

1 Introduction à la théorie de l'estimation et des tests paramétriques

1.1 Position du problème

On se place dans le cadre de l'observation répétée de façon indépendante d'un phénomène aléatoire dont on voudrait découvrir la loi. De fait, dans ce cours, on suppose connu le type de la loi, et inconnus ses paramètres, à valeurs dans Θ sous ensemble de \mathbb{R} ou \mathbb{R}^d , que l'on veut estimer et tester. On parle de "statistique paramétrique". Au vu des observations, on se pose trois problèmes :

- quelle est la valeur du paramètre ,
- ce paramètre appartient-il à un sous-ensemble précisé de Θ ?
- peut-on donner avec une certaine "confiance" un intervalle $I \subset \Theta$ où θ , la vraie valeur du paramètre, se trouve avec une grande probabilité ?

Il s'agit respectivement de l'estimation, de l'estimation par intervalles et du test.

1.2 Définitions

On appelle **modèle statistique** la donnée d'un espace mesurable (Ω, \mathcal{A}) muni d'une famille de probabilités indexée par une famille de paramètres Θ :

$$(\mathbb{P}_\theta, \theta \in \Theta) \text{ sur l'espace mesurable } (\Omega, \mathcal{A}).$$

On appelle **n -échantillon** de ce modèle un ensemble de n variables aléatoires indépendantes $(X_i, i = 1, \dots, n)$ de même loi \mathbb{P}_θ . L'alea ω étant réalisé, on note $(x_i = X_i(\omega), i = 1, \dots, n)$ la réalisation, l'observation de l'échantillon.

On appelle **statistique** toute variable aléatoire fonction du n -échantillon $T : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}^k, T(\omega) = f(X_1(\omega), \dots, X_n(\omega))$ où f est une fonction borélienne de \mathbb{R}^n dans \mathbb{R}^k . Il est important de remarquer que T ne dépend pas de θ mais seulement de l'échantillon.

Exemple de statistique très courante : $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$.

Exemple de modèle : $\Omega = \{0, 1\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\Theta = [0, 1]$, $\mathbb{P}_p(0) = 1 - p$, $\mathbb{P}_p(1) = p$. C'est le modèle de Bernoulli dont un n -échantillon $(X_i, i = 1, \dots, n)$ est à valeurs dans $\{0, 1\}^n$. Une statistique est $T = \sum_i X_i$ qui suit une loi binomiale $B(n, p)$.

1.3 Modèles dominés

Définition 1.1 Un modèle est dit “dominé”, si toute les probabilités \mathbb{P}_θ sont absolument continues par rapport à une mesure μ :

$$(\Omega, \mathcal{A}, (\mathbb{P}_\theta = f_\theta \cdot \mu, \theta \in \Theta)).$$

La donnée de la famille des densités $(f_\theta, \theta \in \Theta)$ est équivalente à celle de la famille des probabilités $(\mathbb{P}_\theta, \theta \in \Theta)$.

Exemples

(i) Si les probabilités $(\mathbb{P}_\theta, \theta \in \Theta)$ sont à support dans les entiers \mathbb{N} , la mesure de comptage $\mu(n) = 1, \forall n$, est une mesure dominante; tous les modèles discrets sont donc dominés.

(ii) Si les probabilités $(\mathbb{P}_\theta, \theta \in \Theta)$ sont à support dans \mathbb{R}^k et absolument continues par rapport à la mesure de Lebesgue, le modèle est dominé. Tous les modèles avec des lois continues sont dominés, par exemple une famille de gaussiennes.

1.4 Introduction à la théorie de l'estimation

Définition 1.2 Soit une application mesurable $g : (\Theta, \mathcal{T}) \rightarrow (\Theta', \mathcal{T}')$. On appelle “estimateur” de g toute statistique à valeurs dans Θ' . Si $g = id$, on parle d'estimateur de θ .

Exemple : dans le modèle de Bernoulli, \bar{X}_n est un estimateur de p . On verra plus tard quelles sont ses qualités.

Quelques méthodes de recherche d'estimateurs.

(i) Méthode des moments : Soit un modèle sur $(\mathbb{R}, \mathcal{B})$ et son image par le n -échantillon, $(\mathbb{R}^n, \mathcal{B}^{\otimes n}, \mathbb{P}_\theta^{\otimes n}(\theta \in \Theta))$. On suppose que $\int |x| \mathbb{P}_\theta(dx) < \infty$. Alors pour tout i , $E_\theta[X_i]$ est une fonction de θ , notée $m(\theta)$. On a que $E_\theta(X_n) = m(\theta)$, la statistique \bar{X}_n et un estimateur de m .

Si la fonction m est inversible, on peut proposer comme estimateur de θ la statistique $m^{-1}(\bar{X}_n)$. Dans le modèle de Bernoulli, $E_\theta(X_n) = \theta$, et \bar{X}_n est un estimateur de θ . On l'appelle l'estimateur “naturel” de θ .

Si la fonction m n'est pas inversible, et si l'échantillon admet un moment d'ordre 2, on peut essayer de résoudre avec $m_2(\theta) = E_\theta[X_i^2]$ de résoudre le système $m(\theta) = \bar{X}_n, m_2(\theta) = \frac{\sum_i X_i^2}{n}$ dont sera solution un estimateur $\hat{\theta}$ du paramètre θ .

Exemple-exercice du modèle gaussien $\mathcal{N}(m, \sigma^2)$: on cherche un couple (m, σ^2) solution du système $m = \bar{X}_n, m^2 + \sigma^2 = \frac{\sum_i X_i^2}{n}$. D'où la proposition de l'estimateur

$$\hat{m} = \bar{X}_n, \hat{\sigma}^2 = \frac{\sum_i X_i^2}{n} - (\bar{X}_n)^2.$$

Lorsque le paramètre θ est vectoriel de dimension d et si l'échantillon admet des moments jusqu'à l'ordre d , on cherche à résoudre le système

$$(1) \quad \begin{aligned} m_1(\theta) &= \bar{X}_n \\ m_2(\theta) &= \frac{\sum_i X_i^2}{n} \\ &\dots \\ m_d(\theta) &= \frac{\sum_i X_i^d}{n}. \end{aligned}$$

(ii) Méthode du maximum de vraisemblance. On s'intéresse à la densité d'un modèle statistique régulier comme fonction de θ sur Θ . Pour un n -échantillon, à θ fixé, l'application $(x_1, \dots, x_n) \mapsto \prod_{i=1}^n f_\theta(x_i)$ représente la probabilité de la réalisation de l'observation (x_1, \dots, x_n) . Maintenant, à (x_1, \dots, x_n) fixé, l'application $\theta \mapsto \prod_{i=1}^n f_\theta(x_i)$ représente la vraisemblance de la valeur θ lorsque l'on a observé (x_1, \dots, x_n) .

Définition 1.3 La statistique $\omega \mapsto \operatorname{argmax}(\theta \mapsto \prod_{i=1}^n f_\theta(X_i(\omega)))$ s'appelle l'estimateur de maximum de vraisemblance de θ .

$L : \theta \mapsto \prod_{i=1}^n f_\theta(x_i)$ s'appelle la fonction vraisemblance du modèle.

$l : \theta \mapsto \sum_{i=1}^n \log f_\theta(x_i)$ s'appelle la fonction log-vraisemblance du modèle.

En pratique, on fait l'étude de l'une des fonctions L ou l . Il n'y a pas forcément unicité. Ces fonctions ne sont pas nécessairement dérivables et.... ce qui annule le gradient ne réalise pas forcément un maximum !!!!

Exemple 1 : le modèle de la loi exponentielle, $\Theta = \mathbb{R}^+$, $f_\theta(x) = \theta e^{-\theta x}$, $L(\theta) = \theta^n e^{-\sum_i x_i}$, $l(\theta) = n \log \theta - \sum_i x_i$. l est une application concave et dérivable qui est maximum au point qui annule sa dérivée, soit $\hat{\theta} = \frac{n}{\sum_i X_i}$.

Exemple 2 : le modèle de la loi uniforme, $\Theta = \mathbb{R}^+$, $f_\theta = \frac{1}{\theta} \mathbf{1}_{[0, \theta]}$, $L(\theta) = \frac{1}{\theta^n} \prod_i \mathbf{1}_{[0, \theta]}(x_i)$. Il s'agit ici d'une fonction non dérivable. On vérifie $\prod_i \mathbf{1}_{[0, \theta]}(x_i) = \mathbf{1}_{[\sup_i x_i, \infty)}(\theta)$ qui est maximum en $\hat{\theta} = \sup_i X_i$.

(iii) Méthode des moindres carrés (cas de la régression linéaire).

On dispose d'un n -échantillon dans \mathbb{R}^2 : (X_i, Y_i) et l'on propose le modèle $Y_i = a + bX_i + \varepsilon_i$ où les variables aléatoires ε_i sont indépendantes, centrées et de variance σ^2 . On estime les paramètres (a, b) par $\operatorname{argmin}((a, b) \mapsto \sum_i (Y_i - a - bX_i)^2)$.

1.5 Introduction à la théorie des tests paramétriques

Dans le cas du modèle de Bernoulli, on cherche à décider si le paramètre $p \in \Theta = [0, 1]$ est égal à 0.25 ou supérieur à 0.5. Il peut par exemple s'agir d'un pourcentage de pièces usinées défectueuses, et cette connaissance sur p est importante à la fois pour l'acheteur et

le vendeur. On a vu qu'un estimateur de p est \bar{X}_n , l'idée est donc de décider que la vraie valeur du paramètre $p_0 = 0.25$ si \bar{X}_n est "petit", et que $p_0 \geq 0.5$ si \bar{X}_n est "grand". Tout le problème sera donc de choisir un seuil en dessous duquel \bar{X}_n est "petit", en dessus duquel \bar{X}_n est "grand". On considère donc l'événement " \bar{X}_n est petit" = $\{\omega : \bar{X}_n(\omega) \leq C\}$. On a alors deux manières de se tromper :

1. si $p_0 = 0.25$ et que l'on observe $\{\bar{X}_n > C\}$, on se trompe ;

2. si $p_0 = 0.5$ et que l'on observe $\{\bar{X}_n \leq C\}$, on se trompe aussi.

