Statistiques mathématiques

R. Petit

année académique 2016 - 2017

Table des matières

Introduction

En probabilités, une variable aléatoire X donnée est entièrement définie par sa loi. On peut l'exprimer par la fonction de répartition F^X ou par la fonction de densité $f^X = \frac{d}{dx}F^X$. Ces fonctions permettent de déterminer :

$$\mathbb{P}[a \leqslant X \leqslant b] = \int_a^b f^X(x)\,dx = F^X(b) - F^X(a).$$

Ou encore :

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

Cependant, les fonctions f^X et F^X ne sont jamais connues précisément. Elles peuvent être approchées par des modélisations, mais les modèles ne sont jamais exacts. En probabilités, on cherche donc les observations sur base de la loi qui est connue, alors qu'en statistiques, on cherche à retrouver la loi sur base de n observations $X_1, \ldots, X_n$.

Nous allons nous intéresser à des *modèles statistiques* sous la forme $\left(\mathbb{R}^n, \mathcal{B}\left(\mathbb{R}^n\right), \mathcal{P}^{(n)}\right)$ où :

$$\mathcal{P}^{(n)} = \left\{\mathbb{P}^{(n)}\right\} = \left\{\mathbb{P}_\theta^{(n)} \text{ t.q. } \theta \in \Theta \subset \mathbb{R}^k\right\},$$

et donc les $\mathbb{P}^{(i)}$ sont chacun une loi possible pour $(X_1, \ldots, X_n)$.

Ces modèles sont dits *paramétriques* car les différentes lois sont les mêmes au paramètre θ près. Nous n'étudierons que des modèles paramétriques où Θ est un espace de dimension $d \in \mathbb{N}$ finie.

Exemple 0.1. Soient $X_1, \ldots, X_n$ des variables aléatoires iid (indépendantes et identiquement distribuées).

— Si les X_i sont de loi normale $\mathcal{N}(\mu, \sigma^2)$, alors le paramètre θ est donné par :

$$\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \in \Theta = \mathbb{R} \times \mathbb{R}^+ \subset \mathbb{R}^2\ ;$$

— si les X_i sont de loi uniforme $\mathrm{Unif}(0, \theta)$, le paramètre θ est donné par $\theta \in \Theta = \mathbb{R}_0^+ \subset \mathbb{R}$;

— si les X_i sont de loi $\mathrm{Bern}(p)$, le paramètre θ est donné par $\theta = p \in \Theta = [0, 1] \subset \mathbb{R}$.

Remarque. Une loi normale $\mathcal{N}(\mu, \sigma^2)$ est déraisonnable car les valeurs observables ne vont empiriquement pas vers les infinis alors que la distribution le permet théoriquement mais n'est pas **complètement** déraisonnable car ces probabilités sont négligeables grâce à l'exponentielle de $(-x^2)$ dans la formule de la densité.

Chapitre 1

Théorie de l'échantillonnage

1.1 Terminologie et définitions

Définition 1.1. On appelle *modèle d'échantillonnage* un modèle d'observations iid.

Définition 1.2. Soit un modèle statistique $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$ où $\mathcal{P}^{(n)} = \left\{\mathbb{P}_\theta^{(n)} \text{ t.q. } \theta \in \Theta \subset \mathbb{R}^k\right\}$. On note ici $\mathbb{P}_\theta^{(n)}$ une loi possible pour $(X_1, \ldots, X_n)$ et $\mathbb{P}_\theta$ une loi possible pour X_i avec i fixé. On dit alors que $\mathbb{P}_\theta^{(n)}$ est déterminé par $\mathbb{P}_\theta$.

Remarque. Ici, deux visions vont s'opposer et se compléter : la vision *population* qui est associée à P_θ et la version *échantillonage* (ou *empirique*), qui, elle, est associée à $P_\theta^{(n)}$.

Définition 1.3. On définit la fonction indicatrice $I_{[\cdot]}$ qui vaut 1 quand l'expression entre crochets est vraie et 0 sinon.

Définition 1.4. Soit $X_1, \ldots, X_n$ une suite de n observations. On définit la i*eme statistique d'ordre* par $X_{(i)} = X_k$ t.q. $\left|\{X_j \text{ t.q. } X_j < X_k, 1 \leqslant j \leqslant n\}\right| = i$. On définit également la *statistique d'ordre* par $\left(X_{(i)}\right)_i$.

Définition 1.5. On définit les fonctions de répartitions comme suit :

— la fonction de répartition population :

$$F_\theta(x) = P_\theta[X_i \leqslant x] \;;$$

— la fonction de répartition empirique :

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{[X_i \leqslant x]}.$$

Remarque. La fonction F_n empirique est une fonction en escaliers. Elle fait des sauts de hauteur $\frac{1}{n}$, et est telle que :

$$\lim_{x\to+\infty} F_n(x) = 1 \qquad \text{et} \qquad \lim_{x\to-\infty} F_n(x) = 0.$$

On peut également remarquer que $F_n(X_{(i)}) = \frac{i}{n}$. En effet, par définition de $X_{(i)}$, il y a exactement i observations inférieures à $X_{(i)}$. Dès lors, la fonction indicatrice donnera i fois la valeur 1 et $(n-i)$ fois la valeur 0. La somme donc donc i et la fonction donne $\frac{i}{n}$.

Définition 1.6. On appelle *statistique* toute fonction mesurable faisant intervenir **uniquement** des observations.

Exemple 1.1. Par exemple F_n est une statistique car seules les valeurs X_i sont utilisée, mais F_θ n'est pas une statistique car la valeur du paramètre θ apparait et n'est pas une observation.

Remarque. Une statistique peut être à valeur scalaire ($X_{(i)}$ par exemple), à valeur vectorielle ($\left(X_{(i)}\right)_{1\leqslant i\leqslant n}$ par exemple), à valeur ensembliste ($[X_i \pm \overline{X}]$ avec i fixé par exemple), ou encore à valeur fonctionnelle (F_n par exemple).
Remarque. L'objectif est de pouvoir approximer la loi régissant les populations (F_θ) à l'aide de la loi observée empiriquement. Par la loi des grands nombres, on a :

$$F_n(x) \xrightarrow[n\to+\infty]{\text{p.s. par } \mathbb{P}_\theta} F_\theta(x).$$

Théorème 1.7 (Théorème de Glivenko-Cantelli)**.** *Si F_n et F_θ sont repsectivement une fonction de répartition empirique et de population, alors :*

$$\sup_{x\in\mathbb{R}}\left|F_n(x) - F_\theta(x)\right| \xrightarrow[n\to+\infty]{\text{p.s.}} 0$$

1.2 Moments

Définition 1.8 (Moments pour populations)**.** On définit $\mu_r'(\theta)$ le *moment **non**-centré* d'ordre r avec $r \in \mathbb{N}^*$ par :

$$\mu_r'(\theta) \coloneqq E_\theta[X_1^r].$$

On définit également $\mu_r(\theta)$, le *moment **centré*** d'ordre r avec $r \in \mathbb{N}^*$ par :

$$\mu_r(\theta) \coloneqq E_\theta\left[\left(X_1 - \mu_1'(\theta)\right)^r\right].$$

Définition 1.9 (Moments pour échantillon)**.** On définit m_r', le *moment **non**-centré* d'ordre r avec $r \in \mathbb{N}^*$ par :

$$m_r' \coloneqq \frac{1}{n}\sum_{i=1}^{n} X_i^r.$$

On définit également le *moment **centré*** d'ordre r avec $r \in \mathbb{N}^*$ par :

$$m_r \coloneqq \frac{1}{n}\sum_{i=1}^{n} \left(X_i - m_1'\right)^r.$$

Remarque. La loi des grands nombres dit que :

$$m_r' \xrightarrow[n\to+\infty]{\text{p.s.}} \mu_r'(\theta),$$

mais on ne peut pas dire que :

$$m_r \xrightarrow[n\to+\infty]{\text{p.s.}} \mu_r(\theta).$$

Ce n'est donc pas possible car pour m_r', il y a une somme de variables iid alors que pour m_r, les variables sommées ne sont pas iid (mais dépendent toutes de tous les X_i).

En réalité, il y a convergence, mais on ne peut pas l'exprimer de manière triviale par la loi des grands nombres.

1.2.1 Indicateurs

On peut observer que $\mu_1'(\theta) = \mathbb{E}_\theta[X_1]$. Pareil pour $m_1' = \overline{X}$. Le moment d'ordre 1 est donc un indice de position. On a alors $\mu := \mu_1(\theta) = \mathbb{E}[(X - \mathbb{E}[X_1])] = \mathbb{E}[X_1] - \mathbb{E}[X_1] = 0$. Cette valeur n'est donc pas intéressante. Par contre :

$$\mu_2(\theta) = \mathbb{E}\left[(X_1 - \mathbb{E}[X_1])^2\right] =: \mathrm{Var}(X) \qquad \text{si} \qquad m_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 =: s^2.$$

Le moment d'ordre 2 est donc un indice de dispersion.

Définition 1.10. On appelle le *coefficient d'asymétrie de Fisher* la quantité :

$$\gamma_1 := \mu_3(\theta) \cdot \left(\mu_2(\theta)\right)^{-\frac{3}{2}}.$$

Remarque. Le dénominateur $\mu_2(\theta)^{\frac{3}{2}}$ apparait afin de rendre invariant le coefficient d'asymétrie de Fisher aux transformations affines.

Définition 1.11. Le coefficient d'asymétrie de Fisher *empirique* est donné par :

$$m_3 \cdot m_2^{-\frac{3}{2}}.$$

Définition 1.12. On appelle *coefficient d'applatissement de Fisher* la quantité :

$$\gamma_2 := \mu_4(\theta) \cdot \left(\mu_2(\theta)\right)^{-2} - 3.$$

Définition 1.13. Le coefficient d'aplatissement de Fisher *empirique* est donné par :

$$m_4 \cdot m_2^{-2} - 3.$$

Remarque. Si $\gamma_2 \gneqq 0$, c'est que les événements extrêmes sont de plus haute probabilité et si $\gamma_2 \lneqq 0$, c'est que les événements extrêmes sont de moins haute probabilité.

À nouveau, le dénominateur y a été ajouté afin de rendre le coefficient invariant aux transformations affines. Et le terme -3 sert à annuler le coefficient d'aplatissement de Fisher pour une normale $\mathcal{N}(\mu, \sigma^2)$.

1.3 Quantile

Définition 1.14. Si F_θ est inversible, alors on définit $x_\alpha(\theta) := F_\theta^{-1}(\alpha)$, et on appelle $x_\alpha(\theta)$ un *quantile*.

Remarque. Il faut cependant faire attention car on peut avoir le cas de F_θ discontinue où on choisit $\alpha = F_\theta^{-1}$(point de discontinuité) ou alors le cas de F_θ admettant un plateau et où on choisit α sur le plateau.

Définition 1.15. On définit alors :

$$x_\alpha(\theta) := \inf\left\{x \in \mathbb{R} \text{ t.q. } F_\theta(x) \geqslant \alpha\right\}.$$

Remarque. On donne les noms de *médiane*, *quartile*, *décile*, *percentile* pour α valant, avec k entier, respectivement $\frac{1}{2}$, $\frac{k}{4}$ avec $k < 4$, $\frac{k}{10}$ avec $k < 10$, et $\frac{k}{100}$ avec $k < 100$.

Définition 1.16. Pour les échantillons, on définit le *quantile empirique d'ordre* α par :

$$x_\alpha^{(n)} := \inf\{x \in \mathbb{R} \text{ t.q. } F_n(x) \geqslant \alpha\}.$$

Remarque. On peut également définir des indices de position, dispersion, asymétrie, aplatissement, etc. sur les quantiles plutôt que sur les moments. Ils auront des propriétés différentes et une robustesse différente aux valeurs aberrantes.

Définition 1.17. La loi échantillonnée de $T(X^{(n)})$ est la loi déterminée par :

$$\mathbb{P}_\theta^{(n)}\left[T(X^{(n)}) \in B\right] = \mathbb{P}_\theta^{(n)}\left[\left\{x^{(n)} \in \mathcal{X}^{(n)} \text{ t.q. } T(x^{(n)}) \in B\right\}\right], B \in \mathcal{B}(\mathbb{R}^m).$$

Exemple 1.2 (Bernoulli). $X^{(n)} = (X_1, \ldots, X_n)$ où les X_i sont iid $\mathrm{Bern}(p)$. On a alors : $T(X^{(n)}) = \sum_{i=1}^n X_i$, sous $\mathbb{P}_\theta^{(n)}$, est de loi $\mathrm{Bin}(n, p)$.

Exemple 1.3 (Normale). $X^{(n)} = (X_1, \ldots, X_n)$ où les X_i sont iid $\mathcal{N}(\mu, \sigma^2)$ et où $\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} \in \Theta = \mathbb{R} \times \mathbb{R}_0^+ \subset \mathbb{R}^2$.

La statistique $T_1(X^{(n)}) = \sum_{i=1}^n X_i$, sous $P_\theta^{(n)}$, est de loi $\mathcal{N}(n\mu, n\sigma^2)$.

La statistique $T_2(X^{(n)}) = \frac{1}{n}\sum_{i=1}^n X_i$, sous $\mathbb{P}_\theta^{(n)}$, est de loi $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Exemple 1.4 (Uniforme). $X^{(n)} = (X_1, \ldots, X_n)$ où les X_i sont iid $\mathrm{Unif}(0, \theta)$, pour $\theta \in \Theta = \mathbb{R}_0^+ \subset \mathbb{R}$. On a donc $f_\theta^{X_i}(x) = \theta^{-1} I_{[0 \leqslant x \leqslant \theta]}$. Et donc :

$$F_\theta^{X_i}(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{\theta} & \text{si } 0 \leqslant x \leqslant \theta \\ 1 & \text{sinon} \end{cases}.$$

La statistique $T(X^{(n)}) = X_{(n)} = \max_{1 \leqslant k \leqslant n}\{X_k\}$ a pour fonction de répartition, sous $P_\theta^{(n)}$:

$$F_\theta^{(n)}(x) = \mathbb{P}[X_{(n)} \leqslant x] = \mathbb{P}[X_1 \leqslant x, X_2 \leqslant x, \ldots, X_n \leqslant x].$$

La seconde forme est plus agréable car on a une intersection d'événements indépendants. Donc :

$$F_\theta^{(n)}(x) = \prod_{i=1}^n \mathbb{P}[X_i \leqslant x] = \prod_{i=1}^n F_\theta^{X_i}(x) = \begin{cases} 0 & \text{si } x < 0 \\ \left(\frac{x}{\theta}\right)^n & \text{si } 0 \leqslant x \leqslant \theta \\ 1 & \text{sinon} \end{cases}.$$

On a alors la fonction de densité :

$$\begin{aligned} f_\theta^{X_{(n)}}(x) = \frac{d}{dx} F_\theta^{X_{(n)}}\Big|_x &= \begin{cases} 0 & \text{si } x < 0 \\ \frac{nx^{n-1}}{\theta^n} I_{[0 \leqslant x \leqslant \theta]} & \text{si } 0 \leqslant x \leqslant \theta \\ 0 & \text{sinon} \end{cases} \\ &= \frac{nx^{n-1}}{\theta^n} I_{[0 \leqslant x \leqslant \theta]}. \end{aligned}$$

Remarque. La loi échantillonnée n'est pas toujours possible à déterminer exactement analytiquement. Dans ce cas, on donne :

(i) les/des moments de la loi échantillonnée exacte ;

(ii) la loi échantillonnée asymptotique.

Et pour de grandes valeurs de n, la loi asymptotique donne une assez bonne approximation de la loi exacte.

Remarque. Ici, les termes *exact* et *asymptotique* s'opposent : on parle d'objet *exact* lorsque l'objet est connu pour n fixé, et d'objet *asymptotique* lorsque l'objet n'est connu que pour $n \to +\infty$.

Exemple 1.5. Voici un cas où on ne peut exprimer de loi exacte mais où il est possible d'exprimer une loi asymptotique. Soit $X^{(n)} = (X_1, \ldots, X_n)$ où les X_i sont iid F avec la fonction F telle que $\mathrm{Var}_F(X_i) = \sigma^2 < +\infty$ et donc $E_F(X_i) = \mu < +\infty$. On peut dès lors appliquer le théorème central limite (TCL) :

$$\sqrt{n}(\overline{X}^{(n)} - \mu) \xrightarrow[n \to +\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

Pour $n \gg$, on peut alors dire :

$$\overline{X}^{(n)} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right),$$

où le symbole $\approx$ se lit *est à peu près de même loi*.

On en conclut donc qu'avec n suffisamment grand, on peut approximer $\overline{X}^{(n)}$, même sans connaitre sa loi exacte.

1.3.1 Lemme de Fisher

Définition 1.18. La variable aléatoire Q est de loi χ^2 (chi-carrée) à $k(\in \mathbb{N}^*)$ degrés de liberté lorsque :

$$Q \stackrel{\mathcal{D}}{=} \sum_{i=1}^{k} Z_i^2,$$

où les Z_i sont iid $\mathcal{N}(0, 1)$ et où « $\stackrel{\mathcal{D}}{=}$ » veut dire *a la même distribution que*. Cela se note :

$$Q \sim \chi_k^2$$

Remarque. Si $Q \sim \chi_k^2$, alors :

$$f^Q(x) = \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} \exp\left(-\frac{x}{2}\right) I_{[x>0]},$$

où Γ est la fonction Gamma d'Euler définie par :

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} \exp(-t)\,\mathrm{d}t.$$

De plus, $\mathrm{Var}(Q) = 2k$, et $\mathbb{E}(Q) = k$.

On peut également noter que les χ^2 sont stables par la somme : si $Q_1 \sim \chi_{k_1}^2$ et $Q_2 \sim \chi_{k_2}^2$, alors :

$$Q_1 + Q_2 \sim \chi_{k_1+k_2}^2.$$

Lemme 1.19. *Soit $W = (W_1, \ldots, W_k)$ un vecteur de variables aléatoires, où $f^W : \mathbb{R}^k \to \mathbb{R}^+$ est la fonction de densité du vecteur W. Alors :*

1. $\mathbb{P}[W \in B] = \int_B f^W(x)\,\mathrm{d}x$;
2. *si $V = AW + b$ où A est une matrice $k \times k$ inversible, alors :*

$$f^V(v) = \left|\det A^{-1}\right| f^W\left(A^{-1}(v - b)\right).$$

Théorème 1.20 (Lemme de Fisher)**.** *Soient $X_1, \ldots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$ où $n \geqslant 2$. Alors :*

(i) $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$;

(ii) $\frac{ns^2}{\sigma^2} \sim \chi_{n-1}^2$;

(iii) $\overline{X} \perp\!\!\!\perp s^2$.

Démonstration. Posons $Z_i \coloneqq \frac{X_i - \mu}{\sigma}$ pour $i \in [\![1, n]\!]$. Puisque les X_i sont iid, les Z_i le sont également (même transformation appliquée à tous les X_i et chaque Z_i ne fait intervenir que le X_i correspondant). Notons que :

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_{i=1}^{n} (\sigma Z_i + \mu) = \sigma \overline{Z} + \mu,$$

où $\overline{Z}$ est la moyenne empirique des Z_i. Notons également que :

$$ns^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 = \sum_{i=1}^{n} \left((\sigma Z_i + \mu) - \left(\sigma \overline{Z} + \mu \right) \right)^2 = \sigma^2 \sum_{i=1}^{n} \left(Z_i - \overline{Z} \right)^2 = n\sigma^2 s_Z^2.$$

Il nous faut alors montrer que $\overline{Z} \sim \mathcal{N}\left(0, \frac{1}{n}\right)$ et $ns_Z^2 \sim \chi^2_{n-1}$, avec $\overline{Z} \perp\!\!\!\perp s_Z^2$.

Pour cela, on sait que le vecteur $Z^{(n)} = (Z_1, \dots, Z_n)$ a pour densité :

$$f^{Z^{(n)}}(z^{(n)}) = \prod_{i=1}^{n} f^{Z_i}(z_i) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}} \exp\left(\frac{z_i^2}{2} \right) \right) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\sum_{i=1}^{n} \frac{z_i^2}{2} \right) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2} \left\| z^{(n)} \right\|^2 \right).$$

Soit O une matrice orthogonale de dimension $n \times n$ telle que $\forall j \in [\![1, n]\!] : O_{1j} = \frac{1}{\sqrt{n}}$. On pose alors :

$$(Y_1, \dots, Y_n) = Y^{(n)} = OZ^{(n)}.$$

Puisque la matrice O est orthogonale, on sait que O^{-1} existe et que $|\det O| = \left|\det O^{-1}\right| = 1$. Par le lemme 1.19, on peut dire :

$$f^{Y^{(n)}}(y^{(n)}) = \left|\det O^{-1}\right| f^{Z^{(n)}}\left(O^{-1} y^{(n)}\right) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2} \left\| O^{-1} y^{(n)} \right\| \right) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2} \left\| y^{(n)} \right\| \right).$$

On a donc $f^{Y^{(n)}} = f^{Z^{(n)}}$, ce qui implique que les Y_i sont iid $\mathcal{N}(0, 1)$.

