de fondamentalement nouveau par rapport au cas du test bilatéral d'une fréquence.

11.2.2. Tests unilatéraux

Considérons à présent le cas du test unilatéral à droite :

$$\mathrm{H}_0 : m \leq m_0 \quad \text{contre} \quad \mathrm{H}_1 : m > m_0$$

Des résultats de la section précédente, on sait que sous H_0 : $m \leq m_0$ avec $m = m_0$, la statistique W_n^m suit (approximativement) une loi normale $\mathcal{N}(0; 1)$ (cf. équation (11.12)), que sous H_1 : $m = m_1 > m_0$, elle est (approximativement) distribuée selon une loi normale d'espérance positive et d'écart-type égal à 1 (cf. équation (11.13)), et on déduit immédiatement que sous H_0 : $m = m_0' < m_0$, on a :

$$W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} \approx \mathcal{N}\left(\frac{m_0'-m_0}{\sqrt{\frac{\sigma^2}{n}}}; 1\right) \tag{11.16}$$

autrement dit que W_n^m suit une loi normale d'espérance négative et d'écart-type également égal à 1.(cf. Graphique 11.11).

On constate à nouveau que le comportement de la statistique de test W_n^m sous H_0 et sous H_1 ressemble comme deux gouttes d'eau à celui de la statistique W_n^f discuté à la Section 11.1.3.1, de sorte que le même raisonnement que celui développé dans cette section conduit à nouveau à retenir, pour choisir de rejeter ou non H_0, le même type de règle de décision :

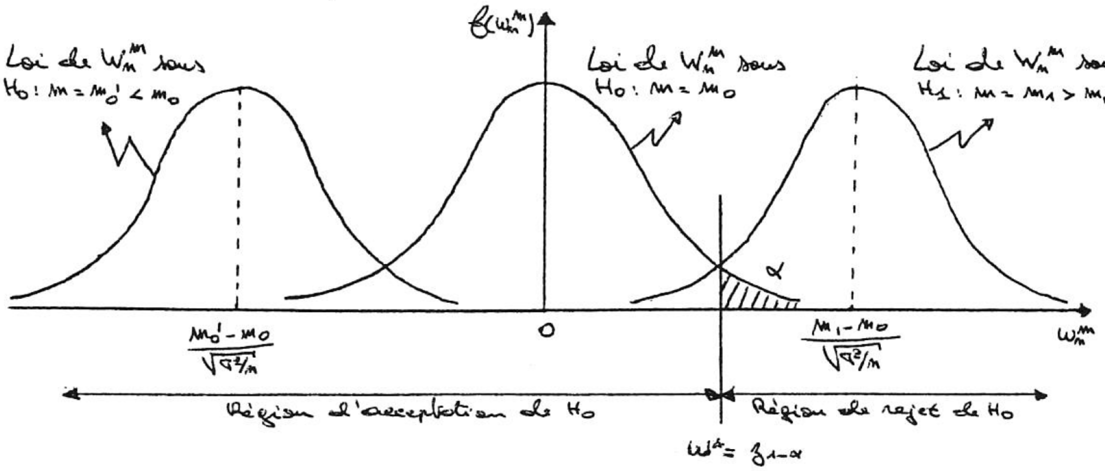
$$\begin{cases} \text{RH}_0 & \text{si } W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} > w^* \\ \text{NRH}_0 & \text{sinon} \end{cases} \tag{11.17}$$

où la valeur critique w^* est telle que la probabilité α — le seuil ou encore le niveau (de signification) du test — de rejeter H_0 alors que H_0 est vraie avec $m = m_0$:

$$\alpha = I\!P(W_n^m > w^* | \text{H}_0 \text{ est vraie avec } m = m_0)$$

est (très) petite.

La statistique W_n^m étant sous H_0 vraie avec $m = m_0$ approximativement normale $\mathcal{N}(0; 1)$, la valeur critique w^* correspondant au seuil α du test est ici pareillement approximativement égale au quantile $z_{1-\alpha}$.



Graphique 11.11 : La règle de décision du test unilatéral à droite d'une moyenne

Le cas du test unilatéral à gauche :

