

Table des matières

1	Rappels et compléments de probabilités	5
1.1	Lois de probabilités	5
1.2	Lois de probabilités usuelles	6
1.2.1	Lois de probabilités discrètes	6
1.2.2	Lois de probabilités continues	7
1.3	Lois multivariées	9
1.3.1	Loi normale multivariée	9
1.3.2	Lois liées à la loi normale	12
1.4	Famille des lois exponentielles	14
1.5	Lois conditionnelles	16
1.6	Fonctions génératrices	18
1.7	Convergences	19
1.7.1	Convergence en loi	19
1.7.2	Convergence en probabilité	20
1.7.3	Convergence presque sûre	20
1.7.4	Convergence en moyenne quadratique	21
1.8	Théorèmes limites	22
2	Statistique : généralités et introduction à l'estimation	27
2.1	Notions de bases	27
2.1.1	Espaces de probabilités	27
2.1.2	Modèles statistiques	27
2.1.3	Statistiques	28
2.2	Inférence statistique	29
2.3	Estimation ponctuelle	29
2.3.1	Définitions	29
2.3.2	Quelques problèmes d'estimation	30
3	Méthodes d'estimation	33
3.1	Méthode des moments	33
3.2	Estimateurs du maximum de vraisemblance	34
3.3	Propriétés asymptotiques des EMV	35
3.4	Estimateurs sans biais de variance minimale	37
3.5	Estimation sans biais : cas vectoriel	39

Table des figures

1.1	Densités de lois gamma	8
1.2	Densités de lois normales	9
1.3	Graphe et courbe de niveaux	10
1.4	Graphe et courbe de niveaux	11
1.5	Graphe et courbe de niveaux	11
1.6	Lois normale et de Student	13
1.7	Densités de lois de Fischer	14
1.8	Evolution des f.r dans la CV en loi	20
1.9	Diagramme des CV	22
1.10	Evolution de la moyenne : n=5000	23
1.11	Evolution de la moyenne : n=10000	23
1.12	Distribution de la moyenne : n=10	24
1.13	Distribution de la moyenne : n=20	24
1.14	Distribution de la moyenne : n=50	25
2.1	Estimateurs sans biais et avec biais	30
2.2	Estimateur convergent	30
2.3	Fonction de répartition empirique	32
2.4	Evolution des f.r empiriques vers la f.r théorique	32
3.1	Fonction de vraisemblance	35
3.2	Fonction de log-vraisemblance	35
3.3	Fonction de vraisemblance	36

Chapitre 1

Rappels et compléments de probabilités

1.1 Lois de probabilités

Définition 1.1. Soit $(\Omega, \mathfrak{A}, P)$ un espace de probabilité et $(\mathfrak{X}, \mathfrak{B})$ un espace probabilisable. On appelle variable aléatoire toute application mesurable

$$\begin{aligned} X : \Omega &\longrightarrow \mathfrak{X} \\ \omega &\longrightarrow X(\omega) \end{aligned}$$

Remarque 1. 1. Si $\mathfrak{X} = \mathbb{R}$ on parle de variable aléatoire réelle (v.a.r), si $\mathfrak{X} = \mathbb{R}^n$ on parle de vecteur aléatoire.

2. Si \mathfrak{X} est discret (fini ou infini dénombrable) on parle de variable aléatoire discrète.

Définition 1.2. On appelle loi de probabilité de la variable aléatoire X , la mesure de probabilité image de P par X , i.e la mesure de probabilité notée P_X définie par

$$\begin{aligned} P_X &: \mathfrak{B} \longrightarrow [0, 1] \\ B &\longrightarrow P_X(B) = P(X^{-1}(B)) \end{aligned}$$

Définition 1.3. On appelle fonction de répartition réelle toute fonction $F : \mathbb{R} \longrightarrow [0, 1]$ vérifiant :

1. F croissante
2. F continue à gauche
3. $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$

Proposition 1.1. Soit P_X la loi de probabilité d'une v.a.r X alors la fonction

$$\begin{aligned} F_X &: \mathbb{R} \longrightarrow [0, 1] \\ x &\longrightarrow F_X(x) = P(X \leq x) \end{aligned}$$

est une fonction de répartition dite fonction de répartition de X (ou de P_X).

Toute mesure de probabilité sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ est caractérisée de manière unique par sa fonction

Définition 1.4. Une variable aléatoire réelle de fonction de répartition F_X est dite absolument continue (ou plus exactement que sa loi de probabilité P_X est absolument continue) s'il existe une fonction $f_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ telle que

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

f_X est alors dite densité de probabilité de X .

Propriétés :

1. F_X est presque partout dérivable et on a $F'_X = f_X$
2. f_X est presque partout positive
3. $\int_{-\infty}^{+\infty} f_X(t)dt = 1$
4. $P(X \in B) = \int_B f_X(t)dt$

Définition 1.5. Toute fonction f vérifiant les propriétés 2 et 3 est dite fonction densité de probabilité .

En pratique, deux types de lois de probabilités sont utilisées ; les lois discrètes et les lois absolument continues.

1.2 Lois de probabilités usuelles

1.2.1 Lois de probabilités discrètes

Loi de Bernoulli

Définition 1.6. Une variable aléatoire X est dite suivre une loi de Bernoulli si elle prend ses valeurs dans $\{0, 1\}$ avec $P(X = 0) = 1 - p = q$ et $P(X = 1) = p$

Notation 1. $X \sim B(p)$

Propriété : $E(X) = p$ et $Var(X) = pq$

Exemple 1.1. Tout résultat d'une expérience aléatoire à deux issues possibles (succès, échec ou encore 0,1) peut-être modélisée à l'aide d'une v.a de Bernoulli. Par exemple le résultat d'un lancer d'une pièce de monnaie, en posant $X = 1$ si pile sort et $X = 0$ si face sort. $p = P(X = 1)$ est la probabilité de pile et $q = 1 - p$ celle de face.

Loi binomiale

Définition 1.7. Une variable aléatoire X est dite suivre une loi binomiale si elle prend ses valeurs dans $\{0, 1, 2, \dots, n\}$ avec $P(X = k) = C_n^k p^k (1 - p)^{n-k}$

Notation 2. $X \sim B(n, p)$

Propriétés :

1. $B(1, p) = B(p)$
2. $E(X) = np$ et $Var(X) = npq$
3. $X_i \sim B(p), 1 \leq i \leq n$ i.i.d $\implies \sum_{i=1}^n X_i \sim B(n, p)$

Exemple 1.2. Considérons, une expérience de type succès-échec répétée n fois dans des conditions identiques et de manière indépendante telle que la probabilité d'avoir à chaque répétition un succès est p . Soit X la v.a qui compte le nombre total de succès obtenus, alors $X \sim B(n, p)$

Loi géométrique

Définition 1.8. Une variable aléatoire X est dite suivre une loi géométrique si elle prend ses valeurs dans N^* avec $P(X = k) = p(1 - p)^{k-1} = pq^{k-1}$

Notation 3. $X \sim G(p)$

Propriétés : $E(X) = 1/p$ et $Var(X) = \frac{q}{p^2}$

Exemple 1.3. Considérons, une expérience de type succès-échec répétée dans des conditions identiques et de manière indépendante jusqu'à obtenir pour la première fois un succès. On suppose que la probabilité d'avoir à chaque répétition un succès est p . Soit X la v.a qui compte le nombre de répétitions nécessaires, alors $X \sim G(p)$

Loi de Poisson

Définition 1.9. Une variable aléatoire X est dite suivre une loi de Poisson si elle prend ses valeurs dans N avec $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$

Notation 4. $X \sim P(\lambda)$

Propriétés :

1. $E(X) = \lambda$ et $Var(X) = \lambda$
2. $X_1 \sim P(\lambda_1), X_2 \sim P(\lambda_2)$ indépendantes $\implies X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$

La loi de Poisson est une loi limite comme le montre le théorème d'approximation de la loi binomiale par la loi de Poisson (voir). Si n grand et p petit

$$B(n, p) \approx P(np)$$

1.2.2 Lois de probabilités continues**Loi Gamma**

Définition 1.10. On appelle fonction Gamma la fonction de R_+ dans R_+ définie par

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx$$

Définition 1.11. On appelle fonction Béta la fonction définie sur $R_+ \times R_+$ par

$$\mathcal{B}(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

Propriétés :

1. $\Gamma(a) = (a-1)\Gamma(a-1) \quad \forall a \in R^*$
2. $\Gamma(n) = (n-1)! \quad \forall n \in N^* \quad \text{avec } \Gamma(0) = 1$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Définition 1.12. Une variable aléatoire réelle X est dite suivre une loi gamma si elle admet la densité :

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} \mathbf{1}_{\mathbf{R}_+}(x)$$

a et b paramètres réels strictement positifs

Notation 5. $X \sim \gamma(a, b)$

Cas particuliers :

1. $\gamma(1, b) = \text{Exp}(b)$: loi exponentielle de paramètre b
2. $\gamma(\frac{n}{2}, \frac{1}{2}) = \chi_n^2$: loi du khi-deux à n degrés de liberté (loi fondamentale en statistique)

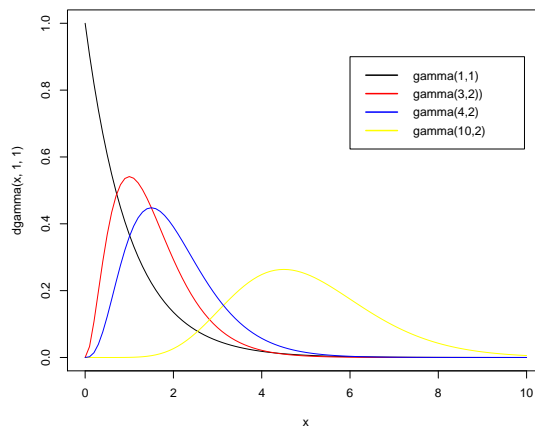


FIGURE 1.1 – Densités de lois gamma

Propriétés :

1. $\int_0^{+\infty} x^{a-1} e^{-bx} dx = \frac{\Gamma(a)}{b^a}$ (intégrale utile à connaître)
2. $E(X) = \frac{a}{b}$, $\text{Var}(X) = \frac{a}{b^2}$ (démonstration directe ou en utilisant les fonctions génératrices)
3. $X_1 \sim \gamma(a_1, b), X_2 \sim \gamma(a_2, b) \implies X_1 + X_2 \sim \gamma(a_1 + a_2, b)$
propriété de stabilité par convolution sur le premier paramètre
(Démonstration directe ou par les fonctions génératrices voir section 1.6)

Loi normale

Définition 1.13. Une variable aléatoire réelle X est dite suivre une loi normale si elle admet la densité :

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

μ, σ^2 paramètres réels respectivement quelconque et strictement positif

Notation 6. $X \sim N(\mu, \sigma^2)$

$N(0,1)$: loi normale centrée et réduite

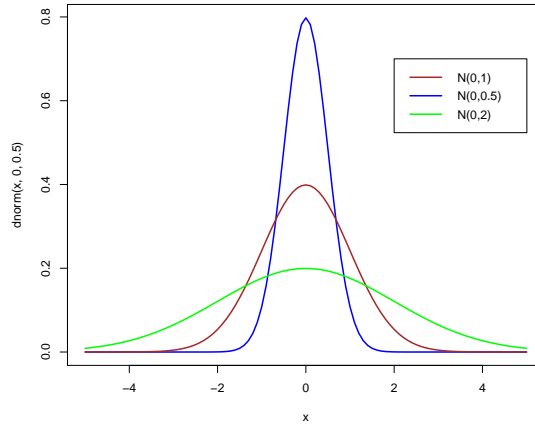


FIGURE 1.2 – Densités de lois normales

Propriétés :

1. $X \sim N(\mu, \sigma^2) \implies Y = \frac{X - \mu}{\sigma} \sim N(0, 1)$ (centrage et réduction)
2. $Y \sim N(0, 1) \implies X = \sigma Y + \mu \sim N(\mu, \sigma^2)$
3. $F_{N(0,1)}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}} dt = \Phi(x)$ n'a pas d'expression analytique (intégrale de Gauss) : tabulée
4. $E(X) = \mu, \text{Var}(X) = \sigma^2$
5. $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ indépendantes $\implies X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
6. $X_i \sim N(\mu_i, \sigma_i^2), 1 \leq i \leq n$ i.i.d $\implies \sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$
propriété de stabilité par convolution sur les deux paramètres
(Démonstration directe ou par les fonctions génératrices)
7. $X \sim N(0, 1) \implies X^2 \sim \chi_1^2$ (khi-deux à un ddl)
8. $X_i \sim N(0, 1), 1 \leq i \leq n$ i.i.d $\implies \sum_{i=1}^n X_i^2 \sim \chi_n^2$ (khi-deux à n ddl)
9. Application : $X_i \sim N(\mu_i, \sigma_i^2), 1 \leq i \leq n$ i.i.d $\implies \sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi_n^2$

1.3 Lois multivariées

1.3.1 Loi normale multivariée

Définition 1.14 (Loi normale multivariée). Un vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ à valeurs dans R^n est dit de loi normale s'il admet la densité

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-m)^t \Sigma^{-1}(x-m)}$$

avec $m \in R^n$ et Σ matrice (n,n) symétrique définie positive.