En particulier, $Y_1 = (Y^{(n)})_1 = (OZ^{(n)})_1 = \sum_{i=1}^{n} O_{1i} Z_i = \sum_{i=1}^{n} \frac{Z_i}{\sqrt{n}} = \sqrt{n}\overline{Z} \sim \mathcal{N}(0, 1)$. On peut alors en déduire que $\overline{Z} \sim \mathcal{N}(0, n^{-1})$.

Montrons alors que $ns_Z^2 \sim \chi^2_{n-1}$:

$$ns_Z^2 = \sum_{i=1}^{n} (Z_i - \overline{Z})^2 = \sum_{i=1}^{n} Z_i^2 - n(\overline{Z})^2 = \left\| Z^{(n)} \right\|^2 - (\sqrt{n}\overline{Z})^2 = \left\| Y^{(n)} \right\| - Y_1^2 = \sum_{i=2}^{n} Y_i^2.$$

Or, les Y_i sont $\mathcal{N}(0, 1)$. On a alors bien $ns_Z^2 \sim \chi^2_{n-1}$ (car la somme sur i commence à 2, il y a donc $(n-1)$ variables sommées).

De plus, puisque les Y_i sont indépendantes deux à deux, que $\overline{Z}$ ne dépend que de Y_1 et que ns_Z^2 ne dépend pas de Y_1, on sait que $\overline{Z} \perp\!\!\!\perp ns_Z^2$. $\square$

Chapitre 2

Estimation ponctuelle

2.1 Introduction

Considérons toujours un modèle statistique $\left(\mathbb{R}^n, \mathcal{B}\left(\mathbb{R}^n\right), \mathcal{P}^{(n)}\right)$ avec θ le paramètre vectoriel $\in \Theta \subset \mathbb{R}^k$.

Définition 2.1. Soit $g : \Theta \to \mathbb{R}^k$. Une statistique est appelée *estimateur de* $g(\theta)$ lorsqu'elle est à valeurs dans $g(\Theta)$.

Définition 2.2. Soit $\theta \in \Theta \subset \mathbb{R}^k$. Si $g : \Theta \to \mathbb{R}^m : \theta \mapsto (\theta_{\varphi(1)}, \dots, \theta_{\varphi(m)})$, on appelle les paramètres $\theta_{\varphi(i)}$ les paramètres *d'intérêt*, et on appelle les autres paramètres les paramètres *de nuisance*.

Exemple 2.1. Soient $X_1, \dots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$. On sait $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times {\mathbb{R}^+}_0 \subset \mathbb{R}^2$. Soit $g : \Theta \to \mathbb{R} : \theta \mapsto \mu$. μ est le paramètre d'intérêt et σ^2 est le paramètre de nuisance.

Remarque. Ne pas connaître le paramètre de nuisance induit une *nuisance* pour déterminer le paramètre d'intérêt.

2.2 Critères d'estimation

Remarque. Afin de définir les estimateurs convergents, il faut définir la notion de convergence, or il n'existe pas une manière canonique de la définir. Il existe donc plusieurs définitions de convergences différentes.

2.2.1 Définitions de convergence

Soient $Z^{(n)} = (Z_1, \dots, Z_n)$ définis pour $n \geqslant 1$ et sur $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 2.3. On dit que $Z^{(n)}$ converge *presque sûrement* (ou *stochastiquement*) vers Z lorsque :

$$\mathbb{P}\left[\left\{\omega \in \Omega \text{ t.q. } Z^{(n)}(\omega) \xrightarrow[n\to+\infty]{} Z(\omega)\right\}\right] = 1.$$

Cela se note :

$$Z^{(n)} \xrightarrow[n\to+\infty]{\text{p.s.}} Z.$$

Définition 2.4. On dit que $Z^{(n)}$ converge *en probabilités* vers Z lorsque :

$$\forall \varepsilon > 0 : \mathbb{P}\left[\left|Z^{(n)} - Z\right| > \varepsilon\right] \xrightarrow[n\to+\infty]{} 0.$$

Cela se note :
$$Z^{(n)} \xrightarrow[n\to+\infty]{\mathbb{P}} Z.$$

Définition 2.5. On dit que $Z^{(n)}$ converge *en* L_r pour $r \in \mathbb{N}^*$ vers Z lorsque :
$$\mathbb{E}\left[\left|Z^{(n)} - Z\right|^r\right] \xrightarrow[n\to+\infty]{} 0.$$

Cela se note :
$$Z^{(n)} \xrightarrow[n\to+\infty]{L_r} Z.$$

Remarque. Lorsque $r = 2$, on parle de convergence en moyenne quadratique.

Définition 2.6. On dit que $Z^{(n)}$ converge *en loi* (ou *en distribution*) vers Z lorsque :
$$\forall z \text{ point de continuité de } F^Z : F^{Z^{(n)}}(z) \xrightarrow[n\to+\infty]{} F^Z(z).$$

Cela se note :
$$Z^{(n)} \xrightarrow[n\to+\infty]{\mathcal{D}} Z.$$

Remarque. Ces définitions sont faites pour des variables aléatoires réelles mais peuvent être étendues à $\mathbb{R}^n$ en appliquant la convergence composante par composante.

2.2.2 Résultats élémentaires sur les convergences

Proposition 2.7. *Les convergences sont induites mutuellement par les assertions suivantes :*

1. *si* $Z^{(n)} \xrightarrow[n\to+\infty]{p.s.} Z$*, alors* $Z^{(n)} \xrightarrow[n\to+\infty]{\mathbb{P}} Z$ *;*
2. *si* $Z^{(n)} \xrightarrow[n\to+\infty]{\mathbb{P}} Z$*, alors* $Z^{(n)} \xrightarrow[n\to+\infty]{\mathcal{D}} Z$ *;*
3. *si* $Z^{(n)} \xrightarrow[n\to+\infty]{L_r} Z$*, alors* $Z^{(n)} \xrightarrow[n\to+\infty]{\mathbb{P}} Z$*.*

Théorème 2.8. *Les convergences presque sûre, en probabilités, et en loi sont stables par transformations continues.*

Théorème 2.9. *Notons* $\to$ *une convergence soit presque sûre, soit en probabilités. Si* $Z^{(n)} \to Z$*, et* $Y^{(n)} \to Y$*, alors :*

(i) $Z^{(n)} + Y^{(n)} \to Z + Y$;
(ii) $Z^{(n)} \cdot Y^{(n)} \to Z \cdot Y$;
(iii) *si* $\mathbb{P}[Y^{(n)} = 0] = 0$*, alors* $\frac{Z^{(n)}}{Y^{(n)}} \to \frac{Z}{Y}$.

Lemme 2.10 (Lemme de Slutzky)**.** *Si* $Z^{(n)} \xrightarrow[n\to+\infty]{\mathcal{D}} Z$*, et* $Y^{(n)} \xrightarrow[n\to+\infty]{\mathcal{D}} c \neq 0$*, alors :*

(i) $Z^{(n)} + Y^{(n)} \xrightarrow[n\to+\infty]{\mathcal{D}} Z + c$;
(ii) $Z^{(n)} \cdot Y^{(n)} \xrightarrow[n\to+\infty]{\mathcal{D}} Z \cdot c$;
(iii) $\frac{Z^{(n)}}{Y^{(n)}} \xrightarrow[n\to+\infty]{\mathcal{D}} \frac{Z}{c}$.

Théorème 2.11 (Loi forte des grands nombres)**.** *Soient* $Z_1, Z_2, \ldots$ *iid avec* $\mathbb{E}\left[|Z_1|\right] < +\infty$*. Alors :*
$$\overline{Z}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} Z_k \xrightarrow[n\to+\infty]{p.s.} \mu = \mathbb{E}[Z_1].$$

Théorème 2.12 (Loi faible des grands nombres)**.** *Soient* $Z_1, Z_2, \ldots$ *iid avec* $\mathbb{E}\left[|Z_1|\right] < +\infty$*. Alors :*
$$\overline{Z}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} Z_k \xrightarrow[n\to+\infty]{\mathbb{P}} \mu = \mathbb{E}[Z_1].$$

Théorème 2.13 (Théorème central limite (TCL))**.** *Soient* $Z_1, Z_2, \ldots$ *iid, avec* $\mathbb{E}[Z_1^2] < +\infty$*. Alors :*

$$\sqrt{n}\left(Z^{(n)} - \mu\right) \xrightarrow[n\to+\infty]{\mathcal{D}} W,$$

où :

$$W \sim \mathcal{N}(0, \sigma^2),$$

avec $\sigma^2 = \mathrm{Var}(Z_1)$.

2.2.3 Estimateurs convergents

Définition 2.14. Un estimateur $T^{(n)}(X^{(n)})$ de $g(\theta)$ est dit *faiblement convergent* lorsque :

$$\forall\theta \in \Theta : T^{(n)}(X^{(n)}) \xrightarrow[n\to+\infty]{\mathbb{P}} g(\theta) \qquad \text{sur } \mathbb{P}_\theta^{(n)}.$$

Définition 2.15. Un estimateur $T^{(n)}(X^{(n)})$ de $g(\theta)$ est dit *fortement convergent* lorsque :

$$\forall\theta \in \Theta : T^{(n)}(X^{(n)}) \xrightarrow[n\to+\infty]{\text{p.s.}} g(\theta) \qquad \text{sur } \mathbb{P}_\theta^{(n)}.$$

Exemple 2.2. Soient $X_1, \ldots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$. Prenons $g(\theta) = \mu$ et $T^{(n)}(X^{(n)}) = \overline{X}$. On a bien :

$$T^{(n)}(X^{(n)}) = \overline{X}^{(n)} \xrightarrow[n\to+\infty]{\text{p.s.}} \mu = \mathbb{E}[X_1] \qquad \text{sur } \mathbb{P}_{\mu,\sigma^2}^{(n)}.$$

$T^{(n)}(X^{(n)})$ est donc un estimateur fortement convergent.

Prenons maintenant $T_2^{(n)}(X^{(n)}) = s^2$. On ne peut pas appliquer la loi des grands nombres car les variables aléatoires $(X_i - \overline{X})^2$ sommées ne sont pas indépendantes. On a alors :

$$\begin{aligned} T_2^{(n)}(X^{(n)}) = s^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 &= \frac{1}{n}\sum_{i=1}^n \left((X_i-\mu)^2 + (\mu - \overline{X})^2 - 2(X_i - \overline{X})(\overline{X} - X_i)\right) \\ &= \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 - (\overline{X} - \mu)^2, \end{aligned}$$

où $\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbb{E}\left[(X_i - \mu)^2\right] = \sigma^2$, par la loi forte des grands nombres, et $(\overline{X} - \mu) \xrightarrow[n\to+\infty]{\text{p.s.}} 0$.
Donc, par le théorème 2.9, on a $T_2^{(n)}(X^{(n)}) \xrightarrow[n\to+\infty]{\text{p.s.}} \sigma^2$

Remarque. On a également :

$$\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{n}{n-1}s^2 \xrightarrow[n\to+\infty]{\text{p.s.}} 1 \cdot \sigma^2.$$

Exemple 2.3. Soient $X_1, \ldots, X_n$ iid $\mathrm{Unif}(0, \theta)$, avec $\theta \in \Theta \subset \mathbb{R}$. On veut estimer $g(\theta) = \theta$. L'estimateur $T^{(n)}(X^{(n)}) = \overline{X}$ n'est pas un estimateur convergent car :

$$\overline{X} \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbb{E}_\theta(X_1) = \frac{\theta}{2} \neq \theta \qquad \text{sur } \mathbb{P}_\theta^{(n)}.$$

Par contre, si on prend $T_2^{(n)}(X^{(n)}) = 2\overline{X}$, on a :

$$2\overline{X} \xrightarrow[n\to+\infty]{\text{p.s.}} 2\mathbb{E}_\theta(X_1) = 2\frac{\theta}{2} = \theta \qquad \text{sur } \mathbb{P}_\theta^{(n)}.$$

Si on prend $T_3^{(n)}(X^{(n)}) = X_{(n)}$, à savoir l'observation maximale, on a :

$$F_\theta^{X_{(n)}}(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x^n}{\theta^n} & \text{si } 0 \leqslant x \leqslant \theta \\ 1 & \text{sinon} \end{cases}.$$

Posons donc $\varepsilon > 0$. On calcule :

$$\begin{aligned} \mathbb{P}_\theta^{(n)}\left[\left|X_{(n)} - \theta\right| > \varepsilon\right] &= \mathbb{P}_\theta^{(n)}\left[X_{(n)} \leqslant \theta - \varepsilon\right] + \mathbb{P}_\theta^{(n)}\left[X_{(n)} \geqslant \theta + \varepsilon\right] = \mathbb{P}_\theta^{(n)}\left[X_{(n)} \leqslant \theta - \varepsilon\right] = F_\theta^{X_{(n)}}(\theta - \varepsilon) \\ &= \begin{cases} 0 & \text{si } \varepsilon \geqslant \theta \\ \left(\frac{\theta - \varepsilon}{\theta}\right)^n & \text{si } 0 < \varepsilon < \theta \end{cases} \longrightarrow 0 \end{aligned}$$

on a alors convergence en probabilité de $X_{(n)}$ vers θ. On en déduit que $T_3^{(n)}(X^{(n)})$ est un estimateur faiblement convergent.
Remarque. L'estimateur $\frac{n+1}{n}X_{(n)}$ est également faiblement convergent.
Remarque. Il n'est pas toujours possible de s'en sortir en invoquant le TCL ou la loi des grands nombres pour déterminer la convergence d'un estimateur. Prenons par exemple $X_1, \ldots, X_n$ iid de densité :

$$f^X(x) = \frac{1}{\pi(1 + (x - \theta)^2)}.$$

On a effectivement $\mathbb{E}[X_1] = +\infty$. En réalité :

$$\neg\left(\overline{X}^{(n)} \xrightarrow[n \to +\infty]{\mathbb{P}} \theta \qquad \text{sur } P_\theta^{(n)}\right),$$

mais bien :

$$\overline{X}^{(n)} \xrightarrow[n \to +\infty]{\mathcal{D}} X_1$$

2.3 Estimateur exhaustif

Définition 2.16. Soit $T^{(n)}(X^{(n)})$, une statistique. On la dit *exhaustive* lorsque :

$$\forall B \in \mathcal{B}(\mathbb{R}^n) : \forall t \in T^{(n)}(\mathbb{R}^n) : \mathbb{P}_\theta^{(n)}\left[X^{(n)} \in B | T^{(n)}(X^{(n)}) = t\right] \text{ ne dépend pas de } \theta.$$

Remarque. Puisque l'on travaille sur des modèles paramétriques $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, il existe toujours un $B \in \mathcal{B}(\mathbb{R}^n)$ tel que $\mathbb{P}_\theta[X^{(n)} \in B]$ dépende de θ.
Remarque. On peut en comprendre qu'une statistique est exhaustive si la valeur prise par $T^{(n)}(X^{(n)})$ donne toutes les informations contenues par $X^{(n)}$ sur θ.
Exemple 2.4. La statistique identité $x^{(n)} \mapsto x^{(n)}$ est une statistique exhaustive car :

$$\mathbb{P}\left[X^{(n)} \in B | X^{(n)} = x^{(n)}\right] = \begin{cases} 1 & \text{si } x^{(n)} \in B \\ 0 & \text{sinon} \end{cases}.$$

Exemple 2.5. Prenons $X_1, \ldots, X_n$ iid $\mathrm{Bern}(p)$ avec la statistique $T^{(n)}(X^{(n)}) = \sum_{i=1}^n X_i$. Pour évaluer la probabilité :

$$\mathbb{P}_p^{(n)}\left[X^{(n)} \in \{x^{(n)}\} | \sum_{i=1}^n X_i = t\right],$$

on est en présence d'une binomiale. Dès lors, si $\sum_{i=1}^n X_i \neq t$, alors la probabilité est nulle. Sinon, la probabilité est $\frac{1}{\binom{n}{t}}$ car il y a $\binom{n}{t}$ moyens d'avoir n observations dont t valant 1 et $n-t$ valant 0. Ces probabilités ne dépendent donc pas de θ, la statistique $T^{(n)}(X^{(n)}) = \sum_{i=1}^n X_i$ est donc une statistique exhaustive.
Remarque. Si $T^{(n)}(X^{(n)})$ est une statistique bijective, alors elle est exhaustive. Cependant, les estimateurs intéressants sont ceux qui « réduisent » l'information de manière à ce qu'elles soient plus facilement analysables.
Définition 2.17. Soit $X^{(n)} = (X_1, \dots, X_n)$. On appelle la *fonction de vraisemblance de* $X^{(n)}$ la fonction :

$$L_\theta^{(n)} : \mathbb{R}^n \to \mathbb{R} : x^{(n)} \mapsto \begin{cases} \mathbb{P}[X^{(n)} = x^{(n)}] & \text{si } X^{(n)} \text{ est de loi discrète} \\ f_\theta^{X^{(n)}}(x^{(n)} & \text{sinon} \end{cases}.$$

Remarque. Dans le cas de variables $X_1, \dots, X_n$ iid, la fonction de vraisemblance correspond toujours à un produit :

$$L_\theta^{(n)}(X^{(n)}) = \mathbb{P}[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n] \overset{\perp\!\!\!\perp}{=} \prod_{i=1}^n \mathbb{P}[X_i = x_i].$$

Théorème 2.18 (Critère de factorisation de Neymann-Fisher)**.** *Dans un modèle paramétrique* $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, *une statistique* $T^{(n)}(X^{(n)})$ *est exhaustive* ***si et seulement si*** *pour tout* $\theta \in \Theta$, *la fonction de vraisemblance* $L_\theta^{(n)}(X^{(n)})$ *est factorisable sous la forme :*

$$L_\theta^{(n)}(X^{(n)}) = \left(g_\theta \circ T^{(n)}\right)(X^{(n)})h(X^{(n)}),$$

et ce $\mathbb{P}_\theta^{(n)}$*-presque sûrement.*
Remarque. Dans cette factorisation, la fonction h ne peut dépendre de θ, et seule la fonction g_θ peut en dépendre, mais uniquement par l'intermédiaire de $T^{(n)}$.
Exemple 2.6. En reprenant l'exemple d'au-dessus : $X_1, \dots, X_n$ iid $\mathrm{Bern}(p)$ et $T^{(n)}(X^{(n)}) = \sum_{i=1}^n X_i$, on a :

$$L_\theta^{(n)}(x^{(n)}) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}.$$

Dès lors, en posant $g_\theta(x) = p^x(1-p)^{n-x}$ et $h(x^{(n)}) = 1$, on a bien une factorisation de Neymann-Fisher, ce qui implique que la statistique est exhaustive.
Remarque. Pour chaque statistique exhaustive, il en existe une infinité définies à bijection près. En effet, si $T(X^{(n)})$ est une statistique exhaustive et si H est une fonction bijective quelconque, alors :

$$L_\theta^{(n)}(x^{(n)}) = g_\theta(T(x^{(n)})h(x^{(n)}) = (g_\theta \circ H^{-1} \circ H \circ T)(x^{(n)})h(x^{(n)}).$$

La fonction $H \circ T$ est donc également une statistique exhaustive.
Remarque. Le critère précédent peut également donner une manière de *deviner* des statistiques exhaustives. Prenons par exemple $X^{(n)} = (X_1, \dots, X_n)$ iid $\mathcal{N}(\mu, \sigma^2)$. Si $\theta = (\mu, \sigma^2) \in \Theta \subset \mathbb{R}^2$, on peut écrire :

$$\begin{aligned} L_\theta^{(n)}(x^{(n)}) = \prod_{i=1}^n f_\theta^{X_i}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{2}{2\sigma^2}\sum_{i=1}^n x_i^2 - \frac{n\mu}{2\sigma^2} + \frac{\mu}{\sigma^2}\sum_{i=1}^n x_i\right). \end{aligned}$$

Dès lors, en prenant $T(X^{(n)}) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2\right)$, on a bien une statistique exhaustive. Alors, de même, on peut dire que $\left(\frac{1}{n}\sum_{i=1}^n x_i, \frac{1}{n}\sum_{i=1}^n x_i^2\right)$ est un estimateur exhaustif (composition avec une bijection).

On peut également dire que $\left(\frac{1}{n}\sum_{i=1}^n x_i, \frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2\right)$ est un estimateur exhaustif (même argument). Or ce dernier vecteur correspond à $(\overline{X}, s^2)$.

