



François Dress

# Les PROBABILITÉS et la STATISTIQUE

de

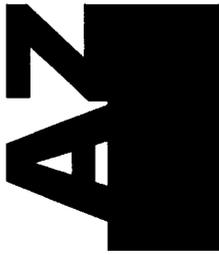
A

à

Z

500 DÉFINITIONS, FORMULES ET TESTS D'HYPOTHÈSE

DUNOD



## AVERTISSEMENT

Ce dictionnaire est un dictionnaire généraliste principalement destiné aux étudiants des premières années d'université et aux utilisateurs professionnels non mathématiciens.

Une attention toute particulière a été portée aux étudiants et aux utilisateurs dans les domaines des sciences expérimentales et des sciences économiques et sociales. Dans le souci de leur fournir un outil de travail adapté, les articles ont été rédigés en restant au niveau mathématique le plus élémentaire possible. En outre, chaque fois que cela apparaît nécessaire, la définition est précédée par une courte introduction en langage « courant ».

Certains termes et concepts mathématiques de base, en nombre très restreint et dont la vertu est de clarifier les définitions et d'éviter les ambiguïtés, sont néanmoins utilisés. En particulier, on se réfère souvent à l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ , notion qui se comprend facilement :  $\Omega$  est l'ensemble de tous les résultats possibles,  $\mathcal{A}$  est l'ensemble de tous les événements auxquels on pourra attribuer une probabilité, et  $P$  est cette mesure de probabilité, dont les valeurs sont comprises entre 0 et 1... le niveau d'abstraction n'est pas insupportable !

Toujours dans le souci de la meilleure efficacité, ce dictionnaire a un parti pris de redondance : redondance de contenu entre les articles, et souvent redondance du vocabulaire à l'intérieur de chaque article.

Enfin, ce dictionnaire inclut un certain nombre d'articles, ou de parties d'articles, dont on pourrait qualifier le niveau mathématique d'intermédiaire, qui lui permettront d'être également utile aux étudiants de mathématiques des premières années.

Nous avons choisi de garder pour certains concepts le vocabulaire usuel même lorsqu'il n'est pas très bien choisi et représente un héritage malheureux des siècles précédents (la « variable » aléatoire !). En outre la terminologie n'est pas complètement fixée. Nous avons systématiquement indiqué les variantes les plus courantes. En tout état de cause, si l'utilisateur trouve insolite l'utilisation d'un mot dans tel ou tel manuel, il ne doit pas se laisser perturber, mais chercher simplement quel sens précis a ce mot dans ce manuel. La même remarque peut être faite pour les notations. Seule exception à notre ouverture aux notations variées, nous noterons  $P(A)$  la probabilité de l'événement  $A$ , et n'utiliserons jamais  $\text{Pr}(A)$  ou  $\text{Prob}(A)$ .

Un détail enfin : en règle générale, toutes les valeurs numériques qui sont données dans les exemples sont des valeurs approchées (par arrondi à la décimale la plus proche) sans que cela soit marqué par la (trop lourde) présence de points de suspension. Bien entendu, une valeur approchée peut aussi être exacte, et il est laissé au lecteur le soin de le détecter.

Quelques conseils matériels d'utilisation

- Lorsqu'une expression contient plusieurs mots, l'article correspondant est répertorié au mot le plus significatif (mais de nombreux renvois facilitent la recherche).
- Les crochets dans une entrée d'article ou un synonyme encadrent une partie optionnelle de l'expression.
- Les mots du texte qui renvoient à une autre entrée sont mis en italiques.
- Enfin, quelques rares expressions n'ont aucun équivalent dans l'usage anglo-saxon (par exemple « épreuves répétées »), elles ne sont donc pas traduites.

François Dress  
dress@math.u-bordeaux.fr





**absolument continue (variable aléatoire)** (*absolutely continuous random variable*)

Voir *variable aléatoire (typologie)*.

**acceptation (région d')** (*acceptance region*)

Voir *région d'acceptation*.

**additivité** (*additivity*)

Soit un espace probabilisable  $(\Omega, \mathcal{A})$ . On s'intéresse aux applications définies sur l'ensemble  $\mathcal{A}$  des évènements et à valeurs réelles positives, et on veut caractériser la propriété qui fera d'elles des mesures (en langage mathématique général) ou des mesures de probabilité alias probabilités (en langage spécifique au calcul des probabilités).

À un niveau élémentaire, on se limite à la réunion finie, avec deux formules qui sont d'un usage extrêmement fréquent ( $P$  est une mesure de probabilité) :

$$\begin{aligned} \text{si } A \text{ et } B \text{ sont disjoints : } & P(A \cup B) = P(A) + P(B) \\ \text{de façon générale : } & P(A \cup B) = P(A) + P(B) - P(A \cap B) \end{aligned}$$

**Exemple** On considère une population humaine  $\mathcal{P}$  et le système ABO de groupes sanguins. Les probabilités qu'un individu tiré au hasard dans la population  $\mathcal{P}$  présente l'un des 6 génotypes possibles dans ce système sont données par le tableau ci-dessous :

OO	OA	AA	OB	BB	AB
0,435	0,364	0,076	0,086	0,004	0,035

Si l'on examine maintenant les phénotypes, l'évènement « phénotype A » est constitué par la réunion des évènements « OA » et « AA », qui sont disjoints, et sa probabilité est donc  $P(OA) + P(AA) = 0,440$ .

À un niveau plus approfondi, il faut envisager la réunion infinie dénombrable, pour laquelle l'axiome de  $\sigma$ -additivité énoncé ci-dessous doit être vérifié.

On dit qu'une application  $\mu : \mathcal{A} \rightarrow [0, +\infty[$  est  $\sigma$ -additive si, pour toute suite  $(A_n) (n = 1, 2, \dots)$  d'éléments de  $\mathcal{A}$ , deux à deux disjoints, on a :

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

On notera une conséquence importante : si les  $A_n$  forment une suite croissante (au sens de la théorie des ensembles, i.e. si  $A_n \subset A_{n+1}$ ), alors  $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mu(A_n)$ .

Voir *espace probabilisé, mesure de probabilité, Poincaré (formule de)*.

**affine (fonction)***(affine function)*

Fonction de la forme  $y = a + bx$  que l'on rencontre notamment en statistique comme courbe de régression. Une telle fonction est souvent appelée « linéaire » par abus de langage.

**ajustement linéaire, polynomial, exponentiel, logarithmique***(fitting)*

On considère deux grandeurs numériques  $x$  et  $y$  intervenant dans l'étude d'un phénomène. On suppose qu'il existe une liaison entre ces deux grandeurs, et que cette liaison est aléatoire (qu'il s'agisse d'un phénomène déterministe perturbé par exemple par des erreurs de mesure, ou d'un phénomène intrinsèquement aléatoire). Le plus souvent, cette liaison sera représentée – ou sera présumée pouvoir être représentée – par une relation mathématique simple, par exemple :  $y = bx$  (liaison linéaire),  $y = a + bx$  (liaison affine, le plus souvent qualifiée aussi de linéaire),  $y = bx^p$  (liaison puissance),  $y = a_0 + a_1x + \dots + a_nx^n$  (liaison polynomiale),  $y = ae^{kx}$  (liaison exponentielle),  $y = a + b \ln x$  (liaison logarithmique).

Le problème se pose alors, connaissant un ensemble de points expérimentaux (un « nuage », dans le vocabulaire statistique), de trouver la forme mathématique de la liaison et les « bons » coefficients (ou paramètres) numériques. Lorsque le phénomène sera déterministe perturbé aléatoirement, le nuage de points sera très voisin de la courbe  $y = f(x)$  que l'on cherche à déterminer. Lorsque le phénomène sera intrinsèquement aléatoire, le nuage sera plus dispersé et la courbe  $y = f(x)$  sera la courbe qui passera au mieux au « milieu » du nuage des points. La courbe ainsi déterminée, dans un cas comme dans l'autre, sera appelée courbe d'ajustement ou, dans le cas particulier affine, *droite d'ajustement*.

Si  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$  est le nuage de points, et si  $f = f_{a, b, \dots}$  est la « forme » connue ou présumée de la relation mathématique, dépendant des paramètres  $a, b, \dots$ , la méthode standard (due à Gauss et Legendre) s'appelle la méthode des moindres carrés : elle consiste à déterminer les valeurs des paramètres  $a, b, \dots$  comme celles qui minimisent la somme des

carrés des écarts  $\sum_{n=1}^{\infty} (y_i - f(x_i))^2$ . Dans le cas particulier affine et dans une situation de type

déterministe perturbé aléatoirement, on obtient ainsi la *droite des moindres carrés*.

Dans une situation de type intrinsèquement aléatoire, la somme des carrés des écarts n'est pas exactement une variance, mais elle est directement liée aux variances des variables aléatoires que l'on doit introduire pour modéliser la situation. L'ensemble des conditions à satisfaire inclut une minimisation de la variance et on aboutit exactement aux mêmes formules et aux mêmes valeurs des paramètres. Le contexte probabiliste–statistique induit un changement de vocabulaire : plutôt que de droite d'ajustement ou de droite des moindres carrés, on parle de *droite de régression* ; plutôt que de courbe d'ajustement, on parle de *courbe de régression*.

Signalons pour terminer que les liaisons linéaire, affine et polynomiale sont directement traitées par les techniques (équivalentes) de moindres carrés ou de régression linéaire, tandis que les liaisons puissance et exponentielle sont le plus souvent ramenées aux cas précédents par passage aux logarithmes :  $y = bx^p$  devient  $\ln y = \ln b + p \ln x$ ,  $y = ae^{kx}$  devient  $\ln y = \ln a + kx$ .

**ajustement**

Voir *droite d'ajustement* et *khi-deux d'ajustement (test du)*.

**aléatoire***(random)*

Cet adjectif qualifie dans la langue courante des phénomènes dont les « résultats » sont variables, non ou mal maîtrisables, imprévisibles, ..., et s'oppose à « déterministe ». Il est égale-

ment employé, avec un sens mathématique codifié, dans des expressions qui désignent certains concepts du calcul des probabilités (expérience aléatoire, variable aléatoire, processus aléatoire, ...).

Le calcul des probabilités a pour but la « modélisation » des phénomènes aléatoires, dès lors qu'ils sont reproductibles dans des conditions identiques. Selon les cas, cette reproduction pourra être provoquée par l'homme (expérimentation), ou sera « naturelle » (observation). La reproductibilité est essentielle notamment pour fonder l'estimation des probabilités, ce qui explique l'extrême difficulté – conceptuelle et pratique – d'évaluer la probabilité d'un évènement très rare.

On notera que la variabilité des phénomènes qualifiés d'aléatoires peut être soit intrinsèque (variabilité d'un caractère biologique, imprévisibilité d'une désintégration radioactive), soit liée à l'imprécision ou à l'erreur de la mesure de la grandeur qui peut être déterministe.

Voir *hasard, modélisation, variable aléatoire*.

## algèbre d'évènements (*field of events*)

Famille (ensemble)  $\mathcal{A}$  d'évènements qui vérifient les propriétés qui caractérisent une « algèbre de Boole » de parties d'un ensemble  $E$  : appartenance de  $E$  à  $\mathcal{A}$ , « stabilité » par réunion finie et par passage au complémentaire.

Si on ajoute la stabilité par réunion dénombrable, on obtient une  $\sigma$ -algèbre ou *tribu*.

## $\sigma$ -algèbre

Voir *tribu*.

## alphabet grec

Voir *grec (alphabet)*.

## amplitude (*amplitude*)

Terme souvent utilisé pour désigner la largeur d'une classe (bornée) : l'amplitude de la classe  $[a_j, a_{j+1}]$  est  $|a_{j+1} - a_j|$ .

## analyse combinatoire (*combinatorial analysis, combinatorics*)

Branche des mathématiques qui étudie les « configurations », formées à partir d'« objets » pris dans un ensemble fini donné et disposés en respectant certaines contraintes ou certaines structures fixées. Les deux problèmes principaux sont l'énumération des configurations, et leur dénombrement.

Les dénombrements (arrangements, combinaisons, permutations) jouent un rôle important en *probabilités combinatoires*, où l'hypothèse d'équiprobabilité ramène la détermination des probabilités à des comptes d'évènements élémentaires.

## analyse de la variance (test d') (*variance analysis test*)

Test paramétrique qui permet de comparer globalement plusieurs espérances mathématiques entre elles. La dénomination du test s'explique par son fonctionnement, qui décompose la variance de l'ensemble des observations en deux variances partielles, la première fournissant une estimation « inter-classes » de la variance commune, et la seconde une estimation « intra-classes » (ou « résiduelle »). On teste ensuite le quotient : si la première estimation est sensiblement plus grande que la deuxième, on rejette alors l'hypothèse  $H_0$  d'égalité de toutes les espérances.

---

test de comparaison  
de  $q$  espérances mathématiques  $\mu_1, \mu_2, \dots, \mu_q$

---

• Données.  $q$  classes ou groupes, soit pour chaque  $k$  ( $1 \leq k \leq q$ ) : un échantillon  $(x_{k1}, x_{k2}, \dots, x_{kn_k})$  de  $n_k$  valeurs observées d'une variable aléatoire numérique  $X_k$  d'espérance mathématique  $\mu_k$ . On note  $N$  le nombre total  $n_1 + n_2 + \dots + n_q$  de valeurs observées.

• Hypothèse testée.  $H_0 = \ll \mu_1 = \mu_2 = \dots = \mu_q \gg$  contre  $H_1 = \ll \text{il existe au moins deux espérances différentes} \gg$ .

• Déroulement technique du test

1a. Pour chaque  $k$ , on calcule la moyenne de l'échantillon n°  $k$  :

$$m_k = \frac{x_{k1} + x_{k2} + \dots + x_{kn_k}}{n_k}.$$

1b. On calcule la moyenne générale :

$$M = \frac{n_1 m_1 + n_2 m_2 + \dots + n_q m_q}{N}.$$

2. On calcule les 2 sommes de carrés qui « analysent » la variance complète :

$$Q_1 = \sum_{k=1}^q n_k (m_k - M)^2,$$

$$Q_2 = \sum_{k=1}^q \left( \sum_{i=1}^{n_k} (x_{ki} - m_k)^2 \right).$$

(nota : la somme  $Q_1 + Q_2$  est la somme complète  $\sum_{k=1}^q \left( \sum_{i=1}^{n_k} (x_{ki} - M)^2 \right)$ ).

On calcule ensuite  $s_1^2 = \frac{Q_1}{q-1}$ , estimation « inter-classes » de la variance commune, et

$s_2^2 = \frac{Q_2}{N-q}$ , estimation « intra-classes » (« résiduelle »).

3. On calcule enfin la valeur observée de la variable de test :

$$F_{q-1, N-q} = \frac{s_1^2}{s_2^2}.$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Fisher-Snedecor. Elles dépendent des deux degrés de liberté  $q-1$  et  $N-q$ , et du risque  $\alpha$ . On notera que les tables de la loi de Fisher-Snedecor sont unilatérales. En effet, si  $H_0$  est fautive, alors l'estimation inter-classes de la variance ne peut être qu'augmentée.

• Conditions et précautions

- En théorie les  $X_k$  doivent être des v.a. normales ; en outre elles doivent avoir même variance, exigence difficile à apprécier et à contrôler... S'il semblait que cela ne soit pas le cas, on peut appliquer le (*test de*) *Bartlett*, ou bien rechercher dans un ouvrage spécialisé d'autres procédures applicables ;
  - lorsque les  $X_k$  ne sont pas normales, le test est robuste et reste applicable si les effectifs des groupes sont « assez grands ».
-

## analyse de la variance à un facteur (one-way ANOVA)

Méthode d'analyse d'une situation où l'on expérimente l'effet d'un facteur sur une variable, facteur supposé agir exclusivement sur l'espérance mathématique. Dans un premier temps de l'analyse, on teste l'existence d'un effet du facteur par le (*test d'*) *analyse de la variance* : la variance inter-classes s'interprète alors comme la part de variance due au facteur considéré, et le terme de variance résiduelle se justifie alors pleinement.

Si l'hypothèse  $H_0$  est rejetée, cela signifie que le facteur a un effet, et dans un deuxième temps de l'analyse, on estime les espérances dont l'ensemble décrit l'effet du facteur.

La présentation globale et les notations données ci-dessous sont très légèrement différentes de celles du test d'analyse de la variance, de façon à pouvoir les généraliser naturellement au cas de deux facteurs.

### analyse de la variance à un facteur A

- Données.  $p$  classes ou groupes, avec pour chaque niveau  $i$  ( $1 \leq i \leq p$ ) du facteur : un échantillon  $(x_{i1}, x_{i2}, \dots, x_{in_i})$  de  $n_i$  valeurs observées d'une variable aléatoire numérique  $X_i$  d'espérance mathématique  $\mu_i$ .

On note  $N$  le nombre total  $n_1 + n_2 + \dots + n_p$  de valeurs observées.

- Modèle étudié

Les observations individuelles sont de la forme  $X_{ij} = \mu_i + E_{ij}$ , où  $E_{ij}$  est un écart de loi normale centrée et d'écart-type  $\sigma$  (indépendant de l'effet du facteur A).

On pose  $\mu_i = \mu + \alpha_i$ , où  $\mu$  représente la moyenne globale et  $\alpha_i$  l'effet du facteur A (en convenant que la moyenne pondérée des effets est nulle).

#### ► Premier temps

- Hypothèse sur l'effet du facteur.  $H_0 =$  « le facteur ne possède aucun effet » = «  $\mu_1 = \mu_2 = \dots = \mu_p$  » contre  $H_1 =$  « il existe au moins deux espérances différentes ».

- Déroulement technique du test

- On calcule la moyenne observée  $m_i$  de l'échantillon n°  $i$  ;
- on calcule la moyenne générale observée  $M$  (moyenne pondérée des  $m_i$ ) ;
- on calcule les sommes de carrés relatives aux parts de variance observées et on regroupe les résultats comme indiqué dans le tableau suivant :

Origine de la dispersion	Somme de carrés	ddl	Estimation de la variance	Test F
totale	$Q = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{ij} - M)^2$	$N - 1$	$\frac{Q}{N - 1}$	
facteur A	$Q_A = \sum_{i=1}^p (m_i - M)^2$	$p - 1$	$s_A^2 = \frac{Q_A}{p - 1}$	$F_{p-1, N-p} = \frac{s_A^2}{s_R^2}$
résiduelle	$Q_R = Q - Q_A$	$N - p$	$s_R^2 = \frac{Q_R}{N - p}$	

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Fisher-Snedecor.



• Conditions et précautions. Voir *analyse de la variance* (test d').

► Deuxième temps (en cas de rejet de  $H_0$ )

• estimations des effets

- $\mu$  est estimée par  $M$ ,
- les  $\alpha_i$  sont estimés par  $m_i - M$ .

## analyse de la variance à deux facteurs (two-way ANOVA)

Méthode d'analyse d'une situation où l'on expérimente l'effet de deux facteurs sur une variable, facteurs supposés agir exclusivement sur l'espérance mathématique. Il faut distinguer les effets séparés des deux facteurs et envisager *a priori* une interaction possible.

Dans un premier temps de l'analyse, on teste l'existence de l'effet des facteurs et l'existence d'une interaction par plusieurs tests de type *analyse de la variance*.

Si l'un ou l'autre de ces tests rejette l'hypothèse de non-existence d'un effet, il faut alors, dans un deuxième temps de l'analyse, estimer les espérances dont l'ensemble décrit les effets des deux facteurs ainsi que leur interaction.

Il existe plusieurs « plans d'expérience » possibles pour préciser le détail de l'expérimentation. Les formules qui suivent concernent les deux plans d'expérience les plus courants, qui sont des plans « complets » et « équilibrés » : toutes les associations entre un « niveau » du premier facteur et un niveau du second sont observées, et toutes les classes présentent le même nombre  $r$  d'observations. Le premier plan est « à répétitions » ( $r \geq 2$ ), le deuxième sans répétition ( $r = 1$ ), auquel cas l'on ne peut pas évaluer les interactions et l'on estime seulement l'effet des deux facteurs.

### 1. analyse de la variance à deux facteurs A et B plan d'expérience complet équilibré avec répétitions

• Données.  $pq$  classes ou groupes, avec pour chaque couple  $(i, j)$  ( $1 \leq i \leq p$ ,  $1 \leq j \leq q$ ) de niveaux des facteurs : un échantillon  $(x_{ij1}, x_{ij2}, \dots, x_{ijr})$  de  $r$  valeurs observées d'une variable aléatoire numérique  $X_{ij}$  d'espérance mathématique  $\mu_{ij}$ .

Le nombre total  $N$  de valeurs observées est égal à  $pqr$ .

• Modèle étudié. Les observations individuelles sont de la forme  $X_{ijk} = \mu_{ij} + E_{ijk}$ , où  $E_{ijk}$  est un écart de loi normale centrée et d'écart-type  $\sigma$  (indépendant de l'effet des facteurs A et B et de leur interaction).

On pose  $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$ , où  $\mu$  représente la moyenne globale et  $\alpha_i$  l'effet du facteur A,  $\beta_j$  l'effet du facteur B,  $\gamma_{ij}$  l'interaction entre A et B (en convenant que la moyenne pondérée des effets et de l'interaction est nulle).

► Premier temps

• Hypothèse globale sur l'existence d'effet(s).  $H_0 =$  « il n'y a aucun effet » = « toutes les  $\mu_{ij}$  sont égales » contre  $H_1 =$  « il existe au moins deux espérances différentes ».

• Déroulement technique du test

- On calcule la moyenne observée  $m_{ij}$  de chaque classe  $(i, j)$  ;
- on calcule les moyennes marginales observées  $m_{i.}$  et  $m_{.j}$  ;
- on calcule la moyenne générale observée  $M$  ;
- on calcule les sommes de carrés relatives aux parts de variance observées et on regroupe les résultats comme indiqué dans le tableau ci-contre.
- on effectue trois tests séparés, les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Fisher-Snedecor.

Origine de la dispersion	Somme de carrés	ddl	Estimation de la variance	Test F
totale	$Q = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (x_{ijk} - M)^2$	$pqr - 1$	$\frac{Q}{pqr - 1}$	
facteur A	$Q_A = qr \sum_{i=1}^p (m_{i.} - M)^2$	$p - 1$	$s_A^2 = \frac{Q_A}{p - 1}$	$F_{p-1, pq(r-1)} = \frac{s_A^2}{s_R^2}$
facteur B	$Q_B = pr \sum_{j=1}^q (m_{.j} - M)^2$	$q - 1$	$s_B^2 = \frac{Q_B}{q - 1}$	$F_{q-1, pq(r-1)} = \frac{s_B^2}{s_R^2}$
interaction AB	$Q_{AB} = r \sum_{i=1}^p \sum_{j=1}^q (m_{ij} - m_{.j} - m_{i.} + M)^2$	$(p - 1)(q - 1)$	$s_{AB}^2 = \frac{Q_{AB}}{(p - 1)(q - 1)}$	$F_{(p-1)(q-1), pq(r-1)} = \frac{s_{AB}^2}{s_R^2}$
résiduelle	$Q_R = Q - Q_A - Q_B - Q_{AB}$	$pq(r - 1)$	$s_R^2 = \frac{Q_R}{pq(r - 1)}$	

• Conditions et précautions. Voir *analyse de la variance* (test d').

➤ Deuxième temps (en cas de rejet de  $H_0$ )

• estimations des effets

- $\mu$  est estimée par  $M$  ;
- les  $\alpha_i$  sont estimés par  $m_{i.} - M$ , les  $\beta_j$  sont estimés par  $m_{.j} - M$  ;
- les  $\gamma_{ij}$  sont estimés par  $m_{ij} - m_{.j} - m_{i.} + M$ .

2. analyse de la variance à deux facteurs A et B  
plan d'expérience complet équilibré sans répétition

• Données.  $pq$  classes ou groupes, avec pour chaque couple  $(i, j)$  ( $1 \leq i \leq p, 1 \leq j \leq q$ ) de niveaux des facteurs : une valeur unique  $x_{ij}$  observée d'une variable aléatoire numérique  $X_{ij}$  d'espérance mathématique  $\mu_{ij}$ .

On note  $N$  le nombre total  $pq$  de valeurs observées.

• Modèle étudié. Les observations individuelles sont de la forme  $X_{ij} = m_{ij} + E_{ij}$ , où  $E_{ij}$  est un écart de loi normale centrée et d'écart-type  $\sigma$  (indépendant de l'effet des facteurs A et B).

On pose  $\mu_{ij} = \mu + \alpha_i + \beta_j$ , où  $\mu$  représente la moyenne globale et  $\alpha_i$  l'effet du facteur A,  $\beta_j$  l'effet du facteur B (en convenant que la moyenne pondérée des effets est nulle).

➤ Premier temps

• Hypothèse globale sur l'existence d'effet(s).  $H_0 = \ll \text{il n'y a aucun effet} \gg = \ll \text{toutes les } \mu_{ij} \text{ sont égales} \gg$  contre  $H_1 = \ll \text{il existe au moins deux espérances différentes} \gg$ .

• Déroulement technique du test

- On calcule les moyennes marginales observées  $m_{i.}$  et  $m_{.j}$  ;
- on calcule la moyenne générale observée  $M$  ;



- on calcule les sommes de carrés relatives aux parts de variance observées et on regroupe les résultats comme indiqué dans le tableau suivant :

Origine de la dispersion	Somme de carrés	ddl	Estimation de la variance	Test F
totale	$Q = \sum_{i=1}^p \sum_{j=1}^q (x_{ij} - M)^2$	$pq - 1$	$\frac{Q}{pq - 1}$	
facteur A	$Q_A = q \sum_{i=1}^p (m_{i.} - M)^2$	$p - 1$	$s_A^2 = \frac{Q_A}{p - 1}$	$F_{p-1, (p-1)(q-1)} = \frac{s_A^2}{s_R^2}$
facteur B	$Q_B = p \sum_{j=1}^q (m_{.j} - M)^2$	$q - 1$	$s_B^2 = \frac{Q_B}{q - 1}$	$F_{q-1, (p-1)(q-1)} = \frac{s_B^2}{s_R^2}$
résiduelle	$Q_R = Q - Q_A - Q_B$	$(p - 1)(q - 1)$	$s_R^2 = \frac{Q_R}{(p - 1)(q - 1)}$	

- on effectue deux tests séparés, les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Fisher-Snedecor.

• Conditions et précautions. Voir *analyse de la variance* (test d').

➤ Deuxième temps (en cas de rejet de  $H_0$ )

• estimation des effets

- $\mu$  est estimée par  $M$  ;
- les  $\alpha_i$  sont estimés par  $m_{i.} - M$ , les  $\beta_j$  sont estimés par  $m_{.j} - M$ .

## analyse des données

(*multivariate analysis*)

Cette expression possède aujourd'hui un sens très précis en mathématiques. Elle désigne globalement l'ensemble des méthodes qui permettent de traiter les situations qui impliquent un grand nombre de caractères et un grand nombre de données. Ces méthodes nécessitent beaucoup de calculs et ont fleuri depuis que l'informatique leur permet d'être mises en œuvre. Souvent très élaborées, elles sont néanmoins très « descriptives » et par conséquent utilisables non seulement dans des situations probabilistes standard, mais aussi dans des situations non probabilisables (données sociologiques ou statistiques économiques par exemple). Elles reposent pour beaucoup sur une analyse géométrique et algébrique de la représentation des données dans un espace (abstrait) de grande dimension (où, par exemple, les coefficients de corrélation peuvent être interprétés comme des cosinus d'angles). Elles s'intéressent à la fois aux caractères (détermination des liaisons) et aux individus (par exemple sous-structures du nuage de points).

La méthode qui se trouve dans le prolongement direct de la corrélation et de la régression linéaire s'appelle l'« analyse en composantes principales ». Les autres méthodes portent les noms d'analyse (factorielle) discriminante, d'analyse des correspondances, de classification hiérarchique, etc. Toutes les méthodes d'analyse des données débouchent (mais non exclusivement) sur des représentations graphiques, et certaines sont fréquemment utilisées dans les médias.

**ANOVA**

(ANalysis Of VAriance)

Voir *analyse de la variance (test d')*, *analyse de la variance à un facteur*, *analyse de la variance à deux facteurs*.

**aplatissement (coefficient d')** (*kurtosis coefficient, excess coefficient*)

Nombre réel sans dimension, indicateur de forme, qui mesure l'aplatissement d'une distribution probabiliste ou statistique.

On considère une variable aléatoire réelle (ou une variable statistique), dont les moments centrés d'ordre 2 et 4 sont :

$$\mu_2 = \sigma^2 \text{ et } \mu_4 \text{ (ou } m_2 = s^2 \text{ et } m_4).$$

Le coefficient d'aplatissement est le quotient

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \quad \left( \text{ou } \frac{m_4}{s^4} - 3 \right).$$

Un coefficient apparemment plus simple :  $\beta_2 = \frac{\mu_4}{\sigma^4}$  (ou  $\frac{m_4}{s^4}$ ), avait été introduit par Pearson.

Comme sa valeur est égale à 3 pour une loi normale, Fisher a retranché 3 pour traduire l'idée que l'aplatissement standard (coefficient nul) est celui de la loi normale. Le coefficient  $\gamma_2$  est compris entre  $-2$  et  $+\infty$ . Ce coefficient est négatif pour une distribution étalée (on dit parfois « platicurtique »), positif pour une distribution pointue (on dit parfois « leptocurtique »). Il faut toutefois noter que son signe n'est vraiment significatif que pour une distribution à peu près symétrique.

**Exemple** Pour une distribution continue uniforme sur  $[0, 1]$ , on a  $\mu_2 = \sigma^2 = \frac{1}{12}$  et  $\mu_4 = \frac{1}{80}$ , d'où  $\gamma_2 = -1,2$  (on aurait bien entendu la même valeur pour une distribution continue uniforme sur un intervalle  $[a, b]$  quelconque).

**appariées (séries) / appariés (échantillons)***(paired samples)*

Terminologie qui s'oppose à échantillons indépendants et qui désigne, dans un test d'hypothèse, la constitution d'un échantillon « par paires ».

Dans un test de comparaison sur des échantillons indépendants, on considère deux variables numériques X et Y et on constitue deux échantillons d'observations  $(x_1, x_2, \dots, x_{n_X})$  et  $(y_1, y_2, \dots, y_{n_Y})$  : les  $n_X$  individus pour lesquels ont été observées les  $n_X$  valeurs de X sont a priori différents des  $n_Y$  individus pour lesquels ont été observées les  $n_Y$  valeurs de Y (et en général  $n_X \neq n_Y$ ) ; les échantillons sont donc indépendants, au sens usuel et au sens probabiliste, et le test compare par exemple les espérances  $\mu_X$  et  $\mu_Y$ .

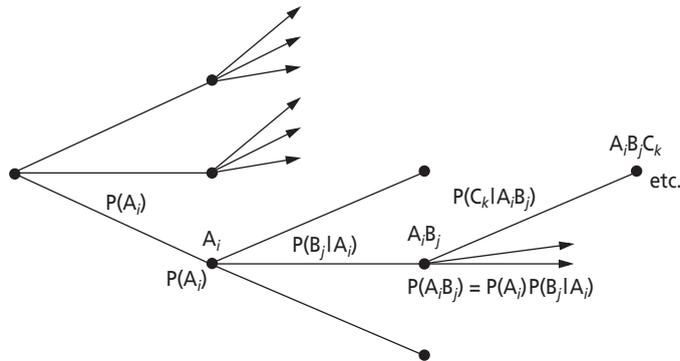
Dans un test de comparaison sur des séries appariées, on considère autant de couples (de « paires ») de variables numériques  $(X_i, Y_i)$  qu'il y a d'individus sur lesquels seront effectuées les observations (dans certains cas les deux observations seront faites non pas sur le même individu mais sur deux individus « similaires »). Les lois des  $X_i$  et des  $Y_i$  ne font pas l'objet d'un présupposé d'uniformité et on teste seulement la différence entre  $X_i$  et  $Y_i$  (certains de ces tests sont parfois appelés tests de *différences couplées*). Le postulat du test est alors qu'il y a une unique variable D dont les différences observées  $d_i = y_i - x_i$  constituent un échantillon, et l'on teste la symétrie de D ou une de ses conséquences, par exemple  $\mu_D = 0$ . Il arrive

que la formulation employée (par exemple «  $\mu_{X_i} = \mu_{Y_i}$  » au lieu de «  $\mu_D = 0$  ») masque cette problématique simple, et donne l'impression que l'on compare deux variables X et Y, dont les  $X_i$  et les  $Y_i$  seraient des exemplaires particuliers aux individus.

**arbre**

(tree)

Technique de représentation des espaces probabilisés lorsque chaque évènement élémentaire peut se décrire comme une suite (temporelle ou logique) d'évènements. Le vocabulaire standard des arbres en combinatoire ou en informatique est le suivant : un arbre est constitué par une « racine », d'où partent des « branches », qui conduisent à des « nœuds » ; aux extrémités se trouvent les « noeuds terminaux », appelés parfois « feuilles ». L'orientation du dessin est le plus souvent avec la racine en haut ou à gauche.



Lorsqu'on utilise un arbre pour représenter un espace probabilisé, on porte les probabilités conditionnelles sur les branches, et les probabilités « absolues » aux nœuds (le « nœud racine » porte la probabilité 1). À chaque nœud, on calcule la probabilité absolue comme la probabilité absolue du nœud précédent multipliée par la probabilité conditionnelle de la branche qui le relie au nœud considéré. Ce calcul est la version « arborescente » de la (formule des probabilités) composées généralisée :

$$P(A_i B_j) = P(A_i) P(B_j | A_i),$$

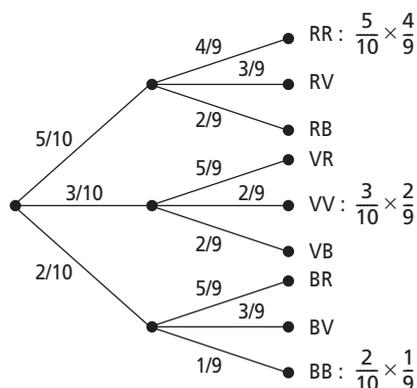
$$P(A_i B_j C_k) = P(A_i) P(B_j | A_i) P(C_k | A_i B_j),$$

etc.

**Exemple** On considère une urne avec 10 boules : 5 Rouges, 3 Vertes et 2 Bleues. On effectue deux tirages sans remise et on cherche la probabilité P que les deux boules tirées soient de même couleur.

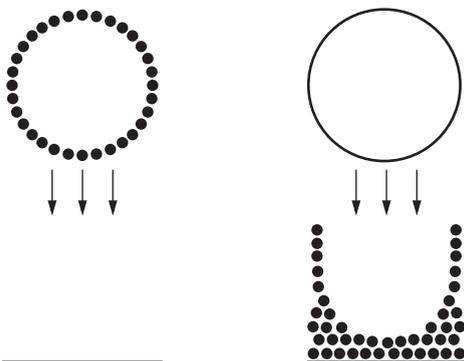
L'évènement « même couleur » est la réunion des 3 évènements élémentaires RR, VV, BB. Leurs probabilités se lisent immédiatement sur l'arbre ci-contre et on a :

$$P = \frac{20}{90} + \frac{6}{90} + \frac{2}{90} = \frac{28}{90} = 0,311.$$



**arc sinus (loi de l')** (*inverse sine distribution, arc sine distribution*)

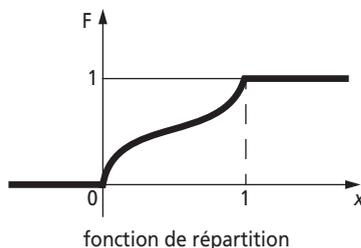
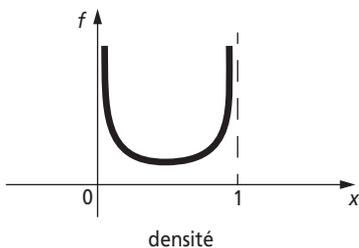
Loi d'une variable aléatoire continue entre 0 et 1, que l'on peut imaginer comme la « projection » sur une droite d'une distribution uniforme sur un cercle.



**Formulaire**

Version standardisée X sans paramètre, à valeurs sur l'intervalle ]0, 1[.

► Loi de probabilité



$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}} \quad (0 < x < 1)$$

$$F(x) = \frac{2}{\pi} \text{Arc sin } \sqrt{x} \quad (0 \leq x \leq 1)$$

$$= \frac{1}{\pi} \text{Arc cos } (1 - 2x) \quad (0 \leq x \leq 1)$$

► Valeurs caractéristiques

- espérance :  $E(X) = \frac{1}{2}$  ;
- variance :  $\text{Var}(X) = \frac{1}{8}$  ;
- écart-type :  $\sigma(X) = \frac{1}{2\sqrt{2}}$  .

Il existe une deuxième version standardisée Y, à valeurs sur l'intervalle ]-1, 1[

► Loi de probabilité

$$f(x) = \frac{1}{\pi\sqrt{1-x^2}} \quad (-1 < x < 1)$$

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \text{Arc sin } x \quad (-1 \leq x \leq 1)$$

► Valeurs caractéristiques

– espérance :  $E(X) = 0$  ;

– variance :  $\text{Var}(X) = \frac{1}{2}$  ;

– écart-type :  $\sigma(X) = \frac{1}{\sqrt{2}}$ .

**Remarque :** les fonctions Arc sin et Arc cos sont les « fonctions réciproques » des fonctions sinus et cosinus :

$$y = \text{Arc sin } x \ (|x| \leq 1) \Leftrightarrow x = \sin y \quad \text{et} \quad -\frac{\pi}{2} \leq y \leq \frac{\pi}{2}$$

$$y = \text{Arc cos } x \ (|x| \leq 1) \Leftrightarrow x = \cos y \quad \text{et} \quad 0 \leq y \leq \pi$$

(ces fonctions apparaissent sur les calculatrices comme  $\sin^{-1}$  et  $\cos^{-1}$ , parfois comme Asn et Acs).

► Utilisations

La loi de l'Arc sinus (première version X) est un cas particulier de la (loi) *bêta de type I* (pour les valeurs  $r = s = \frac{1}{2}$  des deux paramètres).

En théorie, la loi de l'Arc sinus (deuxième version Y) est la loi de  $\sin U$  où  $U$  est une v.a. uniforme dans  $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ , la loi de  $\cos U$  où  $U$  est une v.a. uniforme dans  $[0, \pi]$ , la loi de  $\sin U$  ou de  $\cos U$  où  $U$  est une v.a. uniforme dans  $[0, 2\pi]$ .

Dans la pratique, cette loi, découverte en 1939 par le mathématicien Paul Lévy comme loi limite dans les marches aléatoires et la théorie des jeux, privilégie les situations déséquilibrées au détriment des situations équilibrées. Par exemple, dans un jeu de  $n$  parties de P ou F ( $n$  grand), il arrivera le plus souvent que l'un des deux joueurs soit en situation de gain durant la quasi-totalité des coups, selon une distribution de probabilités qui converge vers la loi de l'Arc sinus.

## arithmétique (triangle)

Voir *triangle de Pascal*.

### arrangements avec répétitions (arrangements with repetitions)

Entiers positifs dépendant de deux paramètres entiers positifs  $n$  et  $k$  ( $n \geq 0$ ,  $0 \leq k \leq n$ ), intervenant dans les dénombrements.

On considère un ensemble  $E$  de  $n$  objets : le nombre d'arrangements avec répétition représente le nombre de suites (ordonnées) de  $k$  objets pris parmi les  $n$ , un même objet pouvant être pris (répété) plusieurs fois.

Il n'y a pas de notation spéciale utilisée en probabilités.

nombre d'arrangements avec répétitions = $n^k$
--

A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

**Exemples 1**  $E = \{P, F\}$  : il y a  $2^k$  suites différentes de  $k$  pile ou face.

**2**  $E = \{As, 2, 3, 4, 5, 6\}$  : il y a  $6^k$  résultats différents pour  $k$  jets d'un dé.

**3**  $E = \{a, b, \dots, z\}$  : on peut former avec les 26 lettres de l'alphabet français  $26^3 = 17\,756$  « mots » différents de 3 lettres.

Voir *binomiaux (coefficients)*.

### arrangements sans répétition (arrangements)

Entiers positifs dépendant de deux paramètres entiers positifs  $n$  et  $k$  ( $n \geq 0, 0 \leq k \leq n$ ), intervenant dans les dénombrements.

On considère un ensemble  $E$  de  $n$  objets : le nombre d'arrangements sans répétition, noté  $A_n^k$ , représente le nombre de suites (ordonnées) de  $k$  objets pris parmi les  $n$ , un même objet ne pouvant pas être pris plusieurs fois.

Lorsque le mot « arrangements » est employé sans précision, il désigne les arrangements sans répétition.

$$A_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

Cas particulier  $k = n$  :

$A_n^n = n!$  est le nombre de *permutations* de  $n$  objets.

**Exemples 1**  $E = \{0, 1, \dots, 9\}$  : il y a  $10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 = 1\,814\,400$  numéros de téléphone de 8 chiffres formés avec des chiffres tous différents.

**2**  $E = \{a, b, \dots, z\}$  : on peut former avec les 26 lettres de l'alphabet français  $26 \times 25 \times 24 = 15\,600$  « mots » différents de 3 lettres toutes différentes.

**3 (permutations)** Le nombre de manières différentes de placer 10 personnes dans une file d'attente est  $10! = 3\,628\,800$ .

Voir *binomiaux (coefficients), permutation*.

### asymétrie (coefficient d') / dissymétrie (coefficient de) ([Fisher] skewness [coefficient])

Nombre réel sans dimension, indicateur de forme, qui mesure l'asymétrie d'une distribution probabiliste ou statistique.

On considère une variable aléatoire réelle (ou une variable statistique), dont les moments centrés d'ordre 2 et 3 sont :

$$\mu_2 = \sigma^2 \text{ et } \mu_3 \quad (\text{ou } m_2 = s^2 \text{ et } m_3).$$

Le coefficient d'asymétrie est le quotient :

$$\gamma_2 = \frac{\mu_3}{\sigma^3} \quad \left( \text{ou } \frac{m_3}{s^3} \right)$$

Ce coefficient a été introduite par Fisher ; pour une distribution régulière, avec un maximum (« mode ») unique, il est positif lorsque ce maximum est à gauche de l'espérance – avec donc une « queue de distribution » importante à droite, nul lorsque la distribution est symétrique, et négatif lorsque le maximum de la distribution est à droite de l'espérance – avec donc une « queue de distribution » importante à gauche.

**auto-corrélation***(autocorrelation)*

Notion employée notamment pour les séries chronologiques  $(y_i)$  où, sous sa forme la plus simple, elle désigne la corrélation « ordinaire » des couples  $(y_i, y_{i+1})$ .

Voir *Durbin-Watson (test de)*.

**axiomatique de Kolmogorov***(Kolmogorov axiom system)*

Formulation donnée en 1930 par Kolmogorov pour caractériser le cadre d'étude des espaces probabilisés, et considérée depuis comme définitive. Cette axiomatique considère le « triplet »  $(\Omega, \mathcal{A}, P)$  formé par un *espace fondamental*, une *tribu* d'évènements (notion introduite précédemment par Borel) et une *mesure de probabilité*, et énonce les axiomes que doivent satisfaire la tribu et la mesure de probabilité.

Voir *tribu, mesure de probabilité*.



## Bartlett (test de)

(Bartlett test)

Test d'hypothèse paramétrique utilisé pour comparer les variances observées de plusieurs échantillons statistiques.

\_\_\_\_\_ test de comparaison de  $q$  variances  $\sigma_1^2, \sigma_2^2, \dots, \sigma_q^2$  \_\_\_\_\_

• Données.  $q$  séries ou groupes, avec pour chaque  $k$  ( $1 \leq k \leq q$ ) : un échantillon de  $n_k$  valeurs observées d'une variable aléatoire numérique  $X_k$  d'espérance mathématique  $\mu_k$  et de variance  $\sigma_k^2$ .

On note  $N$  le nombre total de valeurs observées  $n_1 + n_2 + \dots + n_q$ .

• Hypothèse testée.  $H_0 = \ll \sigma_1^2 = \sigma_2^2 = \dots = \sigma_q^2 \gg$  contre  $H_1 = \ll \text{il existe au moins deux varian-ces différentes} \gg$ .

• Déroulement technique du test

1. On calcule, avec les formules usuelles, la moyenne puis la variance débiaisée  $s_k^2$  de l'échantillon n°  $k$ .

2. On calcule une estimation commune de toutes les variances :

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_q - 1)s_q^2}{N - q}.$$

3. On calcule la valeur observée de la variable de test :

$$t = (N - q) \ln s^2 - \sum_{k=1}^q (n_k - 1) \ln s_k^2.$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi du khi-deux à  $q - 1$  degrés de liberté, pour le risque *unilatéral*  $\alpha$  (qui est le risque standard des tables du khi-deux).

• Conditions et précautions

- Ce test n'est pas robuste : il n'est valable que dans le cas où les lois des échantillons sont toutes normales ;
- il est prudent de demander que chaque effectif  $n_k$  soit  $\geq 5$ .

Il est parfois conseillé de diviser la variable de test  $t$  par le facteur correctif :

$$C = 1 + \frac{1}{3(q-1)} \left( \left( \sum_{k=1}^q \frac{1}{n_k - 1} \right) - \frac{1}{N} \right).$$

**Remarque :** si  $q = 2$ , ce test n'est pas équivalent au test du quotient des variances (sauf si les effectifs sont égaux).

## Bayes (formule de), Bayes (théorème de) (Bayes formula, Bayes theorem)

Nom d'une formule – ou théorème – utilisée pour trouver les « probabilités des causes ».

### Formule de Bayes (version simple)

Soient deux évènements A et B d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , avec  $P(B) \neq 0$ .  
Alors :

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

Le terme  $\frac{P(A \cap B)}{P(B)}$  est la définition de la probabilité conditionnelle, et la formule de Bayes est son expression avec l'autre probabilité conditionnelle. De façon générale, il s'agit d'exprimer l'une quelconque des deux probabilités conditionnelles en fonction de l'autre.

### Formule de Bayes (version composée)

Soient une partition  $(H_j)$  (ou système complet d'évènements) d'un espace  $\Omega$  et un évènement B, avec  $P(B) \neq 0$ . Alors :

$$P(H_j|B) = \frac{P(H_j \cap B)}{P(B)} = \frac{P(H_j)P(B|H_j)}{P(H_1)P(B|H_1) + P(H_2)P(B|H_2) + \dots + P(H_k)P(B|H_k)}.$$

Les évènements  $H_j$  peuvent être considérés comme des causes, et l'évènement B comme un résultat. Il s'agit bien entendu d'une interprétation dans le contexte d'une modélisation, tous les évènements, causes et résultat, étant de même « nature » mathématique.

**Exemple 1** Un dépistage systématique est effectué sur une population dont 6 % des individus présentent une certaine affection A non apparente. Ce dépistage est débuté par un test qui donne 95 % de résultats positifs pour les personnes atteintes par A (les autres étant des « faux négatifs ») et 1 % de résultats positifs pour les personnes non atteintes (les « faux positifs »).

Quelle est la probabilité conditionnelle qu'une personne prise au hasard soit atteinte par A sachant que le test a donné un résultat positif ? Soit indemne sachant que le test a donné un résultat négatif ?

On peut représenter la situation soit par un arbre, soit par un tableau à 4 cases : par exemple, dans cette deuxième représentation (S signifie sain, et A porteur de l'affection A), la probabilité de la case « S et test – » est calculée comme le produit  $P(S)P(\text{test} - | S) = 0,94 \times 0,99 = 0,9306$  ; le tableau est figuré ci-dessous :

	test –	test +	
S	0,9306	0,0094	0,94
A	0,0030	0,0570	0,06
	0,9336	0,0664	1

Ce tableau est complété par les probabilités « marginales », et on peut calculer notamment  $P(\text{test } -) = P(S \text{ et test } -) + P(A \text{ et test } -) = 0,9336$  (somme par colonne) et de même  $P(\text{test } +) = 0,0664$ . On peut alors calculer les probabilités conditionnelles demandées :

$$P(A|\text{test } +) = \frac{P(A \text{ et test } +)}{P(\text{test } +)} = \frac{0,0570}{0,0664} \approx 0,86,$$

valeur à comparer avec la probabilité *a priori* 0,06 d'être porteur de A, et :

$$P(S|\text{test } -) = \frac{P(S \text{ et test } -)}{P(\text{test } -)} = \frac{0,9306}{0,9336} \approx 0,997,$$

valeur à comparer avec la probabilité *a priori* 0,94 d'être sain.

**Exemple 2** On considère une usine où trois machines fabriquent un même modèle de pièce. 40 % des pièces sont fabriquées par la machine A, qui produit 0,1 % de pièces défectueuses ; 30 % des pièces sont fabriquées par la machine B, plus ancienne, qui produit 0,3 % de pièces défectueuses ; 30 % des pièces sont fabriquées par la machine C, encore plus ancienne, qui produit 0,8 % de pièces défectueuses. On demande la probabilité conditionnelle qu'une pièce ait été fabriquée par la machine C, sachant qu'elle est défectueuse.

Appelons A l'évènement « une pièce prise au hasard a été fabriquée par la machine A », B et C les évènements analogues pour les machines B et C. Appelons D l'évènement « une pièce prise au hasard est défectueuse ». Il faut commencer par traduire les pourcentages en probabilités et en probabilités conditionnelles :

$$P(A) = 0,4, P(B) = 0,3, P(C) = 0,3, \\ P(D|A) = 0,001, P(D|B) = 0,003, P(D|C) = 0,008.$$

On peut alors calculer le dénominateur de la formule de Bayes

$$P(D) = P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) \\ = 0,4 \times 0,001 + 0,3 \times 0,003 + 0,3 \times 0,008 = 0,0037.$$

Et on a enfin :

$$P(C|D) = \frac{P(C)P(D|C)}{P(D)} = \frac{0,0024}{0,0037} = 0,65$$

On voit ainsi que, pour employer un vocabulaire ancien, la probabilité *a priori* qu'une pièce (prise au hasard) ait été fabriquée par C est 0,30, et que la probabilité *a posteriori* qu'elle ait été fabriquée par C sachant qu'elle est défectueuse passe à 0,65.

Voir *conditionnelle (probabilité)*.

## Benford (loi de)

(Benford distribution)

Loi empirique qui régit la distribution du premier chiffre des nombres pris dans des ensembles de données présentant des grandes variations d'échelle. Cette loi a été découverte en 1881 par l'astronome S. Newcomb et redécouverte en 1938 par le physicien F. Benford. Elle énonce que la probabilité d'apparition du premier chiffre significatif  $k$  d'un nombre (écrit en base 10) est :

$$P(k) = \log_{10}\left(1 + \frac{1}{k}\right)$$

En particulier  $P(1) = 0,301 \approx 30\%$ . L'une des justifications mathématiques de cette loi est son invariance par un changement arbitraire d'unité de mesure.

### ► Utilisation

Cette loi a été utilisée dans les années 1990 pour détecter des fraudes comptables par utilisation de données inventées.

## Bernoulli (Jacques)

Mathématicien suisse (1654–1705), aîné de la grande lignée des Bernoulli. Il démontra la loi (faible) des grands nombres (qu'il appela le « théorème d'or ») et composa le traité *Ars conjectandi* qui fut publié à titre posthume. Il fit également des travaux en analyse (équations différentielles, cinématique), en géométrie, et est à l'origine du « calcul des variations ».

## Bernoulli (loi de)

(Bernoulli distribution)

Loi d'une variable aléatoire discrète qui prend deux valeurs 1 et 0, dont l'importance théorique et pratique est primordiale.

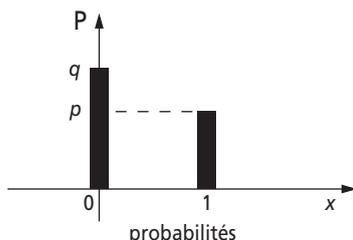
### Formulaire

Un paramètre réel  $p$  ( $0 \leq p \leq 1$ ) qui représente une probabilité et une notation systématiquement utilisée :  $q = 1 - p$ .

Soit  $X$  la variable aléatoire de Bernoulli de paramètre  $p$  ; valeurs prises : 1 ou 0.

#### ► Loi de probabilité

$$P(X = 1) = p, P(X = 0) = q$$



#### ► Valeurs caractéristiques

- espérance :  $E(X) = p$
- variance :  $\text{Var}(X) = pq$
- écart-type :  $\sigma(X) = \sqrt{pq}$ .

#### ► Utilisation

La loi de Bernoulli est la loi de la *variable indicatrice* d'un événement  $A$  :  $X = 1$  si  $\omega \in A$ ,  $X = 0$  si  $\omega \notin A$ .

Conséquence importante : comme  $p = P(A) = E(X)$ , la loi de Bernoulli est souvent utilisée en statistique pour pouvoir traiter une probabilité avec les formules et les tests utilisés pour les espérances mathématiques

**Exemple 1** Le résultat d'un lancer de pièce *codé*, par exemple Pile par 1 et Face par 0, est une variable aléatoire de Bernoulli (de paramètre  $p = 1/2$  si la pièce est homogène et équilibrée).

**Exemple 2** Une question posée dans un sondage, avec deux réponses possibles (par exemple « oui » ou « non » *codées* respectivement 1 et 0), est une variable aléatoire de Bernoulli.

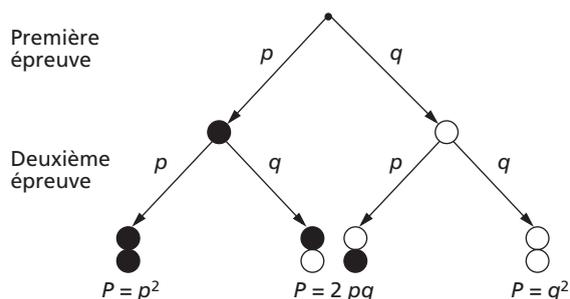
## Bernoulli (schéma de)

(Bernoulli trials)

Synonyme *processus de Bernoulli*.

Modèle mathématique décrivant les « tirages avec remise » ou les épreuves répétées dans le cas particulier où chaque résultat successif ne peut prendre que deux valeurs. Dans l'une ou

l'autre de ces présentations d'un même schéma conceptuel, l'idée essentielle est que les événements antérieurs n'ont aucune influence sur le résultat de la nouvelle épreuve (il y a donc indépendance des épreuves successives).



Le mot « processus », dans sa signification mathématique précise, renvoie à une suite de variables aléatoires, dans le cas présent à la suite de variables aléatoires binomiales  $(B(n, p))_{n=1,2,\dots}$  qui modélise le résultat des tirages aléatoires successifs.

### Berry-Esséen (théorème de) (*Berry-Esseen theorem, Berry-Esseen bound*)

On considère une suite  $(Z_n)$  de variables aléatoires centrées réduites convergeant, dans les conditions du (*théorème*) *central limite*, vers une v.a. normale centrée réduite. Le théorème de Berry-Esséen permet de majorer l'erreur commise en remplaçant la fonction de répartition de  $Z_n$  par celle de la loi normale (à condition de supposer l'existence du moment centré absolu d'ordre 3).

**Théorème.** On considère une suite  $(X_n)$  de variables aléatoires (réelles) indépendantes et identiquement distribuées, d'espérance mathématique  $\mu$  et de variance  $\sigma^2$ . On définit les moyennes :

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n},$$

puis les variables centrées réduites correspondantes :

$$Z_n = \frac{M_n - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)},$$

et enfin les fonctions de répartition :

$$F_n(x) = P(Z_n \leq x).$$

On suppose que le moment centré absolu d'ordre 3 des  $X_n$  existe :  $M_3 = E(|X_n - \mu|^3)$ .

On pose  $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$  (fonction de répartition d'une v.a. normale centrée réduite). Alors, pour tout  $x \in \mathbf{R}$  :

$$|F_n(x) - F(x)| \leq 0,7655 \frac{M_3}{\sigma^3 \sqrt{n}}$$

(Le coefficient numérique « historique » est 3, la valeur 0,7655 est due à Shiganov (1986).)

**Remarque :** lorsque la distribution est symétrique, on peut améliorer ce résultat et donner une borne « en  $\frac{1}{n}$  ».

**bêta de type I (loi)***(beta distribution of type I, beta prime distribution)*

Loi d'une variable aléatoire continue non élémentaire, ajustée par deux paramètres et qui intervient dans divers problèmes, par exemple dans la théorie des intervalles de confiance pour la loi binomiale.

**Formulaire**

Deux paramètres réels  $r > 0$  et  $s > 0$  ; valeurs sur l'intervalle  $[0, 1]$  (éventuellement bornes exclues).

► **Loi de probabilité**

– densité : 
$$f(x) = \frac{1}{B(r, s)} x^{r-1} (1-x)^{s-1} \quad (0 < x < 1)$$

– fonction de répartition : 
$$F(x) = \int_0^x f(t) dt \quad (0 \leq x \leq 1)$$

la fonction B (= bêta majuscule grec) est définie à partir de la fonction gamma par

$$B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}.$$

► **Valeurs caractéristiques**

– espérance : 
$$E(X) = \frac{r}{r+s},$$

– variance : 
$$\text{Var}(X) = \frac{rs}{(r+s+1)(r+s)^2},$$

– écart-type : 
$$\sigma(X) = \frac{\sqrt{\frac{rs}{r+s+1}}}{r+s}$$

► **Cas particulier**

Pour  $r = s = 1$ , la loi bêta est la loi uniforme continue sur  $[0, 1]$ .

Le minimum et le maximum de  $n$  v.a. uniformes sur  $[0, 1]$  et indépendantes suivent des lois bêta de type I, respectivement de paramètres 1 et  $n$ , et  $n$  et 1.

Voir *gamma (fonction)*.

**bêta de type II (loi)***(beta distribution of type II, beta prime distribution)*

Loi d'une variable aléatoire continue non élémentaire, dérivée de la loi bêta de type I, et qui peut modéliser de nombreux phénomènes aléatoires positifs.

Si  $Y$  est une v.a. qui suit une loi bêta de type I et de paramètres  $r$  et  $s$ , le quotient  $X = \frac{Y}{1-Y}$  suit une loi bêta de type II et de mêmes paramètres.

A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

**Formulaire**

Deux paramètres réels  $r > 0$  et  $s > 0$  ; valeurs sur les réels positifs.

► Loi de probabilité

– densité : 
$$f(x) = \frac{1}{B(r, s)} \cdot \frac{x^{r-1}}{(1+x)^{r+s}} \quad (x \geq 0)$$

– fonction de répartition : 
$$F(x) = \int_0^x f(t) dt \quad (x \geq 0)$$

► Valeurs caractéristiques

– espérance :  $E(X) = \frac{r}{s-1}$  (si  $s > 1$ ),

– variance :  $Var(X) = \frac{r(r+s-1)}{(s-1)^2(s-2)}$  (si  $s > 2$ ),

– écart-type :  $\sigma(X) = \frac{\sqrt{\frac{r(r+s-1)}{s-2}}}{s-1}$  (si  $s > 2$ ).

Le quotient de v.a. de lois gamma de paramètres respectivement  $r$  et  $s$  suit une loi bêta de type II, de paramètres  $r$  et  $s$ .

**biais**

(*bias*)

Dans une situation d'estimation d'un paramètre  $\theta$  d'une loi de probabilité par une variable aléatoire « estimateur »  $Y_n = Y_n(X_1, X_2, \dots, X_n)$ , le biais désigne la différence  $E(Y_n) - \theta$ . Voir *débiaisée (estimation)*.

**Bienaymé-Tchebychev (inégalité de)**

(*Chebyshev inequality*)

Soit  $X$  une variable aléatoire réelle d'espérance  $\mu$  et d'écart-type  $\sigma$ . Alors, pour tout  $t > 0$ , on a :

$$P(|X - \mu| \geq t\sigma) \leq \frac{1}{t^2},$$

forme alternative équivalente, pour tout  $a > 0$  :

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Cette inégalité permet de démontrer la loi faible des grands nombres, et son intérêt théorique est donc grand.

Mais si on veut l'utiliser numériquement, son intérêt est limité aux variables aléatoires  $X$  sur lesquelles on ne sait rien (d'autre que l'existence et la valeur de  $\mu$  et de  $\sigma$ ) : cette inégalité est alors le seul renseignement sur la décroissance de  $P(|X - \mu| > t\sigma)$  lorsque  $t$  tend vers l'infini. En revanche, dès que l'on sait quelque chose sur  $X$ , on peut faire mieux... On comparera par exemple  $P(|X - \mu| > 3\sigma) \leq 0,1111$  (inégalité de Bienaymé-Tchebychev) et, pour une v.a. normale  $X$ ,  $P(|X - \mu| > 3\sigma) = 0,0027$ .

Cette inégalité est souvent démontrée à partir de l'*inégalité de Markov* :

Soit  $X$  une variable aléatoire réelle qui possède un moment absolu d'ordre 1 :  $M_1 = E(|X|)$ . Alors, pour tout  $a > 0$ , on a :

$$P(|X| \geq a) \leq \frac{M_1}{a}.$$

(il est équivalent de supposer  $X$  positive et de considérer l'espérance  $\mu = E(X)$ ).

Ces inégalités possèdent des généralisations, désignées le plus souvent sous le nom générique d'*inégalité de Tchebychev*.

## bilatéral

(*two-sided, double-tailed*)

Qualifie un *test paramétrique* où l'on teste l'hypothèse simple  $H_0 = \langle \theta = \theta_0 \rangle$  contre l'hypothèse alternative bilatérale  $H_1 = \langle \theta \neq \theta_0 \rangle$  (à comprendre donc comme  $H_1 = \langle \theta < \theta_0 \rangle$  ou  $\langle \theta > \theta_0 \rangle$ ).

## bimodale (distribution)

Voir *mode*.

## binomial (test)

Nom parfois donné au *test de Student* de comparaison de pourcentages (probabilités).

## binomiale (loi)

(*binomial distribution*)

Loi d'une variable aléatoire discrète de compte qui intervient notamment dans les épreuves répétées.

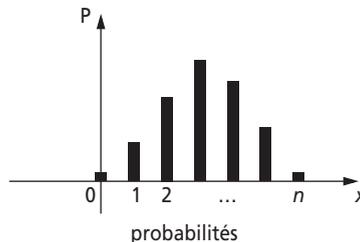
### Formulaire

Deux paramètres réels :  $n$  (entier  $\geq 1$ ) qui représente un nombre d'épreuves ou de tirages,  $p$  ( $0 \leq p \leq 1$ ) qui représente une probabilité, et une notation systématiquement utilisée :  $q = 1 - p$ .

Soit  $N$  la variable aléatoire binomiale  $\mathcal{B}(n, p)$  ; valeurs prises : 0, 1, ...,  $n$ .

#### ► Loi de probabilité

$$P(N = k) = \binom{n}{k} p^k q^{n-k}$$



#### ► Valeurs caractéristiques

- espérance :  $E(N) = np$ ,
- variance :  $\text{Var}(N) = npq$ ,
- écart-type :  $\sigma(N) = \sqrt{npq}$ .

### ► Utilisations

La loi binomiale est la loi de la somme de  $n$  variables aléatoires de Bernoulli semblables et indépendantes. C'est donc la loi du compte des événements dans des épreuves répétées, et c'est pareillement la loi du compte d'un caractère dans des tirages « AVEC remise ».

La loi binomiale est une *approximation* de la loi hypergéométrique (qui est la loi du compte d'un caractère dans des tirages « SANS remise ») lorsque le paramètre  $N$  de cette loi est « grand ».

**Exemple 1** Le nombre de filles dans une famille de 6 enfants, sachant que la probabilité de naissance d'une fille est 0,51 (et en supposant que les sexes des enfants successifs soient indépendants), suit une loi binomiale  $\mathcal{B}(6, 0,51)$ . On peut calculer par exemple la probabilité qu'il y ait (exactement) 4 filles :  $\binom{6}{4} \times 0,51^4 \times 0,49^2 = 15 \times 0,06765 \times 0,2401 = 0,3249$ .

**Exemple 2** Le nombre annuel  $N$  d'accidents à un carrefour donné, sachant qu'il y a chaque jour une chance sur 125 d'accident, suit une loi binomiale  $\mathcal{B}(365, 0,008)$ . On peut calculer par exemple  $E(N) = 365 \times 0,008 = 2,92$  et  $\sigma(N) = \sqrt{365 \times 0,008 \times 0,992} = 1,70$ .

## binomiale négative (loi)

(negative binomial distribution)

Synonyme *loi de Pólya*.

Loi d'une variable aléatoire discrète de compte du nombre d'échecs précédant le  $s$ -ième succès dans des épreuves répétées.

### Formulaire

Deux paramètres réels  $s$  (entier  $\geq 1$ ) qui représente le nombre de succès qui contrôle la loi,  $p$  ( $0 \leq p \leq 1$ ) qui représente une probabilité (notation standard :  $q = 1 - p$ ).

Soit  $N$  la variable aléatoire binomiale négative de paramètres  $s$  et  $p$  ; valeurs prises : 0, 1, 2, ...

#### ► Loi de probabilité

$$P(N = k) = \binom{s+k-1}{s-1} p^s q^k$$

#### ► Valeurs caractéristiques

– espérance :  $E(N) = \frac{sq}{p}$ ,

– variance :  $\text{Var}(N) = \frac{sq}{p^2}$ ,

– écart-type :  $\sigma(N) = \frac{\sqrt{sq}}{p}$

### ► Utilisation

La loi binomiale négative est le nombre d'échecs avant le  $s$ -ième succès dans des épreuves répétées, ou dans des tirages « AVEC remise » (le premier compte possible est  $N = 0$ ).

Si  $N$  est la variable aléatoire binomiale négative de paramètres  $s$  et  $p$ , alors  $T = N + s$  est une variable aléatoire de Pascal de mêmes paramètres  $s$  et  $p$ .

Cette loi s'appelle loi binomiale *négative* parce que, avec la définition généralisée du coefficient binomial  $\binom{x}{m} = \frac{x(x-1)\dots(x-m+1)}{m!}$ , valable pour  $x$  quelconque et  $m$  entier  $\geq 1$ , on peut écrire :

$$P(N = k) = \binom{s+k-1}{s-1} p^s q^k = \binom{-s}{k} p^s (-q)^k.$$

On notera enfin que, si l'on pose  $P = \frac{1}{p'}$ ,  $Q = \frac{q}{p'}$ ,  $P(N = k)$  est égal à  $\binom{s+k-1}{s-1} p^k Q^{-s-k}$ , qui est le coefficient du terme général du développement de  $(P - Q)^{-s}$ .

## binomiaux (coefficients), binôme (coefficients du) (*binomial coefficients*)

Synonyme de *combinaisons (combinations)*.

Entiers positifs dépendant de deux paramètres entiers positifs  $n$  et  $k$  ( $n \geq 0$ ,  $0 \leq k \leq n$ ), intervenant d'une part dans le « développement du binôme », d'autre part dans les dénombrements.

Les coefficients binomiaux possèdent plusieurs définitions combinatoires équivalentes (et très proches).

1. On considère un ensemble  $E$  de  $n$  objets : le nombre de combinaisons, noté  $\binom{n}{k}$ , représente le nombre de manières de prendre (sous-entendu : sans répétition)  $k$  objets parmi les  $n$ .
2. On considère un ensemble  $E$  de  $n$  éléments :  $\binom{n}{k}$  représente le nombre de sous-ensembles de  $k$  éléments de  $E$ .
3. On considère un ensemble  $E$  de  $n$  éléments :  $\binom{n}{k}$  représente le nombre de manières différentes de répartir ces objets en deux classes, contenant respectivement  $k$  et  $n - k$  objets.

$$C_n^k = \binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

### ► Notations

$C_n^k$  est une notation ancienne, mais encore très utilisée dans l'enseignement français,  $\binom{n}{k}$  est la notation moderne, justifiée notamment par le «  $n$  » en regard du numérateur de la formule et le «  $k$  » en regard du dénominateur.