$$\text{H}_0 : m \geq m_0 \quad \text{contre} \quad \text{H}_1 : m < m_0$$

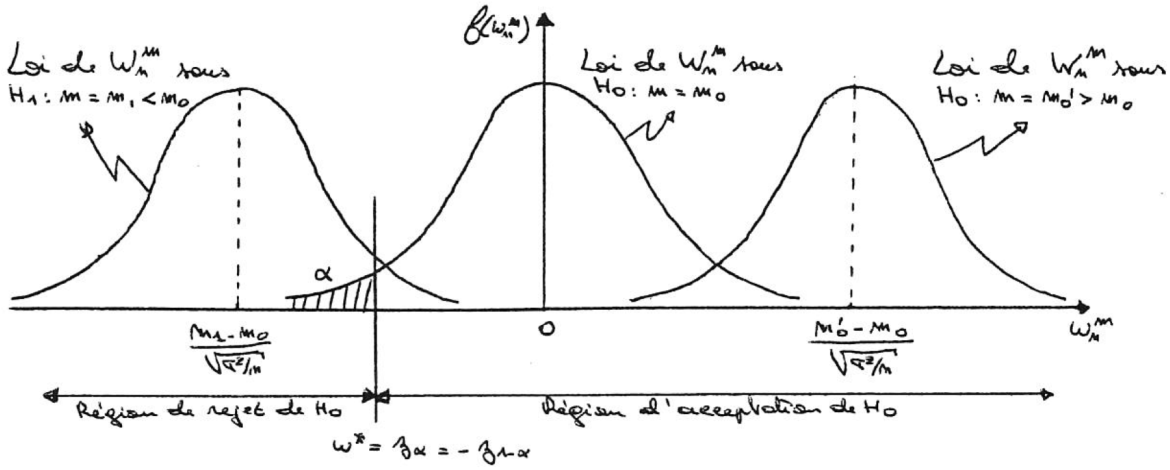
est parfaitement symétrique. La règle de décision est dans ce cas :

$$\begin{cases} \mathrm{RH}_0 & \text{si } W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} < w^* \\ \mathrm{NRH}_0 & \text{sinon} \end{cases} \tag{11.18}$$

où la valeur critique w^* est telle que la probabilité α — le seuil ou encore le niveau (de signification) du test — de rejeter H_0 alors que H_0 est vraie avec $m = m_0$:

$$\alpha = I\!P(W_n^m < w^* | \mathrm{H}_0 \text{ est vraie avec } m = m_0)$$

est (très) petite.



Graphique 11.12 : La règle de décision du test unilatéral à gauche d'une moyenne

La valeur critique w^* correspondant au seuil α du test est dans ce cas, comme dans le cas du test unilatéral à gauche d'une fréquence, approximativement égale au quantile z_α, quantile qui par ailleurs est égal à l'opposé du quantile $z_{1-\alpha}$ $(= -z_{1-\alpha})$.

Propriété 11.5 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire X d'espérance m et de variance σ^2 inconnues. Pour n grand $(n \geq 30)$, un test approximatif au seuil α de $H_0 : m \leq m_0$ contre $H_1 : m > m_0$ est donné par la règle de décision :*

$$\begin{cases} RH_0 & si\ W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} > z_{1-\alpha} \\ NRH_0 & sinon \end{cases} \tag{11.19}$$

et un test approximatif au seuil α de $H_0 : m \geq m_0$ contre $H_1 : m < m_0$ est donné par la règle de décision :

$$\begin{cases} RH_0 & si\ W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} < z_\alpha\ (= -z_{1-\alpha}) \\ NRH_0 & sinon \end{cases} \tag{11.20}$$

Notez bien que, comme la Propriété 11.4, cette propriété est valable quelle que soit la loi de X.

De façon semblable au cas du test unilatéral à droite (resp. à gauche) d'une fréquence, l'application de la règle de décision (11.19) (resp. (11.20)) à l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose consiste à rejeter H_0 si la réalisation observée $\widehat{w}^m = (\widehat{m} - m_0)/\sqrt{\frac{\widehat{s}^2}{n}}$ de la statistique de test W_n^m est supérieure à $z_{1-\alpha}$ (resp. inférieure à $z_\alpha = -z_{1-\alpha}$), c'est-à-dire si l'estimation $\widehat{m}$ obtenue est "assez largement supérieure" à m_0 (resp. "assez largement inférieure" à m_0), et à ne pas rejeter H_0 sinon. Notez à nouveau que le sens du terme "assez largement supérieure" (resp. "assez largement inférieure") est à apprécier en regard de la valeur du dénominateur de $\widehat{w}^m$.

Nous laisserons au lecteur le soin de vérifier que le risque première espèce associé aux règles de décision (11.19) et (11.20) est au maximum (approximativement) égal au seuil α du test, et est donc sous notre contrôle. Nous lui laisserons également le soin de vérifier (faites ces vérifications en raisonnant graphiquement, comme à la Section 11.1.3.1) que la puissance $1 - \beta$ de ces tests (ou ce qui revient au même le risque β de deuxième espèce) dépend du seuil α du test (toute autre chose étant égale, plus α est petit, moins la puissance est grande), de l'importance de la fausseté de l'hypothèse nulle (plus l'écart entre m_1 et m_0 est grand, plus la puissance est grande) et de la précision d'estimation (plus la précision d'estimation est grande, c'est-à-dire plus l'écart-type $\sigma(\overline{X}_n)$ de l'estimateur $\overline{X}_n$ est petit, ce qui est le cas plus n est grand et σ^2 est petit, plus la puissance est grande), que quels que soient $\alpha > 0$ et $m_1 - m_0 \neq 0$ (positif pour le test à droite et négatif pour le test à gauche), elle tend vers 1 lorsque n tend vers l'infini, et qu'elle n'est pas sous notre contrôle : α étant fixé, elle dépend de n, de m_0, de la valeur inconnue de σ^2 et de la valeur inconnue de $m = m_1$ sous H_1. Il est clair que les conclusions à tirer du fait que l'erreur de première espèce soit sous contrôle et que l'erreur de deuxième espèce ne le soit pas sont à nouveau les mêmes que précédemment (pour en être sûr, refaites le raisonnement). Bref, on le voit, à nouveau rien de fondamentalement neuf par rapport aux cas des tests unilatéraux d'une fréquence.

