



Faculté des Sciences  
Département de Mathématique

# Probabilités et statistiques de base

Notes provisoires

Pierre Mathonet

Formation Doctorale à l'Université de Liège  
Année académique 2017-2018

# Chapitre 1

## Probabilités

Il y a eu dans l'histoire des mathématiques plusieurs définitions successives ou parfois concurrentes de la notion de probabilité d'un événement dit aléatoire. J'aborde ici deux définitions possibles. La première approche est appelée *approche fréquentielle*. Elle présente l'avantage d'être intuitive, mais présente rapidement des difficultés mathématiques. L'autre approche, dite axiomatique, fournit des bases mathématiques solides, mais un peu plus abstraites et moins liées à l'expérience concrète que chacun a, par exemple, du lancé d'un dé. Je vais privilégier la deuxième approche, qui permet de mettre en place et d'étudier plus facilement les résultats mathématiques de la théorie. Les deux approches se rejoindront dans le deuxième chapitre, quand nous discuterons de la loi des grands nombres. Je profiterai également du début de ce chapitre pour rappeler des notions de théorie (naïve) des ensembles et d'analyse combinatoire.

Ayant défini la notion de probabilité, on passera aux probabilités conditionnelles, à la notion d'événements indépendants, et aux fameuses formules des probabilités totales et au théorème de Bayes.

Nous passerons ensuite à l'introduction de variables aléatoires et leur distribution de probabilité et fonction de répartition. Nous étudierons les paramètres classiques associés aux variables aléatoires que sont l'espérance mathématique et la variance. Nous étudierons les combinaisons de variables aléatoires, l'indépendance et son implication sur l'espérance et la variance de combinaisons classiques de variables. Les exemples utilisés sont délibérément simples.

### 1.1 Expériences aléatoires et espaces probabilisés

On appelle *expérience aléatoire ou phénomène fortuit* une expérience

- qui a plusieurs résultats possibles,
- telle qu'un seul résultat peut se produire à chaque réalisation de l'expérience,
- telle que le résultat est imprévisible,

L'ensemble des résultats possibles de l'expérience est généralement noté  $\Omega$  et est appelé *univers des possibles* ou encore *catégorie d'épreuve*.

**Exemple 1.1.1.** Voici une liste d'expériences aléatoires et les univers des possibles correspondants.

Expérience	$\Omega$ (Univers)
On lance une pièce	$\{P, F\}$
On lance un dé	$\{1, 2, 3, 4, 5, 6\}$
On lance deux dés discernables	$\{(i, j) : 1 \leq i, j \leq 6\}$
On lance deux dés indiscernables	$\{(i, j) : 1 \leq i \leq j \leq 6\}$
On lance une pièce jusqu'à pile	$\{P, FP, FFP, FFFP, \dots\}$
On mesure la durée d'une course	$[0; +\infty[$

Comme le montre les derniers exemples, l'univers peut contenir un nombre fini ou un nombre infini de points, et dans ce deuxième cas, il peut être dénombrable (on peut numéroter ses éléments) ou non dénombrable.

### 1.1.1 Événements

Dans un premier temps, on pourrait appeler événement tout sous-ensemble  $E$  de  $\Omega$ , c'est-à-dire un certain ensemble fixé de résultats possibles de l'expérience aléatoire que l'on considère. Cependant, pour pouvoir développer la théorie mathématique adéquate (notamment la théorie de la mesure, ou l'intégration), il est parfois nécessaire de ne considérer que certains sous-ensembles. On doit alors *définir* les événements que l'on va considérer, de façon telle que certaines opérations sur les événements soient bien définies. Les événements sont bien sûr les ensembles de résultats auxquels on va associer une probabilité.

**Exemple 1.1.2.** Voici un exemple d'expérience aléatoire et des événements correspondants (parmi beaucoup d'autres que l'on peut considérer).

Expérience	Événement
On lance un dé	le résultat est 6 : $E = \{6\}$
On lance un dé	le résultat est pair : $E = \{2, 4, 6\}$
On mesure un temps	le résultat est entre 1h et 2h : $E = [60\text{min}, 120\text{min}]$

On peut réaliser sur les événements les opérations que l'on définit habituellement sur les sous-ensembles. Ces opérations correspondent au “et”, “ou”, et à la négation qu'on utilise aussi bien en logique qu'en français courant.

1. Si  $E$  et  $F$  sont des événements, l'**union** de  $E$  et  $F$  est définie par

$$E \cup F = \{\omega \in \Omega : \omega \in E \text{ ou } \omega \in F\}.$$

Il est important de remarquer qu'en mathématique, le “ou” n'est *pas exclusif*. L'union de deux événements se réalise quand un des deux *au moins* se réalise.

2. Si  $E$  et  $F$  sont des événements, l'**intersection** de  $E$  et  $F$  est définie par

$$E \cap F = \{\omega \in \Omega : \omega \in E \text{ et } \omega \in F\}.$$

3. Si  $E$  est un événement, on définit l'**événement complémentaire** par

$$\bar{E} = \{\omega \in \Omega : \omega \notin E\}.$$

Cet événement est encore noté  $\Omega \setminus E$ . Le symbole  $\setminus$  est alors le “moins” ensembliste.

4. Il existe un événement particulier, appelé **événement vide** et noté  $\emptyset$ , défini par

$$\omega \notin \emptyset, \quad \forall \omega \in \Omega.$$

Nous dirons que des événements  $E_1$  et  $E_2$  sont **exclusifs** si  $E_1 \cap E_2 = \emptyset$ . Enfin, on appelle **événement élémentaire** un événement qui contient exactement un résultat possible de l'expérience aléatoire.

Notre but est évidemment d'associer à chaque événement sa probabilité, qui sera bien sûr un nombre. Pour des raisons techniques<sup>a</sup>, il est parfois nécessaire de restreindre l'ensemble des événements que l'on prend en considération. L'ensemble des événements, noté  $\mathcal{E}$  n'est alors plus égal à  $\mathcal{P}(\Omega)$  (l'ensemble de tous les sous-ensembles de  $\Omega$ ). On veut quand-même pouvoir associer une probabilité à une union d'événements, à une intersection, ou au complémentaire d'un événement. Si on note  $\mathcal{E}$  l'ensemble des sous-ensembles de  $\Omega$  auxquels on va *pouvoir* associer une probabilité, il faut donc qu'il ait certaines propriétés.

**Définition 1.1.1.** L'ensemble des événements  $\mathcal{E} \subset \mathcal{P}(\Omega)$  doit vérifier les propriétés suivantes

- $\Omega \in \mathcal{E}$ ,
- Si  $E \in \mathcal{E}$ , alors  $\overline{E} \in \mathcal{E}$ ,
- Si  $E_m \in \mathcal{E}, \forall m \in \mathbb{N}_0$ , alors  $\cup_{i=1}^{+\infty} E_m = E_1 \cup E_2 \cup \dots \in \mathcal{E}$ .

Un tel ensemble  $\mathcal{E}$  est appelé **tribu** ou  **$\sigma$ -algèbre**. Le couple  $(\Omega, \mathcal{E})$  est appelé **espace probabilisable**.

*Remarque 1.1.* Si l'univers  $\Omega$  est fini ou dénombrable<sup>b</sup>, alors on peut toujours choisir  $\mathcal{E} = \mathcal{P}(\Omega)$ , l'ensemble de tous les sous-ensembles de  $\Omega$ .

Voici quelques exemples :

1. Si  $\Omega = \{1, 2, 3, 4, 5, 6\}$  alors  $\mathcal{E} = \{\emptyset, \Omega\}$  est une tribu.
2. Si  $\Omega = \{1, 2, 3, 4, 5, 6\}$  alors  $\mathcal{E} = \mathcal{P}(\Omega)$  est une tribu.
3. Si  $\Omega = \{1, 2, 3, 4, 5, 6\}$  alors  $\mathcal{E} = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\}$  est une tribu.
4. Si  $\Omega = \{1, 2, 3, 4, 5, 6\}$  alors

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

est une tribu.

5. Si  $\Omega = \mathbb{R}$ , la **tribu de Borel** est la plus petite tribu qui contient tous les intervalles de type  $]a, b[$ , avec  $a < b$ .

On peut utiliser la théorie des ensembles pour démontrer le résultat suivant.

**Proposition 1.1.1.** Si  $\mathcal{E}$  est une tribu sur  $\Omega$ , alors

1. On a  $\emptyset \in \mathcal{E}$ .
2. Si  $E_m \in \mathcal{E}, \forall m \in \mathbb{N}_0$ , alors  $\cap_{i=1}^{\infty} E_m \in \mathcal{E}$ .
3. Si  $E_1, \dots, E_M \in \mathcal{E}$ , alors  $E_1 \cup \dots \cup E_M \in \mathcal{E}$ .
4. Si  $E_1, \dots, E_M \in \mathcal{E}$ , alors  $E_1 \cap \dots \cap E_M \in \mathcal{E}$ .

---

a. On montre en effet que si l'ensemble des événements est "trop grand", il n'est pas possible d'associer à chaque événement un nombre, de façon telle que des propriétés fondamentales listées ci-dessous soient toutes vraies.

b. Un ensemble est dénombrable si ses éléments peuvent être numérotés, c'est à dire s'il est en bijection (correspondance parfaite) avec une partie de l'ensemble des nombres naturels  $\mathbb{N}$ .

## 1.2 Probabilités : approche fréquentielle

Nous avons mis en place les notions nécessaires à la définition d'une probabilité (expérience – univers des possibles  $\Omega$  – ensemble des événements  $\mathcal{E}$ ). Nous allons maintenant tenter d'associer à chaque événement un nombre, qui définit la probabilité qu'il se réalise.

Considérons l'expérience aléatoire simple consistant à lancer un dé équilibré à 6 faces. Si on cherche à définir intuitivement la probabilité que le résultat obtenu soit 3, chacun répondra naturellement  $\frac{1}{6}$ . C'est très probablement lié aux propriétés de symétrie du problème : toutes les faces ont la même probabilité puisque le dé est symétrique, et donc si la probabilité totale est 1, chaque face a une probabilité  $\frac{1}{6}$ . Ce raisonnement était suivi dans l'*approche classique* des probabilités, qui a été la première approche, mais elle ne peut prendre en compte les expériences non symétriques, comme les points obtenus en lançant deux dés.

Essayons cependant d'analyser cette réponse  $\frac{1}{6}$ . Elle ne veut certainement pas dire que si on lance 6 fois le dé, on obtiendra toujours une fois le résultat 3. Il est évident qu'il est possible de lancer 50 fois le dé sans obtenir le résultat 3, mais c'est très peu "probable".

En fait, si on lance le même dé un certain nombre (disons  $n$ ) de fois et que l'on compte le nombre  $n(E)$  de fois que l'événement  $E$  : "le nombre 3 est sorti" se réalise, alors la fraction

$$\frac{n(E)}{n}$$

compte la proportion de réalisations de l'événement en question sur le nombre d'expériences réalisées. C'est la fréquence relative d'apparition de l'événement  $E$ . Elle mesure en quelque sorte une mesure de la "propension" de l'événement à se réaliser.

Le problème est bien sûr qu'il s'agit d'une fonction de  $n$ . Et que si chacun réalise cette expérience, il aura une fonction différente. Il est donc utile de voir ce que ces fonctions ont en commun.

Chacune d'elle semble tendre vers une valeur limite qui vaut  $\frac{1}{6}$ . Il faut cependant se méfier des semblances quand il s'agit du calcul des limites. On ne peut pas en fait conclure, en regardant l'allure de la fonction et du fait que ses valeurs semblent se rapprocher de  $\frac{1}{6}$ , que la limite est effectivement  $\frac{1}{6}$ .

En utilisant les notations mathématiques habituelles, on écrira

$$\lim_{n \rightarrow +\infty} \frac{n(E)}{n} = \frac{1}{6}.$$

Cette définition se généralise pour tout événement  $E$ , et on écrit alors

$$\mathbb{P}(E) = \lim_{n \rightarrow +\infty} \frac{n(E)}{n},$$

où  $n(E)$  est le nombre de fois que l'événement  $E$  s'est produit sur les  $n$  réalisations de l'expérience aléatoire dans les mêmes conditions, et  $n(E)/n$  la fréquence d'apparition de  $E$  lors des  $n$  premières expériences.

Cette approche a son intérêt, parce qu'au départ, on voit ce que l'on fait, concrètement. Elle est conforme à l'intuition. Cependant, l'introduction de la notion de limite donne une complication importante, qui est assez souvent passée sous silence. Pour avoir une notion de limite, on doit pouvoir faire un nombre arbitrairement grand d'expériences identiques. On peut pour cela recourir à des simulations, mais cela demande du temps de calcul. Il n'est pas sûr que la limite existe toujours, et qu'elle soit conforme à l'intuition.

De plus obtenir des résultats avec cette approche demande le recours systématique à la notion de limite, qui n'est pas le concept le plus évident des mathématiques.

Je privilégierai donc l'approche axiomatique, qui est exposée dans la section suivante, mais on retrouvera bien sûr l'approche fréquentiste au travers de la **loi des grands nombres**.

### 1.3 Probabilités : approche axiomatique

L'approche axiomatique des probabilités que nous étudions ici est essentiellement due à Kolmogorov (1903-1987). L'idée n'est pas de définir une probabilité comme la "limite" d'une fréquence sur un grand nombre d'observations, mais de donner les propriétés qu'une *mesure de probabilité* doit satisfaire. Nous verrons que dans le cas du dé, cela permet de *calculer* la probabilité d'obtenir "3", si on suppose que le dé est équilibré (en utilisant le raisonnement de l'approche classique des probabilités, mais au sein d'un système mathématique). Les propriétés imposées se justifient naturellement par les propriétés que l'on peut constater quand on essaie de définir les probabilités à l'aide de limites de fréquences relatives, comme dans l'approche fréquentielle.

**Définition 1.3.1.** Une probabilité sur un espace probabilisable  $(\Omega, \mathcal{E})$  est la donnée pour chaque événement  $E$  de  $\mathcal{E}$  d'un nombre  $\mathbb{P}(E)$ , appelé probabilité de  $E$ , et satisfaisant les conditions suivantes :

- Pour tout événement  $E$ , on a  $0 \leq \mathbb{P}(E) \leq 1$ .
- On a  $\mathbb{P}(\Omega) = 1$ ,
- Pour toute suite d'événements  $(E_1, E_2, \dots)$  deux à deux exclusifs<sup>c</sup>, on a

$$\mathbb{P}(\cup_{m=1}^{\infty} E_m) = \sum_{m=1}^{+\infty} \mathbb{P}(E_m).$$

Le triplet  $(\Omega, \mathcal{E}, \mathbb{P})$  est alors appelé *espace probabilisé*.

Voici quelques conséquences pratiques de la définition.

**Proposition 1.3.1.** Si  $(\Omega, \mathcal{E}, \mathbb{P})$  est un espace probabilisé alors

- on a toujours  $\mathbb{P}(\emptyset) = 0$  ;
- Si les événements  $E_1, \dots, E_M$  sont deux à deux exclusifs, alors

$$\mathbb{P}(E_1 \cup \dots \cup E_M) = \mathbb{P}(E_1) + \dots + \mathbb{P}(E_M);$$

- On a  $\mathbb{P}(\overline{E}) = 1 - \mathbb{P}(E)$  pour tout événement  $E$  ;
- On a  $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$  pour tous  $E, F \in \mathcal{E}$ .

L'utilité de ces conséquences est évidente. Par exemple, si on considère le jet d'un dé à six faces et équilibré, on a  $\Omega = \{1, 2, 3, 4, 5, 6\}$  et

$$\begin{cases} \mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) = \mathbb{P}(\{6\}) \\ \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}) + \mathbb{P}(\{3\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{5\}) + \mathbb{P}(\{6\}) = \mathbb{P}(\Omega) = 1 \end{cases}$$

donc les faces du dé ont toute une probabilité  $1/6$  d'apparaître. Si on considère l'événement "le dé marque un chiffre pair", on a

$$\mathbb{P}(\{2, 4, 6\}) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = 1/2,$$

tandis que l'événement "le dé marque un chiffre impair" aura également une probabilité  $1/2$ , par la troisième conséquence.

c. Cette suite est encore notée  $(E_m : m \in \mathbb{N}_0)$ . La condition d'exclusivité est  $E_i \cap E_j = \emptyset$  chaque fois que  $i$  et  $j$  sont distincts.

### 1.3.1 Exercices

Voici quelques exercices pour mettre en oeuvre les définitions et prolonger les résultats que nous avons vus jusqu'à présent.

1. Démontrer que si les événements  $E$  et  $F$  satisfont  $E \subset F$ , alors on a  $\mathbb{P}(E) \leq \mathbb{P}(F)$ .
2. Démontrer que pour tous événements  $E, F, G$ , on a

$$\begin{aligned} \mathbb{P}(E \cup F \cup G) &= \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) \\ &\quad - \mathbb{P}(E \cap F) - \mathbb{P}(E \cap G) - \mathbb{P}(F \cap G) + \mathbb{P}(E \cap F \cap G). \end{aligned}$$

3. Je lance un dé à 6 faces pipé (les résultats n'ont pas tous les mêmes probabilités) et je connais les probabilités suivantes  $\mathbb{P}(\{1, 2, 3\}) = 2/3$  et  $\mathbb{P}(\{3, 4, 5, 6\}) = 2/3$ . Puis-je calculer la probabilité que le dé marque le chiffre trois ?
4. Soient deux événements  $E$  et  $F$ . Démontrer que

$$\mathbb{P}(E \cap F) \geq \mathbb{P}(E) + \mathbb{P}(F) - 1.$$

5. Soient deux événements  $E$  et  $F$ . Démontrer que la probabilité qu'un seul (exactement) des deux événements se produise vaut

$$\mathbb{P}(E) + \mathbb{P}(F) - 2\mathbb{P}(E \cap F).$$

## 1.4 Equiprobabilité et analyse combinatoire

Dans le cas d'un espace probabilisé où  $\Omega$  est **fini**, on peut toujours numéroter les éléments de  $\Omega$  et écrire  $\Omega = \{1, \dots, n\}$ , où  $n$  est le cardinal de  $\Omega$ , son nombre d'éléments, encore noté  $|\Omega|$  ou  $\#\Omega$ . On peut considérer  $\mathcal{E} = \mathcal{P}(\Omega)$  et calculer la probabilité de tous les sous-ensembles de  $\Omega$  à partir des probabilités des singletons<sup>d</sup>. En effet, par la proposition 1.3.1, on a

$$\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\{\omega\}) = \sum_{i \in E} \mathbb{P}(\{i\}). \quad (1.1)$$

En d'autres termes, la probabilité d'un événement  $E$  est la somme des probabilités des événements élémentaires qu'il contient.

Un cas particulier important est celui où tous les événements élémentaires sont **équiprobables**, i.e. ont tous la même probabilité. On a alors, comme dans l'exemple du dé équilibré :

$$\sum_{i \in \Omega} \mathbb{P}(\{i\}) = \mathbb{P}(\Omega) = 1 \quad \text{et} \quad \mathbb{P}(\{i\}) = \mathbb{P}(\{j\}), \quad \forall i, j \in \Omega.$$

Puisque  $\Omega$  contient  $n$  éléments, cela implique  $\mathbb{P}(\{i\}) = \frac{1}{n}$  pour tout  $i \in \Omega$ . On peut alors simplifier la relation (1.1) et obtenir

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} \quad \forall E \subset \Omega.$$

On a donc retrouvé l'expression bien connue en probabilités :

*“la probabilité d'un événement est le nombre de cas favorables divisé par le (sur le) nombre de cas possibles”.*

d. Un singleton est un ensemble qui contient exactement un seul élément.

Il faut cependant garder en mémoire que cette simplification n'est valable que lorsque  $\Omega$  contient un nombre fini d'éléments et qu'ils ont tous la même probabilité.

**Exemple 1.4.1.** Une urne contient 10 boules numérotées de 1 à 10. J'en tire deux au hasard consécutivement et sans les remettre. Si on suppose que tous les tirages sont équiprobables, quelle est la probabilité que la première boule tirée marque 7 et la deuxième 9 ?

On est dans un cas fini. L'univers des possibles est

$$\Omega = \{(i, j) : 1 \leq i, j \leq 10, i \neq j\}$$

On a donc  $|\Omega| = 10 \cdot 9 = 90$ , et par suite

$$\mathbb{P}(\{(7, 9)\}) = \frac{1}{90}.$$

Voici deux exemples dans la même veine.

**Exemple 1.4.2.** On lance deux dés à 6 faces équilibrés. Quelle est la probabilité que la somme des points obtenus soit 7 ? Faire de même avec toutes les sommes possibles.

**Exemple 1.4.3.** On lance un dé à 6 faces équilibré, et on tire une carte dans un paquet de 52 cartes. Quel est le nombre de résultats possibles de cette expérience ? Si on suppose que les résultats de cette expérience sont équiprobables, quelle est la probabilité que le dé marque un chiffre pair et que la carte soit un coeur ?

Dans la suite de cette section, nous allons tenter de systématiser ce genre de calculs, en définissant les factorielles, les nombres de permutations, d'arrangements et de combinaisons. Commençons par énoncer le principe<sup>e</sup> que nous avons certainement suivi pour résoudre l'exemple 1.4.3.

**Proposition 1.4.1.** *Si on réalise deux expériences (ou processus), de façon telle que la première expérience peut produire  $m$  résultats, et que pour chacun de ceux-ci, l'expérience 2 produit  $n$  résultats, alors les deux expériences prises ensemble peuvent produire  $m \cdot n$  résultats.*

Il n'est pas nécessaire que les résultats possibles de la deuxième expérience soient identiques quel que soit le résultat de la première expérience.

**Exemple 1.4.4.** Dans une assemblée de 10 personnes, on doit choisir un président et un vice-président (sans cumul des mandats). Combien y a-t-il de choix possibles ?

On commence par choisir le président. Il y a dix choix possibles. Pour chaque choix de président, on ne peut plus choisir que les autres personnes dans l'assemblée. Les choix possibles de vice-président dépendent donc du choix du président, mais le nombre de choix est indépendant : pour chaque choix de président, il y a 9 vice-présidents possibles. La réponse est donc  $10 \cdot 9 = 90$ .

Ce principe élémentaire peut évidemment être généralisé à un nombre quelconque d'expériences successives.

**Exemple 1.4.5.** Dans une assemblée de 10 personnes, on doit choisir un président, un vice-président et un secrétaire (sans cumul). Combien y a-t-il de choix possibles ?

e. La notion de principe n'est pas définie en mathématique. Il s'agit en fait d'un résultat que l'on peut démontrer précisément.



### 1.4.1 Quelques mots d'analyse combinatoire

On appelle permutation de  $n$  objets une façon de ranger ces objets dans un certain ordre. Une permutation est donc également une façon de numéroter ces objets. Par exemple une permutation des lettres  $a, b, c$  peut être  $b, c, a$ . Dans ce cas, on a numéroté  $b$  par 1,  $c$  par 2 et  $a$  par 3. On cherche à savoir combien il y a de permutations différentes de ces  $n$  objets.

**Exemple 1.4.6.** Démontrer que le nombre de permutations de trois objets distinguables est 6.

Ce fait peut être directement généralisé par le résultat suivant.

**Proposition 1.4.2.** *Le nombre de permutations de  $n$  objets distinguables est  $n! = n(n - 1) \cdots 2 \cdot 1$ .*

*Démonstration.* On considère un premier processus qui consiste à choisir le premier objet de la liste. Il y a  $n$  possibilités. Pour chacune de ces possibilités, on a une deuxième processus qui consiste à ranger les  $n - 1$  objets restants. Si on note  $P_n$  le nombre de permutations de  $n$  objets, on a donc la relation  $P_n = n \cdot P_{n-1}$ . On a aussi également la relation  $P_1 = 1$ . Ces deux conditions permettent de déterminer  $P_n$  *récurivement* et on obtient le résultat annoncé.  $\square$

Une situation fondamentale semblable consiste, ayant  $n$  objets distinguables à disposition, à en choisir  $p$ , sans pouvoir les répéter, et à les ranger dans un certain ordre. Ici encore la question est de compter le nombre de façon différentes de procéder. Ce nombre est appelé le nombre d'arrangements (sans répétition) de  $p$  objets parmi  $n$  et est noté  $A_n^p$ .

**Exemple 1.4.7.** Combien de mots de trois lettres *distinctes* peut-on former avec les lettres  $a, b, c, d, e, f$  ?

La réponse est 120, et il est utile de le voir de deux façons différentes. D'une part, on a un premier processus qui consiste à choisir la première lettre, et il y a six façons de faire. Pour chaque choix, il reste à choisir un mot de deux lettres parmi les 5 lettres restantes. On le fait de la même façon, en choisissant d'abord la première lettre, ce qui donne 5 choix, et il reste à choisir la dernière lettre, avec 4 possibilités. Le résultat est donc 6.5.4.

D'autre part, on peut former des mots de trois lettres en considérant tous les mots de 6 lettres distinctes possibles, et en ne regardant que les 3 premières lettres pour former un mot de trois lettres. Par exemple, le mot de 6 lettres  $abcdef$  conduira alors au mot de 3 lettres  $abe$ . Il y a  $6!$  mots de 6 lettres possibles. Mais il est clair que les mots de 6 lettres  $abcdef$  et  $abcedf$  conduisent au même mot de trois lettres. En fait il y a exactement  $3!$  mots de 6 lettres qui commencent par  $abe$ . Et c'est le cas pour n'importe quel préfixe. Le nombre de mots de 3 lettres est donc obtenu comme  $6!/3! = 6 \cdot 5 \cdot 4$ .

En généralisant les raisonnements précédents, nous obtenons le résultat suivant.

**Proposition 1.4.3.** *Le nombre d'arrangements (sans répétition) de  $p$  objets (distinguables) parmi  $n$  et est*

$$A_n^p = \frac{n!}{(n-p)!} = n \cdot (n-1) \cdots (n-p+1), \quad \text{si } p \leq n.$$

*Ce nombre est nul si  $p > n$ .*

*Démonstration.* Profitons de la preuve pour donner encore une construction différente. Considérons l'expérience qui consiste à ordonner les  $n$  objets donnés. Nous savons qu'il y a  $n!$  façons de le faire. On peut compter le nombre de façons de faire en utilisant deux processus successifs. Dans un premier temps, on choisit  $p$  objets que l'on range dans un ordre donné aux  $p$  premières positions. Il y a  $A_n^p$  façons de réaliser ce processus. Pour chaque choix fixé de ces  $p$  objets, il reste  $n - p$  objets que l'on range aux dernières positions de toutes les façons possibles, et il y a  $(n - p)!$  possibilités. On obtient donc la relation

$$A_n^p (n - p)! = n!,$$

qui conduit au résultat annoncé.  $\square$

A titre indicatif, on peut considérer un problème tout à fait similaire.

**Exemple 1.4.8.** Combien de mots de trois lettres peut-on former avec les lettres  $a, b, c, d, e, f$  ?

On utilise le même raisonnement que dans l'exemple précédent. La différence est que l'on peut répéter les lettres. On commence par choisir la première lettre, puis pour chaque choix, on doit compléter avec un mot de deux lettres formé avec  $a, b, c, d, e, f$ . On a de nouveau 6 choix pour la première lettre, puis 6 choix pour la dernière. La réponse est donc  $6^3 = 6.6.6$ .

Le nombre décrit ci-dessus est appelé le nombre d'arrangements (avec répétition) de  $p$  objets parmi  $n$  objets distinguables et est noté  $B_n^p$ . Nous avons calculé ce nombre et on généralise directement au résultat suivant.

**Proposition 1.4.4.** On a  $B_n^p = n^p$ .

Nous aurons encore besoin d'un comptage au moins, qui est classique, et qui est plutôt celui de la loterie nationale plutôt que celui du tiercé.

**Exemple 1.4.9.** Pour jouer au lotto, on doit choisir 6 nombres dans l'ensemble  $\{1, \dots, 42\}$ . Des boules numérotées de 1 à 42 sont extraites au hasard d'une urne. L'ordre dans lequel elles sont tirées n'a pas d'importance, puisqu'on les remet dans l'ordre croissant à la fin, pour comparer avec les grilles qui permettent de parier, qui ne permettent pas de préciser l'ordre de sortie. Un résultat du tirage du lotto est donc un ensemble de 6 numéros. Chaque choix est appelé combinaison de 6 boules parmi 42, et on note  $C_{42}^6$  le nombre de telles combinaisons. Combien y a-t-il de choix d'ensembles de 6 boules parmi les 42 boules ? La réponse est  $C_{42}^6 = \frac{42!}{36!6!}$ . Procédons comme plus haut. On sait qu'il y a  $A_{42}^6$  façons d'arranger 6 nombres ordonnés parmi 42. Mais les différents arrangements comportant les 6 mêmes nombres vont donner la même combinaison, puisque dans une combinaison, l'ordre n'a pas d'importance. Comme il y a  $6!$  arrangements avec les 6 mêmes nombres, on a  $C_{42}^6 = \frac{A_{42}^6}{6!}$ .

On peut généraliser cet exemple et obtenir le résultat suivant.

**Proposition 1.4.5.** Le nombre de combinaisons de  $p$  objets parmi  $n$  objets (distinguables) est

$$C_n^p = \frac{n!}{p!(n - p)!}.$$

*Démonstration.* Comme plus haut, écrivons la preuve pour éviter la division, mais plutôt pour appliquer le principe de comptage. Considérons l'expérience qui définit l'arrangement  $A_n^p$ . On doit ordonner  $p$  objets parmi  $n$ . Ce processus peut être conduit en deux temps :

d'abord on choisit un ensemble de  $p$  objets parmi  $n$  (il y a  $C_n^p$  façons de le faire), et pour chaque résultat de ce processus, on choisit un ordre de ces objets, parmi tous les ordres possibles (il y en a  $p!$ ). On a donc la relation

$$A_n^p = C_n^p p!,$$

qui conduit au résultat annoncé.  $\square$

Nous pouvons maintenant calculer la probabilité de gagner au lotto. Si on joue une grille, c'est à dire 6 numéros, on gagne si les boules qui sont choisies portent ces numéros. Si tous les tirages sont équiprobables, comme il y en a  $C_{42}^6$ , la probabilité de gagner au lotto en remplissant une grille est  $1/C_{42}^6$ .

Terminons par un comptage qui nous sera utile au moment de définir la distribution de probabilités des variables binomiales. Il s'agit du comptage de permutations sans répétitions d'objets partiellement indistinguables.

**Exemple 1.4.10.** Combien de mots différents peut-on écrire avec les lettres du mot MATHONET, sans les répéter ?

Si cela avait été MATHONER, on connaîtrait la réponse, à savoir  $8!$ . Le problème est que les deux T sont indistinguables. La solution consiste à les rendre distinguables dans un premier temps, en les appelant  $T_1$  et  $T_2$ . Il y a donc  $8!$  permutations des lettres du mot  $MAT_1HONET_2$ . Mais bien sûr, quand on enlève les indices, les permutations  $T_1HONET_2MA$  et  $T_2HONET_1MA$  donnent le même résultat. Il faut donc diviser  $8!$  par le nombre de permutations donnant le même résultat quand on enlève les indices, à savoir 2. La solution est donc  $\frac{8!}{2}$ .