Notation 7. $X \sim N_n(m, \Sigma)$

Cas particulier : $m = (0, 0, \dots, 0)$ et $\Sigma = I_n$

$X \sim N_n(0, I_n)$: loi normale centrée, réduite dans R^n

Exemple 1.4. Soit X un vecteur aléatoire de densité

$$f(x, y) = \frac{7}{18\pi} \exp\left(-\frac{4x^2 - 2xy + y^2}{6}\right)$$

On peut écrire :

$$-\frac{4x^2 - 2xy + y^2}{6} = \frac{1}{2}(x, y) \begin{pmatrix} \frac{4}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = -\frac{1}{2}(x, y)\Sigma^{-1} \begin{pmatrix} x \\ y \end{pmatrix}$$

On vérifie que : $\frac{1}{2\pi \det(\Sigma)} = \frac{\det(\Sigma^{-1})}{2\pi} = \frac{7}{18\pi}$.

Le vecteur X est donc un vecteur gaussien avec pour moyenne et matrice de covariances respectivement $m = (0, 0)^t$ et Σ

Exemple 1.5. Loi normale bivariée $X \sim N_2(m, \Sigma)$ avec $m = (m_1, m_2)^t$ et $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$

On a alors :

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\sqrt{1-\rho^2}}\left(\frac{(x-m_1)^2}{\sigma_1^2} - \frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2}\right)\right)$$

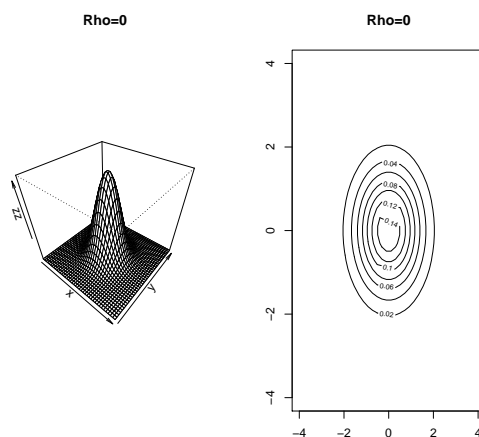


FIGURE 1.3 – Graphe et courbe de niveaux

Théorème 1.2. Soit $X \sim N_n(m, \Sigma)$ et A une matrice (p, n) alors $Y = AX \sim N_p(Am, A\Sigma A^t)$

Démonstration. On utilise les fonctions génératrices (voir section 1.6) $\Psi_Y(u) = E(e^{<u, AX>}) = E(e^{<A^t u, X>}) = \Psi_X(A^t u) = e^{<A^t u, m> + \frac{1}{2}(A^t u)' \Sigma (A^t u)} = e^{<u, Am> + \frac{1}{2}u' (A\Sigma A^t)u}$ \square

Corollaire 1.3. Soit $X = (X_1, X_2, \dots, X_n) \sim N_n(m, \Sigma)$

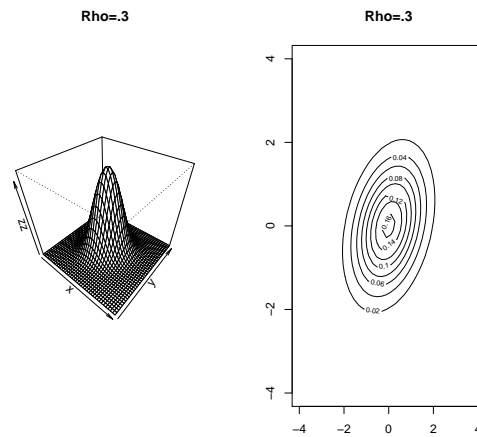


FIGURE 1.4 – Graphe et courbe de niveaux

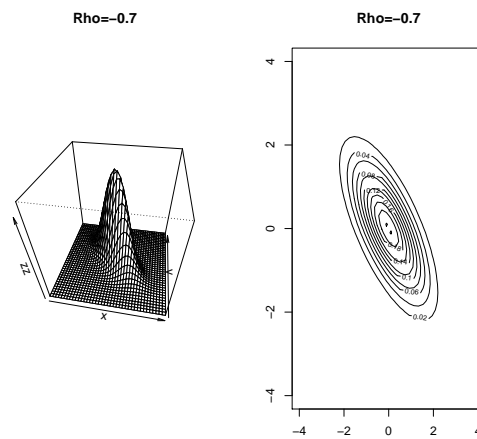


FIGURE 1.5 – Graphe et courbe de niveaux

1. Toute combinaison linéaire $\sum_{i=1}^n a_i X_i$, pour tous réels a_1, \dots, a_n , suit une loi normale
2. En particulier $X_i \sim N(m_i, \sigma_i^2)$ pour tout i

Exemple 1.6. Soit $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \sim N_3\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}\right)$

Loi de $(-X + Y + Z, X - Y + Z, X + Y - Z)$? On a :

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = A(X, Y, Z)^t$$

donc $(U, V, W)^t$ transformé linéaire de $(X, Y, Z)^t$ est un vecteur gaussien dans R^3 de moyenne $A\mu = (0, 0, 0)^t$ et

de matrice de variances-covariances $A\Sigma A^t = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$.

Théorème 1.4. Soit $X = (X_1, X_2, \dots, X_n) \sim N_n(m, \Sigma)$. X_1, X_2, \dots, X_n sont indépendantes si et seulement si Σ est diagonale.

Démonstration. (\Rightarrow) évident

$$(\Leftarrow) \text{ soit } \Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & \sigma_n^2 \end{pmatrix}$$

On a : $\det(\Sigma) = \sigma_1^2 \sigma_2^2 \dots \sigma_n^2$ et $(x - m)^t \Sigma^{-1} (x - m) = \sum_{i=1}^n \frac{(x_i - m_i)^2}{\sigma_i^2}$

$$\text{donc } f_X(x) = \frac{1}{(\sqrt{2\pi})^n \sigma_1 \dots \sigma_n} e^{-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - m_i)^2}{\sigma_i^2}} = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{1}{2} \frac{(x_i - m_i)^2}{\sigma_i^2}} = \prod_{i=1}^n f_{X_i}(x_i) \quad \square$$

1.3.2 Lois liées à la loi normale

Définition 1.15. Soit X_i , $1 \leq i \leq n$ i.i.d de même loi que X

1. (X_1, X_2, \dots, X_n) est dit échantillon de taille n de X

$$2. \overline{X}_n = \frac{\sum_{i=1}^n X_i}{n} : \text{moyenne empirique de la v.a } X$$

$$3. S_n^2 = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n} \text{ et } S_n'^2 = \frac{\sum_{i=1}^n (X_i - \overline{X}_n)^2}{n-1} \text{ variance empirique et variance empirique corrigée de la v.a } X$$

Rappels :

Soit $X_i \sim N(m, \sigma^2)$, $1 \leq i \leq n$ i.i.d alors

$$1. \overline{X}_n \sim N\left(m, \frac{\sigma^2}{n}\right)$$

$$2. \frac{\sum_{i=1}^n (X_i - m)^2}{\sigma^2} = \frac{n}{\sigma^2} \frac{\sum_{i=1}^n (X_i - m)^2}{n} \sim \chi_n^2$$

Théorème 1.5. Soit $X_i \sim N(m, \sigma^2)$, $1 \leq i \leq n$ i.i.d, alors on a :

1. \overline{X}_n et S_n^2 sont indépendantes
2. $\overline{X}_n \sim N\left(m, \frac{\sigma^2}{n}\right)$
3. $\frac{nS_n^2}{\sigma^2} \sim \chi_{n-1}^2$

Théorème 1.6. Soit deux v.a X et Y indépendantes telles que $X \sim N(0, 1)$, $Y \sim \chi_n^2$ alors $Z = \frac{X}{\sqrt{\frac{Y}{n}}}$ a pour densité

$$f_Z(x) = \frac{1}{\sqrt{n}B(\frac{1}{2}, \frac{n}{2})} \frac{1}{(1 + \frac{x^2}{\sigma^2})^{\frac{n+1}{2}}}$$

dite loi de Student à n degrés de liberté.

Notation : $\text{St}(n)$ ou $t(n)$

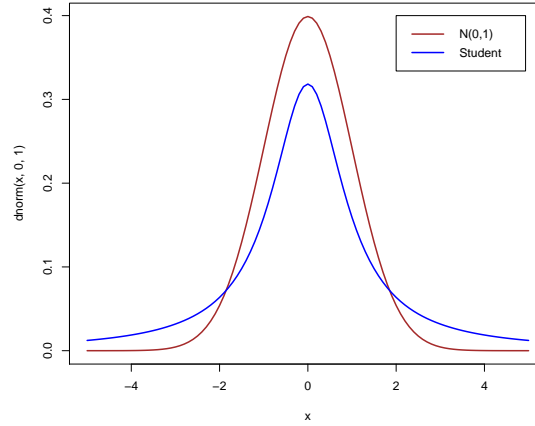


FIGURE 1.6 – Lois normale et de Student

Exemple 1.7. Exemples importants

1. $\frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$
2. $\frac{\sum_{i=1}^n (X_i - m)^2}{\sigma^2} \sim \chi_n^2 \implies$
 $\frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}}$
3. $\sqrt{n} \frac{\frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\sum_{i=1}^n (X_i - m)^2}{\sigma^2}}} = \sqrt{n} \frac{\overline{X_n} - m}{\sqrt{\frac{\sum_{i=1}^n (X_i - m)^2}{n}}} \sim St_n$

Exemple 1.8. 1. $\frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$

2. $\frac{nS_n^2}{\sigma^2} \sim \chi_{n-1}^2 \implies$
 $\frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}}$
3. $\sqrt{n-1} \frac{\frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{nS_n^2}{\sigma^2}}} = \sqrt{n-1} \frac{\overline{X_n} - m}{S_n} \sim St_{n-1}$

Théorème 1.7. Soit deux v.a X et Y indépendantes telles que $U \sim \chi_n^2$, $V \sim \chi_m^2$ alors $F = \frac{U}{V} = \frac{\frac{m}{n} U}{V}$ a pour densité

$$f_F(x) = \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}} (x)^{\frac{n}{2}-1}}{B\left(\frac{n}{2}, \frac{m}{2}\right)} \frac{1}{\left(1 + \frac{n}{m}x\right)^{\frac{n+m}{2}}} 1_{R_+}(x)$$

dite loi de Fischer à n et m degrés de liberté notée $F_{n,m}$.

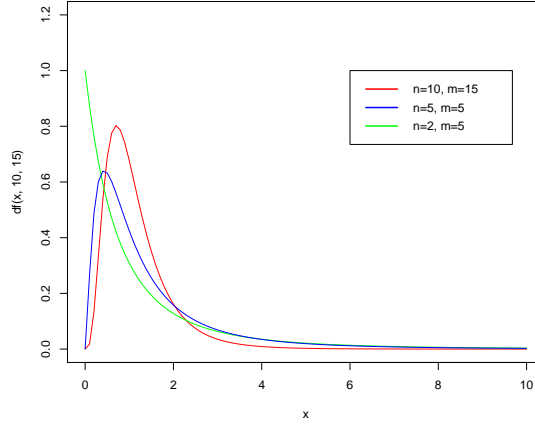


FIGURE 1.7 – Densités de lois de Fischer

Exemple 1.9. Soit $X_i \sim N(m_1, \sigma_1^2)$, $1 \leq i \leq n$ un n -échantillon de X et $Y_j \sim N(m_2, \sigma_2^2)$, $1 \leq j \leq m$ un m -échantillon de Y avec X et Y indépendantes alors

$$1. \frac{\frac{\sum_{i=1}^n (X_i - m_1)^2}{\sigma_1^2}}{\frac{\sum_{j=1}^m (Y_j - m_2)^2}{\sigma_2^2}} \sim F_{n,m}$$

$$2. \frac{\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma_1^2}}{\frac{\sum_{j=1}^m (Y_j - \bar{Y}_m)^2}{\sigma_2^2}} = \frac{n\sigma_2^2}{m\sigma_1^2} \frac{S_1^2}{S_2^2} \sim F_{n-1, m-1}$$

1.4 Famille des lois exponentielles

Définition 1.16. La famille de lois $\mathcal{F} = \{f_\theta(\cdot)/\theta \in \Theta\}$ de \mathbf{R}^p est dite appartenir à la famille des lois exponentielles \mathcal{E}_1 d'ordre 1 s'il existe des fonctions h, a_1 de \mathbf{R}^p dans \mathbf{R} et c, α_1 de Θ dans \mathbf{R} telles que

$$f_{\theta}(x) = c(\theta)h(x) \exp(a(x)\alpha(\theta))$$

Cela est équivalent à

$$f_{\theta} \in \mathcal{E}_1 \iff \log(f_{\theta}(x)) = C(\theta) + H(x) + a(x)\alpha(\theta)$$

en posant $C(\theta) = \log(c(\theta))$, $H(x) = \log(h(x))$

Exemple 1.10. La plupart des lois de probabilités usuelles appartiennent à cette famille