11.2.3. Cas d'une population normale

Dans le cas spécifique où la loi de la population est normale, c'est-à-dire lorsque $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$, on peut établir des versions exactes — valables quel que soit la taille d'échantillon n (et non plus seulement pour n grand) — des trois tests d'hypothèses relatifs à la moyenne (bilatéral, unilatéral à droite et unilatéral à gauche) obtenus ci-dessus.

Pour établir ces trois tests d'hypothèses, nous nous sommes appuyés sur le fait que, pour n grand ($n \geq 30$) et quelle que soit la loi de X, lorsque $m = m_0$ (que $m = m_0$ forme la totalité de H_0 ou en constitue seulement le bord), on a approximativement :

$$W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} \approx \mathcal{N}(0;1) \tag{11.21}$$

alors que, lorsque $m = m_* \neq m_0$ (que $m = m_* \neq m_0$ appartienne à H_0 ou à H_1), on a approximativement :

$$W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} \approx \mathcal{N}\left(\frac{m_* - m_0}{\sqrt{\frac{\sigma^2}{n}}}; 1\right) \tag{11.22}$$

Lorsque la loi de la population est normale $\mathcal{N}(m;\sigma^2)$, on dispose de résultats plus précis. De la Propriété 9.22, sous l'hypothèse de normalité et quel que soit n, lorsque $m = m_0$, on a de façon exacte :

$$W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} \rightsquigarrow t(n-1) \tag{11.23}$$

et on peut montrer que, lorsque $m = m_* \neq m_0$, on a également de façon exacte :

$$W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} \rightsquigarrow t(\delta^*; n-1) \tag{11.24}$$

où $t(\delta^*; n-1)$ désigne une *loi de* STUDENT *non centrale à $n-1$ degrés de liberté*[22], dont le *paramètre de non centralité* δ^* est donnée par : $\delta^* = (m_* - m_0) / \sqrt{\frac{\sigma^2}{n}}$.

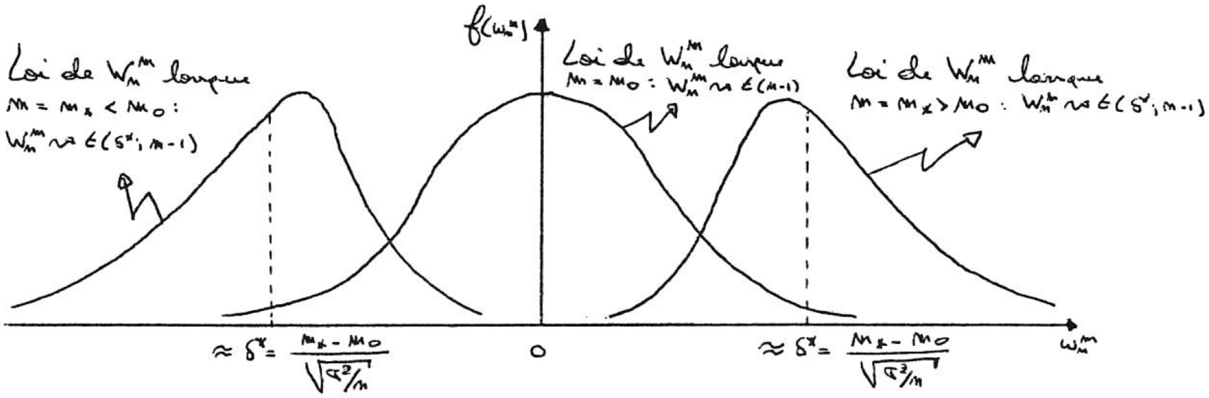
On le sait, la loi de STUDENT $t(\nu)$ ressemble beaucoup à la loi normale $\mathcal{N}(0;1)$: elle a simplement des queues plus épaisses (d'autant plus épaisses que ν est petit, cf. Chapitre 7). De façon similaire, la loi STUDENT non centrale $t(\delta;\nu)$ ressemble à une loi normale $\mathcal{N}(\delta;1)$: elle possède une espérance proche de δ mais a des queues plus épaisses que la loi $\mathcal{N}(\delta;1)$ (d'autant plus épaisses que ν est petit) et est quelque peu asymétrique (d'autant plus asymétrique, à droite si $\delta > 0$, à gauche si $\delta < 0$, que ν est petit ; cf. Graphique 11.13).

On voit que le comportement en échantillon fini sous la normalité de la statistique W_n^m n'est pas très différent de son comportement en grand échantillon, valable quel que soit la loi de X. Lorsque n tend vers l'infini, la loi STUDENT $t(n-1)$ tend vers la loi normale standardisée $\mathcal{N}(0;1)$ (cf. Propriété 9.7) et on peut de même montrer que la loi STUDENT non centrale $t(\delta^*; n-1)$ tend vers la loi normale $\mathcal{N}(\delta^*;1)$, sorte que, comme il se doit, on retrouve les résultats asymptotiques (11.21) et (11.22).

