Université Paris I, Panthéon - Sorbonne

Première année Master M.A.E.F. 2007 – 2008

Statistiques

Ce livret présente :

1. **Les feuilles d'exercices de TD associées au cours;**

2. **Un plan détaillé du cours avec bibliographie;**

3. **Les énoncés (avec corrections) des interrogations et examens des années 2005/2006 et 2006/2007.**

Cours de J.-M. Bardet

Feuille n^o 1 : **Révisions sur les variables et vecteurs aléatoires**

1. **(*)** On considère deux v.a. X et Y définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Supposons que X et Y ont la même loi. Est-t-il nécessairement vrai que $\forall \omega \in \Omega, \quad X(\omega) = Y(\omega)$?

2. **(*)** L'éclairage d'une pièce se compose de deux néons montés en série (ce qui signifie que chaque ampoule ne fonctionne que lorsque les deux fonctionnent). La durée de vie de chaque néon suit une loi de exponentielle de moyenne un an. On note T la durée de vie du système d'éclairage. Donner la loi de T, son espérance et sa variance.

3. **(*)** Un point est choisi au hasard sur un segment de longueur L. Interpréter cet énoncé et trouver la probabilité que le rapport entre le plus petit et le plus grand segment soit inférieure à $1/4$.

4. **(*)** La taille moyenne de 500 élèves de collège est 151 cm et l'écart-type 15 cm. En supposant que la taille soit distribuée suivant une loi normale, trouver combien d'élèves ont leur taille comprise entre (a) 120 et 155 cm (b) 185 cm et plus.

5. **(*)** La durée de vie d'un composant électrique peut être représentée par une variable aléatoire réelle X de densité de probabilité f définie par

$$f(x) = \begin{cases} kx(5-x) & \text{si } 0 \leq x \leq 5 \\ 0 & \text{sinon} \end{cases}$$

   (a) Calculer la valeur de k. Représenter graphiquement la fonction de répartition F_X de X.

   (b) Calculer $\mathbb{E}(X)$ et $\text{var}(X)$. Déterminer la médiane de X, c'est-à-dire le réel m tel que $P(X \leq m) = 0.5$.

6. **(**)** Sur $(\Omega, \mathcal{A}, P)$ un espace de probabilité, on considère une v.a. réelle positive X de fonction de répartition F_X. Déterminer dans les 2 cas suivants l'espérance et la variance de X :

$$\begin{aligned} F_X(t) &= \frac{1}{2}\left(e^t \mathbb{I}_{]-\infty,0[}(t) + (2-e^{-t})\mathbb{I}_{[0,\infty[}(t)\right); \\ F_X(t) &= \frac{1}{4}(t+2)\mathbb{I}_{[-1,0[\cup[1,2[}(t) + \frac{3}{4}\mathbb{I}_{[0,1]}(t) + \mathbb{I}_{[2,\infty[}(t). \end{aligned}$$

7. **(*)** Soit $X \sim \text{U}([0,1])$ (loi uniforme sur $[0,1]$). Trouver la loi de $Y = \exp(X)$ et de $Z = X^{1/n}$, $n \in N$.

8. **(***)** Soit X une v.a. réelle normale centrée réduite. Soit la v.a. $Y = e^X$. On dit que Y suit une loi log-normale.

   (a) Montrer que Y à une mesure de probabilité absolument continue par rapport à la mesure de Lebesgue de densité $f_Y(y) = \dfrac{1}{\sqrt{2\pi}}\dfrac{1}{z}e^{-\ln^2(z)/2}$ si $z > 0$ et 0 sinon.

   (b) Pour $a \in [-1,1]$, soit Y_a la v.a. de densité $f_a(y) = f_Y(y)(1 + a\sin(2\pi \ln(y)))$. Montrer que Y et Y_a ont les mêmes moments, et en déduire que les moments ne caractérisent pas une loi de probabilité.

9. **(*)** Soit X de loi de Poisson $\mathcal{P}(\lambda)$ et Y de loi de Poisson $\mathcal{P}(\mu)$. En supposant que X et Y sont indépendantes, donner la loi de la somme $X + Y$, en justifiant la réponse par les calculs.

10. **(**)** Soit X une v.a. réelle normale centrée réduite et soit Y une v.a. indépendante de X, à valeurs dans $\{-1,1\}$ telle que $P(Y=1)=0.5$. On considère la v.a. $Z = X \cdot Y$. Déterminer la mesure de probabilité de Z, puis celle de $X+Z$. En déduire que la somme de 2 variables gaussiennes peut ne pas être gaussienne.

11. **(**)** Soit X et Y deux v.a. indépendantes définies sur le même espace de probabilité (Ω, P), X suivant une loi de Bernoulli de paramètre p, et Y suivant une loi de Poisson de paramètre λ. On pose pour tout $\omega \in \Omega$,
$$Z(\omega) = \begin{cases} 0 & \text{si} \quad X(\omega) = 0 \\ Y(\omega) & \text{si} \quad X(\omega) = 1 \end{cases}$$
    (a) Montrer que pour tout $\omega \in \Omega \quad Z(\omega) = X(\omega)Y(\omega)$.
    (b) Calculer l'espérance et la variance de Z.

12. **(*)** Soit X et Y deux variables définies sur le même espace de probabilité $(\Omega, \mathcal{A}, P)$ et absolument continues par rapport à la mesure de Lebesgue sur $\mathbb{R}$ (de densité f_X et f_Y). Déterminer la densité de $X+Y$ lorsque X et Y sont indépendantes. La calculer explicitement dans les cas où X et Y sont des v.a. uniformes sur [0,1], puis des lois exponentielles de paramètres respectifs λ et μ.

13. **(**)** Soit X et Y deux variables aléatoires telles que le vecteur $Z = (X,Y)$ soit absolument continu par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$, de densité f_Z telle que :
$$f_Z(x,y) = k \cdot \frac{|x|}{(x^2+|y|+1)^4} \quad \text{pour} \quad (x,y) \in \mathbb{R}^2.$$
    Déterminer la valeur de k. Déterminer la loi de X et celle de Y. Déterminer leurs espérances, leurs variances et $\mathrm{cov}(X,Y)$. Les variables X et Y sont-elles indépendantes ?

14. **(**)** Soit $(\varepsilon_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. gaussiennes centrées réduites. Soit $a \in]-1,1[$. On considère la suite de v.a. $(X_n)_{n\in\mathbb{N}}$ telle que :
    (a) $X_0 = 0$ et $X_{n+1} = a \cdot X_n + \varepsilon_{n+1}$ pour $n \in \mathbb{N}$. Déterminer la loi de X_n, puis celle de $(X_1, \cdots, X_n)$. Les variables X_i sont-elles indépendantes ?
    (b) $X_0 \sim \mathcal{N}(0, (1-a^2)^{-1})$ et $X_{n+1} = a \cdot X_n + \varepsilon_{n+1}$ pour $n \in \mathbb{N}$. Déterminer la loi de X_n, puis celle de $(X_1, \cdots, X_n)$. Les variables (X_i) sont-elles indépendantes ?
    (c) $X_0 = 0$ et $X_{n+1} = \varepsilon_{n+1} - \varepsilon_n$ pour $n \in \mathbb{N}$. Déterminer la loi de X_n, puis celle de $(X_1, \cdots, X_n)$. Les variables (X_i) sont-elles indépendantes ?

15. **(***)** On considère $X = (X_1, X_2)$ un vecteur aléatoire à valeurs dans $\mathbb{R}^2$. On suppose que X est absolument continue, c'est-à-dire que la mesure de probabilité de X est absolument continue par rapport à la mesure de Lebesgue λ_2 sur $\mathbb{R}^2$.
    (a) Montrer alors que la loi de X_1 admet une densité f_1 par rapport à la mesure de Lebesgue λ_1 sur $\mathbb{R}$, que l'on exprimera en fonction de f.
    (b) Calculer f_1 et f_2 pour f telle que :
$$f(x_1,x_2) = \begin{cases} e^{-x_1} & \text{si } x_1 \geq x_2 \geq 0 \\ 0 & \text{sinon.} \end{cases}$$
    A-t-on $f(x_1,x_2) = f_1(x_1)f_2(x_2)$ pour λ_2-presque tout $(x_1,x_2) \in \mathbb{R}^2$? Quelle conclusion en tirer sur X_1 et X_2?
    (c) On suppose maintenant que $X = (X_1, X_1)$ où X_1 est absolument continue par rapport à λ_1. Le vecteur aléatoire X est-il absolument continu par rapport à la mesure de Lebesgue λ_2 sur $\mathbb{R}^2$?

16. **(**)** Soit $(X)_n$ une suite des v.a.i.i.d. de loi de Bernoulli $\mathcal{B}(p)$ avec $0 < p < 1$. Pour $n \geq 1$, on pose $Y_n = X_n \cdot X_{n+1}$.

    (a) Donner la loi de Y_n. Calculer $E(Y_n)$ et $\operatorname{Var}(Y_n)$.

    (b) Calculer $\operatorname{Cov}(Y_n, Y_{n+1})$ et $\operatorname{Cov}(Y_n, Y_m)$ pour $m \neq n+1$.

    (c) Pour $n \geq 1$, soit $S_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$. Calculer $E(S_n)$ et $\operatorname{Var}(S_n)$.

    (d) Soit $\varepsilon > 0$. En utilisant l'inégalité de Tchébychev, montrer que si $n \to \infty$,

    $$P(|S_n - p^2| \geq \varepsilon) \to 0.$$

    Que peut-on en conclure quant à S_n ?

17. **(**)** Soit $(X_1, \cdots, X_n)$ un échantillon de loi exponentielle de paramètre 1.

    (a) Montrer que $P(\exists (i,j) \in \{1, \cdots, n\}^2, i \neq j, X_i = X_j) = 0$.

    (b) On pose $Z = \min_{1 \leq i \leq n} X_i$. Déterminer la loi de Z.

    (c) Soit $N = \min\{1 \leq i \leq n, X_i = Z\}$. Montrer que N est une v.a. et établir que $P(N = K, Z > t) = \dfrac{e^{-nt}}{n}$ pour $k = 1, \cdots, n$ et $t > 0$. En déduire que Z et N sont des v. a. indépendantes et préciser la loi de N.

18. **(***)** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. indépendantes identiquement distribuées suivant une loi de Bernoulli de paramètre p. Pour tout $n \geq 1$ on définit par récurrence $T_n = \inf\{k > T_{n-1}, X_k = 1\}$ et $T_0 = 0$.

    (a) Montrer que pour $n \in \mathbb{N}$, T_n est une v.a.

    (b) Montrer que $(T_{i+1} - T_i)_{i\in\mathbb{N}}$ forme une suite de v. a. indépendantes et identiquement distribuées.

    (c) Calculer la loi de T_1 et sa fonction caractéristique.

    (d) En déduire la loi de T_n.

Feuille n^o 2 : **Révisions sur l'espérance conditionnelle et les théorèmes limites**

1. **(**)** Sur $(\Omega, \mathcal{A}, \mu)$ un espace probabilisé, on considère une v.a. réelle positive X de fonction de répartition F_X. Montrer que pour $n \in \mathbb{N}^*$:
$$\mathbb{E}X^n = \int_0^\infty nt^{n-1}(1 - F_X(t))dt = \int_0^\infty nt^{n-1}P(X > t)dt.$$
Montrer que l'hypothèse X positive est nécessaire.

2. **(*)** Soit X et Y des variables aléatoires indépendantes appartenant à $\mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Comparer les lois des couples $(X, X+Y)$ et $(Y, X+Y)$. En déduire que $\mathbb{E}(X \mid X+Y) = \mathbb{E}(Y \mid X+Y) = \frac{1}{2}\cdot(X+Y)$.

3. **(*)** On jette indépendamment deux dés, dont les résultats sont X_1 et X_2. Quelle est la loi de X_1 sachant que $X_1 + X_2$ est paire ?

4. **(*)** Soit X une variable aléatoire réelle et soit $c \in \mathbb{R}$ une constante. Déterminer la loi de X sachant $\min(X, c)$, puis, en supposant que X est $\mathbb{L}^1(\Omega, \mathcal{A}, P)$, déterminer $\mathbb{E}(X \mid \min(X, c))$.

5. **(**)** Soit X une variable aléatoire suivant une loi exponentielle de paramètre $\beta > 0$. Démontrer que
$$P(X > t + s \mid X > t) = P(X > s), \quad \text{pour tout } (s,t) \in \mathbb{R}_+^2.$$
Prouver que pour tout $t \in \mathbb{R}_+$, $\lim_{h\to 0} \frac{1}{h} \cdot P(t < X < t+h \mid X > t) = \beta$.

6. **(**)** Soient X et Y deux variables telles que $(X+Y)$ soit un vecteur gaussien non dégénéré. Déterminer $\mathbb{E}(X \mid Y)$. Généraliser au cas où Y est un vecteur gaussien de $\mathbb{R}^n$ et de matrice de covariance Σ_Y.

7. **(**)** Soit la chaîne de Markov $(X_n)_{n\in\mathbb{N}}$ à valeur dans $\{x_1, x_2\}$ et de matrice de transition $Q = (q_{ij})_{1\le i,j\le 2}$. Rappeler la définition de (X_n). Déterminer la prédiction optimale au sens des moindres carrés de X_{n+1} et X_{n+2} connaissant $(X_0, \cdots, X_n)$, soit $\mathbb{E}(X_{n+1} \mid (X_0, \cdots, X_n))$ et $\mathbb{E}(X_{n+2} \mid (X_0, \cdots, X_n))$.

8. **(**)** Soit $(X_n)_{n\in\mathbb{Z}}$ une suite de variables aléatoires définie par
$$X_n = \varepsilon_n - \theta\varepsilon_{n-1}, \quad \text{pour } n \in \mathbb{Z},$$
avec $|\theta| < 1$ et (ε_n) une suite de v.a.i.i.d. centrées et de variance σ^2. Montrer que le prédicteur linéaire optimal $\widehat{X}_{n+1}$ au sens des moindres carrés de X_{n+1} à partir de $\{X_{n-j}, j \in \mathbb{N}\}$ est :
$$\widehat{X}_{n+1} = \mathbb{E}(X_{n+1} \mid (X_{n-j}, j \in \mathbb{N})) = -\sum_{j=1}^{\infty} \theta^j X_{n+1-j}.$$

9. **(**)** Dans le modèle de l'exercice précédent, on suppose que X_1, X_2, X_4 et X_5 sont connus, mais pas X_3. Déterminer le prédicteur linéaire optimal $\widehat{X}_3$ au sens des moindres carrés de X_3 à partir de (X_1, X_2), puis à partir de (X_4, X_5) et enfin à partir de (X_1, X_2, X_4, X_5).

10. **(***)** Répondre aux mêmes questions dans le cas du modèle
$$X_n = \phi X_{n-1} + \varepsilon_n, \quad \text{pour } n \in \mathbb{Z},$$
où $\phi \in \mathbb{R}$ et (ε_n) une suite de v.a.i.i.d. centrées de variance σ^2.

11. **(**)** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. indépendantes à valeurs dans $\mathbb{N}$. Montrer que :
    (a) $((X_n)$ converge en probabilité vers $0) \iff$ (la suite $(P(X_n > 0))$ tend vers 0).
    (b) $((X_n)$ converge presque-sûrement vers $0) \iff$ (la série $\sum_{n\ge 1}^{\infty} P(X_n > 0) < \infty$).

(c) On suppose que (X_n) suit une loi de Poisson de paramètre α_n. Etudier la convergence de la suite (X_n) dans $\mathbb{L}^1$ puis presque-sûrement dans les cas où $\alpha_n = 1/n$ et $\alpha_n = 1/n^2$.

12. **(**)** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. non corrélées et dans $\mathbb{L}^2$ à valeurs dans $\mathbb{R}$. On suppose qu'il existe des réels m et C tels que pour tout n, $\mathbb{E}(X_n) = m$ et $\mathrm{var} X_n \leq C$. Montrer que la suite des $\bar{X}_n$ converge vers m dans $\mathbb{L}^2$ et en probabilité.

13. **(***)** Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d., dans $\mathbb{L}^2$ et à valeurs dans $\mathbb{R}$. Soit $Y_k = \bar{X}_k$ pour $k \geq 1$. Peut-on obtenir la loi faible des grands nombres pour la suite (Y_k) ? La loi forte des grands nombres? Un théorème de limite centrale ? Traiter ensuite le cas particulier où les X_k suivent une loi normale centrée réduite.

14. **(*)** Trouver la probabilité d'obtenir 200 fois 1 en lançant 1000 fois un dé honnête.

15. **(*)** On sait que 3% des réservations de places d'avions ne sont pas honorées le jour du départ. Pour un avion de 400 places, combien de réservations au maximum peuvent être acceptées par une compagnie aérienne pour que la probabilité que des personnes ayant réservé n'aient pas de place le jour du départ soit inférieure à 5% ?

16. **(*)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de même loi ayant pour densité par rapport à la mesure de Lebesgue :
$$f(x) = \exp(-x + \theta) \cdot \mathbb{I}_{x \geq \theta},$$
où θ est un nombre réel donné. On définit $m_n = \min\{X_1, \cdots, X_n\}$. Déterminer la fonction de répartition, la densité ainsi que l'espérance de m_n. En utilisant l'inégalité de Markov, vérifier que m_n converge en probabilité vers θ.

17. **(*)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de même loi exponentielle de paramètre β. On pose
$$T_n = \frac{n}{\sum_{i=1}^n X_i}.$$
Montrer que $\sqrt{n}(T_n - \beta)$ converge en loi vers une loi normale dont précisera l'espérance et la variance.

18. **(**)** Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes sur $(\Omega, \mathcal{A}, \mathbb{P})$ et de même loi $\mathbb{P}_X$ (à laquelle est associée la fonction de répartition F_X). Pour tout $n \in \mathbb{N}^*$, on définit la fonction de répartition empirique de $(X_1, \cdots, X_n)$ par :
$$F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{X_i \leq x} \quad \text{pour} \quad x \in \mathbb{R}.$$
(a) Rappeler l'expression de F_X lorsque $\mathbb{P}_X$ est une loi binomiale de paramètre $(3, 1/3)$ et lorsque $\mathbb{P}_X$ est la loi gaussienne centrée réduite.

(b) Montrer que pour tout $x \in \mathbb{R}$, $F_n(x)$ est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $[0, 1]$. Pour tout $n \in \mathbb{N}^*$ et tout $x \in \mathbb{R}$, calculer $\mathbb{E}F_n(x)$ en fonction de F_X.

(c) Démontrer que pour tout $x \in \mathbb{R}$, $F_n(x) \xrightarrow[n\to+\infty]{\mathcal{P}} F(x)$.

(d) Montrer que pour tout $x \in \mathbb{R}$, $F_n(x)$ vérifie un théorème de limite centrale que l'on établira.

(e) On suppose que les $(X_n)_{n\in\mathbb{N}^*}$ ne sont plus indépendants. A partir d'un contre-exemple, montrer que les convergences précédentes n'ont plus lieu en général.

Feuille n^o 3 : **Statistiques exhaustives**

1. **(*)** Soit $((\Omega')^n, \mathcal{A}'_n, (\mathbb{P}_\theta)^{\otimes n}, \theta \in \Theta)$ un modèle paramétrique. Dans tous les cas suivants, préciser Ω', $\mathcal{A}'_n$ et la mesure dominante, et vérifier (sans utiliser les résultats sur les modèles exponentiels) si la statistique $\widehat{S}_n = X_1 + \ldots + X_n$ est une statistique exhaustive pour ce modèle :
   (a) $\mathbb{P}_\theta$ est la loi de Bernoulli de paramètre $\theta = p \in]0,1[$;
   (b) $\mathbb{P}_\theta$ est la loi exponentielle de paramètre $\theta = \lambda \in]0,+\infty[$;
   (c) $\mathbb{P}_\theta$ est la loi normale de moyenne $\theta \in \mathbb{R}$ et de variance 1;
   (d) $\mathbb{P}_\theta$ est la loi normale de paramètre $\theta = (m, \sigma^2) \in \mathbb{R} \times]0,+\infty[$;
   (e) $\mathbb{P}_\theta$ est la loi de Poisson de paramètre $\theta = (\lambda) \in \mathbb{R} \times]0,+\infty[$;
   (f) $\mathbb{P}_\theta$ est la loi uniforme sur $[0,\theta]$ où $\theta \in]0,+\infty[$;

2. **(*)** On considère le modèle paramétrique gaussien $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{N}(0,\sigma^2)^{\otimes n}, \sigma^2 \in]0,+\infty[)$. Montrer (sans utiliser les résultats sur les modèles exponentiels) que la statistique $\widehat{T}_n = X_1^2 + \ldots + X_n^2$ est exhaustive pour ce modèle.

3. **(**)** Soit $((\Omega')^n, \mathcal{A}'_n, (\mathbb{P}_\theta)^{\otimes n}, \theta \in]0,\infty[)$ un modèle paramétrique tel que $\mathbb{P}_\theta$ admette une densité par rapport à la mesure de Lebesgue de densité :
$$f(x) = \frac{1}{\log 2} \cdot \frac{1}{x} \cdot \mathbb{I}_{]\theta, 2\theta[}(x).$$
   Préciser Ω' et $\mathcal{A}'_n$. Montrer que la statistique $\widehat{T} = \big(\min(X_1,\ldots,X_n), \max(X_1,\ldots,X_n)\big)$ est une statistique exhaustive minimale pour ce modèle.

4. **(**)** On considère un n-échantillon de v.a.i.i.d. de loi gamma de paramètres $(a,b) \in]0,\infty[^2$. Préciser le modèle statistique paramétrique induit. Déterminer une statistique exhaustive pour ce modèle. Est-elle minimale ? complète ?

5. **(***)** Soit un n-échantillon $(X_1,\ldots,X_n)$ de v.a. telles que X_1 suit une loi de Bernoulli de paramètre $1/2$, et $\mathbb{P}(X_{n+1} = 1 \mid X_n = 1) = \mathbb{P}(X_{n+1} = 0 \mid X_n = 0) = p$ et $\mathbb{P}(X_{n+1} = 1 \mid X_n = 0) = \mathbb{P}(X_{n+1} = 0 \mid X_n = 1) = 1 - p$ pour $n \in \mathbb{N}^*$, avec $p \in]0,1[$ un paramètre inconnu. Montrer que la statistique d'ordre $\widehat{T}_n = (X_{(1)}, \cdots, X_{(n)})$ n'est pas une famille exhaustive pour ce modèle paramétrique.

6. **(**)** On considère un n-échantillon de v.a.i.i.d. de loi uniforme sur $]\theta, \theta+1[$, où $\theta \in \mathbb{R}$.
   (a) Préciser le modèle statistique induit.
   (b) Montrer que la statistique $\widehat{T} = \big(\min(X_1,\ldots,X_n), \max(X_1,\ldots,X_n)\big)$ est une statistique exhaustive minimale pour ce modèle, mais n'est pas complète.
   (c) Soit $X_{(1)} \leq \cdots \leq X_{(n)}$ l'échantillon ordonné. Soit $\widehat{R}_n = X_{(n)} - X_{(1)}$. Montrer que $\widehat{R}_n$ est une statistique libre pour le modèle.

7. **(*)** En reprenant les différents modèles de l'exercice 1., déterminer l'information de Fisher du modèle et comparer la à celle induite par la statistique $\widehat{S}_n = X_1 + \ldots + X_n$.

8. **(**)** Soit un n-échantillon de v.a.i.i.d. de loi admettant la densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}$ telle que :
$$f(x) = \frac{\exp(x-\theta)}{(1+\exp(x-\theta))^2},$$
où $\theta \in \mathbb{R}$ est un paramètre inconnu. Déterminer une statistique exhaustive minimale et complète pour ce modèle et déterminer son information de Fisher.

9. **(***)** Soit un n-échantillon de v.a.i.i.d. de loi admettant la densité f par rapport à la mesure de Lebesgue telle que :
$$f(x) = \frac{1}{\theta} \cdot x^{-1+1/\theta} \cdot \exp(-x^{1/\theta}) \cdot \mathbb{I}_{]0,\infty[}(x),$$
où $\theta \in]0,\infty[$ est un paramètre inconnu. Déterminer une statistique exhaustive minimale et complète pour ce modèle et déterminer son information de Fisher. Montrer que la statistique $\widehat{T}_2 = \dfrac{\log(X_2)}{\log(X_1)}$ est libre. En déduire que $\widehat{T}_n = \dfrac{\prod_{i=2}^n \log(X_i)}{\log(X_1)}$ l'est également.

10. **(***)** Soit $(\varepsilon_i)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. de loi gaussienne $\mathcal{N}(0,\sigma^2)$, avec $\sigma^2 \in]0,\infty[$ inconnue. On suppose que $(X_1,\cdots,X_n)$ est un n-échantillon tel que :
$$\forall k \in \{1,\cdots,n\}, \quad X_k = \varepsilon_{k+1} - a \cdot \varepsilon_k,$$
où $a \in \mathbb{R}$ est inconnu.
    (a) Montrer que le modèle paramétrique lié à l'échantillon $(X_1,\cdots,X_n)$ (on notera $\theta = (\sigma^2, a)$) appartient à la famille exponentielle.
    (b) Déterminer une statistique exhaustive minimale et complète pour ce modèle.
    (c) Déterminer l'expression de l'information de Fisher pour ce modèle (ne pas faire les calculs).

11. **(***)** Soit $(\varepsilon_i)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. de loi de Bernoulli de paramètre $p \in]0,1[$ inconnu. Soit $(X_1,\cdots,X_n)$ un n-échantillon tel que :
$$\forall k \in \{1,\cdots,n\}, \quad X_k = \varepsilon_{k+1} \cdot \varepsilon_k.$$
    (a) Montrer que le modèle paramétrique lié à l'échantillon $(X_1,\cdots,X_n)$ appartient à la famille exponentielle.
    (b) Déterminer une statistique exhaustive minimale et complète pour ce modèle.
    (c) Soit la statistique d'ordre $\widehat{R}_n = (X_{(1)},\cdots,X_{(n)})$. Est-ce une statistique exhaustive ?

Feuille n^o 4 : **Estimation paramétrique**

1. **(*)** Soit $((\Omega')^n, \mathcal{A}'_n, (\mathbb{P}_\theta)^{\otimes n}, \theta \in \Theta)$ un modèle paramétrique. Dans tous les cas suivants, préciser un estimateur exhaustif complet et sans biais du paramètre θ, déterminer s'il est efficace, et s'il ne l'est pas, donner la fonction du paramètre pouvant être estimé de façon efficace et sans biais :

   (a) $\mathbb{P}_\theta$ est la loi de Bernoulli de paramètre $\theta = p \in]0,1[$;

   (b) $\mathbb{P}_\theta$ est la loi exponentielle de moyenne $\theta = \lambda \in]0,+\infty[$;

   (c) $\mathbb{P}_\theta$ est la loi normale de moyenne $\theta \in \mathbb{R}$ et de variance 1;

   (d) $\mathbb{P}_\theta$ est la loi normale de moyenne 0 et de variance $\theta = \sigma^2 \in]0,+\infty[$;

   (e) $\mathbb{P}_\theta$ est la loi normale de moyenne 0 et d'écart-type $\theta = \sigma \in]0,+\infty[$;

   (f) $\mathbb{P}_\theta$ est la loi de Poisson de paramètre $\theta = \lambda \in]0,+\infty[$;

   (g) $\mathbb{P}_\theta$ est la loi dont la densité par rapport à la mesure de Lebesgue est $f_\theta(x) = \dfrac{\log\theta}{\theta-1} \cdot \theta^x \cdot \mathbb{I}_{]0,1[}(x)$.

2. **(**)** On considère le modèle paramétrique gaussien $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{N}(m,\sigma^2)^{\otimes n}, (m,\sigma^2) \in \mathbb{R}\times]0,+\infty[)$.

   (a) En utilisant les résultats sur les modèles exponentiels, montrer que l'estimateur $\widehat{T}_n = (\overline{X}_n, \overline{\sigma}^2_n)$, où $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i$ et $\overline{\sigma}^2_n = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (X_i - \overline{X}_n)^2$ est exhaustif et complet pour ce modèle.

   (b) En utilisant le théorème de Cochran, montrer que cet estimateur est sans biais. Est-il efficace ? Déterminer une région de confiance à 95% pour (m,σ^2).

   (c) Déterminer l'estimateur du maximum de vraisemblance de (m,σ^2). Est-il biaisé ? Efficace ? Déterminer une région de confiance à 95% pour (m,σ^2).

3. **(**)** Soit $((\Omega')^n, \mathcal{A}'_n, (\mathbb{P}_\theta)^{\otimes n}, \theta \in]0,\infty[)$ un modèle paramétrique tel que $\mathbb{P}_\theta$ admette une densité par rapport à la mesure de Lebesgue de densité :

$$f(x) = \frac{1}{\log 2} \cdot \frac{1}{x} \cdot \mathbb{I}_{]\theta,2\theta[}(x).$$

   (a) Montrer que l'estimateur du maximum de vraisemblance pour ce modèle n'est pas unique.

   (b) On pose $\widehat{\theta}_n^{(1)} = \dfrac{1}{2}\max(X_1,\dots,X_n)$ et $\widehat{\theta}_n^{(2)} = \min(X_1,\dots,X_n)$. Pour chacun de ces estimateurs, déterminer leur loi et montrer qu'ils convergent. Sans rentrer dans les calculs, donner un court raisonnement montrant qu'ils sont biaisés. Déterminer une région de confiance à 95% pour θ.

4. **(***)** Soit $((\Omega')^n, \mathcal{A}'_n, (\mathbb{P}_\theta)^{\otimes n}, \theta \in]0,\infty[)$ un modèle paramétrique tel que $\mathbb{P}_\theta$ admette une densité par rapport à la mesure de Lebesgue de densité :

$$f(x) = \mathbb{I}_{]\theta,\theta+1[}(x).$$

   (a) Montrer que l'estimateur du maximum de vraisemblance pour ce modèle n'est pas unique.

   (b) On pose $\widehat{\theta}_n^{(1)} = \max(X_1,\dots,X_n) - 1$ et $\widehat{\theta}_n^{(2)} = \min(X_1,\dots,X_n)$. Pour chacun de ces estimateurs, déterminer la loi, l'espérance et la variance. Sont-ils indépendants ? Convergents ? Efficaces ?

   (c) Déterminer un estimateur de la forme $\widehat{\theta}_n = \alpha \cdot \widehat{\theta}_n^{(1)} + (1-\alpha) \cdot \widehat{\theta}_n^{(2)}$, avec $\alpha \in]0,1[$, qui soit sans biais. Cet estimateur a-t-il une variance inférieure à celle des précédents ?

5. **(**)** Soit un n-échantillon $(X_1, \ldots, X_n)$ de v.a. telles que X_1 suit une loi de Bernoulli de paramètre $1/2$, et $P(X_{n+1} = 1 \mid X_n = 1) = P(X_{n+1} = 0 \mid X_n = 0) = p$ et $P(X_{n+1} = 1 \mid X_n = 0) = P(X_{n+1} = 0 \mid X_n = 1) = 1 - p$ pour $n \in \mathbb{N}^*$, avec $p \in]0, 1[$ un paramètre inconnu. Déterminer l'estimateur de maximum de vraisemblance de p. Est-il biaisé ? Convergent ?

6. **(*)** On considère un n-échantillon $(X_1, \ldots, X_n)$ de v.a.i.i.d. suivant la loi binomiale $\mathcal{B}(p, k)$, où $p \in]0, 1[$ est inconnu alors que $k \in \mathbb{N}^*$ est connu. On voudrait estimer la probabilité $P(X_1 = 1)$. Montrer que c'est une fonction de p, soit $g(p)$. Déterminer un estimateur sans biais de variance uniformément minimum pour $g(\theta)$. Est-il efficace ?

7. **(**)** Soit un n-échantillon de v.a.i.i.d. de loi admettant la densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}$ telle que :

$$f(x) = (1 - \theta) \cdot \mathbb{I}_{]-1/2,0[}(x) + (1 + \theta) \cdot \mathbb{I}_{]0,1/2[}(x),$$

où $\theta \in]-1, 1[$ est un paramètre inconnu. Est-ce un modèle exponentiel ? Est-il régulier ? Déterminer l'estimateur $\widehat{\theta}_n$ du maximum de vraisemblance. Est-il sans biais ? Converge-t-il ? Quelle est la loi limite de $\sqrt{n}(\widehat{\theta}_n - \theta_0)$, où θ_0 désigne le "vrai" paramètre ? En déduire un intervalle de confiance à 95%.

8. **(***)** Soit $(\varepsilon_i)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. de loi de Bernoulli de paramètre $p \in]0, 1[$ inconnu. Soit $(X_1, \cdots, X_n)$ un n-échantillon tel que :

$$\forall k \in \{1, \cdots, n\}, \quad X_k = \prod_{i=1}^{k} \varepsilon_i$$

Déterminer l'estimateur du maximum de vraisemblance de p. Est-il convergent ? Sans biais ? Efficace ? Comparer ces résultats avec ceux obtenus par l'appartenance du modèle à la famille exponentielle.