1. Loi de Poisson

$$f_{\lambda}(x) = e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \frac{1}{x!} e^{x \log \lambda} = c(\lambda)h(x) \exp(a(x)\alpha(\lambda))$$

ou encore :

$$\log(f_{\lambda}(x)) = -\lambda + \log\left(\frac{1}{x!}\right) + x \log \lambda$$

2. Loi de Bernoulli

$$f_p(x) = p^x(1-p)^{1-x} = (1-p)\left(\frac{p}{1-p}\right)^x = (1-p) \exp\left(x \log\left(\frac{p}{1-p}\right)\right) = c(p) \exp(a(x)\alpha(p)) \text{ ou}$$

bien

$$\log(f_p(x)) = \log(1-p) + x \log\left(\frac{p}{1-p}\right) = C(p) + a(x)\alpha(p)$$

Exemple 1.11. 1. Loi $N(m, \sigma^2)$, σ^2 connu

$$f_m(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{m^2}{\sigma^2}} e^{-\frac{1}{2} \frac{x^2}{\sigma^2}} e^{\frac{xm}{\sigma^2}} = c(m)h(x) \exp(a(x)\alpha(m))$$

2. Loi $N(m, \sigma^2)$, m connu

$$f_m(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}} = c(\sigma^2) \exp(a(x)\alpha(\sigma^2))$$

Définition 1.17. La famille de lois $\mathcal{F} = \{f_{\theta}(\cdot)/\theta \in \Theta\}$ de \mathbf{R}^p est dite appartenir à la famille des lois exponentielles \mathcal{E}_k d'ordre k s'il existe des fonctions $h, a_i (1 \leq i \leq k)$ de \mathbf{R}^p dans \mathbf{R} et $c, \alpha_i (1 \leq i \leq k)$ de Θ dans \mathbf{R} telles que

$$f_{\theta}(x) = c(\theta)h(x) \exp\left(\sum_{i=1}^k a_i(x)\alpha_i(\theta)\right)$$

Cela est équivalent à :

$$f_{\theta} \in \mathcal{E}_k \iff \log(f_{\theta}(x)) = C(\theta) + H(x) + \sum_{i=1}^k a_i(x)\alpha_i(\theta)$$

avec : $C(\theta) = \log(c(\theta))$, $H(x) = \log(h(x))$

Exemple 1.12. Loi $N(m, \sigma^2)$

$$\begin{aligned} f_{m, \sigma^2}(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{m^2}{\sigma^2}} e^{\frac{xm}{\sigma^2}} e^{-\frac{1}{2} \frac{x^2}{\sigma^2}} \\ &= c(m, \sigma^2)h(x) \exp(a_1(x)\alpha_1(m) + a_2(x)\alpha_2(m)) \end{aligned}$$

avec $a_1(x) = x$, $a_2(x) = x^2$

1.5 Lois conditionnelles

Définition 1.18. Soit X et Y deux v.a discrètes à valeurs respectivement dans $\mathcal{X} = \{x_1, x_2, \dots, x_n, \dots\}$ et $\mathcal{Y} = \{y_1, y_2, \dots, y_p, \dots\}$.

On appelle loi conditionnelle de X sachant que $Y = y_p$ la loi définie par :

$$P(X = x_n / Y = y_p) = \frac{P(X = x_n, Y = y_p)}{P(Y = y_p)}$$

pour tout $n \in N$.

Exemple 1.13. $X_1 \sim B(p)$, $X_2 \sim B(p)$ indépendantes. Loi de $X_1 / X_1 + X_2 = 1$?

Il faut déterminer $P(X_1 = 0 / X_1 + X_2 = 1)$ et $P(X_1 = 1 / X_1 + X_2 = 1)$

On a : $X_1 + X_2 \sim B(2, p)$

$$P(X_1 = 0 / X_1 + X_2 = 1) = \frac{P(X_1 = 0, X_1 + X_2 = 1)}{P(X_1 + X_2 = 1)} = \frac{P(X_1 = 0, X_2 = 1)}{P(X_1 + X_2 = 1)} = \frac{P(X_1 = 0)P(X_2 = 1)}{P(X_1 + X_2 = 1)} = \frac{pq}{2pq} = \frac{1}{2}$$

$$P(X_1 = 1 / X_1 + X_2 = 1) = \frac{1}{2} \text{ Donc } X_1 / X_1 + X_2 = 1 \sim B\left(\frac{1}{2}\right)$$

Définition 1.19. Soit (X, Y) un vecteur aléatoire absolument continu de densité $f_{(X,Y)}$.

On appelle loi conditionnelle de X sachant que $Y = y$ la loi définie par la densité :

$$f_X(x / Y = y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}$$

pour tout $x \in R$.

Exemple 1.14. Soit un vecteur aléatoire (X, Y) de densité $f(x, y) = xe^{-xy} 1_{\{x>0, y>1\}}$. Loi de $X / X + Y = y$?

On a après calculs : $f_Y(y) = \frac{1}{y^2} 1_{\{y>1\}} \implies$

$$f_X(x / Y = y) = \frac{xe^{-xy} 1_{\{x>0, y>1\}}}{\frac{1}{y^2} 1_{\{y>1\}}} = xy^2 e^{-xy} 1_{\{x>0\}}$$

Par exemple : $f_X(x / Y = 1) = xe^{-x} 1_{\{x>0\}}$ donc $X / Y = 1 \sim \gamma(2, 1)$

$f_X(x / Y = 2) = 4xe^{-2x} 1_{\{x>0\}}$, donc $X / Y = 2 \sim \gamma(2, 2)$

D'une manière générale on remarque que : $X / Y = y \sim \gamma(2, y)$

Définition 1.20. On appelle espérance conditionnelle de X sachant que $Y = y$ l'espérance de la loi conditionnelle de X sachant que $Y = y$, i.e

1. Dans le cas discret : $E(X / Y = y_p) = \sum_{i=1}^{+\infty} x_i P(X = x_i / Y = y_p)$
2. Dans le cas continu : $E(X / Y = y) = \int_R x f_X(x / Y = y) dx$

Exemple 1.15. 1. Exemple 1 : $E(X_1 / X_1 + X_2 = 1) = \frac{1}{2}$ (puisque $X_1 / X_1 + X_2 = 1 \sim B(\frac{1}{2})$)

2. Exemple 2 : $E(X / Y = 1) = 2$, $E(X / Y = 2) = 1$. D'une manière générale $E(X / Y = y) = \frac{2}{y}$

3. Exemple 3 : $E(X_1 / X_1 + X_2 = 1) = \frac{1}{2}$ (puisque $X_1 / X_1 + X_2 = 1 \sim B(\frac{1}{2})$)

4. Exemple 4 : $E(X/Y = 1) = 2$, $E(X/Y = 2) = 1$. D'une manière générale $E(X/Y = y) = \frac{2}{y}$

Exemple 1.16. $(X, Y) \sim N_2(m, \Sigma)$ avec $m = (m_1, m_2)^t$ et $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$ on a alors :

$$f_X(x/Y = y) = \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\sigma_1^2(1-\rho^2)}\left(x - m_1 - \frac{\rho\sigma_1(y - m_2)}{\sigma_2}\right)^2\right) \text{ donc } X/Y = y \sim N_1\left(m_1 - \frac{\rho\sigma_1(y - m_2)}{\sigma_2}, \sigma_1^2(1 - \rho^2)\right)$$

$$\text{On déduit : } E(X/Y = y) = m_1 - \frac{\rho\sigma_1(y - m_2)}{\sigma_2}$$

$$\text{et } Var(X/Y = y) = \sigma_1^2(1 - \rho^2)$$

la courbe de $y \rightarrow E(X/Y = y)$ est dite courbe de régression de X sur Y. Dans le cas gaussien, c'est donc une droite dite droite de régression.

On peut définir de même les autres moments conditionnels

1. $E(X^k/Y = y) = \int_R x^k f_X(x/Y = y) dx$ moment d'ordre k de la loi conditionnelle

2. notamment la variance conditionnelle $Var(X/Y = y) = E(X^2/Y = y) - (E(X/Y = y))^2$

Définition 1.21. On appelle espérance conditionnelle de X sachant Y la variable aléatoire qui prend la valeur $E(X/Y = y)$ lorsque Y prend la valeur y.

Définition 1.22. On appelle moment conditionnel d'ordre k de X sachant Y la variable aléatoire, notée $E(X^k/Y)$ qui prend la valeur $E(X^k/Y = y)$ lorsque Y prend la valeur y.

En particulier $Var(X/Y) = E(X^2/Y) - E(X/Y)^2$

Remarque 2. Les variables aléatoires $E(X/Y)$, $E(X^k/Y)$, $Var(X/Y)$ sont des fonctions de la v.a Y. Elles admettent à leur tour des moments, et on a par exemple :

$$E(X/Y) = \int_R E(X/Y = y) f_Y(y) dy = \int_R \left(\int_R x f_X(x/Y = y) dx \right) f_Y(y) dy$$

Exemple 1.17. Exemple 2 : $E(X/Y = y) = \frac{2}{y} \implies E(X/Y) = \frac{2}{Y}$

$$Var(X/Y = y) = \frac{2}{y^2} \implies Var(X/Y) = \frac{2}{Y^2}$$

$$\text{Cas gaussien : } E(X/Y = y) = m_1 - \frac{\rho\sigma_1(y - m_2)}{\sigma_2} \implies E(X/Y) = m_1 - \frac{\rho\sigma_1(Y - m_2)}{\sigma_2}$$

$$Var(X/Y = y) = \sigma_1^2(1 - \rho^2) \implies Var(X/Y) = \sigma_1^2(1 - \rho^2)$$

Théorème 1.8. On a $E(E(X/Y)) = E(X)$

Démonstration. (Dans le cas continu)

On a :

$$\begin{aligned} E(E(X/Y)) &= \int_R \left(\int_R x f_X(x/Y = y) dx \right) f_Y(y) dy = \int_R \int_R x f_X(x/Y = y) f_Y(y) dx dy \\ &= \int_R \int_R x \frac{f_{(X,Y)}(x, y)}{f_Y(y)} f_Y(y) dx dy = \int_R \int_R x f_{(X,Y)}(x, y) dx dy \\ &= E(X) \end{aligned}$$

□

Théorème 1.9. On a $Var(X) = E(Var(X/Y)) + Var(E(X/Y))$

1.6 Fonctions génératrices

Les fonctions génératrices des moments, comme les fonctions caractéristiques, sont un outil puissant en probabilités. Elles permettent notamment de simplifier beaucoup de calculs. Elles sont à distinguer de la notion de fonction génératrice des moments factoriels définie uniquement pour les v.a entières (à valeurs dans N), quoique possédant des propriétés identiques.

Définition 1.23. On appelle fonction génératrice des moments de la v.a réelle X la fonction définie sur une partie $D \subset \mathbb{R}$ dans \mathbb{R}_+ par

$$\Psi_X(t) = E(e^{tX})$$

Remarque 3. 1. Le domaine de définition D dépend de la v.a X

2. Si X est discrète $\Psi_X(t) = \sum_{n=0}^{+\infty} e^{tx_i} P(X = x_i)$

3. Si X est absolument continue de densité f_X on a $\Psi_X(t) = \int_{\mathbb{R}} e^{tx} f_X(x) dx$

Exemple 1.18. 1. $X \sim B(p) : \Psi_X(t) = p + qe^t$

2. $X \sim B(n, p) : \Psi_X(t) = (p + qe^t)^n$

3. $X \sim P(\lambda) : \Psi_X(t) = e^{\lambda(e^t - 1)}$

4. $X \sim \gamma(a, b) : \Psi_X(t) = \frac{1}{(1 - \frac{t}{b})^a}$

5. $X \sim N(m, \sigma^2) : \Psi_X(t) = e^{tm + \frac{1}{2}t^2\sigma^2}$ en particulier si $X \sim N(0, 1) : \Psi_X(t) = e^{\frac{1}{2}t^2}$

Théorème 1.10. Soit X et Y deux v.a de lois de probabilités P_X et P_Y . On a :

$$P_X = P_Y \iff \Psi_X = \Psi_Y$$

Théorème 1.11. Soit X et Y deux v.a indépendantes. On a :

$$\Psi_{X+Y}(t) = \Psi_X(t)\Psi_Y(t) \quad \forall t$$

Démonstration. $\Psi_{X+Y}(t) = E(e^{t(X+Y)})$

$$= E(e^{tX})E(e^{tY}) = \Psi_X(t)\Psi_Y(t)$$

□

Exemple 1.19. 1. $X_i \sim B(p), 1 \leq i \leq n$ i.i.d $\implies \sum_{i=1}^n X_i \sim B(n, p)$

2. $X_i \sim P(\lambda_i), 1 \leq i \leq n$ i.i.d $\implies \sum_{i=1}^n X_i \sim P(\sum_{i=1}^n \lambda_i)$

3. $X_i \sim N(\mu_i, \sigma_i^2), 1 \leq i \leq n$ i.i.d $\implies \sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$

4. $X_1 \sim \gamma(a_1, b), X_2 \sim \gamma(a_2, b) \implies X_1 + X_2 \sim \gamma(a_1 + a_2, b)$

Démonstration. 1. $\Psi_{\sum_{i=1}^n X_i}(t) = E(e^{t \sum_{i=1}^n X_i}) = \prod_{i=1}^n E(e^{tX_i}) = (p + qe^t)^n$

$$\implies \sum_{i=1}^n X_i \sim B(n, p)$$

2. $\Psi_{X_1+X_2}(t) = E(e^{t(X_1+X_2)}) = E(e^{tX_1})E(e^{tX_2}) = \frac{1}{(1 - \frac{t}{b})^{a_1}} \frac{1}{(1 - \frac{t}{b})^{a_2}} = \frac{1}{(1 - \frac{t}{b})^{a_1+a_2}} \implies$

$$X_1 + X_2 \sim \gamma(a_1 + a_2, b)$$

Même démonstration pour les autres exemples

□

Théorème 1.12. Soit X une v.a.r, X admet un moment d'ordre k si et seulement si Ψ_X est dérivable à l'ordre k au point 0 et on a :

$$E(X^k) = \Psi_X^{(k)}(0)$$

Définition 1.24. On appelle fonction génératrice des moments du vecteur aléatoire $X = (X_1, X_2, \dots, X_n)$ la fonction définie sur une partie $D \subset \mathbb{R}^n$ dans \mathbb{R}_+ par

$$\Psi_X(t) = E(e^{<t, X>}) = E\left(e^{\sum_{i=1}^n t_i X_i}\right)$$

Exemple 1.20. $X \sim N_n(m, \Sigma) : \Psi_X(t) = e^{<t, m> + \frac{1}{2} t' \Sigma t}$. En particulier : $X \sim N_n(0, I_n) : \Psi_X(t) = e^{\frac{1}{2} t' t}$

1.7 Convergences

Différents types de convergence existent pour les suites de fonctions (CV simple, CV uniforme), certaines étant plus fortes que d'autres. De la même manière on va définir différents types de convergence pour des suites de v.a.