Clairement, les résultats exacts (11.23) et (11.24) ne remettent nullement en cause la forme des règles de décisions adoptées pour nos trois tests relatifs à la

[22] La loi de STUDENT non centrale $t(\delta;\nu)$ est formellement définie de la façon suivante. Si $X \rightsquigarrow \mathcal{N}(\delta;1)$, si $Y \rightsquigarrow \chi^2(\nu)$ et si X et Y sont indépendantes, alors la variable aléatoire $Z = X/\sqrt{\frac{Y}{\nu}} \rightsquigarrow t(\delta;\nu)$.

moyenne : pour le test bilatéral, la règle (11.14), pour le test unilatéral à droite, la règle (11.17), et pour le test unilatéral à gauche, la règle (11.18).



Graphique 11.13 : Loi de W_n^m sous la normalité et n petit ($n < 30$)

En s'appuyant sur le fait que, lorsque $m = m_0$, c'est-à-dire dans le cas du test bilatéral sous H_0 : $m = m_0$ et dans le cas des tests unilatéraux sous H_0 vraie avec $m = m_0$, la statistique W_n^m suit une loi de STUDENT $t(n-1)$, on peut déterminer les valeurs critiques w^* exactes sous la normalité correspondant au seuil α de ces trois tests. Pour $Y \rightsquigarrow t(\nu)$, on a (cf. Chapitres 7 et 10) :

$$I\!P\left(-t_{\nu;1-\frac{\alpha}{2}} \leq Y \leq t_{\nu;1-\frac{\alpha}{2}}\right) = 1-\alpha \quad \Leftrightarrow \quad I\!P\left(|Y| > t_{\nu;1-\frac{\alpha}{2}}\right) = \alpha$$

$$I\!P(Y \leq t_{\nu;1-\alpha}) = 1-\alpha \quad \Leftrightarrow \quad I\!P(Y > t_{\nu;1-\alpha}) = \alpha$$

$$I\!P(Y \leq t_{\nu;\alpha}) = \alpha = I\!P(Y \leq -t_{\nu;1-\alpha})$$

où $t_{\nu;\alpha}$, $t_{\nu;1-\alpha}$ et $t_{\nu;1-\frac{\alpha}{2}}$ sont les quantiles d'ordre respectivement α, $1-\alpha$ et $1-\frac{\alpha}{2}$ de la loi de STUDENT à ν degrés de liberté $t(\nu)$. On en déduit que les valeurs critiques w^* exactes correspondant au seuil α du test bilatéral, du test unilatéral à droite et du test unilatéral à gauche sont sous l'hypothèses de normalité respectivement égales aux quantiles $t_{n-1;1-\frac{\alpha}{2}}$, $t_{n-1;1-\alpha}$ et $t_{n-1;\alpha}$, ce dernier quantile étant par ailleurs égal à l'opposé du quantile $t_{n-1;1-\alpha}$ ($t_{n-1;\alpha} = -t_{n-1;1-\alpha}$).

Propriété 11.6 *Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire relatif à une variable aléatoire $X \rightsquigarrow \mathcal{N}(m; \sigma^2)$, où m et σ^2 sont inconnus. Un test au seuil α de H_0 : $m = m_0$ contre H_1 : $m \neq m_0$ est donné par la règle de décision :*

$$\begin{cases} RH_0 & si\ |W_n^m| = \left|\frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}}\right| > t_{n-1;1-\frac{\alpha}{2}} \\ NRH_0 & sinon \end{cases} \tag{11.25}$$

un test au seuil α de H_0 : $m \leq m_0$ contre H_1 : $m > m_0$ est donné par la règle de décision :

$$\begin{cases} RH_0 & si\ W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} > t_{n-1;1-\alpha} \\ NRH_0 & sinon \end{cases} \tag{11.26}$$

et un test au seuil α de $H_0: m \geq m_0$ contre $H_1: m < m_0$ est donné par la règle de décision :

$$\begin{cases} RH_0 & si\ W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} < t_{n-1;\alpha}\ (= -t_{n-1;1-\alpha}) \\ NRH_0 & sinon \end{cases} \tag{11.27}$$

Ne s'appuyant sur aucune approximation, cette propriété est exacte, valable quelle que soit la taille n de l'échantillon.

L'application des règles de décision (11.25), (11.26) et (11.27) à l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose consiste, de façon semblable aux règles de décision (11.15), (11.19) et (11.20), à rejeter H_0 si la réalisation observée $\widehat{w}^m = (\widehat{m} - m_0)/\sqrt{\frac{\widehat{s}^2}{n}}$ de la statistique de test W_n^m est, respectivement, en valeur absolue supérieure à $t_{n-1;1-\frac{\alpha}{2}}$, supérieur à $t_{n-1;1-\alpha}$ et inférieure à $t_{n-1;\alpha}$ $(= -t_{n-1;1-\alpha})$: seules les valeurs critiques à considérer sont modifiées ($t_{n-1;1-\frac{\alpha}{2}}$, $t_{n-1;1-\alpha}$ et $t_{n-1;\alpha}$ $(= -t_{n-1;1-\alpha})$ au lieu de respectivement $z_{1-\frac{\alpha}{2}}$, $z_{1-\alpha}$ et z_α $(= -z_{1-\alpha})$).

