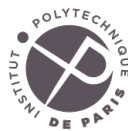


Statistiques Mathématiques
ENSAE - Master Spécialisé

Vincent Cottet

2023-2024



IP PARIS

Table des matières

1	Rappel de Probabilités	5
1.1	Espace Probabilisé	5
1.2	Variabes aléatoires réelles	5
1.3	Moments : espérance, variance, etc.	6
1.4	Couples et vecteurs aléatoires	6
1.5	Les modes de convergence	6
2	Modèle statistique, estimateurs	7
2.1	Modèle statistique	7
2.1.1	Modèle et statistique sur un modèle	7
2.1.2	Quelques types de modèles statistiques	8
2.2	Estimateur, biais et erreur quadratique	11
3	Propriétés asymptotique	17
3.1	Propriétés asymptotiques et théorèmes limites	17
3.2	Estimateur par substitution	17
3.3	Estimateur des moments	19
3.4	Estimateur du maximum de vraisemblance	21
4	Estimation par intervalle de confiance	23
4.1	Intervalle et région de confiance : définitions	23
4.2	Méthodes pour déterminer une région de confiance	23
4.2.1	Fonction pivotale	23
4.2.2	Utilisation des inégalités de Tchebychev et de Hoeffding	25
4.2.3	Inversion d'un test statistique	26
4.3	région de confiance asymptotique	26
4.3.1	Définition	26
4.3.2	Méthode du pivot asymptotique	26
4.3.3	Méthode de Wald	26
4.3.4	Inversion d'un test statistique	27
4.4	Choix des quantiles	27
5	Tests statistiques	29
5.1	Tests paramétriques	29
5.1.1	Hypothèse nulle et hypothèse alternative	29
5.1.2	Risque, niveau, puissance, efficacité	30
5.1.3	Tests asymptotiques	33
6	Travaux Dirigés : Énoncés	39
6.1	TD : Probabilités	39
6.2	TD : Planche 2 : Estimation, construction d'estimateurs	40
6.3	TD : Planche 3 : cadre asymptotique	41
6.4	TD : Planche 4	42

Chapitre 1

Rappel de Probabilités

On commence par des rappels de la théorie des probabilités. Le lecteur pourra se référer au polycopié de Benjamin Jourdain disponible en ligne pour un cours complet.

1.1 Espace Probabilisé

Définition 1 Un espace probabilisé est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ où :

- Ω est un ensemble ;
- \mathcal{A} une tribu sur Ω ;
- \mathbb{P} une mesure de probabilité sur \mathcal{A} :

$$\mathbb{P}(\emptyset) = 0, \quad \forall A \in \mathcal{A}, \mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$$
$$\forall \{(A_i) \in \mathcal{A}^{\mathbb{N}} : \forall i \neq j, A_i \cap A_j = \emptyset\}, \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

1.2 Variables aléatoires réelles

Définition 2 X est une variable aléatoire réelle si X est une fonction mesurable de (Ω, \mathcal{A}) dans $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$.

La fonction de répartition caractérise entièrement la loi d'une variable aléatoire réelle. Elle est la fonction :

$$F_X : \mathbb{R} \longrightarrow [0, 1]$$
$$x \longmapsto F_X(x) = \mathbb{P}(\omega : X(\omega) \leq x) = \mathbb{P}(X \leq x) = \mathbb{P}(X^{-1}(\cdot] - \infty, x])$$

C'est une fonction croissante, qui a pour limite 0 en $-\infty$ et 1 en $+\infty$.

Les variables aléatoires discrètes et continues

Une *variable aléatoire discrète* ne prend qu'un nombre dénombrable ou fini de valeurs, notées ici $(x_i)_{i \in I \subset \mathbb{N}}$. Sa densité par rapport à la mesure de comptage correspond aux probabilités des atomes : $\mathbb{P}(X = x_i) = p_i$. On a alors :

$$F_X(x) = \sum_{i=1}^{\infty} p_i 1_{x_i \leq x}.$$

Une variable aléatoire continue admet une densité définie sur \mathbb{R} , notée f_X par rapport à la mesure de Lebesgue (mesure sur \mathbb{R}). On a alors :

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

1.3 Moments : espérance, variance, etc.

La moyenne, si elle existe, est la quantité :

$$\mu = \mathbb{E}(X) = \int x dF_X(x) \begin{cases} \sum_{i \in I} x_i \mathbb{P}(X = x_i) & \text{si } X \text{ est discrète} \\ \int_{-\infty}^{+\infty} x f_X(x) dx & \text{si } X \text{ est continue} \end{cases}$$

Plus généralement, pour toute fonction φ mesurable, on a :

$$\mathbb{E}(\varphi(X)) = \int \varphi(x) dF_X(x) \begin{cases} \sum_{i \in I} \varphi(x_i) \mathbb{P}(X = x_i) & \text{si } X \text{ est discrète} \\ \int_{-\infty}^{+\infty} \varphi(x) f_X(x) dx & \text{si } X \text{ est continue} \end{cases}$$

La variance est le moment centré d'ordre 2, i.e. $\text{Var}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$

1.4 Couples et vecteurs aléatoires

Un couple aléatoire est une variable aléatoire à valeur dans \mathbb{R}^2 . Un vecteur aléatoire est une variable aléatoire à valeur dans \mathbb{R}^n . Un vecteur peut être discret ou continu (ou les deux, mais on ne traitera pas ce cas ici) et dans ces cas, leur densité est analogue au cas réel.

Un vecteur $X = (X_1, \dots, X_n)$ a des composantes indépendantes si et seulement si :

$$\forall (\mathcal{B}_1, \dots, \mathcal{B}_n) \in \mathbb{B}(\mathbb{R})^n, \mathbb{P}(X_1 \in \mathcal{B}_1, \dots, X_n \in \mathcal{B}_n) = \mathbb{P}(X_1 \in \mathcal{B}_1) \times \dots \times \mathbb{P}(X_n \in \mathcal{B}_n)$$

L'espérance est un vecteur ayant le nombre de composantes correspondantes :

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))^T.$$

La variance est alors une matrice (dite de variance-covariance), symétrique définie positive, qui se définit par :

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))(X - \mathbb{E}(X))^T].$$

1.5 Les modes de convergence

On se donne une suite de variable aléatoire X_1, X_2, \dots . Les modes de convergence suivant sont rangés du plus faible au plus fort.

On dit que la suite converge en loi vers la v.a. X si :

$$\forall t \in \mathbb{R} : F_X(t) \text{ est continu, } F_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} F_X(t).$$

On dit que la suite converge en probabilité vers la v.a. X si :

$$\forall \varepsilon > 0, \mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0.$$

On dit que la suite converge vers la v.a. X presque sûrement si :

$$\mathbb{P}(\omega : X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)) = 1.$$

Les théorèmes *limite*

Loi forte des grands nombres. Soient X_1, X_2, \dots une suite de v.a. i.i.d. telles que $\mathbb{E}(X_1) = \mu < \infty$. Alors :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{p.s.} \mu$$

Théorème de la limite centrale. Soient X_1, X_2, \dots une suite de v.a. i.i.d. telles que $\mathbb{E}(X_1) = \mu$ et $\mathbb{V}(X_1) = \sigma^2 < \infty$. Alors :

$$\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Cela peut se réécrire avec les fonctions de répartition (ϕ représente la f.d.r. d'une loi normale centrée réduite) :

$$\forall t \in \mathbb{R}, \mathbb{P} \left(\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq t \right) \xrightarrow[n \rightarrow \infty]{} \phi(t)$$

Chapitre 2

Modèle statistique, estimateurs

2.1 Modèle statistique

2.1.1 Modèle et statistique sur un modèle

Un modèle statistique est utilisé pour modéliser des observations venant d'un phénomène aléatoire. Par rapport à un espace probabilisé, on ne connaît pas la loi ayant servi à produire les observations (mais on fait l'hypothèse qu'il n'y en a qu'une) mais un ensemble de loi comprenant cette loi d'origine.

Définition 3 *Un modèle statistique $(\mathbb{X}, \mathfrak{F}, \mathcal{P})$ est la donnée de :*

- L'espace \mathbb{X} des observations.
- Une tribu \mathfrak{F} sur \mathbb{X} représentant l'ensemble des évènements sur \mathbb{X} .
- Une famille \mathcal{P} de mesures de probabilités sur \mathfrak{F} .

Un modèle statistique est une famille d'expériences aléatoires ayant même espace d'observations et même tribu d'évènements. Une observation x une réalisation d'une variable aléatoire X qui suit l'une des lois de probabilité (inconnue) de la famille \mathcal{P} . L'un des objectifs de l'inférence statistique est de déterminer cette loi de probabilité à partir des observations x .

On notera $X \sim \mathbb{P}$ pour indiquer que X suit la loi $\mathbb{P} \in \mathcal{P}$.

Exemple : épreuve de Bernoulli.

Soit X une va de loi de Bernoulli $b(\theta)$, où θ est inconnu. Soit $x \in \{0, 1\}$ le résultat d'un lancer à pile ou face ; x est une réalisation de X . Un choix raisonnable de la famille \mathcal{P} est l'ensemble des lois de Bernoulli de paramètre θ . L'espace des observations est $\mathbb{X} = \{0, 1\}$ et \mathfrak{F} est, par exemple, l'ensemble des parties de \mathbb{X} .

Exemple : loi uniforme.

Soit X une va de loi uniforme sur $[0, \theta]$, où θ est inconnu. Soit x une réalisation de X ; si $x = 0.1$, on peut en déduire que $\theta \geq 0.1$. Si $x = 2$, on peut en déduire que $\theta \geq 2$. Un choix raisonnable de la famille \mathcal{P} est l'ensemble des lois uniforme sur $[0, \theta]$. L'espace des observations est $\mathbb{X} = [0, \theta]$ et \mathfrak{F} est, par exemple, l'ensemble des boréliens de \mathbb{X} .

Observons n réalisations x_1, \dots, x_n d'une v.a. X , de manière indépendante. Soient X_1, \dots, X_n les v.a.i.i.d. sous-jacentes à ces observations. Les X_i sont à valeurs dans $(\mathbb{X}, \mathfrak{F})$. Soit \mathbb{P} la loi commune des X_i . (X_1, \dots, X_n) est appelé n -échantillon de X .

Définition 4 *On appelle modèle d'échantillonnage associé au n -échantillon, le triplet*

$$(\mathbb{X}^n, \mathfrak{F}^{\otimes n}, \{\mathbb{P}^{\otimes n}, \mathbb{P} \in \mathcal{P}\}) \quad (2.1)$$

où \mathcal{P} est une famille de probabilité définie sur $(\mathbb{X}, \mathfrak{F})$.

Par indépendance et identique distribution des X_i , $\mathbb{P}^{\otimes n}$ est la loi du vecteur (X_1, \dots, X_n) .

Exemple : répétition de n épreuves de Bernoulli indépendantes.

Le modèle d'échantillonnage associé est $(\{0, 1\}^n, \mathfrak{F}^{\otimes n}, \{b(\theta)^{\otimes n}, \theta \in [0, 1]\})$, avec $\mathfrak{F} = \{\emptyset, \{0\}, \{1\}, \Omega\}$

Ce modèle traduit simplement le fait que dans une expérience répétée de façon indépendante, la loi de probabilité du n -échantillon est la loi produit et l'espace sous-jacent est le produit cartésien de l'espace associé à l'un des éléments de l'échantillon.

Supposons que $n = 10$ et mettons qu'une réalisation du n -échantillon soit $(0, 0, 1, 1, 1, 0, 1, 0, 0, 1)$. Nous souhaitons évaluer θ à partir de cet échantillon. Comment peut-on procéder ?

Exemple : répétition de n épreuves d'une loi uniforme sur $[0, \theta]$.

Supposons que $n = 10$ et que nous disposions de l'échantillon suivant : $(0.1, 0.6, 0.2; 1.2; 3; 0.5; 1.5; 2; 1; 1.9)$. Comment estimer θ ?

Définition 5 Une statistique T sur un modèle $(\mathbb{X}, \mathfrak{F}, \mathcal{P})$ et à valeurs dans $(\mathbb{Y}, \mathfrak{T})$ est une application mesurable

$$T : (\mathbb{X}, \mathfrak{F}) \longrightarrow (\mathbb{Y}, \mathfrak{T}) \quad (2.2)$$

$$x \mapsto y = T(x) \quad (2.3)$$

Pour une réalisation x de X , $y = T(x)$ est la réalisation de la va $T(X)$.

Une statistique est donc simplement une fonction (mesurable) des *seules* observations de l'expérience (et pas du paramètre à estimer). C'est en particulier une variable aléatoire.

Exemple : répétition de n épreuves de Bernoulli indépendantes.

$$S(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i \quad (2.4)$$

Définit une statistique qui est la somme de toutes les observations du n -échantillon.

Toute statistique étant une fonction mesurable des données, elle permet de définir une variable aléatoire à partir des données initiales. La loi image P_S de \mathbb{P} par S permet alors de définir un modèle image à partir du modèle statistique initial.

