# Statistique L3 CPES
# Notes de cours

Angelina Roche

Année universitaire 2022–2023

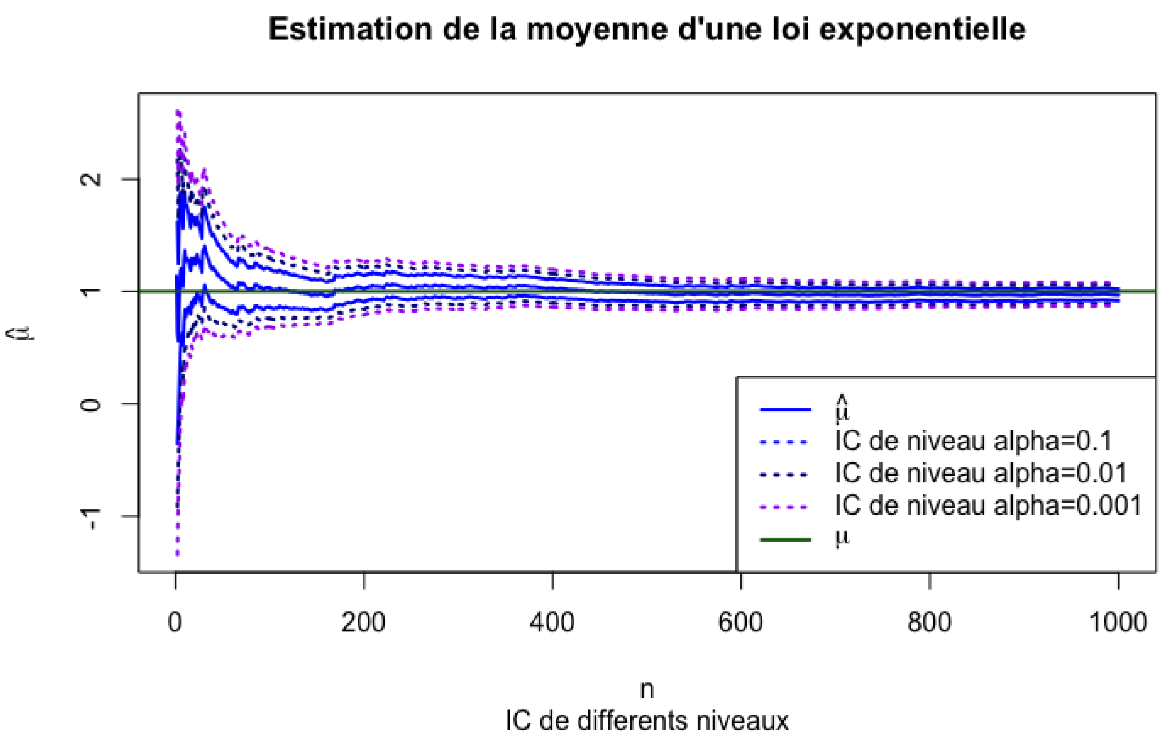


*Fig.1*

Ce polycopié est en cours de rédaction. Il a été construit à partir du cours de Marc Hoffman (2019–2020) et du livre *Statistique et probabilités en économie-gestion*, C. Hurlin et V. Mignon, éd. Dunod, 2018. ainsi que du polycopié *Intégration, Probabilités et Processus Aléatoires* de J.-F. Le Gall. N'hésitez pas à transmettre toute remarque ou à signaler toute coquille ou erreur à `roche@ceremade.dauphine.fr`. Merci à Mohammed Bouabdallah pour avoir relevé une erreur dans le poly et à Lucie Briand pour ses questions qui ont permis d'améliorer la rédaction des exercices.

# Table des matières

# Rappels de probabilité

## Espace de probabilité

**Définition 1** *Une* **expérience aléatoire** *est une expérience renouvelable, en théorie ou en pratique, et qui, renouvelée dans des conditions identiques ne donne pas forcément le même résultat à chaque renouvellement.*

Exemples : le lancer d'un dé, le jeu du pile ou face,....

**Définition 2** *L'***univers des possibles** *(ou* **univers***), noté $\Omega$ est l'ensemble de toutes les éventualités possibles, toutes les déterminations du hasard dans l'expérience considérée.*

Exemples :

1. On lance un dé à 6 faces. L'univers des possibles est l'ensemble des possibilités obtenues pour le résultat du lancer :
$$\Omega = \{1, \ldots, 6\}.$$
2. On lance un dé à 6 faces autant de fois que nécessaire jusqu'à obtenir un 6. Ici c'est plus compliquée, comme il est impossible de savoir combien de fois il faudra lancer le dé pour obtenir un 6, nous sommes obligés de considérer un ensemble $\Omega$ infini, représentant une infinité de lancers :
$$\Omega = \{\omega = (\omega_n)_{n\geq 1} \text{ avec, pour tout } n,\ \omega_n \in \{1, \ldots, 6\}\} = \{1, \ldots, 6\}^{\mathbb{N}\setminus\{0\}},$$
$\omega_n$ représente le résultat du $n$-ème lancer.
3. On mesure la température de l'air à un instant donné. Supposons que nous ne soyons pas limité à la précision de l'instrument de mesure, le résultat peut prendre n'importe quelle valeur réelle supérieure au 0 absolu (−273,15 °C).
$$\Omega = ]-273,15; +\infty[.$$

**Définition 3** *Une* **$\sigma$-algèbre** *ou* **tribu** *est un ensemble $\mathcal{A}$ de sous-ensembles de $\Omega$ qui vérifie les propriétés suivantes*

— $\Omega \in \mathcal{A}$, $\emptyset \in \mathcal{A}$ ;
— **stable par passage au complémentaire** *: si $A \in \mathcal{A}$, $A^c \in \mathcal{A}$ ;*
— **stable par union dénombrable** *: si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{A}$ alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.*

*Un ensemble $\Omega$ munit d'une $\sigma$-algèbre est appelé* **univers probabilisable***. Un élément $A \in \Omega$ est appelé* **évènement***.*

Exemples :

— L'ensemble $\mathcal{P}(\Omega) = \{A, A \subset \Omega\}$ l'**ensemble des parties de** $\Omega$ est une $\sigma$-algèbre appelée **tribu discrète**.

— L'ensemble $\{\emptyset, \Omega\}$ est également une $\sigma$-algèbre, appelée **tribu grossière**.

**Définition 4** *Soit* $(\Omega, \mathcal{A})$ *un univers probabilisable. Une* **probabilité** *sur* $(\Omega, \mathcal{A})$ *est une application*

$$\mathbb{P} : \mathcal{A} \to [0, 1]$$

*telle que*

— $\mathbb{P}(\Omega) = 1$

— *Pour toute suite* $(A_n)_{n\in\mathbb{N}}$ *d'évènements disjoints (i.e.* $A_j \cap A_k = \emptyset$*, si* $j \neq k$*),*

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

*Un ensemble probabilisable* $(\Omega, \mathcal{A})$ *munit d'une probabilité est appelé* **espace de probabilité***.*

Exemples :

1. Si $\Omega$ est fini, une mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est entièrement définie par sa valeur sur les évènements $\{\omega\}$ ne contenant qu'un élément (dits **singletons**). En effet, pour tout $A \in \mathcal{P}(\Omega)$,
$$\mathbb{P}(A) = \sum_{a\in A} \mathbb{P}(\{a\}).$$
Si $\mathbb{P}(\{\omega\}) = \mathbb{P}(\{\omega'\})$ pour tout $\omega, \omega' \in \Omega$, la probabilité est dite **uniforme**. Dans ce cas-là, la seule possibilité est
$$\mathbb{P}(\{\omega\}) = \frac{1}{\mathrm{Card}(\Omega)}.$$
Dans l'exemple du dé à 6 faces où $\Omega = \{1, \ldots, 6\}$, si le dé n'est pas pipé, alors il est raisonnable de penser que la probabilité est uniforme et donc que $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \ldots = \mathbb{P}(\{6\}) = 1/6$ (6 étant égal au cardinal de $\Omega$).
2. Si $\Omega$ est infini, il n'est plus possible de considérer une probabilité uniforme. Dans l'exemple où on lance un dé jusqu'à obtenir un 6, nous avons de plus que, pour tout $\omega = (\omega_n)_{n\in\mathbb{N}}$,
$$\mathbb{P}(\{\omega\}) = 0.$$
Prenons par exemple, $\omega = (1, \ldots, 1, \ldots)$ (l'évènement le dé fait 1 à chacun de ses lancers),
$$\begin{aligned} \mathbb{P}(\{\omega\}) &= \lim_{N\to+\infty} \mathbb{P}(\{\omega_1 = 1\} \cap \{\omega_2 = 1\} \cap \ldots \cap \{\omega_N = 1\}) = \lim_{N\to+\infty} \underbrace{\frac{1}{6} \times \frac{1}{6} \times \ldots \times \frac{1}{6}}_{N \text{ fois}} \\ &= \lim_{N\to+\infty} \left(\frac{1}{6}\right)^N = 0. \end{aligned}$$
Par contre, l'évènement $A = \{$ obtenir un 6 au deuxième lancer $\} = \{\omega_2 = 6\}$ est de probabilité non nulle (égale à 1/6). Dans ce cas-là, l'ensemble $A$ (tout comme l'ensemble $\Omega$) n'est pas dénombrable et
$$\frac{1}{6} = \mathbb{P}(A) \neq \sum_{a\in A} \mathbb{P}(\{\omega\}) = 0.$$

Dans la suite, sauf mention contraire, nous nous plaçons toujours dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

# Variables aléatoires

**Définition 5** *Une* **variable aléatoire** *est une application*

$$X : \Omega \to \mathbb{R}$$

*telle que, pour tout intervalle $I$ de $\mathbb{R}$*

$$X^{-1}(I) \in \mathcal{A}.$$

On distingue les **variables aléatoires continues** prenant leurs valeurs dans un ensemble non-dénombrable (par exemple $\mathbb{R}$ ou un intervalle de $\mathbb{R}$ ou encore $\mathbb{R}^d$) et les **variables aléatoires discrètes** prenant leurs valeurs dans un ensemble fini ou dénombrable (par exemple $\mathbb{N}$).

### 1.2.1 Lois discrètes usuelles

— **Loi de Bernoulli** de paramètre $p \in [0,1]$, notée $\mathcal{B}(p)$. Une variable aléatoire $X \sim \mathcal{B}(p)$, si $X \in \{0,1\}$ et

$$\mathbb{P}(X=1) = p \text{ et } \mathbb{P}(X=0) = 1-p$$

Exemple : lancer d'une pièce (jeu du pile ou face), réussite ou échec d'un lancer de ballon dans un panier,...

— **Loi géomètrique** de paramètre $p \in [0,1]$, notée $\mathcal{G}(p)$. Une variable aléatoire $X \sim \mathcal{G}(p)$, si $X \in \mathbb{N}^*$ et, pour tout $k \in \mathbb{N}^*$,

$$\mathbb{P}(X=k) = p(1-p)^{k-1}.$$

Exemple : si on lance un ballon de basket jusqu'à réussir son lancer, $X \sim \mathcal{G}(p)$ modélise le nombre de lancers que l'on doit réaliser (en notant $p$ la probabilité de réussite à chaque lancer).

— **Loi Binomiale** de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$, notée $\mathcal{B}(n,p)$. Une variable aléatoire $X \sim \mathcal{B}(n,p)$, si $X \in \{0,1,\dots,n\}$ et, pour tout $k \in \{0,\dots,n\}$,

$$\mathbb{P}(X=k) = C_n^k p^k (1-p)^{n-k} \text{ avec } C_n^k = \frac{n!}{k!(n-k)!}.$$

Exemple : si on lance $n$ fois et de manière indépendante un ballon de basket, $X \sim \mathcal{B}(n,p)$ modélise le nombre de paniers réussi où $p$ est la probabilité de réussite à chaque lancer.

— **Loi uniforme** sur un ensemble fini $I$. Une variable aléatoire $X \sim \mathcal{U}(I)$, si $X \in I$ et, pour tout $i \in I$,

$$\mathbb{P}(X=i) = \frac{1}{\mathrm{Card}(I)}.$$

Exemple : lancer d'un dé (loi uniforme sur $\{1,\dots,6\}$), pile ou face (loi uniforme sur $\{P,F\}$),...

— **Loi de Poisson** de paramètre $\lambda > 0$. Une variable aléatoire $X \sim \mathcal{P}(p)$, si $X \in \mathbb{N}$ et, pour tout $k \in \mathbb{N}$,

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

Modélise le nombre d'occurrences d'un évènement de moyenne connue durant un laps de temps donné Exemple : si, en moyenne 3 bateaux arrivent par heure dans un port, le nombre de bateaux arrivant dans un intervalle de temps d'1h donné peut-être modélisé par une loi de Poisson de paramètre 3.

### 1.2.2 Lois continues usuelles

Contrairement aux lois discrètes, les lois continues ne sont pas caractérisées par la valeur de $\mathbb{P}(X = k)$ qui est toujours nulle mais par leur densité.

**Définition 6** — *Une fonction $f : \mathbb{R} \to \mathbb{R}$ est une* **densité de probabilité** *si elle est positive et vérifie*

$$\int_{\mathbb{R}} f(t)dt = 1.$$

— *Une variable aléatoire $X$ est dite* **à densité** *s'il existe une densité de probabilité $f_X$ telle que, pour tous réels $a < b$*

$$\mathbb{P}(X \in [a, b]) = \int_a^b f_X(t)dt.$$

— **Loi normale ou loi gaussienne** de paramètres $\mu \in \mathbb{R}$ et $\sigma > 0$. Loi de densité

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

Exemple : poids et tailles des filles ou des garçons à un âge donné, accroissement du logarithme du prix (formule de Black et Scholes),...

— **Loi uniforme** sur un intervalle $[a, b]$. Loi de densité

$$f_X(t) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(t),$$

où, pour un ensemble $A$, $\mathbf{1}_A(t) = 1$ si $t \in A$, 0 si $t \notin A$.

— **Loi exponentielle** de paramètre $\lambda > 0$. Loi de densité

$$f_X(t) = \lambda e^{-\lambda t} \mathbf{1}_{]0,+\infty[}(t).$$

Exemple : durée de vie d'un atome radioactif/d'une machine, temps avant l'arrivée d'un client dans une file d'attente,...