En prenant cette fois $X^{(n)} = (X_1, \dots, X_n)$ iid $\mathrm{Unif}(0, \theta)$, on peut à nouveau construire des statistiques exhaustives :

$$L_\theta^{(n)}(x^{(n)}) = \prod_{i=1}^n f_\theta^{X_i}(x_i) = \prod_{i=1}^n \frac{1}{\theta} I_{[0 \leqslant x_i \leqslant \theta]} = \frac{1}{\theta^n} I_{[0 \leqslant x_1, \dots, x_n \leqslant \theta]}.$$

Cela garantit bien que $x^{(n)}$ est une statistique exhaustive. Mais de plus, par commutativité du produit :

$$L_\theta^{(n)}(x^{(n)}) = \prod_{i=1}^n f_\theta^{X_{(i)}}(x_i) = \frac{1}{\theta^n} I_{[0 \leqslant x_{(1)}, \dots, x_{(n)} \leqslant \theta)]}.$$

On a donc que la statistique d'ordre est une statistique exhaustive. Or, la condition $0 \leqslant x_{(1)}, \dots, x_{(n)} \leqslant \theta$ revient à la condition $0 \leqslant x_{(1)}, x_{(n)} \leqslant \theta$. La statistique $(x_{(1)}, x_{(n)})$ est donc également une statistique exhaustive. Pour aller plus loin, décomposant la fonction caractéristique $I_{[0 \leqslant x_{(1)}, x_{(n)} \leqslant \theta)]}$ en $I_{[0 \leqslant x_{(1)}]} I_{[x_{(n)} \leqslant \theta]}$, on peut poser $h(x^{(n)}) = I_{[0 \leqslant x_{(1)}]}$, ce qui amène à une nouvelle statistique exhaustive : $x^{(n)} \mapsto x_{(n)}$.

À chaque étape du raisonnement, la statistique exhaustive contient *de moins en moins d'information* générale mais conserve l'information sur θ qui est donc en quelque sorte contenue dans $x_{(n)}$.

2.3.1 Estimateurs non biaisés

En se situant toujours dans un modèle statistique $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, on veut estimer $g(\theta)$, avec $g : \Theta \to \mathbb{R}^m$.

Définition 2.19. Un estimateur $T^{(n)}(X^{(n)})$ de $g(\theta)$ est dit *non biaisé* lorsque :

$$\forall \theta \in \Theta : \mathbb{E}_\theta[T^{(n)}(X^{(n)})] = g(\theta).$$

Définition 2.20. Un estimateur $T^{(n)}(X^{(n)})$ de $g(\theta)$ est dit *asymptotiquement non biaisé* lorsque :

$$\forall \theta \in \Theta : \mathbb{E}_\theta[T^{(n)}(X^{(n)})] \xrightarrow[n \to +\infty]{} g(\theta).$$

Définition 2.21. Le *biais* d'un estimateur $T^{(n)}(X^{(n)})$ est la quantité :

$$b_\theta^{(n)} = b^{(n)}(\theta) = \mathbb{E}_\theta[T^{(n)}(X^{(n)})] - g(\theta).$$

Remarque. On remarque donc qu'un estimateur non biaisé a un biais de 0 et qu'un estimateur asymptotiquement non biaisé a un biais qui tend vers 0 pour n tendant vers $+\infty$.

Exemple 2.7. $X_1, \dots, X_n$ iid $\mathrm{Unif}(0, \theta)$, avec $\theta \in \Theta = \mathbb{R}^+{}_0 \subset \mathbb{R}$.

— Si $T^{(n)}(X^{(n)}) = 2\overline{X}$, on a :

$$\mathbb{E}_\theta^{(n)}\left[2\overline{X}\right] = \frac{2}{n}\mathbb{E}_\theta^{(n)}\left[\sum_{i=1}^n X_i\right] = \frac{2}{n}\sum_{i=2}^n \mathbb{E}_\theta[X_i] = \frac{2}{n}\frac{n\theta}{2} = \theta.$$

— Si $T^{(n)}(X^{(n)}) = X_{(n)}$, on a :

$$\mathbb{E}_\theta^{(n)}[X_{(n)}] = \int_{\mathbb{R}} x f_\theta^{X_{(n)}}(x)\,dx = \int_0^\theta x \frac{n x^{n-1}}{\theta^n}\,dx = \frac{n}{\theta^n}\int_0^\theta x^n\,dx = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n}{n+1}\theta.$$

On en déduit que $T^{(n)}(X^{(n)})$ est biaisé car il *vise en moyenne trop à gauche* et a un biais de $-\frac{\theta}{n+1}$. Il est cependant asymptotiquement non biaisé car $\frac{n}{n+1} \xrightarrow[n \to +\infty]{} 1$.

Remarque. Si le biais d'un estimateur est négatif, alors c'est que l'estimateur sous-estime en moyenne, alors que si le biais est positif, c'est que l'estimateur surestime.
Remarque. On peut tout de même dire que l'estimateur $\frac{n+1}{n}X_{(n)}$ est non biaisé car :

$$\mathbb{E}_\theta^{(n)}\left[\frac{n+1}{n}X_{(n)}\right] = \frac{n+1}{n}\mathbb{E}_\theta^{(n)}[X_{(n)}] = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta.$$

Exemple 2.8. Soient $X_1, \dots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$, avec $\theta = (\mu, \sigma^2)$.

— pour tout $\theta \in \Theta$, on a : $\mathbb{E}_\theta[\overline{X}] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = \frac{n\mu}{n} = \mu$. $\overline{X}$ est donc un estimateur sans biais de μ ;

— $\mathbb{E}_\theta[s^2] = \mathbb{E}_\theta[X_1^2] - \mathbb{E}_\theta[\overline{X}^2]$. On remarque que pour toute variable aléatoire Z, on a :

$$\mathrm{Var}(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 \qquad \iff \qquad \mathbb{E}[Z^2] = \mathrm{Var}(Z) + \mathbb{E}[Z]^2.$$

On peut donc remplacer dans la formule de l'espérance de s^2, et on obtient :

$$\mathbb{E}_\theta[s^2] = \mathrm{Var}_\theta(X_1) + \mathbb{E}(X_1)^2 - \mathrm{Var}_\theta(\overline{X}) - \mathbb{E}(\overline{X})^2.$$

Or on sait $\mathbb{E}_\theta(X_1) = \mathbb{E}_\theta(\overline{X}) = \mu$. On a donc :

$$\mathbb{E}_\theta[s^2] = \mathrm{Var}_\theta(X_1) - \mathrm{Var}_\theta(\overline{X}) = \sigma^2 - \frac{1}{n^2}\mathrm{Var}_\theta\left(\sum_{i=1}^n X_i\right) = \sigma^2 - \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}_\theta(X_i) = \sigma^2 - \frac{n\sigma^2}{n^2} = \frac{n-1}{n}\sigma^2.$$

On en conclut que l'estimateur s^2 est biaisé pour σ^2 et de biais $\frac{-\sigma^2}{n} \xrightarrow[n\to+\infty]{} 0$.

Notons alors $S^2 := \frac{n}{n-1}s^2$. On a que $\mathbb{E}_\theta[S^2] = \sigma^2$, et donc S^2 est un estimateur sans biais.

Remarque. Le non biais est une propriété fragile. Soient $X_1, \dots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$ où l'on veut estimer $g(\theta) = \sigma$ (et pas σ^2).

Que vaut $\mathbb{E}_\theta[S]$? On sait $\mathrm{Var}_\theta(S) = \mathbb{E}_\theta[S^2] - \mathbb{E}_\theta[S]^2 = \sigma^2 - \mathbb{E}_\theta[S]^2$. Or $\mathrm{Var}_\theta(S) \gneqq 0$, et donc $\sigma^2 \gneqq \mathbb{E}_\theta[S]^2$. On en déduit $\mathbb{E}_\theta[S] \lneqq \sigma$.
Remarque. De plus, il n'existe pas toujours d'estimateur sans biais. Soit $X_1 \sim \mathrm{Bern}(p)$. On veut estimer $g(p) = p^2 \in [0,1]$. L'estimateur $T(X^{(n)})$ est entièrement déterminé par $T(0)$ et $T(1)$. Imposons donc pour tout $p \in [0,1]$: $\mathbb{E}_\theta[T(X_1)] = p^2$. On a donc :

$$p^2 = \mathbb{E}_\theta[T(X_1)] = T(0)\mathbb{P}[X_1 = 0] + T(1)\mathbb{P}[X_1 = 1] = T(0)(1-p) + T(1)p.$$

En réarrangeant cette équation du second degré, on obtient :

$$\forall p \in [0,1] : p^2 + (T(0) - T(1))p - T(0) = 0.$$

Or une telle équation ne peut avoir que 2 racines tout au plus, et ici, une infinité non-dénombrable est requise. Il n'existe donc pas de telle fonction T, et donc par extension, il n'existe pas d'estimateur sans biais de p^2.
Exemple 2.9. Prenons $X_1, \dots, X_n$ iid $\mathcal{N}(\mu, 1)$, $g(\theta) = g(\mu) = \mu$. Prenons $T^{(n)}(X^{(n)}) = \overline{X^{(n)}} + Y$, où $Y = \pm 10^9$, avec probabilité $\frac{1}{2}$ pour chaque. On a alors :

$$\mathbb{E}_\mu[T^{(n)}(X^{(n)})] = \mathbb{E}_\mu[\overline{X^{(n)}}] + \mathbb{E}_\mu[Y] = \mu + 0 = \mu.$$

L'estimateur $T^{(n)}(X^{(n)})$ est donc sans biais, mais est très mauvais : sur-/sous-estime toujours à $\simeq 10^9$ près.

2.3.2 Estimateurs à dispersion minimale

Définition 2.22. L'erreur quadratique moyenne d'un estimateur $T^{(n)}(X^{(n)})$ de $g(\theta)$ ($\in \mathbb{R}$) est la quantité :

$$\mathrm{MSE}_\theta[T^{(n)}(X^{(n)})] := \mathbb{E}_\theta\left[\left|T^{(n)}(X^{(n)}) - g(\theta)\right|^2\right].$$

Définition 2.23. On appelle l'erreur absolue moyenne la quantité :

$$\mathsf{MAE}_\theta[T^{(n)}(X^{(n)})] = \mathbb{E}_\theta\left[\left|T^{(n)}(X^{(n)}) - g(\theta)\right|\right]$$

Remarque.

$$\begin{aligned}\mathrm{MSE}_\theta[T^{(n)}(X^{(n)})] &= \mathbb{E}_\theta\left[\left|T^{(n)}(X^{(n)}) - g(\theta)\right|^2\right] = \mathrm{Var}_\theta\left[T^{(n)}(X^{(n)}) - g(\theta)\right] - \mathbb{E}_\theta\left[T^{(n)}(X^{(n)}) - g(\theta)\right]^2 \\ &= \mathrm{Var}_\theta[T^{(n)}(X^{(n)})] - b_\theta^{(n)}(T^{(n)}(X^{(n)}))^2.\end{aligned}$$

Puisque $\mathrm{Var}_\theta(T^{(n)}(X^{(n)})) \geqslant 0$ et $b_\theta^{(n)}(T^{(n)}(X^{(n)})) \geqslant 0$, pour avoir $\mathrm{MSE}[T^{(n)}(X^{(n)})] \xrightarrow[n\to+\infty]{} 0$, il faut :

$$\begin{cases} \mathrm{Var}_\theta(T^{(n)}(X^{(n)})) \xrightarrow[n\to+\infty]{} 0, \\ b_\theta^{(n)}(T^{(n)}(X^{(n)})) \xrightarrow[n\to+\infty]{} 0. \end{cases}$$

Cela dit que sous $\mathbb{P}_\theta^{(n)}$, on a $T^{(n)}(X^{(n)}) \xrightarrow{L^2} g(\theta)$ (et également $T^{(n)}(X^{(n)}) \xrightarrow[n\to+\infty]{\mathbb{P}} g(\theta)$).

On peut également dire que si $\mathrm{MSE}_\theta[T^{(n)}(X^{(n)})] \to 0$, alors $T^{(n)}(X^{(n)})$ est un estimateur faiblement convergent.

Remarque. Le MSE sert d'« arbitrage » entre la variance et le biais. Cet arbitrage est bien souvent nécessaire car il arrive fréquemment que baisser le biais (respectivement la variance) augmente la variance (respectivement le biais).

Exemple 2.10. $X_1, \ldots, X_n$ iid $\mathrm{Unif}(0, \theta)$. Prenons $\theta \in \Theta = \mathbb{R}^+{}_0 \subset \mathbb{R}$. Prenons $T_1^{(n)}(X^{(n)}) = X_{(n)}$. On sait que :

$$\mathbb{E}_\theta[T_1^{(n)}(X^{(n)})] = \frac{n}{n+1}\theta,$$

ce qui nous permet de calculer la variance :

$$\mathrm{Var}_\theta(X_{(n)}) = \mathbb{E}_\theta(X_{(n)}^2) - \mathbb{E}_\theta(X_{(n)})^2,$$

où :

$$\mathbb{E}_\theta(X_{(n)}^2) = \int_{\mathbb{R}} x^2 f^{X_{(n)}}(x)\,dx = \int_0^\theta x^2 \frac{nx^{n-1}}{\theta^2}\,dx = \frac{n}{\theta^n}\int_0^\theta x^{n+1}\,dx = \frac{n}{\theta^n}\left[\frac{x^{n+2}}{n+2}\right]_0^\theta = \frac{n\theta^2}{n+2}.$$

On trouve finalement la variance donnée par :

$$\frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} = n\theta^2\left(\frac{(n+1)^2 - n(n+2)}{(n+2)(n+1)^2}\right) = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

On trouve finalement une erreur quadratique moyenne de :

$$\mathrm{MSE}_\theta[T_1^{(n)}(X^{(n)})] = \frac{2\theta^2}{(n+1)(n+2)} \xrightarrow[n\to+\infty]{} 0.$$

En prenant $T_2^{(n)}(X^{(n)}) = \frac{n+1}{n}X_{(n)}$, on annule le biais, mais on augmente la variance :

$$\mathrm{Var}_\theta[T_2^{(n)}(X^{(n)})] = \left(\frac{n+1}{n}\right)^2 \mathrm{Var}_\theta[X_{(n)}] = \frac{\theta^2}{n(n+2)}.$$

On trouve alors :

$$\forall \theta \in \Theta : \mathrm{MSE}_\theta[T_2^{(n)}(X^{(n)})] = \frac{\theta^2}{n(n+2)} \leqslant \mathrm{MSE}_\theta[T_1^{(n)}(X^{(n)})].$$

Dans le cas présent, annuler le biais donne une erreur quadratique moyenne plus faible. Donc ici, la variance a « moins augmenté que le biais n'a diminué ».

Exemple 2.11. Prenons maintenant $X_1, \dots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$ et $g(\theta) = \sigma^2$.

Pour $T_1^{(n)}(X^{(n)}) = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$, on trouve :

$$\mathrm{MSE}_\theta[T_1^{(n)}(X^{(n)})] = \frac{2(n-1)\sigma^4}{n^2}.$$

À nouveau, il existe un estimateur sans-biais :

$$T_2^{(n)}(X^{(n)}) = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{n-1}{n} T_1^{(n)}(X^{(n)}).$$

Dans ce cas, on trouve :

$$\forall \theta \in \Theta : \mathrm{MSE}_\theta[T_2^{(n)}(X^{(n)})] = \frac{2\sigma^4}{n-1} \geqslant \mathrm{MSE}_\theta[T_1^{(n)}(X^{(n)})].$$

Remarque. Dans les deux exemples ci-dessus, l'inégalité est stricte pour $n \geqslant 2$.

De plus, l'inégalité tient pour tout $\theta \in \Theta$. Cela permet de définir qu'un estimateur est meilleur que l'autre. Avoir deux estimateurs tels que l'erreur quadratique moyenne de l'un est meilleure que l'autre pour certaines valeurs de $g(\theta)$, et inversement pour d'autres ne permet pas de dire que l'un est meilleur que l'autre car la véritable valeur de $g(\theta)$ n'est pas connue. On ne peut donc pas savoir quel estimateur choisir.

Définition 2.24. Soit $\mathcal{C}$, une classe d'estimateurs de $g(\theta)$. On dit que $T_*^{(n)}$ est à erreur quadratique moyenne minimale dans $\mathcal{C}$ lorsque :

- $T_*^{(n)} \in \mathcal{C}$;
- $\forall T^{(n)} \in \mathcal{C} : \forall \theta \in \Theta : \mathrm{MSE}_\theta[T_*^{(n)}] \leqslant \mathrm{MSE}_\theta[T^{(n)}]$.

Remarque. On peut prendre $\mathcal{C} = \{T^{(n)} \text{ t.q. } \mathbb{E}_\theta[{T^{(n)}}^2] < +\infty\}$, mais on n'en prend qu'un sous-ensemble strict bien souvent car dans un tel $\mathcal{C}$, on peut prendre $T_{\theta_0}^{(n)}$, pour un certain $\theta_0 \in \Theta$ défini par $T_{\theta_0}^{(n)}(X^{(n)}) = g(\theta_0)$. On trouve donc :

$$\mathrm{MSE}_\theta[T_{\theta_0}^{(n)}(X^{(n)})] = \mathbb{E}_\theta[(g(\theta_0))g(\theta))^2] = (g(\theta_0) - g(\theta))^2.$$

La fonction d'erreur quadratique moyenne est donc une parabole qui s'annule en $\theta_0 = \theta$. Pour avoir un tel $T_*^{(n)}$, il faut que $\mathrm{MSE}_{\theta_0}[T_*^{(n)}] \leqslant \mathrm{MSE}_{\theta_0}[T_{\theta_0}^{(n)}] = 0$. Et ce, pour tout $\theta_0 \in \Theta$. Il faut donc avoir :

$$\forall \theta \in \Theta : T_*^{(n)} = g(\theta) \qquad \text{sous } \mathbb{P}_\theta^{(n)}.$$

Pour résoudre ce problème, il faut donc réduire $\mathcal{C}$ afin de ne plus avoir de tels estimateurs *pathologiques*.

Définition 2.25. Soit $Z = (Z_1, \dots, Z_m)^\mathsf{T}$, un vecteur aléatoire (vecteur de variables aléatoires). On définit :

$$\mathbb{E}_\theta(Z) \coloneqq (\mathbb{E}_\theta[Z_1], \dots, \mathbb{E}_\theta[Z_m])^\mathsf{T}.$$

On définit également :

$$\mathrm{Var}_\theta[Z] = \mathbb{E}\left[(Z - \mathbb{E}_\theta[Z])(Z - \mathbb{E}_\theta[Z])^\mathsf{T}\right] \in \mathbb{R}^{m \times m}.$$

Remarque. On observe que :

$$\mathrm{Var}_\theta[Z]_{ij} = \mathbb{E}_\theta\left[(Z-\mathbb{E}_\theta[Z])_i(Z-\mathbb{E}_\theta[Z])_j^T\right] = \mathbb{E}_\theta\left[(Z_i-\mathbb{E}_\theta[Z_i])(Z_j-\mathbb{E}_\theta[Z_j])\right] = \mathrm{Cov}[Z_i,Z_j].$$

Cela veut dire :

$$\mathrm{Var}_\theta[Z] = \begin{pmatrix} \mathrm{Var}_\theta[Z_1] & \mathrm{Cov}[Z_1,Z_2] & \dots & \mathrm{Cov}[Z_1,Z_n] \\ \mathrm{Cov}[Z_2,Z_1] & \mathrm{Var}_\theta[Z_2] & \ddots & \mathrm{Cov}[Z_2,Z_n] \\ \vdots & \ddots & \ddots & \vdots \\ \mathrm{Cov}[Z_n,Z_1] & \dots & \dots & \mathrm{Var}_\theta[Z_n] \end{pmatrix}$$

On appelle cette matrice la *matrice de variance/covariance* (pour des raisons évidentes). On voit donc que cette matrice est symétrique car $\mathrm{Cov}[Z_i,Z_j] = \mathrm{Cov}[Z_j,Z_i]$.

Définition 2.26. Soit A une matrice. On dit que $A \geqslant 0$ si la forme bilinéaire associée est semi-définie positive.

Proposition 2.27. *Soit Z un vecteur aléatoire, et soient $A \in \mathbb{R}^{k\times m}$, $b \in \mathbb{R}^k$. Alors :*

- $\mathbb{E}_\theta[AZ+b] = A\mathbb{E}_\theta[Z]+b$;
- $\mathrm{Var}_\theta[AZ+b] = A\,\mathrm{Var}_\theta[Z]A^T$;
- $\mathrm{Var}_\theta[Z] \geqslant 0$.

Démonstration.