Dans l'exemple précédent, la variables aléatoire est $S = T(X) = \sum_{i=1}^n X_i$ et la loi image est la loi binomiale $P_S \sim B(n, \theta)$. Le modèle image est donné par $(\mathbb{Y}, \mathfrak{T}, \{P_S : P \in \mathcal{P}\})$. Considérer S plutôt que X pour mener l'inférence consiste à travailler directement dans le modèle image par S :

$$\mathbb{P}[S \in B] = \mathbb{P}[S(X) \in B] = P_S(B) = \mathbb{P}[X \in S^{-1}(B)] \quad \forall B \in \mathcal{B}(\mathbb{R}) \quad (2.5)$$

Exemple : Statistique définie par le maximum d'une loi uniforme sur $[0, \theta]$.

$$S(X_1, X_2, \dots, X_n) = \sup_{i=1}^n X_i \quad (2.6)$$

L'information concernant la loi inconnue de X au travers de la statistique $T(X)$ est contenue dans la tribu $\mathfrak{F}(T) \subset \mathfrak{F}(X)$, et l'on a égalité si, et seulement si, T est bijective. Habituellement, $\mathfrak{F}(T)$ est plus petite que $\mathfrak{F}(X)$ car on attend d'une statistique qu'elle réduise la tribu et condense l'information.

2.1.2 Quelques types de modèles statistiques

Rappel : mesure dominée et théorème de Radon-Nikodym

Étant donné deux mesures σ -finies μ et ν sur l'espace $(\mathbb{X}, \mathfrak{F})$, on dit que ν est absolument continue par rapport μ et l'on note $\nu \ll \mu$, si

$$\forall B \in \mathfrak{F}, \mu(B) = 0 \Rightarrow \nu(B) = 0 \quad (2.7)$$

On rappelle qu'en ce cas, le théorème de Radon-Nikodym assure que ν possède une densité f par rapport à μ définie par

$$\nu(B) = \int_B f(x) d\mu(x) \quad (2.8)$$

On note $f = d\nu/d\mu$. f est positive, mesurable et s'appelle densité ou dérivée au sens de Radon-Nikodym de ν par rapport à μ .

Modèle dominé

Définition 6 Un modèle statistique $(\mathbb{X}, \mathfrak{F}, \mathcal{P})$ est dominé si, et seulement si, tout $\mathbb{P} \in \mathcal{P}$ est dominé par une même mesure μ . μ est la mesure dominante du modèle.

Si un modèle est dominé par une mesure σ -finie μ , alors le modèle image est dominé par la mesure μ_S . En effet, pour tout ensemble mesurable B ,

$$\mu_S(B) = 0 \iff \mu(S^{-1}(B)) = 0 \Rightarrow (\mathbb{P}(S^{-1}(B)) = 0 \iff \mathbb{P}_S(B) = 0) \quad (2.9)$$

Modèle paramétrique

Définition 7 Un modèle statistique $(\mathbb{X}, \mathfrak{F}, \mathcal{P})$ est paramétré si les éléments de \mathcal{P} peuvent être décrit par un paramètre. Cela veut dire qu'on peut écrire $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$.

Si $\theta \mapsto \mathbb{P}_\theta$ est injective on dit que le modèle est identifiable (cela veut dire que $\theta \neq \theta' \Rightarrow \mathbb{P}_\theta \neq \mathbb{P}_{\theta'}$).

Le modèle est paramétrique si $\Theta \subset \mathbb{R}^p$ pour un entier p donné. Sinon, il est non-paramétrique.

Exemple : épreuve de Bernoulli.

X suit de loi de Bernoulli $b(\theta)$ paramétrée par $\theta \in [0, 1]$. Le modèle est donc paramétrique, identifiable car à chaque valeur de θ correspond une unique loi de Bernoulli. Ce modèle est également dominé par la mesure de comptage discrète sur \mathbb{N} .

Si, au lieu de considérer le paramètre θ , on choisit comme paramètre $\cos^2 \theta$, le modèle n'est plus identifiable car $\mathbb{P}_\theta = \mathbb{P}_{\theta+\pi}$.

Exemple : modèle d'échantillonnage gaussien.

On considère un n -échantillon de loi gaussienne de moyenne m et écart type σ . Le modèle statistique correspondant

$$(\mathbb{R}^n, B(\mathbb{R})^{\otimes n}, \{\mathcal{N}(m, \sigma^2)^{\otimes n} : m \in \mathbb{R}, \sigma^2 > 0\}) \quad (2.10)$$

est paramétrique, dominé et identifiable. Il est paramétré par le couple (m, σ^2) , dominé par la mesure de Lebesgue sur \mathbb{R}^n et identifiable car toute loi gaussienne est caractérisée par sa moyenne et sa variance.

Exemple : modèles non paramétriques.

L'ensemble des lois de probabilités sur \mathbb{R} n'est pas paramétrique. Il en est de même de l'ensemble des lois qui ont une densité par rapport à la mesure de Lebesgue ou bien l'ensemble des lois réelles continues dont la densité est symétrique. De façon générale, lorsque le paramètre n'appartient pas à un espace vectoriel de dimension finie, on parle de modèle non paramétrique.

Exemple : modèles semi-paramétriques.

On peut s'intéresser à des problèmes dépendant d'un paramètre $\theta \in \mathbb{R}^p$, mais également d'un autre paramètre appartenant à un espace de dimension infinie. Ce second paramètre représente souvent un terme de nuisance. On parle en ce cas de problème semi-paramétrique.

Vraisemblance

Définition 8 On considère un modèle statistique $(\mathbb{X}, \mathfrak{F}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ dominé par une mesure μ σ -finie. La vraisemblance du modèle est la fonction $\theta \mapsto L(x, \theta)$ définie par

$$L(x, \theta) = d\mathbb{P}_\theta/d\mu(x) \quad (2.11)$$

La vraisemblance est exactement la densité de la loi \mathbb{P}_θ , vue comme fonction de θ .

Soient $(\mathbb{X}, \mathfrak{B}) = (\mathbb{R}, B(\mathbb{R}))$ et $P_\theta \ll \lambda$ mesure de Lebesgue sur \mathbb{R} ; supposons que P_θ possède une densité continue $f_\theta(x)$ par rapport à la mesure de Lebesgue et plaçons-nous dans un modèle d'échantillonnage. C'est un modèle paramétrique dominé par la mesure de Lebesgue sur \mathbb{R}^n et la vraisemblance de $P_\theta^{\otimes n}$ relativement à $\lambda^{\otimes n}$ est

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f_\theta(x_i) \quad (2.12)$$

Exemple : échantillon uniforme sur $[0, \theta]$.

$$L(x, \theta) = L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n \left(\frac{1}{\theta} \mathbb{1}_{[0, \theta]}(x_i) \right) = \frac{1}{\theta^n} \prod_{i=1}^n \mathbb{1}_{[0, \theta]}(x_i) = \frac{1}{\theta^n} \mathbb{1}_{[0, \theta]^n}(x) \quad (2.13)$$

Où $x = (x_1, \dots, x_n)$ et la dernière indicatrice vaut 1 si, et seulement si, toutes les coordonnées x_i du vecteur x appartiennent à $[0, \theta]$.

Plaçons-nous maintenant dans le cas où $(\mathbb{X}, \mathfrak{B}) = (\mathbb{N}, \mathcal{P}(\mathbb{N}))$, où $P_\theta \ll \delta$ mesure de comptage sur \mathbb{N} ; on a alors, pour tout sous ensemble B de \mathbb{N}

$$P_\theta[X \in B] = \sum_{x \in B} P_\theta[X = x] = \sum_{x \in \mathbb{N}} P_\theta[X = x] \delta_x(B) \quad (2.14)$$

Le modèle d'échantillonnage correspondant est dominé par la mesure discrète sur \mathbb{N}^n et la vraisemblance de $P_\theta^{\otimes n}$ par rapport à cette mesure est donnée par

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n P_\theta[X_i = x_i] \quad (2.15)$$

Exemple : échantillon de Bernoulli de paramètre θ .

$$L(x, \theta) = L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n (\theta^{x_i} (1 - \theta)^{1-x_i} \mathbb{1}_{\{0,1\}}(x_i)) = \theta^s (1 - \theta)^{n-s} \mathbb{1}_{\{0,1\}^n}(x) \quad (2.16)$$

où $x = (x_1, \dots, x_n)$, $s = \sum_{i=1}^n x_i$ et la dernière indicatrice vaut 1 si $x_i = 0$ ou 1 pour tout i .

Modèle homogène

Le support d'une mesure de probabilité \mathbb{P} dont la densité par rapport à une mesure μ est f est par définition

$$\text{supp}(f) = \{x \in \mathbb{R} : f(x) > 0\} \quad (2.17)$$

Ce support n'est défini qu'à un ensemble de mesure nulle près.

Définition 9 Un modèle paramétrique est homogène si le support des lois \mathbb{P}_θ est le même.

Cela signifie également que le support ne dépend pas de θ , ou encore que toutes les lois de \mathcal{P} sont équivalentes, c.-à-d. :

$$\forall \theta, \theta', \mathbb{P}_\theta \ll \mathbb{P}_{\theta'} \quad (2.18)$$

Homogène implique dominé et dans un modèle homogène, les ensembles négligeables sont les mêmes pour toutes les mesures \mathbb{P}_θ . Enfin, un modèle est homogène si, et seulement si,

$$\exists \nu \sigma\text{-finie} / \forall \theta, \mathbb{P}_\theta \ll \nu \text{ et } d\mathbb{P}_\theta/d\nu > 0 \nu\text{-p.p.} \quad (2.19)$$

2.2 Estimateur, biais et erreur quadratique

On se place dans un modèle statistique paramétrique $(\mathbb{X}, \mathfrak{F}, \mathcal{P})$ avec $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ et l'on considère une variable aléatoire $X \sim \mathbb{P}_\theta$.

Définition 10 *Un estimateur T de θ est une statistique à valeurs dans un sur-ensemble de Θ*

Un estimateur doit être une fonction des seules données et ne pas dépendre du paramètre θ qu'il est censé estimer. On attend bien sûr d'un estimateur qu'il donne une valeur proche de la vraie valeur de θ ...

Définition 11 *Le biais d'un estimateur T est la quantité*

$$b(\theta) = \mathbb{E}_\theta[T] - \theta \quad (2.20)$$

L'erreur quadratique moyenne de l'estimateur est

$$\mathbb{E}_\theta[(T - \theta)^2] = \|T - \theta\|_2^2 = \mathbb{V}_\theta(T) + b(\theta)^2 \quad (2.21)$$

Si $b(\theta) \neq 0$, on dit que l'estimateur est biaisé.

Exemple :

Considérons un n -échantillon X_1, \dots, X_n d'une va X d'espérance m et de variance σ^2 , sa moyenne empirique $\bar{X} = (\sum_{i=1}^n X_i)/n$ et sa variance empirique :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (2.22)$$

\bar{X} et S^2 sont des estimateurs respectifs de la moyenne m et de la variance σ^2 de X .

On voit facilement que $\mathbb{E}[\bar{X}] = m$ et un calcul rapide (à faire) donne $\mathbb{E}[S^2] = \frac{n-1}{n}\sigma^2$.

\bar{X} est donc un estimateur sans biais de m tandis que S^2 est un estimateur biaisé de la variance.

Estimateur par substitution

Définition 12 *Soit T un estimateur de θ et $\psi : \Theta \rightarrow \mathbb{R}^d$ une fonction mesurable. Un estimateur par substitution de $\psi(\theta)$ est un estimateur de la forme $\psi(T)$.*

Estimateur des moments

Étant donné une variable aléatoire X , on note $m_k(\theta) = \mathbb{E}_\theta[X^k]$ son moment théorique d'ordre k et

$$\hat{m}_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (2.23)$$

le moment empirique d'ordre k d'un n -échantillon de X .

Définition 13 *L'estimateur des moments de $\theta \in \Theta \subset \mathbb{R}^p$ est la solution $\hat{\theta}$ du système à p équations suivant, d'inconnue θ :*

$$\begin{cases} m_1(\theta) = \hat{m}_1 \\ \vdots \\ m_p(\theta) = \hat{m}_p \end{cases} \quad (2.24)$$

θ est un vecteur à p coordonnées ; dans le cas où $p = 1$, l'estimateur des moments est obtenu en identifiant moyenne théorique et empirique.

Exemple :

Considérons un n -échantillon (X_1, \dots, X_n) de X suivant une loi de Poisson de paramètre θ inconnu. On sait que $\mathbb{E}[X] = \theta$. On va donc poser comme estimateur des moments de θ la statistique $\bar{X} = \hat{m}_1$.

Exemple :

Considérons un n -échantillon (X_1, \dots, X_n) de X suivant une loi exponentielle (anglo-saxonne) de paramètre θ inconnu. On sait que $\mathbb{E}[X] = 1/\theta$. On va donc poser comme estimateur des moments de θ la statistique T_n telle que

$$\bar{X} = \hat{m}_1 = \frac{1}{T_n} \Rightarrow T_n = \frac{1}{\bar{X}} \quad (2.25)$$

Exemple :

L'estimateur des moments d'un n -échantillon de loi uniforme sur $[0, \theta]$ où θ est le paramètre inconnu est donné par la statistique $T_n = 2\bar{X}$ (à faire en exercice).

Exemple :

L'estimateur des moments d'un n -échantillon gaussien de loi $\mathcal{N}(m, \sigma^2)$ pour le paramètre $\theta = (m, \sigma^2)$ est donné par le couple (\bar{X}, S^2) (à faire en exercice).