— **Loi de Cauchy**. Loi de densité

$$f_X(t) = \frac{1}{\pi} \frac{1}{1+t^2}.$$

Il existe des variables aléatoires qui ne sont pas discrètes et qui n'admettent pas de densité. Par exemple soit $X \sim \mathcal{E}(\lambda)$ et $\varepsilon \sim \mathcal{U}(\{0, 1\})$. La variable

$$Y = \begin{cases} 0 & \text{si } \varepsilon\text{=}0 \\ \text{X} & \text{si } \varepsilon\text{=}1. \end{cases}$$

n'est pas discrète (car $X(\Omega) = [0, +\infty[$ est un ensemble non dénombrable) mais n'admet pas de densité. En effet, sinon

$$\mathbb{P}(Y \in [0, +\infty[) = \int_0^{+\infty} f_Y(t)dt = \mathbb{P}(X \in ]0, +\infty[),$$

ce qui n'est pas le cas car

$$\mathbb{P}(X \in [0, +\infty[) = 1$$

par définition de $X$ et

$$\mathbb{P}(X \in ]0, +\infty[) = \mathbb{P}(X > 0) = 1 - \mathbb{P}(X = 0) = 1 - \mathbb{P}(\varepsilon = 1) = 1/2.$$

# Notion d'espérance

**Définition 7** — *Soit $X$ une variable discrète. Notons $X(\Omega) = \{X(\omega), \omega \in \Omega\}$. Nous disons que la variable $X$* **admet une espérance** *si la série*

$$\sum_{x \in X(\Omega)} |x|\mathbb{P}(X = x)$$

*converge. Si la variable $X$ admet une espérance, nous appelons* **espérance mathématique de** $X$ *ou plus simplement* **espérance** *la quantité*

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

— *Soit $X$ une variable aléatoire réelle continue de densité $f_X$. Nous disons que la variable $X$* **admet une espérance** *si l'intégrale*

$$\int_{\mathbb{R}} |t| f_X(t)dt$$

*converge. Si la variable $X$ admet une espérance, nous appelons* **espérance mathématique de** $X$ *ou plus simplement* **espérance** *la quantité*

$$\mathbb{E}[X] = \int_{\mathbb{R}} t f_X(t)dt.$$

Formule d'intégration pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$ continue ou continue par morceaux

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

**Définition 8** — *Soit $k \in \mathbb{N}^*$. Nous disons qu'une variable $X$* **admet un moment d'ordre** $k$ *si la variable $|X|^k$ admet une espérance. Le* **moment d'ordre** $k$*, s'il existe, est la quantité $\mathbb{E}[X^k]$.*

— *Si $X$ admet un moment d'ordre 2, alors la quantité suivante, appelée* **variance**, *est bien définie*

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

*Nous définissons également l'***écart-type**

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}.$$

Exemples :

— La loi gaussienne $X \sim \mathcal{N}(\mu, \sigma^2)$ admet des moments de tous ordres. Nous avons en particulier $\mathbb{E}[X] = \mu$, $\mathbb{E}[X^2] = \sigma^2 + \mu^2$, $\mathbb{E}[X^3] = 2\mu\sigma^3 + \mu^3$, $\mathbb{E}[X^4] = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4$,...

— La loi exponentielle $X \sim \mathcal{E}(\lambda)$ admet des moments de tous ordres

$$\mathbb{E}[X^k] = \frac{k!}{\lambda^k}.$$

— La loi de Cauchy n'admet pas d'espérance, ni de moment d'ordre supérieur.

# Fonction de répartition

**Définition 9** *On appelle* **fonction de répartition** *d'une variable aléatoire $X$, la fonction*

$$F_X(t) = \mathbb{P}(X \leq t).$$

La fonction de répartition d'une variable aléatoire $X$ (discrète ou continue) vérifie toujours les propriétés suivantes :

— $F_X$ est croissante.

— $0 \leq F_X(t) \leq 1$, pour tout $t \in \mathbb{R}$.

— $\lim_{x \to -\infty} F_X(t) = 0$.

— $\lim_{x \to +\infty} F_X(t) = 1$.

La fonction de répartition d'une variable aléatoire continue est continue sur $\mathbb{R}$, dérivable sauf en un nombre fini de points et sa dérivée est égale à la densité en tout point où elle est dérivable. Celle d'une variable discrète est constante par morceaux.

La fonction de répartition de $X$ permet de calculer la probabilité que la variable $X$ soit dans un intervalle $]a, b]$. En effet

$$\mathbb{P}(X \in ]a, b]) = \mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F_X(b) - F_X(a).$$

Elle **caractérise** la loi de $X$ ce qui veut dire notamment que si deux variables aléatoires $X$ et $Y$ sont telles que

$$F_X(t) = F_Y(t), \forall t \in \mathbb{R},$$

alors leurs densités sont égales (si elles sont continues).

**Définition 10** *Soit $\alpha \in ]0, 1[$, le* **quantile d'ordre** *$\alpha$ de $X$, noté $q_\alpha$ ou $F_X^{-1}(\alpha)$ est la plus petite valeur de $X(\Omega)$ associée à une probabilité cumulée supérieure ou égale à $\alpha$ c'est-à-dire*

$$q_\alpha = \inf_{x \in X(\Omega)} \{F_X(x) \geq \alpha\}.$$

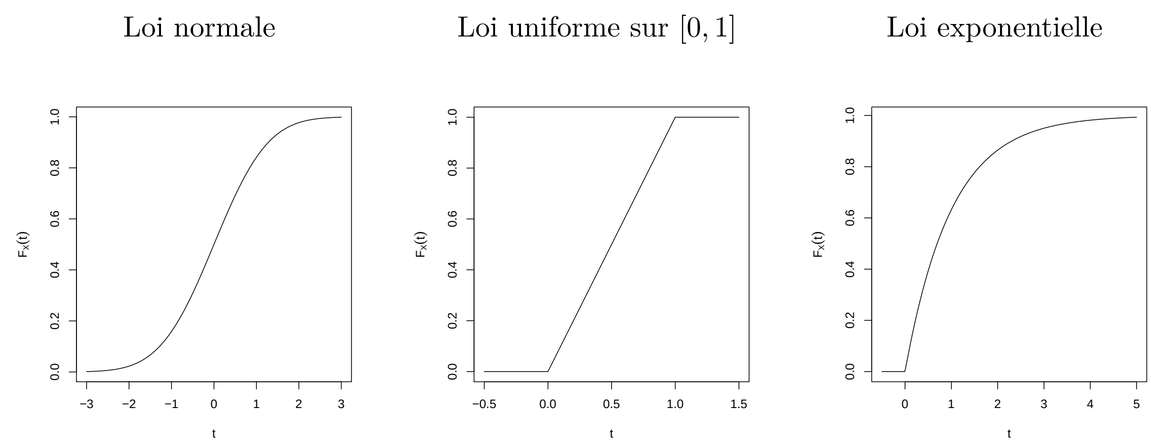


*Fig.2 Fonctions de répartition de quelques lois continues usuelles.*

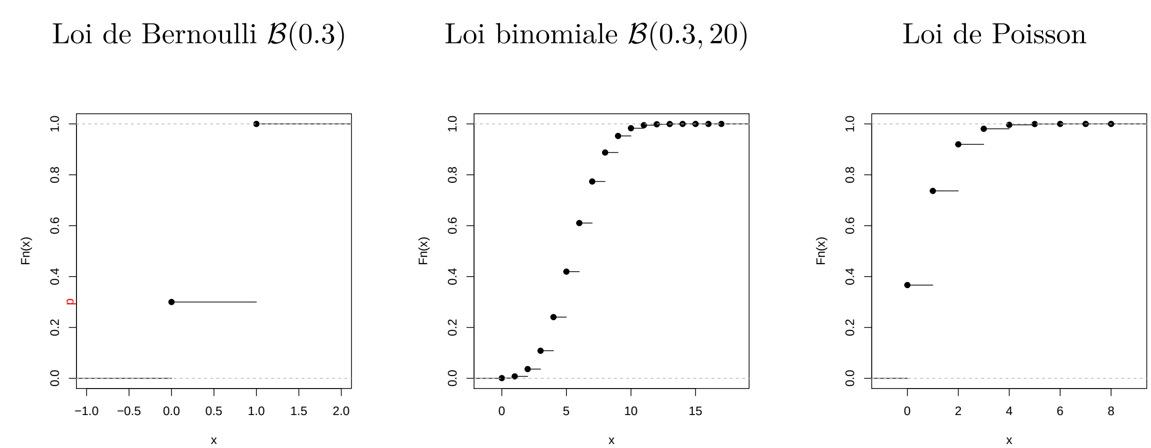


*Fig.3 Fonctions de répartition de quelques lois discrètes usuelles.*

*On appelle* **médiane** *de $X$ et on note* $\mathrm{med}(X) = q_{1/2}$. *Les* **quartiles** *sont les quantiles d'ordre* $\alpha = 0.25$ *et* $\alpha = 0.75$. *Les* **déciles** *les quantiles d'ordre* $\alpha = j/10$ *pour* $j = 1, \ldots, 9$. *Les* **centiles** *ou* **percentiles** *les quantiles d'ordre* $\alpha = j/100$ *pour* $j = 1, \ldots, 99$.

Si $X$ admet une densité, la fonction de répartition étant monotone (croissante) et continue sur $\mathbb{R}$, elle réalise une bijection de $\mathbb{R}$ sur $]0,1[$. Elle admet donc un inverse $F_X^{-1}$. Un cas particulier important est celui de la loi normale centrée (c'est-à-dire de moyenne nulle) et réduite (c'est-à-dire de variance égale à 1). Si $X \sim \mathcal{N}(0,1)$,

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

L'inverse $F_X^{-1}$ existe mais n'est pas explicite, cependant il peut-être approché.

# Indépendance

**Définition 11** — *Deux évènements $A$ et $B$ de $\mathcal{A}$ sont dits* **indépendants** *si*

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

- *Plus généralement, $n$ évènements $A_1, \ldots, A_n$ sont dits* **mutuellement indépendants** *si*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

- *Une suite $(A_n)_{n\in\mathbb{N}}$ d'évènements est une suite d'évènement* **mutuellement indépendants** *si, pour tout sous-ensemble* <u>*fini*</u> *$I$ de $\mathbb{N}$, $\{A_i, i \in I\}$ sont mutuellements indépendants.*
- *Les variables aléatoires $X_1, \ldots, X_n$ sont dites* **mutuellement indépendantes** *si, pour toutes fonctions continues bornées $h_1, \ldots, h_n : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}\left[\prod_{i=1}^{n} h_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}[h_i(X_i)].$$

- *Une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires est une suite de variables aléatoires* **mutuellement indépendantes** *si, pour tout sous-ensemble* <u>*fini*</u> *$I$ de $\mathbb{N}$, $\{X_i, i \in I\}$ sont des variables aléatoires mutuellement indépendantes.*

Pour un couple $(X,Y)$ de variables discrètes, $X$ et $Y$ sont indépendantes (on notera $X \coprod Y$) si et seulement si, pour tout $x \in X(\Omega), y \in Y(\Omega)$,

$$\mathbb{P}(\{X = x\} \cap \{Y = y\}) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$$

## Convergence de variables aléatoires

Si nous lançons $n$ fois un dé non pipé avec $n$ grand, nous nous attendons à "faire 6" environ une fois sur 6. Autrement dit, soit $Y_i$ le résultat du lancer du $i$-ème lancer, le nombre de fois où l'évènement $\{Y_i = 6\}$ est réalisé est à peu près égal à $n/6$ ou autrement dit

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{Y_i=6\}} \approx \frac{1}{6}.$$

Si nous notons $X_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{Y_i=6\}}$. La variable aléatoire $X_n$ va donc "converger" vers la constante $1/6$ lorsque $n$ tend vers l'infini. Mathématiquement, il existe plusieurs façon de donner sens à cette notion de convergence pour les variables aléatoires.

### 1.6.1 Notions de convergence

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires.

— Soit $p \geq 1$. Nous disons que $X_n$ **converge dans** $\mathbb{L}^p$ vers une variable aléatoire $X$ et nous notons

$$X_n \xrightarrow[n\to+\infty]{\mathbb{L}^p} X$$

si

$$\mathbb{E}\left[|X_n - X|^p\right] \underset{n\to+\infty}{\longrightarrow} 0.$$

— Nous disons que $X_n$ **converge p.s.** vers une variable aléatoire $X$ et nous notons

$$X_n \underset{n\to+\infty}{\longrightarrow} X \text{ p.s.}$$

si

$$\mathbb{P}\left(\left\{\omega \in \Omega, \lim_{n\to+\infty} X_n(\omega) = X(\omega)\right\}\right) = 1.$$

— Nous disons que $X_n$ **converge en probabilité** vers une variable aléatoire $X$ et nous notons

$$X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X$$

si, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(|X_n - X| > \varepsilon\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

— Nous disons que $X_n$ **converge en loi** vers une variable aléatoire $X$ et nous notons

$$X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$$

si, pour toute fonction $\varphi : \mathbb{R} \to \mathbb{R}$ continue et bornée,

$$\mathbb{E}\left[\varphi(X_n)\right] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\left[\varphi(X)\right].$$

On appelle **loi limite** de la suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$, la loi de la variable aléatoire $X$.

Implications :

$$X_n \xrightarrow[n\to+\infty]{\mathbb{L}^p} X \quad \Rightarrow \quad X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X \quad \Rightarrow \quad X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$$

$$X_n \xrightarrow[n\to+\infty]{} X \text{ p.s.} \quad \Rightarrow \quad X_n \xrightarrow[n\to+\infty]{\mathbb{P}} X \quad \Rightarrow \quad X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X.$$

**Propriété 1** *Si $\varphi$ est une fonction continue en tous point d'un ensemble $C$ tel que $\mathbb{P}(X \in C) = 1$ alors si $X_n$ converge vers $X$ (en probabilité, dans $\mathbb{L}^p$, p.s. ou en loi) alors $\varphi(X_n)$ converge vers $\varphi(X)$ pour le même mode de convergence.*

Exemple : si $X_n$ converge vers $X$ en loi, $\varphi(X_n)$ converge vers $\varphi(X)$ en loi (mais on ne sait pas si elle converge aussi vers $\varphi(X)$ p.s. ou en probabilité).

**Proposition 1** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires. Les propositions suivantes sont équivalentes :*

1. $$X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$$

2. *En tout point $t \in \mathbb{R}$ où la fonction de répartition $F_X$ est continue*

$$F_{X_n}(t) \underset{n\to+\infty}{\longrightarrow} F_X(t).$$

### 1.6.2 Lois des grands nombres

**Théorème 1 (Loi faible des grands nombres)** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires mutuellement indépendantes telles que*

$$\begin{array}{lcl} \mathbb{E}[X_i] & = & m \\ Var(X_i) & = & \sigma^2 \end{array} \quad \textit{pour tout } i = 1, \ldots, n, \ldots$$

*Nous notons*

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

*la* **moyenne empirique** *de $\{X_1, \ldots, X_n\}$. Nous avons le résultat suivant*

$$\bar{X}_n \xrightarrow[n\to+\infty]{\mathbb{P}} m.$$

**Théorème 2 (Loi forte des grands nombres)** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires mutuellement indépendantes et de mêmes lois (on dit aussi* **indépendantes et identiquement distribuées** *ou i.i.d.) admettant une espérance alors*

$$\bar{X}_n \xrightarrow[n\to+\infty]{} m \textit{ p.s.}$$

### 1.6.3 Théorème Central Limite

**Théorème 3** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires i.i.d. admettant une espérance $\mathbb{E}[X_1] = m$ et une variance $Var(X_1) = \sigma^2$ alors*

$$\sqrt{n}(\bar{X}_n - m) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

*ou, de manière équivalente,*

$$\sqrt{n}\frac{\bar{X}_n - m}{\sigma} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

### 1.6.4 Théorème de Slutsky et méthode delta

La convergence en loi est un type de convergence particulier. En effet, pour la convergence p.s., dans $\mathbb{L}^p$ et en probabilité, si $X_n$ et $Y_n$ convergent vers respectivement $X$ et $Y$, alors $X_n + Y_n$ converge vers $X + Y$, $X_nY_n$ vers $XY$,... **Ce n'est pas le cas pour la convergence en loi !** (cf exercice 9). Le théorème de Slutsky, très utile en statistique, nous donne cependant un cas particulier où cette propriété est vraie.