9. **(**)** Soit (X_i) est une suite de v.a.i.i.d. de même loi que X, où X admet pour densité de probabilité $f(x, \theta)$ par rapport à la mesure Lebesgue, définie par :

$$f(x, \theta) = k \cdot x^{-(p+1)} \cdot \exp(-\frac{\theta}{x}) \cdot \mathbb{I}_{x>0},$$

avec $\theta \in]0, +\infty[$ un paramètre réel inconnu et $p > 0$ un nombre connu.

   (a) Montrer que $k = \dfrac{\theta^p}{\Gamma(p)}$. Déterminer une statistique exhaustive complète minimale. Déterminer le paramètre pouvant être estimer efficacement à partir de cette statistique.

   (b) On pose $U = \dfrac{1}{X}$ et $Z = 2 \cdot \theta \cdot U$. Déterminer $\mathbb{E}(U)$ et $\text{var}(U)$ ainsi que $\mathbb{E}(Z)$ et $\text{var}(Z)$. Déterminer $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ. Montrer que :

$$\mathbb{E}(\widehat{\theta}_n) = \frac{n \cdot p \cdot \theta}{n \cdot p - 1}, \quad \text{var}(\widehat{\theta}_n) = \frac{(n \cdot p \cdot \theta)^2}{(n \cdot p - 1)^2 (n \cdot p - 2)}.$$

   L'estimateur $\widehat{\theta}_n$ est-il sans biais? convergent? efficace?

10. **(**)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de loi de Pareto de densité :

$$f(x; \theta) = \frac{3\theta^3}{x^4} \cdot \mathbb{I}_{x \geq \theta}, \quad \text{où } \theta \text{ est un paramètre positif inconnu.}$$

   (a) Montrer que cet échantillon n'appartient pas à la famille exponentielle.

   (b) Calculer $\widehat{\theta}_n$ l'estimateur de maximum de vraisemblance de θ.

(c) Déterminer la fonction de répartition puis la densité de l'estimateur $\widehat{\theta}_n$. En déduire que $\widehat{\theta}_n$ est un estimateur convergent de θ. Est-il asymptotiquement efficace ?

(d) Calculer $\mathbb{E}(\widehat{\theta}_n)$ et en déduire un estimateur sans biais de θ, puis un intervalle de confiance à 95% de θ.

Feuille n^o 5 : **Tests paramétriques**

1. **(*)** On a lancé 100 fois une pièce de monnaie et obtenu 58 fois pile. Avec un niveau $\alpha = 0.05$, tester si la pièce est équilibrée ? Déterminer la fonction puissance du test. Tester si la pièce est truquée. Aboutit-on au même résultat ?

2. **(*)** Une chaîne des magasins décide d'adopter une nouvelle politique de gestion des stocks pour l'ensemble de ses succursales. Auparavant, le bénéfice mensuel d'une succursale était en moyenne égal à 300000 euros. Après la mise en place de la nouvelle politique pour 100 succursales "pilotes", on observe un bénéfice moyen de 320000 euros avec un écart-type de 20000 euros. Peut-on conclure à l'efficacité de cette politique au niveau 5% ? 1% ?

3. **(*)** On suppose que le nombre de clients N attendant à la caisse d'un grand magasin à 11h00 du matin peut être modélisé par une loi de Poisson de paramètre $\theta > 0$. On sait que pour les clients la limite du supportable en attente est de voir au maximum 5 clients devant une caisse. La question se pose de savoir si l'on doit augmenter ou nom le nombre de personnes travaillant aux caisses : ceci conduit à un problème de test sur l'opportunité d'embaucher.

   (a) Pour la direction, peu désireuse d'embaucher, le problème se pose de la manière suivante : H_0 : $\theta \leq 5$ contre H_1 : $\theta > 5$ à tester avec le niveau 5%; ainsi, la direction veut que la probabilité d'embaucher alors qu'il n'y en a pas besoin est contrôlée (moins de 5%) sans s'intéresser à l'autre erreur possible (qui est de ne pas embaucher alors qu'il le fallait).

   (b) Pour les syndicats, la position est inverse : on veut surtout éviter de ne pas embaucher alors qu'il le fallait. Ils vont tabler sur le fait l'erreur de seconde espèce du test soit de 5%.

   Dans les deux cas, sachant que l'on dispose 1/ d'une seule observation de valeur 6; 2/ de 100 observations de moyenne 5.4, déterminer les résultats des tests du rapport de vraisemblance et de Wald.

4. **(*)** Dans le cas d'un échantillon $(X_1, \cdots, X_n)$ de v.a.i.i.d. de loi $\mathcal{N}(m, 1)$, où m est inconnu, déterminer le test de Wald de niveau 5% (statistique de test, région critique) pour les hypothèses :

   (a) H_0 : $m = 0$ contre H_1 : $m = 2$;

   (b) H_0 : $m = 0$ contre H_1 : $m > 0$;

   (c) H_0 : $m = 0$ contre H_1 : $m \neq 0$.

   Faire ensuite la même chose avec le test du rapport de vraisemblance.

5. **(**)** Dans le cas d'un échantillon $(X_1, \cdots, X_n)$ de v.a.i.i.d. déterminer le test du rapport de vraisemblance de niveau 5% (statistique de test, région critique), pour les hypothèses :

   (a) H_0 : $\theta = \theta_0$ contre H_1 : $\theta = \theta_1 > \theta_0$;

   (b) H_0 : $\theta = \theta_0$ contre H_1 : $\theta > \theta_0$;

   (c) H_0 : $\theta = \theta_0$ contre H_1 : $\theta \neq \theta_0$;

   (d) H_0 : $\theta < \theta_0$ contre H_1 : $\theta > \theta_1 > \theta_0$;

   (e) H_0 : $\theta \neq \theta_0$ contre H_1 : $\theta = \theta_0$,

   où $(\theta_0, \theta_1) \in \Theta^2$ sachant que la loi de X_1 est 1/ une loi de Bernoulli de paramètre $\theta \in]0, 1[$, 2/ une loi uniforme sur $[0, \theta]$ avec $\theta > 0$, 3/ une loi exponentielle de paramètre $\theta > 0$.

6. **(**)** On considère une seule variable aléatoire X observée, de loi binomiale $\mathcal{B}(k, p)$ où $k \in \mathbb{N}^*$ est connu, mais $p \in]0, 1[$ est inconnu.

(a) Montrer que $p \in]0,1[\mapsto L_p(x)$, où $x \in \{0,1,\ldots,k\}$ est une fonction croissante sur $]0,x/k]$, puis décroissante sur $[x/k,1[$.

(b) On veut tester $H_0 \;: p > p_0$ contre $H_1 \;: p \leq p_0$ au niveau α. Vérifier que le rapport de vraisemblance de ce problème est une fonction décroissante de l'observation. En déduire la forme de la région de rejet du test.

7. **(**)** Soit $X_1, \cdots, X_n$ des v.a.i.i.d. de loi exponentielle de paramètre $1/\theta$ avec $\theta > 0$, inconnu. On souhaite faire un choix entre les deux hypothèses suivantes :

$$H_0 :\; \theta = 1, \qquad H_1 : \theta = \theta_1 > 1$$

(a) Appliquer le test du rapport de vraisemblance (ici Neyman-Pearson) de niveau α. Vérifier que la région de rejet de H_0 est de la forme

$$\{\widehat{X}_n > \mu_\alpha,\}$$

où $\mu_\alpha > 0$. Calculer l'erreur de seconde espèce β en fonction de μ_α, θ_1 et de la fonction de répartition G_{2n} de la loi du $\mathcal{X}^2_{2n}$.

(b) Soit N_n le nombre de X_j supérieurs à $\dfrac{1}{2}(1+\theta_1)$. On accepte l'hypothèse H_1 lorsque $N_n \geq \mu'_\alpha$. Pour tout $\alpha \in]0,1[$, ce test est-il moins puissant que celui de Neyman-Pearson ?

8. **(**)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$:

$$f(x) = e^{-(x-\theta)}\, \mathbb{I}_{x \geq \theta},$$

où $\theta \in \mathbb{R}$ est inconnu.

(a) Montrer que la statistique $\widehat{\theta}_n = \min(X_1, \cdots, X_n)$ est exhaustive pour θ. Quelle est le biais et la variance de cet estimateur ? En déduire un test des hypothèses $H_0 \;: \theta \leq \theta_0$ contre $H_1 \;: \theta > \theta_0$ de niveau α utilisant comme statistique de test $\widehat{\theta}_n$.

(b) Déterminer le test $\widehat{T}$ du rapport de vraisemblance de niveau α pour le même problème de test. Conclusion ? Pour α fixé, quelle est explicitement la région critique ? Déterminer la fonction puissance de $\widehat{T}$.

9. **(**)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}_+$:

$$f(x) = \frac{\alpha \cdot r^\alpha}{x^{\alpha+1}}\, \mathbb{I}_{x \geq r},$$

où $\alpha \in \mathbb{R}$ et $r > 0$ sont inconnus. Un tel modèle s'appelle le modèle de Pareto, et pourrait être appliqué en économie pour modéliser des revenus, sachant que r serait une sorte de revenu minimal.

(a) A quelle condition une variable de Pareto admet une espérance ? une variance ?

(b) Donner l'estimateur du maximum de vraisemblance $(\widehat{a}_n, \widehat{r}_n)$ de ces deux paramètres.

(c) On désirer tester $H_0 \;: a = 1$ contre $H_1 \;: a \neq 1$. Montrer que le test $\widehat{T}$ du rapport de vraisemblance de niveau α dans ce cas admet une région critique de la forme $[\widehat{a}_n > s_1] \cup [\widehat{a}_n < s_2]$.

10. **(***)** Soit $(X_1, \cdots, X_n)$ et $(X'_1, \cdots, X'_n)$ deux échantillons de v.a.i.i.d. de loi exponentielle de moyennes respectives θ et θ' où $\theta > 0$ et $\theta' > 0$ sont inconnus. On désirer tester $H_0 \;: \theta = \theta'$ contre $H_1 \;: \theta \neq \theta'$. Déterminer le test du rapport de vraisemblance de niveau α. Montrer que la région critique de ce test peut s'exprimer en fonction de la statistique $\widehat{T} = \dfrac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i + \sum_{i=1}^n Y_i}$. Déterminer la loi de $\widehat{T}$ sous l'hypothèse H_0.

11. **(***)** Soit $(X_1, \cdots, X_n)$ un échantillon de v.a.i.i.d. de loi gaussiennes $\mathcal{N}(m, \sigma^2)$ où $m \in \mathbb{R}$ et $\sigma^2 > 0$ sont deux paramètres inconnus. Montrer que pour le problème de test $H_0 \ : m = m_0$ contre $H_1 \ : m \neq m_0$ de niveau α, le test du rapport de vraisemblance se ramène à utiliser une statistique de Student. Quel est également le test de Wald pour ce problème ?

Université Paris I, Panthéon - Sorbonne

Master M.A.E.F. Première Année 2007 – 2008

Statistique

Plan du cours

1. Quelques rappels la théorie de la mesure.
2. Quelques rappels sur les applications de la théorie de la mesure aux probabilités.
3. Estimation paramétrique.
4. Tests paramétriques.
5. Introduction à la statistique non-paramétrique.

Bibliographie

• **Livres pour revoir les bases....**

1. Baillargeon, B. *Probabilités, statistiques et techniques de régression.* SMG.
2. Bercu, B., Pamphile, P. et Azoulay, E. *Probabilités et Applications - Cours Exercices.* Edisciences.
3. Dress, F. *Probabilités et Statistique.* Dunod.
4. Lecoutre, J.-P. *Statistiques et Probabilités.* Dunod.

Théorie de la mesure et applications aux probabilités

- Ansel et Ducel, *Exercices corrigés en théorie de la mesure et de l'intégration*, Ellipses.
- Barbe, P. et Ledoux, M., *Probabilités*, Belin.
- Dacunha-Castelle, D. et Duflo, M., *Probabilités et Statistiques (I)*, Masson
- Jacod, J., *Cours d'intégration*, http://www.proba.jussieu.fr/pageperso/jacod.html.
- Jacod, J., *Cours de Probabilités*, http://www.proba.jussieu.fr/pageperso/jacod.html.
- Toulouse, P. *Thèmes de probabilités et statistiques*, Masson.

Statistiques inférentielles

- Dacunha-Castelle, D. et Duflo, M., *Probabilités et Statistiques (I)*, Masson.
- Fourdrinier, D., *Statistique inférentielle*, Dunod.
- Lecoutre, J.-M. et Tassi, P., *Statistique non paramétrique et robustesse*, Economica.
- Milhaud, X., *Statistique*, Belin.
- Monfort, A., *Cours de statistique mathématique*, Economica.
- Saporta, G., *Probabilités, analyse des données et statistiques.* Technip.
- Tsybakov, A. *Introduction à la statistique non-paramétrique.* Collection : Mathématiques et Applications, Springer.

Plan détaillé du cours de STATISTIQUES 1

1 Rappels sur la théorie de la mesure

Introduction

Il demeure des choses inconnues à partir des connaissances antérieures en probabilités :

- Qu'est-ce qu'un événement et l'ensemble de tous les événements ?
- Que se passe-t-il pour des probabilités d'événements moins classiques (par exemple l'ensemble des décimaux) ?
- Comment traiter une variable aléatoire qui est continue et discrète à la fois (par exemple le nombre de minutes passées devant la TV) ?

1.1 Mesures

1.1.1 Tribus

Notation. • *Ω est un ensemble (fini ou infini).*

- *$\mathcal{P}(\Omega)$ est l'ensemble de tous les sous-ensembles (parties) de Ω.*

Rappel. *Soit E un ensemble. E est dit dénombrable s'il existe une bijection entre E et $\mathbb{N}$ ou un sous-ensemble de $\mathbb{N}$. Par exemple, un ensemble fini, $\mathbb{Z}$, $\mathbb{D}$, $\mathbb{Z} \times \mathbb{Z}$, $\mathbb{Q}$ sont dénombrables. En revanche, $\mathbb{R}$ n'est pas dénombrable.*

Définition. *Soit une famille $\mathcal{F}$ de parties de Ω (donc $\mathcal{F} \subset \mathcal{P}(\Omega)$). On dit que $\mathcal{F}$ est une algèbre si :*

- *$\Omega \subset \mathcal{F}$;*
- *lorsque $A \in \mathcal{F}$ alors $(\Omega \setminus A) \in \mathcal{F}$;*
- *pour tout $n \in \mathbb{N}^*$, lorsque $(A_1, \cdots, A_n) \in \mathcal{F}^n$ alors $A_1 \cup \cdots \cup A_n \in \mathcal{F}$.*

Définition. *Soit une famille $\mathcal{A}$ de parties de Ω (donc $\mathcal{A} \subset \mathcal{P}(\Omega)$). On dit que $\mathcal{A}$ est une tribu (ou σ-algèbre) sur Ω si :*

- *$\Omega \subset \mathcal{F}$;*
- *lorsque $A \in \mathcal{F}$ alors $(\Omega \setminus A) \in \mathcal{F}$;*
- *pour $I \subset \mathbb{N}$, lorsque $(A_i)_{i \in I} \in \mathcal{F}^I$ alors $\bigcup_{i \in I} A_i \in \mathcal{A}$.*

Exemple. • *Cas du Pile ou Face.*

- *Cas où Ω est infini : $\Omega = \mathbb{N}$ par exemple.*

Propriété. *Avec les notations précédentes :*

1. *$\emptyset \in \mathcal{A}$;*
2. *si A et B sont dans la tribu $\mathcal{A}$, alors $A \cap B$ est dans $\mathcal{A}$;*
3. *si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont deux tribus sur Ω, alors $\mathcal{A}_1 \cap \mathcal{A}_2$ est une tribu sur Ω . Plus généralement, pour $I \subset \mathbb{N}$, si $(\mathcal{A}_i)_{i \in I}$ ensemble de tribus sur Ω, alors $\bigcap_{i \in I} \mathcal{A}_i$ est une tribu sur Ω;*
4. *si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont deux tribus sur Ω, alors $\mathcal{A}_1 \cup \mathcal{A}_2$ n'est pas forcément une tribu sur Ω.*

Définition. *Si $\mathcal{E}$ est une famille de parties de Ω (donc $\mathcal{E} \subset \mathcal{P}(\Omega)$), alors on appelle tribu engendrée par $\mathcal{E}$, notée $\sigma(\mathcal{E})$, la tribu engendrée par l'intersection de toutes les tribus contenant $\mathcal{E}$ (on peut faire la même chose avec des algèbres).*

Remarque. *La tribu engendrée est la "plus petite" tribu (au sens de l'inclusion) contenant la famille $\mathcal{E}$.*

Rappel.
- *Un ensemble ouvert U dans un espace métrique X est telle que pour tout $x \in U$, il existe $r > 0$ tel que $B(x, r) \subset U$.*
- *On dit qu'un ensemble dans un espace métrique X est fermé si son complémentaire dans X est ouvert.*

Définition. *Soit Ω un espace métrique. On appelle tribu borélienne sur Ω, notée, $\mathcal{B}(\Omega)$, la tribu engendrée par les ouverts de Ω. Un ensemble de $\mathcal{B}(\Omega)$ est appelé borélien.*

Exemple.
- *Boréliens sur $\mathbb{R}$, sur $]0, 1[$.*
- *Boréliens sur $\mathbb{R}^2$.*

1.1.2 Espace mesurable

Définition. *Soit Ω un ensemble et soit $\mathcal{A}$ une tribu sur Ω. On dit que $(\Omega, \mathcal{A})$ est un espace mesurable.*

Corollaire. *Quand on s'intéressera aux probabilités, on dira que $(\Omega, \mathcal{A})$ est un espace probabilisable.*

Propriété. *Si $(\Omega_i, \mathcal{A}_i)_i$ sont n espaces mesurables, alors un ensemble élémentaire de $\Omega = \Omega_1 \times \cdots \times \Omega_n$ est une réunion finie d'ensembles $A_1 \times \cdots \times A_n$ où chaque $A_i \in \mathcal{A}_i$. L'ensemble des ensembles élémentaires est une algèbre et on note $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_n$ (on dit $\mathcal{A}_1$ tensoriel $\mathcal{A}_2$ ... tensoriel $\mathcal{A}_n$) la tribu sur Ω engendrée par ces ensembles élémentaires.*

Exemple. *Pavés de $\mathbb{R}^d$.*

Définition. *On appelle espace mesurable produit des $(\Omega_i, \mathcal{A}_i)_i$ l'espace mesurable $\left(\prod_{i=1}^{n} \Omega_i, \bigotimes_{i=1}^{n} \mathcal{A}_i\right)$.*

Exemple. *Pile / Face 2 fois.*

1.1.3 Définitions et Propriétés d'une mesure

Définition. *Soit $(\Omega, \mathcal{A})$ un espace mesurable. L'application $\mu \, : \mathcal{A} \to [0, +\infty]$ est une mesure si :*

- *$\mu(\emptyset) = 0$.*
- *Pour tout $I \subset \mathbb{N}$ et pour $(A_i)_{i \in I}$ famille disjointe de $\mathcal{A}$ (telle que $A_i \cup A_j = \emptyset$ pour $i \neq j$), alors $\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$ (propriété dite de σ-additivité).*

Définition. *Avec les notations précédentes :*

- *Si $\mu(\Omega) < +\infty$, on dit que μ est finie.*
- *Si $\mu(\Omega) < M$ avec $M < +\infty$, on dit que μ est bornée.*
- *Si $\mu(\Omega) = 1$, on dit que μ est une mesure de probabilité.*

Exemple. *Cas de $\Omega = \mathbb{R}$, de $\mathbb{N}$, ou $\mathbb{R}^2$.*

Définition. *Si $(\Omega, \mathcal{A})$ est un espace mesurable (resp. probabilisable) alors $(\Omega, \mathcal{A}, \mu)$ est un espace mesuré (resp. probabilisé quand μ est une probabilité).*

Remarque. *Sur $(\Omega, \mathcal{A})$, on peut définir une infinité de mesures.*

Propriété. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(A_i)_{i \in \mathbb{N}}$, une famille de $\mathcal{A}$.*

1. *Si $A_1 \subset A_2$, alors $\mu(A_1) \leq \mu(A_2)$.*

2. *Si* $\mu(A_1) < +\infty$ *et* $\mu(A_2) < +\infty$*, alors* $\mu(A_1 \cup A_2) + \mu(A_1 \cap A_2) = \mu(A_1) + \mu(A_2)$.

3. *Pour tout* $I \subset \mathbb{N}$*, on a* $\mu\left(\bigcup_{i\in I} A_i\right) \leq \sum_{i\in I} \mu(A_i)$.

4. *Si* $A_i \subset A_{i+1}$ *pour tout* $i \in \mathbb{N}$ *(suite croissante en sens de l'inclusion), alors* $(\mu(A_n))_{n\in\mathbb{N}}$ *est une suite croissante convergente telle que*
$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \lim_{i\to+\infty} \mu(A_i)$ *(même si cette limite est* $+\infty$*).*

5. *Si* $A_{i+1} \subset A_i$ *pour tout* $i \in \mathbb{N}$ *(suite décroissante en sens de l'inclusion) et* $\mu(A_0) < +\infty$*, alors* $(\mu(A_n))_{n\in\mathbb{N}}$ *est une suite décroissante convergente telle que* $\mu\left(\bigcap_{i\in\mathbb{N}} A_i\right) = \lim_{i\to+\infty} \mu(A_i)$.

Exemple. 1. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré. On définit* $\nu(A) = \mu(A \cap B)$ *où* $B \in \mathcal{A}$*.* ν *mesure ?*

2. *Si* μ_1 *et* μ_2 *mesures sur* $(\Omega, \mathcal{A})$*,* $\mu_1 + \mu_2$ *et* $\alpha\mu$ *sont-elles des mesures ?*

Définition. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré et* $(A_i)_{i\in\mathbb{N}}$ *une famille de* $\mathcal{A}$*.*

1. *On définit* $\limsup(A_n)_n = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m$ *(intuitivement,* $\limsup(A_n)_n$ *est l'ensemble des* $\omega \in \Omega$ *tels que* ω *appartienne à une infinité de* A_n*).*

2. *On définit* $\liminf(A_n)_n = \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} A_m$ *(intuitivement,* $\liminf(A_n)_n$ *est l'ensemble des* $\omega \in \Omega$ *tels que* ω *appartienne à tous les* A_n *sauf à un nombre fini d'entre eux).*

Exemple. *Cas des suites croissantes et décroissantes d'ensembles.*

Théorème (Théorème d'extension de Hahn - Caratheodory)**.** *Si* Ω *est un ensemble,* $\mathcal{F}$ *une algèbre sur* Ω*, et* ν *une application de* $\mathcal{F}$ *dans* $[0, +\infty]$ *additive (telle que* $\nu(A\cup B) = \nu(A)+\nu(B)$ *pour* $A\cup B = \emptyset$*), alors si* $\mathcal{A}$ *est la tribu engendrée par* $\mathcal{F}$*, il existe une mesure* $\widehat{\nu}$ *sur la tribu* $\mathcal{A}$ *qui coïncide avec* ν *sur* $\mathcal{F}$ *(c'est-à-dire que pour tout* $F \in \mathcal{F}$*,* $\widehat{\nu(F)} = \nu(F)$*). On dit que* $\widehat{\nu}$ *prolonge* ν *sur la tribu* $\mathcal{A}$*.*

Exemple. *Définition de la mesure de Lebesgue sur* $\mathbb{R}$*,* $\mathbb{R}^n$*,...*

Définition. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré.*

1. *Pour* $A \in \mathcal{A}$*, on dit que* A *est* μ*-négligeable si* $\mu(A) = 0$*.*

2. *Soit une propriété* $\mathcal{P}$ *dépendant des éléments* ω *de* Ω*. On dit que* $\mathcal{P}$ *est vraie* μ*-presque partout (*μ*-presque sûrement sur un espace probabilisé) si l'ensemble des* ω *pour laquelle elle n'est pas vérifiée est* μ*-négligeable.*

Exemple.

- *Mesure de Lebesgue sur* $\mathbb{N}$ *ou* $\mathbb{Q}$*.*
- *La propriété " la suite de fonction* $f_n(x) = x^n$ *converge vers la fonction* $f(x) = 0$ *" est vraie* λ*-presque partout sur* $[0, 1]$*.*
- *Soit* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ *et soit* F *la fonction définie par* $F(x) = \mu(]-\infty, x])$ *pour* $x \in \mathbb{R}$*.*

1.1.4 Fonctions mesurables

Rappel. *Soit* $f : E \mapsto F$*, où* E *et* F *sont 2 espaces métriques.*

- *Pour* $I \subset F$*, on appelle ensemble réciproque de* I *par* f*, l'ensemble* $f^{-1}(I) = \{x \in E, f(x) \in I\}$*.*
- *(*f *continue)* $\Longleftrightarrow$ *(pour tout ouvert* U *de* F *alors* $f^{-1}(U)$ *est un ouvert de* E*).*

Définition. *Soit* $f : E \mapsto F$ *et soit* $\mathcal{I}$ *une tribu sur* F*. On note* $f^{-1}(\mathcal{I})$ *l'ensemble de sous-ensembles de* Ω *tel que* $f^{-1}(\mathcal{I}) = \{f^{-1}(I), I \in \mathcal{I}\}$*.*

Propriété. *Soit* $(\Omega', \mathcal{A}')$ *un espace mesurable et soit* $f : \Omega \mapsto \Omega'$. *Alors* $f^{-1}(\mathcal{A})$ *est une tribu sur* Ω *appelée tribu engendrée par* f.

Définition. *Soit* $(\Omega, \mathcal{A})$ *et* $(\Omega', \mathcal{A}')$ *deux espaces mesurables. Une fonction* $f : \Omega \mapsto \Omega'$ *est dite mesurable pour les tribus* $\mathcal{A}$ *et* $\mathcal{A}'$ *si et seulement si* $f^{-1}(\mathcal{A}') \subset \mathcal{A}$ *(donc si et seulement si* $\forall A' \in \mathcal{A}'$, *alors* $f^{-1}(A') \in \mathcal{A}$*).*

Exemple.
- *Fonction indicatrice.*
- *Combinaison linéaire de fonctions indicatrices.*

Remarque. *Dans le cas où* $(\Omega, \mathcal{A})$ *est un espace probabilisable, et si* $f : \Omega \mapsto \mathbb{R}$, *alors si* f *est une fonction mesurable sur* $\mathcal{A}$ *et* $\mathcal{B}(\mathbb{R})$, *alors* f *est une variable aléatoire.*

Exemple. *Nombre de Piles dans un jeu de Pile/Face.*

Remarque. *Dans le cas où* $(\Omega, \mathcal{A})$ *est un espace mesurable, et si* $f : \Omega \mapsto (\Omega', \mathcal{B}(\Omega'))$, *où* Ω' *est un espace métrique et* $\mathcal{B}(\Omega')$ *l'ensemble des boréliens de* Ω', *si* f *est une fonction mesurable sur* $\mathcal{A}$ *et* $\mathcal{B}(\Omega')$, *alors* f *est dite fonction borélienne.*

Proposition. *Soit* $(\Omega, \mathcal{A})$ *et* $(\Omega', \mathcal{A}')$ *deux espaces mesurables et* $f : \Omega \mapsto \Omega'$. *Soit* $\mathcal{F}$ *une famille de sous-ensembles de* Ω' *telle que* $\sigma(\mathcal{F}) = \mathcal{A}'$. *Alors*

1. $f^{-1}(\mathcal{F})$ *engendre la tribu* $f^{-1}(\mathcal{A})$.
2. *(*f *mesurable)* $\Longleftrightarrow$ *(*$f^{-1}(\mathcal{F}) \subset \mathcal{A}$*)*

Conséquence.
- *Si* $(\Omega, \mathcal{A})$ *et* $(\Omega', \mathcal{A}')$ *sont deux espaces mesurables boréliens, alors toute application continue de* $\Omega \mapsto \Omega'$ *est mesurable.*
- *Pour montrer qu'une fonction* $f : \Omega \mapsto \mathbb{R}$ *est mesurable, il suffit de montrer que la famille d'ensemble* $(\{\omega \in \Omega, f(\omega) \leq a\})_{a \in \mathbb{R}} \in \mathcal{A}$.

Propriété.
- *Soit* f *mesurable de* $(\Omega, \mathcal{A})$ *dans* $(\Omega', \mathcal{A}')$ *et* g *mesurable de* $(\Omega', \mathcal{A}')$ *dans* $(\Omega'', \mathcal{A}'')$. *Alors* $g_0 f$ *est mesurable dans* $\mathcal{A}$ *et* $\mathcal{A}'$.
- *Soit* f_1 *mesurable de* $(\Omega, \mathcal{A})$ *dans* $(\Omega_1, \mathcal{A}_1)$ *et* f_2 *mesurable de* $(\Omega, \mathcal{A})$ *dans* $(\Omega_2, \mathcal{A}_2)$. *Alors* $h : \Omega \mapsto \Omega_1 \times \Omega_2$ *telle que* $h(\omega) = (f_1(\omega), f_2(\omega))$ *est mesurable dans* $\mathcal{A}$ *et* $\mathcal{A}_1 \otimes \mathcal{A}_2$.
- *Soit* $(f_n)_{n \in \mathbb{N}}$ *une suite de fonctions mesurables de* $(\Omega, \mathcal{A})$ *dans* $(\Omega', \mathcal{B}(\Omega'))$, *où* Ω' *est un espace métrique, telle qu'il existe une fonction* f *limite simple de* (f_n) *(donc* $\forall \omega \in \Omega$, $\lim_{n \to \infty} f_n(\omega) = f(\omega)$*). Alors* f *est mesurable dans* $\mathcal{A}$ *et* $\mathcal{B}(\Omega')$.

Définition. *Soit* f *mesurable de* $(\Omega, \mathcal{A}, \mu)$ *dans* $(\Omega', \mathcal{A}')$ *et soit* $\mu_f : \mathcal{A}' \mapsto [0, +\infty]$ *telle que pour tout* $A' \in \mathcal{A}'$, *on ait* $\mu_f(A') = \mu(f^{-1}(A'))$. *Alors* μ_f *est une mesure sur* $(\Omega', \mathcal{A}')$ *appelée mesure image de* μ *par* f.

Cas particulier. *Si* μ *est une mesure de probabilité et si* X *est une variable aléatoire alors* μ_X *est la mesure (loi) de probabilité de la variable aléatoire* X.

1.1.5 Cas des fonctions réelles mesurables

Propriété. *Soit* f *et* g *deux fonctions réelles mesurables (de* $(\Omega, \mathcal{A}, \mu)$ *dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*). Alors* $\alpha.f$, $f + g$, $\min(f, g)$ *et* $\max(f, g)$ *sont des fonctions réelles mesurables.*

Propriété. *Soit* $(f_n)_{n \in \mathbb{N}}$ *une suite de fonctions réelles mesurables. Alors* $\inf(f_n)$ *et* $\sup(f_n)$ *sont des fonctions réelles mesurables.*

Définition. *Soit* $f : \Omega \to \mathbb{R}$. *Alors* f *est dite étagée s'il existe une famille d'ensembles disjoints* $(A_i)_{1 \leq i \leq n}$ *de* Ω *et une famille de réels* $(\alpha_i)_{1 \leq i \leq n}$ *telles que pour tout* $\omega \in \Omega$, *on ait* $f(\omega) = \sum_{i=1}^{n} \alpha_i \mathbb{I}_{A_i}(\omega)$.

Remarque. *Si les* A_i *sont tous dans* $\mathcal{A}$ *tribu sur* Ω, *alors* f *est* $\mathcal{A}$*-mesurable.*

Théorème. *Toute fonction réelle mesurable à valeurs dans* $[0, +\infty]$ *est limite simple d'une suite croissante de fonctions étagées.*

Conséquence. *Soit* f *une fonction réelle mesurable. Alors* f *est limite simple de fonctions étagées.*

1.2 Intégration de Lebesgue

Dans toute la suite, on considère $(\Omega, \mathcal{A}, \mu)$ un espace mesuré.

