Hasard, probabilités et statistique

Louis ESCH[1]

L'objectif de ce chapitre est tout d'abord de dresser rapidement le portrait de ce que l'on entend par *hasard, aléatoire, fortuit*, ensuite d'examiner quels sont les différents sens du mot *probabilité*, et enfin de voir comment ces deux éléments interviennent en *statistique inférentielle*. Il ne s'agit pas de développer les techniques du calcul des probabilités ou de la statistique mais de décrire le cadre et les structures de ces théories.

1. Le hasard

1.1. Sens commun

Plusieurs représentations du hasard existent pour le commun des mortels. Il y a tout d'abord le *destin* : il correspond à une volonté sous-jacente à notre vie de tout les jours qui conditionne celle-ci. C'est de cette représentation qu'il s'agit lorsque l'on parle de la « loi des séries » : des faits se produisent de manière apparemment chaotique mais, sous ce désordre, se dessine une semi-intention qui donnera lieu à l'événement final. Il apparaît également comme une circonstance atténuante des conséquences des décisions que nous prenons et des actions que nous entreprenons, « c'est la faute à pas de chance » : il s'agit ici d'un *refus de responsabilité*. Enfin – et on s'approche ici du sens utilisé dans l'étude scientifique du hasard – on représente le hasard par un ensemble de faits qui n'auraient *pas de cause* : on a ici l'idée du désordre, du chaos, se manifestant au mépris de toute loi.

Cette dernière représentation est caractérisée par le fait que l'événement fortuit qui survient
- paraît n'avoir aucune cause ;
- aurait pu survenir à n'importe quel autre moment, à n'importe quel autre personne, dans n'importe quelle autre circonstance ;
- n'est en aucune mesure prévisible.

Cette analyse s'applique à tout événement fortuit, heureux (gain à un jeu de hasard) ou malheureux (tuile qui tombe d'un toit et que vous recevez sur la tête).

Dans le cas de l'événement « malheureux », le hasard se double d'un sentiment d'injustice, d'intention négative. Qu'il s'agisse d'accidents de la route ou de maladie grave, la personne concernée ne se demande pas quelle était la probabilité de survenance de cet événement ou la proportion de personnes à qui cela peut arriver ; pour cette personne, le hasard a fait le choix de la mettre au nombre de ses victimes.

De manière plus générale, on a ici affaire à l'opposition entre l'*anarchie* et la *constance* : le hasard s'applique aux individus, pour chacun desquels l'événement survient ou ne survient pas, et une loi générale s'applique à la collectivité, pour une proportion donnée de laquelle l'événement survient. Sans le hasard individuel, il n'y aurait pas de constance collective.

[1] HEC-Ecole de Gestion de l'Université de Liège, rue Louvrex, 14 (N1), 4000 Liège, Belgique, louis.esch@ulg.ac.be

1.2. Point de vue philosophique

Deux points de vue s'affrontent relativement à la notion de hasard.

La première approche consiste à considérer que le hasard n'est que le reflet de notre ignorance. Cette manière de voir est confortée par la remarquable efficacité des théories physiques (mécanique newtonienne par exemple). Les lois gouvernant l'évolution d'un système étant connues et les conditions initiales étant fixées, il est théoriquement possible de ce point de vue de déterminer ladite évolution. Ce point de vue *déterministe* a eu de tout temps ses défenseurs, qu'il s'agisse d'Hippocrate[2], de Spinoza, de Kant ou d'Einstein[3].

Même si on admet que la causalité est essentielle – sinon comment faire confiance aux théories physiques efficaces ? – se pose la question, face à des situations aléatoires, de la cause initiale. Ce contre-argument a été utilisé par Aristote déjà. Un homme a mangé le soir des aliments salés ; il a soif et veut boire, mais il n'a plus d'eau en sa maison ; il se rend donc à la fontaine ; des brigands vivant dans la forêt voisine passent par là et le tuent. Quelle est la cause initiale de la mort de cet homme ? Certainement pas d'avoir mangé salé !

L'autre position, celle du *hasard objectif*, consiste à considérer certaines situations comme totalement contingentes : le hasard est simplement la rencontre (incausée) de séries causales indépendantes. Ainsi les séries causales physiologique (ingestion d'aliments salés → sensation de soif → besoin de boire → déplacement à la fontaine), géologico-météorologique (présence d'une source), sociologique (qui fait que des hommes sont devenus brigands, habitent la forêt et n'en sortent que la nuit) sont gouvernées par des lois, mais rien n'explique leur rencontre. Il ne s'agit donc pas d'une négation du déterminisme, mais du caractère imprévisible de la rencontre de séries causales. Ce point de vue est défendu par exemple par Aristote, Cournot ou Wittgenstein.

D'autres points de vue, entre ces positions extrêmes, peuvent être adoptés. C'est ainsi que Poincaré considère le hasard comme la limite de la situation « petites causes, grands effets ». Il l'illustre à l'aide de la théorie cinétique des gaz. Les lois gouvernant l'évolution de chaque molécule de gaz sont connues, mais lorsqu'on est face à énormément de tels systèmes individuels en relation les uns avec les autres, on ne peut qu'en déduire des lois de structure, en décomposant le gaz en cellules et en déterminant la probabilité des états macroscopiques, sans s'inquiéter des molécules individuelles.