Estimateur du maximum de vraisemblance

Définition, propriétés et exemples

Définition 14 *Un estimateur du maximum de vraisemblance de θ (on notera e.m.v.) est, sous réserve d'existence, une solution de l'équation*

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} L(X, \theta) \quad (2.26)$$

Ce maximum peut ne pas exister, n'être pas unique, ou être difficile à calculer.

Déterminer le maximum de vraisemblance est équivalent à déterminer le maximum de la log-vraisemblance (on notera cette log-vraisemblance $\log L(X, \theta) = l(X, \theta)$). Si L est différentiable, l'estimateur du maximum de vraisemblance est alors solution de l'équation

$$\frac{\partial}{\partial \theta} \log L(X, \theta) = \frac{\partial}{\partial \theta} l(X, \theta) = \nabla l(X, \theta) = S(X, \theta) = 0 \quad (2.27)$$

La dérivée S de la log-vraisemblance (ou son gradient dans le cas multidimensionnel) s'appelle le score.

L'équation précédente donne une condition nécessaire, mais non suffisante, pour être un maximum local. On rappelle qu'un maximum est local s'il existe un voisinage de $\hat{\theta}$ dans Θ tel que

$$L(x, \theta) < L(x, \hat{\theta}) \quad (2.28)$$

pour tout θ dans ce voisinage. Un maximum est global si l'inégalité précédente est vérifiée pour tout $\theta \in \Theta$.

Dans le cas où Θ est un intervalle de \mathbb{R} (le paramètre à estimer est scalaire), une condition suffisante d'existence et d'unicité est que $L(x, \theta)$ soit une fonction de θ de classe C^2 sur Θ (x étant fixé) et que les deux conditions suivantes soient réalisées :

$$\begin{cases} \frac{\partial l}{\partial \theta}(x, \theta) = 0 \\ \frac{\partial^2 l}{\partial \theta^2}(x, \theta) < 0 \end{cases} \quad (2.29)$$

La solution doit par ailleurs appartenir à l'intérieur $\overset{\circ}{\Theta}$ de Θ . Les conditions assurent l'existence et l'unicité d'un maximum local uniquement et la valeur $\hat{\theta}$ qui réalise ce maximum n'est donc pas forcément l'estimateur du maximum de vraisemblance (car il peut ne pas être un maximum global).

On a existence et unicité d'un maximum global si la fonction l est strictement concave sur Θ (en tant que fonction de Θ).

Le cas multidimensionnel va être abordé après les quelques exemples qui suivent.

Exemple :

Soit $X = (X_1, \dots, X_n)$ un échantillon gaussien $\mathcal{N}(m, \sigma^2)$ et $\theta = (m, \sigma^2)$.

Le modèle est paramétrique, dominé par la mesure de Lebesgue et la vraisemblance associée est

$$L(x, \theta) = \prod_{i=1}^n f(x_i, \theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2} \log \sigma^2\right) \quad (2.30)$$

Le support du modèle est \mathbb{R}^n ; celui-ci est donc homogène et la vraisemblance est clairement de classe C^2 en θ (à condition que $\sigma > 0$). Par conséquent, l'e.m.v. est solution de l'équation de vraisemblance

$$\frac{d}{d\theta} l(x, \theta) = \nabla_l(\theta) = 0 \quad (2.31)$$

où $l(x, \theta) = \log L(x, \theta)$. La solution est unique si la matrice hessienne $H_l(\hat{\theta})$ est définie négative.

$$l(x, \theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2} \log \sigma^2 - \frac{n}{2} \log(2\pi) \quad (2.32)$$

$l(x, \theta)$ est une fonction vectorielle en θ , car $\theta = (m, \sigma^2)$ est de dimension 2. La dérivée de l est donc à comprendre ici comme le gradient de cette fonction, dont les coordonnées sont les dérivées partielles en m et en σ^2 .

$$\frac{d}{d\theta} l(x, \theta) = \nabla_l(\theta) = \left(\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - m); \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2\sigma^2} \right)' \quad (2.33)$$

Ce gradient s'annule si, et seulement si

$$\sum_{i=1}^n (x_i - m) = 0 \text{ et } \sum_{i=1}^n (x_i - m)^2 = n\sigma^2 \quad (2.34)$$

$$\iff m = \bar{x} \text{ et } \sigma^2 = S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.35)$$

Un point critique de l est donc $\hat{\theta} = (\bar{X}, S^2)$. La matrice hessienne de l en ce point est (à vérifier à titre d'exercice)

$$H_l(\hat{\theta}) = \begin{pmatrix} -n/S^2 & 0 \\ 0 & -n/(2S^4) \end{pmatrix} \quad (2.36)$$

qui est clairement définie négative. Par suite, $\hat{\theta}$ est l'unique e.m.v. Pour vérifier les calculs précédents, on aura intérêt à poser $S^2 = V$ pour éviter les erreurs au moment de la dérivations en S^2 .

Exemple :

Soit $X = (X_1, \dots, X_n)$ un n -échantillon de loi uniforme sur $[0, \theta]$ où $\theta > 0$.

Le modèle est paramétrique, dominé par la mesure de Lebesgue sur \mathbb{R}^n . La vraisemblance de l'échantillon est

$$L(x, \theta) = \prod_{i=1}^n \left(\frac{1}{\theta} \mathbb{1}_{[0, \theta]}(x_i) \right) = \frac{1}{\theta^n} \mathbb{1}_{[0, \theta]^n}(x) = \frac{1}{\theta^n} \mathbb{1}_{[0 \leq x_{(1)} \leq x_{(n)} \leq \theta]}(x) \quad (2.37)$$

où $x_{(1)}$ est le min des x_i et $x_{(n)}$ le max. Le support dépend de θ et le modèle n'est donc pas homogène. Nous ne pouvons pas déterminer le maximum de vraisemblance par la méthode précédente. On remarque par contre que, à x fixé, la fonction de vraisemblance est décroissante en θ à partir de $x_{(n)}$ (avant, elle est nulle). Le maximum est donc atteint en $\hat{\theta} = x_{(n)}$. L'e.m.v. existe et est unique :

$$\hat{\theta}_{MV} = X_{(n)} = \max_{i=1, \dots, n} X_i \quad (2.38)$$

On montre facilement que $\theta X_{(n)}$ suit une loi bêta de paramètres $(n, 1)$. La densité du maximum est donc

$$\frac{ny^{n-1}}{\theta} \mathbb{1}_{[0, \theta]}(y) \quad (2.39)$$

Ainsi, $\mathbb{E}[X_{(n)}] = \frac{n}{n+1}\theta$ et l'e.m.v. a donc un biais égal à $b(\theta) = -\frac{\theta}{n+1}$.

Exemple :

Si X_1, \dots, X_n est un n -échantillon de loi uniforme sur $[\theta, \theta + 1]$, on peut montrer (en exercice) que tout estimateur de θ compris entre $X_{(n)} - 1$ et $X_{(1)}$ est un estimateur du maximum de vraisemblance. Il n'y a alors pas unicité de cet e.m.v.

L'e.m.v. peut ne pas exister, par exemple dans le cas où la vraisemblance L n'est pas une fonction continue de θ . Pour avoir un estimateur intéressant, il faudra donc *a minima* supposer la vraisemblance continue, voire différentiable et également imposer à Θ certaines propriétés :

Théorème 1 (Condition suffisante d'existence de l'e.m.v.) *Si Θ est un espace compact et si l est continue sur Θ , alors l'estimateur du maximum de vraisemblance existe.*

Théorème 2 *Sous les trois hypothèses de régularité (\star) suivantes :*

- *Le modèle est identifiable et homogène.*
- *Θ est un ouvert ou un compact de \mathbb{R}^p .*
- *Pour tout x , la fonction $L(x, \theta)$ est de classe C^2 en θ sur Θ .*

Alors l'estimateur du maximum de vraisemblance $\hat{\theta}$ est solution \mathbb{P}_θ -p.s. du système

$$\begin{cases} \nabla l(x, \theta) = 0 \\ H_l(x, \theta) < 0 \end{cases} \quad (2.40)$$

où $l(x, \theta) = \log L(x, \theta)$ est la log-vraisemblance de l'échantillon, ∇l le gradient de l en θ (x est fixé) et H_l est la matrice hessienne de l en θ (à x fixé).

La condition $H_l(x, \theta) < 0$ signifie que la matrice hessienne est définie négative en θ . Le théorème n'assure toujours pas l'unicité de l'e.m.v. (il est unique localement). Une condition suffisante pour garantir l'unicité est de s'assurer que Θ est un espace convexe et que l est strictement concave sur Θ .

Reparamétrisation

Trouver un estimateur par substitution de l'estimateur du maximum de vraisemblance s'appelle une reparamétrisation. On montre que quelque soit l'application ψ , l'e.m.v. est invariant par reparamétrisation.

Théorème 3 (de Zehna) *Soit $\hat{\theta}$ l'e.m.v. de θ et ψ une fonction quelconque. Alors l'e.m.v. de $\psi(\theta)$ est $\psi(\hat{\theta})$*

Démonstration :

Soit $\eta = \psi(\theta)$. Si ψ est bijective, on a immédiatement

$$\tilde{L}(x, \eta) = L(x, \psi^{-1}(\eta)) \quad (2.41)$$

Dans le cas contraire,

$$\tilde{L}(x, \eta) = \sup_{\theta/\psi(\theta)=\eta} L(x, \theta) \quad (2.42)$$

Soit $\hat{\eta}$ l'e.m.v. de η . Alors par définition, $\hat{\eta}$ maximise la vraisemblance :

$$\tilde{L}(x, \hat{\eta}) = \sup_{\eta} \tilde{L}(x, \eta) = \sup_{\eta} \sup_{\theta/\psi(\theta)=\eta} L(x, \theta) = L(x, \hat{\theta}) = \tilde{L}(x, \psi(\hat{\theta})) \quad (2.43)$$

Par identification, on a alors $\hat{\eta} = \psi(\hat{\theta})$.

□

Chapitre 3

Propriétés asymptotique

3.1 Propriétés asymptotiques et théorèmes limites

Considérons un estimateur T_n de θ dans un modèle d'échantillonnage. On dispose alors d'un échantillon X_1, \dots, X_n de v.a.i.i.d. de loi \mathbb{P}_θ et T_n est une fonction $T(X_1, \dots, X_n)$ de (X_1, \dots, X_n) , de sorte que T_n dépend de n . On dit que :

T_n est asymptotiquement sans biais si

$$\lim_{n \rightarrow +\infty} \mathbb{E}_\theta[T_n] = \theta \quad (3.1)$$

T_n est fortement consistant pour θ si

$$T_n \xrightarrow{\mathbb{P}_\theta\text{-P.s.}} \theta \iff \forall \theta, \mathbb{P}_\theta \left[\lim_{n \rightarrow +\infty} T_n = \theta \right] = 1 \quad (3.2)$$

T_n converge en moyenne quadratique si

$$T_n \xrightarrow{L^2} \theta \iff \lim_{n \rightarrow +\infty} \|T_n - \theta\|_2^2 = \lim_{n \rightarrow +\infty} \mathbb{E} [|T_n - \theta|^2] = 0 \quad (3.3)$$

T_n est consistant pour θ si

$$T_n \xrightarrow{\mathbb{P}_\theta} \theta \iff \forall \epsilon > 0, \lim_{n \rightarrow +\infty} \mathbb{P}_\theta [|T_n - \theta| > \epsilon] = 0 \quad (3.4)$$

On note également $T_n = \theta + o_{\mathbb{P}}(1)$. Dire que $X_n = o_{\mathbb{P}}(1)$ signifie que X_n tend vers 0 en probabilité (sous \mathbb{P}).

La consistance forte implique la consistance et la convergence en moyenne quadratique implique la consistance. On parle de v_n -consistance pour indiquer que v_n est la vitesse de convergence. Dire que T_n est v_n -consistant revient à dire que la suite est bornée en probabilité, c.-à-d. que :

$$v_n(T_n - \theta) = \mathcal{O}_{\mathbb{P}}(1) \quad (3.5)$$

On dit que l'estimateur T_n est v_n -consistant et asymptotiquement normal si

$$v_n(T_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_\theta) \quad (3.6)$$

Σ_θ est une matrice $p \times p$ symétrique et positive (p est la dimension de θ) ; c'est la matrice de covariance de la loi limite.

3.2 Estimateur par substitution

Théorème 4 (de l'application continue) Soit T un estimateur (fortement) consistant de θ . Soit $\psi : \Theta \subset \mathbb{R}^p \rightarrow \mathbb{R}^d$ une fonction continue. Alors $\psi(T)$ est (fortement) consistant pour $\psi(\theta)$.

La propriété est également vraie pour la convergence en loi ; c'est alors une conséquence du théorème de Portmanteau [Portmanteau veut dire *fourre tout* en anglais. Patrick Billingsley a écrit un ouvrage célèbre sur les convergences de variables aléatoires : *Convergence of Probability Measures*. Dans cet ouvrage, il attribue le théorème à Jean-Pierre Portmanteau, de l'université de Felletin, petite commune française de la creuse. L'article est daté de 1915 et intitulé *espoir pour l'ensemble vide*. Il s'agit d'un canular. Le théorème a en fait été prouvé par Alexandre Alexandrov vers 1940. Alexandrov était le directeur de thèse de Grigori Perelman, qui a obtenu la médaille Fields pour avoir démontré la conjecture de Poincaré. Il a refusé cette médaille, ainsi que le prix Wolf d'un million de dollars qui l'accompagnait.]