**Théorème 4 (Théorème de Slutsky)** *Soient $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites telles que*

$$X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X \text{ et } Y_n \xrightarrow[n\to+\infty]{\mathbb{P}} c,$$

*où $c$ est une constante et $c \neq 0$. Alors $(X_n, Y_n) \xrightarrow[n\to+\infty]{\mathcal{L}} (X, c)$ ce qui implique en particulier que*

$$\begin{aligned} X_n + Y_n &\xrightarrow[n\to+\infty]{\mathcal{L}} X + c \\ X_nY_n &\xrightarrow[n\to+\infty]{\mathcal{L}} cX \\ \frac{X_n}{Y_n} &\xrightarrow[n\to+\infty]{\mathcal{L}} \frac{X}{c}. \end{aligned}$$

**Théorème 5 (Méthode delta)** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires telle que*

$$\sqrt{n}(X_n - \mu) \xrightarrow[n\to+\infty]{\mathcal{L}} Z.$$

*Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$ telle que $g'(\mu) \neq 0$, alors*

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow[n\to+\infty]{\mathcal{L}} g'(\mu)Z.$$

# Vecteurs gaussiens

C'est la version multidimensionnelle de la loi normale.

**Définition 12** *Un vecteur aléatoire* $\mathbf{X} = (X_1, \ldots, X_d)$ *est dit* **gaussien** *si, pour tout* $(\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$*,*

$$\alpha_1 X_1 + \ldots + \alpha_d X_d$$

*est une variable aléatoire gaussienne.*

*Un vecteur gaussien est caractérisé par sa moyenne*

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d) = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d]) = \mathbb{E}[\mathbf{X}]$$

*et par sa matrice de variance-covariance*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \ldots & \mathrm{Cov}(X_1, X_d) \\ \mathrm{Cov}(X_1, X_2) & \ddots & & \mathrm{Cov}(X_2, X_d) \\ \vdots & \ddots & \ddots & \vdots \\ \mathrm{Cov}(X_1, X_d) & \mathrm{Cov}(X_2, X_d) & \ldots & \mathrm{Var}(X_d) \end{pmatrix}.$$

**Propriété 2** *Si* $\mathbf{X} = (X_1, \ldots, X_d)$ *est un vecteur gaussien alors, pour tout* $i \neq j$*,* $X_i$ *et* $X_j$ *sont indépendants si et seulement si ils sont décorrelés (c'est-à-dire que* $\mathrm{Cov}(X_i, X_j) = 0$*).*

Cette propriété est spécifique aux vecteurs gaussiens. En général on a juste, indépendant $\Rightarrow$ décorrélé mais pas l'inverse.

### 1.7.1 Exercices

**Exercice 1** *Soit* $\mathcal{A}$ *une* $\sigma$*-algèbre. À partir de la définition d'une* $\sigma$*-algèbre, vérifier que* $\emptyset \in \mathcal{A}$ *et que, pour toute suite* $(A_n)_{n\in\mathbb{N}}$ *d'éléments de* $\mathcal{A}$*,* $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}$*.*

**Exercice 2** *Soit* $X$ *une variable aléatoire à valeurs dans* $\mathbb{N}^*$ *telle qu'il existe* $C > 0$ *et* $\alpha > 0$ *telle que*

$$\mathbb{P}(X = k) = Ck^{-\alpha}.$$

1. *Quelles sont les valeurs possibles pour* $C$ *et* $\alpha$ *?*
2. *Pour quelles valeurs de* $\alpha$ *la variable* $X$ *admet-elle une espérance ? Une variance ?*

**Exercice 3** *Pour chacune des lois usuelles décrites en pp. 5 et 6. Déterminer si elles admettent un espérance ? Un moment d'ordre 2 ? Si oui calculer leurs espérance et, le cas échéant, leur variance.*

**Exercice 4**

1. *Calculer la fonction de répartition :*
   (a) *de la loi de Bernoulli,*
   (b) *de la loi géométrique,*
   (c) *de la loi uniforme sur* $[0, 1]$*,*
   (d) *de la loi exponentielle de paramètre* $\lambda$*,*
   (e) *de la loi de Cauchy.*
2. *En déduire, pour tout* $\alpha \in ]0, 1[$*, le quantile d'ordre* $\alpha$ *associé à chacune des distributions.*

**Exercice 5**

$$f_X(x) = C\exp(-|x|), \textit{ pour tout } x \in \mathbb{R}.$$

1. *Pour quelle valeur de $C$ la fonction $f_X$ est-elle une densité de probabilité ?*
2. *Déterminer la fonction de répartition de la variable $X$ de densité $f_X$ ?*
3. *Calculer l'espérance de la variable $X$.*

**Exercice 6** *Soient $X$ et $Y$ deux variables aléatoires indépendantes et de même loi géomètrique de paramètre $p \in ]0,1[$. On pose*

$$Z = \min\{X, Y\}.$$

1. *Calculer $\mathbb{P}(Z \geq k)$ pour tout entier naturel $k$.*
2. *En déduire la valeur de $\mathbb{P}(Z = k)$ pour tout entier naturel $k$.*
3. *Quelle est la loi de $Z$ ?*

**Exercice 7** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires discrètes telles que*

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2n^2},$$

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2}.$$

1. *Déterminer la limite en probabilité de la suite $(X_n)_{n\in\mathbb{N}}$.*
2. *La suite $(X_n)_{n\in\mathbb{N}}$ converge-t-elle dans $\mathbb{L}^2$ vers cette même limite.*

**Exercice 8 (Application du théorème de Slutsky)** *Supposons que*

$$X_n \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(m, \sigma^2) \textit{ et } Y_n \underset{n\to+\infty}{\xrightarrow{\mathbb{P}}} 2.$$

*Quelle est la loi limite de $X_n + Y_n$, $X_nY_n$ et $X_n/Y_n$ ?*

**Exercice 9** *Soit $X \sim \mathcal{N}(0,1)$. Nous posons $X_n = X$ et $Y_n = -X$ pour tout $n \in \mathbb{N}^*$.*

1. *Montrer que $X_n \longrightarrow_{n\to+\infty} X$ p.s. et $Y_n \longrightarrow_{n\to+\infty} -X$ p.s.*
2. *En déduire que $Y_n \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} X$ et $X_n \underset{n\to+\infty}{\xrightarrow{\mathbb{P}}} X$.*
3. *Quelle est la loi limite de $X_n + Y_n$ ? Ce résultat est-il en contradiction avec le théorème de Slutsky ?*

**Exercice 10** *Soient $X$ et $Y$ deux variables aléatoires indépendantes et de loi $\mathcal{N}(0,\sigma^2)$ et*

$$Z = \sqrt{X^2 + Y^2}.$$

*On admet que $Z$ suit une loi dite* de Rayleigh *c'est-à-dire qu'elle admet comme densité*

$$f_Z(z) = \frac{z}{\sigma^2} e^{-z^2/(2\sigma^2)} \mathbf{1}_{z\geq 0}.$$

1. *Quelle est la valeur de $\mathbb{E}[Z^2]$ ?*

2. *Soit* $(Z_n)_{n\in\mathbb{N}}$ *une suite i.i.d. de variables aléatoires de même loi que* $Z$. *On considère la variable aléatoire*
$$\widehat{\sigma}^2 = \frac{1}{2n}\sum_{i=1}^{n} Z_i^2.$$
*Montrer que cette suite de variables aléatoires converge en probabilité vers* $\sigma^2$.
3. *Vérifier que la suite* $(Z_n^2)_{n\in\mathbb{N}}$ *est également i.i.d.*
4. *Montrer que* $\widehat{\sigma}^2$ *est asymptotiquement normalement distribuée (c'est-à-dire que sa loi limite est gaussienne).*

**Exercice 11** *Un fournisseur d'accès Internet met en place un point local d'accès, qui dessert 5000 abonnés. À un instant donné, chaque abonné a une probabilité égale à 20% d'être connecté. Les comportements des abonnés sont supposés indépendants les uns des autres.*

1. *On note* $X$ *la variable aléatoire égale au nombre d'abonnés connectés à un instant* $t$. *Quelle est la loi de* $X$ *? Quelle est son espérance ? Son écart-type ?*
2. *On pose*
$$Y = \frac{X - 1000}{\sqrt{800}}.$$
*Justifier précisément que l'on peut approcher la loi de* $X$ *par une loi normale.*
3. *Le fournisseur d'accès souhaite savoir combien de connexions simultanées le point d'accès doit pouvoir gérer pour que sa probabilité d'être saturé à un instant donné soit inférieure à 2,5%. En utilisant l'approximation précédente, proposer une valeur approchée de ce nombre de connexions.*

**Exercice 12** *Il arrive assez souvent que le nombre de réservations pour une liaison aérienne soit supérieur au nombre de passagers se présentant effectivement le jour du vol. Pour compenser ce phénomène, une compagnie aérienne exploitant un avion de 300 places décide de faire de la surréservation (on utilise aussi souvent le mot anglais* surbooking*) en prenant pour chaque vol un nombre* $n > 300$ *réservations. S'il se présente plus de 300 passagers à l'embarquement, les personnes ne pouvant pas prendre le vol sont dédommagés financièrement.*

1. *On considère que les passagers sont mutuellement indépendants et que la probabilité de désistement de chacun d'eux est égale à 10%. On note* $S_n$ *le nombre (aléatoire) de passagers se présentant à l'embarquement du vol sur les* $n$ *ayant réservé leur vol (*$n$ *n'est pas aléatoire). Donner la loi de* $S_n$*, son espérance et sa variance ?*
2. *Le directeur de la compagnie voudrait s'assurer que*
$$\mathbb{P}(S_n \leq 300) \geq 0.99.$$
*En utilisant le théorème central limite, proposer une solution approchée à ce problème.*

# Estimation et intervalles de confiance

## Problème d'inférence statistique

Nous observons un phénomène plusieurs fois, et nous notons le résultat $x_1, \ldots, x_n$ (par exemple la taille de $n$ individus pris au hasard). Pour pouvoir extraire de l'information de ces données, l'approche usuelle en statistique consiste à modéliser ces observations comme des réalisations d'une certaine variable aléatoire. Dans le cadre de ce cours, nous supposerons que les observations sont une suite i.i.d. (indépendantes et identiquement distribuées c'est-à-dire de mêmes lois) $X_1, \ldots, X_n$ suivant la même loi qu'une variable aléatoire $X$. Nous appelons également $X_1, \ldots, X_n$ un $n$-échantillon ou un échantillon de taille $n$.

Encore plus généralement, supposons que la loi de la variable $X$ dépende d'un paramètre $\theta$ que nous cherchons à approcher au mieux à partir des données.

**Définition 13** *Un* **estimateur** *du paramètre $\theta$ est une fonction des variables aléatoires de l'échantillon. Cet estimateur, noté $\widehat{\theta}$, est défini par*

$$\widehat{\theta} = g(X_1, \ldots, X_n).$$

Exemples :

— Supposons que $X \sim \mathcal{B}(p)$ et que nous souhaitons estimer le paramètre $p$. La quantité

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + \ldots + X_n}{n} = \frac{\mathrm{Card}\{i, X_i = 1\}}{n}.$$

est un estimateur de $p$ car il existe une fonction $g(x_1, \ldots, x_n) = (x_1 + \ldots + x_n)/n$ telle que $\widehat{p} = g(X_1, \ldots, X_n)$. C'est l'estimateur usuel de $p$.

— Plus généralement, si $X$ est intégrable, l'estimateur usuel de l'espérance $\mathbb{E}[X]$ est la moyenne

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

— Lorsque l'on cherche à estimer la variance de $X$, il existe plusieurs estimateurs usuels, notamment,

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

ou

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

Par contre,
$$\frac{1}{n}\sum_{i=1}(X_i - \mathbb{E}[X])^2$$
n'est pas un estimateur car il dépend de $\mathbb{E}[X]$ inconnu (donc il n'existe pas de fonction déterministe $g$ telle que $g(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^n (X_i - \mathbb{E}[X])^2$).

**Propriété 3** *Un estimateur est une variable aléatoire. Il admet donc une loi. Le calcul de l'estimateur sur les données $(x_1, \dots, x_n)$ est une réalisation de la variable aléatoire $\widehat{\theta}$, on la note $\widehat{\theta}(x)(= g(x_1, \dots, x_n))$, c'est une* **estimation** *du paramètre $\theta$ (à partir des données).*

**Attention** à bien différencier un estimateur $\widehat{\theta}$ qui est une variable aléatoire de sa réalisation sur les données $\widehat{\theta}(x)$ qui est une constante.

Exemple : soient $X_1, \dots, X_n$ des variables aléatoires i.i.d. selon la loi normale $\mathcal{N}(\mu, \sigma^2)$. La moyenne empirique
$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i,$$
est un estimateur du paramètre $\mu$. La somme de variables gaussiennes indépendantes étant aussi gaussienne, elle suit une loi normale de moyenne
$$\mathbb{E}[\widehat{\mu}] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu,$$
et de variance
$$\mathrm{Var}(\widehat{\mu}) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$
En conclusion
$$\widehat{\mu} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$
Si l'on observe une réalisation $(x_1, x_2, x_3) = (2, 2, 3)$ $(n = 3)$ des variables aléatoires $X_1, X_2, X_3$ une estimation du paramètre $\mu$ est donnée par $\widehat{\mu}(x) = \frac{2+2+3}{3} = 7/3$.