1.3. Phénomènes fortuits

Un point de vue qui peut unifier les différents points de vue est de considérer que toutes les situations sont des *situations tendancielles* : à des conditions initiales fixées, sont associées un éventail plus ou moins larges de conséquences, de résultats possibles. L'éventail peut être très serré – voire totalement fermé – pour les cas où la prévisibilité est (presque) parfaite ; il peut être largement ouvert lorsque la prévisibilité n'est que partielle ou même totalement absente. Les *phénomènes fortuits* relèvent de cette seconde catégorie.

---

[2] « Le hasard, quand on vient à l'examiner, se trouve n'être rien. Tout ce qui se fait a une certaine cause, et cette cause se trouve en avoir une autre qui l'a produite ».

[3] « Dieu ne joue pas aux dés ».

Un fait isolé peut bien sûr être fortuit (c'est le cas de l'exemple cité par Aristote) ; mais son observation empirique ne permet pas de découvrir son caractère fortuit[4]. Si par contre on a affaire à une structure répétitive, l'observation empirique peut rendre compte de ce caractère fortuit. Dans ce second cas de figure, un phénomène fortuit est caractérisé par

- la complète *anarchie* des résultats individuels, aucune loi ne présidant à leur occurrence (sinon la Loterie Nationale n'existerait pas) ;
- la *cohérence statistique* de l'ensembles des résultats : stabilité des moyennes calculées sur un grand nombre de résultats successifs[5] (sinon les compagnies d'assurances n'existeraient pas).

Cette description est bien entendu valable pour des phénomènes fortuits constitués d'épreuves successives totalement identiques, mais également dans des cas plus généraux où des phénomènes de dépendance apparaissent (chaîne de Markov, mouvement brownien). Cependant, seul le premier cas d'espèce permet de considérer la définition fréquentiste de la probabilité (voir § 2.3).

1.4. Formalisation

Une théorie mathématique est par essence abstraite et indifférente à l'existence même de l'objet auquel elle s'applique(rait). La physique quant à elle baigne dans l'existentiel. L'étude d'une situation physique[6] à l'aide d'une théorie mathématique – que nous appellerons *théorie physique* – ne peut se réaliser que pour la partie quantifiable du phénomène et n'est jamais que la construction d'une image abstraite de cette partie quantifiable. Elle ne fournit donc pas une connaissance réelle, mais seulement symbolique du phénomène étudié. C'est cet aspect de représentation qui impose à toute théorie physique sa confrontation avec l'observation. En cas d'infirmation[7], la théorie doit être retouchée ou remplacée. C'est ainsi que, dans certaines conditions (vitesses non négligeables par rapport à celle de la lumière), la mécanique newtonienne doit être généralisée par la théorie de la relativité.

C'est à ce titre que la *théorie des probabilités* peut être considérée comme une théorie physique. Les phénomènes fortuits constituent l'objet de cette théorie, un phénomène fortuit étant caractérisé par l'ensemble Ω de ses résultats possibles. Un événement est, d'un point de vue intuitif, un fait qui, lors de chaque réalisation du phénomène fortuit, est susceptible de se produire ou de ne pas se produire et, d'un point de vue formel, une partie A de Ω telle que lors d'une réalisation du phénomène fortuit ayant pour résultat $\omega \in \Omega$, A se produit si et seulement si $\omega \in A$.

La théorie mathématique, applicable à tous les phénomènes fortuits, est basée sur la définition d'une fonction de probabilité Pr, d'une σ-algèbre[8] $\mathcal{F}$ sur Ω dans $[0 ; 1]$, obéissant aux

[4] Ceci est à l'origine des difficultés d'utilisation (et particulièrement de mise en nombre) de la théorie des probabilités dans les sciences humaines (par exemple décision en univers aléatoire en économie).

[5] Cette cohérence est justifiée, en théorie des probabilités par les lois des grands nombres.

[6] Au sens le plus large, sciences humaines comprises.

[7] Attention ! L'observation permet d'infirmer, jamais de confirmer une théorie physique.

[8] Il ne suffit pas de prendre $\mathcal{F} = \mathcal{P}(\Omega)$. En effet, dans certains cas non dénombrables on peut être conduit à des incohérences et, par ailleurs, la limitation à une σ-algèbre est parfois essentiel lorsque l'ensemble des événements observables doit « s'enrichir » au cours du temps ; c'est le cas par exemple en modélisation financière.

axiomes de Kolmogorov[9] et dont on peut déduire toute une série de conséquences, dont les lois des grands nombres.

Le lien entre théorie et expérience est fourni par la notion de propension à l'occurrence des événements, le modèle n'étant pas infirmé lorsque tout événement auquel il attribue une probabilité faible est un événement d'occurrence rare. Cette rareté ne peut bien sûr être constatée que si le phénomène est suffisamment répétitif.

2. Les sens du mot « probabilité »

2.1. Degré de croyance

Cette définition subjective de la probabilité s'applique à des propositions logiques auxquelles on attribue divers degrés de croyance. Il ne s'agit pas ici de l'expression d'une ignorance ou d'un doute, mais on considère qu'une affirmation ayant une certaine vraisemblance à propos de laquelle l'esprit n'atteint pas à la certitude. Nous noterons $\mathrm{Pr}_{\mathrm{DC}}$ cette probabilité.