1.2.1 Intégrale de Lebesgue d'une fonction positive

Définition. 1. *Soit* $f = \mathbb{I}_a$*, où* $A \in \mathcal{A}$*. Alors :*

$$\int f\, d\mu = \int_\omega f(\omega) d\mu(\omega) = \mu(A).$$

2. *Soit* $f = \mathbb{I}_a$*, où* $A \in \mathcal{A}$ *et soit* $B \in \mathcal{A}$*. Alors :*

$$\int_B f\, d\mu = \int_B f(\omega) d\mu(\omega) = \int \mathbb{I}_B \mu(A)(\omega) f(\omega) d\mu(\omega) = \mu(A \cap B).$$

3. *Soit* f *une fonction étagée positive telle que* $f = \sum_{i=1}^n \alpha_i \mathbb{I}_{A_i}$*, où les* $A_i \in \mathcal{A}$ *et* $\alpha_i > 0$ *et soit* $B \in \mathcal{A}$*. Alors :*

$$\int_B f\, d\mu = \int_B f(\omega) d\mu(\omega) = \int \mathbb{I}_B(\omega) f(\omega) d\mu(\omega) = \sum_{i=1}^n \alpha_i \mu(A_i \cap B).$$

Exemple. *Fonction* $\mathbb{I}_{\mathbb{Q}}$*, fonctions en escalier,...*

Définition. *Soit* f *une fonction* $\mathcal{A}$*-mesurable positive et soit* $B \in \mathcal{A}$*. Alors l'intégrale de Lebesgue de* f *par rapport à* μ *sur* B *est :*

$$\int_B f\, d\mu = \int \mathbb{I}_B(\omega) f(\omega) d\mu(\omega) = \sup \left\{ \int_B g\, d\mu, \text{ pour } g \text{ étagée positive telle que } g \leq f \right\}.$$

Propriété. *Soit* f *une fonction* $\mathcal{A}$*-mesurable positive et soit* A *et* $B \in \mathcal{A}$*. Alors :*

1. *Pour* $c \geq 0$, $\int_B cf\, d\mu = c \int_B f\, d\mu$.
2. *Si* $A \subset B$*, alors* $\int_A f\, d\mu \leq \int_B f\, d\mu$.
3. *Si* g *est une fonction* $\mathcal{A}$*-mesurable positive telle que* $0 \leq f \leq g$ *alors* $0 \leq \int_B f\, d\mu \leq \int_B g\, d\mu$.
4. *Si* $\mu(B) = 0$ *alors* $\int_B f\, d\mu = 0$.

Théorème (Théorème de convergence monotone (Beppo-Lévi))**.** *Si* $(f_n)_n$ *est une suite croissante de fonctions mesurables positives convergeant simplement vers* f *sur* Ω*, alors :*

$$\lim_{n\to\infty} \left(\int f_n d\mu \right) = \int f\, d\mu = \int \lim_{n\to\infty} f_n d\mu.$$

Conséquence. *Pour les séries de fonctions mesurables positives, on peut toujours appliquer le Théorème de convergence monotone et donc inverser la somme et l'intégrale.*

Lemme (Lemme de Fatou)**.** *Soit* $(f_n)_n$ *est une suite de fonctions mesurables positives alors :*

$$\int \left(\liminf_{n\to\infty} f_n \right) d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$

Exemple. *Appliquer Fatou à* (f_n) *telle que* $f_{2n} = \mathbb{I}_A$ *et* $f_{2n+1} = \mathbb{I}_B$*.*

1.2.2 Intégrale de Lebesgue d'une fonction réelle et propriétés

Définition. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, $B \in \mathcal{A}$ et soit f une fonction $\mathcal{A}$-mesurable à valeurs réelles telle que $f = f^+ - f^-$ avec $f^+ = \max(f, 0)$ et $f^- = \max(-f, 0)$. On dit que f est μ-intégrable sur B si $\displaystyle\int_B |f|\, d\mu < +\infty$. On a alors*

$$\int_B f\, d\mu = \int_B f^+\, d\mu - \int_B f^-\, d\mu.$$

Notation. *Lorsque f est μ-intégrable sur B, soit $\displaystyle\int |f|\, d\mu < +\infty$, on note $f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ (on dit que f est $\mathcal{L}^1$).*

Exemple. *Intégrale de Riemann et intégrale de Lebesgue.*
Cas de la masse de Dirac.

Propriété. *On suppose que f et $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$. Alors :*

1. $\displaystyle\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$ *pour* $(\alpha, \beta) \in \mathbb{R}^2$.
2. *Si* $f \leq g$ *alors* $\displaystyle\int f\, d\mu \leq \int g\, d\mu$.

Théorème (Théorème de convergence dominée de Lebesgue)**.** *Soit $(f_n)_n$ est une suite de fonctions de $\mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ telles que pour tout $n \in \mathbb{N}$, $|f_n| \leq g$ avec $g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$. Si on suppose que (f_n) converge simplement vers f sur Ω alors :*

$$\lim_{n \to \infty} \int f_n d\mu = \int f\, d\mu.$$

Extension. *Le Théorème de Lebesgue s'applique également dans le cas où $(f_n)_n$ converge presque partout vers f.*

Exemple. *Convergence d'intégrale dépendant d'un paramètre : par exemple* $\displaystyle\int_0^\infty \frac{f(x)}{1 + x^n} dx$.

Théorème (Inégalité de Jensen)**.** *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, soit $\phi : \mathbb{R} \mapsto \mathbb{R}$ une fonction convexe et soit $f : \Omega \mapsto \mathbb{R}$ mesurable telle que $\phi(f)$ soit une fonction intégrable par rapport à P. Alors :*

$$\phi\left(\int f\, d\mathbb{P}\right) \leq \int \phi(f)\, d\mathbb{P}.$$

Exemple. *Soit X une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$. Alors $\phi(\mathbb{E}X) \leq \mathbb{E}(\phi(X))$.*

1.2.3 Mesures induites et densités

Théorème (Théorème du Transport)**.** *Soit f une fonction mesurable de $(\Omega, \mathcal{A}, \mu)$ dans $(\Omega', \mathcal{A}')$ telle que μ_f soit la mesure induite par f (donc $\mu_f(A') = \mu(f^{-1}(A'))$ pour $A' \in \mathcal{A}'$) et soit ϕ une fonction mesurable de $(\Omega', \mathcal{A}')$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors, si $\phi_0 f \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$,*

$$\int_{\Omega'} \phi\, d\mu_f = \int_\Omega \phi_0 f\, d\mu.$$

Définition. *Soit μ et ν deux mesures sur $(\Omega, \mathcal{A})$. On dit que μ domine ν (ou ν est dominée par μ) et que ν est absolument continue par rapport à μ lorsque pour tout $A \in \mathcal{A}$, $\mu(A) = 0 \Longrightarrow \nu(A) = 0$.*

Propriété. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et f une fonction définie sur $(\Omega, \mathcal{A})$ mesurable et positive. On suppose que pour $A \in \mathcal{A}$, $\nu(A) = \displaystyle\int_A f\, d\mu$. Alors, ν est une mesure sur $(\Omega, \mathcal{A})$, dominée par μ. De plus, pour toute fonction g définie sur $(\Omega, \mathcal{A})$ mesurable et positive,*

$$\int g d\nu = \int g.f d\mu.$$

Enfin, g est ν intégrable si et seulement si $g.f$ est μ intégrable.

Définition. *On dit que μ mesure sur $(\Omega, \mathcal{A})$ est σ-finie lorsqu'il existe une famille $(A_i)_{i\in I}$, avec I dénombrable, d'ensembles de $\mathcal{A}$ telle que $\bigcup A_i = \Omega$ et $\mu(A_i) < +\infty$ pour tout $i \in I$.*

Théorème (Théorème de Radon-Nikodym)**.** *On suppose que μ et ν sont deux mesures σ-finies sur $(\Omega, \mathcal{A})$ telles que μ domine ν. Alors il existe une fonction f définie sur $(\Omega, \mathcal{A})$ mesurable et positive, appelée densité de ν par rapport à μ, telle que pour tout $A \in \mathcal{A}$, $\nu(A) = \displaystyle\int_A f\,d\mu$.*

Théorème (Théorème de Fubini)**.** *Soit $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ et $\mu = \mu_1 \otimes \mu_2$ (mesures σ finies), où $(\Omega_1, \mathcal{A}_1, \mu_1)$ et $(\Omega_2, \mathcal{A}_2, \mu_2)$ sont des espaces mesurés. Soit une fonction $f : \Omega \mapsto \mathbb{R}$, $\mathcal{A}$-mesurable et μ-intégrable. alors :*

$$\int_\Omega f d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1) = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2).$$

1.2.4 Espaces $\mathcal{L}^p$

Définition. *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. On appelle espace $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$, où $p > 0$, l'ensemble des fonctions $f : \Omega \mapsto \mathbb{R}$, mesurables et telles que $\int |f|^p d\mu < +\infty$.*

Définition. *Pour $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$, où $p > 0$, on note $\| f \|_p = \left(\displaystyle\int |f|^p d\mu \right)^{1/p}$.*

Propriété (Inégalité de Hölder)**.** *Soit $p > 1$ et $q > 1$ tels que $\dfrac{1}{p} + \dfrac{1}{q} = 1$, et soit $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ et $g \in \mathcal{L}^q(\Omega, \mathcal{A}, \mu)$. Alors, $f\,g \in \mathcal{L}^1(\Omega, \mathcal{A}, \mu)$ et*

$$\| f\,g \|_1 \leq \| f \|_p . \| g \|_q .$$

Propriété (Inégalité de Minkowski)**.** *Soit $p > 1$ et soit f et $g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$. Alors, $f + g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ et*

$$\| f + g \|_p \leq \| f \|_p + \| g \|_p .$$

Remarque. *Pour $p > 1$, $\| . \|_p$ définie ainsi sur une semi-norme sur $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$. Pour obtenir une norme, il faut se place dans l'espace $\mathbb{L}^p(\Omega, \mathcal{A}, \mu)$ obtenu en "quotientant" $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ par la relation d'équivalence $f = g$ μ-presque partout (c'est-à-dire que dans $\mathbb{L}^p(\Omega, \mathcal{A}, \mu)$ on dira que $f = g$ lorsque $f = g$ μ-presque partout).*

Définition. *Pour f et $g \in \mathbb{L}^2(\Omega, \mathcal{A}, \mu)$, on définit le produit scalaire $< f, g >= \displaystyle\int f.g\,d\mu$. On muni ainsi $\mathbb{L}^2(\Omega, \mathcal{A}, \mu)$ d'une structure d'espace de Hilbert. On dira que f est orthogonale à g lorsque $< f, g >= 0$.*

Conséquence. *Si A est un sous-espace vectoriel fermé de $\mathbb{L}^2(\Omega, \mathcal{A}, \mu)$ (par exemple un sous-espace de dimension finie), alors pour tout $f \in \mathbb{L}^2(\Omega, \mathcal{A}, \mu)$, il existe un unique projeté orthogonal de f sur A, noté f_A, qui vérifie $f_A = \mathrm{Arg}\inf_{g\in A} \| g - f \|_2$.*

2 Applications de la théorie de la mesure et de l'intégration en Probabilités

2.1 Espérance de variables aléatoires

Définition. *Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Alors si $X \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, on définit l'espérance de X par le nombre $\mathbb{E}X = \displaystyle\int X d\mathbb{P}$. Plus généralement, si $\phi : \mathbb{R} \mapsto \mathbb{R}$ est borélienne et si $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, on définit l'espérance de $\phi(X)$ par $\mathbb{E}\,\phi(X) = \displaystyle\int \phi(X) d\mathbb{P}$.*

Propriété. *Si X est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, si $\phi : \mathbb{R} \mapsto \mathbb{R}$ est borélienne telle que $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, et si $\mathbb{P}_X$ est la mesure de probabilité de X alors :*

$$\mathbb{E}\,\phi(X) = \int_{\mathbb{R}} \phi(x)\, d\mathbb{P}_X(x).$$

Conséquence. • *Si $\mathbb{P}_X$ est absolument continue par rapport à la mesure de Lebesgue (donc X est une v.a. dite absolument continue), de densité f_X, alors $\mathbb{E}\,\phi(X) = \int_{\mathbb{R}} \phi(x) f_X(x) dx$.*

- *Si $\mathbb{P}_X$ est absolument continue par rapport à la mesure de comptage sur $\mathbb{N}$ (donc X est une v.a. dite discrète), de densité p_X, alors $\mathbb{E}\,\phi(X) = \sum_{k=0}^{\infty} p_X(k)\,\phi(k)$.*

Propriété. 1. *Soit X et Y des variables aléatoires telles que X et $Y \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$. Alors pour tout $(a, b) \in \mathbb{R}^2$, $aX + bY \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ et*

$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y.$$

2. *Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, et soit $A \in \mathcal{A}$. Alors $\mathbb{E}(\mathbb{I}_A(X)) = \mathbb{P}(X \in A)$.*

3. *Soit X et Y des variables aléatoires telles que $X \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ et $Y \in \mathbb{L}^q(\Omega, \mathcal{A}, \mathbb{P})$ avec $\frac{1}{p} + \frac{1}{q} = 1$ et $p > 1$, $q > 1$. Alors $X.Y \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ et*

$$\mathbb{E}|X.Y| \leq (\mathbb{E}|X|^p)^{1/p} (\mathbb{E}|Y|^q)^{1/q}.$$

4. *Soit X et Y des variables aléatoires telles que X et $Y \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$, avec $p \geq 1$. Alors $X + Y \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ et*

$$(\mathbb{E}|X + Y|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}.$$

5. *Soit X une variable aléatoire telle que $X \in \mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ pour $p > 0$. Alors pour tout $0 < r \leq p$, $X \in \mathbb{L}^r(\Omega, \mathcal{A}, \mathbb{P})$ et*

$$(\mathbb{E}|X|^r)^{1/r} \leq (\mathbb{E}|X|^p)^{1/p}.$$

6. *Si X est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, si $\phi : \mathbb{R} \mapsto \mathbb{R}$ est une fonction borélienne convexe telle que X et $\phi(X) \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, alors*

$$\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}X).$$

Définition. *Pour X et Y des variables aléatoires telles que X et $Y \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, on définit la covariance de X et Y par*

$$cov(X, Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right];$$

On appelle variance de X, $var(X) = cov(X, X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.

Propriété. *Sur $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$, $cov(.,.)$ définit un produit scalaire. De plus*

$$|cov(X, Y)|^2 \leq var(X).var(Y).$$

2.2 Fonction de répartition et quantiles d'une loi de probabilité

Il y a une correspondance bijective entre la connaissance de $\mathbb{P}_X$ et celle de $F_X = \mathbb{P}_X(]-\infty, x])$. La fonction de répartition permet également de définir les quantiles qui sont essentiels à la construction d'intervalles de confiance et de test.

Soit $\alpha \in [0, 1]$. Des propriétés de la fonction de répartition, on en déduit qu'il existe $x_\alpha \in \mathbb{R}$, tel que :

$$\lim_{x \to x_\alpha} F_X(x) \leq \alpha \leq F_X(x_\alpha). \qquad (1)$$

Soit $I_\alpha = \{x_\alpha \in \mathbb{R}$ tel que x_α vérifie $(1)\}$. On appelle **quantile** (ou fractile, ou percentile en anglais) d'ordre α de la loi $\mathbb{P}_X$, noté q_α, le milieu de l'intervalle I_α. Evidemment, lorsque X admet une distribution absolument continue par rapport à la mesure de Lebesgue, $q_\alpha = F_X^{-1}(\alpha)$, où F_X^{-1} désigne la fonction réciproque de F_X.

Deux cas particuliers sont à connaître :

1/ pour $\alpha = 0.5$, $q_{0.5}$ est appelé la **médiane** de $\mathbb{P}_X$;

2/ pour $\alpha = 0.25$ et $\alpha = 0.75$ (respectivement), $q_{0.25}$ et $q_{0.25}$ sont appelés premier et troisième **quartile** (respectivement) de $\mathbb{P}_X$.

3/ pour $\alpha = 0.1, \ldots, 0.9$, on parlera de **décile** de $\mathbb{P}_X$.

2.3 Principales lois de probabilités

Loi uniforme discrète :

C'est la loi de probabilité discrète à valeurs dans $\{x_1, \ldots, x_n\}$ telle que

$$\mathbb{P}(X = x_i) = \frac{1}{n}.$$

On alors : $\mathbb{E}X = \frac{1}{n}(x_1 + \ldots + x_n)$ et $\text{var}(X) = \frac{1}{n}(x_1^2 + \ldots + x_n^2) - (\mathbb{E}X)^2$.

Loi de Bernoulli :

C'est la loi de probabilité discrète notée $\mathcal{B}(p)$ à valeurs dans $\{0, 1\}$ telle que

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

On alors : $\mathbb{E}X = p$ et $\text{var}(X) = p(1 - p)$.

Loi binomiale :

C'est la loi de probabilité discrète notée $\mathcal{B}(n, p)$ à valeurs dans $\{0, 1, \ldots, n\}$ telle que

$$\mathbb{P}(X = k) = C_n^k \cdot p^k \cdot (1 - p)^{n-k} \quad \text{pour } k \in \{0, 1, \ldots, n\}.$$

On alors : $X = X_1 + \cdots + X_n$, où (X_i) est une suite de v.a.i.i.d. de loi $\mathcal{B}(p)$, d'où $\mathbb{E}X = n \cdot p$ et $\text{var}(X) = n \cdot p(1 - p)$.

Loi de Poisson :

C'est la loi de probabilité discrète notée $\mathcal{P}(\theta)$ à valeurs dans $\mathbb{N}$ telle que

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

On alors $\mathbb{E}X = \theta$ et $\text{var}(X) = \theta$.

Loi uniforme sur $[a, b]$ **:**

Cette loi est généralement notée $\mathcal{U}([a, b])$, où $-\infty < a < b < \infty$. C'est la loi de probabilité à valeurs dans $[a, b]$ de densité par rapport à la mesure de Lebesgue :

$$f_X(x) = \frac{1}{b - a} \cdot \mathbb{I}_{x \in [a,b]}.$$

On a alors $\mathbb{E}X = \frac{b + a}{2}$ et $\text{var}(X) = \frac{(b - a)^2}{12}$.

Loi Gamma :

Cette loi est généralement notée $\gamma(p, \theta)$, où $p > 0$ et $\theta > 0$. C'est la loi de probabilité à valeurs dans $\mathbb{R}_+$ de densité par rapport à la mesure de Lebesgue :

$$f_X(x) = \frac{\theta^p}{\Gamma(p)} \cdot e^{-\theta \cdot x} \cdot x^{p-1} \cdot \mathbb{I}_{x \in \mathbb{R}_+}.$$

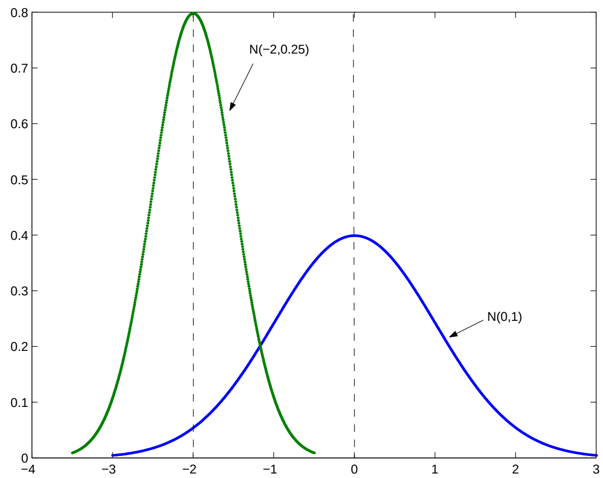

Figure 1: Représentation de la densité de la loi $\mathcal{N}(0,1)$ et de la loi $\mathcal{N}(-2, 0.5^2)$

On a alors $\mathbb{E}X = \dfrac{p}{\theta}$ et $\mathrm{var}(X) = \dfrac{p}{\theta^2}$.
Si $X \sim \gamma(p,\theta)$ et $Y \sim \gamma(q,\theta)$ avec X et Y indépendantes et $p > 0$ et $q > 0$, alors $X + Y \sim \gamma(p+q,\theta)$.
Pour $p = 1$, la loi $\gamma(p,\theta)$ est la loi exponentielle $\mathcal{E}(\theta)$.

Loi Béta :

Cette loi est généralement notée $\beta(p,\theta)$, où $p > 0$ et $q > 0$. C'est la loi de probabilité à valeurs dans $[0,1]$ de densité par rapport à la mesure de Lebesgue :

$$f_X(x) = \frac{x^p(1-x)^{q-1}}{B(p,q)} \cdot x^{p-1} \cdot \mathbb{I}_{x\in[0,1]}, \quad \text{où} \quad B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

On a alors $\mathbb{E}X = \dfrac{B(p+1,q)}{B(p,q)}$ et $\mathrm{var}(X) = \dfrac{p \cdot q}{(p+q)^2(p+q+1)}$.
Si $X \sim \gamma(p,\theta)$ et $Y \sim \gamma(q,\theta)$ avec X et Y indépendantes et $p > 0$ et $q > 0$, alors $\dfrac{X}{X+Y} \sim \beta(p,q)$.
Pour $p = 1$, la loi $\gamma(p,\theta)$ est la loi exponentielle $\mathcal{E}(\theta)$.

Loi normale (ou gaussienne) centrée réduite :

Cette loi est généralement notée $\mathcal{N}(0,1)$. C'est la loi de probabilité à valeurs dans $\mathbb{R}$ de densité par rapport à la mesure de Lebesgue :

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

On a :

$$\mathbb{E}(X) = 0 \quad \text{et} \quad \mathrm{var}(X) = 1.$$

Loi normale (ou gaussienne) de moyenne m et de variance σ^2 :

Si Z suit la loi $\mathcal{N}(0,1)$, $X = m + \sigma Z$ suit par définition la loi $\mathcal{N}(m,\sigma^2)$, loi normale d'espérance m et de variance σ^2. La densité de X est donnée par :

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

La figure A.1. représente la densité de la loi normale centrée réduite et celle d'une loi normale non centrée et non réduite.

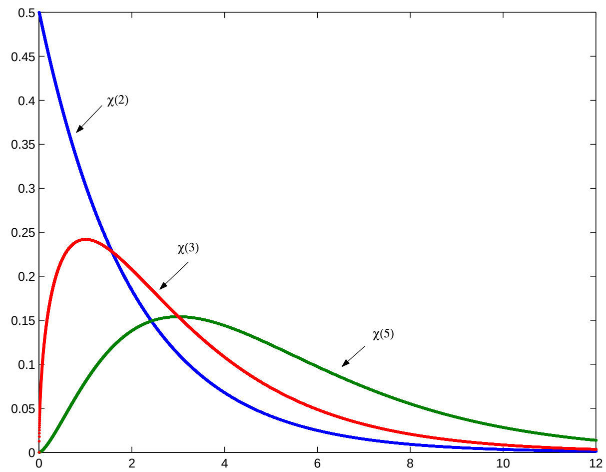

Figure 2: Représentation de la densité des lois $\chi^2(2)$, $\chi^2(3)$ et $\chi^2(5)$

A partir de la loi gaussienne, on peut en déduire les lois suivantes.

Loi du χ^2 à n degrés de libertés :

Soit $X_1, \cdots, X_n$, n variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$, alors

$$S = X_1^2 + \cdots + X_n^2$$

suit une loi du χ^2 à n degrés de libertés, loi notée $\chi^2(n)$. Cette loi est à valeurs dans $\mathbb{R}_+$, d'espérance n et de variance $2n$. C'est aussi la loi Gamma $\gamma(n/2, 1/2)$, c'est-à-dire que $X \sim \chi^2(n)$ admet pour densité par rapport à la mesure de Lebesgue :

$$f_X(x) = \frac{1}{2^{n/2} \cdot \Gamma(n/2)} x^{n/2-1} \exp\left(-\frac{x}{2}\right) \cdot \mathbb{I}_{\{x \geq 0\}},$$

où la fonction Gamma est telle que $\Gamma(a) = \displaystyle\int_0^\infty x^{a-1} \cdot e^{-x}$ pour $a \geq 0$. Enfin, si X suit une loi $\chi^2(n)$, par définition on dira que $Y = \sigma^2 \cdot X$ suit une loi $\sigma^2 \cdot \chi^2(n)$. La figure A.2. exhibe trois tracés différents de densité de loi du χ^2.

Loi de Student à n degrés de libertés :

La loi de Student à n degrés de liberté, notée $T(n)$, est la loi du quotient

$$T = \frac{N}{\sqrt{S/n}}$$

où N suit une loi $\mathcal{N}(0,1)$ et S suit une loi $\chi^2(n)$, N et S étant deux variables aléatoires indépendantes. Il est également possible de déterminer la densité d'une telle loi par rapport à la mesure de Lebesgue, à savoir,

$$f_X(x) = \frac{1}{\sqrt{n} \cdot B(1/2, n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2},$$

où la fonction Beta est telle que $B(a,b) = \dfrac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)}$ pour $a > 0$ et $b > 0$. La figure A.3. illustre deux exemples de cette densité, que l'on compare également avec la densité de la loi normale centrée réduite.

Remarque : Par la loi des grands nombres, plus n est grand, plus S est proche de son espérance qui vaut n. Le dénominateur est donc proche de 1. Il s'ensuit que la loi $T(n)$ est d'autant plus proche d'une loi normale

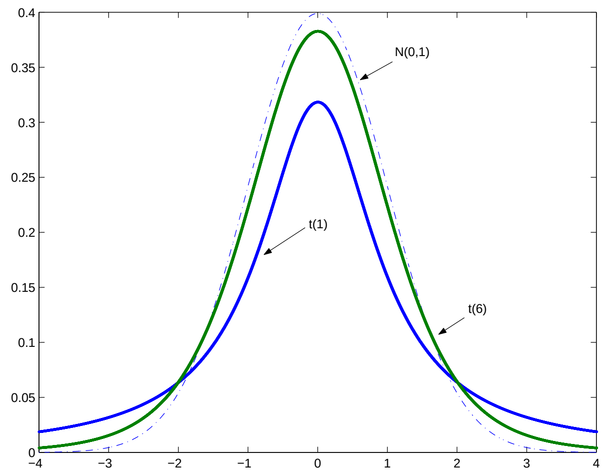

Figure 3: Représentation de la densité des lois $T(2)$, $T(6)$, à comparer avec celle de la loi $\mathcal{N}(0,1)$

que n est grand.

Un des principaux intérêt de la loi de Student réside dans le fait que si $X_1, \cdots, X_n$ sont n variables aléatoires indépendantes de loi $\mathcal{N}(m, \sigma^2)$, si on considère la moyenne et la variance empiriques :

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n) \quad \text{et} \quad \bar{\sigma}_n^2 = \frac{1}{n-1}\left((X_1 - \bar{X}_n)^2 + \cdots + (X_d - \bar{X}_n)^2\right),$$

alors

$$T = \frac{\sqrt{n} \cdot (\bar{X}_n - m)}{\sqrt{\bar{\sigma}_n^2}}$$

suit une loi de Student à $(n-1)$ degrés de liberté.

Loi de Fisher à n_1 et n_2 degrés de liberté :

Soit S_1 et S_2 deux variables aléatoires indépendantes de loi respectives $\chi^2(n_1)$ et $\chi^2(n_2)$. Alors par définition :

$$F = \frac{S_1/n_1}{S_2/n_2}$$

suit une loi de Fisher à n_1 et n_2 degrés de liberté, notée $F(n_1, n_2)$.

Remarque : Par les mêmes considérations que précédemment, la loi F est d'autant plus proche de 1 que les degrés de liberté n_1 et n_2 sont grands.

On a également les propriétés suivantes :

- Si F suit une loi $F(n_1, n_2)$, alors la loi de $\frac{n_1}{n_2} F$ est une loi beta de seconde espèce de paramètres $(n_1/2, n_2/2)$, c'est-à-dire que F est à valeurs dans $\mathbb{R}_+$ et admet la densité par rapport à la mesure de Lebesgue :
$$f_X(x) = \frac{1}{B(n_1/2, n_2/2)} n_1^{n_1/2} \cdot n_2^{n_2/2} \frac{x^{n_1/2-1}}{(n_2 + n_1 \cdot x)^{(n_1+n_2)/2}} \mathbb{I}_{\{x \geq 0\}},$$
la notation B désignant encore la fonction Beta.

- Si $F \sim F(n_1, n_2)$, alors $\mathbb{E}(F) = \dfrac{n_2}{n_2 - 2}$ lorsque $n_2 > 2$ et $\mathrm{var}(F) = \dfrac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 4)(n_2 - 2)^2}$ lorsque $n_2 > 4$.

- Si T suit une loi de Student $T(n)$, alors T^2 suit une loi de Fisher $F(1, n)$.

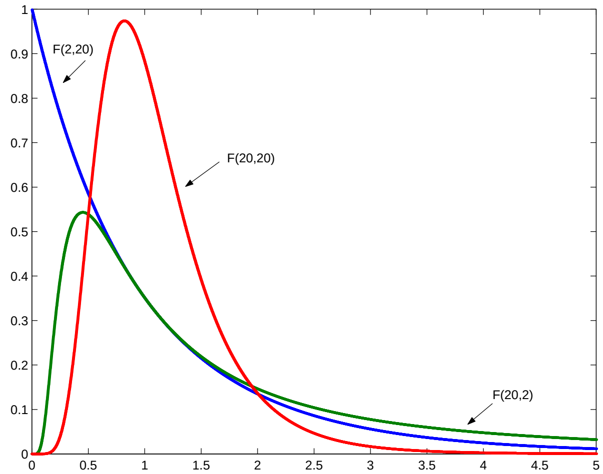

Figure 4: Représentation de la densité des lois $F(2,20)$, $F(20,2)$ et $F(20,20)$

La figure A.4. donne une idée de la distribution d'une loi de Fisher pour différents choix des paramètres.

2.4 Indépendance

Définition. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé.*

- *Soit* $(A_i)_{i\in I}$ *une famille dénombrable d'événements de* $\mathcal{A}$. *On dit que les événements* $(A_i)_{i\in I}$ *sont indépendants si et seulement si pour tous les sous-ensembles finis* $K \subset I$,
$$\mathbb{P}\left(\bigcap_{i\in K} A_i\right) = \prod_{i\in K} \mathbb{P}(A_i).$$
- *Soit* $(\mathcal{A}_i)_{i\in I}$ *une famille de sous-tribus de* $\mathcal{A}$ *(donc pour tout* $i \in I$, $\mathcal{A}_i \subset \mathcal{A}$*). On dit que les tribus* $(\mathcal{A}_i)_{i\in I}$ *sont indépendantes si et seulement si pour tous les sous-ensembles finis* $K \subset I$, *et pour tous les événements* $A_k \in \mathcal{A}_k$ *avec* $k \in K$, *les* A_k *sont indépendants.*
- *Soit* $(X_i)_{i\in I}$ *des variables aléatoires sur* $(\Omega, \mathcal{A})$ *à valeurs dans* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *On dit que les v.a.* $(X_i)_{i\in I}$ *sont indépendantes si et seulement si les tribus engendrées* $(X_i^{-1}(\mathcal{B}(\mathbb{R})))_{i\in I}$ *sont indépendantes.*

Proposition. *Si* $(X_1, \cdots, X_n)$ *sont des variables aléatoires sur* $(\Omega, \mathcal{A}, \mathbb{P})$. *Alors les* (X_i) *sont indépendantes si et seulement si* $\mathbb{P}_{(X_1,\cdots,X_n)} = \bigotimes_{i=1}^{n} \mathbb{P}_{X_i}$.

Proposition. *Si* $(X_i)_{i\in I}$ *sont des variables aléatoires indépendantes sur* $(\Omega, \mathcal{A}, \mathbb{P})$. *Alors les* (X_i) *sont indépendantes si et seulement si pour tout* $J \subset I$, J *fini, pour toutes fonctions boréliennes* $(g_j)_{j\in J}$ *telles que* $g_j(X_j)$ *soit intégrable, alors*
$$\mathbb{E}\left(\prod_{j\in J} g_j(X_j)\right) = \prod_{j\in J} \mathbb{E}(g_j(X_j)).$$

Corollaire. $(X_1, \cdots, X_n)$ *sont des variables aléatoires indépendantes si et seulement si pour tout* $(t_1, \cdots, t_n) \in \mathbb{R}^n$,
$$\phi_{(X_1,\cdots,X_n)}(t_1, \cdots, t_n) = \prod_{j=1}^{n} \phi_{X_j}(t_j).$$

Lemme (Lemme de Borel-Cantelli)**.** *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite d'événements sur* $(\Omega, \mathcal{A}, \mathbb{P})$.

1. *Si* $\sum \mathbb{P}(A_n) < +\infty$ *alors* $\mathbb{P}(\limsup A_n) = 0$.
2. *Si les* (A_n) *sont indépendants,* $\sum \mathbb{P}(A_n) = +\infty$ *implique que* $P(\limsup A_n) = 1$.

2.5 Vecteurs aléatoires

Définition. *On dit que X est un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, un espace probabilisé, si X est une fonction mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.*

Définition. *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$. Alors la loi (ou mesure) de probabilité de X, $\mathbb{P}_X$, est définie de façon univoque à partir de la fonction de répartition de X, telle que pour $x = (x_1, \cdots, x_d)$,*

$$F_X(x) = \mathbb{P}_X(\prod_{i=1}^{d}]-\infty, x_i]) = \mathbb{P}(X \in \prod_{i=1}^{d}]-\infty, x_i]).$$

Propriété. *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$. On suppose que $X = (X_1, \cdots, X_d)$. Alors les X_i sont des variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$, de fonction de répartition*

$$F_{X_i}(x_i) = \lim_{\substack{x_j \to +\infty \\ j \neq i}} F_X(x_1, \cdots, x_i, \cdots, x_d).$$

Les mesures de probabilités P_{X_i} déterminées de façon univoque à partir des F_{X_i} sont appelées lois marginales de X.