### 2.1.1 Biais

**Définition 14** *Le* **biais** *d'un estimateur $\widehat{\theta}$ du paramètre $\theta$ est la quantité*
$$\mathrm{Biais}(\widehat{\theta}) = \mathbb{E}[\widehat{\theta}] - \theta.$$
*Un estimateur est dit* **sans biais** *si son biais est égal à 0 et* **biaisé** *sinon. Il est dit* **asymptotiquement sans biais** *si*
$$\mathrm{Biais}(\widehat{\theta}) \longrightarrow_{n\to+\infty} 0.$$

Reprenons l'exemple précédent, l'estimateur $\widehat{\mu}$ est un estimateur sans biais de $\mu$. En revanche l'estimateur
$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \widehat{\mu})^2,$$

est un estimateur biaisé. En effet

$$\mathbb{E}[\widehat{\sigma}_n^2] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[(X_i - \widehat{\mu})^2].$$

Or

$$\begin{aligned}
\mathbb{E}[(X_i - \widehat{\mu})^2] &= \mathbb{E}[X_i^2] - 2\mathbb{E}[X_i\widehat{\mu}] + \mathbb{E}[\widehat{\mu}^2] = \mathbb{E}[X_i^2] - \frac{2}{n}\sum_{j=1}^n \mathbb{E}[X_iX_j] + \mathrm{Var}(\widehat{\mu}) + \mathbb{E}[\widehat{\mu}]^2 \\
&= \mathbb{E}[X_i^2] - \frac{2}{n}\left(\mathbb{E}[X_i^2] + \sum_{j\neq i}\mathbb{E}[X_i]\mathbb{E}[X_j]\right) + \frac{\sigma^2}{n} + \mu^2 \\
&= \left(1 - \frac{2}{n}\right)\mathbb{E}[X_i^2] - \frac{2}{n}\sum_{j\neq i}\mu \times \mu + \frac{\sigma^2}{n} + \mu^2 \\
&= \left(1 - \frac{2}{n}\right)\left(\mathrm{Var}(X_i) + \mathbb{E}[X_i]^2\right) - \frac{2}{n}\times(n-1)\times\mu^2 + \frac{\sigma^2}{n} + \mu^2 \\
&= \left(1 - \frac{2}{n}\right)\left(\sigma^2 + \mu^2\right) - 2\frac{n-1}{n}\mu^2 + \frac{\sigma^2}{n} + \mu^2 \\
&= \left(1 - \frac{2}{n} + \frac{1}{n}\right)\sigma^2 + \underbrace{\left(1 - \frac{2}{n} - 2\frac{n-1}{n} + 1\right)}_{=0}\mu^2 \\
&= \frac{n-1}{n}\sigma^2.
\end{aligned}$$

Le biais de $\widehat{\sigma}_n^2$ est donc égal à

$$\mathrm{Biais}(\widehat{\sigma}_n^2) = \left(\frac{n-1}{n} - 1\right)\sigma^2 = -\frac{1}{n}\sigma^2.$$

$\widehat{\sigma}_n^2$ est donc asymptotiquement sans biais. L'estimateur $\widehat{S}_n^2$ est, en revanche, un estimateur sans biais de $\sigma^2$. C'est la raison pour laquelle on le préfère usuellement à $\sigma^2$.

### 2.1.2 Variance

Il existe en général plusieurs estimateurs sans biais d'une même quantité. En reprenant l'exemple de $n$ variables aléatoires $X_1, \ldots, X_n$ suivant la loi $\mathcal{N}(\mu, \sigma^2)$, la moyenne empirique $\widehat{\mu}$ est un estimateur sans biais mais aussi $X_1$, $(X_1 + X_2)/2$,... Dans ce cas-là, nous pouvons comparer la variances des estimateurs et prendre celui qui a la variance la plus faible (plus stable). Parmi les trois exemples donnés ci-dessous, nous avons

$$\mathrm{Var}(\widehat{\mu}) = \frac{\sigma^2}{n} \leq \mathrm{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{\sigma^2}{2} \leq \mathrm{Var}(X_1) = \sigma^2$$

c'est $\widehat{\mu}$ qui a la variance la plus faible.

### 2.1.3 Convergence

**Définition 15** *Un estimateur $\widehat{\theta}$ est dit*

— **fortement convergent** *si*

$$\widehat{\theta} \underset{n\to+\infty}{\longrightarrow} \theta \text{ p.s.}$$

— **faiblement convergent** *ou juste* **convergent** *si*

$$\widehat{\theta} \xrightarrow[n\to+\infty]{\mathbb{P}} \theta.$$

**Propriété 4** *Soit $\widehat{\theta}$ un estimateur tel que*

$$\lim_{n\to+\infty} \mathbb{E}[\widehat{\theta}] = \theta \text{ et } \lim_{n\to+\infty} \mathrm{Var}(\widehat{\theta}) = 0$$

*alors $\widehat{\theta}$ est convergent.*

Exemple : $\widehat{\mu}$ est faiblement convergent (par la loi faible des grand nombre), il est également fortement convergent (par la loi forte des grands nombres).

### 2.1.4 Loi asymptotique

**Définition 16** *Un estimateur $\widehat{\theta}$ est dit* **asymptotiquement normalement distribué** *s'il existe $\theta \in \mathbb{R}$ et $\sigma > 0$ tel que*

$$\frac{\sqrt{n}}{\sigma}(\widehat{\theta} - \theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

*Dans ce cas-là, on appelle* **espérance asymptotique** *la quantité $\theta$ et* **variance asymptotique** *la quantité $\sigma^2/n$ et* **loi asymptotique** *la loi $\mathcal{N}(\theta, \sigma^2/n)$.*

Exemple : le théorème central limite implique que $\widehat{\mu}$ est asymptotiquement normalement distribué avec espérance asymptotique $\mu$ et variance asymptotique $\sigma^2/n$.

# Intervalles de confiance

**Définition 17**

— *Un* **intervalle de confiance** *sur le paramètre $\theta$ au niveau de confiance $1-\alpha$ (ou au niveau de risque $\alpha$), avec $\alpha \in ]0, 1[$, est un encadrement du type :*

$$\mathbb{P}(A \leq \theta \leq B) \geq 1 - \alpha,$$

*où $A$ et $B$ sont des variables aléatoires, fonctions de variables de l'échantillon $X_1, \ldots, X_n$.*

— *Un* **intervalle de confiance asymptotique** *sur le paramètre $\theta$ au niveau de confiance $1-\alpha$ (ou au niveau de risque $\alpha$), avec $\alpha \in ]0, 1[$, est un encadrement tel que*

$$\lim_{n\to+\infty} \mathbb{P}(A \leq \theta \leq B) \geq 1 - \alpha,$$

*où $A$ et $B$ sont des variables aléatoires, fonctions de variables de l'échantillon $X_1, \ldots, X_n$.*

— *Une réalisation d'un intervalle de confiance est notée*

$$IC_{1-\alpha} = [a; b]$$

*où a (resp. b) est la réalisation de la variable A (resp. B), obtenues à partir de la réalisation de l'échantillon* $(x_1, \ldots, x_n)$.

### 2.2.1 Construction d'intervalles de confiance dans le cas gaussien (variance connue)

Nous observons $x_1, \ldots, x_n$ une réalisation d'un échantillon $X_1, \ldots, X_n$ suivant la loi normale de paramètres $\mu$ (inconnu) et $\sigma^2$ (connu). Cela peut-être par exemple des mesures réalisées avec une incertitude connue. Un estimateur sans biais de $\mu$ est

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

et nous avons vu précédemment que

$$\widehat{\mu} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

ou, de manière équivalente,

$$\frac{\sqrt{n}}{\sigma}(\widehat{\mu} - \mu) \sim \mathcal{N}(0, 1).$$

Nous connaissons donc, pour tous réels fixés $x < y$,

$$\mathbb{P}\left(x \leq \frac{\sqrt{n}}{\sigma}(\widehat{\mu} - \mu) \leq y\right) = \mathbb{P}(x \leq Z \leq y) = F_Z(y) - F_Z(x).$$

où $Z \sim \mathcal{N}(0, 1)$. Cela revient à écrire

$$\mathbb{P}\left(\widehat{\mu} - \frac{\sigma}{\sqrt{n}}x \geq \mu \geq \widehat{\mu} - \frac{\sigma}{\sqrt{n}}y\right) = F_Z(y) - F_Z(x).$$

Il suffit donc de trouver $x$ et $y$ tels que $F_Z(y) - F_Z(x) = 1 - \alpha$ pour trouver un intervalle de confiance de $\mu$ au niveau $1 - \alpha$. Une possibilité est de poser $y = q_{1-\alpha/2}$ et $x = q_{\alpha/2}$, les quantiles d'ordres $1 - \alpha/2$ et $\alpha/2$ de la loi de $Z$, en effet, par définition

$$F_Z(q_{1-\alpha/2}) = 1 - \alpha/2 \text{ et donc } F_Z(q_{1-\alpha/2}) - F_Z(q_{\alpha/2}) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha.$$

Nous remarquons d'autre part que, par symétrie de la loi normale, $q_{1-\alpha/2} = -q_{\alpha/2}$ et nous avons notre premier intervalle de confiance

$$\boxed{\mathbb{P}\left(\widehat{\mu} - \frac{\sigma}{\sqrt{n}}q_{1-\alpha/2} \leq \mu \leq \widehat{\mu} + \frac{\sigma}{\sqrt{n}}q_{1-\alpha/2}\right) = 1 - \alpha.}$$

Remarquons que si nous ne connaissons pas $\sigma$ à l'avance, l'encadrement n'est plus un intervalle de confiance.

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $x_8$ | $x_9$ | $x_{10}$ |
|---|---|---|---|---|---|---|---|---|---|
| 2.03 | 2.11 | 1.84 | 1.89 | 1.92 | 2.12 | 1.91 | 1.99 | 2.11 | 2 |

*Fig.1*

Application : nous mesurons la taille d'une pièce à l'issue d'une chaîne de fabrication. Les résultats sont donnés dans le tableau 1.

La connaissance du fonctionnement de la chaîne de production, nous permet de modéliser $x_1, \dots, x_{10}$ comme des réalisations $X_1, \dots, X_{10}$ d'une variable aléatoire de loi $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma = 0.1$. Nous calculons $\widehat{\mu}(x) = 1.992$. À partir de la formule précédente et en sachant que $q_{0.975}$ est à peu près égal à 1.96. La réalisation d'un intervalle de confiance de niveau de confiance 95% (donc de risque 5%) est

$$\boxed{IC_{95\%} = \left[\widehat{\mu}(x) - \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2} ; \widehat{\mu}(x) + \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}\right]}$$

ce qui fait dans notre cas

$$IC_{95\%} = [1.93, 2.05]$$

avec une précision de 0.01.

Nous remarquons encore une fois que cette approche n'est plus valable dès que $\sigma$ est inconnu, ou à fortiori, si nous ne connaissons pas la loi de $X_1, \dots, X_n$. Comment donc pouvons-nous procéder dans le cas général ?

### 2.2.2 Approche naïve pour la construction d'intervalle de confiance (cas général)

Si nous connaissons un estimateur sans biais de $\theta$ et que nous connaissons ou pouvons estimer ou majorer sa variance, l'inégalité de Bienaymé-Tchebychev nous fournit un cadre général de construction d'un intervalle de confiance pour estimer l'espérance de $X_1, \dots, X_n$ sans connaître sa loi. En effet, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|\widehat{\theta} - \theta| \geq \varepsilon) \leq \frac{\mathrm{Var}(\widehat{\theta})}{\varepsilon^2}.$$

Il suffit donc de choisir $\varepsilon \geq \sigma(\widehat{\theta})/\sqrt{\alpha}$ pour avoir un intervalle de confiance de risque inférieur à $\alpha$. Prenons par exemple la cas de l'estimation d'une proportion $p$. Nous avons $X_1, \dots, X_n \sim_{i.i.d.} \mathcal{B}(p)$,

$$\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

vérifie $n\widehat{p} = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n, p)$ et donc

$$\mathrm{Var}(\widehat{p}) = \frac{1}{n^2} \mathrm{Var}(n\widehat{p}) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \leq \frac{1}{4n}$$

(étudier les variations de la fonction $x \mapsto x(1-x)$ sur $[0, 1]$). Un intervalle de confiance de risque inférieur à $\alpha$ est donc donné par

$$\mathbb{P}\left(\widehat{p} - \sqrt{\frac{1}{4n\alpha}} \leq p \leq \widehat{p} + \sqrt{\frac{1}{4n\alpha}}\right) \geq 1 - \alpha.$$

Si $\alpha = 5\%$, la précision de l'intervalle (c'est-à-dire la longueur de l'intervalle divisée par 2) est $\sqrt{1/0.05n}$ cela fait environ 0.71 si $n = 10$, 0.22 si $n = 100$,... Sachant que $p$ est compris entre 0 et 1, cet intervalle n'est pas assez précis, ce qui motive une autre approche.

### 2.2.3 Construction d'intervalles de confiance asymptotiques

**Estimation d'une proportion** Nous observons donc $X_1, \ldots, X_n \sim_{i.i.d.} \mathcal{B}(p)$, avec $p \in ]0,1[$, et nous cherchons un intervalle de confiance pour le paramètre $p$. Nous partons de l'estimateur

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Comme $X_1(\Omega)$ est de cardinal fini (égal à 2), $X_1$ admet des moments de tous ordres (car $\sum_{x\in X(\Omega)} |x|^k \mathbb{P}(X = k)$ est toujours finie), donc nous pouvons appliquer le Théorème Central Limite (TCL) et obtenir

$$\frac{\sqrt{n}}{\sigma(X)}(\widehat{p} - p) = \frac{\sqrt{n}}{\sqrt{p(1-p)}}(\widehat{p} - p) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Cela implique que, pour tous $x < y$,

$$\lim_{n\to+\infty} \mathbb{P}\left(x \leq \frac{\sqrt{n}}{\sqrt{p(1-p)}}(\widehat{p} - p) \leq y\right) = \mathbb{P}(x \leq Z \leq y),$$

avec $Z \sim \mathcal{N}(0,1)$. En reprenant le raisonnement de la partie 2.2.1, nous pouvons choisir $x = -q_{1-\alpha/2}$, $y = q_{1-\alpha/2}$ et obtenir le résultat suivant

$$\lim_{n\to+\infty} \mathbb{P}\left(\widehat{p} - \frac{\sqrt{p(1-p)}}{\sqrt{n}}q_{1-\alpha/2} \leq p \leq \widehat{p} + \frac{\sqrt{p(1-p)}}{\sqrt{n}}q_{1-\alpha/2}\right) = 1 - \alpha.$$

Cependant, ce résultat ne donne pas un intervalle de confiance, les quantités $\widehat{p} - \frac{\sqrt{p(1-p)}}{\sqrt{n}}q_{1-\alpha/2}$ et $\widehat{p} + \frac{\sqrt{p(1-p)}}{\sqrt{n}}q_{1-\alpha/2}$ n'étant pas des variables aléatoires car elles dépendent du paramètre inconnu $p$ qui n'est pas une fonction des observations. Nous allons donc remplacer $p$ par son estimateur $\widehat{p}$.