On trouve ce point de vue chez différents auteurs : Platon, Aristote, Mill, de Morgan, Jevons. Nous donnerons ici la formalisation proposée par de Finetti[10]. Il suppose défini, sur un ensemble T de propositions logiques et pour un élément s de T, un préordre total : $(q,s) \prec (r,s)$ si et seulement si, étant donné s, la proposition r est estimée au moins aussi probable que la proposition q. La construction de la probabilité est faite grâce à un ensemble de paris : pour différentes valeurs de a et b[11], « vous gagnez a € si q est vrai et vous perdez b € si q est faux ». La personne confrontée à cette famille de paris acceptera ceux pour lesquels[12]

$$a \cdot \mathrm{Pr}_{\mathrm{DC}}(q,s) > b \cdot \mathrm{Pr}_{\mathrm{DC}}(\widetilde{q},s)$$

et refusera ceux pour lesquels on a l'inégalité inverse. En remplaçant $\mathrm{Pr}_{\mathrm{DC}}(\widetilde{q},s)$ par $1-\mathrm{Pr}_{\mathrm{DC}}(q,s)$ et en résolvant l'équation ci-dessus, on trouve

$$\mathrm{Pr}_{\mathrm{DC}}(q,s) > \frac{b}{a+b}.$$

Donc, si cette personne accepte tous les paris pour lesquels $a/b > x$ et refuse ceux pour lesquels $a/b < x$, c'est que, pour elle,

$$\mathrm{Pr}_{\mathrm{DC}}(q,s) = \frac{1}{1+x}.$$

[9] Contrairement aux autres théories physiques où une unité de mesure pour chaque grandeur est définie une fois pour toutes, un de ces axiomes, $\mathrm{Pr}(\Omega) = 1$, varie d'un phénomène fortuit à l'autre, l'ensemble Ω étant lui-même variable.

[10] DE FINETTI B., La prévision, ses lois logiques, ses sources subjectives, *Annales de l'Institut H. Poincaré*, VII/1, 1937.

[11] En fait, comme on va le voir, le résultat ne dépend que du rapport a/b.

[12] On désigne par $\widetilde{q}$ la négation de la proposition q.

Une objection qui peut être formulée à l'encontre de cette construction est la suivante : dans la réalité, un pari (a, b) peut être accepté et le pari (a', b') refusé par une personne, alors que $a/b = a'/b'$. En effet, il se peut qu'une personne accepte de parier 10 € contre 5 € à propos de la véracité d'une proposition, mais refuse de parier 10 000 € contre 5 000 € ; le concept d'utilité doit alors être introduit.

D'autres objections plus fondamentales sont, d'une part, que cette théorie subjectiviste ne peut pas être confrontée à l'expérience (dans la réalité, un événement survient ou ne survient pas) et, d'autre part, qu'elle ne permet pas d'expliquer des phénomènes comme la cohérence statistique sur un grand nombre de réalisations.

2.2. Inférence partielle

Cette notion s'applique, tout comme la précédente, à des propositions logiques formant un ensemble T. Ici cependant, la probabilité n'est pas une attitude de l'esprit, mais une relation logique objective (pas de variation d'un sujet à l'autre) qui mesure une proximité logique, une connexion déductive liant deux propositions. On notera cette probabilité $\Pr_{\mathrm{IP}}$, l'expression $\Pr_{\mathrm{IP}}(q, r)$ représentant la probabilité que r implique q.

Cette acception est décrite par Keynes, Wittgenstein, Carnap et Jeffreys. Ce dernier[13] construit cette probabilité à partir d'un préordre total sur l'ensemble des couples de propositions en montrant qu'il est possible de définir $\Pr_{\mathrm{IP}}(q, r)$ comme un nombre réel compris entre 0 et 1 vérifiant

$$\begin{gathered}
\text{si } r \Rightarrow q, \text{ alors } \Pr_{\mathrm{IP}}(q,r) = 1 \\
\text{si } r \Rightarrow \widetilde{q}, \text{ alors } \Pr_{\mathrm{IP}}(q,r) = 0 \\
\text{si } r \,\&\, q \Rightarrow s, \text{ alors } \Pr_{\mathrm{IP}}(q \,\&\, s, r) = \Pr_{\mathrm{IP}}(q,r) \\
\Pr_{\mathrm{IP}}(q \,\&\, s, r) = \Pr_{\mathrm{IP}}(q,r) \cdot \Pr_{\mathrm{IP}}(s, q \,\&\, r)
\end{gathered}$$

et qui peut s'interpréter ainsi : entre les deux valeurs extrêmes 0 et 1, correspondant respectivement à la contradiction et à la déductibilité, $\Pr_{\mathrm{IP}}(q, r)$ est d'autant plus proche de 1 que le contenu de q ajoute moins à celui de r.

Cette probabilité, dont la construction formelle est irréprochable, est peu utilisable en raison de la difficulté de définir en pratique le préordre « est au moins aussi probable que ».