Théorème 5 (Méthode delta) Soit T_n un estimateur de θ tel que $v_n(T_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} T$
Soit $\psi : \Theta \subset \mathbb{R}^p \rightarrow \mathbb{R}^d$ une fonction différentiable de matrice jacobienne J . Alors

$$v_n(\psi(T_n) - \psi(\theta)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} d\psi_\theta(T) \quad (3.7)$$

et si T_n est asymptotiquement normal avec

$$v_n(T_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_\theta) \quad (3.8)$$

alors $\psi(T_n)$ est asymptotiquement normal et

$$v_n(\psi(T_n) - \psi(\theta)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, J \times \Sigma_\theta \times J') \quad (3.9)$$

Remarques et cas de la dimension 1 :

$d\psi_\theta(h)$ est la différentielle de ψ au point θ , calculée en h . Lorsque $d = 1$, la différentielle est égale à $\psi'(\theta) \times h$. Si nous notons $J_\psi(\theta)$ la matrice jacobienne de ψ en θ , alors $d\psi_\theta(h) = J_\psi(\theta) \times h$.

$J = J_\psi(\theta) \in \mathbb{R}^{d \times p}$ est la matrice jacobienne de ψ et $J' = J'_\psi(\theta)$ sa transposée. Dans le cas de la dimension 1, on a simplement $J\Sigma J' = \psi'(\theta)^2 \times \sigma_\theta^2$, où σ_θ^2 est la variance de la loi normale asymptotique.

Exemple :

Soit (X_1, \dots, X_n) un n -échantillon d'une v.a. X de loi de Bernoulli de paramètre $\theta \in]0, 1[$. Soit \bar{X}_n la moyenne empirique de l'échantillon. D'après le théorème de la limite centrale,

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \theta(1 - \theta)) \quad (3.10)$$

Posons $\psi(\theta) = 2 \arcsin \sqrt{\theta}$. $\psi(\bar{X}_n)$ est un estimateur de $\psi(\theta)$ et puisque ψ est dérivable, alors $\psi(\bar{X}_n)$ est \sqrt{n} -consistant et asymptotiquement normal.

Comme

$$\psi'(\theta) = \frac{1}{\sqrt{\theta(1 - \theta)}} \quad (3.11)$$

Alors $J\Sigma J' = \psi'(\theta)^2 \times \theta(1 - \theta) = 1$ et l'on a donc

$$\sqrt{n} \left(2 \arcsin \sqrt{\bar{X}_n} - 2 \arcsin \sqrt{\theta} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \quad (3.12)$$

Exemple :

Soit (X_1, \dots, X_n) un n -échantillon d'une v.a. X de loi exponentielle de paramètre $\theta > 0$. Soit \bar{X}_n la moyenne empirique de l'échantillon et soit $\lambda = 1/\theta$. On sait que

$$\mathbb{E}[X] = \lambda \text{ et } \mathbb{V}(X) = \lambda^2 \quad (3.13)$$

D'après le théorème de la limite centrale, $\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \lambda^2)$

Posons $\psi(x) = 1/x$ qui est différentiable sur $]0, +\infty[$ avec $\psi'(x)^2 = 1/x^4$. En appliquant la méthode delta, on en déduit que

$$\frac{\sqrt{n}}{\theta} \left(\frac{1}{\bar{X}_n} - \theta \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \quad (3.14)$$

Exemple :

Considérons un n -échantillon (X_1, \dots, X_n) d'une v.a. X telle que $\mathbb{E}[X^4] < \infty$. On note $\alpha_i = \mathbb{E}[X^i]$ pour $i = 1, \dots, 4$ et $\alpha = (\alpha_1, \alpha_2)'$. On note également

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \quad \text{et} \quad S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (3.15)$$

Soit $T_n = (\bar{X}, \bar{X}^2)'$. Soit $\psi(x, y) = y - x^2$. Alors $S^2 = \psi(\bar{X}, \bar{X}^2)$. Le théorème de la limite centrale multidimensionnel s'applique et donne

$$\sqrt{n}(T_n - \alpha) = \sqrt{n} \left(\begin{pmatrix} \bar{X} \\ \bar{X}^2 \end{pmatrix} - \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \alpha_2 - \alpha_1^2 & \alpha_3 - \alpha_1\alpha_2 \\ \alpha_3 - \alpha_1\alpha_2 & \alpha_4 - \alpha_2^2 \end{pmatrix} \right) \quad (3.16)$$

La fonction ψ est polynomiale donc différentiable et sa différentielle en (x, y) est $d\psi_{(x,y)}(h, k) = -2xh + k$. La méthode delta appliquée à ψ donne alors

$$\sqrt{n}(S^2 - (\alpha_2 - \alpha_1^2)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, -4\alpha_1^4 - \alpha_2^2 + 8\alpha_1^2\alpha_2 - 4\alpha_1\alpha_3 + \alpha_4) \quad (3.17)$$

Cette loi normale est la loi de la variable aléatoire $-2\alpha_1 T_1 + T_2$ où $(T_1, T_2)'$ est le vecteur limite de $\sqrt{n}(T_n - \alpha)$.

Lorsque les variables sont centrées, c'est à dire lorsque $\alpha_1 = 0$, le résultat devient

$$\sqrt{n}(S^2 - \alpha_2) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \alpha_4 - \alpha_2^2) \quad (3.18)$$

Exemple :

Soit $(X_1, Y_1), \dots, (X_n, Y_n)$ un n -échantillon d'un vecteur gaussien (X, Y) dont le coefficient de corrélation linéaire est noté ρ . Le coefficient de corrélation empirique est par définition égal à

$$\hat{\rho}_n = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \times \sum_{i=1}^n (Y_i - \bar{Y})^2 \right]^{1/2}} \quad (3.19)$$

Le théorème de la limite centrale et la méthode delta montrent que $\sqrt{n}(\hat{\rho}_n - \rho)$ est asymptotiquement normale. On peut montrer (à faire en exercice) que

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, (1 - \rho^2)^2) \quad (3.20)$$

La méthode de stabilisation de la variance amène à chercher une fonction différentiable ψ dont la dérivée vaut $(1 - \rho^2)^{-1}$. On pose donc naturellement $\psi(\rho) = \operatorname{arctanh} \rho$ et une nouvelle application de la méthode delta montre que

$$\sqrt{n}(\psi(\hat{\rho}_n) - \psi(\rho)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \quad (3.21)$$

3.3 Estimateur des moments

Soit $M(\theta) = (m_1(\theta), \dots, m_p(\theta))'$ le vecteur des moments d'ordre 1 à p de X et \widehat{M} celui des moments empiriques (les moments empiriques dépendent de θ via les v.a. X_i).

Propriété 1 • Si $M : \Theta \rightarrow \mathbb{R}^p$ est bijective, alors

$$\widehat{\theta} = M^{-1}(\widehat{M}) \quad (3.22)$$

$$\widehat{M} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_{\theta-p.s.}} M(\theta) \quad (3.23)$$

• Si M est bijective et M^{-1} est continue, alors

$$\widehat{\theta} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}_{\theta-p.s.}} \theta \quad (3.24)$$

• Si de plus M^{-1} est différentiable et \mathbb{P}_{θ} a un moment d'ordre 2, alors

$$\sqrt{n} \left(\widehat{M} - M(\theta) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Sigma_{\theta}) \quad (3.25)$$

et

$$\sqrt{n} \left(\widehat{\theta} - \theta \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \widetilde{\Sigma}_{\theta}) \quad (3.26)$$

avec $\Sigma_{\theta} = (\text{cov}(X^i, X^j))_{i,j=1..p}$ et $\widetilde{\Sigma}_{\theta} = J_M^{-1} \Sigma_{\theta} (J_M^{-1})'$

Démonstration

La méthode des moments met en jeu des estimateurs plug-in dont les fonctionnelles statistiques sous-jacentes sont simplement les moments des v.a. X_i . Les propriétés de la méthode des moments découlent immédiatement des propriétés des estimateurs plug-in du paragraphe précédent :

• Si M est bijective, alors $\widehat{\theta}$ est défini de façon unique et la loi forte des grands nombres assure la convergence presque sûre de \widehat{M} vers M (les X_i sont des v.a.i.i.d. et le moment d'ordre p existe par hypothèse).

• Si de plus M^{-1} est continue, alors le théorème de l'application continue montre que $\widehat{\theta}$ converge p.s. vers θ et l'on a donc consistance forte.

• Si M^{-1} est différentiable, le théorème de la limite centrale et la méthode delta s'appliquent (si $p > 1$).

J_M est la matrice jacobienne de M ; en notant m la variable de M , le théorème d'inversion locale permet d'écrire que

$$J_{M^{-1}}(m) = J_M(\theta)^{-1} \quad (3.27)$$

□

Exemple :

Soit (X_1, \dots, X_n) un n -échantillon de $X \sim B(\alpha, \beta)$, dont la densité est

$$g(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbf{1}_{[0,1]}(x) \quad (3.28)$$

avec $\alpha, \beta \in [0, 1]$. L'estimateur des moments de (α, β) est la solution de

$$\begin{cases} \overline{X} = \mathbb{E}[X] = \frac{\alpha}{\alpha + \beta} \\ \overline{X^2} = \mathbb{E}[X^2] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \end{cases} \quad (3.29)$$

La fonction

$$\psi(x, y) = \left(\frac{x}{x+y}, \frac{x(x+1)}{(x+y)(x+y+1)} \right) \quad (3.30)$$

est de classe C^1 et inversible sur $]0, 1[^2$. En inversant cette fonction, on trouve l'expression des estimateurs en fonction de \overline{X} et $\overline{X^2}$:

$$\begin{cases} \widehat{\alpha} = \frac{\overline{X}(\overline{X^2} - \overline{X})}{\overline{X^2} - \overline{X}^2} \\ \widehat{\beta} = \frac{(\overline{X^2} - \overline{X})(1 - \overline{X})}{\overline{X^2} - \overline{X}^2} \end{cases} \quad (3.31)$$

et ces estimateurs sont consistants et asymptotiquement normaux (à faire en exercice).

3.4 Estimateur du maximum de vraisemblance

C'est Fisher qui a le premier étudié, en 1921 et 1925, l'estimateur du maximum de vraisemblance. En 1946, Cramer a donné une première preuve de la consistance de cet estimateur et de sa normalité asymptotique, sous certaines conditions de régularité. Wald a démontré la consistance forte en 1949, sous des hypothèses plus générale. Rappelons que $S = \partial l / \partial \theta$ s'appelle le score du modèle.

Il existe une foultitude de conditions de régularité différentes selon les ouvrages, qui conduisent à la consistance forte ou faible de l'e.m.v. Voici un énoncé dans le cadre multidimensionnel, avec des hypothèses de régularité que nous adopterons par défaut et qui sont suffisamment pratiques pour couvrir la plupart des cas classiquement étudiés.

On note $L(x, \theta)$ la vraisemblance d'une observation, $l(x, \theta)$ son logarithme et $L_n(X, \theta)$ la vraisemblance de l'échantillon de taille n ; $f_\theta(x)$ est la densité. On se place ici dans le cas unidimensionnel mais le cas multidimensionnel se traite de manière identique.

Théorème 6 (Consistance forte de l'e.m.v.) *On considère le modèle d'échantillonnage X_1, \dots, X_n de v.a.i.i.d. $\sim f_{\theta^*}$. Sous les hypothèses de régularité ci-dessous,*

- *Le modèle est identifiable et homogène.*
- *Θ est un ouvert ou un compact de \mathbb{R}^p .*
- *Pour tout x , la fonction $L(x, \theta)$ est différentiable en θ sur Θ .*

Alors l'estimateur du maximum de vraisemblance est fortement consistant.

La normalité asymptotique de l'e.m.v. requiert des conditions supplémentaires. Avant cela, il est nécessaire d'introduire la quantité dite *Information de Fisher*.

Définition 15 *Supposons que la fonction $\theta \mapsto l(x, \theta)$ est différentiable pour presque tout x . Alors la fonction $I(\cdot)$ définie sur Θ :*

$$I(\theta) = \int l'(x, \theta)^2 f_\theta(x) d\mu(x) = \mathbb{E}_\theta([l'(X, \theta)]^2)$$

est appelée Information de Fisher associée à une observation. C'est la variance du score.

Afin d'aller plus loin, il est nécessaire de fixer les conditions supplémentaires. Celles-ci portent sur la régularité du modèle.

Définition 16 *Un modèle paramétrique est régulier si, et seulement si,*

- *Il est dominé, homogène et $\Theta \subset \mathbb{R}^p$ est un ouvert.*
- *Pour presque tout x , les fonctions $\theta \mapsto l(x, \theta)$ et $\theta \mapsto f_\theta(x)$ sont deux fois continuellement dérivables sur Θ .*
- *Pour tout $\theta \in \Theta$ il existe un intervalle ouvert U contenant θ^* et une fonction borélienne $\Lambda(x)$ tels que $|l''(x, \theta)| \leq \Lambda(x)$, $|l'(x, \theta)| \leq \Lambda(x)$, $|l'(x, \theta)^2| \leq \Lambda(x)$, pour tout $\theta \in U$ et presque tout x , et*

$$\int \Lambda(x) \sup_{\theta \in U} f_\theta(x) d\mu(x) < \infty$$

- *L'information de Fisher $I(\theta)$ est positive.*

Les modèles gaussiens ou de Poisson sont réguliers, mais pas le modèle uniforme sur $[0, \theta]$. Habituellement, on ne vérifie pas les conditions techniques et cela ne sera demandé. Elles sont vérifiées dans tous les cas d'exercice sauf la question du support, qui est primordiale et doit être vérifiée.