### ► Cas particulier et propriété importante

$$\binom{n}{0} = \binom{n}{n} = 1$$

$$\binom{n}{n-k} = \binom{n}{k}$$

### Développement (formule) du binôme

$$(x + y)^n = \sum_{k=0}^{k=n} \binom{n}{k} x^{n-k} y^k$$

Dans le contexte probabiliste (dénombrements, loi binomiale), les deux termes, coefficients binomiaux et combinaisons, sont employés concurremment.

**Exemple 1** Le nombre de délégations différentes de 3 personnes prises dans un groupe de 20 est  $\binom{20}{3} = \frac{20 \times 19 \times 18}{1 \times 2 \times 3} = 1\,140$ .

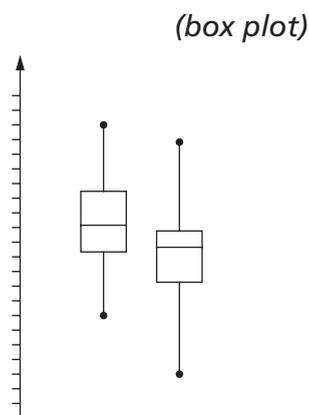
**Exemple 2** Le nombre d'échantillons différents de 900 personnes que l'on peut extraire d'une population de 60 000 000 est  $\binom{60\,000\,000}{900}$  (si on calcule ce nombre, on trouve  $3,18 \times 10^{4\,730}$ ).

Voir *arrangements, permutations, triangle de Pascal*.

## boîte de dispersion, boîte à moustaches

Représentation graphique symbolique de la médiane et des quartiles d'une distribution statistique. Si  $Q_0$ ,  $Q_1$ ,  $Q_2$ ,  $Q_3$  et  $Q_4$  sont respectivement l'extrémité inférieure, le premier quartile, la médiane, le troisième quartile et l'extrémité supérieure, on indique l'axe des ordonnées et on dessine à côté : un trait vertical de  $Q_0$  à  $Q_1$ , trois traits horizontaux en  $Q_1$ ,  $Q_2$  et  $Q_3$ , complétés en « boîte », et enfin un trait vertical de  $Q_3$  à  $Q_4$ . On peut imaginer diverses variantes...

L'expression « boîte à moustaches » se réfère au dessin d'une boîte de dispersion étalée au-dessus d'un axe horizontal.



## Borel (Émile)

Mathématicien et homme politique français (1871-1956). Il développa la notion de mesure, introduisit la convergence « presque sûre », et démontra la loi forte des grands nombres. Il fit également des travaux en analyse (théorie des fonctions, sommation des séries).

## Borel-Cantelli (lemme de)

(Borel-Cantelli lemma)

On considère une suite  $(A_n)$  d'évènements définis sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ . Alors :

- si la série numérique  $\sum P(A_n)$  converge (les mathématiciens écrivent  $\sum_{n=1}^{\infty} P(A_n) < +\infty$ ), il n'existe presque sûrement (*i.e.* avec probabilité 1) qu'un nombre fini de  $A_n$  qui se réalisent ;
- si la série numérique  $\sum P(A_n)$  diverge (les mathématiciens écrivent  $\sum_{n=1}^{\infty} P(A_n) = +\infty$ ), et si de plus les  $A_n$  sont indépendants, il existe presque sûrement (*i.e.* avec probabilité 1) une infinité de  $A_n$  qui se réalisent.

En se restreignant au cas d'une suite d'évènements indépendants, ce lemme est parfois appelé la *loi du zéro-un de Borel-Cantelli*.

**Exemple** Dans une suite infinie de parties indépendantes de Pile ou Face, un « motif » donné, quelle que soit sa longueur, apparaît presque sûrement ; il apparaît même une infinité de fois presque sûrement. De même, si l'on émet (si l'on frappe sur un clavier...) aléatoirement et de façon indépendante les lettres de l'alphabet et les signes de ponctuation, un texte donné, quelle que soit sa longueur, apparaît presque sûrement, il apparaît même une infinité de fois presque sûrement.

Voir *zéro-un (loi du – de Kolmogorov)*.

## boréliens

(*Borel set*)

*Tribu de  $\mathbf{R}$*  engendrée par les intervalles fermés bornés.

## branchement (processus de)

Voir *Galton-Watson (processus de)*.

## Bravais-Pearson (coefficient de)

(*Pearson coefficient,  
Bravais coefficient*)

Ancien nom du coefficient de corrélation linéaire.

Voir *corrélation (coefficient de)*.



**caractère**

Voir *variable statistique*.

**caractéristique (valeur)**

Voir *indicateur*.

**caractéristique (fonction)** *(characteristic function)*

Fonction complexe de variable réelle que l'on peut associer (bijectivement) à toute variable aléatoire réelle et qui est un outil privilégié d'étude des lois de probabilité.

Soit X une variable aléatoire réelle. On appelle fonction caractéristique de X la fonction  $\varphi_X$  de la variable réelle  $t$  définie par

$$\varphi_X(t) = E(e^{itX}).$$

Expression dans le cas discret. X prend les valeurs  $x_k$  avec les probabilités  $p_k$  :

$$\varphi_X(t) = \sum_k p_k e^{itx_k}$$

Expression dans le cas absolument continu. X possède la densité  $f(x)$  :

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx$$

Lorsque la loi est symétrique, la fonction caractéristique est réelle.

La fonction caractéristique est liée aux moments de X. Plus précisément, on a  $\varphi_X(0) = 1$  et, pour tout  $k$  pour lequel la dérivée  $k$ -ième existe :

$$\varphi_X^{(k)}(0) = i^k E(X^k).$$

► Comportement par transformation affine

$$\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at).$$

► Comportement par addition de v.a. indépendantes

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t),$$

► Fonction caractéristique des lois les plus fondamentales

- loi singulière concentrée en  $x_0$  :  $e^{itx_0}$
- loi de Bernoulli de paramètre  $p$  :  $q + pe^{it}$
- loi binomiale de paramètres  $n$  et  $p$  :  $(q + pe^{it})^n$
- loi de Poisson de paramètre  $\mu$  :  $e^{\mu(e^{it}-1)}$
- loi uniforme sur  $[-a, a]$  :  $\frac{\text{sint } at}{at}$

- loi exponentielle de paramètre $\lambda$ :	$\frac{\lambda}{\lambda - it}$
- loi normale centrée réduite	$e^{-t^2/2}$
- loi normale de paramètres $\mu$ et $\sigma$	$e^{i\mu t} e^{-\sigma^2 t^2/2}$

Voir *génératrice des moments (fonction)*.

## Cardan (Gerolamo)

Médecin, mathématicien et astrologue italien (1501–1576). Il publia une étude sur la durée de vie humaine et esquaissa le concept de probabilité dans son ouvrage *Liber de Ludo Aleae* (publié un siècle plus tard !). Il fit également des travaux en algèbre (équation du troisième degré).

## Cauchy (loi de)

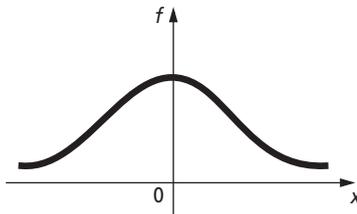
(*Cauchy distribution*)

Loi d'une variable aléatoire continue qui peut être définie comme le quotient de deux v.a. normales indépendantes, centrées et de même écart-type.

### Formulaire

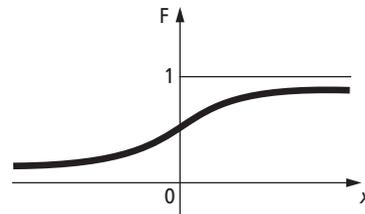
Version standardisée X sans paramètre, à valeurs sur  $\mathbf{R}$ .

#### ► Loi de probabilité



densité

$$f(x) = \frac{1}{\pi(1+x^2)}$$



fonction de répartition

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \text{Arc tan } x$$

Les intégrales qui définissent l'espérance mathématique et la variance de cette loi ne convergent pas, de sorte qu'une v.a. de Cauchy ne possède ni espérance ni variance ni écart-type.

#### ► Valeurs caractéristiques

Les seuls indicateurs de tendance centrale et de dispersion que l'on peut calculer sont la médiane (égale à 0) et les quartiles (égaux à -1 et à 1)

Cette loi est notamment celle du quotient de deux v.a. normales réduites et indépendantes, ainsi que celle de  $\tan U$  où  $U$  est une v.a. uniforme dans  $]-\frac{\pi}{2}, \frac{\pi}{2}[$ . C'est également la loi de la variable de Student  $T_1$  à 1 degré de liberté.

Parmi les « pathologies » de cette loi, on peut signaler que l'inverse d'une v.a. de Cauchy est une v.a. de Cauchy identique, que la loi de la somme de  $n$  v.a. de Cauchy « copies » de  $X$  suit la loi de  $nX$ , et donc que la moyenne suit la loi de  $X$ .

## causes (probabilités des)

Ancienne dénomination du (*théorème de*) Bayes.

## censurées (données)

(*censored data*)

Qualifie une situation où il est impossible (ou trop coûteux) d'observer la totalité des valeurs d'un échantillon. Une situation typique est celle des essais de fiabilité, où l'on considère un échantillon de variables qui sont des durées de vie (ou des temps d'attente de (première) défaillance) de systèmes tous mis en fonctionnement à l'instant  $t = 0$ . La censure peut porter sur la durée d'observation (on arrête les observations à l'instant  $t = \theta$ ), ou sur le nombre d'observations (on arrête les observations après la  $r$ -ième observation).

La théorie de l'estimation des paramètres repose essentiellement sur la méthode du maximum de vraisemblance. On donne ci-dessous les formules de base dans l'unique cas de la loi exponentielle (qui correspond en théorie de la fiabilité à la situation standard à taux de défaillance constant).

### Estimation de l'espérance mathématique $\tau = \frac{1}{\lambda}$ d'une loi exponentielle de paramètre $\lambda$

#### ► Rappel du cas non censuré

- échantillon de  $n$  observations individualisées  $t_1, t_2, \dots, t_n$  ;
- estimation standard de  $\tau$  par la moyenne :

$$\hat{\tau} = \frac{t_1 + t_2 + \dots + t_n}{n}$$

#### ► Censure à l'instant $t = \theta$

- $r$  valeurs ont été observées, formant un échantillon tronqué  $t_1, t_2, \dots, t_r$  ;
- estimation de  $\tau$  par le quotient :

$$\hat{\tau} = \frac{t_1 + \dots + t_r + (n - r)\theta}{r}$$

#### ► Censure au $r$ -ième « décès »

- échantillon tronqué  $t_1, t_2, \dots, t_r$ , que l'on suppose ordonné (donc  $t_r$  est la plus grande valeur observée) ;
- estimation de  $\tau$  par le quotient :

$$\hat{\tau} = \frac{t_1 + \dots + t_r + (n - r)t_r}{r}$$

Dans tous les cas, le paramètre  $\lambda$  sera estimé par l'inverse de  $\hat{\tau}$ .

Malgré leur apparence similaire, les deux dernières formules sont très différentes : dans la première,  $r$  est la valeur observée d'une variable aléatoire  $R$ , dans la seconde, la valeur de  $r$  est fixée par l'expérimentateur.

## centile

(*percentile*)

Indicateur de position attaché à une variable aléatoire réelle, utilisé essentiellement en statistique. Les centiles partagent la série des valeurs en deux parties de fractions  $\alpha$  et  $1 - \alpha$  de l'effectif total, pour  $\alpha = 1\%$ ,  $2\%$ , ...,  $98\%$ ,  $99\%$ . Ils sont au nombre de 99.

Si la signification concrète des centiles est simple et « parlante », la traduction formelle est plus délicate. On adaptera sans difficulté le formulaire détaillé pour la *médiane*.

## central limite (théorème)

(central limit theorem)

**Théorème.** On considère une suite  $(X_n)$  de variables aléatoires (réelles) indépendantes et identiquement distribuées, d'espérance mathématique  $\mu$  et de variance  $\sigma^2$ . On définit les moyennes :

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

puis les variables centrées réduites correspondantes :

$$Z_n = \frac{M_n - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)},$$

et enfin les fonctions de répartition :

$$F_n(x) = P(Z_n \leq x).$$

Alors, pour tout  $x \in \mathbf{R}$  :

$$F_n(x) \rightarrow F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \text{ quand } n \rightarrow \infty.$$

( $F(x)$  est la fonction de répartition d'une v.a. normale centrée réduite)

C'est très exactement la définition de la convergence en loi de la suite  $(Z_n)$  vers une v.a. normale centrée réduite lorsque  $n$  tend vers l'infini.

Une conséquence immédiate est l'expression limite des probabilités d'intervalles :

$$P(a < Z_n \leq b) \rightarrow F(b) - F(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt \text{ quand } n \rightarrow \infty.$$

On notera que l'on peut aussi bien calculer  $Z_n$  à partir de la somme  $S_n = X_1 + \dots + X_n$  :

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Comme pour la loi des grands nombres, il existe des versions du théorème central limite avec des conditions affaiblies.

Il faut enfin signaler que, dans la pratique, on s'autorise à appliquer le théorème central limite pour approcher  $S_n$  et  $M_n$ , lorsque  $n$  est « assez grand », par des variables normales de mêmes paramètres. Cela soulève des problèmes délicats ; il existe des résultats, notamment le (théorème de) *Berry-Esséen*, qui permettent de majorer l'erreur commise.

## centre

(midpoint, class mark, mid-range)

Le centre d'une *série statistique* est la demi-somme des valeurs extrêmes. C'est une caractéristique de tendance centrale médiocre car elle est trop sensible aux valeurs aberrantes (erronées ou exceptionnelles).

Le centre d'une *classe* constituée par un intervalle  $]a_j, a_{j+1}]$  est le milieu  $\frac{a_j + a_{j+1}}{2}$ . Dans un certain nombre de formules (calcul d'une moyenne empirique, calcul d'une variance empirique, ...), on remplace la valeur exacte des éléments par le centre de la classe à laquelle ils appartiennent. Cela introduit un biais, le plus souvent très faible, mais qui peut être corrigé (*correction de Sheppard*).

Synonyme de *milieu*.

**centrée (variable aléatoire)** (*centred random variable*)

Se dit soit d'une variable aléatoire  $X$  dont l'espérance  $E(X)$  est nulle, soit, étant donnée une variable aléatoire quelconque  $X$ , de la variable translatée  $X - E(X)$ .

**centrée réduite (variable aléatoire)** (*standardized random variable, centred and normed random variable*)

Se dit soit d'une variable aléatoire  $X$  dont l'espérance  $E(X)$  est nulle et l'écart-type  $\sigma(X)$  égal à 1, soit, étant donnée une variable aléatoire quelconque  $X$ , de la variable translatée et homogénéisée  $\frac{X - E(X)}{\sigma(X)}$ .

**certain (événement)** (*certain set*)

Évènement *plein*  $\Omega = \ll \text{n'importe quoi s'est passé} \gg$  d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , de probabilité égale à 1 (c'est bien sûr un cas limite, mais il est nécessaire de l'inclure dans l'ensemble  $\mathcal{A}$  de tous les événements envisageables).

**certaine (variable aléatoire)** (*certain random variable*)

Variable aléatoire qui prend une valeur unique avec probabilité 1 (c'est bien sûr un cas limite, mais il est nécessaire de l'inclure dans l'ensemble de toutes les variables aléatoires envisageables – il serait inapproprié de la dénommer variable aléatoire constante !). Le nom mathématique officiel est *variable aléatoire singulière* (ou *dégénérée*).

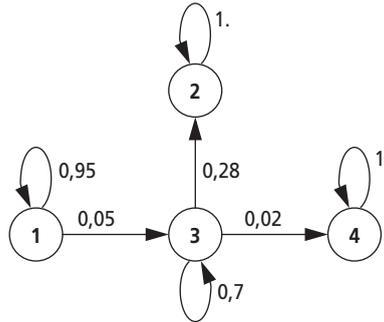
**chaîne de Markov** (*Markov chain*)

Modèle probabiliste d'une suite de « changements d'états » qui se produisent aux instants  $t = 1, 2, \dots, n, \dots$ , et qui vérifient la propriété fondamentale suivante : l'évolution après l'instant  $n$  ne dépend que de la situation à cet instant  $n$  (mais non de la « trajectoire » entre  $t = 0$  et  $t = n - 1$ ).

Une chaîne de Markov est un cas particulier de *processus aléatoire*. On appelle *markoviens* les processus qui satisfont de façon générale la propriété que « le futur ne dépend que du présent mais non du passé », propriété qui s'écrit formellement avec des probabilités conditionnelles. Lorsqu'un processus markovien est à temps discrets et à valeurs discrètes finies ou infinies dénombrables, on l'appelle chaîne de Markov (certains auteurs ont une définition plus extensive).

Une chaîne de Markov  $(X_n)$  est définie par un ensemble  $\mathcal{E} = \{E_i\}$  fini ou dénombrable, appelé *espace des états*, et, pour chaque instant  $n$ , des *probabilités de transition* qui constituent la loi conditionnelle du passage de la situation à  $t = n$  à la situation à  $t = n + 1$ . Lorsque cette loi de passage ne dépend pas de  $n$ , on dit que la chaîne de Markov est homogène (sous-entendu : dans le temps). Dans ce cas, il y a un unique ensemble  $\{p_{ij}\}$  de probabilités de transition :  $P(X_n = E_i \text{ et } X_{n+1} = E_j) = P(X_n = E_i) \times p_{ij}$  (les probabilités de transition vérifient pour tout  $i$  la propriété  $\sum_j p_{ij} = 1$ , simple contrainte de cohérence).

**Exemple** On considère un modèle épidémiologique individuel à 4 états : 1 = sain non immunisé, 2 = sain immunisé, 3 = malade, 4 = mort. On représente généralement une chaîne de Markov par un « graphe » avec des cercles qui figurent les états, des flèches qui figurent les transitions (changements d'état) ayant une probabilité non nulle, et on porte sur les flèches les probabilités de transition.



On peut constater dans cet exemple que les états 2 et 4 sont « absorbants » (un état  $E_i$  est dit absorbant si  $p_{ii} = 1$ ). On peut poser des questions telles que : si on se place initialement dans l'état 1, quelle est la probabilité d'atteindre l'état 2 (et d'y rester), quelle est la probabilité d'atteindre l'état 4 (et d'y rester) ? ou encore : quelle est l'espérance mathématique (si elle existe) du temps d'attente pour aller de l'état 1 à l'état 2, ou de l'état 1 à l'état 4 ?

## chi-deux

Une des deux formes francisées du symbole grec  $\chi^2$  (prononcé ki deux)

Voir *khi-deux*.

## chronologique (série)

Voir *série chronologique*.

## classe

(*class*)

Valeur (ou modalité) ou ensemble de valeurs (ou de modalités) que peut prendre une variable statistique (ou caractère) et qui sert notamment à constituer des tableaux « par classes et effectifs ».

## classe modale

Voir *modale (classe)*.

## classes et effectifs (tableau par)

Voir *tableau par classes et effectifs*.

## coefficient

Voir *aplatissement (coefficient d')*, *asymétrie (coefficient d')*, *corrélation (coefficient de)*, *détermination (coefficient de)*.

## coefficients binomiaux

Voir *binomiaux (coefficients)*.

## cohorte

(*cohort*)

Type particulier d'échantillon, utilisé notamment en épidémiologie, sociologie ou démographie, constitué par un ensemble d'individus choisis à un moment donné, et « suivis » ensuite pendant une période le plus souvent de plusieurs années. Le choix initial est effectué parmi

des individus possédant une même caractéristique (être indemne d'une certaine pathologie, être dans une même année d'études, avoir le même âge, ...). Les études de cohortes sont parfois appelées enquêtes « longitudinales ».

## combinaisons

Voir *binomiaux (coefficients)*.

## combinatoire

Voir *analyse combinatoire*.

## comparaison de moyennes (test de)

Voir *Student (test de)*.

## comparaison de pourcentages / probabilités (test de)

Voir *Student (test de)*.

## comparaison d'une variance à une valeur fixée

Test paramétrique qui compare la variance observée d'un échantillon statistique à une valeur fixée.

— test bilatéral de comparaison d'une variance  $\sigma^2$  à une valeur fixée  $\sigma_0^2$  —

- Données. Un échantillon  $(x_1, x_2, \dots, x_n)$  de  $n$  valeurs observées d'une variable aléatoire normale  $X$  d'espérance mathématique  $\mu$  et de variance  $\sigma^2$ .

- Hypothèse testée.  $H_0 = \langle \sigma^2 = \sigma_0^2 \rangle$  contre  $H_1 = \langle \sigma^2 \neq \sigma_0^2 \rangle$

- Déroulement technique du test

1. On calcule la moyenne  $\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$  de l'échantillon, puis on calcule la variance non

biaisée  $s^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - \bar{x})^2$  de l'échantillon.

2. On calcule la valeur observée de la variable de test :

$$t = n \frac{s^2}{\sigma_0^2}$$

Les valeurs de référence de la variable de test sont deux bornes  $l_1$  et  $l_2$ , à lire dans les tables de la loi du khi-deux à  $n-1$  degrés de liberté :  $l_1$  est la valeur du  $\chi^2$  pour la probabilité  $1 - \frac{\alpha}{2}$ , et  $l_2$  la valeur pour la probabilité  $\frac{\alpha}{2}$ . L'hypothèse  $H_0$  est rejetée si  $t$  est à l'extérieur de l'intervalle  $[l_1, l_2]$ .

- Conditions et précautions. Contrairement au cas d'une espérance ou d'une probabilité, ce test n'est pas robuste : il n'est valable que dans le cas où la loi de l'échantillon est normale.

## comparaison de deux variances

Voir *Fisher-Snedecor (test de)*.

### complémentaire

(*complementary set / event*)

Dans la formalisation ensembliste des espaces probabilisables, les évènements sont des parties de l'espace fondamental  $\Omega$ . Si l'on considère un évènement  $A$ , son complémentaire  $\complement A$  est l'évènement dont la réalisation correspond à la *négation logique* « non- $A$  » :

$$\omega \in \complement A \Leftrightarrow \text{non} (\omega \in A).$$

Outre les deux notations ensembliste et logique, parfaitement synonymes :  $\complement A$ , non- $A$ , on emploie souvent la notation  $\bar{A}$  qui est très commode.

#### Formules (manipulation des évènements)

$$\begin{aligned}\complement(A \cup B) &= \complement A \cap \complement B \\ \complement(A \cap B) &= \complement A \cup \complement B \\ P(\complement A) &= 1 - P(A)\end{aligned}$$

Ces formules ont bien sûr leurs correspondants logiques.

La dernière formule est d'un usage très fréquent ; en effet, il arrive très souvent que la probabilité d'un évènement ne soit pas directement calculable, mais que la probabilité de son complémentaire le soit.

**Exemple** On considère la répartition des sexes des enfants dans les familles de 6 enfants, qui suit en première approximation une loi binomiale de paramètres  $n = 6$  et  $p = 0,5$  :  $p_k = P(k \text{ filles}) = \binom{6}{k} 0,5^6$ . On demande la probabilité  $P$  de l'évènement « au moins une fille ». Il serait très maladroit de calculer  $p_1 + p_2 + p_3 + p_4 + p_5 + p_6$ . Il suffit de constater que l'évènement « au moins une fille » est le complémentaire de « (exactement) 0 fille » et l'on obtient immédiatement  $P = 1 - p_0 = 1 - 0,5^6 = 0,984$ .

### composées (formule des probabilités)

(*composite probabilities formula*)

Version multiplicative de la définition d'une probabilité conditionnelle :

$$P(A \cap B) = P(A) P(B|A) \quad (= \text{aussi } P(B) P(A|B))$$

Cette formule se généralise :  $P(A \cap B \cap C) = P(A) P(B|A) P(C|(A \cap B))$ , etc. C'est ainsi que l'on calcule les probabilités lorsque l'on modélise une suite (logique ou temporelle) d'évènements, et que l'espace fondamental  $\Omega$  global est représenté par un *arbre* : la probabilité à chaque « nœud » de l'arbre s'obtient en multipliant la probabilité associée au nœud précédent par une probabilité conditionnelle.

### compte (variable aléatoire de)

(*counting variable*)

Variable aléatoire qui modélise dans des épreuves successives ou simultanées (épreuves répétées, échantillonnage ou sondage, « processus ») un compte total d'évènements (résultat d'un jeu, réponse d'un certain type à une question ou caractéristique particulière dans une enquête, défaut, panne, naissance, ...). Les lois les plus classiques pour une variable aléatoire de compte sont la *loi binomiale*, la *loi hypergéométrique* et la *loi de Poisson*.

## concentration (courbe de – de Lorentz) ([Lorentz] concentration curve)

On considère une variable réelle  $X$ , en principe positive, définie sur une population, et qui représente généralement un « bien » (salaire, revenu, patrimoine, actif, ...) susceptible d'être « possédé » par les individus. Étant donné une valeur  $x$  de cette variable (de ce bien), on peut lui associer la partie  $A(x)$  de la population constituée par les individus pour lesquels  $X \leq x$  (i.e. les individus qui possèdent chacun une quantité  $\leq x$  du bien). On note  $F(x)$  la proportion de la population de  $A(x)$  par rapport à la population totale, et  $FQ(x)$  la proportion du bien possédé par les individus de  $A(x)$  par rapport au bien total possédé par l'ensemble de la population.

La courbe de concentration est formée par l'ensemble des points d'abscisse  $F(x)$  et d'ordonnée  $FQ(x)$ . Dans le cas où l'on connaît seulement des classes de valeurs de  $X$ , cette courbe est une ligne brisée.

### Construction par points de la courbe

Le cas standard est le cas discret : les valeurs de  $X$  étant réparties en classes d'extrémités  $x_0 < x_1 < \dots < x_k$ , on se donne le nombre  $n_i$  d'individus qui possèdent une valeur de  $X$  entre  $x_{i-1}$  et  $x_i$  ( $i = 1, \dots, k$ ). On représente la valeur de  $X$  dans la classe

$]x_{i-1}, x_i]$  par son centre  $x'_i = \frac{x_{i-1} + x_i}{2}$ , et on pose  $q_i = n_i x'_i$  la quantité de  $X$

possédée par les individus de la classe n°  $i$ . La population totale est  $N = \sum_{i=1}^k n_i =$

$n_1 + n_2 + \dots + n_k$ , et la quantité totale  $Q$  de  $X$  possédée est  $Q = \sum_{i=1}^k n_i x'_i$ .

On définit tout d'abord les proportions (du nombre d'individus et de la fraction possédée) dans chaque classe :

$$f_i = \frac{n_i}{N}, \quad q_i = \frac{n_i x'_i}{Q}$$

On définit enfin, de  $j$  de 0 à  $k$ , les points de la courbe de concentration, qui ont pour coordonnées les proportions cumulées.

► pour  $j = 0$

- abscisse  $F_0 = 0$ ,
- ordonnée  $Q_0 = 0$

► pour  $j \geq 1$

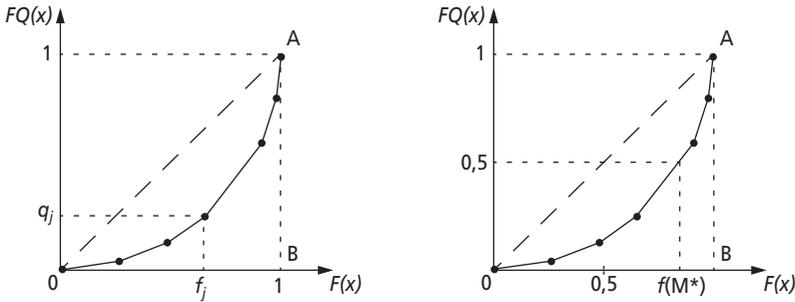
- abscisse  $F_j = \sum_{i=1}^j f_i$ , proportion de la population qui possède jusqu'à la valeur  $x_i$ ,
- ordonnée  $Q_j = \sum_{i=1}^j q_i$ , proportion de  $X$  possédée jusqu'à la valeur  $x_i$ ,

► *remarque* : pour  $j = k$  ces formules donnent

- abscisse  $F_k = 1$ ,
- ordonnée  $Q_k = 1$ .

On peut lire indirectement sur cette courbe la *médiane*  $M^*$  (la valeur de  $x$  qui partage en deux la masse totale « possédée ») :  $M^*$  est la valeur de  $x$  telle que  $FQ(x) = \frac{1}{2}$ . La courbe de concentration est toujours au-dessous de la « première diagonale » OA (cf. figure ci-dessous), et la médiane est toujours supérieure à la médiane  $M$  (caractérisée par  $F(M) = \frac{1}{2}$ ).

Lorsque l’amplitude des variations de  $X$  est faible, et que la « possession » de  $X$  est bien répartie, la courbe de concentration est voisine de OA. Lorsque l’amplitude des variations de  $X$  est grande, et/ou que la « possession » de  $X$  est très concentrée (chez les « gros » possédants), la courbe s’écarte de OA pour se rapprocher des côtés OBA. On peut traduire la distance entre OA et la courbe par un indicateur appelé *indice de concentration de Gini*.



On peut aussi définir la courbe de concentration dans le cas d’une distribution *continue* de  $X$ , de densité  $f(x)$ . On a  $Q = \int_0^\infty tf(t)dt$  et la courbe de concentration est la courbe paramétrique

d’équations  $F(x) = \int_0^x f(t)dt$  (fonction de répartition),  $FQ(x) = \frac{1}{Q} \int_0^x tf(t)dt$ .

**Exemple** On considère une entreprise de 150 personnes où les salaires (en k€ par mois) sont répartis comme indiqué par le tableau ci-dessous :

classe	[1,0, 1,2[	[1,2, 1,5[	[1,5, 2,0[	[2,0, 2,6[	[2,6, 3,4[	[3,4, 4,8[
effectif	15	63	42	18	9	3

Tracer la courbe de concentration.

Il faut commencer par calculer les valeurs et les proportions cumulées, comme détaillé dans le tableau ci-dessous :

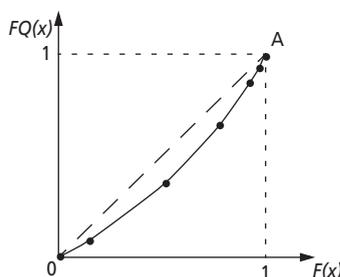
classe $[x_{i-1}, x_i[$	[1,0, 1,2[	[1,2, 1,5[	[1,5, 2,0[	[2,0, 2,6[	[2,6, 3,4[	[3,4, 4,8[
centre de classe $x'_i$	1,1	1,35	1,75	2,3	3,0	4,1
effectif $n_i$	15	63	42	18	9	3
effectif cumulé $\Sigma n_i$	15	78	120	138	147	150
proportion $F_i$	0,10	0,52	0,80	0,92	0,98	1,00

A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

masse $n_i x'_i$	16,5	85,05	73,5	41,4	27,0	12,3
masse cumulée $\Sigma n_i x'_i$	16,5	98,55	172,05	213,45	240,45	252,75
proportion de la masse $Q_i$	0,065	0,40	0,68	0,845	0,95	1,00

On obtient ainsi les points suivants :

- (0,0 , 0,0 )
- (0,10 , 0,065)
- (0,52 , 0,40 )
- (0,80 , 0,68 )
- (0,92 , 0,845)
- (0,98 , 0,95 )
- (1,0 , 1,0 )



**concentration (indice de – de Gini)** ([Gini] concentration index)

Étant donné une variable positive X définie sur une population, on peut définir la *courbe de concentration de Lorentz*.

L'indice de concentration de Gini est le double de l'aire comprise entre la courbe de concentration et la « première diagonale » du carré-unité.

**Calcul pratique de l'indice de concentration**

Le cas standard est le cas discret : les valeurs de X étant réparties en classes d'extrémités  $x_0 < x_1 < \dots < x_k$ , on se donne le nombre  $n_i$  d'individus qui possèdent une valeur de X entre  $x_{i-1}$  et  $x_i$  ( $i = 1, \dots, k$ ). On représente la valeur de X dans la classe

$[x_{i-1}, x_i]$  par son centre  $x'_i = \frac{x_{i-1} + x_i}{2}$ , et on pose  $q_i = n_i x'_i$  la quantité de X

possédée par les individus de la classe n° i. La population totale est  $N = \sum_{i=1}^k n_i =$

$n_1 + n_2 + \dots + n_k$ , et la quantité totale Q de X possédée est  $Q = \sum_{i=1}^k n_i x'_i$ .

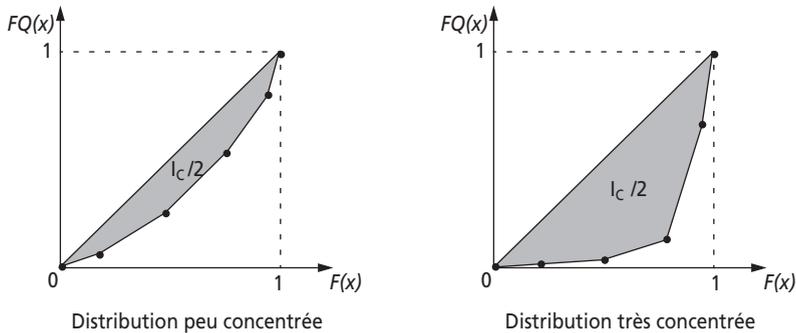
On calcule alors (cf. courbe de concentration) les proportions  $f_j$  et  $q_j$  dans les classes, et les proportions *cumulées*  $F_j$  et  $Q_j$ .

L'indice de concentration  $I_C$  est donné par l'une quelconque des deux formules équivalentes suivantes :

$$I_C = \sum_{j=1}^{k-1} (F_j Q_{j+1} - F_{j+1} Q_j),$$

$$I_C = 1 - \left( f_1 Q_1 + \sum_{j=2}^k f_j (Q_{j-1} + Q_j) \right).$$

C'est un nombre sans dimension (*i.e.* indépendant des unités), compris entre 0 et 1. La valeur de  $I_C$  est d'autant plus grande que l'amplitude des variations de  $X$  est grande, et/ou que la « possession » de  $X$  est très concentrée (chez les « gros » possédants), ces deux circonstances cumulant leurs effets.



**Théorème.** L'indice  $I_C$  de concentration est égal à l'espérance mathématique de la valeur absolue de la différence du caractère  $X$  entre deux individus (choisis indépendamment), soit, si  $X_1$  et  $X_2$  sont deux « copies » indépendantes de  $X$  :

$$I_C = E(|X_1 - X_2|).$$

Outre sa valeur en tant qu'interprétation concrète, ce théorème peut servir au calcul direct de  $I_C$  sans calculer aucune proportion :

$$I_C = \frac{1}{NQ} \sum \sum_{i < j} n_i n_j (x'_j - x'_i)$$

Ce théorème sert enfin à donner dans le cas continu une formule de calcul de  $I_C$  par une intégrale double.

**Exemple** On considère une entreprise de 150 personnes où les salaires (en k€ par mois) sont répartis comme indiqué par le tableau ci-dessous :

classe	[1,0, 1,2[	[1,2, 1,5[	[1,5, 2,0[	[2,0, 2,6[	[2,6, 3,4[	[3,4, 4,8[
effectif	15	63	42	18	9	3

(Les données sont les mêmes que celles de l'exemple donné *ante* pour la courbe de concentration).

Calculer l'indice de concentration.

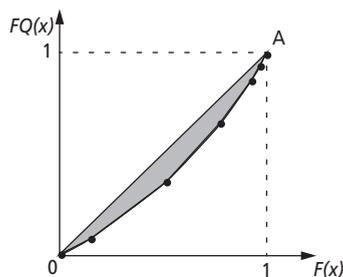
Les valeurs cumulées ont déjà été calculées (en particulier  $N = 150$  et  $Q = 252,75$ ). On redonne la partie du tableau utile pour les calculs :

classe $[x_{j-1}, x_j[$	[1,0, 1,2[	[1,2, 1,5[	[1,5, 2,0[	[2,0, 2,6[	[2,6, 3,4[	[3,4, 4,8[
centre de classe $x'_j$	1,1	1,35	1,75	2,3	3,0	4,1
proportion $F_j$	0,10	0,52	0,80	0,92	0,98	1,00
proportion de la masse $Q_j$	0,065	0,40	0,68	0,845	0,95	1,00

Calcul de  $I_C$  par la première formule :

$$\begin{aligned} I_C &= \sum_{j=1}^5 (F_j Q_{j+1} - F_{j+1} Q_j) \\ &= (f_1 q_2 - f_2 q_1) + \dots + (f_5 q_6 - f_6 q_5) \\ &= 0,0062 + 0,0336 + 0,0504 + 0,0459 + 0,0300 \\ &= 0,1661 \end{aligned}$$

(la concentration n'est pas très forte, comme on peut d'ailleurs le constater visuellement sur la figure).



## conditions de Yule

Voir Yule (*conditions de*).

## conditionnel, le

(*conditional*)

Premier sens : dans un espace probabilisé, qualifie les concepts modifiés par la connaissance d'un évènement fixé : évènements conditionnels, probabilités conditionnelles.

Deuxième sens (voisin pour l'« esprit » mais nettement distinct pour la situation de référence) : dans une situation probabiliste ou statistique « bidimensionnelle » ou « multidimensionnelle », qualifie ce qui concerne une des variables lorsque l'autre est fixée : probabilités conditionnelles (en un sens restreint par rapport au sens général ci-dessus), densité conditionnelle, loi conditionnelle, effectifs conditionnels, fréquences conditionnelles, espérance ou moyenne conditionnelle, variance conditionnelle, écart-type conditionnel.

Dans une situation statistique avec deux variables discrètes représentées par un tableau « de contingence », les effectifs conditionnels sont ceux des lignes ou des colonnes, et l'on calcule les fréquences conditionnelles en effectuant les quotients par l'effectif total de la ligne ou de la colonne (effectif marginal).

## conditionnelle (espérance, variance)

On considère un couple  $(X, Y)$  de variables aléatoires *discrètes*, dont la loi est donnée par les couples  $(x_i, y_j)$  de valeurs prises, et les probabilités ponctuelles (« loi jointe ») :

$$p_{ij} = P(X = x_i \text{ et } Y = y_j).$$

Les *lois marginales* (avec des notations pas très cohérentes mais très utilisées) sont calculées comme suit :

$$p_i = P(X = x_i) = \sum_j p_{ij}, \quad q_j = P(Y = y_j) = \sum_i p_{ij}$$

Pour tous  $i, j$ , on peut définir la probabilité conditionnelle :

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i \text{ et } Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_i}$$

On définit l'*espérance* (théorique) de  $Y$  *conditionnelle* à  $X = x_i$  comme le nombre :

$$E(Y | X = x_i) = \sum_j P(Y = y_j | X = x_i) y_j = \sum_j \frac{p_{ij}}{p_i} y_j, \text{ noté } E(Y | x_i) \text{ en abrégé,}$$

(selon les cas, il s'agira d'une somme finie ou d'une somme infinie).

On définit la *variance* (théorique) de  $Y$  conditionnelle à  $X = x_i$  comme le nombre :

$$\text{Var}(Y|X = x_i) = \sum_j P(Y = y_j|X = x_i)(y_j - E(Y|x_i))^2 = \sum_j \frac{p_{ij}}{p_i}(y_j - E(Y|x_i))^2$$

(selon les cas, il s'agira d'une somme finie ou d'une somme infinie).

À un niveau plus approfondi, on peut remarquer que l'espérance conditionnelle associe un nombre à toute valeur  $x_i$  de la variable aléatoire  $X$ . En associant à ce nombre la probabilité  $p_i$ , on définit ainsi une nouvelle variable aléatoire, appelée « espérance mathématique de  $Y$  en  $X$  » et notée  $E(Y|X)$ . Cette nouvelle variable aléatoire possède elle-même une espérance mathématique (un nombre), et l'on démontre alors la propriété  $E(E(Y|X)) = E(Y)$ .

On peut chercher à étendre les notions d'espérance conditionnelle et de variance conditionnelle au cas d'un couple de variables aléatoires absolument continues. Cela nécessite la définition préalable de la « densité de  $Y$  conditionnelle à  $X = x$  » et l'on obtient les généralisations cherchées. On peut en appréhender intuitivement la signification par un passage à la limite, analogue à celui que l'on effectue pour appréhender la notion de densité de probabilité. Ces notions sont utilisées notamment pour définir la courbe de régression et le rapport de corrélation.

Ces notions se transposent sans difficulté au cas d'un échantillon statistique à condition que l'échantillon soit constitué de  $q$  groupes d'observations relatifs à  $q$  valeurs  $x_1, x_2, \dots, x_q$ , avec pour le groupe n°  $i$ ,  $n_i$  valeurs observées  $y_{i1}, y_{i2}, \dots, y_{in_i}$  de la variable  $Y$  (les formules qui suivent sont valables dès que l'un des  $n_i$  est  $\geq 2$  mais ont peu d'intérêt si ce n'est pas le cas de la plupart !). On utilise les notations définies ci-dessus, et on pose  $n = n_1 + n_2 + \dots + n_q$ . On définit la *moyenne* (empirique) de  $Y$  conditionnelle à  $X = x_i$  comme le nombre :

$$m(Y|X = x_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, \text{ en abrégé } m(Y|x_i), \text{ ou } \bar{y}_i \text{ s'il n'y a pas d'ambiguïté.}$$

On définit la *variance* (empirique) de  $Y$  conditionnelle à  $X = x_i$  comme le nombre :

$$\text{Var}(Y|X = x_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \text{ en abrégé } \text{Var}(Y|x_i).$$

## conditionnelle (probabilité)

(*conditional probability*)

Probabilité définie en restreignant l'ensemble fondamental  $\Omega$  à la partie constituant un événement  $A$ , de façon à modéliser convenablement les probabilités conditionnées par la survenue de  $A$ .

Soient deux événements  $A, B$  d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , avec  $P(A) \neq 0$ . On appelle probabilité conditionnelle de l'événement «  $B$  si  $A$  » (ou «  $B$  sachant  $A$  ») le quotient :

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

On notera que  $P(A|A) = 1$ .

**Remarque** : on peut se demander pourquoi on ne prend pas tout simplement  $P(A \cap B)$  comme définition de la probabilité conditionnelle de  $B$  sachant  $A$ . La réponse est très simple : il faut normer, *i.e.* imposer que la probabilité conditionnelle de  $A$  sachant  $A$  soit égale à 1, ce qui exige le dénominateur  $P(A)$ .

**Exemple** On considère le lancer de 2 dés, représenté par un espace à 36 évènements équiprobables. On considère l'évènement  $A = \text{« somme } = 10 \text{ »}$  et l'évènement  $B = \text{« double »}$ . Calculer la probabilité conditionnelle  $P(B|A)$ .

En vertu de l'équiprobabilité, il suffit de compter les évènements élémentaires :  $A \cap B = \{(5, 5)\}$  donc  $P(A \cap B) = \frac{1}{36}$ ,  $A = \{(4, 6), (5, 5), (6, 4)\}$  donc  $P(A) = \frac{3}{36}$ , de sorte que

$P(B|A) = \frac{1/36}{3/36} = \frac{1}{3}$ , que l'on peut comparer avec la probabilité « *a priori* »  $P(B) = \frac{6}{36} = \frac{1}{6}$ .

Voir *composées (formule des probabilités)*.

## confiance (intervalle de)

(*confidence interval*)

Voir *estimation par intervalle*.

## conjoint, e

Voir *joint, e*.

## conjonction logique

Voir *intersection*.

## contingence (tableau de)

Voir *tableau de contingence*.

## continu (ensemble)

(*continuous set*)

Se dit d'un ensemble de nombres formé par un intervalle (la propriété essentielle – mais non caractéristique – est qu'entre deux nombres distincts quelconques de l'ensemble, il y en a toujours un autre, ce qui entraîne qu'il y en a toujours une infinité d'autres). S'oppose à *discret*, encore qu'il y ait des ensembles de type intermédiaire ou hybride.

## continue (variable aléatoire)

(*continuous random variable*)

Voir *variable aléatoire (typologie)*.

## contraire

Voir *complémentaire*.

## convergence

(*convergence*)

La convergence en calcul des probabilités est une notion multiforme et difficile à définir. Et en même temps, les théorèmes de convergence représentent l'outil mathématique indispensable pour fonder le calcul des probabilités et la statistique mathématique, ainsi que leur connexion.

Voir *convergence en probabilité, convergence presque sûre, convergence en moyenne quadratique*.

## convergence en loi

(*convergence in distribution*)

La *convergence en probabilité* renseigne sur le comportement d'une suite de v.a. en donnant une information sur les probabilités relatives aux valeurs réellement prises « en situation »

d'épreuve aléatoire, mais on peut se contenter d'un renseignement théorique sur la loi des  $X_n$ .

On se donne une suite  $(X_n)$  de variables aléatoires (réelles) définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ , et une variable aléatoire  $X$  définie également sur cet espace. On note  $(F_n)$  les fonctions de répartition des  $X_n$ , et  $F$  celle de  $X$ . On dit la suite  $(X_n)$  converge en loi

vers  $X$  lorsque  $n$  tend vers l'infini, et on écrit  $X_n \xrightarrow{\mathcal{L}} X$ , si, en tout point  $x$  pour lequel  $F$  est continue, on a :

$$F_n(x) \xrightarrow[n \rightarrow \infty]{} F(x)$$

**Théorème.** La convergence en probabilité entraîne la convergence en loi.

À un niveau plus approfondi, on peut montrer que la convergence en loi est équivalente, moyennant certaines conditions supplémentaires qui ne seront pas données ici, à la convergence des fonctions caractéristiques.

Voir *central limite (théorème)*.

## convergence en moyenne quadratique

(convergence in square mean)

Outre les trois convergences « classiques » (en probabilité, en loi, presque sûre), il existe d'autres notions, notamment la convergence en moyenne quadratique, définie par  $E((X_n - X)^2) \rightarrow 0$ , qui joue un rôle technique important en calcul des probabilités (elle entraîne la convergence en probabilité et *a fortiori* la convergence en loi, mais n'a pas de rapport d'implication avec la convergence presque sûre).

## convergence en probabilité

(convergence in probability)

Une théorie mathématique des phénomènes aléatoires n'aurait jamais vu le jour si l'on n'avait pas remarqué la convergence de la *fréquence* d'un évènement lorsque l'on répète indéfiniment l'épreuve : la limite de la fréquence est la *probabilité* de l'évènement. La convergence qui fonctionne ici s'appelle la convergence « en probabilité ».

On se donne une suite  $(X_n)$  de variables aléatoires (réelles) définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On dit que la suite  $(X_n)$  converge en probabilité vers la constante  $x_0$

lorsque  $n$  tend vers l'infini, et on écrit  $X_n \xrightarrow{P} x_0$ , si, pour tout  $\varepsilon > 0$ , on a :

$$P(|X_n - x_0| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0$$

Dans cette définition, le rôle de  $x_0$  n'est pas celui d'un nombre mais celui d'une variable aléatoire certaine. Cette remarque permet de généraliser de façon naturelle la convergence en probabilité.

On se donne une suite  $(X_n)$  de variables aléatoires (réelles) définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ , et une variable aléatoire  $X$  définie également sur cet espace. On dit que la suite  $(X_n)$  converge en probabilité vers la variable aléatoire  $X$  lorsque  $n$  tend vers l'infini,

et on écrit  $X_n \xrightarrow{P} X$ , si la suite  $(X_n - X)$  converge en probabilité vers 0.

Voir *grands nombres (loi des)*.

**convergence presque sûre** (almost certain convergence)

La *convergence en probabilité* oblige des probabilités à tendre vers 0, mais elle n’oblige pas vraiment la suite de variables aléatoires à tendre numériquement vers une valeur limite sous les yeux de l’expérimentateur ou de l’observateur. On peut donc imaginer une convergence plus forte que la convergence en probabilité. Ce type de convergence très technique est essentiellement utilisé par les mathématiciens professionnels.

On se donne une suite  $(X_n)$  de variables aléatoires (réelles) définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On dit que la suite  $(X_n)$  converge presque sûrement vers la constante  $x_0$

lorsque  $n$  tend vers l’infini, et on écrit  $X_n \xrightarrow{p.s.} x_0$ , si la probabilité de l’évènement  $X_n \rightarrow x_0$   
 $n \rightarrow \infty$

est égale à 1.

Pour apprécier la portée de cette affirmation, sous la forme équivalente « la probabilité que  $X_n$  ne tende pas vers  $x_0$  est nulle », il faut faire la différence entre un évènement « logiquement impossible » (en quelque sorte un évènement virtuel qui n’appartient même pas à l’espace probabilisé) et un évènement de cet espace probabilisé mais de probabilité nulle. Ainsi, pour une suite infinie de parties de P ou F, l’évènement PPPPP... = « Pile chaque fois » est imaginable (logiquement possible) mais de probabilité nulle (il ne se produira pas, mais c’est très abstrait car personne n’ira expérimenter jusqu’à l’infini !)

Dans la pratique (théorique), la probabilité de l’évènement global  $X_n \rightarrow x_0$  n’est pas calculable sans introduire par exemple les évènements  $E_n(\epsilon) = \{|X_n - x_0| > \epsilon\}$ , pour écrire

$$\text{comme définition équivalente } \forall \epsilon > 0 \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} E_k(\epsilon)\right) = 0.$$

Comme la convergence en probabilité, on peut généraliser de façon naturelle la convergence presque sûre (convergence vers une v.a. quelconque).

**Théorème.** La convergence presque sûre entraîne la convergence en probabilité (et *a fortiori* la convergence en loi).

**correction de Sheppard** (Sheppard correction)

Correction qui débiaise le calcul d’une variance empirique effectué à partir des centres de classes qui sont des intervalles de longueur fixe.

Voir *variance*.

**correction de Yates**

Voir *Yates (correction [de continuité] de)*.

**corrélation** (correlation)

Lorsqu’il ne qualifie pas un indicateur numérique (le « coefficient de corrélation ») et qu’il est employé de façon « libre », ce terme renvoie à une situation où deux variables aléatoires X et Y – ou deux caractères statistiques – ne sont, ni entièrement liés ( $Y = f(X)$ ) ni indépendants ; la nature de la « liaison » entre les deux variables ou les deux caractères est alors aléatoire (au sens mathématique du mot, qui ne veut pas dire que c’est « n’importe quoi » !), et sa force peut souvent être mesurée par un coefficient numérique adapté.

## corrélation (coefficient de) (correlation coefficient)

Nombre réel sans dimension, compris entre  $-1$  et  $1$ , défini à partir de la covariance, qui mesure la « liaison » (ou la « dépendance ») entre deux variables aléatoires ou deux caractères statistiques.

Le coefficient de corrélation (sans autre précision) devrait être qualifié de coefficient de corrélation *linéaire* – quoique ce qualificatif ne soit que rarement ajouté : il est en effet spécialement adapté à la mesure d'une « liaison » linéaire (ou plutôt affine), et on peut construire d'autres coefficients de corrélation, adaptés à d'autres types de « liaison ».

Voir *corrélation de deux variables aléatoires (coefficient de)*, *corrélation de deux échantillons statistiques (coefficient de)*, *corrélation multiple (coefficient de)*, *corrélation partielle (coefficient de)*, *corrélation des rangs (coefficient de)*.

## corrélation de deux échantillons statistiques (coefficient de)

On considère un couple  $(X, Y)$  de caractères statistiques réels, et un échantillon observé de  $n$  valeurs numériques de ce couple  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$ . On note les moyennes observées  $\bar{x}$  et  $\bar{y}$  ; on note les variances et la covariance observées :

$$s_X^2 = \text{Var}(X), \quad \text{Cov}_{X,Y} = \text{Cov}(X, Y), \quad s_Y^2 = \text{Var}(Y).$$

Le coefficient de corrélation (linéaire) de  $X$  avec  $Y$  est défini comme le quotient :

$$r_{X,Y} = \frac{\text{Cov}_{X,Y}}{s_X s_Y} = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{i=n} (x_i - \bar{x})^2 \sum_{i=1}^{i=n} (y_i - \bar{y})^2}}$$

**Remarque** : si on calcule le coefficient de corrélation à partir d'estimations des variances et de la covariance, il importe peu que l'on utilise les estimations *biaisées* ou *débiaisées* (qui se traduisent par des dénominateurs  $n$  ou  $n-1$  dans les formules qui donnent  $s_X^2$ ,  $\text{Cov}_{X,Y}$  et  $s_Y^2$ ). Le quotient qui définit le coefficient de corrélation fait disparaître ces dénominateurs et annule donc l'effet des biais ou des débiais, donnant le même résultat numérique dans tous les cas.

Le coefficient de corrélation statistique (parfois qualifié de coefficient de corrélation *observé* ou *empirique*) se note le plus souvent  $r_{X,Y}$  ou  $r_{XY}$  ou  $r_{x,y}$  ou  $r_{xy}$  (ou  $r$  s'il n'y a aucun risque de confusion), parfois  $\text{Corr}(X, Y)$  ou  $\text{Corr}(x, y)$ . Pour éviter de mal interpréter des valeurs peu significatives, il est important de savoir que le *poids* de la « liaison » affine révélée par le coefficient de corrélation est correctement mesuré par le carré  $r^2$ .

### Propriétés

- C'est un nombre sans dimension, et on a toujours  $-1 \leq r_{X,Y} \leq 1$ .
- Si  $r_{X,Y} = 1$  ou  $r_{X,Y} = -1$ , il existe une relation affine entre  $X$  et  $Y$  :  $Y = a + bX$  (pour tout  $n$  :  $y_n = a + bx_n$ ,  $b$  étant du signe de  $r_{X,Y}$ ).

Voir *covariance*.

## corrélation de deux échantillons statistiques (rapport de)

On considère un couple  $(X, Y)$  de variables statistiques et on suppose que l'échantillon est composé de  $q$  groupes d'observations relatifs aux valeurs  $x_1, x_2, \dots, x_q$  de  $X$ , avec pour le groupe n°  $i$ ,  $n_i$  valeurs observées  $y_{i1}, y_{i2}, \dots, y_{in_i}$  de la variable  $Y$ , et on pose  $n = n_1 + n_2 + \dots + n_q$  (les formules qui suivent sont valables dès que l'un des  $n_i$  est  $\geq 2$  mais ont peu

d'intérêt si ce n'est pas le cas de la plupart !). On commence par calculer les moyennes conditionnelles :

$$\bar{y}_i = m(Y|X = x_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \quad (i = 1, 2, \dots, q)$$

(nota : la moyenne totale de Y est  $\bar{y} = \sum_{i=1}^q \frac{n_i}{n} \bar{y}_i$ ).

On calcule ensuite les variances conditionnelles

$$\text{Var}_{Y|x_i} = \text{Var}(Y|X = x_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (i = 1, 2, \dots, q)$$

(nota : la variance totale de Y est  $\text{Var}_Y = \frac{1}{n} \sum_{i=1}^q \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$ ).

**Formule de décomposition de la variance (empirique)**

La variance  $\text{Var}_Y$  est la somme de deux termes :

- la *variance expliquée* par la régression en X (ou encore variance inter-groupes), variance des moyennes conditionnelles :

$$\text{Var}(\bar{y}_i) = \sum_{i=1}^q \frac{n_i}{n} (\bar{y}_i - \bar{y})^2$$

- la *variance résiduelle* (ou encore variance intra-groupes), moyenne des variances conditionnelles :

$$m(\text{Var}_{Y|x_i}) = \sum_{i=1}^q \frac{n_i}{n} \text{Var}_{Y|x_i}$$

Avec les notations précédentes, on peut maintenant définir le rapport de corrélation (empirique) de Y en X comme le quotient de la variance expliquée par la régression, par la variance totale :

$$e_{Y|X}^2 = \frac{\text{Var}(\bar{y}_i)}{\text{Var}_Y}$$

Le rapport de corrélation est noté  $e^2$  ou  $\eta^2$ . Lorsqu'il n'y a pas de liaison statistique entre X et Y, il est nul. Lorsque la corrélation est linéaire (affine), il est égal au carré  $r^2$  du coefficient de corrélation linéaire. Lorsqu'il est exactement égal à 1, la variance résiduelle  $m(\text{Var}_{Y|x_i})$  est nulle, ce qui implique que  $\text{Var}_{Y|x_i}$  est nulle pour tout  $i$ , et donc que la liaison entre X et Y est « fonctionnelle » :  $i$  et  $x_i$  étant fixés,  $y_{ij}$  est constante (par rapport à  $j$ ).

**corrélation de deux variables aléatoires (coefficient de)**

On considère un couple (X, Y) de variables aléatoires réelles, et on note les variances et la covariance :

$$\sigma_X^2 = \text{Var}(X), \text{Cov}_{X,Y} = \text{Cov}(X, Y), \sigma_Y^2 = \text{Var}(Y).$$

Le coefficient de corrélation (linéaire) de  $X$  avec  $Y$  est défini comme le quotient :

$$\rho_{X,Y} = \frac{\text{Cov}_{X,Y}}{\sigma_X \sigma_Y}$$

Le coefficient de corrélation se note le plus souvent  $\rho_{X,Y}$  ou  $\rho_{XY}$  (ou  $\rho$  s'il n'y a aucun risque de confusion), parfois  $\text{Corr}(X, Y)$ . Pour éviter de mal interpréter des valeurs peu significatives, il est important de savoir que le *poids* de la « liaison » affine révélée par le coefficient de corrélation est correctement mesuré par le carré  $\rho^2$ .

*En toute rigueur*, il aurait fallu inclure dans la définition que l'on supposait que les espérances  $\mu_X$  et  $\mu_Y$  existaient, de même que les variances et la covariance.

#### Propriétés

- C'est un nombre sans dimension, et on a toujours  $-1 \leq \rho_{X,Y} \leq 1$ .
- Si  $\rho_{X,Y} = 1$  ou  $\rho_{X,Y} = -1$ , il existe une relation affine entre  $X$  et  $Y$  :  $Y = \alpha + \beta X$  (cette relation est vérifiée « presque sûrement »).
- Si  $X$  et  $Y$  sont indépendantes, on a  $\rho_{X,Y} = 0$  (mais la réciproque est fausse).

### corrélation de deux variables aléatoires (rapport de)

Dans le cas d'un couple  $(X, Y)$  de variables aléatoires, on peut utiliser les notions d'espérance conditionnelle  $E(Y|X = x)$  et de variance conditionnelle  $\text{Var}(Y|X = x)$ , (notions simples dans le cas discret, plus délicates dans le cas général) pour démontrer un théorème de *décomposition de la variance* marginale  $\text{Var}(Y)$  en deux termes :

$$\text{Var}(Y) = \text{Var}(E(Y|X)) + E(\text{Var}(Y|X)).$$

La courbe  $y = E(Y|X = x)$  est la courbe de régression, et cette formule signifie que la variance marginale totale de  $Y$  est la somme d'un terme  $\text{Var}(E(Y|X))$  qui est la variance *expliquée* par la régression et d'un terme  $E(\text{Var}(Y|X))$  qui est la variance *résiduelle* de  $Y$ .

Le rapport de corrélation de  $Y$  en  $X$  est défini comme le quotient du premier terme par la variance totale :

$$\eta_{Y|X}^2 = \frac{\text{Var}(E(Y|X))}{\text{Var}(Y)}.$$

Lorsque  $X$  et  $Y$  sont indépendantes, il est nul (mais la réciproque est fausse). Lorsque la courbe de corrélation est une droite, il est égal au carré  $\rho^2$  du coefficient de corrélation linéaire. Lorsqu'il est exactement égal à 1, la variance résiduelle  $E(\text{Var}(Y|X))$  est nulle, ce qui implique que  $\text{Var}(Y|X = x)$  est (presque sûrement) nulle pour tout  $x$ , et donc que la liaison entre  $X$  et  $Y$  est fonctionnelle : il existe une fonction  $\phi$  telle que  $Y = \phi(X)$  (presque sûrement).

### corrélation multiple (coefficient de) (*multiple correlation coefficient*)

Nombre réel dans dimension, compris entre 0 et 1, qui mesure la « liaison » (ou la « dépendance ») globale entre une variable aléatoire (ou caractère statistique) et deux (ou plusieurs) autres.

On considère un triplet  $(X, Y, Z)$  de variables aléatoires réelles, et on définit et on note les coefficients de corrélation (linéaire) simples comme usuellement.

Le coefficient de corrélation multiple de  $Z$  avec  $X$  et  $Y$ , noté généralement  $\rho_{Z,XY}$ , est défini comme le maximum du coefficient de corrélation simple entre  $Z$  et une combinaison linéaire  $\beta X + \gamma Y$  lorsque  $\beta$  et  $\gamma$  varient. On peut calculer son carré par la formule :

A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

$$\rho_{Z.XY}^2 = \frac{\rho_{XZ}^2 + \rho_{YZ}^2 - 2\rho_{XY}\rho_{XZ}\rho_{YZ}}{1 - \rho_{XY}^2}$$

On a toujours  $0 \leq \rho_{Z.XY} \leq 1$  ; lorsque  $\rho_{Z.XY} = 1$ , il existe une relation affine entre X, Y et Z de la forme  $Z = \alpha + \beta X + \gamma Y$  (cette relation est vérifiée « presque sûrement »).

On rappelle que les poids des « liaisons » affines sont correctement mesurés par les carrés des coefficients de corrélation.

On considère maintenant un triplet (X, Y, Z) de caractères statistiques réels et un échantillon observé de n valeurs numériques de ce triplet  $((x_1, y_1, z_1), \dots, (x_n, y_n, z_n))$ . On définit et on note les coefficients de corrélation (linéaire) simples observés comme usuellement.

Le coefficient de corrélation multiple de Z avec X et Y, noté généralement  $r_{Z.XY}$ , est défini par son carré :

$$r_{Z.XY}^2 = \frac{r_{XZ}^2 + r_{YZ}^2 - 2r_{XY}r_{XZ}r_{YZ}}{1 - r_{XY}^2}$$

On a toujours  $0 \leq r_{Z.XY} \leq 1$  ; lorsque  $r_{Z.XY} = 1$ , il existe une relation affine entre X, Y et Z (pour tout n :  $z_n = a + bn_x + cy_n$ ).

Si on généralise la régression linéaire au cas « multidimensionnel », la droite de régression  $a + bX$  (ajustement  $Y = a + bX + E$ ) est remplacée, dans le cas le plus simple de 2 variables « explicatives » X et Y, par un plan de régression  $a + bX + cY$  (ajustement  $Z = a + bX + cY + E$ ). Comme dans le cas simple, la corrélation de la variable « expliquée » avec l'écart E est nulle, et la variance se décompose : si on note  $R^2 = r_{Z.XY}^2$ , la variance (marginale)  $s_Z^2$  de Z est la somme de deux termes : la variance expliquée par la multirégression en X et Y :  $s_{a+bX+cY}^2 = R^2 s_Z^2$ , et la variance résiduelle :  $s_E^2 = (1 - R^2) s_Z^2$ .

Le coefficient de corrélation multiple se généralise au cas d'un nombre quelconque de variables (les formules ne sont pas données ici).

### corrélation partielle (coefficient de) (partial correlation coefficient)

Nombre réel dans dimension, compris entre -1 et 1, qui mesure la « liaison » (ou la « dépendance ») entre deux variables aléatoires (ou caractères statistiques) après « élimination » de l'effet d'une (ou de plusieurs) autres.

Si par exemple X et Y présentent une forte corrélation, il se peut que cette corrélation résulte d'une corrélation commune à X et Y avec une troisième variable Z, et que la corrélation partielle (résiduelle) entre X et Y soit faible. C'est bien entendu une question d'interprétation, les formules mathématiques étant neutres. Mais l'importance pratique peut être très grande dans les domaines (la pathologie par exemple) où la corrélation peut être l'indice d'une causalité : il ne faut pas prendre pour une causalité de X vers Y (ou vice-versa) une causalité provenant de Z pour les deux !

On considère un triplet (X, Y, Z) de caractères statistiques réels et un échantillon observé de n valeurs numériques de ce triplet  $((x_1, y_1, z_1), \dots, (x_n, y_n, z_n))$ . On définit et on note les coefficients de corrélation (linéaire) simples observés comme usuellement.

Le coefficient de corrélation partiel de X et Y avec Z, noté généralement  $r_{XY.Z}$ , est défini par l'expression :

$$r_{XY.Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

On peut vouloir tester la significativité d'un coefficient de corrélation partielle : elle se teste exactement comme la significativité d'un coefficient de corrélation mais en diminuant le

nombre de degrés de liberté : si  $r = r_{XYZ}$ , la variable de test est  $t = \frac{|r|}{\sqrt{1-r^2}} \sqrt{n-3}$ , et il faut se rapporter à la table de la loi de Student à  $n-3$  degrés de liberté (dans le cas simple dont la formule a été donnée ici et où l'on a éliminé l'effet d'une variable).

### corrélation des rangs (coefficient de – de Kendall)

(Kendall tau, Kendall rank  
correlation coefficient)

Nombre réel sans dimension, compris entre  $-1$  et  $1$ , qui mesure la « liaison » (ou la « dépendance ») entre deux variables aléatoires ou deux caractères statistiques *ordinaux*.

Ce coefficient (comme celui de Spearman) est également utilisé pour apprécier, après leur transformation en rangs, la corrélation de deux variables numériques qui sont manifestement non normales, et pour lesquelles donc le coefficient de corrélation linéaire est non approprié. Mais il n'a de pertinence que si la liaison présumée entre les variables est monotone.

Si on dispose de  $n$  « objets » ou individus et de deux variables classées, on peut attribuer dans chacun des classements un rang à chaque objet ou individu :  $r_1, r_2, \dots, r_n$  et  $s_1, s_2, \dots, s_n$ . Le coefficient de corrélation des rangs de Kendall est calculé à partir des concordances et discordances de classement pour chacun des  $\frac{n(n-1)}{2}$  couples distincts  $(i, j)$  d'individus : si  $(r_i, r_j)$  et  $(s_i, s_j)$  sont dans le même ordre, on compte  $1$ , s'ils sont en ordre différent, on compte  $-1$ , puis on effectue la somme  $S$  de ces valeurs.

Le coefficient de corrélation des rangs de Kendall est défini à partir de la somme  $S$  des concordances / discordances de classement des couples d'individus comme le quotient :

$$\tau = \frac{2S}{n(n-1)}.$$

On trouvera dans les manuels spécialisés le détail de la procédure à suivre pour adapter cette formule au cas d'ex æquo. On a  $\tau = 1$  si les deux classements sont identiques,  $\tau = -1$  s'ils sont inverses, et  $\tau = 0$  s'ils sont (empiriquement) indépendants.

Lorsqu'on utilise ce coefficient pour un couple de variables provenant d'une loi normale de coefficient de corrélation linéaire  $\rho$ , et si  $n$  est « grand », on a :

$$\tau \approx \frac{2}{\pi} \text{Arc sin } \rho, \text{ équivalent à } \rho \approx \sin\left(\frac{\pi}{2}\tau\right).$$

Pour autant, on n'utilise pas cette relation pour tester la valeur du coefficient de corrélation des rangs de Kendall. En effet, il est préférable d'utiliser le quotient :

$$z = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}}$$

qui suit approximativement une loi normale centrée réduite, avec une précision excellente dès que  $n \geq 8$ .

### corrélation des rangs (coefficient de – de Spearman)

(Spearman rho, Spearman rank  
correlation coefficient)

Nombre réel dans dimension, compris entre  $-1$  et  $1$ , qui mesure la « liaison » (ou la « dépendance ») entre deux variables aléatoires ou deux caractères statistiques *ordinaux*.

