Table des matières

Première partie
Statistiques

1 Statistique inférentielle

Principe de base :
On dispose de données $x_1, ..., x_n$ issue d'une même variable aléatoire. La valeur de x_i n'est pas influencée par la valeur x_j, $i \neq j$.

Exemple :

x_i : résultat pile ou face ("0" ou "1") d'un jet de pièce dont on ignore si elle est truquée ou non.
Si X_i est la variable aléatoire égale au résultat du ième jet, alors :
- $X_i \perp\!\!\!\perp X_j$ dès que $i \neq j$
- $\mathbb{P}(X_i = 0) = p$ et $\mathbb{P}(X_i = 1) = 1 - p$

En proba, p est supposé connu. En stat, p est inconnu et on souhait "inférer" sur cette valeur sur la base de l'observation $(x_1, ..., x_n)$.

Définition:

Soit $\mathfrak{X}$ un espace dit des observations (muni d'une certaine tribu $\mathcal{A}$ et soit P une loi de probabilité sur $\mathfrak{X}$ régissant le résultat de l'expérience aléatoire.

Soit $(x_1, ..., x_n) \in \mathfrak{X}^n$, on dit que $(x_1, ..., x_n)$ est un échantillon (observé) de la loi P si $(x_1, ..., x_n)$ est régi par la probabilité produit $P^{\otimes n}$ sur $(\mathfrak{X}^n, \mathcal{A}^n)$.

Supposons disposer d'une mesure μ sur $(\mathfrak{X}, \mathcal{A})$ (avec μ σ-finie).
Supposons également que $P << \mu$ (ie $\forall A \in \mathcal{A}$, $\mu(A) = 0 \Rightarrow \P(A) = 0$).
D'après le théorème de Radon-Nykodin, il existe une densité de la loi P par rapport à la mesure μ (ie, il existe une fonction mesurable positive f définie sur $\mathfrak{X}$ tel que

$$P(A) = \int_A f d\mu, \ \forall A \in \mathcal{A}$$

Dans ce cas, la loi $P^{\otimes n}$ admet une densité de probabilité par rapport à $\mu^{\otimes n}$, soit

$$(x_1, ..., x_n) \mapsto \prod_{i=1}^{,} f(x_i)$$

On dit aussi que cette densité de probabilité régit $(x_1, ..., x_n)$ ou encore que $(X_1, ..., X_n)$ admet cette densité comme densité jointe.

Exemple :

Soit $(X_1, ..., X_n)$ un échantillon de variable aléatoire de loi exponentielle de paramètre $\theta > 0$ (ie les $(X_i)_{1 \leq i \leq n}$ sont iid et $X_i \hookrightarrow \mathcal{E}(\theta)$, $\forall i \in \{1, .., n\}$)
Ici, $\mathfrak{X} = \mathbb{R}_+^*$, $\mathcal{A} = \mathcal{B}(\mathbb{R}_+^*)$ et $P_\theta = \mathcal{E}(\theta)$ où $\theta \in \Theta$ où $\Theta = \mathbb{R}_+^*$
Pour tout $\theta \in \Theta$, $\frac{dP_\theta}{d\lambda}(x) = \theta e^{-\theta x} 1_{\mathbb{R}_+^*}(x), \ \forall x \in \mathbb{R}$

$(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ est appelé modèle statistique.
On appelle modèle statistique d'échantillonnage :

$$(\mathfrak{X}^n, (P_\theta^{\otimes n})_{\theta\in\Theta}) = ((\mathbb{R}_+^*)^n, (P_\theta^{\otimes n})_{\theta\in\mathbb{R}_+^*})$$

où

$$L(\theta, x_1, ..., x_n) = \prod_{i=1}^n \theta e^{-\theta x_i} 1_{\mathbb{R}_+^*}(x_i) = \theta^n e^{-\theta\sum_{i=1}^n x_i} \prod_{i=1}^n 1_{\mathbb{R}_+^*}(x_i)$$

est une densité de la loi ${}_\theta^{\otimes n}$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$.

Définition:

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique. On appelle statistique toute application mesurable définie sur l'espace des observations $\mathfrak{X}$ et ne dépendant pas du paramètre inconnu θ.

Définition: Moyenne et Variance empirique

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

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

Propriété:

Soit $(X_1, ..., X_n)$ un échantillon de var d'une loi admettant une moyenne μ et une variance σ^2. Alors $E(\bar{X}) = \mu,\ V(\bar{X}) = \frac{\sigma^2}{n}$ et $E(S^2) = \sigma^2$

Démonstration :

$$E(\bar{X}) = \frac{1}{n}\sum_{i=1}^n E(X_i) = E(X_i) = \mu$$

$$V(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^n V(X_i) = \frac{\sigma^2}{n}$$

Reste le dernier, chiant.

Définition:

Soit X une var de fonction de répartition F_X. On appelle fonction génératrice des moments la fonction (si elle existe) notée M_X et finie dans un voisinage de 0 de $\mathbb{R}$ par $M_X(t) = E(e^{tx})$.

Remarque : – Si f_X est une densité de la loi de X alors $M_X(t) = \int_{\mathbb{R}} e^{tx} f_X(x)dx$
– Si X est discrtèe, $M_X(t) = \sum_{k \in X(\Omega)} e^{tX} \mathbb{P}(X = k)$

Propriété:

Soit X et Y deux variables aléatoires réelles.

