

Cours Commun Scientifique  
de  
Probabilités & Statistiques

---

Résumé de cours

---

Responsable du cours : Laurent Tournier



Ce document, ainsi que les fiches de TD et d'autres documents liés au cours, peut être trouvé au format PDF à l'adresse suivante :

<http://www.math.univ-paris13.fr/~tournier/enseignement/>



# 1 Espaces de probabilités

## 1.1 Définitions

### Définition

Un **espace de probabilité**  $(\Omega, P)$  est constitué de

- $\Omega$ , un ensemble
- $P$ , une probabilité sur  $\Omega$ .

$\Omega$  représente l'**ensemble des résultats possibles d'une expérience aléatoire**.

Un élément  $\omega \in \Omega$  est appelé une **réalisation**, c'est un résultat possible de cette expérience.

Un sous-ensemble  $A \subset \Omega$  est appelé un **événement**. C'est un ensemble de réalisations (par exemple, celles qui vérifient une certaine condition). L'ensemble des événements est donc l'ensemble  $\mathcal{P}(\Omega)$  des sous-ensembles (ou « parties ») de  $\Omega$ .

La **probabilité**  $P$  donne, pour chaque événement  $A$ , la « proportion de chance »  $P(A)$  pour qu'il se produise. Elle doit vérifier deux propriétés :

### Définition

Une **probabilité** sur  $\Omega$  est une application  $P : \mathcal{P}(\Omega) \rightarrow [0,1]$  telle que

1.  $P(\Omega) = 1$
2. pour toute suite finie ou infinie  $(A_n)_n$  d'événements disjoints deux à deux,

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n).$$

(l'union et la somme portent sur toute la suite)

Si un événement  $A$  vérifie  $P(A) = 0$ , on dit que  $A$  est **négligeable** ; et si  $P(A) = 1$ , on dit que  $A$  est **presque sûr**, ou que  $A$  a lieu **presque sûrement** (abrégé « p.s. »).

### Propriétés

- a)  $P(\emptyset) = 0$
- b) Pour tout événement  $A$ ,  $P(A^c) = 1 - P(A)$
- c) Si  $A \subset B$ , alors  $P(A) \leq P(B)$
- d) Pour tous événements  $A$  et  $B$ ,  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- e) Pour toute suite finie ou infinie  $(A_n)_n$  d'événements,  $P\left(\bigcup_n A_n\right) \leq \sum_n P(A_n)$ .

NB. Les opérations usuelles sur des événements  $A$  et  $B$  peuvent s'interpréter intuitivement :

Notation	Sens mathématique	Interprétation en probabilités
$A^c (= \Omega \setminus A)$	complémentaire de $A$	contraire de $A$ , « non $A$ »
$A \cup B$	réunion de $A$ et $B$	« $A$ ou $B$ »
$A \cap B$	intersection de $A$ et $B$	« $A$ et $B$ »
$A \cap B = \emptyset$	$A$ et $B$ sont disjoints	« $A$ et $B$ sont incompatibles »
$A \subset B$	$A$ est inclus dans $B$	« $A$ implique $B$ ».

Pour simplifier, on considère dans ce cours que tout ensemble de réalisations est un événement. En réalité, dans beaucoup d'exemples (comme  $\Omega = \mathbb{R}$  avec une probabilité à densité), on ne pourrait pas définir  $P(A)$  pour tous les ensembles, mais seulement pour des ensembles « mesurables » ; cependant, ceci ne pose pas de problème pratique car tous les ensembles rencontrés seront mesurables.

## 1.2 Cas symétrique : équiprobabilité

On suppose que l'expérience n'a que  $n$  résultats possibles :  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ . Si chacun de ces résultats a la même probabilité alors, comme

$$1 = P(\Omega) = P(\{\omega_1\}) + \dots + P(\{\omega_n\}),$$

on doit avoir  $P(\{\omega_1\}) = \dots = P(\{\omega_n\}) = \frac{1}{n}$ . Ceci définit la probabilité uniforme :

### Définition

La **probabilité uniforme** sur  $\Omega$  (ou *distribution équiprobable*) est la probabilité  $P$  définie par : pour tout  $A = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}\} \subset \Omega$ ,

$$P(A) = \frac{k}{n} = \frac{\text{Card } A}{\text{Card } \Omega}.$$

On résume parfois ceci par : pour la probabilité uniforme,

$$P(\text{événement}) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

NB. L'expression courante « au hasard » fait souvent référence à un tirage selon la probabilité uniforme. Il en va ainsi lorsque les différents résultats jouent des rôles symétriques.

### Rappels de dénombrement :

#### Proposition

Soit  $E$  un ensemble fini.

— Une **permutation** de  $E$  est une façon d'ordonner les éléments de  $E$ .

Le nombre de permutations d'un ensemble à  $n$  éléments est

$$n! = n(n-1) \cdots 2 \cdot 1.$$

— Un **arrangement** de  $k$  éléments de  $E$  est une façon de choisir et d'**ordonner**  $k$  éléments de  $E$  : c'est une suite de  $k$  éléments de  $E$  distincts 2 à 2.

Le nombre d'arrangements de  $k$  éléments parmi  $n$  éléments (où  $0 \leq k \leq n$ ) est

$$A_n^k = \underbrace{n(n-1) \cdots (n-k+1)}_{k \text{ termes}} = \frac{n!}{(n-k)!}.$$

— Une **combinaison** de  $k$  éléments de  $E$  est une façon de choisir  $k$  éléments de  $E$ , **sans spécifier d'ordre** : c'est un sous-ensemble de  $E$  à  $k$  éléments.

Le nombre de combinaisons de  $k$  éléments parmi  $n$  éléments (où  $0 \leq k \leq n$ ) est

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

On peut dire aussi qu'un arrangement correspond à un tirage de  $k$  éléments *successivement* en mémorisant l'ordre de tirage, tandis qu'une combinaison correspond à un tirage de  $k$  éléments *simultanément*.

Un arrangement de  $n$  éléments parmi  $n$  est une permutation, donc  $A_n^n = n!$ .

### 1.3 Probabilités conditionnelles

#### Définition

Soit  $B$  un événement tel que  $P(B) > 0$ . L'application  $P(\cdot|B) : \mathcal{P}(\Omega) \longrightarrow [0,1]$  définie par

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

est une probabilité sur  $\Omega$ .

$P(A|B)$  est appelée la **probabilité conditionnelle de  $A$  sachant  $B$** .

#### Définition

Une **partition** de  $\Omega$  est une suite finie ou infinie  $(A_n)_n$  d'événements disjoints, dont la réunion est  $\Omega$  :

$$\text{pour tous } i \neq j, A_i \cap A_j = \emptyset, \quad \text{et} \quad \Omega = \bigcup_n A_n.$$

Par exemple, pour tout événement  $B$ , le couple  $(B, B^c)$  est une partition de  $\Omega$ .

#### Théorème (Théorème des probabilités totales)

Soit  $(A_n)_n$  une partition de  $\Omega$ . Pour tout événement  $A$ ,

$$P(A) = \sum_n P(A|A_n)P(A_n).$$

En particulier, pour tous événements  $A$  et  $B$ ,

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

#### Théorème (Formule de Bayes)

Soit  $(A_n)_n$  une partition de  $\Omega$ . Pour tout événement  $A$ , et tout événement  $A_i$  de la partition,

$$P(A_i|A) = \frac{P(A|A_i)P(A_i)}{\sum_n P(A|A_n)P(A_n)}.$$

En particulier, pour tous événements  $A$  et  $B$ ,

$$P(B^c|A) = \frac{P(A|B^c)P(B^c)}{P(A|B)P(B) + P(A|B^c)P(B^c)}.$$

## 1.4 Événements indépendants

### Définition

Deux événements  $A$  et  $B$  sont **indépendants** si

$$P(A \cap B) = P(A)P(B).$$

On a alors  $P(A|B) = P(A)$  et  $P(B|A) = P(B)$ , si  $P(A) \neq 0$  et  $P(B) \neq 0$ .

Ainsi,  $A$  et  $B$  sont indépendants si l'information que  $A$  est réalisé n'a pas d'influence sur la probabilité que  $B$  est réalisé aussi : cela correspond à la notion intuitive d'indépendance.

### Proposition (*Indépendance du complémentaire*)

Si  $A$  et  $B$  sont indépendants, alors  $A^c$  et  $B^c$  le sont aussi, de même que  $A$  et  $B^c$ .

On étend la définition à davantage d'événements :

### Définition

Une famille  $(A_i)_{i \in I}$  d'événements est **indépendante** si pour toute sous-famille finie  $A_{i_1}, \dots, A_{i_k}$  on a

$$P(A_{i_1} \cap A_{i_2} \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

En particulier, des événements  $A$ ,  $B$  et  $C$  sont indépendants si

$$P(A \cap B) = P(A)P(B), \quad P(B \cap C) = P(B)P(C), \quad P(A \cap C) = P(A)P(C) \\ \text{et} \quad P(A \cap B \cap C) = P(A)P(B)P(C).$$

### Proposition (*Indépendance du complémentaire*)

Si une famille  $(A_i)_{i \in I}$  d'événements sont indépendants, et que, pour tout  $i \in I$ , on définit un événement  $B_i$  par  $B_i = A_i$  ou  $B_i = A_i^c$ , alors la famille  $(B_i)_{i \in I}$  est indépendante.

Ainsi, si  $A, B, C$  sont indépendants, alors  $A^c, B, C$  aussi, de même que  $A^c, B, C^c$ , etc.