- On observe que $\mathbb{E}_\theta[AZ+b] = \left(\mathbb{E}_\theta[(AZ)_i+b_i]\right)_i$, où :

$$\begin{aligned}\mathbb{E}_\theta[(AZ)_i+b_i] = \mathbb{E}_\theta[(AZ)_i]+b_i = \mathbb{E}_\theta\left[\sum_{j=1}^k A_{ij}Z_j\right]+b_i &= \sum_{j=1}^k \mathbb{E}_\theta[A_{ij}Z_j]+b_i \\ &= \sum_{j=1}^k A_{ij}\mathbb{E}_\theta[Z_j]+b_i = \sum_{j=1}^k A_{ij}(\mathbb{E}_\theta[Z])_j+b_i = (A\mathbb{E}_\theta[Z]+b)_i.\end{aligned}$$

- On remarque premièrement que $\mathrm{Var}_\theta(AZ+b) = \mathbb{E}_\theta\left[(AZ+b-A\mathbb{E}_\theta(Z)-b)(AZ+b-A\mathbb{E}_\theta(Z)-b)^T\right] = \mathbb{E}_\theta\left[(AZ-A\mathbb{E}_\theta(Z))(AZ-A\mathbb{E}_\theta(Z))^T\right] = \mathrm{Var}_\theta[AZ]$. De là, on trouve :

$$\begin{aligned}\mathrm{Var}_\theta[AZ]_{ij} &= \mathbb{E}_\theta\left[(AZ-A\mathbb{E}_\theta(Z))_i(AZ-A\mathbb{E}_\theta(Z))_j\right] = \mathbb{E}_\theta\left[((AZ)_i-\mathbb{E}_\theta((AZ)_i))((AZ)_j-\mathbb{E}_\theta((AZ)_j))\right] \\ &= \mathbb{E}_\theta[(AZ)_i(AZ)_j]-\mathbb{E}_\theta[(AZ)_i]\mathbb{E}_\theta[(AZ)_j] \\ &= \mathbb{E}_\theta\left[\left(\sum_{\delta=1}^m A_{i\delta}Z_\delta\right)\left(\sum_{\gamma=1}^m A_{j\gamma}Z_\gamma\right)\right]-\mathbb{E}_\theta\left[\sum_{\delta=1}^m A_{i\delta}Z_\delta\right]\mathbb{E}_\theta\left[\sum_{\gamma=1}^m A_{j\gamma}Z_\gamma\right] \\ &= \sum_{\delta=1}^m\sum_{\gamma=1}^m A_{i\delta}A_{j\gamma}\mathbb{E}_\theta[Z_\delta Z_\gamma]-\sum_{\delta=1}^m\sum_{\gamma=1}^m A_{i\delta}A_{j\gamma}\mathbb{E}_\theta[Z_\delta]\mathbb{E}_\theta[Z_\gamma] \\ &= \sum_{\delta=1}^m\sum_{\gamma=1}^m A_{i\delta}A_{j\gamma}\left(\mathbb{E}_\theta[Z_\delta Z_\gamma]-\mathbb{E}_\theta[Z_\delta]\mathbb{E}_\theta[Z_\gamma]\right) \\ &= \sum_{\delta=1}^m\sum_{\gamma=1}^m A_{i\delta}A_{j\gamma}\,\mathrm{Var}_\theta[Z]_{\delta\gamma} \\ &= \sum_{\delta=1}^m\sum_{\gamma=1}^m A_{i\delta}\,\mathrm{Var}_\theta[Z]_{\delta\gamma}(A^T)_{\gamma j} \\ &= \left(A\,\mathrm{Var}_\theta[Z]A^T\right)_{ij}\end{aligned}$$

— soit $v \in \mathbb{R}^m$. On remarque que $v^T \operatorname{Var}_\theta[Z] v = \operatorname{Var}_\theta[v^T Z] \geqslant 0$.

$\square$

Définition 2.28. Soient $Z = (Z_1, \ldots, Z_m)^T$, et $Y = (Y_1, \ldots, Y_\ell)^T$. On définit la covariance de Y et Z par :

$$\operatorname{Cov}[Y, Z] \coloneqq \mathbb{E}_\theta \left[(Y - \mathbb{E}_\theta[Y])(Z - \mathbb{E}_\theta[Z]) \right].$$

Proposition 2.29. *Soient* $Z = (Z_1, \ldots, Z_m)^T$*, et* $Y = (Y_1, \ldots, Y_\ell)^T$*. Alors :*
- $\operatorname{Cov}[Y, Z]_{ij} = \operatorname{Cov}[Y_i, Z_j]$;
- $\operatorname{Cov}[Z, Y] = \operatorname{Cov}[Y, Z]$;
- $\operatorname{Cov}[Z, Z] = \operatorname{Var}_\theta[Z]$;
- $\operatorname{Cov}[AY + b, CZ + d] = A \operatorname{Cov}[Y, Z] C^T$

Définition 2.30. Soient $g(\theta) \in \mathbb{R}^m$ et $T^{(n)}(X^{(n)})$, un estimateur de $g(\theta)$. On définit :

$$\operatorname{MSE}_\theta[T^{(n)}(X^{(n)})] \coloneqq \mathbb{E}_\theta \left[(T^{(n)}(X^{(n)}) - g(\theta))(T^{(n)}(X^{(n)}) - g(\theta))^T \right] \in \operatorname{Mat}(m, m).$$

Remarque. $\operatorname{MSE}_\theta[T^{(n)}(X^{(n)})] = \mathbb{E}_\theta \left[\left(T_i^{(n)}(X^{(n)}) - g(\theta) \right) \left(T_j^{(n)}(X^{(n)}) - g(\theta) \right)^T \right]$. En particulier :

$$\operatorname{MSE}_\theta[T^{(n)}(X^{(n)})]_{ii} = \operatorname{MSE}_\theta[T_i^{(n)}].$$

Définition 2.31. Soient $A, B \in \operatorname{Mat}(m, n)$. On dit que $A \geqslant B$ lorsque $A - B \geqslant 0$, c-à-d lorsque la forme bilinéaire définie par $(A - B)$ est semi définie positive.

Définition 2.32. Soit $\mathcal{C}$, une classe d'estimateurs de $g(\theta)$. Alors $T_*^{(n)}$ est à erreur quadratique moyenne minimale dans $\mathcal{C}$ lorsque :

(i) $T_*^{(n)} \in \mathcal{C}$;

(ii) $\forall T^{(n)} \in \mathcal{C} : \forall \theta \in \Theta : \operatorname{MSE}_\theta[T^{(n)}] \geqslant \operatorname{MSE}_\theta[T_*^{(n)}]$.

Proposition 2.33. *Pour tout* $\theta \in \Theta, T^{(n)} \in \mathcal{C}$*, on a :*

$$\operatorname{MSE}_\theta[T^{(n)}] \geqslant 0.$$

Démonstration. Soit $v \in \mathbb{R}^m$. On trouve :

$$\begin{aligned} v^T \operatorname{MSE}_\theta[T^{(n)}] v &= v^T \mathbb{E}_\theta \left[\left(T^{(n)} - g(\theta) \right) \left(T^{(n)} - g(\theta) \right)^T \right] v = \mathbb{E}_\theta \left[v^T \left(T^{(n)} - g(\theta) \right) \left(T^{(n)} - g(\theta) \right)^T v \right] \\ &= \mathbb{E}_\theta \left[\left(v^T (T^{(n)} - g(\theta)) \right) \left(v^T (T^{(n)} - g(\theta)) \right)^T \right] \\ &= \mathbb{E}_\theta \left[\left(v^T (T^{(n)} - g(\theta)) \right)^2 \right] \geqslant 0. \end{aligned}$$

$\square$

Proposition 2.34. *Soient* $T_1^{(n)}, T_2^{(n)}$*, deux estimateurs de* $g(\theta)$ *tels que :*

$$\operatorname{MSE}_\theta[T_1^{(n)}] \geqslant \operatorname{MSE}_\theta[T_2^{(n)}].$$

Alors :

$$\forall v \in \mathbb{R}^m : \operatorname{MSE}_\theta[v^T T_1^{(n)}] \geqslant \operatorname{MSE}_\theta[v^T T_2^{(n)}].$$

Démonstration.

$$\begin{aligned}
\mathrm{MSE}_\theta[\nu^\mathsf{T} T_1^{(n)}] - \mathrm{MSE}_\theta[\nu^\mathsf{T} T_2^{(n)}] &= \mathbb{E}_\theta\left[\left(\nu^\mathsf{T} T_1^{(n)} - \nu^\mathsf{T} g(\theta)\right)^2\right] - \mathbb{E}_\theta\left[\left(\nu^\mathsf{T} T_2^{(n)} - \nu^\mathsf{T} g(\theta)\right)^2\right] \\
&= \mathbb{E}_\theta\left[\nu^\mathsf{T}\left(T_1^{(n)} - g(\theta)\right)\right] - \mathbb{E}_\theta\left[\nu^\mathsf{T}\left(T_2^{(n)} - g(\theta)\right)\right] \\
&= \nu^\mathsf{T}\,\mathrm{MSE}_\theta[T_1^{(n)}]\nu - \nu^\mathsf{T}\,\mathrm{MSE}_\theta[T_2^{(n)}]\nu = \nu^\mathsf{T}\left(\mathrm{MSE}_\theta[T_1^{(n)}] - \mathrm{MSE}_\theta[T_2^{(n)}]\right)\nu \geqslant 0,
\end{aligned}$$

par définition de semi-positivité. □

2.3.3 Estimateurs efficaces

Définition 2.35. On note l'ensemble des valeurs $x^{(n)}$ de $\mathbb{R}^n$ prises par $X^{(n)}$ avec probabilité non-nulle par l'ensemble :

$$\mathcal{X}^{(n)} := \{x^{(n)} \in \mathbb{R}^n \text{ t.q. } L_\theta^{(n)}(x^{(n)}) > 0\}.$$

Définition 2.36. Soit un modèle statistique. On définit la *matrice d'information de Fisher* par :

$$I^{(n)}(\theta) := \int_{\mathcal{X}^{(n)}} \left[\nabla_\theta \ln L_\theta^{(n)}(x^{(n)})\right]\left[\nabla_\theta \ln L_\theta^{(n)}(x^{(n)})\right]^\mathsf{T} L_\theta^{(n)}(x^{(n)})\,dx^{(n)}.$$

Définition 2.37. Soit $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, un modèle statistique. On le dit *régulier* lorsque :

H1 Θ est ouvert ;

H2 $\mathcal{X}^{(n)}$ ne dépend pas de θ ;

H3 $\forall x^{(n)} \in \mathcal{X}^{(n)} : \theta \mapsto L_\theta^{(n)}(x^{(n)})$ est différentiable sur θ ;

H4 l'expression $\int_{\mathcal{X}^{(n)}} L_\theta^{(n)}(x^{(n)})\,dx^{(n)}$ est différentiable sous le signe ;

H5 $\forall\theta \in \Theta : I^{(n)}(\theta)$ existe, est finie, et est inversible.

Remarque. Les intégrales doivent être comprises comme des sommes (voire des séries dans le cas infini) dans les cas discrets.

Remarque. Un artifice fréquemment utilisé ici est le suivant :

$$\frac{\partial}{\partial\theta_i} L_\theta^{(n)}(x^{(n)}) = \frac{\partial}{\partial\theta_i} L_\theta^{(n)}(x^{(n)}) \frac{L_\theta^{(n)}(x^{(n)})}{L_\theta^{(n)}(x^{(n)})} = \frac{\partial}{\partial\theta_i} \ln L_\theta^{(n)}(x^{(n)}) L_\theta^{(n)}(x^{(n)}).$$

Proposition 2.38. *Dans un modèle statistique régulier, on a :*

$$\mathbb{E}_\theta\left[\nabla_\theta \ln L_\theta^{(n)}(x^{(n)})\right] = 0.$$

Démonstration. Par définition de l'espérance :

$$\mathbb{E}_\theta\left[\nabla_\theta \ln L_\theta^{(n)}(x^{(n)})\right] = \int_{\mathcal{X}^{(n)}} \left(\nabla_\theta \ln L_\theta^{(n)}(x^{(n)})\right) L_\theta^{(n)}(x^{(n)})\,dx^{(n)} = \int_{\mathcal{X}^{(n)}} \nabla_\theta L_\theta^{(n)}(x^{(n)})\,dx^{(n)}.$$

Par hypothèse de régularité, on sait :

$$\int_{\mathcal{X}^{(n)}} \nabla_\theta L_\theta^{(n)}(x^{(n)})\,dx^{(n)} = \nabla_\theta \int_{\mathcal{X}^{(n)}} L_\theta^{(n)}(x^{(n)})\,dx^{(n)}.$$

On trouve finalement :

$$\mathbb{E}_\theta\left[\nabla_\theta \ln L_\theta^{(n)}(x^{(n)})\right] = \nabla_\theta \int_{\mathcal{X}^{(n)}} L_\theta^{(n)}(x^{(n)})\,dx^{(n)} = \nabla_\theta 1 = 0.$$

□

Remarque. Par la proposition précédente, on peut dire :

$$I^{(n)}(\theta) = \mathbb{E}_\theta\left[\left[\nabla_\theta \ln L_\theta^{(n)}(x^{(n)})\right]\left[\nabla_\theta \ln L_\theta^{(n)}(x^{(n)})\right]^T\right] = \mathrm{Var}_\theta\left[\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right].$$

Proposition 2.39. *Dans un modèle d'échantillonage, on a :*

$$I^{(n)}(\theta) = nI^{(1)}(\theta).$$

Démonstration. Les variables $X_1, \ldots, X_n$ étant iid (modèle d'échantillonage), on peut exprimer :

$$\begin{aligned}
I^{(n)}(\theta) = \mathrm{Var}_\theta\left[\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right] &= \mathrm{Var}_\theta\left[\nabla_\theta \ln\left(\prod_{i=1}^n L_\theta^{(1)}(X_i)\right)\right] \\
&= \mathrm{Var}_\theta\left[\sum_{i=1}^n \nabla_\theta \ln L_\theta^{(1)}(X_i)\right] = \sum_{i=1}^n \mathrm{Var}_\theta\left[\nabla_\theta \ln L_\theta^{(1)}(X_i)\right] \\
&= n\,\mathrm{Var}_\theta\left[\nabla_\theta \ln L_\theta^{(1)}(X_1)\right] = nI^{(1)}(\theta).
\end{aligned}$$

□

Remarque. On en déduit que n variables aléatoires iid contiennent n fois plus d'information qu'une seule d'entre elles.

Remarque. pour comprendre l'importance de l'hypothèse H5, supposons $k = 1$ et $\forall\theta \in \Theta : I^{(n)}(\theta) = 0$. 0 n'est pas inversible, ce qui viole H5. Alors par l'égalité d'une remarque précédente :

$$\mathrm{Var}_\theta\left[\tfrac{d}{d\theta} \ln L_\theta^{(n)}(X^{(n)})\right] = 0.$$

On en déduit que $\frac{d}{d\theta} \ln L_\theta^{(n)}(X^{(n)}) = 0$, et donc que $\ln L_\theta^{(n)}(X^{(n)})$ est constante sur son espérance, c-à-d 0. On en déduit :

$$\frac{d}{d\theta} L_\theta^{(n)}(X^{(n)}) = 0 \qquad \forall\theta \in \Theta \quad \mathbb{P}_\theta - \text{p.s.}.$$

Et donc, pour tout $\theta \in \Theta$, on a :

$$\mathbb{P}[X^{(n)} \in B] = \int_B L_\theta^{(n)}\, dx^{(n)},$$

et en dérivant les deux membres de l'égalité :

$$\frac{d}{d\theta}\mathbb{P}[X^{(n)} \in B] = \int_B \frac{d}{d\theta} L_\theta^{(n)}(x^{(n)})\, dx^{(n)}.$$

Cela veut dire que la probabilité $\mathbb{P}[X^{(n)} \in B]$ est constante pour toutes valeurs de θ. Les $\mathbb{P}_\theta$ ne sont donc plus différenciables. On a donc perdu toute information sur θ.

Définition 2.40. Dans un modèle statistique régulier, un estimateur $T^{(n)}(X^{(n)})$ de $g(\theta)$ est dit *régulier* lorsque :

(i) $\forall\theta \in \Theta : \mathbb{E}_\theta\left[\left\|T^{(n)}(X^{(n)})\right\|^2\right] < +\infty$;

(ii) $\psi_\theta := \mathbb{E}_\theta[T^{(n)}(X^{(n)})] = \int_{\mathcal{X}^{(n)}} T^{(n)}(x^{(n)}) L_\theta^{(n)}(x^{(n)})\, dx^{(n)}$ est différentiable sous le signe pour tout $\theta \in \Theta$.

Théorème 2.41. *Soient un modèle statistique régulier, et* $T^{(n)}(X^{(n)})$ *un estimateur régulier de* $g(\theta)$*. Alors :*

$$\forall\theta \in \Theta : \mathrm{MSE}_\theta(T^{(n)}(X^{(n)})) = \mathbb{E}_\theta\left[\left[T^{(n)}(X^{(n)}) - g(\theta)\right]\left[T^{(n)}(X^{(n)}) - g(\theta)\right]^T\right] \geqslant \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T,$$

où $\Delta_\theta = \mathrm{Jac}\,\psi = \left[\frac{\partial\psi_i}{\partial\theta_j}(\theta)\right] \in \mathrm{Mat}(m, k)$.

Remarque. Dans le cas où $g(\theta) = \theta = \psi(\theta)$, on a un estimateur non-biaisé, et donc :

$$\mathrm{Jac}\,\psi = \mathrm{Jac}\,\mathrm{Id} = \mathrm{Id}\,.$$

Et donc, on a :

$$\forall\theta \in \Theta : \mathrm{MSE}_\theta(T^{(n)}(X^{(n)})) \geqslant I^{(n)}(\theta)^{-1}.$$

Remarque. Dans le cas d'un modèle d'échantillonage, on a $(I^{(n)})^{-1} = n^{-1}I^{(4)}(\theta)^{-1}$, et donc :

$$\forall\theta \in \Theta : \mathrm{MSE}_\theta(T^{(n)}(X^{(n)})) \geqslant n^{-1}\Delta_\theta I^{(1)}(\theta)^{-1}\Delta_\theta^T \xrightarrow[n\to+\infty]{} 0.$$

Preuve du Théorème 2.41. Notons d'abord :

$$\begin{aligned}(\Delta_\theta)_{ij} = \frac{\partial\psi_i}{\partial\theta_j}(\theta) &= \frac{\partial}{\partial\theta_j}\mathbb{E}_\theta[T_i^{(n)}(X^{(n)}) = \frac{\partial}{\partial\theta_j}\int_{\mathcal{X}^{(n)}} T_i^{(n)}(x^{(n)})L_\theta^{(n)}\,dx^{(n)} \\ &= \int_{\mathcal{X}^{(n)}} T_i(x^{(n)})\frac{\partial}{\partial\theta_j}L_\theta^{(n)}(x^{(n)})\,dx^{(n)} = \int_{\mathcal{X}^{(n)}} T_i(x^{(n)})\frac{\partial}{\partial\theta_j}\ln L_\theta^{(n)}(x^{(n)})L_\theta^{(n)}(x^{(n)})\,dx^{(n)}.\end{aligned}$$

On a donc une espérance donnée par :

$$(\Delta_\theta)_{ij} = \mathbb{E}_\theta\left[T_i(X^{(n)})\frac{\partial}{\partial\theta_j}\ln L_\theta^{(n)}(X^{(n)})\right] = \mathbb{E}_\theta\left[T^{(n)}(X^{(n)})\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right]_{ij}.$$

On a alors l'égalité suivante :

$$\Delta_\theta = \mathbb{E}_\theta\left[T^{(n)}(X^{(n)})\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right].$$

On veut ensuite montrer :

$$\begin{aligned}\mathrm{MSE}_\theta[T^{(n)}] = \mathbb{E}_\theta\left[\left(T^{(n)} - g(\theta)\right)\left(T^{(n)} - g(\theta)\right)^T\right] &\overset{(1)}{=} Var[T^{(n)}] + \left(\psi(\theta) - g(\theta)\right)\left(\psi(\theta) - g(\theta)\right)^T \\ &\overset{(2)}{\geqslant} \mathrm{Var}_\theta[T^{(n)}] \\ &\overset{(3)}{\geqslant} \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T\end{aligned}$$

Montrons d'abord (1). Pour cela :

$$\begin{aligned}\mathrm{MSE}_\theta(T^{(n)}) &= \mathbb{E}_\theta\left[\left((T^{(n)} - \psi(\theta)) + (\psi(\theta) - g(\theta))\right)\left((T^{(n)} - \psi(\theta)) + (\psi(\theta) - g(\theta))\right)^T\right] \\ &= \mathbb{E}_\theta[(T^{(n)} - \psi(\theta))(T^{(n)} - \psi(\theta))^T] + \mathbb{E}_\theta[(T^{(n)} - \psi(\theta))](\psi(\theta) - g(\theta))^T \\ &\quad - (\psi(\theta) - g(\theta))\mathbb{E}_\theta[(T^{(n)} - \psi(\theta))^T] + (\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))^T \\ &= \mathrm{Var}_\theta[T^{(n)}] + 0 + 0 + (\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))^T.\end{aligned}$$

Pour montrer (2), on voit bien que $(\psi(\theta) - g(\theta))(\psi(\theta) - g(\theta))^T$ est semi-définie positive par construction (uu^T est toujours semi-définie positive).