1. Si V est un voisinage de 0 tel que $M_X(t) = M_Y(t)$ pour tout $t \in V$ alors $F_X = F_Y$
2. $\forall k \in \mathbb{N}$, $E(X^k) = M_X^{(k)}(0)$

Définition:

Soit $(\mathfrak{X}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique dominé par une mesure μ σ-finie, ie

$$\forall \theta \in \Theta, P_\theta << \mu$$

On dit que la loi P_θ appartient à la "famille exponentielle" si une densité de probabilité de la loi P_θ par rapport à la dominante μ est de la forme :

$$\forall x \in \mathfrak{X}, \forall \theta \in \Theta, f(x, \theta) = h(x)c(\theta) \exp(\sum_{i=1}^{k} \omega_i(\theta) t_i(x))$$

Théorème:

Soit $(X_1, ..., X_n)$ échantillon de va d'un modèle exponentiel $(\mathfrak{X}, (P_\theta)_{\theta \in \Theta})$ de dominante μ.
Notons $f(x, \theta) = \frac{d_\theta}{d\mu}(x) = h(x)c(\theta) \exp(\sum_{i=1}^{n} \omega_i(\theta) t_i(x)$
Soient $T_1, ..., T_k$ les statistiques $(x_1, ..., x_n) \mapsto T_i(x_1, ..., x_n) = \sum_{j=1}^{n} t_i(x_j)$.
On suppose que les ensembles $\{(\omega_1(x), ..., \omega_k(\theta)); \theta \in \Theta\}$ et $\{T_1(x_1, ..., x_n)..., T_k(x_1, ..., x_n); (x_1, ..., x_n) \in \mathfrak{X}^n\}$ contiennent des ouverts non vides de $\mathbb{R}^k$.
Alors la loi de $T = (T_1, ..., T_k)$ appartient la famille exponentielle sous la forme :

$$f_T(u_1, ..., u_k, \theta) = H(u_1, ..., u_k)(C(\theta))^n \exp(\sum_{i=1}^{n} \omega_i(\theta) u_i)$$

Théorème: de Slutsky

Soit $(X_n)_{n \geq 1}$ une suite de var qui converge en loi vers une va X et $(Y_n)_{n \geq 1}$ une suite de va qui converge en probabilité vers une constante a.
$(Y_n \xrightarrow[n \to +\infty]{\mathbb{P}} a \Leftrightarrow Y_n \xrightarrow[n \to +\infty]{\mathcal{L}} a)$
On a alors $X_n + Y_n \xrightarrow[n \to +\infty]{\mathcal{L}} X + a$ et $X_n Y_n \xrightarrow[n \to +\infty]{\mathcal{L}} aX$

ℹ *Remarque:*

$(X_n)_{n\geq 1}$ iid, de moyenne μ et de variance $\sigma^2 > 0$. D'après le TLC, on sait que :

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

Supposons que σ ne soit pas connue et que $\lim_{n\to+\infty} V(S_n^2) = 0$. Alors :

$$\text{Soit } \varepsilon > 0 : \mathbb{P}(|S_n^2 - \sigma^2| > \varepsilon) = \mathbb{P}(|S_n^2 - E(S_n^2)| > \varepsilon) \leq \frac{V(S_n^2)}{\epsilon} \xrightarrow[n\to+\infty]{} 0$$

Ainsi, $S_n^2 \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma^2$

On en déduit que $\frac{\sigma}{S_n} \xrightarrow[n\to+\infty]{\mathbb{P}} 1$. D'où :

$$\frac{\sigma}{S_n}\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} = \frac{\sqrt{n}(\bar{X}-\mu)}{S_n} \xrightarrow[n\to+\infty]{\mathbb{P}} N \hookrightarrow \mathcal{N}(0,1)$$

Théorème:

Soit $(X_1, ..., X_n)$ un échantillon de loi normale $\mathcal{N}(\mu, \sigma^2)$ avec $(\mu, \sigma^2) \in \mathbb{R}_*^+ \times \mathbb{R}_*^+$ et soient

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ et}$$

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

la moyenne et la variance empririques associées.
Alors :

1. $X \perp\!\!\!\perp S^2$
2. $\bar{X} \hookrightarrow \mathcal{N}(\mu, \frac{\sigma^2}{n})$
3. $\frac{(n-1)S^2}{\sigma^2} \hookrightarrow \chi^2_{n-1}$

2 Principe de réduction des données : l'exhaustivité

Le principe de base est le suivant :
- observation $x = (x_1, ..., x_n)$
- chaque x_i est régie par P_θ
- θ est le paramètre inconnu
- On infère sur θ à partir de l'échantillon

En général, ce n'est pas x qui est considéré mais un résumé de x ; soit $\phi(x) = \phi(x_1, ..., x_n)$.

A priori, on perd de l'information sur θ en considérant $\phi(x)$ à la place de x.

Rappel :
Soit P une probabilité sur un espace $(\mathfrak{X}, \mathcal{A})$ et soit ϕ une application mesurable de $(\mathfrak{X}, \mathcal{A})$ dans un espace $(\mathfrak{Y}, \mathcal{B})$.
On définit, pour tout $A \in \mathcal{A}$ la probabilité conditionnelle de A sachant ϕ comme étant une variable aléatoire notée

$$P(A|\phi) = E(1_A|\phi)$$

et définie presque sûrement relativement à $\phi(P)$ par :

$$\forall B \in \mathcal{B}, P(A \cap (\phi \in B)) = \int_B P(A|\phi = y) d\phi(P)(y)$$

Définition: Exhaustive

On dit qu'une statistique ϕ définie sur un modèle statistique $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ est exhaustive (par θ) s'il existe une version de la probabilité conditionnelle de A sachant ϕ indépendante de θ.
Si tel est le cas, on note $P(\bullet|\phi)$ une telle version qui vérifie alors lorsque $(\mathfrak{Y}, \mathcal{B})$ est l'espace image de ϕ :

$$\forall A \in \mathcal{A}, \forall B \in \mathcal{B}, \forall \theta \in \Theta, P_\theta(A \cap (\phi \in \ B)) = \int_B P(A|\phi = y) d\phi(P_\theta)(y)$$

Remarque :
En pratique, cette définition n'est pas facile à utiliser.