On cherche C en sorte que les deux erreurs ne soient pas trop grandes ni l'une ni l'autre en probabilité :

$\mathbb{P}_{0.25}\{\bar{X}_n > C\}$ que l'on appelle l'erreur de première espèce,

$\sup_{p \geq 0.5}\{\bar{X}_n \leq C\}$ que l'on appelle l'erreur de deuxième espèce.

On verra tout ceci en détail dans le chapitre 5.

Application lorsque n est grand, car alors par le théorème de la limite centrale, la loi de \bar{X}_n est proche d'une loi gaussienne $\mathcal{N}(p, \frac{p(1-p)}{n})$.

2 Statistiques exhaustives, modèles exponentiels, théorème de Rao-Blackwell

2.1 Statistiques exhaustives

On considère un modèle (ou une structure) statistique

$$(*) \quad (\Omega, \mathcal{A}, (\mathbb{P}_\theta, \theta \in \Theta)).$$

Intuitivement une statistique exhaustive est une statistique qui, à elle seule, concentre toute l'information relative à l'état du paramètre $\theta \in \Theta$. Mathématiquement, une statistique T est exhaustive pour le modèle $(*)$ ci-dessus si la loi conditionnelle $\mathbb{P}_\theta(.|T)$ ne dépend pas de θ . Mais ceci n'a pas grand sens puisque cette loi est définie \mathbb{P}_θ presque sûrement. D'où plus précisément :

Définition 2.1 Soit $T : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$. C'est une statistique exhaustive pour le modèle $(*)$ si la loi conditionnelle \mathbb{P}_θ sachant T ne dépend pas de θ au sens suivant : $\forall X$ va bornée ou intégrable sur $(\Omega, \mathcal{A}, \mathbb{P}_\theta)$, il existe f mesurable réelle sur (E, \mathcal{E}) telle que

$$\forall \theta \in \Theta, E_\theta[X/\sigma(T)] = f(T), \quad \mathbb{P}_\theta \text{ presque sûrement .}$$

Lorsque le modèle est dominé, on dispose d'une propriété caractéristique plus maniable, il s'agit du théorème de factorisation de Neymann :

Proposition 2.2 Soit le modèle dominé

$$(\Omega, \mathcal{A}, (\mathbb{P}_\theta = f_\theta \cdot \mu, \theta \in \Theta))$$

et une statistique $T : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$. T est exhaustive si et seulement si il existe g \mathcal{A} -mesurable positive, et pour tout θ il existe g_θ \mathcal{E} -mesurable positive telles que $f_\theta = g_\theta \circ T \cdot g$, ceci μ presque sûrement.

Preuve de la condition suffisante : On peut supposer $\mu(g) = 1$; en effet, si $\mu(g) < \infty$, on récrit $\mathbb{P}_\theta = \mu(g) f_\theta \frac{\mu}{\mu(g)}$, et sinon, il existe une fonction mesurable positive strictement k telle que $kg \in L^1(\mu)$ et l'on récrit $\mathbb{P}_\theta = \mu(kg) \frac{f_\theta}{kg} \frac{kg\mu}{\mu(kg)}$.

Ainsi $g \cdot \mu$ est-elle une probabilité sur (Ω, \mathcal{A}) . Soit X une variable aléatoire bornée sur (Ω, \mathcal{A}) et $Y = E_{g\mu}(X/\sigma(T)) = f(T)$ avec f mesurable par définition de l'espérance conditionnelle et le lemme de Doob.

Soit h borélienne bornée sur (E, \mathcal{E}) , alors $h(T) \in L^\infty(\sigma(T), E_{g\mu}[Yh(T)] = E_{g\mu}[Xh(T)])$. Par ailleurs, $E_\theta[Xh(T)] = \int_\Omega X(\omega)h \circ T(\omega)g_\theta \circ T(\omega)g(\omega)\mu(d\omega) = E_{g\mu}[X(hg_\theta)(T)] = E_{g\mu}[Y(hg_\theta)(T)] = E_\theta[Xh(T)]$ ce qui montre que l'espérance conditionnelle de $X, Y = h(T)$ c'est à dire la définition de l'exhaustivité.

Condition nécessaire : on ajoute à la famille (f_θ) la fonction 1 pour que μ soit dans la famille du modèle (comme dans la CS, on se débrouille pour que μ soit une probabilité) : si $f > 0, E_\mu(f) > 0$, on prend $\mathbb{P}_\theta = \frac{f_\theta}{f} \cdot f \cdot \mu$.

L'hypothèse donne que $E_\theta(X/\sigma(T)) = f(T)$ \mathbb{P}_θ -presque sûrement, en particulier, $E_\mu(X/\sigma(T)) = f(T)$ μ -presque sûrement. On note $T^*\mathbb{P}_\theta$ et $T^*\mu$ les mesures images par T de \mathbb{P}_θ et μ sur (E, \mathcal{E}) . Comme \mathbb{P}_θ est absolument continue par rapport μ , il en est de même pour $T^*\mathbb{P}_\theta$ par rapport $T^*\mu$ et on note g_θ la densité ($T^*\mathbb{P}_\theta = g_\theta \cdot T^*\mu$) : c'est une fonction \mathcal{E} -mesurable positive.

Soit $A \in \mathcal{A}, X = \mathbf{1}_A$; par définition, $h(T) = E_\mu[X/\sigma(T)]$ μ -presque sûrement, donc \mathbb{P}_θ -presque sûrement.

$$\begin{aligned} \mathbb{P}_\theta(A) &= \int_{\Omega} \mathbf{1}_A d\mathbb{P}_\theta(\omega) = \int_{\Omega} h \circ T(\omega) d\mathbb{P}_\theta(\omega) = \\ &= \int_E h(t) dT^*\mathbb{P}_\theta(t) = \int_E h(t) g_\theta(t) dT^*\mu(t) = \\ &= \int_{\Omega} h \circ T(\omega) g_\theta \circ T(\omega) d\mu(\omega) = \int_{\Omega} E_\mu[X/\sigma(T)](\omega) g_\theta \circ T(\omega) d\mu(\omega) = \\ &= \int_{\Omega} X g_\theta \circ T(\omega) d\mu(\omega) \end{aligned}$$

ce qui donne le résultat puisque $X = \mathbf{1}_A$. □

L'intérêt des statistiques exhaustives est double :

- elles permettent de réduire la dimension de l'espace des observations et
- elles vérifient le résultat suivant.

Proposition 2.3 *Soit un modèle statistique*

$$(\Omega, \mathcal{A}, (\mathbb{P}_\theta, \theta \in \Theta))$$

avec $\Theta \subset \mathbb{R}^p$ et T une statistique exhaustive. Soit une autre statistique $S : (\Omega, \mathcal{A}) \rightarrow (\Theta, \mathcal{T})$ et $R(\theta, S) = E_\theta[(S - \theta)^2]$ qui mesure l'erreur quadratique quand on estime θ par S . Alors, $R(\theta, E(S/\sigma(T))) \leq R(\theta, S)$.

R est le "risque quadratique".

Preuve : c'est une conséquence du fait que le carré est une fonction convexe :

$$(E(X/\mathcal{B}))^2 \leq E(X^2/\mathcal{B}).$$

□

On montre de la même façon le

Théorème 2.4 de RAO-BLACWELL : sous les mêmes hypothèses, si L est une fonction sur Θ^2 convexe en sa deuxième coordonnée :

$$E_\theta[L(\theta, E(S/\sigma(T)))] \leq E_\theta[L(\theta, S)].$$

C'est une conséquence de l'inégalité de Jensen (exercice)

L est appelée "fonction de perte" et $R_\theta(S) = E_\theta[L(\theta, S)]$ le risque de l'estimateur S .

2.2 Modèles exponentiels

Ce sont des modèles très courants, où les statistiques exhaustives sont particulièrement faciles à identifier.

Définition 2.5 *Un modèle*

$$(\Omega, \mathcal{A}, (\mathbb{P}_\theta = f_\theta \cdot \mu, \theta \in \Theta))$$

est dit exponentiel s'il est dominé par une mesure μ et s'il existe une fonction $Q : \Theta \rightarrow \mathbb{R}^p$, une constante $c(\theta)$ et une statistique T à valeurs dans \mathbb{R}^p telles que

$$f_\theta = c(\theta) \exp(\langle Q(\theta), T \rangle).$$

On dit que T est la **statistique naturelle**, elle est exhaustive (*exercice*).
On peut faire le changement de paramètre

$$\eta = Q(\theta)$$

car on voit que $c(\theta)$ peut s'écrire uniquement en fonction de η :

$$c(\theta)^{-1} = E_\mu[\exp(\langle \eta, T \rangle)].$$

Enfin, toutes les probabilités \mathbb{P}_θ sont de même support que μ . Une famille dont les supports varient en fonction de θ ne peut être exponentielle.

Exemples et exercices : lesquels de ces modèles sont exponentiels ? si oui, pourquoi ?

loi uniforme sur un intervalle

loi exponentielle

loi Gamma

loi géométrique

loi hypergéométrique

Proposition 2.6 *Soit un modèle exponentiel à paramètre canonique (c'est à dire que $\mathbb{P}_\theta = c(\theta) \exp(\langle \theta, T \rangle)$), on peut étendre le domaine Θ à son enveloppe convexe.*

Ceci se fait de la façon suivante :

$$\hat{\Theta} = \{\lambda\theta_1 + (1 - \lambda)\theta_2, \lambda \in [0, 1], \theta_i \in \Theta\}$$

et pour $\theta \in \hat{\Theta}$, $\mathbb{P}_\theta = \frac{\exp(\langle \theta, T \rangle) \cdot \mu}{\mu(\exp(\langle \theta, T \rangle))}$.

Preuve en exercice

Conséquence : on peut toujours supposer dans les modèles exponentiels que Θ est convexe.