On se place maintenant dans la base canonique orthonormale de $\mathbb{R}^d$. Si Z est un vecteur aléatoire à valeurs sur $\mathbb{R}^d$, on définit $\mathbb{E}(Z)$, le vecteur dont les coordonnées sont les espérances des coordonnées de Z. Ainsi, si dans la base canonique de $\mathbb{R}^d$, $Z = (Z_1, \cdots, Z_d)'$,

$$\mathbb{E}(Z) = \mathbb{E}\begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix} = \begin{pmatrix} \mathbb{E}(Z_1) \\ \vdots \\ \mathbb{E}(Z_d) \end{pmatrix}.$$

De la même manière, on définira l'espérance d'une matrice dont les coordonnées sont des variables aléatoires par la matrice dont les coordonnées sont les espérances de chacune de ces variables aléatoires.

Ceci nous permet de définir la matrice de variance-covariance de Z de la manière suivante :

$$\text{var}(Z) = \mathbb{E}\left[(Z - \mathbb{E}(Z)).(Z - \mathbb{E}(Z))'\right]$$

donc si $Z = (Z_1, \cdots, Z_d)'$,

$$\text{var}\begin{pmatrix} Z_1 \\ \vdots \\ Z_d \end{pmatrix} = \begin{pmatrix} \text{var}(Z_1) & \text{Cov}(Z_1, Z_2) & \cdots & \text{Cov}(Z_1, Z_d) \\ \text{Cov}(Z_1, Z_2) & \text{var}(Z_2) & \cdots & \text{Cov}(Z_2, Z_d) \\ \vdots & \vdots & \cdots & \vdots \\ \text{Cov}(Z_1, Z_d) & \text{Cov}(Z_2, Z_d) & \cdots & \text{var}(Z_d) \end{pmatrix}$$

matrice (d, d) dont les éléments diagonaux sont les variances et les éléments non diagonaux sont les covariances des coordonnées de Z (remarquons que la variance de Z_1 est aussi la covariance de Z_1 et de Z_1).

On vérifie également le résultat suivant : si C est une matrice (p, d) à coordonnées constituées de réels constants et si Z est un vecteur aléatoire à valeurs dans $\mathbb{R}^d$, alors $C \cdot Z$ est un vecteur de taille p de matrice de variance-covariance

$$\text{var}(C \cdot Z) = C \cdot \text{var}(Z) \cdot C'.$$

En particulier, si p vaut 1, alors $C = h'$ où h est un vecteur de taille d, et :

$$\text{var}(h' \cdot Z) = h' \cdot \text{var}(Z) \cdot h.$$

Notez que cette dernière quantité est un scalaire. Soit $Y_1, \cdots, Y_d$ des variables aléatoires indépendantes de même loi $\mathcal{N}(0, \sigma^2)$, indépendantes (ce qui, dans le cas gaussien, est équivalent à $\text{cov}(Y_i, Y_j) = 0$ pour $i \neq j$). On considère le vecteur $Y = (Y_1, \cdots, Y_d)'$. En raison de l'indépendance, Y est un vecteur gaussien admettant une densité f_Y (par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$) qui est le produit des densités de chacune des coordonnées, soit :

$$\begin{aligned} f_Y(y_1,\cdots,y_d) &= f_{Y_1}(y_1)\times f_{Y_2}(y_2)\times\cdots\times f_{Y_d}(y_d) \\ &= \left(2\pi\sigma^2\right)^{-d/2}\exp\left(-\frac{1}{2\sigma^2}(y_1^2+\cdots+y_d^2)\right) \\ &= \left(2\pi\sigma^2\right)^{-d/2}\exp\left(-\frac{\|y\|^2}{2\sigma^2}\right), \end{aligned}$$

avec $y=(y_1,\cdots,y_d)$. On voit donc que la densité de Y ne dépend que de la norme $\|Y\|$: elle est constante sur toutes les sphères centrées en zéro. Cela implique qu'elle est invariante par rotation ou symétrie orthogonale d'axe passant par 0 : elle est invariante par toutes les isométries de $\mathbb{R}^d$: on dira que Y suit une loi gaussienne isotrope. Rappelons que les isométries correspondent à des changements de bases orthonormées (BON). En conséquence, on a la première propriété importante :

Propriété. *Soit Y un vecteur aléatoire de $\mathbb{R}^d$ de loi normale isotrope variance σ^2, c'est-à-dire que dans une BON les coordonnées de Y vérifient $\mathbb{E}(Y)=0$ et $var(Y)=\sigma^2\cdot Id$. Alors les coordonnées de Y dans toute BON sont encore des lois $\mathcal{N}(0,\sigma^2)$ indépendantes.*

Voici maintenant l'un des résultats (encore appelé Théorème de Cochran) que nous utilisons le plus et nous en donnons donc une démonstration.

Théorème (Théorème de Cochran). *Soit E_1 et E_2, deux sous-espaces vectoriels orthogonaux de $E=\mathbb{R}^d$ de dimensions respectives k_1 et k_2 et soit Y un vecteur aléatoire de $\mathbb{R}^d$ de loi normale centrée isotrope de variance σ^2. Alors $P_{E_1}(Y)$ et $P_{E_2}(Y)$ sont deux variables aléatoires gaussienne centrées indépendantes et $\|P_{E_1}(Y)\|^2$ (resp. $\|P_{E_2}(Y)\|^2$) est une loi $\sigma^2\cdot\chi^2(k_1)$ (resp. $\sigma^2\cdot\chi^2(k_2)$). Ce théorème se généralise naturellement pour $2<m\leq d$ sous-espaces vectoriels orthogonaux $(E_i)_{1\leq i\leq m}$ de $E=\mathbb{R}^d$.*

Démonstration : Soit $(e_1,\cdots,e_{k_1})$ et $(e_{k_1+1},\cdots,e_{k_1+k_2})$ deux BON de E_1 et E_2 (respectivement). L'ensemble de ces deux bases peut être complété en

$$(e_1,\cdots,e_{k_1},e_{k_1+1},\cdots,e_{k_1+k_2},e_{k_1+k_2+1},\cdots,e_d)$$

pour former une BON de $\mathbb{R}^d$ (du fait que E_1 et E_2 sont orthogonaux).

Soit $(Y_1,\cdots,Y_d)$, les coordonnées de Y dans cette base; elles sont indépendantes de loi $\mathcal{N}(0,\sigma^2)$ car le changement de base est orthonormal et nous avons vu que la distribution de Y était conservé par transformation isométrique. Comme

$$P_{E_1}(Y)=Y_1e_1+\cdots+Y_{k_1}e_{k_1} \implies \|P_{E_1}(Y)\|^2=\sigma^2\left(\left(\frac{Y_1}{\sigma}\right)^2+\cdots+\left(\frac{Y_{k_1}}{\sigma}\right)^2\right)$$

$$P_{E_2}(Y)=Y_{k_1+1}e_{k_1+1}+\cdots+Y_{k_1+k_2}e_{k_1+k_2} \implies \|P_{E_2}(Y)\|^2=\sigma^2\left(\left(\frac{Y_{k_1+1}}{\sigma}\right)^2+\cdots+\left(\frac{Y_{k_1+k_2}}{\sigma}\right)^2\right).$$

On voit bien ainsi l'indépendance entre les deux projections et le fait que la loi de $\|P_{E_1}(Y)\|^2$ (resp. $\|P_{E_2}(Y)\|^2$) est une loi $\sigma^2\cdot\chi^2(k_1)$ (resp. $\sigma^2\cdot\chi^2(k_2)$). ■

On peut définir plus généralement un vecteur gaussien Y à valeurs dans $\mathbb{R}^d$ (non dégénéré), d'espérance $\mu\in\mathbb{R}^d$ et de matrice de variance-covariance Σ quelconques (du moment que Σ soit une matrice de Toeplitz définie positive). Cela équivaut à définir un vecteur aléatoire de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$,

$$f_Y(y)=\frac{(2\pi)^{-n/2}}{|\Sigma|}\exp\left(-\frac{1}{2}(y-\mu)'.\Sigma^{-1}.(y-\mu)\right),$$

pour $y\in\mathbb{R}^d$, et avec $|\Sigma|$ le déterminant de la matrice Σ. Remarquons une nouvelle fois que l'espérance et la variance définissent complètement la loi de probabilité d'un vecteur gaussien.

A partir des propriétés générales sur les vecteurs aléatoires, on obtient le fait que :

Propriété. *Soit Y un vecteur gaussien à valeurs dans $\mathbb{R}^d$ (non dégénéré), d'espérance $\mu \in \mathbb{R}^d$ et de matrice de variance-covariance Σ. Soit C une matrice réelle de taille (p, d) où $p \in \mathbb{N}^*$. Alors $C \cdot Y$ est un vecteur gaussien tel que :*

$$C \cdot Y \sim \mathcal{N}(C \cdot \mu\,,\, C \cdot \Sigma \cdot C')$$

On en déduit les conséquences suivantes :

- si Y est un vecteur gaussien isotrope de $\mathbb{R}^d$ de variance σ^2 et h un vecteur de $\mathbb{R}^d$, alors $h' \cdot Y$ est une combinaison linéaire des coordonnées de Y tel que :

$$h' \cdot Y \text{ suit la loi } \mathcal{N}(0, \sigma^2 \cdot h' \cdot h) = \mathcal{N}(0, \sigma^2 \cdot \|h\|^2)$$

- si Y est un vecteur gaussien d'espérance μ et de matrice de variance Σ et si h un vecteur de $\mathbb{R}^d$, alors $h' \cdot Y$ est une combinaison linéaire des coordonnées de Y et :

$$h' \cdot Y \text{ suit la loi unidimensionnelle } \mathcal{N}(h' \cdot \mu, h' \cdot \Sigma \cdot h)$$

(Pour une présentation plus détaillée des notions sur les vecteurs gaussiens on peut consulter le livre P. Toulouse, 1999, chap.2)

2.6 Fonctions caractéristiques et génératrices

Définition. *Soit X un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$. La fonction caractéristique de X est la fonction $\phi_X : \mathbb{R}^d \mapsto \mathbb{C}$ telle que*

$$\phi_X(t) = \mathbb{E}\left[\exp(i < t, X >)\right] = \int_{\mathbb{R}^d} e^{i<t,x>} d\mathbb{P}_X(x),$$

où $< . >$ désigne le produit scalaire euclidien sur $\mathbb{R}^d$ tel que $< t, x >= \sum_{i=1}^{d} t_i x_i$ pour $t = (t_1, \cdots, t_d)$ et $x = (x_1, \cdots, x_d)$.

Remarque. *La fonction génératrice existe sur $\mathbb{R}$ et $\phi_X(0) = 1$. ϕ_X est aussi la transformée de Fourier de la mesure $\mathbb{P}_X$.*

Théorème. *Soit X et Y des vecteurs aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$, de lois $\mathbb{P}_X$ et $\mathbb{P}_Y$. Alors $\mathbb{P}_X = \mathbb{P}_Y$ si et seulement si $\phi_X = \phi_Y$.*

Théorème (Théorème d'inversion)**.** *Si X est un vecteur aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$ et si ϕ_X est une fonction intégrable par rapport à la mesure de Lebesgue λ_d sur $\mathbb{R}^d$, alors X admet une densité f_X par rapport à λ_d telle que pour $x \in \mathbb{R}^d$,*

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i<t,x>} \phi_X(t) dt.$$

Proposition. *Si X est une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$ de fonction génératrice ϕ_X. Alors si $\mathbb{E}(|X|^n) < +\infty$ (ou $X \in \mathbb{L}^n(\Omega, \mathcal{A}, \mathbb{P})$), ϕ_X est n fois dérivable et $\phi_X^{(n)}(t) = i^n \mathbb{E}(X^n e^{itX})$.*

Remarque. *Lorsque ces moments existent, on a $i^n \mathbb{E}(X^n) = \phi_X^{(n)}(0)$.*

2.7 Convergence de suites de variables aléatoires

Définition. *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que*

- *(X_n) converge en probabilité vers X, noté $X_n \xrightarrow[n \to +\infty]{\mathcal{P}} X$, lorsque pour tout $\varepsilon > 0$,*

$$\lim_{n \to \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

- (X_n) *converge dans* $\mathbb{L}^p(\Omega, \mathcal{A}, \mathbb{P})$ *vers* X, *noté* $X_n \xrightarrow[n\to+\infty]{\mathbb{L}^p} X$, *avec* $p > 0$, *lorsque*
$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^p = 0.$$
- (X_n) *converge en loi vers* X, *noté* $X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X$, *lorsque,*
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \text{pour tout } x \in \mathbb{R} \text{ tel que } F_X \text{ continue en } x.$$
- (X_n) *converge presque sûrement vers* X, *noté* $X_n \xrightarrow[n\to+\infty]{p.s.} X$, *lorsque pour tout* $\varepsilon > 0$,
$$\lim_{n\to\infty} \mathbb{P}(\sup_{m\geq n} |X_m - X| > \varepsilon) = 0.$$

Propriété. 1. *p.s. et* $\mathbb{L}^p \longrightarrow \mathcal{P} \longrightarrow \mathcal{L}$.

2. *pour* $q \geq p$, $\mathbb{L}^q \longrightarrow \mathbb{L}^p$.
3. *La convergence en loi n'entraîne pas la convergence en probabilité. Mais* $(X_n \xrightarrow[n\to+\infty]{\mathcal{P}} C) \Longleftrightarrow (X_n \xrightarrow[n\to+\infty]{\mathcal{L}} C)$ *pour* C *une constante.*
4. *Si* g *est une fonction borélienne continue alors* $(X_n \xrightarrow[n\to+\infty]{\mathcal{P}} X) \Longrightarrow (g(X_n) \xrightarrow[n\to+\infty]{\mathcal{P}} g(X))$.

Propriété. 1. *Si pour tout* $\varepsilon > 0$, $\sum_{n=0}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon) < +\infty$ *alors* $X_n \xrightarrow[n\to+\infty]{p.s.} X$ *(application du Lemme de Borel-Cantelli).*

2. *Si il existe* $r > 0$ *tel que* $\mathbb{E}(|X_n|^r) < +\infty$ *et* $\sum_{n=0}^{\infty} \mathbb{E}(|X_n - X|^r) < +\infty$ *alors* $X_n \xrightarrow[n\to+\infty]{p.s.} X$.

Théorème (Loi faible des Grands Nombres)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires indépendantes et identiquement distribuées. Alors si* $\mathbb{E}(|X_i|) < +\infty$,
$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to+\infty]{\mathcal{P}} m = \mathbb{E}X_i.$$

Théorème (Loi forte des Grands Nombres)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires indépendantes et identiquement distribuées. Alors si* $\mathbb{E}(|X_i|) < +\infty$,
$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to+\infty]{p.s.} m = \mathbb{E}X_i.$$

Théorème (Théorème de la limite centrale)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires indépendantes et identiquement distribuées. Alors si* $\sigma^2 = \mathbb{E}X_i^2 < +\infty$, *et* $m = \mathbb{E}X_i$,
$$\sqrt{n}\frac{\overline{X}_n - m}{\sigma} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

Théorème (Loi forte des Grands Nombres multidimensionnelle)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de vecteurs aléatoires à valeurs dans* $\mathbb{R}^d$, *indépendants et identiquement distribués. Alors si* $\mathbb{E}(\|X_i\|) < +\infty$ *(pour* $\|.\|$ *une norme sur* $\mathbb{R}^d$*),*
$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to+\infty]{p.s.} m = \mathbb{E}X_i.$$

Théorème (Théorème de la limite centrale multidimensionnel)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de vecteurs aléatoires à valeurs dans* $\mathbb{R}^d$, *indépendants et identiquement distribués. Alors si* Σ *matrice de covariance de chaque* X_i *existe, et* $m = \mathbb{E}X_i$,
$$\sqrt{n}\Big(\overline{X}_n - m\Big) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}_d(0,\Sigma).$$

Théorème (Delta-method)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de vecteurs aléatoires à valeurs dans* $\mathbb{R}^d$, *indépendants et identiquement distribués, telle que* Σ *matrice de covariance de chaque* X_i *existe, et* $m = \mathbb{E}X_i$. *Soit* $g : \mathbb{R}^d \to \mathbb{R}^p$ *une fonction de classe* $\mathcal{C}^1$ *sur un voisinage autour de* m, *de matrice Jacobienne* $J_g(m)$ *en* m. *Alors,*
$$\sqrt{n}\Big(g(\overline{X}_n) - g(m)\Big) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}_d(0, J_g(m)\cdot\Sigma\cdot J_g'(m)).$$

2.8 Espérance conditionnelle

Définition. *Soit Y une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$. Si $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$ et si $Y \in \mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$. Alors on note $\mathbb{E}(Y \,|\, \mathcal{B})$ la projection orthogonale de Y sur $\mathbb{L}^2(\Omega, \mathcal{B}, P)$, appelée espérance conditionnelle de Y sachant $\mathcal{B}$. Ainsi :*

$$\mathbb{E}\,|Y - \mathbb{E}(Y \,|\, \mathcal{B})|^2 = \inf_{Z \in \mathbb{L}^2(\Omega, \mathcal{B}, \mathbb{P})} \left\{ \mathbb{E}\,|Y - Z|^2 \right\}.$$

Par extension, si $Y \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, on définit l'espérance conditionnelle par rapport à $\mathcal{B}$, comme l'unique (p.s.) variable aléatoire, $\mathcal{B}$-mesurable vérifiant p.s. :

$$\int_B \mathbb{E}(Y \,|\, \mathcal{B})\, d\mathbb{P} = \int_B Y \, d\mathbb{P}, \quad \text{pour tout } B \in \mathcal{B}.$$

Définition. *Par convention, si X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ sur $(\Omega, \mathcal{A}, \mathbb{P})$ et si Y une variable aléatoire sur $(\Omega, \mathcal{A}, \mathbb{P})$, on note $\mathbb{E}(Y \,|\, X) = \mathbb{E}(Y \,|\, X^{-1}(\mathcal{B}(\mathbb{R})))$.*

Propriété.

1. *Lemme de Doob : Pour $Y \in \mathbb{L}^1(\Omega, \mathcal{A}, \mathbb{P})$, et X une v.a. de$(\Omega, \mathcal{A}, \mathbb{P})$, alors p.s. $\mathbb{E}(Y \,|\, X) = h(X)$, avec h une fonction borélienne.*
2. *Pour Y_1 et Y_2 deux variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$, et $(a, b, c) \in \mathbb{R}^3$, alors*
$$\mathbb{E}(aY_1 + bY_2 + c \,|\, \mathcal{B}) = a\mathbb{E}(Y_1 \,|\, \mathcal{B}) + b\mathbb{E}(Y_2 \,|\, \mathcal{B}) + c.$$
3. *Si $Y_1 \leq Y_2$, alors $\mathbb{E}(Y_1 \,|\, \mathcal{B}) \leq \mathbb{E}(Y_2 \,|\, \mathcal{B})$.*
4. *Le Lemme de Fatou, les théorèmes de Beppo-Levi, Lebesgue et Jensen s'appliquent avec l'espérance conditionnelle.*
5. *Si $Y \in \mathbb{L}^2(\Omega, \mathcal{B}, \mathbb{P})$, alors $\mathbb{E}(Y \,|\, \mathcal{B}) = Y$; ainsi $\mathbb{E}(g(X) \,|\, X) = g(X)$ pour g une fonction mesurable réelle.*
6. *On a $\mathbb{E}(\mathbb{E}(Y \,|\, \mathcal{B})) = \mathbb{E}Y$.*
7. *Si $Y^{-1}(\mathcal{B}(\mathbb{R}))$ et $\mathcal{B}$ sont indépendantes alors $\mathbb{E}(Y \,|\, \mathcal{B}) = \mathbb{E}Y$; ainsi, si X et Y sont indépendantes, $\mathbb{E}(Y \,|\, X) = \mathbb{E}Y$.*
8. *Si (X, Y) est un couple de v.a. à valeurs dans $\mathbb{R}^2$ possédant une densité $f_{(X,Y)}$ par rapport à la mesure de Lebesgue, alors si X est intégrable ,*
$$\mathbb{E}(Y \mid X = x) = \frac{\int_{\mathbb{R}} y \cdot f_{(X,Y)}(x, y)\, dy}{\int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dy}, \quad \text{pour tout } x \text{ tel que } \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dy > 0.$$

Proposition. *Si $(Y, X_1, \cdots, X_n)$ est un vecteur gaussien, alors $\mathbb{E}(Y \,|\, (X_1, \cdots, X_n)) = a_0 + a_1 X_1 + \cdots + a_n X_n$, où les a_i sont des réels.*

3 Estimation paramétrique

3.1 Définitions

Dans toute la suite, on se place sur $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. On considère $(X_n)_{n\mathbb{N}}$ une suite de variable aléatoire, où chaque X_i est définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et est à valeur dans $\Omega' \subset \mathbb{R}$.

Définition.

- *On appelle modèle statistique de dimension n un espace $((\Omega')^n, \mathcal{A}'_n, \mu)$, où $\mathcal{A}'_n$ est une tribu sur $(\Omega')^n$ et μ une mesure de probabilité sur $((\Omega')^n, \mathcal{A}'_n)$.*
- *On appelle échantillon de taille n du modèle statistique $((\Omega')^n, \mathcal{A}'_n, \mu)$ le vecteur aléatoire $(X_1, \ldots, X_n)$ distribuée selon la loi μ. Pour $\omega \in \Omega$, $(X_1(\omega), \ldots, X_n(\omega))$ vecteur de $\mathbb{R}^n$ est appelé échantillon observé. C'est à partir et sur ce vecteur que le travail statistique s'effectue (en général).*

Définition. *On appelle :*

- *Modèle statistique paramétrique, une famille de modèle de la forme :* $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$, *où* $\Theta \subset \mathbb{R}^p$.
- *Modèle statistique semi-paramétrique, une famille de modèle de la forme :* $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_{(\theta,f)}, \theta \in \Theta, f \in \mathcal{F})$, *où* $\Theta \subset \mathbb{R}^p$ *et* $\mathcal{F}$ *n'est pas de dimension finie.*
- *Modèle statistique non-paramétrique, une famille de modèle de la forme :* $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_f, f \in \mathcal{F})$, *où* $\mathcal{F}$ *n'est pas de dimension finie.*

Définition.
- *On dit que le modèle paramétrique :* $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$, *où* $\Theta \subset \mathbb{R}^p$, *est dominé par une mesure* μ *lorsque* $\mathbb{P}_\theta$ *est absolument continue par rapport à* μ *pour tout* $\theta \in \Theta$.
- *On se place dans le cadre d'un modèle paramétrique* $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$, *où* $\Theta \subset \mathbb{R}^p$, *dominé par une mesure* μ. *Pour* $(x_1, \cdots, x_n) \in (\Omega')^n$, *la fonction* $\theta \in \Theta \mapsto L_\theta(x_1, \cdots, x_n) = \frac{d\mathbb{P}_\theta}{d\mu}(x_1, \cdots, x_n)$ *est appelée une vraisemblance du modèle statistique.*

Exemple.
- *Dans le cas où* μ *est la mesure de Lebesgue sur* $\mathbb{R}^n$, *la vraisemblance sera la densité (classique) en* $(x_1, \cdots, x_n)$.
- *Dans le cas où* μ *est comptage sur* $\mathbb{N}^n$, *la vraisemblance sera la probabilité en* $(x_1, \cdots, x_n)$.
- *Attention ! si le support de* $\mathbb{P}_\theta$ *dépend de* θ, *la mesure qui domine (ainsi que* Ω' *et* $\mathcal{A}'_n$*) ne peut dépendre de* θ *: il ne faut pas oublier de le préciser dans l'expression de la vraisemblance.*

Définition. *Lorsque l'on dispose d'un échantillon* $(X_1, \ldots, X_n)$ *du modèle statistique* $((\Omega')^n, \mathcal{A}'_n, \mu)$, *une statistique* $\widehat{T}_n$ *est une application mesurable de* $((\Omega')^n, \mathcal{A}'_n)$ *dans* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, *donc un vecteur aléatoire défini sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *à valeur dans* $\mathbb{R}^d$, *et telle que :*

$$\widehat{T}_n = h(X_1, \ldots, X_n), \quad \text{où} \quad h : (\Omega')^n \mapsto \mathbb{R}^d \quad \text{est mesurable.}$$

Exemple. *Estimateur du paramètre d'une loi de Bernoulli.*
Estimateur de l'espérance et de la variance par la moyenne et la variance empirique.
Estimateurs du paramètre θ *d'un n-échantillon* $(X_1, \cdots, X_n)$ *de loi uniforme sur* $[0, \theta]$.
Test sur la moyenne.

3.2 Statistiques exhaustives

On se place désormais dans le cadre d'une modèle statistique paramétrique $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$, où $\Theta \subset \mathbb{R}^p$, dominé par une mesure μ.

Exemple. *1. Soit le modèle statistique paramétrique* $\Big([0, \infty[^n, \mathcal{B}([0, \infty[^n), \mathcal{U}([0, \theta])^{\otimes n}, \theta \in]0, +\infty[\Big)$. *On dispose donc d'un n-échantillon* $(X_1, \cdots, X_n)$ *de v.a.i.i.d. suivant une loi uniforme sur* $[0, \theta]$. *Si on considère* $\max\{X_1 \ldots, X_n\}$ *cela semble suffire pour posséder toute l'information sur* θ *que contenait* $(X_1, \ldots, X_n)$ *: on a donc résumé l'"information" sur* θ *contenait* $(X_1, \ldots, X_n)$, *un vecteur de taille n, par une statistique de taille* 1.
2. De même, si on considère le modèle statistique paramétrique $\Big(\{0,1\}^n, \mathcal{P}(\{0,1\}^n), \mathcal{B}(p)^{\otimes n}, p \in [0,1]\Big)$ *(on dispose donc d'un n-échantillon* $(X_1, \cdots, X_n)$ *de v.a.i.i.d. suivant une loi de Bernoulli de paramètre p) alors la statistique* $X_1 + \cdots + X_n$ *contient toute l'"information" sur p contenue dans l'échantillon* $(X_1, \ldots, X_n)$.

Comment exprimer formellement ce fait qu'une statistique puisse résumer à elle seule toute l'information sur le paramètre ?

Définition. *Soit* $\widehat{T}$ *une statistique du modèle statistique paramétrique dominé à valeurs dans* $\mathbb{R}^d$. *On dit que* $\widehat{T}$ *est une statistique exhaustive si pour toute statistique* S *intégrable (donc dans* $\mathbb{L}^1((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta))$ *alors* $\mathbb{E}_\theta(S \mid \widehat{T})$ *ne dépend (*$\mathbb{P}_\theta$*-presque sûrement) pas de* θ.

Théorème (Théorème de factorisation de Neyman)**.** *Soit* $(X_1, \cdots, X_n)$ *un n-échantillon et soit* $\widehat{T}$ *une statistique du modèle statistique paramétrique dominé avec* $\widehat{T}$ *à valeurs dans* $\mathbb{R}^d$, *où* $d \in \mathbb{N}^*$. *La statistique* $\widehat{T}$ *est exhaustive si et seulement s'il existe une fonction* $h : \mathbb{R}^n \to \mathbb{R}_+$ *et une fonction* $g_\theta(.) : \mathbb{R}^d \mapsto \mathbb{R}_+$, *telle que l'on puisse écrire pour tout* $(x_1, \ldots, x_n) \in (\Omega')^n$ *:*

$$L_\theta(x_1, \ldots, x_n) = g_\theta(\widehat{T}(x_1, \ldots, x_n)) \cdot h(x_1, \ldots, x_n) \quad \text{pour tout } \ \theta \in \Theta.$$

Lemme. *Soit le modèle statistique paramétrique $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$, où $\Theta \subset \mathbb{R}^p$. Alors ce modèle est dominé si et seulement si il existe une sous-famille dénombrable $(\mathbb{P}_{\theta_i})_{i\in\mathbb{N}}$ telle que pour tout $A \in \mathcal{A}$, $\forall i \in \mathbb{N}$, $\mathbb{P}_{\theta_i}(A) = 0$ entraîne $\forall\theta$, $\mathbb{P}_\theta(A) = 0$. Toute mesure de probabilité de la forme $\mathbb{P}^* = \sum_{i\in\mathbb{N}} a_i \cdot \mathbb{P}_{\theta_i}$ avec $c_i > 0$ pour tout $i \in \mathbb{N}$ et $\sum_{i\in\mathbb{N}} c_i = 1$ domine le modèle.*

<u>*Démonstration du lemme*</u> *:* $\Longleftarrow$ Il est bien clair que si une telle mesure P^* existe, le modèle est dominé.
$\Longrightarrow$ Montrons maintenant que si le modèle est dominé par une mesure μ alors la famille $(P_{\theta_i})_{i\in\mathbb{N}}$ existe. En premier lieu, si μ est une mesure non finie mais σ-finie (par exemple la mesure de Lebesgue), alors μ_P définie par $\mu_P(A) = \displaystyle\sum_{i=1}^{\infty} \frac{1}{2^i}\frac{\mu(A\cap A_i)}{\mu(A_i)}$ pour tout $A \in \mathcal{A}$, est une mesure de probabilité équivalente à μ (avec $(A_i)_{i\in\mathbb{N}^*}$ une partition de $(\Omega')^n$ telle que $0 < \mu(A_i) < \infty$ pour tout $i \in \mathbb{N}^*$). On travaille donc désormais avec μ_P.
Pour $\theta \in \Theta$, soit B_θ le sous-ensemble de $(\Omega')^n \subset \mathbb{R}^n$ qui est le support de la densité de $\mathbb{P}_\theta$ par rapport à μ. Soit

$$\mathcal{C} = \Big\{ \bigcup_{i\in I} B_{\theta_i},\ I \subset \mathbb{N}, \theta_i \in \Theta \Big\},$$

l'ensemble de toutes les unions dénombrables d'ensembles B_θ. On note $M = \sup_{C\in\mathcal{C}} \mu_P(C)$. Soit $(C_n)_{n\in\mathbb{N}}$ une suite d'ensembles de $\mathcal{C}$ telle que la suite $(\mu_P(C_n))_n$ converge vers M (une telle suite existe forcément sinon M ne serait pas le supremum). Remarquons que chaque C_i étant une union dénombrable de B_{θ_k}, alors une suite (θ_n) de θ suffit pour engendrer la suite $(C_n)_{n\in\mathbb{N}}$. Si on pose :

$$D = \bigcup_{n\in\mathbb{N}} C_n = \bigcup_{k\in\mathbb{N}} B_{\theta_k},$$

alors $M = \mu_P(D)$ et pour tout $\theta \in \Theta$, $B_\theta \cup D \in \mathcal{C}$ et :

$$\mu_P(B_\theta \cup D) \le M \le \mu_P(B_\theta \cup D) = \mu_P(B_\theta \cap D^c) + \mu_P(D)$$

Donc pour tout $\theta \in \Theta$, $\mu_P(B_\theta \cap D^c) = 0$ soit $\forall\theta \in \Theta$, $\mathbb{P}_\theta(B_\theta \cap D^c) = 0$ puisque $\mathbb{P}_\theta << \mu_P$. En conséquence, pour tout $A \in \mathcal{A}'_n$, $A \subset B_\theta \cup B_\theta^c = (\Omega')^n$, soit :

$$\mathbb{P}_\theta(A \cap D^c) = 0, \quad \text{car par définition des } B_\theta,\ \mathbb{P}_\theta(B_\theta^c) = 0.$$

Si on suppose maintenant que $A \in \mathcal{A}'_n$ est tel que $\mathbb{P}_{\theta_k}(A) = 0$, avec la suite (θ_k) précédemment définie, alors $\mu_P(A \cap B_{\theta_k}) = 0$ par définition des B_θ et donc $\mu_P(A \cap D) = 0$ (par la propriété de σ-additivité d'une mesure). Comme $\mathbb{P}_\theta << \mu_P$, on en déduit que $\forall\theta \in \Theta$, $\mathbb{P}_\theta(A \cap D) = 0$ et donc $\mathbb{P}_\theta(A) = \mathbb{P}_\theta(A \cap D) + = \mathbb{P}_\theta(A \cap D^c) = 0$. Ainsi, $\mathbb{P}^*$ domine bien $\mathbb{P}_\theta$ pour tout $\theta \in \Theta$. ■