En général, toute notion de convergence est associée à une notion de distance. Des distances différentes définissent des notions de convergence différentes.

$X_n \xrightarrow[n \rightarrow +\infty]{} X$ dans un certain sens si pour une certaine distance d , $\lim_{n \rightarrow +\infty} d(X_n, X) = 0$.

En probabilité, pour certains types de convergence, cette notion de distance n'apparaît pas de manière explicite, bien qu'elle soit sous-jacente.

1.7.1 Convergence en loi

Définition 1.25. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. La suite $(X_n)_{n \in \mathbb{N}}$ est dite converger en loi vers la v.a X si

$$F_{X_n}(x) \xrightarrow[n \rightarrow +\infty]{} F_X(x)$$

en tout x point de continuité de F_X .

Notation : $X_n \xrightarrow[n \rightarrow +\infty]{Loi} X$

Définition 1.26. $X_n \xrightarrow[n \rightarrow +\infty]{Loi} X$ si et seulement si $\Psi_{X_n}(u) \xrightarrow[n \rightarrow +\infty]{} \Psi_X(u)$ en tout u point du domaine de définition de Ψ_X .

Si $X_n \xrightarrow[n \rightarrow +\infty]{Loi} X$ on a $P(X_n < x) = F_{X_n}(x) \xrightarrow[n \rightarrow +\infty]{} P(X < x) = F_X(x)$ (en tout x point de continuité de F_X) et donc

$$P(a < X_n < b) = F_{X_n}(b) - F_{X_n}(a) \xrightarrow[n \rightarrow +\infty]{} P(a < X < b) = F_X(b) - F_X(a)$$

Conclusion : si n est grand

$$P(X_n < x) \approx P(X < x)$$

$$P(a < X_n < b) \approx P(a < X < b)$$

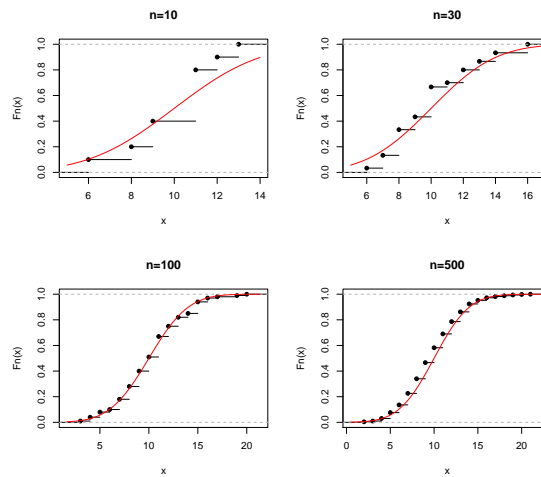


FIGURE 1.8 – Evolution des f.r dans la CV en loi

1.7.2 Convergence en probabilité

Définition 1.27. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. La suite $(X_n)_{n \in \mathbb{N}}$ est dite converger en probabilité vers la v.a X si

$$\forall \epsilon > 0 \quad P(|X_n - X| > \epsilon) \xrightarrow[n \rightarrow +\infty]{} 0$$

Exemple 1.21. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a telles que $X_n \sim B(\frac{1}{n})$. Montrons que $X_n \xrightarrow[n \rightarrow +\infty]{\text{Pr oba}} 0$

$$\begin{aligned} \text{On a : } P(|X_n - 0| > \epsilon) &= \begin{cases} 0 & \text{si } \epsilon \geq 1 \\ \frac{1}{n} & \text{si } \epsilon < 1 \end{cases} \\ \implies \lim_{n \rightarrow +\infty} P(|X_n - 0| > \epsilon) &= 0 \quad \forall \epsilon > 0 \end{aligned}$$

Exemple 1.22. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a telles que $P(X_n = 0) = 1 - \frac{1}{n^2}$ et $P(X_n = n) = \frac{1}{n^2}$

$$\begin{aligned} \text{On a } P(|X_n - 0| > \epsilon) &= \frac{1}{n^2} \\ \lim_{n \rightarrow +\infty} P(|X_n - 0| > \epsilon) &= 0 \quad \forall \epsilon > 0 \implies X_n \xrightarrow[n \rightarrow +\infty]{\text{Pr oba}} 0 \end{aligned}$$

1.7.3 Convergence presque sûre

Définition 1.28. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. La suite $(X_n)_{n \in \mathbb{N}}$ est dite converger presque sûrement vers la v.a X s'il existe un ensemble $A \subset \Omega$, avec $P(A)=1$, telle que :

$$\forall \omega \in A \quad \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$$

Proposition 1.13. Soit $(X_n)_{n \in \mathbb{N}}$ telle que : $X_n \xrightarrow[n \rightarrow +\infty]{P.S.} X$ alors $X_n \xrightarrow[n \rightarrow +\infty]{\text{Pr oba}} X$

Théorème à admettre

1.7.4 Convergence en moyenne quadratique

Les convergences en probabilité et presque sûre ne sont pas généralement faciles à établir. On a donc besoin de critères suffisants. La notion de convergence en moyenne quadratique est plus simple, car c'est la convergence de moments d'ordre 2. Elle implique la C.V en probabilité et donc est une condition suffisante de cette dernière.

Définition 1.29. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. ayant des moments d'ordre 2. La suite $(X_n)_{n \in \mathbb{N}}$ est dite converger en moyenne quadratique vers la v.a. X si

$$\lim_{n \rightarrow +\infty} E((X_n - X)^2) = 0$$

Exemple 1.23. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. telles que $X_n \sim B(\frac{1}{n})$. Montrons que $X_n \xrightarrow[n \rightarrow +\infty]{m.q.} 0$

$$\text{On a } E(|X_n - 0|^2) = E(X_n^2) = \frac{1}{n} \xrightarrow[n \rightarrow +\infty]{} 0$$

Exemple 1.24. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. telles que $P(X_n = 0) = 1 - \frac{1}{n^2}$ et $P(X_n = n) = \frac{1}{n^2}$
 $X_n \xrightarrow[n \rightarrow +\infty]{m.q.} 0$?

$$\text{On a } E(|X_n - 0|^2) = E(X_n^2) = \frac{n^2}{n^2} = 1 \not\xrightarrow[n \rightarrow +\infty]{} 0$$

$$\text{Donc } X_n \not\xrightarrow[n \rightarrow +\infty]{m.q.} 0$$

Proposition 1.14. $X_n \xrightarrow[n \rightarrow +\infty]{m.q.} X \iff \lim_{n \rightarrow +\infty} E(X_n) = E(X) \text{ et } \lim_{n \rightarrow +\infty} \text{Var}(X_n - X) = 0$

Démonstration. $E((X_n - X)^2) = \text{Var}(X_n - X) + (E(X_n - X))^2$ □

Corollaire 1.15. Soit a une constante réelle, on a : $X_n \xrightarrow[n \rightarrow +\infty]{m.q.} a \iff \lim_{n \rightarrow +\infty} E(X_n) = a \text{ et } \lim_{n \rightarrow +\infty} \text{Var}(X_n) = 0$

Démonstration. conséquence directe du théorème précédent $E((X_n - X)^2) = \text{Var}(X_n - X) + (E(X_n - X))^2$ □

Proposition 1.16. Soit $(X_n)_{n \in \mathbb{N}}$ telle que : $X_n \xrightarrow[n \rightarrow +\infty]{m.q.} X$ alors $X_n \xrightarrow[n \rightarrow +\infty]{\text{Pr oba}} X$

Démonstration. Rappel : inégalité de Bienaymé-Tchebychev

si X est une v.a. avec une espérance m et une variance σ^2 finies, on a

$$\forall \epsilon > 0 \quad P(|X - m| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \text{ On applique cet inégalité à } Y = X_n - X \text{ qui est une variable}$$

$$\text{centrée (de moyenne nulle)} \quad P(|X_n - X| > \epsilon) \leq \frac{E((X_n - X)^2)}{\epsilon^2} \implies$$

$$\lim_{n \rightarrow +\infty} P(|X_n - X| > \epsilon) \leq \lim_{n \rightarrow +\infty} \frac{E((X_n - X)^2)}{\epsilon^2} = 0$$

□

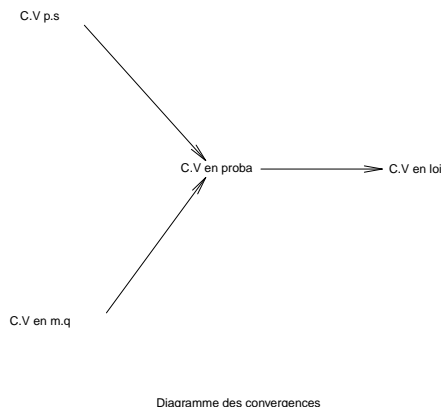


FIGURE 1.9 – Diagramme des CV

1.8 Théorèmes limites

Ce qu'on appelle théorèmes limites, est constitué de deux familles de théorèmes (lois des grands nombres et théorèmes central limite ou de la limite centrale) absolument fondamentaux en probabilités et statistique.

Théorème 1.17 (Loi faible des grands nombres). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a i.i.d admettant une espérance m et une variance σ^2 finies. Alors*

$$\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow[n \rightarrow +\infty]{\text{Pr oba}} m$$

Démonstration. Rappel : inégalité de Bienaymé-Tchebychev

Si X est une v.a avec une espérance m et une variance σ^2 finies, on a :

$$\forall \epsilon > 0 \quad P(|X - m| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

On applique cet inégalité à $\overline{X_n}$

$$\forall \epsilon > 0, \quad P(|\overline{X_n} - m| > \epsilon) \leq \frac{\text{Var}(\overline{X_n})}{\epsilon^2} = \frac{\text{Var}(X)}{n\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow[n \rightarrow +\infty]{} 0$$

□

Théorème 1.18 (Loi forte des grands nombres). *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a i.i.d admettant une espérance m et une variance σ^2 finies. Alors*

$$\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow[n \rightarrow +\infty]{P.S} m$$

Exemple 1.25. 1. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a i.i.d de loi de Bernoulli $B(p)$. On a $\overline{X}_n =$

$$\frac{\sum_{i=1}^n X_i}{n} = \frac{\text{nombre de piles obtenus}}{n} = F_n \xrightarrow[n \rightarrow +\infty]{P.S.} p$$

$$2. X_i \sim N(m, \sigma^2), 1 \leq i \leq n \text{ i.i.d} \implies \overline{X}_n \xrightarrow[n \rightarrow +\infty]{P.S.} m$$

$$3. X_i \sim P(\lambda), 1 \leq i \leq n \text{ i.i.d} \implies \overline{X}_n \xrightarrow[n \rightarrow +\infty]{P.S.} \lambda$$