Théorème 7 *Si le modèle est régulier, l'information de Fisher du modèle est l'espérance de la dérivée seconde de la log-vraisemblance :*

$$\mathbb{I}(\theta) = -\mathbb{E}_\theta[l''(X, \theta)] \tag{3.32}$$

Cela peut se voir comme la courbure de la vraisemblance limite. C'est l'information de Fisher en θ^* qui apparaît après ; si cette quantité est grande, cela veut dire que la vraisemblance limite est "piquée" en ce point, il est donc plus facile d'estimer le paramètre. On peut alors énoncer la normalité asymptotique :

Théorème 8 *On considère le modèle régulier où $X \sim \mathbb{P}_{\theta^*}$. La suite d'estimateurs du maximum de vraisemblance est notée $\hat{\theta}_n$. Alors, pour tout $\theta^* \in \Theta$, on a :*

$$\sqrt{n} \left(\hat{\theta}_n - \theta^* \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(0, I^{-1}(\theta^*) \right)$$

(Hors programme :) L'estimateur du maximum de vraisemblance est dit "asymptotiquement efficace" car on peut montrer qu'aucun estimateur sans biais ne peut avoir une variance plus faible (c'est la borne de Fréchet-Darmois-Cramer-Rao). Néanmoins, ce résultat est peu utile.

Chapitre 4

Estimation par intervalle de confiance

4.1 Intervalle et région de confiance : définitions

Définition 17 Soient T_1, T_2 deux statistiques à valeurs réelles telles que $T_1 \leq T_2$ \mathbb{P}_θ -p.s. pour tout $\theta \in \Theta \subset \mathbb{R}$ et

$$\mathbb{P}_\theta[T_1 \leq \theta \leq T_2] = 1 - \alpha \quad (4.1)$$

Alors $IC_{1-\alpha}(\theta) = [T_1, T_2]$ est un intervalle de confiance de niveau $1 - \alpha$ de θ .

Définition 18 Un sous-ensemble aléatoire $C(X)$ de $\Theta \subset \mathbb{R}^p$ avec $p \geq 2$, vérifiant pour tout θ :

$$\mathbb{P}_\theta[\theta \in C(X)] = 1 - \alpha \quad (4.2)$$

est appelée région de confiance de θ au niveau $1 - \alpha$.

Les deux critères de qualité d'un intervalle de confiance (sa longueur et son niveau de confiance) s'opposent et il faut donc réaliser un compromis entre les deux. On cherche généralement un intervalle de longueur la plus petite possible, pour un niveau de risque α donné.

Étant donné la fonction de répartition F d'une v.a. X sur \mathbb{R} , on rappelle que le quantile d'ordre $\alpha \in]0, 1[$ de X est

$$q_\alpha = \inf \{x \in \mathbb{R} : F(x) \geq \alpha\} \quad (4.3)$$

Si F est continue, $F(q_\alpha) = \alpha$ et si F est strictement croissante, q_α est unique.

4.2 Méthodes pour déterminer une région de confiance

4.2.1 Fonction pivotale

Définition 19 Un pivot $Z = Z(X, \theta)$ est une fonction de la variable d'observation X et du paramètre θ telle que la loi de Z soit libre en θ , c'est à dire que sa loi ne dépend pas de θ .

On dit que θ est un paramètre de position pour T si la loi de $T - \theta$ ne dépend pas de θ , c'est à dire si $T - \theta$ est un pivot.

On dit que θ est un paramètre d'échelle pour T si la loi de $T \div \theta$ ne dépend pas de θ , c'est à dire si $T \div \theta$ est un pivot.

Le pivot se construit généralement à partir d'un estimateur de θ . Une fois déterminé, on cherche, grâce aux quantiles de la loi, un événement B (qui ne dépend pas non plus de θ) tel que $\mathbb{P}_\theta[Z \in B] = 1 - \alpha$. La région de confiance est alors de la forme $IC = \{\theta \in \Theta : Z \in B\}$.

Exemple :

Soit X_1, \dots, X_n un n -échantillon de loi $\mathcal{N}(\theta, 1)$ et soit $q = q_{1-\alpha/2}$ la quantile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite. Comme $\sqrt{n}(\bar{X} - \theta)$ est une v.a. pivotale de loi $\mathcal{N}(0, 1)$, alors

$$\mathbb{P}_\theta [\sqrt{n} |\bar{X} - \theta| \leq q] = \Phi(q) - \Phi(-q) = 1 - \alpha \quad (4.4)$$

où Φ est la fonction de répartition de la loi normale centrée réduite, donnée par

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (4.5)$$

et donc

$$\mathbb{P}_\theta \left(\theta \in \left[\bar{X} - \frac{q}{\sqrt{n}}; \bar{X} + \frac{q}{\sqrt{n}} \right] \leq q \right) = 1 - \alpha \quad (4.6)$$

L'intervalle de confiance est

$$IC_{1-\alpha}(\theta) = \left[\bar{X} - \frac{q_{1-\alpha/2}}{\sqrt{n}}; \bar{X} + \frac{q_{1-\alpha/2}}{\sqrt{n}} \right] \quad (4.7)$$

Exemple :

Soit (X_1, \dots, X_n) un n -échantillon de loi $\mathcal{N}(m, \sigma^2)$.

- Supposons $\theta = m$ inconnu et σ connu.

Un estimateur de θ est donnée par $\bar{X} \sim \mathcal{N}(\theta, \sigma^2/n)$. On peut alors choisir comme pivot

$$Z = \frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1) \quad (4.8)$$

Soit $\alpha \in]0, 1[$. Le choix des quantiles se fait en déterminant q_1, q_2 tels que $\mathbb{P}_\theta[q_1 \leq Z \leq q_2] = 1 - \alpha$.

La symétrie de la loi $\mathcal{N}(0, 1)$ permet de choisir un unique quantile q d'ordre $1 - \alpha/2$ contenant le mode de la densité de la gaussienne pour avoir $\mathbb{P}_\theta[|Z| \leq q] = 1 - \alpha$. L'intervalle de confiance est alors

$$IC_{1-\alpha}(\theta) = \left[\bar{X} - \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}; \bar{X} + \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2} \right] \quad (4.9)$$

- Supposons m et σ inconnu et posons $\theta = m$.

La v.a. $Z = (\bar{X} - \theta) \div (\sigma/\sqrt{n})$ définie précédemment suit une loi $\mathcal{N}(0, 1)$. Posons

$$Y = \frac{n-1}{\sigma^2} S^2 \quad (4.10)$$

où S^2 est la variance empirique non biaisée. Y suit une loi χ_{n-1}^2 et d'après le théorème de Cochran, $S^2 \perp \bar{X}$. Nous savons en outre que

$$T = \frac{Z}{\sqrt{Y/(n-1)}} = \frac{\bar{X} - m}{\sqrt{S^2/n}} \quad (4.11)$$

suit une loi de Student à $n - 1$ degrés de libertés, ne dépend que de m (mais sa loi n'en dépend pas, elle est libre de m) et peut donc être choisi comme pivot. Par symétrie de la loi de Student, le quantile t d'ordre $1 - \alpha/2$ de cette loi de Student est tel que $-t$ est le quantile d'ordre $\alpha/2$. L'intervalle de confiance est donc

$$IC_{1-\alpha}(m) = \left[\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2}; \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2} \right] \quad (4.12)$$

- Supposons m et σ inconnu et posons $\theta = (m, \sigma^2)$.

C'est une région de confiance inclut dans \mathbb{R}^2 que nous allons déterminer. Posons

$$Z = \left(\frac{\bar{X} - m}{\sigma/\sqrt{n}}, \frac{n-1}{\sigma^2} S^2 \right) \sim \mathcal{N}(0, 1) \otimes \chi_{n-1}^2 \quad (4.13)$$

Notons z le quantile d'ordre $1 - \beta/2$ d'une loi normale centrée réduite et c_1, c_2 les quantiles d'ordre $\beta'/2$ et $1 - \beta'/2$ d'une loi du chi deux à $n - 1$ degrés de libertés. Les deux composantes de Z étant indépendantes,

$$\mathbb{P}_\theta (Z \in [-z, z] \times [c_1, c_2]) = \mathbb{P}_\theta (Z_1 \in [-z, z]) \times \mathbb{P}_\theta (Z_2 \in [c_1, c_2]) = (1 - \beta)(1 - \beta') = 1 - \alpha \quad (4.14)$$

en posant $\alpha = (\beta + \beta') + \beta\beta'$. La région de confiance sera donc de la forme

$$RC_{1-\alpha}(\theta) = \left[\bar{X} - \frac{\sigma}{\sqrt{n}}z ; \bar{X} + \frac{\sigma}{\sqrt{n}}z \right] \times \left[\frac{n-1}{c_1}S^2 ; \frac{n-1}{c_2}S^2 \right] \quad (4.15)$$

4.2.2 Utilisation des inégalités de Tchebychev et de Hoeffding

Lorsque la v.a. est discrète ou que la loi n'est pas connue, on ne peut pas construire un intervalle de confiance comme vu précédemment, mais l'on peut introduire la notion plus faible d'intervalle de confiance par excès dans lequel

$$\mathbb{P}_\theta [\theta \in I] \geq 1 - \alpha \quad (4.16)$$

Théorème 9 (Inégalité de Tchebychev)

$$\mathbb{P} [|X - \mathbb{E}[X]| \geq \epsilon] \leq \frac{\mathbb{V}(X)}{\epsilon^2} \quad (4.17)$$

Théorème 10 (Inégalité de Hoeffding) Soient X_1, \dots, X_n n v.a.i.i.d. telles que $a \leq X_1 \leq b$ \mathbb{P} -p.s. et $S_n = \sum_{i=1}^n X_i$. Alors pour tout $t > 0$,

$$\mathbb{P} \left[\left| \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right| \geq t \right] \leq 2 \exp \left(-\frac{2t^2}{n(b-a)^2} \right) \quad (4.18)$$

soit encore

$$\mathbb{P} [|S_n - \mathbb{E}[S_n]| \geq t] \leq 2 \exp \left(-\frac{2t^2}{n(b-a)^2} \right) \quad (4.19)$$

Pour $(a, b) = (0, 1)$, on a :

$$\mathbb{P} (|\bar{X} - \mathbb{E}(X)| \geq t) \leq 2 \exp(-2nt^2).$$

Exemple :

Considérons un n -échantillon X_1, \dots, X_n dont la loi, à support dans un intervalle $[a, b]$ indépendant de θ , admet un moment d'ordre 1. Cherchons un intervalle de confiance pour la moyenne commune $\theta = m$ des X_i .

L'inégalité de Tchebychev donne comme intervalle de confiance (par excès) de niveau $1 - \alpha$

$$I_1 = \left[\bar{X} - \frac{b-a}{\sqrt{n\alpha}} ; \bar{X} + \frac{b-a}{\sqrt{n\alpha}} \right] \quad (4.20)$$

L'inégalité de Hoeffding permet d'améliorer cet intervalle car

$$\mathbb{P}_\theta [|\bar{X} - \theta| \geq t] \leq 2 \exp \left(-\frac{2nt^2}{(b-a)^2} \right) \quad (4.21)$$

En posant

$$t = (b-a) \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}} \quad (4.22)$$

on obtient un second intervalle (par excès également)

$$I_2 = \left[\bar{X} - (b-a) \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}} ; \bar{X} + (b-a) \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}} \right] \quad (4.23)$$

Les contributions de la taille de l'échantillon et de la longueur du support sont les mêmes, mais la contribution de α est nettement meilleure dans le second cas, lorsque α est proche de 0. En effet, quand α tend vers 0, le risque diminue et donc l'intervalle de confiance s'élargit. Mais il s'élargit beaucoup moins vite avec la seconde inégalité (à vitesse logarithmique) qu'avec la première (à la vitesse de $\sqrt{\alpha}$).

4.2.3 Inversion d'un test statistique

Les tests sont le sujet du paragraphe suivant. Nous y verrons le lien entre test et région de confiance et une méthode pour déterminer des intervalles de confiance à partir des régions de rejet des tests.

4.3 région de confiance asymptotique

4.3.1 Définition

Définition 20 Soit X_1, \dots, X_n un n -échantillon de loi \mathbb{P}_θ .

Un pivot asymptotique $Z_n = Z(X_1, \dots, X_n, \theta)$ est une suite de v.a. telle que Z_n converge en loi vers une v.a. qui ne dépend pas de θ .