On généralise ces considérations pour obtenir le résultat suivant.

**Proposition 1.4.6.** Si  $n$  objets sont répartis en  $k$  sous-ensembles de taille  $n_1, \dots, n_k$  au sein desquels les objets sont indistinguables, alors le nombre de permutations sans répétitions de ces objets est

$$\frac{n!}{n_1! \cdots n_k!}.$$

L'exemple suivant est particulièrement utile.

**Exemple 1.4.11.** Je joue à pile ou face 10 fois consécutivement et je note les résultats. Quel est le nombre de mots composés des lettres  $P$  et  $F$  contenant exactement 2 fois pile (par exemple PFFFFFFF ou FFFFPFFFFFFF) ?

D'après le théorème précédent, la réponse est  $\frac{10!}{8!2!}$ . On remarque que c'est également  $C_{10}^8$  ou  $C_{10}^2$ . En effet, un mot de dix lettres composé de  $P$  et de  $F$  est parfaitement déterminé quand on connaît l'ensemble des positions occupées par des  $P$ , qu'il faut choisir de toutes les façons possibles, ou l'ensemble des positions occupées par des  $F$ .

## 1.4.2 Exercices

1. Les anciennes plaques minéralogiques belges étaient formées de 3 lettres suivies de trois chiffres (de 0 à 9). Combien de plaques pouvait-on former de la sorte.
2. On lance une pièce 10 fois et on note le résultat de chaque lancer, pour obtenir un mot composé de 10 symboles  $F$  ou  $P$ , la lettre numéro  $i$  correspondant au résultat du lancer numéro  $i$ . Combien de mots peut-on former ?
3. On a des boules numérotées de 1 à 10 dans une urne. On en tire 3, sans les remettre et on note les résultats dans l'ordre. Quel est le nombre de résultats possibles.
4. On regarde une course hippique avec 15 chevaux au départ. Combien y a-t-il de tiercés possibles à l'arrivée ?

## 1.5 Probabilités conditionnelles

On dispose ici encore d'une expérience aléatoire qui correspond à un espace probabilisé  $(\Omega, \mathcal{E}, \mathbb{P})$ . On s'intéresse maintenant à la probabilité des événements en présence d'une information supplémentaire. Cette information change les probabilités qui pouvaient être calculées a priori. Les nouvelles probabilités ainsi calculées sont alors des *probabilités conditionnelles*. Ces probabilités conditionnelles sont utiles en pratique quand on dispose effectivement de l'information complémentaire, mais elles permettent également un calcul aisé des probabilités dans bien des cas où cette information n'est pas donnée, via la par exemple la **formule des probabilités totales**.

Ces probabilités donnent également lieu à un autre théorème important de la théorie des probabilités, à savoir le *théorème de Bayes*, ainsi qu'au concept d'événements indépendants.

Voici un exemple introductif.

**Exemple 1.5.1.** [Cas équiprobable] On lance un dé à 20 faces équilibré (un icosaèdre<sup>f</sup> régulier par exemple). Les faces sont marquées de 1 à 20. L'univers des possibles est  $\Omega = \{1, \dots, 20\}$ , on choisit  $\mathcal{E} = \mathcal{P}(\Omega)$  et  $\mathbb{P}$  est la probabilité correspondant au cas équiprobable, c'est-à-dire  $\mathbb{P}(\{i\}) = p_i = \frac{1}{20}$  pour tout  $i \in \Omega$ . Si on considère l'événement  $E = \{3, 4, 5, 6, 7, 8\}$ , on a donc  $\mathbb{P}(E) = \frac{6}{20}$ . Notons  $F$  l'événement  $\{1, 2, 3, 4, 5\}$  et supposons que **l'on sache que  $F$  s'est réalisé** lors de la réalisation de l'expérience (ou se réalisera). Cela change évidemment les probabilités a priori que nous avons et on obtient les probabilités "sachant que  $F$  s'est réalisé" (on dit aussi "sachant  $F$ "), ou "sous la condition que  $F$  s'est réalisé." Cette probabilité sera notée  $\mathbb{P}(E|F)$ . Le calcul de cette probabilité est simple si on suit le raisonnement que voici :

1. L'événement  $E$  se produit, sachant que  $F$  se produit, exactement si le résultat du lancé est dans  $\{3, 4, 5\}$ , c'est à dire si le résultat est dans  $E \cap F$ .
2. Les événements élémentaires en dehors de  $F$  ont une probabilité conditionnelle nulle.
3. Les événements élémentaires  $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$  étaient équiprobables a priori. Ils le restent en présence de l'information que  $F$  s'est produit. Leur probabilité est donc  $\frac{1}{5}$  et plus  $\frac{1}{20}$ .

On a donc

$$\mathbb{P}(E|F) = \frac{3}{5} = \frac{|E \cap F|}{|F|}.$$

Ces considérations peuvent être généralisées pour un univers fini  $\Omega$  dans le cas équiprobable et pour tout événement  $F$  non vide : lors du calcul des probabilités conditionnelles sachant que  $F$  s'est produit, dans le cas équiprobable, pour le calcul de la probabilité conditionnelle  $\mathbb{P}(E|F)$ ,

- Les cas favorables sont ceux de  $E \cap F$ .
- L'univers des possibles est dit "réduit" à  $F$ , on peut compter que les cas possibles sont désormais ceux de  $F$ .
- On a donc  $\mathbb{P}(E|F) = \frac{|E \cap F|}{|F|}$ , pour tout  $F \neq \emptyset$ .

Afin de généraliser cette expression naturelle des probabilités conditionnelles, il est utile de l'écrire en termes de probabilités. On a dans le cas équiprobable, puisque  $\mathbb{P}(E \cap F) = \frac{|E \cap F|}{|\Omega|}$  et  $\mathbb{P}(F) = \frac{|F|}{|\Omega|}$ , on a aussi

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)},$$

f. L'icosaèdre régulier est un des cinq solides de Platon, dont toutes les faces sont égales.

pour tout  $F$  non vide (c'est à dire tel que  $\mathbb{P}(F) \neq 0$ ).

Montrons maintenant sur un exemple (académique) que les mêmes raisonnements intuitifs s'étendent au cas général non équiprobable et conduisent à poser la même définition.

**Exemple 1.5.2.** (Cas déséquilibré) Reprenons les conditions et les événements de l'exemple 1.5.1, mais supposons que le dé soit construit de manière telle que les résultats pairs aient tous une probabilité deux fois supérieure à celle des résultats impairs. On a donc a priori

$$p_1 = p_3 = \dots = p_{19} = a \quad \text{et} \quad p_2 = p_4 = \dots = p_{20} = 2a$$

Puisque  $\mathbb{P}(\Omega) = 1 = \sum_{i=1}^{20} p_i$ , on obtient  $10a + 20a = 1$ , donc  $a = \frac{1}{30}$ . Avec les notations de l'exemple précédent, on peut donc calculer les probabilités suivantes :

$$\mathbb{P}(E) = \frac{9}{30} = \frac{3}{10}, \quad \mathbb{P}(F) = \frac{7}{30} \quad \text{et} \quad \mathbb{P}(E \cap F) = \frac{4}{30}.$$

On calcule maintenant les probabilités conditionnelles. On tient le raisonnement suivant :

1. L'événement  $E$  se produit, sachant que  $F$  se produit, exactement si le résultat du lancé est dans  $\{3, 4, 5\}$ , c'est à dire si le résultat est dans  $E \cap F$ .
2. Les événements élémentaires en dehors de  $F$  ont une probabilité conditionnelle nulle.
3. Les événements élémentaires  $\{2\}$  et  $\{4\}$  avaient une probabilité a priori double de celle de  $\{1\}$ ,  $\{3\}$  et  $\{5\}$ . Cela reste encore vrai sachant que  $F$  s'est produit.<sup>g</sup> Si on note  $b$  la probabilité (conditionnelle) de chacun des derniers, la probabilité conditionnelle des premiers est  $2b$  et on est amené à trouver  $b = \frac{1}{7}$ .
4. On calcule donc  $\mathbb{P}(E|F) = \frac{4}{7}$ .

On constate donc que dans ce cas aussi on a obtenu

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}.$$

Ces deux exemples montrent que l'intuition pousse à poser la définition suivante.

**Définition 1.5.1.** Soit  $(\Omega, \mathcal{E}, \mathbb{P})$  un espace probabilisé et soit  $F$  un événement tel que  $\mathbb{P}(F) \neq 0$ . La probabilité de  $E$  conditionnelle à  $F$  est définie par

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}.$$

Cette définition est conforme à l'intuition, comme l'ont montré les exemples (je l'espère). Mais est-elle licite? A-t-on défini une nouvelle mesure de probabilité conditionnelle à  $F$ ? C'est l'objet du résultat suivant.

**Proposition 1.5.1.** Si  $(\Omega, \mathcal{E}, \mathbb{P})$  un espace probabilisé et soit  $F$  un événement tel que  $\mathbb{P}(F) \neq 0$ . Alors l'application

$$Q : \mathcal{E} \rightarrow \mathbb{R} : E \mapsto \mathbb{P}(E|F)$$

définit une mesure de probabilité sur  $(\Omega, \mathcal{E})$ .

---

<sup>g</sup>. Les rapports entre les probabilités des événements élémentaires de  $F$  restent inchangés quand on passe aux probabilités conditionnelles.

*Démonstration.* Il suffit de démontrer que l'application  $Q$  satisfait les trois axiomes de Kolmogorov.

1. On a pour tout  $E \in \mathcal{E}$ ,  $Q(E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} \geq 0$ .
2. Pour tout  $E \in \mathcal{E}$ , on a  $\mathbb{P}(E \cap F) \leq \mathbb{P}(F)$  car  $E \cap F \subset F$ , donc on a  $Q(E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} \leq 1$ .
3. On a  $Q(\Omega) = \frac{\mathbb{P}(\Omega \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F)}{\mathbb{P}(F)} = 1$ .
4. Pour toute suite  $(E_1, E_2, \dots)$  d'événements deux à deux exclusifs, on a

$$Q(\cup_{i=1}^{+\infty} E_i) = \frac{\mathbb{P}((\cup_{i=1}^{+\infty} E_i) \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(\cup_{i=1}^{+\infty} (E_i \cap F))}{\mathbb{P}(F)}.$$

Puisque les événements  $E_i \cap F$  sont également deux à deux exclusifs, on a encore

$$\frac{\mathbb{P}(\cup_{i=1}^{+\infty} (E_i \cap F))}{\mathbb{P}(F)} = \frac{\sum_{i=1}^{+\infty} \mathbb{P}(E_i \cap F)}{\mathbb{P}(F)} = \sum_{i=1}^{+\infty} \mathbb{P}(E_i | F).$$

On a donc enfin

$$Q(\cup_{i=1}^{+\infty} E_i) = \sum_{i=1}^{+\infty} Q(E_i),$$

ce qu'il fallait démontrer. □

Cette proposition, qui peut sembler un exercice de logique pure, nous permet d'utiliser les probabilités conditionnelles avec les mêmes propriétés que nous avons démontrées pour les probabilités en général. On a donc par exemple pour tous événements  $E_1$  et  $E_2$ ,

$$\mathbb{P}(E_1 \cup E_2 | F) = \mathbb{P}(E_1 | F) + \mathbb{P}(E_2 | F) - \mathbb{P}(E_1 \cap E_2 | F).$$

Il est important de noter également que par définition, si  $F$  est un événement tel que  $\mathbb{P}(F) \neq 0$ , on a

$$\boxed{\mathbb{P}(E \cap F) = \mathbb{P}(E | F) \mathbb{P}(F)}. \tag{1.2}$$

La probabilité que  $E$  et  $F$  se produisent conjointement est donc égale à la probabilité que  $F$  se produise, multipliée par la probabilité que  $E$  se produise sachant que  $F$  s'est produit, conformément à l'intuition.

*Remarque 1.2.* Notons également que si  $\mathbb{P}(F) = 0$ , la probabilité conditionnelle  $\mathbb{P}(E | F)$  n'est plus définie, mais on a aussi  $\mathbb{P}(E \cap F) = 0$ , donc la formule (1.2) est pourrait être vraie dans ce cas, quel que soit la valeur donnée à  $\mathbb{P}(E | F)$ . On étendra donc la validité de cette formule, si le besoin s'en fait sentir.

Enfin, le membre de gauche de (1.2) étant symétrique en  $E$  et  $F$ , on peut obtenir une autre expression, si  $\mathbb{P}(E) \neq 0$ , à savoir

$$\mathbb{P}(E \cap F) = \mathbb{P}(F | E) \mathbb{P}(E). \tag{1.3}$$

Il est parfois plus facile dans les applications de calculer les probabilités conditionnelles. Les formules (1.2) et (1.3) permettent alors de calculer  $\mathbb{P}(E \cap F)$  comme le montre l'exemple suivant. Elles sont aussi à l'origine de la formule des probabilités totales et du théorème de Bayes, comme nous le verrons plus loin.

**Exemple 1.5.3.** Dans une urne, il y a 8 boules blanches et 4 noires. On effectue deux tirages consécutifs et sans remise et on suppose que toutes les boules sont identiques en forme (les tirages sont équiprobables parmi les boules présentes dans l'urne). Quelle est la probabilité que les deux boules tirées soient blanches ?

On note  $E_1$  l'événement "la première boule est blanche" et  $E_2$  l'événement "la deuxième boule est blanche". On doit alors calculer  $\mathbb{P}(E_1 \cap E_2)$ . On a deux possibilités de *conditionnement*

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1|E_2)\mathbb{P}(E_2) \quad \text{et} \quad \mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1)\mathbb{P}(E_1).$$

Le premier n'est pas utile pour l'instant, parce qu'il n'est pas facile de calculer  $\mathbb{P}(E_2)$ . Mais il est facile de calculer  $\mathbb{P}(E_1) = \frac{8}{12}$  et  $\mathbb{P}(E_2|E_1) = \frac{7}{11}$ , puisque si  $E_1$  s'est produit, au second tirage il n'y a plus que 11 boules dans l'urne, dont 7 blanches. On a donc

$$\mathbb{P}(E_1 \cap E_2) = \frac{14}{33}.$$

Il est à noter dans cet exemple que le choix du conditionnement n'est pas anodin. Celui qui a l'air inutile pour l'instant nous permet cependant d'obtenir un des deux nombres  $\mathbb{P}(E_1|E_2)$  et  $\mathbb{P}(E_2)$  une fois que nous avons l'autre, puisque nous connaissons maintenant  $\mathbb{P}(E_1 \cap E_2)$ .

### 1.5.1 Exercices

1. Pour l'exemple 1.5.3, calculer les probabilités que la première boule soit noire et la deuxième blanche, que la première soit blanche et la deuxième noire, que les deux boules soient noires, que la deuxième boule soit blanche, que la deuxième boule soit noire.
2. (Ross, Page 93) On lance deux dés équilibrés à 6 faces et discernables. On suppose que les résultats sont équiprobables. Déterminer la probabilité que l'un des deux dés au moins montre 6, sachant que les résultats des deux dés sont différents.
3. (Ross, Page 93) On lance deux dés équilibrés à 6 faces et discernables. On suppose que les résultats sont équiprobables. Déterminer la probabilité que l'un des deux dés au moins montre 6, sachant que la somme des résultats des dés est 8. Faire de même si la somme est 3. Faire de même si la somme est 7.
4. (Ross, Page 93) Une urne contient 6 boules blanches et 9 noires. On tire 4 boules sans remise et on suppose que toutes les boules sont identiques. Quelle est la probabilité que les 2 premières boules soient blanches et les deux suivantes noires.

## 1.6 Événements indépendants

Soit  $(\Omega, \mathcal{E}, \mathbb{P})$  un espace probablisé et  $E$  et  $F$  deux événements. Si  $\mathbb{P}(F) \neq 0$ , on peut définir la probabilité de  $E$  conditionnelle à  $F$ . On aura naturellement tendance à dire que  $E$  est indépendant de  $F$  si le fait que  $F$  se réalise ne modifie pas la probabilité de  $E$  de se produire, c'est-à-dire si

$$\mathbb{P}(E|F) = \mathbb{P}(E).$$

Puisque  $\mathbb{P}(F) \neq 0$ , cette condition est équivalente à

$$\mathbb{P}(E \cap F) = \mathbb{P}(E) \mathbb{P}(F).$$

Cette dernière condition a du sens même si  $\mathbb{P}(F) = 0$ , et c'est donc elle que nous adoptons comme définition. Elle est également symétrique : si  $E$  est indépendant de  $F$ , alors  $F$  est indépendant de  $E$  à ce sens.

**Définition 1.6.1.** Soit  $(\Omega, \mathcal{E}, \mathbb{P})$  un espace probabilisé. Deux événements  $E$  et  $F$  sont indépendants si

$$\mathbb{P}(E \cap F) = \mathbb{P}(E) \mathbb{P}(F).$$

A titre d'illustration voici une propriété, naturelle encore une fois.

**Proposition 1.6.1.** Soit  $(\Omega, \mathcal{E}, \mathbb{P})$  un espace probabilisé et  $E$  et  $F$  deux événements. Si  $E$  et  $F$  sont indépendants, alors  $E$  et  $\overline{F}$  sont indépendants.

*Démonstration.* Il suffit de calculer

$$\mathbb{P}(E \cap \overline{F}) = \mathbb{P}(E) - \mathbb{P}(E \cap F),$$

puisque  $E$  est l'union de  $E \cap F$  et  $E \cap \overline{F}$ . Par indépendance de  $E$  et  $F$ , on a ensuite

$$\mathbb{P}(E) - \mathbb{P}(E \cap F) = \mathbb{P}(E) - \mathbb{P}(E)\mathbb{P}(F) = \mathbb{P}(E)(1 - \mathbb{P}(F)) = \mathbb{P}(E)\mathbb{P}(\overline{F}),$$

et le résultat est démontré.  $\square$

On peut généraliser la notion d'indépendance à plusieurs événements  $E_1, \dots, E_n$ . On pourrait penser qu'il suffit de demander qu'ils soient deux à deux indépendants. Bien que cette définition puisse être posée, elle ne s'avère pas satisfaisante pour les applications. On adopte plutôt la définition suivante, parfois appelée indépendance totale.

**Définition 1.6.2.** Soit  $(\Omega, \mathcal{E}, \mathbb{P})$  un espace probabilisé. Des événements  $E_1, \dots, E_n$  sont indépendants si quel que soit le sous ensemble  $\{a_1, \dots, a_r\}$  de  $\{1, \dots, n\}$  on a

$$\mathbb{P}(E_{a_1} \cap \dots \cap E_{a_r}) = \mathbb{P}(E_{a_1}) \dots \mathbb{P}(E_{a_r}).$$

## 1.7 La formule des probabilités totales et le théorème de Bayes

Revenons à l'exemple 1.5.3. On a une urne avec 8 boules blanches et 4 noires. On a appelé  $E_2$  l'événement "la deuxième boule tirée est blanche". Pour calculer la probabilité de  $E_2$ , puisqu'on n'a pas remis la boule extraite au premier tirage, il faut tenir compte de tous les résultats possibles et utiliser des probabilités conditionnelles. Si  $E_1$  est l'événement "la première boule tirée est blanche", on a visiblement une décomposition en événements exclusifs

$$E_2 = (E_2 \cap E_1) \cup (E_2 \cap \overline{E_1}),$$

et donc

$$\mathbb{P}(E_2) = \mathbb{P}(E_2 \cap E_1) + \mathbb{P}(E_2 \cap \overline{E_1}).$$

Les deux probabilités du membre de droite sont faciles à calculer : on a

$$\mathbb{P}(E_2 \cap E_1) = \mathbb{P}(E_2|E_1)\mathbb{P}(E_1) \quad \text{et} \quad \mathbb{P}(E_2 \cap \overline{E_1}) = \mathbb{P}(E_2|\overline{E_1})\mathbb{P}(\overline{E_1}),$$

ces formules restant acceptables si  $\mathbb{P}(E_1) = 0$  ou  $\mathbb{P}(\overline{E_1}) = 0$ . On a donc en final

$$\mathbb{P}(E_2) = \mathbb{P}(E_2|E_1)\mathbb{P}(E_1) + \mathbb{P}(E_2|\overline{E_1})\mathbb{P}(\overline{E_1}).$$



Cette formule est la *formule des probabilités totales*, elle conduit au résultat  $\mathbb{P}(E_2) = \frac{2}{3}$ .

Comme annoncé plus haut, connaissant  $\mathbb{P}(E_2)$ , on peut calculer  $\mathbb{P}(E_1|E_2)$ . Cela semble ne servir à rien pour le problème en question (mais qui se soucie des boules et des urnes ?), mais c'est en fait un calcul de probabilité en présence d'une information observée, et c'est un cas qui se pose souvent dans la pratique des probabilités, par exemple lorsqu'un test de dépistage est positif, on cherche à évaluer la probabilité de la présence de la maladie chez le patient.

Pour le calcul proprement dit, on a par définition

$$\mathbb{P}(E_1|E_2) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_2)}.$$

Le numérateur est connu, c'est  $\mathbb{P}(E_2|E_1)\mathbb{P}(E_1)$ . Nous venons de calculer le dénominateur. On obtient donc

$$\mathbb{P}(E_1|E_2) = \frac{\mathbb{P}(E_2|E_1)\mathbb{P}(E_1)}{\mathbb{P}(E_2|E_1)\mathbb{P}(E_1) + \mathbb{P}(E_2|\overline{E_1})\mathbb{P}(\overline{E_1})}.$$

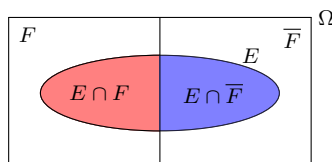
Cette formule est ce que donne le théorème de Bayes dans notre exemple. On obtient  $\mathbb{P}(E_1|E_2) = \frac{7}{11}$ .

Enonçons la formule simple des probabilités totales, dans le cas de deux événements quelconques.

**Proposition 1.7.1.** *Soit  $(\Omega, \mathcal{E}, \mathbb{P})$  un espace probabilisé et  $E$  un événement. Pour tout événement  $F$ , on a*

$$\mathbb{P}(E) = \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|\overline{F})\mathbb{P}(\overline{F}). \tag{1.4}$$

La preuve suit le raisonnement que nous avons développé plus haut. Elle est donc omise. Il est cependant utile de représenter les diagrammes de Venn définis par  $\Omega$ ,  $E$  et  $F$  pour visualiser la situation.



Voici maintenant l'énoncé du théorème de Bayes<sup>h</sup> dans sa forme la plus simple.

**Proposition 1.7.2.** *Soient  $(\Omega, \mathcal{E}, \mathbb{P})$  un espace probabilisé et  $E$  et  $F$  deux événements tels que  $\mathbb{P}(E) \cdot \mathbb{P}(F) \neq 0$ . On a alors*

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(F|E)\mathbb{P}(E)}{\mathbb{P}(F)} = \frac{\mathbb{P}(F|E)\mathbb{P}(E)}{\mathbb{P}(F|E)\mathbb{P}(E) + \mathbb{P}(F|\overline{E})\mathbb{P}(\overline{E})}.$$

Considérons maintenant un exemple plus utile dans la pratique scientifique.

**Exemple 1.7.1.** (Ross, p. 61) Un laboratoire d'analyse assure avec une fiabilité de 95% la détection d'une maladie lorsqu'elle est effectivement présente. Il y a malheureusement 1% des personnes saines qui sont déclarées positives par le test. Enfin, on sait que la prévalence de la maladie dans la population (la proportion de malades) est de 0.5%. Quelle est la probabilité qu'une personne positive au test soit effectivement malade ?

h. Thomas Bayes (1701-1761).

## 1.8 Variables aléatoires

Dans la pratique, on s'intéresse souvent à une fonction du résultat de l'expérience aléatoire, plutôt qu'au résultat lui-même. L'exemple suivant est assez classique.

**Exemple 1.8.1.** Considérons l'exemple du jeu de roulette. La roulette est constituée de 18 cases rouges, 18 cases noires et une case verte. Un joueur décide de placer une mise sur le noir. L'expérience aléatoire consiste à lancer la bille dans la roulette. L'univers des possibles  $\Omega$  est l'ensemble des 37 cases. On suppose que les résultats sont équiprobables : on a  $\mathbb{P}(\{\omega\}) = 1/37$  pour tout  $\omega \in \Omega$ .

Si le résultat est rouge ou vert, le joueur perd sa mise, tandis qu'il gagne une fois sa mise si c'est noir.

Le joueur a ainsi défini une fonction

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto \begin{cases} 1 & \text{si } \omega \text{ est noir} \\ -1 & \text{si } \omega \text{ est vert ou rouge} \end{cases}$$

On s'intéresse alors à la probabilité que le joueur gagne :

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = 1\}) = \mathbb{P}(\{\omega : \omega \text{ est noir}\}) = 18/37.$$

### Quelques remarques et notations :

- L'application  $X$  est appelée variable aléatoire.
- L'ensemble  $\{\omega \in \Omega : X(\omega) = 1\}$  est noté  $(X = 1)$  ou  $X^{-1}(\{1\})$ . Pour que le calcul précédent soit correct, il faut que ce soit un événement.
- L'ensemble des valeurs prises par  $X$  est noté  $\Omega_X$  ou  $X(\Omega)$ . Dans notre exemple il vaut  $\{-1, 1\}$ .

**Exemple 1.8.2.** On considère l'expérience aléatoire consistant à lancer trois fois une pièce équilibrée (de manière indépendante). On note le résultat  $P$  ou  $F$ . On a alors

$$\Omega = \{(P, P, P), (P, P, F), (P, F, P), (F, P, P), (P, F, F), (F, P, F), (F, F, P), (F, F, F)\}.$$

On définit la variable aléatoire "Y : nombre de  $F$  obtenus". On a alors

$$\Omega_Y = Y(\Omega) = \{0, 1, 2, 3\}.$$

On peut calculer la probabilité d'obtenir deux fois face : on constate que tous les résultats sont équiprobables. On a alors

$$\mathbb{P}(Y = 2) = \mathbb{P}(\{\omega \in \Omega : Y(\omega) = 2\}) = \mathbb{P}(\{(P, F, F), (F, P, F), (F, F, P)\}) = 3/8.$$

On peut ainsi définir une probabilité  $\mathbb{P}_Y$  (induite par  $Y$ ) sur l'espace probabilisable  $(\Omega_Y, \mathcal{P}(\Omega_Y))$  par

$$\begin{cases} \mathbb{P}_Y(\{0\}) & = & \mathbb{P}(Y = 0) & = & 1/8 \\ \mathbb{P}_Y(\{1\}) & = & \mathbb{P}(Y = 1) & = & 3/8 \\ \mathbb{P}_Y(\{2\}) & = & \mathbb{P}(Y = 2) & = & 3/8 \\ \mathbb{P}_Y(\{3\}) & = & \mathbb{P}(Y = 3) & = & 1/8. \end{cases}$$

On constate qu'on a bien

$$\mathbb{P}_Y(0) + \mathbb{P}_Y(1) + \mathbb{P}_Y(2) + \mathbb{P}_Y(3) = 1.$$

### 1.8.1 Définitions et notations

Les exemples précédents amènent à la définition suivante.

**Définition 1.8.1.** Une *variable aléatoire* sur un espace probabilisé  $(\Omega, \mathcal{E}, \mathbb{P})$  est une application  $X : \Omega \rightarrow \mathbb{R}$  telle que pour tout intervalle (ouvert)  $I$  de  $\mathbb{R}^1$  l'ensemble

$$(X \in I) = X^{-1}(I) = \{\omega \in \Omega : X(\omega) \in I\}$$

est un événement.

La condition évoquée, qui semble technique, permet de définir une probabilité sur  $\mathbb{R}$ , ou sur  $\Omega_X$ , comme nous l'avons vu dans le deuxième exemple.

Si  $X$  est une variable aléatoire sur  $(\Omega, \mathcal{E}, \mathbb{P})$ , alors l'ensemble des valeurs prises par  $X$  est noté  $\Omega_X$  ou  $X(\Omega)$ . Si il est fini ou dénombrable, on dit que  $X$  est une *variable discrète*, sinon  $X$  est une *variable aléatoire continue*.

Une variable aléatoire  $X$  sur  $(\Omega, \mathcal{E}, \mathbb{P})$  permet de définir une probabilité  $\mathbb{P}_X$  sur  $(\mathbb{R}, \mathcal{B})$  par

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}).$$

La même formule définit une probabilité sur  $\Omega_X$ .

On note souvent  $(X \in B)$  l'événement  $X^{-1}(B)$ . Si  $B = [a, b]$ , on peut aussi le noter  $(a \leq X \leq b)$ . De même, l'événement  $X^{-1}(\{a\})$  est noté  $(X = a)$ .

*Remarque 1.3.* Quand  $\Omega$  est fini ou dénombrable, on choisit  $\mathcal{E} = \mathcal{P}(\Omega)$ , et toute application  $X : \Omega \rightarrow \mathbb{R}$  est alors une variable aléatoire.

## 1.9 Distribution de probabilité d'une variable discrète, espérance et variance

Si  $X$  est une variable discrète, par définition, on a  $\Omega_X = \{x_1, \dots, x_n\}$  (cas fini) ou  $\Omega_X = \{x_1, x_2, \dots\}$  (cas dénombrable). On définit alors les nombres  $p_1, \dots, p_n$  par

$$p_i = \mathbb{P}(X = x_i) = \mathbb{P}_X(x_i) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_i\}), \quad \forall i \leq n.$$

Le nombre  $p_i$  est donc la probabilité que  $X$  prenne la valeur  $x_i$ .

**Définition 1.9.1.** La distribution (ou loi) de probabilité de la variable  $X$  est alors la fonction

$$f : \Omega_X \rightarrow \mathbb{R} : x_i \mapsto p_i.$$

On peut la donner sous forme d'un tableau :

$x_i$	$x_1$	$x_2$	$x_3$	$\dots$
$p_i$	$p_1$	$p_2$	$p_3$	$\dots$

ou par un *diagramme en bâtons* où on portera en abscisses les  $x_i$  et en ordonnées les probabilités  $p_i$ .

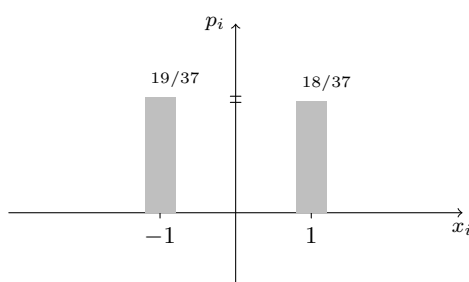
Pour l'exemple 1.8.1, on a le tableau suivant

$x_i$	-1	1
$p_i$	19/37	18/37

Le diagramme en bâtons est donc

---

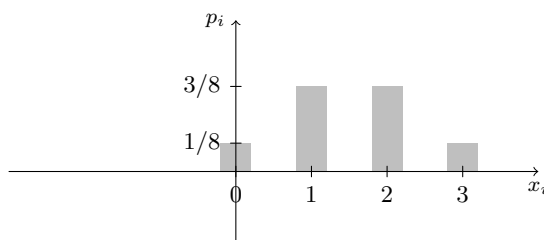
i. Un intervalle ouvert de  $\mathbb{R}$  est un ensemble de la forme  $]a, b[ = \{x \in \mathbb{R} : a < x < b\}$ . Il est souvent noté  $(a, b)$  dans la littérature anglo-saxonne.