### Corollaire (*Loi binomiale*)

Pour toute suite de  $n$  événements **indépendants**  $A_1, \dots, A_n$  ayant tous la **même probabilité**  $P(A_i) = p$  on a, pour  $0 \leq k \leq n$ ,

$$P(\text{exactement } k \text{ événements parmi } A_1, \dots, A_n \text{ se réalisent}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

## 2 Variables aléatoires

Soit  $(\Omega, P)$  un espace de probabilité.

### Définition

Une **variable aléatoire** est une application  $X : \Omega \rightarrow \mathbb{R}$ .

### Définition

Soit  $X$  une variable aléatoire. La **loi** de  $X$  est la probabilité  $P_X$  sur  $\mathbb{R}$  définie par :

$$\text{pour tout } B \subset \mathbb{R}, \quad P_X(B) = P(\{\omega \in \Omega \mid X(\omega) \in B\}) \\ = P(X \in B).$$

$P_X$  peut aussi être vue comme une probabilité sur l'ensemble  $X(\Omega)$  des valeurs prises par  $X$ , aussi appelé **support** de  $P_X$ . On abrège  $X \sim P_X$  et on dit que  $X$  **suit** la loi  $P_X$ .

La seconde égalité est une nouvelle notation : on note  $\{X \in B\}$  l'événement formé des réalisations  $\omega$  pour lesquelles  $X(\omega) \in B$ , et on abrège  $P(\{X \in B\}) = P(X \in B)$ .

### Définition

Si  $A$  est un événement, on introduit la variable aléatoire **fonction indicatrice** de  $A$ , notée  $\mathbf{1}_A$ , qui indique si l'événement  $A$  est réalisé :

$$\text{pour tout } \omega \in \Omega, \quad \mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A. \end{cases}$$

### 2.1 Lois discrètes

Une variable aléatoire  $X$  est dite **discrète** si l'ensemble  $X(\Omega)$  des valeurs qu'elle prend est *dénombrable* (c'est-à-dire que l'on peut trouver une suite qui énumère tous les éléments de  $X(\Omega)$  : c'est le cas notamment si  $X(\Omega)$  est un ensemble fini,  $\mathbb{N}$ ,  $\mathbb{Z}$  ou  $\mathbb{Q}$ , **mais pas** l'intervalle  $[0,1]$  ni  $\mathbb{R}$ ). On dit aussi que la *loi* de  $X$  est discrète.

Si  $X$  est discrète, alors, pour tout  $B \subset \mathbb{R}$ , on peut calculer

$$P_X(B) = P(X \in B) = P\left(\bigcup_{x \in B \cap X(\Omega)} \{X = x\}\right) = \sum_{x \in B \cap X(\Omega)} P(X = x) = \sum_{x \in B} P(X = x).$$

Pour caractériser une loi discrète, il suffit donc de se donner les **probabilités élémentaires**

$$p_X(x) = P(X = x) \quad \text{pour tout } x \in X(\Omega).$$

### Définition-Proposition

Soit  $E \subset \mathbb{R}$ . Une famille  $(p(x))_{x \in E}$  est une **famille de probabilités élémentaires** si

1. pour tout  $x \in E$ ,  $p(x) \geq 0$
2.  $\sum_{x \in E} p(x) = 1$ .

Dans ce cas, on peut définir une variable aléatoire  $X$  (sur un espace de probabilité  $(\Omega, P)$ ), à valeurs dans  $E$ , de probabilités élémentaires  $(p(x))_{x \in E}$ , c'est-à-dire telle que,

$$\text{pour tout } x \in E, \quad P(X = x) = p(x).$$

**Exemple : loi uniforme sur un ensemble fini  $E$ .** Pour un ensemble fini  $E$ , on dit qu'une variable aléatoire  $X$  suit la loi uniforme sur  $E$  si sa loi est la probabilité uniforme sur  $E$  :

$$P(X = x) = \frac{1}{\text{Card } E} \quad \text{pour tout } x \in E.$$

**Exemple : loi de Bernoulli de paramètre  $p$ .** Pour tout événement  $A$  de probabilité  $p$ , la fonction indicatrice  $\mathbf{1}_A$  est une variable aléatoire à valeurs dans  $\{0,1\}$ , donc elle est discrète et sa loi est donnée par les probabilités

$$P(\mathbf{1}_A = 1) = P(A) = p \quad \text{et} \quad P(\mathbf{1}_A = 0) = 1 - p.$$

Cette loi est appelée **loi de Bernoulli** de paramètre  $p$  (où  $p \in [0,1]$ ), notée  $\mathcal{B}(p)$ . Si  $p = 1/2$ , c'est la loi uniforme sur  $\{0,1\}$ .

**Exemple : loi binomiale de paramètres  $n$  et  $p$ .** Si  $A_1, \dots, A_n$  sont  $n$  événements indépendants et ayant la même probabilité  $p$  (c'est-à-dire que  $P(A_1) = \dots = P(A_n) = p$ ), alors la variable aléatoire

$$S_n = \mathbf{1}_{A_1} + \dots + \mathbf{1}_{A_n} = \text{« nombre d'événements parmi } A_1, \dots, A_n \text{ qui se réalisent »}$$

est à valeurs dans  $\{0,1, \dots, n\}$ , donc est discrète, et sa loi est donnée par les probabilités

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{pour } k = 0, 1, \dots, n.$$

Cette loi est appelée **loi binomiale** de paramètres  $n$  et  $p$  (où  $n \in \mathbb{N}$  et  $p \in [0,1]$ ), notée  $\mathcal{B}(n,p)$ . C'est donc la loi du nombre de piles en  $n$  tirages d'une pièce qui a probabilité  $p$  de donner pile.

**Exemple : loi géométrique de paramètre  $p$ .** Si  $A_1, A_2, \dots$  sont des événements indépendants et ayant la même probabilité  $p > 0$ , alors la variable aléatoire

$$N = \min\{n \in \mathbb{N}^* \mid \mathbf{1}_{A_n} = 1\} = \text{« premier indice } n \text{ tel que } A_n \text{ se réalise »}$$

est à valeurs dans  $\mathbb{N}^*$  (on a  $P(N = \infty) = P(A_1^c \cap A_2^c \cap \dots) = (1-p)(1-p)\dots = 0$ ) et sa loi est donnée par

$$P(N = n) = (1-p)^{n-1} p, \quad \text{pour } n = 1, 2, \dots$$

Cette loi est appelée **loi géométrique** de paramètre  $p$  (où  $p \in ]0,1[$ ), notée  $\mathcal{G}(p)$ . C'est donc la loi du nombre de lancers d'une pièce pour obtenir pile, si sa probabilité de donner pile est  $p$ .

**Exemple : loi de Poisson de paramètre  $\lambda$ .** Soit  $\lambda > 0$ . Une variable aléatoire  $X$  suit la loi de Poisson de paramètre  $\lambda$ , notée  $\mathcal{P}(\lambda)$ , si  $X$  est à valeurs dans  $\mathbb{N}$  et

$$\text{pour tout } k \in \mathbb{N}, \quad P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

C'est la loi limite de la loi binomiale  $\mathcal{B}(n,p)$ , avec  $n$  grand et  $np \simeq \lambda$  :

**Proposition**

Si, pour tout  $n$ ,  $S_n$  suit la loi  $\mathcal{B}(n, p_n)$ , et  $np_n \xrightarrow[n \rightarrow \infty]{} \lambda$ , alors

$$\text{pour tout } k \in \mathbb{N}, \quad P(S_n = k) \xrightarrow[n \rightarrow \infty]{} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Dans la pratique, on peut souvent approcher la loi binomiale par une loi de Poisson lorsque les conditions  $n \geq 50$  et  $p \leq 0,1$  sont vérifiées (l'erreur dans les calculs de probabilités est alors inférieure à 5% ; elle devient inférieure à 1% si par exemple  $n \geq 50$  et  $p \leq 0,01$ ).

## 2.2 Lois continues (ou à densité)

Une variable aléatoire  $X$  est dite **continue** ou **à densité** s'il existe une fonction  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  telle que, pour tout  $B \subset \mathbb{R}$ ,

$$P_X(B) = P(X \in B) = \int_B f_X(x) dx.$$

La fonction  $f_X$  est appelée la **densité** de  $X$ .

### Définition-Proposition

Une fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  est une **fonction de densité (de probabilité)** si

1. pour tout  $x \in \mathbb{R}$ ,  $f(x) \geq 0$

2.  $\int_{\mathbb{R}} f(x) dx = 1$ .

Dans ce cas, on peut définir une variable aléatoire  $X$  (sur un espace de probabilité  $(\Omega, P)$ ) de densité  $f_X = f$ .

### Propriétés

Si la variable aléatoire  $X$  a pour densité  $f_X$ , alors

a) pour tout  $x \in \mathbb{R}$ , on a  $P(X = x) = 0$

b)  $P(X \in \text{Supp}(f_X)) = 1$ , où le support de la fonction  $f_X$  est défini par

$$\text{Supp}(f_X) = \{x \in \mathbb{R} \mid f_X(x) > 0\}.$$

Par le point b), on pourra considérer que les valeurs de  $X$  sont dans l'ensemble  $\text{Supp}(f_X)$ .

**Exemple : loi uniforme sur un segment  $[a, b]$ .** Soit  $a < b$ . La loi uniforme sur  $[a, b]$ , notée  $\mathcal{U}([a, b])$ , est la loi de densité

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a, b]}(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{si } x \notin [a, b]. \end{cases}$$

Comme  $f(x) = 0$  si  $x \notin [a, b]$ , une variable aléatoire  $X$  de loi  $\mathcal{U}([a, b])$  est à valeurs dans  $[a, b]$ .