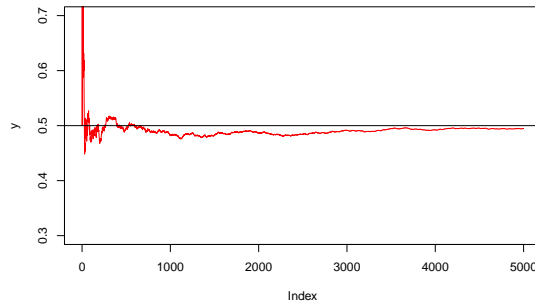


FIGURE 1.10 – Evolution de la moyenne : n=5000

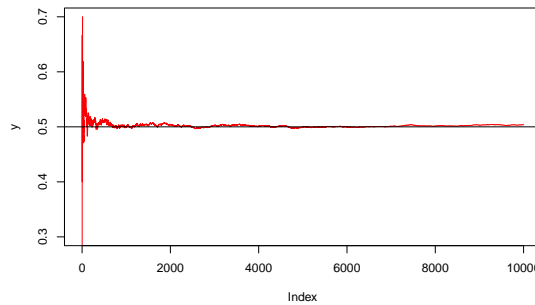


FIGURE 1.11 – Evolution de la moyenne : n=10000

Théorème 1.19 (Théorème central-limite). Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a i.i.d admettant une espérance m et une variance σ^2 finies. Alors

$$\frac{\sum_{i=1}^n X_i - nm}{\sqrt{n}\sigma} \xrightarrow[n \rightarrow +\infty]{Loi} X$$

avec $X \sim N(0, 1)$

Corollaire 1.20. Sous les mêmes hypothèses, si on pose

$$\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n} \text{ on a}$$

$$\frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n \rightarrow +\infty]{\text{Loi}} X$$

avec $X \sim N(0, 1)$

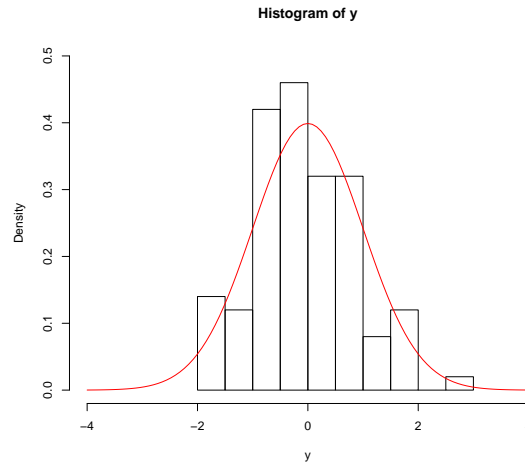


FIGURE 1.12 – Distribution de la moyenne : n=10

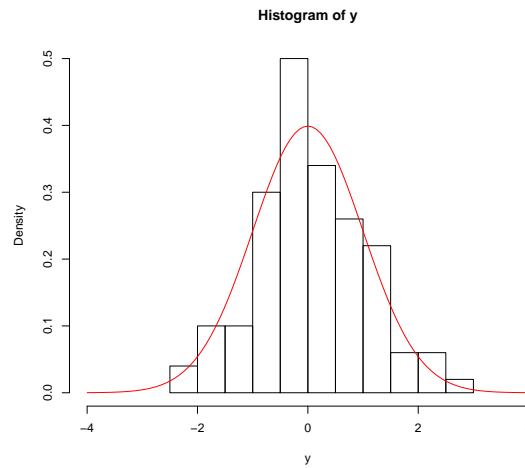


FIGURE 1.13 – Distribution de la moyenne : n=20

Exemple 1.26. Approximation normale de la loi binomiale

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a i.i.d de loi de Bernoulli $B(p)$. On a $\sum_{i=1}^n X_i \sim B(n, p)$

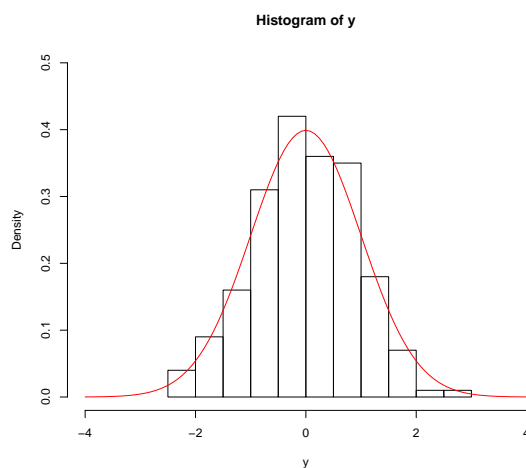


FIGURE 1.14 – Distribution de la moyenne : n=50

et si n grand

$$\sum_{i=1}^n X_i \approx N(np, npq) \implies B(n, p) \approx N(np, npq)$$

Exemple 1.27. Approximation normale de la loi de Poisson Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a i.i.d de loi de Bernoulli $P(n\lambda)$.

On a $\sum_{i=1}^n X_i \sim P(n\lambda)$

et si n grand

$$\sum_{i=1}^n X_i \approx N(\lambda, \lambda) \implies P(n\lambda) \approx N(\lambda, \lambda)$$

Chapitre 2

Statistique : généralités et introduction à l'estimation ponctuelle

2.1 Notions de bases

2.1.1 Espaces de probabilités

En probabilité, un phénomène aléatoire est modélisé par un triplet $(\Omega, \mathfrak{A}, P)$ dit espace de probabilité.

Ω : ensemble des résultats possibles de l'expérience

\mathfrak{A} : ensemble des évènements (tribu de parties de Ω)

P : mesure de probabilité unique et fixée

Exemple 2.1. 1. Lancer d'une pièce de monnaie équilibrée : $(\{\pi, F\}, \mathcal{P}(\{\pi, F\}), B(\frac{1}{2}))$

2. Note obtenue par un étudiant de ST choisi au hasard à l'épreuve de mathématiques $(R, \mathcal{B}_R, \{N(10, 16)\})$

2.1.2 Modèles statistiques

En statistique la loi de probabilité est inconnue. Cette modélisation se fait à l'aide d'un modèle statistique :

Définition 2.1 (Modèle statistique). On appelle modèle statistique un triplet $(\mathfrak{X}, \mathfrak{B}, \mathcal{P})$

\mathfrak{X} : ensemble des observations possibles de l'expérience

\mathfrak{B} : ensemble des évènements (tribu de parties de \mathfrak{X})

\mathcal{P} : ensemble de mesures de probabilités

Exemple 2.2. – Modèle associée à l'observation d'un lancer d'une pièce de monnaie :

$(\{\pi, F\}, \mathcal{P}(\{\pi, F\}), \{B(p)/p \in [0, 1]\})$

– Modèle associée à l'observation de la note obtenue en mathématiques par un étudiant de ST choisi au hasard : $(R, \mathcal{B}_R, \{N(m, \sigma^2)/m \in R, \sigma^2 \in R_+^*\})^{(5)}$

Modèle statistique d'échantillon : $(\mathfrak{X}, \mathfrak{B}, \mathcal{P})^{(n)}$ associé à n observations d'une v.a X de loi dans des conditions d'indépendance.

Exemple 2.3. – $(\{\pi, F\}, \mathcal{P}(\{\pi, F\}), \{B(p)/p \in [0, 1]\})^{(10)}$:

Modèle associée à l'observation de 10 lancers d'une pièce de monnaie

– $(R, \mathcal{B}_R, \{N(m, \sigma^2)/m \in R, \sigma^2 \in R_+^*\})^{(5)}$

Modèle associée à l'observation des notes obtenues par 5 étudiants de ST choisis au hasard.

On distingue deux types de modèles statistiques :

modèles statistiques paramétriques :

$(\mathfrak{X}, \mathfrak{B}, (\mathcal{P}_\theta)_{\theta \in \Theta})$:

$(\mathfrak{X}, \mathfrak{B}, (f_\theta)_{\theta \in \Theta})$

modèles statistiques non paramétriques :

$(\mathfrak{X}, \mathfrak{B}, \{P/P \in \mathcal{P}\})$ \mathcal{P} : ensemble de toutes les lois de probabilités

$(\mathfrak{X}, \mathfrak{B}, \{F/F \in \mathcal{F}\})$ \mathcal{F} : ensemble de toutes les fonctions de répartition.

Conclusion 2.1. *Un modèle paramétrique est un modèle où on fait l'hypothèse que la loi de X appartient à une famille bien déterminée de lois, famille indexée par un nombre fini de paramètres.*

Un modèle non paramétrique est un modèle où on ne fait aucune hypothèse sur la loi de X ou une hypothèse très large, par exemple que la loi est absolument continue.

2.1.3 Statistiques

Définition 2.2. On appelle statistique toute application mesurable définie sur $(\mathfrak{X}, \mathfrak{B}, (\mathcal{P}_\theta)_{\theta \in \Theta})$ à valeurs dans un espace mesurable $(\mathcal{Y}, \mathcal{C})$

$$T : \mathfrak{X} \longrightarrow \mathcal{Y}$$

$$x \longmapsto T(x)$$

Notation 8. : T, S, U...

Si $\mathcal{Y} = \mathbb{R}$: statistique réelle

Si $\mathcal{Y} = \mathbb{R}^p$: statistique vectorielle

Exemple 2.4 (Exemples de statistiques : moments empiriques). 1. $T_1(X_1, X_2, \dots, X_n) = \overline{X_n} =$

$$\frac{\sum_{i=1}^n X_i}{n} : \text{moyenne empirique}$$

$$2. T_2(X_1, X_2, \dots, X_n) = \frac{\sum_{i=1}^n X_i^2}{n} : \text{moment empirique d'ordre 2}$$

$$3. T_k(X_1, X_2, \dots, X_n) = \frac{\sum_{i=1}^n X_i^k}{n} : \text{moment empirique d'ordre } k$$

$$4. S_n^2 = \frac{\sum_{i=1}^n (X_i - \overline{X_n})^2}{n} : \text{variance empirique}$$

$$5. S_n'^2 = \frac{\sum_{i=1}^n (X_i - \overline{X_n})^2}{n-1} : \text{variance empirique corrigée}$$

$$6. S_{X,Y}^2 = \frac{\sum_{i=1}^n (X_i - \overline{X_n})(Y_i - \overline{Y_n})}{n} : \text{covariance empirique}$$

$$7. S_{X,Y}'^2 = \frac{\sum_{i=1}^n (X_i - \overline{X_n})(Y_i - \overline{Y_n})}{n-1} : \text{covariance empirique corrigée}$$

$$8. \tilde{\rho}_{X,Y} = \frac{S_{X,Y}^2}{S_X S_Y} : \text{coefficient de corrélation empirique}$$

Exemple 2.5 (Exemples de statistiques : statistiques d'ordre). – $U_1(X_1, X_2, \dots, X_n) = \min(X_i)_{1 \leq i \leq n} =$

$X_{(1)}$: minimum

- $U_2(X_1, X_2, \dots, X_n) = \max(X_i)_{1 \leq i \leq n} = X_{(n)}$: maximum
- $R(X_1, X_2, \dots, X_n) = \max(X_i)_{1 \leq i \leq n} - \min(X_i)_{1 \leq i \leq n} = X_{(n)} - X_{(1)}$: étendue

2.2 Inférence statistique

Soit $(\mathfrak{X}, \mathfrak{B}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique associé à l'observation d'une v.a X de loi P_θ . θ est inconnu et on cherche à le connaître.

On appelle inférence statistique le processus qui à partir d'observations de X permettent d'avoir une connaissance de θ .

L'hypothèse de base est que les observations faites contiennent de l'information sur θ .

On distingue trois grands problèmes d'inférence statistique :

- Estimation ponctuelle
- Estimation par régions de confiance
- Tests d'hypothèses

2.3 Estimation ponctuelle

2.3.1 Définitions

Soit $(\mathfrak{X}, \mathfrak{B}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique, on veut estimer θ ou une fonction $g(\theta)$ où g est une fonction définie sur Θ .

Définition 2.3. On appelle estimateur de $g(\theta)$ toute statistique à valeurs dans $g(\Theta)$:

$$T : \mathfrak{X} \longrightarrow g(\Theta)$$

$$x \longrightarrow T(x)$$

$T(x)$ est dite estimation de $g(\theta)$

Définition 2.4. Un estimateur T de $g(\theta)$ est dit sans biais si

$$E_\theta(T) = g(\theta)$$

Un estimateur est donc sans biais si en moyenne, il est égal à la quantité qu'il estime.

Définition 2.5. Une suite d'estimateurs $(T_n)_{n \in N}$ de $g(\theta)$ est dite asymptotiquement sans biais si

$$\lim_{n \rightarrow +\infty} E_\theta(T_n) = g(\theta)$$

Définition 2.6. Une suite d'estimateurs $(T_n)_{n \in N}$ de $g(\theta)$ est dite convergent en probabilité (respectivement p.s) si

$$T_n \xrightarrow[n \rightarrow +\infty]{Pr\ oba} g(\theta)$$

(respectivement

$$T_n \xrightarrow[n \rightarrow +\infty]{P.S} g(\theta)$$

)

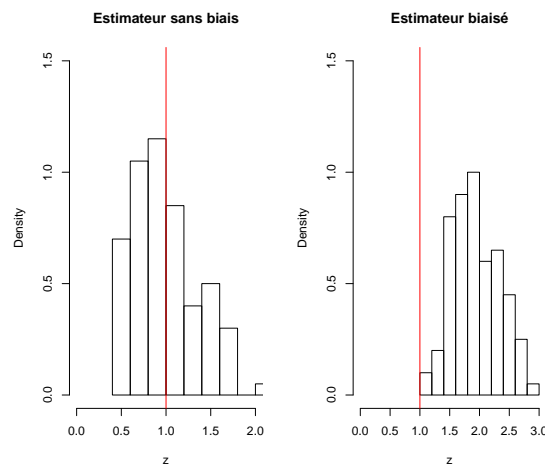


FIGURE 2.1 – Estimateurs sans biais et avec biais

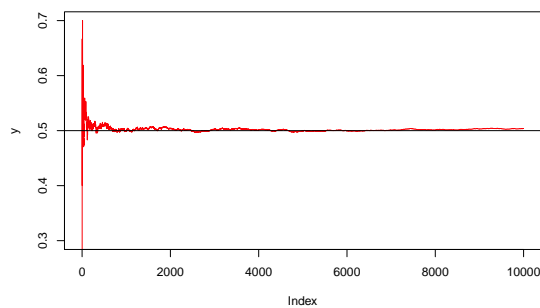


FIGURE 2.2 – Estimateur convergent

2.3.2 Quelques problèmes d'estimation

Soit X_1, X_2, \dots, X_n un n -échantillon d'une v.a X , les moments de cette variable et les probabilités du type $P(X \in A)$ peuvent être estimées de manière simple, et dans un cadre général (i.e sans faire aucune hypothèse sur la loi de X). Les estimateurs obtenus possèdent de bonnes propriétés.