Par construction, le risque de première espèce associé à la règle de décision (11.25) est (exactement) égal au seuil α de ce test. On vérifie par ailleurs aisément (raisonnez graphiquement, en vous appuyant sur le Graphique 11.13) que le risque de première espèce associé aux règles de décision (11.26) et (11.27) est au maximum (exactement) égal au seuil α de ces tests et que les déterminants de la puissance (ou ce qui revient au même du risque de deuxième espèce) des trois tests sont identiques à ceux de leur analogue donnés par les Propriétés 11.4 et 11.5. D'où, pour chacun de ces trois tests, toujours la même interprétation à attacher à un rejet et un non rejet de H_0 pour l'échantillon d'observations $(x_1, x_2, ..., x_n)$ dont on dispose.

A l'aide de la Propriété 11.6, on peut effectuer des tests relatifs à la moyenne m dans le cas de petit échantillons ($n < 30$), ce que ne nous permettait pas les Propriétés 11.4 et 11.5. Evidemment, elle n'offre cette possibilité que pour le cas où la loi de la population est normale.

Pour n est grand ($n \geq 30$), en conséquence de la convergence de loi de STUDENT vers la loi normale, les règles de décisions (11.25), (11.26) et (11.27) ne se différentie guère des règles de décision (11.15), (11.19) et (11.20) : pour n grand, la loi de STUDENT $t(n-1)$ est très proche de la loi normale standardisée $\mathcal{N}(0;1)$, de sorte que les quantiles $t_{n-1;1-\frac{\alpha}{2}}$ et $z_{1-\frac{\alpha}{2}}$, $t_{n-1;1-\alpha}$ et $z_{1-\frac{\alpha}{2}}$, et $t_{n-1;\alpha}$ et z_α sont eux-mêmes très proches. Ainsi par exemple, pour $\alpha = 0,05$ et $n = 30$, on a $t_{29;0{,}975} = 2,042$ et $z_{0{,}975} = 1,96$, $t_{29;0{,}95} = 1,699$ et $z_{0{,}95} = 1,6449$, et $t_{29;0{,}05} = -t_{29;0{,}95} = -1,699$ et $z_{0{,}05} = -z_{0{,}95} = -1,6449$. On le voit, la différence entre les quantiles — qui va en s'amenuisant lorsque n augmente — est très faible. Pour n grand, il est donc pour l'essentiel indifférent d'utiliser l'une ou l'autre forme des tests d'hypothèses relatifs à m. Ainsi, dans la pratique, on constate que les tests donnés par la Propriété 11.6 sont souvent aussi utilisés, lorsque n est grand et quelle que soit la loi de X, en lieu et place de leur analogue donnés par les Propriétés 11.4 et 11.5.

11.2.4. Exercice résolu

La durée moyenne de survie des patients atteints d'une grave maladie et traités par un traitement classique est de $3,75$ années. On suppose que cette durée est distribuée de façon normale. Un nouveau traitement est mis au point et testé sur un échantillon aléatoire de 10 patients atteints de cette grave maladie. Lors de cet essai clinique, on observe une durée de survie moyenne des patients $\widehat{m} = 4,25$ années, avec un écart-type $\widehat{s} = 9$ mois.

1- Sur base de ce résultat préliminaire, le responsable de l'essai clinique se demande s'il y a lieu de poursuivre le développement du nouveau traitement. Suggérez à ce responsable un test d'hypothèse pour évaluer s'il semble valoir la peine de poursuivre ses recherches.

   Notons X la durée de vie d'un patient atteint de la maladie grave et traité à l'aide du nouveau traitement. On a : $X \rightsquigarrow \mathcal{N}(m;\sigma^2)$ (= loi de la population), où m et σ^2 sont inconnus.
   Il n'est raisonnable de poursuivre les recherches concernant ce nouveau traitement que si les résultats de l'essai clinique préliminaire suggère que le nouveau traitement accroît effectivement la durée moyenne de survie des patients. Pour évaluer ce point, le responsable devrait tester :

   $$\mathrm{H}_0 : m \leq m_0 \text{ contre } \mathrm{H}_1 : m > m_0$$

   où $m_0 = 3,75$, et poursuivre ses recherches en cas de rejet de H_0.

2- Sachant que les coûts impliqués par la poursuite des recherches sont très élevés, lequel des deux seuils $\alpha = 0,05$ et $\alpha = 0,01$ conseillez-vous au responsable de choisir pour réaliser ce test ? Pourquoi ?