**Exemple : loi exponentielle de paramètre  $\lambda$ .** Soit  $\lambda > 0$ . La loi exponentielle de paramètre  $\lambda$ , notée  $\mathcal{E}(\lambda)$ , est la loi de densité

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Comme  $f(x) = 0$  si  $x < 0$ , une variable aléatoire  $X$  de loi  $\mathcal{E}(\lambda)$  est à valeurs dans  $\mathbb{R}_+$ .

La loi exponentielle est une loi « sans mémoire ». En effet, si  $X \sim \mathcal{E}(\lambda)$ , pour tous  $s, t \geq 0$ ,

$$P(X \geq s+t \mid X \geq s) = \frac{P(\{X \geq s+t\} \cap \{X \geq s\})}{P(X \geq s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X \geq t),$$

en utilisant le fait que l'événement  $\{X \geq s+t\}$  est inclus dans l'événement  $\{X \geq s\}$ .

De ce fait, les lois exponentielles sont utilisées pour modéliser des durées de vie de machines sans vieillissement.

## 2.3 Fonction de répartition

### Définition

Soit  $X$  une variable aléatoire. La **fonction de répartition** de  $X$  est la fonction  $F_X : \mathbb{R} \rightarrow \mathbb{R}$  définie par

$$\text{pour tout } x \in \mathbb{R}, \quad F_X(x) = P(X \leq x).$$

### Proposition

a) Soit  $X$  une variable aléatoire. Sa fonction de répartition  $F_X$  est une fonction croissante,

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F_X(x) = 1.$$

b) Si  $X$  et  $Y$  sont deux variables aléatoires telles que  $F_X(t) = F_Y(t)$  pour tout  $t \in \mathbb{R}$ , alors  $X$  et  $Y$  ont même loi.

c) Si  $X$  est une variable aléatoire discrète, alors  $F_X$  est une fonction constante par morceaux, dont les sauts se situent aux points de  $X(\Omega)$  et, pour  $x \in X(\Omega)$ , la hauteur du saut au point  $x$  est égale à  $P(X = x)$ .

d) Si  $X$  est une variable aléatoire de densité  $f_X$ , alors on a

$$\text{pour tout } x \in \mathbb{R}, \quad F_X(x) = \int_{-\infty}^x f_X(t) dt,$$

donc en particulier  $F_X$  est continue, et on a la dérivée  $(F_X)'(x) = f_X(x)$  (sauf aux points  $x$  où  $f_X$  n'est pas continue).

Inversement, si  $F_X$  est continue, et dérivable (sauf éventuellement en un nombre fini de points), alors  $X$  a pour densité  $f_X(x) = (F_X)'(x)$ .

## 2.4 Loi de $Y = \varphi(X)$

Soit  $X$  une variable aléatoire à valeurs réelles, et  $\varphi : X(\Omega) \rightarrow \mathbb{R}$  une fonction. On peut définir la variable aléatoire  $Y = \varphi(X)$ .

• Si  $X$  est discrète, alors  $Y$  aussi, et sa loi se calcule directement :  $Y$  est à valeurs dans  $\varphi(X(\Omega))$  et, pour tout  $y \in \varphi(X(\Omega))$ ,

$$P(Y = y) = P(\varphi(X) = y) = \sum_{x \in \mathbb{R} \text{ tel que } \varphi(x)=y} P(X = x).$$

• Si  $X$  a une densité  $f_X$ , on peut souvent montrer que  $Y$  a une densité et la calculer :

- *Via la fonction de répartition.* Si  $\varphi$  est strictement croissante sur un intervalle  $I$  tel que  $X \in I$  presque sûrement, alors  $\varphi$  est une bijection de  $I$  sur  $J = \varphi(I)$  et on a, pour  $y \in J$ ,

$$F_Y(y) = P(Y \leq y) = P(\varphi(X) \leq y) = P(X \leq \varphi^{-1}(y)) = F_X(\varphi^{-1}(y)).$$

On pourra alors vérifier la continuité de  $F_Y$  et dériver (si possible) pour obtenir la densité de  $Y$ .

- *Par changement de variable.* Sous les mêmes conditions : pour tous  $a, b \in J$  avec  $a < b$  on a, en posant  $y = \varphi(x)$ , c'est-à-dire  $x = \varphi^{-1}(y)$ , (si  $\varphi^{-1}$  est de classe  $\mathcal{C}^1$ )

$$P(a < Y < b) = P(\varphi^{-1}(a) < X < \varphi^{-1}(b)) = \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} f_X(x) dx = \int_a^b f_X(\varphi^{-1}(y)) (\varphi^{-1})'(y) dy$$

donc  $Y$  a pour densité  $f_Y(y) = f_X(\varphi^{-1}(y)) (\varphi^{-1})'(y)$ . On retrouverait le même résultat.

## 2.5 Espérance d'une variable aléatoire

### Définition

L'**espérance** d'une variable aléatoire  $X$ , notée  $E[X]$ , est la moyenne de ses valeurs, pondérées par leurs probabilités.

Si  $X$  est discrète,

$$E[X] = \sum_{x \in X(\Omega)} xP(X = x).$$

Si  $X$  est continue, de densité  $f_X$ ,

$$E[X] = \int_{\mathbb{R}} xf_X(x)dx.$$

**Attention.** L'espérance n'est pas toujours définie. Il faut pour cela que la série ou l'intégrale ci-dessus converge absolument.

### Propriétés

a) Si  $X$  est constante, égale à  $c \in \mathbb{R}$  (pour tout  $\omega \in \Omega$ ,  $X(\omega) = c$ ), alors  $E[X] = E[c] = c$ .

b) Pour tout événement  $A \subset \Omega$ ,  $E[\mathbf{1}_A] = P(A)$ .

c) L'espérance est linéaire : pour toutes variables aléatoires  $X$  et  $Y$ , et tout réel  $a$ ,

$$E[aX] = aE[X] \quad \text{et} \quad E[X + Y] = E[X] + E[Y].$$

d) L'espérance est croissante : si  $X \leq Y$ , alors  $E[X] \leq E[Y]$ .

### Proposition

Soit  $X$  une variable aléatoire, et  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction.

— Si  $X$  est discrète, alors

$$E[\varphi(X)] = \sum_{x \in X(\Omega)} \varphi(x)P(X = x).$$

— Si  $X$  est continue, alors

$$E[\varphi(X)] = \int_{\mathbb{R}} \varphi(x)f_X(x)dx.$$

(À condition que la série et l'intégrale soient bien définies)

## 2.6 Variance d'une variable aléatoire

### Définition

Soit  $X$  une variable aléatoire. La **variance** de  $X$  est l'espérance des carrés des écarts de  $X$  à sa moyenne :

$$\text{Var}(X) = E\left[(X - E[X])^2\right] \geq 0.$$

L'**écart type** de  $X$  est  $\sigma(X) = \sqrt{\text{Var}(X)}$ .

**Attention.** La variance n'est pas toujours définie. Il faut que l'espérance  $E[X]$  soit définie et que l'espérance ci-dessus converge. Tout ceci revient à demander à ce que  $E[X^2]$  converge.

NB. L'écart type  $\sigma(X)$  est *homogène* à  $X$  : si par exemple  $X$  est une distance, alors  $\sigma(X)$  est une distance aussi. Ceci justifie l'intérêt de l'écart type par rapport à la variance.

## Propriétés

Pour toutes variables aléatoires  $X$  et  $Y$  ayant une variance, et toute constante  $a$ ,

a)  $\text{Var}(X) = E[X^2] - E[X]^2$

b)  $\text{Var}(aX) = a^2 \text{Var}(X)$

c)  $\text{Var}(X + a) = \text{Var}(X)$

d)  $\text{Var}(X + Y) = \text{Var}(X) + 2 \text{Cov}(X, Y) + \text{Var}(Y)$ , où la **covariance** est définie par

$$\text{Cov}(X, Y) = E\left[(X - E[X])(Y - E[Y])\right] = E[XY] - E[X]E[Y].$$

NB. Si  $X$  a une variance, alors  $Y = \frac{X - E[X]}{\sigma(X)}$  est **centrée** ( $E[Y] = 0$ ) et **réduite** ( $\text{Var}(Y) = 1$ ).

Pour  $r > 0$ , on définit (s'ils existent) le **moment d'ordre  $r$**  et le **moment centré d'ordre  $r$**  :

$$m_r(X) = E[X^r] \quad \text{et} \quad \mu_r(X) = E\left[(X - E[X])^r\right].$$

## 2.7 Inégalités

### Proposition (*Inégalité de Markov*)

Soit  $X$  une variable aléatoire. Pour tout  $a > 0$ ,

$$P(|X| \geq a) \leq \frac{E[|X|]}{a}.$$

Plus généralement, pour tout  $a > 0$  et  $r > 0$ ,

$$P(|X| \geq a) \leq \frac{E[|X|^r]}{a^r}.$$

**Démonstration :** Dans le cas où  $X$  a pour densité  $f_X$ ,

$$\begin{aligned} E[|X|^r] &= \int_{\mathbb{R}} |x|^r f_X(x) dx = \int_{-\infty}^{-a} |x|^r f_X(x) dx + \int_{-a}^a |x|^r f_X(x) dx + \int_a^{+\infty} |x|^r f_X(x) dx \\ &\geq \int_{-\infty}^{-a} a^r f_X(x) dx + 0 + \int_a^{+\infty} a^r f_X(x) dx \\ &= a^r P(X \leq -a) + a^r P(X \geq a) = a^r P(|X| \geq a), \end{aligned}$$

d'où l'inégalité annoncée (la première correspond à  $r = 1$ ). ■

### Proposition (*Inégalité de Tchebychev*)

Soit  $X$  une variable aléatoire ayant une variance. Pour tout  $a > 0$ ,

$$P\left(|X - E[X]| \geq a\right) \leq \frac{\text{Var}(X)}{a^2}.$$

Autre écriture : pour tout  $A > 0$ ,

$$P\left(E[X] - A\sigma(X) \leq X \leq E[X] + A\sigma(X)\right) \geq 1 - \frac{1}{A^2}.$$

**Démonstration :** Appliquer l'inégalité de Markov pour  $r = 2$  à  $X - E[X]$ . ■

## 2.8 Indépendance de variables aléatoires

### Définition

Des variables aléatoires  $X_1, \dots, X_n$  sont **indépendantes** si, pour tous  $B_1, \dots, B_n \subset \mathbb{R}$ ,

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n).$$

Ici, les virgules se lisent « et » :  $P(X_1 \in B_1, \dots, X_n \in B_n) = P(\{X_1 \in B_1\} \cap \dots \cap \{X_n \in B_n\})$ .