Théorème 2.2. Soit X_1, X_2, \dots, X_n un n -échantillon d'une v.a X , admettant un moment d'ordre

k , alors le moment empirique d'ordre k $m_k = \frac{\sum_{i=1}^n X_i^k}{n}$ est un estimateur sans biais et convergent presque sûrement de $E(X^k)$.

Démonstration. Application immédiate de la loi forte des grands nombres (voir le chapitre 1) \square

Corollaire 2.3. $\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n}$ est un estimateur sans biais et convergent presque sûrement de $E(X)$.

Corollaire 2.4. $S_n'^2 = \frac{\sum_{i=1}^n (X_i - \overline{X_n})^2}{n-1}$ est un estimateur sans biais et convergent presque sûrement de $Var(X)$.

Démonstration. On a

$$S_n^2 = \frac{\sum_{i=1}^n (X_i - \overline{X_n})^2}{n} = \frac{\sum_{i=1}^n X_i^2}{n} - \overline{X_n}^2. \text{ Par ailleurs}$$

$$E(S_n^2) = E\left(\frac{\sum_{i=1}^n X_i^2}{n}\right) - E(\overline{X_n}^2) = E(X_i^2) - E(\overline{X_n}^2)$$

$$\begin{aligned} \text{On a } E(\overline{X_n}^2) &= E\left(\left(\frac{\sum_{i=1}^n X_i}{n}\right)^2\right) = E\left(\frac{\sum_{i=1}^n X_i^2}{n^2}\right) - E\left(\frac{\sum_{i \neq j} X_i X_j}{n^2}\right) = \frac{E(X_i^2)}{n} - \frac{n(n-1)}{n^2} E(X_i X_j) \\ &= \frac{E(X_i^2)}{n} - \frac{(n-1)}{n} E(X_i)^2 \implies \end{aligned}$$

$$E(S_n^2) = \frac{(n-1)}{n} (E(X_i^2) - E(X_i)^2) = \frac{(n-1)}{n} Var(X) \implies$$

$$E(S_n'^2) = E\left(\frac{n}{n-1} S_n^2\right) = Var(X)$$

D'autre part :

$$S_n^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \overline{X_n}^2 \xrightarrow[n \rightarrow +\infty]{P.S.} E(X^2) - E(X)^2 = Var(X)$$

(Loi forte des grands nombres)

$$\implies S_n'^2 = \frac{n}{n-1} S_n^2 \xrightarrow[n \rightarrow +\infty]{P.S.} Var(X) \quad \square$$

Corollaire 2.5. $S_{X,Y}'^2 = \frac{\sum_{i=1}^n (X_i - \overline{X_n})(Y_i - \overline{Y_n})}{n-1}$ est un estimateur sans biais et convergent presque sûrement de $Cov(X, Y)$.

Démonstration identique que pour le corollaire 16.

Corollaire 2.6. Soit A un évènement, lié à une expérience aléatoire et soit $F_n(A) = \frac{\sum_{i=1}^n 1_{\{X_i \in A\}}}{n}$ la fréquence de l'évènement $\{X \in A\}$ (le nombre de réalisations de $\{X \in A\}$, lors de n répétitions de cet expérience dans des conditions indépendantes). $F_n(A)$ est un estimateur sans biais et convergent presque sûrement de $P(X \in A)$.

Démonstration. $\forall i, 1_{\{X_i \in A\}} \sim B(P(A))$ et d'après le corollaire 15, $\frac{\sum_{i=1}^n 1_{\{X_i \in A\}}}{n}$ est un estimateur sans biais et convergent de $P(A)$. □

Définition 2.7. Soit (x_1, x_2, \dots, x_n) un n -échantillon d'observations de X , on appelle fonction de répartition empirique de X la fonction F_n de \mathbb{R} dans $[0, 1]$ définie par :

$$F_n(x) = \frac{\sum_{i=1}^n 1_{\{x_i \leq x\}}}{n}$$

Corollaire 2.7. $F_n(x) = \frac{\sum_{i=1}^n 1_{\{x_i \leq x\}}}{n}$ est un estimateur sans biais et convergent de $F(x)$, la fonction de répartition de X

Démonstration. $F(x) = P(X < x)$, le corollaire 18 permet de déduire le résultat.

□

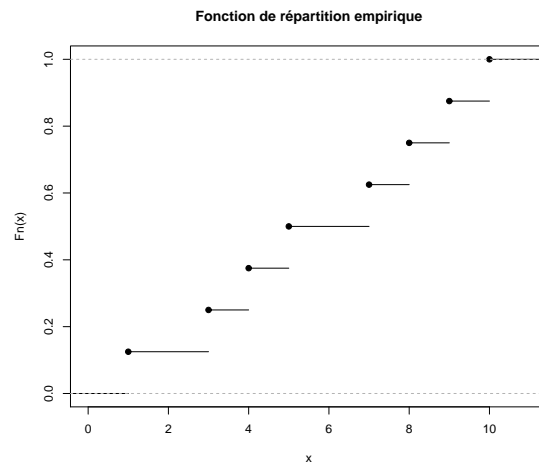


FIGURE 2.3 – Fonction de répartition empirique

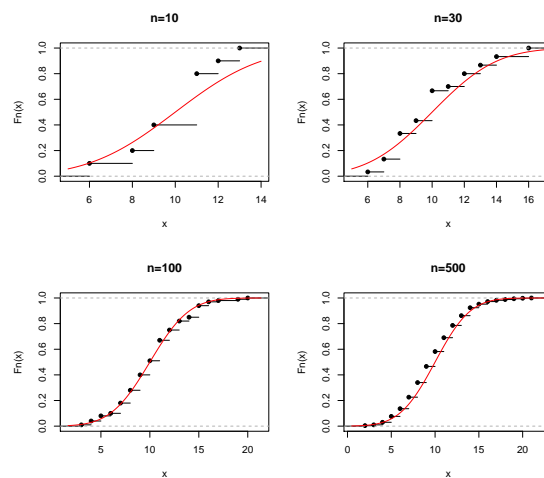


FIGURE 2.4 – Evolution des f.r empiriques vers la f.r théorique

Chapitre 3

Exhaustivité, complétion, information de Fischer

3.1 Fonction de vraisemblance

Définition 3.1. Soit le modèle statistique $(\mathfrak{X}, \mathfrak{B}, (f_\theta)_{\theta \in \Theta})$. On appelle fonction de vraisemblance du modèle (ou de l'observation), la fonction :

$$L : \Theta \rightarrow R_+ \\ \theta \rightarrow L(\theta, x) = f_\theta(x)$$

Si le modèle est un modèle d'échantillon $(\mathfrak{X}, \mathfrak{B}, (f_\theta)_{\theta \in \Theta})^{(n)}$

$$L(\theta, x_1, x_2, \dots, x_n) = L(\theta, \tilde{x}) = \prod_{i=1}^n f_\theta(x_i)$$

Le logarithme de la fonction de vraisemblance est dit fonction de log-vraisemblance notée :

$$l(\theta, \tilde{x}) = \log L(\theta, \tilde{x})$$

Exemple 3.1. 1. n-échantillon loi de Poisson $P(\lambda)$. On a

$$L(\lambda, \tilde{x}) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \\ l(\lambda, \tilde{x}) = \log L(\lambda, \tilde{x}) = -n\lambda + \left(\sum_{i=1}^n x_i\right) \log \lambda - \log\left(\prod_{i=1}^n x_i!\right).$$

2. n-échantillon loi normale, σ^2 connu.

$$L(m, \tilde{x}) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x_i - m)^2}{\sigma^2}} \right) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - m)^2}{\sigma^2}} \\ l(m, \tilde{x}) = \log L(m, \tilde{x})$$

3.2 Exhaustivité

En statistique les observations (données) peuvent être très nombreuses (des fois de l'ordre de plusieurs milliers, par exemple pluviométrie quotidienne ou taux d'inflation mensuel relevés sur plusieurs années. Les statistiques associés à un échantillon d'observations servent à résumer ces observations. Ainsi si on a n observations réelles $\tilde{x} = (x_1, x_2, \dots, x_n)$ $T(\tilde{x}) = \sum_{i=1}^n x_i$ les résume en une seule donnée.

Question : Le résumé est-il un bon résumé ?

En résumant ainsi, i.e en considérant que $T(x)$ au lieu de l'échantillon initiale, perd-on de l'information sur le paramètre θ ?

Quand on lit un livre ou un résumé de ce livre, est-ce la même chose du point de vue de l'information sur un point donné ?

Si oui, on dira que le résumé est exhaustif.

Problème : définir mathématiquement cette notion d'exhaustivité.

Définition 3.2 (Statistique exhaustive). Soit un modèle statistique $(\mathfrak{X}, \mathfrak{B}, (\mathcal{P}_\theta)_{\theta \in \Theta})$ et une statistique T définie sur ce modèle. T est dite exhaustive si la loi conditionnelle $P_{X/T(X)=t}$ ne dépend pas de θ .

Théorème 3.1 (Théorème de factorisation). T est exhaustive si et seulement si :
il existe des fonctions g et h telles que

$$L(\theta, x) = g(\theta, T(x))h(x)$$

Ce théorème est à admettre.

Exemple 3.2. 1. Echantillon d'une loi de Poisson

$$L(\lambda, \tilde{x}) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \frac{1}{\prod_{i=1}^n x_i!} = g(\lambda, \sum_{i=1}^n x_i) h(\tilde{x})$$

donc $T(\tilde{x}) = \sum_{i=1}^n x_i$ est une statistique exhaustive pour λ .

2. Echantillon d'une loi normale à variance connue :

$$L(m, \tilde{x}) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x_i - m)^2}{\sigma^2}} \right) = \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{m^2}{\sigma^2}} \right)^n e^{\frac{m}{\sigma^2} \sum_{i=1}^n x_i} e^{-\frac{1}{2} \frac{\sum_{i=1}^n x_i^2}{\sigma^2}}$$

$$= g(m, \sum_{i=1}^n x_i) h(\tilde{x})$$

donc $T(\tilde{x}) = \sum_{i=1}^n x_i$ est une statistique exhaustive pour m .

3. n-échantillon loi normale

$$L(m, \sigma^2, \tilde{x}) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x_i - m)^2}{\sigma^2}} \right) = \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{m^2}{\sigma^2}} \right)^n e^{\frac{m}{\sigma^2} \sum_{i=1}^n x_i} e^{-\frac{1}{2} \frac{\sum_{i=1}^n x_i^2}{\sigma^2}} = g(m, \sigma^2, \sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$$

avec $h(\tilde{x}) = 1$.

$T(\tilde{x}) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ est une statistique exhaustive pour m et σ^2 . C'est une statistique de dimension 2 (vectorielle).

4. $\tilde{x} = (x_1, x_2, \dots, x_n)$ n-échantillon de $X \sim \mathcal{U}_{[0, \theta]}$ de densité $f_\theta(x) = \frac{1}{\theta} 1_{[0, \theta]}(x)$. On a

$$L(\theta, \tilde{x}) = \prod_{i=1}^n f_\theta(x_i) = \frac{1}{\theta^n} 1_{[0, \theta]}(\max_{1 \leq i \leq n} (x_i)) = g(\theta, \max_{1 \leq i \leq n} (x_i)) h(\tilde{x}) \text{ avec } h(\tilde{x}) = 1 \implies T(\tilde{x}) = \max_{1 \leq i \leq n} (x_i) : \text{statistique exhaustive pour } \theta.$$

Conclusion : Quand on veut faire de l'inférence statistique sur λ , si on a un échantillon poissonien, travailler avec l'échantillon tout entier $\tilde{x} = (x_1, x_2, \dots, x_n)$ ou avec $T(\tilde{x}) = \sum_{i=1}^n x_i$

ne change rien . Toute l'information contenue sur λ dans \tilde{x} se retrouve dans $T(\tilde{x}) = \sum_{i=1}^n x_i$.

Théorème 3.2. Soit f_θ appartenant à la famille des lois exponentielles \mathcal{E}_1 d'ordre 1, i.e

$$f_\theta(x) = c(\theta)h(x) \exp(a(x)\alpha(\theta))$$

Alors $T(x) = a(x)$ est exhaustive pour θ .