3 Critères de choix, qualités d'un estimateur, complétude

3.1 Qualités

On considère un modèle statistique

$$(*) \quad (\Omega, \mathcal{A}, (\mathbb{P}_\theta, \theta \in \Theta)).$$

(i) Biais d'un estimateur T de $g(\theta) : (\Theta, \mathcal{T}) \rightarrow (\mathbb{R}^d, \mathcal{B})$.

Définition 3.1 On appelle *biais* de l'estimateur T de $g(\theta)$ l'application $\theta \mapsto E_\theta[T - g(\theta)]$. On dit que T est sans biais si $\forall \theta, E_\theta[T] = g(\theta)$.

Exemples : soit $g(\theta) = E_\theta(X)$, alors l'estimateur de g \bar{X}_n est sans biais.

Dans le modèle de Bernoulli, \bar{X}_n est estimateur sans biais de p , mais pour $g(p) = p^2$, \bar{X}_n^2 est de biais $E_p[\bar{X}_n^2] - p^2 = \frac{p(1-p)}{n}$. On dit que \bar{X}_n^2 est **asymptotiquement sans biais**. On peut "débiaiser" \bar{X}_n^2 ; en effet $E_p[\bar{X}_n^2] = p^2 + \frac{p(1-p)}{n} = p^2(1 - 1/n) + \frac{E_p[\bar{X}_n]}{n}$, donc $(\bar{X}_n^2 - \frac{\bar{X}_n}{n}) \frac{n}{n-1}$ est estimateur sans biais de p^2 .

(ii) **risque et admissibilité** Soit L une fonction de perte (rappel : L est convexe en sa deuxième coordonnée) : $g(\Theta) \times g(\Theta) \rightarrow \mathbb{R}^+$ et le risque associé à l'estimateur T de $g(\theta) : R_\theta(T) = E_\theta[L(g(\theta), T)]$.

Définition 3.2 Soient deux estimateurs de $g(\theta)$, T_1 et T_2 . on dit que T_1 est **meilleur** que T_2 si $R_\theta(T_1) \leq R_\theta(T_2)$.

Un estimateur T est dit **admissible** si aucun estimateur n'est meilleur que lui.

Exemple : dans le modèle de Bernoulli, \bar{X}_n et X_1 sont deux estimateurs de p . Soit $L : (t, s) \mapsto (s - t)^2$. Les risques avec cette fonction de perte sont :

$$R_p(\bar{X}_n) = \frac{p(1-p)}{n} ; R_p(X_1) = p(1-p).$$

Donc, \bar{X}_n est meilleur que X_1 .

(iii) **convergences** : on regarde le comportement des estimateurs lorsque la taille de l'échantillon tend vers l'infini. Soit donc $T_n = f(X_1, \dots, X_n)$ un estimateur de $g(\theta)$ associé à un n -échantillon.

Définition 3.3 On dit que la suite d'estimateurs (T_n) **converge en probabilité** si

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}_\theta \{ \|T_n - g(\theta)\| > \varepsilon \} = 0.$$

On dit que la suite d'estimateurs (T_n) **converge dans L^p** si $\lim_{n \rightarrow \infty} \|T_n - g(\theta)\|^p = 0$.

Avec l'inégalité de Bienaymé-Tchebichev, on montre que la convergence dans L^p entraîne la convergence en probabilité.

Lorsque $p = 2$, et avec la fonction de perte $L : (t, s) \mapsto (s - t)^2$, la convergence dans L^2 est la convergence du risque vers 0 : on dit que la suite est **asymptotiquement sans risque**.

Si le biais tend vers 0, on dit que la suite est **asymptotiquement sans biais**.

On peut avoir la situation d'un estimateur biaisé, mais meilleur (au sens du risque) qu'un estimateur sans biais.

Définition 3.4 Une suite d'estimateurs (T_n) est **consistante** si elle converge \mathbb{P}_θ -presque sûrement vers $g(\theta)$.

Exemples : par la loi des grands nombres, sous les bonnes hypothèses, la suite (\bar{X}_n) est consistante pour $g(\theta) = E_\theta(X)$.

Exercice Modèle uniforme : montrer que la suite $(T_n = \sup_i X_i)$ est consistante.

$$\mathbb{P}_\theta\{\theta - T_n > \varepsilon\} = \left(\frac{\theta - \varepsilon}{\theta}\right)^n.$$

Et on applique Borel-Cantelli.

3.2 Complétude

Définition 3.5 La statistique T à valeurs dans (E, \mathcal{E}) est dite **complète** si pour toute fonction mesurable $h : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B})$, $h(T) \in L^1(\mathbb{P}_\theta)$, alors

$$E_\theta(h(T)) = 0, \forall \theta \in \Theta, \Rightarrow h(T) = 0, \mathbb{P}_\theta - p.s.$$

exemples :

(a) $T = a$ constante est complète

(b) Soit un 1-échantillon du modèle exponentiel de densité $\theta e^{-\theta x}$, $\theta > 0$. La statistique X est complète. En effet : soit h telle que pour tout θ : $\int_{\mathbb{R}^+} |h(t)| \exp(-\theta t) dt < \infty$ et $\int_{\mathbb{R}^+} h(t) \exp(-\theta t) dt = 0$. Alors $\int_{\mathbb{R}^+} \theta h^+(t) \exp(-\theta t) dt = \int_{\mathbb{R}^+} \theta h^-(t) \exp(-\theta t) dt$ C'est à dire que les fonctions h^+ et h^- ont même transformée de Laplace, donc elles sont égales, et comme elles sont à supports disjoints, elles sont nulles, et h aussi.

De façon plus générale, on a :

Théorème 3.6 Soit un modèle exponentiel de densité $f_\theta(x) = C(\theta) \exp\langle \theta, T(x) \rangle$, $\theta \in \Theta$. Si Θ contient un ouvert de \mathbb{R}^d , alors la statistique T est complète.

Preuve : on considère h telle que $\int_{\mathbb{R}^d} C(\theta) |h(t)| \exp\langle \theta, t \rangle \mu(dt) < \infty$ et $\int_{\mathbb{R}^d} C(\theta) h(t) \exp\langle \theta, t \rangle \mu(dt) = 0$. Puis la preuve est identique à celle de l'exemple. \square

Corollaire 3.7 *Dans un modèle exponentiel, la statistique naturelle est exhaustive et complète.*

Exemple : la loi géométrique de paramètre $p \in]0, 1[$. La loi du n -échantillon est

$$p^n (1-p)^{\sum_{i=1}^n k_i - n} = \left(\frac{p}{1-p}\right)^n \exp[\log(1-p) \sum_{i=1}^n k_i].$$

le paramètre "naturel" (canonique) est $\theta = \log(1-p) < 0$ et la statistique $S = \sum_{i=1}^n X_i$ est exhaustive et complète.

Théorème 3.8 *Soit une statistique U exhaustive et complète pour θ . Alors*

(i) $E_\theta[T/\sigma(U)]$ est un estimateur de θ meilleur que T .

(ii) $E_\theta[T/\sigma(U)]$ est l'unique estimateur de θ de risque minimum parmi les estimateurs de θ de même biais que T .

(iii) Si $\Theta \subset \mathbb{R}$, si la fonction de perte est $L(s, t) = (s - t)^2$, et si T est un estimateur sans biais de θ dans L^2 , $E_\theta[T/\sigma(U)]$ est l'unique estimateur sans biais de θ de risque minimum parmi les estimateurs sans biais de θ .

Dans le (iii), on dit alors que T est de risque uniformément minimum parmi les estimateurs sans biais (en abrégé **UMVSB**). Cette notion sera vue en détail dans le chapitre suivant.

Preuve : (i) on a vu dans un exercice du chapitre précédent que le risque est diminué quand on conditionne l'estimateur (théorème de RAO-BLACKWELL). $E_\theta[T/\sigma(U)]$ est bien sûr de même biais que T puisque de même espérance.

(ii) Soit S un autre estimateur de même biais que T . D'une part, $R_\theta[E_\theta[S/\sigma(U)]] \leq R_\theta(S)$. D'autre part $E_\theta[E_\theta[S/\sigma(U)] - E_\theta[T/\sigma(U)]] = 0$. mais U est complète : par définition on a donc $E_\theta[S/\sigma(U)] - E_\theta[T/\sigma(U)] = 0$ d'où l'unicité.

(iii) C'est un cas particulier de (ii). \square

Théorème 3.9 (BASU) *Soit une statistique U exhaustive et complète pour θ . Soit S une statistique dont la loi ne dépend pas de θ . Alors U et S sont indépendantes.*

Preuve : Soit f borélienne bornée et $h(U) = E_\theta[f(S)/\sigma(U)]$. On a $m = E_\theta[f(S)] = E_\theta[h(U)]$, $\forall \theta$ puisque la loi de S ne dépend pas de θ . Or U est complète, donc $E_\theta[h(U) - m] = 0$ montre que $h(U) = m$ c'est à dire que U et S sont indépendantes. \square

Application dans le modèle gaussien : \bar{X}_n et $S_n^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X}_n)^2$ sont indépendantes. (preuve en exercice ; c'est une autre preuve d'un résultat vu dans le chapitre sur les vecteurs gaussiens.)

3.3 Application aux modèles exponentiels

Soit un modèle exponentiel de densité $f_\theta(x) = C(\theta) \exp\langle\theta, T(x)\rangle$, $\theta \in \Theta$. On a vu que T est exhaustive et complète, de plus

$$E_\theta(T) = C(\theta) \int T(x) \exp\langle\theta, T(x)\rangle \mu(dx) = g(\theta),$$

Donc T est estimateur sans biais de $g(\theta)$. Si l'on dispose d'un n -échantillon, et si T est de carré intégrable, la statistique \bar{T}_n par la loi des grands nombres converge presque sûrement vers $g(\theta)$. Par le (iii) du théorème 3.8, la statistique \bar{T}_n est l'unique estimateur sans biais de risque minimum.

4 Notion de risque minimum, d'efficacité et vraisemblance

C'est encore une recherche de qualités des estimateurs.