Par exemple, deux variables aléatoires  $X$  et  $Y$  sont indépendantes si les événements qui ne dépendent que de  $X$  sont indépendants des événements qui ne dépendent que de  $Y$ .

Ainsi,  $X$  et  $Y$  sont indépendantes si connaître la valeur de  $X$  ne renseigne pas sur  $Y$ . Les résultats de deux tirages à pile-ou-face ou de deux lancers de dés sont en général indépendants. De même, deux tirages de boules numérotées dans une même urne *avec remise* sont indépendants, tandis que deux tirages *sans remise* ne sont pas indépendants car dans ce cas le second tirage ne peut pas être égal au premier.

### Proposition

- Si  $X_1, \dots, X_n$  sont indépendantes, alors les variables aléatoires  $f_1(X_1), \dots, f_n(X_n)$  sont indépendantes, quelles que soient les fonctions  $f_1, \dots, f_n$ .
- « **Indépendance par paquets** ». Si  $X_1, \dots, X_n$  sont indépendantes alors, par exemple, les variables aléatoires  $f(X_1, X_2), g(X_4), h(X_3, X_5, X_6), \dots$  sont indépendantes : les fonctions de « paquets disjoints » de variables sont indépendantes.
- Si des événements  $A_1, \dots, A_n$  sont indépendants, alors leurs fonctions indicatrices  $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$  sont des variables aléatoires indépendantes; et réciproquement.

### Proposition

Si  $X_1, \dots, X_n$  sont des variables aléatoires indépendantes, alors

- si leurs espérances sont bien définies,

$$E[X_1 \cdots X_n] = E[X_1] \cdots E[X_n]$$

- si leurs variances sont bien définies, on a  $\text{Cov}(X_i, X_j) = 0$  pour tous  $i \neq j$ , d'où

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n).$$

NB. Les réciproques sont fausses !

Par le a) de la proposition précédente on déduit, si  $X_1, \dots, X_n$  sont indépendantes,

$$E[f_1(X_1) \cdots f_n(X_n)] = E[f_1(X_1)] \cdots E[f_n(X_n)].$$

On rappelle que l'espérance est toujours linéaire : même si  $X_1, \dots, X_n$  ne sont pas indépendantes,

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n].$$

## 2.9 Théorème (ou « Loi ») des grands nombres

Le résultat suivant est fondamental :

### Théorème (Loi des grands nombres)

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires **indépendantes et de même loi**, d'espérance  $m$  et de variance  $\sigma^2$ . On définit la variable aléatoire  $\bar{X}_n$ , par

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

$\bar{X}_n$  est appelée **moyenne empirique**. On a :

$$\text{pour tout } \varepsilon > 0, \quad P\left(|\bar{X}_n - m| < \varepsilon\right) \xrightarrow[n \rightarrow \infty]{} 1.$$

**Démonstration :** Par linéarité de l'espérance,  $E[\bar{X}_n] = \frac{1}{n}(E[X_1] + \cdots + E[X_n]) = m$ , et les variables sont indépendantes donc

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}.$$

L'inégalité de Tchebychev pour  $\bar{X}_n$  s'écrit donc, pour tout  $\varepsilon > 0$ ,

$$P\left(|\bar{X}_n - m| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2},$$

ce qui donne, en passant au complémentaire,

$$P\left(|\bar{X}_n - m| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

Comme le terme de droite converge vers 1 quand  $n \rightarrow \infty$ , et que le terme de gauche est toujours  $\leq 1$ , on obtient l'énoncé. ■

NB. Si  $(A_n)_{n \geq 1}$  est une suite d'événements *indépendants* et qui ont *même probabilité*  $p$  (par exemple, dans une suite de tirages à pile-ou-face,  $A_n = \{\text{le } n\text{-ième tirage est pile}\}$ , et  $p = \frac{1}{2}$ ), alors en posant  $X_i = \mathbf{1}_{A_i}$ , on a

$$\bar{X}_n = \frac{\mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}}{n} = \frac{\text{nombre d'événements réalisés parmi } A_1, \dots, A_n}{n}$$

donc  $\bar{X}_n$  est la **fréquence de réalisation** des événements  $A_1, \dots, A_n$ , et la loi des grands nombres montre que, si  $n$  est grand, cette fréquence a de grandes chances d'être proche de  $E[X_1] = p$ , qui est la probabilité commune des événements  $A_n$ . On retrouve la définition intuitive d'une probabilité.

Ainsi, la fréquence d'apparition de pile dans une suite de tirages à pile-ou-face indépendants converge vers  $\frac{1}{2}$  (ou vers la probabilité d'obtenir pile, si la pièce est biaisée).

## 2.10 Loi normale de moyenne $m$ et de variance $\sigma^2$

Les lois normales joueront un rôle important dans la partie Statistiques (Section 5.4).

### Définition

La **loi normale centrée** ( $m = 0$ ) **réduite** ( $\sigma = 1$ ), notée  $\mathcal{N}(0,1)$ , est la loi de densité

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

Si  $m \in \mathbb{R}$  et  $\sigma \in ]0, +\infty[$ , la **loi normale de moyenne  $m$  et de variance  $\sigma^2$** , notée  $\mathcal{N}(m, \sigma^2)$ , est la loi de la variable aléatoire  $X = m + \sigma Z$ , où  $Z$  suit la loi  $\mathcal{N}(0,1)$ .

Lorsque  $X$  suit une loi normale, on dit que  $X$  est une variable aléatoire **gaussienne**.

La courbe représentative de  $f$  est la « courbe en cloche » (voir fin du poly). Si  $Z \sim \mathcal{N}(0,1)$ , sa fonction de répartition est

$$\Phi(x) = P(Z \leq x) = \int_{-\infty}^x e^{-t^2/2} \frac{dt}{\sqrt{2\pi}}.$$

Cette intégrale  $\Phi(x)$  ne pouvant pas s'exprimer à l'aide des fonctions usuelles, on utilise une table (imprimée, comme à la fin du photocopié, ou dans un logiciel de calcul numérique), ou une approximation : par exemple,

$$P(Z > x) = 1 - \Phi(x) \underset{x \rightarrow \infty}{\sim} \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}$$

(avec une erreur relative inférieure à 0,2 si  $x > 1,9$  (et inférieure à 0,1 si  $x > 2,9$ )).

Les valeurs d'une variable de loi  $\mathcal{N}(0,1)$  sont fortement concentrées près de 0 : si  $Z$  suit la loi  $\mathcal{N}(0,1)$ , on obtient, avec la table (ou une approximation),

$$P(-2 < Z < 2) \simeq 0,954 \quad \text{et} \quad P(-3 < Z < 3) \simeq 0,997.$$

De la loi normale centrée réduite se déduisent les autres lois normales. Avec la section 2.4, on obtient que la densité de la loi  $\mathcal{N}(m, \sigma^2)$  est

$$f_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Pour les calculs, on se ramène à la loi  $\mathcal{N}(0,1)$  : par exemple, si  $X$  suit la loi  $\mathcal{N}(m, \sigma^2)$ ,

$$P(m - 2\sigma < X < m + 2\sigma) = P\left(-2 < \frac{X - m}{\sigma} < 2\right) \simeq 0,954.$$

### Proposition

Toute combinaison linéaire de variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne.

Plus précisément, si  $X_1, \dots, X_n$  sont indépendantes et  $X_i \sim \mathcal{N}(m_i, \sigma_i^2)$  alors, pour tous  $a_1, \dots, a_n \in \mathbb{R}$ , la variable aléatoire

$$X = a_1 X_1 + \dots + a_n X_n$$

suit la loi  $\mathcal{N}(M, \Sigma^2)$ , où

$$M = E[X] = \sum_{i=1}^n a_i m_i \quad \text{et} \quad \Sigma^2 = \text{Var}(X) = \sum_{i=1}^n a_i^2 \sigma_i^2.$$

### 3 Couples de variables aléatoires

#### 3.1 Loi du couple, lois marginales

##### Définition

Soit  $X, Y$  deux variables aléatoires. La **loi du couple**  $(X, Y)$  est la probabilité  $P_{(X, Y)}$  sur  $\mathbb{R}^2$  qui vérifie :

$$\text{pour tous } A, B \subset \mathbb{R}, \quad P_{(X, Y)}(A \times B) = P(X \in A, Y \in B).$$

Les lois de  $X$  et de  $Y$  se déduisent de  $P_{(X, Y)}$  : ainsi, pour  $A \subset \mathbb{R}$ ,  $P_X(A) = P_{(X, Y)}(A \times \mathbb{R})$ .  $P_X$  et  $P_Y$  sont appelées les **lois marginales** de  $P_{(X, Y)}$ .