<u>*Démonstration du Théorème de factorisation de Neyman*</u> *:* Soit $\mathbb{P}^* = \sum_{i\in\mathbb{N}} a_i \cdot \mathbb{P}_{\theta_i}$ une mesure de probabilité dominante construite comme dans le lemme.
$\Longleftarrow$ Si $g_\theta(\widehat{T}(x)) \cdot h(x)$ avec $x \in (\Omega')^n$ est la densité de $\mathbb{P}_\theta$ par rapport à μ, alors $\sum_{i\in\mathbb{N}} a_i \cdot g_{\theta_i}(\widehat{T}(x)) \cdot h(x) = g_*(\widehat{T}(x)) \cdot h(x)$ est une densité de P^* par rapport à μ. Alors, comme $g_*(\widehat{T}(x)) \cdot h(x) > 0$ P^*-p.s., donc $\mathbb{P}_\theta$-p.s., pour toute variable aléatoire S intégrable, pour tout $\theta \in \Theta$:

$$\begin{aligned}
\mathbb{E}_\theta(S \cdot \mathbb{I}_B) &= \int_B S\, d\mathbb{P}_\theta, \quad \text{pour tout } B \in \sigma(\widehat{T}), \quad \text{tribu engendrée par } \widehat{T} \\
&= \int_B S(x) \cdot g_\theta(\widehat{T}(x)) \cdot h(x)\, d\mu(x) \\
&= \int_B S(x) \cdot \frac{g_\theta(\widehat{T}(x)) \cdot h(x)}{g_*(\widehat{T}(x)) \cdot h(x)}\, d\mathbb{P}^*(x) \\
&= \mathbb{E}_*\Big(\mathbb{I}_B \cdot \frac{g_\theta(\widehat{T})}{g_*(\widehat{T})} \cdot S\Big) \\
&= \mathbb{E}_*\Big(\mathbb{I}_B \cdot \frac{g_\theta(\widehat{T})}{g_*(\widehat{T})} \cdot \mathbb{E}_*(S \mid \widehat{T})\Big) \quad \text{(d'après la définition de l'espérance conditionnelle)} \\
&= \mathbb{E}_\theta\Big(\mathbb{I}_B \cdot \mathbb{E}_*(S \mid \widehat{T})\Big).
\end{aligned}$$

En conséquence, d'après la définition de l'espérance conditionnelle dans $\mathbb{L}^1((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta)$, on a $\mathbb{P}_\theta$-p.s., $\mathbb{E}_*(S \mid \widehat{T}) = \mathbb{E}_\theta(S \mid \widehat{T})$: la statistique $\widehat{T}$ est bien exhaustive.
$\Longrightarrow$ On suppose que $\widehat{T}$ est une statistique exhaustive pour le modèle. Donc pour toute statistique intégrable S, $\forall \theta$, $\mathbb{E}_*(S \mid \widehat{T}) = \mathbb{E}_\theta(S \mid \widehat{T})$. En conséquence, si on note $\phi(x,\theta) = \dfrac{d\mathbb{P}_\theta}{d\mathbb{P}_*}(x)$ la densité de $\mathbb{P}_\theta$ par rapport à $\mathbb{P}_*$,

$$\begin{aligned}
\mathbb{E}_\theta(S) &= \mathbb{E}_\theta\big(\mathbb{E}_*(S \mid \widehat{T})\big), \quad \text{(car } \widehat{T} \text{ est exhaustive et d'après les propriétés de l'espérance conditionnelle)}\\
&= \mathbb{E}_*\big(\phi(X,\theta) \cdot \mathbb{E}_\mu(S \mid \widehat{T})\big), \quad \text{où } X \sim \mathbb{P}_*\\
&= \mathbb{E}_*\Big[\mathbb{E}_*\Big(\phi(X,\theta) \cdot \mathbb{E}_*(S \mid \widehat{T}) \mid \widehat{T}\Big)\Big], \quad \text{(d'après les propriétés de l'espérance conditionnelle)}\\
&= \mathbb{E}_*\Big[\mathbb{E}_*\Big(\phi(X,\theta) \mid \widehat{T}\Big) \cdot \mathbb{E}_*\Big(S \mid \widehat{T}\Big)\Big], \quad \text{(car } \mathbb{E}_*\Big(S \mid \widehat{T}\Big) \text{ est une fonction de } \widehat{T}\text{)}\\
&= \mathbb{E}_*\Big[\mathbb{E}_*\Big(S \cdot \mathbb{E}_*\big(\phi(X,\theta) \mid \widehat{T}\big) \mid \widehat{T}\Big)\Big]\\
&= \mathbb{E}_*\Big[S \cdot \mathbb{E}_*\big(\phi(X,\theta) \mid \widehat{T}\big)\Big]
\end{aligned}$$

Ainsi, la variable aléatoire $\mathbb{E}_*\big(\phi(X,\theta) \mid \widehat{T}\big)$, qui est une fonction de $\widehat{T}$ (qui est elle-même une fonction sur $(\Omega')^n$), est la densité de $\mathbb{P}_\theta$ par rapport à $\mathbb{P}_*$. Par suite, la vraisemblance, qui est la densité de $\mathbb{P}_\theta$ par rapport à μ, s'écrit :

$$L_\theta(x_1,\ldots,x_n) = \frac{d\mathbb{P}_\theta}{d\mu}(x_1,\ldots,x_n) = \frac{d\mathbb{P}_\theta}{d\mathbb{P}_*}(x_1,\ldots,x_n) \cdot \frac{d\mathbb{P}_*}{d\mu}(x_1,\ldots,x_n) = \mathbb{E}_*\big(\phi(X,\theta) \mid \widehat{T}\big) \cdot h(x_1,\ldots,x_n),$$

avec h une fonction mesurable. ■

Exemple. *Différentes statistiques exhaustives pour les modèles paramétriques de loi uniforme, de loi de Bernoulli, de loi gaussienne...*

Propriété. *On se place dans le cadre d'un modèle paramétrique dominé.*

1. *La statistique $\widehat{T} = (X_1,\ldots,X_n)$ est exhaustive.*
2. *Si $\widehat{T}$ est une statistique exhaustive et s'il existe une fonction borélienne h telle qu'une autre statistique $\widehat{U}$ vérifie $\widehat{T} = h(\widehat{U})$, alors $\widehat{U}$ est également exhaustive.*

On vient de voir que l'on peut toujours trouver une statistique exhaustive (l'échantillon lui-même par exemple). Comme on aurait plutôt tendance à vouloir le "maximum d'information" dans une statistique exhaustive, lorsque le paramètre est dans $\mathbb{R}^d$, on aimerait savoir quelle dimension minimale peut avoir cette statistique. En particulier, si $d = 1$, peut-on toujours trouver une statistique exhaustive de taille 1 ? L'exemple suivant montre que ce n'est pas toujours le cas :

Exemple. *Soit le modèle statistique $([0,\infty[^n, \mathcal{B}([0,\infty[^n), (\mathbb{P}_\theta)^{\otimes n}, \theta \in \mathbb{R}_+)$, où la densité de $\mathbb{P}_\theta$ par rapport à la mesure de Lebesgue est : $f_\theta(x) = \theta(e^{\theta^2} - 1) \cdot e^{-\theta \cdot x} \cdot \mathbb{I}_{x \in [0,\theta]}$. Alors les statistiques $\widehat{T}_1 = \max(X_1,\ldots,X_n)$ et $\widehat{T}_2 = X_1 + \ldots + X_n$ ne sont pas chacune exhaustive alors que $\widehat{T} = (\widehat{T}_1, \widehat{T}_2)$ est exhaustive. On pourra même montrer que cette statistique est de taille minimale...*

Définition. *Une statistique exhaustive $\widehat{T}$ du modèle statistique paramétrique dominé avec $\widehat{T}$ est dite minimale si pour toute autre statistique exhaustive $\widehat{U}$ est telle qu'il existe une fonction borélienne h vérifiant :*

$$\widehat{T} = h(\widehat{U}).$$

Proposition. *Soit un modèle statistique paramétrique dominé et soit $L_\theta(x_1,\ldots,x_n)$ sa vraisemblance. Alors $\widehat{T}$ est une statistique exhaustive minimale pour ce modèle lorsque $\forall (x_1,\ldots,x_n) \in (\Omega')^n$ et $\forall (y_1,\ldots,y_n) \in (\Omega')^n$,*

$$\left(\theta \mapsto \frac{L_\theta(x_1,\ldots,x_n)}{L_\theta(y_1,\ldots,y_n)} \quad \textit{ne dépend pas de } \theta\right) \Longleftrightarrow \Big(\widehat{T}(x_1,\ldots,x_n) = \widehat{T}(y_1,\ldots,y_n).\Big) \qquad (2)$$

<u>*Démonstration de la proposition*</u> : On suppose que (2) est vraie et on suppose (sans perte de généralité) que la vraisemblance est strictement positive. Soit $t \in \mathrm{Im}(\widehat{T}((\Omega')^n))$. Notons $x^{(t)} \in \widehat{T}^{-1}(\{t\}) \subset (\Omega')^n$. Alors $\forall x \in \widehat{T}^{-1}(\{t\})$, $\widehat{T}(x) = \widehat{T}(x^{(\widehat{T}(x))})$ et donc d'après (2),

$$h(x) = \frac{L_\theta(x)}{L_\theta(x^{(\widehat{T}(x))})} \quad \text{est indépendant de } \theta.$$

Posons $g_\theta(t) = L_\theta(x^{(t)})$. Alors $L_\theta(x) = g_\theta(\widehat{T}(x)) \cdot h(x)$. Comme ceci est vrai pour tout $x \in (\Omega')^n$, la statistique $\widehat{T}$ est bien exhaustive.
Supposons maintenant que $\widehat{S}$ est une autre statistique exhaustive. Alors par le théorème de factorisation de Neyman, il existe deux fonctions $g_\theta^{(s)}$ et $h^{(s)}$ (ne dépendant pas de θ) telles que pour tout $x \in (\Omega')^n$, $L_\theta(x) = g_\theta^{(s)}(\widehat{S}(x)) \cdot h^{(s)}(x)$. Ainsi pour tout $x \in (\Omega')^n$ et $y \in (\Omega')^n$ tels que $\widehat{S}(x) = \widehat{S}(y)$, alors :

$$\frac{L_\theta(x)}{L_\theta(y)} = \frac{g_\theta^{(s)}(\widehat{S}(x)) \cdot h^{(s)}(x)}{g_\theta^{(s)}(\widehat{S}(y)) \cdot h^{(s)}(y)} = \frac{h^{(s)}(x)}{h^{(s)}(y)}, \quad \text{qui est indépendant de } \theta.$$

Mais d'après (2) ceci n'est possible que si $\widehat{T}(x) = \widehat{T}(y)$. Donc $\widehat{T}$ est une fonction de $\widehat{S}$ et la statistique $\widehat{T}$ est donc minimale. ■.

Qu'elle serait une sorte d'opposée de la notion de statistique exhaustive minimale ? Ce devrait être une statistique ne dépendant pas du paramètre, soit :

Définition. *Une statistique $\widehat{T}$ d'un modèle paramétrique est dite libre si sa loi ne dépend pas du paramètre.*

Or, de façon assez surprenante il peut arriver qu'une statistique exhaustive minimale comprenne une statistique libre, qui intuitivement ne devrait pas être prise en compte pour donner toute l'information sur θ (soit par exemple la loi $\mathbb{P}_\theta$ discrète et équidistribuée sur $\{\theta-1, \theta, \theta+1\}$; pour un échantillon de taille 2, la statistique $(X_{(2)} - X_{(1)}, X_1 + X_2)$ est exhaustive minimale, mais $X_{(2)} - X_{(1)}$ est libre). Aussi peut-on rajouter une autre caractérisation des statistiques exhaustives pour pouvoir atteindre une forme d'optimalité pour ces statistiques, qui serait qu'aucune fonctionnelle non constante de la statistique ne peut être libre. Cela peut également se traduire de la façon suivante :

Définition. *Une statistique exhaustive $\widehat{T}$ du modèle statistique paramétrique dominé avec $\widehat{T}$ à valeur dans $\mathbb{R}^d$ est dite complète si pour toute fonction borélienne $h : \mathbb{R}^d \to \mathbb{R}$ telle que $h(\widehat{T})$ soit intégrable, alors :*

$$\forall \theta \in \Theta, \quad \mathbb{E}_\theta\big(h(\widehat{T})\big) = 0 \quad \Longrightarrow \quad h(\widehat{T}) = 0.$$

Propriété. *Soit un modèle statistique paramétrique dominé.*

1. *si $\widehat{T}$ est une statistique exhaustive complète alors pour toute fonction borélienne h bijective $h(\widehat{T})$ est une statistique exhaustive complète.*
2. *si $\widehat{T}$ est une statistique exhaustive complète alors $\widehat{T}$ est une statistique exhaustive minimale.*
3. *(Théorème de Basu) si $\widehat{T}$ est une statistique exhaustive complète alors $\widehat{T}$ est indépendante de toute statistique libre sur le modèle.*

<u>*Démonstration de la propriété*</u> : 3. Théorème de Basu. Soit $\widehat{S}$ une statistique libre pour le modèle et soit f une fonction telle que $\mathbb{E}_\theta(f(\widehat{S}))$ existe. Comme $\widehat{S}$ est libre, on peut noter $e(f) = \mathbb{E}_\theta(f(\widehat{S}))$ une application linéaire ne dépendant pas de θ. Par suite, la statistique $\mathbb{E}_\mu(f(\widehat{S}) \mid \widehat{T}) - e(f)$ est une fonction de $\widehat{T}$ mesurable telle que $\mathbb{E}_\theta\Big(\mathbb{E}_\mu(f(\widehat{S}) \mid \widehat{T}) - e(f)\Big) = 0$ pour tout $\theta \in \Theta$. Comme on a supposé que $\widehat{T}$ est exhaustive complète, alors $\mathbb{E}_\mu(f(\widehat{S}) \mid \widehat{T}) = e(f)$ presque-sûrement : les statistiques $\widehat{S}$ et $\widehat{T}$ sont indépendantes. ■

Définition. *On suppose un modèle paramétrique $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta \subset \mathbb{R}^p)$ dominé par une mesure μ. Si, pour tout $(x_1, \cdots, x_n) \in (\Omega')^n$ et $\theta \in \Theta$, la vraisemblance de ce modèle par rapport à μ peut s'écrire sous la forme :*

$$L_\theta(x_1, \cdots, x_n) = \exp\left(\beta(\theta) + b(x_1, \cdots, x_n) + \sum_{j=1}^{p} a_j(x_1, \cdots, x_n) \cdot \alpha_j(\theta)\right), \tag{3}$$

avec les fonctions $a_j : (\Omega')^n \to \mathbb{R}$, $b : (\Omega')^n \to \mathbb{R}$, $\alpha_j : \Theta \subset \mathbb{R}^p \to \mathbb{R}$, et $\beta : \Theta \to \mathbb{R}$, alors on dit que le modèle est exponentiel (ou qu'il appartient à la famille exponentielle).

Exemple. *Appartiennent à la famille exponentielle les lois :*

- *Loi discrètes : Lois de Bernoulli, binomiales, de Poisson,...*
- *Loi "continues" : Lois normales, exponentielles, gamma, du chi-deux,...*

Remarque. *Si $(X_1, \cdots, X_n)$ est un n-échantillon d'un modèle exponentiel (avec θ fixé) alors l'ensemble des valeurs prises par $(X_1, \cdots, X_n)$ ne dépend pas du paramètre θ.*

Propriété. *Soit un modèle exponentiel. Si pour tout $\theta \in \Theta$ on note $\alpha(\theta) = (\alpha_1(\theta), \ldots, \alpha_p(\theta))$ et si l'ensemble $\alpha(\Theta)$ est d'intérieur non vide, alors $\widehat{T}(x_1, \cdots, x_n) = (a_1(x_1, \cdots, x_n), \ldots, a_p(x_1, \cdots, x_n))$ est une statistique exhaustive minimale et complète.*

Démonstration de la propriété : Soit $g : \mathbb{R}^p \to \mathbb{R}$ telle que $\mathbb{E}_\theta(g(\widehat{T})) = 0$. Or, $\forall \theta \in \Theta$,

$$\mathbb{E}_\theta(g(\widehat{T})) = \int_{(\Omega')^n} g(\widehat{T}(x)) \cdot \exp\big(\beta(\theta) + b(x) + < \widehat{T}(x), \alpha(\theta) > \big)\, d\mu(x),$$

où $< .,. >$ désigne le produit scalaire. En considérant la mesure ν de densité $\exp(b(x))$ par rapport à μ, on obtient :

$$\begin{aligned} \mathbb{E}_\theta(g(\widehat{T})) = 0 \quad &\Longrightarrow \quad \int_{(\Omega')^n} g(\widehat{T}(x)) \cdot \exp(< \widehat{T}(x), \alpha(\theta) >)\, d\nu(x) = 0 \\ &\Longrightarrow \quad \int_{\widehat{T}((\Omega')^n)} g(y) \cdot \exp(< y, \alpha(\theta) >)\, d\nu_{\widehat{T}}(y) = 0 \end{aligned}$$

pour tout $\theta \in \Theta$, en ayant noté $\nu_{\widehat{T}}$ la mesure image de ν par $\widehat{T}$ et avec $\widehat{T}((\Omega')^n) \in \mathbb{R}^p$. Si on note g^+ et g^- les parties positives et négatives de g (donc $g = g^+ - g^-$), et π^+ et π^- les mesures de densités g^+ et g^- par rapport à $\nu_{\widehat{T}}$, alors, pour tout $\theta \in \Theta$:

$$\int_{\widehat{T}((\Omega')^n)} \exp(< y, \alpha(\theta) >)\, d\pi^+(y) = \int_{\widehat{T}((\Omega')^n)} \exp(< y, \alpha(\theta) >)\, d\pi^-(y).$$

En conséquence sur Θ, donc sur une partie d'intérieure non vide, les mesures π^+ et π^- ont des transformées de Laplace égales : ces deux mesures sont donc égales et donc $g^+ = g^-$ $\nu_{\widehat{T}}$-presque partout (ce qui revient à $g = 0$). A partir des expressions des différentes mesures, on montre que $g = 0$, $\widehat{T}(\mathbb{P}_\theta)$-presque partout. ■

3.3 Information de Fisher

Pour mesurer l'information fournit par un modèle paramétrique dominé (ou une statistique sur ce modèle) au sujet d'un paramètre, une idée naturelle serait de mesurer comment varie localement la mesure de probabilité, ou encore sa vraisemblance. Les fluctuations moyennes de cette vraisemblance serait donc un bon indicateur : pour ce faire on considérera, lorsqu'il existe $\mathrm{grad}_\theta(L_\theta(X_1, \cdots, X_n))$, et on s'intéressera à la matrice de covariance de $\mathrm{grad}_\theta(L_\theta(X_1, \cdots, X_n))$, dont on peut montrer qu'elle ne dépend pas du choix de la mesure dominante choisie. Précisons d'abord la notion de modèle régulier qui nous permettra de définir cette quantité d'information.

Définition. *Dans le cadre d'un modèle statistique paramétrique $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$, où $\Theta \subset \mathbb{R}^p$, dominé par une mesure μ, on dira que ce modèle est régulier lorsque :*

1. *Θ est un ouvert de $\mathbb{R}^d$;*
2. *la vraisemblance $L_\theta(.)$ vérifie $\forall (x_1, \ldots, x_n) \in (\Omega')^n$, $\forall \theta \in \Theta$, $L_\theta(x_1, \ldots, x_n) > 0$;*
3. *$\forall (x_1, \ldots, x_n) \in (\Omega')^n$, la fonction $\theta \in \Theta \mapsto \log(L_\theta(.))$ est différentiable sur Θ par rapport à θ, et son gradient appartient à $\mathbb{L}^2((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta)$ $\forall \theta \in \Theta$;*

4. $\forall\theta \in \Theta$, *pour toute fonction* $h : \mathbb{R}^n \to \mathbb{R}$ *appartenant à* $\mathbb{L}^1((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta)$, *alors :*

$$\frac{\partial}{\partial\theta}\int_{(\Omega')^n} h(x)\cdot L_\theta(x)\,d\mu(x) = \int_{(\Omega')^n} h(x)\cdot \frac{\partial}{\partial\theta}L_\theta(x)\,d\mu(x). \tag{4}$$

Conséquence. *Pour un modèle régulier,* $\mathbb{E}_\theta\Big(grad_\theta(\log L_\theta(.))\Big) = 0$.

<u>*Démonstration*</u> *:* On a $\mathbb{E}_\mu(L_\theta(.)) = 1$ donc $\mathbb{E}_\mu(\mathrm{grad}_\theta L_\theta(.)) = 0$. Par conséquent, $\mathbb{E}_\mu\Big(\dfrac{\mathrm{grad}_\theta(L_\theta(.))}{L_\theta(.)}\Big) = 0$, soit $\mathbb{E}_\theta\big(\mathrm{grad}_\theta(\log L_\theta(.))\big) = 0$. ■

Définition. *Pour un modèle statistique paramétrique dominé régulier, on appelle information de Fisher, la matrice :*

$$I_n(\theta) = \left[\mathbb{E}_\theta\left(\frac{\partial(\log L_\theta(X_1,\cdots,X_N))}{\partial\theta_i}\times\frac{\partial(\log L_\theta(X_1,\cdots,X_N))}{\partial\theta_j}\right)\right]_{1\le i,j\le p}.$$

Propriété. *Pour un modèle statistique paramétrique dominé régulier, et si* $\forall(x_1,\ldots,x_n)\in(\Omega')^n$, *la fonction* $\theta\in\Theta\mapsto\log(L_\theta(.))$ *est* $\mathcal{C}^2(\Theta)$, *alors :*

$$I_n(\theta) = -\left[\mathbb{E}_\theta\left(\frac{\partial^2(\log L_\theta(X_1,\cdots,X_N))}{\partial\theta_i\cdot\partial\theta_j}\right)\right]_{1\le i,j\le p}.$$

Définition. *L'information de Fisher* $I_n^{\widehat{T}}(\theta)$ *associée à une statistique* $\widehat{T}$, *si elle existe, est la matrice de Fisher de la vraisemblance de* $\widehat{T}$ *(déterminée à partir de la vraisemblance de* $\widehat{T}$*).*

Propriété. *Pour un modèle régulier,* $\widehat{T}$ *est une statistique libre si et seulement si* $I_n^{\widehat{T}}(\theta) = 0$.

<u>*Démonstration*</u> *:* $\Longrightarrow$ Si $\widehat{T}$ est libre alors sa loi ne dépend pas de θ donc le gradient du logarithme de sa vraisemblance est nul; l'information de Fisher associée à $\widehat{T}$ est nulle.
$\Longleftarrow$ Si $I_n^{\widehat{T}}(\theta) = 0$, donc la statistique $\mathrm{grad}_\theta(\log L_\theta^{\widehat{T}}(\widehat{T})))$ est centrée et de matrice de covariance nulle. Ainsi, pour tout $\theta\in\Theta$, il existe un ensemble N_θ de mesure 1 pour la mesure de probabilité associée à $\widehat{T}$ (donc, d'après la première hypothèse d'un modèle régulier, tel que $\mu(N_\theta) = 1$)) et tel que pour tout $t\in N_\theta$, $\mathrm{grad}_\theta(\log L_\theta^{\widehat{T}}(t))) = 0$. Pour montrer que $\mathrm{grad}_\theta(\log L_\theta^{\widehat{T}}(t))) = 0$ est bien une variable aléatoire nulle μ-p.s., et donc que $\log L_\theta^{\widehat{T}}(.)$ est une fonction constante en θ, il nous faut montrer que finalement les θ ne dépendent pas de θ. Soit $\Theta^{(d)} = \{\theta_i^{(d)}\}_{i\in\mathbb{N}}$ un sous-ensemble dénombrable de Θ, dense dans Θ. Comme $\Theta^{(d)}$ est dénombrable, il est clair que $N = \bigcup_{i\in\mathbb{N}} N_{\theta_i^{(d)}}$ est tel que $\mu(N) = 1$. De plus, pour tout $\theta\in\Theta$, il existe une sous-suite $(\theta_{\phi(n)}^{(d)})_n$ de $\Theta^{(d)}$ convergeant vers θ et telle que pour tout $t\in N$, pour tout $n\in\mathbb{N}$, $\mathrm{grad}_{\theta_{\phi(n)}^{(d)}}(\log L_{\theta_{\phi(n)}^{(d)}}^{\widehat{T}}(t))) = 0$. Comme une telle fonction de $\theta_{\phi(n)}^{(d)}$ est continue, cette propriété passe à la limite, et donc pour tout $t\in N$, $\forall\theta\in\Theta$, $\mathrm{grad}_\theta(\log L_\theta^{\widehat{T}}(t))) = 0$. Comme N ne dépend pas de θ, alors la fonction $\theta\to\log L_\theta^{\widehat{T}}(.)$ est une constante ne dépendant pas de θ, μ-p.s. : la statistique $\widehat{T}$ est bien libre. ■

Propriété. *Pour un modèle régulier, si* $\widehat{T}$ *est une statistique exhaustive :* $I_n^{\widehat{T}}(\theta) = I_n(\theta)$ *pour tout* $\theta\in\Theta$.

<u>*Démonstration*</u> *:* Comme $\widehat{T}$ est une statistique exhaustive, on peut écrire d'après la démonstration du Théorème de factorisation de Neyman que pour tout $(x_1,\ldots,x_n)\in(\Omega')^n$ et tout $\theta\in\Theta$:

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}^*}(x_1,\ldots,x_n) = g_\theta(\widehat{T}(x_1,\ldots,x_n)).$$

On peut réécrire cela pour la densité de $\widehat{T}$ sous la forme : $\dfrac{d\mathbb{P}_\theta^{\widehat{T}}}{d\mathbb{P}^{*\widehat{T}}}(t) = g_\theta(t))$, pour tout $t\in\widehat{T}((\Omega')^n)$ et tout $\theta\in\Theta$. En conséquence, pour tout $\theta\in\Theta$,

$$I(\theta) \;=\; \left[\mathbb{E}_\theta\left(\frac{\partial(\log L_\theta(X_1,\cdots,X_N))}{\partial\theta_i}\times\frac{\partial(\log L_\theta(X_1,\cdots,X_N))}{\partial\theta_j}\right)\right]_{1\le i,j\le p}$$

$$
\begin{aligned}
&= \left[\int_{(\Omega')^n} \left(\frac{\partial(\log L_\theta(x))}{\partial\theta_i} \times \frac{\partial(\log L_\theta(x))}{\partial\theta_j}\right) d\mathbb{P}_\theta(x)\right]_{1\le i,j\le p} \\
&= \left[\int_{(\Omega')^n} \left(\frac{\partial(\log g_\theta(\widehat{T}(x)))}{\partial\theta_i} \times \frac{\partial(\log g_\theta(\widehat{T}(x)))}{\partial\theta_j}\right) g_\theta(\widehat{T}(x)) d\mathbb{P}^*(x)\right]_{1\le i,j\le p} \quad \text{car } \log L_\theta(x) = \log g_\theta(\widehat{T}(x)) + \log h(x) \\
&= \left[\int_{\widehat{T}(\Omega')^n} \left(\frac{\partial(\log g_\theta(t))}{\partial\theta_i} \times \frac{\partial(\log g_\theta(t))}{\partial\theta_j}\right) g_\theta(t)) d\mathbb{P}^{*\widehat{T}}(x)\right]_{1\le i,j\le p} \quad \text{d'après le théorème du transport} \\
&= \left[\int_{\widehat{T}(\Omega')^n} \left(\frac{\partial(\log g_\theta(t))}{\partial\theta_i} \times \frac{\partial(\log g_\theta(t))}{\partial\theta_j}\right) d\mathbb{P}_\theta(t)\right]_{1\le i,j\le p} \\
&= I_n^{\widehat{T}}(\theta).
\end{aligned}
$$

■

Remarque. *En rajoutant certaines hypothèses de continuité sur la vraisemblance de $\widehat{T}$, on peut montrer que la réciproque est également vraie, et donc que $I_n^{\widehat{T}}(\theta) = 0$ si et seulement si la statistique $\widehat{T}$ est exhaustive.*

Ainsi, on retrouve à l'aide de la notion d'information de Fisher les "intuitions" qui nous avaient guidées dans la section précédentes. Voyons maintenant les applications de la notion d'exhaustivité à l'estimation paramétrique.

3.4 Application à l'estimation paramétrique

On se place dans le cadre d'un modèle statistique paramétrique $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$, où $\Theta \subset \mathbb{R}^p$, dominé par une mesure μ. Par ailleurs, on suppose que Θ est un ouvert.

Définition.

- *Soit $g : \Theta \to \Theta'$, où $\Theta' \subset \mathbb{R}^{p'}$ avec $p' \in \mathbb{N}^*$, une fonction mesurable. On appelle estimateur de la fonction g du paramètre, donc de $g(\theta)$, une statistique $\widehat{T}$ à valeurs dans $\mathbb{R}^{p'}$. En particulier, un estimateur du paramètre θ est une statistique à valeurs dans $\mathbb{R}^p$. Une estimation de $g(\theta)$ est une réalisation de $\widehat{T}$.*
- *On appelle biais d'un estimateur $\widehat{T}$ de $g(\theta)$ le vecteur constant de $\mathbb{R}^{p'}$, $B(\theta) = \mathbb{E}_\theta(\widehat{T}) - g(\theta)$. On dira que l'estimateur est sans biais si $B(\theta) = 0$ pour tout $\theta \in \Theta$.*
- *On appelle risque quadratique de l'estimateur $\widehat{T}$ de $g(\theta)$ le réel positif $R(\theta) = \mathbb{E}_\theta(\|\widehat{T}) - g(\theta)\|^2)$, où $\|.\|$ désigne usuellement la norme euclidienne (mais peut être une autre fonctionnelle positive et convexe). Si l'estimateur est sans biais alors, $R(\theta) = Trace(cov(\widehat{T}))$.*

Pour pouvoir parler du comportement asymptotique d'une statistique, on va devoir se placer dans un "gros" modèle, dans lequel un échantillon est une suite de v.a. En quelque sorte, ce gros modèle pourra s'écrire $((\Omega')^{\mathbb{N}}, \mathcal{A}'_{\mathbb{N}}, \mathbb{P}_\theta^{\mathbb{N}}, \theta \in \Theta)$, où $\Theta \subset \mathbb{R}^p$ (la dimension du paramètre reste constante). Pour un n fixé, une statistique $\widehat{T}_n$ sera d'abord une projection du "gros" modèle sur le modèle de taille n, puis une statistique "normale". On devra donc parler d'une suite d'estimateurs $(\widehat{T}_n)_n$

Définition. *Pour un modèle statistique paramétrique $((\Omega')^{\mathbb{N}}, \mathcal{A}'_{\mathbb{N}}, \mathbb{P}_\theta^{\mathbb{N}}, \theta \in \Theta)$, où $\Theta \subset \mathbb{R}^p$, et pour $(\widehat{T}_n)_n$ une suite d'estimateurs de $g(\theta)$:*

- *Si $\lim_{n\to\infty} B_n(\theta) = 0$, on dit que l'estimateur est asymptotiquement sans biais.*
- *On dit que $(\widehat{T}_n)_n$ est convergent lorsque $\widehat{T}_n \underset{n\to+\infty}{\xrightarrow{\mathcal{P}}} g(\theta)$.*
- *S'il existe (a_n) une suite de réels positifs tels que $a_n(\widehat{T}_n - g(\theta)) \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} Z_\theta$, où Z_θ est une loi centrée non nulle (ne dépendent pas de n), on dit $(\widehat{T}_n)_n$ converge vers $g(\theta)$ à la vitesse a_n.*

A priori, être sans biais n'est pas un bon critère pour garantir une certaine optimalité de la convergence d'un estimateur. On préférera plutôt discriminer entre de potentiels estimateurs à l'aide d'un critère portant sur le risque quadratique ou sur la matrice de variance-covariance. Cependant, il n'existe pas de résultats généraux pour trouver un "meilleur" estimateur en ce sens. Pour en obtenir, on devra se limiter à une certaine classe d'estimateurs, celle des estimateurs sans biais.