Ce coefficient (comme celui de Kendall) est également utilisé pour apprécier, après leur transformation en rangs, la corrélation de deux variables numériques qui sont manifeste-

A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

ment non normales, et pour lesquelles donc le coefficient de corrélation linéaire est non approprié. Mais il n'a de pertinence que si la liaison présumée entre les variables est monotone.

Si on dispose de  $n$  « objets » ou individus et de deux variables classées, on peut attribuer dans chacun des classements un rang à chaque objet ou individu :  $r_1, r_2, \dots, r_n$  et  $s_1, s_2, \dots, s_n$ . Le coefficient de corrélation des rangs de Spearman est simplement le coefficient de corrélation linéaire usuel entre les deux séries de rangs. Compte tenu des valeurs particulières prises par les rangs (les entiers de 1 à  $n$ ), son calcul se simplifie considérablement et on peut l'exprimer en ne faisant intervenir que les différences  $d_i = r_i - s_i$ .

Le coefficient de corrélation des rangs de Spearman, est défini à partir des différences  $d_i$  des classements comme l'expression

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}.$$

On trouvera dans les manuels spécialisés le détail de la procédure à suivre pour adapter cette formule au cas d'ex æquo. On a  $r = 1$  si les deux classements sont identiques,  $r = -1$  s'ils sont inverses, et  $r = 0$  s'ils sont (empiriquement) indépendants.

Pour tester la valeur du coefficient de corrélation des rangs de Spearman, il faut se reporter à des tables spécifiques.

Lorsqu'on utilise ce coefficient pour un couple de variables provenant d'une loi normale de coefficient de corrélation linéaire  $\rho$ , et si  $n$  est « grand », on a :

$$r_s \approx \frac{6}{\pi} \text{Arc sin} \left( \frac{\rho}{2} \right), \text{ équivalent à } \rho \approx 2 \sin \left( \frac{\pi}{6} r_s \right).$$

**corrélation des rangs (tests de)** *(rank correlation tests)*

Tests non paramétriques qui contrôlent l'indépendance de deux variables en testant la nullité d'un coefficient de corrélation *ad hoc* constitué à partir des rangs de classement des deux variables et non à partir de leurs valeurs numériques précises. Ces coefficients et les tests correspondants n'ont de pertinence que si la liaison présumée entre les variables est monotone. Il existe un test pour chacun des deux principaux coefficients de corrélation des rangs, celui de Kendall et celui de Spearman. Leur fonctionnement est très similaire à celui du test de significativité d'un coefficient de corrélation.

\_\_\_\_\_ test de corrélation des rangs de Kendall \_\_\_\_\_

- Données. Un échantillon de  $n$  couples de rangs  $((r_1, s_1), (r_2, s_2), \dots, (r_n, s_n))$  (obtenu, soit par observation directe d'un couple (R, S) de rangs, soit par transformation en rangs des valeurs d'un couple (X, Y) de valeurs réelles).
- Hypothèse testée.  $H_0 = \langle \tau = 0 \rangle$  contre  $H_1 = \langle \tau \neq 0 \rangle$  ( $H_0$  est très proche de l'indépendance si la liaison présumée est monotone).
- Déroulement technique du test
  1. On calcule le coefficient  $\tau$  de corrélation des rangs de Kendall (voir corrélation des rangs (coefficient de – de Kendall))

2. On calcule la variable de test

$$z = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}}$$

la valeur critique est à lire dans la table de la loi normale centrée réduite.

• Conditions et précautions

- $n \geq 8$  ;
- aucune autre, la loi théorique du coefficient de corrélation des rangs de Kendall ne dépendant pas de la loi des variables observées.

---

test de corrélation des rangs de Spearman

---

• Données. Un échantillon de  $n$  couples de rangs  $((r_1, s_1), (r_2, s_2), \dots, (r_n, s_n))$  (obtenu, soit par observation directe d'un couple (R, S) de rangs, soit par transformation en rangs des valeurs d'un couple (X, Y) de valeurs réelles).

• Hypothèse testée.  $H_0 = \ll r_S = 0 \gg$  contre  $H_1 = \ll r_S \neq 0 \gg$  ( $H_0$  est très proche de l'indépendance si la liaison présumée est monotone).

• Déroulement technique du test

1. On calcule le coefficient  $r_S$  de corrélation des rangs de *Spearman*.
- 2a. Si  $n$  est petit ( $n \leq 30$ ), on compare la valeur obtenue à la valeur critique lue dans une table spécifique du *coefficient de Spearman*.
- 2b. Si  $n$  est grand ( $n > 30$ ), on calcule la variable de test  $t = r_S \sqrt{n-1}$  et on compare la valeur obtenue à la valeur critique lue dans la table de la loi normale centrée réduite.

• Conditions et précautions

Aucune, la loi théorique du coefficient de corrélation des rangs de Spearman ne dépendant pas de la loi des variables observées.

---

## corrélation (rapport de)

(*correlation ratio*)

Nombre réel positif sans dimension, compris entre 0 et 1, défini à partir des paramètres d'une régression, qui mesure la « liaison » (ou la « dépendance ») entre deux variables aléatoires ou deux caractères statistiques, y compris dans les cas où cette liaison n'est pas linéaire (affine). Contrairement au coefficient de corrélation linéaire, le rapport de corrélation n'est pas symétrique.

## corrélation (test de significativité d'un coefficient de)

Voir *significativité d'un coefficient de corrélation (test de)*.

## couple de variables aléatoires

(*pair of random variables*)

On peut considérer simultanément deux variables aléatoires X, Y définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ . À chaque événement élémentaire  $\omega \in \Omega$ , correspondent alors deux valeurs  $X(\omega)$  et  $Y(\omega)$ , que l'on peut représenter par un couple  $(X(\omega), Y(\omega))$ .

Il peut aussi arriver que les deux variables X et Y soient préalablement définies, séparément, sur deux espaces probabilisés  $(\Omega_1, \mathcal{A}_1, P_1)$  et  $(\Omega_2, \mathcal{A}_2, P_2)$ . Cette circonstance se produira le plus souvent lorsqu'on supposera que X et Y sont indépendantes. Pour considérer simultanément

ment  $X$  et  $Y$ , il faut commencer théoriquement par faire le *produit* des deux espaces probabilisés. Mais il n'est pas nécessaire dans la pratique, sauf exception, de construire explicitement ce produit pour pouvoir considérer le couple  $(X, Y)$ .

Le problème de base est de définir la loi de probabilité du couple  $(X, Y)$ . On peut imaginer des situations mixtes, mais les deux cas les plus fréquents sont ceux où les v.a. sont toutes deux discrètes, ou bien toutes deux absolument continues.

### ► Cas $X$ et $Y$ discrètes

Si  $X$  et  $Y$  sont discrètes, la *loi du couple* est définie par la donnée (liste ou caractérisation) des couples  $(x_i, y_j)$  de valeurs prises, et les probabilités ponctuelles :

$$p_{ij} = P(X = x_i \text{ et } Y = y_j).$$

Cette loi s'appelle la *loi jointe* (ou *conjointe*) du couple  $(X, Y)$ . On peut si nécessaire expliciter des *lois marginales* :

$$p_i = P(X = x_i) = \sum_j p_{ij}, \quad q_j = P(Y = y_j) = \sum_i p_{ij}$$

(ces notations sont très commodes mais pas très cohérentes, on peut préférer  $p_{i\cdot}$  et  $p_{\cdot j}$ , surtout si on envisage de généraliser à plus de deux v.a.). Si  $X$  et  $Y$  sont réelles, on peut généraliser la notion de fonction de répartition :

$$F(x, y) = P(X \leq x \text{ et } Y \leq y).$$

Cas particulier important : si  $X$  et  $Y$  sont indépendantes, on a :

$$p_{ij} = P(X = x_i \text{ et } Y = y_j) = P(X = x_i) P(Y = y_j) = p_{i\cdot} p_{\cdot j},$$

et

$$F(x, y) = P(X \leq x \text{ et } Y \leq y) = P(X \leq x) P(Y \leq y) = F_1(x) F_2(y).$$

### ► Cas $X$ et $Y$ absolument continues

Si  $X$  et  $Y$  sont absolument continues, la *loi du couple* peut être définie de plusieurs manières. La façon la plus générale est celle par les probabilités d'intervalle :

$$P(a_1 < X \leq b_1 \text{ et } a_2 < Y \leq b_2).$$

On peut aussi utiliser la notion généralisée de densité ; si  $\varphi(x, y)$  est la densité jointe, on a :

$$P(a_1 < X \leq b_1 \text{ et } a_2 < Y \leq b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \varphi(u, v) du dv.$$

Une autre manière consiste à utiliser la notion généralisée de fonction de répartition (même définition formelle que dans le cas des v.a. discrètes) :

$$\Phi(x, y) = P(X \leq x \text{ et } Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y \varphi(u, v) du dv.$$

Quelle que soit la manière dont elle est définie, la loi s'appelle la *loi jointe* (ou *conjointe*) du couple  $(X, Y)$ . On peut si nécessaire en déduire les *lois marginales*, avec leurs densités  $f(x)$  et  $g(y)$ , et leurs fonctions de répartition marginales  $F(x) = P(X \leq x)$  et  $G(y) = P(Y \leq y)$ .

Cas particulier important : si  $X$  et  $Y$  sont indépendantes, on a :

$$\varphi(x, y) = f(x) g(y)$$

et :

$$\Phi(x, y) = F(x) G(y).$$

## couplées (différences)

Voir *appariées (séries)*.

## courbe de concentration [de Lorentz]

Voir *concentration (courbe de - de Lorentz)*.

## courbe de régression

(regression curve)

Courbe  $y = f(x)$  que l'on détermine pour rendre compte de la liaison probabiliste ou statistique entre deux variables.

Étant donné un couple  $(X, Y)$  de variables aléatoires réelles on peut définir pour chaque valeur de  $x$  l'espérance conditionnelle  $E(Y|X = x)$  (notion mathématique délicate à définir dans le cas général, mais qui se comprend intuitivement très bien). Cette espérance conditionnelle est une fonction « ordinaire » de la variable  $x$  et on peut tracer sa courbe représentative appelée courbe (ou ligne) de régression de  $Y$  en  $x$ .

Dans un certain nombre de cas, théoriques et pratiques, la courbe de régression est une droite, la *droite de régression*, et on est alors dans une situation de régression linéaire.

Voir *corrélation (rapport de)*.

## covariance

(covariance)

Nombre réel qui mesure la « liaison » linéaire (ou la « dépendance ») entre deux variables aléatoires ou deux caractères statistiques.

Voir *covariance de deux variables aléatoires*, *covariance de deux échantillons statistiques*, *formule de Huygens-König*.

## covariance de deux échantillons statistiques

### Formule

Pour  $n$  couples d'observations individualisées  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , les moyennes  $\bar{x}$  et  $\bar{y}$  ayant été préalablement calculées :

$$\text{Cov}_{X,Y} = \frac{(x_1 - \bar{x})(y_1 - \bar{y})(x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n} = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - \bar{x})(y_i - \bar{y}).$$

La covariance de  $X$  avec  $Y$  (parfois qualifié de covariance *observée* ou *empirique*) se note le plus souvent  $\text{Cov}_{X,Y}$  ou  $\text{Cov}_{x,y}$  ou  $\text{Cov}_{XY}$  ou  $\text{Cov}_{xy}$  (ou  $\text{Cov}$  s'il n'y a aucun risque de confusion).

**Remarque** : lorsque la valeur de la covariance doit intervenir dans des formules impliquant des lois de probabilité (notamment pour un intervalle de confiance ou un test d'hypothèse), il convient de remplacer l'estimation biaisée donnée par la formule « descriptive » ci-dessus par l'estimation débiaisée obtenue en remplaçant le dénominateur  $n$  par  $n - 1$  (cf. aussi *estimation ponctuelle*).

## covariance de deux variables aléatoires

On considère un couple  $(X, Y)$  de variables aléatoires réelles, et on note leurs espérances mathématiques  $\mu_X$  et  $\mu_Y$ . La *covariance* entre  $X$  et  $Y$  est l'espérance mathématique du produit des variables centrées

$$\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)).$$

Si les deux v.a.  $X$  et  $Y$  sont *discrètes*, caractérisées par les ensembles (finis ou dénombrables) de valeurs  $\{x_i\}$  et  $\{y_j\}$ , avec les probabilités jointes  $p_{ij} = P(X = x_i \text{ et } Y = y_j)$ , on a :

$$\text{Cov}(X, Y) = \sum_{x_i} \sum_{y_j} p_{ij} (x_i - \mu_X)(y_j - \mu_Y)$$

(selon les cas, il s'agira d'une somme finie ou d'une somme infinie).

Si  $X$  et  $Y$  sont *absolument continues*, caractérisées par la densité de probabilité jointe  $\varphi(x, y)$ , on a :

$$\text{Cov}(X, Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - \mu_X)(y - \mu_Y)\varphi(x, y) dx dy .$$

**Remarque** : il n'y a pas de certitude que la somme converge ou que l'intégrale converge avant de l'avoir effectivement vérifié.

La covariance de  $X$  avec  $Y$  se note le plus souvent  $\text{Cov}(X, Y)$  ou  $\text{Cov}_{X, Y}$  ou  $\text{Cov}_{XY}$ . On notera que  $\text{Cov}(X, X) = \text{Var}(X)$ .

#### Propriétés

- On a  $\text{Var}(\alpha X + \beta Y) = \alpha^2 \text{Var}(X) + 2\alpha\beta \text{Cov}(X, Y) + \beta^2 \text{Var}(Y)$   
(cas particuliers intéressants :  $\alpha = 1, \beta = \pm 1$ ).
- Si  $X$  et  $Y$  sont indépendantes, on a  $\text{Cov}(X, Y) = 0$  (mais la réciproque est fausse).

### Cramer–von Mises (test [d'ajustement] de) (Cramer–von Mises test)

Test non paramétrique qui compare la distribution d'un échantillon statistique à une distribution fixée (par exemple : loi exponentielle de paramètre  $\lambda$  spécifié, ou loi normale d'espérance et de variance spécifiées). Les distributions (lois) sont représentées par leurs fonctions de répartition, utilisées pour l'exécution du test.

test bilatéral de comparaison d'une distribution de fonction de répartition  $F(x)$  à une distribution de fonction de répartition fixée  $F_0(x)$

- Données. Un échantillon  $(x_1, x_2, \dots, x_n)$  de  $n$  valeurs observées d'une variable aléatoire numérique  $X$  de fonction de répartition  $F(x)$ .
- Hypothèse testée.  $H_0 = \ll F = F_0 \gg$  contre  $H_1 = \ll F \neq F_0 \gg$ .
- Déroulement technique du test
  - 1a. On ordonne les valeurs observées de l'échantillon – on suppose ce rangement effectué, soit, en gardant les notations initiales :

$$x_1 \leq x_2 \leq \dots \leq x_n .$$

1b. Puis on pose :

$$F(x_1) = \frac{1}{n}, F(x_2) = \frac{2}{n}, \dots, F(x_n) = \frac{n}{n} = 1 ,$$

ce qui définit les « marches » de la fonction de répartition *observée*, qui est une fonction « en escalier ».

2. On calcule la valeur observée de la variable de test :

$$n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left( \frac{2i-1}{2n} - F_0(x_i) \right)^2 .$$

Les valeurs de référence de la variable de test sont à lire dans des tables obtenues par simulation, elles dépendent de la taille  $n$  de l'échantillon et du risque  $\alpha$ . Valeur limite : pour  $\alpha = 0,05$ , la valeur critique de  $n\omega_n^2$  (probabilité de dépassement 0,05) est équivalente à 0,4614.

- Conditions et précautions

- Il n'y a pas d'autre précaution que de fixer complètement la loi de référence (donc en particulier d'éviter toute estimation de paramètres) ;
- Lorsqu'il faut estimer des paramètres, il existe des tables obtenues par simulation. Valeur limite (*Biometrika Tables*) : pour  $\alpha = 0,05$ , la valeur critique de  $n\omega_n^2$  corrigé lorsque l'on fait fonctionner le test de Cramer-von-Mises après estimation de l'espérance et de l'écart-type d'une loi normale, est équivalent à 0,126.

---

Ce test possède les mêmes indications que le test de Kolmogorov. Malgré le peu de connaissances théoriques sur la loi (indépendante de la distribution F) de la variable  $n\omega_n^2 =$

$\int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 dF(x)$  ( $F_n$  désigne la loi empirique pour  $n$  observations de la loi F), ce test est considéré comme plus puissant que le test de Kolmogorov. En tout état de cause, il prend en compte la totalité des écarts et non pas seulement le plus grand.

### critique (région)

Voir *région critique*.



## ddl

Abréviation pour [nombre de] degrés de liberté.

## débiaisée (estimation)

(unbiased estimation)

Qualifie une estimation multipliée par un facteur correctif qui permet d'annuler son biais initial. Les deux cas les plus importants sont les estimations de la variance et de la covariance

qui, sous leur forme primitive :  $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  et  $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ , sont biaisées. On les

« débiaise » en remplaçant le dénominateur  $n$  par  $n - 1$ . La plupart des calechettes qui traitent les séries statistiques notent «  $\sigma_n$  » la racine carrée de la variance biaisée, et «  $\sigma_{n-1}$  » la racine carrée de la variance débiaisée.

La question importante est de savoir quand on utilise l'estimation biaisée, et quand on utilise l'estimation débiaisée. Sauf cas particulier la réponse est simple : si l'on fait de la statistique descriptive, on utilise l'estimation « naturelle » et biaisée ; si l'on fait de la statistique inférentielle (intervalle de confiance, test d'hypothèse), on utilise l'estimation débiaisée.

## décile

(decile)

Indicateur de position attaché à une variable aléatoire réelle, utilisé essentiellement en statistique. Les déciles partagent la série des valeurs en deux parties de fractions  $\alpha$  et  $1 - \alpha$  de l'effectif total, pour  $\alpha = 10\%$ ,  $20\%$ , ...,  $90\%$ . Ils sont au nombre de 9.

Si la signification concrète des déciles est simple et « parlante », la traduction formelle est plus délicate. On adaptera sans difficulté le formulaire détaillé pour la médiane.

## décomposition de la variance

On considère un couple  $(X, Y)$  de variables numériques, soit dans une situation de *modèle linéaire* ( $X$  est alors un ensemble  $\{x_i\}$  de valeurs « maîtrisées » et  $Y$  un ensemble  $\{Y_i\}$  de variables aléatoires normales associées), soit dans une situation de *régression linéaire* ( $(X, Y)$  est alors un couple de v.a. qui suit une loi normale à 2 dimensions). On peut définir dans l'un et l'autre cas la *droite de régression* théorique  $y = \alpha + \beta x$ , et la droite de régression empirique  $y = a + bx$ .

On peut montrer qu'il y a une corrélation nulle entre la variable aléatoire  $\alpha + \beta X$  et la variable aléatoire écart  $E = Y - (\alpha + \beta X)$ . Cette propriété permet de décomposer la variance observée de  $Y$  en deux termes qui traduisent sa double origine (on observera que le coefficient de corrélation intervient par son carré, lequel mesure correctement le poids de la corrélation).

### Décomposition de la variance (empirique)

La variance  $s_Y^2$  de  $Y$  est la somme de deux termes :

- la variance *expliquée* par la régression en  $X$  :  $s_{a+bX}^2 = r^2 s_Y^2$ ,
- la variance *résiduelle* :  $s_E^2 = (1 - r^2) s_Y^2$ .

On notera que la variance résiduelle théorique  $s_E^2 = (1 - \rho^2)\sigma_Y^2$  est très exactement la variance que l'on a appelée conditionnelle dans le modèle linéaire comme dans la régression : variance  $\sigma^2$  commune à toutes les v.a.  $Y_i$  dans le cas d'une variable  $X$  contrôlée à valeurs  $x_i$ , et variance conditionnelle *stricto sensu*  $\sigma^2 = \sigma^2(Y|X = x)$  constante (dans l'hypothèse normale) dans le cas d'un couple  $(X, Y)$  de variables aléatoires.

La variance résiduelle est souvent appelée *variance des résidus*, les résidus étant les écarts  $y_i - (a + bx_i)$ .

Voir *droite de régression, corrélation (rapport de)*.

## défaillance (taux de)

(*failure rate*)

On considère un « système »  $S$  (dispositif industriel, atome radioactif, être vivant, ...) pouvant être affecté par une défaillance (ou panne, ou mort, ...), et on introduit la variable aléatoire réelle  $T =$  « durée de vie de  $S$  » = « temps d'attente de la (première) défaillance ».

On définit le taux de défaillance (ou taux de panne, ou taux de mort, ...)  $\lambda(t)$  de  $S$  comme la « densité conditionnelle » :

$$\begin{aligned}\lambda(t) &= \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} P(\text{défaillance entre } t \text{ et } t + \delta t | \text{pas de défaillance sur } [0, t]) \\ &= \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} P(t < T \leq t + \delta t | T \leq t),\end{aligned}$$

relié à la fonction de répartition  $F(t) = P(T \leq t)$  par  $\lambda(t) = \frac{F'(t)}{1 - F(t)}$ .

Le cas de référence est celui d'un système « sans vieillissement » dont le taux de défaillance est constant : si  $\lambda(t) = \lambda$ ,  $T$  est alors une v.a. exponentielle de paramètre  $\lambda$  (et donc d'espérance « durée de vie » de  $S$  égale à  $\frac{1}{\lambda}$ ).

Pour modéliser les situations avec vieillissement ou usure, on emploie souvent une v.a. de Weibull (avec paramètre  $\beta > 1$ ) ou une loi d'Erlang (avec paramètre  $r > 1$ ).

Voir *fiabilité*.

## dégénérée (loi)

(*degenerate distribution*)

Voir *singulière (loi)*.

## degrés de liberté

(*degrees of freedom*)

Locution qui qualifie le ou l'un des paramètres de certaines lois de probabilité. Lors de l'exécution d'un test d'hypothèse qui se réfère à cette loi, la valeur du paramètre degrés de liberté est déterminée par les conditions du test et notamment la taille et/ou la structure de l'échantillon. Même si cela n'est pas toujours apparent, le nombre de degrés de libertés est égal au nombre de valeurs recueillies ou de classes examinées, diminué du nombre de contraintes numériques, explicites ou implicites.

## de Moivre-Laplace (théorème de)

(*Moivre-Laplace limit theorem*)

Dénomination historique du théorème central limite (voir *central limite (théorème)*) lorsqu'il est énoncé dans le cas particulier d'une somme de v.a. de Bernoulli.

## dénombrable

(*enumerable, countably infinite*)

Qualifie un ensemble infini « numérotable », comme par exemple l'ensemble  $\mathbf{N}$  des entiers naturels ou l'ensemble  $\mathbf{Z}$  des entiers relatifs. Il existe des ensembles infinis « plus gros » que les ensembles dénombrables, par exemple l'ensemble  $\mathbf{R}$  des nombres réels.

**dénombrement***(enumeration, counting)*

Terme générique qui désigne une situation où l'on compte des « configurations », formées à partir d'« objets » pris dans un ensemble fini donné et disposés en respectant certaines contraintes ou certaines structures fixées.

Les dénombrements (arrangements, combinaisons, permutations) jouent un rôle important en *probabilités combinatoires*, où l'hypothèse d'équiprobabilité ramène la détermination des probabilités à des comptes d'événements élémentaires.

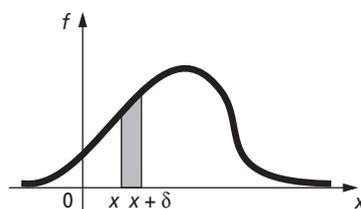
**densité [de probabilité]***([probability] density)*

Fonction réelle positive continue  $f$  associée à toute variable aléatoire réelle absolument continue  $X$ . Elle fournit par intégration les probabilités d'intervalle :

$$P(a < X \leq b) = \int_a^b f(t) dt.$$

La densité est la dérivée de la fonction de répartition :  $f(x) = F'(x)$ , et inversement la fonction de

répartition est sa primitive :  $F(x) = \int_{-\infty}^x f(t) dt$ .



La densité possède la signification intuitive suivante : si deux points  $x$  et  $x + \delta$  sont « très voisins », comme la densité est une fonction continue, sa valeur est « à peu près constante » et égale à  $f(x)$  sur l'intervalle  $[x, x + \delta]$  ; alors la probabilité d'intervalle  $P(x < X \leq x + \delta) =$

$\int_x^{x+\delta} f(t) dt$  est « à peu près » égale au produit  $f(x)\delta$  (cela est conforme à la signification standard du concept de densité).

On notera que la contrainte que la probabilité totale soit égale à 1 s'exprime par l'intégrale de la fonction densité :  $\int_{-\infty}^{\infty} f(x) dx = 1$ .

La notion de densité se généralise naturellement et sans difficulté à un couple puis à un «  $n$ -uple » de variables aléatoires réelles.

**descriptive (statistique)**

Partie de la statistique qui étudie le recueil, la présentation et le résumé des données sans faire intervenir de modèle probabiliste.

Voir *statistique*.

**détermination (coefficient de)***(coefficient of determination)*

Expression qui désigne parfois, notamment dans un contexte de régression, le carré  $r^2$  du *coefficient de corrélation* (on rappelle que le poids de la « liaison » affine révélée par le coefficient de corrélation est précisément mesuré par ce carré).

**diagramme en arbre**

Voir *arbre*.

**diagramme en barres (en tuyaux d'orgue, en bâtons...) (bar chart)**

Ces diagrammes sont des représentations graphiques très « parlantes » visuellement, que l'on peut utiliser aussi bien en calcul des probabilités qu'en statistique, pour des variables quantitatives continues comme pour des variables quantitatives discrètes. Diverses variantes graphiques (et diverses dénominations) sont utilisées.

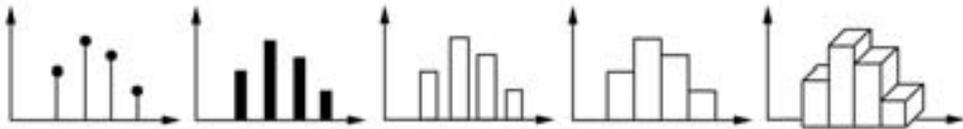
Soit une distribution statistique réelle discrète et finie, *i.e.* un ensemble fini  $\{x_i\}$  de nombres réels, présenté en suite croissante  $x_1 < x_2 < \dots < x_k$ , avec des fréquences associées  $\{f_i\}$ .

On appelle diagramme en bâtons (ou en barres, ou en tuyaux d'orgue) de cette distribution toute représentation graphique imagée qui, pour chaque valeur  $x_i$ , dessine un trait vertical ou un rectangle de hauteur proportionnelle à la fréquence  $f_i$ .

On peut étendre cette définition au cas probabiliste, en remplaçant les fréquences par les probabilités, cela ne soulève aucune difficulté.

On peut étendre cette définition au cas d'une distribution discrète avec beaucoup de valeurs, ou d'une distribution *continue*, et la difficulté est alors celle d'une définition efficace des classes, regroupements de valeurs discrètes ou « tranches » continue de valeurs.

Les figures ci-dessous présentent les variantes graphiques les plus usuelles.



Un certain nombre de conventions garantissent une représentation graphique fidèle de la distribution considérée. Les deux essentielles sont données ci-dessous.

1. L'axe des abscisses porte les valeurs de la variable. Lorsque les valeurs représentées ne sont pas des valeurs ponctuelles mais des classes, elles doivent être marquées sur le graphique par leur milieu. Inversement, lorsqu'on représente des valeurs ponctuelles par des barres ou des tuyaux d'orgue, la valeur représentée doit être marquée au milieu de la base du rectangle barre ou tuyau d'orgue.
2. L'axe des ordonnées porte les valeurs des fréquences ou des effectifs ou des probabilités (entre les fréquences et les effectifs il n'y a pas de différence visuelle, seule change la graduation des ordonnées).

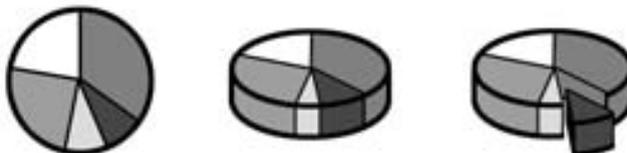
On confond parfois les diagrammes en barre et les *histogrammes* (qui sont toujours dessinés par rectangles contigus – en tuyaux d'orgue). La différence est la suivante : dans un diagramme (en barres ou en en tuyaux d'orgue), les hauteurs des rectangles sont proportionnelles aux effectifs (ou aux fréquences ou aux probabilités), tandis que dans un histogramme, ce sont les aires des rectangles qui sont proportionnelles aux effectifs (ou aux fréquences ou aux probabilités), et les hauteurs représentent alors l'analogie d'une densité. Cela étant, lorsque les tranches sont toutes de largeur égale (ce qui est de toute façon préférable si c'est possible), les deux types de représentation graphique sont identiques.

Signalons enfin la possibilité de représenter des diagrammes cumulés, en utilisant pour chaque valeur ou chaque classe la fréquence ou l'effectif cumulé.

## diagramme en secteurs (en « camemberts »)

(*sector chart, pie chart*)

Représentations graphiques de type circulaire pour des variables qualitatives discrètes. Les figures ci-dessous présentent trois variantes graphiques.

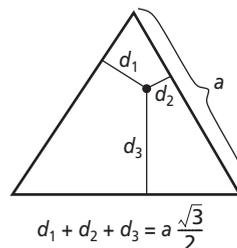


La « logique » de ces représentations est celle des histogrammes : les *aires* des secteurs circulaires sont proportionnelles aux effectifs.

### diagramme triangulaire

Représentation graphique d'un ensemble de données statistiques portant sur trois variables quantitatives dont la somme est constante. La pertinence de la représentation est fondée sur la propriété géométrique suivante : pour tout point intérieur à un triangle équilatéral, la somme des distances de ce point aux trois côtés est constante. (Les utilisations des diagrammes triangulaires vont bien au-delà du domaine de la statistique).

(triangular chart)



### dichotomique

Se dit d'une variable aléatoire ou d'une variable statistique (ou caractère) qualitative qui prend seulement deux modalités (par exemple réponse oui ou non à une question). Selon le but poursuivi, une telle variable sera traitée comme une « vraie » variable qualitative, ou bien un codage par 1 et 0 permettra de lui attribuer des indicateurs numériques utilisables dans une analyse d'ensemble incluant d'autres variables.

(dichotomous)

### différences couplées

Voir *appariées (séries)*.

### discret (ensemble)

Se dit d'un ensemble de nombres formé par des valeurs isolées (*i.e.* de nombres dont chacun est au milieu d'un intervalle qui ne contient que lui comme nombre de l'ensemble). S'oppose à *continu*, mais il y a néanmoins des ensembles de type intermédiaire ou hybride.

Tous les ensembles finis sont discrets, mais il y a aussi des ensembles infinis discrets comme

$\mathbf{N}$  (entiers naturels),  $\mathbf{Z}$  (entiers), ou par exemple  $E = \left\{ 1 - \frac{1}{n} \mid n \in \mathbf{N}^* \right\}$ .

(discrete set)

### discrète (variable aléatoire)

Voir *variable aléatoire (typologie)*.

(discrete random variable)

### disjoints

Se dit d'évènements  $A, B$  dont l'intersection est vide. On a alors  $P(A \cap B) = P(\emptyset) = 0$ . Le qualificatif logique synonyme est *incompatibles*.

Étant donné un nombre supérieur à 2 d'évènements :  $A_1, A_2, \dots, A_n$ , on prendra garde qu'il existe deux manières de généraliser cette propriété : soit en énonçant que les évènements sont « globalement » disjoints :  $A_1 \cap A_2 \cap \dots \cap A_n = \emptyset$ , ce qui ne présente en général guère d'intérêt, soit en énonçant que les évènements sont « 2 à 2 » disjoints :  $\forall i \forall j i \neq j \Rightarrow A_i \cap A_j = \emptyset$ , propriété souvent très utile.

(disjointed sets)

### disjonction logique

Voir *réunion*.

**dispersion (boîte de)**

Voir *boîte de dispersion*.

**dissymétrie (coefficient de)**

Voir *asymétrie (coefficient d')*.

**distribution**

([*probability, statistical*] *distribution*)

Dans un contexte probabiliste, synonyme de *loi de probabilité (d'une variable aléatoire)*.

Dans un contexte statistique, désigne généralement un *échantillon* statistique ou le tableau de valeurs qui le représente.

**données**

(*data*)

Terme générique employé le plus souvent en statistique pour désigner les résultats recueillis lors d'une expérimentation ou d'une observation.

**données (analyse des)**

Voir *analyse des données*.

**données censurées**

Voir *censurées (données)*.

**Doob (Joseph)**

Mathématicien américain (1910). Il a étudié les martingales et les processus stochastiques.

**droite d'ajustement**

(*fitted line*)

Droite d'équation  $y = a + bx$  que l'on détermine pour représenter « au mieux » un *nuage*  $\{(x_i, y_i)\}$  de points observés. Dans un contexte de type déterministe avec erreurs de mesure, un type standard de méthode consiste à définir de façon générale une « distance » entre une droite et un nuage de points, et à déterminer la droite d'ajustement d'un nuage donné comme la droite qui minimise la distance à ce nuage. On prend généralement comme distance la

somme des carrés des écarts  $\sum_{i=1}^n (y_i - (a + bx_i))^2$ , ce qui donne la *droite des moindres*

*carrés*. Dans un contexte de type intrinsèquement aléatoire, où les  $(x_i, y_i)$  sont des observations d'un couple  $(X, Y)$  de v.a., on suppose que l'espérance conditionnelle  $E(Y|X = x)$  est une fonction affine de  $x$  :  $E(Y|X = x) = a + bx$ , et on ajuste la droite en minimisant la variance des écarts, ce qui donne la *droite de régression*. Les deux méthodes ne diffèrent que par la présentation et fournissent la même droite.

**droite de Henry**

(*Henry line*)

Droite d'ajustement, dans une échelle « gaussio-arithmétique », de la distribution statistique d'un échantillon d'une variable aléatoire normale, qui permet de déterminer graphiquement l'espérance et l'écart-type de la loi normale. Cette droite permet en outre de contrôler de façon sommaire l'hypothèse de normalité.

Soit  $X$  une v.a. normale  $\mathcal{N}(\mu, \sigma)$ , et soit  $F$  sa fonction de répartition. Si on désigne par  $\Pi$  la fonction de répartition d'une v.a. normale centrée réduite, et  $g = \Pi^{-1}$  la fonction réciproque,

on a  $F(x) = \Pi\left(\frac{x-\mu}{\sigma}\right)$ , d'où  $g(F(x)) = \frac{x-\mu}{\sigma}$ , équation d'une droite dans un système de coordonnées où l'on porte en abscisses les valeurs de  $x$ , et en ordonnées les valeurs de  $g(F(x))$ . Comme le papier utilisé est « gaussio-arithmétique », les ordonnées représentent l'écart dans une loi normale centrée réduite, mais l'axe est gradué de façon à permettre de reporter les valeurs de  $\Pi(x)$ .

Pour appliquer cette méthode à un échantillon statistique, il faut ordonner les valeurs observées :  $x_1 \leq x_2 \leq \dots \leq x_n$ , puis définir la fonction de répartition observée en posant  $F(x_1) = \frac{0,5}{n}$ ,

$F(x_2) = \frac{1,5}{n}$ , ...,  $F(x_n) = \frac{n-0,5}{n}$ . On reporte alors les points  $(x_i, F(x_i))$  sur le papier gaussio-arithmétique, on trace au mieux la droite d'ajustement, et on détermine graphiquement les valeurs de  $\mu$  et de  $\sigma$ . Lorsque  $n$  est « petit », la « correction de continuité » de  $-0,5$  est nécessaire. Lorsque  $n$  est « grand », cette correction est d'autant moins utile que l'on regroupe généralement les valeurs observées en classes et que l'on se contente de reporter les points correspondant aux extrémités des classes.

Signalons que l'on appelle *transformation en Probit* la transformation  $y = 5 + g(F(x))$  et *Probit* la valeur du résultat : ainsi les probits 3, 4, 5, 6, 7 correspondent aux valeurs  $\mu - 2\sigma$ ,  $\mu - \sigma$ ,  $\mu$ ,  $\mu + \sigma$ ,  $\mu + 2\sigma$  de la loi normale ajustée. Certains papiers gaussio-arithmétiques (papiers « Probit ») portent les Probits sur l'axe des ordonnées. Enfin, on peut utiliser des tables de conversion en Probit pour calculer soi-même les probits et utiliser un papier ordinaire pour reporter les points de la droite de Henry.

## droite de régression [linéaire]

(*regression line*)

Droite d'équation  $y = a + bx$  que l'on détermine pour rendre compte de la liaison linéaire (affine) probabiliste ou statistique entre deux variables. Cette notion concerne deux types différents de situations mais les formules sont identiques.

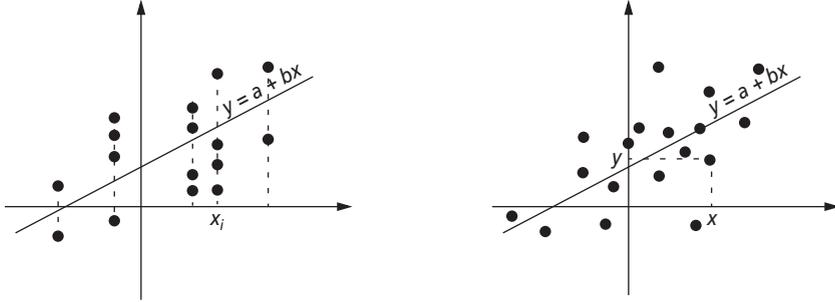
Dans le premier type de situation, on a un ensemble  $\{x_i\}$  de valeurs « maîtrisées » ou « contrôlées » d'une variable  $x$  et, pour chaque  $i$ , une variable aléatoire  $Y_i$ . On peut poser  $Y_i = a + bx_i + E_i$ , où chaque  $E_i$  est une variable aléatoire « écart » ou « fluctuation ». La question posée est de rechercher les valeurs des coefficients numériques  $a$  et  $b$  qui minimisent – en un sens approprié de ce mot – globalement les écarts  $E_i$ . Cette situation est souvent qualifiée de *modèle linéaire*.

Dans le second type de situation, on a un couple  $(X, Y)$  de variables aléatoires et l'on peut poser, de façon similaire au premier type de situation,  $Y = a + bX + E$ , où  $E$  est une variable aléatoire globale « écart ». La question posée est de rechercher les valeurs des coefficients numériques  $a$  et  $b$  qui minimisent – toujours en un sens approprié du mot – l'écart  $E$ . Cette situation constitue la *régression linéaire stricto sensu*.

Pour donner un fondement probabiliste aux formules de la régression linéaire utilisées en statistique, il faut faire certaines hypothèses. Comme très souvent, la méthode est « robuste » et les résultats restent admissibles si la réalité ne s'écarte « pas trop » de ces hypothèses (il existe d'ailleurs des tests qui permettent de contrôler la qualité de l'approximation réalisée).

Dans le premier type de situation, il faut supposer que les variables aléatoires  $Y_i$  sont normales, indépendantes et de même écart-type. Seule leur espérance mathématique  $\mu_i = E(Y_i)$  varie, et il faut enfin supposer que cette variation est une fonction linéaire (*stricto sensu* affine) de

$x_i : \exists \alpha, \exists \beta \mu_i = \alpha + \beta x_i$ . Les coefficients numériques  $\alpha$  et  $\beta$  sont précisément ceux qui minimisent globalement les écarts  $E_i$ , et il faut les estimer à partir des données recueillies.



Si on note globalement  $X$  l'ensemble  $\{x_i\}$  des valeurs de la variable contrôlée, et si on note globalement  $Y$  la famille des variables aléatoires  $(Y_i)$ , on peut par convention appeler l'espérance  $\mu_i = E(Y_i)$  « espérance conditionnelle » (à  $X = x_i$ ) de  $Y$ , et la noter  $E(Y|X = x_i)$ ; et on peut de même par convention appeler la variance  $\sigma_i^2 = \sigma^2(Y_i)$  « variance conditionnelle » (à  $X = x_i$ ) de  $Y$ , et la noter  $\sigma^2(Y|X = x_i)$  (dans le cas présent  $\sigma_i^2 = \sigma^2$  ne dépend pas de  $x_i$ ).

Dans le second type de situation, il faut supposer que le couple  $(X, Y)$  suit une loi normale à 2 dimensions. Quoique ce soit un peu délicat, on peut définir mathématiquement l'espérance conditionnelle  $E(Y|X = x)$  et la variance conditionnelle  $\sigma^2(Y|X = x)$ . En tout état de cause, on peut percevoir intuitivement ce que sont ces notions. La loi normale à 2 dimensions que suit le couple  $(X, Y)$  possède une propriété fondamentale : l'espérance conditionnelle est une fonction linéaire (affine) de  $x$  :  $E(Y|X = x) = a + bx$ , et la variance conditionnelle  $\sigma^2(Y|X = x) = \sigma^2$  ne dépend pas de  $x$  (propriété appelée parfois *homoscedasticité*). Les coefficients numériques  $\alpha$  et  $\beta$  sont précisément ceux qui minimisent l'écart  $E$ . Ils sont fonctions des paramètres de la loi de  $(X, Y)$ , et il faut les estimer à partir des données recueillies.

### Coefficients de la droite de régression (empirique) $y = a + bx$

Formules globales des estimations :

$$a = \bar{y} - \frac{\text{Cov}_{X,Y}}{s_X^2} \bar{x}, \quad b = \frac{\text{Cov}_{X,Y}}{s_X^2}$$

Traduction pour un échantillon de  $n$  couples de valeurs numériques observées  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$  :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{i=n} y_i, \quad b = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}, \quad a = \bar{y} - b\bar{x}$$

(il existe plusieurs variantes (mineures) de ces formules).

La droite de régression passe par le point  $(\bar{x}, \bar{y})$  et peut aussi s'écrire :

$$\frac{y - \bar{y}}{s_Y} = r \frac{x - \bar{x}}{s_X} \quad \text{ou encore} \quad y = \bar{y} + r \frac{s_Y}{s_X} (x - \bar{x})$$

(rappel :  $r = \frac{\text{Cov}_{X,Y}}{s_X s_Y}$ )

A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

Le cas d'un échantillon de  $q$  groupes d'observations relatifs aux valeurs (« contrôlées » ou aléatoires discrètes)  $x_1, x_2, \dots, x_q$  de  $X$ , avec pour le groupe n°  $i$ ,  $n_i$  valeurs observées  $y_{i1}, y_{i2}, \dots, y_{in_i}$  de la variable  $Y_i$ , se traite avec les formules précédentes (on peut trouver dans des ouvrages spécialisés des formules spécifiques, mais les moyens de calcul actuels leur ont fait perdre leur intérêt).

On notera que, malgré leur forme statistique, ces formules sont exactement celles qui donnent les coefficients de la droite des moindres carrés. On notera aussi (cela rendra naturelle la généralisation à plusieurs variables explicatives) que  $a$  et  $b$  sont les solutions du système :

$$\begin{cases} an + b \sum x_i = \sum y_i \\ a \sum x_i + b \sum x_i^2 = \sum x_i y_i \end{cases}$$

(tous les signes  $\sum$  sont à comprendre comme  $\sum_{i=1}^{i=n}$ ), système qui résulte (naturellement !) de

la caractérisation du minimum dans la méthode des moindres carrés.

Le terme de *régression* est dû à Galton et traduit le fait que, si l'on étudie par exemple la taille des enfants en fonction de la taille moyenne des deux parents, il y a « régression » vers la taille moyenne de toute la population ; en effet les enfants de parents grands sont eux-mêmes grands, mais un peu moins (en moyenne) que leurs parents, tendance qui résulte simplement du coefficient  $r < 1$  dans l'équation standard de la droite de régression.

Voir *décomposition de la variance, loi des estimateurs et intervalles de confiance, prédiction, régression à deux variables explicatives*.

### droite des moindres carrés (least square line)

Étant donné un ensemble  $\{(x_i, y_i)\}$  de points observés qui se situent « à peu près » le long d'une droite, la droite des moindres carrés est la droite d'équation  $y = a + bx$  qui minimise la

somme des carrés des écarts  $\sum_{i=1}^n (y_i - (a + bx_i))^2$ .

#### Coefficients de la droite des moindres carrés $y = a + bx$

- Coordonnées du barycentre des points

$$x_G = \frac{1}{n} \sum_{i=1}^{i=n} x_i, \quad y_G = \frac{1}{n} \sum_{i=1}^{i=n} y_i.$$

- Pente de la droite des moindres carrés

$$b = \frac{\sum_{i=1}^{i=n} (x_i - x_G)(y_i - y_G)}{\sum_{i=1}^{i=n} (x_i - x_G)^2}.$$

- Ordonnée à l'origine (la droite des moindres carrés passe par le barycentre)

$$a = y_G - bx_G.$$

**Durbin-Watson (test de)***(Durbin-Watson test)*

Test d'autocorrélation de rang 1 pour contrôler l'indépendance des résidus ( $e_i$ ), classés par valeurs croissantes de la variable  $x_i$ , d'une régression. La variable de test est fabriquée comme

un coefficient de corrélation :  $d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} = 2 - 2 \frac{\sum_{i=2}^n e_i e_{i-1}}{\sum_{i=1}^n e_i^2}$ , et vérifie  $0 \leq d \leq 4$ .

Ce test n'est pertinent que si les valeurs de  $x_i$  ne sont pas trop irrégulièrement espacées. Il est notamment utilisé en économétrie. On trouvera les détails de son fonctionnement dans les ouvrages spécialisés.



## écart (deviation)

Dans une distribution probabiliste ou statistique, étant donné un indicateur de tendance centrale  $C$  de la variable, on appelle écart la différence  $x - C$  entre une valeur prise  $x$  et  $C$ . Le plus souvent, l'indicateur est l'espérance mathématique  $\mu$  (en probabilités) ou la moyenne observée  $m$  (en statistique), et l'écart est donc  $x - \mu$  ou  $x - m$ .

On appelle « écart absolu » la valeur absolue de l'écart  $|x - \mu|$  ou  $|x - m|$ .

## écart intercentiles (interpercentile range)

Indicateur de dispersion attaché à une variable aléatoire réelle, utilisé essentiellement en statistique. C'est l'écart entre le premier *centile* et le dernier centile ; il englobe donc une partie médiane de la distribution de 98 % de l'effectif total.

## écart interdéciles (interdecile range)

Indicateur de dispersion attaché à une variable aléatoire réelle, utilisé essentiellement en statistique. C'est l'écart entre le premier *décile* et le dernier décile ; il englobe donc une partie médiane de la distribution de 80 % de l'effectif total.

## écart interquartiles (interquartile range)

Indicateur de dispersion attaché à une variable aléatoire réelle. C'est l'écart entre le *quartile* inférieur et le quartile supérieur ; il englobe donc une partie médiane de la distribution de probabilité 0,5 ou (en statistique) de 50 % de l'effectif total.

## écart réduit (reduced deviation, normal deviation)

Dans une distribution probabiliste, d'espérance  $\mu$  et d'écart-type  $\sigma$ , ou statistique, de moyenne  $m$  et d'écart-type  $s$ , étant donnée une valeur  $x$ , on définit l'écart par la différence  $x - \mu$  ou

$x - m$ . L'écart réduit est alors le quotient  $t = \frac{x - \mu}{\sigma}$  ou  $\frac{x - m}{s}$  ; on peut aussi le « lire » dans

l'écriture  $x = \mu + t\sigma$  ou  $x = m + ts$ .

## écart-type (standard deviation)

Que l'on soit en probabilités ou en statistique, la dimension « métrologique » de la variance est le carré de la dimension de la variable, et il faut prendre sa racine carrée pour retrouver une valeur interprétable concrètement.

L'écart-type est défini comme la racine carrée de la variance.

En probabilités, on note  $\sigma(X)$  ou  $\sigma_X$  (ou  $\sigma$  s'il n'y a aucun risque de confusion) l'écart-type de la variable aléatoire  $X$  :  $\sigma(X) = \sqrt{\text{Var}(X)} = \sqrt{\sigma^2(X)}$ .

En statistique,  $X$  étant la variable aléatoire dont on étudie un échantillon, on note le plus souvent  $s_X$  ou  $s_x$  (ou  $s$  s'il n'y a aucun risque de confusion) l'écart-type (parfois qualifié d'observé ou d'empirique) :  $s_X = \sqrt{\text{Var}_X} = \sqrt{s_X^2}$ .

## échantillon

([random] sample)

Concept commun au calcul des probabilités et à la statistique, désignant, soit une suite (les mathématiciens disent un «  $n$ -uple »)  $X_1, X_2, \dots, X_n$  de  $n$  variables aléatoires indépendantes et de même loi, soit une suite  $x_1, x_2, \dots, x_n$  de  $n$  valeurs prises par  $n$  variables aléatoires indépendantes et de même loi (dans la pratique, on pourra observer le résultat de  $n$  expérimentations simultanées, ou bien observer successivement les  $n$  résultats d'une expérimentation renouvelée  $n$  fois à l'identique).

Il peut arriver que l'exigence d'indépendance soit affaiblie ou abandonnée. On devra alors préciser explicitement que l'on considère un échantillon indépendant lorsque cela sera le cas.

Étant donnée une « population », le même mot est souvent employé pour désigner le résultat du « tirage » aléatoire de  $n$  « individus » dans cette population, *avant* l'observation d'une variable statistique (ou caractère) défini sur la population.

## échantillon représentatif

(representative sample)

Échantillon permettant d'estimer efficacement et sans biais les caractéristiques de la population dont il est issu. C'est le cas des échantillons *tirés au sort* (avec la même probabilité pour chaque individu de la population), mais c'est également le cas d'échantillons constitués selon des méthodes qui permettent d'améliorer la précision des estimations sans altérer la représentativité.

Voir *sondage*.

## échantillonnage

(sampling)

Opération de recueil de données pour un échantillon d'individus d'une population.

Ce mot est l'exact synonyme de *sondage*, même si les habitudes font utiliser préférentiellement l'un ou l'autre selon les situations.

## échantillons appariés

Voir *appariés (échantillons)*.

## effectif

([sample] size, class number)

Nombre d'individus d'une classe ou d'un échantillon.

L'effectif total d'un échantillon s'appelle souvent la *taille* de l'échantillon.

## élasticité

(elasticity)

Terme utilisé en statistique économique pour qualifier le « quotient des variations relatives »

$\frac{dy/y}{dx/x} = \frac{y'}{y/x}$  : l'élasticité est constante lorsque la liaison (exacte ou approchée) est du type

$y = bx^c$ , l'exposant  $c$  est l'élasticité (ou le coefficient d'élasticité), et ce type de liaison se traduit par une droite sur un graphique logarithmique :  $\ln y = \ln b + c \ln x$ .

## empirique

(empirical)

Lorsque cet adjectif n'est pas employé dans son sens général de la langue courante (procédé ou méthode empirique, ...), il est synonyme d'« observé » et qualifie les paramètres des distributions statistiques (moyenne empirique, variance empirique, ...), par opposition aux paramètres « théoriques » correspondants des lois de probabilités.

## entropie

(entropy [of information source])

Dans le cadre de la *théorie de l'information*, on appelle système simple la donnée d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et d'une partition finie de cet espace, dont on note généralement  $A_i$  les évènements et  $p_i$  les probabilités correspondantes, ou bien d'une variable aléatoire finie  $X$ , dont on note généralement  $A_i$  les valeurs prises et  $p_i$  les probabilités correspondantes. Ces deux présentations sont bien sûr équivalentes, la deuxième a l'avantage de permettre d'employer le concept d'espérance mathématique.

Soit  $X = \{A_i\}$  un système simple. On définit l'entropie de  $X$  comme l'espérance mathématique de la *quantité d'information* apportée par la réalisation d'un évènement du système :

$$H(X) = E(I(X)) = \sum_{A \in X} P(A)I(A) = - \sum_i p_i \log p_i.$$

On utilise généralement le logarithme à base 2, et l'entropie s'exprime alors en *bits*.

Cette notion a été introduite en théorie de l'information par Hartley, puis développée par Shannon. Elle a la même signification concrète que l'entropie introduite un siècle auparavant par Clausius en physique statistique, puis étudiée par Boltzmann : une forte entropie signifie désordre, manque d'information, incertitude...

L'entropie est généralement notée  $H$  par les mathématiciens et  $S$  par les physiciens.

**Théorème.** Si  $X$  est un système simple à  $n$  « états », on a :

$$H(X) \leq \log_2 n,$$

avec égalité si et seulement si les  $n$  états sont équiprobables.

L'entropie est maximale si les  $n$  états sont équiprobables, *i.e.* si le système est dans l'incertitude maximale.

**Exemple** Si une source d'information, représentée par un système simple  $S$ , émet aléatoirement deux symboles (penser par exemple aux pixels Noir et Blanc générés pour une télécopie) avec probabilités  $p$  et  $q = 1 - p$ , l'entropie de  $S$  (*i.e.* la quantité *moyenne* d'information apportée par l'émission d'un symbole) est :

$$h(p) = -p \log_2 p - (1 - p) \log_2 (1 - p).$$

Si par exemple  $p = 0,2$  et  $1 - p = 0,8$ , on a  $H(S) = h(0,2) = 0,722$  bit (quantité qui n'est pas *très* inférieure au maximum 1 bit) ; si par exemple  $p = 0,05$  et  $1 - p = 0,95$ , on a  $H(S) = h(0,05) = 0,286$  bit (quantité qui n'est pas *très* petite).

## épreuve

(trial)

Concept probabiliste qui modélise l'expérimentation ou l'observation d'un phénomène aléatoire.

Formellement, étant donné un espace probabilisable  $(\Omega, \mathcal{A})$ , une épreuve est définie comme le « tirage » d'un évènement élémentaire  $\omega \in \Omega$ , qui est le « résultat » de l'épreuve.

**Exemples** Lancer un dé, tirer une carte dans un jeu, tirer une boule dans une urne, prendre « au hasard » un individu dans une population, prélever un échantillon d'un minerai, croiser deux individus en génétique, mettre en service un dispositif susceptible de tomber en panne, ensemercer une parcelle (dont on mesurera par exemple le rendement), tirer une balle dans une cible, répondre à une question dans un jeu, jouer un match, etc.

## épreuves répétées

Locution qui qualifie une succession d'épreuves indépendantes et identiques. Ces épreuves étant représentées par des variables aléatoires  $X_i$ , on écrit souvent que les  $X_i$  sont « i.i.d. » : indépendantes et identiquement distribuées.

On note  $\mu$  l'espérance mathématique commune aux  $X_i$  réelles et  $\sigma^2$  leur variance commune.

On définit la *somme* :

$$S_n = X_1 + X_2 + \dots + X_n$$

et la *moyenne* :

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

On a alors :

$$E(S_n) = n\mu$$

$$E(M_n) = \mu$$

$$\text{Var}(S_n) = n\sigma^2$$

$$\text{Var}(M_n) = \frac{\sigma^2}{n}$$

$$\sigma(S_n) = \sigma\sqrt{n}$$

$$\sigma(M_n) = \frac{\sigma}{\sqrt{n}}$$

## équiprobabilité

(*equipossibility*)

Hypothèse qui pose que tous les événements élémentaires d'un espace probabilisable *fini* ont la même probabilité. Cette hypothèse permet de ramener la détermination des probabilités à des calculs de *dénombrements*.

Soit un espace fini équiprobabilisé dont l'espace fondamental  $\Omega$  est constitué par  $N$  événements élémentaires, et soit  $A$  un événement constitué par  $k$  événements élémentaires, alors :

$$P(A) = \frac{k}{N}$$

Une formulation imagée ancienne énonce que  $P(A)$  est égal au quotient du « nombre de cas favorables (à  $A$ ) » par le « nombre (total) de cas possibles ».

On justifie souvent l'hypothèse d'équiprobabilité par des raisons de symétrie, géométrique et/ou mécanique. Plus modestement (plus lucidement ?) Bernoulli avançait le principe de « raison insuffisante » (que l'on pourrait traduire plus brutalement par principe d'égal ignorance). Quoi qu'il en soit, on ne démontre pas l'équiprobabilité, on la pose en hypothèse (et on peut envisager ultérieurement de la tester par une étude statistique).

## équitable

(*fair*)

Voir *pari*.

## erf (error function)

Intégrale numérique que l'on trouve dans de nombreux « packages » logiciels et qui est liée à la fonction de répartition de la loi normale :

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

Valeurs particulières :

$$\operatorname{erf}(-\infty) = -1 \quad ; \quad \operatorname{erf}(0) = 0 \quad ; \quad \operatorname{erf}(\infty) = 1$$

Si  $F(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$  est la fonction de répartition de la loi normale centrée réduite, on a :

$$F(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)$$

## erfc (complementary error function)

Intégrale numérique que l'on trouve dans de nombreux « packages » logiciels et qui est liée à la fonction de répartition de la loi normale :

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x).$$

## Erlang (loi d')

(Erlang distribution)

Loi d'une variable aléatoire continue, cas particulier de la *loi gamma* pour les valeurs entières du premier paramètre, et qui est notamment utilisée comme *temps d'attente* dans le processus de Poisson.

### Formulaire

Un paramètre entier et un paramètre réel :  $r \in \mathbf{N}^*$  ;  $\lambda \in \mathbf{R}_+^*$  (paramètre d'échelle de temps). Valeurs sur les réels positifs.

#### ► Loi de probabilité

densité

fonction de répartition

$$f(x) = \frac{\lambda}{(r-1)!} e^{-\lambda x} (\lambda x)^{r-1} \quad (x > 0) \quad F(x) = 1 - e^{-\lambda x} \sum_{k=1}^r \frac{(\lambda t)^{k-1}}{(k-1)!} \quad (x \geq 0)$$

#### ► Valeurs caractéristiques

– espérance :  $E(X) = \frac{r}{\lambda}$

– variance :  $\operatorname{Var}(X) = \frac{r}{\lambda^2}$

– écart-type :  $\sigma(X) = \frac{\sqrt{r}}{\lambda}$

#### ► Cas particulier

Pour  $r = 1$ , la loi d'Erlang est la *loi exponentielle*.

### ► Utilisations

En théorie, la loi d'Erlang de paramètres  $r$  et  $\lambda$  est la loi de la somme de  $r$  variables aléatoires exponentielles de paramètre  $\lambda$  et indépendantes.

Dans la pratique, la loi d'Erlang est principalement utilisée comme loi du temps d'attente qui sépare les évènements numéros  $k$  et  $k + r$  dans un *processus poissonnien* de taux  $\lambda$ , par exemple une succession de pannes à « taux de défaillance » constant.

Elle est également utilisée comme loi de la durée de vie d'un système complexe dont le taux de défaillance augmente avec le temps : à la différence des *lois de Weibull*, les lois d'Erlang (avec  $r \geq 2$ ) permettent de modéliser des situations avec des taux de défaillance croissants mais néanmoins bornés et tendant vers une limite.

Voir *gamma (loi)*.

## espace fondamental (*basic space, event space, sample space*)

Ensemble de tous les évènements élémentaires susceptibles de se réaliser comme résultats de l'expérimentation ou de l'observation d'un phénomène aléatoire.

Pratiquement toujours noté  $\Omega$ , il est parfois appelé univers ou univers des possibles.

## espace probabilisable (*measurable space*)

Association d'un espace fondamental  $\Omega$  d'évènements « élémentaires » possibles et d'une famille  $\mathcal{A}$  de parties (évènements) de  $\Omega$ , dont on pourra ultérieurement définir ou calculer la probabilité.

Lorsque  $\Omega$  est fini ou dénombrable, on peut prendre pour  $\mathcal{A}$  l'ensemble  $\mathcal{P}(\Omega)$  de *toutes* les parties de  $\Omega$  (et en général on le fait). Mais, dès que  $\Omega$  n'est pas fini ou dénombrable, on se heurte à des difficultés mathématiques qui imposent de prendre une famille restreinte de parties.

On définit formellement un espace probabilisable comme un doublet  $(\Omega, \mathcal{A})$  constitué par :

- un ensemble non vide  $\Omega$ , appelé espace fondamental,
- une famille  $\mathcal{A}$  de parties de  $\Omega$  possédant les propriétés d'une *tribu* (ou  $\sigma$ -algèbre).

Les deux cas les plus employés en calcul des probabilités sont, si  $\Omega$  est fini ou dénombrable,  $\mathcal{A} =$  la tribu  $\mathcal{P}(\Omega)$  de toutes ses parties et, si  $\Omega = \mathbf{R}$ ,  $\mathcal{A} =$  la tribu des *boréliens* de  $\mathbf{R}$ .

**Remarque** : on trouvera dans certains manuels de probabilité le terme « espace mesurable », il est parfaitement synonyme d'« espace probabilisable ».

Voir *dénombrable, tribu*.

## espace probabilisé (*probability space*)

Association d'un espace fondamental  $\Omega$  d'évènements « élémentaires » possibles, d'une famille  $\mathcal{A}$  de parties (évènements) de  $\Omega$ , et d'une « mesure » de probabilité.

La mesure de probabilité fournit la probabilité de chaque évènement de  $\mathcal{A}$ , mais doit vérifier un certain nombre de propriétés (ou « axiomes ») destinés à assurer, d'une part la cohérence interne de l'assemblage ainsi composé, d'autre part un fonctionnement permettant de *modéliser* efficacement les situations concrètes aléatoires.

La construction des espaces probabilisés en deux temps : espace probabilisable, puis mesure de probabilité, donne la possibilité de « munir » un espace probabilisé de *plusieurs* mesures de probabilité différentes.

On définit formellement un espace probabilisé comme un triplet  $(\Omega, \mathcal{A}, P)$  constitué par :

- un ensemble non vide  $\Omega$ , appelé espace fondamental ;
- une famille  $\mathcal{A}$  de parties de  $\Omega$  possédant les propriétés d'une tribu (ou  $\sigma$ -algèbre) ;
- une application  $P : \mathcal{A} \rightarrow [0, 1]$  possédant les propriétés d'une mesure de probabilité.

### espérance mathématique (expectation, expected value)

Principal indicateur numérique de tendance centrale attaché à une variable aléatoire réelle. Sa signification concrète est celle d'une moyenne des valeurs prises, pondérée par les probabilités. Il joue un rôle important, associé à la variance ou à l'écart-type, dans de nombreux théorèmes et de nombreuses formules.

L'espérance d'une variable aléatoire  $X$  se note le plus souvent  $E(X)$  ou  $\mu_X$  (ou  $\mu$  s'il n'y a aucun risque de confusion), ou encore  $\bar{X}$  ; si  $\varphi$  est une fonction numérique,  $\varphi(X)$  est une variable aléatoire (cf. définition des variables aléatoires : une fonction de fonction est une fonction), et son espérance se note  $E(\varphi(X))$ .

Si la v.a. réelle  $X$  est discrète, caractérisée par l'ensemble (fini ou dénombrable) de valeurs  $\{x_i\}$ , avec les probabilités ponctuelles  $p_i = P(X = x_i)$ , on a :

$$E(X) = \sum_{x_i} x_i P(X = x_i) = \sum_{x_i} x_i p_i$$

(selon les cas, il s'agira d'une somme finie ou d'une somme infinie).

Si  $X$  est absolument continue, caractérisée par la densité de probabilité  $f(x)$ , on a :

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

Si  $\varphi$  est une fonction numérique, on a, selon le cas :

$$E(\varphi(X)) = \sum_{x_i} \varphi(x_i) P(X = x_i) = \sum_{x_i} \varphi(x_i) p_i$$

ou :

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

Il faut faire deux remarques. *Primo*, il n'y a pas de certitude que, si l'ensemble des valeurs est infini, la somme  $\sum_{x_i} x_i p_i$  converge, ou que l'intégrale  $\int_{-\infty}^{+\infty} x f(x) dx$  converge avant de l'avoir effectivement vérifié. Il existe des variables aléatoires qui n'ont pas d'espérance mathématique, par exemple la « v.a de Cauchy ». *Secundo*, les formules qui donnent  $E(\varphi(X))$  ne sont pas, malgré leur apparence, évidentes, et il faut les démontrer.

**Exemple** On considère une situation d'assurance d'un risque, que l'on modélise de la façon simplifiée suivante : l'assuré paye une prime annuelle  $P$ , et trois éventualités (dans l'année et par assuré) sont possibles :

aucun sinistre	probabilité $p_1 = 0,95$	coût pour l'assureur	0 €
« petit » sinistre	probabilité $p_2 = 0,04$	coût pour l'assureur	500 €
« gros » sinistre	probabilité $p_3 = 0,01$	coût pour l'assureur	5 000 €

On demande à partir de quel montant de la prime l'assureur fait du bénéfice.

Il faut introduire une variable aléatoire  $G = \text{gain}$  de l'assureur, étant entendu que ce « gain » peut être positif ou négatif. La v.a.  $G$  prend la valeur  $g_1 = P$  avec la probabilité  $p_1$ , la valeur  $g_2 = P - 500$  avec la probabilité  $p_2$ , et la valeur  $g_3 = P - 5\,000$  avec la probabilité  $p_3$ , donc :

$$\begin{aligned} E(G) &= p_1 g_1 + p_2 g_2 + p_3 g_3 \\ &= 0,95 P + 0,04 (P - 500) + 0,01 (P - 5\,000) = P - 70. \end{aligned}$$

L'assureur doit donc fixer un montant de prime supérieur à 70 € pour avoir une espérance mathématique de gain positive (le problème de la stratégie de l'assureur n'est pas complètement résolu pour autant et il resterait à examiner, pour  $n$  clients, les fluctuations du gain réel par rapport à  $n$  fois l'espérance mathématique).

L'analogie statistique de l'espérance mathématique est la *moyenne* (éventuellement qualifiée d'observée ou d'empirique).

## estimateur

(*estimator*)

Dans une situation d'épreuves répétées ( $X_1, X_2, \dots, X_n$ ), un estimateur est une variable aléatoire fonction des  $X_i$  :  $Y_n = Y_n(X_1, X_2, \dots, X_n)$ , dont les valeurs observées permettront d'obtenir des estimations de *paramètres* de la loi de probabilité commune aux  $X_i$ .

Par exemple, si  $\mu$  est l'espérance mathématique de la loi des  $X_i$ , la moyenne  $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$  est un estimateur de  $\mu$ .

Il n'y a pas de définition formelle des estimateurs, pour la simple raison qu'il n'est possible, ni techniquement ni conceptuellement, d'introduire des restrictions pour adapter les estimateurs aux paramètres considérés... étant donné un paramètre  $\theta$ , n'importe quelle variable aléatoire fonction des  $X_i$  est un estimateur de  $\theta$ . Par contre, il est possible de préciser mathématiquement les « qualités » d'un estimateur (ou plus précisément de la famille  $(Y_n(X_1, X_2, \dots, X_n))_{n \geq 1}$ ).

Voir *estimateur convergent*, *estimateur sans biais*, *estimateur asymptotiquement sans biais*, *estimateur efficace*.

## estimateur asymptotiquement sans biais (*asymptotically unbiased estimator*)

La variable aléatoire  $Y_n$  est un estimateur asymptotiquement sans biais du paramètre  $\theta$  si  $E(Y_n) \rightarrow \theta$  lorsque  $n$  tend vers  $+\infty$ .

Il est bien entendu préférable qu'un estimateur soit sans biais mais cela est techniquement impossible à obtenir simplement dans de nombreuses situations, et les estimateurs asymptotiquement sans biais peuvent être de très « bons » estimateurs.

## estimateur convergent

(*consistent estimator*)

La variable aléatoire  $Y_n$  est un estimateur convergent du paramètre  $\theta$  si  $E(Y_n) \xrightarrow{P} \theta$  lorsque  $n$  tend vers  $+\infty$  (il s'agit de convergence en probabilité mais dans la pratique la plupart des estimateurs convergent « presque sûrement »). Cette propriété est la plus importante qu'il faille exiger d'un estimateur.