Si  $X$  et  $Y$  sont indépendantes, la loi du couple est fournie par les lois de  $X$  et de  $Y$  :

$$P_{(X, Y)}(A \times B) = P_X(A)P_Y(B).$$

La loi du couple contient davantage d'information que  $P_X$  et  $P_Y$  : elle indique aussi la façon dont les variables dépendent l'une de l'autre (connaître  $X$  peut renseigner sur  $Y$ ).

**Cas de deux variables discrètes.** Si  $X$  et  $Y$  sont discrètes alors la loi de  $(X, Y)$  est donnée par les probabilités élémentaires :

$$p_{(X, Y)}(x, y) = P(X = x, Y = y) \quad \text{pour tous } x \in X(\Omega), y \in Y(\Omega).$$

Elles vérifient  $p_{(X, Y)}(x, y) \in [0, 1]$  pour tous  $x, y$ , et

$$\sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} p_{(X, Y)}(x, y) = 1.$$

Inversement, les lois marginales se déduisent des  $(p_{(X, Y)}(x, y))$  :

$$\text{pour tout } x \in X(\Omega), \quad p_X(x) = P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y) = \sum_{y \in Y(\Omega)} p_{(X, Y)}(x, y),$$

$$\text{pour tout } y \in Y(\Omega), \quad p_Y(y) = P(Y = y) = \sum_{x \in X(\Omega)} P(X = x, Y = y) = \sum_{x \in X(\Omega)} p_{(X, Y)}(x, y).$$

NB.  $X$  et  $Y$  sont indépendantes si, et seulement si  $p_{(X, Y)}(x, y) = p_X(x)p_Y(y)$  pour tous  $x, y$ .

**Cas où  $P_{(X, Y)}$  a une densité.** On dit que le couple  $(X, Y)$  a une densité s'il y a une fonction  $f_{(X, Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}$  telle que

$$\text{pour tout } D \subset \mathbb{R}^2, \quad P_{(X, Y)}(D) = \iint_D f_{(X, Y)}(x, y) dx dy.$$

$f_{(X, Y)}$  est appelée la **densité** du couple  $(X, Y)$ . Alors  $f_{(X, Y)}(x, y) \geq 0$  pour tous  $x, y \in \mathbb{R}$ , et

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f_{(X, Y)}(x, y) dx dy = 1.$$

Inversement, une telle fonction définit la loi d'un couple  $(X, Y)$ .

Comme pour les simples variables aléatoires, si  $(X, Y)$  a pour densité  $f_{(X, Y)}$ , alors, presque sûrement,  $(X, Y) \in \text{Supp}(f_{(X, Y)})$  où le support de la fonction  $f_{(X, Y)}$  est défini par

$$\text{Supp}(f_{(X, Y)}) = \{(x, y) \in \mathbb{R}^2 \mid f_{(X, Y)}(x, y) > 0\}.$$

On déduit les lois marginales de la loi du couple et, *dans le cas indépendant*, on déduit la loi du couple des lois marginales :

**Proposition**

a) Si  $(X,Y)$  a pour densité  $f_{(X,Y)}$ , alors  $X$  et  $Y$  ont des densités  $f_X$  et  $f_Y$  données par

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dy \quad \text{et} \quad f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dx.$$

b) Si  $X$  et  $Y$  ont des densités  $f_X$  et  $f_Y$  et sont indépendantes, alors  $(X,Y)$  a pour densité

$$f_{(X,Y)}(x,y) = f_X(x)f_Y(y).$$

Réciproquement, si  $f_{(X,Y)}(x,y) = f(x)g(y)$  pour deux fonctions  $f$  et  $g$ , alors  $X$  et  $Y$  sont indépendantes, et les densités de  $X$  et  $Y$  sont proportionnelles à  $f$  et  $g$ .

### 3.2 Calculs d'espérances

Connaître la loi du couple  $(X,Y)$  permet de calculer l'espérance de fonctions de  $X$  et  $Y$  :

**Proposition**

Soit  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  une fonction.

— Si  $X$  et  $Y$  sont discrètes, alors

$$E[\varphi(X,Y)] = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} \varphi(x,y)P(X = x, Y = y).$$

— Si  $(X,Y)$  a pour densité  $f_{(X,Y)}$ , alors

$$E[\varphi(X,Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x,y)f_{(X,Y)}(x,y)dx dy.$$

(À condition que les séries et les intégrales soient bien définies)

### 3.3 Somme de variables aléatoires indépendantes à densité.

**Proposition**

Si  $X$  et  $Y$  sont des variables aléatoires indépendantes, de densités  $f_X$  et  $f_Y$ , alors la variable aléatoire  $Z = X + Y$  a une densité, donnée par le produit de convolution de  $f_X$  et  $f_Y$  :

$$f_{X+Y}(z) = (f_X * f_Y)(z) = \int_{\mathbb{R}} f_X(x)f_Y(z-x)dx.$$

**Démonstration :** Comme  $X,Y$  sont indépendantes,  $(X,Y)$  a pour densité  $f_{(X,Y)}(x,y) = f_X(x)f_Y(y)$  donc, pour tout  $A \subset \mathbb{R}$ , en faisant le changement de variable  $y \mapsto z = x + y$ , c'est-à-dire  $y = z - x$ ,

$$\begin{aligned} P(X + Y \in A) &= E[\mathbf{1}_A(X + Y)] = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \mathbf{1}_A(x + y)f_X(x)f_Y(y)dy \right) dx = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \mathbf{1}_A(z)f_X(x)f_Y(z - x)dz \right) dx \\ &= \int_{\mathbb{R}} \mathbf{1}_A(z) \left( \int_{\mathbb{R}} f_X(x)f_Y(z - x)dx \right) dz = \int_{\mathbb{R}} \mathbf{1}_A(z)(f_X * f_Y)(z)dz. \quad \blacksquare \end{aligned}$$

NB. Le même principe permet de calculer la densité de  $Z = \varphi(X,Y)$  pour diverses fonctions  $\varphi$ .

## 4 Statistiques descriptives

Les statistiques descriptives visent à décrire un ensemble, en général important, de données, c'est-à-dire à en résumer certaines particularités, sous la forme de représentations graphiques ou de grandeurs numériques. L'interprétation des résultats est ensuite propre à chaque champ d'application. NB. Cette partie n'utilise pas de probabilités (à la différence de la prochaine).

### 4.1 Vocabulaire, représentations simples

En statistiques, les données dont on dispose associent à chaque **individu** d'un certain ensemble, appelé la **population**, une ou plusieurs **variables** qui quantifient ou qualifient certains caractères des individus (remarque : les individus peuvent être des personnes, mais aussi des objets, des dates, des lieux...). Ces données sont aussi appelées une **série statistique**.

On dispose de données  $(x_1, y_1, \dots), (x_2, y_2, \dots), \dots, (x_n, y_n, \dots)$  où  $n$  est l'**effectif total** (taille de la population), et  $(x_i, y_i, \dots)$  sont les **observations** des variables associées au  $i$ -ième individu.

Une variable à valeurs numériques est dite **quantitative**. Dans le cas contraire, elle est dite **qualitative**, c'est-à-dire que ses valeurs sont des catégories et non pas des nombres : par exemple, des couleurs, ou des pays. Dans ce cas, les valeurs possibles sont appelées les **modalités** de la variable. Une variable quantitative est **discrète** si ses valeurs possibles sont restreintes à un ensemble fini, et elle est **continue** sinon. Une variable quantitative discrète peut d'ailleurs se voir comme une variable qualitative.

Pour une variable qualitative (ou quantitative discrète),

l'**effectif** d'une modalité (ou d'une valeur) est le nombre de fois où elle est présente dans la population. On représente graphiquement les effectifs par un **diagramme en bâtons**.

la **fréquence** d'une modalité (ou d'une valeur) est le quotient de l'effectif de cette valeur par l'effectif total. On représente graphiquement les fréquences par un **diagramme circulaire** (ou par un diagramme en bâtons, éventuellement empilés).

Pour représenter graphiquement des observations de variables quantitatives continues, on choisit en général un nombre fini d'intervalles et on classe les individus selon l'intervalle qui contient leur valeur : ainsi, on approche la variable continue par une variable qualitative. La représentation des effectifs constitue alors un **histogramme**. Attention, le choix des intervalles est essentiel et ne suit pas une règle générale.

### 4.2 Répartition des observations

Supposons que l'on dispose d'observations  $x_1, \dots, x_n$  d'une variable quantitative. Pour tout  $x \in \mathbb{R}$ , la **fréquence cumulée jusqu'à**  $x$  est la proportion d'observations inférieures à  $x$  :

$$F(x) = \frac{1}{n} \text{Card}\{i \in \{1, \dots, n\} \mid x_i \leq x\}.$$

Pour  $\alpha \in [0, 1]$ , un **quantile d'ordre**  $\alpha$  est un réel  $q_\alpha$  tel qu'une proportion  $\geq \alpha$  des données est dans  $] -\infty, q_\alpha]$ , et une proportion  $\geq 1 - \alpha$  des données est dans  $[q_\alpha, +\infty[$ .

En pratique, pour  $0 < \alpha < 1$ , si on ordonne les données  $x_{(1)} \leq \dots \leq x_{(n)}$ ,

- si  $\alpha n \notin \mathbb{N}$ , et  $k - 1 < \alpha n < k$  avec  $k \in \mathbb{N}$ , alors  $x_{(k)}$  est l'unique quantile d'ordre  $\alpha$  ;
- si  $\alpha n = k \in \mathbb{N}^*$ , les quantiles d'ordre  $\alpha$  sont les éléments de  $[x_{(k)}, x_{(k+1)}]$ .