Définition. *Soit un modèle statistique paramétrique* $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$, *et soit* $\widehat{T}$ *un estimateur sans biais de* $g(\theta)$. *On dit que* $\widehat{T}$ *est de variance uniformément minimum parmi les estimateurs sans biais de* $g(\theta)$ *lorsque pour tout estimateur sans biais de* $g(\theta)$, *on a* $\forall \theta \in \theta$, $cov(\widehat{T}) \leq cov(\widehat{S})$ *(au sens où* $cov(\widehat{T}) - cov(\widehat{S})$ *est une matrice positive).*

Propriété. *Si* $\widehat{T}$ *est un estimateur de variance uniformément minimum parmi les estimateurs sans biais, alors il est unique* $\mathbb{P}_\theta$*-p.s.*

<u>*Démonstration*</u> *:* Soit $\widehat{S}$ un autre estimateur que l'on suppose également de variance uniformément minimum parmi les estimateurs sans biais. Montrons d'abord que $E_\theta((\widehat{T} - \widehat{S}) \cdot {}^t\widehat{T}) = 0$. En effet, si $\alpha \in \mathbb{R}$, comme $\widehat{T}$ est de variance minimum, en utilisant des inégalités sur les matrices symétriques :

$$\begin{array}{rcl} \mathrm{cov}(\widehat{T}) & \leq & \mathrm{cov}(\widehat{T} + \alpha(\widehat{T} - \widehat{S})) \\ & \leq & \mathrm{cov}(\widehat{T}) + \alpha^2 \mathrm{cov}(\widehat{T} - \widehat{S}) + 2\alpha \cdot \mathbb{E}_\theta(\widehat{T} \cdot {}^t\widehat{S}) \\ \Longrightarrow \quad 0 & \leq & \alpha \cdot \big(\alpha \cdot \mathrm{cov}(\widehat{S}) + 2\mathbb{E}_\theta(\widehat{T} \cdot {}^t(\widehat{T} - \widehat{S}))\big) \quad \text{pour tout } \alpha \in \mathbb{R}. \end{array}$$

Comme $\mathrm{cov}(\widehat{T} - \widehat{S})$ est une matrice positive, la seule possibilité pour avoir la dernière inégalité est que : $\mathbb{E}_\theta(\widehat{T} \cdot {}^t(\widehat{T} - \widehat{S})) = 0$. Par suite, comme $\mathrm{cov}(\widehat{T} - \widehat{S}) = \mathbb{E}_\theta\big((\widehat{T} - \widehat{S}) \cdot {}^t(\widehat{T} - \widehat{S})\big) = \mathbb{E}_\theta\big(\widehat{T} \cdot {}^t(\widehat{T} - \widehat{S})\big) - \mathbb{E}_\theta\big(\widehat{S} \cdot {}^t(\widehat{T} - \widehat{S})\big)$, et que l'on a supposé $\widehat{T}$ et $\widehat{S}$ de variance minimum, $\mathrm{cov}(\widehat{T} - \widehat{S}) = 0$. Donc $\widehat{T} = \widehat{S}$ sur un ensemble de $\mathbb{P}_\theta$-mesure égale à 1. ■

Théorème (Rao-Blackwell)**.** *Si* $\widehat{T}$ *est un estimateur sans biais de* $g(\theta)$ *et si* $\widehat{S}$ *est une statistique exhaustive, alors* $\widehat{R} = \mathbb{E}_\theta(\widehat{T} \mid \widehat{S})$, *qui ne dépend pas de* θ *car* $\widehat{S}$ *est exhaustive, est un estimateur sans biais de* $g(\theta)$ *de matrice de covariance inférieure ou égale à celle de* $\widehat{T}$.

<u>*Démonstration*</u> *:* il est clair que $\mathbb{E}_\theta(\widehat{R}) = \mathbb{E}_\theta(\widehat{T}) = g(\theta)$. De plus, pour tout $u \in \mathbb{R}^{p'}$ (avec $g : \mathbb{R}^p \to \mathbb{R}^{p'}$),

$$\begin{array}{rcl} \mathrm{cov}({}^tu \cdot \widehat{T}) & = & \mathbb{E}_\theta\Big[\big({}^tu \cdot (\widehat{T} - g(\theta))\big)^2\Big] \\ & = & \mathbb{E}_\theta\Big(\mathbb{E}_\theta\Big[\big({}^tu \cdot (\widehat{T} - g(\theta))\big)^2\Big] \mid \widehat{S}\Big) \\ & \geq & \mathbb{E}_\theta\Big(\mathbb{E}_\theta\Big[{}^tu \cdot (\widehat{T} - g(\theta)) \mid \widehat{S}\Big]^2\Big) \quad \text{d'après l'inégalité de Jensen,} \\ & \geq & \mathrm{cov}({}^tu \cdot \widehat{R}). \end{array}$$

Cela revient bien à écrire que $\mathrm{cov}(\widehat{T}) \geq \mathrm{cov}(\widehat{R})$. ■

Théorème (Lehmann-Scheffé)**.** *Si* $\widehat{T}$ *est un estimateur sans biais de* $g(\theta)$ *et si* $\widehat{S}$ *est une statistique exhaustive et complète, alors l'unique estimateur de* $g(\theta)$ *sans biais uniformément de variance minimale est* $\widehat{R} = \mathbb{E}_\theta(\widehat{T} \mid \widehat{S})$ *(c'est-à-dire que* $\widehat{R}$ *est une fonction de* $\widehat{S}$*).*

<u>*Démonstration*</u> *:* Soit $\widehat{T}'$ un autre estimateur sans biais de $g(\theta)$. Si $\widehat{R}' = \mathbb{E}_\theta(\widehat{T}' \mid \widehat{S})$, on sait que $\mathrm{cov}(\widehat{T}') \geq \mathrm{cov}(\widehat{R}')$ d'après le Théorème de Rao-Blackwell. Or $\mathbb{E}_\theta(\widehat{R} - \widehat{R}') = 0$ pour tout $\theta \in \Theta$ car les deux estimateurs sont sans biais. De plus comme $\widehat{R}$ et $\widehat{R}'$ sont des fonctions de $\widehat{S}$, $\widehat{R} - \widehat{R}'$ l'est aussi, et du fait que $\widehat{S}$ est une statistique exhaustive et complète, alors pour tout $\theta \in \Theta$, $\widehat{R} = \widehat{R}'$, $\mathbb{P}_\theta$-p.s. Par conséquent, pour tout $\theta \in \Theta$, $\mathrm{cov}(\widehat{R}') = \mathrm{cov}(\widehat{R})$ et donc $\mathrm{cov}(\widehat{R}) \leq \mathrm{cov}(\widehat{T}')$: $\widehat{R}$ est bien l'estimateur sans biais de variance uniformément minimale. ■

Retenons donc de tout ceci que l'estimateur sans biais de $g(\theta)$ et de variance uniformément minimale est une unique fonction d'une statistique exhaustive et complète, lorsqu'une telle statistique existe. On aimerait maintenant connaître un peu mieux la covariance d'un tel estimateur.

Théorème (Inégalité de Cramer-Rao)**.** *Soit un modèle statistique paramétrique* $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$ *dominé et régulier, et soit* $\widehat{T}$ *un estimateur sans biais de* $g(\theta)$, *tel que* $\mathbb{E}_\theta\|\widehat{T}\|^2 < +\infty$. *Si on suppose que l'information de Fisher est une matrice définie positive, alors, en notant* $\dfrac{\partial g}{\partial \theta}(\theta)$ *la matrice jacobienne de* g, *pour tout* $\theta \in \Theta$ *:*

$$cov(\widehat{T}) \geq \frac{\partial g}{\partial \theta}(\theta) \cdot (I_n(\theta))^{-1} \cdot {}^t\frac{\partial g}{\partial \theta}(\theta) \quad (\textit{au sens des matrices symétriques}).$$

En particulier, si $\widehat{T}$ est un estimateur sans biais de θ, alors :

$$cov(\widehat{T}) \geq (I_n(\theta))^{-1} \quad (\textit{au sens des matrices symétriques}).$$

<u>*Démonstration*</u> *:* Soit $Z_\theta(x) = \text{grad}(\log L_\theta(x))$ où $x \in (\Omega')^n$ suit $\mathbb{P}_\theta$. On sait que comme le modèle est régulier, $\mathbb{E}_\theta(Z_\theta) = 0$ pour tout $\theta \in \Theta$ et donc :

$$\text{cov}(Z_\theta) = I(\theta) \quad \text{pour tout } \theta \in \Theta.$$

De plus, $\widehat{T}$ est un estimateur sans biais de $g(\theta)$ donc pour tout $\theta \in \Theta$:

$$\begin{aligned}
\mathbb{E}_\theta(\widehat{T}) = g(\theta) \quad & \Longrightarrow \quad \int_{(\Omega')^n} \widehat{T}(x) \cdot \frac{\partial L_\theta}{\partial \theta}(x)\, d\mu(x) = \frac{\partial g}{\partial \theta}(\theta) \ \ \text{(en dérivant)} \\
& \Longrightarrow \quad \int_{(\Omega')^n} \widehat{T}(x) \cdot \frac{\partial L_\theta}{\partial \theta}(x) \cdot (L_\theta(x))^{-1}\, d\mathbb{P}_\theta(x) = \frac{\partial g}{\partial \theta}(\theta) \\
& \Longrightarrow \quad \mathbb{E}_\theta(\widehat{T} \cdot {}^t Z_\theta) = \frac{\partial g}{\partial \theta}(\theta).
\end{aligned}$$

Ainsi, d'après ce qui précède,

$$\begin{aligned}
\text{cov}_\theta(\widehat{T} - \frac{\partial g}{\partial \theta}(\theta) \cdot I^{-1}(\theta) \cdot Z_\theta) &= \text{cov}_\theta(\widehat{T}) - 2\frac{\partial g}{\partial \theta}(\theta) \cdot I^{-1}(\theta) \cdot {}^t\frac{\partial g}{\partial \theta}(\theta) + \frac{\partial g}{\partial \theta}(\theta) \cdot I^{-1}(\theta) \cdot {}^t\frac{\partial g}{\partial \theta}(\theta) \\
&= \text{cov}_\theta(\widehat{T}) - \frac{\partial g}{\partial \theta}(\theta) \cdot I^{-1}(\theta) \cdot {}^t\frac{\partial g}{\partial \theta}(\theta).
\end{aligned}$$

En conséquence, comme $\text{cov}_\theta(\widehat{T} - \frac{\partial g}{\partial \theta}(\theta) \cdot I^{-1}(\theta) \cdot Z_\theta)$ est une matrice positive, l'inégalité de Cramer-Rao est prouvée. ■

Corollaire. *Deux cas particuliers méritent attention :*

- *Si le modèle est de la forme $((\Omega')^n, \mathcal{A}'_n, (f_\theta \cdot d\mu)^{\otimes n}, \theta \in \Theta)$, alors $I_n(\theta) = n \cdot I_1(\theta)$, où $I_1(\theta)$ est la matrice d'information de Fisher d'une seule variable aléatoire X distribuée suivant $f_\theta \cdot d\mu$ et l'Inégalité de Cramer-Rao devient donc :*

  $$cov(\widehat{T}) \geq \frac{1}{n} \cdot \left(\frac{\partial g}{\partial \theta}(\theta) \cdot (I_1(\theta))^{-1} \cdot {}^t\frac{\partial g}{\partial \theta}(\theta) \right) \quad (\textit{au sens des matrices symétriques}).$$

  On voit donc que pour un échantillon de variables indépendantes et identiquement distribuées, si la vraisemblance est régulière, alors la vitesse de convergence de tout estimateur sans biais est au mieux en $\sqrt{n}$.

- *Si le modèle n'est pas régulier, mais que sous la probabilité $\mathbb{P}_\theta$, la matrice d'information de Fisher existe et est inversible, et surtout si la propriété (4) est vérifiée, alors l'Inégalité de Cramer-Rao est vérifiée.* **Cela exclut cependant les modèles dont le support de $\mathbb{P}_\theta$ dépend de** θ*, comme par exemple le simple modèle de v.a.i.i.d. de loi $\mathcal{U}(]0, \theta[)$, avec $\theta > 0$.*

Définition. *Si un estimateur sans biais atteint (respectivement asymptotiquement) la borne de Cramer-Rao (qui ne dépend pas de l'estimateur), on dit qu'il est (resp. asymptotiquement) efficace.*

Remarque. *Un estimateur peut être sans biais, de variance minimale, mais ne pas atteindre la borne de Cramer-Rao, donc ne pas être efficace. De la même manière, il peut exister des estimateurs biaisés atteignant la borne de Cramer-Rao.*

Nous allons voir que les modèles exponentiels jouent un rôle central pour l'estimation paramétrique puisque sous certaines conditions ils sont les seuls pour lesquels on aura une estimation sans biais efficace.

Théorème. *Soit un modèle statistique paramétrique $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$, avec $\Theta \subset \mathbb{R}^p$, dominé et régulier. Soit $g : \mathbb{R}^d \to \mathbb{R}^d$ de classe $\mathcal{C}^1$ sur Θ telle que la matrice carrée de taille p, $\frac{\partial g}{\partial \theta}(\theta)$ soit de rang p*

pour tout $\theta \in \Theta$. Alors $\widehat{T} = {}^t(\widehat{T}_1, \ldots, \widehat{T}_d)$ est un estimateur sans biais de $g(\theta)$ atteignant la borne de Cramer-Rao si et seulement si le modèle est exponentiel et plus précisément s'il existe des fonctions $a : (\Omega')^n \to \mathbb{R}$, $\beta : \Theta \to \mathbb{R}$ et $\alpha_j : \Theta \to \mathbb{R}$ $(1 \leq j \leq p)$, telles que pour tout $\theta \in \Theta$, $\quad g(\theta) = -\left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)^{-1}_{1 \leq i,j \leq p} \cdot \frac{\partial \beta}{\partial \theta}(\theta)$ *et*

$$L_\theta(x_1, \cdots, x_n) = \exp\left(\beta(\theta) + b(x_1, \cdots, x_n) + \sum_{j=1}^d T_j(x_1, \cdots, x_n) \cdot \alpha_j(\theta)\right).$$

Démonstration : $\Longleftarrow$ On suppose donc le modèle exponentiel décrit dans le théorème. Si on dérive par rapport à θ un tel modèle, on obtient que pour μ-presque tout $x \in (\Omega')^n$:

$$\frac{\partial}{\partial \theta}(\log L_\theta(x)) = \left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)_{1 \leq i,j \leq p} \cdot \widehat{T} + \frac{\partial \beta}{\partial \theta}(\theta), \quad \text{pour tout } \theta \in \Theta. \tag{5}$$

En conséquence, comme $I(\theta) = \mathbb{E}_\theta\left(\left(\frac{\partial}{\partial \theta}(\log L_\theta(.))\right) \cdot {}^t\left(\frac{\partial}{\partial \theta}(\log L_\theta(.))\right)\right)$, on en déduit que :

$$I(\theta) = \left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)_{1 \leq i,j \leq d} \cdot \mathrm{cov}_\theta(\widehat{T}) \cdot {}^t\left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)_{1 \leq i,j \leq p} \quad \Longrightarrow \quad \mathrm{cov}_\theta(\widehat{T}) = \left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)^{-1}_{1 \leq i,j \leq p} \cdot I(\theta) \cdot {}^t\left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)^{-1}_{1 \leq i,j \leq p}$$

Par ailleurs, comme $\widehat{T}$ est un estimateur sans biais de $g(\theta)$ d'après la preuve de l'Inégalité de Cramer-Rao,

$$\mathbb{E}_\theta\left(\widehat{T}(.) \cdot {}^t\left(\frac{\partial}{\partial \theta}(\log L_\theta(.))\right)\right) = \frac{\partial g}{\partial \theta}(\theta)$$

et en utilisant (5) que l'on multiplie par $\left(\frac{\partial}{\partial \theta}(\log L_\theta(.))\right)$, on obtient :

$$\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial \theta}(\log L_\theta(.))\right) \cdot {}^t\left(\frac{\partial}{\partial \theta}(\log L_\theta(.))\right)\right) = \mathbb{E}_\theta\left(\left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)_{1 \leq i,j \leq p} \cdot \widehat{T} \cdot {}^t\left(\frac{\partial}{\partial \theta}(\log L_\theta(.))\right)\right) + \mathbb{E}_\theta\left(\frac{\partial \beta}{\partial \theta}(\theta) \cdot {}^t\left(\frac{\partial}{\partial \theta}(\log L_\theta(.))\right)\right),$$

et donc $I(\theta) = \left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)_{1 \leq i,j \leq p} \cdot \frac{\partial g}{\partial \theta}(\theta)$. A l'aide de cette égalité, et en reprenant le calcul précédent, on en arrive à ce que :

$$\mathrm{cov}_\theta(\widehat{T}) = \frac{\partial g}{\partial \theta}(\theta) \cdot I^{-1}(\theta) \cdot {}^t\frac{\partial g}{\partial \theta}(\theta),$$

donc $\widehat{T}$ atteint bien la borne de Cramer-Rao. De plus, grâce à (5),

$$\begin{aligned} \mathbb{E}_\theta\left(\frac{\partial}{\partial \theta}(\log L_\theta(x))\right) &= \mathbb{E}_\theta\left(\left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)_{1 \leq i,j \leq p} \cdot \widehat{T} + \frac{\partial \beta}{\partial \theta}(\theta)\right) \\ \text{soit} \qquad 0 &= \left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)_{1 \leq i,j \leq p} \cdot g(\theta) + \frac{\partial \beta}{\partial \theta}(\theta) \\ \text{et donc} \qquad g(\theta) &= -\left(\frac{\partial \alpha_j}{\partial \theta_i}(\theta)\right)^{-1}_{1 \leq i,j \leq p} \cdot \frac{\partial \beta}{\partial \theta}(\theta). \end{aligned}$$

$\Longrightarrow$ D'après la preuve de l'Inégalité de Cramer-Rao, si $\widehat{T}$ est un estimateur sans biais de $g(\theta)$ atteignant la borne de Cramer-Rao, alors

$$\mathrm{cov}_\theta(\widehat{T} - \frac{\partial g}{\partial \theta}(\theta) \cdot I^{-1}(\theta) \cdot Z_\theta) = 0.$$

Ainsi, pour tout $\theta \in \Theta$, il existe un ensemble $N_\theta \subset (\Omega')^n$ tel que $\mathbb{P}_\theta(N_\theta) = 1$ et tel que pour tout $x \in N_\theta$, $\widehat{T}(x) - g(\theta) = \frac{\partial g}{\partial \theta}(\theta) \cdot I^{-1}(\theta) \cdot Z_\theta(x)$. Par le même procédé que celui de la preuve de la nullité de l'information de Fisher pour une statistique libre, on peut déterminer un ensemble N ne dépendant pas de θ, tel que cette propriété soit également vraie, avec $\mu(N) = 1$, ce qui revient à écrire que $\forall x \in N$,

$$I(\theta) \cdot \left(\frac{\partial g}{\partial \theta}(\theta)\right)^{-1} \cdot \left(\widehat{T}(x) - g(\theta)\right) = \frac{\partial}{\partial \theta}(\log L_\theta(x)), \quad \text{pour tout } \theta \in \Theta.$$

Alors en intégrant par rapport à θ, et en notant $\begin{cases} \alpha(\theta) \text{ le vecteur colonne "intégrant" } I(\theta)\cdot\left(\frac{\partial g}{\partial\theta}(\theta)\right)^{-1} \\ \beta(\theta) \text{ la fonction "intégrant" } -I(\theta)\cdot\left(\frac{\partial g}{\partial\theta}(\theta)\right)^{-1}\cdot g(\theta) \\ b(x) \text{ une fonction ne dépendant pas de } \theta \end{cases}$

on a $\log L_\theta(x) = \alpha(\theta)\cdot\widehat{T}(x)+\beta(\theta)+b(x)$, d'où l'écriture de la vraisemblance sous forme d'un modèle exponentiel, et on retrouve l'expression de $g(\theta)$ par le même raisonnement que plus haut. ■

Corollaire. *A l'inverse, si l'on dispose d'un modèle exponentiel régulier (3), alors il n'existe qu'une seule fonction (à une transformation affine près) du paramètre pouvant être estimer efficacement, il s'agit de* $g(\theta) = -\frac{1}{n}\cdot\left(\frac{\partial\alpha_j}{\partial\theta_i}(\theta)\right)^{-1}_{1\le i,j\le p}\cdot\frac{\partial\beta}{\partial\theta}(\theta)$ *(noter que cette fonction semble dépendre de n; dans le cas de v.a.i.i.d. ce n'est pas le cas). L'estimateur est alors :* $\widehat{T} = \frac{1}{n}\cdot(a_1(X_1,\dots,X_n),\dots,a_p(X_1,\dots,X_n))$ *et sa matrice de covariance minimale est donnée par sa borne de Cramer-Rao, soit :*

$$cov_\theta(\widehat{T}) = \frac{1}{n}\cdot\frac{\partial g}{\partial\theta}(\theta)\cdot\left(\frac{\partial\alpha_j}{\partial\theta_i}(\theta)\right)^{-1}_{1\le i,j\le d}.$$

3.5 Estimateur du maximum de vraisemblance

Nous allons voir une méthode permettant d'obtenir aisément et dans la plupart des cas un estimateur possédant de très bonnes qualités... Par la suite on se place une nouvelle fois dans le cadre d'un modèle statistique paramétrique $((\Omega')^n,\mathcal{A}'_n,\mathbb{P}_\theta,\theta\in\Theta)$, avec $\Theta\subset\mathbb{R}^p$, dominé.

Définition. *Pour* $(x_1,\dots,x_n)\in(\Omega')^n$*, soit* $\theta\in\Theta\mapsto L_\theta(x_1,\dots,x_n)$ *la vraisemblance du modèle. On appelle estimateur du maximum de vraisemblance une statistique* $\widehat{\theta}_n$ *telle que pour* $(X_1,\dots,X_n)$ *un n-échantillon quelconque du modèle :*

$$L_{\widehat{\theta}_n}(X_1,\cdots,X_n) = \sup_{\theta\in\Theta} L_\theta(X_1,\cdots,X_n).$$

Remarque. *Il n'y a pas de garantie de l'unicité d'un tel estimateur. Une méthode pour l'obtenir (mais pas toujours) est de rechercher un extremum local de* L_θ *sur* Θ*, ce qui pourra être fait en annulant les dérivées partielles de* L_θ *par* θ_i*. De même, il est clair que l'estimateur du maximum de vraisemblance pourra être également obtenu en maximisant le logarithme de la vraisemblance, appelé encore la log-vraisemblance. Enfin, si l'on désire estimer* $g(\theta)$ *avec g une fonction bijective, alors* $g(\widehat{\theta})$ *sera l'estimateur du maximum de vraisemblance de* $g(\theta)$*.*

Propriété. *S'il existe une statistique exhaustive* $\widehat{T}$ *pour le modèle, alors* $\widehat{\theta}$ *est une fonction mesurable de* $\widehat{T}$ *pour tout* $\theta\in\Theta$*.*

Démonstration : Si $\widehat{T}$ est exhaustive, d'après le théorème de factorisation, la vraisemblance du modèle par rapport à la mesure dominante P^* est $g_\theta(\widehat{T}(x_1,\dots,x_n))$ pour tout $\theta\in\Theta$ et $\mathbb{P}_\theta$-presque tout $(x_,\dots,x_n)\in(\Omega')^n$, ce qui revient à P^*-presque tout $(x_,\dots,x_n)\in(\Omega')^n$ par la même démonstration que celle de la nullité de l'information de Fisher d'une statistique libre. Ainsi, prendre l'argument maximal de $\theta\to L_\theta$ revient à prendre l'argument maximal de $\theta\to g_\theta(\widehat{T}(x_1,\dots,x_n))$, et $\widehat{\theta}$ sera donc une fonction de $\widehat{T}$. ■

Propriété. *On suppose que le modèle est régulier. Si on suppose qu'il existe un estimateur sans biais efficace de* θ *alors c'est l'estimateur du maximum de vraisemblance de* θ*.*

Démonstration : D'après ce qui précède, si le modèle est régulier et que $\widehat{T}$ est un estimateur sans biais efficace de θ , alors le modèle est exponentiel et l'égalité (5) a encore lieu, soit pour tout $\theta\in\Theta$,

$$\frac{\partial}{\partial\theta}(\log L_\theta(x)) = \left(\frac{\partial\alpha_j}{\partial\theta_i}(\theta)\right)_{1\le i,j\le p}\cdot\widehat{T}+\frac{\partial\beta}{\partial\theta}(\theta) \implies \left(\frac{\partial\alpha_j}{\partial\theta_i}(\theta)\right)_{1\le i,j\le p}\cdot\mathbb{E}_\theta(\widehat{T})+\frac{\partial\beta}{\partial\theta}(\theta)=0.$$

Comme $\widehat{T}$ est un estimateur sans biais de θ, on a donc $\left(\frac{\partial\alpha_j}{\partial\theta_i}(\theta)\right)_{1\le i,j\le p}\cdot\theta+\frac{\partial\beta}{\partial\theta}(\theta)=0$, pour tout $\theta\in\Theta$, ce qui s'applique également à $\widehat{\theta}$ et donc :

$$\left(\frac{\partial\alpha_j}{\partial\theta_i}(\widehat{\theta})\right)_{1\le i,j\le p}\cdot\widehat{\theta}+\frac{\partial\beta}{\partial\theta}(\widehat{\theta})=0.$$

Mais d'après sa définition, le modèle étant régulier $\widehat{\theta}$ minimise la log-vraisemblance et annule donc sa dérivée, ce qui implique que :

$$\left(\frac{\partial \alpha_j}{\partial \theta_i}(\widehat{\theta})\right)_{1\leq i,j\leq p} \cdot \widehat{T} + \frac{\partial \beta}{\partial \theta}(\widehat{\theta}) = 0.$$

En conséquence, obtient :

$$\left(\frac{\partial \alpha_j}{\partial \theta_i}(\widehat{\theta})\right)_{1\leq i,j\leq p} \cdot \left(\widehat{T} - \widehat{\theta}\right) = 0 \implies \widehat{T} = \widehat{\theta},$$

car la matrice des dérivées des α_j est supposée de rang d. Enfin, l'unicité de $\widehat{\theta}$ est liée à l'écriture du modèle exponentiel. ■

Nous allons nous intéresser maintenant au comportement asymptotique de l'estimateur du maximum de vraisemblance (lorsqu'il existe), donc quand la taille n de l'échantillon tend vers l'infini. Il est clair que pour chaque n l'expression de l'estimateur est différente et, surtout, le modèle statistique change. Pour palier à cela, on se placera dans un "gros" modèle, $((\Omega')^{\mathbb{N}}, \mathcal{A}'_{\mathbb{N}}, \mathbb{P}_\theta^{\mathbb{N}}, \theta \in \Theta)$, où $\Theta \subset \mathbb{R}^p$ (la dimension du paramètre reste constante) dans lequel un échantillon est une suite de v.a. Par ailleurs, on supposera désormais que **tout échantillon de ce modèle est constitué de v.a.i.i.d.**, et que $d\mathbb{P}_\theta^{\mathbb{N}} = (f_\theta \cdot d\mu)^{\otimes \mathbb{N}}$, le modèle étant dominé par la mesure μ, et f_θ étant la densité de chaque X_i par rapport à μ.

Théorème (Convergence de l'estimateur du maximum de vraisemblance)**.** *On suppose le modèle paramétrique* $((\Omega')^{\mathbb{N}}, \mathcal{A}'_{\mathbb{N}}, (f_\theta \cdot d\mu)^{\otimes \mathbb{N}}, \theta \in \Theta)$*, où* $\Theta \subset \mathbb{R}^d$ *dominé par une mesure* μ *et régulier. On suppose en plus que le modèle est identifiable (au sens où* $f_{\theta_1} = f_{\theta_2}$*,* μ*-presque partout, entraîne* $\theta_1 = \theta_2$*). Alors si la suite* $(X_n)_{n\in\mathbb{N}}$ *est issue du modèle avec pour paramètre* $\theta_0 \subset \Theta$*,*

$$\widehat{\theta}_n \xrightarrow[n\to+\infty]{p.s.} \theta_0 \quad \textit{pour la mesure } (f_{\theta_0} \cdot d\mu)^{\otimes \mathbb{N}}.$$

<u>Démonstration</u> : En premier lieu, pour n fixé, il est clair que pour tout $\theta \in \Theta$:

$$\log(L_\theta(x_1, \ldots, x_n)) - \log(L_{\theta_0}(x_1, \ldots, x_n)) = \sum_{i=1}^n \log\left(\frac{f_\theta(x_i)}{f_{\theta_0}(x_i)}\right).$$

Par ailleurs, pour tout $i \in \mathbb{N}$, les X_i ont tous la même loi et pour $\theta \in \Theta$,

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left[\log\left(\frac{f_\theta(X_i)}{f_{\theta_0}(X_i)}\right)\right] &\leq \log\left(\mathbb{E}_{\theta_0}\left[\frac{f_\theta(X_i)}{f_{\theta_0}(X_i)}\right]\right) \quad \text{(Inégalité de Jensen pour la fonction } -\log\text{)}\\
&\leq \log\left(\mathbb{E}_\mu\left[f_\theta(X_i)\right]\right)\\
&\leq 0.
\end{aligned}$$

En fait, du fait que la fonction $-\log$ est strictement convexe, la borne 0 ne peut être atteinte que si $f_\theta = f_{\theta_0}$. Ainsi, avec la contrainte d'un modèle identifiable, dès que $\theta \neq \theta_0$, alors :

$$\mathbb{E}_{\theta_0}\left[\log\left(\frac{f_\theta(X_i)}{f_{\theta_0}(X_i)}\right)\right] < 0.$$

On peut appliquer la loi forte des grands nombres pour les variables aléatoires $\left(\log\left(\frac{f_\theta(X_i)}{f_{\theta_0}(X_i)}\right)\right)_{i\in\mathbb{N}}$ (qui sont bien i.i.d. et $\mathbb{L}^1$ car le modèle est régulier), et ainsi :

$$\begin{aligned}
\frac{1}{n}\left(\log(L_\theta(X_1, \ldots, X_n)) - \log(L_{\theta_0}(X_1, \ldots, X_n))\right) &= \frac{1}{n}\sum_{i=1}^n \log\left(\frac{f_\theta(X_i)}{f_{\theta_0}(X_i)}\right)\\
&\xrightarrow[n\to+\infty]{p.s.} \mathbb{E}_{\theta_0}\left[\log\left(\frac{f_\theta(X_i)}{f_{\theta_0}(X_i)}\right)\right] < 0,
\end{aligned}$$

la convergence presque sûre ayant lieu pour la mesure $(f_{\theta_0} \cdot d\mu)^{\otimes \mathbb{N}}$. Considérons maintenant pour tout $\varepsilon > 0$ une famille dénombrable $(\theta_i^{(\varepsilon)})_{i\in I}$ dense sur la sphère de centre θ_0 et de rayon ε. Du fait du caractère dénombrable de cette famille, pour tout $\varepsilon > 0$, il existe n_ε tel que pour tout $n \geq n_\varepsilon$, pour tout $i \in I$:

$$\log(L_{\theta_i^{(\varepsilon)}}(X_1, \ldots, X_n)) < \log(L_{\theta_0}(X_1, \ldots, X_n)) \quad \text{p.s. pour la mesure } (f_{\theta_0} \cdot d\mu)^{\otimes \mathbb{N}}.$$

Comme le modèle est régulier, pour tout $n \in \mathbb{N}^*$, la log-vraisemblance de $X_1, \ldots, X_n$ est continue sur Θ. De plus pour tout n elle atteint son unique maximum en θ_0. En conséquence, pour $n \geq n_\varepsilon$, $\widehat{\theta}_n$ sera à l'intérieur de la boule de centre θ_0 et de rayon ε (toujours p.s. pour la mesure $(f_{\theta_0} \cdot d\mu)^{\otimes \mathbb{N}}$). Le raisonnement étant vrai pour tout $\varepsilon > 0$, le théorème s'en déduit. ■

Théorème (Normalité asymptotique de l'estimateur du maximum de vraisemblance)**.** *On suppose le modèle paramétrique* $((\Omega')^{\mathbb{N}}, \mathcal{A}'_{\mathbb{N}}, (f_\theta \cdot d\mu)^{\otimes \mathbb{N}}, \theta \in \Theta)$*, où* $\Theta \subset \mathbb{R}^p$*, dominé par une mesure* μ *et régulier. On suppose en plus que le modèle est identifiable et que la fonction* $\theta \in \Theta \mapsto L_\theta$ *est de classe* $\mathcal{C}^2(\Theta)$*. Alors si la suite* $(X_n)_{n \in \mathbb{N}}$ *est issue du modèle avec pour paramètre* $\theta_0 \subset \Theta$ *:*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}_d(0, I_1^{-1}(\theta_0)),$$

où $I_1(\theta)$ *est la matrice de Fisher de taille* p *(supposée inversible) pour la variable* X_1*.*

Démonstration : Comme le modèle est régulier, on peut différencier la vraisemblance et pour tout $\theta \in \Theta$, noter :

$$M_\theta(X_1, \ldots, X_n) = \frac{1}{n} \frac{\partial}{\partial \theta} \log L_\theta(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log \left(f_\theta(X_i)\right).$$

Un développement limité d'ordre 1 de M_θ autour de θ_0 est possible (toujours en raison du modèle régulier) et donc pour tout tout $\theta \in \Theta$:

$$M_\theta(X_1, \ldots, X_n) = M_{\theta_0}(X_1, \ldots, X_n) + (\theta - \theta_0) \cdot \frac{\partial}{\partial \theta} M_{\theta^*}(X_1, \ldots, X_n),$$

avec θ^* dans le segment $[\theta, \theta_0]$ (remarquons que $\dfrac{\partial}{\partial \theta} M_{\theta^*}(X_1, \ldots, X_n)$ est une matrice carrée de taille d). Ainsi en remplaçant θ par $\widehat{\theta}_n$, on obtient pour chaque n l'existence de θ_n^* appartenant au segment $[\widehat{\theta}_n, \theta_0]$ tel que :

$$M_{\widehat{\theta}_n}(X_1, \ldots, X_n) = M_{\theta_0}(X_1, \ldots, X_n) + (\widehat{\theta}_n - \theta_0) \cdot \frac{\partial}{\partial \theta} M_{\theta_n^*}(X_1, \ldots, X_n). \tag{6}$$

Pour un modèle régulier, on a vu que $\mathbb{E}_{\theta_0}\left(\dfrac{\partial^2}{\partial \theta^2} \log f_{\theta_0}(X_i)\right) = -I_1(\theta_0)$, matrice de Fisher pour n'importe quelle variable X_i. Ainsi, $\dfrac{\partial}{\partial \theta} M_\theta(.)$ étant une moyenne empirique, on a par la loi forte des grands nombres :

$$\frac{\partial}{\partial \theta} M_{\theta_0}(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \log f_{\theta_0}(X_i) \xrightarrow[n \to +\infty]{p.s.} -I_1(\theta_0) \text{ pour la mesure } (f_{\theta_0} \cdot d\mu)^{\otimes \mathbb{N}}.$$

Maintenant, en utilisant le fait que les densités f_θ sont de classe $\mathcal{C}^2(\Theta)$ et en utilisant la convergence presque sûre de $\widehat{\theta}_n$ vers θ_0 démontrée au théorème précédent, on a :

$$\frac{\partial}{\partial \theta} M_{\theta_n^*}(X_1, \ldots, X_n) \xrightarrow[n \to +\infty]{p.s.} -I_1(\theta_0) \text{ pour la mesure } (f_{\theta_0} \cdot d\mu)^{\otimes \mathbb{N}}.$$

Finalement, comme $\widehat{\theta}_n$ est le maximum d'une fonction de classe $\mathcal{C}^1$, cet estimateur annule $M_{\widehat{\theta}_n}(X_1, \ldots, X_n)$, et donc l'égalité (6) devient :

$$M_{\theta_0}(X_1, \ldots, X_n) \cdot I_1^{-1}(\theta_0) = (\widehat{\theta}_n - \theta_0).$$

Enfin, comme $M_{\theta_0}(X_1, \ldots, X_n)$ est une moyenne empirique, ce vecteur aléatoire vérifie un théorème de la limite centrale :

$$\sqrt{n}\left(M_{\theta_0}(X_1, \ldots, X_n) - \mathbb{E}_{\theta_0}\left(\frac{\partial}{\partial \theta} \log f_{\theta_0}(X_i)\right)\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}_d(0, I_1(\theta_0)),$$

d'après la première définition de l'information de Fisher. Comme $\mathbb{E}_{\theta_0}\left(\dfrac{\partial}{\partial \theta} \log f_{\theta_0}(X_i)\right) = 0$ (voir les propriétés précédentes), on obtient la normalité asymptotique de $\widehat{\theta}_n$. ■

Remarque. *Sous ces hypothèses, l'estimateur du maximum de vraisemblance est asymptotiquement sans biais et efficace. Cependant, à* n *fixé, il peut avoir un biais et ne pas être un estimateur efficace.*

3.6 Régions de confiance

En pratique, estimer un paramètre le plus souvent ne suffit pas. On aimerait connaître plus précisément quelle marge de sécurité on a sur la connaissance de ce paramètre.