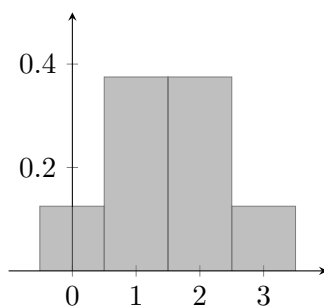
Pour l'exemple 1.8.2, on a le tableau suivant

$x_i$	0	1	2	3
$p_i$	1/8	3/8	3/8	1/8

Le diagramme en bâtons est donc



La forme des bâtons n'est pas importante pour représenter une fonction. Il est cependant utile pour la suite de leur donner une base unitaire. On obtient alors le diagramme suivant.



### 1.9.1 Espérance mathématique

Si on considère à nouveau l'exemple 1.8.1, on peut se demander ce que le joueur peut espérer gagner en jouant à la roulette. Il est naturel de définir l'espérance mathématique de gain par

$$E(X) = 19/37 \cdot (-1) + 18/37 \cdot 1 = -1/37.$$

En effet, le joueur risque de remporter une fois sa mise, avec une probabilité  $18/37$  et il risque de perdre sa mise avec une probabilité  $19/37$ . L'espérance est alors le "centre de masse des gains, pondérés par leur probabilité d'apparition."

Si dans l'exemple 1.8.2, on reçoit un euro chaque fois que la pièce montre pile, quelle somme peut-on espérer recevoir ?

La réponse est

$$E(Y) = 1/8 \cdot 0 + 3/8 \cdot 1 + 3/8 \cdot 2 + 1/8 \cdot 3 = 12/8 = 3/2.$$

En général, on a la définition suivante.

**Définition 1.9.2.** L'espérance mathématique de la variable aléatoire (discrète)  $X$  est donnée par

$$E(X) = \sum_i p_i x_i.$$

Il peut arriver que  $E(X)$  n'existe pas (dans le cas infini dénombrable). L'espérance est également notée  $\mu_X$  ou  $\mu$  s'il n'y a pas d'ambiguïté possible.

*Remarque 1.4.* — Il est clair que dans les deux exemples considérés, le joueur ne reçoit jamais une somme correspondant à l'espérance calculée. En effet, dans le premier jeu, il reçoit  $+1$  ou  $-1$  fois sa mise tandis que dans le deuxième jeu, il reçoit  $0, 1, 2$  ou  $3$  euros. Cependant, on peut concevoir qu'en répétant le jeu un nombre suffisant de fois, **la moyenne arithmétique** des gains va tendre vers l'espérance de la variable. Cette constatation, que tous les joueurs avertis devraient connaître, fait l'objet d'un théorème appelé *la loi des grands nombres*.

- Dans le cas où tous les  $p_i$  sont égaux (les  $x_i$  sont équiprobables), l'espérance mathématique de  $X$  n'est rien d'autre que la moyenne arithmétique des  $x_i$ .
- L'espérance est un *paramètre de tendance centrale*, puisqu'elle indique le centre de masse de la distribution. Il existe d'autres paramètres de tendance centrale, le plus célèbre étant sans doute la médiane.

### 1.9.2 Variance

Il est également utile de mesurer la dispersion des gains possibles autour de l'espérance. Donnons à titre d'exemple la situation élémentaire suivante.

On considère le jeu de pile ou face, sur lequel on parie. Le fait de fixer les paris revient à définir une variable aléatoire sur l'univers  $\Omega = \{F, P\}$ . On considère trois cas :

$$X : \Omega \rightarrow \mathbb{R} : \begin{cases} F \mapsto 0 \\ P \mapsto 0 \end{cases}$$

$$Y : \Omega \rightarrow \mathbb{R} : \begin{cases} F \mapsto 1 \\ P \mapsto -1 \end{cases}$$

$$Z : \Omega \rightarrow \mathbb{R} : \begin{cases} F \mapsto 100 \\ P \mapsto -100 \end{cases}$$

On vérifie que dans les trois cas, l'espérance mathématique vaut zéro. Cependant le troisième jeu semble beaucoup plus risqué. Cela s'explique par les écarts des  $x_i$  à l'espérance, qui sont beaucoup plus grands dans le troisième cas. La différence entre ces trois situations est la *dispersion* de la distribution autour de l'espérance. Il pourrait paraître raisonnable d'additionner les écarts, mais alors les écarts positifs et négatifs se compenseraient. On prend donc naturellement les carrés des écarts de  $x_i$  à  $E(X)$ , et comme pour l'espérance, on pondère par la probabilité de  $x_i$ . On arrive donc à la définition suivante.

**Définition 1.9.3.** La *variance* de la variable aléatoire (discrète)  $X$  est donnée par

$$\sigma^2(X) = \sum_i p_i (x_i - E(X))^2.$$

L'*écart type* de  $X$  est alors

$$\sigma(X) = \sqrt{\sigma^2(X)}.$$

La variance de  $X$  est également notée  $V[X]$ ,  $\sigma_X^2$  ou simplement  $\sigma^2$ .

*Remarque 1.5.* La variance n'est définie que si  $E(X)$  existe et même dans ce cas, il peut arriver qu'elle n'existe pas (dans le cas infini dénombrable bien sûr).

Dans les trois exemples précédents, on vérifie que la variance est donnée par

$$\sigma^2(X) = 0 \quad \sigma^2(Y) = 1, \quad \sigma^2(Z) = 10000.$$

## 1.10 Un exemple fondamental en statistique

L'exemple que nous allons développer dans cette section sera fondamental, car il explique deux façons de lier les probabilités en général et les variables aléatoires en particulier aux statistiques (descriptives) pour permettre le développement de la statistique inférentielle (ou inductive).

On considère une classe  $\Omega = \{\omega_1, \dots, \omega_{20}\}$  de 20 étudiants et on collecte leur notes obtenues à une interrogation. On obtient ainsi la *série statistique* (de données) suivante.

Étudiant	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$	$\omega_6$	$\omega_7$	$\omega_8$	$\omega_9$	$\omega_{10}$
Note	8	8	10	10	12	12	12	10	10	10
Étudiant	$\omega_{11}$	$\omega_{12}$	$\omega_{13}$	$\omega_{14}$	$\omega_{15}$	$\omega_{16}$	$\omega_{17}$	$\omega_{18}$	$\omega_{19}$	$\omega_{20}$
Note	8	16	16	10	10	16	16	10	12	12

La statistique descriptive nous enseigne comment étudier ce genre de série statistique en groupant les données dans un tableau et en calculant la moyenne  $m_x$  des notes, et la variance  $V_x$  de la série statistique, où  $x_i$  représente la note obtenue par l'étudiant  $\omega_i$ . On recense les données dans un tableau :

$x_i$	$n_i$	$f_i$	$f_i x_i$	$x_i - m_x$	$(x_i - m_x)^2$	$f_i (x_i - m_x)^2$
8	3	3/20	24/20	-3.4	11.56	1.734
10	8	8/20	80/20	-1.4	1.96	0.784
12	5	5/20	60/20	0.6	0.36	0.09
16	4	4/20	64/20	4.6	21.16	4.232

Les trois premières colonnes (où les  $n_i$  sont les effectifs observés et les  $f_i$  les fréquences observées) permettent de calculer la moyenne arithmétique des notes :

$$m_x = \frac{1}{20} \sum_i n_i x_i = \sum_i f_i x_i = \frac{228}{20} = 11.4.$$

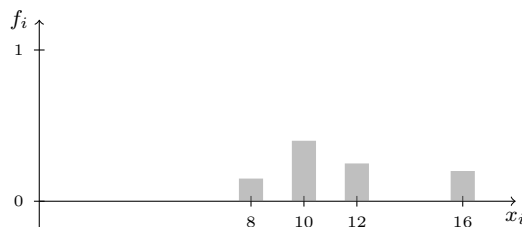
Les colonnes suivantes donnent la variance de la série statistique :

$$V_x = \frac{1}{20} \sum_i n_i (x_i - m_x)^2 = \sum_i f_i (x_i - m_x)^2 = 6.84$$

et l'écart-type de la série statistique,

$$\sqrt{V_x} = 2.61.$$

On peut également représenter la série statistique par un diagramme en bâtons :



On considère maintenant l'*expérience aléatoire* qui consiste à choisir au hasard et de manière équiprobable un étudiant de la classe. On a comme univers  $\Omega$  l'ensemble des étudiants. Puisque  $\Omega$  est fini, on choisit  $\mathcal{E} = \mathcal{P}(\Omega)$ . Bien sûr, si le choix est équiprobable, on a la probabilité uniforme, c'est à dire

$$\mathbb{P}(\{\omega_i\}) = 1/20 \quad \forall i \in \{1, \dots, 20\}.$$

On définit sur  $\Omega$  la variable aléatoire

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega) = \text{la note obtenue par } \omega.$$

Il s'agit d'une variable aléatoire. En effet, puisque  $\mathcal{E} = \mathcal{P}(\Omega)$ , la condition technique de la définition est toujours vérifiée.

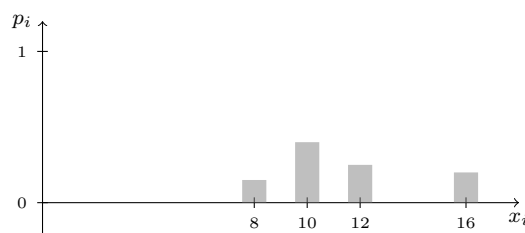
On a alors

$$\Omega_X = \{8, 10, 12, 16\}.$$

La distribution de  $X$  est donnée par le tableau suivant

$x_i$	8	10	12	16
$p_i$	3/20	8/20	5/20	4/20

et se représente par un diagramme en bâtons



On constate donc que la variable aléatoire  $X$  ainsi construite admet pour univers des possibles  $\Omega_X$  l'ensemble des valeurs de  $x$ , et a une distribution de probabilité qui correspond à la distribution des fréquences de la série statistique  $x$ . On peut également calculer l'espérance mathématique de  $X$  par

$$E(X) = \sum_i p_i x_i = 11.4,$$

tandis que la variance est donnée par

$$\sigma^2(X) = \sum_i p_i (x_i - E(X))^2 = 6.84.$$

De la même façon, l'écart-type vaut

$$\sigma(X) = \sqrt{\sigma^2(X)} = 2.61$$

En conclusion, on remarque que si on dispose d'une série statistique  $x$  numérique, en réalisant l'expérience aléatoire consistant à tirer un individu au hasard et de manière uniforme dans la population et en considérant la variable aléatoire "valeur de la statistique sur l'individu choisi", on obtient une variable aléatoire  $X$  distribuée comme  $x$  et telle que

$$E(X) = m_x, \quad \sigma^2(X) = V_x \quad \text{et} \quad \sigma(X) = \sigma_x.$$

Cette première vision permet de donner un lien fort entre statistique descriptive et théorie des probabilités. Cependant, c'est un autre point de vue qui est utilisé en statistique inférentielle. Il renverse les rôles de la vision précédente.

Dans les considérations ci-dessus, on a une population fixe, les 20 étudiants, qui détermine la *distribution statistique* d'une variable statistique  $x$ . L'expérience aléatoire consiste à choisir un individu parmi ceux qui forment la population des 20 étudiants.

Si on connaît les valeurs de  $x$  pour tous les individus de la population, il n'y a pas besoin de statistique inférentielle ou de théorie des tests. On peut mesurer tout ce qu'il y a à savoir sur la variable  $x$  dans la population : on la connaît parfaitement, et la statistique descriptive suffit. Nul besoin donc de théorie des probabilités.

En règle générale, on n'a pas accès à toutes les valeurs de  $x$ , parce que la population est trop grande, ou parce qu'on ne peut pas l'observer (on pourrait parler de la série statistique des tailles des hommes belges entre 1980 et 2020, on ne pourra dès lors pas faire de mesures pour toute cette population).

On est donc forcé de renverser les rôles. On postule qu'il existe une distribution théorique  $\mathcal{D}$ , idéale, de la "note de l'étudiant qui passe le test". On peut réaliser une expérience aléatoire consistant "à choisir au hasard l'étudiant 1 dans la population possible (passée, présente et à venir)", et à lui faire passer le test. La fonction qui à l'étudiant choisi associe sa note est alors une variable aléatoire  $X_1$ , qui a par définition la distribution de probabilité  $\mathcal{D}$ . Quand on réalise l'expérience, l'étudiant choisi est  $\omega_1$  et la donnée  $x_1$  est  $X(\omega_1)$ , la réalisation de  $X_1$ . On peut répéter l'expérience avec le choix d'un deuxième étudiant. On a alors une deuxième variable aléatoire  $X_2$ , qui quand on réalise cette deuxième expérience, donne lieu à la variable  $x_2$ . On fait de même jusque  $X_{20}$ . Bien sûr,  $X_1, \dots, X_{20}$  sont des variables aléatoires ayant toutes la même distribution de probabilité ( $\mathcal{D}$ ). On dit qu'elles sont identiquement distribuées. Leur distribution de probabilité commune  $\mathcal{D}$  est appelée **distribution sous-jacente** ou **distribution théorique** de la variable dans la population.

Dans cette meilleure vision, c'est la distribution de probabilité sous-jacente qui existe en premier, et les séries statistiques viennent en second. Elles servent à répondre à des questions sur la distribution théorique sous-jacente, qui est en général inconnue.

## 1.11 Exemples supplémentaires, exercices

Voici maintenant un exemple où l'expérience aléatoire considérée est telle que  $\Omega$  est infini mais dénombrable.

L'expérience consiste à lancer un dé équilibré jusqu'à obtenir un 4. On peut considérer l'univers des possibles

$$\Omega = \{(4), (a, 4), (a, a, 4), (a, a, a, 4), \dots\} \cup \{(a, a, a, a, \dots)\},$$



où  $a$  signifie qu'on a obtenu un nombre autre que 4 lors du lancer en question. L'univers des possibles est donc dénombrable mais pas fini. On calcule les probabilités des événements élémentaires et on a

$$\mathbb{P}(\{(a, \dots, a, 4)\}) = \frac{1}{6} \left(\frac{5}{6}\right)^n$$

si  $a$  apparaît  $n$  fois. Alors on a

$$\mathbb{P}(\{(a, a, a \dots)\}) = 1 - 1/6 \sum_{n=0}^{+\infty} \left(\frac{5}{6}\right)^n = 1 - \frac{1}{6} \frac{1}{1 - \frac{5}{6}} = 0.$$

On peut alors se limiter au cas où 4 apparaît effectivement et poser

$$\Omega = \{(4), (a, 4), (a, a, 4), (a, a, a, 4), \dots\},$$

ou conserver le résultat possible où 4 n'apparaît pas, avec une probabilité nulle. On choisira  $\mathcal{E} = \mathcal{P}(\Omega)$ .

On définit alors la variable aléatoire “ $X$  : nombre total de lancers effectués” et on a alors  $\Omega_X = \{1, 2, 3, \dots\} = \mathbb{N}_0$ . La loi de probabilité de  $X$  est donnée par

$x_i$	1	2	...	$n$	...
$p_i$	1/6	1/6.5/6	...	1/6.(5/6) <sup><math>n</math></sup>	...

On peut représenter la distribution par un diagramme en bâtons. Enfin, on a

$$E(X) = 6 \quad \text{et} \quad \sigma^2(X) = 30.$$

Terminons cette section par quelques exercices.

1. Un étudiant subit un examen comportant 5 questions à choix multiples avec chaque fois trois réponses possibles. On suppose qu'il n'a pas étudié et répond au hasard. Quelle est la distribution de probabilité de la variable aléatoire “nombre de bonnes réponses”. Calculer son espérance.
2. On lance un dé équilibré et dont les faces sont numérotées 1,1,2,2,3,3, jusqu'à obtenir un total d'au moins trois points. Étudier la variable aléatoire  $X$  : “nombre de lancers effectués”. (Calculer  $\Omega_X$ , la distribution, l'espérance et la variance de  $X$ ). Calculer  $\mathbb{P}(X \leq 2)$ .
3. On considère une urne contenant 4 boules blanches et six boules noires. On tire trois fois une boule, on note sa couleur et on la replace dans l'urne. Étudier la variable aléatoire  $X$  : “nombre de boules blanches obtenues”.
4. On considère une urne contenant 4 boules blanches et six boules noires. On tire successivement trois boules, sans remise. Étudier la variable aléatoire “nombre de boules blanches tirées.”
5. On considère une urne contenant 4 boules blanches et six boules noires. On considère l'expérience aléatoire consistant à tirer une boule et à la replacer dans l'urne jusqu'à obtenir une blanche. Étudier la variable aléatoire “nombre de tirages nécessaires” (comme pour l'exemple du début de cette section, on peut négliger le cas où il n'arrive jamais de boule blanche, car sa probabilité est nulle).
6. On considère une urne contenant 4 boules blanches et six boules noires. On tire deux fois une boule, avec remise. Étudier la variable aléatoire “nombre de boules blanches au premier tirage fois nombre de boules noires au deuxième.”
7. Refaire l'exercice 6, si le tirage se fait sans remise.

## 1.12 Fonction de répartition

Dans le cas de l'exemple fondamental de la section 1.10, nous avons vu que la variable aléatoire considérée admettait comme loi de probabilité la distribution des fréquences de la série statistique. Un autre outil classique en statistique descriptive est l'étude des fréquences cumulées de la série statistique. Il est donc naturel de définir l'analogue dans le cadre des variables aléatoires, à savoir la *Fonction de répartition*.

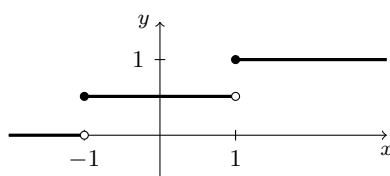
**Définition 1.12.1.** Soit  $X$  une variable aléatoire (discrète ou continue). La fonction de répartition de  $X$ , notée  $F$  est définie par

$$F : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \mathbb{P}(X \in ]-\infty, x]) = \mathbb{P}(X \leq x) = \mathbb{P}_X(]-\infty, x]).$$

**Exemple 1.12.1.** Traitons l'exemple de la roulette. On a directement

$$F(x) = \begin{cases} 0 & \text{si } x < -1 \\ 19/37 & \text{si } -1 \leq x < 1 \\ 1 & \text{si } 1 \leq x \end{cases}$$

Sa représentation graphique est donnée par

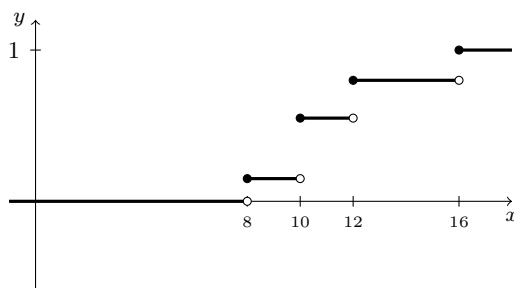


On a pris comme convention de représenter par un cercle noir la valeur prise par  $F$  en  $-1$  et en  $1$  et d'indiquer par un cercle blanc la valeur qui n'est pas prise par  $F$ .

De même pour l'exemple fondamental de la section 1.10, la fonction de répartition est donnée par

$$F(x) = \begin{cases} 0 & \text{si } x < 8 \\ 3/20 & \text{si } 8 \leq x < 10 \\ 11/20 & \text{si } 10 \leq x < 12 \\ 16/20 & \text{si } 12 \leq x < 16 \\ 1 & \text{si } 16 \leq x. \end{cases}$$

Sa représentation graphique est donnée par



On constate que la fonction de répartition n'est pas continue : il y a des sauts dans la représentations graphique. Ces sauts correspondent exactement aux éléments de  $\Omega_X$ , qui ont une probabilité non nulle.

Traitons maintenant un exemple de variable aléatoire admettant une fonction de répartition continue.

**Exemple 1.12.2.** On considère l'expérience aléatoire qui consiste à choisir au hasard un nombre dans l'intervalle  $[3, 7]$ . L'ensemble des événements  $\mathcal{E}$  sera l'ensemble des parties  $A$  de  $[3, 7]$  qui admettent une mesure  $L(A)$ . Si le choix est uniforme, la probabilité d'une partie  $A$  est *proportionnelle* à sa mesure. On obtient alors

$$\mathbb{P} : \mathcal{E} \rightarrow [0, 1] : A \mapsto \frac{L(A)}{L([3, 7])} = \frac{L(A)}{4}.$$

On définit alors la variable aléatoire

$$X : [3, 7] \rightarrow \mathbb{R} : x \mapsto \text{distance de } x \text{ à } 5.$$

Analytiquement, on a

$$X : [3, 7] \rightarrow \mathbb{R} : x \mapsto |x - 5|.$$

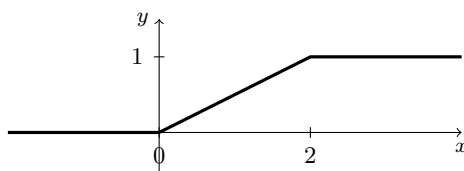
On peut vérifier que  $X$  est une variable aléatoire. Par exemple si  $0 < a < b$ , alors on a

$$X^{-1}(]a, b[) = [3, 7] \cap (]5 - b, 5 - a[ \cup ]5 + a, 5 + b[).$$

On a de plus  $\Omega_X = [0, 2]$  car tout point de  $[3, 7]$  a une distance à 5 comprise entre 0 et 2. Enfin, la fonction de répartition de  $X$  est donnée par

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x/2 & \text{si } 0 \leq x < 2 \\ 1 & \text{si } 2 \leq x. \end{cases}$$

Sa représentation graphique est donnée par



La fonction de répartition calculée dans ce dernier exemple est *continue*, mais dérivable seulement sur  $\mathbb{R} \setminus \{0, 2\}$ .

Les fonctions de répartition que nous venons de calculer ont des propriétés communes qu'il est utile de rassembler. C'est l'objet de la proposition suivante.

**Proposition 1.12.1.** *La fonction de répartition  $F$  d'une variable aléatoire  $X$  jouit des propriétés suivantes :*

- a) Si  $a < b$ , alors on a  $\mathbb{P}(a < X \leq b) = F(b) - F(a)$  ;
- b) Elle est croissante ;
- c) On a  $\lim_{x \rightarrow +\infty} F(x) = 1$  ;
- d) On a  $\lim_{x \rightarrow -\infty} F(x) = 0$  ;
- e) On a  $\lim_{x \rightarrow a^+} F(x) = F(a)$  pour tout réel  $a$ , i.e. la fonction  $F$  est continue à droite sur  $\mathbb{R}$  ;
- f) Si  $F$  est continue en  $a$ , alors  $\mathbb{P}(X = a) = 0$ .

On pourrait faire la preuve de cette proposition, qui n'est pas difficile, bien qu'un peu technique. Il semble plus pertinent d'en discuter les tenants et aboutissants.

La première propriété indique qu'il est possible de récupérer les probabilités d'intervalles à partir de la fonction de répartition. deuxième vient du fait que si on passe de

$F(x)$  à  $F(x')$  avec  $x < x'$ , on calcule dans le deuxième cas la probabilité d'un événement plus grand. La dernière est importante : si la fonction de répartition est continue, les événements élémentaires ont une probabilité nulle. Ce dernier fait peut paraître bizarre : la distribution de la taille est continue, et on a l'impression que la probabilité qu'un individu mesure un mètre 80 n'est pas nulle. C'est évidemment dû au fait que l'on associe à 1,8m toutes les mesures comprise entre 1 mètre 79 et demi et un mètre 80 et demi : on discrétise alors une distribution continue. Cette proposition admet une réciproque, qui permet de définir des variables aléatoires, sans pour autant définir explicitement l'expérience aléatoire qui permet de les définir.

**Proposition 1.12.2.** *Toute fonction  $F$  satisfaisant les propriétés b),c),d),e) de la proposition précédente est la fonction de répartition d'une unique variable aléatoire.*

### 1.13 Densité de probabilité d'une variable continue

Nous avons défini la loi ou distribution de probabilité d'une variable aléatoire discrète. Nous avons également pu remarquer qu'elle détermine en un certain sens la variation de la fonction de répartition de la variable discrète.

Quand la variable aléatoire est continue, il est donc naturel de définir la loi de probabilité (ou densité, ou distribution) comme la mesure de la variation de la fonction de répartition. Il serait donc logique de la définir comme la dérivée de  $F$ . Cependant, comme nous l'avons vu dans l'exemple 1.12.2, la fonction  $F$  n'est pas toujours dérivable. Il convient donc de prendre quelques précautions.

**Définition 1.13.1.** Soit  $X$  une variable aléatoire continue dont la fonction de répartition  $F$  est continue sur  $\mathbb{R}$ . On suppose qu'il existe un nombre fini de points  $x_1 < \dots < x_n$  tels que  $F$  soit continûment dérivable sur  $\mathbb{R} \setminus \{x_1, \dots, x_n\}$ . La *densité de probabilité de  $X$*  est la fonction  $f = F'$ .

*Remarque 1.6.* La densité de probabilité est encore appelée distribution ou loi de probabilité. Cette densité est définie et continue sur  $\mathbb{R} \setminus \{x_1, \dots, x_n\}$ .

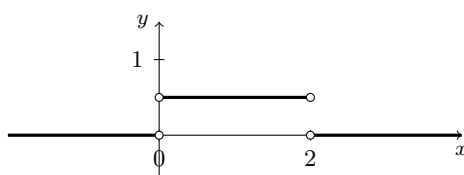
**Exemple 1.13.1.** Nous avons traité l'exemple d'une variable aléatoire continue en 1.12.2. La fonction de répartition est donnée par

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x/2 & \text{si } 0 \leq x < 2 \\ 1 & \text{si } 2 \leq x \end{cases}$$

Elle est donc dérivable sur  $\mathbb{R} \setminus \{0, 2\}$  et la densité de  $X$  est donnée par

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/2 & \text{si } 0 < x < 2 \\ 0 & \text{si } x > 2 \end{cases}$$

Sa représentation graphique est donnée par



Il s'agit d'un exemple de *variable aléatoire uniforme*.

Passons maintenant aux propriétés des densités de probabilités.

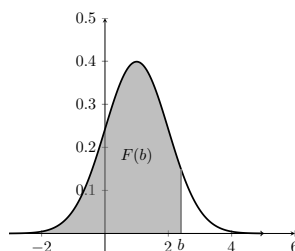
**Théorème 1.13.1.** *La densité de probabilité d'une variable  $X$  jouit de propriétés suivantes*

- a) *Si  $a, b \in \mathbb{R}$  sont tels que  $a < b$ , alors  $\int_a^b f(t)dt = F(b) - F(a)$ ,*
- b) *Pour tout réel  $b$ , on a  $F(b) = \int_{-\infty}^b f(t)dt$ ,*
- c) *On a  $\int_{-\infty}^{+\infty} f(t)dt = 1$ ,*
- d) *Pour tout  $x \in D_f$ , on a  $f(x) \geq 0$ .*

La démonstration de ce théorème repose sur le fait que  $f$  est la dérivée de  $F$ , sauf quand celle-ci n'est pas définie. Il faut alors être prudent en les points  $x_1, \dots, x_n$  en lesquels  $f$  n'est pas définie. On utilise alors la continuité de  $F$ .

*Remarque 1.7.* La deuxième propriété est également importante. Elle montre que la valeur de la fonction de répartition en  $b$  est l'aire comprise entre l'axe des abscisses et le graphe de  $f$ , limitée par la droite d'équation  $x = b$ .

Illustrons cette propriété sur un exemple classique :



Les deux dernières propriétés admettent une réciproque.

**Théorème 1.13.2.** *Si  $f : \mathbb{R} \rightarrow \mathbb{R}$  est définie et continue sur  $\mathbb{R}$  sauf en un nombre fini de points et vérifie*

- *Pour tout  $x \in D_f$ , on a  $f(x) \geq 0$ ;*
- *$\int_{-\infty}^{+\infty} f(t)dt = 1$ ;*

*alors  $f$  est la densité de probabilité d'une variable aléatoire continue.*

*Remarque 1.8.* La variable aléatoire ainsi définie admet pour fonction de répartition

$$F(x) = \int_{-\infty}^x f(t)dt.$$

Il nous reste pour terminer cette section, à définir l'espérance, la variance et l'écart-type d'une variable aléatoire continue. Ces concepts sont définis par analogie avec ceux du cas discret.

**Définition 1.13.2.** Soit  $X$  une variable aléatoire continue admettant une densité de probabilité  $f$ . L'**espérance mathématique** de  $X$  est définie par

$$E(X) = \int_{-\infty}^{+\infty} x f(x)dx.$$

La **variance** de  $X$  est définie par

$$\sigma_X^2 = V[X] = \int_{-\infty}^{+\infty} f(x)(x - E(X))^2 dx.$$

L'**écart-type** est défini par

$$\sigma_X = \sigma(X) = \sqrt{\sigma^2(X)}.$$

*Remarque 1.9.* Comme dans le cas discret, il peut arriver que l'espérance n'existe pas. Même quand celle-ci existe, il arrive que la variance n'existe pas. Comme dans le cas discret, l'espérance mathématique peut être interprétée comme un centre de gravité. C'est un *paramètre de tendance centrale ou de position* de  $X$ . De même la variance et l'écart-type sont des *paramètres de dispersion*.

**Exemple 1.13.2.** L'espérance mathématique, la variance et l'écart-type de l'exemple 1.12.2 sont donnés par

$$E(X) = \int_{-\infty}^{+\infty} f(x) \cdot x dx = \int_0^2 \frac{x}{2} dx = 1,$$

et

$$\sigma_X^2 = \int_{-\infty}^{+\infty} f(x)(x-1)^2 dx = \int_0^2 \frac{1}{2}(x-1)^2 dx = \frac{1}{3}.$$

On a donc  $\sigma_X = \frac{\sqrt{3}}{3}$ .

### 1.13.1 Exercices

1) Soit la fonction  $F$  définie par

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/2 & \text{si } 0 \leq x < 1 \\ 3/5 & \text{si } 1 \leq x < 2 \\ 9/10 & \text{si } 2 \leq x < 3.5 \\ 1 & \text{si } 3.5 \leq x \end{cases}$$

1. Donner la représentation graphique de la fonction  $F$ .
2. Déterminer la distribution d'une variable aléatoire  $X$  dont  $F$  serait la fonction de répartition.
3. Calculer  $\mathbb{P}(1 < X \leq 3.5)$  et  $\mathbb{P}(1 \leq X \leq 3)$ .
- 2) On considère l'expérience aléatoire consistant à lancer un dé équilibré et on note  $X$  la variable aléatoire "résultat obtenu". Déterminer  $\Omega_X$ , la loi de  $X$ , son espérance et sa variance.
- 3) On considère l'expérience aléatoire consistant à lancer deux dés équilibrés discernables et on note  $X$  la variable aléatoire "somme des points obtenus".
  1. Déterminer l'univers  $\Omega$  de l'expérience.
  2. Étudier  $\Omega_X$ , la loi de  $X$ , son espérance et sa variance.
  3. Comparer avec le résultat de l'exercice 2).
- 4) On considère l'expérience aléatoire consistant à lancer deux dés équilibrés discernables et on note  $X$  la variable aléatoire "produit des points obtenus".
  1. Déterminer l'univers  $\Omega$  de l'expérience.
  2. Étudier  $\Omega_X$ , la loi de  $X$ , son espérance et sa variance.
  3. Comparer avec le résultat de l'exercice 2).
- 5) Soit  $X$  une variable aléatoire continue dont la densité de probabilité est donnée par la fonction

$$f : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto \begin{cases} k(4x - 2x^2) & \text{si } 0 < x < 2 \\ 0 & \text{sinon} \end{cases}$$

Déterminer la valeur de  $k$ ,  $\sigma_X^2$  et  $\mathbb{P}(X > 1)$ .