Une **médiane** est un quantile d'ordre  $1/2$  : elle sépare les données en deux parties égales.

Un **premier quartile** est un quantile d'ordre  $1/4$ .

Un **troisième quartile** est un quantile d'ordre  $3/4$ .

Le **minimum** est le plus grand quantile d'ordre  $0$ , et le **maximum** est le plus petit d'ordre  $1$ .

On peut représenter ces valeurs par une **boîte de dispersion** (ou « boîte à moustaches »).

### 4.3 Moyenne, variance

Si  $x_1, \dots, x_n$  sont les observations d'une variable quantitative, la **moyenne** de cette série statistique est

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}.$$

Pour une variable quantitative discrète, de valeurs  $v_1, \dots, v_r$ , ayant pour effectifs  $n_1, \dots, n_r$  et fréquences  $f_1, \dots, f_r$ , on a aussi

$$\bar{x} = \frac{n_1 v_1 + \dots + n_r v_r}{n} = f_1 v_1 + \dots + f_r v_r.$$

Moyenne et médiane sont des valeurs qui « approchent » au mieux toutes les données :

#### Proposition

La moyenne est l'unique réel  $c$  qui minimise  $\sum_{i=1}^n (x_i - c)^2$ .  
Les médianes sont les réels  $c$  qui minimisent  $\sum_{i=1}^n |x_i - c|$ .

Remarquons que la moyenne est sensible aux valeurs aberrantes, à la différence de la médiane. La **variance** de la série statistique est

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (= \overline{x^2} - \bar{x}^2),$$

et son **écart type** est  $\sigma_x = \sqrt{s_x^2}$ . C'est une mesure de la dispersion des valeurs autour de la moyenne : si le quotient  $\frac{\sigma_x}{|\bar{x}|}$  est petit (et  $\bar{x} \neq 0$ ), les données sont concentrées autour de  $\bar{x}$ .

### 4.4 Observation d'un couple de variables, régression linéaire

On suppose que l'on observe deux variables pour chaque individu, si bien que l'on dispose de données  $(x_1, y_1), \dots, (x_n, y_n)$ .

La **covariance** des deux variables est

$$s_{x,y} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \quad (= \overline{xy} - \bar{x} \bar{y}),$$

(donc  $s_x^2 = s_{x,x}$ ) et leur **corrélation** est  $\rho_{x,y} = \frac{s_{x,y}}{\sigma_x \sigma_y}$ . Une corrélation positive (et proche de 1) indique que les variables ont tendance à être simultanément grandes, ou simultanément petites, tandis qu'une corrélation négative (et proche de -1) indique des variations en sens opposés.

Si les variables sont fortement corrélées (corrélation proche de  $\pm 1$ ), on peut penser qu'une variable est la cause de l'autre, et on peut chercher à approcher  $y$  par une fonction de  $x$ . Le cas le plus simple est le cas affine ( $y_i \simeq ax_i + b$ ), qui peut s'aborder par **régression linéaire** :

#### Proposition (*Droite des moindres carrés*)

La droite  $y = ax + b$  qui minimise la somme des carrés des erreurs

$$E(a,b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

est donnée par

$$a = \frac{s_{x,y}}{s_x^2} = \frac{\overline{xy} - \bar{x} \bar{y}}{\overline{x^2} - \bar{x}^2} \quad \text{et} \quad b = \bar{y} - a \bar{x}.$$

NB. Si on cherche une approximation de la forme  $y = Cx^\alpha$ , alors  $\ln(y) = \alpha \ln(x) + \ln(C)$ , ce qui permet de déterminer  $C$  et  $\alpha$  par régression linéaire entre  $\ln(x)$  et  $\ln(y)$ .

## 5 Estimation

### 5.1 Principe, statistiques classiques

#### Définition

Soit  $X$  une variable aléatoire. Un **échantillon de taille  $n$  de  $X$**  est une famille  $X_1, \dots, X_n$  de  $n$  variables aléatoires indépendantes et de même loi que  $X$ .

On souhaite étudier la loi de  $X$ . Par exemple,  $X$  est la taille en centimètres d'un individu choisi uniformément dans la population adulte française. Son espérance est donc la taille moyenne d'un Français adulte, que l'on peut vouloir estimer.

On ne dispose pour cela que d'une **observation** d'un échantillon de taille  $n$  : une réalisation  $(x_1, \dots, x_n) \in \mathbb{R}^n$  de  $n$  variables aléatoires indépendantes  $(X_1, \dots, X_n)$  qui ont la même loi que  $X$ . À défaut de pouvoir mesurer toute la population, ce qui serait long, coûteux et compliqué, on se contente de mesurer la taille de  $n$  personnes **choisies au hasard** parmi les Français adultes. L'objectif de l'estimation statistique est de déduire certaines propriétés de la loi de  $X$  (son espérance, sa variance, ses paramètres...) à partir des valeurs observées  $x_1, \dots, x_n$ . On va notamment appliquer des méthodes de statistiques descriptives à ces données.

Pour étudier la distribution (la loi) des quantités qui dépendent des données, on les considère comme des variables aléatoires, donc des fonctions de  $X_1, \dots, X_n$ . Lors de l'utilisation pratique, on remplacera  $X_1, \dots, X_n$  par l'observation  $x_1, \dots, x_n$  dont on dispose.

**Statistiques simples.** Les quantités les plus classiques pour décrire un échantillon sont

— la **moyenne empirique** :

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

— la **variance empirique** :

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

— la **variance empirique modifiée** :

$$\Sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

#### Proposition

Si  $X$  a pour espérance  $m$  et pour écart type  $\sigma$ , et  $X_1, \dots, X_n$  est un échantillon de  $X$ , alors

$$E[\bar{X}_n] = m, \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n},$$

$$E[S_n^2] = \frac{n-1}{n} \sigma^2, \quad E[\Sigma_n^2] = \sigma^2.$$

## 5.2 Estimateurs

On suppose que  $X$  suit une loi  $P_\theta$  qui dépend d'un paramètre  $\theta \in \Theta$ , où  $\Theta \subset \mathbb{R}$  est l'ensemble des valeurs a priori possibles du paramètre. On ignore la valeur de  $\theta$ , et on souhaite l'estimer. NB. Dans ce qui suit, on pourrait noter  $E_\theta$  au lieu de  $E$  pour indiquer que la loi de  $X$  et donc de  $X_1, \dots, X_n$  dépend de  $\theta$ ; on ne le fait pas, afin d'alléger la notation, mais il faut le garder à l'esprit.

### Définition

Un **estimateur** de  $\theta$  est une variable aléatoire  $T_n = f(X_1, \dots, X_n)$  qui dépend d'un échantillon  $X_1, \dots, X_n$  de  $X$ . On utilise souvent la notation  $\hat{\theta}$  pour un estimateur de  $\theta$ .

Une **estimation** de  $\theta$  est la valeur réelle  $t_n = f(x_1, \dots, x_n)$  prise par une réalisation particulière de l'échantillon.

NB. La définition d'estimateur peut paraître curieuse à plusieurs titres. On note que  $n$  n'apparaît pas dans l'appellation « estimateur de  $\theta$  » mais apparaît dans la définition. En fait, un estimateur peut être vu comme une suite  $(T_n)_n$  de variables aléatoires où  $T_n$  dépend de  $X_1, \dots, X_n$ , et on utilisera la variable  $T_n$  adaptée à la taille de l'échantillon dont on dispose. De plus,  $\theta$  n'apparaît pas dans la définition : n'importe quelle fonction de  $X_1, \dots, X_n$  est donc un estimateur de  $\theta$ . En revanche  $\theta$  intervient pour mesurer la qualité de l'estimateur :

### Définition

Soit  $T_n$  un estimateur de  $\theta$ .

Le **biais** de  $T_n$  est la différence  $E[T_n] - \theta$ .

On dit que  $T_n$  est **sans biais** si  $E[T_n] = \theta$ , quel que soit  $\theta \in \Theta$ .

On dit que  $T_n$  est **asymptotiquement sans biais** si  $E[T_n] \xrightarrow[n \rightarrow \infty]{} \theta$ , quel que soit  $\theta \in \Theta$ .

On dit que  $T_n$  est **convergent** si, quel que soit  $\theta \in \Theta$ ,

$$\text{pour tout } \alpha > 0, \quad P(|T_n - \theta| > \alpha) \xrightarrow[n \rightarrow \infty]{} 0.$$

### Proposition

Tout estimateur asymptotiquement sans biais dont la variance tend vers 0 est convergent.

### Définition

Le **risque quadratique** d'un estimateur  $T_n$  de  $\theta$  est

$$R_{T_n}(\theta) = E[(T_n - \theta)^2].$$

On dit que l'estimateur  $S_n$  est **meilleur** que  $T_n$  si, quel que soit  $\theta$ ,

$$R_{S_n}(\theta) \leq R_{T_n}(\theta).$$

Par l'inégalité de Markov, un estimateur dont le risque quadratique tend vers 0 (quel que soit  $\theta$ ) est convergent.

NB. Si  $T_n$  est sans biais, alors  $R_{T_n}(\theta) = \text{Var}(T_n)$ .

NB. On peut parfois vouloir estimer non pas un paramètre réel, mais une fonction de la loi de  $X$  : par exemple, les fréquences cumulées de l'échantillon (voir p. 16) estiment la fonction de répartition de  $X$ . Et un histogramme peut servir à estimer la densité de  $X$ , si  $X$  a une densité.