Montrons alors (3). Posons pour cela :

$$S_\theta \coloneqq T^{(n)}(X^{(n)}) - \Delta_\theta I^{(n)}(\theta)^{-1}\nabla_\theta \ln L_\theta^{(n)}(X^{(n)}) \in \mathbb{R}^m.$$

Calculons ensuite :

$$\mathbb{E}_\theta[S_\theta] = \mathbb{E}_\theta[T^{(n)}(X^{(n)})] - \Delta_\theta I^{(n)}(\theta)^{-1}\mathbb{E}_\theta[\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})] = \mathbb{E}_\theta[T^{(n)}(X^{(n)})] - \Delta_\theta I^{(n)}(\theta)^{-1} \cdot 0.$$

On sait également que :

$$\begin{aligned}
\mathrm{Var}_\theta[S_\theta] &= \mathbb{E}_\theta[S_\theta S_\theta^T] - \mathbb{E}_\theta[S_\theta]\mathbb{E}_\theta[S_\theta]^T \\
&= \mathbb{E}_\theta\left[\left(T^{(n)} - \Delta_\theta I^{(n)}(\theta)^{-1}\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right)\left(T^{(n)} - \Delta_\theta I^{(n)}(\theta)^{-1}\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right)^T\right] - \psi(\theta)\psi(\theta)^T \\
&= \mathrm{Var}_\theta[T^{(n)}] - \mathbb{E}_\theta\left[T^{(n)}(X^{(n)})\left(\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right)\right]\left(I^{(n)}(\theta)^{-1}\right)^T\Delta_\theta^T \\
&\quad - \Delta_\theta I^{(n)}(\theta)^{-1}\mathbb{E}_\theta\left[\left(\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right)T^{(n)}(X^{(n)})\right] \\
&\quad + \Delta_\theta I^{(n)}(\theta)^{-1}\mathbb{E}_\theta\left[\left(\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right)\left(\nabla_\theta \ln L_\theta^{(n)}(X^{(n)})\right)^T\right]\left(I^{(n)}(\theta)^{-1}\right)^T\Delta_\theta^T \\
&= \mathrm{Var}_\theta[T^{(n)}(X^{(n)})] - \Delta_\theta\left(I^{(n)}(\theta)^{-1}\right)^T\Delta_\theta^T - \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T + \Delta_\theta I^{(n)}(\theta)^{-1}I^{(n)}(\theta)\left(I^{(n)}(\theta)^{-1}\right)^T\Delta_\theta^T,
\end{aligned}$$

par définition de $I^{(n)}(\theta)$. Puisque $I^{(n)}(\theta)$ est symétrique et inversible, on sait que $I^{(n)}(\theta)^{-1}$ est également symétrique. On trouve alors :

$$\mathrm{Var}_\theta[S_\theta] = \mathrm{Var}_\theta[T^{(n)}(X^{(n)})] - \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T - \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T + \Delta_\theta I^{(n)}(\theta)^{-1}I^{(n)}(\theta)I^{(n)}(\theta)^{-1}\Delta_\theta^T.$$

En simplifiant les produits de matrice avec leur inverse, on obtient :

$$\begin{aligned}
\mathrm{Var}_\theta[S_\theta] &= \mathrm{Var}_\theta[T^{(n)}(X^{(n)})] - \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T - \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T + \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T \\
&= \mathrm{Var}_\theta[T^{(n)}(X^{(n)})] - \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T
\end{aligned}$$

Or $\mathrm{Var}_\theta[S_\theta]$ est semi-définie positive par construction. Dès lors, on sait $\mathrm{Var}_\theta[T^{(n)}(X^{(n)})] - \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T \geqslant 0$, ou encore $\mathrm{Var}_\theta[T^{(n)}(X^{(n)})] \geqslant \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T$. □

Définition 2.42. Dans un modèle statistique régulier, l'estimateur régulier $T^{(n)}(X^{(n)})$ de $g(\theta)$ est dit *efficace* lorsque :

$$\forall\theta \in \Theta : \mathrm{MSE}_\theta[T^{(n)}(X^{(n)})] = \Delta_\theta I^{(n)}(\theta)^{-1}\Delta_\theta^T.$$

Remarque. Pour avoir $T^{(n)}(X^{(n)})$ efficace, il faut $g(\theta) = \psi(\theta)$ et $\mathrm{Var}_\theta[T^{(n)}(X^{(n)})] = \Delta_\theta I^{(n)}(\theta)^{-1}\Delta^T$. L'estimateur doit donc être sans biais. De plus, la variance est définie par la borne de Cramer.

Exemple 2.12. Soient $X_1, \ldots, X_n$ iid $\mathcal{N}(\mu, \sigma_0^2)$, $g(\theta = \mu) = \mu$. On a donc $k = m = 1$. On sait que $T^{(n)}(X^{(n)}) = \overline{X}^{(n)}$ est un estimateur sans biais. Comparons alors la variance de $T^{(n)}(X^{(n)})$ avec la borne de Cramer.

$$\Delta_\mu = \frac{d}{d\mu}\psi(\mu) = \frac{d}{d\mu}\mathbb{E}_\theta[\overline{X}^{(n)}] = \frac{d}{d\mu}\mu = 1.$$

Calculons alors $I^{(n)}(\theta)$, qui est donné par :

$$L_\mu^{(n)}(x^{(n)}) = \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left(-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i - \mu)^2\right),$$

et donc :

$$I^{(n)}(\theta) = \mathrm{Var}_\mu\left[\tfrac{d}{d\mu}\ln L_\mu^{(n)}(X^{(n)})\right] = \mathrm{Var}_\mu\left[\tfrac{d}{d\mu}\left(n\ln\frac{1}{\sqrt{2\pi}\sigma_0} - \frac{1}{2\sigma_0^2}\sum_{i=2}^n (X_i-\mu)^2\right)\right]$$
$$= \frac{1}{(2\sigma_0^2)^2}\mathrm{Var}_\mu\left[-2\sum_{i=1}^n (X_i-\mu)\right] = \frac{1}{\sigma_0^4}\sum_{i=1}^n \mathrm{Var}_\mu(X_i) = \frac{\sigma_0^2}{\sigma_0^4} = \frac{n}{\sigma_0^2} = \mathrm{Var}[T^{(n)}(X^{(n)})]^{-1} = \mathrm{MSE}_\theta[T^{(n)}(X^{(n)})].$$

On en déduit que l'estimateur $\overline{X}^{(n)}$ est efficace pour $g(\theta) = \mu$.
Remarque. On observe que $I^{(n)}(\theta) \propto \mathrm{Var}_\theta[T^{(n)}(X^{(n)})]^{-1}$, ce qui est cohérent avec l'intuition de la notion d'information : au plus la variance est faible, au plus les observations sont groupées, et au plus il y a d'information à déduire sur θ, alors qu'au plus la variance est élevée, au plus les observations sont dispersées, et au moins il y a d'information contenue sur θ.
Exemple 2.13. $X_1, \dots, X_n$ iid $\mathrm{Bern}(p)$, avec $p \in (0,1)$. On sait que $\mathbb{E}_p[X_i] = p$ et $\mathrm{Var}_p[X_i] = p(1-p)$. On pose $g(p) = p$. On calcule ensuite :

$$I^{(n)}(\theta) = \mathrm{Var}_p\left[\tfrac{d}{dp}\ln L_p^{(n)}(X^{(n)})\right] = \mathrm{Var}_p\left[\tfrac{d}{dp}\ln\left(p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i}\right)\right]$$
$$= \mathrm{Var}_p\left[\tfrac{d}{dp}\left(\sum_{i=1}^n X_i\ln p + \left(n - \sum_{i=1}^n X_i\right)\ln(1-p)\right)\right]$$
$$= \mathrm{Var}_p\left[\frac{1}{p}\sum_{i=1}^n X_i - \frac{1}{1-p}\left(n - \sum_{i=1}^n X_i\right)\right] = \mathrm{Var}_p\left[\left(\frac{1}{p} + \frac{1}{1-p}\right)\sum_{i=1}^n X_i - \frac{n}{1-p}\right]$$
$$= \frac{1}{(p(1-p)p^2}\sum_{i=1}^n \mathrm{Var}_p[X_i] = \frac{1}{(p(1-p))^2}n(p(1-p)) = \frac{n}{p(1-p)}.$$

En prenant $T^{(n)}(X^{(n)}) = \overline{X}^{(n)}$, on a $\mathbb{E}_p[T^{(n)}(X^{(n)})] = p$, et $\mathrm{Var}_p[T^{(n)}(X^{(n)})] = \frac{\mathrm{Var}[X_1]}{n} = \frac{p(1-p)}{n} = I^{(n)}(\theta)^{-1}$.
À nouveau, on en déduit que $\overline{X}^{(n)}$ est un estimateur efficace.

2.4 Méthodes d'estimation

2.4.1 Méthode des moments

Plaçons-nous dans un modèle paramétrique $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, avec $\theta \in \Theta \subseteq \mathbb{R}^k$. On suppose que pour tout $1 \leqslant r \leqslant k$, on a : $\mathbb{E}_\theta[|X_i|^r] < +\infty$.
Définition 2.43. Soit $M : \Theta \to M(\Theta) : \theta \mapsto \left[\mu_i'(\theta)\right]_{1\leqslant i\leqslant k}$. Supposons M inversible. Soit $X^{(n)}$ un vecteur d'observations de loi $\mathbb{P}_\theta^{(n)}$. On appelle *estimateur de* θ *par la méthode des moments* la statistique :

$$T^{(n)}(X^{(n)}) = M^{-1}\left([m_i']_{1\leqslant i\leqslant k}\right),$$

où m_i' est le moment empirique d'ordre i de θ

Exemple 2.14. Soient $X_1, \dots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$, avec $\theta \in \Theta \subset \mathbb{R}^2$. On connaît les moments d'ordre 1 et 2 :

$$\begin{aligned}\mu_1'(\theta) &= \mathbb{E}_\theta[X_1] = \mu\\ \mu_2'(\theta) &= \mathbb{E}_\theta[X_1^2] = \mathrm{Var}_\theta[X_i] + \mathbb{E}_\theta[X_1]^2 = \mu^2 + \sigma^2\end{aligned}$$

On a alors la fonction M donnée par :

$$M : \Theta \to M(\Theta) = \Theta : \theta = \begin{pmatrix}\mu\\ \sigma\end{pmatrix} \mapsto \begin{pmatrix}\mu\\ \mu^2+\sigma^2\end{pmatrix},$$

qui est bien inversible. Il faut alors résoudre le système :

$$\begin{cases}\mu_1'(\theta) = m_1'\\ \mu_2'(\theta) = m_2',\end{cases}$$

ce qui donne :

$$\begin{cases}\mu = \mu_1'(\theta) = m_1' = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i = \overline{X}\\[2ex] \sigma^2 + \mu^2 = \mu_2'(\theta) = m_2' = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i^2.\end{cases}$$

On prend alors pour estimateur :

$$T^{(n)}(X^{(n)}) = \begin{pmatrix}\overline{X}\\ \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2\end{pmatrix} = \begin{pmatrix}\overline{X}\\ s^2\end{pmatrix}.$$

Cet estimateur est convergent (car convergent en chaque composante), est exhaustif, est biaisé (car s^2 est un estimateur biaisé de σ^2), et est asymptotiquement efficace.

Exemple 2.15. Soient $X_1, \dots, X_n$ iid $\mathrm{Bern}(p)$, $p \in [0,1] \subset \mathbb{R}$. On obtient :

$$p = \mu = \mu_1'(\theta) = m_1' = \overline{X},$$

et donc on prend l'estimateur $T^{(n)}(X^{(n)}) = \overline{X}$.

Exemple 2.16. Soient $X_1, \dots, X_n$ iid $\mathrm{Unif}(0,\theta)$, $\theta \in {\mathbb{R}^+}_0 \subset \mathbb{R}$. Le système nous donne donc :

$$\frac{\theta}{2} = \mathbb{E}_\theta[X_1] = \mu_1'(\theta) = m_1' = \overline{X}.$$

L'estimateur de θ par la méthode des moments est alors donné par :

$$T^{(n)}(X^{(n)}) = 2\overline{X}.$$

Remarque.

$$\mathrm{MSE}_\theta[2\overline{X}] = 0^2 + \mathrm{Var}_\theta[2\overline{X}] = \frac{4}{n}\mathrm{Var}[X_1] \propto \frac{c}{n},$$

alors qu'un estimateur efficace a un $\mathrm{MSE}_\theta \propto \frac{c}{n^2}$.

2.4.2 Méthode du maximum de vraisemblance

Remarque. Une notation usuelle pour noter un estimateur de θ est $\widehat{\theta}$. Cette notation remplace l'estimateur $T^{(n)}(X^{(n)})$ pour lequel il faut explicitement préciser la variable estimée. $\widehat{\theta}$ est une statistique, et ne dépend aucunement de θ !

L'idée d'un tel estimateur est de se baser sur la maximisation (selon θ) de la vraisemblance. On est donc tenté d'écrire :

$$\widehat{\theta} = \arg\max_{\theta \in \Theta} L_\theta^{(n)}(X^{(n)}).$$

Or il peut arriver qu'un tel θ de soit pas unique, et donc qu'il existe plusieurs estimateurs de θ par le maximum de vraisemblance.

Définition 2.44. L'estimateur $\widehat{\theta}$ est un estimateur de θ par la méthode du maximum de vraisemblance lorsque :

$$\forall \theta \in \Theta : L_{\widehat{\theta}}^{(n)}(X^{(n)}) \geqslant L_\theta^{(n)}(X^{(n)}).$$

Exemple 2.17. Soient $X_1, \ldots, X_n$ iid $\mathrm{Unif}\left(\left[\theta \pm \frac{1}{2}\right]\right)$, avec $\theta \in \mathbb{R}$. On peut écrire la vraisemblance :

$$L_\theta^{(n)}(X^{(n)}) = \prod_{i=1}^n f_\theta^{X_i}(X_i) = \prod_{i=1}^n I_{[\theta - \frac{1}{2} \leqslant X_i \leqslant \theta + \frac{1}{2}]} = I_{[\theta - \frac{1}{2} \leqslant X_{(1)}]} I_{[X_{(n)} \leqslant \theta + \frac{1}{2}]} = I_{[X_{(n)} - \frac{1}{2} \leqslant \theta \leqslant X_{(1)} + \frac{1}{2}]}.$$

La fonction de vraisemblance valant ici soit 1 soit 0, toute valeur de θ telle que $\theta \in \left[X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2}\right]$ maximise la vraisemblance.

Exemple 2.18. Soient $X_1, \ldots X_n$ iid $\mathrm{Unif}(0, \theta)$, avec $\theta \in \mathbb{R}^+{}_0$. Ici, la fonction de vraisemblance ne vaut plus uniquement 0 ou 1 :

$$_\theta^{(n)}(X^{(n)}) = \frac{1}{\theta^n} I_{[0 \leqslant X_{(1)}]} I_{[X_{(n)} \leqslant \theta]} = \frac{1}{\theta^n} I_{[0 \lneqq \theta \leqslant X_{(n)}]}.$$

Pour $\theta \in [0, X_{(n)}]$, la fonction de vraisemblance est nulle, et puis vaut θ^{-n} pour $\theta \geqslant X_{(n)}$. Le maximum est donc en $\theta = X_{(n)}$.

De plus, la fonction $\theta \mapsto \theta^{-n}$ est strictement décroissante, donc le maximum de vraisemblance est unique.

Remarque. Par le critère de factorisation de Neymann-Fisher, si $S(X^{(n)})$ est une statistique exhaustive, on sait :

$$L_\theta^{(n)}(X^{(n)}) = (g_\theta \circ S)(X^{(n)}) \cdot h(X^{(n)}).$$

Donc (en considérant $\arg\max$ comme l'ensemble des arguments maximisant la valeur), on a :

$$\widehat{\theta} \in \arg\max_{\theta \in \Theta} L_\theta^{(n)}(X^{(n)}) = \arg\max_{\theta \in \Theta} (g_\theta \circ S)(X^{(n)}) = (\alpha \circ S)(X^{(n)}).$$

Cela implique que dans le cas du maximum de vraisemblance, l'estimateur dépend toujours des observations à travers une statistique exhaustive.

Proposition 2.45. *Si $\theta \mapsto L_\theta^{(n)}(X^{(n)})$ est différentiable sur $\mathrm{int}(\Theta)$, alors tout estimateur du maximum de vraisemblance à valeurs dans $\mathrm{int}(\Theta)$ vérifie :*

$$\nabla_\theta L_\theta^{(n)}(X^{(n)}) \Big|_{\theta = \widehat{\theta}} = 0.$$

Remarque. Par continuité et croissance stricte de $\ln(\cdot)$, on sait également que :

$$\nabla_\theta \ln\left(L_\theta^{(n)}(X^{(n)})\right) \Big|_{\theta = \widehat{\theta}} = 0.$$

L'intérêt de passer par un logarithme est de transformer les produits en somme, ce qui est plus agréable à dériver.
Exemple 2.19. Soient $X_1, \dots, X_n$ iid $\mathrm{Bern}(p)$. On peut calculer la fonction de vraisemblance :

$$L_p^{(n)}(X^{(n)}) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} = p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} = p^{n\overline{X}}(1-p)^{n(1-\overline{X})}.$$

Annulons ensuite le gradient de $\ln \circ L_p^{(n)}$:

$$0 = \frac{\partial}{\partial p} = n\overline{X}\frac{\partial}{\partial p}\ln p + n(1-\overline{X})\frac{\partial}{\partial p}\ln(1-p) = n\overline{X}\frac{1}{p} + n(\overline{X}-1)\frac{1}{1-p}.$$

En multipliant par $p(1-p)$ de part et d'autre ,on obtient :

$$0 = n\overline{X}(1-p) + n(\overline{X}-1)p = n\overline{X} - n\overline{X}p + n\overline{X}p - np = n\overline{X} - np,$$

ou encore $p = \overline{X}$. Dès lors, $\widehat{p} = \overline{X}$ est un estimateur de p par le maximum de vraisemblance.
Proposition 2.46. *Soit* $g : \Theta \to g(\Theta)$. *Alors :*

$$\forall\theta \in \Theta : \widehat{g(\theta)} = g(\widehat{\theta}).$$

Démonstration. Distinguons les deux cas où g est bijective, et où g n'est pas bijective.

Si g est bijective, alors c'est évident car $g(\theta)$ transforme juste le paramètre de manière univoque.

Si g n'est pas bijective, on observe que g est surjective par construction. Et donc g n'est pas injective. On pose alors :

$$G : \Theta \to g(\Theta) \times \Theta : \theta \mapsto \begin{pmatrix} g(\theta) \\ \theta \end{pmatrix}.$$

G est injective par construction, et donc est bijective (car surjective en chaque composante). On sait donc que :

$$\widehat{G(\theta)} = G(\widehat{\theta}),$$

ce qui implique l'égalité composante par composante, et donc :

$$\widehat{g(\theta)} = g(\widehat{\theta}).$$

□

Définition 2.47. Soit $\widehat{\theta}$ un estimateur. On dit qu'il est *asymptotiquement efficace* lorsque :

$$\mathrm{Var}_\theta[\widehat{\theta}] = \frac{1}{nI^{(1)}(\theta)}.$$

Exemple 2.20. Prenons $X_1, \dots, X_n$ iid $\mathbb{P}_\theta^{(1)}$, avec $\theta \in \Theta \subseteq \mathbb{R}$. Par la loi des grands nombres, on a :

$$\forall\theta \in \Theta : \widehat{\theta} \xrightarrow[n\to+\infty]{\text{p.s.}} \theta \quad \text{sous } \mathbb{P}_\theta^{(1)}.$$

On a également :

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{I^{(1)}(\theta)}\right).$$

Pour $n \gg$, on a :

$$\widehat{\theta} \approx \mathcal{N}\left(\theta, \frac{1}{nI^{(1)}(\theta)}\right),$$

et donc :

$$\begin{aligned}\mathbb{E}_\theta[\widehat{\theta}] &\approx \theta \\ \mathrm{Var}_\theta[\widehat{\theta}] &\approx \frac{1}{nI^{(1)}(\theta)}.\end{aligned}$$

On sait dès lors que pour n grand, on a :

$$\mathrm{MSE}_\theta[\widehat{\theta}] \approx b_\theta(\widehat{\theta})^2 + \mathrm{Var}_\theta[\widehat{\theta}] = \mathrm{Var}_\theta[\widehat{\theta}] = \frac{1}{nI^{(1)}(\theta)},$$

qui est la borne de Cramer-Rao. $\widehat{\theta}$ est donc asymptotiquement efficace.

2.4.3 Comportement asymptotique des estimateurs du maximum de vraisemblance

On se place dans un modèle statistique $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)} = \{\mathbb{P}_\theta^{(n)} \text{ t.q. } \theta \in \Theta \subseteq \mathbb{R}\})$ d'échantillonage.

On considère les hypothèses suivantes :

H1 Θ est ouvert ;

H2 $\mathcal{X} = \{x \in \mathbb{R} \text{ t.q. } L_\theta^{(1)}(x) > 0\}$ ne dépend pas de θ ;

H3 $\forall x \in \mathcal{X} : \theta \mapsto L_\theta^{(1)}(x)$ est dérivable sur Θ et $\theta \mapsto \frac{d}{d\theta} L_\theta^{(1)}(x)$ est dérivable sur Θ ;

H4 l'expression $\int_{\mathcal{X}} L_\theta^{(1)}(x)\,dx$ peut être dérivée sous le signe deux fois ;

H5 $I^{(1)}(\theta) = \int_{\mathcal{X}} \left(\frac{d}{d\theta} \ln L_\theta^{(1)}(x)\right)^2 dx$ est fini et $\neq 0$;

H6 Pour toute suite $(\theta_n)_n$ convergente en $\theta \in \Theta$, on a :

$$\sup_{x \in \mathcal{X}} \left| B_{\theta_n}^{(1)}(x) - B_\theta^{(1)}(x) \right| \xrightarrow[n \to +\infty]{} 0,$$

avec $B_\theta^{(1)}(x) = \frac{d^2}{d\theta^2} \ln L_\theta^{(1)}(x)$.