## 1.14 Variables aléatoires indépendantes

On considère deux variables aléatoires  $X$  et  $Y$  sur un même espace probabilisé  $(\Omega, E, \mathbb{P})$ .

**Définition 1.14.1.** Les variables aléatoires  $X$  et  $Y$  sont indépendantes si pour tous intervalles  $I$  et  $J$ , on a

$$\mathbb{P}((X \in I) \cap (Y \in J)) = \mathbb{P}(X \in I) \cdot \mathbb{P}(Y \in J).$$

On constate que les variables aléatoires  $X$  et  $Y$  sont indépendantes si les événements  $(X \in I)$  et  $(Y \in J)$  sont indépendants quels que soient les intervalles  $I$  et  $J$  considérés. On a donc la même interprétation pour l'indépendance des variables aléatoires que pour celle de l'indépendance des événements. Si par exemple  $\mathbb{P}(Y \in J) \neq 0$ , cette condition s'écrit

$$\mathbb{P}((X \in I)|(Y \in J)) = \mathbb{P}(X \in I),$$

ce qui veut dire que la connaissance de la valeur prise par  $Y$  n'influence pas la probabilité que  $X$  prenne une certaine valeur.

Bien sûr, cette définition se généralise pour le cas de plusieurs variables aléatoires  $X_1, \dots, X_n$  définies sur un même espace probabilisé.

**Exemple 1.14.1.** On lance deux dés discernables et on note les points obtenus. Soit alors  $X$  la variable aléatoire "points marqués par le premier dé" et  $Y$  la variable "points marqués par le second dé". Ces variables sont indépendantes.

**Exemple 1.14.2.** Dans les mêmes conditions, soit  $X$  la variable "plus petit point marqué par les dés" et  $Y$  la variable "plus grand point marqué par les dés". Ces variables ne sont pas indépendantes.

**Exemple 1.14.3.** Dans les mêmes conditions, soit  $X$  la variable "points marqués par le premier dé" et  $Y$  la variable "somme des points marqués par les dés". Ces variables ne sont pas indépendantes.

*Remarque 1.10.* On note souvent l'événement  $((X \in I) \cap (Y \in J))$  par  $(X \in I \text{ et } Y \in J)$ , ou encore  $(X \in I, Y \in J)$ .

Dans le cas de variables aléatoires discrètes, la condition d'indépendance peut être plus facilement vérifiée : on peut se passer de regarder tous les intervalles et se concentrer sur les points où se concentre la densité de probabilité. C'est l'objet de la proposition suivante.

**Proposition 1.14.1.** Les variables aléatoires discrètes  $X$  et  $Y$  sont indépendantes si

$$\mathbb{P}(X = x \text{ et } Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y), \quad \forall x \in \Omega_X, y \in \Omega_Y.$$

## 1.15 Somme produit, combinaisons de variables aléatoires

Il est parfois utile de pouvoir combiner deux ou plusieurs variables aléatoires sur un même espace probabilisé pour n'en former qu'une seule. Par exemple, dans le cas du lancer de deux dés, si on note  $X_1$  la variable "résultat du premier dé" et  $X_2$  la variable "résultat du second dé", la somme des points obtenus s'écrira  $X_1 + X_2$ , la moyenne des points obtenus  $(X_1 + X_2)/2$  et le produit  $X_1 X_2$ . Si on reçoit 5\$ par point, la variable "gain" sera  $5(X_1 + X_2)$ .

Dans le cas où l'univers  $\Omega$  de l'expérience considérée n'est pas dénombrable, il faut toutefois vérifier que de telles combinaisons de variables aléatoires en sont encore. Nous admettrons la proposition suivante.

**Proposition 1.15.1.** *Si  $X$  et  $Y$  sont des variables aléatoires sur  $(\Omega, E, \mathbb{P})$  et si  $a$  est un nombre réel, alors les fonctions*

$$\begin{aligned} X + Y &: \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega) + Y(\omega) \\ aX &: \Omega \rightarrow \mathbb{R} : \omega \mapsto a \cdot X(\omega) \\ XY &: \Omega \rightarrow \mathbb{R} : \omega \mapsto X(\omega) \cdot Y(\omega) \end{aligned}$$

*sont des variables aléatoires.*

Cette proposition peut être généralisée. Celle qui suit permet de faire des transformations de variables aléatoires.

**Proposition 1.15.2.** *Si  $X$  est une variable aléatoire sur  $(\Omega, E, \mathbb{P})$  et si  $g : \mathbb{R} \rightarrow \mathbb{R}$  est une application continue, alors*

$$g \circ X : \Omega \rightarrow \mathbb{R} : \omega \mapsto g(X(\omega))$$

*est une variable aléatoire sur  $(\Omega, E, \mathbb{P})$ .*

### 1.15.1 Les théorèmes

Nous allons maintenant voir à quelles conditions on peut exprimer l'espérance et la variance de ces nouvelles variables aléatoires en fonctions de celles de  $X$  et  $Y$ . Les démonstrations sont omises. On pourrait bien sûr donner quelques exemples et contre-exemples pour se convaincre de la validité de ces théorèmes, et montrer l'importance des hypothèses. Je ne le ferai pas pour ne pas alourdir cet exposé. On considère deux variables aléatoires  $X$  et  $Y$  sur un même espace probabilisé  $(\Omega, E, \mathbb{P})$  et un nombre réel  $a$ . On a alors les résultats suivants.

**Théorème 1.15.1** (Espérance). *Si  $E(X)$  et  $E(Y)$  existent, alors  $E(X + Y)$  existe et on a*

$$E(X + Y) = E(X) + E(Y).$$

*Si  $E(X)$  existe, alors  $E(aX)$  existe et on a*

$$E(aX) = aE(X).$$

*Si  $X$  et  $Y$  sont indépendantes et si  $E(X)$  et  $E(Y)$  existent, alors  $E(XY)$  existe et on a*

$$E(XY) = E(X) \cdot E(Y).$$

Les deux premières conditions signifient que l'espérance mathématique est une application linéaire sur l'espace vectoriel des variables aléatoires sur  $(\Omega, E, \mathbb{P})$  qui admettent une espérance.

**Théorème 1.15.2** (Variance). *Si  $X$  et  $Y$  sont indépendantes et si  $\sigma^2(X)$  et  $\sigma^2(Y)$  existent, alors  $\sigma^2(X + Y)$  existe et on a*

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y).$$

*Si  $\sigma^2(X)$  existe, alors  $\sigma^2(aX)$  existe et on a*

$$\sigma^2(aX) = a^2\sigma^2(X), \quad \sigma(aX) = |a|\sigma(X).$$



### 1.15.2 Résultats supplémentaires

On peut tout d'abord étendre les combinaisons en considérant les variables aléatoires constantes :

**Définition 1.15.1.** Si  $c \in \mathbb{R}$ , la variable aléatoire constante  $c$  est définie sur  $(\Omega, E, \mathbb{P})$  par

$$c : \Omega \rightarrow \mathbb{R} : \omega \mapsto c.$$

Les caractéristiques de cette variables sont simples : On a

$$\Omega_c = \{c\}, \quad E(c) = c, \quad \sigma^2(c) = 0.$$

De plus, puisque  $c$  est visiblement indépendante de toute autre variable  $X$  définie sur le même espace, les théorèmes précédents s'appliquent :

**Théorème 1.15.3.** Si  $a$  et  $b$  sont deux nombres réels et  $X$  une variable aléatoire sur  $(\Omega, E, \mathbb{P})$  admettant une espérance et une variance, alors  $aX + b$  admet une espérance et une variance et on a

$$E(aX + b) = aE(X) + b, \quad \sigma^2(aX + b) = a^2\sigma^2(X), \quad \sigma(aX + b) = |a|\sigma(X).$$

Le résultat suivant permet de simplifier le calcul de la variance. Son importance se mesure quand il s'agit de calculer une variance "à la main".

Dans le cas d'une variable discrète, d'après la définition, on a

$$E(X^2) = \sum_i p_i x_i^2.$$

Dans le cas d'une variable continue, ceci se généralise par

$$E(X^2) = \int_{-\infty}^{+\infty} f(x) \cdot x^2 dx.$$

On a alors la formule de Koenig, donnée par le théorème suivant.

**Théorème 1.15.4.** Soit  $X$  une variable aléatoire sur  $(\Omega, E, \mathbb{P})$  admettant une espérance et une variance, alors

$$\sigma^2(X) = E(X^2) - (E(X))^2.$$

Dans la même veine, on a le résultat suivant, qui lie variance de la somme et covariance.

**Proposition 1.15.3.** Si  $X$  et  $Y$  sont deux variables sur le même espace probabilisé admettant une espérance et une variance, on a

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) + 2[E(XY) - E(X)E(Y)].$$

En particulier,  $\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y)$  si et seulement si  $E(XY) - E(X)E(Y) = 0$ .

Enfin, pour conclure cette section, citons la généralisation au cas de plusieurs variables aléatoires  $X_1, \dots, X_n$  définies sur le même espace  $(\Omega, E, \mathbb{P})$ .

**Définition 1.15.2.** Les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes si pour tous intervalles  $I_1, \dots, I_n$ , les événements  $(X_1 \in I_1), \dots, (X_n \in I_n)$  sont indépendants.

Les théorèmes se généralisent alors comme suit.

**Proposition 1.15.4.** Si  $X_1, \dots, X_n$  sont indépendantes, et si les espérances  $E(X_1), \dots, E(X_n)$  existent alors

$$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n).$$

Si  $X_1, \dots, X_n$  sont indépendantes, et si  $\sigma^2(X_1), \dots, \sigma^2(X_n)$  existent, alors

$$\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n).$$

Cette proposition sera importante dans le cadre de la statistique inductive.

## Chapitre 2

# Quelques distributions et variables usuelles

Nous allons dans ce chapitre étudier quelques variables aléatoires classiques. Il sera utile de pouvoir reconnaître les situations dans lesquelles elles apparaissent, les formules donnant leur espérance et leur variance, ainsi que les liens entre les différentes variables, qui se présentent sous forme de théorèmes d'approximation.

### 2.1 La loi uniforme discrète

La variable aléatoire  $X$  suit une loi uniforme discrète si  $\Omega_X$  est fini ( $\Omega_X = \{x_1, \dots, x_n\}$ ) et si  $\mathbb{P}_X$  est la probabilité uniforme sur  $\Omega_X$ , c'est-à-dire si on a  $\mathbb{P}(x_i) = \frac{1}{n}$  pour tout  $i \leq n$ . La loi de  $X$  est donnée par le tableau suivant :

$x_i$	$x_1$	$\cdots$	$x_n$
$p_i$	$1/n$	$\cdots$	$1/n$

L'espérance et la variance d'une telle variable sont données par

$$E(X) = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma^2(X) = \frac{1}{n} \sum_{i=1}^n (x_i)^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2.$$

En effet, le premier résultat découle de la définition, tandis que le deuxième utilise la formule de Koenig.

### 2.2 La loi uniforme continue

La variable aléatoire  $X$  suit une loi uniforme continue si  $\Omega_X$  est un intervalle  $[a, b]$  et si  $\mathbb{P}_X$  est la probabilité uniforme sur  $[a, b]$ . Nous avons rencontré une telle variable dans l'exemple 1.12.2. La fonction de répartition est la suivante :

$$F(x) = \begin{cases} 0 & \text{et si } x < a \\ \frac{x-a}{b-a} & \text{et si } a \leq x < b \\ 1 & \text{et si } b \leq x. \end{cases}$$

On peut alors calculer la densité de probabilité en dérivant  $F$  :

$$f(x) = \begin{cases} 0 & \text{et si } x < a \\ \frac{1}{b-a} & \text{et si } a < x < b \\ 0 & \text{et si } x > b \end{cases}$$

L'espérance et la variance de  $X$  sont alors données par

$$E(X) = \frac{a+b}{2}, \quad \sigma^2(X) = \frac{(a-b)^2}{12}.$$

## 2.3 Variables de Bernoulli

On considère un espace probabilisé  $(\Omega, \mathcal{E}, \mathbb{P})$  et  $A \in \mathcal{E}$ . La variable indicatrice de  $A$ , encore appelée variable de Bernoulli associée à  $A$ , est

$$X : \Omega \rightarrow \mathbb{R} : \omega \mapsto \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases}$$

On la note également  $1_A$ . L'événement  $A$  est appelé succès tandis que son complémentaire est appelé échec. L'expérience en question est également appelée épreuve de Bernoulli. On note  $p$  la probabilité de succès ( $p = \mathbb{P}(A)$ ).

Les caractéristiques de la variables de Bernoulli sont données par le résultat suivant.

**Proposition 2.3.1.** *On a  $E(1_A) = p$  et  $\sigma^2(1_A) = p(1-p)$ .*

Cette variable aléatoire va intervenir dans la section suivante. On la rencontre souvent quand on fait des comptages, ou quand on s'intéresse au caractère particulier dans la population qui peut être présent ou absent, ou qui n'a que deux valeurs possibles.

## 2.4 La loi binomiale

Cette variable très importante apparaît dans la situation suivante.

*On répète  $n$  fois, de manière indépendante, une épreuve de Bernoulli pouvant conduire à un succès  $S$  avec une probabilité  $p$  et à un échec avec une probabilité  $1-p$ . On considère alors la variable aléatoire  $X$  : "nombre de succès enregistrés".*

Une variable  $X$  définie de la sorte est une variable binomiale de paramètres  $n$  et  $p$ . On note  $X \sim B(n, p)$ . L'exemple le plus classique d'une telle variable est le tirage *avec remise* de boules de couleurs dans une urne et la variable  $X$  qui compte le nombre de boules d'une certaine couleur qui sont sorties. Chaque tirage constitue alors une épreuve de Bernoulli où un succès correspond au tirage d'une boule de la couleur souhaitée.

Le théorème suivant donne les caractéristiques principales des variables binomiales.

**Théorème 2.4.1.** *Si  $X \sim B(n, p)$ , alors on a*

$$\Omega_X = \{0, \dots, n\}, \quad \text{et } \forall k \in \Omega_X, \quad \mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}.$$

*De plus on a*

$$E(X) = np \quad \text{et} \quad \sigma^2(X) = np(1-p).$$

*Démonstration.* Le nombre de succès enregistrés sur  $n$  expériences indépendantes est évidemment un nombre entier entre 0 et  $n$ . L'événement  $(X = k)$  est alors formé des  $n$ -uplets formés de  $S$  et  $\bar{S}$  et contenant exactement  $k$  fois  $S$  (par exemple  $(\underbrace{S, \dots, S}_{k \text{ fois}}, \underbrace{\bar{S}, \dots, \bar{S}}_{n-k \text{ fois}})$ ). La

probabilité de chacun d'eux est  $p^k(1-p)^{n-k}$ . Leur nombre correspond au nombre de choix possibles des numéros des expériences conduisant à un succès,  $C_n^k$  donc.

Si on note  $1_{A_i}$  la variable de Bernoulli associée à l'événement "la  $i$ -ème expérience donne un succès, alors  $X = 1_{A_1} + \dots + 1_{A_n}$  et les théorèmes sur l'espérance et la variance s'appliquent pour conduire au résultat annoncé.  $\square$

## 2.5 La loi hypergéométrique

Les variables de type hypergéométrique sont définies par le processus suivant. On considère une population de  $N$  individus parmi lesquels on distingue  $n_1$  individus de type 1 et  $n_2 = N - n_1$  individus de type 2. On pose alors  $p = \frac{n_1}{N}$  et  $q = \frac{n_2}{N}$ . On tire au hasard sans remise un échantillon de  $n$  individus et on considère la variable aléatoire  $X$  : “Nombre d’individus de type 1 choisis”

On dit que  $X$  suit une loi hypergéométrique de paramètres  $N, n$  et  $p$ . On note alors  $X \sim H(N, n, p)$  et on a alors

**Proposition 2.5.1.** Si  $X$  suit une loi  $H(N, n, p)$  alors on a

—  $\Omega_X = \{m_1, \dots, m_2\}$  où  $m_1 = \max\{0, n - n_2\}$  et  $m_2 = \min\{n, n_1\}$ .

—  $\forall k \in \Omega_X, \mathbb{P}(X = k) = \frac{C_{n_1}^k C_{n_2}^{n-k}}{C_N^n}$

De plus l’espérance et la variance de  $X$  sont données par

$$E(X) = np, \sigma^2(X) = np(1-p) \frac{N-n}{N-1}.$$

*Démonstration.* On numérote les individus de type 1 de 1 à  $n_1$  et ceux de type 2 de  $n_1 + 1$  à  $N$ . On considère le nombre d’échantillons de taille  $n$  qu’on peut tirer (l’ordre dans lequel on choisit les individus n’ayant pas d’importance) dans une population de  $N$  individus. Nous savons que ce nombre est  $C_N^n$ . D’autre part, l’événement  $(X = k)$  est formé des échantillons contenant exactement  $k$  individus de type 1 et  $n - k$  individus de type 2. On considère que tous les échantillons sont équiprobables. Il suffit donc de dénombrer le nombre d’échantillons de l’événement  $(X = k)$ . Cela correspond aux différentes possibilités pour choisir  $k$  individus parmi  $n_1$  et  $n - k$  parmi  $n_2$ . On obtient donc bien

$$\mathbb{P}(X = k) = \frac{C_{n_1}^k C_{n_2}^{n-k}}{C_N^n}.$$

D’autre part, si on note  $A_i$  l’événement “l’individu numéro  $i$  est dans l’échantillon choisi” on obtient

$$X = 1_{A_1} + \dots + 1_{A_{n_1}}.$$

On a de plus, puisque le nombre d’échantillons dans l’événement  $A_i$  est le nombre de choix de  $n - 1$  individus dans les  $N - 1$  individus différents de  $i$

$$\mathbb{P}(A_i) = \frac{C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N}.$$

On obtient ainsi

$$E(X) = \sum_{i=1}^{n_1} E(1_{A_i}) = n_1 \frac{n}{N} = np.$$

Pour la variance, on utilise la formule de Koenig :

$$\sigma^2(X) = E(X^2) - (E(X))^2.$$

Il faut cependant calculer la variable  $X^2$  :

$$X^2 = \sum_i \sum_j (1_{A_i} 1_{A_j}) = \sum_i (1_{A_i})^2 + \sum_{i \neq j} 1_{A_i} 1_{A_j}$$

On voit tout de suite que les variables  $(1_{A_i})^2$  et  $(1_{A_i})$  coïncident. De plus, on a  $\Omega_{1_{A_i}1_{A_j}} = \{0, 1\}$  et

$$\begin{aligned} \mathbb{P}(1_{A_i}1_{A_j} = 1) &= \mathbb{P}(1_{A_i} = 1 \text{ et } 1_{A_j} = 1) \\ &= \mathbb{P}(1_{A_i} = 1 | 1_{A_j} = 1) \mathbb{P}(1_{A_j} = 1) = \frac{n-1}{N-1} \frac{n}{N}. \end{aligned}$$

On obtient alors

$$E(X^2) = n_1 \frac{n}{N} + (n_1^2 - n_1) \frac{n-1}{N-1} \frac{n}{N}.$$

Ceci permet de montrer que

$$\sigma^2(X) = E(X^2) - (E(X))^2 = np(1-p) \frac{N-n}{N-1},$$

ce qui achève la démonstration. □

*Remarque 2.1.* Il est utile de noter que l'espérance de la variable  $H(N, n, p)$  coïncide avec celle de la variable  $B(n, p)$ , et que leurs variances ne diffèrent que par le facteur  $\frac{N-n}{N-1}$ .

### 2.5.1 Exercices

- 1) Dans un groupe formé de 4 garçons et 6 filles on choisit 5 fois un enfant au hasard (avec remise) et on considère la variable : “nombre de garçons choisis”. Déterminer la loi de  $X$ , son espérance et sa variance.
- 2) Dans un groupe formé de 4 garçons et 6 filles on choisit une équipe de 5 enfants et on considère la variable  $Y$  : “nombre de garçons choisis”. Déterminer la loi de  $Y$ , son espérance et sa variance.
- 3) Faire de même avec un choix au hasard de 5 enfants parmi 400 garçons et 600 filles.

### 2.5.2 Solutions

- 1) Par définition,  $X$  suit une loi binomiale de paramètres  $n = 5$  et  $p = 0,4$  ( $X \sim B(5; 0,4)$ ). On a donc  $E(X) = np = 2$  et  $\sigma^2(X) = np(1-p) = 1,2$ . On a de plus  $\Omega_X = \{0, \dots, 5\}$  et la distribution est donnée par le tableau suivant :

$x_i$	0	1	2	3	4	5
$p_i$	0,07776	0,2592	0,3456	0,2304	0,0768	0,01024

- 2) La variable  $Y$  suit une loi hypergéométrique de paramètres  $N = 10$ ,  $n = 5$ , et  $p = 0,4$  ( $Y \sim H(10; 5; 0,4)$ ), on a donc  $\Omega_Y = \{0, \dots, 4\}$  (le nombre maximal de garçons est 4), la loi est donnée par le tableau

$x_i$	0	1	2	3	4
$p_i$	1/42	10/42	20/42	10/42	1/42

l'espérance est donnée par  $E(Y) = np = 5 \cdot 0,4 = 2$  et la variance est donnée par  $\sigma^2(Y) = 5 \cdot 0,4 \cdot \frac{5}{9} = \frac{2}{3}$ .

- 3) La variable  $X$  suit toujours une loi binomiale  $B(5; 0,4)$ . Seules la probabilité de succès et le nombre d'expériences comptent. La variable  $Y$  suit par contre une loi  $H(100; 5; 0,4)$ , dont la distribution est donnée par

$x_i$	0	1	2	3	4	5
$p_i$	0,0725	0,259	0,3545	0,2322	0,0728	0,0087

Son espérance est donnée par  $E(Y) = 5 \cdot 0,4 = 2$  et sa variance est donnée par  $\sigma^2(Y) = 5 \cdot 0,4 \cdot 6 \cdot \frac{95}{99} = 1.151$ .

### 2.5.3 Théorème d'approximation

Il ressort des exercices précédents que la loi hypergéométrique peut être approchée par la loi Binomiale. Nous admettrons le résultat suivant.

**Proposition 2.5.2.** *La loi hypergéométrique  $H(N, n, p)$  tend vers<sup>a</sup> la loi  $B(n, p)$  quand  $N$  tend vers l'infini,  $n$  et  $p$  étant fixés.*

En pratique, on considère que l'approximation est bonne quand  $\frac{n}{N} < 0,1$ . Si tel est le cas, on note  $Y \approx B(n, p)$ .

### 2.5.4 Exercices supplémentaires

1. On considère un jeu de cartes contenant 48 cartes dont 4 as. On choisit au hasard 12 cartes et on considère  $X$  : nombre d'as obtenus. Déterminer la loi suivie par  $X$ , son espérance et sa variance, ainsi que  $\mathbb{P}(X = 2)$ .  
Peut-on approcher  $X$  par une binomiale ?
2. Avec le même jeu qu'en 1. (contenant 12 coeurs), on tire une main de 4 cartes au hasard et on considère la variable  $Y$  : nombre de coeurs obtenus. Déterminer la loi suivie par  $Y$ , son espérance et sa variance, ainsi que  $\mathbb{P}(Y = 2)$ .  
Peut-on approcher  $Y$  par une binomiale ? Si oui, donner la valeur approchée de  $\mathbb{P}(Y = 2)$ .

## 2.6 La loi géométrique

Ce type de variable apparaît dans les circonstances suivantes.

*On considère une épreuve de Bernoulli donnant un succès  $S$  avec une probabilité  $p$ . On répète l'expérience de manière indépendante jusqu'à obtenir un succès et on considère la variable  $X$  : rang du succès.*

La variable aléatoire ainsi définie suit une loi géométrique de paramètre  $p$ . On note  $X \sim G(p)$ . On a alors le résultat suivant

**Proposition 2.6.1.** *Si  $X \sim G(p)$ , alors si  $q = 1 - p$ , on a*

1.  $\Omega_X = \mathbb{N}_0$ ,
2. Pour tout  $k$ ,  $\mathbb{P}(X = k) = q^{k-1}p$ ,
3.  $E(X) = \frac{1}{p}$ ,  $\sigma^2(X) = \frac{1-p}{p^2}$ .

*Démonstration.* On a, comme dans l'exemple du dé (voir section 1.11),  $\Omega_X = \mathbb{N}_0$  car on néglige l'événement où le succès n'apparaîtrait pas. De plus,

$$\mathbb{P}(X = k) = \mathbb{P}(\bar{S}, \dots, \bar{S}, S) = q^{k-1}p.$$

Pour l'espérance, on utilise la formule

$$\sum_{k=1}^{+\infty} kq^k = \frac{q}{(1-q)^2}$$

a. Bien sûr, je n'ai pas tenté d'expliquer la notion de convergence qui est utilisée dans cette affirmation. Sachez cependant qu'il existe plusieurs notions de convergence en probabilité. Dans notre étude, nous nous contenterons de remplacer la distribution  $H$  par la distribution  $B$  dans certaines conditions pratiques.

qui fournit

$$E(X) = \sum_{k=1}^{+\infty} kq^{k-1}p = \frac{p}{q} \sum_{k=1}^{+\infty} kq^k = \frac{p}{q} \frac{q}{(1-q)^2} = \frac{1}{p}.$$

Pour la variance, on utilise

$$\sum_{k=1}^{+\infty} k^2 q^k = \frac{q(q+1)}{(1-q)^3}.$$

On a alors

$$E(X^2) = \sum_{k=1}^{+\infty} k^2 q^{k-1} p = \frac{p}{q} \frac{q(q+1)}{(1-q)^3} = \frac{q+1}{p^2}$$

et

$$\sigma^2(X) = \frac{q+1}{p^2} - \frac{1}{p^2},$$

ce qui donne la conclusion. □

### 2.6.1 Exercices

1. On lance deux dés équilibrés dont les faces portent les nombres 1,1,1,2,2,3. Le succès  $S$  consiste à obtenir un nombre total d'au moins 5 points. Calculer  $\mathbb{P}(S)$ .
2. On lance ces deux dés jusqu'à obtenir un succès et considère  $X$  : nombre de lancers. Déterminer la loi de  $X$ ,  $E(X)$ ,  $\sigma^2(X)$ ,  $\mathbb{P}(X = E(X))$ ,  $\mathbb{P}(X \leq E(X))$ . Trouver le plus petit entier  $n$  tel que  $\mathbb{P}(X \leq n) \geq \frac{1}{2}$ .

## 2.7 La loi de Poisson

On considère le problème suivant :

Une caisse d'un magasin est ouverte 400 minutes par jour. Il y a en moyenne 100 clients par jour à cette caisse. On choisit une minute donnée (ex : 12h-12h01). On considère alors la variable aléatoire  $X$  : nombre de clients se présentant à cette caisse durant cette minute.

On réalise en fait 100 épreuves de Bernoulli (une par client)

- Succès de l'épreuve numéro  $i$  : le client numéro  $i$  se présente à la caisse dans la minute choisie.
- Échec : il s'y présente durant une autre minute

On considère les hypothèses de travail suivantes :

- Les épreuves sont indépendantes.
- Pour un client donné, toutes les minutes sont équiprobables de sorte que la probabilité de succès est  $1/400$ .

La variable  $X$  suit alors une loi  $B(100, 1/400)$ . Si on considère le même problème sur deux jours, on obtient une loi  $B(200, 1/800)$ , sur une semaine de 6 jours,  $B(600, 1/2400)$ .

On constate que ce sont des variables binomiales avec  $n$  grand et  $p$  petit, et telles que le produit  $pn$  soit une constante  $\lambda = 1/4$ . Cette constante représente le nombre moyen de clients se présentant à la caisse par unité de temps (une minute). Il est donc naturel d'étudier la limite pour  $n$  tendant vers l'infini de  $B(n, \lambda/n)$  On a alors le résultat suivant.

**Proposition 2.7.1.** *Si  $X_n \sim B(n, \lambda/n)$ , alors on a*

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

**Définition 2.7.1.** Soit  $\lambda$  une constante strictement positive. Une variable aléatoire  $X$  suit une loi de Poisson de paramètre  $\lambda$ , ce qu'on note  $X \sim P(\lambda)$ , si on a  $\Omega_X = \mathbb{N} = \{0, 1, 2, \dots\}$  et

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

pour tout  $k \in \mathbb{N}$ .

On montre le résultat suivant, en utilisant les mêmes formules que pour la loi géométrique.

**Proposition 2.7.2.** Si  $X \sim P(\lambda)$ , alors  $E(X) = \lambda$  et  $\sigma^2(X) = \lambda$ .

La définition de la loi de Poisson conduit à un résultat d'approximation.

**Proposition 2.7.3.** Si  $n \geq 50$ ,  $p < 0, 1$ , la loi binomiale  $B(n, p)$  peut être approchée par  $P(np)$  (la loi de Poisson ayant la même espérance).

### 2.7.1 Exercices

- Dans un livre de 400 pages, il y a 100 fautes d'impression. On considère une page du livre et la variable aléatoire  $X$  : nombre de fautes dans la page choisie. Quelle est la loi suivie par  $X$ . Peut-on approcher cette loi par une loi de Poisson. Comparer les valeurs  $p_0$ ,  $p_1$  et  $p_2$  pour les deux lois.
- Un vaccin provoque un accident grave pour 10000 vaccinations. On administre le vaccin à 25000 individus. Étudier la loi de la variable aléatoire  $X$  : nombre d'accidents survenus. Par quelle loi peut-on approcher la loi de  $X$ ? Utiliser cette approximation pour calculer  $\mathbb{P}(X > 6)$ .
- Un standard téléphonique reçoit en moyenne 1,8 appels par minute. Quel est la probabilité qu'entre 11h23 et 11h24 il y ait au moins deux appels?  
Quelle est la probabilité qu'entre 18h15 et 18h25 il y ait au moins 20 appels?
- Vous participez à 50 tirages consécutifs et indépendants d'une loterie. A chaque tirage, la probabilité de remporter un prix est de  $\frac{1}{100}$ . On note  $X$  la variable aléatoire "nombre de prix remportés".
  - Quelle est la loi de probabilité de  $X$ ?
  - Par quelle variable aléatoire la variable  $X$  peut-elle être approchée?
  - Calculer les probabilités exactes et approchée(s) de gagner au moins un prix, exactement un prix, au moins deux prix.
- Les clients entrent dans un magasin au rythme d'une personne toutes les deux minutes.
  - Quelle est la probabilité qu'il n'entre personne entre 12h et 12h05?
  - Quelle est la probabilité que 4 clients au moins se présentent durant cette même période?
- Un journaliste se voit remettre une liste de personnes à interviewer. Il doit interroger 5 personnes au moins. Les personnes de la liste acceptent de lui répondre avec une probabilité  $\frac{2}{3}$ , indépendamment les uns des autres.
  - Quelle est la probabilité qu'il réalise un nombre suffisant d'entretiens si la liste contient 5 noms, 8 noms?
  - Dans ce dernier cas, quelle est la probabilité qu'il puisse parler à 6 personnes exactement, à 7 personnes?
- Afin de tester une solution toxique, on fait des injections à 81 souris. On admet l'hypothèse que l'injection est mortelle dans 80 % des cas. Étudier la variable aléatoire "nombre de souris mortes". Le fait d'observer que 22 souris ne soient pas mortes est-il compatible avec cette hypothèse?