4.1 Notion de risque minimum

On considère toujours un modèle statistique

$$(\Omega, \mathcal{A}, (\mathbb{P}_\theta, \theta \in \Theta)).$$

Définition 4.1 Soit $T : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$. On dit qu'une statistique (un estimateur) est **de risque minimum** pour ce modèle si c'est un estimateur sans biais, et de variance minimum dans l'ensemble des estimateurs sans biais.

Le théorème 3.8 vu dans le chapitre précédent dit que si un estimateur T est sans biais et que U est exhaustive et complète, alors $E[T/U]$ est de risque minimum.

Dans le cas des modèles exponentiels, on prend toujours un estimateur fonction de la statistique naturelle car elle est exhaustive et complète.

Proposition 4.2 Il existe au plus un estimateur de risque minimum.

Preuve : soient X et Y de risque minimums : $\forall T$ sans biais, $V_\theta(T) \geq V_\theta(X) = V_\theta(Y)$. Donc, pour tout $\alpha \in [0, 1]$, $V_\theta(\alpha X + (1 - \alpha)Y) \geq V_\theta(X)$ soit en développant :

$$[\alpha^2 + (1 - \alpha)^2 - 1]V_\theta(X) + 2\alpha(1 - \alpha)cov_\theta(X, Y) \geq 0$$

d'où il vient, puisque $\alpha^2 + (1 - \alpha)^2 - 1 = 2\alpha(1 - \alpha) \geq 0$, $cov_\theta(X, Y) \geq V_\theta(X) = \sqrt{V_\theta(X)V_\theta(Y)}$ ce qui montre par Cauchy-Schwarz que X et Y sont colinéaires ; comme ils ont même variance, ils sont égaux. \square

Proposition 4.3 Soit un modèle statistique $(\Omega, \mathcal{A}, (\mathbb{P}_\theta, \theta \in \Theta))$ avec $\Theta \subset \mathbb{R}$ et T un estimateur sans biais de θ dans $L^2(\Omega)$.

Alors, T est de risque minimum si et seulement si

$$\forall \theta \in \Theta, \forall S \in L^2 \text{ tel que } E_\theta(S) = 0, \text{ alors } E_\theta[TS] = 0.$$

Preuve : (CN) puisque T est de risque minimum, et que $T + \alpha S$ est sans biais, pour tout $\alpha \in [0, 1]$, $V_\theta(T + \alpha S) \geq V_\theta(T)$. Le développement de cette inégalité conduit à une contradiction sauf si $cov_\theta(T, S) = 0$, c'est à dire $E_\theta[TS] = 0$.

Réciproquement (CS), soit T vérifiant la condition à droite, et Y sans biais et de carré intégrable : $E_\theta(Y - T) = 0$, donc $E_\theta[T(Y - T)] = 0$,

$$0 = cov_\theta(T, Y - T), Y = T + (Y - T), V_\theta(Y) = V_\theta(T) + V_\theta(Y - T) \geq V_\theta(T)$$

et T est de risque minimum. \square

Proposition 4.4 Soit un modèle statistique exponentiel

$$(\Omega, \mathcal{A}, (C(\theta)e^{\langle \theta, X \rangle}, \theta \in \Theta))$$

Alors, X est un estimateur efficace de $g(\theta) = -\frac{\nabla C(\theta)}{C(\theta)}$.

Preuve : On dérive par rapport à θ^i l'égalité

$$\frac{1}{c(\theta)} = \int_{\Omega} e^{\langle \theta, X \rangle} d\mathbb{P}(\omega)$$

c'est à dire

$$\frac{-\partial_i c(\theta)}{c^2(\theta)} = \int_{\Omega} X_i e^{\langle \theta, X \rangle} d\mathbb{P}(\omega) = E_{\theta}(X_i)/c(\theta).$$

Donc X est un estimateur sans biais de la fonction de θ annoncée. De plus, on est dans un modèle exponentiel : X est exhaustive et complète : elle est de risque minimum. \square

4.2 Notion d'efficacité

On considère maintenant un modèle dominé tel que $\mathbb{P}_{\theta} = f_{\theta} \cdot \mu$ avec une densité dérivable en θ et de carré intégrable sous \mathbb{P}_{θ} .

Définition 4.5 Alors on appelle **score du modèle** le gradient de $\log f_{\theta}$, $\nabla_{\theta}(\log f_{\theta})$; la matrice de variance du score, $V_{\theta}(\nabla_{\theta} \log f_{\theta})$ s'appelle le **tenseur d'information du modèle** (information de Fisher).

On peut alors montrer :

Théorème 4.6 Rao-Cramer : soit S un estimateur sans biais de $g(\theta)$. Alors pour tout $\theta \in \Theta$,

- (i) $\text{cov}_{\theta}(S, \nabla_{\theta} \log f_{\theta}) = \nabla g(\theta)$,
- (ii) $V_{\theta}(S) \geq \nabla g(\theta) J^{-1}(\theta) \nabla g(\theta)$.

Définition 4.7 Si T estimateur sans biais de $g(\theta)$, $T \in L^2(\mathbb{P}_{\theta})$, réalise le minimum ci-dessus, on dit qu'il est **efficace**.

4.3 Vraisemblance

L'idée est que à θ fixé, la densité du modèle $x \mapsto f_\theta(x)$ mesure une densité d'occupation de x par X , θ étant donné, et pour un n -échantillon, $(x_i) \mapsto \prod_i f_\theta(x_i)$ mesure la probabilité d'observer les valeurs (x_i) sachant que le paramètre est θ . Mais, une fois que les valeurs (x_i) **ont été** observées, a posteriori, $\theta \mapsto \prod_i f_\theta(x_i)$ mesure la vraisemblance que le paramètre soit θ sachant que **l'on a vu** (x_i) .

Définition 4.8 La vraisemblance d'un modèle statistique est l'application sur Θ qui, une fois les observations (x_1, \dots, x_n) d'un n -échantillon faites, est donnée par :

$$\theta \mapsto \prod_{i=1}^n f(x_i, \theta)$$

où $f(\cdot, \theta)$ est la distribution du modèle quand le paramètre est θ .

On appelle **log-vraisemblance** l'application sur Θ qui, une fois les observations du n -échantillon (x_1, \dots, x_n) faites, est donnée par :

$$\theta \mapsto \sum_{i=1}^n \log f(x_i, \theta).$$

4.3.1 Définitions et exemples

Définition 4.9 Un estimateur de maximum de vraisemblance d'un modèle statistique (noté EMV) réalise le maximum sur Θ de la vraisemblance ou de la log-vraisemblance.

Il n'y a pas forcément ni existence, ni unicité.

Exemples

(i) modèle de la loi uniforme sur $[0, \theta]$, $\theta \in \mathbb{R}^+$

(ii) double exponentielle (voir livre de Milhaud, chapitre III, 3.4) et faire le lien avec la médiane : la densité de ce modèle est $f_\theta(x) = \frac{1}{2} \exp(-|x - \theta|)$, $\theta > 0$ sur \mathbb{R} . La vraisemblance d'un n -échantillon est l'application sur \mathbb{R}_*^+ : $\theta \mapsto \exp(-\sum_i |x_i - \theta|)$ et la log-vraisemblance d'un est l'application sur \mathbb{R}_*^+ : $\theta \mapsto -\sum_i |x_i - \theta| = -\sum_{\{x_i \leq \theta\}} (\theta - x_i) - \sum_{\{x_i \geq \theta\}} (x_i - \theta)$.

Cette fonction est négative et l'on peut voir qu'elle est maximum lorsque θ est une médiane de l'ensemble $\{x_i\}$, c'est à dire $\hat{\theta}$ tel que les deux ensembles $\{x_i \leq \theta\}$ et $\{x_i \geq \theta\}$ ont même cardinal.

4.3.2 Cas vectoriel

Les définitions sont les mêmes, mais la recherche d'un maximum est plus délicate. Il ne suffit pas de dériver et d'annuler le gradient pour obtenir un EMV !!!!! Lorsque la fonction

de vraisemblance L est de classe C^2 , on peut rechercher D^2L , sinon, il faut faire l'étude "à la main".....

Exemple corrigé en amphi en détail :

un modèle gaussien de moyenne μ et de variance $a = \sigma^2$.

$$L(x_1, \dots, x_n; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi a}^n} \exp\left[-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{a}\right],$$

$$l(x_1, \dots, x_n; \mu, \sigma^2) = -\frac{1}{2}n \log a - \frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{a}.$$

Le gradient de l est donné par

$$\frac{1}{a} \left(\sum_{i=1}^n x_i - n\mu \right); \quad \frac{-n}{2a} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2a^2}.$$

Ce gradient s'annule pour la statistique

$$\hat{\mu} = \bar{X}; \quad \hat{a} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}.$$

Reste à vérifier sur le tableau de variations que ceci réalise bien un maximum :

. d'une part, la vraisemblance est concave en μ et ceci pour tout a , donc $\hat{\mu}$ est bien EMV pour μ .

. d'autre part si l'on considère l'application $a \mapsto l(x_1, \dots, x_n; \mu, a)$ à μ fixé, elle n'est pas concave mais elle est de dérivée

$$\frac{-na + \sum_{i=1}^n (x_i - \mu)^2}{2a^2}.$$

Un simple tableau de variations donne le résultat à faire ci-dessous :

4.4 Bilan pratique

1. Pour trouver un estimateur, on peut toujours essayer le max de vraisemblance ou encore la méthode des moments.

2. Puis on en cherche les propriétés :

biais ? sans biais ?

erreur quadratique à calculer

recherche de stat exhaustives, complètes

si on peut débiaiser, chercher un estimateur efficace

enfin rechercher le comportement asymptotique (convergence en loi ? en proba ? presque sûre ? dans L^2 ?)