2.3. Fréquence

Cette notion, notée $\Pr_{\mathrm{F}}$, s'applique aux événements d'un phénomène fortuit à nombre fini de résultats possibles : $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. La probabilité d'un événement A est ici définie comme étant la proportion de résultats constituant l'événement A parmi l'ensemble de tous les résultats :

$$\Pr_{\mathrm{F}}(A) = \frac{\#(A)}{n}$$

ou, si l'on veut définir une fonction à deux arguments comme dans les cas précédents,

[13] JEFFREYS H., *Theory of Probability*, Oxford University Press, 1939.

$$\Pr_{\mathrm{F}}(A,B)=\begin{cases}\dfrac{\#(A\cap B)}{\#(B)} & \text{si} \quad \#(B)\neq 0\\ 1 & \text{sinon}\end{cases}$$

et qui rend la définition à un seul argument dans le cas particulier où $B = \Omega$.

On a affaire ici à la définition – qu'on trouve chez de Moivre, Montmort et Laplace – exprimée par la sempiternelle formule « quotient du nombre de cas favorables par le nombre de cas possibles » qui, dans l'enseignement secondaire, limite l'étude des probabilités aux seuls modèles finis équiprobables, quand elle ne laisse pas croire que la théorie des probabilités est simplement une application de l'analyse combinatoire !

Bien entendu, pour lesdits modèles finis équiprobables, la théorie basée sur cette définition convient et est efficace, mais elle limite son utilisation à des applications quelque peu artificielles telles que certains jeux de hasard. Elle ne permet pas de traiter les situations discrètes non équiprobables ou dénombrables et encore moins les situations continues.

Une généralisation de cette approche fréquentiste a été donnée par von Mises[14] qui traduit l'idée que la théorie des probabilités a pour objet des phénomènes répétitifs et qu'une probabilité se mesure comme une fréquence sur une suite suffisamment longue de réalisations. Les difficultés techniques de construction de la fonction de probabilité dans ce contexte sont considérables et conduisent à la théorie des collectifs, cohérente mais pas élémentaire du tout.

2.4. Propension

Cette dernière approche s'applique aux événements d'un phénomène fortuit (voir §§ 1.3 et 1.4) et représente la tendance qu'a un événement à se produire, sa *propension à l'occurrence*. Cette notion ne nécessite nullement la répétitivité de la situation aléatoire ; elle est a rapprocher d'une notion telle que la fragilité d'un objet en cristal. Il n'est pas nécessaire de donner un choc à cet objet ni de le briser pour définir sa fragilité, c'est une notion qui lui est intimement associée ; il en va de même pour la propension à l'occurrence des événements d'un phénomène fortuit. Cette probabilité sera notée $\Pr_{\mathrm{P}}$.

Ce point de vue plus moderne apparaît chez Fréchet, Cramér, Popper ainsi qu'au sein de l'école russe (Chebyshev, Markov, Lyapounov, …) et il a été formalisé par Kolmogorov[15] par l'axiomatique que voici.

Etant donné un phénomène fortuit caractérisé par un ensemble Ω de résultats possibles et une σ-algèbre $\mathcal{F}$ sur Ω correspondant aux événements observables, on définit la probabilité comme une mesure de la tendance qu'ont les événements à ce produire. Il s'agit de la fonction $\Pr_{\mathrm{P}}$ de $\mathcal{F}$ dans $[0 ; 1]$ qui obéit aux axiomes suivants : $\Pr_{\mathrm{P}}(\Omega) = 1$ et si $\{A_j : j \in J\}$ est un ensemble d'éléments de $\mathcal{F}$ deux à deux disjoints, alors

[14] VON MISES R., *Probability, Statistics and Truth*, Dover, 1981.
[15] KOLMOGOROV A.N., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, 1933.

$$\mathrm{Pr}_{\mathrm{P}}\left(\bigcup_{j\in J} A_j\right) = \sum_{j\in J} \mathrm{Pr}_{\mathrm{P}}(A_j)\,.$$

A nouveau, si on veut une définition à deux arguments, on définit à partir de là

$$\mathrm{Pr}_{\mathrm{P}}(A,B) = \begin{cases} \dfrac{\mathrm{Pr}(A\cap B)}{\mathrm{Pr}(B)} & \text{si} \quad \mathrm{Pr}(B)\neq 0 \\ 1 & \text{sinon} \end{cases}$$

et le cas particulier $B=\Omega$ permet de retrouver la formulation à un argument.

Dans cette approche, le triplet $(\Omega, \mathcal{F}, \mathrm{Pr}_{\mathrm{P}})$ est appelé *espace probabilisé*.

2.5. Unification

Popper[16] a proposé un formalisme abstrait général qui s'applique aux différents sens rencontrés. Les éléments en sont

- un ensemble S ;
- une application de S dans S : $a \mapsto \sim a$;
- une application de $S\times S$ dans S : $(a,b)\mapsto a\otimes b$
- une application de $S\times S$ dans $\mathbb{R}$: $(a,b)\mapsto p(a,b)$

liés par les axiomes

$$\begin{aligned}
&\forall a,b\in S,\\
&\qquad \exists c,d\in S : p(a,b)\neq p(c,d)\\
&\qquad [\forall c\in S,\ p(a,c)=p(b,c)] \Rightarrow [\forall d\in S,\ p(d,a)=p(d,b)]\\
&\qquad p(a,a)=p(b,b)\\
&\forall a,b,c\in S,\\
&\qquad p(a\otimes b,c)\leq p(a,c)\\
&\qquad p(a\otimes b,c)=p(a,b\otimes c)\cdot p(b,c)\\
&\forall a,b\in S,\\
&\qquad [\exists c\in S : p(b,b)\neq p(c,b)] \Rightarrow p(a,b)+p(\sim a,b)=p(b,b)
\end{aligned}$$