Définition: Vraisemblance

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique dominé par une mesure σ-finie μ.
Ce modèle est caractérisé par la donnée d'une vraisemblance :

$$\begin{aligned} L : \mathfrak{X} \times \Theta \quad &\to \quad \mathbb{R}^+ \\ (x, \theta) \quad &\mapsto \quad L(\theta, x) = \frac{dP_\theta}{d\mu}(x) \end{aligned}$$

Exemple :

Soit $(X_1, ..., X_n)$ un échantillon de la loi $\mathcal{N}(\theta, \sigma^2)$. Ici, le modèle statistique est $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$
Ici, $\mathfrak{X} = \mathbb{R}$ et $P_\theta = \mathcal{N}(\theta, \sigma^2)$ et μ est la mesure de Lebesgue sur $\mathbb{R}$.

$$\forall x \in \mathbb{R},\ L(\theta, x) = \frac{dP_\theta}{d\mu}(x) = \frac{1}{\sqrt{2\pi}\sigma} \ \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)$$

Si on considère le modèle $(\mathfrak{X}^n, P_\theta^{\otimes n})_{\theta\in\Theta}$, une vraisemblance est : $\forall \theta \in \mathbb{R}, \forall x = (x_1, \cdots, x_n) \in \mathbb{R}^n$:

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

Théorème: de factorisation de Neymann-Fisher :

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique dominé par une mesure σ-finie μ et soit L une vraisemblance. Alors toute statistique ϕ à valeur dans un espace $\mathfrak{Y}$ est exhaustive si et seulement s'il existe une application

$$g : \mathfrak{Y} \times \Theta \to \mathbb{R}^+$$

et une application

$$h : \mathfrak{X} \to \mathbb{R}^+$$

tel que

$$\forall x \in \mathfrak{X}, \forall \theta \in \Theta, L(\theta, x) = g(\phi(x), \theta)h(x)$$

Exemple :

Soit $(X_1, ..., X_n)$ un échantillon de la loi $\mathcal{N}(\theta, \sigma^2)$.

$$\begin{aligned} L(\theta, x) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\theta)^2\right)\right) \\ &= \underbrace{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\right)}_{h(x)} \underbrace{\exp\left(\frac{-n(\bar{x}-\theta)^2}{2\sigma^2}\right)}_{g(\phi(x),\theta)} \end{aligned}$$

D'après le théorème de Neymann-Fischer, la statistique $\phi(x) = \phi(x_1, \cdots, x_n) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ est une statistique exhaustive par le paramètre θ.

Définition: Famille complète

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique. On dit que la famille $(P_\theta)_{\theta\in\Theta}$ est complète ou que le modèle statistique est complet si pour toute fonction g numérique, P_θ intégrable $(\theta \in \Theta)$ définie sur $\mathfrak{X}$, on a :

$$\forall \theta \in \Theta, \int_{\mathfrak{X}} g dP_\theta = 0 \Rightarrow \forall \theta \in \Theta, g = 0 \ P_\theta\text{-p.s.}$$

Définition: Statistique complète

Une statistique $Y : \mathfrak{X} \to \mathfrak{Y}$ est dite complete si son modèle image est complet, ie :
Pour toute fonction g définie sur $\mathfrak{Y}$, $\phi(P_\theta)$-intégrable $(\forall\theta \in \Theta)$, on a :

$$\forall \theta \in \Theta, \int_{\mathfrak{Y}} g d\phi(P_\theta) = 0 \Rightarrow \forall \theta \in \Theta, g = 0 \ \phi(P_\theta)\text{-p.s.}$$

ou encore :

$$\forall\theta\in\Theta,\int_{\mathfrak{Y}} g\circ\phi dP_\theta=0\Rightarrow\forall\theta\in\Theta, g\circ\phi=0\ P_\theta\text{-p.s.}$$

Exemple :

Soit ϕ une variable aléatoire de la loi binomiale $P_\theta=\mathcal{B}(n,\theta),\ \theta\in]0,1[$
Soit g d'intégrale nulle par rapport à la loi P_θ, ie :

$$0=\sum_{k=0}^{n} g(k)P_\theta(\phi=k)=\sum_{k=0}^{n} g(k)\binom{n}{k}\theta^k(1-\theta)^{n-k}=(1-\theta)^k\sum_{k=0}^{n} g(k)\binom{n}{k}\left(\frac{\theta}{1-\theta}\right)^k$$

Lorsque $\theta\in]0,1[$, $\frac{\theta}{1-\theta}$ parcourt $\mathbb{R}_*^+$. Par conséquent, $\forall k\in\{1,\cdots,n\}$, $\binom{n}{k}g(k)=0$
D'où $\forall k\in\{1,\cdots,n\}, g(k)=0$, donc $g=0$ P_θ-ps.

Théorème:

Considérons le cas d'un modèle exponentiel de vraisemblance L donnée par :

$$L(x,\theta)=h(x)c(\theta)\exp(u(\theta)t(\theta))$$

Alors la statistique $T:(x_1,...,x_n)\mapsto\sum_{i=1}^n t(x_i)$ est complète.

Exemple :

$P_\theta=\mathcal{B}(n,\theta)$ avec $\theta\in]0,1[$

$$\forall x\in\{0,\cdots,n\}, L(x,\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x}=\binom{n}{x}(1-\theta)^n\exp\left(x\ln\left(\frac{\theta}{1-\theta}\right)\right)$$

Ainsi, la statistique T définie pour tout $x=(x_1,\cdots,x_n)\in\{0,\cdots,n\}^n$ par :

$$T(x)=\sum_{i=1}^n t(x_i)=\sum_{i=1}^n x_i$$

est complète.