— Par la loi faible des grands nombres, nous avons

$$\widehat{p} \xrightarrow[n\to+\infty]{\mathbb{P}} p.$$

— Comme la fonction $x \mapsto \sqrt{x(1-x)/(p(1-p))}$ est continue sur $[0,1]$, nous avons donc également que

$$\sqrt{\frac{\widehat{p}(1-\widehat{p})}{p(1-p)}} \xrightarrow[n\to+\infty]{\mathbb{P}} 1.$$

— Nous appliquons donc le théorème de Slutsky et nous obtenons

$$\frac{\frac{\sqrt{n}}{\sqrt{p(1-p)}}(\widehat{p} - p)}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{p(1-p)}}} = \frac{\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}(\widehat{p} - p) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

— Ce qui nous permet d'obtenir l'intervalle de confiance asymptotique suivant

$$\boxed{\lim_{n\to+\infty} \mathbb{P}\left(\widehat{p} - \frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}} q_{1-\alpha/2} \leq p \leq \widehat{p} + \frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}} q_{1-\alpha/2}\right) = 1-\alpha}$$

et sa réalisation sur les observations $x_1, \ldots, x_n$,

$$\boxed{IC_{1-\alpha}^{(n)} = \left[\widehat{p}(\mathbf{x}) - \frac{\sqrt{\widehat{p}(\mathbf{x})(1-\widehat{p}(\mathbf{x}))}}{\sqrt{n}} q_{1-\alpha/2}; \widehat{p}(\mathbf{x}) + \frac{\sqrt{\widehat{p}(\mathbf{x})(1-\widehat{p}(\mathbf{x}))}}{\sqrt{n}} q_{1-\alpha/2}\right].}$$

**Estimation d'une moyenne** Nous observons maintenant $X_1, \ldots, X_n$ i.i.d. mais nous ne connaissons plus sa loi. Nous supposons juste que $X_1$ admet un moment d'ordre 2 et nous notons $\sigma^2$ sa variance et $\mu$ son espérance, les deux paramètres étant inconnus. La procédure de construction d'intervalle de confiance pour $\mu$ est similaire à celle décrite dans le paragraphe précédent.

— Le Théorème Central Limite nous donne la convergence en loi suivante

$$\frac{\sqrt{n}}{\sigma}(\widehat{\mu} - \mu) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

— Le paramètre $\sigma$ étant inconnu, nous devons le remplacer par un estimateur pour construire un intervalle de confiance. Nous posons

$$S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{\mu})^2},$$

l'estimateur usuel de $\sigma$ (variance corrigée). Nous avons le résultat suivant (admis)

$$\frac{S_n}{\sigma} \xrightarrow[n\to+\infty]{\mathbb{P}} 1.$$

— Le théorème de Slutsky nous permet d'écrire ensuite

$$\frac{\sqrt{n}}{S_n}(\widehat{\mu} - \mu) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

— Et nous obtenons l'intervalle de confiance asymptotique suivant

$$\boxed{\lim_{n\to+\infty} \mathbb{P}\left(\widehat{\mu} - \frac{S_n}{\sqrt{n}} q_{1-\alpha/2} \leq \mu \leq \widehat{\mu} + \frac{S_n}{\sqrt{n}} q_{1-\alpha/2}\right) = 1-\alpha}$$

et sa réalisation sur les observations $x_1, \ldots, x_n$,

$$\boxed{IC_{1-\alpha}^{(n)} = \left[\widehat{\mu}(x) - \frac{S_n(x)}{\sqrt{n}} q_{1-\alpha/2}; \widehat{\mu}(x) + \frac{S_n(x)}{\sqrt{n}} q_{1-\alpha/2}\right].}$$

En reprenant l'exemple de l'estimation du tableau 1, nous avons

$$S_n = 0.100,$$

avec une précision de 0.001. L'intervalle asymptotique coïncide donc ici avec l'intervalle non-asymptotique précédent.

**Cas général** Cette approche générale permet la construction d'intervalles de confiance dès que la loi asymptotique est connue. La méthode delta permet par exemple ensuite de généraliser l'approche à toute fonction de la moyenne des observations par exemple.

# Exercices

**Exercice 13** *Parmi les quantités suivantes, lesquelles sont des estimateurs ?*

1. $\frac{1}{n}\sum_{i=1}^{n} X_i$ ;
2. $\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$ ;
3. $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])^2}$ ;
4. $\min\{X_1, \ldots, X_n\}$ ;
5. $\max\{x_1, \ldots, x_n\}$.

**Exercice 14** *Sur 12 000 individus d'une espèce, on a dénombré 13 albinos. Estimer la proportion d'albinos dans l'espèce. On comparera sur cet exemple l'intervalle de confiance non asymptotique donné par l'inégalité de Bienaymé-Tchebychev et l'intervalle asymptotique donné par le TCL.*

**Exercice 15** *Nous observons* $n = 100$ *réalisations d'une loi normale de moyenne* $\mu$ *et de variance* $\sigma^2$ *toutes deux inconnues. La moyenne empirique*

$$\widehat{\mu}(x) = 4.3$$

*et la variance empirique corrigée*

$$S_n^2(x) = 6.76.$$

*Donner un intervalle de confiance asymtotique au niveau de confiance* $\alpha = 5\%$ *? Au niveau* $\alpha = 10\%$ *? Comment varie la taille de l'intervalle de confiance en fonction de* $\alpha$ *? En fonction de* $n$ *?*
*Données :* $q_{0.975} = 1.96$, $q_{0.95} = 1.64$.

**Exercice 16** *Une compagnie aérienne a demandé des statistiques afin d'améliorer la sûreté au décollage et définir un poids limite de bagages. Pour l'estimation du poids des voyageurs et du poids des bagages, un échantillon est constitué de 300 passagers qui ont accepté d'être pesés : on a obtenu une moyenne* $\widehat{\mu}_p$ *de 68kg, avec un écart-type* $\widehat{S}_p$ *de 7 kg.*

1. *Définir un intervalle de confiance pour la moyenne du poids des passagers.*
2. *Par la suite, on modélisera le poids des passagers comme une variable aléatoire* $X$ *de moyenne* $\mu_p = 68$*kg, d'écart-type* $\sigma_p = 7$*kg. On procédera de même pour le poids des bagages que l'on modélisera comme une variable aléatoire* $Y$ *de moyenne 15 kg, d'écart-type 5 kg. Nous supposons de plus que* $X$ *et* $Y$ *sont indépendantes. La capacité de l'avion est de 300 passagers ; l'avion pèse, à vide, 250 tonnes. Le décollage est interdit si le poids total dépasse 276.2 tonnes. Calculer ou approcher la probabilité pour que le décollage soit interdit dans le cas où*

(a) *les variables aléatoires $X$ et $Y$ sont supposées gaussiennes.*

(b) *les variables aléatoires $X$ et $Y$ ne sont pas supposées gaussiennes.*

(c) *Que se passe-t'il si $X$ et $Y$ ne sont plus supposées indépendantes ?*

**Exercice 17 (Estimation et loi de Rayleigh (inspiré de HEC Lausanne, 2013))** *Soient deux variables aléatoires réelles $X_1$ et $X_2$ indépendantes et distribuées chacune selon une loi $\mathcal{N}(0, \sigma^2)$ avec $\sigma > 0$ inconnu. On considère la variable aléatoire*

$$Y = \sqrt{X_1^2 + X_2^2}.$$

*Nous observons $Y_1, \ldots, Y_n$ i.i.d. de même loi que $Y$ et nous étudions l'estimateur*

$$\widehat{\sigma}^2 = \frac{1}{2n} \sum_{i=1}^{n} Y_i^2$$

*de $\sigma^2$.*

1. *Quel est le biais de l'estimateur $\widehat{\sigma}^2$ ?*
2. *Montrer que l'estimateur $\widehat{\sigma}^2$ est convergent.*
3. *Montrer que l'estimateur $\widehat{\sigma}^2$ est asymptotiquement normal et donner sa variance asymptotique.*
4. *Donner un (ou plusieurs) intervalles de confiances asymptotiques du paramètre $\sigma$ ?*

*On pourra utiliser dans l'exercice le résultat suivant : pour une variable gaussienne $X \sim \mathcal{N}(0, \sigma^2)$, $\mathbb{E}[X^4] = 3\sigma^4$.*

**Exercice 18** *Soit $X$ une variable aléatoire telle que*

$$\mathbb{E}[X] = \mathbb{E}[X^2] = \theta$$

*avec $\theta \in ]0, 1[$ un paramètre inconnu. Nous observons $X_1, \ldots, X_n$ i.i.d. de même loi que $X$ et nous considérons deux estimateurs de $\theta$ :*

$$\widehat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \text{ et } \widehat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

1. *Montrer que les estimateurs $\widehat{\theta}_1$ et $\widehat{\theta}_2$ sont sans biais.*
2. *Montrer que $\widehat{\theta}_1$ est convergent.*
3. *Sous quelle condition $\widehat{\theta}_2$ est-il convergent ?*
4. *Sous quelle condition avons-nous*

$$\mathrm{Var}(\widehat{\theta}_1) \leq \mathrm{Var}(\widehat{\theta}_2)?$$

5. *En supposant que $X$ admet un moment d'ordre 4 donner les lois asymptotiques des estimateurs $\widehat{\theta}_1$ et $\widehat{\theta}_2$ ?*

# Construction d'estimateurs

Nous supposons que nous observons un échantillon $X_1, \ldots, X_n$ et que nous souhaitons estimer du paramètre $\theta$ de cette loi. Il existe une infinité de manière de définir un estimateur de $\theta$. Dans le cadre de ce cours, nous allons nous concentrer sur deux méthodes génériques de construction d'estimateurs : la méthode des moments et l'estimation par maximum de vraisemblance.

## Méthode des moments

Le principe général de la méthode des moments est le suivant : supposons qu'il existe un entier non nul $k$ et une fonction $h : \mathbb{R} \to \mathbb{R}$ connue telle que le paramètre $\theta$ que l'on souhaite estimer vérifie

$$\theta = h(\mathbb{E}[X^k])$$

L'estimateur des moments est l'estimateur

$$\widehat{\theta}^{Mom} = h\left(\frac{1}{n}\sum_{i=1}^n X_i^k\right).$$

Par la suite, nous noterons, pour simplifier :

$$\mu_k = \mathbb{E}[X^k] \text{ et } \widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^n X_i^k$$

ainsi que, dans la cas $k = 1$, $\mu = \mu_1$, $\widehat{\mu} = \widehat{\mu}_1$. Exemples :

1. Supposons $X_1, \ldots, X_n \sim_{i.i.d.} \mathcal{N}(\mu, \sigma^2)$, nous avons

$$\mu = \mathbb{E}[X_1] \text{ donc } \widehat{\mu}^{Mom} = \frac{1}{n}\sum_{i=1}^n X_i = \widehat{\mu}.$$

2. Supposons $X_1, \ldots, X_n \sim_{i.i.d.} \mathcal{E}(\lambda)$, nous avons

$$\lambda = \frac{1}{\mathbb{E}[X_1]} = \frac{1}{\mu} \text{ donc } \widehat{\lambda}^{Mom} = \frac{1}{\widehat{\mu}}.$$

Mais nous avons aussi

$$\mathrm{Var}(X) = \frac{1}{\lambda^2} = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

et donc

$$\mu_2 = \mathbb{E}[X^2] = \frac{2}{\lambda^2} \text{ i.e. } \lambda = \sqrt{\frac{2}{\mu_2}}$$

Nous avons donc un deuxième estimateurs des moments

$$\tilde{\lambda}^{Mom} = \sqrt{\frac{2}{\widehat{\mu}_2}}.$$

L'avantage de la méthode des moments est que, par construction, les estimateurs héritent des propriétés de convergence de la moyenne empirique. Nous avons en particulier les propriétés suivantes :

**Proposition 2** *Soit $X_1, \dots, X_n$ suite i.i.d. de variables aléatoires admettant un moment $\mu_k$ d'ordre $k$ tel que*

$$\theta = h(\mu_k).$$

*Alors l'estimateur des moments*

$$\widehat{\theta}^{Mom} = h\left(\widehat{\mu}_k\right) = h\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right)$$

*vérifie les propriétés suivantes.*

— *Si $h$ est continue au point $t = \mu_k$, alors $\widehat{\theta}^{Mom}$ est un estimateur fortement convergent (donc convergent) de $\theta$.*
— *Si $h(t) = t$ pour tout $t$, alors $\widehat{\theta}^{Mom}$ est sans biais.*
— *Si $h$ est continue et bornée, alors $\widehat{\theta}^{Mom}$ est asymptotiquement sans biais.*
— *S $X_1$ admet un moment $\mu_{2k}$ d'ordre $2k$ et $h$ est de fonction de classe $\mathcal{C}^1$ sur $I$ telle que $h'(\mu_k) \neq 0$, alors $\widehat{\theta}^{Mom}$ est asymptotiquement normal avec variance asymptotique*

$$\frac{h'(\mu_k)^2 \mathrm{Var}(X^k)}{n} = \frac{h'(\mu_k)^2(\mu_{2k} - \mu_k^2)}{n}.$$

# Estimation par maximum de vraisemblance

### 3.2.1 Cas univarié

Nous supposons que la loi de $X$ dépend du paramètre $\theta$. Nous observons $X_1, \dots, X_n$ i.i.d. et de même loi que $X$. Nous supposons que la loi de $X$ dépend d'un paramètre $\theta \in \mathbb{R}$ inconnu.