Les conséquences de ces axiomes sont trop nombreuses pour être toutes citées ici. Signalons simplement que, dans les arguments de la fonction p, l'opération $\otimes$ est commutative et associative, que

$$\begin{aligned}
&\exists a\in S : p(\sim a,a)=0\\
&\exists a,b\in S : p(a,b)\neq 1
\end{aligned}$$

et que, quels que soient a, b, c et d dans S, on a

[16] POPPER K., *The Logic of Scientific Discovery*, Hutchinson, 1959.

$$0 \leq p(a,b) \leq 1 \qquad\qquad p(a,a) = 1$$
$$p\big(\sim(\sim a),b\big) = p(a,b)$$
$$p(a,b) + p(\sim a,b) = 1 + p(\sim b,b)$$
$$p(a \otimes b,c) + p\big((\sim a) \otimes b,c)\big) = p(b,c) + p(\sim c,c)$$

Par ailleurs, si on définit $a \oplus b$ par $\sim\big((\sim a) \otimes (\sim b)\big)$, alors

$$p\big(a \oplus (\sim a),b\big) = 1$$
$$p(a \oplus b,c) = p(a,c) + p(b,c) - p(a \otimes b,c)$$
$$p\big(a \otimes (b \oplus c),d\big) = p\big((a \otimes b) \oplus (a \otimes c),d\big)$$

Maintenant, si on définit $p_0(a)$ par $p\big(a,a \oplus (\sim a)\big)$, on peut montrer que

$$p_0(b) \neq 0 \;\Rightarrow\; \begin{cases} p(a,b) = \dfrac{p_0(a,b)}{p_0(b)} \\ p(\sim b,b) = 0 \end{cases}$$

Le nombre $p(a, b)$ est la probabilité de a si b, et le nombre $p_0(a)$ est la probabilité de a par rapport à $(S, \sim, \otimes)$. Les relations ci-dessus définissent un calcul des probabilités.

Les définitions de la probabilité comme degré de croyance, inférence partielle, fréquence ou propension entrent toutes dans ce moule (on a noté $\overline{a}$ le complémentaire de l'ensemble a) :

	S	$\sim a$	$a \otimes b$	$p(a, b)$
degré de croyance	T	$\tilde{a}$	a & b	$\mathrm{Pr}_{\mathrm{DC}}(a, b)$
inférence partielle	T	$\tilde{a}$	a & b	$\mathrm{Pr}_{\mathrm{IP}}(a, b)$
fréquence	$\mathcal{P}(\Omega)$	$\overline{a}$	$a \cap b$	$\mathrm{Pr}_{\mathrm{F}}(a, b)$
propension	$\mathcal{F}$	$\overline{a}$	$a \cap b$	$\mathrm{Pr}_{\mathrm{P}}(a, b)$

Bien entendu, l'opération $a \oplus b$ correspond, dans les deux premiers cas, à « a ou b » et, dans les deux autres, à $a \cup b$.

L'interprétation qui sera adoptée dans la suite du texte est celle des propensions et on notera, plus classiquement,

$$\mathrm{Pr}_{\mathrm{P}}(A) = \mathrm{Pr}(A) \qquad\qquad \mathrm{Pr}_{\mathrm{P}}(A, B) = \mathrm{Pr}(A \mid B)$$

3. La statistique inférentielle

3.1. Schéma général

Dans les applications simples de la théorie des probabilités (nombre d'as obtenu lors d'un nombre donné de lancers indépendants d'un dé par exemple), la situation se décrit simplement et complètement et le formalisme de la théorie des probabilités permet de répondre à des questions relatives à cette situation. Dans des situations réelles plus complexes (analyse et gestion de la problématique des files d'attente aux caisses d'une grande surface), après avoir décrit la situation – ce qui n'est pas toujours simple – tout ce qu'on peut proposer, c'est une famille de modèles plausibles, dont un est plus adéquat que les autres mais qui est inconnu. Nous nous limiterons, dans ce qui suit, où tous les modèles ont la même forme mais varient de l'un à l'autre en fonction d'un paramètre ; on parle alors de *statistique paramétrique*.

L'objectif de la *statistique inférentielle* peut prendre diverses formes. Sur base d'une ou plusieurs observation(s), on peut

- choisir une action à exécuter (problème de décision) ;
- préciser la valeur ou un ensemble de valeurs pour le paramètre adéquat (problème d'estimation) ;
- exprimer un avis sur une affirmation relative à la situation (problème de test d'hypothèse).

Le schéma général est le suivant. Il comprend un ensemble Θ des valeurs paramétriques possibles θ et, pour chacune d'entre elles, un espace probabilisé $(\Omega, \mathcal{F}, \Pr_\theta)$ pour lequel seule la fonction de probabilité dépend du paramètre[17]. Il est toujours admis qu'il existe une valeur paramétrique $\theta_0 \in \Theta$ pour laquelle le modèle est adéquat. Cette valeur θ_0 est appelée *état du monde* ; les conclusions, quelles qu'elles soient, n'en font jamais état, l'état du monde étant par définition inconnu.

Le schéma général que nous venons de décrire $(\Omega, \mathcal{F}, \Theta, \Pr_\theta)$ sera, suivant le problème à résoudre, complété ou particularisé.