4.5 Cas des modèles exponentiels

La vraisemblance d'un n -échantillon s'écrit de manière particulièrement simple :

$$L(x_1, \dots, x_n; \theta) = (c(\theta))^n \exp[-\langle \theta, \sum_i x_i \rangle],$$

$$l(x_1, \dots, x_n; \theta) = -n \log c(\theta) - \langle \theta, \sum_i x_i \rangle.$$

Lorsque par exemple $\Theta \subset \mathbb{R}$, la dérivée logarithmique est donnée par :

$$\frac{l'}{l}(\theta) = \frac{nc'(\theta)}{c(\theta)} + \sum_i x_i$$

et la solution $\hat{\theta}$ de l'équation $\bar{X} = \frac{-c'(\theta)}{c(\theta)}$ est un EMV dès que (par exemple !) l est concave.

5 Estimation par intervalles

On va utiliser dans ce chapitre le théorème de la limite centrale pour un système statistique $(\Omega, \mathcal{A}, (\mathbb{P}_\theta, \theta \in \Theta))$.

5.1 Estimation par intervalles

cf Jaffard, chapitre XI.3

5.1.1 Généralités

On considère toujours un modèle statistique

$$(\Omega, \mathcal{A}, (\mathbb{P}_\theta, \theta \in \Theta)), \Theta \subset \mathbb{R}.$$

On cherche, au lieu d'un estimateur ponctuel, un intervalle qui donne une mesure de la précision de l'estimation proposée.

Définition 5.1 On appelle "intervalle de confiance" de niveau α une couple de var T_1, T_2 tel que, $\forall \theta \in \Theta$,

$$\mathbb{P}_\theta\{T_1 \leq \theta \leq T_2\} \geq 1 - \alpha.$$

Quelques moyens pratiques pour trouver des intervalles de confiance :

si on connaît la loi de $T - \theta$, on peut trouver une constante C tel que $\mathbb{P}_\theta(T - C \leq \theta \leq T + C) \geq 1 - \alpha$,

si on connaît la loi de $\frac{T - \theta}{Z}$, (par exemple si T est un estimateur de la moyenne et Z un estimateur de l'écart-type pour une famille gaussienne), on cherche K tel que $\mathbb{P}_\theta\{|(T - \theta)/Z| \leq K\} \geq 1 - \alpha$, auquel cas, l'intervalle $[T - KZ, T + KZ]$ est un intervalle de confiance de niveau α pour θ .

si on connaît la loi de $\frac{T}{\theta}$, avec C_1 et C_2 tels que $\mathbb{P}_\theta\{C_1 \leq T/\theta \leq C_2\} \geq 1 - \alpha$, l'intervalle $[T/C_2, T/C_1]$ est un intervalle de confiance de niveau α pour θ .

l'inégalité de B.T., s'il existe p tel que $\|T - \theta\|_p$ est uniformément majoré par M sur Θ , donne

$$\mathbb{P}_\theta\{|T - \theta| \geq K\} \leq (M/K)^p,$$

et si K est tel que $(M/K)^p = \alpha$, on obtient un intervalle de confiance.

Exemple : loi de Bernoulli $\theta \in [0, 1]$. La variance de la statistique \bar{X}_n est $\|\bar{X}_n - \theta\|_2^2 = \frac{\theta(1-\theta)}{n} \leq \frac{1}{4n}$. De là on tire un intervalle de confiance.

5.2 Intervalles de confiance dans le cas des modèles gaussiens

On considère dans ce paragraphe le modèle statistique gaussien réel : $\mathcal{N}(m, \sigma^2)$.

(i) On cherche un intervalle de confiance pour m lorsque σ^2 est connue, alors, on sait que pour un n -échantillon, la statistique \bar{X}_n est un estimateur sans biais de m et que la loi de $\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ est la loi $\mathcal{N}(0, 1)$. D'où l'intervalle de confiance donné par

$$\left[\bar{X}_n - \sigma \frac{1.96}{\sqrt{n}}, \bar{X}_n + \sigma \frac{1.96}{\sqrt{n}} \right]$$

est de niveau 0.05 puisque l'événement

$$\{\omega : \left| \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}(\omega) \right| \leq 1.96\}$$

est de probabilité 0.95.

(ii) On cherche un intervalle de confiance pour m lorsque σ^2 est inconnue, alors, on sait que pour un n -échantillon, la statistique \bar{X}_n est un estimateur sans biais de m et que la loi de $\frac{\sqrt{n}(\bar{X}_n - m)}{\sqrt{Q_n/(n-1)}}$ est la loi de Student à $n - 1$ degrés de liberté (Jaffard page 101 et chapitre de proba sur les lois gaussiennes avec $Q_n = \sum_i (X_i - \bar{X})^2$). D'où l'intervalle de confiance donné par

$$\left[\bar{X}_n - C \frac{\sqrt{Q_n}}{\sqrt{n(n-1)}}, \bar{X}_n + C \frac{\sqrt{Q_n}}{\sqrt{n(n-1)}} \right]$$

est de niveau 0.05 avec C tel que l'événement

$$\{\omega : \left| \frac{\sqrt{n}(\bar{X}_n - m)}{\sqrt{Q_n/(n-1)}}(\omega) \right| \leq 1.96\}$$

soit de probabilité 0.95.

(iii) On cherche un intervalle de confiance pour σ^2 . Alors, on sait que pour un n -échantillon, la statistique $Q_n/(n-1)$ est un estimateur sans biais de σ^2 et que la loi de Q_n/σ^2 est une loi de Khi-deux à $n - 1$ degrés de liberté (Jaffard page 101 et cours de proba). D'où l'intervalle de confiance (par exemple)

$$\left[\frac{Q_n}{C_2}, \frac{Q_n}{C_1} \right]$$

de niveau α avec C_1 et C_2 tels que l'événement

$$\{\omega : C_1 \leq \frac{Q_n}{\sigma^2} \leq C_2\}$$

soit de probabilité $1 - \alpha$;

(exercices à faire en amphi ou TD)

5.3 Intervalles de confiance dans le cas de lois limites

Si n la taille de l'échantillon observé est assez grande pour pouvoir appliquer les lois limites du théorème limite central, on a des possibilités de trouver des intervalles de confiance assez bons pour estimer par intervalle $\theta = E(X)$ ou $\theta = V(X)$.

(i) Si $\theta = E(X)$ et $V(X) = f(\theta)$, on utilise le fait que, pour n assez grand, la loi de $\frac{\sqrt{n}(\bar{X}_n - \theta)}{\sqrt{f(\theta)}}$ est proche de la loi gaussienne centrée réduite.

Exercices :

1. Trouver un intervalle de confiance de 0.95 pour p paramètre d'un n -échantillon d'une loi de Bernoulli .

On est ramené à résoudre, puisque $f(\theta) = \theta(1 - \theta)$, l'inéquation

$$\frac{\sqrt{n}|\bar{X}_n - \theta|}{\sqrt{\theta(1 - \theta)}} \leq 1.96.$$

c'est une inéquation du second degré qui donne l'intervalle

$$\left[\bar{X}_n + \frac{1.96^2}{2n} - \frac{1.96}{\sqrt{n}} \sqrt{\bar{X}_n(1 - \bar{X}_n)}, \bar{X}_n + \frac{1.96^2}{2n} + \frac{1.96}{\sqrt{n}} \sqrt{\bar{X}_n(1 - \bar{X}_n)} \right]$$

Comparer ce résultat avec celui obtenu à l'aide de l'inégalité de B.T.

2. Trouver un intervalle de confiance pour λ paramètre d'un n -échantillon d'une loi de Poisson. Dans ce cas , $E(X) = Var(X) = \lambda$ et la statistique $\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}}$ suit approximativement une loi gaussienne centrée réduite.

$$\mathbb{P}_\lambda \left\{ \left| \sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}} \right| \leq 1.96 \right\} = 0.95$$

et là encore on a une inéquation en λ à résoudre; On obtient l'intervalle

$$\bar{X}_n + \frac{1.96^2}{2n} - 1.96 \sqrt{\frac{\bar{X}_n}{n}}, \bar{X}_n + \frac{1.96^2}{2n} + 1.96 \sqrt{\frac{\bar{X}_n}{n}}$$

(ii) Si l'on cherche à estimer $\theta = V(X)$, on peut montrer que la loi limite pour n assez grand de $T_n = \frac{\sum_i (X_i - \bar{X})^2}{\sigma^2}$ est celle d'un $\chi^2(n - 1)$; on peut en déduire un intervalle de confiance, mais il est assez mauvais lorsque $\frac{E(X^4)}{\sigma^4}$ est trop grand.

Exemples et exercices

5.4 Régression linéaire

Cf. l'exo 3 de la feuille 6, on considère le modèle

$$Y_i = a + bx_i + \varepsilon_i, i = 1, \dots, n$$

avec ε_i variables gaussiennes centrées de variance σ^2 .

On trouve les estimateurs aux moindres carrés de $(a, b) : (A, B)$ par projection sur F engendré par les deux vecteurs $\mathbf{1} = (1, \dots, 1)$ et $x = (X_1, \dots, X_n)$.

$$A\mathbf{1} + Bx = P_F Y.$$

et on obtient par application du chapitre (cours de probabilité) sur les vecteurs gaussiens :

Proposition 5.2 *Les estimateurs*

$$B = \frac{\frac{1}{n} \sum_i x_i Y_i - \bar{x} \bar{Y}}{s_x^2} \text{ et } A = \bar{Y} - B\bar{x} \text{ sont de loi respectivement}$$

$$\mathcal{N}\left(b, \frac{\sigma^2}{ns_x^2}\right) \text{ et } \mathcal{N}\left(a, \frac{\sigma^2}{n}\left(1 + \frac{\bar{x}^2}{s_x^2}\right)\right) \text{ où } \bar{x}_n = \frac{1}{n} \sum_i x_i \text{ et } s_x^2 = \frac{1}{n} \sum_i x_i^2 - (\bar{x})^2.$$

Preuve en exercice

Quand on ne connaît pas la variance σ^2 , pour obtenir des intervalles de confiance, on peut utiliser le théorème suivant.

Théorème 5.3

$$S_n^2 = \frac{1}{n-2} \sum_i (Y_i - A - Bx_i)^2$$

est un estimateur sans biais de σ^2 ; $\frac{n-2}{\sigma^2} S_n^2$ est de loi de $\chi^2(n-2)$, et c'est une var indépendante de A et B .