Si un estimateur est convergent, il est asymptotiquement sans biais. Inversement, on peut démontrer que, si un estimateur est asymptotiquement sans biais et si sa variance tend vers 0 lorsque  $n$  tend vers  $+\infty$ , il est convergent.

## estimateur efficace

(*efficient estimator*)

La précision d'un estimateur est liée à sa variance, mais il ne suffit pas d'imposer une variance minimale pour obtenir le meilleur estimateur, car tous les estimateurs certains (*i.e.* de valeur

constante) sont de variance nulle ! Il faut donc ajouter une condition sur l'espérance mathématique.

La variable aléatoire  $Y_n$  est un estimateur efficace du paramètre  $\theta$  s'il est sans biais et s'il est de variance minimale parmi les estimateurs sans biais.

### estimateur sans biais

(unbiased estimator)

La variable aléatoire  $Y_n$  est un estimateur sans biais du paramètre  $\theta$  si  $E(Y_n) = \theta$ .

### estimation [ponctuelle]

(estimation, estimate [value])

Lorsque ce mot ne désigne pas la théorie générale de l'estimation, il désigne la valeur numérique prise par une variable aléatoire estimateur sur un échantillon statistique. Par exemple,

la variable aléatoire « moyenne »  $M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$  est un estimateur de l'espérance

mathématique, et la moyenne observée  $m$  ou  $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$  est une estimation de cette espérance.

Les estimations (ponctuelles) classiques dont la liste suit sont produites par les meilleurs estimateurs pour une grande classe de lois de probabilités (incluant notamment la loi normale). Néanmoins, pour certaines lois particulières, il peut exister des estimateurs meilleurs.

Voir *loi des estimateurs dans une régression et intervalles de confiance*.

#### ► estimation d'une espérance mathématique

Loi d'espérance  $\mu$  ; échantillon de  $n$  observations individualisées  $x_1, x_2, \dots, x_n$ .

Estimation standard de  $\mu$  par la moyenne :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$$

#### ► estimation d'une probabilité

Évènement  $A$  de probabilité  $p$  ; échantillon : sur  $n$  épreuves répétées,  $A$  a été observé  $n_A$  fois. Estimation standard de  $p$  par la fréquence :

$$\frac{n_A}{n}$$

#### ► estimation d'une variance

Loi d'espérance  $\mu$  et de variance  $\sigma^2$  ; échantillon de  $n$  observations individualisées  $x_1, x_2, \dots, x_n$ . Estimation préalable de  $\mu$  par  $\bar{x}$  (cf. ci-dessus).

Estimation biaisée de  $\sigma^2$  par l'écart quadratique moyen :

$$V_{(n)} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - \bar{x})^2$$

(l'estimateur correspondant a pour espérance mathématique  $\frac{n-1}{n} \sigma^2$ ).

Estimation *débiaisée* de  $\sigma^2$  par l'écart quadratique moyen corrigé :

$$V_{(n-1)} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - \bar{x})^2$$

(l'estimateur correspondant a pour espérance mathématique  $\sigma^2$ ).

► estimation d'un écart-type

Loi d'espérance  $\mu$  et de variance  $\sigma^2$ .

Estimation de  $\sigma$  par la racine carrée de l'estimation de la variance :

$$\sqrt{V_{(n)}} \text{ ou } \sqrt{V_{(n-1)}} \text{ (selon l'utilisation).}$$

► estimation d'une covariance

Formules analogues aux estimations de la variance. Voir *covariance*.

► estimation d'un coefficient de corrélation

Loi à deux variables d'espérances  $\bar{x}$  et  $\bar{y}$  (n'interviennent pas directement dans la formule), de variances  $s_X^2$  et  $s_Y^2$ , de covariance  $\text{Cov}_{X,Y}$ .

Estimation de  $\rho$  par le quotient :

$$r = \frac{\text{Cov}_{X,Y}}{s_X s_Y}$$

## estimation par intervalle

(*interval estimation*)

Remplacement, dans une estimation, de la donnée d'une valeur ponctuelle par celle d'un intervalle entourant cette valeur, dont les extrémités dépendent d'un seuil de confiance fixé *a priori*. Cet intervalle s'appelle *intervalle de confiance* du paramètre. Dans la présentation de résultats au public, on parle souvent de *fourchette* (malheureusement trop souvent sans préciser ni la taille de l'échantillon ni le seuil de confiance).

La détermination théorique d'un intervalle de confiance pour un paramètre  $\theta$  d'une loi de probabilité nécessite d'utiliser un estimateur  $Y = Y_n(X_1, X_2, \dots, X_n)$  de ce paramètre. Si on se donne le seuil de confiance  $1 - \alpha$ , et si on cherche un intervalle à risque symétrique, on peut déterminer, pour toute valeur  $\theta$  du paramètre, l'intervalle « d'acceptation »  $I_a(\theta) = [y_1, y_2]$  caractérisé par :

$$P(Y < y_1) = \frac{\alpha}{2}, P(y_1 \leq Y \leq y_2) = 1 - \alpha, P(Y > y_2) = \frac{\alpha}{2}.$$

On obtient ainsi un encadrement probabiliste de l'estimation lorsque le paramètre  $\theta$  est fixé. La détermination de d'intervalle de confiance  $I_c(y)$  nécessite d'inverser le procédé, afin d'obtenir un encadrement probabiliste du paramètre lorsque l'estimation  $y = Y_n(x_1, x_2, \dots, x_n)$  est fixée.

————— intervalle de confiance  $(1 - \alpha)$  d'une espérance mathématique —————

• Données. Un échantillon  $(x_1, x_2, \dots, x_n)$  de  $n$  valeurs observées d'une variable aléatoire numérique  $X$  d'espérance mathématique  $\mu$ .

A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

• Calculs

1. On calcule la moyenne  $\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$  de l'échantillon, puis on calcule la variance non

$$\text{biaisée } s^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - \bar{x})^2 \text{ de l'échantillon, et enfin l'écart-type } s = \sqrt{s^2}.$$

2. On détermine un écart réduit de Student  $t_\alpha$ , à lire dans la table de la loi de Student, pour le nombre de degrés de liberté de ddl =  $n - 1$  et le risque  $\alpha$ .

3. L'intervalle de confiance est :

$$I_c = \left[ \bar{x} - t_\alpha \frac{s}{\sqrt{n}}, \bar{x} + t_\alpha \frac{s}{\sqrt{n}} \right].$$

• Conditions et précautions

- En théorie X doit être une v.a. normale, donc aucune précaution si c'est le cas ;
- Lorsque ce n'est pas le cas, l'intervalle de confiance est robuste et reste utilisable si n est « assez grand », la condition  $n \geq 30$  est traditionnelle (en fait, on peut descendre en-dessous si la loi de X est continue et/ou symétrique).

————— intervalle de confiance  $(1 - \alpha)$  d'une probabilité —————

• Données. Un échantillon de n observations, sur lesquelles A a été observé k fois.

• Calculs

1. On calcule l'estimation de P(A) :  $p = \frac{k}{n}$ .

2. On détermine un écart réduit normal  $u_\alpha$ , à lire dans la table de la loi normale réduite, pour la probabilité  $\alpha$  de dépassement de l'écart absolu.

3. L'intervalle de confiance est :

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

• Conditions et précautions

Il faut que la distribution ne soit pas « trop » dissymétrique, ce qui se traduit par la double condition traditionnelle  $np \geq 10, n(1 - p) \geq 10$  (que l'on peut sans grand risque affaiblir en  $np \geq 5, n(1 - p) \geq 5$ ).

Dans le cas contraire, on peut recourir, soit à des tables spécifiques, soit à des calculs adaptés que l'on trouvera dans les ouvrages spécialisés.

————— intervalle de confiance  $(1 - \alpha)$  d'une variance —————

• Données. Un échantillon  $(x_1, x_2, \dots, x_n)$  de n valeurs observées d'une variable aléatoire normale X d'espérance mathématique  $\mu$  et de variance  $\sigma^2$ .

• Calculs

1. On calcule la moyenne  $\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$  de l'échantillon, puis on calcule la variance non

$$\text{biaisée } s^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - \bar{x})^2 \text{ de l'échantillon.}$$

2. On détermine deux bornes  $l_1$  et  $l_2$ , à lire dans la table de la loi du khi-deux à  $n - 1$  degrés de liberté :  $l_1$  est la valeur du  $\chi^2$  pour la probabilité  $1 - \frac{\alpha}{2}$ , et  $l_2$  la valeur pour la probabilité  $\frac{\alpha}{2}$ .

3. L'intervalle de confiance est :

$$I_c = \left[ \frac{n}{l_2} s^2, \frac{n}{l_1} s^2 \right].$$

• Conditions et précautions

Contrairement au cas d'une espérance ou d'une probabilité, l'intervalle de confiance de la variance n'est pas robuste : la formule ci-dessus est valable exclusivement dans le cas où la loi de l'échantillon est normale.

En comparant avec les tests d'hypothèses, on constatera que les intervalles de confiance sont fabriqués avec les mêmes formules que celles qui donne les variables de test : il est entièrement équivalent de rejeter une hypothèse  $H_0 = \langle \theta = \theta_0 \rangle$  ou de constater que  $\theta_0$  n'appartient pas à l'intervalle de confiance construit à partir des valeurs observées de l'échantillon.

## étendue

(range)

L'étendue d'une *série statistique* est l'écart entre ses valeurs extrêmes. C'est une caractéristique de dispersion médiocre car elle est trop sensible aux valeurs aberrantes (erronnées ou exceptionnelles).

On emploie aussi le mot étendue pour désigner la largeur d'une *classe*.

## évènement

([random] event, composite event)

Ensemble de résultats possibles de l'expérimentation ou de l'observation d'un phénomène aléatoire, *i.e.* sous-ensemble de l'ensemble de tous les résultats possibles.

Lorsqu'on a fixé un évènement A, l'épreuve effectuée donnant pour résultat un évènement élémentaire  $\omega \in \Omega$ , ou bien l'évènement A « s'est réalisé », si  $\omega \in A$ , ou bien l'évènement A « ne s'est pas réalisé », si  $\omega \notin A$ .

Formellement, étant donné l'espace probabilisable  $(\Omega, \mathcal{A})$  « associé » à une épreuve, un évènement est une partie A, partie qui doit appartenir à l'ensemble  $\mathcal{A}$  (dont on rappelle qu'il est égal à l'ensemble de toutes les parties de  $\Omega$  dans le cas simple où  $\Omega$  est fini ou dénombrable, mais qu'il est en général plus restreint).

Parmi les évènements, on trouve l'évènement vide  $\emptyset$  « rien du tout ne s'est passé » qui est donc impossible (et sa probabilité est 0), les « singletons », constitués par un seul évènement élémentaire, et l'évènement plein  $\Omega$  « n'importe quoi s'est passé » qui est donc certain (et sa probabilité est 1).

On peut spécifier un évènement par l'énumération (finie ou infinie) de tous les évènements élémentaires qui le constituent, mais le plus souvent on le définira par une propriété ou une caractéristique (*cf.* exemples ci-dessous).

**Exemples** « pair » pour un dé lancé et retombé, « trèfle » pour une carte tirée d'un jeu, « rouge » pour une boule tirée d'une urne, « mauvaise réponse » dans un jeu, « gain » dans un match, etc.

## évènement élémentaire

(simple event)

Résultat de l'expérimentation ou de l'observation d'un phénomène aléatoire.

Le vocabulaire n'est pas entièrement fixé : on parle aussi d'éventualité, parfois d'issue, plus rarement d'atome.

Formellement, étant donné l'espace probabilisable  $(\Omega, \mathcal{A})$  « associé » à une épreuve, un évènement élémentaire est un élément  $\omega$  de l'espace fondamental  $\Omega$ .

**Exemples** Le 3 montré par un dé lancé et retombé, le roi de trèfle tiré d'un jeu, une boule rouge tirée d'une urne, un individu pris « au hasard » dans une population, une plante produite par un croisement, une désintégration détectée par un compteur Geiger, un impact de balle dans la cible, la réponse C dans un jeu, le gain d'un match avec le score 4-2, etc.

### expérience aléatoire

Expression parfois employée pour désigner une épreuve.

### exponentielle (loi)

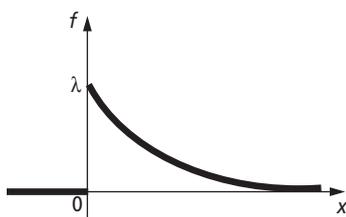
(exponential distribution)

Loi d'une variable aléatoire continue « temps d'attente » qui intervient notamment dans le processus de Poisson et la modélisation de la fiabilité.

#### Formulaire

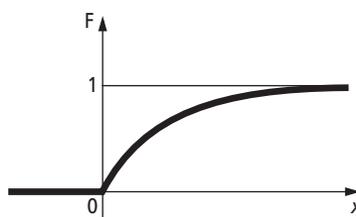
Un paramètre réel,  $\lambda \in \mathbf{R}_+^*$  ; valeurs sur les réels positifs.

#### ► Loi de probabilité



densité

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ \lambda e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$



fonction de répartition

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-\lambda x} & \text{si } x \geq 0 \end{cases}$$

#### ► Valeurs caractéristiques

- espérance :  $E(X) = \frac{1}{\lambda}$
- variance :  $\text{Var}(X) = \frac{1}{\lambda^2}$
- écart-type :  $\sigma(X) = \frac{1}{\lambda}$

#### ► Utilisations

En théorie, la loi exponentielle est la loi du temps d'attente d'un processus poissonnien de taux  $\lambda$  : temps d'attente du premier évènement, ou intervalle entre deux évènements successifs (dans certaines situations concrètes, on parle de « durée de vie » plutôt que de temps d'attente).

Dans la pratique, la loi exponentielle est la loi des situations concrètes modélisées par un processus poissonnien, comme une succession de pannes à « taux de défaillance » constant, ou la désintégration d'un atome radioactif (dans ce cas le processus est un processus « de mort » qui s'interrompt après le premier évènement).

Une erreur à ne pas commettre : la somme de deux variables aléatoires exponentielles indépendantes n'est jamais une variable aléatoire exponentielle. La loi de la somme de  $n$  variables aléatoires exponentielles indépendantes de même paramètre  $\lambda$  est une *loi d'Erlang* de paramètres  $n$  et  $\lambda$  (c'est le cas notamment du temps d'attente qui sépare les évènements numéros  $k$  et  $k + n$  dans un processus poissonnien).

**Exemple 1** On considère une variable aléatoire exponentielle  $X$ , qui modélise par exemple un processus de mort ou de désintégration, de « vie moyenne »  $\tau$ . Quelle est la « période » (au sens par exemple des physiciens en radioactivité) de  $X$  ?

Par vie moyenne, il faut toujours comprendre l'espérance mathématique. Comme  $E(X) = \frac{1}{\lambda}$ ,

on en déduit la valeur du paramètre  $\lambda$  :  $\lambda = \frac{1}{\tau}$ . Par période, il faut comprendre la valeur de  $T$

déterminée par  $F(T) = P(X \leq T) = \frac{1}{2}$  (avec la signification statistique suivante : au bout du temps  $T$ , environ la moitié d'un grand nombre de particules semblables se sont désintégrées).

Comme  $F(t) = 1 - e^{-\lambda t}$ , on a finalement  $T = \frac{\ln 2}{\lambda} = \tau \ln 2$ .

**Exemple 2** On considère un ensemble de 4 dispositifs identiques et indépendants, susceptibles de tomber en panne selon une loi exponentielle de « vie moyenne » égale à 200 jours, donc de paramètre  $\lambda = 0,005 \text{ jour}^{-1}$ . Quelle est la probabilité qu'aucun des dispositifs ne soit tombé en panne au bout de 90 jours ?

La probabilité qu'un dispositif tombe en panne *après* 90 jours est  $P(X > t) = 1 - P(X \leq t) = e^{-\lambda t}$  pour  $\lambda = 0,005$  et  $t = 90$ , soit  $e^{-0,45} = 0,6376$ . La probabilité que les 4 dispositifs tombent en panne après 90 jours est  $(P(X > 90))^4 = (e^{-0,45})^4 = e^{-1,8} = 0,165$ .



## factorielle

(factorial)

Produit des  $n$  premiers entiers :

$$n! = 1 \times 2 \times 3 \times \dots \times n$$

(prononcer : « factorielle ènne »)

Convention importante :

$$0! = 1$$

► Signification combinatoire

$n!$  représente le nombre de *permutations* d'un ensemble de  $n$  objets.

► Valeur approchée (*formule de Stirling*) :

- équivalent :  $n! \sim n^n e^{-n} \sqrt{2\pi n}$
- approximation à deux termes :  $n! \approx n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{12n}\right)$

## factoriels (moments)

Voir *moments factoriels*.

## fiabilité

(reliability)

De façon générale, la théorie de la fiabilité étudie les défaillances (ou pannes) qui peuvent affecter un système (ou dispositif) destiné à fonctionner. Le mot défaillance (ou panne) ne signifie pas arrêt complet du fonctionnement mais cessation du « bon » fonctionnement (*i.e.* du fonctionnement conforme au cahier des charges).

Une partie de la théorie de la fiabilité est consacrée à la modélisation probabiliste des défaillances, permettant notamment d'étudier les problèmes de sûreté (probabilités ponctuelles) et de qualité/rentabilité (espérances mathématiques).

Le concept précis de fiabilité se définit comme l'aptitude du dispositif étudié à accomplir la fonction requise dans des conditions données et de façon ininterrompue durant une période de temps donnée (selon les spécifications du « cahier des charges »).

Enfin, et de façon « technique », la fiabilité désigne la fonction, généralement notée  $R(t)$ , égale à la probabilité qu'un système  $S$  (considéré comme mis en service ou observé en bon fonctionnement au temps 0), ait fonctionné sans défaillance jusqu'au temps  $t$ . Cette fonction est directement liée aux éléments les plus fondamentaux du modèle probabiliste de la fiabilité.

Si on introduit la variable aléatoire réelle  $T$  = « durée de vie de  $S$  » = « temps d'attente de la (première) défaillance », de fonction de répartition  $F(t) = F(T \leq t)$ , et de densité de probabilité  $f(t)$ , on a  $F(t) = 1 - R(t)$  et  $f(t) = -R'(t)$ .

On peut ensuite définir le *taux de défaillance*  $\lambda(t)$  de S comme la « densité conditionnelle » :

$$\begin{aligned}\lambda(t) &= \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} P(\text{défaillance entre } t \text{ et } t + \delta t \mid \text{pas de défaillance sur } [0, t]) \\ &= \lim_{\delta t \rightarrow 0} \frac{1}{\delta t} P(t < T \leq t + \delta t \mid T \leq t)\end{aligned}$$

d'où l'on déduit la relation  $\lambda(t) = -\frac{R'(t)}{R(t)}$ .

## Fisher (Ronald)

Biologiste et mathématicien britannique (1890–1962). Développa les techniques d'estimation par la méthode du maximum de vraisemblance, et introduisit la méthode d'analyse de la variance.

## Fisher (test de)

(Fisher test)

Nom donné selon les auteurs à plusieurs tests paramétriques :

1. Le test « d'analyse de variance » d'une régression  $Y = \alpha + \beta x + E$ , qui compare à zéro la pente  $\beta$  de la droite de régression ; ce test est décrit ci-dessous.

2. Une version « exacte » du test du khi-deux à 4 cases où l'on introduit les probabilités multinomiales  $\frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$  (valeurs  $a, b, c, d$  dans les 4 cases,  $n = a + b + c + d$ ).

Si les effectifs sont très faibles, il n'existe qu'un petit nombre de configurations sous la condition que les sommes des lignes et des colonnes soient constantes, configurations dont on calcule explicitement les probabilités (il semble que dans le monde anglophone, *Fisher Test* désigne le plus souvent ce test).

3. Le test de Fisher–Snedecor du rapport des variances (voir *Fisher-Snedecor (test de)*).

### ———— test de Fisher d'analyse de la variance d'une régression linéaire ————

- Données. Un échantillon de  $n$  couples de valeurs observées  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  d'un couple  $(X, Y)$  de variables aléatoires numériques, modélisé par une régression  $Y = \alpha + \beta x + E$ .

- Hypothèse testée.  $H_0 = \langle \beta = 0 \rangle$  contre  $H_1 = \langle \beta \neq 0 \rangle$

- Déroulement technique du test

1. On calcule avec les formules usuelles les moyennes observées  $\bar{x}$  et  $\bar{y}$ , puis les estimations  $a$  et  $b$  des coefficients de la droite de régression.

2. On calcule les sommes de carrés :  $Q_1 = \sum_{i=1}^{i=n} ((a + bx_i) - \bar{y})^2$  et  $Q_2 = \sum_{i=1}^{i=n} ((a + bx_i) - y_i)^2$ .

3. On calcule la valeur observée de la variable de test :

$$F_{1, n-2} = \frac{Q_1}{\left(\frac{Q_2}{n-2}\right)}$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Fisher-Snedecor, elles dépendent des deux degrés de liberté 1 et  $n - 2$ , et du risque  $\alpha$ .

• Conditions et précautions

- En théorie (X, Y) doit suivre une loi normale à 2 dimensions, donc aucune précaution si on présume que c'est le cas ;
- Lorsque ce n'est pas le cas, le test est robuste et reste applicable si n est « assez grand », la condition  $n \geq 30$  est traditionnelle

Ce test est parfois dénommé *test de la significativité d'une régression*, expression calquée sur celle de test de significativité d'un coefficient de corrélation. De fait, il s'agit de deux présentations différentes du même test. La valeur de la variable de test de Fisher est exactement le carré  $\frac{r^2}{1-r^2}(n-2)$  de la variable de test pour le coefficient de corrélation, et la loi « du  $F_{1, n-2}$  » est exactement la loi du carré « du  $S_{n-2}$  ».

**Fisher (z-transformation de)** (Fisher z-transformation)

Voir *significativité d'un coefficient de corrélation (test de)*.

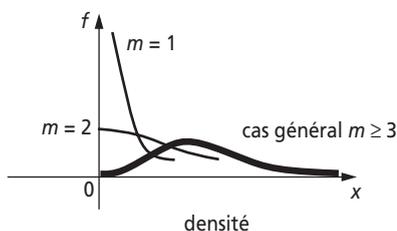
**Fisher-Snedecor (loi du F de)** (Fisher z distribution)

Loi d'une variable aléatoire continue utilisée pour le contrôle des tests de comparaison de deux variances ainsi que dans le test d'« analyse de la variance » qui permet de comparer plusieurs espérances mathématiques.

**Formulaire**

Deux paramètres entiers  $m, n \geq 1$  qui représentent des « degrés de liberté » ; valeurs sur les réels positifs.

► Loi de probabilité



fonction de répartition

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}$$

$$F(x) = \int_0^x f(t)dt \quad (x \geq 0)$$

► Valeurs caractéristiques

- espérance :  $E(X) = \frac{n}{n-2}$  (si  $n \geq 3$ )
- variance :  $\text{Var}(X) = \left(\frac{n}{n-2}\right)^2 \frac{2(m+n-2)}{m(n-4)}$  (si  $n \geq 5$ )
- écart-type :  $\sigma(X) = \frac{n}{n-2} \sqrt{\frac{2(m+n-2)}{m(n-4)}}$  (si  $n \geq 5$ )

Techniquement, la loi du F de Fisher-Snedecor est une loi bêta de type II (avec homothétie de la variable).

### ► Utilisations

En théorie, la loi de Fisher -Snedecor est la loi du quotient normalisé de deux v.a. khi-deux à

$m$  et  $n$  degrés de liberté :  $F(m, n) = \frac{\left(\frac{\chi_m^2}{m}\right)}{\left(\frac{\chi_n^2}{n}\right)}$ . Cette loi vérifie dans un certain nombre de cas

particuliers des relations intéressantes : si  $X$  suit une loi bêta de type I de paramètres  $r$  et  $s$ , alors  $\frac{s-X}{r(1-X)}$  est un  $F(2r, 2s)$  ; si  $Y$  suit une loi bêta de type II de paramètres  $r$  et  $s$ , alors  $\frac{sY}{r}$  est un  $F(2r, 2s)$ .

Dans la pratique, cette loi est utilisée dans de très nombreux tests d'hypothèses, notamment dans le test de comparaison de deux variances (de façon directe, en effectuant le quotient des deux estimations débiaisées), et dans le test dit d'« analyse de la variance », qui permet de comparer globalement plusieurs espérances mathématiques entre elles (de façon indirecte, en décomposant la variance totale de l'ensemble des observations en deux variances « partielles » puis en testant leur quotient).

On trouvera dans les ouvrages spécialisés les tables étendues du  $F$  de Fisher-Snedecor qui nécessitent plusieurs pages imprimées serrées. En effet, d'une part elles doivent prendre en compte trois paramètres, la probabilité niveau du test et les deux degrés de liberté, d'autre part les plages de valeurs relatives aux deux degrés de liberté sont très étendues et il n'y a pas de loi limite.

## Fisher-Snedecor (test de)

(Fisher-Snedecor test)

Synonymes : test  $F$  ( $F$  test), test du rapport des variances (*variance ratio test*).

Test paramétrique qui compare les variances observées de deux échantillons statistiques (ce test détient le record des appellations différentes : il est aussi appelé *test de Snedecor-Fisher* ou *test de Fisher* !).

————— test bilatéral de comparaison de deux variances  $\sigma_X^2$  et  $\sigma_Y^2$  —————

### • Données

Deux séries :

- un échantillon  $(x_1, x_2, \dots, x_{n_X})$  de  $n_X$  valeurs observées d'une variable aléatoire numérique  $X$  d'espérance mathématique  $\mu_X$  et de variance  $\sigma_X^2$  ;
- un échantillon  $(y_1, y_2, \dots, y_{n_Y})$  de  $n_Y$  valeurs observées d'une variable aléatoire numérique  $Y$  d'espérance mathématique  $\mu_Y$  et de variance  $\sigma_Y^2$ .

• Hypothèse testée.  $H_0 = \langle \sigma_X^2 = \sigma_Y^2 \rangle$  contre  $H_1 = \langle \sigma_X^2 \neq \sigma_Y^2 \rangle$ .

• Déroulement technique du test

1. On calcule les moyennes observées  $m_X$  et  $m_Y$  des deux échantillons.
2. On calcule les variances *non biaisées*  $s_X^2$  et  $s_Y^2$  des deux échantillons.
3. On permute éventuellement les notations  $X$  et  $Y$  pour que  $s_X^2$  soit *la plus grande* des deux variances observées, et on calcule la valeur observée de la variable de test :

$$F_{n_X-1, n_Y-1} = \frac{\left(\frac{n_X s_X^2}{n_X - 1}\right)}{\left(\frac{n_Y s_Y^2}{n_Y - 1}\right)}$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Fisher-Snedecor, elles dépendent des deux degrés de liberté  $n_X - 1$  et  $n_Y - 1$ , et du risque  $\alpha$ . Comme les tables de la loi de Fisher-Snedecor sont construites pour leur utilisation directe dans le test d'analyse de la variance, elles donnent une probabilité unilatérale de dépassement ; dans le cas présent, l'artifice du quotient de la plus grande variance observée par la plus petite doit faire lire la valeur critique de la variable de test pour la probabilité de dépassement  $\frac{\alpha}{2}$ .

• Conditions et précautions

En théorie  $X$  et  $Y$  doivent être des v.a. normales ; lorsque ce n'est pas le cas, le test est robuste et reste applicable si les effectifs  $n_X$  et  $n_Y$  sont « assez grands ».

## fluctuation d'échantillonnage (sampling fluctuation)

Cette expression, employée surtout dans le contexte des tests d'hypothèse, désigne une variation autour de la valeur moyenne (ou « théorique »), lorsque son amplitude est suffisamment limitée pour qu'il soit probable qu'elle provienne de la dispersion inhérente à tout échantillonnage (et non pas d'un décalage de la valeur moyenne). Affirmer qu'un écart est une fluctuation d'échantillonnage est bien entendu un *pari*, qui peut être assorti d'une probabilité d'erreur, d'autant plus élevée que l'écart est grand.

## fonction

Voir *caractéristique (fonction)*, *génératrice (fonction)*, *génératrice des moments (fonction)*, *répartition (fonction de)*.

## fondamental (espace)

Voir *espace fondamental*.

## formule

Voit *Bayes (formule de)*, *Poincaré (formule de)*, *sommation (formules de)*, *totales (formule des probabilités)*.

## formule de Huygens-König

► Dans le cas de la variance

En probabilités

$$\text{Var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2.$$

En statistique (formule pour  $n$  observations *individualisées*  $x_1, x_2, \dots, x_n$ )

$$s^2 = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - \bar{x})^2 = \left( \frac{1}{n} \sum_{i=1}^{i=n} x_i^2 \right) - \left( \frac{1}{n} \sum_{i=1}^{i=n} x_i \right)^2$$

La signification de cette formule est la suivante. Sous la forme  $E(X^2) = (E(X))^2 + \text{Var}(X)$ , elle est l'analogie du théorème de König pour l'inertie en mécanique :  $E(X^2)$  est la somme de deux termes, *primo*  $(E(X))^2$  qui dépend de la distance entre l'origine et la valeur centrale de la distribution, et *secundo*  $\text{Var}(X)$  qui exprime la dispersion intrinsèque de la distribution (laquelle ne dépend pas de l'emplacement de l'origine).

L'utilisation de cette formule sous la forme  $\text{Var}(X) = E(X^2) - (E(X))^2$  (ou son équivalent statistique) est une utilisation légitime si l'on effectue des calculs numériques exacts (ou tout au moins avec une approximation suffisante). C'est en revanche une utilisation très dangereuse si l'on effectue des calculs numériques avec une approximation inadaptée. Elle ne doit donc être utilisée qu'avec prudence, sans compter (!) que les calculettes lui ont fait perdre beaucoup de son intérêt pratique.

► Dans le cas de la covariance

Formule en probabilités :

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y).$$

Formule en statistique (pour  $n$  couples d'observations individualisées  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ ) :

$$\text{Cov} = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - \bar{x})(y_i - \bar{y}) = \left( \frac{1}{n} \sum_{i=1}^{i=n} x_i y_i \right) - \left( \frac{1}{n} \sum_{i=1}^{i=n} x_i \right) \left( \frac{1}{n} \sum_{i=1}^{i=n} y_i \right)$$

C'est l'extension de la formule de même nom pour la variance. Elle appelle le même commentaire.

## formule du binôme

Voir *binômiaux (coefficients)*.

## fourchette

Nom souvent donné à l'*intervalle de confiance* dans la présentation de résultats au public. Il est malheureusement trop fréquent que l'on omette de préciser la taille de l'échantillon et le seuil de confiance.

Voir *estimation par intervalle*.

## fréquence

(frequency)

On se donne une épreuve et un évènement A (éventuellement défini par une valeur d'une variable aléatoire ou d'un caractère). Si l'on répète  $n$  fois l'épreuve, et si l'évènement A se réalise  $k$  fois, on définit la fréquence de A comme le quotient  $\frac{k}{n}$ .

À la base du calcul des probabilités, le concept de probabilité *modélise* la notion de fréquence. Ensuite, la « manière » dont la fréquence (expérimentale) converge vers la probabilité (théorique) lorsque le nombre de répétitions de l'épreuve tend vers l'infini fait l'objet d'un théorème essentiel du calcul des probabilités, la *loi des grands nombres*, qui sert de « pont » vers la statistique mathématique.

Dans un contexte statistique, on emploie indifféremment fréquence ou fréquence observée ou fréquence empirique.



## Galilée

Mécanicien, physicien et astronome italien (1564–1642). Il aborda les questions d'erreurs de mesure d'un point de vue qui préfigure le calcul des probabilités. Initiateur de la mathématisation de la physique, il fit des travaux importants en mécanique. Auteur des premières observations astronomiques avec une lunette, il révolutionna la conception du cosmos.

## Galton (Francis)

Généticien et voyageur britannique (1822–1911). Il développa et utilisa les méthodes statistiques en anthropologie et définit la régression. Inventeur en outre du sac de couchage et du terme anticyclone.

## Galton-Watson (processus de)

(*Galton-Watson process*)

Modèle probabiliste de l'évolution d'une population dont les individus se reproduisent par générations distinctes, chaque individu se reproduisant indépendamment des autres, et son nombre de descendants suivant une loi qui ne dépend ni de l'individu ni de la génération.

Un tel modèle est un cas particulier de *chaîne de Markov* (homogène), elle-même cas particulier d'un *processus stochastique*. Il est complètement caractérisé par l'ensemble des probabilités  $\{p_k\}$  qu'un individu de la population à la génération  $n$  ait  $k$  descendants à la génération  $n + 1$ . Historiquement introduit pour l'étude de l'extinction des noms de famille (on parle parfois de *processus généalogique* ou encore de *processus en cascade*), ce modèle s'applique notamment à des populations de gènes, de bactéries, ...

On peut démontrer que presque sûrement 1 un processus de branchement « s'éteint » ou « explose », sans possibilité d'oscillations infinies ni de convergence vers une limite. Si  $X$  est la v.a. nombre de descendants d'un individu et si  $G(s) = E(e^{sX})$  est la *fonction génératrice* de  $X$ , on peut démontrer le résultat suivant : si  $G'(1) = E(X) \leq 1$ , la probabilité d'extinction est égale à 1, si  $G'(1) > 1$ , la probabilité d'extinction est égale à la plus petite racine positive de l'équation  $G(s) = s$ .

Synonyme de *processus de branchement*.

## gamma (fonction)

(*gamma function*)

Fonction définie par une intégrale et qui fournit notamment la valeur des moments (espérance mathématique, variance, ...) de plusieurs lois de probabilité. Elle possède également un intérêt puissant dans d'autres domaines des mathématiques.

### ► Définition

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (\alpha > 0).$$

Pour tout  $\alpha$  réel positif, on a la récurrence  $\alpha \Gamma(\alpha) = \Gamma(\alpha + 1)$ .

Pour tout entier  $n$  positif, on a  $\Gamma(n) = (n - 1) !$

► Quelques valeurs numériques

$$\Gamma(0) = +\infty \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} = 1,773$$

$$\Gamma(1) = 1 = 0! \qquad \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2} = 0,886$$

$$\Gamma(2) = 1 = 1! \qquad \Gamma\left(\frac{5}{2}\right) = \frac{3\sqrt{\pi}}{4} = 1,329$$

$$\Gamma(3) = 2 = 2! \qquad \Gamma\left(\frac{7}{2}\right) = \frac{15\sqrt{\pi}}{8} = 3,323$$

$$\Gamma(4) = 6 = 3! \qquad \text{etc.}$$

► Utilisation en vue des moments des lois de probabilité

$$\int_0^{\infty} t^k e^{-t} dt = \Gamma(k+1) = k!$$

$$2 \int_0^{\infty} t^k e^{-t^2/2} dt = 2^{\frac{k+1}{2}} \Gamma\left(\frac{k+1}{2}\right) \quad (\text{attention aux bornes d'intégration})$$

## gamma (loi)

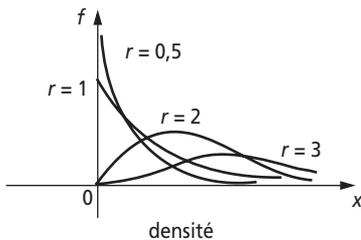
(gamma distribution)

Loi d'une variable aléatoire continue qui possède de nombreuses applications.

### Formulaire

Deux paramètres réels :  $r \in \mathbf{R}_+^*$  (paramètre « de forme ») ;  $\lambda \in \mathbf{R}_+^*$  (paramètre d'échelle). Valeurs sur les réels positifs.

► Loi de probabilité



fonction de répartition

$$f(x) = \frac{\lambda}{\Gamma(r)} e^{-\lambda x} (\lambda x)^{r-1} \quad (x > 0)$$

$$F(x) = \int_0^x f(t) dt \quad (x \geq 0)$$

► Valeurs caractéristiques

– espérance :  $E(X) = \frac{r}{\lambda}$

– variance :  $\text{Var}(X) = \frac{r}{\lambda^2}$

– écart-type :  $\sigma(X) = \frac{\sqrt{r}}{\lambda}$

► Cas particulier important

Lorsque  $r$  est entier, la loi gamma (souvent appelée alors *loi d'Erlang*), est la loi de la somme de  $r$  variables aléatoires exponentielles de paramètre  $\lambda$  et indépendantes. (pour  $r = 1$ , la loi gamma est la loi exponentielle elle-même)

► Loi gamma et loi du khi-deux

La loi gamma de paramètres  $\frac{1}{2}$  et  $\frac{1}{2}$  est la loi du carré d'une v.a. normale centrée réduite  $\mathcal{N}(0, 1)$ , et de façon générale, si  $X$  suit une loi gamma de paramètres  $r$  et 1,  $2X$  suit une loi du khi-deux de paramètre  $2r$ .

► Utilisation

Outre ses applications pour  $r$  entier (cf. loi d'Erlang) – la loi gamma est assez souvent utilisée en théorie de la fiabilité, avec le paramètre  $r$  supérieur à 1, pour modéliser la durée de vie d'un dispositif qui « vieillit », donc avec un « taux de défaillance » non pas constant (cas  $r = 1$  : loi exponentielle) mais qui augmente au cours du temps.

**Théorème d'addition.** La somme de deux variables aléatoires gamma indépendantes  $\Gamma(r_1, \lambda)$  et  $\Gamma(r_2, \lambda)$  est une variable aléatoire normale gamma  $\Gamma(r_1 + r_2, \lambda)$  (dans le cas où  $r_1$  et  $r_2$  sont entiers, ce théorème est bien entendu cohérent avec l'interprétation de la loi gamma alias loi d'Erlang comme somme de v.a. exponentielles).

## Gauss (Carl–Friedrich)

Mathématicien, physicien et astronome allemand (1777–1855). Il introduisit la méthode des moindres carrés dans la théorie des erreurs d'observation et fit des travaux importants en théorie des nombres, en algèbre, en géométrie, en optique et en théorie du magnétisme.

## Gauss (loi de)

Voir *normale (loi)*.

## gausso-arithmétique (papier)

Papier gradué avec une échelle « gaussienne » en abscisse et une échelle arithmétique en ordonnée, qui permet de tracer une *droite de Henry*.

## génératrice (fonction)

(*generating function*)

Série de puissances que l'on peut associer aux variables aléatoires à valeurs entières positives et qui est en relation directe avec les moments factoriels.

Soit  $X$  une variable aléatoire à valeurs entières positives. On appelle *fonction génératrice* de  $X$  la fonction  $G_X$  de la variable réelle  $s$  définie par :

$$G_X(s) = E(s^X) = \sum_{k=0}^{\infty} s^k P(X = k)$$

Si la v.a.  $X$  admet des *moments factoriels* de tous les ordres, la fonction  $G_X$  est développable en série entière au voisinage de 1 et l'on a :

$$G_X(1 + u) = 1 + E(X) \frac{u}{1!} + E(X(X-1)) \frac{u^2}{2!} + \dots + E(X(X-1)\dots(X-k+1)) \frac{u^k}{k!} + \dots$$

ce qui signifie en particulier que, pour tout  $k \geq 1$ , on a  $E(X(X-1)\dots(X-k+1)) = G_X^{(k)}(1)$  (relation vraie même si tous les moments factoriels n'existent pas dès lors que celui d'ordre  $k$  existe et la dérivée de même).

Comportement par addition de v.a. indépendantes :

$$G_{X+Y}(s) = G_X(s) G_Y(s).$$

### **génératrice des moments (fonction)** (*moment generating function*)

Fonction réelle de variable réelle que l'on peut associer à certaines variables aléatoires réelles et qui est en relation directe avec les moments.

Soit  $X$  une variable aléatoire réelle. On appelle *fonction génératrice* des moments de  $X$  la fonction  $g_X$  de la variable réelle  $u$  définie par

$$g_X(u) = E(e^{uX})$$

sur l'ensemble des valeurs de  $u$  pour lesquelles cette espérance mathématique existe.

– Expression dans le cas discret :  $X$  prend les valeurs  $x_k$  avec les probabilités  $p_k$  :

$$g_X(t) = \sum_k p_k e^{ux_k}$$

– Expression dans le cas absolument continu :  $X$  possède la densité  $f(x)$  :

$$g_X(t) = \int_{-\infty}^{\infty} e^{ux} f(x) dx$$

Lorsque la v.a. est bornée, la fonction génératrice des moments existe et est continue pour tout  $u$ .

Si la v.a.  $X$  admet des moments de tous les ordres, la fonction  $g_X$  est développable en série entière au voisinage de 0 et l'on a :

$$g_X(u) = 1 + E(X) \frac{u}{1!} + E(X^2) \frac{u^2}{2!} + \dots + E(X^k) \frac{u^k}{k!} + \dots$$

ce qui signifie en particulier que, pour tout  $k \geq 1$ , on a  $E(X^k) = G_X^{(k)}(0)$ .

Comportement par transformation affine :

$$g_{aX+b}(t) = e^{bt} g_X(at).$$

Comportement par addition de v.a. indépendantes :

$$g_{X+Y}(t) = g_X(t) g_Y(t).$$

Malgré la relation directe de la fonction génératrice avec les moments, les mathématiciens préfèrent travailler avec la *fonction caractéristique*  $E(e^{itX})$ , qui peut être définie pour toute variable aléatoire  $X$ , bijectivement, et qui existe pour toute valeur de  $t$ .

### **géométrique (loi)**

(*geometric distribution*)

Loi d'une variable aléatoire discrète « temps d'attente » du premier succès dans des épreuves répétées.

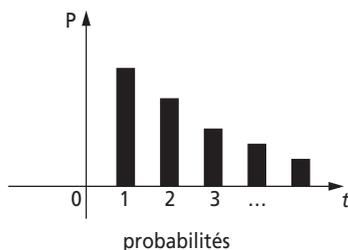
#### Formulaire

Un paramètre réel  $p$  ( $0 \leq p \leq 1$ ) qui représente une probabilité (notation standard :  $q = 1 - p$ ).

Soit  $T$  la variable géométrique de paramètre  $p$  ; valeurs prises : 1, 2, 3, ...

► Loi de probabilité

$P(T = k) = pq^{k-1}$  (une probabilité commode dans les applications est  $P(T > k) = q^k$ )



► Valeurs caractéristiques

- espérance :  $E(T) = \frac{1}{p}$
- variance :  $\text{Var}(T) = \frac{q}{p^2}$
- écart-type :  $\sigma(T) = \frac{\sqrt{q}}{p}$

► Utilisation

La loi géométrique est le temps d'attente (alias le rang) du premier succès dans des épreuves répétées, ou dans des tirages « AVEC remise ».

Attention ! le premier temps possible est  $T = 1$ . Dans certaines applications, le premier temps possible est  $T' = 0$  : on a alors  $P(T' = k) = pq^k$  et  $E(T') = \frac{q}{p}$  (la variance et l'écart-type restent les mêmes).

$p$  représente une probabilité conditionnelle de succès :

$$p = P(T = k \mid T \geq k)$$

On parle parfois de processus « poissonnien discret » et on qualifie parfois la loi géométrique de loi « exponentielle discrète ».

**Exemple 1** Si l'on joue à Pile ou Face ( $p = q = \frac{1}{2}$ ) et si l'on décide de s'arrêter après le premier Pile obtenu, l'espérance du temps d'attente est égale à 2.

**Exemple 2** Si on lance 3 dés et que l'on recommence jusqu'à ce que l'on obtienne 3 As ( $p = \frac{1}{216}$ ,  $q = \frac{211}{216}$ , la probabilité d'insuccès à la 100-ième fois est  $P(T > 100) = \left(\frac{211}{216}\right)^{100} = 0,99537^{100} \approx 0,629$ , la probabilité d'insuccès à la 1 000-ième fois est  $P(T > 1\,000) = \left(\frac{211}{216}\right)^{1000} = 0,99537^{1\,000} \approx 0,0097$ .

Voir Pascal (loi de).

## géométrie (généralisation de la loi) : loi « sans nom »

Loi d'une variable aléatoire discrète « temps d'attente » du premier succès dans des tirages « SANS remise ».

### Formulaire

Deux paramètres entiers  $N, K \leq N$

Les notations standard introduisent  $p = \frac{K}{N}$  ( $0 \leq p \leq 1$ ) et  $q = 1 - p$ .

Soit T la variable aléatoire « sans nom » ; valeurs prises : 1, 2, ...,  $N - K + 1$ .

#### ► Loi de probabilité

$$P(T = k) = \frac{K}{k} \cdot \frac{\binom{N-K}{k-1}}{\binom{N}{k}}$$

#### ► Valeurs caractéristiques

- espérance :  $E(T) = \frac{N+1}{K+1} = \frac{1}{p} \cdot \frac{1 + \frac{1}{N}}{1 + \frac{1}{K}}$
- variance :  $\text{Var}(T) = \left(\frac{N}{K+1}\right)^2 \frac{(N+1)pq}{K+2}$
- écart-type :  $\sigma(T) = \frac{N}{K+1} \sqrt{\frac{(N+1)pq}{K+2}}$

## Gini (indice de concentration de)

Voir *concentration (indice de – de Gini)*.

## Gosset (William)

Brasseur et mathématicien britannique (1876–1937). Il étudia, sous le pseudonyme de « Student », les techniques statistiques.

## grands nombres (loi des)

Historiquement, la « loi des grands nombres » désigne le constat fait par les premiers probabilistes (Pascal, Fermat, Huygens) de la convergence de la fréquence d'un événement vers sa probabilité lorsque le nombre d'épreuves (indépendantes) augmente indéfiniment. Cette convergence était alors perçue comme une loi de la nature – d'où le mot loi – et ce n'est que plus tard, lorsqu'on a commencé à concevoir le calcul des probabilités comme un modèle mathématique des phénomènes aléatoires, que l'on a pris conscience qu'il s'agissait en réalité d'un vrai théorème de mathématiques, parfaitement démontrable (Bernoulli, De Moivre). L'usage a conservé l'ancienne appellation.

**grands nombres (loi faible)***(weak law of large numbers)*

**Théorème.** On considère une suite  $(X_n)$  de variables aléatoires (réelles) indépendantes et identiquement distribuées, d'espérance mathématique  $\mu$ . On définit les moyennes :

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Alors, pour tout  $\varepsilon > 0$ , on a :

$$P(|X_n - \mu| > \varepsilon) \rightarrow 0$$

$n \rightarrow \infty$

C'est très exactement la définition de la convergence en probabilité de la suite  $(X_n)$  vers (la v.a certaine)  $\mu$  lorsque  $n$  tend vers l'infini.

La première démonstration a été donnée par Bernoulli dans le cas particulier où les  $X_n$  sont des v.a. de Bernoulli (que lui-même n'appelait pas ainsi !), et la somme une v.a. binomiale : si la v.a. de Bernoulli est l'indicatrice d'un évènement, il s'agit alors de la convergence de la fréquence vers la probabilité de l'évènement. Plusieurs versions de la loi faible des grands nombres avec des conditions affaiblies ont été démontrées au XX<sup>e</sup> siècle (par exemple en ne supposant pas les  $X_n$  indépendantes mais seulement leurs covariances 2 à 2 nulles, ou bien en abandonnant l'exigence de loi identique).

**grands nombres (loi forte)***(strong law of large numbers)*

**Théorème.** On considère une suite  $(X_n)$  de variables aléatoires (réelles) indépendantes et identiquement distribuées, d'espérance mathématique  $\mu$ . On définit les moyennes :

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Alors la probabilité de l'évènement  $X_n \rightarrow \mu$  est égale à 1.

$n \rightarrow \infty$

C'est très exactement la définition de la convergence presque sûre de la suite  $(X_n)$  vers (la v.a certaine)  $\mu$  lorsque  $n$  tend vers l'infini. Ce résultat est très technique et est essentiellement utilisé par les mathématiciens professionnels.

Comme pour la loi faible, plusieurs versions de la loi forte des grands nombres avec des conditions affaiblies ont été démontrées au XX<sup>e</sup> siècle.

**grec (alphabet)**

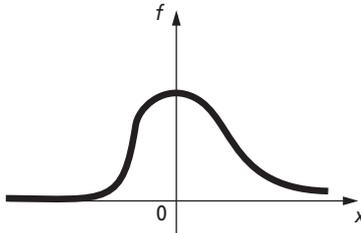
alpha	$\alpha$	A	iota	$\iota$	I	rhô	$\rho$	P
bêta	$\beta$	B	kappa	$\kappa$	K	sigma	$\sigma, \varsigma$	$\Sigma$
gamma	$\gamma$	$\Gamma$	lambda	$\lambda$	$\Lambda$	tau	$\tau$	T
delta	$\delta$	$\Delta$	mu	$\mu$	M	upsilon	$\upsilon$	Y
epsilon	$\varepsilon$	E	nu	$\nu$	N	phi	$\varphi$	$\Phi$
dzêta	$\zeta$	Z	xi	$\xi$	$\Xi$	khi	$\chi$	X
êta	$\eta$	H	omicron	$o$	O	psi	$\psi$	$\Psi$
thêta	$\theta$	$\Theta$	pi	$\pi$	$\Pi$	oméga	$\omega$	$\Omega$

**Gumbel (loi de)***(Gumbel distribution)*

Loi d'une variable aléatoire continue utilisée pour la distribution des valeurs extrêmes.

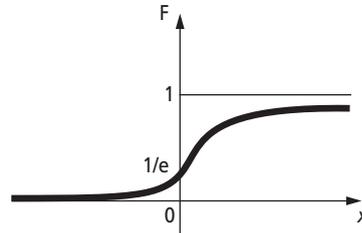
**Formulaire**Version standardisée  $X$  sans paramètre, à valeurs sur  $\mathbf{R}$ .

## ► Loi de probabilité



densité

$$f(x) = e^{-x} \exp(-e^{-x}) = \exp(-x - e^{-x})$$



fonction de répartition

$$F(x) = \exp(-e^{-x})$$

## ► Valeurs caractéristiques

- espérance :  $E(X) = \gamma = 0,5772\dots$  (« constante d'Euler »)
- variance :  $\text{Var}(X) = \frac{\pi}{\sqrt{6}}$
- écart-type :  $\sigma(X) = \frac{\pi^2}{6}$

## ► Utilisation

En théorie, la loi de Gumbel est la loi limite, lorsque  $n$  tend vers l'infini, du maximum de  $n$  v.a. identiques et indépendantes, dans le cas où la queue de distribution est de type exponentiel.

Soit  $X$  une v.a. et  $F(x) = P(X \leq x)$  sa fonction de répartition. On dit que « la queue de distribution est de type exponentiel » si, lorsque  $x$  tend vers l'infini,  $1 - F(x)$  tend vers 0 au moins aussi vite qu'une fonction exponentielle  $e^{-kx}$  (c'est le cas notamment de la loi normale et la loi exponentielle).

La loi de Gumbel utilisée comme approximation de la loi du maximum doit bien sûr être

dénormalisée en introduisant des paramètres de position et d'échelle :  $F(x) = \exp\left(-e^{-\frac{x-a}{b}}\right)$ ,

$f(x) = F'(x)$ . Sous cette forme générale, la loi de Gumbel est parfois appelée *loi de Fisher-Tippett* ou *loi log-Weibull* par les Anglo-Saxons.

Cas particulier : si les v.a. identiques et indépendantes sont des v.a. normales centrées, leur maximum suit approximativement, pour  $n$  grand, une loi de Gumbel de paramètres

$$a = \sqrt{2 \ln n} \quad \text{et} \quad b = \frac{1}{\sqrt{2 \ln n}}.$$

Dans la pratique, la loi de Gumbel est notamment utilisée en hydrologie et en climatologie pour qualifier les valeurs extrêmes.

**Exemple** On considère la variable  $X$  = moyenne sur le mois d'août des températures journalières maximales, dans une ville donnée. On suppose que  $X$  suit approximativement une loi normale d'espérance  $\mu$  et d'écart-type  $\sigma$  (valeurs typiques dans une région tempérée :  $\mu = 26$  °C,  $\sigma = 2$  °C). Quelle est la valeur que le maximum de  $X$  sur 50 années consécutives risque de dépasser avec une probabilité 0,5 ? Il est sous-entendu que les valeurs de  $X$  sur les années qui se succèdent sont indépendantes.

Si on raisonnait sur une variable normale réduite, il faudrait résoudre  $F(x) = \exp\left(-e^{\frac{a-x}{b}}\right) = 0,5$

avec  $a = \sqrt{2 \ln 50} = 2,80$  et  $b = \frac{1}{\sqrt{2 \ln 50}} = 0,358$ . La solution est  $x = 2,93$ . La réponse à la question posée est donc  $\mu + 2,93 \sigma$  (soit 31,9 °C avec les valeurs typiques données).



## Hardy-Weinberg (loi de)

Théorème qui exprime les proportions des génotypes dans le modèle probabiliste de transmission aléatoire des gènes.

On exprime usuellement la loi de Hardy-Weinberg en ne considérant que deux allèles, notés par exemple A et a (la généralisation ne soulève pas de difficulté conceptuelle). Les génotypes sont donc AA, Aa et aa. On représente les distributions de probabilité par des expressions symboliques :

– la distribution génotypique est représentée par  $uAA + 2vAa + waa$ , où :

$$\begin{cases} u \text{ est la fréquence du génotype AA} \\ 2v \text{ est la fréquence du génotype Aa} \\ w \text{ est la fréquence du génotype aa} \end{cases}$$

(on a bien sûr  $u + 2v + w = 1$ ).

– la distribution génique est représentée par  $pA + qa$ , où :

$$\begin{cases} p = u + v & \text{est la fréquence de l'allèle A} \\ q = v + w & \text{est la fréquence de l'allèle a} \end{cases}$$

(on a bien sûr  $p + q = 1$ ).

Ces expressions symboliques seront indicées par  $n$  pour représenter les distributions de la  $n$ -ième génération.

La validité de la loi de Hardy-Weinberg dépend de nombreuses conditions : absence de migrations, absence de mutations, existence de générations distinctes et sans croisements entre elles (c'est une condition qui peut paraître contraignante, mais on peut montrer qu'elle s'évanouit dès que la population est en « équilibre génotypique »), absence de sélection, panmixie, et enfin « grand » effectif de la population, qui permet d'assimiler probabilités individuelles théoriques et fréquences réelles.

**Théorème** (loi de Hardy-Weinberg). Sous les conditions exprimées ci-dessus,

– d'une part la distribution génique est constante :

$$\text{pour tout } n : p_n A + q_n a = p_0 A + q_0 a,$$

– d'autre part la distribution génotypique est stable à partir de la première génération de descendants :

$$\text{pour tout } n \geq 1 : u_n AA + 2v_n Aa + w_n aa = (p_0 A + q_0 a)^2,$$

– et les fréquences génotypiques ne sont pas quelconques mais liées par la relation :

$$u_n w_n = v_n^2.$$

Cette dernière relation, dite « d'équilibre panmictique », permet notamment, en cas d'allèles dominant/récessif, de calculer les fréquences génotypiques alors qu'on ne connaît que les fréquences phénotypiques.

## hasard

(*randomness*)

Terme que l'on rencontre, soit dans la langue courante (« le hasard ») pour désigner le *deus ex machina* des phénomènes aléatoires ou pour désigner notre impuissance à prévoir ou à contrôler, soit dans des locutions du calcul des probabilités (« tirer au hasard », « choisir au hasard »). Dans ce deuxième type d'usage, le sens est, quoiqu'implicite, très précis et renvoie à des critères techniques d'« honnêteté » du process, par exemple équiprobabilité ou indépendance.

Voir *aléatoire, stochastique*.

## Henry (droite de)

Voir *droite de Henry*.

## histogramme

(*histogram*)

Un histogramme est une représentation graphique très « parlante » visuellement, que l'on peut utiliser aussi bien en calcul des probabilités qu'en statistique. Sa figuration standard, aussi bien pour une variable discrète que pour une variable continue, est celle d'un diagramme en tuyaux d'orgue, avec des rectangles contigus.

Soit une distribution statistique réelle, rangée dans un nombre fini de classes. On donne la suite croissante des extrémités des classes  $x_0 < x_1 < \dots < x_k$ , avec pour chaque classe  $[x_{i-1}, x_i[$  sa fréquence  $f_i$ .

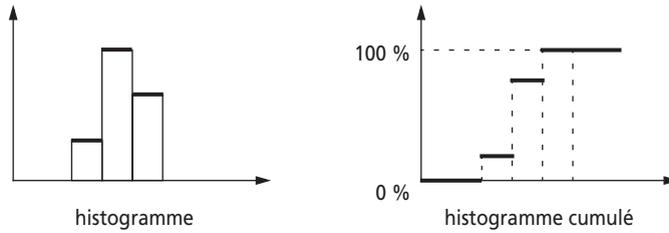
On appelle histogramme la représentation graphique imagée qui, pour chaque classe  $[x_{i-1}, x_i[$ , dessine un rectangle d'aire proportionnelle à sa fréquence  $f_i$ .

On obtient le même résultat visuel en prenant les effectifs au lieu des fréquences (seule change la graduation des ordonnées). On peut étendre cette définition au cas probabiliste, en remplaçant les fréquences par les probabilités, cela ne soulève aucune difficulté.

Le point important, et qui différencie l'histogramme d'un *diagramme en barres*, est la proportionnalité de l'aire (*i.e.* de la surface) des rectangles – et non pas de leur hauteur – aux fréquences ou aux effectifs. Bien entendu, le résultat est le même dans le cas où toutes les classes sont de largeur égale (ce qui est souhaitable, mais pas toujours possible). Cette proportionnalité de l'aire à la fréquence fait de l'histogramme un concept voisin de la courbe qui représente la densité.

Lorsqu'un histogramme est dessiné pour une distribution continue (ou discrète à valeurs nombreuses regroupées), on marque généralement les extrémités des classes sur l'axe des abscisses. Lorsqu'un histogramme est dessiné pour une distribution discrète, il est très souhaitable de marquer la valeur ponctuelle de chaque classe au milieu de la base du rectangle qui la représente.

La détermination du nombre de classes approprié d'un histogramme est délicate et il n'existe pas de règle universelle (il faut faire des essais...). Un nombre trop faible de classes fait perdre de l'information et atténue ou fait disparaître les particularités caractéristiques de la distribution. À l'inverse, un nombre trop grand de classes conduit à des graphiques chaotiques dont les irrégularités ne sont pas significatives.



On peut construire des *histogrammes cumulés*, en utilisant pour chaque classe la fréquence ou l'effectif cumulé. Le graphique a un aspect « en escalier » et le concept est voisin de la courbe qui représente la fonction de répartition.

Voir *pyramide des âges*.

## homogénéité (test du khi-deux d')

Voir *khi-deux d'homogénéité (test du)*.

## homoscédasticité

(*homoscedasticity*)

Propriété relative à un couple  $(X, Y)$  de variables aléatoires numériques, qui énonce que la variance conditionnelle  $\sigma^2(Y|X = x)$  est constante (ne dépend pas de  $x$ ). Le contraire s'appelle *hétérosédasticité*.

Voir *Bartlett (test de)*, *droite de régression*.

## Huygens (Christiaan)

Mathématicien, physicien et astronome hollandais (1629–1695). Il écrit *De Raticiniis in Ludo Aleae* qui constitue le premier traité de Calcul des probabilités, définit l'espérance mathématique et calcula avec son frère la première table de mortalité. Il fit des travaux importants en mécanique, en optique et en géométrie.

## Huygens-König (formule de)

Voir *formule de Huygens-König*.

## hypergéométrique (loi)

(*hypergeometric distribution*)

Loi d'une variable aléatoire discrète de « compte » qui intervient dans les tirages « SANS remise ».

### Formulaire

Trois paramètres entiers :  $N$ ,  $K \leq N$  et  $n \leq N$ , ce troisième paramètre représentant le nombre d'épreuves ou de tirages. Les notations standard introduisent  $p = \frac{K}{N}$  ( $0 \leq p \leq 1$ )

et  $q = 1 - p$ .

Soit  $X$  la variable aléatoire hypergéométrique ; valeurs prises :  $0, 1, \dots, n$ . La valeur prise  $k$  doit satisfaire les contraintes  $k \leq N$  et  $n - k \leq N - K$ . (évidentes d'après l'interprétation, cf. ci-dessous).

➤ Loi de probabilité

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

► Valeurs caractéristiques

- espérance :  $E(X) = np$
- variance :  $\text{Var}(X) = npq \frac{N-n}{N-1}$
- écart-type :  $\sigma(X) = \sqrt{npq} \sqrt{\frac{N-n}{N-1}}$

► Utilisations

La loi hypergéométrique est la loi du compte d'un caractère dans des tirages « SANS remise » : il y a initialement  $N$  objets dont  $K$  possèdent « le » caractère : on en tire  $n$ , et le caractère est observé  $k$  fois.

Dans de très nombreuses situations réelles (sondages, enquêtes, ...), la *vraie* loi est la loi hypergéométrique. Mais, sauf si les valeurs de  $N$  et  $K$  sont faibles, on la remplace systématiquement par la loi binomiale pour utiliser des formules plus simples. L'espérance est la même, la variance et l'écart-type sont très légèrement augmentés. Ainsi, pour un échantillon par exemple de 900 personnes sur 40 millions d'électeurs, il faudrait corriger la variance de

la loi binomiale par le facteur  $\frac{N-n}{N-1} = \frac{39\,999\,100}{39\,999\,999} = 0,99998$  !

**Exemple** On considère 20 objets dont 10 présentent un caractère donné. On effectue 8 tirages, et on appelle  $X$  le nombre d'objets possédant le caractère. Comparer les probabilités  $P_1(X=4)$  pour des tirages AVEC remise et  $P_2(X=4)$  pour des tirages SANS remise.

Le premier cas est celui de la loi binomiale avec  $p = q = 0,5$  :  $P_1(X=4) = \binom{8}{4} \times 0,5^4 \times 0,5^4$   
 $= 70 \times 0,5^8 = 0,2734$ . Le deuxième cas est celui de la loi hypergéométrique avec  $N = 20$ ,

$K = 10, n = 8, k = 5$  :  $P_2(X=4) = \frac{\binom{10}{4} \binom{10}{4}}{\binom{20}{8}} = \frac{210 \times 210}{125\,990} = 0,3501$ . On notera que, comme

on pouvait s'y attendre, la loi hypergéométrique favorise les situations moyennes.



### **i.i.d.**

Abréviation parfois employée pour signifier que des variables aléatoires sont **identiques** et **indépendamment distribuées** (situation des épreuves répétées).

### **impossible (événement)**

(*impossible event*)

Évènement vide  $\emptyset$  = « rien du tout ne s'est passé » d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , de probabilité égale à 0 (c'est bien sûr un cas limite, mais il est nécessaire de l'inclure dans l'ensemble  $\mathcal{A}$  de tous les évènements envisageables).

### **incompatibles (événements)**

(*[mutually] exclusive events*)

Se dit d'évènements  $A, B$  qui ne peuvent pas se réaliser simultanément. On a alors  $P(A \text{ et } B) = 0$ . Le qualificatif ensembliste synonyme est *disjoints*.

Étant donné un nombre supérieur à 2 d'évènements :  $A_1, A_2, \dots, A_n$ , on prendra garde qu'il existe deux manières de généraliser cette propriété : soit en énonçant que les évènements sont « globalement » incompatibles :  $A_1$  et  $A_2$  et ... et  $A_n$  ne peuvent pas se réaliser simultanément, ce qui ne présente en général guère d'intérêt, soit en énonçant que les évènements sont « 2 à 2 » incompatibles :  $\forall i \forall j \quad i \neq j \Rightarrow A_i$  et  $A_j$  ne peuvent pas se réaliser simultanément, propriété souvent très utile.

### **indépendance de deux épreuves**

On dit que deux épreuves sont indépendantes si, dans l'« épreuve produit » qui les représente simultanément, tout évènement « qui ne dépend que de la première épreuve » est indépendant avec tout évènement « qui ne dépend que de la deuxième ».

Cette notion est en pratique très claire dans les cas simples, mais on peut vouloir la formaliser. Il faut pour cela construire le produit des espaces probabilisés qui représentent les deux épreuves. Dans la pratique, l'indépendance de deux épreuves sera rarement une propriété qui doit être vérifiée, et le plus souvent un décret que l'on impose *a priori*.

### **indépendance de deux évènements**

(*stochastic independence*)

On dit que deux évènements  $A, B$  sont indépendants si :

$$P(AB) = P(A) P(B)$$

*Nota* : si  $P(A) \neq 0$ ,  $P(AB) = P(A) P(B)$  équivaut à  $P(B|A) = P(B)$ .

On peut généraliser à une suite  $(A_n)$  finie ou infinie d'évènements : les  $A_n$  sont mutuellement indépendants si, pour tout sous-ensemble fini d'indices  $(i_1, i_2, \dots, i_k)$  on a :

$$P(A_{i_1} A_{i_2} \dots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k})$$

(cette condition est beaucoup plus forte qu'une simple indépendance deux à deux).

L'indépendance « stochastique » (terme consacré qui veut dire « en probabilité ») de deux évènements peut être une propriété que l'on vérifie, mais le plus souvent ce sera une propriété que l'on décrète pour modéliser une indépendance « physique ».

#### Propriété

Si A et B sont indépendants, A et  $\bar{B}$ ,  $\bar{A}$  et B,  $\bar{A}$  et  $\bar{B}$  sont également indépendants

## indépendance de deux variables aléatoires

*(independence of random variables)*

Soient deux variables aléatoires X et Y définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ , à valeurs respectivement dans  $(E_X, \mathcal{A}_X)$  et  $(E_Y, \mathcal{A}_Y)$ . On dit qu'elles sont indépendantes si on a, pour tout évènement  $A \in \mathcal{A}_X$  et tout évènement  $B \in \mathcal{A}_Y$  :

$$P(X \in A \text{ et } Y \in B) = P(X \in A) P(Y \in B)$$

On peut ici aussi généraliser à une suite finie ou infinie de variables aléatoires (définies sur le même espace probabilisé).

Dans la pratique, l'indépendance des variables aléatoires sera rarement une propriété qui doit être vérifiée, et le plus souvent un décret que l'on impose *a priori* (épreuves répétées notamment).

## indépendance (test du khi-deux d')

Voir *khi-deux d'indépendance (test du)*.

## indicateur

Ce mot, qui possède de nombreux synonymes, notamment [valeur] caractéristique et résumé numérique, parfois *paramètre* (mais ce dernier mot n'est pas très approprié dans la situation présente), désigne un certain nombre de valeurs numériques qui résument et synthétisent une distribution, en probabilités ou en statistique. Ils sont utilisés aussi bien pour « se faire une idée » de la distribution que pour figurer dans des formules et des énoncés de théorèmes. Les *conditions de Yule* proposent des critères généraux de qualité globale pour les indicateurs.

Les indicateurs se répartissent en trois grandes catégories.

Les *indicateurs de position* (ou de *tendance centrale* ou encore de *localisation*) caractérisent généralement le « centre » ou le « milieu » d'une distribution. Ils sont nombreux et présentent des qualités et des usages très variés. On peut citer le milieu (trop sommaire), l'espérance mathématique (appelée moyenne en statistique), la médiane, le mode (mais il peut y en avoir plusieurs...), la médiale. Certains indicateurs de position ne caractérisent pas le centre mais des points remarquables, ainsi les quartiles et autres quantiles.

Les *indicateurs* (ou caractéristiques) *de dispersion* caractérisent la manière dont la distribution s'écarte de sa valeur centrale. Les principaux sont liés mathématiquement à un indicateur de position, ainsi la variance et l'écart-type complètent l'espérance (ou moyenne), l'écart inter-quartiles complète la médiane.

La troisième catégorie est celle des *indicateurs de forme*, qui caractérisent notamment l'asymétrie et l'aplatissement.