## 2.8 La loi normale, ou loi gaussienne

### 2.8.1 Définition

La loi Normale est encore appelée loi de Laplace-Gauss. Donnons tout de suite la définition.

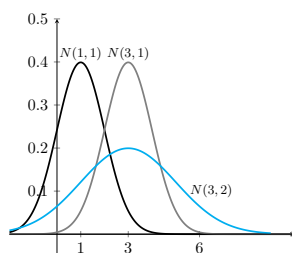
**Définition 2.8.1.** Une variable aléatoire continue  $Y$  suit une loi normale  $N(m, \sigma)$  si elle admet pour densité de probabilité la fonction

$$f : \mathbb{R} \rightarrow [0, 1] : x \mapsto \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

On peut calculer l'espérance et la variance des variables  $N(m, \sigma)$  en utilisant le calcul intégral. On obtient le théorème suivant.

**Proposition 2.8.1.** Si  $Y \sim N(m, \sigma)$ , alors  $E(Y) = m$  et  $\sigma^2(Y) = \sigma^2$ .

Il est utile de tracer le graphe de la fonction  $f$ . On vérifie aisément qu'elle prend des valeurs positives, que son graphe est symétrique par rapport à la droite d'équation  $x = m$  et qu'elle admet l'axe  $y = 0$  comme asymptote horizontale en  $+\infty$  et en  $-\infty$ . De plus elle admet un maximum en  $x = m$ . Sa représentation graphique est donc donnée par une courbe "en cloche", dont voici quelques exemples.



Comme attendu, dans  $N(m, \sigma)$ , le paramètre  $m$  détermine la position de la distribution (tendance centrale) et le paramètre  $\sigma$  son aplatissement, déterminant également sa dispersion autour de l'espérance.

### 2.8.2 Normale centrée réduite

Les lois normales dépendent de deux paramètres, mais on voit aisément sur les graphiques précédents qu'elles sont translatées et dilatées l'une de l'autre. Elles sont donc décrites à partir d'une seule, que l'on choisit la plus simple possible. C'est la loi normale centrée réduite.

**Définition 2.8.2.** La variable aléatoire  $Y$  suit une loi normale centrée réduite si  $Y \sim N(0, 1)$ . On note encore  $f_0$  la densité de probabilité d'une telle variable et  $\pi$  sa fonction de répartition.

La proposition suivante donne le lien fondamental entre les lois normales et la loi normale centrée réduite.

**Proposition 2.8.2.** Si  $Y \sim N(m, \sigma)$  alors  $\frac{Y-m}{\sigma} \sim N(0, 1)$ .

*Démonstration.* Soit  $Y$  une variable suivant la loi  $N(m, \sigma)$ . On a alors

$$\begin{aligned} \mathbb{P}\left(a < \frac{Y-m}{\sigma} \leq b\right) &= \mathbb{P}(a\sigma + m < Y \leq b\sigma + m) \\ &= \int_{a\sigma+m}^{b\sigma+m} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}t^2} dt \\ &= \pi(b) - \pi(a), \end{aligned}$$

ce qui conclut la preuve. □

Cette dernière formule était classiquement très utile en pratique puisqu'il suffit de connaître les valeurs de la fonction de répartition de la normale centrée réduite (qui étaient données dans les tables) pour connaître ces valeurs pour toutes les lois normales. Elle est moins utile en pratique maintenant que les ordinateurs permettent d'avoir les valeurs immédiatement. Citons encore une dernière propriété de  $\pi$  qui permet de lire les tables.

**Proposition 2.8.3.** *Pour tout réel  $t$  (non négatif), on a  $\pi(-t) = 1 - \pi(t)$ .*

### 2.8.3 Approximation de la loi Binomiale

Considérons une variable aléatoire  $X_n$  suivant une loi binomiale  $B(n, p)$  et analysons son comportement quand  $n$  devient grand. La figure suivante reprend les distributions de quelques variables binomiales et les variables normales ayant la même espérance et la même variance.

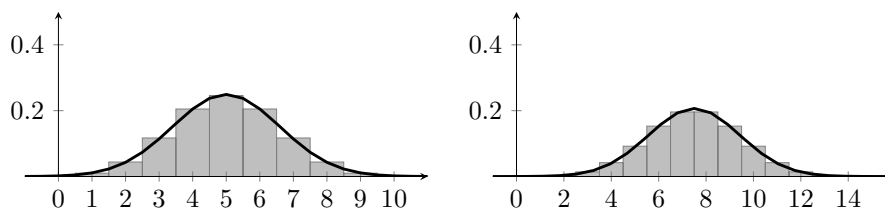


FIGURE 2.1 – Les distributions binomiales  $B(10, \frac{1}{2})$  et  $B(15, \frac{1}{2})$

Dans la figure suivante, on considère les distributions binomiales correspondant à  $n = 30$  et  $n = 50$

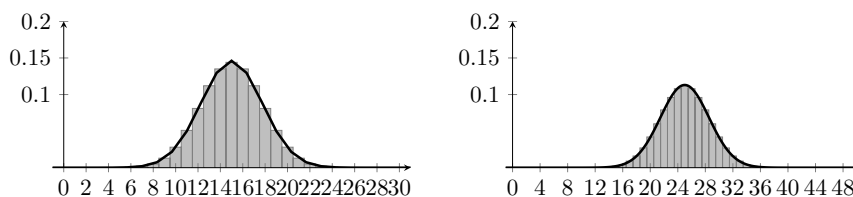


FIGURE 2.2 – Les distributions binomiales  $B(30, \frac{1}{2})$  et  $B(50, \frac{1}{2})$

On constate qu'on peut approximer, quand  $n$  est suffisamment grand, la somme des longueurs des bâtons par les aires qu'ils définissent. Cela amène le résultat suivant, qui est déjà une transposition pratique d'un théorème de convergence asymptotique.

**Proposition 2.8.4.** *Si  $np(1-p) > 9$ , on peut approximer la loi binomiale  $B(n, p)$  par la loi normale de même espérance et de même variance ( $N(m, \sigma)$  où  $m = np$  et  $\sigma = \sqrt{np(1-p)}$ ). Plus précisément, si  $X \sim B(n, p)$  et si  $Y \sim N(m, \sigma)$  alors*

$$\forall a \leq b \in \{0, \dots, n\}, \mathbb{P}(a \leq X \leq b) \approx \mathbb{P}(a - \frac{1}{2} < Y < b + \frac{1}{2}).$$

Il est important de noter que le résultat original est un théorème de convergence asymptotique, qui donne un résultat sur une limite pour  $n$  tendant vers  $+\infty$ . D'un point de vue mathématique, on ne peut pas déterminer quand  $n$  est assez grand. On a donc dû fixer une convention pratique, qui tient compte de la grandeur de  $n$ , mais qui tient compte aussi de  $p$ , parce que la convergence est moins bonne quand  $p$  est proche de 0 ou de 1.

Quant à la dernière égalité, elle tient compte d'une correction de continuité : on passe en effet d'une loi discrète à une loi continue. Le fait que  $X$ , discrète, prenne une valeur  $k$  est donc traduit par le fait que  $Y$ , continue, prenne une valeur dans  $]k - \frac{1}{2}, k + \frac{1}{2}[$ .

## 2.8.4 Exercices

1. Soit  $X$  suivant une loi binomiale  $B(50, 1/4)$ . Calculer exactement  $\mathbb{P}(11 \leq X < 13)$ . Donner une valeur approchée au moyen d'une loi normale.
2. Soit  $X$  une variable aléatoire suivant la loi  $N(12, 3)$ . Calculer  $\mathbb{P}(X = 12)$ ,  $\mathbb{P}(8 < X < 16)$ ,  $\mathbb{P}(7 \leq X \leq 10)$ . Trouver  $\alpha \in \mathbb{R}$  tel que  $\mathbb{P}(12 - \alpha < X < 12 + \alpha) \approx 0,7$ . Trouver  $\beta \in \mathbb{R}$  tel que  $\mathbb{P}(X < \beta) \approx 0,7$ .

## 2.9 Exercices récapitulatifs

### 2.9.1 Les énoncés

1. Sur 10000 individus d'une population, il y a une moyenne de 108 albinos.
  - a. Quelle est la probabilité qu'un individu choisi au hasard soit albinos ?
  - b. On choisit au hasard un échantillon de 100 individus et on désigne par  $X$  la variable aléatoire "nombre d'albinos dans l'échantillon". Quelle est la loi suivie par  $X$  (on note  $N$  l'effectif de la population considérée). Calculer  $E(X)$ ,  $\sigma^2(X)$  et  $\mathbb{P}(X \leq 1)$ . Par quelles lois peut-on approcher  $X$  ? Calculer une approximation de  $\mathbb{P}(X \leq 1)$  au moyen de chacune de ces lois.
  - c. Quelle est la probabilité que sur un échantillon de 1000 individus, il y ait 11 albinos (au plus 10 albinos, au plus 11 albinos) ?
2. On lance 2000 fois une pièce ordinaire et on considère la variable aléatoire  $X$  : "nombre de piles obtenues".
  - a. Quelle est la loi suivie par  $X$  ?
  - b. Calculer  $\mu = E(X)$  et  $\sigma = \sigma(X)$ .
  - c. Par quelle loi peut-on approcher  $X$  ?
  - d. Approcher  $\mathbb{P}(X = \mu)$ ,  $\mathbb{P}(|X - \mu| \leq 45)$  et  $\mathbb{P}(|X - \mu| \geq 100)$ .
3. Le sang humain contient 2 types de globules blancs, les polynucléaires et les lymphocytes. Pour un individu sain, il y a 65 % de polynucléaires. Un taux strictement supérieur à 70 % ou strictement inférieur à 50 % est pathologique. On fait des analyses sanguines qui portent sur un échantillon de 100 globules. Quelle est la probabilité qu'une étude sanguine réalisée sur un sujet sain fasse croire à un état pathologique ?

4. Un pépiniériste reçoit des graines en vrac dont 3% ne germeront pas. Il met les graines en sachets de 200 et garantit à sa clientèle 95 % de bonnes graines. Quelle est la probabilité qu'un sachet choisi au hasard ne respecte pas la garantie ?
5. On lance 12000 fois un dé ordinaire et on considère la variable aléatoire  $X$  : "nombre de six obtenus".
  - a. Quelle est la loi suivie par  $X$  ? Calculer  $\mu = E(X)$  et  $\sigma = \sigma(X)$ .
  - c. Par quelle loi peut-on approcher  $X$  ?
  - d. Approcher  $\mathbb{P}(X < 2000)$ ,  $\mathbb{P}(|X - \mu| \leq 45)$  et  $\mathbb{P}(|X - \mu| \geq 100)$ .
  - e. Trouver le plus petit intervalle fermé  $I$  centré sur  $\mu$  tel que  $\mathbb{P}(X \in I) \geq 0,9$ .
6. On considère un dé à six faces. Trois faces sont marquées d'un point, deux faces de deux points et une face de trois points. On considère l'expérience aléatoire consistant à lancer deux fois le dé dans les mêmes conditions et la variable aléatoire  $X$  : "somme des points obtenus aux deux lancés."  
Déterminer la loi de probabilité, l'espérance mathématique et la variance de  $X$ .
7. On considère une pièce équilibrée que l'on jette 40 fois dans les mêmes conditions. Soit  $X$  la variable aléatoire "nombre de piles obtenues".
  - a) Quelle est la distribution de probabilité de  $X$  ?
  - b) Par quelle(s) variable(s) aléatoire(s)  $X$  peut-elle être approchée ?
  - c) Calculer pour chacune de ces variables la probabilité d'obtenir exactement 20 piles.

## 2.9.2 Les solutions

### Exercice 1

- a. Puisque l'individu est choisi au hasard, on considère que tous les individus ont la même probabilité d'être choisis. La probabilité cherchée est donc  $p = \frac{108}{10000} = 0,0108$ .
- b. Puisqu'il s'agit d'un prélèvement, sans remise, de 100 individus parmi  $N$ , la variable  $X$  suit une loi hypergéométrique  $H(N; 100; 0,0108)$ . Son espérance est donnée par  $E(X) = 100 \cdot 0,0108 = 1,08$ , sa variance est donnée par  $\sigma^2(X) = 100 \cdot 0,0108 \cdot 0,9892 \cdot \frac{N-100}{N-1}$ .  
On a enfin, en appliquant directement la formule pour la loi hypergéométrique

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \frac{C_{0,0108N}^0 C_{0,9892N}^{100}}{C_N^{100}} + \frac{C_{0,0108N}^1 C_{0,9892N}^{99}}{C_N^{100}}.$$

Cette formule est lourde à calculer, et il faut supposer que  $0,9892N$  est entier supérieur à 100, faire une approximation d'une manière ou d'une autre, ou étendre la définition des combinaisons.

On peut supposer que la population étudiée a un effectif  $N$  supérieur à 1000. Dès lors  $\frac{n}{N} < 0,1$  et  $X \approx B(100; 0,0108)$ .

En utilisant cette approximation, on obtient

$$\begin{aligned} \mathbb{P}(X \leq 1) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) \\ &\approx C_{100}^0 (0,0108)^0 (0,9892)^{100} + C_{100}^1 (0,0108)^1 (0,9892)^{99} \\ &= 0,7062. \end{aligned}$$

D'autre part, comme  $n = 100 > 50$  et  $p = 0,0108 < 0,1$ , on peut approcher la loi  $B(100; 0,0108)$  par une loi de Poisson  $P(1,08)$ .

On utilise cette dernière approximation pour obtenir

$$\begin{aligned} \mathbb{P}(X \leq 1) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) \\ &\approx e^{-1,08} + 1,08 \cdot e^{-1,08} \\ &= 0,7063. \end{aligned}$$

- c. En suivant le même raisonnement et en supposant cette fois  $N > 10000$ , on arrive à approcher la loi de la variable aléatoire  $Y$  "nombre d'albinos dans l'échantillon" par une loi de Poisson  $P(10, 8)$ . On obtient alors

$$P(Y = 11) \approx \frac{(10, 8)^{11}}{11!} e^{-10,8} = 0,1191.$$

En utilisant la table de  $P(11)$ , ou en faisant le calcul explicite, on trouverait

$$\mathbb{P}(Y \leq 10) \approx 0,4599.$$

On peut aussi utiliser l'approximation normale de la binomiale, puisque  $1000 \cdot 0,0108 \cdot 0,9892 = 10,68 \dots > 9$ . On a donc  $Y \approx Y'$  si la variable  $Y'$  est telle que  $Y' \sim N(10,8; 3,2685)$ . Alors on peut calculer

$$\begin{aligned} \mathbb{P}(0 \leq Y \leq 10) &\approx \mathbb{P}(-0,5 \leq Y' \leq 10,5) = \pi(-0,091) - \pi(-3,457) \\ &= \pi(3,457) - \pi(0,091) = 0,4637. \end{aligned}$$

### Exercice 2

- a. Par définition  $X$  suit une loi  $B(2000; \frac{1}{2})$ .  
 b. On a  $E(X) = 2000 \cdot \frac{1}{2} = 1000$  et  $\sigma(X) = \sqrt{2000 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 10\sqrt{5} = 22,36$ .  
 c. On ne peut pas approcher  $X$  par une loi de Poisson car  $p = \frac{1}{2}$  est trop grand. Par contre, puisque  $np(1-p) = 500 > 9$ , on a une approximation normale par  $Y \sim N(1000; 10\sqrt{5})$ .  
 d. On a

$$\begin{aligned} \mathbb{P}(X = \mu) &= \mathbb{P}(X = 1000) \approx \mathbb{P}(999,5 < Y < 1000,5) \\ &= \pi(0,0223) - \pi(-0,0223) = 2 \cdot \pi(0,0223) = 0,018 \end{aligned}$$

De la même façon, on obtient

$$\begin{aligned} \mathbb{P}(|X - \mu| \leq 45) &= \mathbb{P}(955 \leq X \leq 1045) \\ &\approx \mathbb{P}(954,5 < Y < 1045,5) = 0,9581, \end{aligned}$$

et

$$\mathbb{P}(|X - \mu| \geq 100) = 1 - \mathbb{P}(|X - \mu| \leq 100) \approx 0,000118.$$

### Exercice 3

On tire au hasard 100 globules blancs parmi les  $N$  globules disponibles chez un individu sain. La variable aléatoire  $X$  "nombre de polynucléaires obtenus" suit alors une loi hypergéométrique  $H(N; 100; 0,65)$ . Puisque  $N$  est très grand par rapport à  $n = 100$ , on peut approcher la loi hypergéométrique par la loi Binomiale  $B(100; 0,65)$ . Enfin, puisque  $100 \cdot 0,65 \cdot 0,35 = 22,75 > 9$ , on a une approximation de cette dernière loi par une loi  $Y \sim N(65; 4,76)$ . Donc on a

$$\begin{aligned} \mathbb{P}(X < 50 \text{ ou } X > 70) &= 1 - \mathbb{P}(50 \leq X \leq 70) \\ &\approx 1 - \mathbb{P}(49,5 < Y < 70,5) \\ &= 1 - (\pi(1,155) - \pi(-3,256)) = 0,125 \end{aligned}$$

### Exercice 4

On considère l'expérience qui consiste à extraire au hasard un choix de 200 graines parmi les  $N$  graines du pépiniériste. La variable aléatoire  $X$  "nombre de mauvaises graines obtenues" suit alors une loi  $H(N; 200; 0,03)$ . Un sachet ne respecte pas la garantie si il contient plus de 10 mauvaises graines. Nous sommes donc amenés à calculer  $\mathbb{P}(X > 10) = 1 - \mathbb{P}(X \leq 10)$ . Puisqu'on peut supposer que le stock contient plus de 2000 graines, on peut approcher  $X$  par une variable suivant une loi  $B(200; 0,03)$ . Cette dernière peut à son tour être approchée par une loi de Poisson  $P(6)$ . On a alors

$$\mathbb{P}(X > 10) = 1 - 0,9574 = 0,0426.$$

## 2.10 Quelques théorèmes

Dans cette section, nous allons énoncer et démontrer l'inégalité de Bienaymé et Tchébychef, la loi faible des grands nombres et nous énoncerons le théorème central limite.

L'inégalité à l'avantage de pouvoir être démontrée facilement dans le cas des distributions discrètes. Elle reste valable dans le cas continu, mais la démonstration est plus technique. Elle permet de quantifier à quel point la distribution est concentrée autour de l'espérance, en termes de l'écart-type. Elle est valable pour toute distribution.

La loi des grands nombres se déduit de l'inégalité de Bienaymé et Tchébychef. Elle permet entre autres de faire un lien entre l'approche axiomatique et l'approche fréquentiste des probabilités. Enfin, elle explique pourquoi il vaut mieux ne pas jouer aux jeux d'argent.

Le théorème central limite est central dans la théorie des statistiques inférentielles, d'où son nom. C'est un théorème de convergence asymptotique, c'est-à-dire "pour une limite à l'infini". Il permet d'utiliser la loi de Gauss dans bien des situations.

### 2.10.1 L'inégalité de Bienaymé et Tchébychef

Nous avons constaté sur les exemples de la loi binomiale et de la loi normale que la probabilité qu'une variable aléatoire prenne une valeur éloignée de son espérance devient vite assez faible. Ce fait est énoncé précisément dans la proposition suivante.

**Proposition 2.10.1.** *Si  $X$  est une variable aléatoire sur un espace probabilisé  $(\Omega, E, \mathbb{P})$ , admettant une espérance  $\mu$  et un écart type  $\sigma > 0$ , alors on a*

$$\mathbb{P}(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2}, \quad \forall t > 0.$$

*Démonstration.* Nous donnons la démonstration dans le cas discret. On a alors  $\Omega_X = \{x_i \in I\}$ , avec  $I$  fini ou dénombrable. On pose alors  $J = \{i \in I : |x_i - \mu| \geq t\sigma\}$ . L'événement  $(|X - \mu| \geq t\sigma)$  est alors aussi  $(X \in \{x_i : i \in J\})$ , donc

$$\mathbb{P}(|X - \mu| \geq t\sigma) = \sum_{i \in J} p_i,$$

et enfin

$$\begin{aligned} \sigma^2 = \sigma^2(X) &= \sum_{i \in I} p_i (x_i - \mu)^2 \\ &\geq \sum_{i \in J} p_i (x_i - \mu)^2 \geq \sum_{i \in J} p_i t^2 \sigma^2 = t^2 \sigma^2 \mathbb{P}(|X - \mu| \geq t\sigma), \end{aligned}$$

et un peu d'algèbre suffit à conclure. □

*Remarque 2.2.* En passant à l'événement complémentaire, on obtient l'inégalité

$$\forall t > 0, \mathbb{P}(|X - \mu| < t\sigma) \geq 1 - \frac{1}{t^2}.$$

En fixant  $t = 3$ , on voit que la probabilité que  $X$  prenne une valeur entre  $\mu - 3\sigma$  et  $\mu + 3\sigma$  est supérieure à  $\frac{8}{9}$ .

A titre d'exercice, on peut considérer le problème suivant : le nombre de pièces sortant d'une usine donnée par semaine est une variable aléatoire d'espérance 50 et de variance 25. Estimer la probabilité que la production soit entre 40 et 60 pièces.

### 2.10.2 La loi faible des grands nombres

Ce théorème est une conséquence de l'inégalité de Bienaymé et Tchébychef. On considère une suite  $(X_1, X_2, \dots)$  de variables aléatoires *indépendantes* sur un même espace probabilisé  $(\Omega, E, \mathbb{P})$ . On suppose que ces variables suivent toutes une même loi d'espérance  $\mu$  et d'écart-type  $\sigma > 0$ . On définit la variable aléatoire "moyenne arithmétique"  $\bar{X}_n$  par

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n).$$

On a alors le résultat suivant.

**Proposition 2.10.2.** *Pour tout nombre  $\varepsilon$  strictement positif et tout entier positif  $n$ ,*

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

*En particulier, on a pour tout  $\varepsilon$  strictement positif*

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) = 0.$$

A titre d'exemple important, on peut considérer une expérience aléatoire et un événement  $A$ . On répète  $n$  fois l'expérience aléatoire et on note  $A_i$  l'événement "A s'est produit lors de la  $i$ -ème réalisation de l'expérience". On note  $X_i$  la variable indicatrice de  $A_i$ , qui vaut donc 1 si  $A_i$  se réalise. Alors  $X_1 + \dots + X_n$  est une variable aléatoire qui compte le nombre de fois que  $A$  se réalise sur les  $n$  expériences successives, tandis que  $\bar{X}_n$  est une variable aléatoire qui enregistre la fréquence de la réalisation de  $A$  sur ces  $n$  expériences. Enfin, dans ce cas, on sait que  $\mu = \mathbb{P}(A)$ . La loi des grands nombres donne donc une estimation, une borne supérieure pour la probabilité que la fréquence observée de  $A$  soit loin de  $\mathbb{P}(A)$ . Cette probabilité tend vers 0 quand  $n$  tend vers l'infini.

Dans le même ordre d'idée, au casino, les gains moyens enregistrés par un joueur tendent donc inexorablement vers l'espérance de gain, qui est négative...

### 2.10.3 Le théorème central limite

On considère les mêmes hypothèses que dans la section précédente :  $(X_1, X_2, \dots)$  est une suite de variables aléatoires *indépendantes* sur un même espace probabilisé  $(\Omega, E, \mathbb{P})$ , qui suivent toutes la même loi d'espérance  $\mu$  et un écart-type  $\sigma$ . La variable aléatoire  $\bar{X}_n$  admet donc également  $\mu$  pour espérance et  $\frac{\sigma}{\sqrt{n}}$  comme écart type. On note alors  $Z_n$  la variable centrée réduite correspondante, c'est à dire

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu).$$

On note enfin  $F_n$  la fonction de répartition de  $Z_n$ .

**Proposition 2.10.3.** *La suite de fonctions  $F_n$  converge vers la fonction  $\pi$  (la fonction de répartition de la normale  $N(0, 1)$ ) : on a*

$$\forall x \in \mathbb{R}, \lim_{n \rightarrow +\infty} F_n(x) = \pi(x).$$

*En particulier,*

$$\forall a < b \in \mathbb{R}, \lim_{n \rightarrow +\infty} \mathbb{P}(a < Z_n \leq b) = \pi(b) - \pi(a).$$

On peut conclure de ce résultat que la variable  $Z_n$  suit une loi  $N(0, 1)$  pour  $n$  assez grand. En pratique, il est d'usage de considérer que l'approximation est bonne quand  $n \geq 30$ . Enfin, comme nous l'avons vu dans le paragraphe concernant les lois normales, si  $Z_n$  suit une loi  $N(0, 1)$ , alors  $\bar{X}_n$  suit une loi  $N(m, \frac{\sigma}{\sqrt{n}})$ .

Ce résultat va s'avérer fondamental, que dis-je, central, en statistique inductive. En effet, dans ce cadre, on ne connaît pas la distribution commune des variables aléatoires indépendantes  $X_1, \dots, X_n$  et c'est précisément elle que l'on souhaite étudier. Le passage à  $\bar{X}_n$  pour  $n$  assez grand permet de régler ce problème, puisque la distribution est alors normale.



## Chapitre 3

# Statistique inductive : estimation

### 3.1 Introduction

Voici quelques questions que l'on se pose dans la pratique et auxquelles la statistique fournit une réponse. Les premières questions concernent les paramètres d'une distribution. Les voici telles que vous pourriez les rencontrer dans la littérature, ou dans les journaux.

- Quelle est la taille moyenne de la population belge ?
- Quelle est la dispersion des tailles de la population belge ?

On peut également vouloir connaître la distribution d'une caractéristique  $x$  associée à chaque individu de la population (fréquences ou fréquences cumulées, histogramme). On souhaite également pouvoir tester des hypothèses sur une ou plusieurs populations :

- Les Belges sont-ils en moyenne plus grands que les Allemands ?
- Le médicament A est-il plus efficace que le médicament B ?
- La taille est-elle plus variable en Belgique qu'en Allemagne ?

Ce sont des tests de comparaison que nous détaillerons dans les chapitres qui suivent.

En général, on ne dispose pas de la mesure du caractère que l'on souhaite étudier sur toute la population. En effet, il y a généralement des problèmes de coût ou de logistique pour organiser une mesure de tous les individus. C'est parfois simplement impossible : si une entreprise teste son produit pour déterminer son espérance de vie, elle ne peut évidemment pas tester tous les produits qui sont fabriqués.

Enfin, il n'est souvent pas possible d'avoir toute la population à disposition : si on teste les effets d'un médicament agissant sur une maladie, on se pose des questions sur la population des individus qui sont ou seront malades et prennent ou prendront le médicament. La population n'existe pas encore entièrement.

Tout le monde connaît la procédure à suivre : on extrait alors un **échantillon** de taille  $n$  de la population (ou les populations) que l'on souhaite étudier. Pour chaque individu choisi, on mesure la caractéristique que l'on considère. On obtient ainsi des observations  $(x_1, \dots, x_n)$  que l'on va utiliser de manière assez naturelle pour répondre aux questions que l'on se pose, du type de celles listées plus haut.

Le reproche qui est souvent fait à l'utilisation de la statistique est que les résultats peuvent varier d'un échantillon à l'autre, et qu'on ne peut donc rien conclure. Ce que nous allons faire ici, c'est précisément quantifier cette dépendance au choix d'un échantillon et tenter de déterminer s'il est raisonnable de généraliser les résultats obtenus sur les observations à toute la population.

Pour ce faire, nous allons d'abord formaliser un peu mieux les questions posées plus haut en termes de la **distribution de probabilité sous-jacente** à la population (voir la section 1.10).

## 3.2 La distribution sous-jacente et l'échantillon aléatoire simple

Nous avons vu dans l'exemple fondamental (voir la section 1.10) qu'il y a deux façons de voir une distribution de probabilité associée à une population. Je le rappelle ici.

- Si on a une population finie  $\Omega$  sur laquelle on peut mesurer une *variable statistique* (taille poids,...). On considère l'expérience aléatoire "choix d'un individu de manière aléatoire et équiprobable dans la population" et la variable aléatoire  $X$  : "valeur de  $x$  sur l'individu choisi", alors la distribution de probabilité de  $X$  est exactement la distribution des fréquences de  $x$  dans la population. Autrement dit, la distribution de probabilité (la loi) de  $X$  est exactement la distribution statistique de  $x$ . Cette loi  $\mathcal{L}$  est *en général inconnue*. Se poser des questions sur la distribution statistique de  $x$ , sa moyenne, son écart-type, revient à se poser des questions sur les paramètres de la distribution de probabilité  $\mathcal{L}$  de  $X$ , puisque ces paramètres coïncident.
- Nous avons également discuté un deuxième point de vue, qui est le point de vue de la statistique mathématique. Il prend comme point de départ que si la variable statistique  $x$  que l'on observe a une telle distribution, c'est parce que la valeur de  $x$  pour chaque individu est obtenue à partir d'un *processus aléatoire*, c'est une variable aléatoire, qui suit une distribution donnée  $\mathcal{L}$ , *inconnue*. Cette distribution est appelée distribution *sous-jacente* à la population.

Nous adopterons ce deuxième point de vue, mais le premier mène aux mêmes considérations. Les questions que nous nous posons sont donc des questions sur la distribution  $\mathcal{L}$  sous-jacente, dont nous en connaissons rien a priori.

Formulons les questions énoncées plus haut en fonction de cette distribution sous-jacente. Les questions

- Quelle est la taille moyenne de la population belge ?
- Quelle est la dispersion des tailles de la population belge ?

sont formulées correctement comme suit

- Quelle est l'espérance de la distribution  $\mathcal{L}_{TB}$  de la variable aléatoire "taille d'un belge".
- Quelle est la variance de  $\mathcal{L}_{TB}$ .

Les questions

- Les Belges sont-ils en moyenne plus grands que les Allemands ?
- La taille est-elle plus variable en Belgique qu'en Allemagne ?

sont reformulées en

- L'espérance de  $\mathcal{L}_{TB}$  est elle supérieure à l'espérance de  $\mathcal{L}_{TA}$  ?
- La variance de  $\mathcal{L}_{TB}$  est elle supérieure à la variance de  $\mathcal{L}_{TA}$  ?