4.3.2 Méthode du pivot asymptotique

Soit $\hat{\theta}_n$ un estimateur \sqrt{n} -consistant et asymptotiquement normal de θ (ou plus généralement de $g(\theta)$) :

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad (4.24)$$

Alors un pivot asymptotique est donné par

$$Z_n = \frac{\sqrt{n} (\hat{\theta}_n - \theta)}{\sigma} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \quad (4.25)$$

Si σ est connu, on calcule le quantile q d'ordre $1 - \alpha/2$ de la loi normale centrée réduite et l'on a alors, par définition de la convergence en loi,

$$\begin{cases} \lim_{n \rightarrow +\infty} \mathbb{P}_\theta[Z_n \leq q] = 1 - \alpha/2 \\ \lim_{n \rightarrow +\infty} \mathbb{P}_\theta[Z_n \leq -q] = \alpha/2 \end{cases} \quad (4.26)$$

L'intervalle de confiance asymptotique de niveau $1 - \alpha$ pour θ est alors donné par

$$IC_{1-\alpha}(\theta) = \left[\left| \frac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}} \right| \leq q \right] \quad (4.27)$$

4.3.3 Méthode de Wald

Lorsque σ n'est pas connu ou dépend de θ , la méthode de Wald consiste à substituer dans la méthode précédente $\sigma(\theta)$ par un estimateur consistant $\hat{\sigma}(\theta)$ de l'écart type et d'utiliser ensuite le lemme de Slutsky.

$$Z_n = \frac{\sqrt{n} (\hat{\theta}_n - \theta)}{\hat{\sigma}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \frac{1}{\sigma(\theta)} \mathcal{N}(0, \sigma(\theta)^2) = \mathcal{N}(0, 1) \quad (4.28)$$

Si q est le quantile d'ordre $1 - \alpha/2$ d'une loi $\mathcal{N}(0, 1)$, alors l'intervalle de confiance asymptotique est

$$IC_{1-\alpha}(\theta) = \left[\hat{\theta} - q \frac{\hat{\sigma}}{\sqrt{n}} ; \hat{\theta} + q \frac{\hat{\sigma}}{\sqrt{n}} \right] \quad (4.29)$$

Exemple :

On considère un n -échantillon de bernoulli de paramètre θ à estimer. Le paramètre d'intérêt est la moyenne de chaque v.a. et un estimateur est donné par la moyenne empirique. Le pivot asymptotique est donné par

$$\sqrt{\frac{n}{\bar{X}(1-\bar{X})}} (\bar{X} - \theta) \quad (4.30)$$

En effet, le théorème de la limite centrale montre que

$$\sqrt{n} \frac{\bar{X} - \theta}{\sqrt{\theta(1-\theta)}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1) \quad (4.31)$$

Et comme $\mathbb{E}[\bar{X}] = \theta$, en remplaçant θ par cet estimateur et en utilisant le lemme de Slutsky, le pivot asymptotique converge en loi vers une gaussienne centrée réduite. L'intervalle de confiance asymptotique est alors

$$IC_{1-\alpha}(\theta) = \left[\bar{X} - \frac{q}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})}; \bar{X} + \frac{q}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})} \right] \quad (4.32)$$

où q est la quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$.

4.3.4 Inversion d'un test statistique

Considérons un test asymptotique de région critique R relativement à un paramètre θ , pour un risque α fixé. Alors on peut considérer la région d'acceptation R comme l'intervalle de confiance au niveau $1 - \alpha$ pour θ , grâce à l'équivalence

$$X \in \bar{R}_\theta \iff \theta \in IC(X) \quad (4.33)$$

4.4 Choix des quantiles

Une loi de probabilité de fonction de répartition F est unimodale s'il existe x_M tel que F soit convexe avant x_M et concave après. x_M est alors le mode de F .

Proposition 1 *Soit f la densité du pivot Z . Si f est unimodale de mode M et vérifie, pour $a \leq b$,*

$$\bullet \mathbb{P}[a \leq Z \leq b] = \int_a^b f(x) dx = 1 - \alpha \quad (4.34)$$

$$\bullet f(a) = f(b) > 0 \quad (4.35)$$

$$\bullet a < x_M < b \quad (4.36)$$

Alors $[a, b]$ est l'intervalle de longueur minimale vérifiant la première condition

Ainsi, par exemple, $q_{1-\alpha/2}$ est le meilleur choix de quantile pour une loi symétrique. $q_{\alpha/2}$ et $q_{1-\alpha/2}$ ne sont pas optimaux pour une loi du chi deux, mais sont souvent utilisés par défaut.

Chapitre 5

Tests statistiques

5.1 Tests paramétriques

5.1.1 Hypothèse nulle et hypothèse alternative

On dispose d'un n -échantillon $X = (X_1, \dots, X_n)$ et l'on veut construire une fonction de décision ϕ (appelée également fonction de test) à valeurs binaire, de la forme

$$\phi(X) = \mathbb{1}[T(X) \in R] \quad (5.1)$$

$R \subset \mathbb{X}^n$ est la région de rejet du test et $T(X)$ est une statistique des observations X . On identifie souvent la fonction de décision et la zone de rejet.

Définition 21 *Un test pur est une statistique $\phi : \mathbb{X} \rightarrow \{0, 1\}$*

$R = \phi^{-1}(1) = \{x \in \mathbb{X} : \phi(x) = 1\}$ est la région de rejet du test.

$\bar{R} = \phi^{-1}(0) = \{x \in \mathbb{X} : \phi(x) = 0\}$ est la région d'acceptation du test.

On suppose que l'on se place dans le cas d'un modèle paramétrique où le paramètre $\theta \in \Theta$. Θ est partitionné en deux sous-ensembles Θ_0 et Θ_1 de telle sorte que $\Theta_0 \cap \Theta_1 = \emptyset$. $\theta_0 \in \Theta_0$, $\theta_1 \in \Theta_1$.

Définition 22 *Faire une hypothèse \mathcal{H}_0 sur θ , appelée hypothèse nulle, consiste à se donner un sous-ensemble Θ_0 de Θ . On note cette hypothèse :*

$$\mathcal{H}_0 : \theta \in \Theta_0 \quad (5.2)$$

On note \mathcal{H}_1 l'hypothèse alternative :

$$\mathcal{H}_1 : \theta \in \Theta_1 \quad (5.3)$$

où $\Theta_1 \subset \Theta$ et $\Theta_0 \cap \Theta_1 = \emptyset$ (souvent $\bar{\Theta}_0 = \Theta_1$).

Définition 23 *L'erreur de première espèce consiste à rejeter \mathcal{H}_0 à tort.*

L'erreur de seconde espèce consiste à ne pas rejeter \mathcal{H}_0 à tort.

Dans l'approche de Neyman-Pearson, les deux hypothèses ne sont pas symétriques : on part de l'idée que \mathcal{H}_0 est vraie *a priori* et l'on va essayer de se convaincre (ou pas) à l'aide des observations, que ce n'est pas le cas. \mathcal{H}_0 est l'hypothèse privilégiée, en ce sens qu'elle est supposée vérifiée tant que les observations ne conduisent pas à la rejeter. Les observations amènent alors à accepter \mathcal{H}_0 (on dit plutôt *ne pas rejeter*) ou bien à rejeter cette hypothèse au profit de \mathcal{H}_1 .

Il existe différentes formes d'hypothèses :

- Hypothèses simples : $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$ sont des singletons.

- Hypothèse simple contre hypothèse composite :

Test bilatère : $\Theta_0 = \{\theta_0\}$ et $\Theta_1 \neq \{\theta_0\}$ ou test unilatère : $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta \in \Theta : \theta > \theta_0\}$.

- Hypothèses composites : lorsque les deux ensembles Θ_0 et Θ_1 contiennent au moins deux éléments.

5.1.2 Risque, niveau, puissance, efficacité

Le risque de première espèce (fausse alerte) s'écrit

$$\alpha(\theta) = \mathbb{P}_\theta(R) = \mathbb{P}_\theta[\phi(X) = 1] = \mathbb{E}_\theta[\phi(X)] , \text{ pour } \theta \in \Theta_0 \quad (5.4)$$

Le risque de seconde espèce (non détection d'une erreur) s'écrit

$$\beta(\theta) = \mathbb{P}_\theta(\overline{R}) = \mathbb{P}_\theta[\phi(X) = 0] = \mathbb{E}_\theta[1 - \phi(X)] , \text{ pour } \theta \in \Theta_1 \quad (5.5)$$

α est donc une fonction défini sur Θ_0 et à valeurs dans $[0, 1]$, tandis que β est une fonction définie sur Θ_1 à valeurs dans $[0, 1]$.

On souhaiterait minimiser ces deux risques simultanément mais ce n'est pas possible et il faut faire un compromis (minimiser α revient à minimiser la taille de R , tandis que minimiser β revient à augmenter la taille de R). On peut privilégier le risque de fausse alerte et demander qu'il soit aussi petit que possible.

Définition 24 *Le niveau α^* d'un test est*

$$\alpha^* = \sup_{\theta \in \Theta_0} \alpha(\theta) \quad (5.6)$$

Ce supremum existe (il est défini sur un ensemble non vide et borné) mais n'est pas toujours atteint. On dira que le test est de niveau α si $\alpha = \alpha^*$ et qu'il est de seuil α si $\alpha < \alpha^*$.

$$\mathbb{P}_{\theta_0}(R) \leq \alpha^* \quad (5.7)$$

Pour tout test, on dispose donc d'une famille de régions critiques $(R_\alpha)_{\alpha \in]0,1[}$ indicées par le niveau α du test.

Supposons qu'il existe une statistique à valeurs réelles $T(X)$ (appelée statistique de test) telle que

$$\phi = \mathbb{1}_{[s, \infty[}(T(X)) = \mathbb{1}_{[T(x) \geq s]} \quad (5.8)$$

Définition 25 *La valeur critique (ou p -valeur) d'un tel test ϕ est, pour toute observation x donnée,*

$$p(x) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta[T(X) > T(x)] \quad (5.9)$$

$p(x)$ est le plus petit niveau à partir duquel on rejète \mathcal{H}_0 sur la base de l'observation x .

$$\phi(x) = 1 \iff T(x) > s \iff p(x) < \alpha \quad (5.10)$$

Dans le cas où $\Theta = \{\theta_0\}$, la p -valeur est donc la v.a. fonction de X égale à la plus petite valeur de α à partir de laquelle on rejète \mathcal{H}_0 au vue des observations. C'est à dire :

- $\forall \alpha < p(x)$, on ne rejète pas \mathcal{H}_0 .
- $\forall \alpha > p(x)$, on rejète \mathcal{H}_0 .

Interprétation de la p -valeur : on calcule la probabilité (sous \mathcal{H}_0) d'obtenir pour T une observation aussi extrême que $T(x)$. Cette probabilité est la p -valeur. Avec la p -valeur, on cherche à quantifier le fait d'être bon ou pas dans le fait de rejeter \mathcal{H}_0 . La p -valeur représente donc le point de basculement entre rejet et non-rejet, pour une observation donnée. Si elle est petite, on peut rejeter \mathcal{H}_0 . Nous allons revenir rapidement sur cette définition dans les exemples ci-après.

On peut démontrer (à titre d'exercice) que $p(X)$ suit une loi uniforme sur $[0, 1]$.

À partir du risque de seconde espèce β , on calcule la puissance γ du test :

$$\gamma(\theta) = 1 - \mathbb{P}_\theta[T = 0] = 1 - \beta(\theta) \quad (5.11)$$

de sorte que γ prolonge à Θ_1 la fonction α .

$$\alpha(\theta_0) = \mathbb{P}_{\theta_0}[T = 1] = \mathbb{E}_{\theta_0}[T] \quad (5.12)$$

$$\gamma(\theta_1) = \mathbb{P}_{\theta_1}[T = 1] = \mathbb{E}_{\theta_1}[T] \quad (5.13)$$

Et l'on peut résumer les liens entre α , β et γ dans le tableau ci-dessous :

Décision \ Réalité	\mathcal{H}_0	\mathcal{H}_1
\mathcal{H}_0	$1 - \alpha$	β
\mathcal{H}_1	α	$\gamma = 1 - \beta$

Définition 26 La courbe d'efficacité d'un test T est la fonction

$$h : \Theta_0 \cup \Theta_1 \longrightarrow [0, 1] : \theta \mapsto h(\theta) = 1 - \gamma(\theta) \quad (5.14)$$

Nous donnons maintenant plusieurs exemples pour illustrer toutes ces définitions.

Exemple :

Le ministère de la santé étudie régulièrement la nécessité de prendre des mesures contre la consommation d'alcool et l'efficacité de ces mesures. L'Insee fournit à cet effet des données annuelles de consommation moyenne d'alcool par personne et par jour. En 1991, la loi Évin interdit la publicité sur les boissons alcoolisées et lance une campagne de sensibilisation sous forme de spots publicitaires. Avant la loi Évin, la consommation d'alcool moyenne chez les personnes de plus de 15 ans était de 35 g par jour. L'objectif premier de la loi était de baisser cette consommation journalière moyenne à 33 g. En se basant sur l'observation des consommations moyennes de 1991 à 1994, le ministère a fixé la règle de décision suivante : si la moyenne des consommations journalières sur ces quatre années est supérieure à 34.2 g, alors les mesures prises ont été inefficaces.

On suppose que les données recueillies sont des réalisations de v.a. supposées i.i.d. et de loi normale $\mathcal{N}(\theta, \sigma^2)$ avec $\sigma = 2$. On note X_i , $i = 1 \dots 4$, les moyennes de consommation journalières pour chacune des quatre années de l'enquête et l'on note \bar{X} la moyenne empirique des X_i .

Le test est donc le suivant :

$$\begin{cases} \mathcal{H}_0 : \theta = 33 \\ \mathcal{H}_1 : \theta = 35 \end{cases} \quad (5.15)$$

L'hypothèse nulle est $\theta = 33$ car on souhaite que la politique ait été efficace. On suppose donc l'effet significatif *a priori*.