### 5.3 Construction d'estimateurs

**Méthode des moments** Le principe est d'utiliser la loi des grands nombres pour estimer les moments, et d'utiliser ensuite ces estimateurs des moments pour estimer  $\theta$ .

Par la loi des grands nombres, on a :

**Proposition**

Soit  $X$  une variable aléatoire d'espérance  $m$  et de variance  $\sigma^2$ .

- a) La moyenne empirique est un estimateur sans biais et convergent de  $m$ .
- b) La variance empirique est un estimateur asymptotiquement sans biais et convergent de  $\sigma^2$ , et la variance empirique modifiée est un estimateur sans biais et convergent de  $\sigma^2$ .
- c) Pour tout  $r > 0$ , le moment empirique d'ordre  $r$ ,

$$\widehat{m}_r = \frac{1}{n} \sum_{i=1}^n (X_i)^r$$

est un estimateur sans biais et convergent de  $m_r = E[X^r]$ . (Si  $m_r$  est bien défini)

On en déduit la méthode des moments : exprimer (si possible)  $\theta$  à l'aide des moments  $(m_r)_{r>0}$ , puis remplacer dans cette expression les moments par les moments empiriques. Ceci fournit un estimateur convergent de  $\theta$ . L'expression peut aussi faire intervenir  $\sigma^2$ , remplacé par  $S_n^2$  ou  $\Sigma_n^2$ . En pratique, on calcule  $E[X]$ ,  $E[X^2]$ , etc., jusqu'à obtenir une expression faisant intervenir  $\theta$  (souvent,  $E[X]$  suffit), et on inverse pour obtenir  $\theta$  en fonction de  $E[X]$ ,  $E[X^2]$ , etc. Il ne reste plus qu'à remplacer  $m_1 = E[X]$  par  $\bar{X}_n$ ,  $m_2 = E[X^2]$  par  $\widehat{m}_2$ , etc.

**Méthode du maximum de vraisemblance** Le principe est d'estimer  $\theta$  par la valeur qui maximise la densité du vecteur  $(X_1, \dots, X_n)$ .

**Définition**

La **vraisemblance** de l'échantillon  $(X_1, \dots, X_n)$  est la fonction  $L$  définie par :

- si  $X$  est discrète, de probabilité élémentaire  $P_\theta$ , pour tous  $x_1, \dots, x_n$ ,

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n P_\theta(x_i)$$

- si  $X$  est continue, de densité  $f_\theta$ , pour tous  $x_1, \dots, x_n$ ,

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_\theta(x_i).$$

Un **estimateur du maximum de vraisemblance** (ou EMV) pour  $\theta$  est un estimateur  $h(X_1, \dots, X_n)$  tel que, pour toutes les valeurs  $x_1, \dots, x_n \in X(\Omega)$ ,

$$L(x_1, x_2, \dots, x_n; h(x_1, \dots, x_n)) = \max_{\theta \in \Theta} L(x_1, \dots, x_n; \theta).$$

NB. Pour tous  $x_1, \dots, x_n$ ,  $h(x_1, \dots, x_n)$  est la ou l'une des valeurs de  $\theta$  où  $L(x_1, \dots, x_n; \theta)$  est maximum. Ceci définit l'estimation. L'*estimateur* est la variable aléatoire  $h(X_1, \dots, X_n)$ .

Sous des hypothèses assez générales, on montre que ceci définit un bon estimateur convergent.

Pour le calcul, on est amené à maximiser  $L(x_1, \dots, x_n; \theta)$  selon  $\theta$ . Vu que le logarithme est strictement croissant, c'est équivalent à maximiser la **log-vraisemblance**  $\ln(L(x_1, \dots, x_n; \theta))$ , souvent plus pratique. Pour maximiser, on étudie les variations de la fonction  $\theta \mapsto \ln(L(x_1, \dots, x_n; \theta))$ .

## 5.4 Intervalles de confiance

### Définition

Un **intervalle de confiance de niveau**  $1 - \alpha$  pour  $\theta$  est un intervalle  $I$ , qui dépend de  $X_1, \dots, X_n$ , tel que

$$P(\theta \in I) \geq 1 - \alpha.$$

**Intervalle de confiance pour la moyenne  $m$ .** Soit  $X$  une variable aléatoire de moyenne  $m$  et de variance  $\sigma^2$ .

Soit  $(X_1, X_2, \dots, X_n)$  un échantillon de  $X$ . On cherche un intervalle de confiance à partir des estimateurs sans biais  $\bar{X}_n$  et  $\Sigma_n^2$ .

### Théorème (Théorème central limite)

Soit  $(X_n)_n$  une suite de variables aléatoires indépendantes, de même loi de moyenne  $m$  et d'écart type  $\sigma$ . Soit  $Z_n$  la variable aléatoire définie par

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

Lorsque  $n \rightarrow +\infty$ ,  $Z_n$  converge en loi vers une variable  $Z \sim \mathcal{N}(0,1)$ . On a donc

$$\text{pour tout intervalle } I \subset \mathbb{R}, \quad P(Z_n \in I) \xrightarrow{n \rightarrow \infty} P(Z \in I) = \frac{1}{\sqrt{2\pi}} \int_I e^{-\frac{x^2}{2}} dx.$$

Dans la pratique on applique ce résultat dès que  $n$  est suffisamment grand ( $n \geq 30$ ).

Soit  $Z$  une variable aléatoire de loi  $\mathcal{N}(0,1)$  et  $a$  et  $\alpha$  définis par

$$P(-a \leq Z \leq a) = 1 - \alpha$$

Par exemple, on sait que (voir la table)

$$P(|Z| \leq 1,96) = 95\% \quad \text{et} \quad P(|Z| \leq 2,576) = 99\%.$$

Si  $n \geq 30$ , le théorème central limite permet d'écrire

$$P(-a \leq \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \leq a) \simeq 1 - \alpha.$$

On obtient, vu que  $\Sigma_n^2$  est un estimateur convergent de  $\sigma^2$ ,

$$P(-a \leq \frac{\sqrt{n}(\bar{X}_n - m)}{\sqrt{\Sigma_n^2}} \leq a) \simeq 1 - \alpha,$$

ce qui se réécrit

$$P\left(\bar{X}_n - a \frac{\sqrt{\Sigma_n^2}}{\sqrt{n}} \leq m \leq \bar{X}_n + a \frac{\sqrt{\Sigma_n^2}}{\sqrt{n}}\right) \simeq 1 - \alpha.$$

Si  $\bar{x}_n$  est la moyenne observée et  $\sigma_n^2$  la variance modifiée observée, on en déduit que

$$[\pi_1, \pi_2] = \left[ \bar{x}_n - a \frac{\sqrt{\sigma_n^2}}{\sqrt{n}}, \bar{x}_n + a \frac{\sqrt{\sigma_n^2}}{\sqrt{n}} \right]$$

est un intervalle de confiance de  $m$  de niveau  $1 - \alpha$ .

**Intervalle de confiance pour une proportion  $p$ .** Ici  $X$  est une variable aléatoire de Bernoulli de paramètre  $p$ , d'où  $E[X] = p$  et  $\text{Var}(X) = p(1 - p)$ .

On a, pour  $n$  grand,

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1 - p)}} \sim \mathcal{N}(0,1)$$

et l'intervalle de confiance pour  $p$  de niveau  $1 - \alpha$  précédent est donc

$$I = \left[ \bar{x}_n - a \frac{\sqrt{\bar{x}_n(1 - \bar{x}_n)}}{\sqrt{n}}, \bar{x}_n + a \frac{\sqrt{\bar{x}_n(1 - \bar{x}_n)}}{\sqrt{n}} \right].$$

*Remarques.*

1. Attention : pour que ces approximations soient justifiées, les valeurs de  $n\pi_1$ ,  $n\pi_2$ ,  $n(1 - \pi_1)$  et  $n(1 - \pi_2)$  doivent être toutes les quatre supérieures ou égales à 5.
2. Les intervalles de confiance donnés ci-dessus permettent aussi de déterminer la taille  $n$  de l'échantillon nécessaire pour avoir une précision donnée pour l'estimation d'une proportion.
3. Vu que, pour tout  $p \in [0,1]$ ,  $p(1 - p) \leq 1/4$ , on a  $I \subset J$  où  $J$  est donc aussi de niveau  $\alpha$  et plus simple mais un peu plus large :

$$J = \left[ \bar{x}_n - \frac{a}{2\sqrt{n}}, \bar{x}_n + \frac{a}{2\sqrt{n}} \right].$$

## 5.5 Tests d'hypothèse

On souhaite, à partir de l'échantillon observé  $x_1, \dots, x_n$ , savoir si l'on peut raisonnablement conclure qu'une certaine hypothèse sur la loi de  $X$  est fautive (en vue de prendre une décision). L'hypothèse est appelée **hypothèse nulle**, et notée  $\mathcal{H}_0$ . C'est une hypothèse que l'on veut avoir « peu de chance » de rejeter si elle est vraie (erreur de première espèce).

### Définition

Un **test de  $\mathcal{H}_0$  au seuil de risque  $\alpha$**  est une condition portant sur  $X_1, \dots, X_n$  permettant de décider si  $\mathcal{H}_0$  est rejetée ou non, et telle que,

$$\text{si } \mathcal{H}_0 \text{ est vraie, } \quad P(\text{rejet de } \mathcal{H}_0) \leq \alpha.$$

Le risque  $\alpha$  doit être petit, en général on demande  $\alpha = 5\%$ .