   Les coûts impliqués par la poursuite des recherches étant très élevés, le rejet de H_0 est une décision très coûteuse. Pour cette raison, il convient donc de restreindre la probabilité de prendre à tort cette décision, c'est-à-dire de limiter le risque de première espèce. Le responsable devrait donc choisir $\alpha = 0,01$ plutôt que $\alpha = 0,05$.

3- Si le responsable suit vos conseils, que décidera-t-il ?

   De la Propriété 11.6, un test au seuil α de $\mathrm{H}_0 : m \leq m_0$ contre $\mathrm{H}_1 : m > m_0$ est donné par la règle de décision :

   $$\begin{cases} \mathrm{RH}_0 & \text{si } W_n^m = \frac{\overline{X}_n - m_0}{\sqrt{\frac{S_n^2}{n}}} > t_{n-1;1-\alpha} \\ \mathrm{NRH}_0 & \text{sinon} \end{cases}$$

   Pour $\alpha = 0,01$ et $n = 10$, on trouve dans les tables $t_{9;0,99} = 2,821$. La réalisation observée $\widehat{w}^m$ de la statistique W_n^m est :

$$\widehat{w}^m = \frac{\widehat{m} - m_0}{\sqrt{\frac{\widehat{s}^2}{n}}} = \frac{4,25 - 3,75}{\sqrt{\frac{(0,75)^2}{10}}} \approx \frac{0,5}{0,2371} \approx 2,108$$

On a $\widehat{w}^m \approx 2,108 < 2,821 = t_{9;0,99}$. On ne rejette pas H_0. Le responsable ne poursuivra donc pas ses recherches.

4- Qu'aurait-il décidé si il avait choisi l'autre seuil ?

Pour $\alpha = 0,05$ et $n = 10$, on trouve dans les tables $t_{9;0,95} = 1,833$. Pour ce risque de première espèce (maximum) plus important, on a $\widehat{w}^m \approx 2,108 > 1,833 = t_{9;0,95}$, de sorte qu'on rejette H_0. A ce seuil de 5%, le responsable aurait donc choisi de poursuivre ses recherches.

5- Le responsable de l'essai clinique hésite car il est persuadé, même si cela n'apparaît pas de façon très nette, que le nouveau traitement est plus efficace que l'ancien. Que lui conseillez-vous de faire ?

De lever l'ambiguïté du résultat obtenu dans l'essai préliminaire en le complétant par un essai sur un échantillon aléatoire supplémentaire de patients. La puissance du test s'accroissant avec n, si le nouveau traitement est effectivement meilleur ($m = m_1 > 3,75$), la probabilité que cela apparaisse lors du test sera plus grande (ou ce qui revient au même, la probabilité que cela n'apparaisse pas sera plus petite).

11.3. Exercices

1- Soient une population représentée par variable aléatoire X suivant une loi de Bernoulli $\mathcal{B}(p)$, où p est inconnu, et un échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à X. On considère le test unilatéral au seuil $\alpha = 5\%$ de $\text{H}_0 :\ p \geq 0,5$ contre $\text{H}_1 :\ p < 0,5$ (cf. Propriété 11.3). Calculez le risque de deuxième espèce et la puissance du test pour :

a- $n = 100$ et $p = p_1 = 0,35$.
b- $n = 100$ et $p = p_1 = 0,45$.
c- $n = 1000$ et $p = p_1 = 0,40$.
d- $n = 1000$ et $p = p_1 = 0,45$.
e- $n = 1000$ et $p = p_1 = 0,48$.
f- $n = 10\,000$ et $p = p_1 = 0,48$.
g- $n = 10\,000$ et $p = p_1 = 0,49$.

Interprétez les résultats de vos calculs.

2- Lors des dernières élections, un parti a recueilli 16,6% des voix. Six mois plus tard, on tire un échantillon aléatoire de 144 électeurs et on compte le nombre d'électeurs déclarant qu'il voterait à nouveau pour ce parti en cas de nouvelles élections.

a- Proposez un test d'hypothèse pour évaluer si, en cas de nouvelles élections, ce parti referait le score qu'il avait fait lors des dernières élections.

b- Sur les 144 électeurs interrogés, 32 déclarent qu'il voterait à nouveau pour ce parti en cas de nouvelles élections.
   i- Testez au seuil $\alpha = 5\%$ si, en cas de nouvelles élections, ce parti referait le score qu'il avait fait lors des dernières élections.
   ii- Comment interprétez-vous α ?

3- Une loi de l'Etat de New-York stipule que les jurys d'assises doivent être composés d'au moins 15% de membres des minorités. Une association de défense des minorités a constaté que sur un échantillon aléatoire de 85 personnes tiré parmi les individus constituant les jurys de cet Etat au cours de l'année écoulée, seulement 9 appartenaient à une minorité. L'association se demande si elle doit ou non intenter un procès à l'Etat qui affirme que la loi est respectée.
   a- Effectuez un test au seuil $\alpha = 0,05$ et formulez vos recommandations à l'association.
   b- Comment interprétez-vous α ?