Corollaire 5.4

$$\frac{B - b}{\sqrt{\frac{S_n^2}{ns_x^2}}} \text{ et } \frac{A - a}{\sqrt{\frac{S_n^2(s_x^2 + (\bar{x})^2)}{ns_x^2}}}$$

suivent des lois de Student $(n-2)$.

6 Généralités sur les tests, test de Neyman-Pearson

On dispose toujours d'un n-échantillon (X_1, \dots, X_n) d'une variable aléatoire X , de dimension 1 ou 2, dont la loi est de type connu ou non, loi à paramètres ou non. Plusieurs types de problèmes peuvent se poser :

. lorsque le **type de la loi est inconnu**, on peut décider de "tester" si la loi est d'un certain type (hypothèse nulle H_o) contre le fait qu'elle n'est pas de ce type (hypothèse alternative H_1),

- ou bien, lorsque le type de la loi est connu et que cette **loi est définie par un paramètre p** appartenant à un ensemble P , on peut décider de "tester" si p appartient à une partie P_o de P (hypothèse nulle H_o) contre le fait que p appartient à P_1 , partie de P disjointe de P_o (hypothèse alternative H_1) ; par exemple, une pièce est-elle équilibrée ou non ? c'est à dire : le paramètre de la loi de Bernoulli est-il égal à $\frac{1}{2}$ ou non ?

- ou bien, si X est de dimension 2, **les deux composantes sont-elle indépendantes ?** ont-elles même moyenne (hypothèse nulle H_o) ou non (hypothèse alternative H_1)?

6.1 Exemples

6.1.1 Le test d'indépendance

sera fait au second semestre, je l'ai laissé ici pour la culture générale.....

On prend l'exemple du tri croisé de deux variables qualitatives, c'est à dire que X prend un nombre fini de valeurs i et Y prend un nombre fini de valeurs j , on note $N_{i,j}$ le nombre de fois que $X = i$ et $Y = j$, $N_{i.}$ le nombre de fois que $X = i$, soit $N_{i.} = \sum_j N_{i,j}$ et $N_{.j}$ le nombre de fois que $Y = j$, soit $N_{.j} = \sum_i N_{i,j}$. On veut tester l'indépendance de X et Y :

(H_o) : X et Y sont indépendantes,

(H_1) : X et Y ne sont pas indépendantes.

On définit la variable aléatoire :

$$\chi^2 = \sum \frac{(N_{ij} - \frac{N_{i.}N_{.j}}{n})^2}{\frac{N_{i.}N_{.j}}{n}}$$

qui suit approximativement, quand H_o est vraie, une loi de khi-deux à $(p-1)(r-1)$ degrés de liberté si p est le nombre de modalités de X et r celui de Y . Par exemple, on trouve pour $p = 3$ et $r = 2$, une estimation de la variable statistique χ^2 valant 9 au vu des valeurs observées n_{ij} de N_{ij} dans le tableau de tri croisé. On lit dans la table de la loi du khi-deux :

$$\mathbb{P}\{\chi^2 > 5.99\} = 0.05$$

Au seuil de 0.05, on doit rejeter l'hypothèse d'indépendance : on conclut H_1 en se trompant avec une probabilité de 0.05. Si dans une autre expérience, on obtient une estimation de χ^2 valant 4, alors on accepte H_o , on accepte l'hypothèse d'indépendance . Mais il est difficile de calculer la probabilité de se tromper : la loi de χ^2 est inconnue si X et Y ne sont pas indépendantes.

6.1.2 Un deuxième exemple

Soit X suivant une loi de Bernoulli ; \bar{X} , la moyenne empirique est un estimateur de l'espérance p . On veut tester $p < \frac{1}{2}$ (H_o) contre $p \geq \frac{1}{2}$ (H_1). Soit $P_o = [0, \frac{1}{2}[$ et $P_1 = [\frac{1}{2}, 1]$. On propose, de façon naturelle, pour région de rejet de (H_o) l'événement :

$$R = \{\bar{X} > b\}.$$

Le calcul de la probabilité de R , quand le paramètre est p , utilise la loi binomiale $B(n, p)$ qui est celle de $n\bar{X}$:

$$\mathbb{P}_p(R) = \mathbb{P}\{n\bar{X} > nb\} = \sum_{k > nb} C_n^k p^k (1-p)^{n-k}$$

et si $p < \frac{1}{2}$, c'est à dire si (H_o) est vraie, rejeter (H_o) quand on est dans la région R conduit à l'erreur :

$$\sup\{\mathbb{P}_p(P); 0 \leq p < 1/2\}$$

que l'on précisera plus tard.

6.1.3 Principe

On peut tirer de ces exemples le principe suivant : au vu des valeurs observées du n-échantillon, on cherche à choisir entre l'hypothèse nulle (H_o) et l'hypothèse alternative (H_1). On donne ici le principe pour tester :

$$H_o = \{p \in P_o\} \text{ contre : } H_1 = \{p \in P_1\}$$

ou bien :

X et Y sont indépendantes, contre : ' X et Y ne sont pas indépendantes

De façon générale, on construit une variable aléatoire (une "statistique") $T(X_1, \dots, X_n)$ fonction de l'échantillon qui, par exemple, "reflète" le paramètre à tester p (voir le chapitre Estimation) ou l'hypothèse H_o comme le χ^2 du premier exemple, et on recherche un ensemble de valeurs C conforme à l'intuition tel que,

- . si $T(X_i) \in C$, alors on rejette H_0
- . si $T(X_i)$ n'appartient pas à C , alors on accepte H_0 .

Lorsque l'échantillon est celui d'une loi paramétrée, pour un modèle statistique $(\Omega, \mathcal{A}, (\mathbb{P}_\theta, \theta \in \Theta))$, on se pose la question (sans nécessairement estimer le paramètre) de savoir, étant donnés deux sous-ensembles disjoints de l'ensemble des paramètres $\Theta : \Theta_0$ et Θ_1 , si $\theta \in \Theta_0$, ou si $\theta \in \Theta_1$. Par exemple, dans le modèle de Bernoulli de paramètre $p \in [0, 1]$, p représentant la proportion de votes pour le oui dans un referendum, les instituts de sondage sont intéressés de savoir si $p < 0.5$ ou $p \geq 0.5$.

Supposer que $\theta \in \Theta_0$ s'appelle l'hypothèse nulle, supposer que $\theta \in \Theta_1$ s'appelle l'hypothèse alternative.

Passons aux définitions précises concernant les tests.

6.2 Définitions

Définition 6.1 *Un test est une application mesurable Φ sur (Ω, \mathcal{A}) à valeurs dans $\{0, 1\}$. L'événement $R = \{\omega : \Phi(\omega) = 1\}$ s'appelle la **région critique**.*

Lorsque l'observation ω du n -échantillon est telle que $\Phi(\omega) = 1$, c'est à dire que l'on est dans la région critique, on conclut que $\theta \in \Theta_1$. Sinon, on conclut que $\theta \in \Theta_0$. Il y a deux manières de se tromper :

- quand on est dans R , on conclut $\theta \in \Theta_1$ alors qu'en fait $\theta \in \Theta_0$,
- quand on est dans R^c , on conclut $\theta \in \Theta_0$ alors qu'en fait $\theta \in \Theta_1$.

Exemple du sondage : avec un 1000-échantillon, supposons choisie la région critique $R = \{\omega : \sum_{i=1}^{1000} X_i > 560\}$, où X_i est la réponse (1 ou 0) du i ème sondé. Si plus de 560 personnes ont dit "oui", on conclut $p \geq 0.5$, sinon, on conclut $p < 0.5$. Dans le premier cas, on s'est trompé si en vrai $p < 0.5$, dans le second cas on s'est trompé si en vrai $p \geq 0.5$.

Puisque Φ est une var, $R \in \mathcal{A}$ et on peut calculer sa probabilité dans les deux cas, donc calculer la probabilité de ces deux types d'erreurs.

Définition 6.2 (i) *risque de première espèce est l'application sur $\Theta_0 : \theta \mapsto \mathbb{P}_\theta(R)$*
(ii) *risque de deuxième espèce est l'application sur $\Theta_1 : \theta \mapsto \mathbb{P}_\theta(R^c)$.*

Un test est "bon" si ces deux erreurs sont petites.

Définition 6.3 *à sauter en première lecture.... On appelle **fonction de perte** L définie sur $\Theta \times \{0, 1\}$ par*

$$L(\theta, 0) = \mathbf{1}_{\Theta_1}(\theta) ; L(\theta, 1) = \mathbf{1}_{\Theta_0}(\theta).$$

la fonction de risque est l'espérance de cette fonction de perte :

$$R : \theta \mapsto E_\theta[L(\theta, \Phi)] = \mathbb{P}_\theta(R)\mathbf{1}_{\Theta_0} + \mathbb{P}_\theta(R^c)\mathbf{1}_{\Theta_1}.$$

Définition 6.4 Le **niveau** d'un test est $\alpha_\Phi = \sup_{\theta \in \Theta_0} E_\theta[\Phi]$.

La fonction **puissance** d'un test est l'application mesurable sur Θ_1 : $\beta_\Phi : \theta \mapsto E_\theta[\Phi]$.

On dit qu'un test est de **seuil** α si α est un majorant de son niveau.

On peut dire maintenant plus précisément qu'un test est bon si on a à la fois un niveau faible et une fonction puissance grande. En général, il faut composer entre ces deux exigences a priori contradictoires. Remarquer que le niveau est le maximum des risques de première espèce.

Exercice : test sur le modèle gaussien $\mathcal{N}(m, 1)$ et les hypothèses $m = 0$ contre $m \neq 0$. On propose le test déterministe de région critique $R = \{|\bar{X}_n| > C\}$. Pour déterminer C , on se donne un "seuil", par exemple 0.05, et on cherche C en sorte que le niveau du test soit ≤ 0.5 . Puis, on calcule la fonction puissance.