3.2. Trois problèmes en statistique inférentielle

Problème de décision

Dans un problème de décision, le schéma général ci-dessus est complété par deux ensembles Δ et Γ et par deux fonctions γ et σ. L'ensemble Δ est l'ensemble des *décisions* possibles et Γ est l'ensemble des *conséquences* ; ce dernier est muni d'une structure de préordre total, permettant au décideur d'exprimer ses *préférences* quant auxdites conséquences. La fonction

$$\gamma : \left| \begin{array}{l} \Theta \times \Delta \to \Gamma \\ (\theta, \delta) \mapsto \gamma(\theta, \delta) \end{array} \right.$$

fournit les différentes valeurs de $\gamma(\theta, \delta)$, conséquence de la décision δ si l'état du monde est θ. L'objectif d'un problème de décision est de trouver la fonction

$$\sigma : \left| \begin{array}{l} \Omega \to \Delta \\ \omega \mapsto \sigma(\omega) \end{array} \right.$$

appelée *stratégie*, pour laquelle la fonction de conséquence

[17] Ce paramètre peut par exemple être une moyenne de population d'individus, ou une variance, ou le couple formé par ces deux éléments, …

$$\left| \begin{array}{l} \Omega \times \Theta \to \Gamma \\ (\omega,\theta) \mapsto \gamma(\theta,\sigma(\omega)) \end{array} \right.$$

a des propriétés globalement optimales (c'est-à-dire pour l'ensemble des valeurs possibles de ω).

Un exemple simple est celui où un client a le choix entre deux lots de N pièces, un à un prix élevé mais de qualité parfaite (ou garantie) et l'autre moins cher mais pour lequel chaque pièce a une probabilité θ d'être défectueuse. Le schéma général $(\Omega, \mathcal{F}, \Theta, \Pr_\theta)$ est celui de l'échantillonnage de n pièces que l'on teste, l'ensemble Δ comporte deux éléments (chacune des décisions possibles), la fonction $\gamma(\delta, \theta)$ représente, pour chacune des décisions, le coût (achat et rebut estimé) en fonction de la probabilité de défaut θ. On peut donc déduire de l'étude de cette fonction les valeurs de θ pour lesquelles l'une ou l'autre des décisions est meilleure et en déduire une stratégie de choix de la décision.

Dans certains cas, les décisions sont multiples et échelonnées dans le temps ; dans ce cas, au lieu d'utiliser simplement la matrice $\gamma(\delta, \theta)$ des conséquences, il convient de représenter la situation par un *arbre de décision*, aux nœuds duquel se succèderont des manifestation du hasard et des décision à prendre, toujours suivant un principe d'optimalisation globale des conséquences. La stratégie retenue sera, dans ce cas, une succession de décisions.

Problème d'estimation

L'objectif d'un problème d'estimation est de donner des informations sur l'état du monde. Une *région de confiance* pour le paramètre[18] θ au *niveau d'incertitude* α est un sous-ensemble aléatoire R de Θ pour lequel

$$\inf_{\theta\in\Theta} \Pr_\theta[\theta \in R] = 1 - \alpha .$$

Pour toutes les valeurs de θ, la probabilité $\Pr_\theta[\theta \in R]$ est donc au moins égale à $1 - \alpha$. Pour une réalisation du phénomène fortuit (résultat ω), l'ensemble (non aléatoire) $R(\omega)$ est la réalisation observée de la région de confiance.

Le niveau d'incertitude α doit bien sûr être suffisamment petit pour caractériser un événement (affirmer que l'état du monde est un élément de R alors qu'il n'en est rien) rare. Bien entendu, pour tout ω, $R(\omega)$ est d'autant plus grand que α est petit ; c'est l'expression technique du conflit entre précision (taille de la région de confiance) et sécurité (niveau d'incertitude).

Lorsque $\Theta \subset \mathbb{R}$, la notion de région de confiance prend presque toujours la forme d'un *intervalle de confiance* ; il s'agit d'un intervalle aléatoire $[u ; v]$ tel que

$$\inf_{\theta\in\Theta} \Pr_\theta[u \le \theta \le v] = 1 - \alpha .$$

[18] Si les éléments de Θ sont des couples (ou des triplets) de paramètres, la notion de région de confiance peut s'appliquer au couple (ou au triplet), mais également aux composantes individuelles.

Problème de test d'hypothèse

On appelle *hypothèse* d'un schéma général $(\Omega, \mathcal{F}, \Theta, \Pr_\theta)$ toute proposition H_0 de la forme « $\theta \in \Theta_0$ », où $\Theta_0 \subset \Theta$. Si l'ensemble Θ_0 ne comporte qu'un seul élément, l'hypothèse est dite *simple*, sinon, elle est dite *composite*. La question qui se pose dans un problème de test d'hypothèse est de savoir dans quelle mesure le résultat observé ω de la réalisation du phénomène fortuit infirme ou confirme[19] l'hypothèse H_0.