Dans le cas de la question sur le médicament, on fait de même, mais on doit encore définir à partir de quelle mesure on décide son effet.

Quel que soit le point de vue choisi ci-dessus, l'expérience aléatoire consistant à choisir un individu dans la population et à considérer "valeur de la variable  $x$  sur cet individu" détermine une variable aléatoire  $X$  admettant la distribution  $\mathcal{L}$ . On répète  $n$  fois cette expérience, dans les mêmes conditions et de manière indépendante. *On définit ainsi  $n$  variables aléatoires  $X_1, \dots, X_n$ , indépendantes et de même distribution  $\mathcal{L}$ .*

**Définition 3.2.1.** Le vecteur ou  $n$ -uplet  $(X_1, \dots, X_n)$  ainsi défini est appelé *échantillon aléatoire simple*.

Il faut bien voir que  $(X_1, \dots, X_n)$  est un *vecteur de variables aléatoires*, donc de fonctions mathématiques. Il ne s'agit donc pas d'un vecteur de données.

La juxtaposition des  $n$  expériences indépendantes peut être considérée comme une unique expérience aléatoire "choix d'un échantillon de taille  $n$ " (avec remise).

Une *réalisation* de l'expérience "choix d'un échantillon" fournit alors un  $n$ -uplet d'individus  $(\omega_1, \dots, \omega_n)$ . Le vecteur aléatoire  $(X_1, \dots, X_n)$  prend sur ce  $n$ -uplet la valeur  $(x_1, \dots, x_n)$ , qui est le vecteur de *données*.

Le vecteur  $(x_1, \dots, x_n)$  est encore appelé *réalisation de l'échantillon aléatoire*.

### Conséquences pour la réalisation pratique :

- On fixe l'effectif de l'échantillon à l'avance.
- Les tirages doivent se faire au hasard, de manière équiprobable. Si on a la possibilité de numéroter les individus, on peut recourir à des tables de nombres aléatoires ou à un générateur de nombres aléatoires.
- Les tirages successifs se font dans une même population et de manière indépendante. Cela implique notamment que le tirage se fait avec remise, un individu pouvant alors être tiré plusieurs fois.

## 3.3 Estimation ponctuelle

Il s'agit ici d'estimer un paramètre de la distribution  $\mathcal{L}$  sous-jacente à la population. On s'intéressera plus particulièrement aux espérances et variances, ou aux paramètres d'une variable de Bernoulli (la probabilité de succès  $\pi_0$ ) ou d'une binomiale  $B(n', \pi_0)$ . On notera  $\mu_0, \sigma_0^2$  ou  $\pi_0$  ces paramètres, qui sont *inconnus*.

### 3.3.1 Estimation ponctuelle d'une espérance

On souhaite estimer l'espérance, le centre de gravité de la loi  $\mathcal{L}$  sous-jacente. Vu la construction de l'espérance, il semble naturel d'estimer  $\mu_0$  par la moyenne arithmétique des données sur l'échantillon prélevé.

Ce nombre dépend évidemment du résultat de l'expérience aléatoire "choix de l'échantillon" et est donc la *réalisation d'une variable aléatoire* fabriquée avec  $X_1, \dots, X_n$ . Nous l'avons déjà rencontrée : elle s'écrit effectivement

$$\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n).$$

Analysons maintenant les qualités de  $\overline{X}_n$  :

1. On a

$$E(\overline{X}_n) = \frac{1}{n}(E(X_1) + \dots + E(X_n)) = \frac{1}{n}(\mu_0 + \dots + \mu_0) = \mu_0.$$

L'espérance que nous cherchons à estimer  $\mu_0$  est donc le centre de gravité (l'espérance) de la distribution de la variable aléatoire  $\overline{X}_n$ .

2. On a, vu l'indépendance des variables  $X_1, \dots, X_n$ ,

$$\sigma^2(\overline{X}_n) = \frac{1}{n^2}(\sigma^2(X_1) + \dots + \sigma^2(X_n)) = \frac{\sigma_0^2}{n}$$

si  $\sigma_0^2$  est la variance de  $\mathcal{L}$ . Donc on a

$$\sigma(\overline{X}_n) = \frac{\sigma_0}{\sqrt{n}} \quad \text{et} \quad \lim_{n \rightarrow \infty} \sigma(\overline{X}) = 0.$$

Ces égalités ont une interprétation simple : la dispersion des observations de  $\overline{X}_n$  peut être rendue aussi petite que l'on veut, pour autant que la taille de l'échantillon choisi soit suffisamment grande. En d'autres termes la précision de l'estimation peut être aussi grande que l'on veut pour autant que  $n$  soit suffisamment grand.

3. Cette précision peut être visualisée à l'aide de l'ingalité de Bienaymé et Tchébychev. Remarquons toutefois que pour diminuer de moitié l'écart type de  $\overline{X}_n$ , il faut multiplier par 4 la taille de l'échantillon, cela a donc une répercussion sur le coût de l'expérience.

Ces propriétés sont caractéristiques de bons estimateurs. Elles motivent la définition suivante.

**Définition 3.3.1.** Soit  $\theta_0$  un paramètre de la distribution  $\mathcal{L}$  sous-jacente que l'on veut étudier.

- Un estimateur *sans biais (ou fidèle)* de  $\theta_0$  est une variable aléatoire  $\Theta$  (dépendant de  $(X_1, \dots, X_n)$ ) telle que  $E(\Theta) = \theta_0$ .
- Si de plus, on a

$$\lim_{n \rightarrow \infty} \sigma(\Theta) = 0,$$

on dit que  $\Theta$  est un estimateur *absolument convergent* de  $\theta_0$ .

Lorsqu'on dispose d'un tel estimateur  $\Theta$ , on réalise l'expérience "choix d'un échantillon" pour obtenir un  $n$ -uplet d'individus  $(\omega_1, \dots, \omega_n)$  puis une réalisation  $(x_1, \dots, x_n)$  de l'échantillon aléatoire. On calcule la valeur de  $\Theta$  sur ces données et on obtient une bonne estimation de  $\theta_0$  (d'autant meilleure que  $n$  est grand).

On a montré plus haut le résultat suivant.

**Proposition 3.3.1.** La variable aléatoire  $\overline{X} = \frac{1}{n}(X_1 + \dots + X_n)$  est un estimateur absolument convergent de l'espérance  $\mu_0$  de  $\mathcal{L}$ .

### 3.3.2 Estimation ponctuelle d'une proportion/probabilité

On souhaite estimer la proportion  $\pi_0$  des individus d'une population présentant un caractère, une propriété  $P$ , par exemple, le nombre de fumeurs dans la population. Si on exprime le problème en termes de distribution sous-jacente, cela revient à estimer l'espérance de la variable de Bernoulli  $1_P$  dans la population, qui est la variable aléatoire définie par

$$1_P : \Omega \rightarrow \{0, 1\} : \omega \mapsto \begin{cases} 1 & \text{si } \omega \text{ présente le caractère } P \\ 0 & \text{si } \omega \text{ ne présente pas le caractère } P \end{cases}$$

On se ramène alors au cas précédent, et il suffit de prendre la variable aléatoire  $\overline{X}_n$  comme estimateur. On obtient alors le résultat suivant, en interprétant  $\overline{X}_n$ .

**Proposition 3.3.2.** La variable aléatoire "proportion d'individus de l'échantillon présentant  $P$ " est un estimateur absolument convergent du paramètre  $\pi_0$ .

Pour estimer la proportion d'individus présentant le caractère  $A$ , on calcule donc la proportion d'individus présentant  $A$  dans l'échantillon choisi.

### 3.3.3 Estimation ponctuelle d'une variance

On souhaite estimer ici la variance  $\sigma_0^2$  de la loi  $\mathcal{L}$ . Il est encore naturel de considérer la variable aléatoire  $\Sigma^2$  : "Variance mesurée sur l'échantillon choisi", c'est à dire

$$\Sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2.$$

Malheureusement, cet estimateur est biaisé, on a en effet

$$E(\Sigma^2) = \frac{n-1}{n} \sigma_0^2.$$

On considère alors la variable aléatoire  $S^2 = \frac{n}{n-1}\Sigma^2$  pour obtenir un estimateur sans biais. On a en effet le résultat suivant.

**Proposition 3.3.3.** *La variable aléatoire  $S^2 = \frac{n}{n-1}\Sigma^2$  est un estimateur absolument convergent de la variance  $\sigma_0^2$ .*

En conséquence, si on veut estimer la variance d'une population, on réalise l'expérience "choix d'un échantillon" qui conduit au vecteur  $(x_1, \dots, x_n)$  et on calcule la valeur

$$s^2 = \frac{n}{n-1}\Sigma^2(x_1, \dots, x_n) = \frac{1}{n-1}\left(\sum_{i=1}^n (x_i - \bar{x})^2\right) = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}^2\right). \quad (3.1)$$

La façon dont l'estimateur  $S^2$  a été construit ainsi que ses propriétés nous permettent de penser que la valeur calculée est une bonne estimation, même si, malheureusement, cette valeur calculée peut être loin de la vraie valeur. Terminons par une remarque élémentaire, mais qui peut surprendre au premier abord.

*Remarque 3.1.* La variable  $S = \sqrt{S^2}$  n'est pas un estimateur sans biais de l'écart-type de la distribution  $\mathcal{L}$ .

## 3.4 Estimation par intervalle de confiance

Dans la section précédente, on a appris à estimer un paramètre  $\theta_0$  d'une population/distribution à partir d'un échantillon. Le résultat de cette estimation est un nombre, dont on a de bonnes raisons de penser qu'il est proche du paramètre inconnu, puisqu'il est extrait d'une distribution centrée sur ce paramètre, et dont la variance est petite. C'est parfois insuffisant : prenons par exemple le cas d'un sondage sur des intentions de votes entre deux candidats. Si le sondage donne le premier candidat vainqueur avec 52% des votes (il s'agit d'une estimation ponctuelle d'un paramètre d'une distribution de Bernoulli), comment peut-on être (presque) sûr que la vraie valeur du paramètre de la distribution n'est pas 48% ?

Il faut donc pouvoir maîtriser plus finement la variabilité due au choix de l'échantillon. L'idée est alors de construire une fourchette de valeurs, un intervalle, qui dépend de l'expérience "choix d'un échantillon" et est donc un **intervalle aléatoire** et qui a une probabilité  $1 - \alpha$  (proche de 1 : on choisit généralement  $\alpha = 0,05$  ou  $0,01$ ) de *recouvrir* la vraie valeur de  $\theta_0$ . La valeur  $\alpha$  représente alors le *risque* que l'intervalle construit ne contienne pas la vraie valeur à estimer. La **réalisation** de cet intervalle aléatoire à partir d'une réalisation de l'échantillon aléatoire est alors appelé **Intervalle de confiance pour  $\theta_0$  au seuil de confiance  $1 - \alpha$** .

On souhaite que cet intervalle soit le plus petit possible.

### 3.4.1 Intervalle de confiance pour une espérance

Nous considérons encore une distribution  $\mathcal{L}$  dont nous souhaitons estimer l'espérance  $\mu_0$ . Nous notons  $\sigma_0^2$  la variance de  $\mathcal{L}$ . D'après la section sur l'estimation ponctuelle, nous savons que la variable aléatoire  $\bar{X}_n$  est un estimateur absolument convergent de l'espérance  $\mu_0$ . Il y a également deux cas où l'on en sait plus :

- Si  $n$  est grand, alors  $\frac{\bar{X}_n - \mu_0}{\frac{\sigma_0}{\sqrt{n}}}$  suit une loi normale  $N(0, 1)$ , par le théorème central limite. L'approximation est considérée comme bonne si  $n \geq 30$ .
- Si  $\mathcal{L}$  est normale, alors  $\frac{\bar{X}_n - \mu_0}{\frac{\sigma_0}{\sqrt{n}}}$  suit une loi  $N(0, 1)$ , pour tout  $n$ .

Si on est dans un de ces cas, il est possible de construire un intervalle de confiance pour  $\mu_0$ .

Il est naturel de centrer l'intervalle sur la valeur que nous avons estimée ponctuellement. On cherche alors un intervalle aléatoire de la forme  $]\bar{X} - A, \bar{X} + A[$  où  $A$  est un nombre positif (éventuellement aléatoire) et le plus petit possible tel que

$$\mathbb{P}(\bar{X} - A < \mu_0 < \bar{X} + A) = 1 - \alpha.$$

On a alors

$$\begin{aligned} \mathbb{P}(\bar{X} - A < \mu_0 < \bar{X} + A) &= \mathbb{P}(\mu_0 - A < \bar{X} < \mu_0 + A) \\ &= \mathbb{P}(-A < \bar{X} - \mu_0 < A) \\ &= \mathbb{P}\left(-A \frac{\sqrt{n}}{\sigma_0} < \frac{\bar{X} - \mu_0}{\frac{\sigma_0}{\sqrt{n}}} < A \frac{\sqrt{n}}{\sigma_0}\right) \end{aligned}$$

La variable aléatoire intervenant dans la dernière équation suit une loi normale  $N(0, 1)$ . On utilise alors les quantiles de la loi normale pour obtenir la valeur de  $A$ , qui n'est dans ce cas-ci pas aléatoire.

**Proposition 3.4.1.** *Pour tout nombre  $\beta \in ]0, 1[$ , il existe un nombre  $u_\beta \in \mathbb{R}$  tel que, si  $X \sim N(0, 1)$ ,  $\mathbb{P}(X < u_\beta) = \beta$ .*

Si on souhaite avoir

$$\mathbb{P}\left(-A \frac{\sqrt{n}}{\sigma_0} < \frac{\bar{X} - \mu_0}{\frac{\sigma_0}{\sqrt{n}}} < A \frac{\sqrt{n}}{\sigma_0}\right) = 1 - \alpha,$$

il suffit de prendre

$$-A \frac{\sqrt{n}}{\sigma_0} = u_{\frac{\alpha}{2}}$$

et

$$A \frac{\sqrt{n}}{\sigma_0} = u_{1-\frac{\alpha}{2}},$$

ce qui donne

$$A = u_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}.$$

On a ainsi obtenu le résultat suivant.

**Proposition 3.4.2** (Grands échantillons,  $\sigma_0$  connu). *Si l'échantillon extrait de la population est grand ( $n \geq 30$ ) et si on note  $\bar{x}$  la moyenne arithmétique sur l'échantillon choisi (la réalisation de  $\bar{X}_n$ ), alors l'intervalle*

$$\left] \bar{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, \bar{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \right[$$

*est un intervalle de confiance pour  $\mu_0$  au seuil  $1 - \alpha$ .*

Ce résultat est important. Cependant en général, on ne connaît pas  $\sigma_0$  au moment d'estimer l'espérance  $\mu_0$ . Si l'échantillon est suffisamment grand ( $n \geq 30$ ), on peut sans problème le remplacer par un estimateur. On obtient alors le résultat suivant.

**Proposition 3.4.3** (Grands échantillons,  $\sigma_0$  inconnu). *Si l'échantillon extrait de la population est grand ( $n \geq 30$ ) et si on note  $\bar{x}$  la moyenne arithmétique sur l'échantillon choisi (la réalisation de  $\bar{X}_n$ ), et  $s^2$  la réalisation de  $S^2$  sur l'échantillon choisi, alors l'intervalle*

$$\left] \bar{x} - u_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + u_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right[$$

*est un intervalle de confiance pour  $\mu_0$  au seuil  $1 - \alpha$ .*

### 3.4.2 Le cas des petits échantillons

Dans le cas où l'échantillon extrait est de taille inférieure à 30, on ne peut plus utiliser le théorème central limite et, si  $\sigma_0$  n'est pas connu, il n'est plus possible de le remplacer par  $s$  sans introduire une erreur importante. On traite ces deux problèmes séparément. On doit bien sûr compenser cette petite taille de l'échantillon par des hypothèses qui réduisent le champ d'application des constructions qui sont faites.

Si  $\sigma_0$  est connu et si la distribution  $\mathcal{L}$  est normale, alors on peut procéder comme dans la section précédente puisque la variable aléatoire  $\frac{\bar{X}-\mu_0}{\frac{\sigma_0}{\sqrt{n}}}$  suit une loi  $N(0, 1)$ . On obtient donc le résultat suivant.

**Proposition 3.4.4** (Petits échantillons,  $\sigma_0$  connu). *Si  $\mathcal{L}$  est une distribution normale et si on note  $\bar{x}$  la moyenne arithmétique sur l'échantillon choisi, alors l'intervalle*

$$\left] \bar{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}}, \bar{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma_0}{\sqrt{n}} \right[$$

*est un intervalle de confiance pour  $\mu_0$  au seuil  $1 - \alpha$ .*

Si l'échantillon est petit et si on ne connaît pas  $\sigma_0$ , on ne peut plus utiliser la loi normale. Puisqu'on n'a pas  $\sigma_0^2$ , il faut utiliser son estimateur  $S^2$ . On est donc amené à considérer la variable aléatoire  $\frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$ . On a sur cette variable le résultat suivant.

**Proposition 3.4.5.** *Si la distribution sous-jacente à la population est une loi  $N(\mu_0, \sigma_0)$ , alors la variable aléatoire*

$$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

*suit une loi de Student à  $n - 1$  degrés de libertés ( $T \sim t_{n-1}$ ).*

La distribution de Student à  $n$  degrés de libertés jouit des propriétés suivantes

**Proposition 3.4.6.** *Si  $X \sim t_n$  alors  $E(X) = 0$ ,  $\sigma^2(X) = \frac{n}{n-2}$  (si  $n > 2$ ).*

*La densité de probabilité de  $X$  est représentée par une courbe en cloche, symétrique par rapport à la droite d'équation  $x = 0$ , et plus aplatie que la normale. Enfin, la distribution  $t_n$  peut être approchée par une normale  $N(0, 1)$  quand  $n$  tend vers l'infini. En pratique, on considère que l'approximation est valable si  $n \geq 30$ .*

On suit alors les mêmes développements théoriques que dans la section précédente.

**Proposition 3.4.7.** *Si  $X \sim t_n$  et si  $\beta \in ]0, 1[$ , il existe un nombre unique  $t_{n,\beta}$  tel que  $\mathbb{P}(X < t_{n,\beta}) = \beta$ .*

On a alors

$$\begin{aligned} \mathbb{P}(\bar{X} - A < \mu_0 < \bar{X} + A) &= \mathbb{P}(\mu_0 - A < \bar{X} < \mu_0 + A) \\ &= \mathbb{P}(-A < \bar{X} - \mu_0 < A) \\ &= \mathbb{P}\left(-A \frac{\sqrt{n}}{S} < \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} < A \frac{\sqrt{n}}{S}\right). \end{aligned}$$

La variable aléatoire intervenant dans la dernière équation suit une loi  $t_{n-1}$ . Si on souhaite avoir

$$\mathbb{P}\left(-A \frac{\sqrt{n}}{S} < \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} < A \frac{\sqrt{n}}{S}\right) = 1 - \alpha,$$

il suffit de prendre  $A$  tel que

$$-A \frac{\sqrt{n}}{S} = t_{n-1, \frac{\alpha}{2}}$$

et

$$A \frac{\sqrt{n}}{S} = t_{n-1, 1-\frac{\alpha}{2}},$$

ce qui donne

$$A = t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}.$$

On a ainsi obtenu le résultat suivant.

**Proposition 3.4.8** (Petits échantillons,  $\sigma_0$  inconnu). *Si la variable  $x$  considérée suit une loi normale et si on note  $\bar{x}$  la moyenne arithmétique sur l'échantillon choisi, alors l'intervalle*

$$\left] \bar{x} - t_{n-1, 1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1, 1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right[$$

*est un intervalle de confiance pour  $\mu_0$  au seuil  $1 - \alpha$ .*

**Exercice :**

Dans un lot de 500 cocons de papillons, on prélève au hasard et avec remise 20 cocons que l'on pèse. Les poids obtenus sont donnés dans le tableau suivant :

0,64	0,65	0,73	0,60	0,65
0,77	0,82	0,64	0,66	0,72
0,87	0,84	0,66	0,76	0,63
0,52	0,66	0,45	0,74	0,79

Donner une estimation de l'espérance de la distribution des poids des 500 cocons.  
 Donner une estimation de la variance de la distribution des poids des 500 cocons.  
 Si on suppose que le poids est distribué de manière normale dans la population, donner un intervalle de confiance à 95% pour l'espérance de la distribution des poids des 500 cocons.

### 3.4.3 Intervalle de confiance pour une proportion/probabilité $\pi_0$

On se propose de déterminer un intervalle de confiance pour la proportion  $\pi_0$  d'individus d'une populations possédant une propriété  $P$ . C'est l'estimation que l'on fait pour les sondages d'opinion. Un autre exemple de question classique : estimer la proportion de fumeurs parmi les individus de 15 à 25 ans. La question peut-être reformulée comme ceci : estimer la probabilité qu'un individu choisi au hasard dans cette population soit fumeur.

Comme nous l'avons vu dans le cadre de l'estimation ponctuelle, le problème doit être posé en termes de distribution sous-jacente. Il s'agit d'estimer l'espérance de la variable de Bernoulli  $1_P$ , celle qui associe 1 aux individus possédant la propriété et 0 aux autres. La situation est ici encore différente selon que l'échantillon est suffisamment grand (en pratique  $n \geq 30$ ) ou pas. Nous ne traiterons que le cas des grands échantillons. Même dans ce cas, il faut faire une certaine approximation. Le résultat clef est une conséquence directe du théorème central limite.

On considère les variables aléatoires de Bernoulli  $X_i$  définies par

$$X_i = \begin{cases} 1 & \text{si l'individu } i \text{ a la propriété } P \\ 0 & \text{sinon} \end{cases}$$

On considère encore la variable aléatoire  $\mathcal{P} = \frac{1}{n}(X_1 + \dots + X_n)$ . C'est la variable "proportion d'individus de l'échantillon possédant la propriété  $P$ ".



**Proposition 3.4.9.** *Si  $n > 30$ , alors la variable aléatoire*

$$\frac{\mathcal{P} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

*suit une loi normale  $N(0, 1)$ .*

Nous savons également que  $\mathcal{P}$  est un estimateur absolument convergent de  $\pi_0$ . Il paraît donc naturel de centrer l'intervalle aléatoire sur  $\mathcal{P}$ . Nous considérons alors un intervalle aléatoire de la forme  $]\mathcal{P} - A, \mathcal{P} + A[$  et nous avons

$$\begin{aligned} \mathbb{P}(\mathcal{P} - A < \pi_0 < \mathcal{P} + A) &= \mathbb{P}(\pi_0 - A < \mathcal{P} < \pi_0 + A) \\ &= \mathbb{P}(-A < \mathcal{P} - \pi_0 < A) \\ &= \mathbb{P}\left(\frac{-A}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} < \frac{\mathcal{P} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} < \frac{A}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}\right). \end{aligned}$$

Si on souhaite avoir

$$\mathbb{P}\left(\frac{-A}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} < \frac{\mathcal{P} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} < \frac{A}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}\right) = 1 - \alpha,$$

il suffit de choisir  $A$  tel que

$$\frac{-A}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = u_{\frac{\alpha}{2}}$$

et

$$\frac{A}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = u_{1-\frac{\alpha}{2}}.$$

Ces deux équations sont équivalentes et conduisent à la construction de l'intervalle aléatoire suivant :

$$\left] \mathcal{P} - u_{1-\frac{\alpha}{2}} \sqrt{\frac{\pi_0(1-\pi_0)}{n}}, \mathcal{P} + u_{1-\frac{\alpha}{2}} \sqrt{\frac{\pi_0(1-\pi_0)}{n}} \right[.$$

Cet intervalle pose encore problème, puisqu'on ne connaît pas  $\pi_0$  : c'est le paramètre que l'on veut estimer. Deux possibilités s'offrent à nous : soit on remplace  $\pi_0$  par son estimation ponctuelle  $p$  obtenue à partir de l'échantillon. Cela ne pose pas de problème car, d'une part l'estimation peut être considérée comme bonne, puisque l'échantillon est grand, et d'autre part  $\pi_0$  et  $p$  sont des nombres compris entre 0 et 1. La différence entre  $\sqrt{\frac{\pi_0(1-\pi_0)}{n}}$  et la valeur estimée  $\sqrt{\frac{p(1-p)}{n}}$  est alors très petite, quand  $n$  est grand. On obtient alors le résultat suivant.

**Proposition 3.4.10.** *Si l'échantillon extrait de la population est de taille suffisante ( $n > 30$ ) et si  $p$  est la proportion d'individus de l'échantillon possédant la propriété  $P$ , alors l'intervalle*

$$\left] p - u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}, p + u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right[$$

*est un intervalle de confiance pour  $\pi_0$  au seuil  $1 - \alpha$ .*

Cependant, au sens strict, si on remplace  $\pi_0$  par sa valeur estimée  $p$ , on pourrait être amené à définir un intervalle un peu trop petit. On peut prendre la décision de prendre la valeur la plus grande possible pour le nombre  $\sqrt{\pi_0(1-\pi_0)}$ , c'est à dire  $\frac{1}{2}$ . On est alors sûr d'avoir pris un intervalle assez grand. C'est le résultat suivant.

**Proposition 3.4.11.** *Si l'échantillon extrait de la population est de taille suffisante ( $n \geq 30$ ) et si  $p$  est la proportion d'individus de l'échantillon possédant la propriété  $P$ , alors l'intervalle*

$$\left] p - u_{1-\frac{\alpha}{2}} \frac{1}{2\sqrt{n}}, p + u_{1-\frac{\alpha}{2}} \frac{1}{2\sqrt{n}} \right[$$

*est un intervalle de confiance pour  $\pi_0$  au seuil  $1 - \alpha$ .*

Le cas des petits échantillons requiert d'utiliser la distribution exacte de  $\mathcal{P}$ , et ne nous permet pas d'utiliser les approximations pour les variances et écarts-types. Il est plus technique et nous ne l'aborderons pas ici.

### 3.4.4 Intervalle de confiance pour une variance

Nous souhaitons donner un intervalle de confiance pour la variance, notée  $\sigma_0^2$ , (et pour l'écart-type) d'une variable aléatoire  $X$ . Nous savons que la variable aléatoire  $S^2$  est un estimateur absolument convergent de  $\sigma_0^2$ . Il est donc raisonnable de construire un intervalle de confiance au moyen de cet estimateur.

Nous aurons également besoin d'une nouvelle distribution théorique, appelée distribution de Pearson ou distribution de chi-carré à  $\nu$  degrés de libertés. Si une variable aléatoire  $X$  suit une telle distribution, on note  $X \sim \chi_\nu^2$ . La proposition suivante décrit les propriétés d'une variable aléatoire ayant cette distribution.

**Proposition 3.4.12.** *Si  $X \sim \chi_\nu^2$ , alors  $E(X) = \nu$  et  $\sigma^2(X) = 2\nu$ .*

*Les densités de probabilité de  $\chi_1^2$  et  $\chi_2^2$  sont décroissantes sur  $]0, +\infty[$ . Pour  $\nu \geq 3$ , les densités de  $\chi_\nu^2$  admettent un maximum et sont plus symétriques quand  $\nu$  augmente. Enfin, quand  $\nu \geq 30$ , on a  $\chi_\nu^2 \approx N(\nu, 2\nu)$ .*

La figure 3.4.4 montre les représentations graphiques de quelques densités de probabilités de variables  $\chi_\nu^2$ .

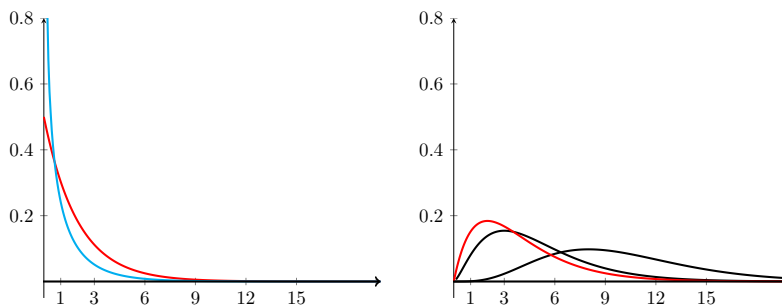


FIGURE 3.1 – Les distributions Chi-carré de paramètres 1,2,4, 5 et 10.

Les théorèmes permettant d'obtenir un intervalle de confiance pour une variance sont alors les suivants. Le premier précise l'existence de quantile.

**Proposition 3.4.13.** *Pour tout  $\nu$  naturel positif et tout nombre  $\beta \in ]0, 1[$ , il existe un unique nombre, noté  $\chi_{\nu,\beta}^2$  tel que si  $X \sim \chi_\nu^2$ , alors  $\mathbb{P}(X < \chi_{\nu,\beta}^2) = \beta$*

Le suivant donne la distribution théorique de  $S^2$ .

**Proposition 3.4.14.** *Si la distribution sous-jacente à la population est gaussienne (normale) de variance  $\sigma_0^2$ , alors la variable aléatoire*

$$\frac{n-1}{\sigma_0^2} S^2,$$

définie à partir d'un échantillon aléatoire de taille  $n$ , suit une loi de chi carré à  $n - 1$  degrés de liberté.

Nous sommes maintenant en mesure de construire un intervalle de confiance au seuil  $1 - \alpha$  pour la variance  $\sigma_0^2$ . On a en effet, pour  $A_1$  et  $A_2$  positifs

$$\begin{aligned} \mathbb{P}(A_1 < \sigma_0^2 < A_2) &= \mathbb{P}\left(\frac{1}{A_2} < \frac{1}{\sigma_0^2} < \frac{1}{A_1}\right) \\ &= \mathbb{P}\left(\frac{(n-1)S^2}{A_2} < \frac{(n-1)S^2}{\sigma_0^2} < \frac{(n-1)S^2}{A_1}\right) \end{aligned}$$

Si on souhaite que cette probabilité soit égale à  $1 - \alpha$ , il suffit de satisfaire

$$\frac{(n-1)S^2}{A_2} = \chi_{n-1, \frac{\alpha}{2}}^2$$

et

$$\frac{(n-1)S^2}{A_1} = \chi_{n-1, 1-\frac{\alpha}{2}}^2.$$

Nous obtenons donc le résultat final suivant.

**Proposition 3.4.15.** *Si la distribution sous-jacente à la population est normale et si l'échantillon considéré est de taille  $n$ , alors l'intervalle*

$$\left] \frac{n-1}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} s^2, \frac{n-1}{\chi_{n-1, \frac{\alpha}{2}}^2} s^2 \right[$$

est un intervalle de confiance au seuil  $1 - \alpha$  pour  $\sigma_0^2$ . De même, l'intervalle

$$\left] \sqrt{\frac{n-1}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}} s, \sqrt{\frac{n-1}{\chi_{n-1, \frac{\alpha}{2}}^2}} s \right[$$

est un intervalle de confiance au seuil  $1 - \alpha$  pour  $\sigma_0$ .

### 3.4.5 Exercices

- On choisit au hasard 12 nourrissons dans une population  $\mathcal{P}$  suivant un certain régime durant une période donnée. On observe les gains de poids suivants :

70 118 101 85 107 132 94 105 98 107 89 114.