Il semble naturel de choisir \bar{X} comme statistique pour mesurer l'effet moyen de la consommation journalière (pourquoi?). Nous commençons par calculer l'erreur de première espèce α du test, en notant $s = 34.2$ le seuil de significativité choisi (arbitrairement) :

$$\alpha = \mathbb{P}_{33}[\bar{X} > s] = \mathbb{P}_{33}[\bar{X} - 33 > s - 33] = 1 - \Pi(34.2 - 33) \approx 0,1151 \quad (5.16)$$

Où Π est la fonction de répartition d'une loi normale centrée réduite :

$$\Pi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (5.17)$$

\bar{X} suit une loi normale de moyenne θ inconnue et de variance 1, car la somme de v.a. normales indépendantes est également normale et nous avons choisi l'écart type afin que la variance de \bar{X} vaille 1.

Le risque de seconde espèce est

$$\beta = \mathbb{P}_{35}[\bar{X} \leq s] = \mathbb{P}_{35}[\bar{X} - 35 \leq s - 35] = \Pi(s - 35) \approx 0,212 \quad (5.18)$$

Les deux hypothèses jouent des rôles dissymétriques car les conséquences liées aux deux types d'erreurs ne sont pas les mêmes. Lorsque la valeur du seuil s augmente, il est clair que α décroît vers 0 tandis que β croît vers 1. On ne peut donc pas minimiser les deux simultanément (on remarque par ailleurs que $\alpha + \beta \neq 1$).

La valeur de α par défaut est souvent 5%. Le seuil correspondant, que l'on notera $s^* = s(\alpha)$ (ou parfois $c^* = c(\alpha)$) vérifie :

$$\mathbb{P}_{33}[\bar{X} > s^*] = 0,05 = \alpha \iff 1 - \mathbb{P}_{33}[\bar{X} - 33 < s^* - 33] = 0,05 \quad (5.19)$$

$$\iff \mathbb{P}[Z < s^* - 33] = 0,95 \quad (5.20)$$

$$\iff \Pi(s^* - 33) = \Pi(1,65) \quad (5.21)$$

$$\iff s^* = 33 + 1,65 = 34,65 \quad (5.22)$$

Ainsi, si l'on pose $s = 34.65$, le test défini par $\phi(x) = \mathbb{1}_{[\bar{x} > 34.65]}$ est donc de niveau $\alpha = 5\%$.

Si l'on a observé $x = (34.7, 34.4, 33.7, 33.3)$, alors la moyenne est $\bar{x} = 34.025$ et la p -valeur du test (pour cette observation) est

$$p(x) = \mathbb{P}_{33}[\bar{X} \leq 34.025] \approx 0.153 \quad (5.23)$$

0.153 est le niveau minimum auquel on va rejeter \mathcal{H}_0 pour ces observations. 0.05 est plus petit que cette p -valeur et effectivement, on accepte \mathcal{H}_0 à ce niveau de 5% car la moyenne observée de 34.025 est bien inférieure à 34.645.

Exemple :

On considère un n -échantillon gaussien de loi $\mathcal{N}(\theta, 1)$ dont la moyenne θ est inconnue. On définit le test :

$$\begin{cases} \mathcal{H}_0 : \theta \leq \theta_0 \\ \mathcal{H}_1 : \theta > \theta_0 \end{cases} \quad (5.24)$$

Où $\theta \in \mathbb{R}$ est fixé. Une idée naturelle est de considérer la moyenne empirique $\bar{X} \sim \mathcal{N}(\theta, 1/n)$. En ce cas, et sous l'hypothèse \mathcal{H}_0 , la v.a. $\sqrt{n}(\bar{X} - \theta)$ suit une loi $\mathcal{N}(0, 1)$.

Une région de rejet serait de la forme $R(q) = [x : \sqrt{n}(\bar{X} - \theta_0) > q]$. Le quantile q doit être déterminé pour que le test soit de niveau α . On considère donc, sous l'hypothèse \mathcal{H}_0 , un $\theta \leq \theta_0$:

$$\mathbb{P}_\theta[R(q)] = \mathbb{P}_\theta[\sqrt{n}(\bar{X} - \theta_0) > q] = \mathbb{P}_\theta[\sqrt{n}(\bar{X} - \theta) + \sqrt{n}(\theta - \theta_0) > q] \quad (5.25)$$

$$\leq \mathbb{P}_\theta[\sqrt{n}(\bar{X} - \theta) > q] \quad (5.26)$$

$$\leq \mathbb{P}[Z > q] \quad (5.27)$$

L'égalité n'est obtenue que si $\theta = \theta_0$. Ainsi, $\alpha^* = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta[R(q)] = \mathbb{P}[Z > q] = 1 - \Pi(q)$

où Π est la fonction de répartition de la loi normale centrée réduite et $Z \sim \mathcal{N}(0, 1)$. q doit donc être égal au quantile $q_{1-\alpha}$ d'une loi normale centrée réduite pour que le test soit de niveau α .

Exemple :

On effectue n lancers à pile ou face indépendants avec une même pièce.

On souhaite savoir si la pièce est équilibrée. On dispose donc d'un n -échantillon de Bernoulli de paramètre θ et le test est défini par

$$\begin{cases} \mathcal{H}_0 : \theta = 1/2 \\ \mathcal{H}_1 : \theta \neq 1/2 \end{cases} \quad (5.28)$$

Nous avons $\theta = \mathbb{E}_\theta[X_1]$ et $n\bar{X}$ suit une loi binomiale $B(n, \theta)$. Sous l'hypothèse \mathcal{H}_0 , cette v.a. suit donc une loi $B(n, 1/2)$. On propose comme région de rejet

$$R(a, b) = [b \leq \bar{x} \text{ ou } \bar{x} \leq a] \quad (5.29)$$

avec $a < 1/2 < b$. Il faut également calculer a et b de telle sorte que $\mathbb{P}_{[\theta=1/2]}[R(a, b)] = \alpha$. On a

$$\mathbb{P}_{1/2}([\bar{X} < a] \cup [\bar{X} > b]) = \mathbb{P}_{1/2}([n\bar{X} < na] \cup [n\bar{X} > nb]) = \mathbb{P}_{1/2}[\bar{X} < a] + \mathbb{P}_{1/2}[\bar{X} > b] = \alpha \quad (5.30)$$

chacune des deux probabilités ne peut prendre qu'un nombre fini de valeurs car $n\bar{X}$ est à valeurs dans $\{0, \dots, n\}$. On ne peut pas choisir α quelconque : il n'existe pas toujours de test de niveau α . Par contre, on peut toujours définir un seuil. Dans l'exemple, on peut montrer (à titre d'exercice) que si $n = 1000$, $a = 0,469$ et $b = 0,531$ donne un test de seuil $\alpha = 5\%$. C'est d'ailleurs le test le plus puissant parmi tous les tests dont le milieu de $[a, b]$ est $1/2$.

Construction pratique des tests

L'idée de Neyman et Pearson (en 1933) est d'avoir une approche en deux temps : on commence par contrôler le risque de première espèce α , puis une fois ce risque fixé, on cherche parmi les tests restants ceux qui minimisent β . Pour construire un test, on procède suivant les étapes suivantes :

1. Détermination du modèle statistique représentant le phénomène aléatoire étudié.
2. Détermination des hypothèses \mathcal{H}_0 et \mathcal{H}_1 .
3. Détermination d'une statistique de test T .
4. Détermination de la forme de la fonction de test ϕ et de la forme de la région critique.
5. Détermination des constantes et quantiles intervenant dans ϕ de telle façon que le test soit de niveau ou de seuil donné. Cela revient à déterminer la frontière de la région critique.
6. Conclusion au vu de l'observations x .
7. Évaluation de la puissance du test.

5.1.3 Tests asymptotiques

Test de Wald unidimensionnel

On dispose d'un estimateur T_n de θ déterminé à partir du n -échantillon.

Définition 27 *Un test de région de rejet R est de niveau asymptotique α si*

$$\sup_{\theta \in \Theta_0} \lim_{n \rightarrow +\infty} \mathbb{P}_\theta(R) = \alpha \quad (5.31)$$

Théorème 11 (Test de Wald en dimension 1) *On suppose que $\Theta \subset \mathbb{R}$ et que les deux hypothèses suivantes sont vérifiées :*

$$\sqrt{n}(T_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2(\theta)) \quad (5.32)$$

$$\sigma^2 : \Theta \rightarrow]0, \infty[\text{ est continue} \quad (5.33)$$

Alors le test asymptotique $\mathcal{H}_0 : \theta = \theta_0$ vs $\mathcal{H}_1 : \theta \neq \theta_0$ de région de rejet

$$R = \left[\left| \sqrt{n} \frac{T_n - \theta_0}{\sigma(T_n)} \right| > q \right] \quad (5.34)$$

où q est le quantile d'ordre $1 - \alpha/2$ d'une loi $\mathcal{N}(0, 1)$, est un test de niveau α qui est consistant.

Démonstration :

Si $Z_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Z$ avec $Z \mathcal{N}(0, 1)$, alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(Z_n \in [a, b]) = \mathbb{P}(Z \in [a, b]) \quad (5.35)$$

car la frontière de l'intervalle $[a, b]$ est de mesure nulle. Ainsi,

$$\lim_{n \rightarrow +\infty} \mathbb{P}_{\theta_0}(R) = \lim_{n \rightarrow +\infty} \mathbb{P}_{\theta_0} \left[\left| \sqrt{n} \frac{T_n - \theta_0}{\sigma(T_n)} \right| > q \right] \quad (5.36)$$

$$= \mathbb{P}[Z > q] = \alpha \quad (5.37)$$

Pour démontrer la constistance du test, nous devons montrer que $\forall \theta \in \Theta_1$, $\lim_n \mathbb{P}_{\theta}(R) = 1$. Ceci revient à démontrer que la limite de la probabilité du complémentaire tend vers 0. Nous avons

$$\bar{R} = \left[\left| \sqrt{n} \frac{T_n - \theta_0}{\sigma(T_n)} \right| \leq q \right] \quad (5.38)$$

En utilisant l'inégalité $|x - y| \geq |x| - |y|$, il vient :

$$\mathbb{P}_{\theta}(\bar{R}) = \mathbb{P}_{\theta} \left[\left| \sqrt{n} \frac{T_n - \theta}{\sigma(T_n)} + \sqrt{n} \frac{\theta - \theta_0}{\sigma(T_n)} \right| \leq q \right] \quad (5.39)$$

$$\leq \mathbb{P}_{\theta} \left[\sqrt{n} \frac{|\theta - \theta_0|}{\sigma(T_n)} - \sqrt{n} \frac{|T_n - \theta|}{\sigma(T_n)} \leq q \right] \quad (5.40)$$

$$\leq \mathbb{P}_{\theta} \left[|\theta - \theta_0| \leq \frac{\sigma(T_n)}{\sqrt{n}} \times \frac{\sqrt{n}}{\sigma(T_n)} |T_n - \theta| + \frac{\sigma(T_n)}{\sqrt{n}} q \right] \quad (5.41)$$

Le terme $\sigma(T_n)/\sqrt{n}$ tend vers 0 en probabilité quand n tend vers l'infini, tandis que $\sqrt{n}|T_n - \theta|/\sigma(T_n)$ converge en loi vers une gaussienne centrée réduite. L'expression à droite de l'inégalité tend donc vers 0 en loi et en probabilité et par suite, la probabilité tend vers 0 avec n .

□

Exemple :

On considère un modèle d'échantillonnage avec comme statistique $T_n = \hat{\theta}_n^{\text{MV}} = \hat{\theta}_n$ l'estimateur du maximum de vraisemblance de θ . On suppose que $\Theta \subset \mathbb{R}$ et que le modèle est dominé de vraisemblance $f(x, \theta)$.

Sous les conditions de régularités suffisantes, nous savons que

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \mathbb{I}(\theta)^{-1}) \quad (5.42)$$

L'application $\theta \rightarrow \mathbb{I}(\theta)$ est continue et définie positive (ici, c'est donc un scalaire > 0). Le test de Wald avec $\mathcal{H}_0 : \theta = \theta_0$ vs $\mathcal{H}_1 : \theta \neq \theta_0$ a pour région de rejet

$$R = \left[\left| \sqrt{n} \frac{\hat{\theta}_n - \theta_0}{\sigma(\hat{\theta}_n)} \right| > q \right] \quad (5.43)$$

Test de Wald multidimensionnel

On considère le test

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 = \{\theta \in \mathbb{R}^d : g(\theta) = 0\} \\ \mathcal{H}_1 : \theta \notin \Theta_0 \end{cases} \quad (5.44)$$

où $g : \mathbb{R}^d \rightarrow \mathbb{R}^n$ où $n \leq d$.