Remarque. À nouveau :

$$0 = \frac{d}{d\theta} 1 = \frac{d}{d\theta} \int_{\mathcal{X}} L_\theta^{(1)}(x)\,dx = \int_{\mathcal{X}} \frac{d}{d\theta} L_\theta^{(1)}(x)\,dx = \int_{\mathcal{X}} \left(\tfrac{d}{d\theta} \ln L_\theta^{(1)}(x)\right) L_\theta^{(1)}(x)\,dx = \mathbb{E}_\theta\left[\tfrac{d}{d\theta} \ln L_\theta^{(1)}(X_1)\right],$$

et donc :

$$I^{(1)}(\theta) = \mathbb{E}_\theta\left[\left(\tfrac{d}{d\theta} \ln L_\theta^{(1)}(X_1)\right)^2\right] = \mathrm{Var}_\theta\left[\tfrac{d}{d\theta} \ln L_\theta^{(1)}(X)\right].$$

Également, on sait :

$$0 = \frac{d}{d\theta} \int_{\mathcal{X}} \left(\ln L_\theta^{(1)}(x)\right) L_\theta^{(1)}(x)\,dx = \int_{\mathcal{X}} \left(\left(\tfrac{d^2}{d\theta^2} \ln L_\theta^{(1)}(x)\right) L_\theta^{(1)}(x) + \left(\tfrac{d}{d\theta} \ln L_\theta^{(1)}(x)\right) \tfrac{d}{d\theta} L_\theta^{(1)}(x)\right) dx.$$

On en déduit :

$$0 = \mathbb{E}_\theta\left[\tfrac{d^2}{d\theta^2} \ln L_\theta^{(1)}(X_1)\right] + \mathbb{E}_\theta\left[\left(\tfrac{d}{d\theta} \ln L_\theta^{(1)}(X_1)\right)^2\right] = \mathbb{E}_\theta\left[\tfrac{d^2}{d\theta^2} \ln L_\theta^{(1)}(X_1)\right] + I^{(1)}(\theta),$$

ou encore :

$$I^{(1)}(\theta) = -\mathbb{E}_\theta\left[\tfrac{d^2}{d\theta^2} \ln L_\theta^{(1)}(X_1)\right].$$

Théorème 2.48. *Si le modèle satisfait (H1)-(H6), alors :*

(i) *il existe une suite* $(T^{(n)}(X^{(n)}))_{n\in\mathbb{N}}$ *de solutions de l'équation de vraisemblance :*

$$\frac{d}{d\theta}\ln L_\theta^{(n)}(X^{(n)}) = 0$$

qui est fortement convergente pour θ*, c-à-d :*

$$\forall\theta\in\Theta : T^{(n)}(X^{(n)}) \xrightarrow[n\to+\infty]{p.s.} \theta \textit{ sous } \mathbb{P}_\theta^{(n)}$$

(ii) *pour toute telle suite* $(T^{(n)}(X^{(n)}))_n$*, on a :*

$$\forall\theta\in\Theta : \sqrt{n}(T^{(n)}(X^{(n)})-\theta) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0, I^{(1)}(\theta)^{-1}).$$

Remarque. (ii) implique que $T^{(n)}(X^{(n)})$ est asymptotiquement efficace.

Démonstration. Supposons que (i) est vrai et montrons (ii). Posons :

$$\begin{aligned} A_\theta^{(n)}(x^{(n)}) &:= \frac{1}{n}\tfrac{d}{d\theta}\ln L_\theta^{(n)}(x^{(n)}) \\ B_\theta^{(n)}(x^{(n)}) &:= \frac{1}{n}\tfrac{d^2}{d\theta^2}\ln L_\theta^{(n)}(x^{(n)}). \end{aligned}$$

On remarque :

$$A_\theta^{(n)}(x^{(n)}) = \frac{1}{n}\sum_{i=1}^n A_\theta^{(1)}(x_i) \qquad \text{et} \qquad B_\theta^{(n)}(x^{(n)}) = \frac{1}{n}\sum_{i=1}^n B_\theta^{(1)}(x_i).$$

Par définition de $T^{(n)}(X^{(n)})$, pour tout n, on a :

$$\frac{d}{d\theta}\ln L_\theta(X^{(n)})\bigg|_{\theta=T^{(n)}(X^{(n)})} = 0,$$

et donc :

$$A_{T^{(n)}(X^{(n)})}(X^{(n)}) = \frac{1}{n}\frac{d}{d\theta}\ln L_\theta(X^{(n)}) = 0.$$

On en déduit :

$$\forall n : 0 = A^{(n)}_{T^{(n)}(X^{(n)})}(X^{(n)}) = A_\theta^{(n)}(X^{(n)}) + \frac{d}{d\theta}A_\theta^{(n)}(X^{(n)})\bigg|_{T_*^{(n)}(X^{(n)})}(X^{(n)})\cdot(T^{(n)}(X^{(n)})-\theta),$$

et donc (théorème des accroissements finis) :

$$T_*^{(n)} = \frac{-A_\theta^{(n)}(X^{(n)})}{\frac{d}{d\theta}A_\theta^{(n)}(X^{(n)})\Big|_{\theta=T_*^{(n)}(X^{(n)})}} = \frac{-A_\theta^{(n)}(X^{(n)})}{B^{(n)}_{T_*^{(n)}(X^{(n)})}(X^{(n)})},$$

pour $T_*^{(n)}(X^{(n)})\in[0, T^{(n)}(X^{(n)})]$

On a alors :

$$\sqrt{n}\left(T^{(n)}(X^{(n)})-\theta\right) = \frac{-\sqrt{n}A_\theta^{(n)}(X^{(n)})}{B_{T_*^{(n)}(X^{(n)})}(X^{(n)})}.$$

Montrons indépendamment la convergence du numérateur et celle du dénominateur.

Convergence du numérateur :

$$-\sqrt{n}A_\theta^{(n)}(X^{(n)}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(-A_\theta^{(n)}(X^{(n)})\right) \xrightarrow[n\to+\infty]{\mathcal{D} \text{ sous } \mathbb{P}_\theta^{(n)}} \mathcal{N}(-\mathbb{E}_\theta[A_\theta^{(1)}(X_1)], I^{(1)}(\theta)) = \mathcal{N}(0, \mathbb{E}_\theta[(A_\theta^{(1)}(X^{(n)}))^2]).$$

Convergence du dénominateur :

$$\begin{aligned} B^{(1)}_{T_*^{(n)}(X^{(n)})}(X^{(n)}) &= \frac{1}{n}\sum_{i=1}^{n} B^{(1)}_{T_*^{(n)}}(X_i) = \frac{1}{n}\sum_{i=1}^{n} B_\theta^{(1)}(X_i) + \frac{1}{n}\sum_{i=1}^{n}\left(B^{(1)}_{T_*^{(n)}}(X_i) - B_\theta^{(1)}(X_i)\right) \\ &= -I^{(1)}(\theta) + \frac{1}{n}\sum_{i=1}^{n}\left(B^{(1)}_{T_*^{(n)}(X_i)}(X^{(n)}) - B_\theta^{(1)}(X_i)\right). \end{aligned}$$

Par inégalité triangulaire, on peut dire :

$$\begin{aligned} \left|\frac{1}{n}\sum_{i=1}^{n}\left(B^{(1)}_{T_*^{(n)}(X^{(n)})}(X_i) - B_\theta^{(1)}(X_i)\right)\right| &\leqslant \frac{1}{n}\sum_{i=1}^{n}\left|B^{(1)}_{T_*^{(n)}(X^{(n)})}(X_i) - B_\theta^{(1)}(X_i)\right| \\ &\leqslant \frac{1}{n}\sum_{i=1}^{n}\sup_{x\in\mathcal{X}}\left|B^{(1)}_{T_*^{(n)}(X^{(n)})}(x) - B_\theta^{(1)}(x)\right| \\ &= \sup_{x\in\mathcal{X}}\left|B^{(1)}_{T_*^{(n)}(X^{(n)})}(x) - B_\theta^{(1)}(x)\right|, \end{aligned}$$

ce qui, par (H6) converge presque sûrement vers 0.

Le lemme de Slutzky nous affirme alors que :

$$\sqrt{n}\left(T^{(n)}(X^{(n)}) - \theta\right) \xrightarrow[n\to+\infty]{\mathcal{D} \text{ sous } \mathbb{P}_\theta(n)} \frac{\mathcal{N}(0, I^{(1)}(\theta))}{\left(I^{(1)}(\theta)\right)^2} = \mathcal{N}(0, I^{(1)}(\theta)^{-1}).$$

□

Chapitre 3

Tests d'hypothèse

En se plaçant dans le modèle statistique $\left(\mathbb{R}^n, \mathcal{B}\left(\mathbb{R}^n\right), \mathcal{P}^{(n)}\right)$, avec $\mathcal{P}^{(n)} = \{\mathbb{P}_\theta^{(n)} \text{ t.q. } \theta \in \Theta\}$. On partitionne Θ en $\Theta_0 \subseteq \Theta$ et $\Theta = \Theta \setminus \Theta_0$.

On parle de *problème de test d'hypothèse* pour :

$$\begin{cases} H_0 : \theta \in \Theta_0 & \text{hypothèse nulle,} \\ H_1 : \theta \in \Theta_1 & \text{contre-hypothèse.} \end{cases} \tag{3.1}$$

Il y a deux choix possibles, à savoir RH0, le rejet de H_0, et ~~RH0~~, le non-rejet de H_0.
Exemple 3.1. Soient $X_1, \ldots, X_n$ iid $\mathcal{N}(\mu, \sigma^2)$, avec $\theta = [\mu, \sigma^2]^\mathsf{T} \in \Theta \subsetneq \mathbb{R}^2$. Voici une partition de Θ :

$$\Theta : \Theta_0 \sqcup \Theta_1 = \left\{[\mu, \sigma^2]^\mathsf{T} \in \Theta \text{ t.q. } \mu > 15\right\} \sqcup \left\{[\mu, \sigma^2]^\mathsf{T} \in \Theta \text{ t.q. } \mu \leqslant 15\right\}.$$

$$\begin{cases} \theta \in \Theta_0 \\ \theta \in \Theta_1 \end{cases}$$

Exemple 3.2. Soient $X_1, \ldots, X_n$ iid $\mathrm{Bern}(p)$, pour $p \in [0, 1] \subset \mathbb{R}$. Soit $p_0 \in [0, 1]$. Voici un test d'hypothèse :

$$\begin{cases} H_0 : p \in [0, p_0], \\ H_1 : p \in (p_0, 1]. \end{cases}$$

Définition 3.1. On dit que le problème est *unilatéral* s'il est sous la forme :

$$\begin{cases} H_0 : \theta \leqslant \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

(une partition de Θ en deux espaces de mesure non-nulle).

On dit que le problème est *bilatéral* s'il est sous la forme :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} \tag{3.2}$$

(une partition de Θ en un singleton et Θ moins un point).

Dans ce cas, on dit que $H_0 : \theta = \theta_0$ est l'*hypothèse simple*, et $H_1 : \theta \neq \theta_0$ est l'*hypothèse composée*.

Définition 3.2. Un test pur est une fonction mesurable $\phi^{(n)} : \mathcal{X}^{(n)} \to \{0,1\}$, où $\mathcal{X}$ est l'ensemble des valeurs observables pour $X^{(n)}$.
Définition 3.3. On appelle *zone critique* l'ensemble :

$$\left(\phi^{(n)}\right)^{-1}(\{1\}).$$

Définition 3.4. Un *test randomisé* (ou *test impur*) est une fonction mesurable $\phi^{(n)} : \mathcal{X}^{(n)} \to [0,1]$.
Remarque. Soit $\gamma := \phi^{(n)}(x^{(n)})$. Si $\gamma = 0$, alors ~~RH0~~. Si $\gamma = 1$, alors RH0. Et si $\gamma \in (0,1)$, alors RH0 avec probabilité γ.

De plus :

$$\mathbb{E}_\theta[\phi^{(n)}(X^{(n)})] = \int_{\mathcal{X}^{(n)}} \phi^{(n)}(x^{(n)}) L_\theta^{(n)}(x^{(n)})\, dx^{(n)} = \int_{\mathcal{X}^{(n)}} \mathbb{P}[\text{RH0}\,|\,X^{(n)} = x^{(n)}] L_\theta^{(n)}(x^{(n)})\, dx^{(n)} = \mathbb{P}[\text{RH0}].$$

Définition 3.5. Le *risque de première espèce* en $\theta \in \Theta_0$ est la valeur :

$$\mathbb{P}_\theta[\phi^{(n)}(X^{(n)}) \text{ rejette}] = \mathbb{E}_\theta[\phi^{(n)}(X^{(n)})].$$

Le *risque de seconde espèce* en $\theta \in \Theta_1$ est la valeur :

$$\mathbb{P}_\theta[\phi^{(n)}(X^{(n)}) \text{ ne rejette pas}] = 1 - \mathbb{E}_\theta[\phi^{(n)}(X^{(n)})].$$

Définition 3.6. La *dimension* d'un test $\phi^{(n)}$ en $\theta \in \Theta_0$ est $\mathbb{E}_\theta[\phi^{(n)}(X^{(n)})]$.

La *puissance* d'un test $\phi^{(n)}$ en $\theta \in \Theta_1$ est $\mathbb{E}_\theta[\phi^{(n)}(X^{(n)})]$.

3.1 Principe de Neyman

Définition 3.7. Soit $\phi^{(n)}$ un test (pur ou randomisé). On dit que $\phi^{(n)}$ est de niveau $\alpha \in (0,1)$ lorsque :

$$\forall \theta \in \Theta : \mathbb{E}_\theta^{(n)}[\phi^{(n)}(X^{(n)})] \leqslant \alpha.$$

Définition 3.8. Le test ϕ_* est dit *à puissance uniformément maximale* (PUM) dans la classe C_α lorsque :
— $\phi_* \in C_\alpha$;
— $\forall \phi \in C_\alpha : \forall \phi \in \Theta_1 : \mathbb{E}_\theta^{(n)}[\phi_*(X^{(n)})] \geqslant \mathbb{E}_\theta^{(n)}[\phi^{(n)}(X^{(n)})]$.

Considérons ici le test le plus simple possible :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases} \tag{3.3}$$

où $\Theta = \{\theta_0, \theta_1\}$. Pour $\alpha \in (0,1)$, la classe C_α est donc :

$$C_\alpha = \left\{ \phi : \mathcal{X}^{(n)} \to [0,1] \text{ t.q. } \mathbb{E}_{\theta_0}^{(n)}[\phi(X^{(n)})] \leqslant \alpha \right\}.$$

Théorème 3.9 (Lemme fondamental de Neyman)**.** *Soit le problème d'hypothèse* (3.3)*. Soit* $\alpha \in (0,1)$ *et notons* $L_\theta^{(n)}$ *la vraisemblance associée à* $\mathbb{P}_\theta^{(n)}$*. Alors :*
(i) *il existe* $(k_\alpha, \gamma_\alpha) \in \mathbb{R}^+ \times [0,1]$ *tels que le test* ϕ_* *défini par :*

$$\phi_*(x^{(n)}) = \begin{cases} 1 & \text{si } L_{\theta_1}^{(n)}(x^{(n)}) > k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \\ \gamma_\alpha & \text{si } L_{\theta_1}^{(n)}(x^{(n)}) = k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \\ 0 & \text{si } L_{\theta_1}^{(n)}(x^{(n)}) < k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \end{cases}$$

vérifie $\mathbb{E}_{\theta_0}[\phi_*]$ *;*

— ϕ_* *est à puissance uniformément maximale dans* C_α.

Démonstration. Soit F_{θ_0} la fonction de répartition de la valeur aléatoire $Y^{(n)} := \frac{L_{\theta_1}^{(n)}(X^{(n)})}{L_{\theta_0}^{(n)}(X^{(n)})}$ sous $\mathbb{P}_{\theta_0}^{(n)}$, et posons $L_{\theta_0}^{(n)}(X^{(n)}) = 0 \Rightarrow Y^{(n)} = +\infty$.
Pour $z \in \mathbb{R}$, on peut alors écrire :

$$\mathbb{E}_{\theta_0}^{(n)}[\phi_*] = \mathbb{P}_{\theta_0}[Y^{(n)} > k_\alpha] + \gamma_\alpha \mathbb{P}[Y = k^\alpha] = 1 - F_{\theta_0}(k_\alpha) + \lim_{\varepsilon \to 0} \gamma_\alpha \left(F_{\theta_0}(k_\alpha) - F_{\theta_0}(k_\alpha - \varepsilon)\right).$$

Distinguons deux cas :
— s'il existe $x \in \mathbb{R}$ tel que $F_{\theta_0}(z) = 1 - \alpha$, alors on pose=

$$(k_\alpha, \gamma_\alpha) = (z, 0).$$

On a alors bien :

$$\mathbb{E}_{\theta_0}^{(n)}[\phi_*(X^{(n)})] = 1 - F_{\theta_0}(z) + 0 = \alpha.$$

— Sinon, on pose :

$$k_\alpha := \inf\{z \in \mathbb{R} \text{ t.q. } F_{\theta_0}(z) \geqslant 1 - \alpha\}.$$

Puisque $Y \geqslant 0$ (quotient de vraisemblances, qui sont positives), on sait $k_\alpha \geqslant 0$. Par continuité de F_{θ_0} (induite par la dérivabilité de F_{θ_0}), on a :

$$F_{\theta_0}(k_\alpha - \varepsilon) < 1 - \alpha < F_{\theta_0}(k_\alpha).$$

On pose donc :

$$\gamma_\alpha := \lim_{\varepsilon \to 0} \frac{F_{\theta_0}(k_\alpha) - (1-\alpha)}{F_{\theta_0}(k_\alpha) - F_{\theta_0}(k_\alpha - \varepsilon)} \in (0, 1).$$

On a alors en effet :

$$\begin{aligned}
\mathbb{E}_{\theta_0}^{(n)}[\phi_*(X^{(n)})] &= 1 - F_{\theta_0}(k_\alpha) + \lim_{\varepsilon \to 0} \frac{F_{\theta_0}(k_\alpha) - (1-\alpha)}{F_{\theta_0}(k_\alpha) - F_{\theta_0}(k_\alpha - \varepsilon)} \left(F_{\theta_0}(k_\alpha) - F_{\theta_0}(k_\alpha - \varepsilon)\right) \\
&= 1 - F_{\theta_0}(k_\alpha) + (F_{\theta_0}(k_\alpha) - (1-\alpha)) = 1 - 1 + \alpha = \alpha.
\end{aligned}$$

Montrons alors le point 2. Prenons $\phi \in C_\alpha$, et considérons la fonction g définie par :

$$g : \mathcal{X}^{(n)} \to \mathbb{R} : x^{(n)} \mapsto \left((\phi_* - \phi)(x^{(n)})\right) \left((L_{\theta_1} - k_\alpha L_{\theta_0})(x^{(n)})\right).$$

Observons que g est définie positive car :

$$\begin{aligned}
&\forall x^{(n)} \in \mathcal{X}^{(n)} : && \left(L_{\theta_1}^{(n)}(x^{(n)}) - k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \gneqq 0\right) \Rightarrow \left(\phi_*(x^{(n)}) - \phi(x^{(n)}) = 1 - \phi(x^{(n)}) \geqslant 0\right), \\
&\text{et} && \left(L_{\theta_1}^{(n)}(x^{(n)}) - k_\alpha L_{\theta_0}^{(n)}(x^{(n)}) \lneqq 0\right) \Rightarrow \left(\phi_*(x^{(n)}) - \phi(x^{(n)}) = 0 - \phi(x^{(n)}) \leqslant 0\right).
\end{aligned}$$

On peut dès lors écrire, en distribuant g :

$$\begin{aligned}
0 \leqslant \int_{\mathcal{X}^{(n)}} g(x^{(n)})\, dx^{(n)} &= \int_{\mathcal{X}^{(n)}} \left(\phi_*(x^{(n)}) - \phi(x^{(n)})\right) L_{\theta_1}(x^{(n)})\, dx^{(n)} - k_\alpha \left(\int_{\mathcal{X}^{(n)}} \left(\phi_*(x^{(n)}) - \phi(x^{(n)})\right) L_{\theta_0}\, dx^{(n)}\right) \\
&= \mathbb{E}_{\theta_1}^{(n)}[\phi_*(X^{(n)}) - \phi(X^{(n)})] + k_\alpha(\mathbb{E}_{\theta_0}^{(n)}[\phi_*(X^{(n)}) - \phi(X^{(n)})]) \\
&= \mathbb{E}_{\theta_1}^{(n)}[\phi_*(X^{(n)})] - \mathbb{E}_{\theta_1}[\phi(X^{(n)})] - k_\alpha(\alpha - \mathbb{E}_{\theta_0}[\phi(X^{(n)})]).
\end{aligned}$$

Or, comme $\mathbb{E}_{\theta_0}[\phi(X^{(n)})] < \alpha$ (car $\phi \in C_\alpha$), on a :

$$\mathbb{E}_{\theta_1}^{(n)}[\phi_*(X^{(n)})] \geqslant \mathbb{E}_{\theta_0}^{(n)}[\phi(X^{(n)})] + k_\alpha(\alpha - \mathbb{E}_{\theta_0}[\phi(X^{(n)})] \geqslant \mathbb{E}_{\theta_1}[\phi(X^{(n)})].$$

$\square$

Théorème 3.10. *Selon les hypothèses du lemme fondamental de Neyman, si ϕ est de niveau $\alpha \in (0,1)$ et est de même puissance que ϕ_*, i.e. ils ont la même espérance sous θ_1, alors il existe une fonction $\gamma_\alpha : \mathcal{X}^{(n)} \to [0,1]$ telle que :*

$$\forall x^{(n)} \in \mathcal{X}^{(n)} : \phi(x^{(n)}) = \begin{cases} 1 & \text{si } L_{\theta_1}(x^{(n)}) < k_\alpha L_{\theta_0}(x^{(n)}) \\ \gamma_\alpha(x^{(n)}) & \text{si } L_{\theta_1}(x^{(n)}) = k_\alpha L_{\theta_0}(x^{(n)}) \\ 0 & \text{sinon}, \end{cases}$$

où k_α est celui de ϕ_*, et où le $\forall x^{(n)} \in \mathcal{X}^{(n)}$ signifie pour tout $x^{(n)}$ dans $\mathcal{X}^{(n)}$, sauf un ensemble de mesure nulle.