**Définition 18** *Supposons que l'on dispose d'une réalisation $x_1, \dots, x_n$ d'une échantillon $X_1, \dots, X_n \sim_{i.i.d.} X$. La* **vraisemblance** *des données est la fonction $L(\cdot; x_1, \dots, x_n)$ définie de la façon suivante.*

— *Si $X$ admet une densité $f_\theta$,*

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

— *Si $X$ est une variable discrète, nous notons, pour tout $k \in X(\Omega)$, $\mathbb{P}_\theta(k) = \mathbb{P}(X = k)$*

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

*La* **log-vraisemblance** *des données est la fonction*

$$\ell(\theta; x_1, \dots, x_n) = \ln(L(\theta; x_1, \dots, x_n)).$$

**Définition 19** *Soit $\Theta$ l'ensemble des valeurs possibles du paramètre $\theta$. L'***estimateur du maximum de vraisemblance***, lorsqu'il existe, est la solution du problème de minimisation*

$$\max_{\theta\in\Theta} L(\theta; X_1, \dots, X_n)$$

*ou, de manière équivalente,*

$$\max_{\theta\in\Theta} \ell(\theta; X_1, \dots, X_n).$$

*Nous le noterons $\widehat{\theta}^{EMV}$. Nous pouvons écrire également, lorsque cela à un sens,*

$$\widehat{\theta}^{EMV} = \arg\max_{\theta\in\Theta} L(\theta; X_1, \dots, X_n) = \arg\max_{\theta\in\Theta} \ell(\theta; X_1, \dots, X_n).$$

Il est souvent plus aisé de maximiser la log-vraisemblance que la vraisemblance. Dans les cas usuels, la fonction $\theta \mapsto \ell(\theta; x_1, \dots, x_n)$ est de classe $\mathcal{C}^2$ et l'estimateur du maximum de vraisemblance existe sous deux conditions dites **condition nécessaire (CN)** et **condition suffisante (CS)**. La **condition nécessaire** du programme de minimisation est l'existence d'un **point critique**, noté $\theta^*$, pour la fonction $\ell(\cdot; x_1, \dots, x_n)$ c'est-à-dire qu'il existe $\theta^* \in \mathrm{int}(\Theta)$ ($\mathrm{int}(\Theta)$ est le plus grand ouvert inclus dans $\Theta$) tel que

$$\left.\frac{\partial \ell}{\partial \theta}(\theta; x_1, \dots, x_n)\right|_{\theta=\theta^*} = 0.$$

La condition suffisante permet de vérifier que le point critique $\theta^*$ est un maximum ce qui est vrai si

$$\left.\frac{\partial^2 \ell}{\partial \theta^2}(\theta; x_1, \dots, x_n)\right|_{\theta=\theta^*} < 0.$$

Le point optimal $\theta^*$ est une fonction des observations $x_1, \dots, x_n$ i.e. $\theta^* = \theta^*(x_1, \dots, x_n)$ et nous posons ensuite

$$\widehat{\theta}^{EMV} = \theta^*(X_1, \dots, X_n).$$

Prenons l'exemple de la loi de Poisson de paramètre $\lambda$ (Exercice 20), nous avons $\mathbb{P}_\lambda(k) = \frac{\lambda^k}{k!}e^{-\lambda}$, donc

$$L(\lambda; x_1, \dots, x_n) = \prod_{i=1}^n \mathbb{P}_\lambda(x_i) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}.$$

La log-vraisemblance est égale à

$$\ell(\lambda; x_1, \dots, x_n) = \sum_{i=1}^n \left(x_i \ln(\lambda) - \lambda - \ln(x_i!)\right).$$

Nous avons

$$\frac{\partial \ell}{\partial \lambda}(\lambda; x_1, \dots, x_n) = \sum_{i=1}^n \left(\frac{x_i}{\lambda} - 1\right).$$

Donc la condition (CN) s'écrit

$$\sum_{i=1}^n \left(\frac{x_i}{\lambda^*} - 1\right) \Leftrightarrow \lambda^* = \frac{1}{n}\sum_{i=1}^n x_i.$$

De plus
$$\frac{\partial^2 \ell}{\partial \lambda^2}(\lambda; x_1, \ldots, x_n) = -\sum_{i=1}^{n} \frac{x_i}{\lambda^2} < 0,$$
donc la conditions (CS) est vérifiée également. Nous en déduisons
$$\widehat{\lambda}^{EMV} = \lambda^*(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \widehat{\mu}.$$

Dans l'exemple de l'exercice 18, l'estimateur du maximum de vraisemblance existe même si ni la condition nécessaire, ni la condition suffisante sont vérifiées (la vraisemblance n'est pas dérivable en $\theta$). C'est également un exemple où l'estimateur du maximum de vraisemblance n'est pas un estimateur des moments.

Les deux sections suivantes, marquées d'une *, ne seront pas traitées en cours et ne sont pas au programme de l'examen. Il s'agit de deux extensions classiques et utiles de l'estimateur du maximum de vraisemblance, pour votre culture personnelle.

### 3.2.2 Extension au cas à plusieurs paramètres*

Nous supposons que la loi de $X$ ne dépend plus d'un seul paramètre $\theta \in \mathbb{R}$ mais de plusieurs $\theta_1, \ldots, \theta_d$. Nous devons donc estimer le vecteur $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$. Pour cela, nous pouvons définir de la même façon l'estimateur du maximum de vraisemblance :
$$\widehat{\theta}^{EMV} = \arg\max_{\theta \in \Theta} L(\theta; X_1, \ldots, X_n) = \arg\max_{\theta \in \Theta} \ell(\theta; X_1, \ldots, X_n).$$

Nous l'écrivons ici dans le cas où $X \sim \mathcal{N}(\mu, \sigma^2)$. Le vecteur $\theta$ à estimer est $\theta = (\mu, \sigma^2)$ et la densité s'écrit
$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$
donc la vraisemblance des données $x_1, \ldots, x_n$ est
$$L((\mu, \sigma^2); x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{(\mu,\sigma^2)}(x_i).$$

La log-vraisemblance s'écrit
$$\ell((\mu, \sigma^2); x_1, \ldots, x_n) = \ln\left(\prod_{i=1}^{n} f_{(\mu,\sigma^2)}(x_i)\right) = \sum_{i=1}^{n} \ln(f_{(\mu,\sigma^2)}(x_i)) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 + \frac{n}{2}\ln(2\pi\sigma^2).$$

La condition nécessaire (CN) porte maintenant sur les dérivées partielles par rapport aux différents paramètres
$$\left.\frac{\partial \ell}{\partial \theta_j}(\theta; x_1, \ldots, x_n)\right|_{\theta=\theta^*} = 0, j = 1, \ldots, d.$$

Dans le cas gaussien, nous avons
$$\frac{\partial \ell}{\partial \mu}((\mu, \sigma^2); x_1, \ldots, x_n) = \sum_{i=1}^{n}(x_i - \mu) = n(\bar{x}_n - \mu) \text{ avec } \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

et

$$\frac{\partial \ell}{\partial \sigma^2}((\mu,\sigma^2);x_1,\dots,x_n) = \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i-\mu)^2 + n\frac{1}{2\sigma^2}.$$

(CN) est vérifiée pour $\theta^* = (\mu^*, (\sigma^2)^*)$ tel que

$$\mu^* = \bar{x}_n \text{ et } \frac{1}{2((\sigma^2)^*)^2}\sum_{i=1}^n (x_i-\mu^*)^2 + n\frac{1}{2(\sigma^2)^*} = 0$$

ce qui donne

$$(\sigma^2)^* = \frac{1}{n}\sum_{i=1}^n (x_i-\mu^*)^2.$$

La condition suffisante (CS) porte sur la matrice hessienne des dérivées seconde

$$\left.\left(\frac{\partial^2 \ell}{\partial\theta_j\partial\theta_k}(\theta;x_1,\dots,x_n)\right)_{j,k=1,\dots,d}\right|_{\theta=\theta^*},$$

qui doit être définie négative c'est-à-dire que toutes ses valeurs propres sont négatives (si $d=2$, cela équivaut à dire que son déterminant et sa trace sont tous deux négatifs).

Nous avons

$$\frac{\partial^2 \ell}{\partial \mu^2}((\mu,\sigma^2);x_1,\dots,x_n) = -n$$

$$\frac{\partial^2 \ell}{\partial \mu\partial\sigma^2}((\mu,\sigma^2);x_1,\dots,x_n) = 0$$

$$\frac{\partial^2 \ell}{\partial (\sigma^2)^2}((\mu,\sigma^2);x_1,\dots,x_n) = -\frac{1}{\sigma^6}\sum_{i=1}^n (x_i-\mu)^2 - n\frac{1}{2\sigma^4}.$$

La matrice hessienne est donc diagonale, avec ses deux coefficients diagonaux négatifs, elle est donc définie négative.

Nous obtenons donc l'estimateur du maximum de vraisemblance $\widehat{\theta}^{EMV} = (\widehat{\mu}^{EMV}, \widehat{\sigma^2}^{EMV})$

$$\widehat{\mu}^{EMV} = \widehat{\mu} \text{ et } \widehat{\sigma^2}^{EMV} = \frac{1}{n}\sum_{i=1}^n (X_i-\widehat{\mu})^2.$$

### 3.2.3 Vraisemblance conditionnelle*

Nous pouvons étendre l'estimateur du maximum de vraisemblance à des cas de vraisemblance conditionnelle. C'est le cas par exemple dans un modèle de régression à design aléatoire où nous observons $\{(Y_i, X_i), i=1,\dots,n\}$ tels que

$$Y_i = \beta X_i + \varepsilon_i, \quad i=1,\dots,n,$$

où $\beta\in\mathbb{R}$ est le paramètre inconnu à estimer, $X_1,\dots,X_n$ est i.i.d. de loi inconnue, $\varepsilon_1,\dots,\varepsilon_n \sim_{i.i.d.} \mathcal{N}(0,\sigma^2)$ avec $\sigma^2>0$ supposé connu et $\varepsilon_i$ indépendant de $X_i$. La densité conditionnelle de $Y_i$ sachant $X_i$ est la loi normale de moyenne $\mathbb{E}[Y_i|X_i] = \mathbb{E}[\beta X_i + \varepsilon_i|X_i] = \beta X_i$ et de variance

$\mathrm{Var}(Y_i|X_i) = \mathbb{E}[(Y_i - \mathbb{E}[Y_i|X_i])^2|X_i] = \mathrm{Var}(\varepsilon_i) = \sigma^2$. La densité conditionnelle de $Y_i$ sachant $X_i = x$ s'écrit

$$f_{Y_i|X_i=x}(y,x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\beta x)^2/(2\sigma^2)}.$$

La vraisemblance conditionnelle des données $y_1, \ldots, y_n$ sachant $X_1 = x_1, \ldots, X_n = x_n$ est la quantité

$$L(\beta; y_1, \ldots, y_n|x_1, \ldots, x_n) = \prod_{i=1}^n f_{Y_i|X_i=x}(y_i, x_i),$$

et la log-vraisemblance s'écrit

$$\ell(\beta; y_1, \ldots, y_n|x_1, \ldots, x_n)) = \ln\left(L(\beta; y_1, \ldots, y_n|x_1, \ldots, x_n)\right) = \sum_{i=1}^n \ln(f_{Y_i|X_i=x}(y_i, x_i)).$$

ce qui donne

$$\ell(\beta; y_1, \ldots, y_n|x_1, \ldots, x_n) = -\sum_{i=1}^n \left(\frac{1}{2}\ln(2\pi\sigma^2) + \frac{(y_i - \beta x_i)^2}{2\sigma^2}\right).$$

Écrivons pour simplifier $\ell(\beta; \mathbf{y}|\mathbf{x}) = \ell(\beta; y_1, \ldots, y_n|x_1, \ldots, x_n)$. Nous avons

$$\frac{\partial \ell}{\partial \beta}(\beta; \mathbf{y}|\mathbf{x}) = \frac{1}{\sigma^2}\sum_{i=1}^n x_i(y_i - \beta x_i) = \frac{n}{\sigma^2}(\bar{xy} - \beta\bar{x^2}),$$

avec $\bar{xy} = \frac{1}{n}\sum_{i=1}^n x_i y_i$ et $\bar{x^2} = \frac{1}{n}\sum_{i=1}^n x_i^2$. La condition nécessaire est donc possible si et seulement si $\bar{x^2} \neq 0$ et s'écrit

$$\frac{n}{\sigma^2}(\bar{xy} - \beta^*\bar{x^2}) = 0 \Longrightarrow \beta^* = \frac{\bar{xy}}{\bar{x^2}}.$$

De plus, la condition suffiisante est vérifiée car

$$\frac{\partial^2 \ell}{\partial \beta^2}(\beta; \mathbf{y}|\mathbf{x}) = -\frac{1}{\sigma^2}\sum_{i=1}^n x_i^2 = -\frac{n}{\sigma^2}\bar{x^2} < 0.$$

L'estimateur du maximum de vraisemblance de $\beta$ conditionnellement à $X_1 = x_1, \ldots, X_n = x_n$ s'écrit donc

$$\widehat{\beta}^{EMVcond} = \frac{\frac{1}{n}\sum_{i=1}^n X_i Y_i}{\frac{1}{n}\sum_{i=1}^n X_i^2}.$$

### 3.2.4 Information de Fisher

La vraisemblance associe naturellement la notion d'information de Fisher définie comme cela.

**Définition 20** *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires de fonction de log-vraisemblance $\ell(\theta; x_1, \ldots, x_n)$.*

— *Le* **score** *des observations est la variable aléatoire*

$$\frac{\partial \ell}{\partial \theta}(\theta; X_1, \ldots, X_n).$$

— *L'***information de Fisher** *est le moment d'ordre 2 du score*

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}(\theta; X_1, \dots, X_n)\right)^2\right].$$

Le score est une variable aléatoire d'espérance nulle (propriété admise), l'information de Fisher est donc aussi la variance du score. Lorsque $\theta \mapsto \ell(\theta; x_1, \dots, x_n)$ est deux fois dérivable, on peut aussi écrire

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}(\theta; X_1, \dots, X_n)\right]$$

La notion de score est d'information de Fisher peut-être généralisée au cas où la loi dépend de plusieurs paramètres $\theta_1, \dots, \theta_d$. Dans ce cas, le score est un vecteur aléatoire de coordonnées

$$\frac{\partial \ell}{\partial \theta_j}(\theta; X_1, \dots, X_n), \quad j = 1, \dots, d$$

est l'information de Fisher se présente sous la forme d'une matrice, la matrice de variance-covariance du vecteur de score

$$I(\theta) = \left(\mathbb{E}\left[\frac{\partial \ell}{\partial \theta_j}(\theta; \mathbf{X})\frac{\partial \ell}{\partial \theta_k}(\theta; \mathbf{X})\right]\right)_{1 \leq j,k \leq d}.$$

L'information de Fisher nous donne des informations sur la variance minimale d'un estimateur via la borne dite FDCR.

**Théorème 6 (Inégalité de Fisher-Darmois-Cramer-Rao (FDCR))** *Soit $X_1, \dots, X_n \sim_{i.i.d.} X$. Supposons que le support de $X$ ne dépende pas de $\theta$ et que l'information de Fisher existe alors, pour tout estimateur sans biais $\widehat{\theta}$ de $\theta$,*

$$\mathrm{Var}(\widehat{\theta}) \geq \frac{1}{I(\theta)}.$$

*On dit qu'un estimateur est* **efficace** *s'il atteint la borne FDCR c'est-à-dire si*

$$\mathrm{Var}(\widehat{\theta}) = \frac{1}{I(\theta)}.$$

### 3.2.5 Propriétés du maximum de vraisemblance

Nous appelons l'ensemble d'hypothèses suivante $H_{reg}$.

— Pour tout $x$, la fonction $\theta \mapsto \ln(f_\theta(x))$ (ou $\theta \mapsto \ln(\mathbb{P}_\theta(x))$ si $X$ est discrète) est trois fois différentiable. Ses dérivées sont continues et finies pour toutes valeurs de $x$ et de $\theta$.

— Les dérivées premières et seconde de $\theta \mapsto \ln(f_\theta(X))$ (ou $\theta \mapsto \ln(\mathbb{P}_\theta(X))$ si $X$ est discrète) admettent une espérance.

— $\theta$ appartient à un ensemble compact $\Theta$.

**Théorème 7** *Sous les hypothèses $H_{reg}$ l'estimateur $\widehat{\theta}^{EMV}$ est*

— *convergent*

$$\widehat{\theta}^{EMV} \xrightarrow[n\to+\infty]{\mathbb{P}} \theta$$

,

— *asymptotiquement normal de variance asymptotique*

$$\mathrm{Var}^{(n)}(\widehat{\theta}^{EMV}) = \frac{1}{I(\theta)}.$$

*Il est dit aussi* asymptotiquement efficace.