Dans l'approche de Neyman-Pearson, le problème du test d'hypothèse apparaît comme un cas particulier d'un problème de décision avec $\Delta = \{\delta_1, \delta_2\}$ où δ_1 (resp. δ_2) correspond au rejet (resp. au non rejet) de H_0 et où la stratégie σ est définie à partir de la notion de *région critique* Z_0, partie de Ω telle que

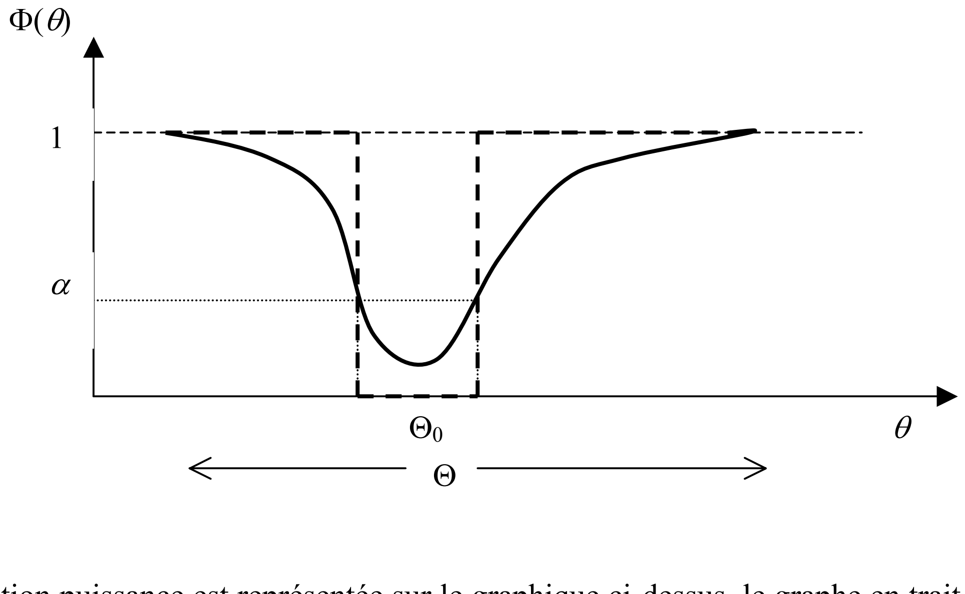
$$\sigma(\omega) = \begin{cases} \delta_1 & \text{si} \quad \omega \in Z_0 \\ \delta_2 & \text{si} \quad \omega \in \Omega \setminus Z_0 \end{cases}$$

Le *niveau d'incertitude* du test est le nombre

$$\alpha = \sup_{\theta \in \Theta_0} \Pr_\theta(Z_0)$$

et la *puissance* du test est la fonction

$$\Phi : \left| \begin{array}{l} \Theta \to [0;1] \\ \theta \mapsto \Pr_\theta(Z_0) \end{array} \right.$$



La fonction puissance est représentée sur le graphique ci-dessus, le graphe en trait interrompu étant la finesse « idéale ».

[19] On devrait plutôt dire « infirme ou n'infirme pas », une observation isolée permettant de rejeter une affirmation, pas de la prouver.

Le niveau d'incertitude α représente donc la probabilité de rejeter l'hypothèse H_0 si celle-ci est vraie. Cette erreur est appelée erreur de type I. L'erreur de type II est de ne pas rejeter H_0 si celle-ci est fausse ; sa probabilité est

$$1 - \inf_{\theta \in \Theta \setminus \Theta_0} \Pr_\theta(Z_0).$$

Lorsque l'on est amené à rejeter H_0, on peut être intéressé de connaître l'importance de l'écart. Cet aspect s'appréhende à l'aide de la notion de *probabilité de dépassement*. Définissons sur Ω le préordre total $\prec$ par : $\omega_1 \prec \omega_2$ si et seulement si ω_1 est au moins aussi défavorable à H_0 que ω_2. Définissons ensuite, pour chaque $\omega \in \Omega$, l'événement (que nous supposerons observable)

$$W_0(\omega) = \{t \in \Omega \ : \ t \prec \omega\}.$$

La probabilité de dépassement est la variable aléatoire PD_0, définie par

$$PD_0(\omega) = \sup_{\theta \in \Theta_0} \Pr_\theta\left(W_0(\omega)\right).$$

Il est clair que, quel que soient ω_1 et ω_2, on a

$$W_0(\omega_1) \subset W_0(\omega_2) \quad \text{ou} \quad W_0(\omega_2) \subset W_0(\omega_1)$$
$$W_0(\omega_1) \subset W_0(\omega_2) \quad \Leftrightarrow \quad PD_0(\omega_1) \le PD_0(\omega_2)$$

et que, pour l'implication $\Leftarrow$ ci-dessus, une inégalité stricte entre les probabilités de dépassement implique une inclusion stricte entre les événements.

Etant donné une observation (résultat ω observé lors d'une réalisation du phénomène fortuit), on peut dire que, si l'hypothèse H_0 est vraie, alors les observations au moins aussi défavorables à H_0 que l'observation ω ont une probabilité au plus égale à $PD_0(\omega)$.

3.3. Utilisation pratique

Dans la pratique, presque tous les problèmes de statistique inférentielle concernent une *population* de N individus, dont on extrait un *échantillon* à partir duquel on tente d'inférer des conclusions s'appliquant à la population.

Population

Si chaque individu de la population est porteur d'un (ou de plusieurs) nombre(s), la population est dite univariée (ou multivariée). Si les différents individus sont affectés chacun à une et une seule catégorie, on parle de population nominale.