4- Une minorité forme 20% de la population d'un pays. On se demande si cette minorité est représentée fidèlement (c'est-à-dire ni sous-représentée, ni sur-représentée) dans une grande administration. On prélève un échantillon aléatoire de 200 employés de cette administration et on observe le nombre de membres de la minorité qu'il contient.
   a- Proposez un test pour évaluer si la minorité est représentée fidèlement dans l'administration.
   b- Pour quelles valeurs de la proportion observée du nombre de membres de la minorité dans l'échantillon dont on dispose ne rejettera-t-on pas, au seuil de 10%, l'hypothèse nulle H_0 ?
   c- On constate que l'échantillon comprend 30 personnes appartenant à la minorité. Sur base de votre réponse au point b, quelle conclusion tirez-vous ?
   d- Quelle aurait été cette conclusion si on avait choisi un seuil égal à 5% ? Expliquez pourquoi le choix de ce seuil plus petit entraîne une conclusion différente.
   e- Si la taille de l'échantillon avait été de 1 000, pour quelles valeurs de la proportion observée de membres de la minorité dans l'échantillon n'aurait-on pas, au seuil de 10%, rejeté l'hypothèse nulle H_0 ? Comparez votre réponse à celle trouvée au point b et interprétez.

5- L'administration affirme que la minorité en question n'est pas sous-représentée. Un organe de presse de la minorité décide de tester, au seuil de 5%, l'assertion de l'administration et fait procéder à un sondage. Un échantillon aléatoire de 500 employés de l'administration est prélevé et le nombre de membres de la minorité qu'il contient est compté.
   a- Définissez les hypothèses H_0 et H_1 du test auquel l'organe de presse a décidé de procéder.
   b- Pour quelles valeurs de la proportion observée de membres de la minorité dans l'échantillon dont il dispose l'organe de presse rejettera-t-il H_0 ?
   c- L'échantillon prélevé contient 83 membres de la minorité. Quelle est dans le cas présent la conclusion de l'organe de presse ?

d- Quelle aurait été sa conclusion si il avait choisi un seuil de 1% ? Expliquez pourquoi le choix de ce seuil plus petit entraîne une conclusion différente.
e- Si la taille de l'échantillon avait été de 2 000, pour quelles valeurs de la proportion observée de membres de la minorité dans l'échantillon l'organe de presse aurait-il rejeté H_0 ? Comparez votre réponse à celle trouvée au point b et interprétez.

6- Un candidat à une élection présidentielle pense qu'il va obtenir une majorité absolue dès le premier tour, mais il voudrait en avoir une confirmation avant solliciter le soutien de certains sponsors. Il fait réaliser un sondage auprès de 200 électeurs. 31 personnes sans opinion sont écartées. Il reste 169 personnes dont 96 affirment avoir l'intention de voter pour le candidat considéré.
   a- L'institut de sondage peut-il lui confirmer, au seuil de 5%, qu'en effet il devrait obtenir une majorité absolue dès le premier tour ?
   b- Et avec un seuil égal à 1% ?

   Interprétez vos résultats.

7- Un institut d'études agronomiques étudie la résistance des poules à une certaine maladie. On inocule le virus de la maladie à 100 poules choisies de façon aléatoire, et on constate que 62 survivent.
   a- Donnez une estimation du taux p de résistance des poules à cette maladie.
   b- Donnez un intervalle de confiance de niveau 98% pour le taux p de résistance des poules à cette maladie.
   c- Dix ans plus tard, la maladie ayant été bien étudiée, on considère le taux de résistance comme connu : il vaut $0,6$. Un chercheur de l'institut a mis au point une substance chimique, la dextro-AP3, dont il pense que, mélangée à la nourriture des poules, elle doit augmenter leur résistance. N'ayant pas obtenu des crédits suffisants pour travailler sur un échantillon plus grand, il administre la dextro-AP3 pendant un mois à un échantillon aléatoire de 60 poules, puis il leur inocule le virus. Il constate que 41 survivent.
      i- Peut-il conclure, au seuil de 5%, que de la dextro-AP3 est efficace ?
      ii- Le chercheur reste persuadé la dextro-AP3 est efficace Que devrait-t-il faire, à votre avis ?

8- Dans le cadre d'une étude de marketing, on analyse la consommation annuelle de lessive des ménages. Sur la base d'un échantillon aléatoire de 100 ménages, on a observé une consommation moyenne $\widehat{m} = 14$ kilos avec un écart-type $\widehat{s} = 2$ kilos. D'autre part, 30% des personnes interrogées accordaient leur préférence à un produit sans phosphates.
   a- Construisez un intervalle de confiance de niveau 95% pour la consommation annuelle moyenne de lessive des ménages dans la population.
   b- Supposez que pour le même niveau de confiance (95%) on souhaite obtenir une estimation de cette consommation annuelle moyenne dans la population précise à $\pm$ 100 grammes près. Quelle devrait être approximativement la taille de l'échantillon n ?
   c- Un fabricant affirme qu'il n'y a pas plus d'un ménage sur quatre qui préfèrent la lessive sans phosphates. Testez au seuil $\alpha = 0,05$ l'assertion

de ce fabriquant. Que concluez-vous ?