On peut aussi spécifier un condition alternative  $\mathcal{H}_1$  (si elle n'est pas précisée, c'est que  $\mathcal{H}_1$  est le contraire de  $\mathcal{H}_0$ ), et on parle de **test de  $\mathcal{H}_0$  contre  $\mathcal{H}_1$** . Le risque  $\alpha$  ne dépend que de  $\mathcal{H}_0$ .

### Définition

Rejeter  $\mathcal{H}_0$  alors que  $\mathcal{H}_0$  est vraie est l'**erreur de première espèce**.

Ne pas rejeter  $\mathcal{H}_0$  alors que  $\mathcal{H}_1$  est vraie est l'**erreur de deuxième espèce**.

On dit que le test de  $\mathcal{H}_0$  contre  $\mathcal{H}_1$  a une **puissance**  $1 - \beta$  si

$$\text{si } \mathcal{H}_1 \text{ est vraie, } \quad P(\text{rejet de } \mathcal{H}_0) \geq 1 - \beta.$$

NB. Le « test » qui consiste à ne *jamais* rejeter  $\mathcal{H}_0$  a un seuil de risque 0. Par contre, sa puissance est nulle. Il ne fait jamais d'erreur de première espèce mais toujours une erreur de deuxième espèce. Un bon test a un risque faible et une puissance proche de 1.

## Test de comparaison à une moyenne théorique

Soit  $\bar{x}_n$  la moyenne observée sur un échantillon de taille  $n$ . On veut savoir si cette moyenne est conforme à la moyenne théorique (c'est-à-dire l'espérance) annoncée,  $\mu$ .

Hypothèse  $\mathcal{H}_0$  : la moyenne théorique est  $\mu$ .

Sous l'hypothèse  $\mathcal{H}_0$ ,  $E[X] = \mu$ , si bien que, si  $n$  est suffisamment grand, approximativement,

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\Sigma_n^2}} \sim \mathcal{N}(0,1),$$

d'où par exemple

$$P\left(-1,96 \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\Sigma_n^2}} \leq 1,96\right) \simeq 95 \%.$$

Soit  $\sigma_n^2$  la variance empirique (modifiée ou non) observée sur l'échantillon. Alors on considère le test suivant :

$$\text{Test bilatéral : si } \left| \frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma_n} \right| > 1,96, \text{ alors on rejette l'hypothèse } \mathcal{H}_0.$$

C'est un test dont le seuil de risque est 5%. (On remplace 1,96 par  $a$  pour un seuil de risque  $\alpha$ ) Si l'hypothèse est rejetée, on peut donc conclure que l'espérance n'est pas  $\mu$ , et on se trompe dans seulement 5% des cas.

NB. Si le test est faux, on ne peut pas conclure que l'espérance est  $\mu$ , mais simplement que l'échantillon *ne permet pas d'exclure* qu'elle vaut  $\mu$ . On utilise ce test pour détecter (avec grande probabilité) les cas où la moyenne observée *n'est pas conforme* à  $\mu$ . Par exemple pour vérifier si un fabricant fournit bien des pièces qui ont une précision donnée, ou des médicaments qui ont une certaine efficacité, etc.

*Néanmoins*, ce test a heureusement tendance à conduire au rejet de  $\mathcal{H}_0$  lorsque  $\mathcal{H}_0$  est fautive : en effet, si  $E[X] = m$  avec  $m \neq \mu$ , alors  $\frac{\bar{X}_n - \mu}{\sqrt{\Sigma_n^2}} \simeq \frac{m - \mu}{\sigma} \neq 0$  donc  $\left| \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\Sigma_n^2}} \right| \xrightarrow[n \rightarrow \infty]{} +\infty$  et par conséquent le test rejette  $\mathcal{H}_0$  avec une probabilité qui tend vers 1.

Ceci revient à dire que le test précédent est un test de l'hypothèse  $\mathcal{H}_0$  « l'espérance est  $\mu$  » contre  $\mathcal{H}_1$  « l'espérance est  $\neq \mu$  » de risque 5% et dont la puissance tend vers 1 quand  $n \rightarrow \infty$ .

Souvent, le seul cas considéré à risque est lorsque  $E[X] > \mu$ , c'est-à-dire que l'hypothèse alternative est  $\mathcal{H}_1$  « l'espérance est  $> \mu$  ». On utilise alors un test **unilatéral** pour augmenter la puissance : par la table,

$$P(Z_n \leq 1,65) \simeq 95 \%,$$

d'où

$$\text{Test unilatéral : si } \frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma_n} > 1,65, \text{ alors on rejette l'hypothèse } \mathcal{H}_0.$$

C'est toujours un test de  $\mathcal{H}_0$  de risque 5%, mais par rapport au test précédent il a davantage tendance à conduire au rejet de  $\mathcal{H}_0$  si  $\mathcal{H}_1$  est vrai ; il est donc plus puissant.

On définit symétriquement un test de  $\mathcal{H}_0$  «  $E[X] = \mu$  » contre  $\mathcal{H}_1$  «  $E[X] < \mu$  ».

## Test de comparaison à une fréquence théorique

Soit  $\bar{x}_n$  la fréquence observée d'un caractère  $C$  sur un échantillon de taille  $n$ . On cherche à savoir si cette proportion paraît conforme à une fréquence théorique (probabilité) annoncée,  $\pi$ .

Hypothèse  $\mathcal{H}_0$  : la fréquence théorique est  $\pi$ .

Sous l'hypothèse  $\mathcal{H}_0$ ,  $X$  suit la loi de Bernoulli de paramètre  $\pi$ . Si  $n$  est suffisamment grand,

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \pi)}{\sqrt{\pi(1 - \pi)}} \sim \mathcal{N}(0,1),$$

d'où par exemple

$$P\left(-1,96 \leq \frac{\sqrt{n}(\bar{X}_n - \pi)}{\sqrt{\pi(1 - \pi)}} \leq 1,96\right) \leq 95\%.$$

Comme  $\bar{x}_n$  est l'estimation associée à  $\bar{X}_n$ , ceci conduit au test suivant, qui a un seuil de risque de 5% :

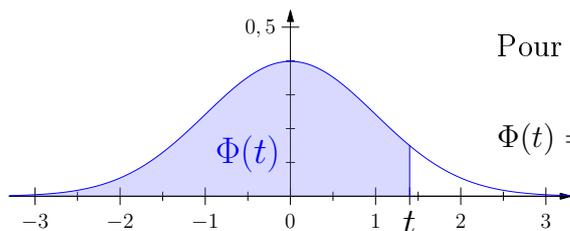
$$\text{Test bilatéral : si } \left| \frac{\sqrt{n}(\bar{x}_n - \pi)}{\sqrt{\pi(1 - \pi)}} \right| > 1,96 \text{ alors on rejette l'hypothèse } \mathcal{H}_0.$$

On peut ici aussi considérer le test de « la fréquence théorique est  $\pi$  » contre « la fréquence théorique est  $> \pi$  », auquel cas on utilise le test suivant, de risque 5% et plus puissant :

$$\text{Test unilatéral : si } \frac{\sqrt{n}(\bar{x}_n - \pi)}{\sqrt{\pi(1 - \pi)}} > 1,65 \text{ alors on rejette l'hypothèse } \mathcal{H}_0.$$

et de même symétriquement si  $\mathcal{H}_1$  est « la fréquence théorique est  $< \pi$  ».

**Table de la fonction de répartition de la loi normale centrée réduite  $\mathcal{N}(0,1)$**



Pour  $Z$  de loi  $\mathcal{N}(0,1)$ ,

$$\Phi(t) = F_Z(t) = P(Z \leq t) = \int_{-\infty}^t e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}}$$

$t$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

**Exemples :**  $\Phi(0,25) \simeq 0,5987$ ,  $\Phi(-0,32) = 1 - \Phi(0,32) \simeq 1 - 0,6255 = 0,3745$

**Memento des lois usuelles**

**Lois discrètes**

Nom	Paramètres	Support	Définition : $P(A) = \sum_{a \in A} p(a)$	Espérance	Variance
Loi de Dirac $\delta_a$	$a \in \mathbb{R}$	$\{a\}$	$p(a) = 1$	$a$	$0$
Loi de Bernoulli $\mathcal{B}(p)$	$p \in [0,1]$	$\{0,1\}$	$p(0) = 1 - p, p(1) = p$	$p$	$p(1 - p)$
Loi binomiale $\mathcal{B}(n,p)$	$n \in \mathbb{N}, p \in [0,1]$	$\{0, \dots, n\}$	$p(k) = \binom{n}{k} p^k (1 - p)^{n-k}$	$np$	$np(1 - p)$
Loi géométrique $\mathcal{G}(p)$	$p \in ]0,1]$	$\mathbb{N}^*$	$p(k) = (1 - p)^{k-1} p$	$\frac{1}{p}$	$\frac{1 - p}{p^2}$
Loi de Poisson $\mathcal{P}(\lambda)$	$\lambda \in ]0, +\infty[$	$\mathbb{N}$	$p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$	$\lambda$	$\lambda$

**Lois continues**

Nom	Paramètres	Support	Définition : $P(A) = \int_A f(x) dx$	Espérance	Variance
Loi uniforme $\mathcal{U}([a,b])$	$a < b$	$[a,b]$	$f(x) = \frac{1}{b - a} \mathbf{1}_{[a,b]}(x)$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
Loi exponentielle $\mathcal{E}(\lambda)$	$\lambda \in ]0, \infty[$	$]0, +\infty[$	$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{]0, +\infty[}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Loi normale/gaussienne $\mathcal{N}(m, \sigma^2)$	$m \in \mathbb{R}, \sigma^2 \in ]0, +\infty[$	$\mathbb{R}$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right)$	$m$	$\sigma^2$