Définition. *On se place dans le cadre d'un modèle paramétrique* $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$*, où* $\Theta \subset \mathbb{R}^p$*. Soit* $\alpha \in]0,1[$ *un nombre fixé a priori. On appelle région de confiance du paramètre* θ *de niveau* $1-\alpha$ *un sous-ensemble aléatoire* $R_{1-\alpha}$ *inclus dans* $\mathbb{R}^p$ *et défini sur* $((\Omega')^{\mathbb{N}}, \mathcal{A}'_{\mathbb{N}})$*, tel que pour tout* $\theta \in \Theta$*,* $\{(x_1,\ldots,x_n) \in (\Omega')^{\mathbb{N}}, \theta \in R_{1-\alpha}(x_1,\ldots,x_n)\} \in \mathcal{A}'_n$ *et :*

$$\inf_{\theta\in\Theta} \{\mathbb{P}_\theta(\theta \in R_{1-\alpha})\} \geq 1-\alpha. \tag{7}$$

Si un échantillon observé $(X_1(\omega),\ldots,X_n(\omega))$ *est connu,* $R_{1-\alpha}(X_1(\omega),\ldots,X_n(\omega))$ *est appelé région de confiance observé. Dans le cas où le paramètre est un réel* ($p=1$)*, on pourra obtenir un intervalle de confiance.*

Comment déterminer une région de confiance ? En premier lieu, il est clair que pour tout $\alpha \in]0,1[$, $R_{1-\alpha} \subset \Theta$ (en général, on choisit α proche de 0, et en particulier $\alpha = 0.05$ est très souvent utilisé). Une démarche possible pour la construction de région de confiance est la suivante : naturellement, on désirerait utiliser un estimateur $\widehat{T}$ convergent de θ, mais sa loi dépend en général de θ ce qui rend difficile (à part quelques exceptions) son utilisation directe. On préférera donc utiliser ce que l'on appelle une fonction pivotale $\pi(\widehat{T},\theta)$, qui est une fonction mesurable d'un estimateur et de θ et qui est une statistique libre. On essayera alors d'écrire la propriété (7) sous la forme

$$\inf_{\theta\in\Theta} \left\{\mathbb{P}_\theta(\pi(\widehat{T},\theta) \in C_\alpha)\right\} \geq 1-\alpha,$$

où C_α est une région déterministe. Aussi pourra-t-on ensuite construire la région de confiance en fonction des quantiles (souvent à $\alpha/2$ et $1-\alpha/2$) de la loi de la fonction pivotale.

Exemple. *Si le modèle est régulier, sous les conditions du théorème de normalité asymptotique du maximum de vraisemblance, on peut également montrer (théorème de Sltuski) que*

$$\pi(\widehat{\theta}_n,\theta_0) = \sqrt{n}\cdot(I_1(\widehat{\theta}_n))^{1/2}\cdot\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}_d(0,I_p),$$

où I_d *est la matrice identité de taille* p *et* $(I_1(\theta))^{1/2}\cdot(I_1(\theta))^{1/2} = I_1(\theta)$ *pour tout* $\theta\in\Theta$*. Ainsi, si* n *est grand, on pourra assimiler la loi de* $\pi(\widehat{\theta}_n,\theta_0)$ *avec la loi normale centrée réduite multidimensionnelle. Or si* $Z \sim \mathcal{N}_p(0,I_p)$*, avec* $q_{1-\alpha/2} > 0$ *le quantile d'une loi normale centrée réduite réelle de niveau* $1-\alpha/2$*, tel que* $P(Z \in [-q_{1-\alpha/2}, q_{1-\alpha/2}]^d) \geq 1-\alpha$*. Aussi le polyèdre* $n^{-1/2}\cdot(I_1(\widehat{\theta}_n))^{-1/2}\cdot[-q_{1-\alpha/2},q_{1-\alpha/2}]^d$ *recentré autour de* $\widehat{\theta}_n$ *formera la région de confiance cherchée.*

4 Tests paramétriques

4.1 Principes d'un test

Un test permet, à partir d'une réalisation d'un échantillon, de décider entre deux hypothèses, en mettant en avant une hypothèse privilégiée, appelée hypothèse H_0, et une hypothèse alternative, appelée H_1. On associe à un test un niveau α (avec souvent $\alpha \simeq 0.05$) et une puissance $1-\beta$. La plupart du temps, α est fixé a priori et β s'en déduit. Plus précisément,

Définition. *On se place dans le cadre d'un modèle paramétrique dominé* $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta\in\Theta)$*, où* $\Theta\subset\mathbb{R}^p$ *et soit* θ *la "vraie" valeur du paramètre. Un problème de test est un choix entre deux hypothèses :*

$$\begin{cases} H_0 : \theta\in\Theta_0 & : \textit{hypothèse dite nulle} \\ \\ H_1 : \theta\in\Theta_1 & : \textit{hypothèse dite alternative,} \end{cases} \tag{8}$$

où $\Theta_0 \subset \mathbb{R}^p$*,* $\Theta_1\subset\mathbb{R}^d$ *et* $\Theta_0\cap\Theta_1 = \emptyset$*.*

Ceci posé, on peut préciser deux types de problèmes de tests suivant les constitutions de Θ_0 et Θ_1 :

Définition. *Une hypothèse (H_0 ou H_1) est dite simple si elle est associée à un singleton (Θ_0 ou Θ_1). Sinon, elle sera dite composite. Dans le cas réel ($\Theta \subset \mathbb{R}$), si H_0 est simple de la forme $\theta = \theta_0$, et si H_1 est composite de la forme $\theta > \theta_0$ ou $\theta < \theta_0$, on parlera de test unilatéral; si H_1 est composite de la forme $\theta \neq \theta_0$, on parlera de test bilatéral.*

Comment faire pour choisir entre les deux hypothèses H_1 et H_2 ? Il faudra partir de ce que l'on peut connaître du modèle, c'est-à-dire généralement un échantillon observé $(X_1, \ldots, X_n)$. Pour cela, on définit une statistique qui sera la clé de voûte du test :

Définition. *Dans le cadre du problème de test (8, soit $\widehat{T}$ une statistique (donc une fonction mesurable d'un échantillon $(X_1, \cdots, X_n)$ issu du modèle) à valeurs dans $\mathbb{R}^d$, qui sera appelée statistique du test. Le test sera défini par la fonction $\widehat{\phi} = \mathbb{I}_{\widehat{T} \in W}$, où W est une partie de $\mathbb{R}^p$ appelée région critique du test (et sa partie complémentaire dans $\mathbb{R}^p$ est appelée région d'acceptation du test). Si $\widehat{\phi} = 1$, on choisira H_1, sinon on décidera plutôt H_0.*

Donc, à chaque hypothèse H_0 et H_1, on associe une partie de $\mathbb{R}^p$ pour la statistique de test $\widehat{T}$. En général, ces parties ne sont pas Θ_0 et Θ_1. Pour pouvoir précisément déterminer la région W, dans un cadre théorique (qui n'est pas le même que le cadre pratique, voir plus bas), on peut commencer par associer une fonction puissance à la statistique de test, puis définir les erreurs de premier espèce α et de deuxième espèce β :

Définition. *Pour la statistique de test $\widehat{T}$, on associe :*

- *une fonction puissance, qui est la probabilité de choisir H_1 : $\theta \in \Theta_1 \mapsto \mathbb{P}_\theta(\widehat{T} \notin W)$.*
- *une erreur de première espèce : $P_{H_0}(\text{Choisir } H_1) = \alpha = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\widehat{T} \in W)$;*
- *une erreur de seconde espèce : $P_{H_1}(\text{Choisir } H_0) = \beta = \sup_{\theta \in \Theta_1} \mathbb{P}_\theta(\widehat{T} \notin W)$.*

La puissance du test est $1 - \beta$.

Cependant, ce qui vient d'être écrit reste théorique. En pratique, on utilisera plutôt la démarche suivante :

Construction concrète d'un test : On suppose le problème de test (8). On pose également a priori α qui dépend du problème posé (mais en général $\alpha = 0.05$), et $1 - \alpha$ est appelé le niveau du test. Par la suite, on réalise :

1. L'expression quantitative des hypothèses H_0 et H_1.
2. Le choix de la statistique $\widehat{T}$ du test.
3. La construction d'une région critique W à l'hypothèse H_1 par rapport à $\widehat{T}$.
4. La détermination explicite de W en fonction de α.
5. Le calcul (si possible) de la puissance du test $1 - \beta$.
6. Pour la réalisation de l'échantillon, rejet ou acceptation de H_0.

Remarque : Cependant, en pratique on ne procède pas ainsi. On a donc deux types d'erreur. Le choix de l'hypothèse privilégiée est donc fondamental car le résultat d'un test n'est pas symétrique. Par exemple, supposons que l'on ait pour modèle $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{N}(\theta, 1)^{\otimes n}, \theta \in \mathbb{R})$ et que l'on veuille tester $H_0 : \theta = 0$ contre $H_1 : \theta = 1$ à partir d'un échantillon $(X_1, \cdots, X_n)$ du modèle. Nous verrons pourquoi un peu plus loin, $\overline{X}_n$ est une statistique de test pertinente. Par exemple, si $n = 1$, et $X_1(\omega) = \overline{X}_1(\omega) = 0.8$, que va-t-on choisir entre H_0 et H_1 ? Naturellement, une région critique sera de la forme $[s, +\infty[$, où $s \in \mathbb{R}$, car $\overline{X}_n$ est un estimateur de θ. On détermine s à l'aide de α, puisque $P_{H_0}(\text{Choisir } H_1) = \alpha = P_0(\overline{X}_1 \geq s)$, donc par exemple, si $\alpha = 0.05$, $s \simeq 1.65$. Par suite, si $\overline{X}_1(\omega) = 0.8$, on accepte H_0 et l'erreur de seconde espèce est $P_1(\overline{X}_1 < s) \simeq 0.74$, donc très élevée : le test n'est pas très discriminant. Maintenant, si on inverse H_0 et H_1, soit $H_0 : \theta = 1$ contre $H_1 : \theta = 0$, le même résultat $X_1(\omega) = 0.8$, conduit à accepter H_0, avec une erreur de second espèce encore $\simeq 0.74$. On obtient donc deux résultats opposés pour la même expérience aléatoire. Les hypothèses H_0 et H_1 ne sont clairement pas interchangeable.

La question qui se pose maintenant est de savoir comment trouver une statistique de test. Une idée naturelle dans ce cadre paramétrique serait d'utiliser un estimateur du maximum de vraisemblance.

4.2 Test de Wald

Un estimateur du maximum de vraisemblance permet d'associer à chaque hypothèse du test un ensemble de même "forme" que Θ_0 et Θ_1. Cependant, la difficulté est trouver la loi de l'estimateur du maximum de vraisemblance $\widehat{\theta}$ à n fixé. Si cela est possible, on utilisera directement $\widehat{\theta}$ comme statistique de test.

Sinon, de manière plus générale, on connaît la loi asymptotique de $\widehat{\theta}_n$ quand le modèle est régulier. Donc quand n est grand, on pourrait utiliser une loi normale comme approximation de la loi de $\widehat{\theta}_n$. Mais, un nouvel obstacle apparaît : la matrice de covariance asymptotique, qui est la matrice d'information de Fisher inverse, dépend du paramètre θ. Aussi va-t-on préférer utiliser la statistique de test $\widehat{T}$ suivante :

Définition. *Pour un modèle paramétrique dominé* **régulier** $((\Omega')^n, \mathcal{A}'_n, \mathbb{P}_\theta, \theta \in \Theta)$*, où* $\Theta \subset \mathbb{R}^p$*. La statistique de Wald* $\widehat{T}$ *pour le test* $H_0 \ : \theta = \theta_0$ *contre* $H_1 \ : \theta \in \Theta_1$ *est :* $\widehat{T}_n = n \cdot {}^t(\widehat{\theta}_n - \theta) \cdot I(\theta) \cdot (\widehat{\theta}_n - \theta)$.

Pour montrer "théoriquement" la pertinence de ce test, on va donc considérer la suite de tests $(\widehat{T}_n)$ en se plaçant dans le "grand" modèle asymptotique :

Théorème. *Dans le cadre d'un modèle paramétrique* $((\Omega')^{\mathbb{N}}, \mathcal{A}'_{\mathbb{N}}, (f_\theta \cdot d\mu)^{\otimes \mathbb{N}}, \theta \in \Theta)$*, où* $\Theta \subset \mathbb{R}^p$*, dominé par une mesure* μ *et régulier, pour le problème de test* $H_0 \ : \theta = \theta_0$ *contre* $H_1 \ : \theta \neq \theta_0$*, alors, en notant* $\widehat{T}_n$ *la statistique de test de Wald pour le modèle projeté de taille* n *sous l'hypothèse* H_0*,*

$$\widehat{T}_n \xrightarrow[n \to +\infty]{\mathcal{L}} \chi^2(p).$$

La région de rejet asymptotique du test sera donc de la forme $\widehat{T}_n > s_\alpha$*, où* s_α *est le quantile d'ordre* $1 - \alpha$ *de la loi du* $\chi^2(p)$*. La suite de test* $(\widehat{T}_n)_n$ *a donc une puissance qui tend vers* 1 *lorsque* α *est fixé.*

<u>Démonstration</u> : La loi asymptotique de $\widehat{\theta}_n$ induit la loi asymptotique de $\widehat{T}_n$, car $\sqrt{n} \cdot I(\theta)^{1/2} \cdot (\widehat{\theta}_n - \theta)$ suit asymptotiquement une loi $\mathcal{N}(0, I_d)$ sous l'hypothèse H_0 et $\widehat{T}_n = \| \sqrt{n} \cdot I(\theta)^{1/2} \cdot (\widehat{\theta}_n - \theta) \|^2$. ■

Voici donc un premier type de test, qui sous certaines conditions de régularités du modèle et pour certaines hypothèses de tests est intéressant. Mais pourrait-on faire mieux ? Et en quel sens ? Désormais, il nous faut donc définir un moyen de comparaison entre deux tests.

4.3 Test du rapport de vraisemblance

Définition. *Sous les hypothèses et notations précédentes, on dira qu'un test* ϕ *est uniformément le plus puissant (U.P.P.) au seuil* α *si le niveau de* $\widehat{\phi}$ *associé à la statistique* $\widehat{T}$ *est inférieur ou égal à* α *et si pour tout autre test* $\widehat{\phi}'$ *associé à la statistique* $\widehat{T}'$ *de niveau inférieur ou égal à* α*,* $\forall \theta \in \Theta_1$*,*

$$\mathbb{E}_\theta(\widehat{\phi}) = 1 - \mathbb{P}_\theta(\widehat{T} \notin W) \leq 1 - \mathbb{P}_\theta(\widehat{T}' \notin W') = \mathbb{E}_\theta(\widehat{\phi}').$$

Définition. *Sous les hypothèses précédentes, si* $L_\theta(.)$ *est la vraisemblance, on appellera test du rapport de vraisemblance (test de Neyman-Person dans le cas d'hypothèses simples) un test de statistique* $\widehat{T}$ *telle que :*

$$\widehat{T} = \frac{\sup_{\theta \in \Theta_0} L_\theta(X_1, \cdots, X_n)}{\sup_{\theta \in \Theta_1} L_\theta(X_1, \cdots, X_n)}.$$

La région critique W *associée à un tel test est de la forme* $W =]+\infty, K[$ *(donc si* $\widehat{T} < K$*, on rejette* H_0*).*

Une des vertus du test du rapport de vraisemblance par rapport au test de Wald est qu'il peut être utilisé dans un modèle non régulier (mais la question de sa loi, ou de la loi d'une fonctionnelle de ce test, demeure). De plus, la propriété suivante confirme l'intérêt de cette statistique de test :

Propriété (Principe de Lehmann)**.** *Dans le cas du test de deux hypothèses simples, ou d'un test unilatéral (*$\Theta \subset \mathbb{R}$*), ce test est U.P.P. Dans le cas d'un test bilatéral, il n'existe pas forcément de test U.P.P.*

<u>Démonstration</u> : ■

Enfin, un tel test pour un modèle régulier, va pouvoir être traité de manière générale grâce à la normalité asymptotique de l'estimateur du maximum de vraisemblance :

Théorème. *Dans le cadre d'un modèle paramétrique* $((\Omega')^{\mathbb{N}}, \mathcal{A}'_{\mathbb{N}}, (f_\theta \cdot d\mu)^{\otimes \mathbb{N}}, \theta \in \Theta)$*, où* $\Theta \subset \mathbb{R}^p$*, dominé par une mesure* μ *et régulier, pour le problème de test* $H_0 \; : \theta = \theta_0$ *contre* $H_1 \; : \theta \neq \theta_0$*, alors, en notant* $\widehat{T}_n$ *la statistique du rapport de vraisemblance pour le modèle projeté de taille* n*,*

$$-2\log(\widehat{T}_n) \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2(p).$$

La région de rejet asymptotique du test sera donc de la forme $-2\log(\widehat{T}_n) > s_\alpha$*, où* s_α *est le quantile d'ordre* $1-\alpha$ *de la loi du* $\chi^2(p)$*. La suite de test* $(\widehat{T}_n)_n$ *a donc une puissance qui tend vers 1 lorsque* α *est fixé.*

Démonstration : la démonstration reprend un peu celle de la normalité asymptotique du maximum de vraisemblance. ■

5 Introduction à la statistique non-paramétrique

On se place donc un modèle semi-paramétrique, soit $((\Omega')^n, \mathcal{A}'_n, P_{(\theta,f)}, \theta \in \Theta, f \in \mathcal{F})$, où $\Theta \subset \mathbb{R}^p$ et $\mathcal{F}$ n'est pas de dimension finie, ou dans un modèle non-paramétrique, soit $((\Omega')^n, \mathcal{A}'_n, P_f, f \in \mathcal{F})$, où $\mathcal{F}$ n'est pas de dimension finie. **Dans un tel cadre, les estimations et tests construits à partir de la vraisemblance ne sont plus directement utilisables.**

5.1 Estimation semi-paramétrique et non-paramétrique

• Quelques estimations semi-paramétriques

Soit $(X_1, \cdots, X_n)$ un n-échantillon issu d'une suite de v.a.i.i.d. réelles.

1. Estimation de la moyenne : si on suppose que $\mathbb{E}X_i = m$ existe, un estimateur convergent (grâce à la L.F.G.N.) de m est : $\overline{X}_n = \dfrac{1}{n}\sum_{j=1}^n X_j$. De plus, si $\mathrm{var}X_i < +\infty$, le T.L.C. de Slutsky montre qu'avec $\overline{\sigma}_n^2 = \dfrac{1}{n}\sum_{j=1}^n (X_j - \overline{X}_n)^2$,
$$\sqrt{n}\,\frac{(\overline{X}_n - m)}{\overline{\sigma}_n} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$
Ceci permet d'obtenir des intervalles de confiance sur m.
2. Estimation de la variance : si on suppose que $\mathrm{var}X_i = \sigma^2$ existe, un estimateur convergent (grâce à la L.F.G.N.) de σ^2 est : $\overline{\sigma}_n^2 = \dfrac{1}{n}\sum_{j=1}^n (X_j - \overline{X}_n)^2$. De plus, si $\mathbb{E}X_i^4 < +\infty$, le T.L.C. de Slutski montre qu'avec $\overline{\mu}_{4,n} = \dfrac{1}{n}\sum_{j=1}^n ((X_j - \overline{X}_n)^2 - \overline{\sigma}_n^2)^2$,
$$\sqrt{n}\,\frac{(\overline{\sigma}_n^2 - \sigma^2)}{\sqrt{\overline{\mu}_{4,n}}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$
Ceci permet d'obtenir des intervalles de confiance sur σ^2.

• Quelques estimations non-paramétriques

Soit $(X_1, \cdots, X_n)$ un n-échantillon issu d'une suite de v.a.i.i.d. réelles.

1. Estimation de la fonction de répartition :
$$F_n(x) = \frac{1}{n}\sum_{j=1}^n \mathbb{I}_{X_j \leq x},$$
pour $x \in \mathbb{R}$ est un estimateur convergent de $F(x)$.

2. Estimation de la densité par histogramme :

   - si X est une v.a. discrète, on utilise les valeurs $(x_i)_{i\in I}$ possibles des X_j, et

   $$f_n(x) = \sum_{i\in I} \frac{1}{x_{i+1}-x_i} \mathbb{I}_{x\in[x_i,x_{i+1}[} \left(\frac{1}{n}\sum_{j=1}^n \mathbb{I}_{x_i\le X_j<x_{i+1}} \right), \quad \text{pour } x\in \mathbb{R}.$$

   - si X est une v.a. continue, on découpe $\mathbb{R}$ en m-classe **a priori**, $[x_i, x_{i+1}]$. On estime f par

   $$f_n(x) = \sum_{i=1}^{m} \frac{1}{x_{i+1}-x_i} \mathbb{I}_{x\in[x_i,x_{i+1}]} \left(\frac{1}{n}\sum_{j=1}^n \mathbb{I}_{x_i\le X_j\le x_{i+1}} \right), \quad \text{pour } x\in \mathbb{R}.$$

3. Estimation de la densité par noyau pour une variable **continue**. On utilise un noyau $K(x)$ qui vérifie :

   - $K(x)\ge 0$ pour $x\in\mathbb{R}$ et $K(0)>0$.
   - $\int_{\mathbb{R}} K(u)du = 1$ et $\int_{\mathbb{R}} K^2(u)du < +\infty$.

   Par exemple, $K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (dit "noyau gaussien") ou $K(x) = (1-|x|)\cdot \mathbb{I}_{|x|\le 1}$ (fonction "triangle"). On définit alors un estimateur de f, densité de X_j par rapport à la mesure de Lebesgue sur $\mathbb{R}$, par

   $$f_{n,h}(x) = \frac{1}{n}\sum_{j=1}^n \frac{1}{h} K\left(\frac{X_j - x}{h}\right),$$

   pour $x\in\mathbb{R}$ et $h>0$. Si $h=h_n$ avec $h_n\to 0$ et $n\cdot h_n \to +\infty$ quand $n\to\infty$, alors $f_n(x)$ est un estimateur convergent de $f(x)$.

5.2 Tests semi et non-paramétriques

Quelques tests semi-paramétriques

- Test sur la moyenne (ou test de Student) : on suppose un échantillon $(X_1,\cdots,X_n)$ issu d'une suite (X_i) de v.a.i.i.d. d'espérance m et de variance finie. On veut tester $H_0 \ : \ m = m_0$ contre $H_1 \ : \ m\in\Theta_1$. Alors la statistique

  $$T_n = \sqrt{n}\,\frac{\overline{X}_n - m}{\widehat{\sigma}_n}$$

  est la statistique du test de Student, et :

  1. pour tout $n\ge 2$, on a $T_n \sim t(n-1)$ si les X_i sont gaussiennes (auquel cas le modèle est paramétrique);
  2. $T_n \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$ (dans le cadre semi-paramétrique).

- Test de comparaison de moyenne (ou test de Student) : on suppose deux échantillons $(X_1,\cdots,X_n)$ et $(Y_1,\cdots,Y_n)$ issus de suites (X_i) et Y_i de v.a.i.i.d. d'espérance m_X et m_Y et de même variance σ finie. On veut tester $H_0 \ : \ m_X = m_Y$ contre $H_1 \ : \ m_X - m_Y \in\Theta_1$. Alors la statistique :

  $$T_n = \sqrt{n-1}\,\frac{(\overline{X}_n - \overline{Y}_n) - (m_X - m_Y)}{\sqrt{\widehat{\sigma}_X^2 + \widehat{\sigma}_Y^2}}$$

  est la statistique du test de comparaison de Student, et :

  1. pour tout $n\ge 2$, on a $T_n \sim t(2n-2)$ si les X_i sont gaussiennes (auquel cas le modèle est paramétrique);;
  2. $T_n \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$ (dans le cadre semi-paramétrique).

- Test sur la variance : on suppose un échantillon $(X_1, \cdots, X_n)$ issu d'une suite (X_i) de v.a.i.i.d. de variance σ^2 et de moment d'ordre 4 fini. On veut tester $H_0 \ : \ \sigma^2 = \sigma_0$ contre $H_1 \ : \ \sigma^2 \in \Theta_1$. Alors la statistique de test,
$$T_n = \sqrt{n}\,\frac{(\widehat{\sigma}_n^2 - \sigma^2)}{\sqrt{\overline{\mu}_{4,n}}},$$
où $\overline{\mu}_{4,n} = \frac{1}{n}\sum_{j=1}^{n}((X_j - \overline{X}_n)^2 - \overline{\sigma}_n^2)^2$ est telle que $T_n \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1)$, d'où l'utilisation du test.

Quelques tests non-paramétriques

- Test de signe : on suppose n individus dont on peut connaître les valeurs de deux variables indépendantes X et Y, donc deux échantillons $(X_1, \cdots, X_n)$ et $(Y_1, \cdots, Y_n)$ de v.a.i.i.d. On veut tester $H_0 \ : \ P(X_i > Y_i) = P(X_i < Y_i)$ contre $H_1 \ : \ P(X_i < Y_i) > 1/2$. Pour cela, on considère la statistique :
$$T_n = \sum_{i=1}^{n} \mathbb{I}_{X_i < Y_i}.$$
Sous l'hypothèse H_0, $T_n \sim \mathcal{B}(n, 1/2)$, d'où le test.

- Test de signe et de rang : on suppose n individus dont on peut connaître les valeurs de deux variables indépendantes X et Y, donc deux échantillons $(X_1, \cdots, X_n)$ et $(Y_1, \cdots, Y_n)$ de v.a.i.i.d. On veut tester $H_0 \ : \ P(X_i > Y_i) = P(X_i < Y_i)$ contre $H_1 \ : \ P(X_i < Y_i) > 1/2$. Pour cela, on considère la statistique :
$$T_n = \sum_{i=1}^{n} \mathrm{rang}(|X_i - Y_i|) \cdot \mathbb{I}_{X_i < Y_i},$$
où $\mathrm{rang}(|X_i - Y_i|)$ est le classement dans l'ordre croissant (du plus petit au plus grand) des différentes valeurs de $|X_i - Y_i|$. Sous l'hypothèse H_0, on peut montrer que $T_n \sim \mathcal{L}_n$, où $\mathcal{L}_n$ est une loi dont on connaît les quantiles. Ceci permet l'utilisation du test.

- Test d'ajustement : Tests du χ^2 et tests de Kolmogorov-Smirnov.

Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2005 – 2006

Statistiques I

Contrôle continu n°1, novembre 2005

Examen de 2 h 00. Tout document ou calculatrice est interdit.

1. On considère une suite $(X_i)_{i\in\mathbb{N}}$ de variables aléatoires indépendantes et identiquement distribuées suivant une loi $\mathcal{N}(m,\sigma^2)$, où $m\in\mathbb{R}$ et $\sigma^2>0$ sont des paramètres inconnus. Soit également pour $n\in\mathbb{N}^*$, $\overline{X}_n=\dfrac{1}{n}(X_1+\cdots+X_n)$ et $\widehat{\sigma}_n^2=\dfrac{1}{n}\displaystyle\sum_{j=1}^n(X_j-\overline{X}_n)^2$.

   (a) Pour n fixé, déterminer le modèle statistique paramétrique.

   (b) Pour $n\in\mathbb{N}^*$, quelle est la loi de $\overline{X}_n$?

   (c) Quelles sont les limites (en probabilité et en loi) de $\overline{X}_n$ et de $\widehat{\sigma}_n^2$ (justifier...) ?

   (d) Montrer que la connaissance de $(\overline{X}_1,\cdots,\overline{X}_n)$ induit celle de $(X_1,\cdots,X_n)$. Déterminer la loi du vecteur $(\overline{X}_1,\cdots,\overline{X}_n)$. Les $(\overline{X}_k)$ sont-elles indépendantes ?

   (e) Soit $\overline{\overline{X}}_n=\dfrac{1}{n}(\overline{X}_1+\cdots+\overline{X}_n)$. Quelle est la loi de $\overline{\overline{X}}_n$ pour $n\in\mathbb{N}^*$? En déduire que $\overline{\overline{X}}_n \xrightarrow[n\to+\infty]{\mathcal{P}} m$. Montrer également que $\overline{\overline{X}}_n \xrightarrow[n\to+\infty]{p.s.} m$.

   (f) Comment peut-on faire pour savoir quelle suite de variables aléatoires, $(\overline{X}_k)_{k\geq1}$ ou $(\overline{\overline{X}}_k)_{k\geq1}$, s'approche le plus vite de m ? Conclusion ?

   (g) Pour le modèle paramétrique de taille n où σ^2 est supposé connue, montrer que la statistique $\overline{X}_n$ est exhaustive complète. Et la statistique $(\overline{X}_1,\cdots,\overline{X}_n)$? Enfin, la statistique $\overline{\overline{X}}_n$ est-elle exhaustive ?

2. En fait, on ne connaît pas explicitement chaque X_i, mais plutôt pour tout $k\in\mathbb{N}^*$, $T_k=\max(X_1,\cdots,X_k)$.

   (a) La connaissance de $(T_1,\cdots,T_n)$ induit-elle celle de $(X_1,\cdots,X_n)$?

   (b) Déterminer la fonction de répartition F_k de T_k, puis, après avoir montré son existence, sa densité f_k par rapport à la mesure de Lebesgue, le tout en fonction de la fonction de répartition F et de la densité f de X_1.

   (c) Déterminer, en justifiant, le comportement asymptotique (quand $n\to\infty$) de $(T_n)_n$.

   (d) Pour $k\in\mathbb{N}^*$, montrer que T_k et T_{k+1} ne sont pas indépendantes. Montrer que $\mathbb{P}(T_{k+1}=T_k)=\dfrac{k}{k+1}$. En déduire la mesure de probabilité de la variable $T_{k+1}-T_k$. La loi de probabilité de la variable T_k est-elle "continue" ? "Discrète" ?