Définition 6.5 On dit qu'un test Φ est **uniformément plus puissant** dans l'ensemble des tests de niveau α si tout test Ψ de niveau α vérifie $\beta_\Psi \leq \beta_\Phi$.

Un test Φ est dit **sans biais** si $\alpha_\Phi \leq \beta_\Phi(\theta)$, $\forall \theta \in \Theta_1$.

Ce qui suit est hors programme.

Dans le cas discret, quand on se donne un seuil, il y a peu de chances de trouver une région critique telle que le niveau soit juste égal au seuil. Pour "tomber juste", on s'autorise à prendre comme test, au lieu de $\Phi = \mathbf{1}_R$ l'indicatrice de la région critique, une variable aléatoire Φ à valeurs dans $[0, 1]$ au lieu de $\{0, 1\}$. On peut donc définir plus généralement un test aléatoire, c'est à dire que c'est une application mesurable Φ sur (Ω, \mathcal{A}) à valeurs dans l'intervalle $[0, 1]$. Alors le lemme suivant a un sens.

Lemme 6.6 Si le modèle admet une statistique exhaustive T pour θ , le test $E_\theta(\Phi/T)$ a même niveau et même puissance que Φ .

Preuve : en exercice. De fait, c'est une conséquence immédiate de ce que $E_\theta[E_\theta(\Phi/T)] = E_\theta[\Phi]$. \square

Donc, en conclusion, on prendra plutôt un test Φ fonction d'une statistique exhaustive.

6.3 Test de Neyman-Pearson

(Jersy NEYMAN, Moldavie 1894-Californie 1981 ; Karl PEARSON, Londres 1857-1936, fondateur de la revue Biometrika)

Il s'agit du cas où $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$ pour un modèle dominé.

Théorème 6.7 . Pour tout $\alpha \in]0, 1[$ il existe $K \in \mathbb{R}^+$ et $\gamma \in [0, 1[$ tels que le test

$$(*) \quad \Phi = \mathbf{1}_{\{f_{\theta_1} > K f_{\theta_0}\}} + \gamma \mathbf{1}_{\{f_{\theta_1} = K f_{\theta_0}\}}$$

(i) est de niveau α ,

(ii) vérifie $E_{\theta_0}[\Phi] < E_{\theta_1}[\Phi]$ (donc est sans biais),

(iii) est UPP dans l'ensemble des tests de seuil α ,

(iv) Réciproquement, tout test UPP dans l'ensemble des tests de niveau α est de la forme définie en (*).

Preuve : preuve en amphi uniquement dans le cas où F est surjective (loi continue) : dans ce cas $\gamma = 0$.

(i) on cherche K et γ tels que $\mathbb{P}_{\theta_0}\{f_{\theta_1} > K f_{\theta_0}\} + \gamma \mathbb{P}_{\theta_0}\{f_{\theta_1} = K f_{\theta_0}\} = \alpha$.

La statistique $T = \frac{f_{\theta_1}}{f_{\theta_0}}$ est définie \mathbb{P}_{θ_0} presque sûrement ; on note F sa fonction de répartition sous \mathbb{P}_{θ_0} : $F(x) = \mathbb{P}_{\theta_0}(T \leq x)$. On est ramené au problème

$$1 - F(K) + \gamma(F(K) - F(K^-)) = \alpha.$$

Si'il existe K tel que $F(K) = 1 - \alpha, \gamma = 0$. Sinon, K tel que $F(K^-) < 1 - \alpha < F(K)$, $\gamma = \frac{F(K) - 1 + \alpha}{F(K) - F(K^-)}$ et le test est entièrement défini.

(ii) et (iii) On montre d'abord que si un autre test Ψ est de seuil α (donc il vérifie $E_{\theta_0}(\Psi) \leq \alpha = E_{\theta_0}(\Phi)$) alors $E_{\theta_1}(\Psi) \leq E_{\theta_1}(\Phi)$. En effet, l'hypothèse est $\int (\Phi - \Psi) f_{\theta_0} d\mu \geq 0$.

On remarque que

sur l'ensemble $\{f_{\theta_1} > K f_{\theta_0}\}$, $\Phi = 1$ donc $\Phi - \Psi \geq 0$

sur l'ensemble $\{f_{\theta_1} < K f_{\theta_0}\}$, $\Phi = 0$ donc $\Phi - \Psi \leq 0$

et globalement $(\Phi - \Psi)(f_{\theta_1} - K f_{\theta_0}) \geq 0$, et donc aussi $\int (\Phi - \Psi)(f_{\theta_1} - K f_{\theta_0}) d\mu \geq 0$. C'est à dire que

$$E_{\theta_1}(\Phi) - E_{\theta_1}(\Psi) \geq K(E_{\theta_0}(\Phi) - E_{\theta_0}(\Psi)) \geq 0.$$

On a bien $E_{\theta_1}(\Psi) \leq E_{\theta_1}(\Phi)$.

Pour (ii), on prend Ψ constant égal à α , d'où le résultat.

Pour (iii), si Ψ est de seuil α , $E_{\theta_0}(\Psi) \leq \alpha$ et comme on a toujours $\int (\Phi - \Psi)(f_{\theta_1} - K f_{\theta_0}) d\mu \geq 0$, il vient encore

$$E_{\theta_1}(\Phi) - E_{\theta_1}(\Psi) \geq K(E_{\theta_0}(\Phi) - E_{\theta_0}(\Psi)) \geq 0.$$

Donc, $E_{\theta_1}(\Phi) - E_{\theta_1}(\Psi) \geq 0$ d'où la propriété UPP.

(iv) Réciproquement soit un test Ψ UPP de niveau α et le test Φ de niveau α construit en (i) : d'abord $\alpha_\Phi = \alpha_\Psi$. Par (iii), Φ est UPP donc $E_{\theta_1}(\Phi) \geq E_{\theta_1}(\Psi)$. Mais Ψ l'est aussi : $E_{\theta_1}(\Phi) \geq E_{\theta_1}(\Psi)$ et ainsi $E_{\theta_1}(\Phi) = E_{\theta_1}(\Psi)$. Il vient alors par combinaison linéaire :

$$\int (\Phi - \Psi)(f_{\theta_1} - K f_{\theta_0}) d\mu = 0.$$

Comme l'intégrand est positif, si l'intégrale est nulle, c'est qu'il est nul, soit

$$\Phi = \Psi \text{ sur l'ensemble où } f_{\theta_1} \neq K f_{\theta_0}$$

et Ψ est de la même forme que Φ . □

7 Tests lorsque $\Theta \subset \mathbb{R}$

cf. Jaffard, chapitre XIII.

On va traiter deux sortes de tests :

$$H_0 : \{\theta \leq \theta_0\} ; H_1 : \{\theta > \theta_0\},$$

qu'on appelle **test unilatère** ;

$$H_0 : \{\theta \in I\} ; H_1 : \{\theta \in I^c\}, \quad I \text{ intervalle de } \mathbb{R},$$

qu'on appelle **test bilatère** (hypothèse simple contre hypothèse composite lorsque I est un singleton), et ceci dans le cadre des modèles à rapport de vraisemblance monotone et lorsque la statistique T est de loi continue. On donnera quelques indications pour le mode opératoire dans le cas de lois discrètes.

7.1 Modèle à rapport de vraisemblance monotone

Définition 7.1 *Ce sont des modèles dominés, tels qu'il existe une statistique T telle que la probabilité sur (Ω, \mathcal{A}) est de la forme $\mathbb{P}_\theta = f_\theta(T) \cdot \mu$, et si $\theta < \theta'$, le rapport $\frac{f_{\theta'}}{f_\theta}(T)$ est une fonction croissante de T .*

Exemples :

(i) la loi gaussienne $\mathcal{N}(m, 1)$ et $\theta_1 > \theta_2$:

$$\frac{f_{\theta_1}}{f_{\theta_2}}(T) = \exp\left(-\frac{1}{2} \sum_i [(x_i - \theta_1)^2 - (x_i - \theta_2)^2]\right) = \exp\left[-\frac{1}{2}(\theta_2 - \theta_1) \sum_i (2x_i - \theta_1 - \theta_2)\right] = \exp(\theta_1 - \theta_2)(n\bar{x} - \frac{1}{2}(\theta_1 + \theta_2)),$$
 ce qui est une application croissante par rapport à $T = \bar{x}$.

(ii) Plus généralement, tout modèle exponentiel : $f_\theta = e^{-C(\theta)} e^{T \cdot Q(\theta)}$ lorsque la fonction Q est croissante.

Application du lemme de Neyman-Pearson dans ce cas : la région critique

$$\{f_{\theta_1}(T) > K f_{\theta_0}(T)\} \text{ est } \{T > C\} \text{ puisque par la monotonie } \frac{f_{\theta_1}(T)}{f_{\theta_0}(T)} > K \text{ équivaut à } T > C.$$

7.2 Tests unilatères

Théorème 7.2 *Soit un modèle statistique $(\Omega, \mathcal{A}, (\mathbb{P}_\theta, \theta \in \Theta))$ à rapport de vraisemblance monotone pour la statistique réelle T . Alors, $\forall \alpha \in]0, 1[$, $\exists C \in \mathbb{R}$, $\exists \gamma \in [0, 1[$, tels que le test*

$$\Phi = \mathbf{1}_{\{T > C\}} + \gamma \mathbf{1}_{\{T = C\}}$$

est UPP parmi les tests de niveau α de Θ_0 contre Θ_1 .

De plus, la fonction $\theta \mapsto E_\theta[\Phi]$ est croissante et $\alpha = E_{\theta_0}[\Phi]$. Ainsi ce test est-il sans biais.

Remarque : lorsque la loi est continue, comme pour le test de Neyman-Pearson, $\gamma = 0$.

Preuve dans le cas continu : dans ce cas la fonction de répartition F_θ de la statistique T est strictement croissante de 0 à 1 pour tout θ donc inversible et on choisit $\gamma = 0$ et $C = F_{\theta_0}^{-1}(1 - \alpha)$.