Si on a un n -échantillon alors $T(\tilde{x}) = \sum_{i=1}^n a(x_i)$ est exhaustive pour θ .

Démonstration. Théorème de factorisation. □

Théorème 3.3. Soit f_θ appartenant à la famille des lois exponentielles \mathcal{E}_k d'ordre k , i.e

$$f_\theta(x) = c(\theta)h(x) \exp\left(\sum_{i=1}^k a_i(x)\alpha_i(\theta)\right)$$

Alors $T(x) = (a_1(x), a_2(x), \dots, a_k(x))$ est exhaustive pour θ .

Si on a un n -échantillon alors $T(\tilde{x}) = \left(\sum_{i=1}^n a_1(x_i), \sum_{i=1}^n a_2(x_i), \dots, \sum_{i=1}^n a_k(x_i)\right)$ est exhaustive pour θ .

3.3 Complétion

Problème :

Il peut exister pour un paramètre donné plusieurs statistiques exhaustives (résumés exhaustifs). On a vu que n données peuvent résumées de manière exhaustive dans 3, 4 données ou même 1 donnée. Il y a donc réduction plus ou moins importante.

Question : Quelle est la plus grande réduction que l'on peut avoir ?

C'est la notion de statistique exhaustive minimale liée à la notion de statistique exhaustive complète.

Définition 3.3 (Statistique complète). Une statistique T est dite complète si :

$$E(h_1(T)) = E(h_2(T)) \implies h_1(T) = h_2(T) \text{ } P_\theta\text{-presque sûrement } \forall \theta \in \Theta.$$

Théorème 3.4. Soit f_θ appartenant à la famille des lois exponentielles \mathcal{E}_k d'ordre k .

Si $\dim(\Theta) = k$, alors la statistique $T(\tilde{x}) = \left(\sum_{i=1}^n a_1(x_i), \sum_{i=1}^n a_2(x_i), \dots, \sum_{i=1}^n a_k(x_i)\right)$ exhaustive est complète.

3.4 Information de Fischer

L'expression d'information sur un paramètre contenue dans des observations est souvent employée. Cette notion est jusqu'à présent intuitive.

Question : peut-on définir mathématiquement cette notion

3.4.1 Cas d'un paramètre réel

Il existe plusieurs définitions possibles (Fisher, Shannon, Kullback....). Nous introduisons dans la suite la notion d'information au sens de Fisher.

L'information de Fisher est définie sous certaines hypothèses dites : conditions de régularité de Fisher :

1. Le support de $f_\theta : \text{Supp}(f_\theta) = \{x / f_\theta(x) \neq 0\}$ ne dépend pas de θ
2. La fonction f_θ est dérivable par rapport à θ et on peut dériver sous le signe somme dans $\int f_\theta(x)dx$
i.e $\frac{\partial}{\partial \theta} \int f_\theta(x)dx = \int \frac{\partial}{\partial \theta} f_\theta(x)dx$.

Définition 3.4 (Information de Fisher). Soit un modèle statistique associé à l'observation d'une v.a de loi P_θ de densité f_θ .

On appelle quantité d'information (de Fisher) ramenée par le modèle(ou par l'observation) sur θ le nombre

$$I(\theta) = \text{Var}\left(\frac{\partial}{\partial \theta} \log L(\theta, x)\right)$$

Exemple 3.3. Information ramenée par une observation d'une variable de loi $N(m, \sigma^2)$ sur m (σ^2 connu)

$$\begin{aligned} L(m, x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}} \implies \log L(m, x) = \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2} \frac{(x-m)^2}{\sigma^2} \\ \implies \frac{\partial}{\partial m} \log L(m, x) &= \frac{(X-m)}{\sigma^2} \implies I(m) = \text{Var}\left(\frac{\partial}{\partial m} \log L(m, x)\right) = \frac{1}{\sigma^2} \text{Var}\left(\frac{X-m}{\sigma}\right) \\ \implies I(m) &= \frac{1}{\sigma^2} \end{aligned}$$

Propriétés :

1. $I(\theta) = E\left(\left(\frac{\partial}{\partial \theta} \log L(\theta, x)\right)^2\right)$
2. Si on suppose en plus $\frac{\partial^2}{\partial \theta^2} \int f_\theta(x)dx = \int \frac{\partial^2}{\partial \theta^2} f_\theta(x)dx$. on a
$$I(\theta) = -E\left(\frac{\partial^2}{\partial \theta^2} \log L(\theta, x)\right)$$
3. Si on a un n -échantillon l'information ramenée par le n -échantillon $I_n(\theta)$ est égale à $nI(\theta)$
4. Si T est exhaustive $I_T(\theta) = I(\theta)$

Démonstration. 1) On a $E\left(\frac{\partial}{\partial \theta} \log L(\theta, x)\right) = 0$ $E\left(\frac{\partial}{\partial \theta} \log L(\theta, x)\right) = \int \frac{\partial}{\partial \theta} \log f_\theta(x)dx = \int \frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} f_\theta(x)dx =$

$$\int \frac{\partial}{\partial \theta} f_\theta(x)dx = \frac{\partial}{\partial \theta} \int f_\theta(x)dx = 0$$

$$\begin{aligned}
2) \quad E\left(\frac{\partial^2}{\partial \theta^2} \log L(\theta, x)\right) &= \int \frac{\partial}{\partial \theta} \left(\frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} \right) f_\theta(x) dx = \int \frac{f_\theta(x) \frac{\partial^2}{\partial \theta^2} f_\theta(x) - \left(\frac{\partial}{\partial \theta} f_\theta(x) \right)^2}{f_\theta(x)^2} f_\theta(x) dx = \\
&= \int \frac{\partial^2}{\partial \theta^2} f_\theta(x) dx - \int \left(\frac{\frac{\partial}{\partial \theta} f_\theta(x)}{f_\theta(x)} \right)^2 f_\theta(x) dx = -E\left(\left(\frac{\partial}{\partial \theta} \log L(\theta, x) \right)^2 \right) \quad \square
\end{aligned}$$

3.4.2 Cas d'un paramètre vectoriel

Cas d'un paramètre vectoriel $\theta = (\theta_1, \theta_2, \dots, \theta_p)$

L'information de Fischer est alors donnée par la matrice des variances-covariances

$$\mathcal{I}(\theta) = \mathcal{V}(\text{Grad} \log L(\theta, X))$$

$$\text{du vecteur } \text{Grad} \log L(\theta, X) = \begin{cases} \frac{\partial}{\partial \theta_1} \log L(\theta, X) \\ \frac{\partial}{\partial \theta_2} \log L(\theta, X) \\ \vdots \\ \frac{\partial}{\partial \theta_p} \log L(\theta, X) \end{cases}$$

Comme pour le cas réel on peut montrer que

$$\mathcal{I}(\theta) = -\left(E\left(\frac{\partial}{\partial \theta_i} \frac{\partial}{\partial \theta_j} \log L(\theta, X) \right) \right)_{1 \leq i \leq p, 1 \leq j \leq p},$$

Exemple 3.4. Soit x une observation de $X \sim N(m, \sigma^2)$

$$\begin{aligned}
\mathcal{I}(m, \sigma^2) &= - \begin{pmatrix} E\left(\frac{\partial^2}{\partial m^2} \log L(\theta, X) \right) & E\left(\frac{\partial}{\partial m} \frac{\partial}{\partial \sigma^2} \log L(\theta, X) \right) \\ E\left(\frac{\partial}{\partial m} \frac{\partial}{\partial \sigma^2} \log L(\theta, X) \right) & E\left(\frac{\partial^2}{\partial (\sigma^2)^2} \log L(\theta, X) \right) \end{pmatrix} \\
\mathcal{I}(m, \sigma^2) &= - \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}
\end{aligned}$$

Chapitre 4

Méthodes d'estimation ponctuelle

Pour certaines quantités, moments, variances, covariances, probabilités il a été vu qu'il existe des estimateurs simples ayant de bonnes qualités : sans biais et convergents. Pour des paramètres quelconques, il est nécessaire d'avoir des méthodes d'estimation conduisant à des estimateurs de bonne qualité. Il existe plusieurs grandes méthodes d'estimation

1. Méthode des moments
2. Méthode du maximum de vraisemblance
3. Recherche des estimateurs sans biais de variance minimale

Autres méthodes - estimation par les moindres carrés qui sera utilisée dans le contexte de la régression linéaire - estimation bayésienne qui ne sera pas abordée ici

4.1 Méthode des moments

Les moments empiriques sont des estimateurs sans biais et convergents des moments théoriques. La méthode des moments consiste à évaluer les uns aux autres. Soit le modèle d'échantillon $(\mathcal{X}, \mathcal{B}, (\mathcal{P}_\theta)_{\theta \in \Theta})^{(n)}$ associé à n observations (x_1, x_2, \dots, x_n) d'une v.a de loi P_θ .

On suppose $\theta = (\theta_1, \theta_2, \dots, \theta_p)$. On veut estimer θ i.e p paramètres.

$$E(X) = \overline{X_n} = \frac{\sum_{i=1}^n X_i}{n}$$
$$E(X^2) = \frac{\sum_{i=1}^n X_i^2}{n}$$
$$\dots\dots\dots$$
$$E(X^p) = \frac{\sum_{i=1}^n X_i^p}{n} :$$

On a ainsi un système de p équations à p inconnues à résoudre.

Exemple 4.1. 1. Soit le modèle d'échantillon poissonien $(N, \mathcal{P}(N), \{P(\lambda), \lambda \in R_+^*\})^{(n)}$ associé à n observations (x_1, x_2, \dots, x_n) d'une v.a de loi $P(\lambda)$.

Estimer λ par la méthode des moments.

$$E(X) = \lambda = \overline{x} \implies \hat{\lambda} = \overline{x}$$

2. Soit le modèle d'échantillon gaussien $(R, \mathcal{P}(R), \{N(m, \sigma^2), m \in R\})^{(n)}$. On suppose σ^2 connu.

Estimer m .

$$E(X) = m = \overline{x} \implies \hat{m} = \overline{x}$$

3. Soit le modèle d'échantillon gaussien $(R, \mathcal{P}(R), \{N(m, \sigma^2), m \in R\})^{(n)}$

Estimer m et σ^2

$$E(X) = m = \bar{x}$$

$$E(X^2) = \sigma^2 + m^2 = m_2 \implies$$

$$\hat{m} = \bar{x} \text{ et } \hat{\sigma}^2 = m_2 - \bar{x}^2 = S_n^2$$

4.2 Estimateurs du maximum de vraisemblance

Considérons une pièce de monnaie pour laquelle la probabilité d'obtenir pile est inconnue. Cette pièce est lancée 10 fois et on obtient 6 piles. Une méthode d'estimation intuitive est de considérer que l'évènement réalisé, ici 6 piles et 4 faces, a une forte probabilité (puisque'il s'est réalisé contrairement aux autres).

On estime alors p par la valeur \hat{p} de $[0,1]$ qui attribue à cet évènement, la plus forte probabilité, ou en d'autres termes par la valeur \hat{p} qui maximise la vraisemblance.

On a $L(p, \tilde{x}) = p^6(1-p)^4$

$$\frac{\partial}{\partial p} L(p, \tilde{x}) = 0 \iff \frac{\partial}{\partial p} \log L(p, \tilde{x}) = 0$$

$$\iff \hat{p} = \frac{6}{10}$$

Définition 4.1. Soit $\tilde{x} = (x_1, x_2, \dots, x_n)$ un n -échantillon d'une v.a X de loi $(f_\theta)_{\theta \in \Theta}$. On appelle estimateur du maximum de vraisemblance de θ , s'il existe, la statistique $\hat{\theta}$ telle que

$$L(\hat{\theta}, \tilde{x}) = \max_{\theta \in \Theta} L(\theta, \tilde{x})$$

La maximisation de la fonction de vraisemblance peut se faire de différentes manières

Cas 1 : θ réel et $L(\theta, \tilde{x})$ dérivable par rapport à θ

$\hat{\theta}$ est alors solution du système d'équations dit équations du maximum de vraisemblance

$$\begin{cases} \frac{\partial}{\partial \theta} L(\theta, \tilde{x}) = 0 : \text{condition du } 1^{er} \text{ ordre} \\ \frac{\partial^2}{\partial \theta^2} L(\theta, \tilde{x}) < 0 : \text{condition du } 2^{ème} \text{ ordre} \end{cases}$$

$$\iff \begin{cases} \frac{\partial}{\partial \theta} \log L(\theta, \tilde{x}) = 0 : \text{condition du } 1^{er} \text{ ordre} \\ \frac{\partial^2}{\partial \theta^2} \log L(\theta, \tilde{x}) < 0 : \text{condition du } 2^{ème} \text{ ordre} \end{cases}$$

Car la fonction log est continue et strictement monotone.