3 Estimation ponctuelle

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique. Sur la base d'une observation x dans $\mathfrak{X}$, on souhaite donner une "estimation" de la valeur inconnue du maramètre θ. Si on note $\phi(x)$ une telle estimation, ce procédé conduit à la détermination d'une fonction ϕ qui par nature est indépendante de θ, qui sera appelé "estimateur".

Définition:

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique. On appelle estimateur tout statistique ϕ à valeur dans Θ.

3.1 Méthodes de construction

3.1.1 Méthode des moments

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique, où Θ est une partie de $\mathbb{R}^k$ $(k \in \mathbb{N}^*)$ et $\mathfrak{X}$ une partie de $\mathbb{R}$. On suppose que pour tout $\theta = (\theta_1, \cdots, \theta_n) \in \Theta \subset \mathbb{R}^k$, la loi P_θ admet un moment d'ordre k, noté :

$$\mu_k = \int_{\mathbb{R}} x^k dP_\theta(x) < \infty$$

Ainsi, pour tout $1 \leq j \leq k$, le moment d'ordre j de P_θ existe et est une fonction μ_j de θ. Soit $\mu_j(\theta_1, ..., \theta_k)$.

Soit $(x_1, ..., x_n)$ une échantillon de la loi P_θ. La méthode des moments consiste à fournir comme estimation de $\theta = (\theta_1, ..., \theta_k)$ le k-uplet

$$\phi(x_1, ..., x_n) = (\hat{\theta}_1, ..., \hat{\theta}_n)$$

solution du système d'équations :

$$\begin{cases} m_1 & \overset{def}{=} & \frac{1}{n}\sum_{i=1}^n x_i & = & \mu_1(\theta_1, ..., \theta_k) & \overset{def}{=} & \mu_1 \\ \vdots & & & & & & \\ m_k & \overset{def}{=} & \frac{1}{n}\sum_{i=1}^n x_i^k & = & \mu_k(\theta_1, ..., \theta_k) & \overset{def}{=} & \mu_k \end{cases}$$

Solution en $\theta_1, ..., \theta_k$

Exemple :

Supposons que $\Theta = \mathbb{R} \times \mathbb{R}_*^+$ et $P_\theta = \mathcal{N}(\lambda, \sigma^2)$ (ie $\theta = (\theta_1, \theta_2) = (\lambda, \sigma^2)$)
Modèle statistique : $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta}) = \mathbb{R}(, (\mathcal{N}(\lambda, \mu))_{(\lambda,\mu)\in\mathbb{R}\times\mathbb{R}_*^+})$ Soit $(X_1, ..., X_n)$ un échantillon de la loi P_θ. Construisons un estimateur de $\theta = (\lambda, \sigma^2)$ via la méthode des moments.

$$\begin{cases} \frac{1}{n}\sum_{i=1}^n X_i & = & E_\theta(X_i) \\ \frac{1}{n}\sum_{i=1}^n X_i^2 & = & E_\theta(X_1)^2 \end{cases} \Leftrightarrow \begin{cases} \bar{X} & = & \lambda \\ \frac{1}{n}\sum_{i=1}^n X_i^2 & = & \sigma^2 + \lambda^2 \end{cases} \Leftrightarrow \begin{cases} \bar{X} & = & \lambda \\ \sigma^2 & = & \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \end{cases}$$

La méthode des moments nous donne $\bar{X}$ comme estimateur de λ et $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ comme estimateur de σ^2. On remarque que $\bar{X}$ est l'estimateur "naturel" sans biais pour λ mais $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ n'est pas l'estimateur sans biais "habituel" pour σ^2.

3.1.2 Méthode du maximum de vraisemblance

Définition:

Soit $(\mathfrak{X}, P_\theta)_{\theta\in\Theta})$ un modèle statistique dominé, de vraisemblace $L : \Theta \times \mathfrak{X} \to \mathbb{R}^+$.
Un estimateur du maximum de vraisemblance (EMV) de θ est une statistique $\hat{\theta} : \mathfrak{X} \to \Theta$ tel que :

$$\forall x \in \mathfrak{X}, L(\hat{\theta}(x), x) = \sup_{\theta\in\Theta} L(\theta, x)$$

Remarque :
On écrit aussi :

$$\forall x \in \mathfrak{X}, \hat{\theta}(x) = \text{argmax}_{\theta\in\Theta} L(\theta, x)$$

Exemple :

Soit $x = (x_1, ..., x_n)$ un échantillon observé de lal oi $\mathcal{N}(\theta, 1)$. La vraisemblance du modèle est définie pour tout $x = (x_1, ..., x_n) \in \mathbb{R}^n$ par :

$$\forall\theta \in \mathbb{R}, L(\theta, x) = \prod_{i=1}^{n} f(\theta, x_i) \text{ où } f(\theta, x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2}}$$

$$L(\theta, x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right)$$

Pour x fixé, il faut étudier $\theta \mapsto L(\theta, x)$. En fait, il suffit d'étudier $\theta \mapsto \log L(\theta, x)$.
Or, $\log L(\theta, x) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2$, donc :

$$\frac{\partial}{\partial\theta} \log L(\theta, x) = \sum_{i=1}^{n}(x_i - \theta) = 0 \Leftrightarrow \theta = \bar{x}$$

Comme $\frac{\partial^2}{\partial\theta^2} \log L(\theta, x)_{|_{\theta=\bar{x}}} = -n < 0$, donc il s'agit bien d'un maximum.

Technique plus directe :

$$\min_{a\in\mathbb{R}} \sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

Ainsi :

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right) \leq \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$$

avec égalité si et seulement si $\theta = \bar{x}$.
Donc $\hat{\theta}(x) = \bar{x}$ est l'EMV pour θ

3.2 Méthodes d'évalutation des estimateurs

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique où Θ est une partie de $\mathbb{R}$. On désigne par E_θ l'espérance par rapport à P_θ pour tout $\theta \in \Theta$.