- Donner une estimation ponctuelle de l'espérance du gain de poids et de la variance des gains de poids dans la population  $\mathcal{P}$ .
  - Donner un intervalle de confiance au seuil de 95% pour l'espérance du gain de poids dans la population  $\mathcal{P}$ .
- La gendarmerie a contrôlé la vitesse de 15 automobilistes pris au hasard sur une section d'autoroute. Voici les vitesses enregistrées (en km/h) :

112	125	105	132	130	95	136	125
145	128	110	136	127	115	142	

- Donner une estimation ponctuelle l'espérance de la distribution des vitesses des conducteurs, et de la variance de la distribution des vitesses, sur cette section d'autoroute.

- b. Donner un intervalle de confiance au seuil de 95% l'espérance de la distribution des vitesses sur cette section d'autoroute.
3. Donner une estimation ponctuelle et une estimation par intervalle de confiance au seuil 0,95 pour la variance d'une population normale, sachant que l'étude d'un échantillon de taille 30 a fourni une variance de l'échantillon égale à 2,9.
4. On a mesuré le taux de glucose dans le sang de 11 individus. Les résultats sont les suivants :  
si on note  $x_i$  le taux de glucose pour l'individu  $i$ , on a  $\sum x_i = 10,45$  et  $\sum x_i^2 = 13,5$ .  
Déterminer un intervalle de confiance au seuil 0,95 pour l'espérance du taux de glucose  $\mu_0$  de la population dont l'échantillon a été extrait. Quelle hypothèse de travail faut-il faire ?
5. Dans une clinique, on a examiné un échantillon de 500 bébés et observé que 15 d'entre eux présentaient une luxation congénitale de la hanche. Donner un intervalle de confiance au seuil 0,95 pour la proportion  $\pi_0$  de nouveau-nés présentant cette pathologie, dans la population dont l'échantillon est extrait.

# Chapitre 4

## Statistique inductive : les tests

### 4.1 Introduction

Dans ce chapitre, nous allons étudier quelques tests statistiques, encore appelés tests d'hypothèses. Nous commencerons par des exemples concrets, qui permettront de mettre en place quelques concepts et notations et qui mettront en évidence le principe des tests. Ensuite nous construirons les tests classiques couramment utilisés.

#### 4.1.1 Pouvoir surnaturel ou pas ?

Nous allons considérer les deux exemples suivants.

##### **Exemple 1**

*Une personne prétend pouvoir prédire l'avenir. On teste ce pouvoir de la manière suivante. On demande à la personne de prédire le résultat de dix lancers d'une pièce équilibrée. Elle donne 7 réponses correctes. Peut-on en conclure qu'elle possède le don de prédire l'avenir ?*

##### **Exemple 2**

*Une personne prétend pouvoir prédire l'avenir. On teste ce pouvoir de la manière suivante. On demande à la personne de prédire le résultat de 100 lancers d'une pièce équilibrée. Elle donne 70 réponses correctes. Peut-on en conclure qu'elle possède le don de prédire l'avenir ?*

Comme on le voit, on peut utiliser plusieurs méthodologies pour tester l'hypothèse selon laquelle la personne aurait un pouvoir extra-sensoriel. Commençons par quelques définitions.

#### 4.1.2 L'hypothèse nulle

C'est l'hypothèse que l'on souhaite tester. On la note  $H_0$ . Dans l'exemple précédent, l'hypothèse nulle est "la personne choisit au hasard", on note

$$H_0 : \text{"choix au hasard"}$$

On formule en général l'hypothèse de façon telle que le test soit positif quand on rejette l'hypothèse nulle. Nous verrons également que l'hypothèse nulle correspond (en général) au cas où on connaît la distribution d'une variable aléatoire. Si le test conduit à rejeter l'hypothèse  $H_0$ , on note  $RH_0$ . Dans le cas contraire, on dit que le test ne permet pas de rejeter  $H_0$  (on dit parfois par abus de langage qu'on accepte  $H_0$ ), on note  $AH_0$ , bien qu'il serait plus correct d'écrire  $NRH_0$ . On dit aussi que le test est non significatif.

### 4.1.3 L'hypothèse alternative

Elle est notée  $H_1$ . Dans notre exemple, on a  $H_1$  : “le médium présumé *fait mieux* que choisir au hasard”. L'hypothèse alternative représente en général ce que l'on veut effectivement prouver en effectuant le test.

Ici, on va bien sûr considérer que le médium fait mieux que choisir au hasard s'il a un nombre de bonnes réponses très élevé, étonnamment élevé, élevé de manière extraordinaire. Bien entendu, on ne dira pas que la personne est un médium s'il donne un nombre de bonnes réponses très bas. Il s'agit d'un test *unilatéral*. Le domaine de résultat de l'expérience qui conduiront à *rejeter  $H_0$  au profit de  $H_1$*  est situé d'un seul côté de la zone de non rejet de  $H_0$ .

Il faut maintenant savoir ce que l'on considère comme extraordinaire (pour un prétendu médium qui choisirait au hasard). C'est l'objet de la section suivante.

### 4.1.4 Le théorème

Afin de pouvoir mettre en oeuvre un test, il faut généralement avoir un résultat théorique sur la variable dont on observera une réalisation au cours du test, *dans le cas où  $H_0$  est vraie*. Ce résultat fait intervenir les probabilités. Dans notre exemple, si on lance 10 fois la pièce, on a le résultat suivant :

*Si  $H_0$  est vraie, alors la variable aléatoire  $X$  : “nombre de bonnes réponses” suit une loi  $B(10, \frac{1}{2})$ .*

Pour le deuxième exemple, on a également un théorème :

*Si  $H_0$  est vraie, alors la variable aléatoire  $X$  : “nombre de bonnes réponses” suit une loi  $B(100, \frac{1}{2})$ .*

Dans ce deuxième cas, on a même un théorème d'approximation asymptotique :

*Si  $H_0$  est vraie, alors la variable aléatoire  $X$  : “nombre de bonnes réponses” peut être approchée par une loi  $N(50, 5)$ .*

Ces théorèmes permettent de mesurer si le résultat observé est *raisonnable* sous l'hypothèse  $H_0$  en calculant la probabilité d'obtenir une valeur *plus extrême* pour la variable  $X$  quand  $H_0$  est vraie. Dans le premier cas, on calcule  $p = P(X \geq 7|H_0)$  tandis que dans le second, on calcule  $p = P(X \geq 70|H_0)$ . Ces nombres sont appelés *Probabilités de dépassement*, ou en anglais *p-values*. Ils représentent la probabilité de observer une valeur “plus extrême” (que 7 ou 70) au cours de la réalisation du test, si  $H_0$  est vraie.

Dans le premier cas, il semble naturel de penser que la valeur 7 n'est pas incompatible avec l'hypothèse  $H_0$  : quelqu'un qui choisit au hasard pile ou face a une probabilité égale à  $\mathbb{P}(X = 7) + \mathbb{P}(X = 8) + \mathbb{P}(X = 9) + \mathbb{P}(X = 10) = 0,171$  de répondre au moins aussi bien que la personne testée. C'est (beaucoup) trop. On dira que le test ne permet pas de rejeter  $H_0$ . Une autre façon de voir les choses est de dire que si, avec un résultat de 7 sur 10, on déclare que la personne a un pouvoir de prédiction, alors on a un risque égal à 17 % de se tromper, et que le résultat soit dû au hasard.

Dans le deuxième exemple, l'inspection directe de la position de 70 par rapport au graphe de la loi de  $X$  semble indiquer que la valeur 70 est peu compatible avec l'hypothèse  $H_0$ . Plus précisément, sous  $H_0$  (si  $H_0$  est vraie), on peut calculer

$$\mathbb{P}(X \geq 70) \approx \mathbb{P}(N(50, 5) > 69, 5) \approx \mathbb{P}(N(0, 1) > 3, 9) \approx 1 - \pi(3, 9) \approx 0,00005.$$

On constate que la valeur observée de 70 bonnes réponses est incompatible avec l'hypothèse  $H_0$  au sens que, si  $H_0$  est vraie (si on choisissait au hasard), la probabilité de donner autant de bonnes réponses est de 5 sur 100000. On dira que le test permet de rejeter  $H_0$ , et donc de conclure que la personne a bel et bien des pouvoirs de perception

extra-sensorielle. En effet, en faisant une telle affirmation, on n'a qu'un risque de 5 sur 100000 que le résultat soit dû au hasard.

#### 4.1.5 Erreurs et risques

Dans la conclusion d'un test, il y a deux types d'erreurs possibles :

— On rejette l'hypothèse  $H_0$  alors qu'en réalité elle est vraie. C'est une erreur de première espèce. On dit aussi un faux positif. La probabilité de commettre une telle erreur au cours d'un test est appelée *risque de première espèce* et est notée  $\alpha$ . On fixe cette valeur à l'avance. On a donc

$$\alpha = \mathbb{P}(\text{Rejeter } H_0 \mid H_0 \text{ est vraie})$$

— On ne rejette pas l'hypothèse nulle alors qu'elle est fautive. C'est une erreur de seconde espèce. La probabilité de commettre une telle erreur est appelée *risque de seconde espèce* et est généralement notée  $\beta$ .

En règle générale, une erreur de première espèce est plus grave qu'une erreur de seconde espèce puisque cette dernière conduit à formuler "Le test ne permet pas de rejeter  $H_0$ " ou encore "le test ne permet pas de conclure". On s'attache donc à maîtriser le risque de première espèce. Dans le cas de la fabrication d'un médicament,  $H_0$  sera généralement formulée "Le médicament n'agit pas". Rejeter  $H_0$  correspondra à déclarer que le médicament agit, et l'erreur de première espèce consistera donc à déclarer que le médicament agit, alors qu'il n'agit pas.

#### 4.1.6 La mise en oeuvre classique d'un test

Le test est classiquement réalisé de la manière suivante :

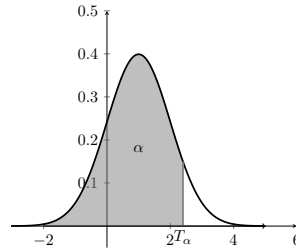
1. On définit les hypothèses  $H_0$  et  $H_1$  (dans notre cas "choix au hasard" et "choix meilleur que le hasard").
2. On se fixe un risque de première espèce maximum acceptable. La valeur  $\alpha = 0,05$  est généralement considérée, mais on peut aussi prendre  $\alpha = 0,01$  ou  $\alpha = 0,001$ .
3. On considère une variable aléatoire  $T$  appelée "Statistique du test" (Test statistic en anglais) ou "Statistique" à partir de l'échantillon aléatoire (dans les exemples  $X$  = "nombre de bonnes réponses" est la somme des variables indicatrices de bonne réponse du premier au dernier lancer, de  $X_1$  à  $X_{10}$  (ou  $X_{100}$ )).
4. On s'arrange pour connaître la distribution de la variable  $T$  si  $H_0$  est vraie. Dans nos exemples  $T$  suit une loi binomiale de paramètres  $(10, 1/2)$  ou  $(100, 1/2)$ .
5. On réalise l'expérience et on calcule la valeur  $T(x_1, \dots, x_n)$  de  $T$  sur les données collectées. On note cette valeur  $T_{obs}$ .
6. On analyse si la valeur observée  $T_{obs}$  est compatible avec la distribution de probabilité de  $T$  quand  $H_0$  est vraie. Si  $T_{obs}$  est extrême par rapport à la distribution (exemple 2), on dit qu'il est *défavorable* à  $H_0$ , si il n'est pas extrême (exemple 1), on dit qu'il est *favorable* à  $H_0$ .

Tout le problème revient alors à préciser ce qu'est "être extrême par rapport à la distribution". Il faut d'abord distinguer si le test est *bilatéral* ou *unilatéral*. Dans un test bilatéral, l'ensemble des valeurs défavorables à  $H_0$  sont les valeurs extrêmes des deux côtés de la distribution, tandis que dans un test unilatéral, les valeurs défavorables à  $H_0$ , c'est à dire celles qui correspondent à  $H_1$ , sont des valeurs extrêmes d'un seul côté de la distribution. Dans nos exemples, le test est unilatéral.

On définit alors a priori une zone de rejet de  $H_0$  (zone des valeurs extrêmes pour  $T$ ) et une zone de non rejet de  $H_0$  en utilisant les quantiles de la distribution  $T$ , et en fonction du risque de première espèce choisi  $\alpha$ .

**Définition 4.1.1.** Si  $T$  est une variable aléatoire dont la fonction de répartition est strictement croissante et continue, alors pour tout nombre  $\alpha \in ]0, 1[$ , il existe un unique nombre  $T_\alpha$  (aussi noté  $Q_T(\alpha)$ ) tel que,  $\mathbb{P}(X < T_\alpha) = \alpha$ . Le nombre  $T_\alpha$  est appelé *quantile d'ordre  $\alpha$  de  $T$*

Graphiquement, la situation est la suivante.



Dans le cas d'un test bilatéral, la zone de non rejet de  $H_0$  est définie par

$$T_{\frac{\alpha}{2}} < T_{obs} < T_{1-\frac{\alpha}{2}},$$

donc la zone de rejet est

$$T_{obs} < T_{\frac{\alpha}{2}} \text{ ou } T_{obs} > T_{1-\frac{\alpha}{2}}$$

Dans le cas d'un test unilatéral, la zone de non rejet de  $H_0$  est donnée par

$$T_\alpha < T_{obs}$$

ou

$$T_{obs} < T_{1-\alpha}$$

selon la définition de l'hypothèse alternative. La zone de rejet est alors la zone complémentaire.

#### 4.1.7 Le lien avec les $p$ -values

Dans les exemples, nous avons calculé mes probabilités de dépassement associées à une valeur observée. Cela permet de ne pas devoir calculer plusieurs zones de rejet correspondant à plusieurs risques  $\alpha$ . En effet, la zone de rejet de  $H_0$  est plus grande pour  $\alpha = 0.05$  que pour  $\alpha = 0.01$ , ou  $\alpha = 0.001$ ...

On réalise le test de la même façon, et on calcule la probabilité d'avoir une valeur plus extrême que  $T_{obs}$  si  $H_0$  est vraie, c'est à dire une valeur "plus en faveur de  $H_1$ ". Dans le cas d'un test unilatéral, on calculera alors  $p = \mathbb{P}(T < T_{obs})$  ou  $p = \mathbb{P}(T > T_{obs})$  selon l'hypothèse considérée. Dans le cas d'un test bilatéral, si la distribution de  $T$  est symétrique sous  $H_0$ , on calculera  $p = 2\mathbb{P}(T > |T_{obs}|)$ .

On aura ainsi l'équivalence avec le formalisme précédent :  $T_{obs}$  est dans la zone de rejet de  $H_0$  pour un risque  $\alpha$  si, et seulement si  $p \leq \alpha$ .

Dans le cas de notre premier exemple, on a calculé  $p = 0.17$ , donc la valeur observée n'est pas dans la zone de rejet de  $H_0$  pour le niveau le plus grand habituel, à savoir  $\alpha = 0.05$ . Le test n'est pas significatif. Dans le deuxième exemple, le calcul de  $p$  montre



que la valeur observée se trouve dans la zone de rejet de  $H_0$  au profit de  $H_1$  quel que soit  $\alpha = 0.05$  ou  $\alpha = 0.01$  ou  $\alpha = 0.001$ .

Terminons cette section en insistant sur le fait qu'il faut se méfier des mauvaises interprétations que l'on eut accorder à  $p$ . Par exemple, il arrive couramment que l'on entende que  $p = 0,001$  signifie qu'il y a une chance sur mille que l'hypothèse  $H_0$  soit vraie. Ce n'est pas la bonne interprétation. Je rappelle que cela signifie qu'il y a une chance sur mille d'obtenir une valeur observée plus extrême, si  $H_0$  est vraie.

## 4.2 Les tests de conformité

Les tests présentés dans cette section ont pour but de tester si un paramètre d'une population est *conforme* à une valeur théorique donnée.

**Exemple 4.2.1.** On sait qu'une race de rats présente un gain moyen de poids de 65 grammes pendant les trois premiers mois de leur vie. Douze rats de cette espèce sont choisis au hasard et reçoivent un régime spécial jusqu'à trois mois. On observe alors les gains de poids suivants

55	62	54	58	65	64
60	62	59	67	62	61

Est-il raisonnable de penser que le régime modifie significativement le gain de poids ?

Il s'agit d'un test sur l'espérance  $\mu_0$  de la distribution sous-jacente des gains de poids des rats traités suivant le régime spécial. Il s'agit d'un test bilatéral, vu la façon dont la question est posée : on a  $H_0 : \mu_0 = 65$  et  $H_1 : \mu_0 \neq 65$ .

On pourrait aussi vouloir démontrer que le régime induit un gain de poids plus élevé (ou moins élevé). On aurait dans ce cas un test unilatéral, et  $H_0 : \mu_0 = 65$  et  $H_1 : \mu_0 > 65$ .

### 4.2.1 Test de conformité pour une espérance, grands échantillons

Nous souhaitons tester si l'espérance  $\mu_0$  d'une distribution est égale à une valeur  $a$  (dite théorique) donnée à l'avance.

1. *Les hypothèses* : nous avons bien sûr  $H_0 : \mu_0 = a$ . Pour  $H_1$ , il y a plusieurs possibilités. Nous commencerons par le cas bilatéral, c'est-à-dire  $H_1 : \mu_0 \neq a$ .
2. *Le fonctionnement intuitif<sup>a</sup>* : On extrait un échantillon sur lequel on mesure  $\bar{x}$  (la valeur de l'estimateur de  $\mu_0$  sur l'échantillon). Si  $\bar{x}$  est "proche" de  $a$ , on accepte  $H_0$ . S'il est loin de  $a$ , on rejette  $H_0$ .
3. *Le théorème et ses corollaires* : Puisque l'échantillon que nous allons extraire est grand ( $n \geq 30$ ), le théorème central limite donne

$$\frac{\bar{X} - \mu_0}{\frac{\sigma_0}{\sqrt{n}}} \sim N(0, 1)$$

On obtient alors un "théorème" de type 4.1.4.

**Théorème 4.2.1.** *Si  $H_0$  est vraie alors*

$$T = \frac{\bar{X} - a}{\frac{\sigma_0}{\sqrt{n}}} \sim N(0, 1).$$

a. On peut se passer de cette vision intuitive, mais elle montre bien que les zones de rejet de  $H_0$  doivent être faites de valeurs nettement plus grandes et nettement plus petites que  $a$ .

Ce qui fournit la zone de rejet et de non rejet de  $H_0$  pour un risque  $\alpha$  donné.

**Théorème 4.2.2.** *Pour tout  $\alpha \in ]0, 1[$ , si  $H_0$  est vraie*

$$\mathbb{P}\left(-u_{1-\frac{\alpha}{2}} < \frac{\bar{X} - a}{\frac{\sigma_0}{\sqrt{n}}} < u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Généralement on choisit le risque de première espèce  $\alpha = 0,05$ , mais on peut également fixer  $\alpha = 0,01$  ou  $0,001$ .

Le test est alors simple :

- on extrait de l'échantillon de taille  $n$
- on calcule la valeur  $\bar{x}$  de  $\bar{X}$  sur l'échantillon
- on en déduit la valeur de  $T_{obs} = \frac{\bar{x} - a}{\frac{\sigma_0}{\sqrt{n}}}$

On décide alors en fonction de la valeur de  $T_{obs}$

- Si  $T_{obs} = \frac{\bar{x} - a}{\frac{\sigma_0}{\sqrt{n}}} \in ] -u_{1-\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}} [$ , le test ne permet pas de rejeter  $H_0$ .
- Si  $T_{obs} \notin ] -u_{1-\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}} [$ , on rejette  $H_0$ .

**Remarques :**

- a) La région de rejet de  $H_0$  est située symétriquement de part et d'autre de 0. C' est naturel puisqu'on n'a aucune raison de penser a priori que la valeur de  $\mu_0$  est plus grande que  $a$ , ou plus petite que  $a$  : on est dans le cas d'un test bilatéral.
- b) En pratique, il arrive souvent qu'on ne connaisse pas  $\sigma_0$ . Comme pour les intervalles de confiance, on le remplace alors par son estimateur  $s$ , sans problème puisque la taille de l'échantillon est suffisante.
- c) Le critère de décision du test peut aussi s'écrire  
Si  $|\frac{\bar{x} - a}{\frac{\sigma_0}{\sqrt{n}}}| < u_{1-\frac{\alpha}{2}}$ , on ne rejette pas  $H_0$ , sinon on rejette  $H_0$ .

Considérons maintenant l'hypothèse alternative  $H_1 : \mu_0 > a$ . Cela signifie qu'on s'attend a priori à trouver une espérance plus grande que  $a$  et qu'on veut le prouver par le test.

**Exemple 4.2.2.** Il y a dix ans, une entreprise avait mesuré le montant des achats moyens de ses clients et avait obtenu une moyenne de 250 euros. La firme pense que l'achat moyen par client a baissé. Afin de le démontrer, on étudie un échantillon de 100 clients pris au hasard et on obtient les valeurs  $\bar{x} = 242,5$  et  $s = 66$ .

Peut-on affirmer avec un risque de première espèce de 5% que l'achat moyen a réellement baissé ?

Il s'agit ici encore d'un test sur l'espérance  $\mu_0$  de la distribution de la variable "montant de l'achat". Mais on ne rejettera  $H_0$  que si cette espérance a baissé. Cela conduit donc aux hypothèses  $H_0 : \mu_0 = a = 250$  et  $H_1 : \mu_0 < a = 250$ . On utilise toujours le théorème 4.2.1. Mais intuitivement, on rejette l'hypothèse  $H_0$  si  $\bar{x}$  se situe suffisamment en dessous de 250. On construit donc une zone de rejet de  $H_0$  non symétrique : il s'agit d'un test unilatéral. On utilise la même variable  $T$  que précédemment pour obtenir le théorème.

**Théorème 4.2.3.** *Pour tout  $\alpha \in ]0, 1[$ , si  $H_0$  est vraie*

$$\mathbb{P}\left(u_\alpha < \frac{\bar{X} - a}{\frac{\sigma_0}{\sqrt{n}}}\right) = 1 - \alpha$$

Le test fonctionne alors comme suit :

- Si  $T_{obs} = \frac{\bar{x} - a}{\frac{\sigma_0}{\sqrt{n}}} > u_\alpha$ , on ne rejette pas  $H_0$ .

— Si  $T_{obs} < u_\alpha$ , on rejette  $H_0$ .

Ici encore, puisque la taille de l'échantillon est grande, on peut remplacer  $\sigma_0$  par son estimation  $s$ .

### Solution de l'exemple 4.2.2

Pour  $\alpha = 0,05$ , on trouve  $u_\alpha = -1,65$ , de plus on a

$$T_{obs} = \frac{\bar{x} - a}{\frac{s}{\sqrt{n}}} = \frac{242,5 - 250}{66/10} = -1,136$$

On ne peut donc pas rejeter  $H_0$ , donc le test ne permet pas de prouver que le montant moyen des achats par clients a diminué de manière significative.

**Remarque :** Le test unilatéral défini par  $H_1 : \mu_0 > a$  se traite de manière analogue

— Si  $T_{obs} < u_{1-\alpha}$ , on ne rejette pas  $H_0$ .

— Si  $T_{obs} > u_{1-\alpha}$ , on rejette  $H_0$ .

### 4.2.2 Test de conformité pour une espérance, petits échantillons

Comme nous l'avons vu pour la construction des intervalles de confiance pour une espérance, quand  $n$  est petit ( $n < 30$ ), il n'est plus possible d'utiliser le théorème central limite ni de remplacer  $\sigma_0$  par son estimation. Nous procédons comme pour la construction des intervalles de confiance.

On suppose que la distribution sous-jacente est normale (*gaussienne*). Il n'est alors plus nécessaire d'utiliser le théorème central limite.

Si  $\sigma_0$  est connu, on a alors comme plus haut, pour obtenir le théorème valide si  $H_0$  est vrai.

**Théorème 4.2.4.** *Si  $H_0$  est vraie alors*

$$\frac{\bar{X} - a}{\frac{\sigma_0}{\sqrt{n}}} \sim N(0, 1).$$

On a alors exactement les mêmes résultats que dans la section précédente. Si  $\sigma_0$  n'est pas connu, il faut (toujours comme pour les intervalles de confiance) utiliser la loi de Student, et l'estimateur de la variance.

**Théorème 4.2.5.** *Si la distribution dont on teste l'espérance est normale et si  $H_0$  est vraie alors*

$$T = \frac{\bar{X} - a}{\frac{S}{\sqrt{n}}} \sim t_{n-1}.$$

Les résultats de la section précédente s'adaptent alors aisément en remplaçant  $\sigma_0$  par  $s$  et les quantiles  $u_\beta$  par leurs analogues  $t_{n-1,\beta}$ .

### Solution de l'exemple 4.2.1

Il s'agit d'un test de conformité pour une espérance. C'est un test bilatéral puisqu'on n'a pas d'information a priori sur le sens dans lequel le gain de poids est modifié par le régime : on ne cherche pas à prouver que le régime fait augmenter le gain de poids, ou diminuer, mais seulement qu'il modifie le gain de poids. On est dans le cas des petits échantillons, puisque  $n = 12$ . On a  $H_0 : \mu_0 = 65$  et  $H_1 : \mu_0 \neq 65$ .

On calcule  $\bar{x} = 60,75$  g et  $s = 3,8406$  g. On rejette alors  $H_0$  pour un risque de première espèce  $\alpha$  si

$$\frac{\bar{x} - a}{\frac{s}{\sqrt{n}}} \notin \left] -t_{n-1,1-\frac{\alpha}{2}}, t_{n-1,1-\frac{\alpha}{2}} \right[$$

et on l'accepte sinon.

Pour  $\alpha = 0,05$ , les quantiles de la loi  $t_{11}$  sont donnés par

$$t_{11;0,975} = 2,20 = -t_{11;0,025}.$$

Autrement dit, on ne rejette pas  $H_0$  si

$$\frac{60,75 - 65}{3,8406/\sqrt{11}} = -3,670 \in ] -2,20; 2,20[.$$

Ce n'est pas le cas et on rejette donc l'hypothèse  $H_0$  pour un risque de première espèce de 0,05 : le test permet de prouver que le régime modifie significativement<sup>b</sup> le gain de poids.

### 4.3 Tests d'ajustement

Il s'agit ici de tester si la variable aléatoire sous-jacente à la population suit une distribution donnée, dite *distribution théorique*. L'intérêt de ce genre de question a déjà été démontré dans les sections précédentes où on devait faire l'hypothèse : "la distribution sous-jacente est normale". Un des tests les plus célèbres pour résoudre cette question est le test de Pearson, encore appelé test du chi-carré. Voici un exemple concret.

**Exemple 4.3.1.** Un joueur pense qu'un dé à 6 faces n'est pas équilibré. Afin de tester son impression, il lance 150 fois le dé et obtient les résultats suivants

Points	1	2	3	4	5	6
Répétitions	23	26	23	21	31	26

Ces résultats permettent-ils de conclure que le dé n'est pas bien équilibré ?

Intuitivement, le nombre 5 est apparu un peu trop souvent, tandis que le nombre 4 est apparu trop peu souvent. On s'attend en effet à observer un nombre de résultats proches de 25 pour chaque face. Il faut alors savoir si ces différences sont significatives ou si elles sont dues au hasard.

Les résultats varient selon que la distribution théorique suit une loi discrète ou continue, et selon que l'on connaît tous ses paramètres ou non.

#### 4.3.1 Le cas discret et non paramétrique

On considère une variable aléatoire discrète (cas fini) pouvant prendre  $m$  modalités  $a_1, \dots, a_m$ . On s'intéresse alors à la distribution de probabilité de  $X$  :

Modalités	$a_1$	$\cdots$	$a_m$
$p_i$	$p_1$	$\cdots$	$p_m$

L'hypothèse nulle s'écrit alors  $H_0 : p_i = p_{0,i}$ , où les  $p_{0,i}$  sont donnés. Dans le cas de notre exemple, l'hypothèse nulle est  $H_0 : p_i = \frac{1}{6}$ , pour tout  $i \leq 6$ .

#### Le fonctionnement intuitif

L'idée est de comparer les répétitions observées pour chacune des modalités au répétitions attendues (théoriques) si l'hypothèse  $H_0$  est vraie. Dans l'exemple 4.3.1, ces répétitions

<sup>b</sup>. Le terme significativement est réservé et correspond à un rejet de  $H_0$  pour un risque de 0,05. Pour un risque de 0,01, le rejet, s'il intervient, sera dit très significatif.

théoriques sont toutes égales à 25. On mesure la distance en calculant la somme des carrés des écarts (à peu de choses près) entre les répétitions théoriques et les répétitions que l'on observe sur l'échantillon. Ces dernières sont des réalisations de variables aléatoires, notées  $R_1, \dots, R_m$ . Si la distance est trop grande, on rejette l'hypothèse  $H_0$ . Il faut évidemment disposer d'un théorème pour définir ce qu'est une trop grande distance.

**Théorème 4.3.1** (Pearson). *Si l'hypothèse  $H_0$  est vraie, et si  $np_{0,i} > 5$  pour tout  $i$ , alors la variable aléatoire*

$$\sum_{i=1}^m \frac{(R_i - np_{0,i})^2}{np_{0,i}}$$

suit une loi  $\chi_{m-1}^2$ .

On voit donc bien que la variable utilisée mesure, lors de la réalisation de l'expérience, la somme (pondérée) des carrés de écarts des répétitions observées aux répétitions théoriques. Il s'agit ici encore d'un théorème asymptotique. La condition  $np_{0,i} > 5$  dans l'énoncé est la condition pratique qui implique que  $n$  doit être "grand".

#### Le test :

On extrait un échantillon de taille  $n$  et on note  $r_i$  les répétitions observées. On aura évidemment tendance à ne pas rejeter  $H_0$  si la valeur  $\chi_{\text{obs}}^2 = \sum_{i=1}^m \frac{(r_i - np_{0,i})^2}{np_{0,i}}$  est proche de zéro, puisqu'alors les répétitions observées seront proches des valeurs attendues (théoriques). On rejette  $H_0$  si la quantité  $\chi_{\text{obs}}^2$  est supérieure à la valeur  $\chi_{m-1,1-\alpha}^2$ . Dans le cas contraire, on accepte  $H_0$ .

#### Solution de l'exemple 4.3.1 :

On a  $n = 150$  et  $p_{0,i} = 1/6$ , donc  $np_{0,i} = 25 > 5$ . On peut donc appliquer le test du chi-carré. On a le tableau suivant

Rép. Obs.	Rép. Th.	$\frac{(r_i - np_{0,i})^2}{np_{0,i}}$
23	25	0,16
26	25	0,04
23	25	0,16
21	25	0,64
31	25	1,44
26	25	0,04

On obtient une valeur  $\chi_{\text{obs}}^2 = 2,48$  tandis que la valeur  $\chi_{5;0,95}^2$  est 11,07. On ne doit donc pas rejeter l'hypothèse nulle : les différences observées ne sont pas significatives.

#### 4.3.2 Le cas discret, paramétrique

Il s'agit de tester si la variable  $X$  suit une loi de probabilité connue aux paramètres près. Par exemple, on peut se demander si une variable suit une loi binomiale, ou une loi de Poisson, sans préciser laquelle.