À cette liste il faut ajouter les *moments* d'une loi de probabilité, qui jouent un rôle théorique important, ainsi que les divers *indices* adaptés à tel ou tel usage particulier et que l'on peut calculer à partir d'une distribution.

Enfin, il existe des indicateurs adaptés à des distributions « à plusieurs dimensions », notamment la covariance et le coefficient de corrélation.

## indicateur

Indicateur d'un événement : variable aléatoire qui prend la valeur 1 ou 0 selon que l'événement se réalise ou ne se réalise pas (*cf. Bernoulli (loi de)*). Indicateur d'un intervalle : fonction  $I(x)$  qui vaut 1 ou 0 selon que  $x$  appartient ou non à l'intervalle.

## indice

(*index*)

Premier sens : dans une situation probabiliste ou statistique, nombre réel sans dimension, indicateur d'une distribution, ou de la « liaison » entre deux distributions (dans ce sens, on emploie plutôt le terme coefficient).

Deuxième sens : dans le domaine des sciences économiques et sociales, nombre réel positif sans dimension, indicateur de la variation (notamment dans le temps) d'une ou plusieurs variables. Il existe deux types d'indices : les indices élémentaires, obtenus par quotient de deux valeurs de la même variable en deux situations (temps, lieu, ...) différentes, et les indices synthétiques, qui se rapportent à un ensemble complexe de variables, le plus souvent non homogènes, pour donner un résumé global de l'évolution. Les indices s'expriment généralement en pourcentages et sont donnés avec une situation de référence (« base 100 en ... »). Les indices synthétiques (indice de Laspeyres, indice de Paasche, indice de Fisher) sont construits dans le respect de critères convenables mais aucun ne peut les satisfaire tous.

## indice de concentration [de Gini]

Voir *concentration (indice de – de Gini)*.

## individu

« Objet » ou « unité » statistique, pris dans une *population* donnée. Il ne faut pas prendre individu et population au sens biologique de ces mots ; il s'agit ici de tout élément sur lequel on peut faire une étude statistique : valeur numérique, mot, document, bien matériel, fait ou phénomène, ... mais incluant bien sûr individu biologique, notamment être humain.

Cette notion est le concept statistique correspondant au concept probabiliste d'événement élémentaire (élément d'un espace fondamental).

## inégalité de Bienaymé-Tchebychev

Voir *Bienaymé-Tchebychev (inégalité de)*.

## inférentielle (statistique)

(*inferential statistics*)

Partie de la statistique qui analyse, dans un cadre explicitement probabiliste, des données préalablement recueillies de façon à en déduire les paramètres des lois de probabilité et à tester la validité du modèle probabiliste ainsi reconstitué.

Voir *statistique*.

## information (quantité d')

(*self-information*)

Soit un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et soit un événement  $A \in \mathcal{A}$ . On appelle quantité d'information apportée par la réalisation de  $A$  la quantité :

$$I(A) = -\log P(A)$$

La base du logarithme importe peu en théorie mais la pratique utilise aujourd'hui, sauf exception, le logarithme à base 2 : l'unité correspondante de quantité d'information est le bit, adaptée aux codes binaires.

## Propriété

Si A, B sont deux évènements indépendants, on a :  

$$I(AB) = I(A) + I(B).$$

En fait, cette propriété est une contrainte plus qu'une conséquence : elle est « naturelle » et nécessaire si l'on veut que la quantité d'information soit une « bonne » mesure de l'information, et le respect de cette contrainte oblige à définir la quantité d'information par le logarithme de la probabilité.

**Exemple théorique** Soit un espace constitué par  $2^n$  évènements élémentaires équiprobables (donc de probabilité  $\frac{1}{2^n}$ ). Si A est l'un d'entre eux,  $I(A) = -\log_2\left(\frac{1}{2^n}\right) = n$  bits.

La *théorie de l'information* (dite parfois de *Shannon-Weaver*) utilise la notion de quantité d'information pour définir l'*entropie*, puis élaborer une théorie destinée à modéliser l'efficacité de la représentation et du stockage de l'information (théorie du codage, théorèmes de Kraft et de Macmillan), et l'efficacité de la transmission de l'information par un « canal » éventuellement « bruité » (théorèmes de Shannon).

## interquartiles (écart)

Voir *écart interquartiles*.

## intersection

([logical] conjunction)

Synonyme de *conjonction logique*.

Dans la formalisation ensembliste des espaces probabilisables, les évènements sont des parties de l'espace fondamental  $\Omega$ . Si l'on considère deux évènements A, B, leur intersection est un évènement dont la réalisation correspond à la conjonction logique « A et B » :

$$\omega \in A \cap B \Leftrightarrow (\omega \in A \text{ et } \omega \in B)$$

Cela se généralise sans difficulté à un nombre supérieur d'évènements. Outre les deux notations ensembliste et logique, parfaitement synonymes :  $A \cap B$ , A et B, on emploie souvent la notation AB qui est très commode.

## Propriété

Si A et B sont indépendants :  $P(A \cap B) = P(A) P(B)$ .  
 De façon générale :  $P(A \cap B) = P(A) P(B|A) = P(B) P(A|B)$ .

**Exemple** Dans un travail de psychologie appliquée, on effectue l'étude croisée de deux « caractères » : la réaction face à un certain signal et la réponse à un certain test. Soient S1, S2, S3 et S4 les réactions au signal, et T1, T2 et T3 les réponses au test. On considère les évènements :

« un sujet pris au hasard montre la réaction S3 au signal »

et : « un sujet pris au hasard donne la réponse T2 au test ».

L'intersection de ces deux évènements est l'évènement « un sujet pris au hasard à la fois montre la réaction S3 au signal et donne la réponse T2 au test ».

## intervalle de confiance

(confidence interval)

Voir *estimation par intervalle*.

## intervalle (estimation par)

Voir *estimation par intervalle*.

## intervalle de prévision, de prédiction

Dans un modèle linéaire ou une régression  $Y = \alpha + \beta X + E$ , intervalle qui encadre avec un certain seuil de confiance les valeurs possibles de  $Y$  pour une valeur « ultérieure » de  $X$ . Sa détermination associe une estimation et une prévision *stricto sensu*.

Voir *prédiction*.

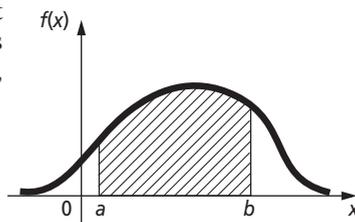
## intervalle (probabilités d')

Lorsqu'une variable aléatoire réelle  $X$  est (absolument) continue, toutes les probabilités ponctuelles  $P(X = x)$  sont nulles et les probabilités « de base » sont les probabilités d'intervalle  $P(a \leq X \leq b)$ . Elles peuvent être calculées, soit à partir de la densité  $f$  :

$$P(a < X \leq b) = \int_a^b f(t) dt ,$$

soit à partir de la fonction de répartition  $F$  :

$$P(a < X \leq b) = F(b) - F(a).$$



## issue

(*outcome*)

Terme parfois employé pour désigner un évènement élémentaire (résultat d'une épreuve).

## joint, e

(*joint*)

Synonyme de conjoint, e.

Dans une situation probabiliste ou statistique « bidimensionnelle » ou « multidimensionnelle », qualifie ce qui concerne globalement les variables : probabilités jointes, densité jointe, loi jointe, effectifs joints, fréquences jointes, espérance ou moyenne jointe, variance jointe, écart-type joint.

Voir *couple de variables aléatoires*.



## Kendall (coefficient de corrélation des rangs de Kendall)

Voir *corrélation des rangs (coefficient de – de Kendall)*.

## khi-deux de Pearson (loi du)

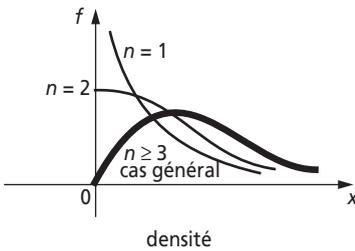
(*chi-squared distribution*)

Loi d'une variable aléatoire continue positive utilisée pour le contrôle des tests du khi-deux.

### Formulaire

Version standardisée à un paramètre entier  $n \geq 1$  qui représente le nombre de « degrés de liberté ». Valeurs sur les réels positifs.

#### ► Loi de probabilité



$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad (x \geq 0)$$

fonction de répartition

$$F(x) = \int_0^x f(t) dt \quad (x \geq 0)$$

#### ► Valeurs caractéristiques

- espérance :  $E(X) = n$
- variance :  $\text{Var}(X) = 2n$
- écart-type :  $\sigma(X) = \sqrt{2n}$

Techniquement, la v.a. khi-deux à  $n$  degrés de liberté est le double d'une v.a. gamma de paramètres  $r = \frac{n}{2}$  et  $\lambda = 1$ .

#### ► Utilisations

En théorie, la v.a. khi-deux  $X$  à  $n$  degrés de liberté (souvent notée  $\chi_n^2$ ) peut être définie comme

la somme des carrés de  $n$  v.a. normales  $U_i$  centrées réduites indépendantes :  $X = \sum_{i=1}^n U_i^2$  (à

noter que l'on a à l'évidence le même « théorème d'addition » que pour les v.a. gamma).

Cette loi intervient dans l'estimation de la variance d'un échantillon : si  $X_1, X_2, \dots, X_n$  sont  $n$  v.a. normales identiques (d'espérance  $\mu$  et d'écart-type  $\sigma$ ) indépendantes, si  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$  est la

variable aléatoire moyenne, et si  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2$  est la variable aléatoire estimateur « débiaisé » de la variance, alors  $\frac{(n-1)S_n^2}{\sigma^2}$  suit une loi du khi-deux à  $n-1$  degrés de liberté.

Enfin, on peut démontrer que, si  $(N_1(n), N_2(n), \dots, N_k(n))$  est un « vecteur multinomial » de paramètres  $n$  et  $(p_1, p_2, \dots, p_k)$ , alors la loi de la variable aléatoire  $\sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}$  converge vers une loi du khi-deux à  $k-1$  degrés de liberté. Dans la pratique, cette propriété fonde le test du khi-deux qui permet notamment de comparer une distribution multinomiale observée à une distribution « théorique ».

### khi-deux d'ajustement (test du) (*chi-squared test of goodness of fit*)

Test qui compare globalement, pour une variable discrète ou discrétisée, la distribution « observée » d'un échantillon statistique à une distribution « théorique » fixée. La variable est quelconque, quantitative ou qualitative, mais le nombre de classes doit être fini (ou rendu fini par regroupements).

Les cas les plus classiques sont :

- 2 classes de probabilités données  $p$  et  $q = 1 - p$  ;
- $k$  classes équiprobables ;
- $k$  classes dont les probabilités sont données *a priori* ;
- $n + 1$  classes associées à une variable binomiale de paramètres  $n$  et  $p$  ;
- $n + 1$  classes associées à une variable à valeurs entières (ou *codée* par des valeurs entières), la  $(n + 1)$ -ième classe regroupant les valeurs  $\geq n$ .

#### test du khi-deux standard sans paramètres estimés

• Notations.  $k$  classes  $A_1, A_2, \dots, A_k$  :

- probabilités réelles des  $k$  classes :  $p_1^* = P(A_1), \dots, p_k^* = P(A_k)$ ,
- probabilités « théoriques » des  $k$  classes :  $p_1, \dots, p_k$ .

• Données.  $n$  observations, avec les effectifs « observés » :  $n_1$  dans la classe  $A_1, \dots, n_k$  dans la classe  $A_k$ .

• Hypothèse testée.  $H_0 = \langle p_1^* = p_1 \text{ et } \dots \text{ et } p_k^* = p_k \rangle$  contre  $H_1 = \langle \text{il existe au moins deux probabilités } p_j^* \text{ et } p_j \text{ différentes} \rangle$ .

• Déroulement technique du test

1. On calcule les effectifs « théoriques »  $np_j$ .
2. On calcule la valeur observée de la variable de test :

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j}.$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi du khi-deux, elles dépendent du nombre de degrés de liberté de l'échantillon :  $ddl = k - 1$ , et du risque  $\alpha$ .

• Conditions et précautions

La loi du khi-deux est la loi limite pour la variable de test, ce qui induit une condition de taille : il est classique de demander que chaque effectif théorique  $np_j$  (insistons :  $np_j$ , non pas  $n_j$ ) soit  $\geq 5$ , exigence que l'on peut descendre à 2, mais pour une seule classe (cas fréquent de la dernière classe d'une distribution décroissante).

Tantôt les probabilités « théoriques » des classes sont explicitement données, tantôt elles sont à déduire de l'énoncé (par exemple classes équiprobables).

**Remarque** : On peut utiliser le test du khi-deux avec deux classes, ce qui revient à comparer un pourcentage (une probabilité) à un pourcentage (une probabilité) théorique. On pourrait faire le même contrôle avec un test (de Student) de comparaison de pourcentages : les valeurs de la variable de test du khi-deux, aussi bien celle observée que celle de la table, sont exactement les carrés des valeurs de la variable du test de Student, et les conclusions sont identiques !

test du khi-deux avec  $h$  paramètres estimés

La procédure de test et les calculs sont identiques au cas standard à une seule exception près : le nombre de degrés de liberté est  $ddl = k - h - 1$ .

Les deux cas les plus fréquents de paramètres estimés sont pour le paramètre  $p$  d'une loi binomiale ou pour le paramètre  $\mu$  d'une loi de Poisson, qui se déduisent l'un comme l'autre du tableau par classes et effectifs (dans ces cas,  $h = 1$  et  $ddl = n - 2$ ).

Les tests du khi-deux sont habituellement rangés parmi les tests non paramétriques, c'est clair si l'on considère le test du khi-deux d'indépendance (ou d'homogénéité), c'est moins évident si l'on considère le test du khi-deux d'ajustement...

**khi-deux d'homogénéité (test du),  
khi-deux d'indépendance (test du)**

*(chi-squared test of  
homogeneity, chi-squared  
test of independence)*

Test qui fonctionne sur un tableau d'effectifs à double entrée et qui contrôle, soit l'homogénéité de sous-populations par rapport à une variable discrète (ou discrétisée), soit l'indépendance de deux variables discrètes (ou discrétisées). La ou les variables sont quelconques, quantitatives ou qualitatives, mais le nombre de classes doit être fini (ou rendu fini par regroupements).

Il n'y a pas d'autre différence entre le test d'homogénéité et le test d'indépendance que la présentation de la situation concrète sur laquelle porte le test.

Dans la présentation test d'indépendance, on considère deux variables définies sur une même population, on dispose d'un tableau croisé d'effectifs (ou tableau de contingence), et on teste l'indépendance des variables.

Dans la présentation test d'homogénéité, on considère d'une part une répartition de la population en sous-populations, d'autre part une variable définie sur la population globale, on dispose donc comme précédemment d'un tableau croisé d'effectifs, et on teste l'homogénéité des sous-populations par rapport à la seule variable explicite. Mais répartir une popula-

tion en sous-populations revient à définir une variable qualitative, dont les modalités sont précisément l'appartenance aux sous-populations. On voit donc qu'il y avait une variable *cachée*, et tester l'homogénéité n'est rien d'autre que tester l'indépendance des deux variables, la cachée et l'explicite.

La description ci-dessous du déroulement du test se fera sous l'unique présentation test d'indépendance.

---

test du khi-deux d'indépendance

---

• Notations

- $nl$  classes  $A_1, A_2, \dots, A_{nl}$  pour la première variable  $X$  ( $nl$  comme nombre de lignes) ;
- $nc$  classes  $B_1, B_2, \dots, A_{nc}$  pour la deuxième variable  $Y$  ( $nc$  comme nombre de colonnes) ;
- probabilités « théoriques » des  $nc \times nl$  classes  $A_i B_j$  :  $p_{ij}$ .

• Données. Effectif total  $n$ , distribué en  $nc \times nl$  effectifs « observés » :  $n_{ij}$  dans la classe  $A_i B_j$ .

• Hypothèse testée.  $H_0$  = «  $X$  et  $Y$  sont indépendantes » contre  $H_1$  alternative.

• Déroulement technique du test

1. On « borde » le tableau des effectifs pour effectuer les sommes marginales :  $L_i$  pour la ligne n°  $i$ ,  $C_j$  pour la colonne n°  $j$ . La disposition pratique des préparatifs du calcul se fait comme indiqué dans le tableau modèle ci-dessous :

		colonne $B_j$		$\Sigma$
	...	...	...	...
ligne $A_i$	...	$n_{ij}$	...	$L_i$
	...	...	...	...
$\Sigma$	...	$C_j$	...	$n$

2. On calcule les effectifs « théoriques »  $np_{ij}$  par la formule  $np_{ij} = \frac{L_i C_j}{n}$ .

3. On calcule la valeur observée de la variable de test :

$$\chi^2 = \sum_{i=1}^{nl} \sum_{j=1}^{nc} \frac{(n_{ij} - np_{ij})^2}{np_{ij}}$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi du khi-deux, elles dépendent du nombre de degrés de liberté de l'échantillon :  $ddl = (nl - 1)(nc - 1)$ , et du risque  $\alpha$ .

• Conditions et précautions

La loi du khi-deux est la loi *limite* pour la variable de test, ce qui induit une condition de taille : il est classique de demander que chaque effectif théorique  $np_{ij}$  (insistons :  $np_{ij}$ , non pas  $n_{ij}$ ) soit  $\geq 5$ , exigence que l'on peut descendre à 2, mais pour une seule classe.

---

On peut utiliser la formule de calcul des effectifs « théoriques » mécaniquement, mais on peut aussi vouloir comprendre « d'où elle sort ». Si  $p_{ij}$  est la probabilité de la classe  $A_i B_j$

« sous l'hypothèse  $H_0$  », l'effectif théorique de cette classe est  $np_{ij}$ . Le problème revient donc à déterminer ces probabilités  $p_{ij}$ . Mais l'hypothèse  $H_0$  énonce que les 2 variables sont indépendantes, donc  $p_{ij} = p_i q_j$ , où  $p_i$  est la probabilité « marginale » de la classe  $A_i$ , et  $q_j$  la probabilité « marginale » de la classe  $B_j$ . Il faut alors estimer ces probabilités marginales. Cela se fait naturellement en totalisant les effectifs par lignes et par colonnes : si  $L_i$  est la somme des effectifs observés de la ligne n°  $i$ , on peut estimer  $p_i$  par  $\frac{L_i}{n}$  ; de même, si  $C_j$  est la somme des effectifs observés de la colonne n°  $j$ , on peut estimer  $q_j$  par  $\frac{C_j}{n}$ . Finalement, on prend pour valeur de  $np_{ij} = np_i q_j$  ; son estimation est  $n \times \frac{L_i}{n} \times \frac{C_j}{n} = \frac{L_i C_j}{n}$ .

Il n'y a pas de notations à la fois évidentes, simples et cohérentes... On trouvera dans certains manuels les notations  $n_{i\cdot}$  et  $n_{\cdot j}$  pour désigner les sommes marginales des effectifs par lignes et par colonnes (voire  $n_{\cdot\cdot}$  pour désigner l'effectif total), c'est parfaitement cohérent mais un tout petit peu lourd. On trouvera aussi  $\widehat{n}_{ij}$  pour désigner l'effectif théorique (que nous avons appelé  $np_{ij}$ ), cela conviendra à ceux qui sont familiers avec la notation « chapeau » des mathématiciens pour désigner les estimations. Enfin, on trouvera parfois  $o_{ij}$  et  $c_{ij}$  (ou  $e_{ij}$ ) pour désigner respectivement les effectifs observés et théoriques ( $c$  comme calculé ou  $e$  comme *estimated*), c'est particulièrement simple, mais cela gêne si l'on veut expliquer et justifier la règle  $c_{ij} = \frac{L_i C_j}{n}$ .

**Remarque** : On peut utiliser le test du khi-deux d'indépendance avec quatre classes ( $2 \times 2$ ), ce qui revient à comparer deux pourcentages (deux probabilités) entre eux. On pourrait faire le même contrôle avec un test (de Student) de comparaison de pourcentages : les valeurs de la variable de test du khi-deux, aussi bien celle observée que celle de la table, sont exactement les carrés des valeurs de la variable du test de Student, et les conclusions sont identiques ! On notera bien que pour ce test  $2 \times 2$ , on a  $ddl = 1$ .

## Khintchine (Alexandre)

Mathématicien russe (1894–1989). Il obtint des résultats profonds sur les théorèmes limites du calcul des probabilités.

## Kolmogorov (Andrei)

Mathématicien russe (1903–1987). Il formula l'axiomatique moderne du calcul des probabilités et fit également des travaux en topologie et en théorie des systèmes dynamiques.

## Kolmogorov (axiomatique de)

Voir *axiomatique de Kolmogorov*.

## Kolmogorov (test [d'ajustement] de)

(Kolmogorov test)

Test non paramétrique qui compare la distribution d'un échantillon statistique à une distribution fixée (par exemple : loi exponentielle de paramètre  $\lambda$  spécifié, ou loi normale d'espérance et de variance spécifiées). Les distributions (lois) sont représentées par leurs fonctions de répartition, utilisées pour l'exécution du test.

test bilatéral de comparaison  
d'une distribution de fonction de répartition  $F(x)$   
à une distribution de fonction de répartition fixée  $F_0(x)$

- Données. Un échantillon  $(x_1, x_2, \dots, x_n)$  de  $n$  valeurs observées d'une variable aléatoire numérique  $X$  de fonction de répartition  $F(x)$ .
- Hypothèse testée.  $H_0 = \langle F = F_0 \rangle$  contre  $H_1 = \langle F \neq F_0 \rangle$
- Déroulement technique du test

1a. On ordonne les valeurs observées de l'échantillon – on suppose ce rangement effectué, soit, en gardant les notations initiales :

$$x_1 \leq x_2 \leq \dots \leq x_n$$

1b. Puis on pose :

$$F(x_1) = \frac{1}{n}, F(x_2) = \frac{2}{n}, \dots, F(x_n) = \frac{n}{n} = 1,$$

ce qui définit les « marches » de la fonction de répartition observée, qui est une fonction « en escalier ».

2. On pose alors :

$$K^+ = \sup_x (F(x) - F_0(x)) = \max_{1 \leq j \leq n} \left( \frac{j}{n} - F_0(x_j) \right)$$

$$K^- = -(\inf_x (F(x) - F_0(x)))$$

$$= \sup_x (F_0(x) - F(x)) = \max_{1 \leq j \leq n} \left( F_0(x_j) - \frac{j-1}{n} \right)$$

La valeur observée de la variable de test est :

$$K = \sup_x |F(x) - F_0(x)| = \max(K^+, K^-)$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi « du  $\Delta$  » de Kolmogorov–Smirnov, elles dépendent de la taille  $n$  de l'échantillon et du risque  $\alpha$ .

Valeurs limites : si  $D_n$  est la variable aléatoire tabulée, telle que  $\sqrt{n} D_n$  converge en loi vers la loi « du  $\Delta$  » de Kolmogorov–Smirnov lorsque  $n \rightarrow \infty$ , et si  $d_n(\alpha)$  est la valeur critique

de  $D_n$  définie par  $P(D_n > d_n(\alpha)) = \alpha$ , on a  $D_n(0,05) \sim \frac{1,358}{\sqrt{n}}$  et  $D_n(0,01) \sim \frac{1,629}{\sqrt{n}}$ .

- Conditions et précautions

- Il n'y a pas d'autre précaution que de fixer complètement la loi de référence (donc en particulier d'éviter toute estimation de paramètres) ;
- lorsqu'il faut estimer des paramètres, la loi de référence de la variable de test n'est pas connue mais il existe des tables obtenues par simulation... Valeurs limites (*Biometrika Tables*) : si  $D'_n$  est la variable aléatoire *corrigée* lorsque l'on fait fonctionner le test de

Kolmogorov après estimation de l'espérance et de l'écart-type d'une loi normale, et si  $d'_n(\alpha)$  est la valeur critique de  $D'_n$  définie comme ci-dessus, on a  $D'_n(0,05) \sim \frac{0,895}{\sqrt{n}}$  et

$$D'_n(0,01) \sim \frac{1,035}{\sqrt{n}}.$$

On pourrait faire fonctionner le test en unilatéral en considérant seulement  $K^+$  ou  $K^-$ , mais il ne semble pas y avoir d'application pratique.

## Kolmogorov-Smirnov (loi du $\Delta$ de)

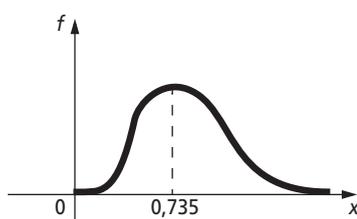
(Kolmogorov-Smirnov  
 $\Delta$  distribution)

Loi d'une variable aléatoire continue utilisée pour le contrôle des tests de Kolmogorov et de Smirnov qui permettent de comparer deux fonctions de répartition.

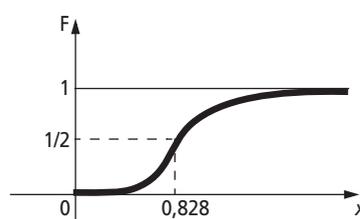
### Formulaire

Version standardisée  $X$  sans paramètre, à valeurs réelles positives.

#### ► Loi de probabilité



densité



fonction de répartition

$$F(x) = 1 - 2 \sum_{k=1}^{\infty} (-1)^k \exp(-2k^2 x^2)$$

#### ► Valeurs caractéristiques

- espérance :  $E(X) = 0,8687$
- variance :  $\text{Var}(X) = 0,06777$
- écart-type :  $\sigma(X) = 0,2603$

#### ► Autres valeurs caractéristiques

mode :  $M_0 = 0,7354$

médiane :  $M = 0,8276$

#### ► Utilisation

Si  $D_n$  est le sup (sur  $x$ ) de la valeur absolue de la différence entre deux fonctions de répartition,  $F(x)$  « exacte » et  $F_n(x)$  empirique (observée), la loi de la variable aléatoire  $\sqrt{n} D_n$  converge vers la loi du  $\Delta$  de Kolmogorov-Smirnov lorsque  $n$  tend vers l'infini.

Cette loi intervient à ce titre dans les tests de Kolmogorov et de Smirnov.

## [Kolmogorov-]Smirnov (test de) ([Kolmogorov-]Smirnov test)

Test non paramétrique qui compare entre elles les distributions de deux échantillons statistiques. Les distributions (lois) sont représentées par leurs fonctions de répartition, utilisées pour l'exécution du test.

---

### test bilatéral de comparaison de deux distributions de fonctions de répartition $F_1(x)$ et $F_2(x)$

---

• Données. Deux séries :

- un échantillon  $(x_1, x_2, \dots, x_{n_X})$  de  $n_X$  valeurs observées d'une variable aléatoire numérique  $X$  de fonction de répartition  $F_1(z)$  ;
- un échantillon  $(y_1, y_2, \dots, y_{n_Y})$  de  $n_Y$  valeurs observées d'une variable aléatoire numérique  $Y$  de fonction de répartition  $F_2(z)$ .

• Hypothèse testée.  $H_0 = \langle F_1 = F_2 \rangle$  contre  $H_1 = \langle F_1 \neq F_2 \rangle$

• Déroulement technique du test

1a. On ordonne toutes les valeurs observées des deux échantillons :

$$z_1 \leq z_2 \leq \dots \leq z_{n_X + n_Y}.$$

1b. Puis on construit les deux fonctions de répartition observées, qui sont des fonctions « en escalier », avec des « marches » pour certaines des  $z_i$ .

2. On calcule alors  $K = \sup_z |F_1(z) - F_2(z)|$  en s'inspirant de la méthode exposée dans le cas du test de Kolmogorov simple de comparaison d'une fonction de répartition à une fonction de répartition fixée (avec notamment le « dédoublement » des valeurs aux bords des marches – les formules ne sont pas données ici).

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi « du  $\Delta$  » de Kolmogorov–Smirnov, elles dépendent de la taille  $n$  de l'échantillon et du risque  $\alpha$ . La lecture de la valeur critique n'est pas directe sauf si l'on se contente d'assimiler la loi à la loi limite : dans le cas de  $H_0 = \langle F = F_0 \rangle$  avec un échantillon de  $n$  valeurs, la variable  $\sqrt{n} \sup |F(x) - F_0(x)|$  converge en loi vers le «  $\Delta$  », et dans le cas

de  $H_0 = \langle F_1 = F_2 \rangle$  avec deux échantillons de  $n_X$  et  $n_Y$  valeurs, la variable  $\sqrt{\frac{n_X n_Y}{n_X + n_Y}} \sup |F_1(z) - F_2(z)|$  converge en loi vers le «  $\Delta$  ».

---

**Remarque :** Le calcul de  $K = \sup |F_1(z) - F_2(z)|$ , qui est extrêmement lourd dans le cas général, se simplifie considérablement s'il est possible, sans trop altérer l'information contenue dans les observations, de ranger celles-ci dans des classes de mêmes limites pour  $X$  et  $Y$ .

## König-Huygens (formule de)

Voir formule de Huygens-König.

## Kruskall-Wallis (test de)

(Kruskall-Wallis test)

Test non paramétrique utilisé pour comparer les distributions de plusieurs échantillons statistiques. Comme le test de Wilcoxon, il fonctionne, non pas à partir des valeurs précises observées, mais à partir des rangs de ces valeurs interclassées.

Si les variables aléatoires  $X_1, X_2, \dots, X_q$  dont proviennent respectivement les  $q$  échantillons ont toutes même loi, elles ont en particulier même espérance mathématique, et c'est très souvent comme test de l'hypothèse dérivée «  $\mu_1 = \mu_2 = \dots = \mu_q$  » que le test de Kruskal-Wallis est utilisé.

———— test non paramétrique de comparaison de  $q$  lois de probabilité, ———  
également utilisé pour comparer  $q$  espérances mathématiques  $\mu_1, \mu_2, \dots, \mu_q$

- Données.  $q$  séries ou groupes, soit pour chaque  $k$  ( $1 \leq k \leq q$ ) : un échantillon  $(x_{k1}, x_{k2}, \dots, x_{kn_k})$  de  $n_k$  valeurs observées d'une variable aléatoire numérique  $X_k$  d'espérance mathématique  $\mu_k$ . On note  $N$  le nombre total de valeurs observées  $n_1 + n_2 + \dots + n_q$ .
- Hypothèse réellement testée.  $H_0 =$  « toutes les  $X_k$  ont même loi » contre  $H_1$  alternative.
- Hypothèse dérivée.  $H_0 =$  «  $\mu_1 = \mu_2 = \dots = \mu_q$  » contre  $H_1 =$  « il existe au moins deux espérances différentes ».
- Déroulement technique du test
  1. On classe les  $N$  valeurs observées, tous groupes confondus, par ordre croissant.
  2. Pour chaque  $k$  on calcule la somme  $T_k$  des rangs des valeurs de la variable  $X_k$  (s'il y a des ex æquo, on leur attribue le rang moyen).
  3. On calcule la valeur observée de la variable de test par l'une ou l'autre des formules :

$$K = \frac{12}{N(N+1)} \sum_{k=1}^q \frac{T_k^2}{n_k} - 3(N+1) = \frac{12}{N(N+1)} \sum_{k=1}^q n_k \left( \frac{T_k}{n_k} - \frac{N+1}{2} \right)^2.$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi du khi-deux, elles dépendent du nombre de degrés de liberté :  $ddl = q - 1$ , et du risque  $\alpha$ .

- Conditions et précautions
  - Il n'y a aucune condition sur la loi commune à toutes les  $X_k$  ;
  - par contre, la loi du khi-deux est la loi limite pour la variable de test, ce qui induit une condition de taille : il est classique de demander que l'effectif  $n_k$  de chaque groupe soit  $\geq 5$ .



## de Laplace (Pierre-Simon)

Astronome, mathématicien et physicien français (1749–1827). Il publia une somme monumentale, *Théorie analytique des probabilités* (corrélation, convergence stochastique, théorème central limite, ...) et fit des travaux importants en mécanique céleste et en physique (théorie de la chaleur, électromagnétisme).

## Laplace (loi de)

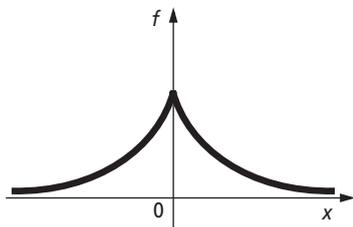
(Laplace distribution)

Loi d'une variable aléatoire continue qui avait été proposée par Laplace pour rendre compte des erreurs d'expérience, rôle où elle a été abandonnée au profit de la *loi normale alias* loi de Laplace-Gauss.

### Formulaire

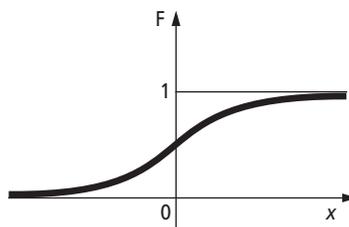
Version standardisée sans paramètre ; valeurs sur les réels.

#### ► Loi de probabilité



densité

$$f(x) = \frac{1}{2}e^{-|x|}$$



fonction de répartition

$$F(x) = \begin{cases} \frac{1}{2}e^x & \text{si } x < 0 \\ 1 - \frac{1}{2}e^{-x} & \text{si } x \geq 0 \end{cases}$$

#### ► Valeurs caractéristiques

- espérance :  $E(X) = 0$
- variance :  $\text{Var}(X) = 2$
- écart-type :  $\sigma(X) = \sqrt{2}$

Cette loi est la loi de la différence entre deux v.a. exponentielles de paramètre  $\lambda = 1$  et indépendantes.

Synonyme de *première loi de Laplace, loi double-exponentielle*.

## Laplace–Gauss (loi de)

Synonyme de *seconde loi de Laplace. Voir normale (loi)*.

## Legendre (Adrien-Marie)

Mathématicien français (1752–1833). Il introduisit la méthode des moindres carrés (utilisée par Gauss en astronomie) pour l'ajustement en statistique et fit des travaux importants en théorie des nombres et en analyse.

### lemme

Voir *Borel-Cantelli (lemme de)*, *Neyman-Pearson (lemme de)*.

## Lévy (Paul)

Mathématicien français (1886–1971). Il approfondit la théorie des processus et du mouvement brownien.

### libre (test)

(*free test, distribution-free test*)

Voir *test d'hypothèse*.

### linéaire (fonction)

*Stricto sensu*, fonction de la forme  $y = bx$ . En fait, les fonctions « affines »  $y = a + bx$  que l'on rencontre notamment en statistique comme courbes de régression sont le plus souvent qualifiées de « linéaires ».

### linéarité d'une régression (test de)

(*test of linearity*)

Test paramétrique qui contrôle la linéarité d'une régression dans le cas d'un couple de variables (X, Y) pour lequel on dispose de plusieurs observations de Y pour chaque valeur  $x_i$  de X (il suffit en théorie qu'il y ait au moins deux observations pour une seule des valeurs de X, mais la pertinence du test devient douteuse s'il n'y a pas plusieurs observations pour beaucoup de valeurs de X). Dans ce cas il est possible de calculer le rapport de corrélation  $e_{Y|X}^2$ , et la linéarité se contrôle en testant l'égalité de  $e_{Y|X}^2$  avec le carré  $r^2$  du coefficient de corrélation linéaire.

#### test de contrôle de la linéarité d'une régression

• Données. Un échantillon de  $q$  groupes d'observations relatifs aux valeurs  $x_1, x_2, \dots, x_q$  de X, avec pour le groupe n°  $i$ ,  $n_i$  valeurs observées  $y_{i1}, y_{i2}, \dots, y_{in_i}$  de la variable Y.

On pose  $n = n_1 + n_2 + \dots + n_q$ .

• Hypothèse testée.  $H_0 = \ll \text{l'espérance conditionnelle est affine : } E(Y|x) = \alpha + \beta X \gg = \ll \eta_{Y|X}^2 = \rho^2 \gg$  contre  $H_1 = \ll \eta_{Y|X}^2 \neq \rho^2 \gg$

• Déroulement technique du test

1. On calcule avec les formules usuelles le coefficient de corrélation linéaire observé  $r$  et le rapport de corrélation observé  $e^2 = e_{Y|X}^2$ .
2. On calcule la variable de test :

$$F_{q-2, n-q} = \frac{\left( \frac{e^2 - r^2}{q-2} \right)}{\left( \frac{1 - e^2}{n-q} \right)}$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Fisher-Snedecor, elles dépendent des deux degrés de liberté  $q-2$  et  $n-q$ , et du risque  $\alpha$ .

• Conditions et précautions

- En théorie le couple  $(X, Y)$  doit être une v.a. normale (à deux dimensions) ; en outre, la variance conditionnelle  $\text{Var}(Y|x)$  doit être constante, exigence difficile à apprécier et à contrôler... ;
- lorsque le couple  $(X, Y)$  n'est pas normal, le test est robuste et reste applicable si les effectifs  $n_i$  des groupes d'observations par valeurs de  $X$  sont « assez grands »

**Remarque :** On trouve parfois la formule « type analyse de la variance »  $F_{q-2, n-q}$

$$= \frac{s_1^2}{s_2^2}, \text{ où } s_1^2 \text{ est la variance entre valeurs de } x \text{ de l'écart (à la linéarité) et } s_2^2 \text{ la}$$

variance résiduelle. C'est une simple variante de la présentation, et la valeur de la variable de test est la même (le premier nombre de degrés de liberté est ici  $q-2$  et non pas  $q-1$  car la régression estime 2 paramètres et non pas 1).

## logarithme itéré (loi du)

(law of iterated logarithm)

Lorsque l'on considère par exemple une suite de parties de Pile ou Face, avec chaque partie un gain de 1 ou une perte de 1 selon le résultat, le somme des gains est une variable aléatoire  $S_n$  centrée, dont on sait qu'elle vérifie les deux propriétés : d'une part  $\frac{S_n}{n}$  tend vers 0 (« en probabilité » et aussi « presque sûrement »), et d'autre part l'ordre de grandeur moyen de  $\frac{S_n}{n}$  est  $\frac{1}{\sqrt{n}}$ . À un niveau approfondi, la « loi » du logarithme itéré (qui est tout aussi un théorème que la « loi » des grands nombres !) donne un renseignement supplémentaire sur les valeurs extrêmes de  $\frac{S_n}{n}$ .

**Théorème.** On considère une suite  $(X_n)$  de variables aléatoires (réelles) indépendantes et identiquement distribuées, d'espérance mathématique  $\mu$  et de variance  $\sigma^2$ . On définit les sommes  $S_n = X_1 + X_2 + \dots + X_n$ , puis les variables centrées réduites correspondantes :

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Alors on a presque sûrement :

$$\limsup_{n \rightarrow \infty} \left| \frac{Z_n}{\sqrt{2 \ln \ln n}} \right| \leq 1.$$

On peut aussi dire, en « traduisant » la notion mathématique de limite supérieure, que, pour tout  $c > 1$ , il n'y aura presque sûrement qu'un nombre fini d'évènements «  $|Z_n| > c \sqrt{2 \ln \ln n}$  » qui se réaliseront.

Sur la présentation donnée en introduction de gains +1 ou -1, cela veut dire que, pour  $c > 1$ ,  $\frac{|S_n|}{n}$  ne dépassera  $\frac{c\sqrt{2 \ln \ln n}}{\sqrt{n}}$  qu'un nombre fini de fois (et  $\sqrt{2 \ln \ln n}$  est une fonction qui tend vers l'infini très très lentement !).

**logit (transformation en)***(logit transformation)*

Transformation  $y = \ln\left(\frac{F(x)}{1-F(x)}\right)$ , appliquée à la fonction de répartition empirique d'une distribution statistique d'un échantillon d'une variable aléatoire, et qui permet pour certains types de lois de représenter graphiquement cette fonction de répartition par une droite (*nota* : le « log » de logit est le début du mot logistique).

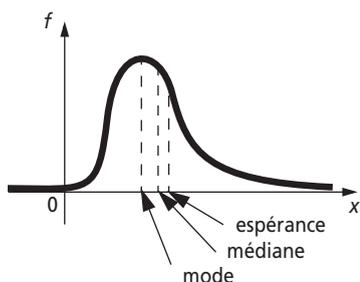
**log-normale (loi)***(log-normal distribution)*

Loi d'une variable aléatoire continue dont le logarithme est une v.a. normale.

**Formulaire**

Deux paramètres réels :  $\mu \in \mathbf{R}$  et  $\sigma \in \mathbf{R}_+^*$  ; valeurs sur les réels positifs.

## ► Loi de probabilité



densité

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \quad (x > 0)$$

fonction de répartition

$$F(x) = \int_0^x f(t) dt \quad (x > 0)$$

## ► Valeurs caractéristiques

- espérance :  $E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$
- variance :  $\text{Var}(X) = \exp(2\mu + \sigma^2) (\exp(\sigma^2) - 1)$
- écart-type :  $\sigma(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right) \sqrt{\exp(\sigma^2) - 1}$

## ► Autres valeurs caractéristiques

mode :  $M_0 = \exp(\mu - \sigma^2)$

médiane :  $M = \exp(\mu)$

## ► Utilisation

Si  $X$  suit une loi log-normale de paramètres  $\mu$  et  $\sigma$ ,  $Z = \ln X$  suit une loi normale  $\mathcal{N}(\mu, \sigma)$ . Attention !  $\mu$  et  $\sigma$  ne sont donc pas l'espérance et l'écart-type de  $X$ .

Dans la pratique, la loi normale est une approximation de la « résultante multiplicative » de grandeurs aléatoires positives, petites et nombreuses, et pas trop mutuellement dépendantes...

Cette loi est très employée par les utilisateurs professionnels. Néanmoins, lorsque l'écart-type est faible devant l'espérance (inférieur au quart, par exemple), une telle résultante multiplicative est souvent approchée par une loi normale « ordinaire » (la probabilité artificielle que la variable aléatoire soit négative est alors négligeable).

**Remarque** : la loi log-normale est parfois appelée *loi de Gilbrat* lorsque le paramètre  $\mu$  est nul ; ce cas particulier est d'un intérêt très limité.

**loi conjointe**

Voir *couple de variables aléatoires*.

## loi des estimateurs dans une régression et intervalles de confiance

On considère un couple  $(X, Y)$  de variables numériques, soit dans une situation de *modèle linéaire* ( $X$  est alors un ensemble  $\{x_i\}$  de valeurs « maîtrisées » et  $Y$  un ensemble  $\{Y_i\}$  de variables aléatoires normales associées), soit dans une situation de *régression linéaire* ( $(X, Y)$  est alors un couple de v.a. qui suit une loi normale à 2 dimensions). On peut définir dans l'un et l'autre cas la *droite de régression* théorique  $y = \alpha + \beta x$ , et la droite de régression empirique  $y = a + bx$ .

On donne ci-dessous des estimations non biaisées de tous les paramètres et coefficients qui interviennent (on peut employer au choix le langage des estimations, ou celui des estimateurs (dont les estimations sont des « réalisations »)).

### ► Variances marginales et covariance

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - \bar{x})^2, s_Y^2 = \frac{1}{n-1} \sum_{i=1}^{i=n} (y_i - \bar{y})^2, \text{Cov}_{X,Y} = \frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - \bar{x})(y_i - \bar{y})$$

sont des estimations sans biais de  $\sigma_X^2$ ,  $\sigma_Y^2$  et  $\text{Cov}_{X,Y}$ .

### ► Coefficients de la droite de régression, coefficient de corrélation

$$b = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}, a = \bar{y} - b\bar{x}, r = \frac{\text{Cov}_{X,Y}}{s_X s_Y}$$

sont des estimations sans biais de  $\beta$ ,  $\alpha$  et  $\rho = \text{Corr}(X, Y)$  (quel que soit le mode de calcul : les quotients qui définissent ces paramètres font disparaître les dénominateurs, qu'ils soient  $n$  ou  $n-1$  (sous réserve de cohérence) et annulent donc l'effet des biais ou des débiais).

### ► Valeur de l'espérance conditionnelle

$a + bx$  est une estimation sans biais de  $E(Y|X = x)$

### ► Variance résiduelle

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{i=n} (y_i - (a + bx_i))^2 \text{ est une estimation sans biais de } \sigma_E^2 = \sigma^2.$$

On donne maintenant les lois des estimateurs avec leurs paramètres, qui seront utilisées soit pour calculer des intervalles de confiance, soit pour effectuer des tests d'hypothèses (les deux démarches étant, présentation exceptée, entièrement équivalentes).

Dans toutes les formules qui suivent,  $s$  représente la racine carrée de l'estimation non biaisée (ci-dessus) de la variance résiduelle.

### ► Pente de la droite de régression

Si  $s_b = \frac{s}{\sqrt{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}}$ , le quotient  $\frac{b - \beta}{s_b}$  suit une loi de Student à  $n - 2$  degrés de

liberté.

- Ordonnée à l'origine de la droite de régression

Si  $s_a = s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$ , le quotient  $\frac{a - \alpha}{s_a}$  suit une loi de Student à  $n - 2$  degrés

de liberté.

- Coefficient de corrélation

$\frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2}$  suit une loi de Student à  $n - 2$  degrés de liberté.

- Valeur de l'espérance conditionnelle

Si  $s_{a+bx} = s \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$ , le quotient  $\frac{(a + bx) - E(Y|X = x)}{s_{a+bx}}$   
 $= \frac{(a + bx) - (\alpha + \beta x)}{s_{a+bx}}$  suit une loi de Student à  $n - 2$  degrés de liberté

## loi d'une variable aléatoire / loi de probabilité d'une variable aléatoire ([probability] distribution)

Mesure de probabilité d'une variable aléatoire, qui peut être explicitement décrite, ou bien caractérisée en donnant les valeurs prises et la liste ou l'expression de la mesure de probabilité, ou encore définie par référence à une variable aléatoire « modèle » classique.

**Exemple 1** cas discret, liste des valeurs et liste des probabilités : soit X la variable aléatoire qui prend les valeurs 1, 3 et 5 avec les probabilités 0,1, 0,8 et 0,1.

**Exemple 2** cas discret, caractérisation des valeurs et expression mathématique des probabilités : soit X la variable aléatoire qui prend les valeurs entières de  $k = 0$  à  $k = n - 1$ , avec les probabilités définies par  $P(X = k) = p_k = 2 \frac{n - k}{n(n - 1)}$  (il faut bien sûr contrôler que  $\sum_{k=0}^{n-1} p_k = 1$ ).

**Exemple 3** cas discret, référence à une variable aléatoire classique : soit X la variable de Poisson  $\mathcal{P}(\mu)$  de paramètre  $\mu = 2,5$ .

**Exemple 4** cas continu, caractérisation des valeurs et caractérisation de la mesure de probabilités par l'expression de sa densité : soit X la variable aléatoire qui prend ses valeurs sur l'intervalle  $[-1, 1]$  avec la densité de probabilité  $f(x) = \frac{3}{4}(1 - x^2)$  (il faut bien sûr contrôler

que  $\int_{-1}^1 f(x) dx = 1$ ).

**Exemple 5** cas continu, référence à une variable aléatoire classique : soit X la variable normale  $\mathcal{N}(\mu, \sigma)$  de paramètres  $\mu = 1,39$  et  $\sigma = 0,08$ .

Voir *variable aléatoire (typologie)*.

### loi jointe

Voir *couple de variables aléatoires*.

### loi marginale

Voir *couple de variables aléatoires*.

### Lorentz (courbe de concentration de)

Voir *concentration (courbe de - de Lorentz)*.



## Mann–Whitney (test de)

(Mann-Whitney test, *u*-test)

Test d'hypothèse non paramétrique utilisé pour comparer les distributions de deux échantillons statistiques. Il fonctionne, non pas à partir des valeurs précises observées, mais à partir des rangs de ces valeurs interclassées.

Si les variables aléatoires  $X$  et  $Y$  dont proviennent respectivement les deux échantillons ont même loi, elles ont en particulier même espérance mathématique, et c'est très souvent comme test de l'hypothèse dérivée «  $\mu_X = \mu_Y$  » que le test de Mann–Whitney est utilisé. L'hypothèse (réellement testée)  $H_0 = \ll X \text{ et } Y \text{ ont même loi} \gg$  a pour conséquence immédiate la symétrie  $P(X \leq Y) = P(X \geq Y)$  (si les lois sont continues, on a par surcroît  $P(X = Y) = 0$ , et donc  $P(X \leq Y) = P(X \geq Y) = \frac{1}{2}$ ). La mise en œuvre du test de Mann-Whitney est une simple exploitation de cette égalité des probabilités symétriques.

—— test non paramétrique de comparaison de deux lois de probabilité, ——  
également utilisé pour comparer deux espérances mathématiques  $\mu_X$  et  $\mu_Y$

• Données. Deux séries :

- un échantillon  $(x_1, x_2, \dots, x_{n_X})$  de  $n_X$  valeurs observées d'une variable aléatoire numérique  $X$  d'espérance mathématique  $\mu_X$  ;
- un échantillon  $(y_1, y_2, \dots, y_{n_Y})$  de  $n_Y$  valeurs observées d'une variable aléatoire numérique  $Y$  d'espérance mathématique  $\mu_Y$ .

• Hypothèse réellement testée.  $H_0 = \ll X \text{ et } Y \text{ ont même loi} \gg$  contre  $H_1$  alternative.

• Hypothèse dérivée.  $H_0 = \ll \mu_X = \mu_Y \gg$  contre  $H_1 = \ll \mu_X \neq \mu_Y \gg$

• Déroulement technique du test

1. On classe les  $n_X + n_Y$  valeurs observées par ordre croissant.
2. On calcule le nombre  $U_{YX}$  d'inversions  $y_i < x_j$  (s'il y a des ex-æquos, on compte  $\frac{1}{2}$  pour tout couple  $y_i = x_j$ ).
3. On calcule la valeur observée de la variable de test :

$$U = \frac{\left| U_{YX} - \frac{n_X n_Y}{2} \right|}{\sqrt{\frac{n_X n_Y (n_X + n_Y + 1)}{12}}}$$

Les valeurs de référence de la variable de test sont à lire soit dans des tables spécifiques pour les petites valeurs de  $n_X$  et  $n_Y$ , soit dans la table de la loi normale (centrée réduite), pour le risque bilatéral  $\alpha$ .

## • Conditions et précautions

- Il n'y a aucune condition sur la loi commune à X et Y ;
- par contre, la loi normale (centrée réduite) est la loi limite pour la variable de test, ce qui induit une condition de taille si l'on ne dispose pas de table spécifique ; il est classique de demander  $n_X$  et  $n_Y \geq 10$  pour pouvoir se référer à la table de la loi normale.

Il règne un certain flottement dans l'appellation de ce test. Il existe en effet un *test de Wilcoxon* qui teste les mêmes hypothèses dans la même situation, en calculant la somme des rangs au lieu de compter les inversions du classement. Ces deux tests sont complètement équivalents (la variable  $W_X$  du test de Wilcoxon et la variable  $U_{YX}$  du test de Mann-Whitney sont liées par la relation  $U_{YX} = W_X - \frac{1}{2}n_X(n_X + 1)$ ). Dans certains ouvrages, les appellations sont permutées.

**marginal, e***(marginal)*

Dans une situation probabiliste ou statistique « bidimensionnelle » ou « multidimensionnelle », qualifie ce qui concerne une seule des variables : probabilités marginales, densité marginale, loi marginale, effectifs marginaux, fréquences marginales, espérance ou moyenne marginale, variance marginale, écart-type marginal.

Le mot s'inspire des tableaux « de contingence » pour deux variables, où traditionnellement les effectifs marginaux sont calculés dans la marge droite (sommation des lignes) et la marge inférieure (sommation des colonnes).

Voir *couple de variables aléatoires*.

**martingale***(martingale)*

Terme qui désigne dans la langue courante, dans un contexte de jeux de hasard, une stratégie qui permet (qui permettrait...) de gagner à coup sûr. Si le jeu n'est pas « équitable » (*i.e.* si l'espérance de gain du joueur est négative – c'est le cas notamment pour tous les jeux de casino !), c'est bien sûr totalement impossible. Et si le jeu est « équitable » (*i.e.* si l'espérance de gain du joueur est nulle), un théorème de calcul des probabilités énonce qu'il ne peut exister de martingale que si la fortune du joueur est infinie.

Ce terme a été repris par les mathématiciens pour désigner une catégorie de processus stochastiques  $(X_n)$  dont la propriété essentielle est que l'espérance conditionnelle (à tout ce qui précède) de  $X_{n+1}$  (à ce niveau élaboré, l'espérance conditionnelle est elle-même une variable aléatoire) est  $X_n$ . Ces processus peuvent notamment modéliser les gains cumulés d'un joueur dans un jeu équitable.

**Markov (Andrei)**

Mathématicien russe (1856–1922). Il prépara la modernisation du calcul des probabilités et introduisit les chaînes d'évènements.

**Markov (chaîne de)**

Voir *chaîne de Markov*.

**Markov (inégalité de)**

Voir *Bienaymé-Tchebychev (inégalité de)*.

## maximum de vraisemblance (méthode du)

(*maximum likelihood method*)

Méthode théorique d'estimation des paramètres. On considère  $n$  valeurs observées  $(x_1, x_2, \dots, x_n)$  d'un échantillon de  $n$  variables aléatoires indépendantes et de même loi, loi dépendant d'un paramètre  $\theta$ . On définit alors une fonction  $L(\theta; x_1, x_2, \dots, x_n)$  qui est, pour la valeur  $\theta$  du paramètre, la probabilité de la loi jointe de l'échantillon  $(X_1, X_2, \dots, X_n)$  dans le cas discret, ou sa densité jointe dans le cas continu. Considérée comme fonction de  $\theta$  seul, cette fonction s'appelle la « vraisemblance ».

Dans son principe, la méthode du maximum de vraisemblance consiste à choisir comme estimation de  $\theta$  la valeur qui maximise la vraisemblance  $L(\theta; x_1, x_2, \dots, x_n)$ , *i.e.* la valeur pour laquelle les observations effectuées avaient « le plus de chance » de se produire. Cette justification empirique est confirmée par des théorèmes de mathématiques et cette méthode générale fournit des estimateurs qui sont très souvent les meilleurs.

## Maxwell (loi de)

(*Maxwell distribution*)

Synonyme de *Maxwell–Boltzmann (loi de)*.

Loi d'une variable aléatoire continue qui peut être définie comme la racine carrée de la somme des carrés de trois variables aléatoires normales indépendantes  $Y_1, Y_2$  et  $Y_3$ , centrées et de même écart-type :

$$X = \sqrt{Y_1^2 + Y_2^2 + Y_3^2}$$

### Formulaire

Un paramètre réel  $c \in \mathbf{R}_+^*$  (qui est l'écart-type des v.a. normales indépendantes  $Y_i$ ).

#### ► Loi de probabilité

densité

fonction de répartition

$$f(x) = \sqrt{\frac{2}{\pi}} \frac{x^2}{c^2} \exp\left(-\frac{x^2}{2c^2}\right) \quad (x \geq 0)$$

$$F(x) = \int_0^x f(t) dt \quad (x \geq 0)$$

#### ► Valeurs caractéristiques

– espérance :  $E(X) = \sqrt{\frac{8}{\pi}} c \approx 1,596c$

– variance :  $\text{Var}(X) = \frac{3\pi - 8}{\pi} c^2 \approx 0,454c^2$

– écart-type :  $\sigma(X) = \sqrt{\frac{3\pi - 8}{\pi}} c \approx 0,673c$

#### ► Utilisations

De par sa définition même, la loi de Maxwell est la loi de la distance à l'origine d'un point aléatoire de l'espace, dont les coordonnées sont trois variables aléatoires normales indépendantes, centrées et de même écart-type.

C'est également, et « naturellement », la loi de la distribution des vitesses des molécules dans un gaz ; on trouvera dans l'article juste suivant *mécanique statistique* une justification de cette distribution comme conséquence de la modélisation effectuée par la mécanique statistique.

Voir *Rayleigh (loi de)*.

## mécanique statistique

(statistical mechanics)

Branche de la physique ayant pour but d'expliquer le comportement macroscopique d'un système composé d'un grand nombre d'éléments ou de de particules à partir de la modélisation microscopique de ses particules. Les grandeurs physiques macroscopiques sont modélisées par des moyennes (des espérances mathématiques) relatives au système des particules.

Cette discipline est intimement liée aux concepts fondamentaux de la thermodynamique (entropie notamment) ainsi qu'à la théorie de l'information. Son premier objet a été la théorie cinétique des gaz (Boltzmann, Maxwell, Gibbs).

Considérons un système physique  $X$  constitué de  $N$  particules  $x_i$  « distinguables » :  $X = \{x_i\} \leq i \leq N$ . Sa dynamique est régie par la loi de Newton ( $\forall i m_i \mathbf{a}_i = \mathbf{F}_i$ ), et l'état microscopique des  $N$  particules est représenté dans un espace abstrait appelé « espace des phases ». Mais certaines grandeurs macroscopiques attachées à ce système (température, pression, aimantation, ...) ne dépendent pas « en détail » de son état microscopique. Il est donc possible de calculer ces grandeurs sans connaître l'état microscopique.

En mécanique statistique, on formule les postulats suivants :

1. Hypothèse ergodique : les grandeurs « moyennes temporelles » (correspondant aux mesures physiques qui s'effectuent durant un temps très long à l'échelle des mouvements microscopiques) sont égales aux grandeurs moyennes calculées sur tous les états du système selon une loi de probabilité (dénommée « ensemble ») qui attribue la probabilité  $p_i$  à chaque état  $i$  du système ( $1 \leq i \leq N$ ).

2. Principe du maximum d'entropie : la loi de probabilité des états du système est d'entropie  $S(X) = -\sum_i p_i \ln(p_i)$  maximale.

La maximisation de l'entropie s'effectue sous des contraintes reliées à l'énergie des états. Si on considère un système isolé de particules toutes de même énergie  $E_0$ , en nombre  $N = \Omega(E_0)$ ,

on trouve que tous les états sont équiprobables et donc  $\forall i p_i = \frac{1}{\Omega(E_0)} = \exp(-S(X))$  : la distribution de probabilité obtenue s'appelle « ensemble microcanonique ».

Si on considère maintenant un système isolé de  $N$  particules et si on attribue une énergie  $E(i)$  à chaque état du système, l'énergie moyenne du système est  $\langle E \rangle = \sum_i p_i E(i)$ . La maximisation de l'entropie du système sous les deux contraintes, la contrainte triviale  $\sum_i p_i = 1$  et la

contrainte  $\langle E \rangle$  fixée, conduit à une loi de probabilité de la forme  $p_i = \frac{1}{Z_\beta} \exp(-\beta E(i))$  : la

distribution obtenue s'appelle « ensemble canonique » ou encore « distribution de Gibbs ». Le paramètre  $\frac{1}{\beta}$  est identifié à  $kT$ , où  $k$  est la constante de Boltzmann et  $T$  la température (en

degrés Kelvin) du système, et le coefficient  $Z_\beta = \sum_u \exp(-\beta E(u))$  s'appelle la fonction de partition. C'est à partir de cette fonction que l'on peut calculer les grandeurs macroscopiques : par exemple l'énergie  $\langle E \rangle$  est égale à l'opposé de la dérivée partielle de  $Z_\beta$  par rapport à  $\beta$ .

Il faut ajouter que ces distributions sont établies dans le cadre de la mécanique classique et que d'autres distributions interviennent dans le cadre de la mécanique quantique (statistique de Fermi-Durac, statistique de Bose-Einstein).

C'est la distribution de Gibbs qui est à la base de la distribution de Maxwell-Boltzmann des vitesses des molécules dans un gaz. Dans ce cas, l'énergie considérée est l'énergie cinétique. Le problème est un peu plus compliqué car l'énergie varie continûment ; il faut considérer des densités de probabilité et remplacer les sommes par des intégrales. Pour une particule de

masse  $m$ , l'état est défini par sa vitesse  $\mathbf{v} = (v_x, v_y, v_z)$  et on peut montrer que la densité de probabilité vérifie une formule analogue :  $f(\mathbf{v}) = \frac{1}{Z_\beta} \exp(-\beta u)$  où  $u = \frac{1}{2} m \mathbf{v}^2 = \frac{1}{2} m (v_x^2 + v_y^2 + v_z^2)$ .

En reportant l'expression de  $Z_\beta$  (non donnée ici) et celle de  $u$  dans  $f(\mathbf{v})$ , on obtient la densité de la loi de Maxwell-Boltzmann relative à la racine de la somme des carrés de 3 v.a. normales de variance  $\sigma^2 = \frac{1}{\beta m} = \frac{kT}{m}$ . On peut ensuite utiliser cette distribution pour calculer des grandeurs statistiques et établir des relations : la vitesse moyenne d'une particule est liée à la température du système par  $\langle \mathbf{v}^2 \rangle = \frac{3kT}{m}$ , le produit de la pression par le volume vaut  $nkT$ , etc.

## médiale

(*medial [value]*)

Indicateur de tendance centrale attaché à une variable aléatoire réelle  $X$ , en principe positive, et qui représente généralement un « bien » susceptible d'être « possédé » par des individus. La médiale est la (ou une des ...) valeurs qui partage la distribution (en probabilités) ou la série des valeurs (en statistique) en deux parties qui « accumulent » la même quantité du bien (50 % de la quantité totale). Elle se « lit » de façon très simple sur la *courbe de concentration de Lorentz*.

## médiane

(*median [value]*)

Indicateur de tendance centrale attaché à une variable aléatoire réelle. La médiane est la (ou une des ...) valeurs qui partage la distribution (en probabilités) ou la série des valeurs (en statistique) en deux parties de même probabilité (0,5) ou de même effectif (50 % de l'effectif total). Elle se note le plus souvent  $M$  ou  $Me$ . L'information qu'elle fournit peut être complétée par les quartiles, les déciles, etc.

Si la signification concrète de la médiane est simple et « parlante », la traduction formelle est plus délicate.

### Formules (calcul des probabilités)

Si la v.a. réelle  $X$  est *discrète*, caractérisée par l'ensemble (fini ou dénombrable) de valeurs  $\{x_i\}$ , avec les probabilités ponctuelles  $p_i = P(X = x_i)$ , la médiane est toute valeur  $M$  telle que :

$$\sum_{x_i < M} p_i < 0,5 \quad \text{et} \quad \sum_{x_i > M} p_i < 0,5 .$$

Le plus souvent, cela caractérisera une valeur unique  $x_k$  pour laquelle on aura simultanément  $\sum_{x_i < x_k} p_i < 0,5$  et  $\sum_{x_i \leq x_k} p_i > 0,5$ . Il pourra arriver exceptionnellement qu'il existe deux valeurs consécutives  $x_k$  et  $x_{k+1}$  telles que l'on ait exactement  $\sum_{x_i \leq x_k} p_i = 0,5$  et  $\sum_{x_i \geq x_{k+1}} p_i = 0,5$ . Dans ce cas on peut prendre pour médiane toute valeur comprise entre  $x_k$  et  $x_{k+1}$  (dans la pratique, on prend le plus souvent le milieu ou une valeur « ronde »).

Si  $X$  est *absolument continue*, caractérisée par la densité de probabilité  $f(x)$ , la médiane  $M$  est définie par :

$$\int_{-\infty}^M f(x) dx = 0,5$$

(Cela caractérise une valeur unique si la densité ne s'annule pas sur un intervalle au centre de la distribution – circonstance exceptionnelle dans la pratique).

**Formules (statistique)**

Si la série statistique présente des observations individualisées en nombre impair  $2m + 1 : x_1 < x_2 < \dots < x_{2m+1}$ , la médiane M est la valeur centrale :

$$M = x_{m+1}.$$

Si la série statistique présente des observations individualisées en nombre pair  $2m : x_1 < x_2 < \dots < x_{2m}$ , on peut prendre pour médiane M est toute valeur de l'intervalle central :

M au choix entre  $x_m$  et  $x_{m+1}$ .

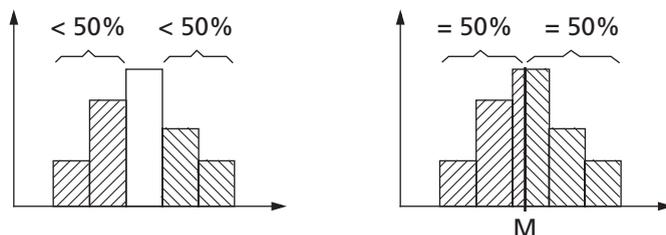
(le choix est le plus souvent le milieu de cet intervalle central :  $M = \frac{x_m + x_{m+1}}{2}$ ).

Si la série statistique présente  $n$  observations individuelles regroupées selon  $k$  classes d'effectif  $n_j$  pour la classe  $]a_j, a_{j+1}[$ , il peut exister (cas exceptionnel) une valeur  $a_k$  telle que l'effectif des classes jusqu'à  $a_k$  soit exactement la moitié de l'effectif total (et de même pour l'effectif des classes au-delà de  $a_k$ ), alors :

$$M = a_k.$$

Si la série statistique présente  $n$  observations individuelles regroupées selon  $k$  classes d'effectif  $n_j$  pour la classe  $]a_j, a_{j+1}[$ , et s'il existe (cas général) une classe  $]a_k, a_{k+1}[$  telle l'effectif des classes avant  $a_k$  soit inférieur à la moitié de l'effectif total, et l'effectif des classes après  $a_{k+1}$  soit inférieur à la moitié de l'effectif total, alors :

- d'une part on dit que  $]a_k, a_{k+1}[$  est la classe médiane,
- d'autre part, si la variable est continue, on peut prendre pour valeur ponctuelle M de la médiane la valeur qui partage l'histogramme en deux parties d'aire égale (on rappelle que la caractéristique de l'histogramme est d'avoir les aires proportionnelles aux effectifs).



Calcul de la valeur médiane (par « interpolation linéaire ») : la classe médiane étant  $]a_k, a_{k+1}[$ , on note  $F_{k-1}$  la fraction d'effectif cumulée jusqu'à la classe précédente (donc jusqu'à  $a_k$ ) et  $f_k$  la fraction d'effectif de la classe médiane. Alors :

$$M = a_k + (a_{k+1} - a_k) \left( \frac{0,5 - F_{k-1}}{f_k} \right).$$

**médiane (test de la – de Mood)**

(Mood median test)

Test non paramétrique rapide qui compare deux distributions numériques par un khi-deux à 4 cases : on mêle les deux échantillons numériques observés, on calcule la médiane globale M et on porte comme effectifs observés le nombre de valeurs supérieures/inférieures à M pour le premier puis pour le deuxième échantillon.

**médiane (test de la valeur d'une)**

Voir signes (test des).

## mémoire

« Le hasard n'a pas de mémoire » (Joseph Bertrand, repris par Émile Borel). S'il se trouve (par hasard !) que l'on a obtenu 10 fois Pile en lançant 10 fois de suite une pièce (« honnête » *i.e.* non faussée !), Face n'a toujours qu'une chance sur deux de sortir au 11<sup>e</sup> lancer, pas plus (même pas un tout petit peu plus !!!).

## mesure de probabilité

(*probability measure*)

Pour être une mesure de probabilité, une application définie sur l'ensemble des évènements d'un espace probabilisable  $(\Omega, \mathcal{A})$  et à valeurs réelles doit vérifier des propriétés sur les valeurs prises, et être « additive ». À un niveau élémentaire, on limite la vérification de l'additivité à la réunion finie.

On dit qu'une application  $P : \mathcal{A} \rightarrow \mathbf{R}$ , définie sur l'ensemble des évènements associés à un espace fondamental  $\Omega$  est une mesure de probabilité si elle satisfait les trois propriétés (ou « axiomes ») :

- pour tout  $A \in \mathcal{A}$ ,  $0 \leq P(A) \leq 1$  ;
- $P(\Omega) = 1$  ;
- $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ .

### ► Commentaires

1. «  $0 \leq P(A) \leq 1$  » signifie que les probabilités sont toujours positives, et inférieures ou égales à 1.
2. «  $P(\Omega) = 1$  » est une autre manière de dire que l'unité des probabilités est la mesure de  $\Omega$  dont on a noté, qu'en tant qu'évènement, il est l'« évènement certain ».
3. «  $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$  » signifie que, si deux évènements sont disjoints, leurs probabilités s'ajoutent. C'est très exactement la caractéristique de toute « mesure » (de longueur, d'aire, de masse, ...).

À un niveau plus approfondi, il faut considérer la réunion infinie dénombrable, et vérifier la «  $\sigma$ -additivité ».

### ► Définition formelle complète

On dit qu'une application  $P : \mathcal{A} \rightarrow \mathbf{R}$ , définie sur l'ensemble de évènements associés à un espace fondamental  $\Omega$  est une mesure de probabilité si elle satisfait les trois axiomes :

1. pour tout  $A \in \mathcal{A}$ ,  $0 \leq P(A) \leq 1$  ;
2.  $P(\Omega) = 1$  ;
3. si  $(A_n) (n = 1, 2, \dots)$  est une suite d'éléments de  $\mathcal{A}$ , deux à deux disjoints, alors :

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Voir *additivité*.

## méthodologie des tests d'hypothèses

Voir *tests d'hypothèses (méthodologie des)*.

## milieu

Voir *centre*.

## modale (classe)

Dans une présentation où les classes d'une distribution statistique sont d'égale amplitude, classe d'effectif maximal.

Voir *mode*.

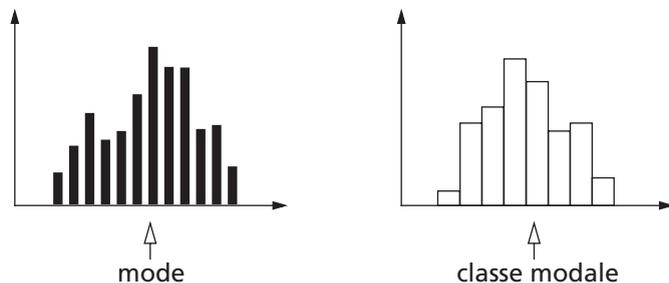
## modalité

Valeur que peut prendre une variable aléatoire ou une variable statistique (ou caractère) dans les cas où cette valeur est élément d'un ensemble sur lequel on ne peut pas effectuer d'opération mathématique (addition notamment).

## mode

(*mode, modal value*)

Le mode d'une série statistique est en principe la valeur la plus fréquente, *i.e.* la valeur de la variable pour laquelle l'effectif est maximal. Lorsque la variable est discrète et ne peut prendre qu'un nombre restreint de valeurs, le mode est « très bien » défini et c'est une caractéristique de tendance centrale intéressante. Lorsque la variable est discrète mais à valeurs très nombreuses, ou bien continue, la présentation se fait avec des regroupements en classes, et le mode est en réalité une classe modale. Dans ce cas, il est obligatoire que les classes soient d'égale amplitude ; en outre, la définition du mode peut varier selon le choix des classes pour le regroupement des valeurs, et il faut vérifier que les classes ne sont ni trop petites (entraînant des irrégularités non significatives) ni trop larges (« lissant » artificiellement la distribution).



On emploie également le mode en probabilités pour désigner la valeur de la variable de probabilité maximale (pour une loi discrète) ou la valeur de densité maximale (pour une loi continue).

Enfin, on peut imaginer que l'on définisse des modes relatifs, maximums locaux de l'effectif. Cette perspective se retrouve dans les adjectifs *unimodale*, *bimodale* et *plurimodale*, utilisés pour qualifier des distributions présentant, soit un unique maximum, soient plusieurs maximums importants et nettement séparés (le cas bimodal par exemple est souvent l'indice que la série statistique a été constituée par un échantillonnage portant, volontairement ou non, sur deux populations différentes et mélangées).

## modèle linéaire [gaussien]

(*linear model*)

Méthode de représentation géométrique et d'analyse statistique d'une situation probabiliste ou statistique « bidimensionnelle » ou « multidimensionnelle ». Dans le cas simple à deux variables, la situation décrite et analysée par le modèle linéaire est celle d'un ensemble  $\{x_i\}$  de valeurs « maîtrisées » ou « contrôlées » d'une variable  $x$  et, pour chaque  $i$ , une variable aléatoire  $Y_i$  « dépendante ». Le modèle linéaire ajuste la liaison entre les deux variables par

une relation  $Y_i = a + bx_i + E_i$ , où chaque  $E_i$  est une variable aléatoire « écart » (*a priori* gaussienne). Le problème fondamental est de rechercher les valeurs des coefficients numériques  $a$  et  $b$  qui minimisent – en un sens approprié de ce mot – globalement les écarts  $E_i$ .

Dans le cas multiple à  $p + 1$  variables ( $p \geq 2$ ), il y a  $p$  variables « contrôlées »  $x_j$ , de valeurs  $\{x_{ij}\}$ , et la liaison entre les  $p$  variables contrôlées  $x_j$  et la  $(p + 1)$ -ième variable dépendante est ajustée par une relation  $Y_i = a + b_1x_{i1} + \dots + b_px_{ip} + E_i$ , où chaque  $E_i$  est une variable aléatoire « écart » (*a priori* gaussienne). Le problème fondamental est de rechercher les valeurs des coefficients numériques  $a, b_1, \dots, b_p$  qui minimisent – en un sens approprié de ce mot – globalement les écarts  $E_i$ .

Voir *droite de régression linéaire*.

## modélisation

(*modelization*)

Démarche intellectuelle par laquelle on représente une situation concrète dans le cadre et avec les « outils » d'une théorie scientifique. Cette démarche a pour but de décrire et de comprendre la situation concrète, et si possible de donner des éléments de prévision sur son évolution.

Comme toute représentation, une modélisation nécessite une simplification de la situation concrète (réduction de la complexité, diminution du nombre de variables et du nombre de paramètres, approximations diverses), et sa pertinence – à tout le moins sa compatibilité – doit dès lors être évaluée.

## moindres carrés (droite des)

Voir *droite des moindres carrés*.

## de Moivre (Abraham)

Mathématicien français, huguenot réfugié en Angleterre (1667–1750). Il démontra le théorème central limite pour la loi binomiale et écrivit *Doctrine of Chances* qui fut pendant près d'un siècle le traité classique du calcul des probabilités (on y trouve notamment l'indépendance et les probabilités conditionnelles).

## moments, moments absolus, moments centrés

(*moments, absolute moments, moments about the mean*)

Indicateurs numériques associés à une variable aléatoire réelle, et qui fournissent de nombreux renseignements sur sa distribution. L'espérance mathématique et la variance sont des moments particuliers (*cf.* ci-dessous). En calcul des probabilités, les moments sont reliés à la *fonction caractéristique*, et en statistique, ils sont notamment utilisés pour définir des *indicateurs* de dispersion et de forme.

Étant donnée une variable aléatoire réelle  $X$  d'espérance mathématique  $\mu$  et un entier  $r \geq 1$ , on définit les moments de  $X$  par les espérances mathématiques suivantes :

- moment (centré en 0) d'ordre  $r$  :  $m_r(X) = E(X^r)$  ;
- moment centré d'ordre  $r$  :  $\mu_r(X) = E((X - \mu)^r)$  ;
- moment absolu d'ordre  $r$  :  $M_r(X) = E(|X|^r)$ .

### ► Cas particuliers

L'espérance mathématique  $\mu$  est le moment  $m_1$ , la variance  $\sigma^2$  est le moment centré  $m_2$  ; par ailleurs, par la définition même de  $\mu$ , le moment centré d'ordre 1 est égal à 0.

Si la variable aléatoire est discrète (donnée par ses valeurs  $\{x_i\}$  et les probabilités ponctuelles  $p_i = P(X = x_i)$ ), les moments se calculent par les sommes :

$$\sum_{x_i} x_i^r p_i, \quad \sum_{x_i} (x_i - \mu)^r p_i, \quad \sum_{x_i} |x_i|^r p_i.$$

Si la variable aléatoire est absolument continue (donnée par sa densité de probabilité  $f(x)$ ), les moments se calculent par les intégrales :

$$\int_{-\infty}^{+\infty} x^r f(x) dx, \quad \int_{-\infty}^{+\infty} (x - \mu)^r f(x) dx, \quad \int_{-\infty}^{+\infty} |x|^r f(x) dx.$$

Il faut remarquer qu'il n'y a pas de certitude que les sommes convergent, si l'ensemble des valeurs est infini, ou que les intégrales convergent, avant de l'avoir effectivement vérifié. Certaines variables aléatoires n'ont pas de moment si  $r$  dépasse une certaine valeur. Enfin les moments (définis ci-dessus dans le cadre du calcul des probabilités) ont tous un analogue statistique, qui se calcule en adaptant de façon évidente les formules avec les sommes.

## moments factoriels

(factorial moments)

Indicateurs numériques associés à une variable aléatoire réelle, et qui fournissent des renseignements sur sa distribution. Dans le cas particulier où la variable est à valeurs entières positives, ils sont reliés à la *fonction génératrice*.

Étant donnée une variable aléatoire réelle  $X$  et un entier  $r \geq 1$ , on définit le *moment factoriel* d'ordre  $r$  de  $X$  comme l'espérance mathématique :

$$E(X(X-1) \dots (X-r+1))$$

## moments (fonction génératrice des)

Voir *génératrice des moments (fonction)*.

## moustaches (boîte à)

Voir *boîte de dispersion*.

## moyenne

(mean [value])

La moyenne, telle qu'elle est définie de façon standard en statistique, est adaptée à l'étude des grandeurs à « sensibilité » additive (comprendre : grandeurs qui sont structurellement des sommes de grandeurs « élémentaires », ou bien qui ont un comportement « remarquable » par addition). Mais ce n'est pas le cas de toutes les grandeurs. Par exemple, de nombreuses grandeurs positives (la taille ou le poids en biologie, le revenu ou la production en économie, etc.) ont une « sensibilité » multiplicative, attestée par exemple par la significativité du concept de « taux » de croissance. Dans ce cas, la moyenne géométrique sera le concept le mieux adapté. Il est donc utile de définir les différents types de moyenne que l'on peut être amené à utiliser.

### ► Moyenne arithmétique

- de 2 nombres  $x$  et  $y$  :

$$m_A = \frac{x+y}{2}$$

- de  $n$  nombres  $x_1, x_2, \dots, x_n$  :

$$m_A = \frac{x_1 + x_2 + \dots + x_n}{n}$$

► Moyenne géométrique

- de 2 nombres positifs  $x$  et  $y$  :

$$m_G = \sqrt{xy}$$

- de  $n$  nombres positifs  $x_1, x_2, \dots, x_n$  :

$$m_G = (x_1 x_2 \dots x_n)^{1/n}$$

► Moyenne harmonique

- de 2 nombres non nuls  $x$  et  $y$  :

$$m_H \text{ définie par } \frac{2}{m_H} = \frac{1}{x} + \frac{1}{y}$$

- de  $n$  nombres non nuls  $x_1, x_2, \dots, x_n$  :

$$m_H \text{ définie par } \frac{n}{m_H} = \frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}$$

► Moyenne quadratique

- de 2 nombres  $x$  et  $y$  :

$$m_Q = \sqrt{\frac{x^2 + y^2}{2}}$$

- de  $n$  nombres  $x_1, x_2, \dots, x_n$  :

$$m_Q = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}}$$

► Inégalités

Lorsque les moyennes sont toutes définies, on a :

$$m_H \leq m_G \leq m_A \leq m_Q$$

(inégalités dès qu'il existe deux nombres différents, égalités si et seulement si tous les  $x_i$  sont égaux)

**Exemple** Si un véhicule roule pendant le temps  $t$  à la vitesse  $v_1$ , et ensuite pendant le même temps  $t$  à la vitesse  $v_2$ , sa vitesse moyenne est la moyenne arithmétique de  $v_1$  et de  $v_2$ . Si un véhicule roule sur une distance  $x$  à la vitesse  $v_1$ , et ensuite sur la même distance  $x$  à la vitesse  $v_2$ , sa vitesse moyenne est la moyenne harmonique de  $v_1$  et de  $v_2$ .

On peut si nécessaire généraliser ces formules pour effectuer des moyennes pondérées.

## moyenne d'un échantillon statistique

Dans une situation d'observation d'un échantillon statistique, la moyenne est le principal indicateur numérique de tendance centrale. Parfois qualifiée de *moyenne observée* ou de *moyenne empirique* ou encore de *moyenne statistique*, elle est définie comme la moyenne numérique des valeurs observées.

La moyenne d'un échantillon d'une variable aléatoire  $X$  se note le plus souvent  $m_X$  ou  $m_x$  (ou  $m$  ou  $m_{\text{obs}}$  s'il n'y a aucun risque de confusion), ou encore  $\bar{x}$ .

Formule pour  $n$  observations individualisées  $x_1, x_2, \dots, x_n$  :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{i=n} x_i.$$

Formule pour  $n$  observations individuelles regroupées selon  $k$  classes d'effectif  $n_j$  pour la valeur  $\xi_j$  :

$$\bar{x} = \frac{n_1 \xi_1 + n_2 \xi_2 + \dots + n_k \xi_k}{n} = \sum_{j=1}^{j=k} f_j \xi_j \quad \text{où } f_j = \frac{n_j}{n}$$

Lorsque les classes sont des intervalles  $]a_j, a_{j+1}]$ , la valeur « typique »  $\xi_j$  qui est utilisée dans la dernière formule est celle du centre  $\frac{a_j + a_{j+1}}{2}$  de la classe.

### moyenne d'un échantillon de variables aléatoires

Dans une situation probabiliste d'épreuves répétées, avec un échantillon  $(X_1, X_2, \dots, X_n)$  d'une variable aléatoire numérique, la moyenne est est la *variable aléatoire* :

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

### moyenne mobile

(moving average)

Étant donné une *série chronologique* à intervalles de temps réguliers  $(x_1, x_2, \dots, x_n, \dots)$ , on appelle moyennes mobiles d'ordre  $k = 2p + 1$  impair les moyennes arithmétiques de  $k$  valeurs consécutives. Ces moyennes sont généralement rapportées au temps médian, de sorte que l'on a :

$$m_{2p+1}(t) = \frac{1}{2p+1} \sum_{i=-p}^p x_{t+i}$$

Cette définition peut être étendue au cas  $k = 2p$  pair, par exemple en comptant pour  $\frac{1}{2}$  les valeurs extrêmes :

$$m_{2p}(t) = \frac{1}{2p} \left( \frac{1}{2} x_{t-p} + \sum_{i=-(p-1)}^{p-1} x_{t+i} + \frac{1}{2} x_{t+p} \right)$$

Les moyennes mobiles sont utilisées pour « lisser » les courbes en atténuant l'effet des fluctuations accidentelles. Elles peuvent aussi être utilisées pour « corriger » globalement la série chronologique des « variations saisonnières ».

### moyennes (test de comparaison de)

Voir *Student (test de)*.

### multinomiale (loi)

(multinomial distribution)

Loi d'une variable aléatoire discrète de « multi-compte » à  $k$  dimensions (vecteur aléatoire) – ou d'un ensemble ordonné de  $k$  variables aléatoires discrètes – qui intervient dans les tirages *sans remise* à  $k$  éventualités.

#### Formulaire

Un paramètre entier positif  $k$  qui est la « dimension » ;  $k$  paramètres réels (probabilités des  $k$  éventualités) :  $p_1, p_2, \dots, p_k \in [0, 1]$  et qui vérifient  $p_1 + p_2 + \dots + p_k = 1$  ; un paramètre entier positif  $n$  qui est le nombre d'épreuves.

Deux présentations possibles de la v.a. multinomiale :

- un vecteur aléatoire  $(N_1, N_2, \dots, N_k)$  à composantes entières ; valeurs prises : des «  $k$ -uplets »  $(n_1, n_2, \dots, n_k)$  d'entiers vérifiant  $n_1 + n_2 + \dots + n_k = n$  ;
- un ensemble ordonné  $(N_1, N_2, \dots, N_k)$  de  $k$  variables aléatoires entières ; valeurs prises : les même  $k$ -uplets  $(n_1, n_2, \dots, n_k)$ , appelés parfois « multi-entiers ».

Deux notations possibles :

- $P((N_1, N_2, \dots, N_k) = (n_1, n_2, \dots, n_k))$
- $P(N_1 = n_1 \text{ et } N_2 = n_2 \text{ et } \dots \text{ et } N_k = n_k)$

► Loi de probabilité

$$P(N_1 = n_1 \text{ et } N_2 = n_2 \text{ et } \dots \text{ et } N_k = n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

(nota : le coefficient  $\frac{n!}{n_1! \dots n_k!}$  s'appelle le coefficient multinomial et possède la

notation  $\binom{n}{n_1 \dots n_k}$ )

► Valeurs caractéristiques

- espérance :  $E(N_j) = np_j$
- variance :  $\text{Var}(N_j) = np_j(1 - p_j)$
- écart-type :  $\sigma(N_j) = \sqrt{np_j(1 - p_j)}$
- covariance ( $j \neq h$ ) :  $\text{Cov}(N_j, N_h) = -np_j p_h$
- coefficient de corrélation ( $j \neq h$ ) :  $\rho(N_j, N_h) = -\sqrt{\frac{p_j p_h}{(1 - p_j)(1 - p_h)}}$

► Cas particulier

Lorsque  $k = 2$ , on obtient la *loi binomiale* (avec les notations  $p, q$  au lieu de  $p_1, p_2$ , et  $k, n - k$  au lieu de  $n_1, n_2$ ).

► Utilisations

La loi multinomiale est la loi du  $k$ -uplet variable de « multi-compte » de  $k$  caractères dans des épreuves répétées – ou dans des tirages « AVEC remise ». On peut faire une présentation à partir d'une variable aléatoire  $X$  pouvant prendre  $k$  valeurs, codées 1, 2, ...,  $k$  avec les probabilités  $p_j = P(X = j)$  : les  $n$  épreuves répétées sont représentées par l'échantillon  $(X_1, X_2, \dots, X_n)$ , les  $k$  composantes de la v.a. multinomiale sont définies par  $N_j = \text{Card} \{i | X_i = i\}$ .

Chaque  $N_j$  considérée isolément est une v.a. binomiale de paramètres  $n$  et  $p_j$ .

**Exemple** On lance 12 fois un dé. Quelle est la probabilité  $P$  d'obtenir exactement deux fois 1, deux fois 2, ... deux fois 6 ? On a  $k = 6, p_1 = \dots = p_6 = \frac{1}{6}, n = 12, n_1 = n_2 = \dots = n_6 = 2$ ,

$$\text{d'où } P = \frac{12!}{2! \dots 2!} \left(\frac{1}{6}\right)^{12} = 0,00344.$$

## multinormale (loi)

Voir *normale à  $p$  dimensions (loi)*.

## multiple (coefficient de corrélation)

Voir *corrélation multiple (coefficient de)*.

## multirégression

Voir *droite de régression, régression à deux variables explicatives, corrélation multiple (coefficient de)*.



## **négation logique**

Voir *complémentaire*.

## **Neyman-Pearson (lemme de, théorème de) (Neyman-Pearson lemma)**

Théorème qui permet de construire le « meilleur » test d'hypothèse lorsque  $H_0$  et  $H_1$  sont des hypothèses simples. Étant donné une loi de densité  $L(\theta ; x)$ , un niveau  $\alpha$  (probabilité de rejet à tort de  $H_0$ ), et un échantillon  $(X_1, X_2, \dots, X_n)$ , le théorème de Neyman-Pearson permet de définir une région  $\mathcal{R}$  de  $\mathbf{R}^n$  qui est une « région critique optimale » pour le test de  $H_0 = \langle \theta = \theta_0 \rangle$  contre  $H_1 = \langle \theta = \theta_1 \rangle$ . Soit  $1 - \beta$  la puissance du test (probabilité de rejet justifié de  $H_0$ ). Alors, en conséquence du théorème, le test avec la région  $\mathcal{R}$  est le plus puissant, il est sans biais (notion spécifique aux tests d'hypothèse qui signifie que  $1 - \beta > \alpha$ ), et il est convergent ( $1 - \beta \rightarrow 1$  lorsque  $n \rightarrow \infty$ ).

## **niveau**

(level)

Mot ambigu qui est utilisé aussi bien dans l'expression niveau de risque (notation traditionnelle  $\alpha$ ) que dans l'expression niveau de confiance (notation traditionnelle  $1 - \alpha$ ). Employé de façon absolue, ce mot désigne la probabilité de risque de rejet à tort de l'hypothèse  $H_0$  (appelé aussi risque de première espèce) pour un test d'hypothèse.

## **nombre de degrés de liberté**

Voir *degrés de liberté*.

## **nominal, e**

Se dit parfois d'une variable statistique (ou caractère) qualitative.

## **non paramétrique (test)**

(non-parametric test)

Voir *test d'hypothèse*.

## **normale (loi)**

(gaussian distribution, normal distribution)

Synonymes *Gauss (loi de)*, *Laplace-Gauss (loi de)*.

Loi de la variable aléatoire continue qui tient la place centrale dans le calcul des probabilités.

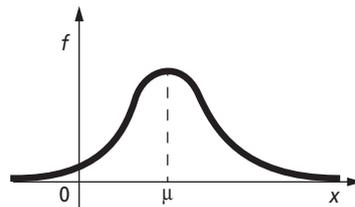
### Formulaire

Deux paramètres réels :  $\mu \in \mathbf{R}$  et  $\sigma \in \mathbf{R}_+^*$  ; valeurs sur les réels.

#### ► Loi de probabilité

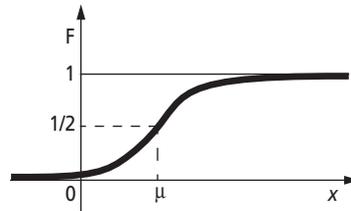
densité

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



fonction de répartition

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



► Valeurs caractéristiques

- espérance :  $E(X) = \mu$
- variance :  $\text{Var}(X) = \sigma^2$
- écart-type :  $\sigma(X) = \sigma$

**Cas particulier fondamental pour la théorie  
et pour les tables numériques : variable normale centrée réduite**

► Loi de probabilité

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

► Valeurs caractéristiques

- espérance :  $E(X) = 0$
- variance :  $\text{Var}(X) = 1$
- écart-type :  $\sigma(X) = 1$

► Utilisations

En théorie, la loi normale est la loi limite de la somme (ou de la moyenne) dans une suite infinie d'épreuves répétées – c'est également la loi limite de la somme (ou de la moyenne) d'une suite infinie de variables aléatoires vérifiant des conditions « raisonnables ».

Dans la pratique, la loi normale est une approximation de la somme – ou « résultante additive » – de grandeurs aléatoires petites et nombreuses, et pas trop mutuellement dépendantes...

La loi normale est en outre utilisée comme approximation de la « résultante mutiplicative » de grandeurs aléatoires positives nombreuses, lorsque l'espérance est grande devant l'écart-type (à partir par exemple de  $\mu > 4\sigma$ ).

La loi normale possède une propriété extrêmement importante, la décroissance rapide de la probabilité en fonction de l'écart. Quelques valeurs numériques ( $\mathcal{N}$  v.a. normale centrée réduite) montrent cette décroissance :

$$\begin{aligned} P(|\mathcal{N}| > 3) &= 4,55 \times 10^{-2}, \\ P(|\mathcal{N}| > 3) &= 2,70 \times 10^{-3}, \\ P(|\mathcal{N}| > 4) &= 6,34 \times 10^{-5}, \\ P(|\mathcal{N}| > 5) &= 5,72 \times 10^{-7}, \\ P(|\mathcal{N}| > 6) &= 1,97 \times 10^{-9}. \end{aligned}$$

Une première conséquence est que dans la pratique on peut faire comme si le dépassement de 4 ou 5 écarts-type était quasiment impossible. On peut indiquer, en anticipant sur la statistique, une seconde conséquence : lorsque pour l'analyse d'une situation ayant conduit à une hypothèse, l'observation se trouve à plus de 4 ou 5 écarts-type de ce qui se passerait si l'hypothèse était vraie, alors celle-ci peut être rejetée sans aucune précaution de langage ni état d'âme.

**Théorème d'addition.** La somme de deux variables aléatoires normales indépendantes  $\mathcal{N}(\mu_1, \sigma_1)$  et  $\mathcal{N}(\mu_2, \sigma_2)$  est une variable aléatoire normale  $\mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ .

Ce résultat est fondamental dans toute la théorie du calcul des probabilités.

### Calculs numériques

#### ► Réduction

Si  $X$  suit une loi normale  $\mathcal{N}(\mu, \sigma)$  et si  $\mathcal{N}$  suit la loi normale centrée réduite  $\mathcal{N}(0, 1)$ , on a

$$P(a < X \leq b) = P\left(\frac{a - \mu}{\sigma} < \mathcal{N} \leq \frac{b - \mu}{\sigma}\right),$$

ce qui permet de ramener tout calcul relatif à des valeurs de  $\mathcal{N}(\mu, \sigma)$  à des calculs sur les valeurs de  $\mathcal{N}$ , que l'on peut trouver dans les tables.

#### ► Tables de la loi normale

Si  $F$  est la fonction de répartition de la loi normale centrée réduite  $\mathcal{N}$ , on trouve des tables notamment de  $F(x) = P(\mathcal{N} \leq x)$ , de  $2(1 - F(x)) = P(|\mathcal{N}| > x)$ , et de la fonction réciproque de cette dernière : écart réduit absolu  $x$  en fonction de sa probabilité de dépassement  $\alpha$  (soit donc  $P(|\mathcal{N}| > x) = \alpha$ ).



**Exemple 1** On contrôle la justesse d'une balance en effectuant la pesée d'un objet-test de 250 g exactement. Si la balance est juste, le résultat doit être représenté par une variable aléatoire  $X$  de loi normale  $\mathcal{N}(250, 0,6)$ . Quelle est la probabilité que le résultat de la pesée dépasse 251 g ?

Il faut calculer  $P(X > 251) = P\left(\frac{X - 250}{0,6} > \frac{251 - 250}{0,6} = 1,667\right)$  où  $\frac{X - 250}{0,6}$  suit une v.a.

normale réduite. Si  $F$  est la fonction de répartition d'une v.a. normale réduite, la probabilité demandée est  $1 - F(1,667) = 0,048$ .

**Exemple 2** On jette 1 000 fois une pièce « honnête », quelle est la probabilité que le nombre de Piles soit compris entre 490 et 510 ?

La vraie loi du nombre  $X$  de Piles est une loi binomiale  $\mathcal{B}(1\,000, 0,5)$ , qu'il est raisonnable d'approcher par une loi normale de même espérance et de même écart-type, soit  $\mathcal{N}(500, 15,81)$ . On pourrait calculer  $P(490 \leq X \leq 510)$  mais on commettrait une légère erreur :  $X$  ne prend que des valeurs entières et il faut, dans l'approximation par la loi normale qui est continue, étaler une valeur entière  $k$  entre  $k - 0,5$  et  $k + 0,5$ . Soit donc

$$\begin{aligned} P(489,5 < X < 510,5) &= P\left(\frac{489,5 - 500}{15,81} < \frac{X - 500}{15,81} < \frac{510,5 - 500}{15,81}\right) \\ &= P(-0,664 < \mathcal{N} < 0,664). \end{aligned}$$

On peut prendre une table de la fonction de répartition  $F$  et calculer  $F(0,664) - F(-0,664) = 0,493$ , ou bien, comme les valeurs sont symétriques par rapport à 500, prendre une table de l'écart absolu (le résultat numérique est le même !).

**normale à 2 dimensions (loi)** (2-dimensional gaussian distribution)

Loi d'un vecteur aléatoire à deux dimensions (X, Y) qui possède la propriété fondamentale suivante : pour tout a réel et tout b réel, la combinaison linéaire aX + bY est une v.a. normale (scalaire).

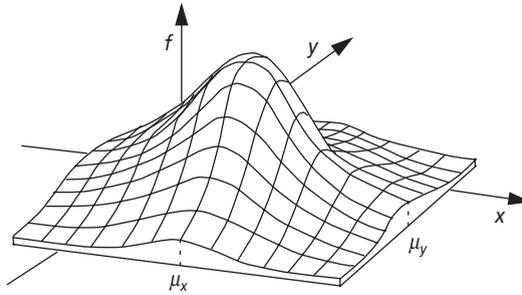
**Formulaire**

Cinq paramètres réels :  $\mu_X, \mu_Y \in \mathbf{R}$ ,  $\sigma_X, \sigma_Y \in \mathbf{R}_+^*$ ,  $\rho \in ]-1, 1[$ .

nota : si  $\rho$ , qui représente le coefficient de corrélation entre les variables « marginales » à 1 dimension, était égal à -1 ou à 1, la loi serait « dégénérée » et sa densité serait nulle sauf sur une droite du plan (x, y).

Valeurs sur les couples de réels (i.e. les vecteurs de  $\mathbf{R}^2$ ).

➤ Loi de probabilité



$$\varphi(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right\}$$

➤ Valeurs caractéristiques

Elles sont représentées de façon standard par :

- le vecteur espérance mathématique  $\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$  ;
- et la matrice de variance-covariance  $\begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$ .

**normale à p dimensions (loi)** (p-dimensional gaussian distribution)

Synonyme *multinormale (loi)*.

Loi d'un vecteur aléatoire à p dimensions (X<sub>1</sub>, X<sub>2</sub>, ..., X<sub>p</sub>) qui possède la propriété fondamentale suivante : pour tous a<sub>1</sub>, a<sub>2</sub>, ..., a<sub>p</sub> réels, la combinaison linéaire a<sub>1</sub>X<sub>1</sub> + a<sub>2</sub>X<sub>2</sub> + ... + a<sub>p</sub>X<sub>p</sub> est une v.a. normale (scalaire).

**Définition**

Il faut utiliser un formalisme vectoriel et matriciel :

- on note  $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{bmatrix}$  le vecteur v.a. normale à p dimensions et  $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{bmatrix}$  le vecteur de ses composantes numériques,

– on note  $\mathbf{E}_{\mathbf{X}} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_p \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \dots \\ E(X_p) \end{bmatrix}$  le vecteur espérance mathématique,

on note enfin  $\mathbf{V}_{\mathbf{X}} = (\text{Cov}(X_j, X_k))_{1 \leq j \leq p, 1 \leq k \leq p}$  la « matrice de variance-covariance » (qui doit être une matrice « définie positive » pour que la loi ne soit pas dégénérée).

La densité de la loi de probabilité de  $\mathbf{X}$  est :

$$\varphi(x_1, x_2, \dots, x_p) = (2\pi)^{-p/2} (\det \mathbf{V}_{\mathbf{X}})^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{E}_{\mathbf{X}}) \mathbf{V}_{\mathbf{X}}^{-1} (\mathbf{x} - \mathbf{E}_{\mathbf{X}}) \right\}.$$

## nuage de points

(*scatter diagram, dispersion diagram*)

Dans une situation de description ou d'analyse statistique d'un échantillon  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$  d'un couple de variables numériques, désigne à la fois l'ensemble de points de l'espace  $\mathbf{R}^2$  formé par l'échantillon, et sa représentation graphique dans un système d'axes.

Cette notion peut se généraliser à un nombre quelconque de dimensions, notamment en analyse des données.

## numérique

(*numerical*)

Se dit des fonctions et des variables aléatoires qui prennent pour valeurs des nombres réels (incluant bien sûr et notamment les valeurs entières).



**observé, e**

(*observed, sample*)

Lorsque cet adjectif n'est pas employé dans son sens général de la langue courante, il est synonyme d'*empirique* et qualifie les paramètres des distributions statistiques (moyenne observée, variance observée, ...), par opposition aux paramètres « théoriques » des lois de probabilités.

**odds, odds ratio**

Expressions anglaises (le mot *odds* signifie cote dans le monde des bookmakers) non traduites en français qui désignent, notamment en épidémiologie et essais cliniques, des concepts concurrents des concepts de probabilité et de risque relatif. Si  $p$  est la probabilité d'un évènement, son « odds » est le quotient  $\frac{p}{1-p}$ . Ainsi, un évènement de probabilité 0,20 aura un

« odds » de 0,25, correspondant à l'expression de « 1 chance contre 4 ». La probabilité varie de 0 à 1, l'« odds » de 0 à l'infini (il est néanmoins très voisin de la probabilité lorsque celle-ci est faible). Étant donné un risque de référence  $p_0$  et un risque  $p$  (par exemple calculé comme

moyenne d'une série d'observations), le risque relatif est  $\frac{p}{p_0}$ , et l'« odd ratio » le quotient de

quotients  $\frac{p/(1-p)}{p_0/(1-p_0)}$  (très voisin du risque relatif lorsque les risques sont faibles).

**opérations sur les variables aléatoires**

Effet sur les indicateurs (espérance mathématique, variance, écart-type) d'un changement de variable affine, *i.e.* d'un changement d'origine (ou translation) et/ou d'échelle (ou homothétie).

Translation de l'origine seule	Changement d'échelle seul	Cas général
$E(X + b) = E(X) + b$ $Var(X + b) = Var(X)$ $\sigma(X + b) = \sigma(X)$	$E(aX) = aE(X)$ $Var(aX) = a^2Var(X)$ $\sigma(aX) =  a  \sigma(X)$	$E(aX + b) = aE(X) + b$ $Var(aX + b) = a^2Var(X)$ $\sigma(aX + b) =  a  \sigma(X)$

Ces formules permettent notamment d'associer à une v.a. réelle quelconque une v.a. centrée

$X - E(X)$ , d'espérance nulle, et une v.a. centrée réduite  $\frac{X - E(X)}{\sigma(X)}$ , d'espérance nulle et

d'écart-type égal à 1.

Voir *somme de deux variables aléatoires*.

**ordinal, e**

Se dit d'une variable aléatoire ou d'une variable statistique (ou caractère) dont les valeurs possibles sont des éléments que l'on peut classer entre eux (auxquels donc on peut attribuer un numéro d'ordre ou « rang »).

**ordre (statistique d')**

(*[rank] order statistic*)

Un échantillon numérique  $(x_1, x_2, \dots, x_n)$  étant donné, désigne l'échantillon des valeurs rangées par ordre croissant.



## paramètre

(parameter)

En mathématique, mot qui qualifie de façon générale un *statut* mathématique de nombre, intermédiaire entre celui de constante ( $2, \pi, \dots$ ) et celui de variable (qui indique aussi bien un choix variable qu'une réelle variation spatiale ou temporelle). Un paramètre est en quelque sorte une variable dont on a « bloqué » la valeur pour étudier un problème, quitte à la « débloquer » ultérieurement.

En calcul des probabilités, ce mot est employé pour désigner les valeurs numériques qui permettent de caractériser complètement une loi de probabilités.

Enfin, ce mot est parfois utilisé comme synonyme d'indicateur numérique ou de valeur caractéristique (d'une distribution). Cet usage, trop ambigu (et trop voisin de l'usage précédent) est à déconseiller.

## paramétrique (test)

(parametric test)

Voir *test d'hypothèse*.

## parente (loi)

Désigne parfois la loi de probabilité commune à toutes les variables aléatoires d'un échantillon.

## parente (population)

Désigne parfois la population dont est issu un échantillon.

## Pareto (loi de)

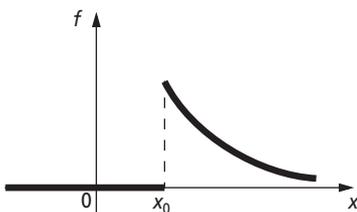
(Pareto distribution)

Loi d'une variable aléatoire positive continue utilisée pour modéliser l'inégalité des richesses, imaginée en 1897 par l'économiste et statisticien italien Vilfredo Pareto.

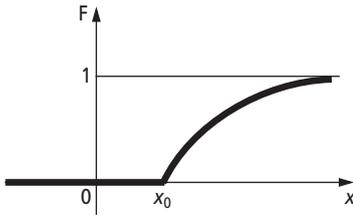
### Formulaire

Deux paramètres réels :  $\alpha \in \mathbf{R}_+^*$  (paramètre « de forme ») ;  $x_0 \in \mathbf{R}_+^*$  (paramètre d'origine et d'échelle).

#### ► Loi de probabilité



$$f(x) = \begin{cases} 0 & \text{si } x < x_0 \\ \frac{\alpha}{x_0} \left(\frac{x_0}{x}\right)^{\alpha+1} & \text{si } x \geq x_0 \end{cases}$$



fonction de répartition

$$F(x) = \begin{cases} 0 & \text{si } x < x_0 \\ 1 - \left(\frac{x_0}{x}\right)^\alpha & \text{si } x \geq x_0 \end{cases}$$

► Valeurs caractéristiques

- espérance :  $E(X) = \frac{\alpha}{\alpha - 1}x_0$  ( $\alpha > 1$ )
- variance :  $\text{Var}(X) = \frac{\alpha}{(\alpha - 1)^2(\alpha - 2)}x_0^2$  ( $\alpha > 2$ )
- écart-type :  $\sigma(X) = \frac{1}{\alpha - 1}\sqrt{\frac{\alpha}{\alpha - 2}}x_0$  ( $\alpha > 2$ )

► Utilisation

Dans son utilisation économique, la v.a. de Pareto  $X$  représente le revenu d'un individu pris au hasard,  $x_0$  étant le revenu minimum ( $F(x) = 1 - \left(\frac{x_0}{x}\right)^\alpha$  est donc la proportion des individus ayant un revenu inférieur ou égal à  $x$ ) ; le paramètre  $\alpha$  est généralement voisin de 2.

Cette loi est également utilisée en décalant l'origine ( $F(x) = 1 - \left(\frac{x_0}{x + x_0}\right)^\alpha$  ( $x \geq 0$ )), de façon à modéliser une situation où le revenu minimal est nul.

En théorie, la variable aléatoire  $Y = \ln\left(\frac{X}{x_0}\right)$  suit une loi exponentielle de paramètre  $-\alpha$ .

Ainsi cette loi, qui modélise une variable continue positive de densité décroissante, peut être très facilement « ajustée » grâce à son logarithme, ce qui pousse à son utilisation comme loi approchée dans des circonstances variées.

## pari

(bet)

Convention entre deux parties (par exemple deux joueurs dans un « jeu d'argent »), qui engage chacune une somme d'argent, et qui se dénouera lors d'une épreuve aléatoire à venir. Selon le résultat de l'épreuve, l'une des parties versera à l'autre la somme convenue. Un pari se modélise par une variable aléatoire « gain » d'une partie (et donc « perte » de l'autre), ces deux mots pouvant désigner des sommes de signe quelconque. On dit que le pari est équitable si l'espérance mathématique de cette variable aléatoire est nulle.

On peut également engager des paris sur des événements qui ne sont pas aléatoires, mais sur lesquels l'information n'est pas complète au moment du pari. Dans pareil cas, la modélisation utilise la théorie probabiliste de l'information.

## partielle (coefficient de corrélation)

Voir *corrélation partielle (coefficient de)*.

## partition

Voir *système complet d'évènements*.

## Pascal (Blaise)

Mathématicien et physicien, puis écrivain français (1623–1662). En répondant à des problèmes posés par le chevalier de Méré à lui-même et à Robertval, il commença l'étude systématique des probabilités. Il entreprit ensuite une correspondance avec Fermat sur la combinatoire et les probabilités et fit également des travaux en géométrie et en hydrostatique.

## Pascal (loi de)

(*Pascal distribution*)

Loi d'une variable aléatoire discrète « temps d'attente » du  $s$ -ième succès dans des épreuves répétées.

### Formulaire

Deux paramètres réels :  $s$  (entier  $\geq 1$ ) qui représente le nombre de succès recherché ;  $p$  ( $0 \leq p \leq 1$ ) qui représente une probabilité (notation standard :  $q = 1 - p$ ).

Soit  $T$  la variable aléatoire de Pascal de paramètres  $s$  et  $p$  ; valeurs prises :  $s, s + 1, s + 2, \dots$

#### ► Loi de probabilité

$$P(T = k) = \binom{k-1}{s-1} p^s q^{k-s}$$

#### ► Valeurs caractéristiques

- espérance :  $E(T) = \frac{s}{p}$
- variance :  $\text{Var}(T) = \frac{sq}{p^2}$
- écart-type :  $\sigma(T) = \frac{\sqrt{sq}}{p}$

#### ► Utilisation

La loi de Pascal est le temps d'attente du  $s$ -ième succès dans des épreuves répétées, ou dans des tirages « AVEC remise » (le premier temps possible est  $T = s$ ).

Une variable aléatoire de Pascal de paramètres  $s$  et  $p$  est la somme de  $s$  variables aléatoires géométriques de paramètre  $p$ .

## Pascal (triangle de)

Voir *triangle de Pascal*.

## Pearson (Karl)

Mathématicien britannique (1857–1936). Il établit la théorie générale de la corrélation et inventa le test du khi-deux.

## permutation

Étant donné un ensemble  $E$  de  $n$  objets, une permutation de  $E$  est une suite (ou un rangement) ordonnée de ces  $n$  objets.

Si l'on suppose les  $n$  objets numérotés de 1 à  $n$ , une permutation est alors représentée par une suite  $(s_1, s_2, \dots, s_n)$ , où les  $s_i$  sont les chiffres de 1 à  $n$  pris dans un ordre qui caractérise la permutation.

Le nombre de permutations de  $n$  objets est  $n!$  (« factorielle »  $n$ ).

**Exemple**  $E = \{a, b, c\}$  : il y a  $3! = 6$  permutations de  $E$ , notées de façon simplifiée et évidente :  $abc, acb, bac, bca, cab, cba$ .

Voir *arrangements sans répétition*.

## plan d'expérience, plan factoriel (sample design, factorial design)

Description du détail d'une expérimentation lorsque l'on étudie l'effet de plusieurs facteurs sur une variable (les facteurs étant le plus souvent supposés agir exclusivement sur l'espérance mathématique). Selon la nature de l'étude et le « coût » unitaire de l'observation, les plans d'expérience seront de types très divers, variant notamment par les combinaisons entre modalités (« niveaux ») des facteurs et le nombre de répétitions.

Voir *analyse de la variance à deux facteurs*.

## plurimodale (distribution)

Voir *mode*.

## Poincaré (formule de)

Formule qui généralise la formule d'additivité :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{1 \leq i \leq n} P(A_i) - \sum_{1 \leq j < k \leq n} P(A_j \cap A_k) + \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n).$$

## Poisson (Simon-Denis)

Mathématicien français (1781–1840). Il étendit plusieurs résultats du calcul des probabilités et fit également des travaux en physique mathématique.

## Poisson (loi de)

(Poisson distribution)

Loi d'une variable aléatoire discrète qui représente le nombre total d'évènements survenus jusqu'à un temps fixé dans un processus de Poisson.

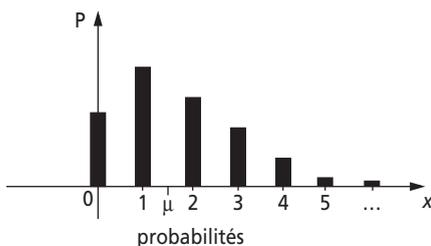
### Formulaire

Un paramètre réel positif  $\mu$ .

Soit  $N$  la variable aléatoire de Poisson  $\mathcal{P}(\mu)$  ; valeurs sur  $\mathbf{N}$ .

► Loi de probabilité

$$P(N = k) = \frac{\mu^k}{k!} e^{-\mu}$$



► Valeurs caractéristiques

- espérance :  $E(N) = \mu$
- variance :  $\text{Var}(N) = \mu$
- écart-type :  $\sigma(N) = \sqrt{\mu}$

► Utilisations

C'est la loi (exacte) du compte des évènements dans un processus poissonnien : si le taux du processus est  $\lambda$ , le compte au temps  $t$  est une loi de Poisson de paramètre  $\mu = \lambda t$ .

La loi de Poisson  $\mathcal{P}(np)$  est une approximation de la loi binomiale  $\mathcal{B}(n, p)$  lorsque le paramètre  $n$  est « grand », le paramètre  $p$  « petit », et que le produit  $np$  (égal au paramètre  $\mu$  de la loi de Poisson) reste dans une zone « moyenne » de valeurs (à apprécier en fonction de  $n$  et de  $p$ ). On traduit souvent cette approximation en disant que la loi de Poisson est la loi de compte des évènements « rares ».

**Théorème d'addition.** La somme de deux variables aléatoires de Poisson *indépendantes*  $\mathcal{P}(\mu_1)$  et  $\mathcal{P}(\mu_2)$  est une variable aléatoire de Poisson  $\mathcal{P}(\mu_1 + \mu_2)$ .

**Exemple 1** (processus poissonnien). On suppose que, dans un process industriel, il se produit en moyenne un incident tous les 3 jours. Quelle est la probabilité  $P$  qu'il y ait au moins 2 incidents dans une journée ?

On peut supposer que la survenue des incidents est un processus de Poisson. Le taux journalier est de  $1/3$ . Si  $X$  est la v.a. de Poisson de paramètre  $\mu = 1/3$ , l'évènement « au moins 2 incidents dans une journée » est le complémentaire de « 0 ou 1 incident dans une

journée », de probabilité  $P(X = 0) + P(X = 1) = \left( \frac{(1/3)^0}{0!} + \frac{(1/3)^1}{1!} \right) e^{-1/3} = \frac{4}{3} e^{-1/3} = 0,955$ .

La probabilité demandée est donc  $P = 1 - 0,955 = 0,045$ .

**Exemple 2** (évènements rares – microbiologie). On désire contrôler la faible concentration en bactéries d'une solution. Pour cela on prend 1 ml de la solution que l'on dilue 20 fois, puis on prélève dans la solution diluée 20 gouttes de 10  $\mu\text{l}$  chacune avec lesquelles on ensemence 20 boîtes de Petri. Au bout de 48 h, on constate que des colonies de bactéries se sont développées dans 14 des 20 boîtes de Petri. Peut-on en déduire la concentration en bactéries de la solution mère ?

Si  $v$  est le nombre moyen de bactéries par goutte de 10  $\mu\text{l}$ , et  $X$  la v.a. nombre de bactéries par goutte,  $X$  suit une loi de Poisson de paramètre  $v$ . On a donc  $P(X = 0) = \frac{v^0}{0!} e^{-v} = e^{-v}$ . Or on

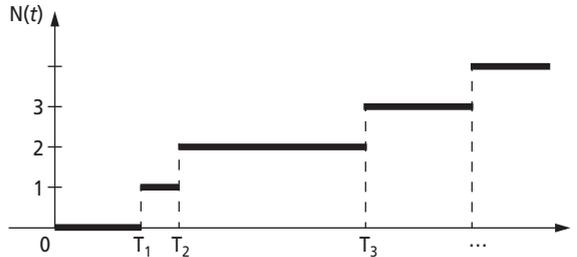
connaît une estimation expérimentale de cette probabilité :  $\frac{6}{20} = 0,3$ . On en déduit  $v = -\ln 0,3 = 1,20$ , puis, en multipliant par 100 (rapport entre 1 ml et 10  $\mu\text{l}$ ) puis par 20 (dilution), la concentration de la solution mère : 2 400 bactéries par ml.

## Poisson (processus de), poissonnien (processus)

(Poisson process, Poisson trials)

Modèle mathématique décrivant et caractérisant la réalisation d'évènements aléatoires indépendants se succédant au cours du temps : des arrivées, des naissances, des pannes ou défaillances ou incidents, des désintégrations, des décès... Parmi tous les processus, celui de Poisson peut se caractériser de diverses façons, dont la principale pour l'utilisateur est d'être « sans vieillissement » (ou « sans mémoire »).

Un processus de Poisson fait intervenir en premier lieu la suite croissante des temps d'arrivée :  $T_1, T_2, \dots, T_n, \dots$ , à laquelle on peut associer les intervalles  $T_1, T_2 - T_1, \dots, T_{n+1} - T_n, \dots$  qui sont des variables aléatoires continues positives.



À cette suite de temps d'arrivée (on posera  $T_0 = 0$  par convention), on associe une « fonction aléatoire » croissante de « comptage »  $N(t)$  qui est, pour chaque  $t \geq 0$ , une variable aléatoire discrète à valeurs entières positives (on dit aussi que  $N(t)$  est une « famille » de variables aléatoires) :  $N(t)$  fait un saut de 1 chaque fois que  $t$  franchit une valeur  $T_n$  (ainsi  $N(t) = n$  pour  $t \in [T_n, T_{n+1}[$ ).

Vu du point de vue des intervalles entre les temps d'arrivée, le processus de Poisson est caractérisé par deux propriétés, d'une part que les v.a.  $T_1, T_2 - T_1, \dots, T_{n+1} - T_n, \dots$  sont indépendantes et de même loi, d'autre part que cette loi commune est une loi exponentielle de paramètre  $\lambda$  (appelé le *taux* du processus). On prendra garde que, pour  $h \geq 2$ , les différences  $T_{n+h} - T_n$  (en particulier  $T_h = T_{0+h} - T_0$ ) ne suivent pas une loi exponentielle mais une loi d'Erlang !

Vu du point de vue de la fonction de comptage  $N(t)$ , le processus de Poisson est caractérisé par un ensemble de trois propriétés portant sur les probabilités d'accroissement de  $N(t)$  :

- $P(N(t+u) = N(t) + k)$  ne dépend ni de  $t$  ni de  $N(t)$ , mais seulement de  $u$  et de  $k$  ;
- $P(N(t + \delta t) = N(t) + 1)$  est équivalent à  $\lambda \delta t$  lorsque  $\delta t$  tend vers 0 ;
- $P(N(t + \delta t) \geq N(t) + 2)$  est  $o(\delta t)$  lorsque  $\delta t$  tend vers 0 (la notation «  $o(\delta t)$  » signifie « tend vers 0 plus vite que  $\delta t$  »).

En d'autres termes, entre  $t$  et  $t + \delta t$  ( $\delta t$  très petit), la fonction de comptage a une probabilité voisine de  $1 - \lambda \delta t$  de rester constante, une probabilité voisine de  $\lambda \delta t$  d'augmenter de 1, et une probabilité infime d'augmenter de plus de 1.

Le paramètre  $\mu$  de la loi de Poisson suivie par  $N(t)$  est égal à  $\lambda t$ .

Dans une caractérisation mathématique rigoureuse du processus de Poisson, on établit l'équivalence entre les propriétés des temps d'arrivée et les propriétés du processus de comptage associé.

Pour les *temps d'arrivée* : le processus de Poisson est caractérisé par les deux propriétés :

- les v.a.  $T_1, T_2 - T_1, \dots, T_{n+1} - T_n, \dots$  sont indépendantes, et les v.a.  $T_2 - T_1, \dots, T_{n+1} - T_n, \dots$  sont positives et de même loi (on dit que l'on a un « processus de renouvellement ») ;
- les v.a.  $T_1, T_2 - T_1, \dots, T_{n+1} - T_n, \dots$  suivent toutes une même loi exponentielle de paramètre  $\lambda$ .

Pour le *processus de comptage*  $N(t)$ , noté plutôt  $N_t$  : le processus de Poisson est caractérisé par les deux propriétés :

- le processus est « à accroissements indépendants » : pour toute suite croissante  $0 = t_0 < t_1 < \dots < t_n$ , les  $n$  variables aléatoires  $N_{t_k} - N_{t_{k-1}}$  sont indépendantes ;
- pour tout  $t$  et tout  $u$ , la variable aléatoire entière  $N_{t+u} - N_t$  suit une loi de Poisson de paramètre  $\mu = \lambda u$  (dans certaines présentations, cette deuxième propriété est elle-même découpée en trois propriétés moins explicites dont l'ensemble force le couple loi exponentielle – loi de Poisson).

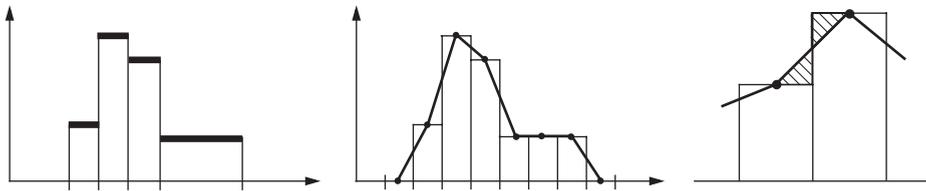
## Pólya (loi de)

Voir *binomiale négative (loi)*.

## polygone des effectifs / des fréquences (frequency polygon)

Procédé de « lissage » polygonal d'un histogramme : on ajoute aux deux extrémités de l'histogramme des classes d'effectif ou de fréquence nulle, et on trace le polygone qui joint les milieux successifs des sommets des rectangles qui figurent les classes. De cette façon, l'aire totale sous le polygone est égale à l'aire totale de l'histogramme, de façon à respecter la proportionnalité des aires aux effectifs ou aux fréquences qui caractérise les histogrammes (cf. 3<sup>e</sup> figure ci-dessous).

Pour obtenir cette conservation de l'aire, il est impératif que toutes les classes soient de largeur égale (si ce n'est pas le cas, il faut subdiviser artificiellement les classes les plus larges pour obtenir l'égalité de toutes les largeurs).



## population (population)

Ensemble de référence, dans lequel seront pris des *individus* ou des *échantillons* d'individus. Il ne faut pas prendre population et individus au sens biologique de ces mots, il s'agit ici de tout ensemble d'« objets (ou unités) statistiques », valeurs numériques, mots, documents, biens matériels, faits ou phénomènes, ..., mais incluant bien sûr individus biologiques, notamment êtres humains.

Cette notion est le concept statistique correspondant au concept probabiliste d'espace fondamental.

**Exemple 1** Le corpus des mots anglais du vocabulaire technique de l'industrie chimique.

**Exemple 2** Le parc des automobiles européennes en circulation au 1<sup>er</sup> janvier 2004.

**Exemple 3** L'ensemble des habitants de la commune de Bègles en Gironde enregistrés lors du recensement de mars 1999.

## pourcentage (percentage)

Manière de nommer et d'écrire un nombre réel positif dans un contexte particulier (taux d'accroissement, grandeur économique, fréquence, probabilité, ...).

Pour parler de l'estimation, de l'intervalle de confiance et du test de Student relatifs à une probabilité, un vocabulaire encore très employé utilise « pourcentage » à la place de « probabilité ».

### prédiction, prédiction (intervalle de) (prediction interval)

On considère un couple  $(X, Y)$  de variables numériques, soit dans une situation de *modèle linéaire* ( $X$  est alors un ensemble  $\{x_i\}$  de valeurs « maîtrisées » et  $Y$  un ensemble  $\{Y_i\}$  de variables aléatoires normales associées), soit dans une situation de *régression linéaire* ( $(X, Y)$  est alors un couple de v.a. qui suit une loi normale à 2 dimensions). On peut définir dans l'un et l'autre cas la *droite de régression* théorique  $y = \alpha + \beta x$ , et estimer les coefficients  $a$  et  $b$  de la droite de régression empirique  $y = a + bx$ .

On peut alors prédire la valeur de  $Y$  lorsque l'on donne  $X = x$ . L'intérêt majeur est bien entendu dans le cadre du *modèle linéaire*, où la variable  $x$  est contrôlée. La valeur ponctuelle de la prédiction est l'espérance mathématique (conditionnelle) empirique  $y = a + bx$ .

Si l'on veut donner l'analogie d'un intervalle de confiance, le problème est mixte et il faut tenir compte à la fois de l'incertitude sur l'estimation de l'espérance (mesurée par la variance propre de cette espérance : variance expliquée par la régression, cf. plus haut) et de la variabilité d'une valeur individuelle (mesurée par la variance résiduelle). On obtient ainsi un « intervalle de prévision ou de prédiction ».

#### Valeur « probable » de $y$ en fonction de $x$

Si  $s_y = s \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$ , le quotient  $\frac{y - E(Y|X=x)}{s_y} = \frac{y - (\alpha + \beta x)}{s_y}$  suit une loi de Student à  $n - 2$  degrés de liberté.

On notera que l'imprécision augmente assez rapidement lorsque l'on extrapole en dehors de la zone des abscisses des points observés.

Voir aussi *décomposition de la variance*.

### presque sûrement (almost certainly)

Locution utilisée lorsqu'un événement d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , *a priori* différent de  $\Omega$ , possède une probabilité égale à 1. Par exemple, dans l'espace qui représente une infinité de parties indépendantes de Pile ou Face, l'évènement « au moins une fois F » est de probabilité égale à 1 quoiqu'il ne soit pas « logiquement » certain : on peut imaginer que l'on ait PPPPP... indéfiniment (on peut imaginer... mais la probabilité est nulle !).

### prévalence (prevalence)

Terme de statistique médicale, qui désigne la proportion d'individus atteints d'une affection à un moment donné dans une population (anglicisme).

### prévision, prévision (intervalle de)

Voir *prédiction*.

### probabilité (probability)

Nombre réel, compris entre 0 et 1, attribué à un évènement.

Voir *mesure de probabilité*.

**probabilités (calcul des)** (*probability theory, theory of probabilities*)

Branche des mathématiques qui définit les concepts et les outils fondamentaux adaptés à la modélisation des phénomènes aléatoires, qui élabore une « panoplie » étendue de modèles (« lois ») adaptés aux diverses situations concrètes à représenter, et qui énonce et démontre les théorèmes de convergence relatifs à la répétition des expériences aléatoires.

Les théorèmes de convergence (« loi » des grands nombres, théorème central limite ...) permettent ensuite de relier le calcul des probabilités et l'analyse statistique (statistique « inférentielle »).

**probabilités combinatoires** (*combinatorial [theory of] probabilities*)

Partie du calcul des probabilités qui étudie les espaces finis et équiprobabilisés. La détermination des probabilités se ramène alors à des comptes d'évènements élémentaires et utilise intensivement les *dénombrements* étudiés en analyse combinatoire.

**probabilité conditionnelle**

Voir *conditionnelle (probabilité)*.

**probabilités composées (formule des)**

Voir *composées (formule des probabilités)*.

**probit (transformation en)** (*probit transformation*)

Transformation appliquée à la fonction de répartition empirique d'une distribution statistique d'un échantillon d'une variable aléatoire normale, qui permet de représenter graphiquement cette fonction de répartition par une droite.

Voir *droite de Henry*.

**processus [aléatoire], processus stochastique** (*stochastic process*)

Concept du calcul des probabilités qui fournit un cadre général pour l'étude des suites temporelles de variables aléatoires. On peut considérer qu'un processus aléatoire est une « fonction aléatoire » où la variable est le temps. Plutôt que d'écrire  $f(t)$ , la « valeur » de  $f(t)$  étant une variable aléatoire, on écrit généralement  $(X_n)$  lorsque le temps est *discret* (et on parle de la suite  $(X_n)$  de v.a), ou bien  $(X_t)$  lorsque le temps est continu (et on parle de la famille  $(X_t)$  de v.a). Les valeurs prises par le processus sont les éléments d'un ensemble  $\mathcal{E}$ , appelé « espace des états » du processus, et qui peut être fini, dénombrable ou infini continu. Une « réalisation » du processus, *i.e.* une suite (discrète ou continues) de valeurs prises s'appelle une *trajectoire*.

**Exemple 1** On considère une suite de parties de Pile ou Face indépendantes et on note  $S_n$  le nombre de Pile (par exemple) après la  $n$ -ième partie. La suite croissante  $(S_n)$  est un processus à temps discrets et à valeurs discrètes appelé *processus* ou *schéma de Bernoulli*.

**Exemple 2** On considère une « particule » sur une droite qui part de l'origine et qui, à chaque unité de temps, effectue un saut de 1 vers la droite ou vers la gauche avec égales probabilités  $\frac{1}{2}$ . On note  $X_n$  l'abscisse atteinte à l'instant  $n$ . La suite  $(X_n)$  est un processus à

temps discrets et à valeurs discrètes appelé *marche aléatoire* (ce processus est très proche d'un processus de Bernoulli et la principale différence est dans la présentation).

**Exemple 3** On considère un ensemble de « particules ». À chaque unité de temps (qui représente une « génération »), chaque particule « engendre », de façon indépendante et selon une loi de probabilité fixée, un certain nombre de « descendants ». On note  $A_n$  le

nombre de particules à la génération  $n$ . La suite  $(A_n)$  est un processus à temps discrets et à valeurs discrètes appelé *processus de branchement* ou *processus de Galton-Watson*.

**Exemple 4** On considère une succession de « pannes », modélisée par une famille croissante  $(N_t)$  de variables aléatoires : pour chaque temps  $t$ , la valeur de  $N_t$  est égale au nombre total de pannes survenues entre 0 et  $t$ . Si l'accroissement de  $N_t$  vérifie un certain nombre de règles mathématiques qui modélisent la survenue de pannes à taux constant et « sans mémoire », la famille  $(N_t)$  est un processus à temps continus et à valeurs discrètes appelé *processus de Poisson*.

**Exemple 5** On considère une généralisation de la marche aléatoire, à temps continus et à valeurs dans un espace géométrique  $(\mathbf{R}, \mathbf{R}^2, \mathbf{R}^3, \dots)$ , et on appelle  $X(t)$  la position de la « particule » à l'instant  $t$ . On dit que ce processus (partant de l'origine à l'instant 0) est un *mouvement brownien* s'il vérifie les deux propriétés : 1° il est à « accroissements indépendants et stationnaires » (les lois de  $X(t_2) - X(t_1)$  sont indépendantes et ne dépendent que de  $t_2 - t_1$ ) ; 2° pour tout  $t$  la loi de  $X(t)$  est une loi normale centrée en 0 et d'« extension » proportionnelle à  $\sqrt{t}$  (dans le cas de la dimension 1, cela veut dire loi normale « ordinaire » d'écart-type proportionnel à  $\sqrt{t}$ , dans le cas général, cela veut dire loi normale « multidimensionnelle » de « matrice de variance-covariance » proportionnelle à  $t$ ).

Voir *chaîne de Markov*.

## processus de Bernoulli

Voir *Bernoulli (schéma de)*.

## processus de Poisson

Voir *Poisson (processus de)*.

## produit de deux variables aléatoires (réelles)

Des formules générales existent pour les probabilités et les densités, mais on se limitera ici au cas de v.a. indépendantes.

### Variables aléatoires réelles discrètes et indépendantes

On suppose  $X$  définie par l'ensemble  $\{x_i\}$  des valeurs prises et par les probabilités ponctuelles  $p_i = P(X = x_i)$ .

On suppose  $Y$  définie par l'ensemble  $\{y_j\}$  des valeurs prises et par les probabilités ponctuelles  $q_j = P(Y = y_j)$ .

On pose  $Z = XY$ .

Formule 1 :

$$P(Z = z) = \sum_{x_i y_j = z} p_i q_j$$

(comprendre : sommer sur l'ensemble des couples  $(i, j)$  tels que  $x_i y_j = z$ ).

Formule 2 équivalente :

$$P(Z = z) = \sum_i P(X = x_i) P\left(Y = \frac{z}{x_i}\right)$$

(comprendre : prendre  $P\left(Y = \frac{z}{x_i}\right) = P(Y = y_j)$  lorsqu'il existe une valeur  $y_j$  telle que

$y_j = \frac{z}{x_i}$ , prendre  $P\left(Y = \frac{z}{x_i}\right) = 0$  lorsqu'il n'en existe pas).

**Variables aléatoires absolument continues et indépendantes**

On suppose X définie par sa densité  $f(x)$ .

On suppose Y définie par sa densité  $g(y)$ .

On pose  $Z = XY$ .

La densité  $h(z)$  de Z est donnée par :

$$h(z) = \int_{-\infty}^{\infty} f(x)g\left(\frac{z}{x}\right)\frac{dx}{|x|} = \int_{-\infty}^{\infty} g(y)f\left(\frac{z}{y}\right)\frac{dy}{|y|}$$

(nota : si X et/ou Y prend ses valeurs sur une partie seulement de  $\mathbf{R}$ , avec donc sa densité nulle à l'extérieur, ne pas oublier de réduire en conséquence l'intervalle d'intégration).

**Exemple** On considère deux v.a.  $X_1$  et  $X_2$  uniformes sur  $[0, 1]$  et on définit leur produit  $Z = X_1 X_2$ . Donner la densité de probabilité  $h$  de Z.

Les densités  $f_1$  et  $f_2$  de  $X_1$  et  $X_2$  sont nulles hors de l'intervalle  $[0, 1]$  et égales à 1 sur cet intervalle. Donc le facteur  $f_1(x)$  restreint l'intervalle d'intégration à  $[0, 1]$  et oblige  $z$  à appartenir également à cet intervalle (pour que  $h(z)$  soit  $\neq 0$ ). Le facteur  $f_2\left(\frac{z}{x}\right)$  restreint quant à lui l'intervalle d'intégration à  $[z, 1]$ . On en déduit finalement la valeur de la densité :

$$h(z) = \int_z^1 \frac{dx}{x} = -\ln z \quad (0 < z \leq 1),$$

d'où l'on peut déduire l'expression de la fonction de répartition :

$$H(z) = \begin{cases} 0 & \text{si } z \leq 0 \\ z - z \ln z & \text{si } 0 < z \leq 1 \\ 1 & \text{si } z > 1 \end{cases}$$

**Indicateurs**

Aucune propriété dans le cas général.

Cas où X et Y sont indépendantes :

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

Voir opérations sur les variables aléatoires.

**produit d'espaces probabilisés** (product of probability spaces)

Concept permettant de modéliser efficacement l'étude globale de la survenue (simultanée ou séquentielle) de plusieurs phénomènes aléatoires.

On disjointra le cas discret du cas général.

► Cas discret

Soient  $(\Omega_1, \mathcal{P}(\Omega_1), P_1)$  et  $(\Omega_2, \mathcal{P}(\Omega_2), P_2)$  deux espaces probabilisés discrets (i.e. finis ou dénombrables). On se place dans la situation standard où les tribus sont les ensembles de toutes les parties. On définit l'espace probabilisé produit avec :

- pour espace fondamental le produit cartésien  $\Omega_1 \times \Omega_2$ ,
- pour tribu d'évènements l'ensemble de toutes les parties  $\mathcal{P}(\Omega_1 \times \Omega_2)$ ,
- pour mesure de probabilité la mesure P définie par :

$$\forall x_i \in \Omega_1 \quad \forall y_j \in \Omega_2 \quad P((x_i, y_j)) = P_1(x_i) P_2(y_j).$$

Cette indépendance « ponctuelle » entraîne l'indépendance générale :

$$\forall A_i \subset \Omega_1 \quad \forall B_j \subset \Omega_2 \quad P(A_i \times B_j) = P_1(A_i) P_2(B_j).$$

**Exemple** Si  $\Omega_1 = \Omega$  est l'espace fondamental à 6 évènements élémentaires qui représente le lancement d'un premier dé, et  $\Omega_2 = \Omega$  est l'espace fondamental à 6 évènements élémentaires qui représente le lancement d'un deuxième dé, le produit cartésien  $\Omega_1 \times \Omega_2 = \Omega^2$  représente le lancement des 2 dés. Il contient  $6 \times 6 = 36$  évènements élémentaires. Et l'équiprobabilité

$$\forall i \quad \forall j \quad P(x_i) = P(y_j) = \frac{1}{6} \text{ entraîne l'équiprobabilité } \forall (i, j) \quad P((x_i, y_j)) = \frac{1}{36}.$$

Cette construction se généralise sans difficulté à un nombre fini d'espaces probabilisés discrets.