Définition:

On appelle erreur quadratique moyenne (ou risque quadratique) d'un estimateur w de θ la fonction

$$R_w : \theta \mapsto E_\theta\left((w - \theta)^2\right)$$

Remarque:

$R_w(\theta) = V_\theta(w) + b_w^2(\theta)$ où $b_w(\theta) = |E(w) - \theta|$ est le biais de w.

Dans toute la suite, on va se restreindre aux estimateurs sans biais d'ordre 2 (ie les estimateurs pour lesquels l'EQM existe) de $g(\theta)$ où g est une fonction arbitraire.

Définition: UVMSB

Un estimateur sans biais w^* de $g(\theta)$ est dit meilleur estimateur si pour tout autre estimateur w sans biais de $g(\theta)$ on a :

$$\forall \theta \in \Theta,\ V_\theta(w^*) \leq V_\theta(w)$$

On dit aussi que w^* est uniformément de variance minimale parmi les estimateurs sans biais de $g(\theta)$ (UVMSB)

Un tel estimateur exists-t-il ? Comment en déterminer un ?

Propriété:

Il existe une borne inférieure pour la variance des estimateurs sans biais. Si cette borne est atteinte, on a trouvé l'estimateur UVMSB.

Remarque:

1. Cette borne inférieure existe toujours !
2. Mais elle n'est pas toujours atteinte !

D'après l'inégalité de Cauchy-Schwarz :

$$\mathrm{cov}(X, Y) \leq \sqrt{V(X)}\sqrt{V(Y)}$$

$$\Rightarrow V(X) \geq \frac{\mathrm{cov}^2(X, Y)}{V(Y)} \text{ (si } V(Y) > 0)$$

SUpposons que Θ soit un ouvert de $\mathbb{R}$ et appliquons cette inégalité à une quantité ne dépendant que du modèle statistique : la log-vraisemblance.

Supposons que $\frac{\partial}{\partial \theta} \log L(\theta, x)$ existe pour tout $\theta \in \Theta$ et pour P_θ-presque tout $x \in \mathfrak{X}$.

De plus, on suppose une ondition de régularité qui permet de dériver sous le signe d'intégration. C'est raisonnable puisque les modèles exponentiels le vérifient. (cf TD)

Fixons $\theta \in \Theta$ et calculons $E_\theta\left(\frac{\partial}{\partial\theta}\log L(\theta,X)\right)$ où $X \hookrightarrow P_\theta$.

$$\begin{aligned}
E_\theta\left(\frac{\partial}{\partial\theta}\log L(\theta,X)\right) &= E_\theta\left(\frac{\frac{\partial}{\partial\theta}L(\theta,X)}{L(\theta,X)}\right)\\
&= \int_{\mathfrak{X}} \frac{\frac{\partial}{\partial\theta}L(\theta,x)}{L(\theta,x)}dP_\theta(x)\\
&= \int_{\mathfrak{X}} \frac{\frac{\partial}{\partial\theta}L(\theta,x)}{L(\theta,x)}L(\theta,x)d\mu(x) \quad \mu \text{ étant la domniante du modèle statistique } (\mathfrak{X},(P_\theta)_{\theta\in\Theta})\\
&= \int_{\mathfrak{X}} \frac{\partial}{\partial\theta}L(\theta,x)d\mu(x)\\
&= \frac{d}{d\theta}\int_{\mathfrak{X}} L(\theta,x)d\mu(x)\\
&= \frac{d}{d\theta}\underbrace{\int_{\mathfrak{X}} dP_\theta(x)}_{=1}\\
&= 0
\end{aligned}$$

Ainsi, $\frac{\partial}{\partial\theta}\log L(\theta,X)$ est P_θ-centrée. Donc :

$$\begin{aligned}
\operatorname{cov}\left(W,\frac{\partial}{\partial\theta}\log L(\theta,X)\right) &= E\left(W\times\frac{\partial}{\partial\theta}\log L(\theta,X)\right)\\
&= E\left(W\times\frac{\frac{\partial}{\partial\theta}L(\theta,X)}{L(\theta,X)}\right)\\
&= \int_{\mathfrak{X}} W(x)\frac{\frac{\partial}{\partial\theta}L(\theta,x)}{L(\theta,x)}dP_\theta(x)\\
&= \int_{\mathfrak{X}} W(x)\frac{\partial}{\partial\theta}L(\theta,x)d\mu(x)\\
&= \frac{d}{d\theta}\int_{\mathfrak{X}} W(x)L(\theta,x)d\mu(x)\\
&= \frac{d}{d\theta}\int_{\mathfrak{X}} W(x)dP_\theta(x)\\
&= \frac{d}{d\theta}E_\theta(W)
\end{aligned}$$

De plus, comme $E\left(\frac{\partial}{\partial\theta}\log L(\theta,X)\right)=0$, on a :

$$V_\theta\left(\frac{\partial}{\partial\theta}\log L(\theta,X)\right) = E_\theta\left(\left(\frac{\partial}{\partial\theta}\ \log(\theta,X)\right)^2\right)$$

L'inégalité de Cauchy-Schwarz appliquée) W et $\frac{\partial}{\partial\theta}\ \log L(\theta,X)$ donne :

$$V_\theta(W) \geq \frac{\left(\frac{d}{d\theta}E_\theta(W)\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log L(\theta,X)\right)^2\right)}$$