De plus, $\mathbb{E}_{\theta_1}[\phi_*(X^{(n)})] > \alpha$.

Démonstration. On sait que :

$$-k_\alpha(\alpha - \mathbb{E}_{\theta_0}[\phi(X^{(n)})] = \mathbb{E}_{\theta_1}^{(n)}[\phi_*(X^{(n)})] - \mathbb{E}_{\theta_1}^{(n)}[\phi(X^{(n)})] - k_\alpha(\alpha - \mathbb{E}_{\theta_0}^{(n)}[\phi(X^{(n)})] \geqslant 0.$$

Or, on sait que $k_\alpha(\alpha - \mathbb{E}_{\theta_0}^{(n)}[\phi(X^{(n)})]) \geqslant 0$. On a donc $\mathbb{E}_{\theta_0}^{(n)}[\phi(X^{(n)})] = \alpha$, et donc :

$$\int_{\mathcal{X}^{(n)}} g(x^{(n)})\, dx^{(n)} = 0,$$

c'est à dire, pour tout $x^{(n)} \in \mathcal{X}^{(n)} \setminus T$, tel que T est de mesure nulle, on a :

$$\Big(\phi_*(x^{(n)}) - \phi(x^{(n)})\Big) \cdot \Big(L_{\theta_1}^{(n)}(x^{(n)}) - k_\alpha L_{\theta_0}^{(n)}(x^{(n)})\Big) = 0.$$

Soit un tel $x^{(n)} \in \mathcal{X}^{(n)} \setminus T$. Si $L_{\theta_1}^{(n)}(x^{(n)}) \neq k_\alpha L_{\theta_0}^{(n)}(x^{(n)})$, alors :

$$0 = \phi_*(x^{(n)}) - \phi(x^{(n)}).$$

De plus, on pose $\gamma_\alpha = \phi$ sur $\{x^{(n)} \in \mathcal{X}^{(n)} \text{ t.q. } L_{\theta_1}^{(n)}(x^{(n)}) = k_\alpha L_{\theta_0}^{(n)}(x^{(n)})\}$.

Montrons que $\mathbb{E}_{\theta_1}[\phi_*(X^{(n)})] \gneqq \alpha$. Posons $\phi_\alpha : x^{(n)} \mapsto \alpha$. On sait que ϕ_* est de puissance maximale, donc :

$$\mathbb{E}_{\theta_1}[\phi_*(X^{(n)})] \geqslant \mathbb{E}_{\theta_1}[\phi_\alpha(X^{(n)})] = \alpha.$$

Supposons par l'absurde que $\mathbb{E}_{\theta_1}[\phi_*(X^{(n)})] = \alpha \in (0,1)$. Par la première partie du théorème, on a :

$$\alpha = \phi_\alpha(x^{(n)}) \notin \{0,1\}.$$

Donc :

$$\forall x^{(n)} \in \mathcal{X}^{(n)} \setminus T : L_{\theta_1}^{(n)}(x^{(n)}) = k_\alpha L_{\theta_0}^{(n)}(x^{(n)}).$$

En intégrant, on trouve :

$$1 = \int_{\mathcal{X}^{(n)}} L_{\theta_1}^{(n)}(x^{(n)})\, dx^{(n)} = k_\alpha \int_{\mathcal{X}^{(n)}} L_{\theta_0}^{(n)}(x^{(n)})\, dx^{(n)} = k_\alpha,$$

et donc $L_{\theta_0}^{(n)} = L_{\theta_1}^{(n)}$ sur $\mathcal{X}^{(n)} \setminus T$.

Pour $B \in \mathcal{B}(\mathbb{R}^n)$, on a alors $\mathbb{P}_{\theta_0}[B] = \mathbb{P}_{\theta_1}[B]$, ce qui est une contradiction car $\theta_0 \neq \theta_1$. $\square$

3.2 Tests unilatéraux à puissance uniformément maximale

Définition 3.11. Soit le modèle statistique $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, avec $\mathcal{P} = \{\mathbb{P}_\theta \text{ t.q. } \theta \in \Theta \subseteq \mathbb{R}\}$. Ce modèle est dit *à vraisemblance monotone en la statistique* $T(x^{(n)})$ lorsque :

$$\forall \theta', \theta'' \in \Theta : \theta' < \theta'' \Rightarrow \exists h_{\theta',\theta''} \text{ strict. croissante } \text{ t.q. } h_{\theta',\theta''}(T(x^{(n)})) = \frac{L_{\theta''}^{(n)}(x^{(n)})}{L_{\theta'}^{(n)}(x^{(n)})}.$$

Théorème 3.12. *Soit un modèle statistique à vraisemblance monotone en une statistique* $T(x^{(n)})$ *et soit* $\alpha \in (0,1)$. *Alors :*

— $\exists (t_\alpha^+, \gamma_\alpha) \in \mathbb{R} \times [0,1]$ *t.q.* ϕ_{*T}^+ *défini par :*

$$\phi_{*T}^+(x^{(n)}) = \begin{cases} 1 & \text{si } T(x^{(n)}) > t_\alpha^+ \\ \gamma_\alpha & \text{si } T(x^{(n)}) = t_\alpha^+ \\ 0 & \text{sinon} \end{cases}$$

vérifie $\mathbb{E}_{\theta_0}^{(n)}[\phi_{*T}^+(X^{(n)})] = \alpha$ *;*

— *la fonction* $\theta \mapsto \mathbb{E}_\theta[\phi_{*T}^+(X^{(n)})]$ *est strictement croissante ;*

— *le test* ϕ_{*T}^+ *est à PUM dans la classe des tests de niveau* α *pour le problème d'hypothèse :*

$$\begin{cases} H_0 : \theta \leqslant \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \tag{3.4}$$

avec $\theta_0 \in \operatorname{int} \Theta$.

Démonstration. Soit $F_{\theta_0}^T$ la fonction de répartition de $T(X^{(n)})$ sous $\mathbb{P}_{\theta_0}$. Comme pour le Théorème 3.9, on a :

$$\mathbb{E}_{\theta_0}^{(n)}[\phi_{*T}^+(X^{(n)})] = 1 - F_{\theta_0}^T(t_\alpha^+) + \gamma_\alpha(F_{\theta_0}(t_\alpha^+) - \lim_{\varepsilon \to 0} F_{\theta_0}^T(t_\alpha^+ - \varepsilon),$$

d'où on distingue deux cas :

— s'il existe z_α t.q. $F_{\theta_0}^T(z_\alpha) = 1 - \alpha$, on prend :

$$(t_\alpha^+, \gamma_\alpha) := (z_\alpha, 0)),$$

— et sinon, on pose :

$$t_\alpha^+ := \inf\{z \in \mathbb{R} \text{ t.q. } F_{\theta_0}^T(z) \geqslant 1 - \alpha\},$$

et :

$$\gamma_\alpha := \frac{F_{\theta_0}^T(t_\alpha^+) - (1-\alpha)}{\lim_{\varepsilon \to 0} F_{\theta_0}^T(t_\alpha^+) - F_{\theta_0}^T(t_\alpha^+ - \varepsilon)} \in (0,1).$$

Pour le second point, on prend $\theta', \theta'' \in \Theta$ tels que $\theta' < \theta''$. On réécrit :

$$\phi_{*T}^+(x^{(n)}) = \begin{cases} 1 & \text{si } h_{\theta',\theta''}(T(x^{(n)})) > h_{\theta',\theta''}(t_\alpha^+) \\ \gamma_\alpha & \textit{si } h_{\theta',\theta''}(T(x^{(n)})) = h_{\theta',\theta''}(t_\alpha^+) \\ 0 & \text{sinon} \end{cases}$$

$$= \begin{cases} 1 & \text{si } L_{\theta''}(x^{(n)}) > h_{\theta',\theta''}(t_\alpha^+) L_{\theta'}(x^{(n)}) \\ \gamma_\alpha & \text{si } L_{\theta''}(x^{(n)}) = h_{\theta',\theta''}(t_\alpha^+) L_{\theta'}(x^{(n)}) \\ 0 & \text{sinon} \end{cases}$$

On trouve que ϕ^{+}_{*T} est le test de Neyman du problème d'hypothèse :

$$\begin{cases} H_0 : \theta = \theta' \\ H_1 : \theta = \theta'' \end{cases}$$

au niveau $\mathbb{E}_{\theta'}[\phi^{+}_{*T}(X^{(n)})]$. Le Le théorème précédent donne donc :

$$\mathbb{E}_{\theta'}[\phi^{+}_{*T}(X^{(n)})] < \mathbb{E}_{\theta''}[\phi^{+}_{*T}(X^{(n)})].$$

Pour le dernier point, soient $\theta_1 > \theta_0$ et ϕ, un test de niveau α pour le problème d'hypothèse (3.4). θ_1 étant choisi arbitrairement, on a que ϕ^{+}_{*T} est de niveau α pour le problème d'hypothèse :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \tag{3.5}$$

car ϕ^{+}_{*T} est à PUM de niveau α pour le problème d'hypothèse (3.3).

Par choix de ϕ, on sait en particulier que ϕ est de niveau α pour (3.5). ϕ^{+}_{*T} domine donc ϕ uniformément en puissance. Puisque ϕ^{+}_{*T} est de niveau α pour (3.4), on a ϕ^{+}_{*T} à PUM pour (3.4). $\square$

Théorème 3.13. *Soit un modèle statistique à rapport de vraisemblance pour la statistique* $T(x^{(n)})$ *et soit* $\alpha \in (0,1)$. *Alors :*

— $\exists(t^{-}_{\alpha}, \gamma_\alpha) \in \mathbb{R} \times [0,1]$ *t.q.* ϕ^{-}_{*T} *défini par :*

$$\phi^{-}_{*T}(x^{(n)}) = \begin{cases} 1 & \textit{si } T(x^{(n)}) < t^{-}_{\alpha} \\ \gamma_\alpha & \textit{si } T(x^{(n)}) = t^{-}_{\alpha} \\ 0 & \textit{sinon} \end{cases}$$

vérifie $\mathbb{E}^{(n)}_{\theta_0}[\phi^{-}_{*T}(X^{(n)})] = \alpha$ *;*

— *la fonction* $\theta \mapsto \mathbb{E}^{(n)}_{\theta}[\phi^{-}_{*T}]$ *est strictement décroissante ;*

— *le test* ϕ^{-}_{*T} *est à PUM dans la classe des tests de niveau* α *pour le problème :*

$$\begin{cases} H_0 : \theta \geqslant \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} \tag{3.6}$$

3.3 Tests bilatéraux

Théorème 3.14. *En considérant un modèle à rapport de vraisemblance monotone, pour tout* $\alpha \in (0,1)$, *il n'existe pas de test à PUM au niveau* α *pour le problème d'hypothèse* (3.2).

Démonstration. Supposons par l'absurde qu'il existe ϕ_{*T} à PUM maximal au niveau α. On fixe $\theta_1 > \theta_0$. ϕ_{*T} est de niveau α pour le problème, et donc, on a :

$$\mathbb{E}_{\theta_1}[\phi_{*T}] \geqslant \mathbb{E}_{\theta_1}[\phi^{+}_{*T}],$$

avec ϕ^{+}_{*T}, test de Neyman pour le problème (3.3) au niveau α. On déduit donc :

$$\mathbb{E}_{\theta_1}[\phi_{*T}] = \mathbb{E}_{\theta_1}[\phi^{+}_{*T}].$$

Et puisque ϕ_{*T} est de niveau α pour le problème (3.3), et est de même puissance que ϕ^+_{*T}, on déduit (Théorème 3.10) qu'il existe $\gamma_\alpha : \mathcal{X}^{(n)} \to [0,1]$ et $k_\alpha \in \mathbb{R}$ tels que :

$$\phi_{*T}(x^{(n)}) = \begin{cases} \quad\quad\quad 1 \text{ si } L^{(n)}_{\theta_1}(x^{(n)}) > k_\alpha L^{(n)}_{\theta_0}(x^{(n)}) \\ \gamma_\alpha(x^{(n)}) \quad \text{si } L^{(n)}_{\theta_1}(x^{(n)}) = k_\alpha L^{(n)}_{\theta_0}(x^{(n)}) \\ 0 \quad\quad\quad\quad \text{sinon.} \end{cases}$$

Le modèle étant à RVM, on peut montrer que $\theta \mapsto \mathbb{E}_\theta[\phi_{*T}]$ est strictement croissante (voir Théorème 3.12). Par conséquent, pour $\theta < \theta_0$:

$$\mathbb{E}_\theta[\phi_{*T}] \lneqq \mathbb{E}_{\theta_0}[\phi_{*T}] = \alpha = \mathbb{E}_{\theta_0}[\phi^+_{*T}] < \mathbb{E}_\theta[\phi^+_{*T}],$$

ce qui contredit le fait que ϕ_{*T} soit de PUM. □

3.3.1 Tests sans biais

Définition 3.15. Le test ϕ de niveau $\alpha \in (0,1)$ pour le problème (3.1) est dit *sans biais* lorsque :

$$\forall \theta \in \Theta_1 : \mathbb{E}_\theta[\phi] \geqslant \alpha.$$

Remarque. Pour les problèmes considérés ici (tests bilatéraux sous la forme (3.2)), les tests ϕ admissibles sont donc les tests sans biais, c-à-d tels que $\mathbb{E}_{\theta_0}[\phi] \leqslant \alpha$ et pour tout $\theta \neq \theta_0$ (et donc $\in \Theta_1 = \Theta_0 \setminus \{\theta_0\}$) $\mathbb{E}_\theta[\phi] \geqslant \alpha$. On note donc que $\theta \mapsto \mathbb{E}_\theta[\phi]$ continue implique $\mathbb{E}_{\theta_0}[\phi] = 0$.

Cette définition permet de ne plus avoir les estimateurs ϕ^+_{*T} et ϕ^-_{*T}. En effet, leur fonction de puissance $\theta \mapsto E_\theta[\Phi]$ pour $\Phi \in \{\phi^+_{*T}, \phi^-_{*T}\}$ sont strictement croissantes. Donc pour tout $\theta < \theta_0$, on a $\mathbb{E}_\theta[\Psi] < \alpha$.

Remarque. Notons alors, pour $\alpha \in (0,1)$, $\mathfrak{B}_\alpha$ la classe des tests sans biais au niveau α.

Théorème 3.16. *Soit un modèle statistique à RVM tel que les vraisemblances s'écrivent sous la forme :*

$$L^{(n)}_\theta(x^{(n)}) = c_\theta h(x^{(n)}) \exp(\theta T(x^{(n)})).$$

Soit $\alpha \in (0,1)$ et soit le problème (3.2). Alors :

— $\exists (t^1_\alpha, t^2_\alpha) \in \mathbb{R}^2$ *et* $(\gamma^1_\alpha, \gamma^2_\alpha) \in [0,1]^2$ *tels que le test* ϕ_{*T} *défini par :*

$$\phi_{*T}(x^{(n)}) = \begin{cases} 1 & \textit{si } T(x^{(n)}) \notin [t^1_\alpha, t^2_\alpha] \\ \gamma^1_\alpha & \textit{si } T(x^{(n)}) = t^1_\alpha \\ \gamma^2_\alpha & \textit{si } T(x^{(n)}) = t^2_\alpha \\ 0 & \textit{sinon} \end{cases}$$

vérifie $\mathbb{E}_{\theta_0}[\phi_{*T}] = \alpha$ *et* $\mathbb{E}_{\theta_0}[\phi_{*T} \cdot T] = \alpha \cdot \mathbb{E}_{\theta_0}[T]$ *;*

— ϕ_{*T} *est à PUM dans* $\mathfrak{B}_\alpha$ *pour le problème donné.*

Remarque. Un tel modèle paramétrique est donc à RVM en la statistique T.

3.4 Tests de rapport de vraisemblance

Définition 3.17. On définit le rapport de vraisemblance suivant, noté Λ :

$$\Lambda : \mathcal{X}^{(n)} \to [0,1] : x^{(n)} \mapsto \Lambda(x^{(n)}) := \frac{\sup_{\theta \in \Theta_0} L^{(n)}_\theta(x^{(n)})}{\sup_{\theta \in \Theta} L^{(n)}_\theta(x^{(n)})}.$$

Théorème 3.18. *Soit un modèle statistique avec $\Theta \subset \mathbb{R}^k$, avec $r \leqslant k$ paramètres indépendants. Sous certaines hypothèses de régularité, on a :*

$$\forall \theta \in \Theta_0 : -2 \log \Lambda(X^{(n)}) \xrightarrow[n \to +\infty]{\mathcal{D}} \chi^2_{k-r}.$$

3.5 Tests χ^2

3.5.1 Loi multinomiale

Soit une expérience aléatoire E à valeurs dans $\{A_1, \ldots, A_k\}$ de probabilités $p_1, \ldots, p_k$.
Définition 3.19. Soit $N = (N_1, \ldots, N_k) \in \mathbb{N}^k$ tel que $\sum_{i=1}^k N_i = n$, où n est le nombre d'expériences indépendantes de l'expérience E. (N_i désigne le nombre d'occurrences de A_i). La loi du vecteur aléatoire N est dite *multinomiale* de paramètres $n, p_1, \ldots, p_{k-1}$. On note alors :

$$N \sim \mathrm{Mult}(n, p_1, \ldots, p_{k-1}).$$

Remarque. Pour $1 \leqslant i \leqslant k$, on a $N_i \sim \mathrm{Bin}(n, p_i)$, d'où le nom de multinomiale. On en déduit :

$$\mathrm{Var}[N_i] = np_i(1 - p_i).$$

Prenons $i \neq j$. Par stabilité des binomiales, on trouve :

$$\mathrm{Var}[N_i + N_j] = n(p_i + p_j)(1 - (p_i + p_j)) = np_i - np_i^2 - 2np_ip_j + np_j - np_j^2,$$

et :

$$\begin{aligned}\mathrm{Var}[N_i + \mathbb{N}_j] = \mathrm{Var}[N_i] + \mathrm{Var}[N_j] + 2\,\mathrm{Cov}[N_i, N_j] &= np_i(1 - p_i) + np_j(1 - p_j) + 2\,\mathrm{Cov}[N_i, N_j] \\ &= np_i - np_i^2 + np_j - np_j^2 + 2\,\mathrm{Cov}[N_i, N_j].\end{aligned}$$

On en déduit :

$$\mathrm{Cov}[N_i, N_j] = -np_ip_j.$$

On observe alors une dépendance entre N_i et N_j, et avec une corrélation négative.
Théorème 3.20. *Soit* $N \sim \mathrm{Mult}(n, p_1, \ldots, p_{k+1})$. *Alors :*

$$\sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \xrightarrow[n \to +\infty]{\mathcal{D}} \chi^2_{k-1}.$$

Théorème 3.21. *Soit* $N \sim \mathrm{Mult}(n, p_1(\theta), \ldots, p_{k-1}(\theta))$, *avec* $\theta \in \Theta \subseteq \mathbb{R}^d$ *t.q.* $\operatorname{int} \Theta \neq \emptyset$. *Si* $\hat{\theta}$ *est l'estimateur du maximum de vraisemblance de* θ, *alors :*

$$\sum_{i=1}^{k} \frac{(N_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})} \xrightarrow[n \to +\infty]{\mathcal{D}} \chi^2_{k-1-d} \ \textit{sous } \mathbb{P}_\theta \ \forall \theta \in \Theta.$$

Remarque. Il est commun d'utiliser ces propriétés de convergence pour approximer N à l'aide de χ^2 à partir du moment où :
- $n \geqslant 30$;
- $\forall i \in \{1, \ldots, k\} : np_i \geqslant 1$;
- $\left|\left\{i \in \{1, \ldots, k\} \text{ t.q. } np_i \geqslant 5\right\}\right| \geqslant 0.8k$.