► **Cas général**

Soient  $(\Omega_1, \mathcal{A}_1, P_1)$  et  $(\Omega_2, \mathcal{A}_2, P_2)$  deux espaces probabilisés quelconques. On définit l'espace probabilisé produit avec :

- pour espace fondamental le produit cartésien  $\Omega_1 \times \Omega_2$ ,
- pour tribu d'évènements la « tribu produit » (notée  $\mathcal{A}_1 \otimes \mathcal{A}_2$ ), qui est la tribu engendrée par les produits cartésiens  $A_i \times B_j$ , pour  $A_i \in \mathcal{A}_1, B_j \in \mathcal{A}_2$ ,
- pour mesure de probabilité la mesure  $P$  (parfois notée  $P_1 \otimes P_2$ ) caractérisée par :

$$\forall A_i \in \mathcal{A}_1 \quad \forall B_j \in \mathcal{A}_2 \quad P(A_i \times B_j) = P_1(A_i) P_2(B_j).$$

Cette mesure est faite pour rendre indépendants les évènements « qui ne dépendent que du premier espace » avec les évènements « qui ne dépendent que du deuxième espace », et c'est un résultat mathématique essentiel de montrer qu'elle existe et est unique.

Cette construction se généralise sans difficulté à un nombre fini d'espaces probabilisés, mais le problème se complique lorsque l'on veut effectuer le produit d'un nombre infini d'espaces probabilisés (discrets ou non), de manière à donner un cadre convenable à l'étude des suites infinies  $(\omega_1, \omega_2, \dots, \omega_n, \dots)$ . La théorie mathématique qui permet de construire ce cadre repose sur la considération privilégiée des évènements « observables », qui ne font intervenir qu'un nombre fini d'indices.

**puissance [d'un test]**

(*power [of a test], strenght*)

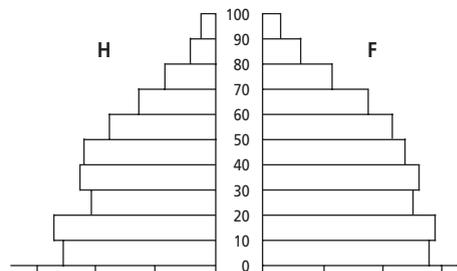
Désigne, pour un test d'hypothèse, la probabilité de rejet justifié de l'hypothèse  $H_0$  (le contraire, non-rejet de  $H_0$  alors qu'elle est fausse, est le risque de manque de puissance, appelé aussi risque de seconde espèce).

Lorsque l'hypothèse  $H_0$  est du type «  $\theta = \theta_0$  », où  $\theta$  est un paramètre de la loi testée, la puissance est une fonction de la vraie valeur  $\theta$  du paramètre, d'autant plus voisine de 1 que  $\theta$  est éloigné de  $\theta_0$ .

**pyramide des âges**

(*age pyramid*)

Double histogramme en miroir représentant pour les hommes et pour les femmes la distribution du caractère « âge ». Dans sa version traditionnelle, la pyramide des âges est figurée avec permutation de l'orientation des axes : l'axe des abscisses (pour l'âge en années, ou l'année de naissance, selon la présentation) est vertical, et l'axe des ordonnées (pour le nombre ou la proportion des personnes, ce qui revient visuellement au même) est horizontal.





## qualitatif, ive

Se dit d'une variable aléatoire ou d'une variable statistique (ou caractère) dont les valeurs possibles – souvent appelées dans ce cas « modalités » – sont des éléments d'un ensemble sur lequel on ne peut pas effectuer d'opération mathématique (addition notamment). Il arrive néanmoins qu'un codage d'une variable qualitative donne à ses modalités l'apparence de nombres.

## quantile [d'ordre $\alpha$ ]

(*quantile [of order  $\alpha$ ],  $\alpha$ -quantile*)

Indicateur de position attaché à une variable aléatoire réelle, utilisé essentiellement en statistique. Concrètement le quantile d'ordre  $\alpha$  est la (ou une des ...) valeur qui partage la série des valeurs en deux parties de fractions  $\alpha$  et  $1 - \alpha$  de l'effectif total. Certains quantiles portent des noms spécifiques : quartiles, déciles, centiles.

Si la signification concrète des quantiles est simple et « parlante », la traduction formelle est plus délicate. On adaptera sans difficulté le formulaire détaillé pour la *médiane*.

Synonyme de *fractile [d'ordre  $\alpha$ ]*.

## quantitatif, ive

Se dit d'une variable aléatoire ou d'une variable statistique (ou caractère) dont les valeurs possibles sont des « grandeurs », éléments d'un ensemble mathématique, de nombres ou de vecteurs le plus souvent, sur lequel peut effectuer des opérations mathématiques, des additions notamment (ce qui permet de définir espérance mathématique ou moyenne, etc.).

On peut ou non considérer comme des variables quantitatives les variables « ordinales », dont les valeurs possibles sont des éléments que l'on peut classer entre eux (auxquels donc on peut attribuer un numéro d'ordre).

## quantité d'information

Voir *information (quantité d')*.

## quartile

(*quartile*)

Indicateur de position attaché à une variable aléatoire réelle. Concrètement c'est la (ou une des ...) valeur qui partage la distribution (en probabilités) ou la série des valeurs (en statistique) en deux parties de probabilités 0,25 et 0,75, ou d'effectif 25 % et 75 % de l'effectif total. On parle parfois de quartile inférieur et de quartile supérieur.

Si la signification concrète des quartiles est simple et « parlante », la traduction formelle est plus délicate. On adaptera sans difficulté le formulaire détaillé pour la *médiane*.

## Quételet (Adolphe)

Astronome et mathématicien belge (1796–1874). Il appliqua les probabilités aux sciences humaines et à l'anthropométrie et fonda la statistique sur le calcul des probabilités.

**quotient de deux variables aléatoires (réelles)** *(quotient of random variables)*

Seul le cas absolument continu possède des applications pratiques. En outre, quoique des formules générales existent, on se limitera ici au cas de v.a. indépendantes.

**Variables aléatoires absolument continues et indépendantes**

On suppose  $X$  définie par sa densité  $f(x)$ , et on pose  $F(x)$  sa fonction de répartition.

On suppose  $Y$  définie par sa densité  $g(y)$ , et on pose  $G(y)$  sa fonction de répartition.

On pose  $Z = \frac{X}{Y}$ .

La densité  $h(z)$  de  $Z$  est donnée par :

$$h(z) = \int_{-\infty}^{\infty} f(yz)g(y)|y|dy.$$

(Nota : si  $X$  et/ou  $Y$  prend ses valeurs sur une partie seulement de  $\mathbf{R}$ , avec donc sa densité nulle à l'extérieur, ne pas oublier de réduire en conséquence l'intervalle d'intégration.)



## randomisation

(*randomization*)

Introduction volontaire d'un élément aléatoire dans un processus (au sens courant du mot). Procédé utilisé en médecine pour les tests de comparaison d'efficacité de deux traitements d'une pathologie : pour éviter toute perturbation du test notamment par effet placebo, on tire au sort le traitement donné à chaque patient.

## rangs (coefficient de corrélation des)

Voir *corrélation des rangs (coefficient de)*.

## rapport de corrélation

Voir *corrélation (rapport de)*.

## rapport des variances (test du)

Voir *Fisher-Snedecor (test de)*.

## Rayleigh (loi de)

(*Rayleigh distribution*)

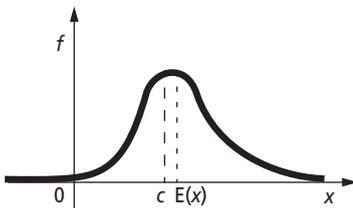
Loi d'une variable aléatoire continue qui peut être définie comme la racine carrée de la somme des carrés de deux variables aléatoires normales indépendantes  $Y_1$  et  $Y_2$ , centrées et de même écart-type :

$$X = \sqrt{Y_1^2 + Y_2^2}$$

### Formulaire

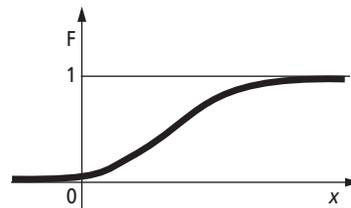
Un paramètre réel :  $c \in \mathbf{R}_+^*$  (qui est l'écart-type des v.a. normales indépendantes  $Y_1$  et  $Y_2$ ).

#### ► Loi de probabilité



densité

$$f(x) = \frac{x}{c} \exp\left(-\frac{x^2}{2c^2}\right) \quad (x \geq 0)$$



fonction de répartition

$$F(x) = \int_0^x f(t) dt \quad (x \geq 0)$$

► Valeurs caractéristiques

- espérance :  $E(X) = c \sqrt{\frac{\pi}{2}} \approx 1,253c$
- variance :  $\text{Var}(X) = \frac{4-\pi}{2}c^2 \approx 0,429c^2$
- écart-type :  $\sigma(X) = \sqrt{\frac{4-\pi}{2}}c \approx 0,655c$

► Utilisations

De par sa définition même, la loi de Rayleigh est la loi de la distance à l'origine d'un point aléatoire d'un plan, dont les coordonnées sont deux variables aléatoires normales indépendantes, centrées et de même écart-type.

Dans la pratique, la loi de Rayleigh est utilisée par les spécialistes de théorie du signal pour étudier le filtrage d'un « bruit gaussien ».

C'est enfin un cas particulier de la *loi de Weibull* (pour la valeur  $\beta = 2$  du paramètre de forme) ; à ce titre elle est utilisée en théorie de la fiabilité des systèmes.

Voir *Maxwell (loi de)*.

**recensement**

([complete] census)

Recueil de données sur la totalité des individus d'une population.

**rectangulaire (loi)**

(rectangular distribution)

Voir *uniforme continue (loi)*.

**réduit (écart)**

Voir *écart réduit*.

**réduite (variable aléatoire)**

Se dit d'une variable aléatoire  $X$  dont l'écart-type est égal à 1. Dans la pratique, on ne considère que rarement des variables aléatoires réduites mais non centrées.

Voir *centrée réduite (variable aléatoire)*.

**réel (nombre)**

(real)

Nombre « ordinaire », tel que 19,  $-1$ ,  $\frac{355}{113}$ ,  $e^{-0,5}$ , 0,123456789,  $\frac{\pi}{6}$ ,  $-0,875$ , 3, ...

Dans des expressions telles que « fonction réelle » ou « variable aléatoire réelle », les adjectifs « réel(le) » et « numérique » sont entièrement synonymes.

**région d'acceptation**

(acceptance region)

Voir *région critique*.

**région critique**

(critical region)

On considère un test d'hypothèse, auquel est associé une hypothèse  $H_0$  et une « variable de test » ou « variable de décision »  $T$  à valeurs dans un ensemble  $E$  (le plus souvent  $\mathbf{R}$  ou  $\mathbf{R}_+$ ).

Étant donné un risque  $\alpha$ , la procédure de test définit une région  $W = W(\alpha)$  de  $E$  : si la valeur observée de  $T$  appartient à  $W$ , on rejette  $H_0$ , si  $T$  n'appartient pas à  $W$  (donc appartient à la région complémentaire), on ne rejette pas  $H_0$ . On appelle  $W$  *région critique* (pour le test considéré et le niveau  $\alpha$ ), et son complémentaire *région d'acceptation*.

## régression

Voir *courbe de régression, droite de régression, décomposition de la variance, loi des estimateurs dans une régression et intervalle de confiance, prédiction*.

## régression à deux variables explicatives

On se limite à donner ci-dessous les formules relatives à la régression linéaire à deux variables explicatives, les formules les plus générales – qui utilisent nécessairement le formalisme matriciel – pourront être trouvées dans les ouvrages spécialisés.

On suppose que l'on a un triplet  $(X_1, X_2, Y)$  de variables aléatoires qui suit une loi normale à 3 dimensions,  $X_1$  et  $X_2$  sont les variables « explicatives » et  $Y$  la variable « dépendante ». On a  $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + E$ , où  $E$  est une variable aléatoire globale « écart »; l'espérance conditionnelle est une fonction linéaire (affine) de  $x$  :  $E(Y|X_1 = x_1 \text{ et } X_2 = x_2) = \alpha + \beta_1 x_1 + \beta_2 x_2$ , et la variance conditionnelle est la constante  $\sigma^2$ .

### Formulaire

Cas d'un échantillon observé de  $n$  triplets de valeurs numériques  $((x_{11}, x_{21}, y_1), (x_{12}, x_{22}, y_2), \dots, (x_{1n}, x_{2n}, y_n))$ .

Les estimations (non biaisées)  $a$ ,  $b_1$  et  $b_2$  de  $\alpha$ ,  $\beta_1$  et  $\beta_2$  sont les solutions du système :

$$\left\{ \begin{array}{l} an + b_1 \sum x_{1i} + b_2 \sum x_{2i} = \sum y_i \\ a \sum x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i} = \sum x_{1i} y_i \\ a \sum x_{2i} + b_1 \sum x_{1i} x_{2i} + b_2 \sum x_{2i}^2 = \sum x_{2i} y_i \end{array} \right.$$

Une estimation non biaisée de la variance résiduelle  $\sigma^2$  est :

$$s^2 = \frac{1}{n-3} \sum_{i=1}^{i=n} (y_i - (a + b_1 x_{1i} + b_2 x_{2i}))^2.$$

## régression, régression linéaire (regression, linear regression)

Méthode de représentation géométrique et d'analyse statistique d'une situation probabiliste ou statistique « bidimensionnelle » ou « multidimensionnelle ». Dans le cas simple à deux variables, la situation décrite et analysée par la régression est celle d'un couple  $(X, Y)$  de variables aléatoires qui suit *a priori* une loi normale à 2 dimensions. La variable  $X$  est souvent appelée « explicative » et la variable  $Y$  « dépendante ». La *régression linéaire* ajuste la liaison entre les deux variables par une relation  $Y = a + bX + E$ , où  $E$  est une variable aléatoire « écart » (qui est normale à 1 dimension). Le problème fondamental est de rechercher les valeurs des coefficients numériques  $a$  et  $b$  qui minimisent – en un sens approprié de ce mot – l'écart  $E$ . Simultanément, les coefficients  $a$  et  $b$  sont ceux de la droite de régression  $y = E(Y|X = x) = a + bx$ , qui est, si l'hypothèse de linéarité est correcte, la courbe de régression.

Le cas général, où la liaison est de la forme  $Y = a + bf(X) + E$ , avec  $f$  fixée (et  $E$  normale), est le plus souvent traité par adaptation du cas linéaire, qui reste le cas fondamental.

Dans le cas *multiple* à  $p + 1$  variables ( $p \geq 2$ ), il y a  $p$  variables « explicatives »  $X_j$ , et la liaison entre ces  $p$  variables et la  $(p + 1)$ -ième variable « dépendante » est ajustée par une relation  $Y = a + b_1X_1 + \dots + b_pX_p + E$ , où  $E$  est une variable aléatoire « écart » (*a priori* gaussienne). Le problème fondamental est de rechercher les valeurs des coefficients numériques  $a, b_1, \dots, b_p$  qui minimisent – en un sens approprié de ce mot – globalement l'écart  $E$ .

Le cas général est, comme précédemment, le plus souvent traité par adaptation du cas linéaire.

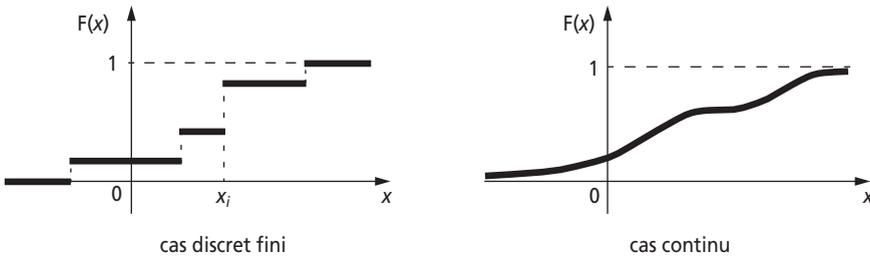
Voir *droite de régression linéaire, corrélation (rapport de), corrélation multiple (coefficient de)*.

**répartition (fonction de)** *(distribution function)*

Fonction réelle  $F$  associée à toute variable aléatoire réelle  $X$  :

$$F(x) = P(X \leq x).$$

Cette fonction est croissante (au sens large) et varie de 0 pour  $x = -\infty$  à 1 pour  $x = \infty$ . Si  $X$  est une v.a. discrète, elle varie par sauts, si  $X$  est une v.a. continue, elle varie continuellement.



fonction de répartition

Si  $X$  est *discrète*, caractérisée par l'ensemble (fini ou dénombrable) de valeurs  $\{x_i\}$ , avec les probabilités ponctuelles  $p_i = P(X = x_i)$ , on a :

$$F(x) = \sum_{x_i \leq x} p_i,$$

(selon les cas, il s'agira d'une somme finie ou d'une somme infinie).

Si  $X$  est *absolument continue*, caractérisée par la densité de probabilité  $f(x)$ , on a :

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

Certains manuels définissent la fonction de répartition comme la probabilité  $P(X < x)$ , la différence est minime (et même nulle lorsque la v.a. est continue).

La notion de fonction de répartition se généralise naturellement et sans difficulté à un couple puis à un «  $n$ -uple » de variables aléatoires réelles.

**répétées (épreuves)**

Voir *épreuves répétées*.

**représentatif (échantillon)**

Voir *échantillon représentatif*.

**résumé numérique**

Voir *indicateur*.

**réunion***(logical sum)*

Dans la formalisation ensembliste des espaces probabilisables, les événements sont des parties de l'espace fondamental  $\Omega$ . Si l'on considère deux événements  $A$ ,  $B$ , leur réunion  $A \cup B$  est un événement dont la réalisation correspond à la *disjonction logique* «  $A$  ou  $B$  » :

$$\omega \in A \cup B \Leftrightarrow (\omega \in A \text{ ou } \omega \in B)$$

Cela se généralise sans difficulté à un nombre supérieur d'événements. Les deux notations ensembliste et logique :  $A \cup B$ ,  $A$  ou  $B$ , sont parfaitement synonymes.

**Exemple** On considère le tirage d'une carte dans un jeu ordinaire de 52 cartes. La réunion de l'événement  $A$  = « trèfle » (13 événements élémentaires) et de l'événement  $B$  = « Roi » (4 événements élémentaires) comprend 16 événements élémentaires : les 12 trèfles autres que le Roi, le Roi de trèfle, et les 3 Rois autres que celui de trèfle.

On voit sur cet exemple que le connecteur logique « ou » tel qu'il est codifié par la logique mathématique, et contrairement à certains de ses usages dans la langue courante, est non exclusif. Les événements  $A$  et  $B$  ne sont pas exclusifs (en termes ensemblistes : ne sont pas disjoints), et l'événement « Roi de trèfle », qui leur est commun, appartient à leur réunion.

$$\begin{aligned} \text{Si } A \text{ et } B \text{ sont disjoints : } P(A \cup B) &= P(A) + P(B). \\ \text{De façon générale : } P(A \cup B) &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Voir *Poincaré (formule de)*.

Synonyme de *disjonction logique*.

**risque de première espèce,  
de deuxième espèce***(risk of the first kind,  
of the second kind)*

Pour un test d'hypothèse, le risque de première espèce (ou risque d'« erreur »), généralement noté  $\alpha$ , est le risque de rejet à tort de l'hypothèse  $H_0$  lorsqu'elle est vraie. Le risque de deuxième espèce (ou risque de « manque de puissance »), généralement noté  $\beta$ , est le risque de non-rejet de l'hypothèse  $H_0$  lorsqu'elle est fausse.

**robuste, robustesse [d'un test]***(robust, robustness [of a test])*

Qualité d'un test d'hypothèse qui lui permet de rester applicable lorsque les conditions nominales de son utilisation (notamment loi de probabilité et taille de l'échantillon) ne sont pas vérifiées. Bien entendu, cet écart des conditions nominales doit rester dans des limites raisonnables (et qui peuvent d'ailleurs être précisées).



## schéma de Bernoulli

Voir *Bernoulli (schéma de)*.

## série chronologique, série temporelle

(*time serie*)

Série statistique qui donne l'évolution d'un caractère numérique  $Y$  en fonction du temps  $T$ . Une telle série peut se présenter, soit sous l'aspect d'une série double  $((y_i, t_i))$ , soit, lorsque le temps prend des valeurs régulièrement espacées, sous l'aspect d'une série simple  $(y_i)$  où l'indice représente le temps avec pour unité l'espacement constant. Dans l'un et l'autre cas, les graphiques portent le temps en abscisse et le caractère  $Y$  en ordonnée.

Le problème fondamental de l'analyse statistique des séries chronologiques est de décomposer la série en trois ou quatre composantes adaptées aux facteurs de variation. Dans le modèle additif,  $y_i$  est la somme de ces composantes, dans le modèle multiplicatif (fréquent en économie),  $y_i$  est le produit des composantes. Comme on peut ramener le modèle multiplicatif au modèle additif en prenant le logarithme des valeurs (et, graphiquement, en utilisant une échelle « semi-logarithmique »), on présentera ci-dessous la problématique de l'analyse des séries chronologiques sur le modèle additif. On écrit :

$$y_i = y(t_i) = f(t_i) + S(t_i) + e(t_i).$$

Le terme  $f(t_i)$  s'appelle la *tendance (trend)*. Il se détermine à partir d'une hypothèse sur sa « forme ». Dans le modèle additif, on suppose le plus souvent que la tendance est linéaire (affine) :  $f(t_i) = a + bt_i$ , on calcule ses coefficients  $a$  et  $b$  par ajustement linéaire. À la tendance peut s'ajouter un (ou plusieurs) cycle(s) de périodes longues, difficiles à isoler tant sur le plan théorique que sur le plan pratique (il faut notamment disposer de statistiques sur une très grande durée).

Le terme  $S(t_i)$  s'appelle la *variation saisonnière*. Elle doit satisfaire aux deux contraintes d'être de moyenne nulle sur une année et de se répéter identiquement d'année en année. On la détermine en considérant, après calcul de la tendance, la différence  $y(t_i) - f(t_i)$ . Diverses méthodes peuvent être utilisées, par exemple un ajustement linéaire, calculé pour chaque mois ou trimestre à partir des valeurs échelonnées d'année en année. Selon la nature du caractère analysé, on fera ou non un lissage préalable par moyennes mobiles.

Le terme  $e(t_i)$  s'appelle la *variation accidentelle (ou résiduelle)*. Défini comme la différence  $y(t_i) - (f(t_i) + S(t_i))$ , il est purement aléatoire (pour autant que ce mot soit approprié pour le phénomène étudié) ; il est de moyenne nulle sur une année.

## séries appariées

Voir *appariées (séries)*.

## série statistique

(*sample values*)

Suite  $(x_i)$  de données individuelles (ce qualificatif s'oppose à « regroupées ou totalisées par classes ») recueillies en vue d'une analyse statistique. Synonyme d'échantillon.

On appelle parfois série statistique double une suite  $((x_i, y_i))$  de *couples* de données individuelles.

**seuil**

Mot ambigu qui est utilisé aussi bien dans l'expression *seuil de risque* (notation traditionnelle  $\alpha$ ) que dans l'expression *seuil de confiance* (notation traditionnelle  $1 - \alpha$ ).

**Sheppard (correction de)***(Sheppard correction)*

Correction qui débiaise le calcul d'une variance empirique effectué à partir des centres de classes qui sont des intervalles de longueur fixe.

Voir *variance d'un échantillon statistique*.

**sigma-additivité,  $\sigma$ -additivité**

Voir *additivité*.

**sigma-algèbre,  $\sigma$ -algèbre**

Voir *tribu*.

**signes (test des)***(sign test)*

Test d'hypothèse non paramétrique utilisé pour contrôler l'absence de « différence systématique » dans un échantillon statistique constitué « par paires » (*Sign Test for Matched Pairs*). Il est aussi utilisé pour contrôler la valeur de la médiane.

La situation de base est celle de couples de variables numériques  $(X_i, Y_i)$ , dont chacun donnera lieu à une unique observation couplée. Les lois des  $X_i$  et des  $Y_i$  ne font pas l'objet d'un présupposé d'uniformité et l'on postule seulement que les différences  $D_i = Y_i - X_i$  suivent toutes la loi d'une même variable  $Z$ , que l'on supposera symétrique. Le test des signes contrôle l'égalité  $P(Z \leq 0) = P(Z \geq 0)$  (si les lois sont continues, on a par surcroît  $P(Z = 0) = 0$ , et donc  $P(Z \leq 0) = P(Z \geq 0) = \frac{1}{2}$ ). Il fonctionne très simplement par vérification de la répartition

binomiale  $\mathcal{B}(n, \frac{1}{2})$  des signes de  $Z$ , après éventuelle répartition conventionnelle des valeurs nulles.

Un premier type d'utilisation porte sur des observations couplées qui sont présentées soit comme des observations d'une même variable sur plusieurs individus dans deux situations différentes, soit comme provenant de deux familles de variables  $X_i$  et  $Y_i$ . Le but pratique est d'examiner l'efficacité d'un apprentissage, l'efficacité d'une thérapeutique, la discordance de deux corrections, la différence systématique entre deux méthodes de mesure, etc. L'hypothèse  $H_0$  est que la loi de la variable  $Z$  qui représente la différence des observations est symétrique. Elle est parfois présentée comme «  $\mu_Z = 0$  » quoique le rejet de la symétrie n'interdise pas en théorie que l'espérance soit nulle (mais la présentation plus rarement employée  $H_0 = \langle \forall_i \mu_{X_i} = \mu_{Y_i} \rangle$  n'est pas appropriée).

Un deuxième type d'utilisation considère une seule variable  $Z$  de médiane  $M$ , et teste  $H_0 = \langle M = 0 \rangle$  (ou, par décalage, de  $H_0 = \langle M = M_0 \rangle$ )... Dans cette utilisation, le test devient paramétrique (mais reste « libre », puisqu'il n'y aucune exigence sur la loi).

————— test bilatéral non paramétrique de symétrie —————  
de la loi d'une différence de variables

• Données. Deux séries appariées : un échantillon double  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$  de  $n$  valeurs couplées de deux variables aléatoires numériques dont les différences suivent toutes la loi d'une même variable  $Z$ .

• Hypothèse testée.  $H_0 = \langle \text{la loi de } Z \text{ est symétrique} \rangle$  contre  $H_1$  alternative.

- Déroulement technique du test

1. On compte le nombre  $n_1$  des différences  $z_i = y_i - x_i$  positives (s'il y a des différences nulles, on les compte positives pour moitié, négatives pour moitié ; autre méthode : on les retire de l'échantillon, ce qui diminue bien sûr la valeur de  $n$ ).

2a. Si  $n$  est petit ( $n < 20$ ), on calcule directement la probabilité que  $\left| \mathcal{B}\left(n, \frac{1}{2}\right) - \frac{n}{2} \right|$  dépasse  $\left| n_1 - \frac{n}{2} \right|$  et on la compare au risque bilatéral  $\alpha$  (ou bien on se réfère à une table spécifique).

2b. Si  $n$  est grand ( $n \geq 20$ ), on utilise l'approximation normale de la loi binomiale et on calcule :

$$t = \frac{|2n_1 - n| - 1}{\sqrt{n}},$$

que l'on compare au  $t_\alpha$  (écart absolu de probabilité de dépassement  $\alpha$ ) de la table de la loi normale.

- Conditions et précautions

Aucunes

**Remarque :** Pour approcher la loi binomiale par la loi normale qui est continue, il faut étaler la valeur entière  $n_1$  entre  $n_1 - 0,5$  et  $n_1 + 0,5$  : sous la forme  $|2n_1 - n| - 1$ , le numérateur de la formule qui donne  $t$  inclut cette « correction de continuité » dans le sens approprié.

————— test bilatéral de comparaison —————  
de la valeur d'une médiane  $M$  à une valeur fixée  $M_0$

- Données. Un échantillon  $(z_1, z_2, \dots, z_n)$  de  $n$  valeurs observées d'une variable aléatoire numérique  $X$  de médiane  $M$ .

- Hypothèse testée.  $H_0 = \langle M = M_0 \rangle$  contre  $H_1 = \langle M \neq M_0 \rangle$ .

- Déroulement technique du test

1. On compte le nombre  $n_1$  des différences  $z_i - M$  positives (s'il y a des différences nulles, on les compte positives pour moitié, négatives pour moitié ; autre méthode : on les retire de l'échantillon, ce qui diminue bien sûr la valeur de  $n$ ).

2a ou 2b comme ci-dessus.

- Conditions et précautions

Aucunes.

————— test unilatéraux —————

Comme les tests bilatéraux avec pour seule différence :

Si 2a : on se réfère à la probabilité unilatérale que l'écart  $\left| \mathcal{B}\left(n, \frac{1}{2}\right) - \frac{n}{2} \right|$  soit inférieur ou supérieur (selon l'hypothèse  $H_0$ ) à l'écart observé.

Si 2b : on prend dans la table de la loi normale l'écart dont la probabilité unilatérale de dépassement est  $\alpha$  (ou, ce qui est équivalent, l'écart absolu dont la probabilité de dépassement est  $2\alpha$ ).

Dans le cas standard du test des signes pour les séries appariées, on peut aussi tester la même hypothèse « la loi de  $Z$  est symétrique » avec une ancienne version du *test de Wilcoxon*, ou

bien l'hypothèse dérivée «  $\mu_Z = 0$  » avec un *test de Student* (il est souvent suggéré que ces tests sont plus puissants, mais que le test des signes est plus rapide « à la main »).

## significativité d'un coefficient de corrélation (test de)

(*correlation coefficient significance test, testing  $\rho = 0$* )

Test paramétrique qui compare la valeur observée d'un coefficient de corrélation à zéro (qui contrôle donc la « significativité » d'une corrélation non nulle).

— test bilatéral de comparaison d'un coefficient de corrélation à zéro —

- Données. Un échantillon de  $n$  couples de valeurs observées  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  d'un couple  $(X, Y)$  de variables aléatoires numériques de coefficient de corrélation  $\rho$ .

- Hypothèse testée.  $H_0 = \langle \rho = 0 \rangle$  contre  $H_1 = \langle \rho \neq 0 \rangle$

- Déroulement technique du test

- 1a. On calcule les moyennes observées  $\bar{x}$  et  $\bar{y}$  avec les formules usuelles.

- 1b. On calcule la valeur observée du coefficient de corrélation avec une des formules usuelles, par exemple

$$r = \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{i=n} (x_i - \bar{x})^2 \sum_{i=1}^{i=n} (y_i - \bar{y})^2}}$$

2. On calcule la valeur observée de la variable de test :

$$t = \frac{|r|}{\sqrt{1 - r^2}} \sqrt{n - 2}$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Student, elles dépendent du nombre de degrés de liberté  $ddl = n - 2$ , et du risque  $\alpha$  (on peut aussi utiliser une table spécifique qui donne directement la valeur *critique* de  $|r|$  en fonction du risque  $\alpha$ ).

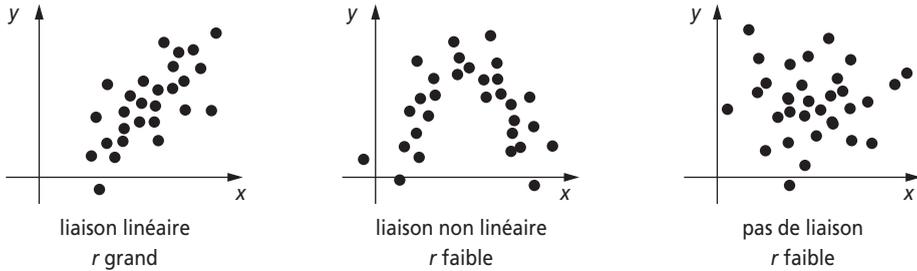
- Conditions et précautions

- En théorie  $(X, Y)$  doit suivre une loi normale à 2 dimensions, donc aucune précaution si on présume que c'est le cas ;
- lorsque ce n'est pas le cas, le test est robuste et reste applicable si  $n$  est « assez grand », la condition  $n \geq 30$  est traditionnelle.

La valeur critique de  $|r|$  peut paraître assez grande... par exemple  $|r| = 0,36$  pour  $n = 30$  et  $\alpha = 0,05$ . Mais on se souviendra que le poids du coefficient de corrélation est correctement mesuré par son carré : dans cet exemple  $0,36^2 = 0,13$ , et c'est une valeur raisonnablement faible.

On n'oubliera pas enfin que le coefficient de corrélation « mesure » une liaison *linéaire* (ce qui est toujours le cas si  $(X, Y)$  suit une loi normale) mais qu'il perd toute signification lorsqu'il existe une liaison non linéaire !

Cela étant noté, si la liaison est présumée linéaire, le test « de significativité » du coefficient de corrélation vaut test d'indépendance.



On peut aussi tester une hypothèse  $H_0 = \langle \rho = \rho_0 \rangle$ . La loi du coefficient de corrélation observé sous cette hypothèse est compliquée et mal utilisable mais la *z-transformation de Fisher* permet d'utiliser (si  $n \geq 30$ ) la loi normale comme loi de test.

---

### test bilatéral de comparaison d'un coefficient de corrélation à une valeur fixée

---

- Données. Un échantillon de  $n$  couples de valeurs observées  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  d'un couple  $(X, Y)$  de variables aléatoires numériques de coefficient de corrélation  $\rho$ .
- Hypothèse testée.  $H_0 = \langle \rho = \rho_0 \rangle$  contre  $H_1 = \langle \rho \neq \rho_0 \rangle$ .
- Déroulement technique du test
  - 1a. On calcule les moyennes observées  $\bar{x}$  et  $\bar{y}$  avec les formules usuelles ;
  - 1b. On calcule la valeur observée du coefficient de corrélation avec une des formules usuelles (cf. plus haut).
  2. On calcule la valeur observée de la variable de test, d'abord :

$$z = \arg \operatorname{th} r = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) \text{ et } \zeta_0 = \arg \operatorname{th} \rho_0 = \frac{1}{2} \ln \left( \frac{1+\rho_0}{1-\rho_0} \right),$$

puis :

$$t = \frac{z - \zeta_0}{\sqrt{n-3}}.$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi normale centrée réduite, elles ne dépendent que du risque  $\alpha$ .

- Conditions et précautions

En théorie  $(X, Y)$  doit suivre une loi normale à 2 dimensions, et il faut de plus  $n \geq 30$ .

---

## significativité d'une régression (test de)

Voir *Fisher (test de)*.

## simulation de nombres [pseudo]-aléatoires

Le principe de base consiste à générer une suite d'entiers compris entre 0 et  $M - 1$  par des opérations « modulo  $M$  ». On divise ensuite par  $M$  pour obtenir une suite de réels compris entre 0 et 1. Si  $M$  est très grand et si la méthode de génération de la suite d'entiers est convenable, avec des paramètres appropriés, on obtient finalement une suite qui a toutes les propriétés d'un échantillon d'une variable aléatoire réelle continue de loi uniforme sur  $[0, 1[$ . Un type classique et efficace de générateurs définit la suite d'entiers par récurrence linéaire :

$$x_{n+1} \equiv a + bx_n \pmod{M},$$

pour un choix convenable de  $M$ ,  $a$  et  $b$ . Une telle suite est périodique mais cela est sans inconvénient dès lors que sa période est très grande par rapport au nombre de termes utilisés. Cette

génération est faite automatiquement par les bibliothèques des logiciels et des « packages » logiciels, l'utilisateur a en général la possibilité de choisir un « germe », i.e. un élément de départ pour la suite récurrente.

### simulation d'un échantillon aléatoire de loi donnée

Un théorème d'« inversion de la fonction de répartition » permet des tirages de valeurs d'une v.a. continue  $X$  de loi donnée à partir de valeurs au hasard d'une v.a.  $U$  de loi uniforme sur  $[0, 1[$ . Si la fonction de répartition  $F$  de  $X$  est continue et possède une fonction réciproque  $F^{-1}$ , alors  $F^{-1}(U)$  suit la loi de  $X$ . Si  $F^{-1}$  est d'expression simple, cette méthode est très efficace.

**Exemple** Soit la loi exponentielle de paramètre  $\lambda$  et de fonction de répartition  $F(x) = 1 - e^{-\lambda x}$  (sur  $x \geq 0$ ). Alors  $F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u)$ . Donc, si  $U$  suit une loi uniforme sur  $[0, 1[$ ,  $-\frac{1}{\lambda} \ln(1 - U)$ , et aussi  $-\frac{1}{\lambda} \ln U$  par symétrie, suivent une loi exponentielle de paramètre  $\lambda$ .

Pour la loi normale, la fonction  $F^{-1}$  ne possède pas d'expression analytique et l'on peut par exemple utiliser la méthode suivante :

Si  $U$  et  $V$  sont uniformes sur  $[0, 1[$  et indépendantes,  $X = \sqrt{-2 \ln U} \sin 2\pi V$  et  $Y = \sqrt{-2 \ln U} \cos 2\pi V$  sont normales centrées réduites et indépendantes.

Pour une loi discrète il faut procéder autrement. Pour simuler par exemple une v.a.  $X$  de Bernoulli de paramètre  $p$ , on prend une v.a.  $U$  uniforme sur  $[0, 1[$  et on pose  $X = 1$  si  $U \leq p$ ,  $X = 0$  sinon (ce procédé s'étend naturellement à la simulation d'une v.a. multinomiale).

### singleton

(one-element set)

Partie d'un ensemble réduite à un seul élément.

En calcul des probabilités, si  $\{a\}$  est un singleton ( $a \in \Omega$ ), on écrit le plus souvent  $P(a)$  au lieu de  $P(\{a\})$  pour simplifier les notations.

### singulière (loi)

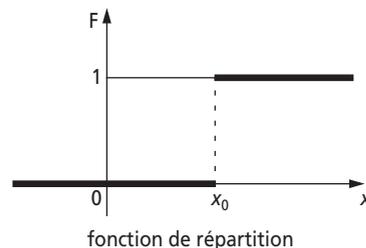
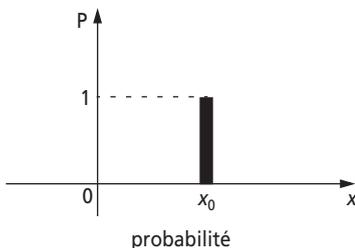
(degenerate distribution, singular distribution)

Synonyme de *dégénérée (loi)*.

Loi d'une variable aléatoire certaine.

#### Formulaire

Un paramètre réel  $x_0$ .  $X$  est la variable aléatoire qui prend la valeur  $x_0$  avec probabilité 1.



#### ► Valeurs caractéristiques

- espérance :  $E(X) = x_0$
- variance :  $\text{Var}(X) = 0$
- écart-type :  $\sigma(X) = 0$

► Utilisation

Cette loi intervient dans un certain nombre de situations comme cas limite ; elle peut également intervenir comme variable aléatoire limite (au sens de la « convergence en loi ») : par exemple comme limite de la moyenne dans les épreuves répétées.

### Smirnov (test de)

Voir [Kolmogorov-] Smirnov (test de).

### sommation (formules de)

(summation formulae)

#### Formules utiles pour des calculs relatifs à des probabilités discrètes

- Somme des  $n$  premiers entiers

$$1 + 2 + 3 + \dots + n = \sum_{k=1}^n k = \frac{n(n+1)}{2}$$

- Somme des  $n$  premiers termes d'une progression arithmétique

$$a + (a+b) + (a+2b) + \dots + (a+(n-1)b) = \sum_{k=0}^{n-1} (a+kb) = na + \frac{n(n-1)}{2}b$$

- Somme des carrés des  $n$  premiers entiers

$$1^2 + 2^2 + 3^2 + \dots + n^2 = \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}$$

- Somme des cubes des  $n$  premiers entiers

$$1^3 + 2^3 + 3^3 + \dots + n^3 = \sum_{k=1}^n k^3 = \frac{n^2(n+1)^2}{4}$$

- Somme des  $n$  premiers termes d'une progression géométrique ( $\alpha \neq 1$ )

$$1 + \alpha + \alpha^2 + \dots + \alpha^{n-1} = \sum_{k=0}^{n-1} \alpha^k = \frac{1-\alpha^n}{1-\alpha}$$

- Somme *infinie* des termes d'une progression géométrique ( $|\alpha| < 1$ )

$$1 + \alpha + \alpha^2 + \dots + \alpha^n + \dots = \sum_{k=0}^{\infty} \alpha^k = \frac{1}{1-\alpha}$$

- Somme *infinie* liée aux termes d'une progression géométrique ( $|\alpha| < 1$ )

$$\alpha + 2\alpha^2 + \dots + n\alpha^n + \dots = \sum_{k=0}^{\infty} (k+1)\alpha^{k+1} = \frac{\alpha}{(1-\alpha)^2}$$

- Somme des inverses des  $n$  premiers entiers (encadrement numérique)

$$\ln\left(n + \frac{1}{2}\right) + 0,5772 \leq \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} = \sum_{k=1}^n \frac{1}{k} \leq \ln\left(n + \frac{1}{2}\right) + 0,5946 \quad (n \geq 1)$$

$$\ln\left(n + \frac{1}{2}\right) + 0,577\ 215 \leq \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} = \sum_{k=1}^n \frac{1}{k} \leq \ln\left(n + \frac{1}{2}\right) + 0,577\ 593 \quad (n \geq 10)$$

**somme de deux variables aléatoires (réelles)**

(sum ...)

Des formules générales existent pour les probabilités et les densités, mais on se limitera ici au cas de v.a. indépendantes.

**Variables aléatoires réelles discrètes et indépendantes**

On suppose X définie par l'ensemble  $\{x_i\}$  des valeurs prises et par les probabilités ponctuelles  $p_i = P(X = x_i)$ .

On suppose Y définie par l'ensemble  $\{y_j\}$  des valeurs prises et par les probabilités ponctuelles  $q_j = P(Y = y_j)$ .

On pose  $S = X + Y$ .

Formule 1 :

$$P(S = s) = \sum_{x_i + y_j = s} p_i q_j$$

(comprendre : sommer sur l'ensemble des couples  $(i, j)$  tels que  $x_i + y_j = s$ ).

Formule 2 équivalente :

$$P(S = s) = \sum_i P(X = x_i)P(Y = s - x_i)$$

(comprendre : prendre  $P(Y = s - x_i) = P(Y = y_j)$  lorsqu'il existe une valeur  $y_j$  telle que  $y_j = s - x_i$ , prendre  $P(Y = s - x_i) = 0$  lorsqu'il n'existe pas).

**Variables aléatoires absolument continues et indépendantes**

On suppose X définie par sa densité  $f(x)$ , et on pose  $F(x)$  sa fonction de répartition.

On suppose Y définie par sa densité  $g(y)$ , et on pose  $G(y)$  sa fonction de répartition.

On pose  $S = X + Y$ .

La densité  $h(s)$  de S est donnée par :

$$h(s) = \int_{-\infty}^{\infty} f(x)g(s - x)dx = \int_{-\infty}^{\infty} g(y)f(s - y)dy$$

(nota : si X et/ou Y prend ses valeurs sur une partie seulement de  $\mathbf{R}$ , avec donc sa densité nulle à l'extérieur, ne pas oublier de réduire en conséquence l'intervalle d'intégration).

La fonction de répartition  $H(s)$  de S est donnée par :

$$H(s) = \int_{-\infty}^{\infty} f(x)G(s - x)dx = \int_{-\infty}^{\infty} g(y)F(s - y)dy$$

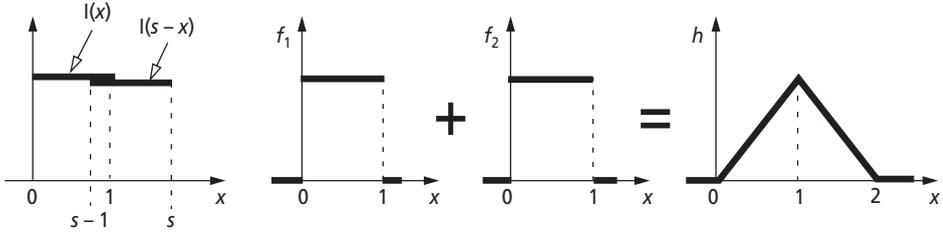
**Exemple** On considère deux v.a.  $X_1$  et  $X_2$  uniformes sur  $[0, 1]$  et on définit leur somme  $S = X_1 + X_2$ . Donner la densité de probabilité  $h$  de S.

Il existe une notation mathématique spécifique  $I_{[a, b]}(x)$  pour représenter la « fonction caractéristique » (ou indicatrice) de l'intervalle  $[a, b]$ , i.e. la fonction égale à 1 sur cet intervalle et nulle en dehors. On a donc ici, pour  $f_1$  et  $f_2$  les densités de  $X_1$  et  $X_2$  :  $f_1 = f_2 = I_{[0, 1]}$ . On peut donc écrire :

$$h(s) = \int_{-\infty}^{\infty} I_{[0, 1]}(x)I_{[0, 1]}(s - x)dx$$

Donc  $h(s)$  est simplement égale à la longueur de l'intervalle sur lequel  $I_{[0, 1]}(x)$  et  $I_{[0, 1]}(s - x)$  sont simultanément non nulles. Si  $s < 0$  ou  $s > 2$ , cet intervalle est vide et  $h(s) = 0$  ;

si  $0 \leq s \leq 1$ , cet intervalle est  $[0, s]$  et  $h(s) = s$  ; si  $1 \leq s \leq 2$ , cet intervalle est  $[s - 1, 1]$  et  $h(s) = 2 - s$ .



À un niveau plus approfondi, il faut signaler que l'intégrale  $\int_{-\infty}^{\infty} f(x)g(s-x)dx$  représente le « produit de convolution » des fonctions  $f$  et  $g$ , qui se traduit de façon intéressante sur les fonctions caractéristiques. Si  $\varphi_X, \varphi_Y$  et  $\varphi_S$  sont respectivement les fonctions caractéristiques de  $X, Y$  et  $S$ , on a :

$$\varphi_S(t) = \varphi_X(t) \varphi_Y(t),$$

formule qui se généralise immédiatement à la somme d'un nombre fini quelconque de v.a. indépendantes.

**Indicateurs**

Cas général :

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

$$\text{Var}(X + Y) = \text{Var}(X) + 2 \text{Cov}(X, Y) + \text{Var}(Y).$$

Cas où  $X$  et  $Y$  sont indépendantes :

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

La formule  $E(X + Y) = E(X) + E(Y)$ , pour laquelle l'indépendance n'est pas nécessaire, est, pour cette raison, tout à fait remarquable et souvent l'unique moyen d'aborder l'étude de certains problèmes compliqués.

Voir opérations sur les variables aléatoires.

**somme de  $n$  variables aléatoires**

**Indicateurs**

Cas général :

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

Cas où  $X_1, X_2, \dots, X_n$  sont 2 à 2 indépendantes :

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)$$

Même remarque que plus haut sur la formule qui donne  $E(X_1 + X_2 + \dots + X_n)$ .

Voir épreuves répétées.

**sondage**

(sampling)

Opération de recueil de données pour un échantillon d'individus d'une population.

Ce mot est l'exact synonyme d'échantillonnage, même si les habitudes font utiliser préférentiellement l'un ou l'autre mot selon les situations.

Un des chapitres de la statistique mathématique étudie les techniques de constitution d'échantillons qui, tout en restant « représentatifs » de la population, permettent d'améliorer

la précision des estimations. Certaines de ces techniques sont dites *aléatoires* (tirage aléatoire simple, sondage stratifié, sondage par grappes, sondage systématique), les autres *empiriques* (méthode des quotas, méthode des unités-types).

## Spearman

Voir *corrélation des rangs (coefficient de – de Spearman)*.

## statistique

(*statistics*)

### ► Sens général

Branche des mathématiques qui traite de l'étude de « données », résultats obtenus lors d'expérimentation ou d'observations de phénomènes aléatoires ou mal prévisibles. La statistique se divise en trois grandes parties :

- recueillir, décrire, présenter, résumer les données sous une forme qui rend leur exploitation commode et efficace (*statistique descriptive*),
- analyser ces données de façon à obtenir des informations sur le modèle probabiliste (« loi » de probabilité avec ses paramètres) qui a régi leur production (*estimation des paramètres*),
- contrôle et validation du modèle probabiliste reconstitué (*tests d'hypothèses*),

Les parties 2 et 3 sont parfois globalement qualifiées de *statistique inférentielle*.

Chacune de ces parties débouche, à son niveau propre, sur des possibilités de prévision et de décision.

L'usage le plus courant est d'écrire en français statistique « sans s », en référence à « méthode statistique », pour différencier *la* statistique *des* statistiques (sens n° 2 ci-dessous).

### ► Sens technique faible

Ensemble ou tableau de données recueillies lors d'expérimentation ou d'observations de phénomènes aléatoires ou mal prévisibles.

### ► Sens technique fort

Désigne en calcul des probabilités une variable aléatoire fonction d'un échantillon ( $X_1, X_2, \dots, X_n$ ) de variables aléatoires. Ce pourra être dans certains cas un « résumé » de l'échantillon, auquel on appliquera les outils mathématiques de la statistique, dans d'autres cas la « variable de décision » d'un test d'hypothèse.

## statistique d'ordre

Voir *ordre (statistique d')*.

## statistique (variable)

Voir *variable statistique*.

## Stirling (formule de)

Voir *factorielle*.

## stochastique

(*stochastic*)

Adjectif qui signifie aléatoire ou lié au calcul des probabilités, et qui figure dans des locutions consacrées par l'usage. En particulier, indépendance stochastique signifie indépendance « en probabilité », par opposition à indépendance physique, et processus stochastique est l'appellation officielle des « fonctions aléatoires » dépendant du temps.

## Student (William)

Voir Gosset (William).

### Student (loi du t de)

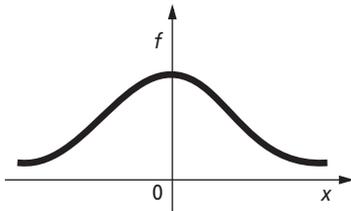
(Student t distribution)

Loi d'une variable aléatoire continue utilisée pour le contrôle des tests de comparaison de deux espérances mathématiques.

#### Formulaire

Un paramètre entier  $n \geq 1$  qui représente le nombre de « degrés de liberté » ; valeurs réelles.

- Loi de probabilité de X (qui sera notée ultérieurement  $T_n$ )



densité

fonction de répartition

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

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

- Valeurs caractéristiques

- espérance :  $E(X) = 0$
- variance :  $\text{Var}(X) = \frac{n}{n-2}$  (si  $n \geq 3$ )
- écart-type :  $\sigma(X) = \sqrt{\frac{n}{n-2}}$  (si  $n \geq 3$ )

- Cas particulier : lorsque  $n = 1$ , on obtient la *loi de Cauchy* standard.

**Théorème.** Lorsque  $n$  tend vers l'infini, la variable aléatoire de Student à  $n$  degrés de liberté converge en loi vers une v.a. normale centrée réduite.

- Utilisations

En théorie, une v.a. de Student à  $n$  degrés de liberté peut être définie comme le quotient

$$T_n = \frac{X}{\sqrt{\frac{Y_n}{n}}}, \text{ où } X \text{ est une v.a. normale centrée réduite et } Y_n \text{ une v.a. du khi-deux à } n \text{ degrés de}$$

liberté,  $X$  et  $Y_n$  indépendantes.

Cette loi intervient dans les tests de comparaison de deux espérances en raison de la propriété fondamentale suivante : si  $X_1, X_2, \dots, X_n$  sont  $n$  v.a. normales identiques (d'espérance  $\mu$  et

d'écart-type  $\sigma$ ) indépendantes, si  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$  est la variable aléatoire moyenne, et si

$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2$  est la variable aléatoire estimateur « débiaisé » de la variance,

alors, d'une part  $M_n$  et  $S_n^2$  sont des v.a. indépendantes, d'autre part le quotient  $\frac{M_n - \mu}{\left(\frac{S_n}{\sqrt{n}}\right)}$  suit

une loi de Student à  $n - 1$  degrés de liberté.

**Remarque :** si l'on prenait pour  $S_n^2$  l'estimateur biaisé de la variance, le dénominateur devrait être  $\frac{S_n}{\sqrt{n-1}}$ .

## Student (test de), test t

([Student] t test)

Test *paramétrique* qui compare, soit la moyenne observée d'un échantillon statistique à une valeur fixée, soit les moyennes observées de deux échantillons statistiques (en fait, un vocabulaire rigoureux devrait faire référence aux espérances mathématiques des lois et non pas aux moyennes (observées)). Il permet également de comparer, soit la probabilité observée à partir d'un échantillon statistique à une valeur donnée, soit les probabilités observées à partir des deux échantillons.

Chacun de ces tests peut être effectué, soit en bilatéral (cas général), soit en unilatéral (cas où l'on sait, ou bien où l'on postule, que l'une des inégalités imaginables est exclue).

### ► Comparaison de moyennes

Dans ce cas, ce test ne s'applique en toute rigueur qu'à des échantillons issus de variables aléatoires normales, mais il est « robuste », ce qui signifie qu'il demeure applicable à d'autres lois – à condition toutefois de prendre certaines précautions qui sont décrites ci-dessous.

#### test bilatéral de comparaison

d'une espérance mathématique  $\mu$  à une valeur fixée  $\mu_0$

- Données. Un échantillon  $(x_1, x_2, \dots, x_n)$  de  $n$  valeurs observées d'une variable aléatoire numérique  $X$  d'espérance mathématique  $\mu$ .

- Hypothèse testée.  $H_0 = \langle \mu = \mu_0 \rangle$  contre  $H_1 = \langle \mu \neq \mu_0 \rangle$

- Déroulement technique du test

1. On calcule la moyenne  $m_{\text{obs}}$  de l'échantillon.

2. On calcule la variance non biaisée  $s_{\text{obs}}^2$  de l'échantillon.

3. On pose  $s^* = \frac{s_{\text{obs}}}{\sqrt{n}}$  et on calcule la valeur observée de la variable de test :

$$t = \frac{|m_{\text{obs}} - \mu_0|}{s^*}.$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Student, elles dépendent du nombre de degrés de liberté de l'échantillon :  $ddl = n - 1$ , et du risque  $\alpha$ .

• Conditions et précautions

- En théorie X doit être une v.a. normale, donc aucune précaution si c'est le cas ;
- lorsque ce n'est pas le cas, le test est robuste et reste applicable si n est « assez grand », la condition  $n \geq 30$  est traditionnelle (en fait, on peut descendre en-dessous si la loi de X est continue et/ou symétrique).

— test bilatéral de comparaison de deux espérances mathématiques  $\mu_X$  et  $\mu_Y$  —

• Données. Deux séries :

- un échantillon  $(x_1, x_2, \dots, x_{n_X})$  de  $n_X$  valeurs observées d'une variable aléatoire numérique X d'espérance mathématique  $\mu_X$  ;
- un échantillon  $(y_1, y_2, \dots, y_{n_Y})$  de  $n_Y$  valeurs observées d'une variable aléatoire numérique Y d'espérance mathématique  $\mu_Y$ .

• Hypothèse testée.  $H_0 = \langle \mu_X = \mu_Y \rangle$  contre  $H_1 = \langle \mu_X \neq \mu_Y \rangle$ .

• Déroulement technique du test

1. On calcule les moyennes observées  $m_X$  et  $m_Y$  des deux échantillons.
- 2a. On calcule les variances non biaisées  $s_X^2$  et  $s_Y^2$  des deux échantillons.
- 2b. On calcule une variance commune pondérée :

$$s^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}.$$

3. On pose  $s^* = s \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$  et on calcule la valeur observée de la variable de test :

$$t = \frac{|m_X - m_Y|}{s^*}.$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi de Student, elles dépendent du nombre total de degrés de liberté de l'échantillon :  $ddl = n_X + n_Y - 2$ , et du risque  $\alpha$ .

• Conditions et précautions

- En théorie X et Y doivent être des v.a. normales, mais le test est robuste et reste applicable à des v.a. non normales si  $n_X$  et  $n_Y$  sont « assez grands » (cf. ci-dessus) ;
- en outre X et Y doivent avoir même variance : lorsque  $n_X$  et  $n_Y$  ne sont pas très grands et qu'il semble que les variances ne soient pas égales, il faut faire précéder le test de comparaison des deux espérances par un test de comparaison des variances ; lorsque  $n_X$  et  $n_Y$  sont « grands », on peut de nouveau utiliser la robustesse et passer outre à l'inéga-

lité des variances (certains manuels proposent l'estimation  $s^* = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$ ).

---

test unilatéral de comparaison  
d'une espérance mathématique  $\mu$  à une valeur fixée  $\mu_0$

---

• Données. Un échantillon  $(x_1, x_2, \dots, x_n)$  de  $n$  valeurs observées d'une variable aléatoire numérique  $X$  d'espérance mathématique  $\mu$ .

• Hypothèse testée. *Par exemple*  $H_0 = \langle \mu \leq \mu_0 \rangle$  contre  $H_1 = \langle \mu > \mu_0 \rangle$ .

• Déroulement technique du test

On calcule la moyenne  $m_{\text{obs}}$  de l'échantillon :

- si  $m_{\text{obs}} \leq \mu_0$  il y a NON-REJET de l'hypothèse  $H_0$  et le test est terminé ;
- si  $m_{\text{obs}} > \mu_0$  on poursuit à l'identique du cas bilatéral jusqu'au calcul de la variable de test  $t$ .

La seule différence est que – pour le risque  $\alpha$  – les valeurs de référence de  $t$  sont à lire dans les tables de la loi de Student dans la colonne  $2\alpha$  (avec le même nombre de degrés de libertés  $\text{ddl} = n - 1$ ).

• Conditions et précautions : Les mêmes que ci-dessus.

---

— test unilatéral de comparaison de deux espérances mathématiques  $\mu_X$  et  $\mu_Y$  —

---

• Données. Deux séries :

- un échantillon  $(x_1, x_2, \dots, x_{n_X})$  de  $n_X$  valeurs observées d'une variable aléatoire numérique  $X$  d'espérance mathématique  $\mu_X$  ;
- un échantillon  $(y_1, y_2, \dots, y_{n_Y})$  de  $n_Y$  valeurs observées d'une variable aléatoire numérique  $Y$  d'espérance mathématique  $\mu_Y$ .

• Hypothèse testée. *Par exemple*  $H_0 = \langle \mu_X \leq \mu_Y \rangle$  contre  $H_1 = \langle \mu_X > \mu_Y \rangle$ .

• Déroulement technique du test

On calcule les moyennes observées  $m_X$  et  $m_Y$  des deux échantillons :

- si  $m_X \leq m_Y$  il y a NON-REJET de l'hypothèse  $H_0$  et le test est terminé ;
- si  $m_X > m_Y$  on poursuit à l'identique du cas bilatéral jusqu'au calcul de la variable de test  $t$ .

La seule différence est que – pour le risque  $\alpha$  – les valeurs de référence de  $t$  sont à lire dans les tables de la loi de Student dans la colonne  $2\alpha$  (avec le même nombre de degrés de libertés  $\text{ddl} = n - 1$ ).

• Conditions et précautions : Les mêmes que ci-dessus.

---

Voir *moyenne, variance*.

➤ Comparaison de pourcentages / probabilités

Comme le paramètre  $p$  d'une v.a. de Bernoulli  $X$  est à la fois la probabilité que  $X = 1$  et l'espérance  $E(X)$ , le fonctionnement du test de Student pour les probabilités est identique au fonctionnement du test de Student pour les espérances, à deux détails près : l'absence d'écart-type à estimer, et quelques adaptations des conditions et précautions.

Un usage ancien utilise le mot pourcentage (à connotation statistique) plutôt que probabilité (à connotation... probabiliste, si l'on ose dire).

Ce test peut être effectué, soit en bilatéral (cas général), soit en unilatéral (cas où l'on sait, ou bien où l'on postule, que l'une des inégalités imaginables est exclue).

— test bilatéral de comparaison d'une probabilité  $P(A)$  à une valeur fixée  $p_0$  —

- Données. Un échantillon de  $n$  observations, sur lesquelles  $A$  a été observé  $k$  fois.
- Hypothèse testée.  $H_0 = \langle P(A) = p_0 \rangle$  contre  $H_1 = \langle P(A) \neq p_0 \rangle$ .
- Déroulement technique du test
  - Si  $n$  est petit, on utilise directement la valeur de  $k$  et on se réfère soit à des tables numériques spéciales soit au calcul de la probabilité correspondante de la loi binomiale  $\mathcal{B}(n, p_0)$ .
  - Si  $n$  est grand :

1. On calcule l'estimation de  $P(A)$  :  $p = \frac{k}{n}$  ;

2. On pose  $s^* = \sqrt{\frac{p(1-p)}{n}}$  et on calcule la valeur observée de la variable de test :

$$t = \frac{|p - p_0|}{s^*}.$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi normale, elles ne dépendent que du risque  $\alpha$ .

- Conditions et précautions

Il faut que la distribution ne soit pas « trop » dissymétrique, ce qui se traduit par la double condition traditionnelle  $np \geq 10$ ,  $n(1-p) \geq 10$  (que l'on peut sans grand risque affaiblir en  $np \geq 5$ ,  $n(1-p) \geq 5$ ).

Lorsqu'on fait fonctionner ce test avec une valeur de  $n$  « moyennement grande » mais néanmoins avec la loi normale comme loi de référence, il faut faire une correction qui « étale » la probabilité binomiale d'obtenir la valeur entière  $k$  en une probabilité pour la loi normale (avant centrage et réduction) d'obtenir une valeur entre  $k - 0,5$  et  $k + 0,5$  : on prend donc

$$t = \frac{|p - p_0| + \frac{1}{2n}}{s^*}.$$

———— test bilatéral de comparaison de deux probabilités  $P(A_1)$  et  $P(A_2)$  ————

- Données
  - Un premier échantillon de  $n_1$  observations, sur lesquelles  $A_1$  a été observé  $k_1$  fois ;
  - un second échantillon de  $n_2$  observations, sur lesquelles  $A_2$  a été observé  $k_2$  fois.
- Hypothèse testée.  $H_0 = \langle P(A_1) = P(A_2) \rangle$  contre  $H_1 = \langle P(A_1) \neq P(A_2) \rangle$ .
- Déroulement technique du test

1a. On calcule les estimations de  $P(A_1)$  :  $p_1 = \frac{k_1}{n_1}$ , et de  $P(A_2)$  :  $p_2 = \frac{k_2}{n_2}$ .

1b. On calcule une estimation globale commune :  $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{k_1 + k_2}{n_1 + n_2}$ .

2. On pose  $s^* = \sqrt{p(1-p)} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$  et on calcule la valeur observée de la variable de test :

$$t = \frac{|p_1 - p_2|}{s^*}.$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi normale, elles ne dépendent que du risque  $\alpha$ .

• Conditions et précautions

- Il n'y a pas de tables numériques spéciales pour les petites valeurs des effectifs, et il faut donc supposer  $n_1$  et  $n_2 \geq 30$  ;
- il faut que les distributions ne soient pas « trop » dissymétriques, ce qui se traduit par la quadruple condition  $n_1 p_1 \geq 5$  ou  $10$ ,  $n_1(1-p_1) \geq 5$  ou  $10$ ,  $n_2 p_2 \geq 5$  ou  $10$ ,  $n_2(1-p_2) \geq 5$  ou  $10$ .

— test unilatéral de comparaison d'une probabilité  $P(A)$  à une valeur fixée  $p_0$  —  
 Test unilatéral de comparaison de deux probabilités  $P(A_1)$  et  $P(A_2)$

La procédure de test et les calculs sont identiques au cas bilatéral à une seule exception près : la valeur de référence de  $t$  – pour le risque  $\alpha$  – est à lire dans les tables de la loi normale dans la colonne  $2\alpha$ .

## suites (test des)

(run test)

Test non paramétrique qui contrôle l'indépendance des valeurs successives d'un échantillon d'une variable dichotomique en examinant le nombre de suites de valeurs identiques, dans le cas où la succession des valeurs observées est pertinente. Ce test est parfois employé pour contrôler l'indépendance des valeurs successives des résidus dans une régression en testant les suites de leurs signes.

— test d'indépendance globale d'un échantillon dichotomique séquentiel —

- Données. Un échantillon séquentiel de  $n_1$  observations de l'évènement  $A_1$  et  $n_2$  observations de l'évènement contraire  $A_2$ .
- Hypothèse testée.  $H_0 =$  « les valeurs successives sont indépendantes » contre  $H_1$  alternative.
- Déroulement technique du test

1. On compte le nombre  $r$  de suites d'observations identiques qui composent les  $n_1 + n_2$  observations totales ;

2. On pose  $m_r = \frac{2n_1 n_2}{n_1 + n_2} + 1$ ,  $s_r = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$  et on calcule la valeur observée de la variable de test :

$$t = \frac{|r - m_r|}{s_r}$$

Les valeurs de référence de la variable de test sont à lire dans les tables de la loi normale, elles ne dépendent que du risque  $\alpha$ .

• Conditions et précautions

Aucunes.

**système complet d'évènements** (*complete set of events*)

Synonyme de *partition*.

(*partition, disjoint decomposition*)

Étant donné un espace probabilisable  $(\Omega, \mathcal{A})$ , on appelle système complet d'évènements – ou encore partition de  $\Omega$  – un ensemble fini  $(A_i)$  d'évènements 2 à 2 disjoints (2 à 2 incompatibles) et dont la réunion est  $\Omega$ .

Ainsi, chaque fois que l'on effectue l'épreuve, un et un seul des  $A_i$  est réalisé.

Voir *Bayes (formule de)*, *totales (formule des probabilités)*.



## t (test)

Voir *Student (test de)*.

## tableau

(*table*)

Mode de présentation des données recueillies en vue d'une analyse statistique.

## tableau par classes et effectifs

Présentation des données statistiques regroupées par classes (qui peuvent chacune correspondre à une ou plusieurs valeurs ponctuelles, ou bien être une tranche continue de valeurs). Un tableau par classes et effectifs se dispose en deux lignes ou deux colonnes, l'une contenant l'identification des classes, l'autre les effectifs des classes. Il est possible d'adjoindre des lignes ou des colonnes supplémentaires, par exemple pour donner les fréquences relatives.

## tableau de contingence

(*contingency table*)

Tableau à double entrée qui permet de représenter les *effectifs* d'une population ou d'un échantillon répartie selon les classes (en nombre fini) de deux variables.

Synonyme de *tableau croisé*.

## taille [d'un échantillon]

(*[sample] size*)

Nombre d'individus de l'échantillon. Synonyme d'*effectif total*.

## taux de défaillance, de panne, de mort

Voir *défaillance (taux de)*.

## Tchebychev (Pafnouti)

Mathématicien russe (1821–1894). Il contribua au calcul de nombreuses approximations numériques et fit également des travaux en théorie des nombres.

## Tchebychev (inégalité de)

Voir *Bienaymé-Tchebychev (inégalité de)*.

## temporelle (série)

Voir *série chronologique*.

## temps d'attente (variable aléatoire) (*waiting time, first passage time*)

Variante aléatoire qui modélise dans des « processus » le temps ou le délai avant la survenue d'un évènement fixé (a priori le premier, mais éventuellement le  $k$ -ième,  $k$  donné). Un temps

d'attente peut être continu (temps physique ordinaire), ou discret (temps physique discrétisé, ou bien un numéro d'ordre ou un nombre de parties). Les lois les plus classiques pour une variable aléatoire temps d'attente sont la *loi exponentielle*, la *loi d'Erlang*, la *loi de Weibull* (cas continu), et la *loi géométrique* (cas discret).

## test d'hypothèse

(*hypothesis testing*)

Procédure, basée sur l'analyse statistique de résultats expérimentaux, qui permet de décider (avec un risque d'erreur) entre deux hypothèses.

Dans la pratique, diverses raisons conduisent à privilégier l'une des hypothèses, appelée « hypothèse zéro (ou nulle) » et notée  $H_0$  (l'autre hypothèse étant simplement l'alternative). Par surcroît, des raisons mathématiques font que dans la plupart des tests l'hypothèse  $H_0$  est une hypothèse ponctuelle et que la conclusion ne peut jamais être son acceptation *stricto sensu*.