On commence par estimer les paramètres inconnus en utilisant la théorie de l'estimation ponctuelle (remarquons cependant qu'on pourrait estimer les paramètres en utilisant d'autres estimateurs, dits du chi-carré minimum). Les développements théoriques sont identiques à ceux de la section précédente sauf le théorème, qui tient compte du nombre de paramètres estimés.

**Théorème 4.3.2.** Si l'hypothèse  $H_0$  est vraie, et si  $np_{0,i} > 5$ , alors la variable aléatoire

$$\sum_{i=1}^m \frac{(R_i - np_{0,i})^2}{np_{0,i}}$$

suit une loi  $\chi_{m-r-1}^2$ , où  $r$  est le nombre de paramètres estimés.

Le test est alors le même sauf qu'il faut utiliser les quantiles de  $\chi_{m-r-1}^2$ .

**Exemple 4.3.2.** Sur une étude portant sur 100 familles comportant 4 enfants choisies au hasard, on compte le nombre de filles. On obtient la table suivante

Nombre de filles	0	1	2	3	4
nombre de familles	10	31	41	15	3

Tester l'hypothèse selon laquelle le nombre de filles par familles suit une loi binomiale  $B(4, 1/2)$ .

Tester l'hypothèse selon laquelle le nombre de filles par familles suit une loi binomiale.

Dans les deux cas, il s'agit d'un test sur la distribution du nombre de filles dans les familles de 4 enfants. Dans le premier cas, on connaît le paramètre de la distribution théorique. Ce n'est pas le cas dans la deuxième situation.

### 4.3.3 Le cas continu

Dans ce cas la variable aléatoire considérée a une distribution continue. Il s'agit encore de tester si elle admet une distribution connue (complètement ou aux paramètres près). Il existe plusieurs tests pour répondre à ce genre de questions. Citons par exemple le test de Kolmogorov-Smirnov. Nous allons nous contenter ici de voir comment il convient d'adapter le test du Chi-carré en *discrétisant* le problème.

Nous connaissons bien ce genre d'opération en statistique descriptive. En effet, pour résumer l'information contenue dans une série statistique d'une variable continue, nous partitionnons l'ensemble de variation de  $x$  en classes et nous comptons le nombre de répétitions de la variable dans la classe, pour réaliser un histogramme. On est alors ramené à traiter un problème discret.

**Exemple 4.3.3.** On a extrait un échantillon de taille  $n = 100$  d'une population et reporté le poids de chaque individu. Les données obtenues sont les suivantes :

Classes	[59, 5; 62, 5[	[62, 5; 65, 5[	[65, 5; 68, 5[	[68, 5; 71, 5[	[71, 5; 74; 5[
Répétitions	5	18	42	27	8

Peut-on affirmer que le poids des individus de cette population suit une loi normale ?

#### Solution

La loi normale est déterminée par deux paramètres qu'il faut estimer à partir de l'échantillon. De plus, on ne dispose pas de toute l'information puisque les données sont déjà réparties en classes.

On peut toutefois estimer la moyenne

$$\bar{x} \approx \frac{5 * 61 + 18 * 64 + 42 * 67 + 27 * 70 + 8 * 73}{100} = 67,45,$$

et la variance

$$s \approx \sqrt{\frac{1}{n-1} \sum_i n_i (c_i - \bar{x})^2} = 2,93,$$

si  $c_i$  est le centre de la classe  $C_i$ . Nous calculons ensuite les probabilités théoriques correspondant aux différentes classes : Si  $X \sim N(67,45; 2,93)$  et si on note  $\pi$  la fonction de répartition de la normale  $N(0,1)$  alors

$$\begin{aligned} \mathbb{P}(X \in [62,5; 65,5[) &= \mathbb{P}\left(\frac{X-67,45}{2,93} \in [-1,68; -0,66[) = \pi(-0,66) - \pi(-1,68) \\ \mathbb{P}(X \in [65,5; 68,5[) &= \mathbb{P}\left(\frac{X-67,45}{2,93} \in [-0,66; 0,35[) = \pi(0,35) - \pi(-0,66) \\ \mathbb{P}(X \in [68,5; 71,5[) &= \mathbb{P}\left(\frac{X-67,45}{2,93} \in [0,35; 1,38[) = \pi(1,38) - \pi(0,35). \end{aligned}$$

Pour les deux classes extrêmes, il faut être attentif. On remplace la probabilité théorique  $\mathbb{P}(X \in [59,5; 62,5[)$  par  $\mathbb{P}(X < 62,5)$  pour tenir compte du fait que  $X$  peut prendre des valeurs inférieures à 59,5, même si elles ne sont pas apparues dans l'échantillon prélevé. On fait de même avec la classe supérieure.

$$\begin{aligned} \mathbb{P}(X < 62,5) &= \mathbb{P}\left(\frac{X-67,45}{2,93} < -1,68\right) = \pi(-1,68) \\ \mathbb{P}(X > 71,5) &= \mathbb{P}\left(\frac{X-67,45}{2,93} > 1,38\right) = 1 - \pi(1,38). \end{aligned}$$

On peut alors réaliser le test du chi-carré :

Classes	Rép. Obs. $r_i$	Rép. Th. $(np_{0,i})$	$\frac{(r_i - np_{0,i})^2}{np_{0,i}}$
$] -\infty; 62,5[$	5	4,55	0,044
$[62,5; 65,5[$	18	20,59	0,325
$[65,5; 68,5[$	42	38,91	0,245
$[68,5; 71,5[$	27	27,56	0,011
$[71,5; +\infty[$	8	8,38	0,017

On obtient donc

$$\chi_{\text{obs}}^2 = 0,044 + 0,325 + 0,245 + 0,011 + 0,017 = 0,642$$

Si on choisit  $\alpha = 0,1$ , alors il convient de comparer cette valeur observée à  $\chi_{2;0,9}^2 = 4,6$  ( $2 = 5 - 2 - 1$ ). On ne doit donc pas rejeter l'hypothèse nulle et on ne rejette donc pas l'hypothèse selon laquelle la variable suit effectivement une loi gaussienne.

## 4.4 Comparaison de deux populations

On se propose ici de comparer les paramètres de deux populations distinctes.

**Exemple 4.4.1.** Une série de 15 nourrissons reçoit pendant une période donnée un traitement riche en protéines et on observe les gains de poids suivants

134	146	104	119	124	161	107	83
113	129	97	123	123	119	121	

Durant la même période, 12 nourrissons reçoivent un traitement pauvre en protéines et on observe les gains de poids suivants

70 118 101 85 107 132 94 105 98 107 89 114

Peut-on conclure que le régime riche en protéines assure en moyenne un gain de poids supérieur à celui que donne l'autre régime ?

Il s'agit ici de comparer les espérances  $\mu_{0,1}$  et  $\mu_{0,2}$  du gain de poids pour deux populations distinctes.

On a ici les hypothèses  $H_0 : \mu_{0,1} = \mu_{0,2}$  et  $H_1 : \mu_{0,1} > \mu_{0,2}$ .

#### 4.4.1 Comparaison de deux espérances

Nous supposons ici que les populations étudiées sont indépendantes, plus précisément, que les variables aléatoires  $X_1$  correspondant aux mesures effectuées sur la population 1 et  $X_2$  correspondant aux mesures effectuées sur la population 2 sont indépendantes. Cela doit être testé (voir les tests d'indépendance). Ces méthodes ne s'appliquent donc pas pour un test avant-après qui mesure la même variable sur les individus d'un échantillon par exemple avant et après l'administration d'un traitement. Nous notons  $n_1$  et  $n_2$  les effectifs des échantillons,  $\bar{x}_1$  et  $\bar{x}_2$  les valeurs observées de  $\bar{X}_1$  et  $\bar{X}_2$  sur les échantillons choisis,...

L'hypothèse nulle est évidemment  $H_0 : \mu_{0,1} = \mu_{0,2}$ . Les hypothèses alternatives sont  $H_1 : \mu_{0,1} \neq (>, <) \mu_{0,2}$ .

Pour le fonctionnement intuitif, l'idée est de comparer les estimations des espérances :  $\bar{x}_1$  et  $\bar{x}_2$  et de voir si la différence est proche de zéro ou non. Comme d'habitude, nous avons besoin d'une variable aléatoire et d'un théorème pour préciser l'expression "proche de zéro".

##### Le cas des grands échantillons

Si les deux échantillons ont un effectif supérieur à 30, on peut appliquer le théorème central limite pour obtenir

$$\bar{X}_1 \sim N\left(\mu_{0,1}, \frac{\sigma_{0,1}}{\sqrt{n_1}}\right) \quad \text{et} \quad \bar{X}_2 \sim N\left(\mu_{0,2}, \frac{\sigma_{0,2}}{\sqrt{n_2}}\right)$$

On en déduit le théorème suivant, vu l'indépendance de  $\bar{X}_1$  et  $\bar{X}_2$ .

**Théorème 4.4.1.** *On a*

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_{0,1} - \mu_{0,2}, \sqrt{\frac{\sigma_{0,1}^2}{n_1} + \frac{\sigma_{0,2}^2}{n_2}}\right).$$

*En particulier, si  $H_0$  est vraie, alors*

$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_{0,1}^2}{n_1} + \frac{\sigma_{0,2}^2}{n_2}}} \sim N(0, 1).$$

Le test se construit alors comme les précédents. Détaillons le cas bilatéral  $H_1 : \mu_{0,1} \neq \mu_{0,2}$ .

Si  $H_0$  est vraie, on a la relation

$$\mathbb{P}\left(\left|\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_{0,1}^2}{n_1} + \frac{\sigma_{0,2}^2}{n_2}}}\right| < u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

L'alternative est donc la suivante.

- On rejette  $H_0$  si  $\left|\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_{0,1}^2}{n_1} + \frac{\sigma_{0,2}^2}{n_2}}}\right| > u_{1-\frac{\alpha}{2}}$

- On ne rejette pas  $H_0$  si ce n'est pas le cas.

**Remarque :** Si les variances des deux populations ne sont pas connues, puisque la taille des échantillons est suffisante, on peut les remplacer par leurs estimateurs.



### Le cas des petits échantillons

Si un au moins des deux échantillons a un effectif inférieur à trente, on ne peut plus arriver à la conclusion que la variable  $\bar{X}$  correspondante est normale en utilisant le théorème central limite. On suppose alors que **la distribution de  $X$  est gaussienne**. Si tel est le cas, on peut procéder comme plus haut, puisque  $\bar{X}_1 - \bar{X}_2$  suit également une loi gaussienne.

De même, il n'est plus possible de remplacer les variances par leurs estimateurs et obtenir les mêmes conclusions. On doit, comme dans la construction des intervalles de confiance, utiliser la variable de Student. Pour cela il faut également supposer que **les variances dans les deux populations sont égales** ( $\sigma_{0,1} = \sigma_{0,2} = \sigma$ )<sup>c</sup> On définit alors la variable

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

et on obtient le résultat suivant.

**Théorème 4.4.2.** *Si  $H_0$  est vraie, alors la variable aléatoire*

$$\frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

*suit une loi de Student  $t_{n_1+n_2-2}$ .*

Traitons par exemple le cas unilatéral  $H_1 : m_{0,1} > m_{0,2}$ .  
On calcule la quantité

$$\frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

On rejette l'hypothèse nulle si cette quantité est plus grande que  $t_{n_1+n_2-2;1-\alpha}$ .

## 4.5 Exercices

1. Sur un dé à six faces, on marque deux faces avec le point 1, deux faces avec le point 2, une face avec le point 3, une face avec le point 4.
  - a. Si le dé est équilibré (toutes les faces ont la même probabilité d'apparition lors d'un lancé), donner les probabilités d'apparition des points 1,2,3 et 4 lors d'un lancer.
  - b. Un joueur pense que le dé n'est pas équilibré. Pour vérifier son hypothèse, il jette 36 fois le dé et obtient les observations suivantes :

$x_i$	1	2	3	4
Répétitions	20	9	4	3

Le test permet-il de conclure, en acceptant un risque de première espèce de 5% que le dé n'est pas équilibré ?

Même question avec un risque de première espèce de 1%.

2. On considère l'expérience aléatoire consistant à jeter 3 fois une pièce de monnaie équilibrée. Donner la distribution de probabilité du nombre de "piles" obtenues.

---

c. Les deux suppositions sur la distribution de  $X$  et sur l'égalité des variances doivent être testées préalablement.

On choisit une pièce et on réalise effectivement 48 fois l'expérience. On obtient les résultats suivants

Nombre de piles	0	1	2	3
Répétitions	9	27	10	2

Peut-on déduire de cette étude que la pièce choisie n'est pas adéquate pour réaliser l'expérience, en admettant un risque de première espèce de 5%. Même question avec un risque de 1%.

3. On veut tester l'efficacité d'un médicament agissant sur la pression artérielle. On sait que la pression artérielle moyenne dans la population est de 115 mmHg. On observe alors la pression artérielle chez 11 individus traités avec le médicament et on obtient les résultats suivants :

73 81 103 88 130 106 106 111 122 108 99

- Donner une estimation ponctuelle de la pression artérielle moyenne dans la population traitée avec le médicament.
  - Donner une estimation ponctuelle de la variance de la pression artérielle dans la population traitée.
  - Donner un intervalle de confiance au seuil de 95% pour la pression artérielle moyenne dans la population traitée.
  - Peut-on déduire de cette expérience que la pression artérielle dans la population traitée est différente de 115 mmHg, en admettant un risque de première espèce de 5% ? Même question pour un risque de 1%.
4. Une roulette possède 18 cases rouges, 18 cases noires et 4 cases vertes. On lance 200 fois une bille dans la roulette et on note à chaque fois la couleur de la case dans laquelle la bille arrive. Les résultats sont les suivants :

Couleur	Répétition
Rouge	88
Noir	102
Vert	10

En acceptant un risque de première espèce de 5%, peut-on admettre que la roulette est bien équilibrée ? Que se passe-t-il pour un risque de première espèce de 1% ?

## 4.6 Comparaison de deux espérances, populations appariées

On considère maintenant le cas de deux populations où les données sont naturellement groupées par paires. C'est notamment le cas quand on veut mesurer l'efficacité d'un traitement en faisant des mesures avant-après sur les mêmes sujets. Il n'est alors plus question de supposer que les populations sont indépendantes puisqu'on mesure sur les mêmes sujets deux valeurs de la même variable considérée.

**Exemple :** On souhaite tester l'efficacité d'un régime amaigrissant. On le fait subir à 10 sujets choisis au hasard. Les résultats sont les suivants :

Sujet	1	2	3	4	5	6	7	8	9	10
Poids après	127	195	162	170	143	205	168	175	197	136
Poids avant	135	200	160	182	147	200	172	186	194	141

On a alors les hypothèses suivantes  $H_0 : \mu_1 = \mu_2$  et  $H_1 : \mu_1 \neq \mu_2$  ou  $H_1 : \mu_1 > \mu_2$  (l'autre cas pouvant encore être ramené à celui-ci en changeant les noms des populations).

On remarque que les deux échantillons ont le même nombre d'individus que l'on note  $n$ . On note  $X_{1,1}, \dots, X_{1,n}$  les variables aléatoires "première mesure sur l'individu  $1 \dots n$ ". On fait de même pour  $X_{2,1}, \dots, X_{2,n}$ . L'idée est de définir la variable aléatoire  $D$

$$D(\omega) = X_1(\omega) - X_2(\omega) :$$

on associe à chaque individu la différence des mesures. On a également  $\mu_d = \mu_1 - \mu_2$ .

Comme d'habitude, en mesurant  $d$  sur les individus de l'échantillon, on obtient  $n$  variables aléatoires  $D_i = X_{1,i} - X_{2,i}$  pour  $i \in \{1, \dots, n\}$ . On utilise ces variables pour tester l'hypothèse  $H_0 : \mu_d = 0$  contre  $H_1 : \mu_d \neq 0$  ou  $H_1 : \mu_d > 0$ . On est alors face à un test de conformité pour une espérance. Rappelons que la marche à suivre dépend de la taille de l'échantillon. Le test est dans les deux cas basé sur la variable aléatoire  $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$ .

#### 4.6.1 Le cas des grands échantillons

Puisque les  $D_i$  peuvent être considérées comme indépendantes et qu'elles ont même distribution, on a par le théorème central limite :

$$\frac{\bar{D} - \mu_d}{\frac{\sigma_D}{\sqrt{n}}} \sim N(0, 1),$$

donc si  $H_0$  est vraie, on a

$$T = \frac{\bar{D}}{\frac{\sigma_D}{\sqrt{n}}} \sim N(0, 1).$$

Dans le cas bilatéral, on rejette  $H_0$  si  $|T_{obs}| > u_{1-\frac{\alpha}{2}}$ , on l'accepte sinon. La probabilité de dépassement associée à  $T_{obs}$  est donnée par

$$p = \mathbb{P}(T > |T_{obs}|) + \mathbb{P}(T < -|T_{obs}|) = 2 \mathbb{P}(T > |T_{obs}|).$$

Dans le cas unilatéral, avec  $H_1 : \mu_1 > \mu_2$ , on rejette  $H_0$  si  $T_{obs} > u_{1-\alpha}$ . La probabilité de dépassement associée à  $T_{obs}$  est donnée par

$$p = \mathbb{P}(T > T_{obs}).$$

Il est évident que  $\sigma_D$  n'est en général pas connu. On le remplace comme d'habitude par  $s_D$  et les résultats restent valables puisque l'échantillon est grand.

#### 4.6.2 Le cas des petits échantillons

On doit supposer que  $D$  suit une loi normale dans la population. On peut alors utiliser la variable aléatoire

$$T = \frac{\bar{D}}{\frac{s}{\sqrt{n}}}.$$

Si  $H_0$  est vraie, alors  $T$  suit une loi de Student  $t_{n-1}$ . Les résultats de la section précédente restent valables en remplaçant  $N(0, 1)$  par  $t_{n-1}$ .

##### Solution de l'exemple :

On doit supposer que les différences de poids suivent une loi normale, puisque l'échantillon est petit. L'hypothèse  $H_0$  est  $\mu_1 = \mu_2$  (ou  $\mu_D = 0$ ) : le régime n'a pas d'effet. L'hypothèse  $H_1$  est  $\mu_1 > \mu_2$ , ou ( $\mu_d > 0$ ). C'est donc un test unilatéral. On calcule

$$\bar{d} = 3.9, \quad s_D = 5.74, \quad T_{obs} = 2.13.$$

On doit comparer cette quantité à  $t_{9;0,95} = 1.83$ . On rejette donc  $H_0$  pour un risque de 5% : le régime fait effectivement maigrir. Cependant, on constate que  $t_{9;0,975} = 2.23$ , donc la probabilité de dépassement se situe entre 0,025 et 0,05. En d'autres termes, pour un risque de 2,5%, on n'aurait pas pu conclure que le régime fait maigrir.

Remarquons également que si on avait voulu tester l'hypothèse "le régime affecte les poids" on aurait dû comparer, pour un risque de 5%,  $T_{obs}$  à  $t_{9;0,975} = 2.23$  et on n'aurait pas pu rejeter  $H_0$ .

### 4.6.3 Exercices

- Deux compagnies productrices de semences prétendent que leurs semences sont les meilleures. Afin de trancher la dispute, on décide de planter dans 8 localités un champs de superficie donnée avec chaque semence. Les résultats sont les suivants

Localité	1	2	3	4	5	6	7	8
Comp. 1	114	86	93	75	102	89	64	95
Comp. 2	107	94	86	70	90	82	73	81

Peut-on conclure de ces données que les résultats sont différents, en acceptant un risque de première espèce de 5%? Calculer la probabilité de dépassement des observations.

Peut-on décider à partir de ces données qu'une compagnie est meilleure que l'autre?

- On veut comparer la sévérité de deux examinateurs. On fait corriger 30 copies par chacun d'eux et on obtient les résultats suivantes

Exam. A	13	15	12	16	18	15	14	18	17	20	15	16	15	17	16
Exam. B	14	16	13	16	17	15	15	17	16	17	17	17	15	18	16

Exam. A	13	15	11	14	15	17	15	16	18	14	16	15	17	14	15
Exam. B	14	16	12	14	18	16	15	18	20	15	15	15	19	16	16

Peut-on déduire de ces données, en admettant un risque de première espèce de 5% que les examinateurs ont une sévérité différente?

## 4.7 Comparaison de deux variances

Nous souhaitons étudier ici si les variances d'une variable statistique  $x$  sur deux populations sont égales ou différentes. Nous supposons qu'il s'agit de deux populations indépendantes. On utilise les notations de la section précédente. On a alors l'hypothèse nulle

$$H_0 : \sigma_1^2 = \sigma_2^2,$$

et les hypothèses alternatives

$$H_1 : \sigma_1^2 \neq \sigma_2^2, \quad \text{ou} \quad \sigma_1^2 > \sigma_2^2,$$

qui correspondent respectivement à un test bilatéral ou unilatéral. L'hypothèse alternative  $\sigma_1^2 < \sigma_2^2$  peut être ramenée à  $\sigma_1^2 > \sigma_2^2$  en changeant le nom des populations.

**Exemple :** Un professeur enseigne une même matière à deux classes  $A$  et  $B$ , de 16 et 25 élèves. Au cours d'un même examen, il ne note aucune différence significative de

moyennes, mais l'écart-type mesuré de la classe  $A$  est 9, alors que celui de la classe  $B$  est 12. Peut-on conclure en admettant un risque de première espèce de 5% qu'il y ait plus de dispersion dans la classe  $B$  que dans la classe  $A$ .

Ce test est basé sur la distribution  $F$  de Fischer-Snedecor. Elle est définie de la façon suivante

**Définition 4.7.1.** Soit  $X$  une variable aléatoire suivant une distribution  $\chi_m^2$  et  $Y$  une variable suivant une distribution  $\chi_n^2$ , telles que  $X$  et  $Y$  soient indépendantes. La variable aléatoire

$$F_{m,n} = \frac{X/m}{Y/n}$$

suit une loi de Snedecor (loi  $F$ ) à  $m$  et  $n$  degrés de libertés.

Si  $X$  a une distribution normale, de variance  $\sigma^2$ , alors quand on choisit un échantillon de taille  $n$ , la variable

$$\frac{(n-1)S^2}{\sigma^2}$$

suit une loi  $\chi_{n-1}^2$ . Dans le cas qui nous occupe, cela donne tout de suite le résultat suivant.

**Théorème 4.7.1.** Si la variable  $X$  étudiée suit une loi normale dans les deux populations alors la variable aléatoire

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2}$$

suit une loi de Fischer-Snedecor à  $n_1 - 1$  et  $n_2 - 1$  degrés de libertés.

En particulier si  $H_0$  est vraie, alors

$$F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}.$$

On peut alors construire le test comme d'habitude :

Dans le cas bilatéral, on a l'alternative suivante, pour un risque de première espèce  $\alpha$  :

- Si  $F_{obs} \in ]F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}, F_{n_1-1, n_2-1, \frac{\alpha}{2}}[$ , le test ne permet pas de rejeter  $H_0$ ,
- Si  $F_{obs} < F_{n_1-1, n_2-1, \frac{\alpha}{2}}$  ou  $F_{obs} > F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}$ , on rejette  $H_0$  au profit de  $H_1$  : les variances sont significativement différentes.

Dans le cas unilatéral  $H_1 : \sigma_1^2 > \sigma_2^2$ , on aura tendance à accepter  $H_1$  quand  $F_{obs}$  prend des valeurs grandes. On a donc l'alternative suivante

- Si  $F_{obs} < F_{n_1-1, n_2-1, 1-\alpha}$ , le test ne permet pas de rejeter  $H_0$ ,
- Si  $F_{obs} > F_{n_1-1, n_2-1, 1-\alpha}$ , on rejette  $H_0$  au profit de  $H_1$  : la variance  $\sigma_1^2$  est significativement plus grande que  $\sigma_2^2$ .

**Solution de l'exemple :**

Ici le choix d'un échantillon consiste à faire passer l'examen. La variable mesurée est la performance des étudiants de la classe  $A$  et de la classe  $B$  à ce type d'examens en général. On a un test unilatéral. Pour cadrer avec l'alternative que nous avons détaillée, nous appelons échantillon 1 les élèves de la classe  $B$  et échantillon 2 les élèves de la classe  $A$ . On a alors  $n_1 = 25$  et  $n_2 = 16$ . Il reste à calculer  $F_{obs}$ . Pour cela nous avons besoin de  $s_1^2$  et de  $s_2^2$  :

$$s_1^2 = \frac{25}{24}(12^2) = 150 \quad \text{et} \quad s_2^2 = \frac{16}{15}(9^2) = 86,4.$$

Alors  $F_{obs} = 150/86,4 = 1,74$  doit être comparé à  $F_{24,15;0,95} = 2,28$ , donc le test ne permet pas de rejeter  $H_0$  : la variance n'est pas significativement plus grande dans la classe  $B$ .

**Remarque :** On peut comme d'habitude calculer la probabilité de dépassement dans le cas d'un test unilatéral. On a

$$p = \mathbb{P}(F_{n_1-1, n_2-1} > F_{obs}).$$

Pour obtenir une valeur précise, on utilise un ordinateur. Dans le cas de l'exemple, on obtient une probabilité de dépassement égale à 0,13.

# Bibliographie

- A. Albert, Biostatistique, notes de cours ULG.
- P. Gérard, Travaux pratiques de statistiques, ULG, 1997.
- S.A. Glantz, Primer of Biostatistics, McGraw-Hill, 1997
- G. Kass, Notes du cours de mathématiques et statistiques, Univ. Luxembourg.
- N. Poncin, Eléments de biostatistique, Notes de cours, CUNLUX, 1995.
- S.M. Ross, Initiation aux probabilités, Presses polytechniques Romandes, 1987.
- M.R. Spiegel, Theory and problems of statistics, Schaum's outline series.

# Table des matières

<b>1</b>	<b>Probabilités</b>	<b>1</b>
1.1	Expériences aléatoires et espaces probabilisés . . . . .	1
1.1.1	Événements . . . . .	2
1.2	Probabilités : approche fréquentielle . . . . .	4
1.3	Probabilités : approche axiomatique . . . . .	5
1.3.1	Exercices . . . . .	6
1.4	Equiprobabilité et analyse combinatoire . . . . .	6
1.4.1	Quelques mots d'analyse combinatoire . . . . .	8
1.4.2	Exercices . . . . .	10
1.5	Probabilités conditionnelles . . . . .	11
1.5.1	Exercices . . . . .	14
1.6	Événements indépendants . . . . .	14
1.7	La formule des probabilités totales et le théorème de Bayes . . . . .	15
1.8	Variables aléatoires . . . . .	17
1.8.1	Définitions et notations . . . . .	18
1.9	Distribution de probabilité d'une variable discrète, espérance et variance . . . . .	18
1.9.1	Espérance mathématique . . . . .	19
1.9.2	Variance . . . . .	20
1.10	Un exemple fondamental en statistique . . . . .	21
1.11	Exemples supplémentaires, exercices . . . . .	23
1.12	Fonction de répartition . . . . .	25
1.13	Densité de probabilité d'une variable continue . . . . .	27
1.13.1	Exercices . . . . .	29
1.14	Variables aléatoires indépendantes . . . . .	30
1.15	Somme produit, combinaisons de variables aléatoires . . . . .	30
1.15.1	Les théorèmes . . . . .	31
1.15.2	Résultats supplémentaires . . . . .	32
<b>2</b>	<b>Quelques distributions et variables usuelles</b>	<b>33</b>
2.1	La loi uniforme discrète . . . . .	33
2.2	La loi uniforme continue . . . . .	33
2.3	Variables de Bernoulli . . . . .	34
2.4	La loi binomiale . . . . .	34
2.5	La loi hypergéométrique . . . . .	35
2.5.1	Exercices . . . . .	36
2.5.2	Solutions . . . . .	36
2.5.3	Théorème d'approximation . . . . .	37
2.5.4	Exercices supplémentaires . . . . .	37
2.6	La loi géométrique . . . . .	37



2.6.1	Exercices . . . . .	38
2.7	La loi de Poisson . . . . .	38
2.7.1	Exercices . . . . .	39
2.8	La loi normale, ou loi gaussienne . . . . .	40
2.8.1	Définition . . . . .	40
2.8.2	Normale centrée réduite . . . . .	40
2.8.3	Approximation de la loi Binomiale . . . . .	41
2.8.4	Exercices . . . . .	42
2.9	Exercices récapitulatifs . . . . .	42
2.9.1	Les énoncés . . . . .	42
2.9.2	Les solutions . . . . .	43
2.10	Quelques théorèmes . . . . .	45
2.10.1	L'inégalité de Bienaymé et Tchébychef . . . . .	45
2.10.2	La loi faible des grands nombres . . . . .	46
2.10.3	Le théorème central limite . . . . .	46
<b>3</b>	<b>Statistique inductive : estimation</b>	<b>48</b>
3.1	Introduction . . . . .	48
3.2	Exchantillon aléatoire simple . . . . .	49
3.3	Estimation ponctuelle . . . . .	50
3.3.1	Estimation ponctuelle d'une espérance . . . . .	50
3.3.2	Estimation ponctuelle d'une proportion/probabilité . . . . .	51
3.3.3	Estimation ponctuelle d'une variance . . . . .	51
3.4	Estimation par intervalle de confiance . . . . .	52
3.4.1	Intervalle de confiance pour une espérance . . . . .	52
3.4.2	Le cas des petits échantillons . . . . .	54
3.4.3	Intervalle de confiance pour une proportion/probabilité $\pi_0$ . . . . .	55
3.4.4	Intervalle de confiance pour une variance . . . . .	57
3.4.5	Exercices . . . . .	58
<b>4</b>	<b>Statistique inductive : les tests</b>	<b>60</b>
4.1	Introduction . . . . .	60
4.1.1	Pouvoir surnaturel ou pas? . . . . .	60
4.1.2	L'hypothèse nulle . . . . .	60
4.1.3	L'hypothèse alternative . . . . .	61
4.1.4	Le théorème . . . . .	61
4.1.5	Erreurs et risques . . . . .	62
4.1.6	La mise en oeuvre classique d'un test . . . . .	62
4.1.7	Le lien avec les $p$ -values . . . . .	63
4.2	Les tests de conformité . . . . .	64
4.2.1	Test de conformité pour une espérance, grands échantillons . . . . .	64
4.2.2	Test de conformité pour une espérance, petits échantillons . . . . .	66
4.3	Tests d'ajustement . . . . .	67
4.3.1	Le cas discret et non paramétrique . . . . .	67
4.3.2	Le cas discret, paramétrique . . . . .	68
4.3.3	Le cas continu . . . . .	69
4.4	Comparaison de deux populations . . . . .	70
4.4.1	Comparaison de deux espérances . . . . .	71
4.5	Exercices . . . . .	72
4.6	Comparaison de deux espérances, populations appariées . . . . .	73

4.6.1	Le cas des grands échantillons . . . . .	74
4.6.2	Le cas des petits échantillons . . . . .	74
4.6.3	Exercices . . . . .	75
4.7	Comparaison de deux variances . . . . .	75