Théorème: de la borne de Carmer-Rao

Soit $(\mathfrak{X},(P_\theta)_{\theta\in\Theta})$ un modèle statistique dominé par une mesure σ-finie μ et de vraisemblance L où Θ est un ouvert de $\mathbb{R}$. Supposons que, pour toute fonction $(P_\theta)_{\theta\in\Theta}$-intégrable h, nous ayons la propriété :

$$\frac{d}{d\theta}\int_{\mathfrak{X}} h(x)L(\theta,x)d\mu(x) = \int_{\mathfrak{X}} h(x)\frac{\partial}{\partial\theta}L(\theta,x)d\mu(x)$$

Alors pour tout estimateur W, la fonction $\theta \mapsto E_\theta(W)$ est dérivable sur Θ et de plus,

$$\forall \theta \in \Theta, V_\theta(W) \geq \frac{\left(\frac{d}{d\theta} E_\theta(W)\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta} \log L(\theta, X)\right)^2\right)}$$

En particulier, si W est un estimateur sans biais de $g(\theta)$, alors :

$$V_\theta(W) \geq \frac{\left(\frac{d}{d\theta} g(\theta)\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta} \log L(\theta, \bullet)\right)^2\right)}$$

Définition:

$\frac{\partial}{\partial\theta} \log L(\theta, \bullet)$ est appellé le "score" du modèle statistique et $E_\theta\left(\left(\frac{\partial}{\partial\theta} \log L(\theta, \bullet)\right)^2\right)$ l'information de Fischer du modèle en θ, qu'on note $I(\theta)$

Lemme: technique

Si la vraisemblance L du modèle statistique vérifie :

$$\frac{\partial}{\partial\theta} E_\theta\left(\frac{\partial}{\partial\theta} \log L(\theta, \bullet)\right) = \int_{\mathfrak{X}} \frac{\partial^2}{\partial\theta^2} \log L(\theta, x) d\mu(x)$$

alors

$$E_\theta\left(\left(\frac{\partial}{\partial\theta} \log L(\theta, \bullet)\right)^2\right) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2} \log L(\theta, \bullet)\right)$$

Démonstration :

$$\begin{aligned}
E_\theta\left(\frac{\partial^2}{\partial\theta^2} \log L(\theta, \bullet)\right) &= E_\theta\left(\frac{\partial}{\partial\theta}\left(\frac{\frac{\partial}{\partial\theta)L(\theta,\bullet)}}{L(\theta, \bullet)}\right)\right) \\
&= E_\theta\left(\frac{\left(\frac{\partial^2}{\partial\theta^2} L(\theta, \bullet)\right) L(\theta, \bullet) - \left(\frac{\partial}{\partial\theta} L(\theta, \bullet)\right)^2}{(L(\theta, \bullet))^2}\right) \\
&= E_\theta\left(\frac{\frac{\partial^2}{\partial\theta^2} L(\theta, \bullet)}{L(\theta, \bullet)}\right) - E_\theta\left(\left(\frac{\frac{\partial}{\partial\theta} L(\theta, \bullet)}{L(\theta, \bullet)}\right)^2\right)
\end{aligned}$$

$$\begin{aligned}
\text{Or, } E_\theta\left(\frac{\frac{\partial^2}{\partial\theta^2}L(\theta,\bullet)}{L(\theta,\bullet)}\right) &= \int_{\mathfrak{X}} \frac{\frac{\partial^2}{\partial\theta^2}L(\theta,x)}{L(\theta,x)}dP_\theta(x)\\
&= \int_{\mathfrak{X}} \frac{\partial^2}{\partial\theta^2}L(\theta,x)d\mu(x)\\
&= \frac{d}{d\theta}\int_{\mathfrak{X}} \frac{\partial}{\partial\theta}L(\theta,x)d\mu(x)\\
&= \frac{d}{d\theta}\int_{\mathfrak{X}} \frac{\partial}{\partial\theta}\log L(\theta,x)dP_\theta(x)\\
&= \frac{d}{d\theta}\underbrace{E_\theta\left(\frac{\partial}{\partial\theta}\log L(\theta,x)\right)}_{=0}\\
&= 0
\end{aligned}$$

$$\text{D'autre part } E_\theta\left(\left(\frac{\frac{\partial}{\partial\theta}L(\theta,\bullet)}{L(\theta,\bullet)}\right)^2\right) = E_\theta\left(\left(\frac{\partial}{\partial\theta}\log L(\theta,\bullet)\right)^2\right)$$

$$\text{D'où } E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta,\bullet)\right) = -E_\theta\left(\left(\frac{\partial}{\partial\theta}\log L(\theta,\bullet)\right)^2\right)$$

Lemme:

Suppsons que pour tout $\theta\in\Theta$ et tout $x=(x_1,...,x_n)\in\mathfrak{X}^n$, on ait $L(\theta,x)=\prod_{i=1}^n f(\theta,x_i)$. alors :

$$I(\theta) = E_\theta\left(\left(\frac{\partial}{\partial\theta}\log L(\theta,\bullet)\right)^2\right) = -nE_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(\theta,\bullet)\right)$$

Démonstration :

$$\begin{aligned}
E_\theta\left(\left(\frac{\partial}{\partial\theta}\log L(\theta,\bullet)\right)^2\right) &= E_\theta\left(\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(\theta,X_i)\right)^2\right)\\
&= E_\theta\left(\left(\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f(\theta,X_i)\right)^2\right)\\
&= \sum_{i=1}^n E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\theta,X_i)\right)^2\right) + 2\sum_{1\le i\le j\le n}\underbrace{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\theta,X_i)\right)\left(\frac{\partial}{\partial\theta}\log f(\theta,X_j)\right)\right)}_{E_\theta\left(\frac{\partial}{\partial\theta}\log f(\theta,X_i)\right)E_\theta\left(\frac{\partial}{\partial\theta}\log f(\theta,X_j)\right)=0}\\
&= nE_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\theta,\bullet)\right)^2\right)\\
&= -nE_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(\theta,\bullet)\right)
\end{aligned}$$

Exemple :

A REPRENDRE

Définition: Estimateur efficace

Un estimateur sans biais du paramètre $g(\theta)$ qui atteint la borne de Cramer-Rao est dit efficace.