   (e) La statistique T_n est-elle exhaustive pour le modèle paramétrique de taille n où σ^2 est supposé connue ? Et la statistique $(T_1,\cdots,T_n)$?

Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2005 – 2006

Statistiques I

Contrôle continu n°2, janvier 2006

Examen de 2 h 00. Tout document ou calculatrice est interdit.

1. Soit la variable X qui suit une loi dont la densité f_X par rapport à la mesure de Lebesgue sur $]0,1]$ est, avec β et $K \in \mathbb{R}$:
$$f_X(x) = K \cdot x^\beta \quad \text{pour tout } x \in]0,1],$$
   (a) Déterminer K en fonction de β, en précisant quelle condition doit vérifier β. En déduire $\mathbb{E}(X)$ et $\mathrm{var}(X)$, en précisant également des conditions sur β.
   (b) On suppose que la suite $(X_i)_{i\in\mathbb{N}}$ est constituée de variables aléatoires indépendantes et identiquement distribuées suivant la même loi que X. Soit un échantillon observé $(X_1, \ldots, X_n)$. On désire estimer β à partir de cet échantillon. Quel est le modèle statistique ? Montrer que ce modèle appartient à la famille exponentielle.
   (c) En déduire qu'il n'existe pas d'estimateur sans biais efficace de β.
   (d) Montrer que que $-\log(X)$ suit une loi connue dont on précisera le paramètre. En déduire que $\tilde{\beta}_n = -1 - n \cdot \Big(\sum_{i=1}^n \log(X_i)\Big)^{-1}$ est un estimateur sans biais de β (utiliser les lois gammas...), puis qu'il est de variance uniformément minimale parmi les estimateurs sans biais (Lehmann-Scheffé...).

2. Soit Y une variable suivant une loi de Bernoulli de paramètre $p \in]0,1[$ et indépendante de X. On définit une variable Z de la manière suivante : si $Y = 1$, alors $Z = X$, et si $Y = 0$ alors $Z = -X$.
   (a) Montrer que Z suit une loi absolument continue par rapport à la mesure de Lebesgue sur $[-1,0[\cup]0,1]$ et que sa densité f_Z est :
$$f_Z(z) = (\beta+1)\cdot|x|^\beta\Big(p\cdot \mathbb{I}_{x\in]0,1]} + (1-p)\cdot \mathbb{I}_{x\in[-1,0[}\Big) \quad \text{pour tout } x \in [-1,0[\cup]0,1].$$
   Calculer $\mathbb{E}(Z)$ et $\mathrm{var}(Z)$ (en précisant les conditions sur β).
   (b) On suppose que la suite $(Z_i)_{i\in\mathbb{N}}$ est constituée de variables aléatoires indépendantes et identiquement distribuées suivant la même loi que Z. Soit un échantillon observé $(Z_1, \ldots, Z_n)$. On désire estimer (β, p) à partir de cet échantillon. Quel est le modèle statistique ? Montrer que ce modèle appartient à la famille exponentielle.
   (c) En déduire une statistique exhaustive minimale complète pour ce modèle. Déterminer la matrice d'information de Fisher du modèle, puis la borne de Cramer-Rao. Déterminer une fonction g de (β, p) que l'on peut estimer sans biais et de manière efficace.

(d) Déterminer, après avoir montré son unicité, l'estimateur $(\widehat{\beta}_n, \widehat{p}_n)$ du maximum de vraisemblance de (β, p). Les estimateurs $\widehat{\beta}_n$ et $\widehat{p}_n$ sont-ils indépendants ? Déterminer un théorème de la limite centrale vérifié par $(\widehat{\beta}_n, \widehat{p}_n)$. Est-ce un estimateur asymptotiquement efficace ?

(e) Déterminer une région de confiance de niveau 95% sur (β, p), en utilisant 1/ l'estimateur efficace de $g(\beta, p)$ 2/ l'estimateur de maximum de vraisemblance dans un cadre asymptotique.

Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2005 – 2006

Statistiques I

Examen terminal, janvier 2006

Examen de 3 h 00. Tout document ou calculatrice est interdit.

1. On considère $(X_k)_{k\in\mathbb{N}}$ et $(X'_k)_{k\in\mathbb{N}}$ deux suites indépendantes de variables aléatoires définies sur le même espace de probabilité, indépendantes et identiquement distribuées suivant les lois respectives $\mathcal{N}(\mu,\sigma^2)$ (pour les X_i) et $\mathcal{N}(\mu',\sigma^2)$ (pour les X'_i), où $(\mu,\mu')\in\mathbb{R}^2$ et $\sigma^2>0$. Le but du problème est de tester si $\mu=\mu'$ à partir d'un échantillon de chacune de ces suites.

   Soit $(X_1,\dots,X_n)$ et $(X'_1,\dots,X'_{n'})$, où $n\in\mathbb{N}^*$ et $n'\in\mathbb{N}^*$, deux échantillons issus de $(X_k)_{k\in\mathbb{N}}$ et $(X'_k)_{k\in\mathbb{N}}$. On pose $Z=(Z_1,\dots,Z_{n+n'})=(X_1,\dots,X_n,X'_1,\dots,X'_{n'})$

   (a) Déterminer le modèle statistique associé à $Z=(Z_1,\dots,Z_{n+n'})$.

   (b) Montrer que ce modèle est exponentiel. En déduire que μ et μ' peuvent être estimés efficacement (on notera $\widehat{\mu}$ et $\widehat{\mu}'$ leurs estimateurs respectifs).

   (c) Montrer qu'un estimateur du maximum de vraisemblance de σ^2 est $\widehat{\sigma}^2$, avec :

   $$\widehat{\sigma}^2=\frac{1}{n+n'}\Big(\sum_{j=1}^{n}\big(Z_j-\widehat{\mu}\big)^2+\sum_{j=n+1}^{n+n'}\big(Z_j-\widehat{\mu}'\big)^2\Big).$$

   Est-ce un estimateur biaisé de σ^2 ? Est-il convergent ? Efficace ?

   (d) Lorsque n et n' sont "grands", déduire de ce qui précède, des intervalles de confiance à 95% pour μ et μ'.

   (e) Soit le problème de test :

   $$H_0\ :\ \mu=\mu'\qquad\text{contre}\qquad H_1\ :\ \mu\neq\mu'$$

   (σ^2 restant inconnu). Démontrer que la statistique $\widehat{T}$ du rapport de vraisemblance vérifie :

   $$\widehat{T}=\left(\frac{\widehat{\sigma}^2}{\overline{\sigma}^2}\right)^{(n+n')/2}\qquad\text{avec}\qquad\begin{cases}\overline{\sigma}^2 &= \dfrac{1}{n+n'}\displaystyle\sum_{j=1}^{n+n'}\big(Z_j-\overline{Z}_{n+n'}\big)^2\\[2mm] \overline{Z}_{n+n'} &= \dfrac{1}{n+n'}\displaystyle\sum_{i=1}^{n+n'}Z_i\end{cases}.$$

   En déduire que la région d'acceptation du test peut s'écrire sous la forme $\widehat{\sigma}^2\geq K\cdot\overline{\sigma}^2$, avec K dépendant du niveau du test.

(f) Pour déterminer la valeur K en fonction du niveau $1-\alpha$ du test, on peut considérer la statistique, dite de Fisher,

$$\widehat{T}' = \frac{(n+n')\cdot \overline{\sigma}^2 - (n+n')\cdot \widehat{\sigma}^2}{\Big(\dfrac{n+n'}{n+n'-2}\Big)\,\widehat{\sigma}^2}$$

i. Soit les vecteurs de $\mathbb{R}^{n+n'}$, $u_1 = (1,\dots,1)$, $u = (1,\dots,1,0,\dots,0)$ (soit n fois 1 et n' fois 0) et $u' = u_1 - u$. Montrer que u et u' sont orthogonaux.

ii. Sous l'hypothèse H_0, déterminer une expression plus simple de $P_{<u_1>}(Z)$, projeté orthogonal de Z sur le sous-espace vectoriel (s.e.v.) engendré par u_1, et $P_{<u,u'>}(Z)$, projeté orthogonal de Z sur le s.e.v. engendré par u et u'.

iii. Montrer que sous l'hypothèse H_0, le vecteur Z peut s'écrire $Z = \mu \cdot u_1 + \sigma \cdot \varepsilon$, où ε est un vecteur aléatoire gaussien composé de $n + n'$ variables gaussiennes centrées réduites.

iv. Montrer que sous l'hypothèse H_0, $(n + n')\cdot \widehat{\sigma}^2 = \sigma^2 \cdot \|P_A(\varepsilon)\|^2$, où $\|.\|$ est la norme euclidienne classique sur $\mathbb{R}^{n+n'}$ et A est un s.e.v. de $\mathbb{R}^{n+n'}$ que vous préciserez.

v. En utilisant le Théorème de Pythagore, montrer que sous l'hypothèse H_0 et avec B est s.e.v. de $\mathbb{R}^{n+n'}$ que vous préciserez, $(n+n')\cdot\overline{\sigma}^2-(n+n')\cdot\widehat{\sigma}^2 = \sigma^2\cdot\|P_B(\varepsilon)\|^2$.

vi. En utilisant le Théorème de Cochran, montrer que sous l'hypothèse H_0, $\widehat{T}'$ suit une loi de Fisher à $(1,(n + n' - 2))$ degrés de liberté. Lorsque n et n' sont "grands", quelle loi suit approximativement $\widehat{T}'$?

vii. Pour finir, déterminer K en fonction d'un quantile de la loi de Fisher à $(1,(n+n'-2))$ degrés de liberté.

2. Soit X une variable aléatoire dont la mesure de probabilité est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et de densité :

$$f(x) = \frac{\lambda}{2}\exp\big(-\lambda|x-m|\big) \quad \text{pour } x \in \mathbb{R},$$

avec $m \in \mathbb{R}$ et $\lambda > 0$, des paramètres inconnus.

(a) Calculer l'espérance et la variance de X.

(b) Calculer $\mathbb{P}(X = m)$ et $\mathbb{P}(X < m)$. En déduire la médiane (théorique) de la loi de X.

(c) Soit une suite $(X_i)_{i\in\mathbb{N}}$ de variables aléatoires indépendantes et identiquement distribuées suivant la même loi que X, dont on extrait un échantillon observé $(X_1,\dots,X_{2n+1})$. Par ailleurs, on note $X_{(1)} \le X_{(2)} \le \dots \le X_{(2n+1)}$ la statistique d'ordre associée. Soit :

$$\widehat{H}_n(a) = \frac{1}{2n+1}\sum_{i=1}^{2n+1}|X_i - a| \quad \text{pour } a \in \mathbb{R}.$$

Calculer $\widehat{H}_n(X_{(n+1)})$ en fonction des $X_{(i)}$. Montrer que la fonction $a \mapsto \widehat{H}_n(a)$ est minimale en $X_{(n+1)}$ (on pourra développer $\widehat{H}_n(X_{(n+k)})$ en fonction des $X_{(i)}$ pour $k > 1$).

(d) On suppose ici que $m = 1$, donc que m est connu ($\lambda > 0$ restant inconnu). Quel est alors le modèle statistique ? Montrer que ce modèle appartient à la famille exponentielle, et en déduire une statistique exhaustive dont vous montrerez qu'elle est complète. Déterminer la matrice d'information de Fisher du modèle. Quelle est la fonction de λ (à une transformation affine près) que l'on peut estimer efficacement ? Déterminer l'estimateur de maximum de vraisemblance de λ et montrer qu'il vérifie un théorème de la limite centrale.

(e) On suppose désormais que $m \in \mathbb{R}$ est inconnu, tout comme $\lambda > 0$. Quel est alors le modèle statistique ? Montrer que ce modèle n'appartient pas à la famille exponentielle. A l'aide de la question 2.(c), déterminer un estimateur $(\widehat{m}_n, \widehat{\lambda}_n)$ du maximum de vraisemblance du couple (m, λ).

(f) Pour $a \in \mathbb{R}$, démontrer que $\widehat{H}_n(a)$ converge presque sûrement quand $n \to \infty$ vers $\mathbb{E}(|X-a|)$. Montrer que la fonction $a \in \mathbb{R} \mapsto \mathbb{E}(|X-a|)$ est minimale en $a = m$. En déduire que $\widehat{m}_n \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} m$, puis que $\widehat{\lambda}_n \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} \lambda$.

Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2005 – 2006

Statistiques I

Examen de septembre 2006

Examen de 3 h 00. Tout document ou calculatrice est interdit.

1. On considère $(X_k)_{k\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant la loi $\mathcal{N}(m,1)$, où $m \in \mathbb{R}$. Soit la suite de variables $(Y_k)_{k\in\mathbb{N}}$ telle que pour tout $k \in \mathbb{N}^*$,
$$Y_k = X_1 + \cdots + X_k.$$
Soit $(Y_1,\dots,Y_n)$, où $n \in \mathbb{N}^*$ un échantillon issus de $(Y_k)_{k\in\mathbb{N}}$.
   (a) Déterminer la loi de Y_k pour $k \in \mathbb{N}^*$.
   (b) Déterminer la loi du vecteur $(Y_1,\dots,Y_n)$. Montrer que pour $i \neq j$, Y_i n'est pas indépendante de Y_j.
   (c) Déterminer le modèle statistique associé à $(Y_1,\dots,Y_n)$.
   (d) Montrer que ce modèle est exponentiel.
   (e) Soit J_n la matrice de covariance de $(Y_1,\dots,Y_n)$. Vérifier que
$$J_n^{-1} = \begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}.$$
En déduire que m peut être estimé par un estimateur $\widehat{m}$ (que l'on précisera) sans biais et efficace.
   (f) Déterminer l'estimateur du maximum de vraisemblance de m. Est-ce un estimateur biaisé ? Est-il convergent ? Efficace ? Quel est son risque quadratique ?
   (g) Déterminer la statistique du test de rapport de vraisemblance pour le test
$$H_0 \ : \ m = m_0 \qquad \text{contre} \qquad H_1 \ : \ m \neq m_0,$$
où m_0 est une constante connue. On réalise une application numérique de ce test au niveau 5% pour $n = 100$. On trouve que $\widehat{m} = m_0 + 1$. Accepte-t-on alors l'hypothèse H_0 ?

2. Soit X une variable aléatoire dont la mesure de probabilité est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et de densité :
$$f(x) = k \cdot \frac{1}{\sqrt{x}} \cdot \mathbb{I}_{0\leq x\leq\theta} \quad \text{pour } x \in \mathbb{R},$$
avec $k \in \mathbb{R}$ et $\theta > 0$, des paramètres inconnus.

(a) Déterminer l'expression de k en fonction de θ. Calculer l'espérance et la variance de X.

(b) Soit une suite $(X_i)_{i\in\mathbb{N}}$ de variables aléatoires indépendantes et identiquement distribuées suivant la même loi que X, dont on extrait un échantillon observé $(X_1,\dots,X_n)$. Par ailleurs, on note $X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$ la statistique d'ordre associée. On désire estimer θ à partir $(X_1,\dots,X_n)$. Quel est alors le modèle statistique ? Quel est la vraisemblance L_θ du modèle ?

(c) Montrer que $\widehat{T}_n = X_{(n)} = \max(X_1,\dots,X_n)$ est une statistique exhaustive pour ce modèle.

(d) Montrer que cette statistique est minimale.

(e) Soit $(x_1,\dots,x_n)\in\mathbb{R}^n$ et le même n-uplet ordonné $\min(x_i) = x_{(1)} \le \cdots \le x_{(n)} = \max(x_i)$. Montrer que :

$$\mathbb{P}(X_{(1)} \le x_{(1)},\dots,X_{(n)} \le x_{(n)}) = n!\cdot\mathbb{P}(X_1 \le x_{(1)} \cap X_1 \le X_2 \le x_{(2)} \cap\cdots\cap X_{n-1} \le X_n \le x_{(n)}).$$

Montrer par itération sur les dérivées partielles que :

$$\frac{\partial^n}{\partial x_{(1)}\dots\partial x_{(n)}}\mathbb{P}(X_1 \le x_{(1)} \cap X_1 \le X_2 \le x_{(2)} \cap\cdots\cap X_{n-1} \le X_n \le x_{(n)}) = \prod_{i=1}^n f(x_{(i)}).$$

Soit $L_\theta^{(n)}$ la vraisemblance de $(X_{(1)},\dots,X_{(n)})$. Déduire de ce qui précède que :

$$L_\theta^{(n)}(x_{(1)},\dots,x_{(n)}) = n!\cdot L_\theta(x_1,\dots,x_n).$$

(f) Déterminer la densité puis le biais de $\widehat{T}_n$.

(g) Montrer que $\widehat{T}_n$ est une statistique exhaustive et complète.

(h) Déduire de ce qui précède un estimateur $\widehat{T}'_n$ de θ, sans biais et uniformément de variance minimale.

(i) Calculer le risque quadratique de $\widehat{T}'_n$ et en déduire que $\widehat{T}'_n \xrightarrow[n\to+\infty]{\mathcal{P}} \theta$, puis, que pour tout $\alpha\in[0,1[$, $n^\alpha(\widehat{T}'_n-\theta) \xrightarrow[n\to+\infty]{\mathcal{P}} 0$.

(j) Déterminer explicitement un intervalle de confiance à 95% de θ en fonction de $\widehat{T}'_n$.

Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2006 – 2007

Statistiques I

Contrôle continu n°1, novembre 2006

Examen de 2 h 00. Tout document ou calculatrice est interdit.

1. On considère une suite $(\varepsilon_i)_{i\in\mathbb{N}}$ de variables aléatoires indépendantes et identiquement distribuées suivant une loi $\mathcal{N}(0,\sigma^2)$, où $\sigma^2>0$ est un paramètre inconnu. Pour tout $n\in\mathbb{N}^*$, on définit :
$$X_n=\varepsilon_n-\alpha\varepsilon_{n-1},$$
où $\alpha\in\mathbb{R}$ est inconnu. On notera par la suite $\theta=(\sigma^2,\alpha)$.
   (a) Montrer que $\mathrm{var}(\varepsilon_i^2)=2\sigma^4$.
   (b) Pour tout $i\in\mathbb{N}^*$, déterminer $\mathbb{E}(X_i)$ et $\mathrm{var}(X_i)$. Montrer que $(X_i)_i$ est une suite de variables identiquement distribuées dont vous préciserez la loi.
   (c) Montrer que $\mathrm{cov}(X_i,X_j)=(1+\alpha^2)\sigma^2$ si $i=j$, $\mathrm{cov}(X_i,X_j)=-\alpha\sigma^2$ si $|i-j|=1$ et $\mathrm{cov}(X_i,X_j)=0$ sinon.
   (d) Pour n fixé, en déduire la loi du vecteur $(X_1,\dots,X_n)$, puis déterminer le modèle statistique paramétrique associé en précisant une mesure dominante.
   (e) Soit $\widehat{\sigma}_n^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i^2$. Montrer que $\widehat{\sigma}_n^2$ est un estimateur non biaisé de σ^2. Soit
$$Z_n^{(1)}=\frac{1}{n}\sum_{k=1}^{[n/2]}X_{2k}^2\quad\text{et}\quad Z_n^{(2)}=\frac{1}{n}\sum_{k=1}^{[(n+1)/2]}X_{2k-1}^2,$$
avec $[x]$ la partie entière de x. Montrer que les suites de variables $(Z_n^{(1)})_n$ et $(Z_n^{(2)})_n$ convergent presque sûrement (préciser leurs limites) et qu'elles vérifient un théorème de la limite centrale.
   (f) Montrer que si deux suites de variables convergent presque sûrement, la suite composée de leurs sommes converge presque sûrement. En déduire que $(\widehat{\sigma}_n^2)_n$ converge presque sûrement vers σ^2.
   (g) Soit $\widehat{\rho}_n=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n-1}X_iX_{i+1}$. En utilisant le même type d'argument que précédemment, montrer que $(\widehat{\rho}_n)_n$ converge presque sûrement vers $-\alpha\sigma^2$. En déduire un estimateur de α convergeant presque sûrement.

2. Soit X une variable aléatoire dont la loi est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et telle que sa densité par rapport à cette mesure soit :
$$f_{\alpha,a}(x)=K\cdot\frac{1}{|x|^\alpha}\mathbb{I}_{-a\le x\le a},$$

avec $a \in \mathbb{R}$ et $\alpha \in \mathbb{R}$. Soit $(X_1, \ldots, X_n)$ un échantillon de n v.a.i.i.d. de même loi que X.

(a) Après avoir préciser l'ensemble des valeurs Θ pour $\theta = (a, \alpha)$, déterminer K en fonction de a et α.

(b) Calculer $\mathbb{E}(X)$ après avoir vérifier que ce calcul peut être effectué pour $\theta \in \Theta$.

(c) Quel est le modèle statistique associé à $(X_1, \ldots, X_n)$.

(d) Montrer que $\widehat{S} = (|X_1|, \ldots, |X_n|)$ est une statistique exhaustive. Montrer que pour $n \geq 3$ cette statistique n'est pas minimale.

(e) Déterminer une statistique $\widehat{T} = (\widehat{T}_1, \widehat{T}_2)$ à valeurs dans $[0, +\infty[^2$ qui soit exhaustive minimale pour tout $n \in \mathbb{N}$.

Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2006 – 2007

Statistiques I

Contrôle continu n°2, janvier 2007

Examen de 2 h 00. Tout document ou calculatrice est interdit.

1. Soit X une variable aléatoire suivant la loi suivante:

$$\mathbb{P}(X=1)=\mathbb{P}(Y=-1)=p \quad \text{et} \quad \mathbb{P}(X=0)=1-2p,$$

où p est un paramètre réel inconnu.

(a) Déterminer l'ensemble des valeurs possibles pour p. Calculer EX et $\mathrm{var}X$.

(b) On suppose que la suite $(X_i)_{i\in\mathbb{N}}$ est constituée de variables aléatoires indépendantes et identiquement distribuées suivant la même loi que X. Soit un échantillon $(X_1,\ldots,X_n)$. Déterminer le modèle statistique associé à cet échantillon et déterminer une mesure dominant ce modèle. Montrer que le modèle appartient à la famille exponentielle. En déduire une statistique exhaustive complète pour ce modèle. Montrer que p peut être estimé efficacement et donner un tel estimateur. Calculer la borne de Cramer-Rao et vérifier qu'elle est bien atteinte par cet estimateur.

(c) On définit la suite $(Y_i)_{i\in\mathbb{N}^*}$ à partir de $(X_i)_{i\in\mathbb{N}}$ de la manière suivante:

$$Y_{i+1}=X_i\cdot X_{i+1} \quad \text{pour} \quad i\in\mathbb{N}.$$

Déterminer la loi de Y_i. Montrer que $\mathrm{cov}(Y_i,Y_{i+1})=0$. Les $(Y_i)_i$ sont-elles indépendantes ?

(d) Montrer que $(|Y_1|,\ldots,|Y_n|)$ est une statistique exhaustive pour le modèle statistique induit par $(Y_1,\cdots,Y_n)$.

2. Soit la variable X qui suit une loi dont la densité f_X par rapport à la mesure de Lebesgue sur $\mathbb{R}$ est, avec $\theta>0$ et $\alpha>0$:

$$f_X(x)=K\cdot x^\alpha \mathbb{I}_{0\le x\le\theta} \quad \text{pour tout } x\in\mathbb{R},$$

(a) Déterminer K en fonction de α et θ.

(b) Montrer que $Y=\log(\theta/X)$ suit une loi exponentielle dont on précisera le paramètre.

(c) On suppose que la suite $(X_i)_{i\in\mathbb{N}}$ est constituée de variables aléatoires indépendantes et identiquement distribuées suivant la même loi que X. Soit un échantillon $(X_1,\ldots,X_n)$. On suppose que (θ,α) est inconnu. Préciser alors le modèle statistique formé par cet échantillon et la mesure dominante. Ce modèle appartient-il à la famille exponentielle ?

(d) Dans cette question, et uniquement dans cette question, on suppose que θ est connu. Préciser alors le modèle statistique. Ce modèle appartient-il à la famille exponentielle ? Montrer que l'estimateur du maximum de vraisemblance $\tilde{\alpha}_n$ de α existe, est unique et s'écrit:

$$\tilde{\alpha}_n = \frac{1}{\frac{1}{n}\sum_{i=1}^n \log(\theta/X_i)} - 1$$

Montrer que $\tilde{\alpha}_n$ converge presque sûrement vers α et qu'il vérifie un théorème de la limite centrale que l'on précisera. En déduire un intervalle de confiance à 95% sur α pour n grand.

(e) Dans cette question, θ et α sont inconnus. Déterminer une statistique exhaustive pour le modèle. En vous aidant de la question précédente, déterminer l'estimateur du maximum de vraisemblance $(\widehat{\theta}_n, \widehat{\alpha}_n)$ de (θ, α). Déterminer la fonction de répartition de $\log(\theta/\widehat{\theta}_n)$ et en déduire que $\widehat{\theta}_n \xrightarrow[n\to+\infty]{\mathcal{P}} \theta$, puis que $\sqrt{n}\log(\theta/\widehat{\theta}_n) \xrightarrow[n\to+\infty]{\mathcal{P}} 0$.

(f) Soit $(U_n)_{n\in\mathbb{N}}$ et $(V_n)_{n\in\mathbb{N}}$ deux suites de variables aléatoires définies sur le même espace de probabilité. Montrer que si $(U_n)_n$ converge vers une loi P_0 et $(V_n)_n$ converge en probabilité vers 0, alors $(U_n + V_n)_n$ converge en loi vers P_0 (on pourra par exemple majorer la différence de fonctions caractéristiques). En déduire que $\widehat{\alpha}_n$ suit le même théorème de la limite centrale que $\tilde{\alpha}_n$.

Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2006 – 2007

Statistiques I

Examen terminal, janvier 2007

Examen de 3 h 00. Tout document ou calculatrice est interdit.

1. On considère une suite de variables aléatoires $(X_k)_{k\in\mathbb{N}}$ définies sur le même espace de probabilité, indépendantes et identiquement distribuées suivant une loi de Bernoulli de paramètre p. On définit:
$$Y = \min\{k \in \mathbb{N}^*,\ X_k = 0\}.$$
   (a) Comment peut-on interpréter la variable Y? Montrer que la loi de Y est:
$$\mathbb{P}(Y = k) = p^{k-1}(1-p) \quad \text{pour} \quad k \in \mathbb{N}^*.$$
   (b) On suppose que $(Y_1, \ldots, Y_n)$ est un échantillon de variables aléatoires indépendantes et identiquement distribuées suivant la loi de Y avec $p \in]0,1[$ est inconnu. Déterminer le modèle statistique et sa mesure dominante. Montrer que ce modèle est exponentiel. En déduire, un estimateur $\widehat{p}_n$ de p sans biais et efficace. Déterminer la borne de Cramer-Rao et vérifier que cette borne est bien atteinte par $\widehat{p}_n$.
   (c) La variable Y est tronquée lorsqu'elle est trop "grande", par un paramètre $T \in \mathbb{N}^*$, c'est-à-dire que l'on définit une variable Y_T telle que $Y_T = \min(Y, T)$.

2. Soit X une variable aléatoire dont la mesure de probabilité est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et de densité :
$$f(x) = \frac{\lambda}{2}\exp\big(-\lambda|x-m|\big) \quad \text{pour } x \in \mathbb{R},$$
avec $m \in \mathbb{R}$ et $\lambda > 0$, des paramètres inconnus.
   (a) Calculer l'espérance et la variance de X.
   (b) Calculer $P(X = m)$ et $P(X < m)$. En déduire la médiane (théorique) de la loi de X.
   (c) Soit une suite $(X_i)_{i\in\mathbb{N}}$ de variables aléatoires indépendantes et identiquement distribuées suivant la même loi que X, dont on extrait un échantillon observé $(X_1, \ldots, X_{2n+1})$. Par ailleurs, on note $X_{(1)} \le X_{(2)} \le \ldots \le X_{(2n+1)}$ la statistique d'ordre associée. Soit :
$$\widehat{H}_n(a) = \frac{1}{2n+1}\sum_{i=1}^{2n+1}|X_i - a| \quad \text{pour } a \in \mathbb{R}.$$
   Calculer $\widehat{H}_n(X_{(n+1)})$ en fonction des $X_{(i)}$. Montrer que la fonction $a \mapsto \widehat{H}_n(a)$ est minimale en $X_{(n+1)}$ (on pourra développer $\widehat{H}_n(X_{(n+k)})$ en fonction des $X_{(i)}$ pour $k > 1$).

(d) On suppose ici que $m = 1$, donc que m est connu ($\lambda > 0$ restant inconnu). Quel est alors le modèle statistique ? Montrer que ce modèle appartient à la famille exponentielle, et en déduire une statistique exhaustive dont vous montrerez qu'elle est complète. Déterminer la matrice d'information de Fisher du modèle. Quelle est la fonction de λ (à une transformation affine près) que l'on peut estimer efficacement ? Déterminer l'estimateur de maximum de vraisemblance de λ et montrer qu'il vérifie un théorème de la limite centrale.

(e) On suppose désormais que $m \in \mathbb{R}$ est inconnu, tout comme $\lambda > 0$. Quel est alors le modèle statistique ? Montrer que ce modèle n'appartient pas à la famille exponentielle. A l'aide de la question 2.(c), déterminer un estimateur $(\widehat{m}_n, \widehat{\lambda}_n)$ du maximum de vraisemblance du couple (m, λ).

(f) Pour $a \in \mathbb{R}$, démontrer que $\widehat{H}_n(a)$ converge presque sûrement quand $n \to \infty$ vers $\mathbb{E}(|X - a|)$. Montrer que la fonction $a \in \mathbb{R} \mapsto \mathbb{E}(|X - a|)$ est minimale en $a = m$. En déduire que $\widehat{m}_n \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} m$, puis que $\widehat{\lambda}_n \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} \lambda$.

Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2006 – 2007

Statistiques I

Examen de septembre 2007

Examen de 3 h 00. Tout document ou calculatrice est interdit.

1. Soit la fonction $f_a(x) = \frac{1}{2}\,(a - a^2 \cdot x) \cdot \mathbb{I}_{\{-1/a \leq x \leq 1/a\}}$ où $a > 0$.
   (a) Démontrer que f_a est une densité de probabilité par rapport à la mesure de Lebesgue et la tracer.
   (b) On suppose que X est une variable aléatoire de densité f_a. Déterminer $\mathbb{E}X$ et $\mathrm{var}X$.
   (c) Soit $(X_k)_{k\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes, telle que la densité de X_n soit f_n pour tout $n \in \mathbb{N}^*$. Déterminer la limite en probabilité de $(X_k)_{k\in\mathbb{N}^*}$ lorsque $n \to \infty$.
   (d) Soit $(X_1, \ldots, X_n)$ un n-échantillon de v.a.i.i.d. de même densité f_a. On suppose que a est inconnu. Déterminer l'estimateur du maximum de vraisemblance $\widehat{a}_n$ pour a. Calculer la fonction de répartition de $\widehat{a}_n$ et en déduire sa convergence en probabilité vers a.
   (e) Pour $\alpha > 0$, déterminer un intervalle de confiance de niveau $1 - \alpha$ pour a.
   (f) Déterminer le test du rapport de vraisemblance de niveau α pour tester l'hypothèse $H_0 \ : \ a = a_0$, contre l'hypothèse $H_1 \ : \ a \neq a_0$ et déterminer la zone d'acceptation du test en fonction de α.
   (g) Proposer un autre estimateur convergent de a.

2. Une compagnie fabrique des piles et s'intéresse à savoir quelle est leur durée de vie moyenne T. Pour ce faire, on considère 1000 piles produites le même jour que l'on soumet à la même activité. Comme on ne veut pas attendre que toutes les piles soient usées, on décide d'arrêter l'expérience au bout de 10 jours et de compter combien sont encore "en vie". Soit N_{10} ce nombre.
   (a) Dans une première approximation, on suppose que la durée de vie d'une pile peut être modélisée par une loi exponentielle de paramètre $\beta > 0$. Quelle est alors la durée de vie moyenne (théorique) T d'une pile en fonction de β ? Quelle est, en fonction de T, la probabilité qu'une pile meure avant 10 jours ? Montrer alors que $N_{10}/1000$ suit approximativement un théorème de la limite centrale dont on précisera les paramètres en fonction de T. En déduire alors un estimateur $\widehat{T}$ de T en fonction de $N_{10}/1000$ dont on donnera un théorème de la limite centrale.

(b) Montrer que N_{10} n'est pas une statistique exhaustive pour le paramètre T par rapport à l'échantillon des 1000 durées de vie des piles. Et si l'on avait attendu x jours au lieu de 10 ? Déterminer alors une équation vérifiée par x tel que $\widehat{T}$ estime "le mieux" T (donc trouver x tel que $\widehat{T}$ soit de variance minimale). Ce résultat vous semble t-il en pratique intéressant ?