Théorème 12 (Test de Wald multidimensionnel) *On suppose que les hypothèses suivantes sont vérifiées :*

- $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \mathbb{I}(\theta)^{-1})$
- $\mathbb{I}(\theta)^{-1}$ est une fonction continue de θ
- g est de classe C^1 sur \mathbb{R}^d et sa matrice jacobienne $J_g(\theta)$ est continue en θ
- $\text{rg}(J_g(\theta)) = r$

Alors le test $\mathcal{H}_0 : g(\theta) = 0$ vs $\mathcal{H}_1 : g(\theta) \neq 0$ est de niveau asymptotique α et sa région de rejet est

$$R = \left[n g(\hat{\theta})' B(\hat{\theta})^{-1} g(\hat{\theta}) > q \right] \quad (5.45)$$

Annexe : Lois usuelles

Nom	paramètres	$\mathbb{P}[X = k]$	$\mathbb{E}[X]$	$\mathbb{V}(X)$	À valeurs dans
Bernoulli	$p \in [0, 1]$	$\mathbb{P}[X = 1] = p$ $\mathbb{P}[X = 0] = 1 - p = q$	p	pq	$\{0, 1\}$
Équiprobable		$\mathbb{P}[X = k] = 1/n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\llbracket 1 \dots n \rrbracket$
Binomiale	$n \in \mathbb{N}^*$ et $p \in [0, 1]$	$\mathbb{P}[X = k] = C_n^k p^k q^{n-k}$	np	npq	$\llbracket 0 \dots n \rrbracket$
Géométrique	$p \in [0, 1]$	$\mathbb{P}[X = k] = q^{k-1} p$	$\frac{1}{p}$	$\frac{q}{p^2}$	\mathbb{N}^*
Poisson	$\lambda > 0$	$\mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ	\mathbb{N}
Binomiale négative	$n \in \mathbb{N}^*$ et $p \in [0, 1]$	$\mathbb{P}[X = k] = C_{k-1}^{n-1} p^n q^{k-n}$	$\frac{n}{p}$	$\frac{np}{p^2}$	$k \geq n$

- Une v.a. de loi binomiale $\mathcal{B}(n, p)$ est la somme de n v.a. de Bernoulli indépendantes de paramètre p .
- Une v.a. de loi binomiale négative $\mathcal{B}(n, p)$ est la somme de n v.a. géométriques indépendantes de paramètre p . Attention à cette loi qui connaît plusieurs définitions différentes.
- Les lois binomiales, de Poisson et binomiales négatives sont stables par additions indépendantes.

Nom	paramètres	densité $f(x)$	$\mathbb{E}[X]$	$\mathbb{V}(X)$	À valeurs dans
Loi uniforme	α, β	$\frac{1}{\beta - \alpha} \times \mathbb{1}_{[\alpha, \beta]}(x)$	$\frac{\alpha + \beta}{2}$	$\frac{(\beta - \alpha)^2}{12}$	$[\alpha, \beta]$
Loi exponentielle VF	λ	$\lambda e^{-\lambda x} \mathbb{1}_{[0, +\infty[}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$[0, +\infty[$
Loi exponentielle VO	θ	$\frac{1}{\theta} e^{-x/\theta} \mathbb{1}_{[0, +\infty[}(x)$	θ	θ^2	$[0, +\infty[$
Loi de Laplace		$\frac{1}{2} e^{- x }$	0	2	\mathbb{R}
Loi de Cauchy		$\frac{1}{\pi} \frac{1}{1+x^2}$	∞	∞	\mathbb{R}
Loi triangulaire		$(1 - x) \mathbb{1}_{[-1, 1]}(x)$	0	1/6	$[-1, 1]$
Loi normale	$m \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2\right)$	m	σ^2	\mathbb{R}
Loi gamma	$a > 0, \lambda > 0$	$\frac{x^{a-1} \lambda^a}{\Gamma(a)} e^{-\lambda x} \mathbb{1}_{[0, \infty[}(x)$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$	$[0, \infty[$
Loi bêta	$a > 0, b > 0$	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{[0, 1]}(x)$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	$[0, 1]$

Tout le matériel de ce polycopié a été fourni par Claude Petit, en charge de ce cours pour les années 2016 et 2017. Qu'il en soit chaleureusement remercié.

Chapitre 6

Travaux Dirigés : Énoncés

6.1 TD : Probabilités

1.1 Lois discrètes usuelles

1. On dit que X suit une loi de Bernoulli de paramètre p , noté $X \sim \mathcal{B}(p)$, si $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = 0) = 1 - p$. Calculer $\mathbb{E}[X]$ et $\text{Var}[X]$.

2. Soient X_1, \dots, X_n des variables indépendantes identiquement distribuées de loi $\mathcal{B}(p)$, et soit $S = X_1 + \dots + X_n$. Quelle est la loi de S (le montrer par récurrence)? On dit que S suit une loi binomiale de paramètres n et p , noté $S \sim \mathcal{B}(n, p)$. Calculer $\mathbb{E}[S]$ et $\text{Var}[S]$.

3. Soit N une variable aléatoire à valeurs dans \mathbb{N} . On dit que N suit une loi de Poisson de paramètre λ , noté $N \sim \mathcal{P}(\lambda)$, s'il existe $C > 0$ telle que $\forall n \in \mathbb{N}, \mathbb{P}(N = n) = C \frac{\lambda^n}{n!}$. Que vaut C ? Calculer $\mathbb{E}[N]$ et $\text{Var}[N]$. Calculer $\mathbb{E}\left(\frac{1}{N+1}\right)$.

1.2 Lois à densité usuelles

1. On dit que X suit une loi normale de paramètres μ et σ , noté $X \sim \mathcal{N}(\mu, \sigma)$, si elle admet pour densité $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Calculer $\mathbb{E}[X]$ et $\text{Var}[X]$.

2. On dit que X suit une loi exponentielle de paramètres λ , noté $X \sim \mathcal{E}(\lambda)$, si elle admet pour densité $f(x) = C e^{-\lambda x} \mathbb{1}_{x>0}$ pour une certaine constante C . Calculer C . Quelle est sa fonction de répartition? Calculer $\mathbb{E}[X]$ et $\text{Var}[X]$.

3. Soient X, Y des variables aléatoires indépendantes telles que $X \sim \mathcal{E}(\lambda)$ et $Y \sim \mathcal{E}(\lambda)$. Quelle est la loi de $X + Y$?

4. On dit que X suit une loi de Cauchy si elle admet la densité $f(x) = \frac{1}{\pi(1+x^2)}$. Une variable de Cauchy est-elle intégrable (i.e. a-t-elle une espérance)? Quelle est la loi de $1/X$?

5. Soit U une variable aléatoire uniforme sur $[0, 1]$ (de densité $f(x) = \mathbb{1}_{0<x<1}$). Pour quelles valeurs de α la variable U^α est-elle intégrable? Que vaut alors son espérance? Déterminer la loi de $-\frac{1}{\lambda} \ln(U)$.

1.3 Autour de l'espérance

1. Soit X une variable aléatoire de carré intégrable à valeurs dans \mathbb{R} . Montrer que $\mathbb{E}[(a - X)^2]$ atteint son minimum pour $a = \mathbb{E}[X]$.

2. Soit X une variable aléatoire intégrable à densité à valeurs dans \mathbb{R}^+ . Montrer que $\mathbb{E}[X] = \int \mathbb{P}(X > t) dt$.

3. Soit A un événement. Montrer que $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]$. En déduire que pour tout $p \geq 1$, toute variable aléatoire positive X de moments X^p intégrable et tout $a > 0$, $\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X^p]}{a^p}$. Montrer en particulier que pour toute variable X de carré intégrable, $\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}$.

6.2 TD : Planche 2 : Estimation, construction d'estimateurs

2.1 Loi uniforme : différents estimateurs.

Soit (X_1, \dots, X_n) un n-échantillon de loi uniforme sur $[0, \theta]$.

1. Déterminez un estimateur $\hat{\theta}_1$ de θ par la méthode des moments. Étudiez son biais, son erreur quadratique, sa convergence.

2. Déterminez un estimateur $\hat{\theta}_2$ de θ par la méthode du maximum de vraisemblance. Étudiez son biais, son erreur quadratique, sa convergence. En déduire un estimateur $\hat{\theta}_3$ de θ , sans biais.

4. Quel est le meilleur des estimateurs précédents?

2.2 Deux estimateurs du paramètre d'une loi de Poisson

Soit X une va suivant une loi de Poisson de paramètre θ . Soit (X_1, \dots, X_n) un n-échantillon de X .

1. Déterminer deux estimateurs de θ à partir de la moyenne et de la variance de l'échantillon.

2. Calculer l'estimateur du maximum de vraisemblance.

2. Comparer ces estimateurs.

2.3 Estimation du paramètre d'une loi de Pareto

La loi de Pareto est utilisée pour modéliser les salaires par exemple. Le paramètre est de dimension 2 et on le note : $\theta = (x_m, \lambda) \in \mathbb{R}^{2+}$. C'est une loi continue dont le support est $[x_m, \infty[$ et la distribution est caractérisée par :

$$\forall x > x_m, \mathbb{P}(X > x) = \left(\frac{x_m}{x}\right)^\lambda.$$

1. Déterminer la vraisemblance du modèle.

2. Donner l'estimateur du maximum de vraisemblance.

3. Donner l'estimateur des moments.

2.4 Loi binomiale négative : estimation sans biais.

Soit $(X_n)_n$ une suite de v.a.i.i.d. de loi de Bernoulli de paramètre p et soit T_n le nombre d'essais nécessaires pour obtenir n succès.

1. Déterminer la loi de T_n , son espérance, sa variance.

2. Montrer que $\hat{p} = (n-1)/(T_n-1)$ est un estimateur sans biais de p .

3. Montrer que $\mathbb{E}(n/t_n) > p$.

6.3 TD : Planche 3 : cadre asymptotique

3.1 Loi Uniforme

On reprend le cadre de l'exercice 2.1 et les deux estimateurs.

1. Trouver la loi asymptotique de l'estimateur par la méthode des moments. Est-il asymptotiquement biaisé.

2. Le modèle est-il régulier ?

3. Calculer la loi de $n(\hat{\theta} - \theta_2)$.

4. Quel est l'estimateur à privilégier dans un cadre asymptotique ?

3.2 Loi de Poisson

On reprend le cadre de l'exercice 2.2 et on considère l'estimateur du maximum de vraisemblance.

1. Calculer l'information de Fisher et trouver la loi limite de l'estimateur.

2. Calculer la loi limite de l'estimateur des moments utilisant la variance.

3.3 Loi Gamma

La famille des lois Gamma, notée $\Gamma(\alpha, \beta)$, englobe la loi exponentielle et sert à modéliser un phénomène à support sur \mathbb{R}^+ . La densité est donnée, pour $\alpha, \beta > 0$, par :

$$f_{\alpha, \beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) \mathbb{1}_{\mathbb{R}^+}(x),$$

où $\Gamma(x) = \int t^{x-1} \exp(-t) dt$ est la fonction dite *Gamma d'Euler* et a la propriété : $\Gamma(x+1) = x\Gamma(x)$.

1. Donner l'estimateur par la méthode des moments (une utilisation intelligente de la constante de normalisation devrait aider).

2. Donner les équations vérifiées par l'estimateur du maximum de vraisemblance.

3. Donner un algorithme d'optimisation permettant de calculer une valeur approchée de l'estimateur.

4. Donner la loi limite de l'estimateur de la méthode des moments.

5. L'information de Fisher n'a pas d'expression explicite. Trouver une valeur approchée et donner la loi limite.

6.4 TD : Planche 4

4.1 Intervalle de confiance non asymptotique pour une loi exponentielle

Soit X_1, \dots, X_n un échantillon iid de loi $\mathcal{E}(\lambda_0)$. La loi $\mathcal{E}(\lambda)$ est continue et a pour densité $f_\lambda(x) = \lambda \exp(-\lambda x) \mathbb{1}_{\mathbb{R}^+}(x)$ pour $\lambda > 0$. La loi $\mathcal{E}(\lambda)$ est aussi la loi $\Gamma(1, \lambda)$.

Une somme de variables aléatoires **indépendantes** de loi $\mathcal{E}(\lambda)$ suit une loi Gamma $\Gamma(n, \lambda)$.

1. Montrer que $\lambda_0 \sum_{i=1}^n X_i$ suit une loi $\Gamma(n, 1)$. Comment appelle-t-on la variable aléatoire $Z(\lambda_0, X_1, \dots, X_n) = \lambda_0 \sum_{i=1}^n X_i$?
2. Construire un intervalle de confiance pour λ_0 non asymptotique de niveau 95%.
3. Quelle est la loi de $\lambda_0 \min X_i$?
4. Construire un autre intervalle de confiance pour λ_0 non asymptotique de niveau 95%.
5. Calculer l'estimateur du maximum de vraisemblance, l'information de Fisher du modèle. Justifier la normalité asymptotique de l'EMV et donner la loi limite.
6. Donner un intervalle de confiance asymptotique à 95%.
7. Comparer sur un exemple numérique les intervalles en faisant évoluer n .

4.2 Intervalle de confiance asymptotique pour la loi de Bernoulli

On cherche à estimer une proportion p à partir d'un échantillon X_1, \dots, X_n de v.a. de loi de Bernoulli $\text{Be}(p)$.

1. En utilisant l'inégalité de Bienaymé-Tchebychev, donner un intervalle de confiance à 95% (il faudra majorer $p(1-p)$).
2. Utiliser l'inégalité de Hoeffding pour donner un intervalle de confiance à 95%.
3. En utilisant le TCL, donner un intervalle de confiance asymptotique de même niveau.
4. Comparer les trois intervalles pour $\bar{X} = .4$ en faisant varier n de 10 à 10^6 .