Cas 2 : θ réel et $L(\theta, \tilde{x})$ non dérivable par rapport à θ

Il faut maximiser la vraisemblance directement

Cas 3 : $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ vectoriel et $L(\theta, \tilde{x})$ dérivable par rapport à θ

$$\begin{cases} \frac{\partial}{\partial \theta_1} \log L(\theta, \tilde{x}) = 0 \\ \frac{\partial}{\partial \theta_2} \log L(\theta, \tilde{x}) = 0 \\ \dots \\ \frac{\partial}{\partial \theta_p} \log L(\theta, \tilde{x}) = 0 \end{cases} \quad \text{conditions du } 1^{er} \text{ ordre}$$

La condition du $2^{ème}$ ordre s'exprime à l'aide de la matrice Hessienne qui doit être définie négative.

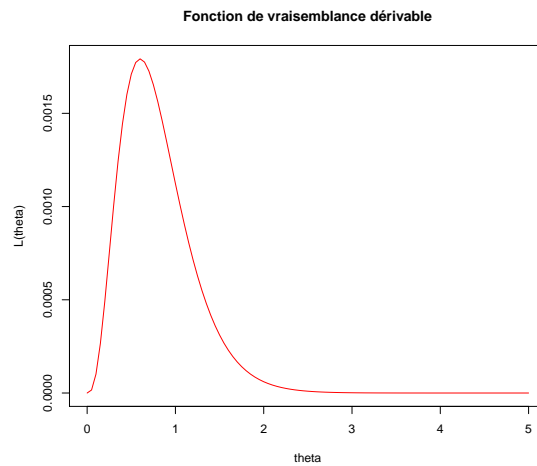


FIGURE 4.1 – Fonction de vraisemblance

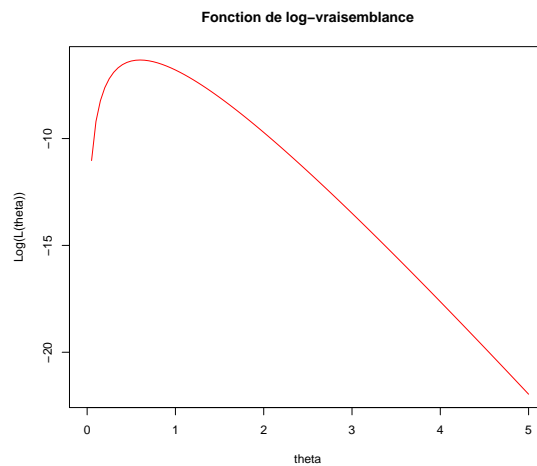


FIGURE 4.2 – Fonction de log-vraisemblance

Théorème 4.1. Soit $\tilde{x} = (x_1, x_2, \dots, x_n)$ un n -échantillon d'une v.a X de loi $(f_\theta)_{\theta \in \Theta}$. S'il existe une statistique exhaustive T pour θ alors l'estimateur du maximum de vraisemblance de θ est fonction de T .

Conséquence du théorème de factorisation

Théorème 4.2 (Invariance fonctionnelle). Soit $\hat{\theta}$ l'estimateur du maximum de vraisemblance de θ alors l'estimateur du maximum de vraisemblance de $g(\theta)$ est $g(\hat{\theta})$ pour toute fonction g .

4.3 Propriétés asymptotiques des EMV

Les estimateurs du maximum de vraisemblance n'ont pas de propriétés connus pour des échantillons de taille finie (sauf la propriété d'invariance fonctionnelle).

Cependant ils ont de très bonnes propriétés asymptotiques.

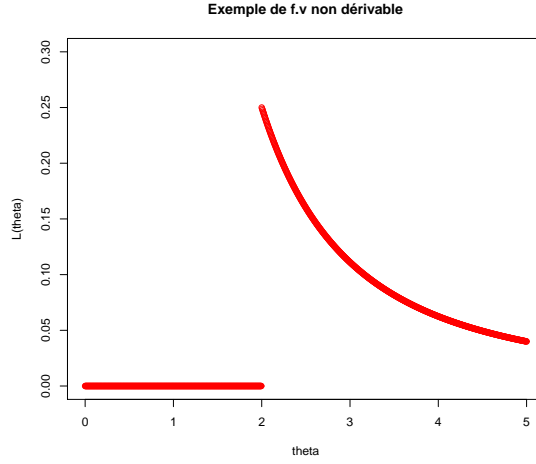


FIGURE 4.3 – Fonction de vraisemblance

Théorème 4.3. Soit $\widehat{\theta}_n$ l'EMV de θ basé sur un n -échantillon et $I_n(\theta)$ l'information de Fischer ramenée par les observations sur θ . (On suppose les conditions de régularité de Fischer vérifiées). On a

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{\frac{1}{I_n(\theta)}}} \xrightarrow[n \rightarrow +\infty]{\text{Loi}} X$$

Théorème 4.4. Soit g une fonction dérivable et $g(\widehat{\theta})$ l'EMV de $g(\theta)$ On a sous les hypothèses du théorème précédent

$$\frac{g(\widehat{\theta}_n) - g(\theta)}{\sqrt{\frac{g'(\theta)^2}{I_n(\theta)}}} \xrightarrow[n \rightarrow +\infty]{\text{Loi}} X$$

Les estimateurs du maximum de vraisemblance sont donc asymptotiquement gaussiens, asymptotiquement sans biais et asymptotiquement efficaces.

On a si n est grand

$$\widehat{\theta}_n \approx N\left(\theta, \frac{1}{I_n(\theta)}\right)$$

et

$$g(\widehat{\theta}_n) \approx N\left(g(\theta), \frac{g'(\theta)^2}{I_n(\theta)}\right)$$

Pour les applications le fait que la variance soit inconnue (dépend de θ) est gênant. On a cependant les théorèmes suivants obtenus en remplaçant la variance par la variance estimée.

Théorème 4.5. On a sous certaines hypothèses de régularité

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{\frac{1}{I_n(\widehat{\theta}_n)}}} \xrightarrow[n \rightarrow +\infty]{\text{Loi}} X$$

Théorème 4.6. *On a sous certaines hypothèses de régularité*

$$\frac{\widehat{g(\theta_n)} - g(\theta)}{\sqrt{\frac{g'(\widehat{\theta_n})^2}{I_n(\widehat{\theta_n})}}} \xrightarrow[n \rightarrow +\infty]{Loi} X$$

On peut alors écrire si n est grand

$$\widehat{\theta}_n \approx N\left(\theta, \frac{1}{I_n(\widehat{\theta}_n)}\right)$$

et

$$g(\widehat{\theta}_n) \approx N\left(g(\theta), \frac{g'(\widehat{\theta}_n)^2}{I_n(\widehat{\theta}_n)}\right)$$

4.4 Estimateurs sans biais de variance minimale

Définition 4.2. Soit T un estimateur de $g(\theta)$

On appelle risque quadratique de l'estimateur T , l'erreur quadratique moyenne de T , i.e

$$R_\theta(T) = E_\theta((T(X) - g(\theta))^2)$$

Proposition 4.7. *On a*

$$R_\theta(T) = E_\theta((T(X) - E_\theta(T(X)))^2) + (E_\theta(T(X)) - g(\theta))^2$$

i.e

$$R_\theta(T) = Var(T) + \text{biais}(T)^2$$

Définition 4.3. Un estimateur T_1 est dit meilleur qu'un autre estimateur T_2 au sens du risque quadratique si

$$R_\theta(T_1) \leq R_\theta(T_2) \quad \forall \theta \in \Theta$$

(uniformément en θ)

Proposition 4.8. *Parmi tous les estimateurs sans biais le meilleur au sens du risque quadratique est celui qui a la plus petite variance.*

$$Var_\theta(T_1) \leq Var_\theta(T_2) \quad \forall \theta \in \Theta$$

pour tout T_2 autre estimateur sans biais.

Si un tel estimateur existe, il est alors dit estimateur sans biais de variance minimale (ou uniformément minimale) : ESBVUM.

Théorème 4.9. *Soit T un estimateur sans biais de $g(\theta)$ et S une statistique exhaustive pour θ . Alors*

- 1) $E_\theta(T/S)$ est un estimateur sans biais de $g(\theta)$
- 2) $Var_\theta(E_\theta(T/S)) \leq Var_\theta(T)$

Conclusion : $E_\theta(T/S)$ est un meilleur estimateur sans biais de $g(\theta)$

Démonstration. $E_\theta(T/S)$ est un estimateur de $g(\theta)$ car ne dépend pas de θ .

$$1) E_\theta(E_\theta(T/S)) = E_\theta(T) = g(\theta)$$

(Théorème de l'espérance conditionnelle)

$$2) Var_\theta(T) = Var_\theta(E_\theta(T/S)) + E_\theta(Var_\theta(T/S))$$

(Théorème de la variance conditionnelle)

$$\implies Var_\theta(E_\theta(T/S)) = Var_\theta(T) - E_\theta(Var_\theta(T/S)) \leq Var_\theta(T)$$

□

Remarque : Si $T = h(S)$ alors $E_\theta(h(S)/S) = h(S) = T$

Dans ce cas il n'y a pas d'amélioration.

Théorème 4.10. *Théorème de Lehmann-Scheffé Soit T un estimateur sans biais de $g(\theta)$ et S une statistique exhaustive et complète pour θ . Alors*

$E_\theta(T/S)$ est l'unique estimateur sans biais de $g(\theta)$ dit estimateur sans biais de variance uniformément minimale (ESBVUM)

Démonstration

Théorème 4.11 (Théorème de Rao-Cramer). *Soit T un estimateur sans biais de $g(\theta)$. On suppose les conditions de régularité de Fischer vérifiées, alors*

$$Var_\theta(T) \geq \frac{(g'(\theta))^2}{I(\theta)}$$

La quantité $\frac{(g'(\theta))^2}{I(\theta)}$ est dite borne de Rao-Cramer (BRC)

$$Var_\theta(T) = \frac{(g'(\theta))^2}{I(\theta)}$$

Démonstration : Rappel :

$$|\rho_{X,Y}| = \left| \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \right| \leq 1 \implies Cov(X,Y)^2 \leq \sigma_X^2 \sigma_Y^2$$

$$E_\theta(T(X)) = \int_{\mathcal{X}} T(x) f_\theta(x) dx = \int_{\mathcal{X}} T(x) L(\theta, x) dx = g(\theta) \implies$$

$$\frac{\partial}{\partial \theta}(E_\theta(T(X))) = \int_{\mathcal{X}} T(x) \frac{\partial}{\partial \theta}(L(\theta, x)) dx = g'(\theta)$$

$$= \int_{\mathcal{X}} T(x) \frac{\partial}{\partial \theta}(\log L(\theta, x)) L(\theta, x) dx = Cov(T(X), \frac{\partial}{\partial \theta}(\log L(\theta, X)))$$

$$\implies Cov^2(T(X), \frac{\partial}{\partial \theta}(\log L(\theta, X))) \leq Var_\theta(T(X)) Var_\theta(\frac{\partial}{\partial \theta}(\log L(\theta, X)))$$

$$\implies Var_\theta(T(X)) \geq \frac{Cov^2(T(X), \frac{\partial}{\partial \theta}(\log L(\theta, X)))}{Var_\theta(\frac{\partial}{\partial \theta}(\log L(\theta, X)))} = \frac{g'(\theta)^2}{I(\theta)}$$

Cas particulier : $g(\theta) = \theta$

$$BCR = \frac{1}{I(\theta)}$$

Remarque 4. Pour que $Var_{\theta}(T(X)) \geq \frac{Cov^2(T(X), \frac{\partial}{\partial \theta}(\log L(\theta, X)))}{Var(\frac{\partial}{\partial \theta}(\log L(\theta, X)))} = \frac{g'(\theta)^2}{I(\theta)}$

Il est nécessaire et suffisant que $\frac{\partial}{\partial \theta}(\log L(\theta, X)) = k(\theta)T(X) + l(\theta)$

$$\implies E_{\theta}(\frac{\partial}{\partial \theta}(\log L(\theta, X))) = k(\theta)E_{\theta}(T(X)) + l(\theta)$$

$$\implies E_{\theta}(T(X)) = g(\theta) = -\frac{l(\theta)}{k(\theta)}$$

$$\iff \frac{\partial}{\partial \theta}(\log L(\theta, X)) = k(\theta)(T(X) - g(\theta))$$

Définition 4.4. Un estimateur sans biais T de $g(\theta)$ est dit efficace si

$$Var_{\theta}(T) = \frac{(g'(\theta))^2}{I(\theta)}$$

4.5 Estimation sans biais : cas vectoriel

Cas d'un paramètre vectoriel $\theta = (\theta_1, \theta_2, \dots, \theta_p)$. Les théorèmes de Rao-Blackwell et de Lehmann-Schéffé restent vraies dans ce cas. Le théorème de Rao-Cramer nécessite une adaptation.

Soit deux matrices définies positives A et B , on dit que $A \geq B$ si

$$Q_A(x) = x^t A x \geq Q_B(x) = x^t B x$$

On rappelle que les matrices de variances-covariances sont définies positives.

Théorème 4.12 (Théorème de Rao-Cramer : cas vectoriel). *Soit T un estimateur sans biais de $g(\theta)$ (à valeurs dans R^p). On suppose les conditions de régularité de Fischer vérifiées, alors*

$$\mathcal{V}_{\theta}(T) \geq Grad(g(\theta))^t \mathcal{I}(\theta)^{-1} Grad(g(\theta))$$

$\mathcal{V}_{\theta}(T)$ est la matrice de variances-covariances de T .