Reprenons l'exemple d'un échantillon suivant la loi de Poisson de paramètre $\lambda$, le score est

$$\frac{\partial \ell}{\partial \lambda}(\lambda; X_1, \ldots, X_n) = \sum_{i=1}^{n} \left( \frac{X_i}{\lambda} - 1 \right).$$

Nous avons bien

$$\mathbb{E}\left[\frac{\partial \ell}{\partial \lambda}(\lambda; X_1, \ldots, X_n)\right] = \sum_{i=1}^{n} \left( \frac{1}{\lambda}\mathbb{E}[X_i] - 1 \right) = 0.$$

et l'information de Fisher

$$I(\lambda) = \mathrm{Var}\left(\frac{\partial \ell}{\partial \lambda}(\lambda; X_1, \ldots, X_n)\right) = \sum_{i=1}^{n} \mathrm{Var}\left(\frac{X_i}{\lambda} - 1\right) = n\mathrm{Var}(X/\lambda) = \frac{n}{\lambda^2}\mathrm{Var}(X) = \frac{n}{\lambda}.$$

L'estimateur du maximum de vraisemblance

$$\widehat{\lambda}^{EMV} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

est donc asymptotiquement normal, de variance asymptotique $\lambda/n$ (comme il coïncide avec la moyenne empirique nous retrouvons ici le résultat du TCL). L'information en plus, c'est que nous avons maintenant que cet estimateur est efficace puisqu'il atteint la borne FDCR.

# Exercices

**Exercice 19** *On observe $X_1, \ldots, X_n$ un échantillon i.i.d. À partir des moments d'ordre 1 et 2, trouver des estimateurs des moments, lorsque c'est possible, du paramètre $\theta$ dans le cas où*

1. *$X_1$ suit une loi de Bernoulli de paramètre $\theta$.*
2. *$X_1$ suit une loi de Poisson de paramètre $\theta$.*
3. *$X_1$ suit une loi exponentielle de paramètre $\theta$.*

*Étudier les propriétés (convergence, biais, variance asymptotique, normalité asymptotique) de ces estimateurs. En déduire des intervalles de confiances sur ces estimateurs.*

**Exercice 20** *Soient $X_1, \ldots, X_n \sim_{i.i.d.} \mathcal{P}(\lambda)$ et $x_1, \ldots, x_n$ une réalisation de $X_1, \ldots, X_n$.*

1. *Calculer la vraisemblance et la log-vraissemblance des données.*
2. *Supposons que*
$$(x_1, \dots, x_n) = (5, 0, 1, 1, 0, 3, 2, 3, 4, 1).$$
*Donner la valeur de $L(\theta; x_1, \dots, x_n)$ en fonction de $\theta$. Comment interprétez-vous cette quantité ?*
3. *Calculer la valeur de $\theta$ maximisant $\ell(\theta; x_1, \dots, x_n)$.*

**Exercice 21** *Soit $\theta > 0$ et $X \sim \mathcal{U}([0, \theta])$. Nous observons $X_1, \dots, X_n$ i.i.d. de même loi que $X$.*

1. *Proposer un estimateur des moments de $\theta$.*
2. *Quelles sont ses propriétés (biais, variance, loi asymptotique) ?*
3. *Calculer l'estimateur du maximum de vraisemblance de $\theta$.*

**Exercice 22** *Calculer l'estimateur du maximum de vraisemblance :*

1. *du paramètre $p$ de la loi de Bernoulli $\mathcal{B}(p)$ ;*
2. *du paramètre $p$ de la loi binomiale $\mathcal{B}(n, p)$ ($n$ connu) ;*
3. *du paramètre $\lambda$ de la loi exponentielle $\mathcal{E}(\lambda)$.*

**Exercice 23** *Calculer le score et l'information de Fisher*

1. *de la loi de Poisson de paramètre $\lambda$ ;*
2. *de la loi normale de moyenne $\mu$ inconnue et de variance 1 ;*
3. *de la loi exponentielle de paramètre $\theta$.*

*Quelles sont les propriétés asymptotiques de l'estimateur du maximum de vraisemblance ?*

**Exercice 24** *On considère un $n$-échantillon $(X_1, \dots, X_n)$de variables aléatoires de même loi que $X$. On suppose que $X$ admet une distribution log-normale de paramètres $\mu$ et $\sigma^2$, c'est-à-dire que $\ln(X) \sim \mathcal{N}(\mu, \sigma^2)$. La fonction de densité de la variable $X$ est donnée par*

$$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}.$$

*On suppose que $\mu$ est connu et l'on cherche à estimer $\theta = \sigma^2$.*

1. *Déterminer la log-vraisemblance associée à la réalisation de l'échantillon $(x_1, \dots, x_n)$.*
2. *Déterminer l'estimateur du maximum de vraisemblance du paramètre $\theta$.*
3. *Montrer que cet estimateur est sans biais.*
4. *Montrer qu'il est convergent (on supposera que les hypothèses de régularité sont vérifiées).*
5. *Calculer le score de l'échantillon. Vérifier que son espérance est nulle.*
6. *Calculer l'information de Fisher. On pourra s'aider du fait qu'une variable aléatoire $Y \sim \mathcal{N}(0, \sigma^2)$ admet comme moment d'ordre 4 $\mathbb{E}[Y^4] = 3\sigma^4$.*
7. *Déterminer la loi asymptotique de l'estimateur du maximum de vraisemblance (EMV).*
8. *Montrer que l'EMV est efficace au sens de la borne FDCR.*
9. *Proposer un estimateur convergent de la variance asymptotique de cet estimateur.*

# Tests statistiques

## Test de comparaison de proportion

Nous observons $n$ naissances $x_1, \dots, x_n \in \{0,1\}$ ($x_1 = 0$ signifie que la $i$-ème naissance est celle d'un garçon et $x_i = 1$ si c'est une fille). Nous souhaitons déterminer si le sexe (fille ou garçon) d'un enfant qui vient de naître est équiprobable.

Nous modélisons les observations $X_1, \dots, X_n$ comme une suite i.i.d. selon la loi de Bernoulli de paramètre $p$ inconnu. Nous estimons $p$ par la méthode usuelle

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Une idée naturelle est de comparer $\widehat{p}(\mathbf{x})$ à $p_0 = \frac{1}{2}$ et de décider que le nombre de naissances n'est pas équiprobable si $\widehat{p}(\mathbf{x})$ est loin de $p_0$. Mais à quel moment décide-t'on que $\widehat{p}(\mathbf{x})$ est suffisamment loin de $p_0$ ? Cela dépend déjà du nombre $n$ d'observations. En effet, si nous observons une proportion $\widehat{p}(\mathbf{x}) = 0.66$, si $n = 3$ naissances (2 filles et 1 garçons), nous estimons que nous n'avons pas assez de données pour conclure. Et si nous en observons 100, est-ce suffisant pour conclure ?

Nous savons, par la loi des grands nombres, que

$$\widehat{p} \underset{n\to+\infty}{\longrightarrow} p \text{ p.s.}$$

et, par continuité de la fonction $x \mapsto |x - p_0|$,

$$|\widehat{p} - p_0| \underset{n\to+\infty}{\longrightarrow} |p - p_0| \text{ p.s.}$$

Cela veut dire que, si $p = p_0$,

$$|\widehat{p} - p_0| \underset{n\to+\infty}{\longrightarrow} 0 \text{ p.s.}$$

et, sinon $|\widehat{p} - p_0|$ converge p.s. vers une quantité non nulle mais cela ne nous dit pas à partir de quand, avec une seule observation $\widehat{p}(\mathbf{x})$, considère-t'on que $|\widehat{p}(\mathbf{x}) - p_0|$ est suffisamment grand ?

Dans le chapitre précédent, nous avons vu que

$$\frac{\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}(\widehat{p} - p) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0,1),$$

nous avons vu également que cela impliquait, pour tout $\alpha \in ]0,1[$, que

$$\mathbb{P}\left(\left|\frac{\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}(\widehat{p} - p)\right| \leq q_{1-\alpha/2}\right) \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathbb{P}(|Z| \leq q_{1-\alpha/2}) = 1 - \alpha.$$

Cela implique, qu'avec probabilité "proche" de $1 - \alpha$,

$$|\widehat{p} - p| \leq \frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}} q_{1-\alpha/2}. \tag{1}$$

Si $p = p_0$, et $n$ suffisamment grand, la probabilité que (1) soit vérifiée est donc très grande (de l'ordre de $1-\alpha$ avec $\alpha$ petit). Si (1) n'est pas vérifiée, cela indique que la valeur observée de $\widehat{p}(\mathbf{x})$ est en contradiction avec le fait que $p = p_0$.

En terme de **théorie des tests**, nous testons

$$H_0 : p = p_0 \text{ versus } H_1 : p \neq p_0.$$

La quantité

$$T = \frac{\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}(\widehat{p} - p_0)$$

est appelé **statistique de test** (c'est une fonction des observations). Lorsque

$$|T| > q_{1-\alpha/2}$$

nous **rejetons** $H_0$ : cela veut dire que les observations sont en contradiction avec le fait que $p = 1/2$. Nous appelons **région critique** de niveau $\alpha$

$$\mathcal{R}_\alpha = \{\mathbf{x} = (x_1, \ldots, x_n), |T(\mathbf{x})| > q_{1-\alpha/2}\},$$

l'ensemble des $n$-uplets d'observations possibles conduisant au rejet de l'hypothèse nulle $H_0$.

Etant donné des observations $x_1, \ldots, x_n$, on peut chercher la valeur de $\alpha$ la plus petite possible pour que $H_0$ soit rejetée, c'est ce que l'on appelle ***p*-valeur** du test qui s'interprète usuellement comme suit :

- $p$-valeur $< 0.01$ : très forte présomption que $H_0$ soit fausse,
- $0.01 \leq p$-valeur $< 0.05$ : forte présomption que $H_0$ soit fausse (on rejette $H_0$ au niveau 5% mais pas au niveau 1%),
- $0.05 \leq p$-valeur $< 0.1$ : faible présomption que $H_0$ soit fausse,
- $p$-valeur $\geq 0.1$ : pas de preuve que $H_0$ soit fausse.

Exemples :

- Nous observons $\widehat{p}(\mathbf{x}) = 0.66$ avec $n = 3$ naissances. La statistique de test vaut environ

$$T(x) = 0.59$$

qui est plus petite que les quantiles usuels de la loi normale ($q_{0.95} = 1.64$, $q_{0.975} = 1.96$, $q_{0.99} = 2.33$). On ne rejette donc pas $H_0$. Les données ne sont pas en contradiction avec le fait que $p = 1/2$. Cela ne veut pas dire que $p = 1/2$ mais que potentiellement on n'a pas assez de données pour conclure. La $p$-valeur vaut

$$\begin{aligned}
\min\{\alpha > 0; |T| > q_{1-\alpha/2})\} &= \min\{\alpha > 0; F_Z(|T|) > F_Z(q_{1-\alpha/2}) = 1 - \alpha/2\} \\
&= \min\{\alpha > 0; \alpha > 2(1 - F_Z(|T|))\} \\
&= 2(1 - F_Z(|T|)) = 0.56.
\end{aligned}$$

- Nous observons $\widehat{p}(\mathbf{x}) = 0.66$ mais cette fois avec $n = 100$ naissances. La statistique de test vaut environ

$$T(x) = 3.38$$

qui est plus **grande** en valeur absolue que les quantiles usuels de la loi normale, nous rejetons donc $H_0$ aux niveaux usuels $\alpha = 10\%$, $\alpha = 5\%$, $\alpha = 2\%$,...). Cela veut dire que la valeur $\widehat{p}(\mathbf{x})$ observée contredit $H_0 : p = 1/2$. La $p$-valeur vaut environ

$$\min\{\alpha > 0; |T| > q_{1-\alpha/2})\} = 7.31.10^{-4}.$$

— Nous observons $\widehat{p}(\mathbf{x}) = 0.51$ avec $n = 100$ naissances. La statistique de test vaut environ

$$T(x) = 0.20$$

qui est plus petite en valeur absolue que les quantiles usuels de la loi normale, nous acceptons donc $H_0$. La $p$-valeur vaut environ

$$\min\{\alpha > 0; |T| > q_{1-\alpha/2})\} = 0.84.$$

— Nous observons $\widehat{p}(\mathbf{x}) = 0.51$ avec $n = 736000$ naissances (nombre de naissances en France en 2020). La statistique de test vaut environ

$$T(x) = 17.16$$

qui est plus grande en valeur absolue que les quantiles usuels de la loi normale, nous rejetons donc $H_0$ bien que $\widehat{p}(\mathbf{x})$ soit relativement proche de $p_0$, cela veut dire que la différence en $\widehat{p}$ est $p_0$ est significative.

Supposons maintenant que nous testions un vaccin. Sans vaccin, la probabilité d'apparition de la maladie après exposition est de 20%. Parmi $n$ personnes vaccinées et exposées au virus, nous observons la proportion $\widehat{p}(\mathbf{x})$ de personnes malades. Cette fois-ci, nous testons

$$H_0 : p = 0.2 \text{ versus } H_1 : p < 0.2.$$

c'est ce que l'on appelle un **test unilatéral à gauche** (par opposition au **test bilatéral** précédent). Cette fois-ci, nous rejetons $H_0$ lorsque $\widehat{p}(\mathbf{x})$ est trop petit. La statistique de test est la même mais la région de rejet est

$$\mathcal{R}_\alpha = \{\mathbf{x} = (x_1, \ldots, x_n), T(\mathbf{x}) \leq q_\alpha\}$$

car, sous $H_0$,

$$\mathbb{P}(\mathcal{R}_\alpha) = \mathbb{P}\left(T \leq q_\alpha\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathbb{P}(Z \leq q_\alpha) = \alpha.$$

Si $\widehat{p}(\mathbf{x}) > 0.2$, nous ne rejetons jamais $H_0$ mais, si nous rejetons $H_0$, cela veut dire que les données observées sont en contradiction avec une inefficacité du vaccin, ou, dit de manière différente, que les observations indiquent que le vaccin réduit la probabilité d'apparition de la maladie de manière significative. La $p$-valeur est liée à la région de rejet et change donc aussi

$$p\text{-valeur} \quad = \quad \min\{\alpha > 0; T \leq q_\alpha\} = \min\{\alpha > 0; F_Z(T) \leq F_Z(q_\alpha) = \alpha\} = F_Z(T).$$

À l'inverse, un **test unilatéral à droite** serait de la forme

$$H_0 : p = p_0 \text{ versus } H_1 : p > p_0,$$

avec une région de rejet de la forme

$$\mathcal{R}_\alpha = \{\mathbf{x} = (x_1, \ldots, x_n), T(\mathbf{x}) \geq q_{1-\alpha}\}$$

car, sous $H_0$,

$$\mathbb{P}(\mathcal{R}_\alpha) = \mathbb{P}\left(T \geq q_{1-\alpha}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathbb{P}(Z \geq q_{1-\alpha}) = 1 - \mathbb{P}(Z < q_{1-\alpha}) = 1 - (1-\alpha) = \alpha$$

et

$$
\begin{aligned}
p\text{-valeur} &= \min\{\alpha > 0; T \geq q_{1-\alpha}\} = \min\{\alpha > 0; F_Z(T) \geq F_Z(q_{1-\alpha}) = 1-\alpha\} \\
&= \min\{\alpha > 0; \alpha \geq 1 - F_Z(T)\} = 1 - F_Z(T).
\end{aligned}
$$

En résumé nous avons :

| Test | $H_1$ | Région critique $\mathcal{R}_\alpha$ | $p$-valeur |
|---|---|---|---|
| bilatéral | $p \neq p_0$ | $\{\mathbf{x}; \lvert T(\mathbf{x})\rvert > q_{1-\alpha/2}\}$ | $2(1 - F_Z(\lvert T(\mathbf{x})\rvert)$ |
| unilatéral à gauche | $p < p_0$ | $\{\mathbf{x}; T(\mathbf{x}) < q_\alpha\}$ | $F_Z(T(\mathbf{x}))$ |
| unilatéral à droite | $p > p_0$ | $\{\mathbf{x}; T(\mathbf{x}) > q_{1-\alpha}\}$ | $1 - F_Z(T(\mathbf{x}))$ |

où, nous avons toujours, l'hypothèse nulle $H_0 : p = p_0$, la statistique de test

$$
T = \frac{\sqrt{n}}{\sqrt{\widehat{p}(1-\widehat{p})}}(\widehat{p} - p_0)
$$

et $F_Z$ désigne la fonction de répartition de $Z \sim \mathcal{N}(0,1)$.