Pour une population univariée l'ensemble des nombres $\{x_1, x_2, \ldots, x_N\}$ peut faire l'objet de la définition d'une *ogive*, fonction F de $\mathbb{R}$ dans $[0\,;\,1]$ telle que $F(t)$ représente la proportion des individus de la population porteur d'un nombre inférieur ou égal à t. A partir de cette ogive, on peut également définir différents paramètres statistiques de population tels que moyenne, écart type, … Pour une population nominale, les paramètres statistiques de population sont simplement les proportions $p_1, p_2, \ldots, p_m$ d'individus relevant de chacune des catégories.

La statistique inférentielle s'applique uniquement aux populations dont les *paramètres statistiques de population* sont (partiellement) inconnus, l'objectif étant de suppléer à cette absence de connaissance.

Echantillonnage

Nous nous limiterons ici à la notion d'*échantillon aléatoire simple*, c'est-à-dire pour lequel le nombre n d'individus extraits est fixé, les extractions se faisant de la même population, de manière aléatoire et indépendante.

Le phénomène fortuit $(\Omega, \mathcal{F})$ dont on parle dans le schéma général est ce mécanisme d'échantillonnage, l'ensemble Ω étant l'ensemble de tous les échantillons possibles.

La population n'est pas totalement connue. Nous supposerons que cette indétermination est de type paramétrique : il manque la connaissance de la valeur paramétrique[20] θ, prenant ses valeurs dans l'ensemble Θ. C'est la troisième composante du schéma général $(\Omega, \mathcal{F}, \Theta, \Pr_\theta)$.

Enfin, la loi de probabilité est définie, pour une population univariée, en considérant les éléments de l'échantillon comme n variables aléatoires indépendantes et identiquement distribuées de commune fonction de répartition donnée par l'ogive de la population et, pour une population nominale, si on note C_j les catégories et e_i les individus de l'échantillon, par

$$\Pr\left(\bigcap_{i=1}^{n}\left[e_i \in C_{j_i}\right]\right) = \prod_{i=1}^{n} p_{j_i},$$

quels que soient $j_1, \ldots, j_n \in \{1, \ldots, m\}$.

Pour chaque échantillon, il est possible de calculer des *paramètres statistiques d'échantillon* : moyenne, écart type, … pour une population univariée, fréquence des catégories pour une population nominale[21]. Si on considère ces paramètres statistiques d'échantillon calculés pour chaque échantillon possible (chaque élément de Ω), ils définissent des variables aléatoires dans le schéma général. Pour ces variables aléatoires, on peut définir des valeurs typiques (paramètres de localisation, de dispersion, …). Ce sont les *paramètres statistiques d'échantillonnage*.

Clairement, les paramètres statistiques d'échantillon et d'échantillonnage sont liés à ceux de la population[22]. C'est ainsi qu'il est possible de démontrer, pour une population univariée, que la moyenne d'échantillonnage de la moyenne d'échantillon est égale à la moyenne de population.

Estimateurs

---

[20] Eventuellement multivariée.
[21] Ceci relève de la statistique descriptive.
[22] Mais il convient de soigneusement les distinguer !

Les problèmes rencontrés plus haut

- définir d'une stratégie de choix de décision en fonction des conséquences liées aux différentes valeurs d'un paramètre θ ;
- estimer un paramètre θ ;
- tester une hypothèse relative à la valeur d'un paramètre θ ;

utilisent la notion d'*estimateur* du paramètre θ. Il s'agit d'une variable aléatoire $\hat{\theta}$ sur le schéma général (paramètre statistique d'échantillon) dont un paramètre de localisation (d'échantillonnage) est égal au paramètre (de population) θ.

Un estimateur est dit *sans biais* si sa moyenne d'échantillonnage est égale à la valeur du paramètre de population $E(\hat{\theta}) = \theta$ et *efficace* si sa variance est inférieure à celle de tous les autres estimateurs. Pour une population univariée de moyenne μ et d'écart type σ dont les éléments de l'échantillon sont les variables aléatoires $X_1, \ldots, X_n$, si on définit moyenne et variance d'échantillon par

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad\qquad s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$

on a affaire, pour le premier, à un estimateur sans biais car, comme on l'a déjà dit, $E(m) = \mu$, alors que le second n'est qu'asymptotiquement sans biais puisqu'on peut démontrer que

$$E(s^2) = \frac{n-1}{n}\sigma^2 .$$

3.4. Et ensuite ?

Cette troisième section est le cadre conceptuel et le point de départ de tous les développements de la statistique inférentielle paramétrique, qu'il s'agisse des méthodes de construction des estimateurs (maximum de vraisemblance, khi carré minimum, …) ou du développement de méthodes d'estimation et de test dans des cas divers et variés. Cet important chapitre technique n'a évidemment pas sa place ici.

Il convient de signaler que ce qui vient d'être exposé ne se veut pas complet, bien loin de là. On n'a en rien abordé les aspects non paramétriques, les techniques robustes, les méthodes bayesiennes, … On s'est également limité volontairement aux aspects classiques des fondements du calcul des probabilités et de la statistique ; par exemple, à côté de l'approche de Neyman-Pearson des tests, on aurait pu envisager celles qui utilisent les modèles probabilistes relatifs aux propositions logiques $\Pr_{DC}$ et $\Pr_{IP}$.