La décision devient alors fortement dissymétrique : ou bien l'on rejette  $H_0$ , ou bien l'on ne rejette pas  $H_0$ .

Et le risque devient lui aussi fortement dissymétrique : ou bien l'on rejette  $H_0$  alors qu'elle était vraie, risque appelé risque de première espèce ou *risque d'erreur* ; ou bien l'on ne rejette pas  $H_0$  alors qu'elle était fautive, risque appelé risque de seconde espèce ou *risque de manque de puissance*. Bien entendu, tout gain sur l'un des risques se « paye » d'une perte sur l'autre.

On introduit souvent une distinction entre tests *paramétriques* et tests *non paramétriques*. Lorsque l'hypothèse  $H_0$  à tester nécessite une hypothèse préalable sur la loi de probabilité (ce sera le plus souvent une hypothèse de normalité) et implique des paramètres de cette loi, on dit que le test est paramétrique. Pour autant, la plupart des tests paramétriques sont robustes, *i.e.* supportent (dans des limites raisonnables que l'on peut préciser) que la loi réelle s'écarte de la loi nominale du test. Lorsque l'hypothèse  $H_0$  ne nécessite aucune hypothèse préalable sur la loi de probabilité, on dit que le test est non paramétrique. À risque d'erreur égal, un test non paramétrique est en général moins puissant que le test paramétrique concurrent, mais cette perte de puissance est souvent assez faible. Cette distinction est surtout pertinente pour les tests de comparaison (d'espérances ou moyennes, de variances). Néanmoins, il y a quelque contradiction à qualifier de non paramétriques certains des tests qui contrôlent l'égalité de deux espérances (qui sont des paramètres !). Aussi, certains auteurs préfèrent parler de test libre lorsque la validité du test ne dépend d'aucune condition sur la loi.

Voir *Neyman-Pearson (lemme de)*.

## tests d'hypothèses (méthodologie des)

La situation de « départ » est la suivante : on dispose d'une série ou d'un tableau de valeurs observées, et on se pose une question, éventuellement en langage « ordinaire ».

1. Étape préliminaire. Il faut tout d'abord identifier la ou les variables aléatoires sous-jacentes aux valeurs observées, et leur loi, puis traduire la question posée en une hypothèse  $H_0$  énoncée formellement en référence, soit à un (ou plusieurs) paramètre(s) de la loi identifiée, soit globalement à la nature de la loi. Cela étant fait, on peut alors choisir un test d'hypothèse adapté au problème étudié.

2. Exécution du test : calculs. À chaque « variété » de test est associée une variable aléatoire positive qui sera d'autant plus grande que les observations s'« écartent » de l'hypothèse  $H_0$ . Cette variable de test est souvent représentée par une notation traditionnelle :  $t$ ,  $\chi^2$ ,  $F$ ,  $U$ , etc. Les formules qui sont indiquées pour chaque variété de test permettent de calculer la valeur observée de la variable de test, que l'on notera ci-dessous  $y_{\text{obs}}$ .

3. Exécution du test : formulation de la conclusion. De façon schématique (la pratique pourra être plus souple et plus détaillée), la conclusion d'un test d'hypothèse est la décision de rejet ou de non-rejet de l'hypothèse  $H_0$ . À cette décision est toujours associé un risque d'erreur, appelé niveau du test et traditionnellement noté  $\alpha$ . Ce niveau mesure le risque de se tromper dans le pari qu'est la conclusion du test. Il existe deux manières différentes de « gérer » ce risque d'erreur.

Conclusion si le risque d'erreur est fixé *a priori*. Dans ce cas, on se fixe le niveau  $\alpha$  avant de calculer  $y_{\text{obs}}$ . On doit alors chercher dans les tables numériques du test, en fonction de  $\alpha$  et éventuellement d'autres paramètres (nombre de degrés de liberté notamment) la valeur critique  $y(\alpha)$  de la variable de test. Trois conclusions sont alors possibles :

- si  $y_{\text{obs}} \ll y(\alpha)$  : non-rejet de  $H_0$  ;
- si  $y_{\text{obs}}$  est inférieure à  $y(\alpha)$  mais néanmoins assez grande (par exemple comprise entre  $y(0,20)$  et  $y(\alpha)$ ) : affirmation que l'hypothèse  $H_0$  est probablement fausse, mais que néanmoins on ne peut pas la rejeter sans courir un risque supérieur à  $\alpha$  (le non-rejet de  $H_0$  sans commentaire ne serait pas très judicieux) ;
- si  $y_{\text{obs}} \geq y(\alpha)$  : rejet de  $H_0$ .

Conclusion si le risque d'erreur n'est pas fixé *a priori*. Dans ce cas, et après avoir calculé  $y_{\text{obs}}$ , on cherche dans les tables numériques du test la valeur exacte (mais approximative – une grande précision n'a aucun intérêt) d'un risque  $\beta$  tel que  $y(\beta) = y_{\text{obs}}$  (cette valeur  $\beta$  s'appelle la *probabilité critique*). La conclusion est alors unique mais s'énonce par deux phrases complémentaires :

- si on prend un risque d'erreur  $\alpha < \beta$ , on ne peut pas rejeter  $H_0$  ;
- si on prend un risque d'erreur  $\alpha > \beta$ , on peut rejeter  $H_0$  ;

Selon la nature du problème et les enjeux de la situation réelle, on en restera à cette conclusion « académique », ou bien on choisira *a posteriori* un risque et on donnera une conclusion tranchée.

### théorique

(theoretical)

Lorsque cet adjectif n'est pas employé dans son sens général de la langue courante, il qualifie les paramètres des lois de probabilités (espérance théorique, variance théorique, ...), par opposition aux paramètres « observés » alias « empiriques » des distributions statistiques.

### tirage

(drawing)

Opération qui consiste en probabilités à effectuer une « épreuve » dont le résultat est un *évènement élémentaire*, et en statistique à extraire (par un procédé en principe aléatoire) un *individu* d'une population.

Dans la présentation « extraction », on distingue les tirages AVEC remise, où les probabilités restent inchangées, et les tirages SANS remise, où les probabilités varient (un peu) au fur et à mesure des tirages.

Voir *binomiale (loi)*, *hypergéométrique (loi)*.

### totales (formule des probabilités)

(total probabilities formula)

Étant donné un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , une partition (ou système complet d'évènements)  $H_1, H_2, \dots, H_k$  de  $\Omega$ , et un évènement  $B \in \mathcal{A}$ , on a :

$$P(B) = P(H_1)P(B|H_1) + P(H_2)P(B|H_2) + \dots + P(H_k)P(B|H_k)$$

On notera que  $B$  est la réunion des évènements  $B \cap H_i$ , qui sont 2 à 2 disjoints.





## uniforme continue (loi)

(uniform distribution, rectangular distribution)

Synonyme de *rectangulaire (loi)*.

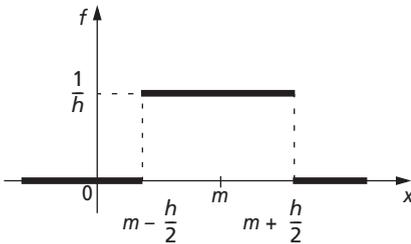
Loi d'une variable aléatoire continue à densité constante sur un intervalle.

### Formulaire

Deux paramètres réels :  $m \in \mathbf{R}$  ;  $h \in \mathbf{R}_+^*$ .

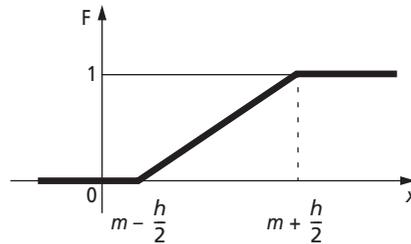
Valeurs concentrées sur l'intervalle  $I = \left[ m - \frac{h}{2}, m + \frac{h}{2} \right]$ .

#### ► Loi de probabilité



densité

$$f(x) = \begin{cases} 0 & \text{si } x < m - \frac{h}{2} \\ \frac{1}{h} & \text{si } x \in I \\ 0 & \text{si } x > m + \frac{h}{2} \end{cases}$$



fonction de répartition

$$F(x) = \begin{cases} 0 & \text{si } x < m - \frac{h}{2} \\ \frac{1}{2} + \frac{1}{h}(x - m) & \text{si } x \in I \\ 1 & \text{si } x > m + \frac{h}{2} \end{cases}$$

#### ► Valeurs caractéristiques

- espérance :  $E(X) = m$
- variance :  $\text{Var}(X) = \frac{h^2}{12}$
- écart-type :  $\sigma(X) = \frac{h}{2\sqrt{3}}$

**uniforme discrète (loi)***(discrete uniform distribution)*

Loi d'une variable aléatoire discrète équirépartie entre un nombre fini de valeurs.

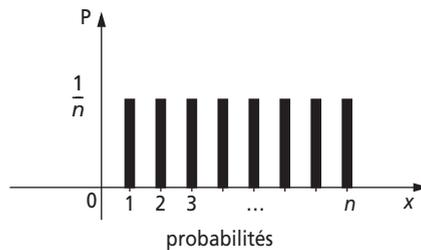
**Formulaire**

Version standardisée (valeurs équidistantes sur les entiers).

Un paramètre réel  $n$  (nombre de valeurs prises). Soit  $X$  la variable uniforme discrète de paramètre  $n$  : valeurs prises :  $1, 2, \dots, n$ .

## ► Loi de probabilité

$$P(X = k) = \frac{1}{n} \text{ pour } 1 \leq k \leq n.$$



## ► Valeurs caractéristiques

- espérance :  $E(X) = \frac{n+1}{2}$
- variance :  $\text{Var}(X) = \frac{n^2-1}{12}$
- écart-type :  $\sigma(X) = \frac{\sqrt{n^2-1}}{2\sqrt{3}}$

**unilatéral***(one-sided, single-tail)*Qualifie un *test paramétrique* où l'on teste l'hypothèse simple  $H_0 = \langle \theta = \theta_0 \rangle$  contre une hypothèse alternative unilatérale, soit  $H_1 = \langle \theta < \theta_0 \rangle$ , soit  $H_1 = \langle \theta > \theta_0 \rangle$ .Les tables sont généralement adaptées au fonctionnement des tests bilatéraux, et la valeur critique de la variable de test pour le risque unilatéral  $\alpha$  est égale à la valeur critique donnée par la table pour le risque bilatéral  $2\alpha$ .**unimodale (distribution)**Voir *mode*.



## v.a.

Abréviation universelle (en français) pour « variable aléatoire ».

## valeur

(value)

Quantité ou modalité que peut « prendre » une variable aléatoire ou une variable statistique (ou caractère), ou les divers paramètres et indicateurs des distributions probabilistes ou statistiques.

## valeur caractéristique

Voir *indicateur*.

## variable aléatoire

(random variable)

### ► Définitions de base

Grandeur (nombre réel, numéro d'ordre) ou « modalité » dont la valeur dépend (varie « en fonction ») du résultat d'une *épreuve* en probabilités. À chaque événement élémentaire  $\omega \in \Omega$ , correspond une valeur bien définie : cette « correspondance » est très exactement ce qu'on appelle dans le vocabulaire ensembliste une fonction ou une application (les deux mots sont synonymes, l'utilisation tantôt de l'un tantôt de l'autre est un simple problème d'usage).

De façon formelle, on se donne, d'une part un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , d'autre part un ensemble  $E$  (l'ensemble des « valeurs prises ») ; on appelle alors variable aléatoire (définie sur  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $E$ ) toute application  $X$  de  $\Omega$  dans  $E$  :

$$X : \omega \in \Omega \rightarrow X(\omega) \in E.$$

Il ne faut se laisser perturber ni par le vocabulaire ni par les notations. Une variable aléatoire n'est pas une variable mais une application (au sens ensembliste du mot). Le mot variable est malencontreux mais son usage est consacré... Par surcroît, l'application variable aléatoire ne se note jamais  $f$  ou  $g$ , mais  $X$  ou  $Y$  (ou par exemple  $N$  pour une variable aléatoire « de compte »). Ces notations, également consacrées par l'usage, sont particulièrement commodes et pédagogiques : les variables aléatoires se notent en lettres latines majuscules, et les valeurs prises (ou observées, en statistique) se notent par les minuscules correspondantes.

**Exemple** Le lancer de 3 pièces et le nombre de Pile(s). On considère l'espace  $\Omega = \{PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF\}$  constitué par les 8 événements élémentaires du lancer de 3 pièces. On définit l'application  $X$  de  $\Omega$  dans  $\mathbf{R}$  :

$$X(\omega) = \text{nombre de Pile(s) de l'évènement } \omega.$$

C'est une variable aléatoire. On a par exemple  $X(PPF) = 2$ ,  $X(FFF) = 0$ .

Étant donné un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et une variable aléatoire à valeurs dans  $E$ , il faudra ensuite « induire » une mesure de probabilité sur  $E$ . Pour cela, il faut préalablement faire de  $E$  l'espace fondamental d'un espace probabilisable  $(E, \mathcal{B})$ , et ensuite utiliser la

notion d'application réciproque étendue à l'ensemble des parties. Soit  $f$  une application d'un ensemble  $E$  dans un ensemble  $F$  : on définit  $f^{-1}$  application de  $\mathcal{P}(F)$  dans  $\mathcal{P}(E)$  par :

$$\text{si } B \subset F : f^{-1}(B) = \{x \in E \mid f(x) \in B\} = \{x \in E \mid \exists y \in B \text{ et } y = f(x)\}.$$

La partie  $f^{-1}(B)$  s'appelle l'image réciproque de  $B$ . Elle est toujours définie (s'il n'existe aucun  $x \in E$  tel que  $f(x) \in B$ ,  $f^{-1}(B)$  est la partie vide).

On peut alors compléter la définition formelle. On se donne un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et un espace probabilisable  $(E, \mathcal{B})$ . On définit une variable aléatoire comme une application  $X$  de  $\Omega$  dans  $E$  qui vérifie la propriété :

$$\forall B \in \mathcal{B}, f^{-1}(B) \in \mathcal{A}.$$

On dit que  $X$  est « mesurable ». On notera que la mesure de probabilité  $P$  ne joue aucun rôle dans cette définition (elle est « en réserve » pour la suite).

### ► Probabilité image

Étant donné un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , un espace probabilisable  $(E, \mathcal{B})$ , et une variable aléatoire  $X : \Omega \rightarrow E$ , la mesure  $P$  induit sur  $(E, \mathcal{B})$  une mesure, appelée mesure de probabilité image  $P_X$ , par :

$$\text{pour tout } B \in \mathcal{B} : P_X(B) = P(X^{-1}(B)).$$

La notation la plus courante est  $P(X \in B)$  plutôt que  $P_X(B)$  ou  $P(X^{-1}(B))$ . Cette notation s'adapte notamment lorsque  $B$  est un point ou un intervalle de  $\mathbf{R}$ , on écrira alors  $P(X = 3)$ ,  $P(a < X \leq b)$ ,  $P(X > 10)$ , etc.

**Exemple** On reprend le lancer de 3 pièces, décrit par l'espace  $\Omega = \{\text{PPP, PPF, PFP, PFF, FPP, FPF, FFF}\}$ . Ces 8 événements élémentaires sont supposés équiprobables (de probabilité  $\frac{1}{8}$  donc). On considère la variable aléatoire  $X =$  nombre de Pile(s), et on cherche la probabilité des événements  $X = 2$ ,  $X = 0$ . On a  $X^{-1}(\{2\}) = \{\text{PPF, PFP, FPP}\}$ , événement de  $\Omega$  de probabilité  $\frac{3}{8}$ , donc  $P(X = 2) = \frac{3}{8}$ . On a  $X^{-1}(\{0\}) = \{\text{FFF}\}$ , événement de  $\Omega$  de probabilité  $\frac{1}{8}$ , donc  $P(X = 0) = \frac{1}{8}$ .

### ► Typologie

Une première distinction sépare les v.a. *quantitatives*, dont les valeurs sont des « grandeurs », nombres réels, nombres complexes ou vecteurs le plus souvent, et les v.a. *qualitatives*, dont les valeurs sont des « modalités ». La différence fondamentale est que l'on peut effectuer sur les premières des opérations mathématiques, des additions notamment (ce qui permet de définir espérance mathématique ou moyenne, etc.), et que par contre on ne peut pas dépasser le stade descriptif pour les secondes.

Parmi les variables aléatoires quantitatives, les v.a. réelles sont d'une importance primordiale (d'autant que la généralisation aux v.a. complexes et aux v.a. vectorielles des concepts définis pour les v.a. réelles est le plus souvent très naturelle). Une étude théorique poussée conduit à identifier trois types fondamentaux de v.a. réelles, qui peuvent d'ailleurs se combiner. Mais dans la pratique, la quasi-totalité des variables aléatoires réelles utilisées appartient à deux catégories seulement.

Première catégorie : les *variables aléatoires réelles discrètes*, dont les valeurs forment un ensemble fini ou dénombrable. La loi de ces variables aléatoires est définie par la donnée de l'ensemble  $\{x_i\}$  des valeurs prises et des probabilités ponctuelles  $p_i = P(X = x_i)$ , qui seront données tantôt par leur liste, tantôt par une formule permettant de les calculer.

Deuxième catégorie : les *variables aléatoires réelles absolument continues*, appelées aussi *variables aléatoires réelles à densité*, cas particulier des v.a. réelles continues. On précise

généralement l'ensemble exact des valeurs prises ( $\mathbf{R}$  tout entier, les réels positifs, l'intervalle  $[0, 1]$ , etc.). La loi de ces variables aléatoires est définie, soit par leur densité  $f(x)$ , soit par leur fonction de répartition  $F(x)$ . La donnée de ces deux fonctions est bien sûr équivalente, puisque  $f(x) = F'(x)$  et inversement  $F(x) = \int_{-\infty}^x f(t) dt$ . Néanmoins, il arrive souvent que l'une des deux seulement ait une expression mathématique explicite.

## variable [statistique], caractère (variable, variate)

Grandeur (nombre réel, numéro d'ordre) ou « modalité » définie sur une « population » d'« individus » et susceptible d'être observée.

Une définition formelle introduirait un premier ensemble  $P$  (une « population » d'« individus ») et un deuxième ensemble  $E$  (l'ensemble des « valeurs prises »), puis énoncerait qu'une variable statistique est une application (au sens ensembliste) de  $P$  dans  $E$ . Cette définition formelle fait de la variable statistique (ou caractère) le concept statistique correspondant au concept probabiliste de variable aléatoire.

Lorsqu'une étude statistique s'effectue dans un contexte clairement aléatoire (incluant la reproductibilité dans des conditions identiques), les mots variable aléatoire, variable statistique et caractère sont rigoureusement synonymes et interchangeables. Lorsqu'une étude statistique s'effectue dans un contexte privilégiant l'aspect ensemble fixé de données (par exemple en économie), le mot caractère est le plus fréquemment utilisé.

Bien entendu, la typologie des variables statistiques est la même que celle des variables aléatoires, notamment en ce qui concerne la distinction quantitative/qualitative et la distinction discrète/continue.

**Exemple** Si on prend pour population le parc des automobiles européennes en circulation au 1<sup>er</sup> janvier 2004, on peut définir sur cette population de nombreux caractères, « qualitatifs » (constructeur, nationalité du possesseur, etc.) ou « quantitatifs » (puissance, âge, valeur marchande, etc.).

## variance (décomposition de la)

Voir *décomposition de la variance*.

## variance d'un échantillon statistique

Dans une situation d'observation d'un échantillon statistique, la variance est le principal indicateur numérique de dispersion. Parfois qualifiée de *variance observée* ou de *variance empirique*, elle est définie comme la moyenne numérique des carrés des écarts entre les valeurs observées et leur moyenne  $\bar{x}$ .

La variance d'un échantillon d'une variable aléatoire  $X$  se note le plus souvent  $\text{Var}_X$  ou  $\text{Var}_x$  ou  $s_X^2$  ou  $s_x^2$  ou  $\text{Var}$  ou  $s^2$  ou  $s_{\text{obs}}^2$  s'il n'y a aucun risque de confusion).

Formule pour  $n$  observations individualisées  $x_1, x_2, \dots, x_n$ , la moyenne  $\bar{x}$  ayant été préalablement calculée :

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{1}{n} \sum_{i=1}^{i=n} (x_i - \bar{x})^2$$

Formule pour  $n$  observations individuelles regroupées selon  $k$  classes d'effectif  $n_j$  pour la valeur  $\xi_j$ , la moyenne ayant été préalablement calculée :

$$s^2 = \frac{n_1(\xi_1 - \bar{x})^2 + n_2(\xi_2 - \bar{x})^2 + \dots + n_k(\xi_k - \bar{x})^2}{n} = \sum_{j=1}^{j=k} f_j(\xi_j - \bar{x})^2 \text{ où } f_j = \frac{n_j}{n}$$

**Remarque :** lorsque la valeur de la variance doit intervenir dans des formules impliquant des lois de probabilité (notamment pour un intervalle de confiance ou un test d'hypothèse), il convient de remplacer l'estimation biaisée donnée par la formule « descriptive » ci-dessus par l'estimation *débiaisée* obtenue en remplaçant le dénominateur  $n$  par  $n - 1$  (cf. aussi *estimation ponctuelle*).

Lorsque les classes sont des intervalles  $[a_j, a_{j+1}]$ , la valeur « typique »  $\xi_j$  qui est utilisée dans la dernière formule est celle du centre  $\frac{a_j + a_{j+1}}{2}$  de la classe. Cela introduit un biais, le plus

souvent très faible (ce biais est sans rapport avec le biais systématique évoqué juste au-dessus), mais qui peut être compensé par la *correction de Sheppard* : si  $a$  est l'amplitude (obli-

gatoirement supposée fixe) de chaque classe, on remplace la valeur calculée  $\sum_{j=1}^{j=k} f_j(\xi_j - \bar{x})^2$

$$\text{par } \sum_{j=1}^{j=k} f_j(\xi_j - \bar{x})^2 - \frac{a^2}{12}.$$

Voir aussi *formule de Huygens-König*.

## variance d'une variable aléatoire

La variance est le principal indicateur numérique de dispersion attaché à une variable aléatoire réelle. Il est associé à l'espérance mathématique. Sa signification est celle d'une moyenne du carré de l'écart entre la variable et son espérance, pondérée par les probabilités. Comme la dimension « métrologique » de la variance est le carré de la dimension de la variable aléatoire, il faut prendre sa racine carrée (qui est l'écart-type) pour retrouver une valeur interprétable concrètement.

La variance d'une variable aléatoire  $X$  se note le plus souvent  $\text{Var}(X)$  ou  $\sigma^2(X)$  ou  $\sigma_X^2$  (ou  $\text{Var}$  ou  $\sigma^2$  s'il n'y a aucun risque de confusion), parfois  $V(X)$ .

Si la v.a. réelle  $X$  est *discrète*, caractérisée par l'ensemble (fini ou dénombrable) de valeurs  $\{x_i\}$ , avec les probabilités ponctuelles  $p_i = P(X = x_i)$ , et si son espérance mathématique est  $\mu = E(X)$ , on a :

$$\text{Var}(X) = E((X - E(X))^2) = \sum_{x_i} p_i (x_i - \mu)^2,$$

(selon les cas, il s'agira d'une somme finie ou d'une somme infinie).

Si  $X$  est *absolument continue*, caractérisée par la densité de probabilité  $f(x)$ , et si son espérance mathématique est  $\mu = E(X)$ , on a :

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

**Remarque :** si l'ensemble des valeurs est infini, il n'y a pas de certitude que la somme

$\sum_{x_i} p_i (x_i - \mu)^2$  converge, ou que l'intégrale  $\int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$  converge, avant de

l'avoir effectivement vérifié. Il existe des variables aléatoires qui ont une espérance mathématique mais qui n'ont pas de variance.

Voir aussi *formule de Huygens-König*.

## vraisemblance (méthode du maximum de)

Voir *maximum de vraisemblance (méthode du)*.



## Weibull (loi de)

(Weibull distribution)

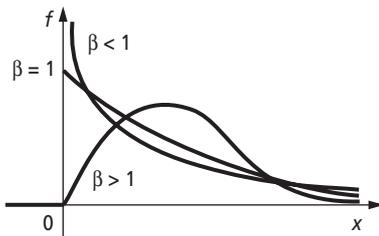
Loi d'une variable aléatoire continue qui intervient en théorie de la fiabilité pour modéliser la durée de vie d'un système complexe.

### Formulaire

Deux paramètres réels :  $\beta \in \mathbf{R}_+^*$  (paramètre « de forme »),  $\tau \in \mathbf{R}_+^*$  (paramètre d'échelle de temps).

Valeurs sur les réels positifs.

#### ► Loi de probabilité



densité

$$f(x) = \frac{\beta}{\tau} \left(\frac{x}{\tau}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\tau}\right)^\beta\right) \quad (x \geq 0)$$

fonction de répartition

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\tau}\right)^\beta\right) \quad (x \geq 0)$$

#### ► Valeurs caractéristiques

– espérance :  $E(X) = \tau \Gamma\left(1 + \frac{1}{\alpha}\right)$ ;

– variance :  $\text{Var}(X) = \tau^2 \left\{ \Gamma^2\left(1 + \frac{1}{\alpha}\right) - \Gamma\left(1 + \frac{2}{\alpha}\right) \right\}$ ;

– écart-type :  $\sigma(X) = \tau \sqrt{\Gamma^2\left(1 + \frac{1}{\alpha}\right) - \Gamma\left(1 + \frac{2}{\alpha}\right)}$ .

#### ► Cas particulier

Pour  $\beta = 1$ , on obtient la variable aléatoire exponentielle de paramètre  $\beta = \frac{1}{\tau}$  (et d'espérance  $\tau$ ).

#### ► Utilisations

Dans la pratique, la loi de Weibull est la loi de la durée de vie d'un système complexe dont le

taux de défaillance est  $\lambda(x) = \frac{f(x)}{1 - F(x)} = \frac{\beta}{\tau} \left(\frac{x}{\tau}\right)^{\beta-1}$ .

Dans le cas particulier  $\beta = 1$  de la loi exponentielle, le taux de défaillance est constant (et égal à  $\beta$ ) : le système est « sans vieillissement ».

Lorsque  $\beta > 1$ , le taux de défaillance augmente avec le temps : le système s'use ou vieillit.

## Wilcoxon (test de)

(Wilcoxon [rank sum] test)

Test d'hypothèse non paramétrique utilisé pour comparer les distributions de deux échantillons statistiques. Aussi appelé « test de la somme des rangs », il fonctionne, non pas à partir des valeurs précises observées, mais à partir des *rangs* de ces valeurs interclassées.

Si les variables aléatoires X et Y dont proviennent respectivement les deux échantillons ont même loi, elles ont en particulier même espérance mathématique, et c'est très souvent comme test de l'hypothèse *dérivée* «  $\mu_X = \mu_Y$  » que le test de Wilcoxon est utilisé. L'hypothèse (réellement testée)  $H_0 =$  « X et Y ont même loi » a pour conséquence immédiate la symétrie  $P(X \leq Y) = P(X \geq Y)$  (si les lois sont continues, on a par surcroît  $P(X = Y) = 0$ , et donc  $P(X \leq Y) = P(X \geq Y) = \frac{1}{2}$ ). La mise en œuvre du test de Wilcoxon est une simple exploitation de cette égalité des probabilités symétriques.

\_\_\_\_\_ test non paramétrique de comparaison de deux lois de probabilité, \_\_\_\_\_ également utilisé pour comparer deux espérances mathématiques  $\mu_X$  et  $\mu_Y$

• Données. Deux séries :

- un échantillon  $(x_1, x_2, \dots, x_{n_X})$  de  $n_X$  valeurs observées d'une variable aléatoire numérique X d'espérance mathématique  $\mu_X$  ;
- un échantillon  $(y_1, y_2, \dots, y_{n_Y})$  de  $n_Y$  valeurs observées d'une variable aléatoire numérique Y d'espérance mathématique  $\mu_Y$ .

• Hypothèse réellement testée.  $H_0 =$  « X et Y ont même loi » contre  $H_1$  alternative.

• Hypothèse dérivée.  $H_0 =$  «  $\mu_X = \mu_Y$  » contre  $H_1 =$  «  $\mu_X \neq \mu_Y$  ».

• Déroulement technique du test

1. On classe les  $n_X + n_Y$  valeurs observées par ordre croissant.
2. On calcule la somme  $W_X$  des rangs des valeurs de la variable X (s'il y a des ex æquo, on leur attribue le rang moyen).
3. On calcule la valeur observée de la variable de test :

$$W = \frac{\left| W_X - \frac{n_X(n_X + n_Y + 1)}{2} \right|}{\sqrt{\frac{n_X n_Y (n_X + n_Y + 1)}{12}}}$$

Les valeurs de référence de la variable de test sont à lire, soit dans des tables spécifiques pour les petites valeurs de  $n_X$  et  $n_Y$ , soit dans la table de la loi normale (centrée réduite), pour le risque bilatéral  $\alpha$ .

• Conditions et précautions

- Il n'y a aucune condition sur la loi commune à X et Y ;
- par contre, la loi normale (centrée réduite) est la loi limite pour la variable de test, ce qui induit une condition de taille si l'on ne dispose pas de table spécifique ; il est classique de demander  $n_X$  et  $n_Y \geq 10$  pour pouvoir se référer à la table de la loi normale.

Il règne un certain flottement dans l'appellation de ce test. Il existe en effet un *test de Mann-Whitney* qui teste les mêmes hypothèses dans la même situation, en comptant les inversions du classement au lieu de faire la somme des rangs. Ces deux tests sont complètement équivalents (la variable  $W_X$  du test de Wilcoxon et la variable  $U_{YX}$  du test de Mann-Whitney sont liées par la relation  $U_{YX} = W_X - \frac{1}{2}n_X(n_X + 1)$ ). Dans certains ouvrages, les appellations sont permutées. Par ailleurs, on trouve aussi l'appellation « test de Wilcoxon » sans précision pour un précurseur historique qui fonctionne sur des *séries appariées*, cf. juste ci-dessous.

### Wilcoxon (test de – pour les séries appariées) (*Wilcoxon signed rank test*)

Test d'hypothèse non paramétrique utilisé pour contrôler l'absence de « différence systématique » dans un échantillon statistique constitué « par paires ».

La situation de base est celle de couples de variables numériques  $(X_i, Y_i)$ , dont chacun donnera lieu à une observation couplée. Les lois des  $X_i$  et des  $Y_i$  ne font pas l'objet d'un présupposé d'uniformité et l'on postule seulement que les différences  $D_i = Y_i - X_i$  suivent toutes la loi d'une même variable  $Z$ , et l'hypothèse  $H_0$  est la symétrie de cette loi. Cette symétrie peut être contrôlée par le test des signes, qui ne prend en compte que les signes des différences. Comme le test de Wilcoxon prend aussi en compte l'ampleur des différences (à travers leur rang de classement), il est meilleur que le test des signes (mais il n'en a pas la simplicité et la rapidité « manuelle »).

Pour l'utilisation pratique du test de Wilcoxon pour les séries appariées, voir *test des signes*.

---

#### test bilatéral non paramétrique de symétrie de la loi d'une différence de variables

---

- Données. Deux séries appariées : un échantillon double  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$  de  $n$  valeurs couplées de deux variables aléatoires numériques dont les différences suivent toutes la loi d'une même variable  $Z$ .
- Hypothèse testée.  $H_0 =$  « la loi de  $Z$  est symétrique » contre  $H_1$  alternative.
- Déroulement technique du test
  1. On calcule les différences  $d_i = y_i - x_i$  et on les range par valeurs absolues croissantes.
  2. On calcule la somme  $W$  des rangs des différences positives (s'il y a des ex æquo, on leur attribue le rang moyen).
  3. On calcule la valeur observée de la variable de test :

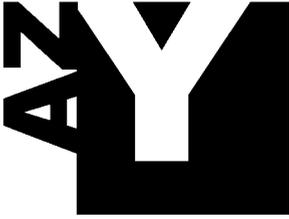
$$W = \frac{\left| W - \frac{n(n+1)}{4} \right|}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Les valeurs de référence de la variable de test sont à lire soit dans des tables spécifiques pour les petites valeurs de  $n$ , soit dans la table de la loi normale (centrée réduite), pour le risque bilatéral  $\alpha$ .

- Conditions et précautions. Aucunes.
- 

### de Witt (Jan)

Homme d'état hollandais (1625–1672). Il effectua le premier calcul de rentes viagères.



### Yates (correction [de continuité] de)

(*Yates correction*)

Dans un test du khi-deux d'indépendance ou d'homogénéité à quatre classes ( $2 \times 2$ ), correction qui permet de faire fonctionner le test même lorsque la condition sur l'effectif  $np_{ij} \geq 5$  n'est pas respectée : si tous les  $np_{ij}$  sont  $\geq 3$ , on remplace :

$$\sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - np_{ij})^2}{np_{ij}} \quad \text{par} \quad \sum_{i=1}^2 \sum_{j=1}^2 \frac{\left(|n_{ij} - np_{ij}| - \frac{1}{2}\right)^2}{np_{ij}}.$$

Cette correction, qui est le simple décalque de certaines formules où l'on approxime une loi discrète (notamment binomiale) par une loi continue, n'a pas de justification théorique et ne fait pas l'unanimité. Elle a néanmoins pour avantage de diminuer la valeur du khi-deux et de prémunir contre un rejet trop hâtif de l'hypothèse  $H_0$ .

### Yule (conditions de)

(*Yule conditions*)

Critères de qualité globale pour un indicateur statistique (de tendance centrale, de dispersion, ...). Dans un ordre qui est à peu près celui de l'importance, ce sont :

- être défini de façon objective ;
- dépendre de toutes les observations (pour avoir une signification « exhaustive » et aussi pour assurer une convergence de fait lorsque la taille de l'échantillon augmente) ;
- être peu sensible aux valeurs extrêmes (sur ce critère comme sur le précédent, le centre et l'étendue sont médiocres) ;
- avoir une signification concrète ;
- se prêter au calcul algébrique (ou plutôt figurer naturellement dans les théorèmes et formules du calcul des probabilités : supériorité par exemple de la moyenne sur la médiane ou le mode) ;
- être simple à calculer ;
- être peu sensible aux fluctuations d'échantillonnage.



## **z (test)**

Nom parfois donné au *test de Student* lorsqu'il est appliqué à un échantillon de « grande taille » (classiquement  $n \geq 30$ ) et que la table de référence du test est la table de la loi normale et non la table de la loi de Student.

## **zéro-un (loi du – de Kolmogorov) (zero-one Kolmogorov law)**

Théorème qui énonce que, sous certaines conditions, certains évènements (qui sont souvent des évènements très intéressants dans la pratique) relatifs à une suite infinie de variables aléatoires indépendantes ne peuvent pas avoir une probabilité autre que 0 ou 1.

Pour donner l'énoncé de ce théorème, il faut préalablement définir la notion d'« évènement de queue » : étant donnée une suite infinie  $(X_n)$  de variables aléatoires, on appelle évènement de queue tout évènement global qui n'est pas modifié si l'on modifie un nombre fini d'évènements de la suite (par exemple « pour  $n$  assez grand (traduction formelle :  $\exists n_0 \ n \geq n_0 \dots$ ), on a telle ou telle propriété de  $X_n$  »).

L'énoncé du théorème (« loi ») de Kolmogorov est alors : étant donnée une suite infinie de variables aléatoires indépendantes, un évènement de queue ne peut pas avoir une probabilité autre que 0 ou 1.

Voir *Borel-Cantelli (lemme de)*.

## **Zipf (loi de) (Zipf distribution)**

Loi empirique qui régit notamment la fréquence des mots dans une langue. Cette loi a été découverte en 1936 par le sociologue américain G. K. Zipf. Elle énonce que, si l'on classe les mots par ordre de fréquences décroissantes, la fréquence du  $k$ -ième mot est approximativement proportionnelle à  $\frac{1}{k}$ .

Cette loi semble particulièrement adaptée aux évènements qui sont très nombreux et de probabilité ou de fréquence très faiblement décroissantes au-delà des tout premiers. Par exemple, elle rend bien compte des populations des villes d'un pays (ou du Globe) – étant entendu que l'on peut convertir une population en probabilité d'appartenance d'un citoyen tiré au hasard.

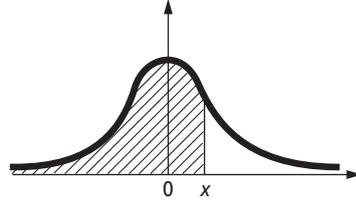


Table 1

**Loi normale centrée réduite  $\mathcal{N}(0, 1)$**   
**Table de la fonction de répartition**

Probabilité d'avoir une valeur inférieure à  $x$  :

$$\Pi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$



x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,00	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,60	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,70	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,80	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,90	0,9981	0,9982	0,9982	0,9984	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Pour  $x < 0$  prendre le complément à 1 de la valeur lue dans la table pour  $-x$  :

$$\Pi(x) = 1 - \Pi(-x)$$

Table pour les grandes valeurs de  $x$ 

$x$	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7
$\Pi(x)$	0,998 650	0,999 032	0,999 313	0,999 517	0,999 663	0,999 767	0,999 841	0,999 892
$1 - \Pi(x)$	0,001 350	0,000 968	0,000 687	0,000 483	0,000 337	0,000 233	0,000 159	0,000 108

$x$	3,8	3,9	4,0	4,1	4,2	4,3	4,4	4,5
$\Pi(x)$	0,998 928	0,999 952	0,999 968	0,999 979	0,999 987	0,999 991	0,999 995	0,999 997
$1 - \Pi(x)$	0,000 072	0,000 048	0,000 032	0,000 021	0,000 013	0,000 009	0,000 005	0,000 003

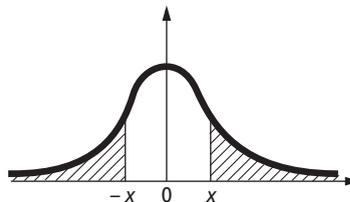
Table 2

Loi normale centrée réduite  $\mathcal{N}(0, 1)$

Table de dépassement de l'écart absolu

En fonction d'une probabilité  $\alpha$  ; valeur de l'écart  $x$  qui possède la probabilité  $\alpha$  d'être dépassé en valeur absolue :

$$P(|X| > x) = \alpha$$



$\alpha$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	$\infty$	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

Table pour les petites valeurs de  $\alpha$

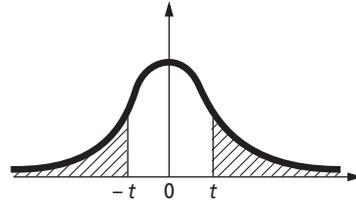
$\alpha$	$x$	$\alpha$	$x$
0,000 000 001	6,109	0,000 1	3,891
0,000 000 01	5,730	0,000 2	3,719
0,000 000 02	5,612	0,000 5	3,481
0,000 000 05	5,451	0,001	3,291
0,000 000 1	5,327	0,002	3,090
0,000 000 2	5,199	0,003	2,968
0,000 000 5	5,026	0,004	2,878
0,000 001	4,892	0,005	2,807
0,000 002	4,753	0,006	2,748
0,000 005	4,565	0,007	2,696
0,000 01	4,417	0,008	2,652
0,000 02	4,265	0,009	2,612
0,000 05	4,056	0,010	2,576

Table 3

Loi de Student

Table de dépassement de l'écart absolu

En fonction du nombre ddl de degrés de liberté et d'une probabilité  $\alpha$  : valeur de l'écart  $t$  qui possède la probabilité  $\alpha$  d'être dépassé en valeur absolue.



$\alpha$ ddl	0,50	0,20	0,10	0,05	0,02	0,01	0,005	0,002	0,001	0,0001
1	1,000	3,078	6,314	12,706	31,821	63,657	127,32	318,31	636,62	6366,2
2	0,816	1,886	2,920	4,303	6,965	9,925	14,089	22,327	34,599	99,992
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,215	12,924	28,000
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610	15,544
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,893	6,869	11,178
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959	9,082
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408	7,885
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041	7,120
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781	6,594
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587	6,211
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437	5,921
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318	5,694
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221	5,513
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140	5,363
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073	5,239
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015	5,134
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965	5,044
18	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922	4,966
19	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883	4,897
20	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850	4,837
21	0,686	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819	4,784
22	0,686	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792	4,736
23	0,685	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768	4,693
24	0,685	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745	4,654
25	0,684	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725	4,619
30	0,683	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646	4,482
35	0,682	1,306	1,690	2,030	2,438	2,724	2,996	3,340	3,591	4,389
40	0,681	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551	4,321
45	0,680	1,301	1,679	2,014	2,412	2,690	2,952	3,281	3,520	4,269
50	0,679	1,299	1,676	2,009	2,403	2,678	2,937	3,261	3,496	4,228
60	0,679	1,296	1,671	2,000	2,390	2,660	2,915	3,232	3,460	4,169
70	0,678	1,294	1,667	1,994	2,381	2,648	2,899	3,211	3,435	4,127
80	0,678	1,292	1,664	1,990	2,374	2,639	2,887	3,195	3,416	4,096
90	0,677	1,291	1,662	1,987	2,368	2,632	2,878	3,183	3,402	4,072
100	0,677	1,290	1,660	1,984	2,364	2,626	2,871	3,174	3,390	4,053
150	0,676	1,287	1,655	1,976	2,351	2,609	2,849	3,145	3,357	3,998
200	0,676	1,286	1,653	1,972	2,345	2,601	2,839	3,131	3,340	3,970
300	0,675	1,284	1,650	1,968	2,339	2,592	2,828	3,118	3,323	3,944
500	0,675	1,283	1,648	1,965	2,334	2,586	2,820	3,107	3,310	3,922
1 000	0,675	1,282	1,646	1,962	2,330	2,581	2,813	3,098	3,300	3,906
$\infty$	0,674	1,282	1,645	1,960	2,326	2,576	2,807	3,090	3,291	3,891

Table 4

Valeurs critiques du coefficient de corrélation linéaire  $\rho$ 

## Table de la valeur absolue qui possède une probabilité donnée d'être dépassée (échantillon normal)

En fonction du nombre ddl de degrés de liberté (égal à  $n - 2$  pour une corrélation simple) et d'une probabilité  $\alpha$  : valeur de  $r$  qui possède la probabilité  $\alpha$  d'être dépassée en valeur absolue, soit  $P(|\rho| > r) = \alpha$ .

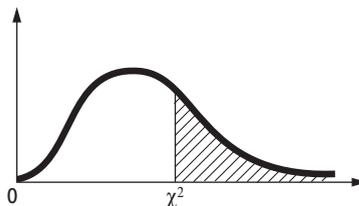
ddl \ $\alpha$	0,10	0,05	0,01
1	0,9877	0,9969	0,9999
2	0,9000	0,9500	0,9900
3	0,8054	0,8783	0,9587
4	0,7293	0,8114	0,9172
5	0,6694	0,7545	0,8745
6	0,6215	0,7067	0,8343
7	0,5822	0,6664	0,7977
8	0,5494	0,6319	0,7646
9	0,5214	0,6021	0,7348
10	0,4973	0,5760	0,7079
11	0,4762	0,5529	0,6835
12	0,4575	0,5324	0,6614
13	0,4409	0,5139	0,6411
14	0,4259	0,4973	0,6226
15	0,4124	0,4821	0,6055
16	0,4000	0,4683	0,5897
17	0,3887	0,4555	0,5751
18	0,3783	0,4438	0,5614
19	0,3687	0,4329	0,5487
20	0,3598	0,4227	0,5368
21	0,3515	0,4132	0,5256
22	0,3438	0,4044	0,5151
23	0,3365	0,3961	0,5052
24	0,3297	0,3882	0,4958
25	0,3233	0,3809	0,4869
30	0,2960	0,3494	0,4487
35	0,2746	0,3246	0,4182
40	0,2573	0,3044	0,3932
45	0,2428	0,2875	0,3721
50	0,2306	0,2732	0,3541
60	0,2108	0,2500	0,3248
70	0,1954	0,2319	0,3017
80	0,1829	0,2172	0,2830
90	0,1726	0,2050	0,2673
100	0,1638	0,1946	0,2540
ddl > 100	$\frac{1,645}{\sqrt{\text{ddl} + 1}}$	$\frac{1,960}{\sqrt{\text{ddl} + 1}}$	$\frac{2,576}{\sqrt{\text{ddl} + 1}}$

Table 5

### Loi du khi-deux

#### Table de dépassement de l'écart

En fonction du nombre ddl de degrés de liberté et d'une probabilité  $\alpha$  : valeur de l'écart  $\chi^2$  qui possède la probabilité  $\alpha$  d'être dépassée.



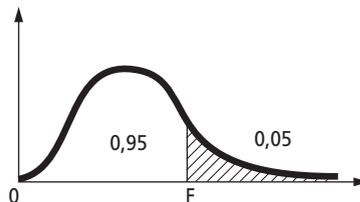
ddl \ $\alpha$	0,999	0,99	0,95	0,90	0,50	0,10	0,05	0,01	0,001
1	0,000002	0,00016	0,00393	0,0158	0,455	2,706	3,841	6,635	10,828
2	0,00200	0,0201	0,103	0,211	1,386	4,605	5,991	9,210	13,816
3	0,0243	0,115	0,352	0,584	2,366	6,251	7,815	11,345	16,266
4	0,0908	0,297	0,711	1,064	3,357	7,779	9,488	13,277	18,467
5	0,210	0,554	1,145	1,610	4,351	9,236	11,070	15,086	20,515
6	0,381	0,872	1,635	2,204	5,348	10,645	12,592	16,812	22,458
7	0,598	1,239	2,167	2,833	6,346	12,017	14,067	18,475	24,322
8	0,857	1,646	2,733	3,490	7,344	13,362	15,507	20,090	26,124
9	1,152	2,088	3,325	4,168	8,343	14,684	16,919	21,666	27,877
10	1,479	2,558	3,940	4,865	9,342	15,987	18,307	23,209	29,588
11	1,834	3,053	4,575	5,578	10,341	17,275	19,675	24,725	31,264
12	2,214	3,571	5,226	6,304	11,340	18,549	21,026	26,217	32,909
13	2,617	4,107	5,892	7,042	12,340	19,812	22,362	27,688	34,528
14	3,041	4,660	6,571	7,790	13,339	21,064	23,685	29,141	36,123
15	3,483	5,229	7,261	8,547	14,339	22,307	24,996	30,578	37,697
16	3,942	5,812	7,962	9,312	15,338	23,542	26,296	32,000	39,252
17	4,416	6,408	8,672	10,085	16,338	24,769	27,587	33,409	40,790
18	4,905	7,015	9,390	10,865	17,338	25,989	28,869	34,805	42,312
19	5,407	7,633	10,117	11,651	18,338	27,204	30,144	36,191	43,820
20	5,921	8,260	10,851	12,443	19,337	28,412	31,410	37,566	45,315
21	6,447	8,897	11,591	13,240	20,337	29,615	32,671	38,932	46,797
22	6,983	9,542	12,338	14,041	21,337	30,813	33,924	40,289	48,268
23	7,529	10,196	13,091	14,848	22,337	32,007	35,172	41,638	49,728
24	8,085	10,856	13,848	15,659	23,337	33,196	36,415	42,980	51,179
25	8,649	11,524	14,611	16,473	24,337	34,382	37,652	44,314	52,620
30	11,59	14,95	18,49	20,60	29,34	40,26	43,77	50,89	59,70
35	14,69	18,51	22,47	24,80	34,34	46,06	49,80	57,34	66,62
40	17,92	22,16	26,51	29,05	39,34	51,81	55,76	63,69	73,40
45	21,25	25,90	30,61	33,35	44,34	57,51	61,66	69,96	80,08
50	24,67	29,71	34,76	37,69	49,33	63,17	67,50	76,15	86,66
60	31,74	37,48	43,19	46,46	59,33	74,40	79,08	88,38	99,61
70	39,04	45,44	51,74	55,33	69,33	85,53	90,53	100,43	112,32
80	46,52	53,54	60,39	64,28	79,33	96,58	101,88	112,33	124,84
90	54,16	61,75	69,13	73,29	89,33	107,57	113,15	124,12	137,21
100	61,92	70,06	77,93	82,36	99,33	118,50	124,34	135,81	149,45

Nota : pour effectuer un test du khi-deux, seule la partie droite de la table est utile ; pour calculer un intervalle de confiance pour une variance (échantillon normal) ou pour effectuer un test de quotient de variances (échantillons normaux), les valeurs pour les probabilités complémentaires  $\alpha$  et  $1-\alpha$  sont simultanément utilisées.

Table 6a  
Loi du F de Fisher-Snedecor

Table de l'écart ayant une probabilité 0,05 de dépassement

En fonction des nombres de degrés de liberté  $v_1$  et  $v_2$  : valeur de l'écart de la variable  $F(v_1, v_2)$  qui possède la probabilité 0,05 d'être dépassée.



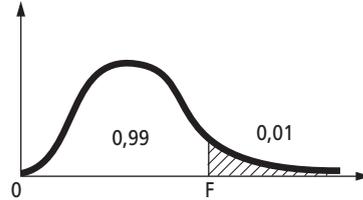
$v_1 \backslash v_2$	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40	50	100	$\infty$
1	161	200	216	225	230	234	237	239	241	242	244	246	248	250	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,62	8,59	8,58	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,75	5,72	5,70	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,50	4,46	4,44	4,41	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,81	3,77	3,75	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,38	3,34	3,32	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,08	3,04	3,02	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,27	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,86	2,83	2,80	2,76	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,70	2,66	2,64	2,59	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,57	2,53	2,51	2,46	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,47	2,43	2,40	2,35	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,38	2,34	2,31	2,26	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,31	2,27	2,24	2,19	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,25	2,20	2,18	2,12	2,07
16	4,49	3,63	3,23	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,19	2,15	2,12	2,07	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,15	2,10	2,08	2,02	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,11	2,06	2,04	1,98	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,07	2,03	2,00	1,94	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,04	1,99	1,97	1,91	1,84
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,92	1,87	1,84	1,78	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,84	1,79	1,76	1,70	1,62
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,04	1,96	1,88	1,79	1,74	1,70	1,63	1,56
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,74	1,69	1,66	1,59	1,51
45	4,05	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	1,97	1,89	1,81	1,71	1,66	1,63	1,55	1,47
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,95	1,87	1,78	1,69	1,63	1,60	1,52	1,44
55	4,02	3,16	2,77	2,54	2,38	2,27	2,18	2,11	2,06	2,01	1,93	1,85	1,76	1,67	1,61	1,58	1,50	1,41
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,65	1,59	1,56	1,48	1,39
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,03	1,98	1,90	1,82	1,73	1,63	1,58	1,54	1,46	1,37
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,89	1,81	1,72	1,62	1,57	1,53	1,45	1,35
75	3,97	3,12	2,73	2,49	2,34	2,22	2,13	2,06	2,01	1,96	1,88	1,80	1,71	1,61	1,55	1,52	1,44	1,34
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,88	1,79	1,70	1,60	1,54	1,51	1,43	1,32
85	3,95	3,10	2,71	2,48	2,32	2,21	2,12	2,05	1,99	1,94	1,87	1,79	1,70	1,59	1,54	1,50	1,42	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,86	1,78	1,69	1,59	1,53	1,49	1,41	1,30
95	3,94	3,09	2,70	2,47	2,31	2,20	2,11	2,04	1,98	1,93	1,86	1,77	1,68	1,58	1,53	1,48	1,40	1,29
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85	1,77	1,68	1,57	1,52	1,48	1,39	1,28
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,82	1,73	1,64	1,54	1,48	1,44	1,34	1,22
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,80	1,72	1,62	1,52	1,46	1,41	1,32	1,19
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,78	1,70	1,61	1,50	1,43	1,39	1,30	1,15
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,77	1,69	1,59	1,48	1,42	1,38	1,28	1,11
1 000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,76	1,68	1,58	1,47	1,41	1,36	1,26	1,08
$\infty$	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,46	1,39	1,35	1,24	1,00

Table 6b

Loi du F de Fisher-Snedecor

Table de l'écart ayant une probabilité 0,01 de dépassement

En fonction des nombres de degrés de liberté  $v_1$  et  $v_2$  : valeur de l'écart de la variable  $F(v_1, v_2)$  qui possède la probabilité 0,01 d'être dépassée.



$v_1 \backslash v_2$	1	2	3	4	5	6	7	8	9	10	12	15	20	30	40	50	100	$\infty$
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6261	6287	6303	6334	6366
2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	26,9	26,7	26,5	26,4	26,4	26,2	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,2	14,0	13,8	13,7	13,7	13,6	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,89	9,72	9,55	9,38	9,29	9,24	9,13	9,02
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,23	7,14	7,09	6,99	6,88
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	5,99	5,91	5,86	5,75	5,65
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,20	5,12	5,07	4,96	4,86
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,65	4,57	4,52	4,41	4,31
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,25	4,17	4,12	4,01	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	3,94	3,86	3,81	3,71	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,70	3,62	3,57	3,47	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,51	3,43	3,38	3,27	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,35	3,27	3,22	3,11	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,21	3,13	3,08	2,98	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,10	3,02	2,97	2,86	2,75
17	8,40	6,11	5,19	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,00	2,92	2,87	2,76	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	2,92	2,84	2,78	2,68	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,84	2,76	2,71	2,60	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,78	2,69	2,64	2,54	2,42
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,54	2,45	2,40	2,29	2,17
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,39	2,30	2,25	2,13	2,01
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88	2,74	2,60	2,44	2,28	2,19	2,14	2,02	1,89
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,20	2,11	2,06	1,94	1,80
45	7,23	5,11	4,25	3,77	3,45	3,23	3,07	2,94	2,83	2,74	2,61	2,46	2,31	2,14	2,05	2,00	1,88	1,74
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,56	2,42	2,27	2,10	2,01	1,95	1,82	1,68
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,53	2,38	2,23	2,06	1,97	1,91	1,78	1,64
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,03	1,94	1,88	1,75	1,60
65	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,80	2,69	2,61	2,47	2,33	2,17	2,00	1,91	1,85	1,72	1,57
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,45	2,31	2,15	1,98	1,89	1,83	1,70	1,54
75	6,99	4,90	4,05	3,58	3,27	3,05	2,89	2,76	2,65	2,57	2,43	2,29	2,13	1,96	1,87	1,81	1,67	1,52
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,42	2,27	2,12	1,94	1,85	1,79	1,65	1,49
85	6,94	4,86	4,02	3,55	3,24	3,02	2,86	2,73	2,62	2,54	2,40	2,26	2,10	1,93	1,83	1,77	1,64	1,47
90	6,93	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,39	2,24	2,09	1,92	1,82	1,76	1,62	1,46
95	6,91	4,84	3,99	3,52	3,22	3,00	2,83	2,70	2,60	2,51	2,38	2,23	2,08	1,90	1,81	1,75	1,61	1,44
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,37	2,22	2,07	1,89	1,80	1,74	1,60	1,43
150	6,81	4,75	3,91	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,31	2,16	2,00	1,83	1,73	1,66	1,52	1,33
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,27	2,13	1,97	1,79	1,69	1,63	1,48	1,28
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38	2,24	2,10	1,94	1,76	1,66	1,59	1,44	1,22
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,22	2,07	1,92	1,74	1,63	1,57	1,41	1,16
1 000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,20	2,06	1,90	1,72	1,61	1,54	1,38	1,11
$\infty$	6,63	4,61	3,78	3,31	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,70	1,59	1,52	1,36	1,00

Table 7

## Statistique des rangs de Wilcoxon

## Table des valeurs critiques

On désigne par  $n_1$  la taille du plus petit des deux échantillons et par  $n_2$  la taille du plus grand. La variable aléatoire  $W (= W(n_1, n_2))$  est la somme des rangs du plus petit échantillon.

## 1. Valeurs critiques unilatérales inférieures

Les valeurs  $w (= w(n_1, n_2, \alpha))$  données dans la table sont définies par :

$$P(W \leq w) \leq \alpha \quad \text{et} \quad P(W \leq w + 1) > \alpha$$

Table pour  $\alpha = 0,025$ 

$n_1 \backslash n_2$	1	2	3	4	5	6	7	8	9	10
1	-									
2	-	-								
3	-	-	-							
4	-	-	-	10						
5	-	-	6	11	17					
6	-	-	7	12	18	26				
7	-	-	7	13	20	27	36			
8	-	3	8	14	21	29	38	49		
9	-	3	8	14	22	31	40	51	62	
10	-	3	9	15	23	32	42	53	65	78
11	-	3	9	16	24	34	44	55	68	81
12	-	4	10	17	26	35	46	58	71	84
13	-	4	10	18	27	37	48	60	73	88
14	-	4	11	19	28	38	50	62	76	91
15	-	4	11	20	29	40	52	65	79	94
16	-	4	12	21	30	42	54	67	82	97
17	-	5	12	21	32	43	56	70	84	100
18	-	5	13	22	33	45	58	72	87	103
19	-	5	13	23	34	46	60	74	90	107
20	-	5	14	24	35	48	62	77	93	110
21	-	6	14	25	37	50	64	79	95	113
22	-	6	15	26	38	51	66	81	98	116
23	-	6	15	27	39	53	68	84	101	119
24	-	6	16	27	40	54	70	86	104	122
25	-	6	16	28	42	56	72	89	107	126

Table pour  $\alpha = 0,05$ 

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10
1	–									
2	–	–								
3	–	–	6							
4	–	–	6	11						
5	–	3	7	12	19					
6	–	3	8	13	20	28				
7	–	3	8	14	21	30	39			
8	–	4	9	15	23	31	41	51		
9	–	4	10	16	24	33	43	54	66	
10	–	4	10	17	26	35	45	56	69	82
11	–	4	11	18	27	37	47	59	72	86
12	–	5	11	19	28	38	49	62	75	89
13	–	5	12	20	30	40	52	64	78	92
14	–	6	12	21	31	42	54	67	81	96
15	–	6	13	22	33	44	56	69	84	99
16	–	6	14	24	34	46	58	72	87	103
17	–	6	15	25	35	47	61	75	90	106
18	–	7	15	26	37	49	63	77	93	110
19	1	7	16	27	38	51	65	80	96	113
20	1	7	17	28	40	53	67	83	99	117
21	1	8	17	29	41	55	69	85	102	120
22	1	8	18	30	43	57	72	88	105	123
23	1	8	19	31	44	58	74	90	108	127
24	1	9	19	32	45	60	76	93	111	130
25	1	9	20	33	47	62	78	96	114	134

## 2. Valeurs critiques unilatérales supérieures

Elles sont égales à

$$n_1(n_1 + n_2 + 1) - w,$$

où  $w$  est la valeur critique inférieure (donnée par la table précédente).

## 3. Valeurs critiques bilatérales inférieures

Pour le risque  $\alpha$ , ce sont les couples constitués par la valeur critique inférieure et la valeur critique supérieure relatives au risque  $\frac{\alpha}{2}$ . Les deux tables précédentes sont donc utilisables pour le risque bilatéral  $\alpha = 0,05$  et  $\alpha = 0,10$ .

## Statistique d'inversions de Mann-Whitney

La variable aléatoire  $U (= U(n_1, n_2))$  est le nombre d'inversions : valeurs du plus grand échantillon devant des valeurs du plus petit échantillon.

La relation

$$w = u + \frac{1}{2} n_1(n_1 + 1)$$

permet de se rapporter aux tables de la statistique de Wilcoxon.

Table 8

Valeurs critiques pour la variable aléatoire  $\Delta_n$  de Kolmogorov

## Table de la valeur absolue qui possède une probabilité donnée d'être dépassée

En fonction de la taille  $n$  de l'échantillon et d'une probabilité  $\alpha$  : valeurs de  $\delta$  qui possèdent la probabilité  $\alpha$  d'être dépassées en valeur absolue.

$$\text{Si } \Delta_n = \sup_x |F_n(x) - F_0(x)|, P(\Delta_n > \delta) = \alpha.$$

$n \backslash \alpha$	0,10	0,05	0,01
1	0,9500	0,9750	0,9950
2	0,7764	0,8419	0,9293
3	0,6360	0,7076	0,8290
4	0,5652	0,6239	0,7342
5	0,5095	0,5633	0,6685
6	0,4680	0,5193	0,6166
7	0,4361	0,4834	0,5758
8	0,4096	0,4543	0,5418
9	0,3875	0,4300	0,5133
10	0,3697	0,4092	0,4889
11	0,3524	0,3912	0,4677
12	0,3381	0,3754	0,4491
13	0,3255	0,3614	0,4325
14	0,3142	0,3489	0,4176
15	0,3040	0,3376	0,4042
16	0,2947	0,3273	0,3920
17	0,2863	0,3180	0,3809
18	0,2785	0,3094	0,3706
19	0,2714	0,3014	0,3612
20	0,2647	0,2941	0,3524
21	0,2586	0,2872	0,3443
22	0,2528	0,2809	0,3367
23	0,2475	0,2749	0,3295
24	0,2424	0,2693	0,3229
25	0,2377	0,2640	0,3166
30	0,2176	0,2417	0,2899
35	0,2019	0,2242	0,2690
40	0,1891	0,2101	0,2521
45	0,1786	0,1984	0,2380
50	0,1696	0,1884	0,2260
60	0,1551	0,1723	0,2067
70	0,1438	0,1598	0,1917
80	0,1347	0,1496	0,1795
90	0,1271	0,1412	0,1694
100	0,1207	0,1340	0,1608
$n > 100$	$\frac{1,223}{\sqrt{n}}$	$\frac{1,358}{\sqrt{n}}$	$\frac{1,629}{\sqrt{n}}$

# Bibliographie

## Ouvrages essentiellement de probabilités

FOATA Dominique, FUCHS Aimé – *Calcul des probabilités*, Dunod, 1998–2003.

Cours, exercices et problèmes corrigés. Orienté Licence de mathématiques 3<sup>e</sup> année et écoles d'ingénieurs. Très progressif et très pédagogique.

OUVRARD Jean-Yves – *Probabilités 1 : CAPES – Agrégation*, Cassini, 1998.

Beaucoup plus élémentaire que le titre ne le suggère (mais le tome 2 (*Martingales, chaînes de Markov*) est d'un niveau plus élevé).

BOULEAU Nicolas – *Probabilités de l'ingénieur, variables aléatoires et simulation*, Hermann, 1986–2002.

Orienté modélisation et simulation.

MAZLIAK Laurent – *Cours de probabilités de l'ingénieur, exercices et problèmes corrigés*, Hermann, 1998.

Exercices gradués, jusqu'aux niveaux convergences et fonction caractéristique.

LECOUTRE Jean-Pierre – *Statistique et probabilités*, Dunod, 2<sup>e</sup> édition, 2002.

Orienté économie-gestion. Intéressant surtout pour sa partie probabilités (70 % de l'ensemble). Plutôt compact, assez poussé en théorie, nombreux exercices.

## Ouvrages mixtes

SAPORTA Alain-Jacques – *Probabilités, analyse des données et statistique*, Technip, 1999, 4<sup>e</sup> réimpression 2002.

Une véritable bible très claire, très détaillée. Le niveau monte souvent mais le néophyte s'y retrouvera. Avec en prime une initiation très intéressante à l'« analyse des données ».

WONNACOTT Thomas H. et WONNACOTT Ronald J. – *Statistique*, Economica, 4<sup>e</sup> édition, 1991.

Orienté économie-gestion-sciences-médecine. Avec de très nombreux exemples concrets et de très nombreux exercices. Un gros pavé, riche et complet, aussi agréable à lire qu'à utiliser. Malgré son absence du titre, la partie probabilités est raisonnablement développée.

DRESS François – *TD Probabilités Statistique pour les sciences de la vie*, Dunod, 2<sup>e</sup> édition, 2002.

Orienté sciences de la vie et de la terre. Cours et commentaires, exercices corrigés.

FÉMÉNIAS Jean-Louis – *Probabilités et statistique pour les sciences physiques*, Dunod, 2003.

Avec de très nombreux exemples théoriques et concrets pour la physique.

COUTY Françoise, DEBORD Jean, FREDON Daniel – *Probabilités et statistiques pour DEUG SV ET ST, PCEM*, Pharmacie, Dunod, 1999.

Résumés de cours, exercices et problèmes résolus, de la statistique descriptive à la régression linéaire.

VALLERON Alain-Jacques – *Probabilités et statistique*, Masson, 2001.

Manuel d'enseignement orienté médecine-pharmacie. Aéré, pédagogique, nombreux exemples et exercices.

### Ouvrages essentiellement de statistique

CHARUELLE Pascal, PINAULT Yves – *Statistique descriptive*, Montchrestien, 2000.

Orienté économie et administration. Compact, très clair, très complet sur le sujet, nombreux exemples détaillés.

PY Bernard – *Statistique descriptive*, Economica, 4<sup>e</sup> édition, 1996.

Orienté économie. Avec QCM et exercices corrigés.

SCHWARTZ Daniel – *Méthodes statistiques à l'usage des médecins et biologistes*, Flammarion Médecine-Sciences, 4<sup>e</sup> édition, 1993, actualisée 1996.

Un livre ancien qui n'a pas vieilli. Élémentaire et complet, très bien expliqué, nombreux exemples détaillés et commentés.

MEOT Alain – *Introduction aux statistiques inférentielles*, De Boeck, 2003.

Orienté sciences humaines avec une réflexion de haute qualité.

FRONTIER Serge, DAVOULT Dominique, GENTILHOMME Valérie, LAGADEUC Yvan – *Statistiques pour les sciences de la vie et de l'environnement*, Dunod, 2001.

Présente un panorama très étendu des techniques, avec de nombreux commentaires et exemples très concrets. Une ouverture intéressante vers l'analyse des données.

VALLERON Alain-Jacques – *Introduction à la biostatistique*, Masson, 1998.

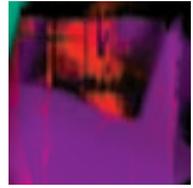
Beaucoup plus et beaucoup mieux qu'une introduction ! Les notions et techniques de base sont présentées et très intelligemment commentées, leur application est détaillée sur des exemples, le panorama est étendu même si certains choix sont personnels.

DAGNELIE Pierre – *Statistique théorique et appliquée, tome 1 : statistique descriptive et base de l'inférence statistique*, De Boeck et Larcier, 1998.

Un pavé très complet avec de bonnes explications. La structuration interne en 3 niveaux est un peu complexe mais tout à fait pertinente. Bibliographie très (trop ?) abondante.

MORTON Richard F., HEBEL J. Richard, MCCARTER Robert J. – *Épidémiologie et biostatistique*, Aspen Publishers/Doin, 4<sup>e</sup> édition, 1996, réimpression 2002.

Très élémentaire et très pédagogique, pour tout comprendre (plutôt que pour pratiquer).



François Dress

# LES PROBABILITÉS ET LA STATISTIQUE DE A à Z

## 500 définitions, formules et tests d'hypothèse

Ce dictionnaire présente, en près de 500 entrées, toutes les notions de probabilités et de statistique abordées dès les premières années d'université (sciences expérimentales, sciences économiques et sociales, mathématiques).

L'ouvrage propose tout à la fois les définitions des concepts de base, la présentation détaillée de plus de 40 lois de probabilité, et un formulaire très riche. Les définitions ont été rédigées en restant à chaque fois au niveau mathématique le plus élémentaire possible, et en les faisant précéder très souvent par une courte introduction en langage courant. On trouvera enfin le fonctionnement commenté de plus de 25 tests d'hypothèses.

Ce dictionnaire est destiné aux étudiants en Licence, aux élèves des classes préparatoires, aux candidats aux concours de l'enseignement, et il sera aussi utile aux professionnels non-mathématiciens.

FRANÇOIS DRESS  
est professeur à  
l'Université Bordeaux 1.



6647861

ISBN 978-2-10-051403-8



[www.dunod.com](http://www.dunod.com)