3.5.2 Tests d'ajustement et d'adéquation

Notons $p = (p_1, \ldots, p_k)$ le vecteur de probabilités d'une variable multinomiale. On dénotera par $\overline{p} \in [0,1]^k$ un vecteur de probabilités fixé (donc tel que $\langle \overline{p}, \mathbf{1} \rangle = 1$). Intéresons-nous au problème d'hypothèse :

$$\left\{ \begin{array}{l} H_0 : p = \overline{p} \\ H_1 : p \neq \overline{p} \end{array} \right. \tag{3.7}$$

Définition 3.22. Désignons par $\chi^2_{\ell,\alpha}$ le quantile d'ordre α d'une variable χ^2 à ℓ degrés de liberté.

Par le Théorème 3.20, on définit le statistique suivante :

$$Q^{(n)} = Q^{(n)}(N) := \sum_{i=1}^{k} \frac{(N_i - \overline{p}_i)^2}{n\overline{p}_i},$$

justifiée par le fait que sous H_0, $n\overline{p}_i = \mathbb{E}[N_i]$.
Remarque. Cette statistique peut-être vue comme une certaine *distance* entre N et H_0.

Pour $\alpha \in (0,1)$, on définit alors le test pur :

$$\phi_{\alpha, Q^{(n)}} : \mathcal{N} \to \{0,1\} : N \mapsto \begin{cases} 1 & \text{si } Q^{(n)}(N) > \chi^2_{k-1, 1-\alpha} \\ 0 & \text{sinon,} \end{cases}$$

qui donc rejette H_0 lorsque la valeur de $Q^{(n)}$ est *trop grande*, c-à-d au-delà du seuil déterminé par le quantile d'ordre $1-\alpha$.
Remarque. Le théorème 3.20 implique que ce test $\phi_{\alpha, Q^{(n)}}$ est asymptotiquement de niveau α.

Afin de ne pas recalculer systématiquement $\chi^2_{k-1,1-\alpha}$ pour différentes valeurs de $\alpha \in (0,1)$, on définit la p-valeur.
Définition 3.23. La p-valeur définie par :

$$\text{p-valeur} := P^{(n)}_{H_0}[Q^{(n)} \geqslant q^{(n)}],$$

avec $q^{(n)}$ la valeur calculée de la statistique $Q^{(n)}$ est un indicateur d'acceptation de l'hypothèse nulle.
Remarque. La p-valeur $\in [0,1]$ donne une idée de la présomption d'une hypothèse nulle fausse. De faibles valeurs pour p-valeur (de l'ordre de $< 10^{-2}$) indiquent un très haute présomption d'hypothèse nulle fausse, alors que de hautes valeurs (de l'ordre de $> 10^{-1}$) indiquent qu'il n'y a pas de présomption d'hypothèse nulle fausse.

Attention : une haute p-valeur ne veut pas dire que l'hypothèse nulle est vraie, mais qu'il est peu probable qu'elle soit à rejeter.

Le problème d'hypothèse (3.7) peut être élargi à un modèle paramétrique de la forme :

$$\left\{ \begin{array}{l} H_0 : \exists \theta \in \Theta \text{ t.q. } p = p(\theta) \\ H_1 : \forall \theta \in \Theta : p \neq p(\theta) \end{array} \right. \tag{3.8}$$

avec $\Theta \subset \mathbb{R}^d$ tel que $\operatorname{int} \Theta \neq \emptyset$.

On définit alors la statistique :

$$Q^{(n)} = Q^{(n)}(N) := \sum_{i=1}^{k} \frac{(N_i - np_i(\hat{\theta}))^2}{np_i(\hat{\theta})}.$$

Pour cette nouvelle statistique, le test $\phi_{\alpha, Q^{(n)}}$ est à nouveau asymptotiquement de niveau α pour le problème (3.8), par le Théorème 3.21.

3.5.3 Tests d'homogénéité

Les tests d'homogénéité correspondent aux situations de vecteurs de multinomiales :

$$N = (N_1, \ldots, N_\ell) \quad \text{t.q. } \forall 1 \leqslant j \leqslant \ell : N_j = (N_{1j}, \ldots, N_{kj}).$$

Définition 3.24. Un test d'homogénéité est un problème d'hypothèse sous la forme :

$$\begin{cases} H_0 : \forall i \in [\![1, k]\!] : \forall j \in [\![2, \ell]\!] : p_{i\,j-1} = p_{i\,j} \\ H_1 : \exists (i, j) \in [\![1, k]\!] \times [\![2, \ell]\!] \text{ t.q. } p_{i\,j-1} \neq p_{ij}. \end{cases} \tag{3.9}$$

Dans un tel cas, par indépendance des N_j, on étend le Théorème 3.20 en :

$$\sum_{j=1}^{\ell} \sum_{i=1}^{k} \frac{(N_{ij} - n_j p_{ij})^2}{n_j p_{ij}} \xrightarrow[n \to +\infty]{\mathcal{D}} \chi^2_{\ell \cdot (k-1)}.$$

Les χ^2 étant des sommes de normales, sommer des χ^2, donne une χ^2 avec pour nombre de degrés de libertés la somme des degrés de libertés des variables sommées.

Sous l'hypothèse nulle, les p_{ij} sont constantes selon j, et donc la convergence peut être reformulée en :

$$\sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{(N_{ij} - n_j p_i)^2}{n_j p_i} \xrightarrow[n \to +\infty]{\mathcal{D}} \chi^2_{\ell \cdot (k-1)},$$

avec p_i l'unique valeur selon les j tels que $\sum_i p_i = 1$.

Les estimateurs de maximum de vraisemblance des p_i sont donnés par :

$$\hat{p_i} = \frac{N_{i\cdot}}{n} := \frac{1}{n} \sum_{j=1}^{\ell} N_{i\,j}.$$

En considérant les $k - 1$ valeurs p_i comme indépendantes, sous H_0, et par le Théorème 3.21, on a :

$$\sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{(N_{ij} - n_j \hat{p_i})^2}{n_j \hat{p_i}} = \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{\left(N_{ij} - \frac{n_j N_{i\cdot}}{n}\right)^2}{\frac{n_j N_{i\cdot}}{n}} \xrightarrow[n \to +\infty]{\mathcal{D}} \chi^2_{(k-1)(\ell-1)}.$$

On définit alors le test pur de niveau asymptotique $\alpha \in (0, 1)$:

$$\phi_\alpha : N \mapsto \begin{cases} 1 & \text{si } \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{n\left(N_{ij} - \frac{n_j N_{i\cdot}}{n}\right)^2}{n_j N_{i\cdot}} > \chi^2_{(k-1)(\ell-1), 1-\alpha} \\ 0 & \text{sinon} \end{cases}$$

3.5.4 Tests d'indépendance

Pour un test d'indépendance, il faut un vecteur aléatoire (X, Y) avec X ayant k valeurs possibles, et Y, ℓ valeurs pour $k, \ell \geqslant 2$.

On définit la distribution jointe par les p_{ij}, $(i, j) \in [\![1, k]\!] \times [\![1, \ell]\!]$.

Définition 3.25. On définit les distributions marginales par :

$$p_{i\cdot} = \mathbb{P}[X = x_i] = \sum_{j=1}^{\ell} p_{ij} \qquad \text{et} \qquad p_{\cdot j} = \mathbb{P}[Y = y_j] = \sum_{i=1}^{k} p_{ij}.$$

Le but d'un test d'indépendance est d'avoir pour hypothèse nulle l'indépendance des n observations iid $(X_t, Y_t), 1 \leqslant t \leqslant n$:

$$\begin{cases} H_0 : \forall (i,j) \in [\![1,k]\!] \times [\![1,\ell]\!] : p_{ij} = p_{i\cdot}p_{\cdot j} \\ H_1 : \exists (i,j) \in [\![1,k]\!] \times [\![1,\ell]\!] \text{ t.q. } p_{ij} \neq p_{i\cdot}p_{\cdot j}. \end{cases} \tag{3.10}$$

À nouveau, selon H_0, on a :

$$\sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{(N_{ij} - np_{i\cdot}p_{\cdot j})^2}{np_{i\cdot}p_{\cdot j}} \xrightarrow[n\to+\infty]{\mathcal{D}} \chi^2_{(k\ell)-1}.$$

Les estimateurs de maximum de vraisemblance étant :

$$\hat{p_{i\cdot}} = \frac{N_{i\cdot}}{n} \qquad \text{et} \qquad \hat{p_{\cdot j}} = \frac{N_{\cdot j}}{n},$$

on a alors, sous H_0 toujours :

$$\sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{N\left(N_{ij} - \frac{N_{i\cdot}N_{\cdot j}}{N}\right)^2}{N_{i\cdot}N_{\cdot j}} \xrightarrow[n\to+\infty]{\mathcal{D}} \chi^2_{(k-1)(\ell-1)}.$$

Pour $\alpha \in (0,1)$, le test est alors défini par :

$$\phi_\alpha : N \mapsto \begin{cases} 1 & \text{si } \sum_{i=1}^{k} \sum_{j=1}^{\ell} \frac{N\left(N_{ij} - \frac{N_{i\cdot}N_{\cdot j}}{N}\right)^2}{N_{i\cdot}N_{\cdot j}} > \chi^2_{(k-1)(\ell-1),1-\alpha} \\ 0 & \text{sinon} \end{cases}$$

et est de niveau asymptotique α.

Chapitre 4

Intervalles de confiance

Définition 4.1. Soit le modèle statistique à paramètre θ scalaire $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P}^{(n)}\right)$, où $\mathcal{P}^{(n)} = \{\mathbb{P}_\theta^{(n)} \text{ t.q. } \theta \in \Theta \subset \mathbb{R}\}$ et fixons $\alpha \in (0,1)$. Alors :

$$\left[L(X^{(n)}), U(X^{(n)})\right]$$

est un *intervalle de confiance* pour θ au *niveau de confiance* $1-\alpha$ lorsque les statistiques L et U vérifient :

— $\forall\theta \in \Theta : U(X^{(n)}) \leqslant U(X^{(n)})$ presque sûrement sous $\mathbb{P}_\theta^{(n)}$;

— $\forall\theta \in \Theta : \mathbb{P}_\theta^{(n)}\left[\theta \in \left[L(X^{(n)}), U(X^{(n)})\right]\right] \geqslant 1-\alpha$.

4.1 Intervalles exacts

4.1.1 Intervalles sans paramètre de nuisance

Définition 4.2. La fonction $g : \mathcal{X}^{(n)} \times \Theta \to \mathbb{R}$ est *pivotale* pour θ lorsque :

— $\forall\theta \in \Theta : x^{(n)} \mapsto g(x^{(n)}, \theta)$ est mesurable ;

— $\forall x^{(n)} \in \mathcal{X}^{(n)} : \theta \mapsto g(x^{(n)}, \theta)$ est inversible ;

— la distribution de $g(X^{(n)}, \theta$ sous $\mathbb{P}_\theta^{(n)}$ ne dépend pas de θ.

Remarque. L'hypothèse d'inversibilité de $\theta \mapsto g(X^{(n)}, \theta)$ permet de réécrire :

$$\forall\theta \in \Theta : \mathbb{P}_\theta^{(n)}\left[a \leqslant g(X^{(n)}, \theta) \leqslant b\right] \geqslant 1-\alpha,$$

en :

$$\forall\theta \in \Theta : \mathbb{P}_\theta^{(n)}\left[\theta \in g^{-1}(X^{(n)}, \cdot)([a,b])\right] \geqslant 1-\alpha,$$

où $g^{-1}(X^{(n)}, \cdot)([a,b])$ est l'image de $[a,b]$ par la réciproque de $\theta \mapsto g(X^{(n)}, \theta)$, que l'on peut également écrire :

$$g^{-1}(X^{(n)}, \cdot)([a,b]) = g^{-1}(X^{(n)}, [a,b]).$$

On veut également pouvoir lever le degré de liberté dû à a et b en imposant que l'intervalle de confiance soit de longueur minimale.

4.1.2 Intervalles avec paramètres de nuisance

Considérons maintenant le modèle statistique où $\mathcal{P}^{(n)} = \{\mathbb{P}_\theta^{(n)} \text{ t.q. } \theta \in \Theta \subset \mathbb{R}^d\}$.

Définition 4.3. Fixons $\alpha \in (0,1)$. L'intervalle :

$$\left[L(X^{(n)}), U(X^{(n)})\right]$$

est un *intervalle de confiance* pour θ_1 au *niveau de confiance* $1-\alpha$ lorsque les statistiques L et U vérifient :

- $\forall\theta \in \Theta : L(X^{(n)}) \leqslant U(X^{(n)})$ presque sûrement sous $\mathbb{P}_\theta^{(n)}$;
- $\forall\theta \in \Theta : \mathbb{P}_\theta^{(n)}\left[\theta_1 \in \left[L(X^{(n)}), U(X^{(n)})\right]\right] > 1-\alpha$.

Définition 4.4. Notons Θ_1 l'ensemble des valeurs prises par θ_1 pour $\theta \in \Theta$ (donc $\pi(\Theta)$ pour $\pi : \theta \mapsto \theta_1$). La fonction $g : \mathcal{X}^{(n)} \times \Theta_1 \to \mathbb{R}$ est dite *pivotale* pour θ_1 lorsque :

- $\forall\theta_1 \in \Theta_1 : x^{(n)} \mapsto g(x^{(n)}, \theta_1)$ est mesurable ;
- $\forall x^{(n)} \in \mathcal{X}^{(n)} : \theta_1 \mapsto g(x^{(n)}, \theta_1)$ est inversible ;
- la distribution de $g(X^{(n)}, \theta_1)$ sous $\mathbb{P}_\theta^{(n)}$ ne dépend pas de θ.

Remarque. Cela dit que dans le cas de θ de dimension $\gneqq 1$, on détermine séparément un intervalle de confiance pour chaque coordonnée, ce qui permet d'avoir une estimation pour chacune des coordonnées, et donc détermine un pavé de $\mathbb{R}^k$.

4.2 Intervalles asymptotiques

Lorsqu'il n'est pas possible de construire explicitement une fonction pivotale, il est toujours possible de construire des intervalles de confiances dits asymptotiques, qui sont, en pratique (en supposant que l'échantillon est suffisamment grand), aussi bon que leur équivalents exacts.

Définition 4.5. Soit un modèle statistique paramétrique avec $\Theta \subseteq \mathbb{R}^k$ et fixons $\alpha \in (0,1)$.

$$\left[L(X^{(n)}), U(X^{(n)})\right]$$

est un *intervalle de confiance asymptotique* pour θ_1 au *niveau de confiance* $1-\alpha$ lorsque les statistiques L et U vérifient :

- $\forall\theta \in \Theta : L(X^{(n)}) \leqslant U(X^{(n)})$ presque sûrement sous $\mathbb{P}_\theta^{(n)}$;
- $\forall\theta \in \Theta : \mathbb{P}_\theta^{(n)}\left[\theta_1 \in \left[L(X^{(n)}), U(X^{(n)})\right]\right] \xrightarrow[n\to+\infty]{} 1-\alpha$.

Remarque. Dans le cas où on possède un estimateur $T^{(n)}(X^{(n)})$ de θ faiblement convergent et tel que :

$$\forall\theta \in \Theta : \frac{\sqrt{n}\left(T_1^{(n)}(X^{(n)}) - \theta_1\right)}{\sigma(\theta)} \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0,1) \quad \text{sous } \mathbb{P}_\theta^{(n)},$$

avec $\sigma : \Theta \to \mathbb{R}^+$ continue, on peut déterminer une fonction *asymptotiquement pivotale pour* θ_1 par :

$$g : \mathcal{X}^{(n)} \times \Theta_1 : (x^{(n)}, \theta_1) \mapsto \frac{\sqrt{n}\left(T_1^{(n)}(x^{(n)}) - \theta_1\right)}{\sigma(T^{(n)}(x^{(n)}))}.$$

On observe également que :

$$g(X^{(n)}, \theta) = \frac{\sigma(\theta)}{\sigma(T^{(n)}(X^{(n)}))}\frac{\sqrt{n}\left(T_1^{(n)}(X^{(n)}) - \theta_1\right)}{\sigma(\theta)} \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0,1) \quad \text{sous } \mathbb{P}_\theta^{(n)}.$$

Pour $\alpha \in (0,1)$, on a alors :

$$\forall \theta \in \Theta : \mathbb{P}_\theta^{(n)} \left[-\Phi^{-1}\left(1-\frac{\alpha}{2}\right) \leqslant g(X^{(n)}, \theta_1) \leqslant \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \right] \xrightarrow[n \to +\infty]{} 1-\alpha,$$

ce qui donne pour intervalle de confiance **asymptotique** pour θ_1 au niveau $1-\alpha$:

$$\left[T_1^{(n)}(X^{(n)}) \pm \frac{\sigma(T_1^{(n)}(X^{(n)}))}{\sqrt{n}} \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \right].$$

4.3 Liens avec les tests d'hypothèses

On peut lier tests d'hypothèses et intervalles de confiance en considérant un problème d'hypothèse sous la forme (3.2) pour lequel on dispose d'un intervalle de confiance $[L(X^{(n)}), U(X^{(n)})]$ pour θ au niveau $1-\alpha$. On peut construire un test :

$$\phi : \mathcal{X}^{(n)} \to \{0,1\} : x^{(n)} \mapsto \begin{cases} 1 & \text{si } \theta_0 \notin [L(x^{(n)}), U(x^{(n)})] \\ 0 & \text{sinon.} \end{cases}$$

Remarque. On peut calculer l'erreur de première espèce :

$$\mathbb{E}_{\theta_0}[\phi(X^{(n)})] = \mathbb{P}_{\theta_0}^{(n)} \left[\theta_0 \notin \left[L(X^{(n)}), U(X^{(n)}) \right] \right] = 1 - \mathbb{P}_\theta^{(n)} \left[\theta_0 \in \left[L(X^{(n)}), U(X^{(n)}) \right] \right] \leqslant 1-(1-\alpha) = \alpha.$$

Ce test est donc de niveau α pour le problème (**??**).

Inversement, si ϕ_{θ_0} est un test pur de niveau $\alpha \in (0,1)$ pour ce même problème, on peut définir :

$$A(x^{(n)}) := \{\theta_0 \in \Theta \text{ t.q. } \phi_{\theta_0}(x^{(n)}) = 0\}.$$

Cet ensemble correspond à l'ensemble des valeurs de θ_0 qui ne seront pas rejetées si $x^{(n)}$ est observé.

En prenant $A(X^{(n)})$ pour intervalle de confiance (pour peu que $A(X^{(n)})$ soit un intervalle) pour θ, on a un intervalle de confiance de niveau $1-\alpha$ car :

$$\mathbb{P}_\theta^{(n)} \left[\theta \in A(X^{(n)}) \right] = \mathbb{P}_\theta^{(n)}[\phi_\theta(X^{(n)}) = 0] = 1 - \mathbb{P}_\theta[\phi_\theta(X^{(n)}) = 1] \geqslant 1-\alpha.$$

Remarque. Ce lien permet de construire des intervalles de confiance *optimaux* en un sens puisque au chapitre précédent, des tests optimaux (à PUM) ont été définis et démontrés alors qu'il n'y a pas cette notion directement dans les intervalles de confiance.

4.4 Zones de confiance

Définition 4.6. Soit un modèle statistique paramétrique à paramètre vectoriel dans $\Theta \subseteq \mathbb{R}^k$ et fixons $\alpha \in (0,1)$. La statistique $R^{(n)}(X^{(n)})$ à valeurs dans $\mathcal{P}(\Theta)$ est une *zone de confiance* pour $\theta \in \Theta \subset \mathbb{R}^k$ au niveau de confiance $1-\alpha$ lorsque :

$$\forall \theta \in \Theta : \mathbb{P}_\theta^{(n)}[\theta \in R^{(n)}(X^{(n)})] \geqslant 1-\alpha.$$

Remarque. Cette définition n'impose pas que $R(X^{(n)})$ soit connexe. En dimension 1, une zone de confiance n'est donc pas forcément un intervalle de confiance. De plus, en dimension $\geqslant 2$, les zones de confiance ne sont pas forcément des produits d'intervalles.

Il peut alors être intéressant de prendre un hyper-rectangle défini par le produit des intervalles obtenus indépendamment pour les θ_i. Notons tout de même que souvent :

$$\mathbb{P}_\theta^{(n)}[\theta \in R(X^{(n)})] \lneqq \mathbb{P}_\theta^{(n)}\left[\forall i \in [\![1, k]\!] : \theta_i \in [L_i(X^{(n)}), U_i(X^{(n)})]\right].$$

Le niveau de confiance d'une telle zone de confiance est alors difficile à définir. Cependant, on observe que si tous les intervalles $[L_i(X^{(n)}), U_i(X^{(n)})]$ sont de niveau $1 - \frac{\alpha}{k}$ pour k, la dimension de Θ, on a une zone de confiance de niveau $1 - \alpha$. En effet :

$$\begin{aligned}\mathbb{P}_\theta^{(n)}[\theta \in R(X^{(n)})] &= 1 - \mathbb{P}_\theta^{(n)}\left[\bigcup_{i=1}^{k}\left(\theta_i \in \left[L_i(X^{(n)}), U_i(X^{(n)})\right]\right)\right] \\ &\geqslant 1 - \sum_{i=1}^{k}\mathbb{P}_\theta^{(n)}\left[\theta_i \in \left[L_i(X^{(n)}), U_i(X^{(n)})\right]\right].\end{aligned}$$