Théorème:

Sans les conditions du théorème précédnt, si W est un estimateur sans biais de $g(\theta)$ alors W atteint la borne de Cramer-Rao si et seulement s'il existe une fonction a de θ tel que :

$$\frac{\partial}{\partial\theta} \log L(\theta, \bullet) = a(\theta)(W - g(\theta)) \; P_\theta\text{-ps}$$

Propriété:

Soient X et Y deux variables aléatoires réelles de arré intégrable. On a :

1. E(X)=E(E(X|Y))
2. V(X)=V(E(X|Y))+E(V(X|Y)) où V(X|Y)=$E\left((X - E(X))^2|Y\right)$ = E(X^2|Y)+$(E(X|Y))^2$

Théorème: de Rao-Blackwell

Soit W un estimateur sans biais de $g(\theta)$ et soit T une statistique exhaustive pour θ. Alors $E(W|T)$ est un estimateur sans biais de $g(\theta)$ de variance inférieure ou égale à celle de W.

Conséquence Pour obtenir un extimateur UVMSB pour $g(\theta)$, il faut considérer les estimateur fonction d'une statstique exhaustive.

Théorème:

Si W est UVMSB alors il est unique.

Théorème:

Soit T une statistique exhaustive et complète pour le paramètre θ.
Alors toute statistique $\phi(T)$ fondée sur T constitue l'unique estimateur UVMSB de son espérance $(E(W|T) = \phi(T))$

Remarque:

Si W est un estimateur sans biais pour le paramètre $g(\theta)$ et si T est une statistique exhaustive et complète, alors l'améliorée de Rao-Blackwell $E(W|T)$ est l'unique estimateur UVMSB pour $g(\theta)$.

Exemple :

A reprendre

Deuxième partie

Tests d'hypothèse

1 Introduction

Exemple :

Une pièce de monnaie est truquée. On cherche à savoir sur la base de deux lancers de cette pièce si cette dernière donne avec une probabilité $\frac{1}{4}$ "FACE" ou "PILE".

On associe à cette expérience aléatoire un modèle statistique (la réussite étant d'obtenir FACE).

$$(\{0,1\},(P_\theta)_{\theta\in\{\frac{1}{4},\frac{3}{4}\}}) \text{ où } P_\theta = \mathcal{B}(\theta)$$

On note $H_0 : \theta = \frac{1}{4}$ (Hypothèse nulle)
et $H_1 : \theta = \frac{3}{4}$ (Hypothèse alternative).

On cherche une règle de décision permettant au vu des résultats de l'expérience de décider laquelle des deux hypothèses est vérfiée.
Deux erreurs sont possibles :

- Décider que H_1 est vraie à tort (erreur de 1ère espèce)
- Décider que H_0 est vraie à tort (erreur de 2nde espèce)

En pratique, on contrôle l'erreur de 1ère espèce qui ne soit pas dépasser un risque $\alpha \in]0,1[$ fixé à l'avance (en général, on choisit $\alpha = 0,05$ ou $\alpha = 0,1$).
Puis parmi les règles de décisions qui respectent cette contrainte, on choisira celle qui rend l'erreur de 2nde espèce minimale.

Prenons $\alpha = 0,1$. Considérons la règle de décision suivante :

"Si on obtient deux fois FACE, on décide H_1"

Dans le cas contraire, on décide évidemment H_0.

Cette règle respècte la contrainte :

$$P_{H_0}(H_1) = P_{H_0}(F,F) = \frac{1}{16} < \alpha$$

Dans ce cas :

$$P_{H_1}(H_1) = P_{H_0}(F,F) = \frac{9}{16}$$

Autre règle de décision :

"Si on obtient deux fois PILE, on décide H_0, si on obtient deux fois FACE, on décide H_1, et dans les autres cas, on tire au sort la décisio ; on décide H_1 avec une probabilité $\gamma \in]0,1[$ indépendemment du résultat obtenu."

Pour satisfaire à la contrainte, il faut choisir γ tel que :

$$P_{H_0}(H_1) = \frac{1}{6} + \frac{6\gamma}{16} \leq 0,1$$

Soit $\gamma \leq 0,1$.

On choisit ensuite *gamma* tel que :

$$P_{H_1}(H_1) = \frac{9}{16} + \frac{6\gamma}{16}$$

soit maximal. On en déduit $\gamma = 0,1$. On obtient ainsi :

$$P_{H_1}(H_1) = \frac{9,6}{16}$$

On remarque que cette probabilité est bien plus haute que la précédente. On dit donc que cette règle de décision est plus puissante que la première.
En d'autres termes, elle a une erreur de 2nde espèce plus petite que celle de la 1ère règle de décision.

Définition: Plus généralement

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique.
On fait une hypothèse sur θ et on désire savoir si l'observation x_i est cohérente avec l'hypothèse ou au contraire la contredit.
Formellement, on note souvent cette hypothèse H_0 et celle-ci est comparée à son omplémentaire, notée H_1.

H_0 est dite "hypothèse nulle"
H_1 est dit "hypothèse alternative"

L'espace des paramètres Θ est subdivisée en deux parties disjointes :
- une partie Θ_0 correspondant à H_0
- une partie Θ_1 correspondant à H_1

On note $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta_1$.

Exemple :

Supposons que θ désigne le changement de pression arterielle provoquée par la prise d'un médicament et que l'on souhaite tester si $H_0 : \theta = 0$. Dans ce cas, $H_1 : \theta \neq 0$ traduit la présence d'un effet du médicament.