Remarquons que, pour le test bilatéral, $\mathbf{x} \in \mathcal{R}_\alpha$ si et seulement si $p_0 \in IC_{1-\alpha}^{(n)}$ où $IC_{1-\alpha}^{(n)}$ est la réalisation sur les données de l'intervalle de confiance asymptotique sur une proportion défini p. 25.

# Théorie des tests

Nous reprenons ici, de manière générale, les différentes notions vues dans la section précédente.

**Définition 21** — *Un* **test statistique** *est une règle de décision relative à une hypothèse sur la distribution d'une variable d'intérêt dans la population, qui se fonde sur les observations d'un échantillon.*

- *Une* **hypothèse** *est une assertion concernant la population. Une test statistique permet de tester la validité d'une hypothèse de référence, dite* **hypothèse nulle***, et notée $H_0$, contre une* **hypothèse alternative***, notée $H_1$.*
- *Une* **statistique de test** *est une variable aléatoire définie comme une fonction de l'échantillon $T = T(X_1, \ldots, X_n)$.*
- *La* **région critique** *d'un test, est un ensemble de réalisations possibles de l'échantillon pour lesquelles l'hypothèse nulle du test est rejetée*

$$
\mathcal{R} = \{\mathbf{x} = (x_1, \ldots, x_n); T(x) \in \Gamma(c)\},
$$

$\Gamma(c)$ *est un ensemble dépendant de l'hypothèse alternative $H_1$.*

Exemple : $\Gamma_\alpha(c) = ]-\infty; -q_{1-\alpha/2}[\cup]q_{1-\alpha/2}; +\infty[$ pour le test de proportion bilatéral de niveau $\alpha$.

Pour évaluer les performances d'un test, nous considérons deux erreurs :

— le **risque de première espèce** correspond au risque de rejeter l'hypothèse nulle $H_0$ alors qu'elle est vraie. La probabilité de cet évènement est le **niveau du test**

$$\alpha = \mathbb{P}_{H_0}(\mathcal{R}).$$

— Le **risque de seconde espèce** correspond au risque de ne pas rejeter l'hypothèse nulle $H_0$ alors qu'elle est fausse. La probabilité de cet évènement est la quantité

$$\beta = \mathbb{P}_{H_1}(\mathcal{R}^c).$$

La probabilité de l'évènement complémentaire est appelée la **puissance d'un test**

$$\text{Puissance} = 1 - \beta = \mathbb{P}_{H_1}(\mathcal{R}).$$

Il n'est pas possible d'avoir un test qui soit à la fois d'un niveau et d'une puissance élevée. Par exemple :

— le test qui rejette tout le temps l'hypothèse nulle $\mathcal{R}^c = \emptyset$ est de niveau $\alpha = 1$ (maximal) mais de puissance nulle.

— À l'inverse, le test qui ne rejette jamais l'hypothèse nulle $\mathcal{R} = \emptyset$ est de puissance $\beta = 1$ (maximale) mais de niveau $\alpha = 0$.

Ces deux tests ne sont d'aucune utilité en pratique. Les statisticiens ont généralement adopté le **principe de Neyman**, du nom du mathématicien Jerzy Neyman (1894–1981).

1. On fixe le niveau du test $\alpha$ ($\alpha = 1\%$, $\alpha = 5\%$,...).
2. On en déduit une région critique $\mathcal{R}_\alpha$ de niveau $\alpha$ ou de niveau asymptotiquement égal à $\alpha$ c'est-à-dire tel que
$$\lim_{n\to+\infty} \mathbb{P}_{H_0}(\mathcal{R}_\alpha) = \alpha.$$
Si plusieurs régions critiques sont possibles, on choisit celle qui donne un test le plus puissant possible.

**Définition 22** *Un test est dit* **uniformément plus puissant (UPP)** *de niveau $\alpha$ si sa puissance est plus élevée que celle de tous les autres tests de niveau $\alpha$.*

**Définition 23** *Soient $x_1, \ldots, x_n$ une réalisation d'une suite i.i.d. selon une loi de paramètre $\theta$ inconnu. Soient $\theta_0$ et $\theta_1$ deux valeurs possibles du paramètre $\theta$ (choisies par le statisticien). Nous testons*

$$H_0 : \theta = \theta_0 \text{ contre } H_1 : \theta = \theta_1.$$

*Pour tout niveau $\alpha \in [0,1]$ le* **test du rapport de vraisemblance** *est le test de région critique*

$$\mathcal{R}_\alpha = \left\{ \mathbf{x} = (x_1, \ldots, x_n), \frac{L_n(\theta_0; \mathbf{x})}{L_n(\theta_1; \mathbf{x})} \leq k_\alpha \right\}$$

*où $k_\alpha$ est le quantile d'ordre $\alpha$ sous $H_0$ (c'est-à-dire lorsque $\theta = \theta_0$) de la variable aléatoire*

$$T = \frac{L_n(\theta_0; \mathbf{X})}{L_n(\theta_1; \mathbf{X})},$$

*où $\mathbf{X} = (X_1, \ldots, X_n)$.*

**Théorème 8 (Lemme de Neyman-Pearson (admis))** *Pour tout $\alpha \in [0,1]$, le test du rapport de vraisemblance est UPP de niveau $\alpha$.*

# Tests du $\chi^2$

**Test d'adéquation** Le test d'adéquation (ou d'ajustement) du $\mathcal{X}^2$ permet de tester si des observations $x_1, \ldots, x_n$ d'un échantillon $X$ discrète à valeurs dans l'ensemble $A = \{a_1, \ldots, a_J\}$ sont issues d'une distribution $D$ donnée. Il s'écrit

$$H_0 : X \sim D \text{ contre } H_1 : X \text{ ne suit pas la distribution } D.$$

Pour réaliser ce test, nous comparons les effectifs réels observés $n_1, \ldots, n_s$ où

$$n_j = \sum_{i=1}^{n} \mathbf{1}_{\{x_i = a_j\}}$$

et les effectifs théoriques

$$m_j = n\mathbb{P}_D(X = a_j),$$

où $(\mathbb{P}_D(X = a_j))_{j=1,\ldots,s}$ caractérise la distribution de $X$ lorsque $X \sim D$ (donc sous $H_0$). Sous $H_0$, nous nous attendons à ce que les effectifs observés soient proches des effectifs théoriques mais il ne sont jamais complètements égaux. Nous mesurons donc l'écart entre les effectifs observés et théoriques et souhaitons rejeter $H_0$ lorsque cet écart est très grand. Cela conduit à s'appuyer sur la statistique de test

$$C_n(\mathbf{x}) = \sum_{j=1}^{s} \frac{(n_j - m_j)^2}{m_j}.$$

Sous $H_0$, nous connaissons la distribution asymptotique de $C_n$ qui est **la loi du** $\chi^2$ à $s-1$ degrés de liberté, notée $\chi^2(s-1)$, qui est une distribution dont nous pouvons calculer les quantiles. Nous notons par la suite $q_\alpha^{\chi^2(s-1)}$ le quantile d'ordre $\alpha$ de cette loi. La région critique du test du $\chi^2$ est notée :

$$\mathcal{R}_\alpha = \left\{ \mathbf{x} : C_n(\mathbf{x}) > q_{1-\alpha}^{\chi^2(s-1)} \right\}.$$

Pour pouvoir réaliser un test du $\chi^2$ il convient de s'assurer qu'aucune classe ne contient un effectif trop faible. Une règle communément admise consiste à s'assurer que $n_j > 5$ pour tout $j = 1, \ldots, s$.

Il peut arriver que la distribution $D$ dépende d'un ou plusieurs paramètres inconnu que nous devons estimer. Dans ce cas, si $\ell$ est le nombre de paramètres à estimer, la distribution asymptotique de la statistique $C_n$ sous $H_0$ est la loi du $\chi^2$ à $s - \ell - 1$ degrés de liberté.

**Test d'indépendance** Le test d'indépendance du $\chi^2$ permet de tester si deux variables discrètes $X$ à valeurs dans $\{a_1, \ldots, a_J\}$ et $Y$ à valeurs dans $\{b_1, \ldots, b_K\}$ sont indépendantes. Il s'écrit

$$H_0 : X \text{ et } Y \text{ sont indépendantes contre } H_1 : X \text{ et } Y \text{ne sont pas indépendantes.}$$

Il s'agit d'un cas particulier du test d'adéquation où la loi $D$ s'écrit

$$\mathbb{P}_D((X, Y) = (a_j, b_k)) = \mathbb{P}(X = a_j)\mathbb{P}(Y = b_k),$$

et où nous estimons les $(J-1)+(K-1) = J+K-2$ [1] paramètres de cette loi $p_X(a_1), \ldots, p_X(a_J), p_Y(b_1), \ldots, p_Y(b_K)$ où $p_X(a_j) = \mathbb{P}(X = a_j)$ et $p_Y(b_k) = \mathbb{P}(Y = b_k)$ de la façon suivante

$$\widehat{p}_X(a_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i=a_j\}} \text{ et } \widehat{p}_Y(b_k) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i=b_k\}}.$$

Notons

$$n_{j,k} = \sum_{i=1}^{n} \mathbf{1}_{\{x_i=a_j, y_i=b_k\}}$$

les effectifs observés et

$$m_{j,k} = n\widehat{p}_X(a_j)\widehat{p}_Y(a_k),$$

les effectifs théoriques. La statistique de test s'écrit

$$D_n(\mathbf{x}) = \sum_{j=1}^{J}\sum_{k=1}^{K} \frac{(n_{j,k} - m_{j,k})^2}{m_{j,k}},$$

qui suit asymptotiquement la loi du $\chi^2$ à $KJ - (J+K-2) - 1 = (K-1)(J-1)$ degrés de liberté. La région critique s'écrit

$$\mathcal{R}_\alpha = \left\{\mathbf{x} : D_n(\mathbf{x}) > q_{1-\alpha}^{\chi^2((K-1)(J-1))}\right\}.$$

# Exercices

**Exercice 25** *Vérifier que les tests de comparaison de proportion satisfont le principe de Neyman.*

**Exercice 26** *On observe $n = 100$ variables i.i.d. $X_1, \ldots, X_n$ selon la loi $\mathcal{N}(\mu, \sigma^2)$ avec $\sigma = 1$. Le paramètre $\mu$ est inconnu et nous souhaitons tester*

$$H_0 : \mu = \mu_0 = 1.2 \text{ contre } H_1 : \mu = \mu_1 = 1.4.$$

1. *Écrire la vraisemblance de l'échantillon.*
2. *Montrer que la région critique du test du rapport de vraisemblance s'écrit*

$$\mathcal{R}_\alpha = \{\mathbf{x} : \widehat{\mu}(\mathbf{x}) > c_\alpha\},$$

*où $c_\alpha$ est à déterminer.*

3. *Vérifier que*

$$c_\alpha = \mu_0 + \frac{\sigma}{\sqrt{n}} q_{1-\alpha}.$$

1. le nombre de paramètres à estimer est $(J-1)+(K-1)$ et non $J+K$ car une fois estimé $p_X(a_1), \ldots, p_X(a_{J-1})$, nous avons $p_X(a_J) = 1 - p_X(a_1) - \ldots - p_X(a_{J-1})$.

| Nombre de voitures $X$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 et + | Total |
|---|---|---|---|---|---|---|---|---|---|
| Nombre de relevés | 15 | 25 | 25 | 12 | 10 | 8 | 4 | 1 | 100 |

*Fig.2 Tableau de répartition empirique.*

| Montant/média | Emailing | Courriers | Appels | Total |
|---|---|---|---|---|
| 50–100 € | 220 | 200 | 50 | 470 |
| 100–200 € | 140 | 250 | 100 | 490 |
| + 200 € | 140 | 50 | 350 | 540 |
| Total | 500 | 500 | 500 | 1500 |

*Fig.3 Tableau de contingence empirique.*

**Exercice 27** *Afin d'ajuster au mieux la gestion du personnel des gares de péage, une société d'autoroute souhaite modéliser le nombre de voitures, noté $X$, se présentant à un péage par tranche de 30 minutes. On souhaite tester si la variable aléatoire $X$ admet une distribution de Poisson, pour un niveau de risque de 5%. On dispose pour cela d'un échantillon de 100 relevés consécutifs durant lesquels on a compté le nombre de voitures passant le péage. La répartition des effectifs empiriques est donnée dans le tableau Fig. 2.*

1. *Tester si les données peuvent suivre une loi de Poisson de paramètre $\lambda = 2$.*
2. *Tester si les données peuvent suivre une loi de Poisson de paramètre $\lambda$ inconnu.*

**Exercice 28** *Une entreprise souhaite analyser l'impact d'une campagne marketing suivant les trois canaux de diffusion utilisés : emailing, courriers et appels téléphoniques. Pour cela elle dispose d'un échantillon de 1500 clients ayant été contactés par l'un des trois médias, représenté par le tableau de contingence de la Fig. 3.*

*Faire un test du $\chi^2$ d'indépendance entre les variables médias et montant.*