9- Soient une population représentée par variable aléatoire X d'espérance m et de variance σ^2 inconnues et un échantillon aléatoire $(X_1, X_2, ..., X_n)$ relatif à X. On considère le test bilatéral au seuil $\alpha = 5\%$ de $\mathrm{H}_0 : m = 10$ contre $\mathrm{H}_1 : m \neq 10$ (cf. Propriété 11.4). Calculez le risque de deuxième espèce et la puissance du test pour :

a- $n = 100$, $m = m_1 = 9$ et $\sigma^2 = 25$.
b- $n = 100$, $m = m_1 = 11,5$ et $\sigma^2 = 25$.
c- $n = 100$, $m = m_1 = 9$ et $\sigma^2 = 100$.
d- $n = 300$, $m = m_1 = 9$ et $\sigma^2 = 25$.

Interprétez les résultats de vos calculs.

10- Vous êtes chargé(e) de vérifier une livraison de 5 000 sacs de terreau. Le cahier des charges impose au fournisseur que le poids moyen des sacs qu'il livre soit de 10 kg. Un échantillon de 100 sacs est tiré aléatoirement dans la livraison. Les 100 sacs sont pesés. On trouve $\widehat{m} = 10,1$ kg et $\widehat{s} = 0,5$ kg.

a- Testez au seuil de 5% si la livraison est conforme.
b- Qu'auriez-vous conclu, toujours au seuil de 5%, si on avait observé dans l'échantillon $\widehat{m} = 9,95$ kg et $\widehat{s} = 0,5$ kg ?

11- Dans un sol normal, la hauteur moyenne atteinte par les plants d'une certaine variété de maïs est de 85 cm. On plante un échantillon aléatoire de 100 semences de cette variété de maïs dans un sol enrichi pour voir si la hauteur peut être améliorée.

a- Proposez un test d'hypothèse permettant de vérifier si l'enrichissement du sol augmente effectivement la hauteur des plants de maïs.
b- On observe dans l'échantillon une hauteur moyenne $\widehat{m} = 88$ cm avec un écart-type $\widehat{s} = 10$ cm. Que peut-on en conclure, au seuil $\alpha = 0,01$?

12- Le propriétaire d'un commerce de détail change de stratégie de marketing. Il espère ainsi augmenter le montant moyen de ses transactions. Jusqu'à présent ce montant moyen était de 500 F. Il constate qu'après avoir mis en oeuvre sa nouvelle stratégie de marketing le montant moyen des transactions d'un échantillon aléatoire de 200 clients est de $\widehat{m} = 537$ F avec un écart-type $\widehat{s} = 100$ F. Peut-il en conclure, au seuil de 5%, que sa nouvelle stratégie de marketing augmente effectivement le montant moyen de ses transactions ?

13- La durée de vie moyenne d'un certain type de pile est de 100 h. Un chercheur a mis au point un nouveau procédé de fabrication destiné à augmenter notablement cette durée, et il la propose à une entreprise. Celle-ci est intéressée, et est disposée à exploiter le procédé sauf si elle a de bonnes raisons de penser que l'augmentation de durée de vie est inférieure à 50%. Pour le vérifier, elle teste un échantillon aléatoire de 50 piles fabriquées selon le nouveau procédé, et trouve une durée de vie moyenne $\widehat{m} = 148,31$ h, avec un écart-type $\widehat{s} = 12,05$ h. Effectuez un test au seuil $\alpha = 0,05$ et formulez vos recommandations à l'entreprise.

14- Le taux d'une certaine hormone dans le sang est de 2 mg/l en moyenne. On admet que d'un individu à l'autre, ce taux varie autour de cette valeur suivant une loi normale. Un médecin, qui a appliqué à 14 malades pris au hasard un nouveau traitement, se demande si celui-ci n'aurait pas pour effet secondaire de modifier ce taux (alors qu'on sait que la maladie pour laquelle il les soigne ne le modifie pas). Il constate sur ces quatorze malades un taux moyen $\widehat{m} = 1,80$ mg/l avec un écart-type $\widehat{s} = 0,4$ mg/l. Que peut-il, au seuil $\alpha = 5\%$, en conclure ?

15- Un fabricant de câbles d'acier garantit que la tension de rupture moyenne des câbles qu'il produit est de 3 tonnes. Une entreprise cliente voudrait s'assurer que cette tension de rupture moyenne n'est pas inférieure à la valeur garantie. Pour ce faire, elle fait l'essai de 10 câbles pris au hasard et constate qu'ils cassent pour des tensions dont la moyenne est $\widehat{m} = 2\,980$ kg, avec un écart-type $\widehat{s} = 40$ kg. On admet que la tension de rupture des câbles est distribuée de façon normale.

a- Quel test lui proposez-vous d'effectuer ?

b- Compte tenu du fait que la rupture d'un câble peut avoir des conséquences graves, lequel des deux seuils $\alpha = 0,05$ et $\alpha = 0,10$ lui conseillez-vous de choisir ? Pourquoi ?

c- Que conclura-t-elle si elle suit vos conseils ?

d- Qu'aurait-elle conclu si elle avait choisi l'autre seuil ?