Définition:

Une procédure de test d'hypothèse ou un test est une règle de décision qui spécifie :
- la partie de l'espace $\mathfrak{X}$ des observations pour laquelle on accepte l'hypothèse nulle (région d'acceptation)
- la partie de l'espace $\mathfrak{X}$ des observations pour laquelle on rejette l'hypothèse nulle (région critique)

Typiquement, la région d'acceptation et la région critique sont définie par l'intermédiaire d'une certaine statistique ϕ.
Souvent, la région critique est de la forme $\{\phi > a\}$ où a est une certaine constante.

2 Test du rapport de vraisemblance

Soit $(\mathfrak{X}, (P_\theta)_{\theta\in\Theta})$ un modèle statistique dominé pa une mesure σ-finie μ et soit L une vraisemblance de ce modèle par rapport à μ. On souhaite tester l'hypotèhse $H_0 : \theta \in \Theta_0$ contre l'hpothèse $H_1 : \theta \notin \Theta_0$

Définition:

On appelle statistique de test du rapport de vraisemblance (TRV) la statistique ϕ définie pour tout $x \in \mathfrak{X}$ par :

$$\phi(x) = \frac{\sup_{\theta\in\Theta_0} L(\theta, x)}{\sup_{\theta\in\Theta} L(\theta, x)}$$

Un test de rapport de vraisemblance est un test dont la région critique est de la forme $\{\phi \leq c\}$ où c est une constante dans [0,1].

Exemple :

Supposons disposer d'un échantillon $(x_1, ..., x_n)$ de la loi $\mathcal{N}(\theta, 1)$ et que l'on souhaite tester l'hypothèse $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ où θ_0 est fixé.

La statistque du TRV est :

$$\phi(x) = \frac{L(\theta_0, x)}{\sup_{\theta\in\Theta} L(\theta, x)} = \frac{L(\theta_0, x)}{L(\bar{x}, x)}$$

Car $\bar{X}$ est l'EMV pour θ.

$$\Rightarrow \phi(x) = \exp\left(\frac{1}{2}\left(-\sum_{i=1}^{n}(x_i - \theta_0)^2 + \sum_{i=1}^{n}(x_i - \bar{x}^2)\right)\right)$$

Or, $\sum_{i=1}^{n}(x_i - \theta_0)^2 = \sum_{i=1}^{n}(x_i - \bar{x}^2 + n(\bar{x} - \theta_0)^2$, d'où :

$$\phi(x) = \exp\left(-\frac{1}{2}n(\bar{x} - \theta_0)^2\right)$$

Un TRV est un test qui rejette h_0 pour de "petites valeurs" de $\phi(x)$. Autrement dit, un test de région critique :

$$R = \{x \in \mathbb{R}^n;\ |\bar{x} - \theta_0| \geq \left(-\frac{2}{n}\log c\right)^{\frac{1}{2}}\}$$

2.1 Probabilité d'erreur et fonction puissance

On teste $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$. Notons R la région critique du test.

	On accèpte H_0	On accèpte H_1
H_0 est vraie	Décision correcte	Erreur de première espèce : $P_\theta(R)$ avec $\theta \in \Theta_0$
H_0 est fausse	Erreur de 2nde espèce : $P_\theta(R)$ avec $\theta \in \Theta_1$	Decision correcte

Définition:

On appelle fonction puissance d'un teste de région critique R la fonction β définie pour tout $\theta \in \Theta$ par :

$$\beta(\theta) = \begin{cases} \text{Proba de 1ère espèce si } \theta \in \Theta_0 \\ 1 - (\text{proba de 2ème espèce) si } \theta \in \Theta_1 \end{cases}$$

Remarque:

L'idéal serait que β soit nulle sur θ_0 e maximal sur θ_1.

Un "bon" test apprcherait cet idéal :

1. β proche de 0 sur θ_0
2. β proche de 1 sur θ_1

En général, il est impossible de minimiser simultanément la probabilité des erreur de 1ère et de 2nde espèce. Classiquement, on choisit de borner l'erreur de 1ère espèce et de minimiser l'erreur de 2nde espèce.

Définition:

Pour tout $\alpha \in [0,1]$, un test de fonction puissance β est dit de niveau (respectivement de seuil) α si :

$$\sup_{\theta\in\Theta_0} \beta(\theta) = \alpha \text{ (respectivement : } \sup_{\theta\in\Theta_0} \beta(\theta) \leq \alpha)$$

Définition:

Un test de fonction puissance β est dit sans biais si :

$$\forall \theta' \in \Theta_1, \forall \theta'' \in \Theta_0, \beta(\theta') \geq \beta(\theta'')$$

3 Tests les plus puissants

Définition:

Soit τ une classe de tests de $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$.
Un test (dans τ) de fonction puissance β est dit uniformément le plus puissant (UPP) dans la classe τ si, quelque soit le test H dans τ de fonction puissance β', on a :

$$\forall \theta \in \Theta_0^c, \beta'(\theta) \leq \beta(\theta)$$

Remarque:

"UPP" est une exigence forte !
Dans le cas d'un test d'hypothèse simple contre une alternative simple, on dispose d'une caractérisation.

⚮ *Lemme: de Neymann-Pearson*

Supposons que $\theta = \{\theta_0, \theta_1\}$ et que l'on souhaite tester $H_0 : \theta = \theta_0$ conte $h_1 : \theta = \theta_1$.
Soit L une vraisemblance du modèle. Pour $k > 0$ fixé, considérons le reste de région critique R vérifiant :

1. $x \in R$ si $L(\theta, x) > kL(\theta_0, x)$
2. $x \notin R$ si $L(\theta, x) < kL(\theta_0, x)$
3. $P_{\theta_0}(R) = \alpha$

alors ce test est UPP de niveau α