(i) On procède comme dans la preuve du lemme de Neyman-Pearson. Soit $\theta < \theta'$. Comme la fonction $\frac{f_{\theta'}}{f_\theta}$ est croissante, il existe $K = \frac{f_{\theta'}}{f_\theta}(C)$ tel que $T > C$ équivaut à $\frac{f_{\theta'}}{f_\theta}(T) > K$. C'est à dire que le test Φ est un test de Neyman-Pearson de θ contre θ' . On a montré qu'alors $E_\theta(\Phi) \leq E_{\theta'}(\Phi)$ et on a bien la croissance de l'application $\theta \mapsto E_\theta[\Phi]$ en même temps que le fait que $\alpha = E_{\theta_0}(\Phi) = \sup_{\theta \leq \theta_0} E_\theta(\Phi)$: ce test est bien de niveau α et sans biais.

(ii) Soit un autre test Ψ de niveau α : $\sup_{\theta \leq \theta_0} E_\theta(\Psi) = \alpha$. En particulier, $E_{\theta_0}(\Psi) \leq \alpha$. Le test Φ est un test de Neyman-pearson de θ_0 contre $\theta', \forall \theta' > \theta_0$. Donc il est UPP et $\forall \theta' > \theta_0, E_{\theta'}(\Phi) \geq E_{\theta'}(\Psi)$. \square

Exercice : on dispose d'un n -échantillon de la loi gaussienne $\mathcal{N}(m, 1)$ et on construit un test UPP de niveau 0.05 de l'hypothèse $m \leq m_0$ contre $m > m_0$. La région critique est $\{\bar{X}_n > m_0 + \frac{1.645}{\sqrt{n}}\}$. Tracer la courbe de la fonction puissance $\beta : m \mapsto 1 - F(1.645 - \sqrt{n}(m - m_0))$ et étudier cette courbe quand n varie. Remarquer la valeur de la pente de cette fonction au point m_0 en fonction de n .

7.3 Tests d'une hypothèse simple contre une hypothèse composite

On est toujours dans le cas de modèles à vraisemblance monotone, et plus précisément de modèles exponentiels où la fonction Q est croissante. En effet, dans ce cas, $\theta_0 \leq \theta_1$ équivaut à $Q(\theta_0) \leq Q(\theta_1)$ et le problème revient après un changement de paramètre au modèle avec $f_\theta(T) = C(\theta)e^{\theta T}$, avec Θ une partie de \mathbb{R}^d . Il s'agit de tester $\Theta_0 = \{\theta_0\}$ contre $\Theta_1 = \Theta - \Theta_0$.

Rappel : On dit qu'un test Φ de niveau α est **sans biais** si $\forall \theta \in \Theta_1, E_\theta[\Phi] \geq \alpha$.

Théorème 7.3 Soit un modèle statistique exponentiel où $f_\theta(T) = C(\theta)e^{\theta T}$ avec Θ un intervalle de \mathbb{R} . Soit $\Theta_0 = \{\theta_0\}$. Alors $\forall \alpha \in]0, 1[$, il existe a et b tels que

$$(*) \quad \mathbb{P}_{\theta_0}\{T \in [a, b]\} = 1 - \alpha \text{ et } \frac{d}{d\theta}\mathbb{P}_\theta\{T \in [a, b]\}(\theta_0) = 0.$$

(i) Le test $\Phi = \mathbf{1}_{\{T > b\}} + \mathbf{1}_{\{T < a\}}$ est alors un test sans biais, de niveau α de l'hypothèse $\theta \in \Theta_0$ contre l'hypothèse $\theta \in \Theta_0^c$.

(ii) De plus, il est UPP parmi les tests sans biais de niveau α vérifiant (*).

Preuve : on admet l'existence. Mais on peut analyser la condition (*) dans le cas de modèles réguliers : l'application $\theta \mapsto \mathbb{P}_\theta\{T \in [a, b]\} = C(\theta) \int_a^b e^{\theta x} \mu(dx)$ vaut $1 - \alpha$ en θ_0 , elle est de dérivée $C'(\theta) \int_a^b e^{\theta x} \mu(dx) + C(\theta) \int_a^b x e^{\theta x} \mu(dx)$ qui est nulle en θ_0 . Cette deuxième condition montre que $\frac{\int_a^b x e^{\theta_0 x} \mu(dx)}{\int_a^b e^{\theta_0 x} \mu(dx)} = -\frac{C'(\theta_0)}{C(\theta_0)}$. On peut montrer, mais c'est assez délicat, que l'application $\theta \mapsto C(\theta) \int_a^b e^{\theta x} \mu(dx)$ est convexe et donc admet un maximum en θ_0 : c'est à dire que le test proposé est bien de niveau α et sans biais. Cette première partie du théorème se montre en revanche sans problème sur les exemples les plus courants de modèles exponentiels comme on le verra dans le chapitre suivant consacré aux modèles gaussiens.

(ii) Soit un autre test Ψ de niveau α et sans biais vérifiant (*) et soit par ailleurs λ et μ solution du système

$$C(\theta)e^{\theta x} - \lambda C(\theta_0)e^{\theta_0 x} - \mu \frac{C^2}{C'}(\theta_0) x e^{\theta_0 x} = 0$$

pour $x = a$ et b . L'existence de ces deux constantes est facile à montrer, en particulier, puisque $a < b$, le déterminant de ce système est positif, et μ est du signe de $e^{\theta_0 a} e^{\theta b} - e^{\theta_0 b} e^{\theta a}$, qui est du signe de $\theta - \theta_0$: $\mu(\theta - \theta_0) > 0$.

On pose $G(x) = C(\theta) - \lambda C(\theta_0)e^{(\theta_0 - \theta)x} - \mu \frac{C^2}{C'}(\theta_0) x e^{(\theta_0 - \theta)x}$. Par définition de λ et μ , $G(a) = G(b) = 0$, et la dérivée de G , $G'(x)$ est du signe de $\mu(\theta - \theta_0) \frac{C^2}{C'}(\theta_0) x - (\theta_0 - \theta) \lambda C(\theta_0) - \mu \frac{C^2}{C'}(\theta_0)$ qui est croissante puisque $\mu(\theta - \theta_0) > 0$. Ainsi, G' est négative puis positive, donc G est décroissante puis croissante : $G(x) \leq 0$ entre a et b , et positive à l'extérieur.

On remarque que pour tout fonction intégrable f :

$$\frac{d}{d\theta} E_\theta(f) = C'(\theta) \int f(x) e^{\theta x} \mu(dx) + C(\theta) \int f(x) x e^{\theta x} \mu(dx),$$

d'où

$$\int \frac{C^2}{C'}(\theta_0) (\Phi - \Psi)(x) x e^{(\theta_0 - \theta)x} = \frac{C}{C'}(\theta_0) \left[\frac{d}{d\theta} E_{\theta_0}(\Phi - \Psi) \right] - \frac{C'}{C}(\theta_0) E_{\theta_0}(\Phi - \Psi).$$

On introduit alors

$$\int e^{\theta x} G(x) [\Phi(x) - \Psi(x)] \mu(dx) = E_\theta(\Phi - \Psi) - \lambda E_{\theta_0}(\Phi - \Psi) - \mu \frac{C}{C'}(\theta_0) \left[\frac{d}{d\theta} E_{\theta_0}(\Phi - \Psi) \right] - \frac{C'}{C}(\theta_0) E_{\theta_0}(\Phi - \Psi).$$

L'intégrand est positif car si $x \in [a, b]$, $\Phi - \Psi = -\Psi \leq 0$, et sinon, $\Phi - \Psi = 1 - \Psi \geq 0$. Par ailleurs, par hypothèse (*), $E_{\theta_0}(\Phi - \Psi) = \frac{d}{d\theta} E_{\theta}(\Phi - \Psi)(\theta_0) = 0$, et il reste $E_{\theta}(\Phi - \Psi) \geq 0$, c'est à dire que Φ est UPP. \square

Exercice : montrer que (*) est possible dans le cas du modèle gaussien de moyenne θ et de variance 1. Donner le test sans biais de niveau α UPP parmi les tests sans biais de niveau α UPP de l'hypothèse $\theta = \theta_0$ contre l'alternative.

8 Tests sur les paramètres de lois gaussiennes

On considère un n -échantillon (X_1, \dots, X_n) du modèle gaussien

$$(\Omega, \mathcal{A}, (\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}, m \in \mathbb{R}, \sigma^2 > 0)).$$

Ici, l'ensemble des paramètres $\Theta = \mathbb{R} \times \mathbb{R}_*^+$ est à deux dimensions. On rappelle qu'il s'agit d'un modèle exponentiel dont le couple (\overline{X}_n, Q_n) est la statistique naturelle exhaustive, (cf. cours de probabilités, chapitre sur les lois associées aux vecteurs gaussiens) avec

$$\overline{X}_n = \frac{1}{n} \sum_i X_i, \quad Q_n = \sum_i (X_i - \overline{X}_n)^2.$$

8.1 Tests sur le paramètre σ^2 .

(i) Test unilatère.

Théorème 8.1 *Le test de région critique $\{Q_n > C\sigma_0^2\}$, où C est tel que $\mathbb{P}\{\chi_{n-1}^2 > C\} = \alpha$, est un test de niveau α de l'hypothèse $\sigma^2 \leq \sigma_0^2$ contre l'hypothèse $\sigma^2 > \sigma_0^2$, uniformément plus puissant parmi les tests de cette hypothèse de niveau α . De plus, la fonction puissance $\sigma^2 \mapsto \mathbb{P}_{\sigma^2}\{Q_n > C\sigma_0^2\}$ est croissante.*

Preuve : c'est une simple conséquence du fait qu'il s'agit d'un modèle exponentiel à rapport de vraisemblance monotone, puisque la statistique $\frac{Q_n}{\sigma^2}$ suit une loi de chi-deux, soit une loi Gamma de paramètres $((n-1)/2, \frac{1}{2})$. La loi de Q_n est donnée par :

$$\mathbb{P}\{Q_n \leq x\} = C_n \int_0^{x/\sigma^2} t^{\frac{n-3}{2}} e^{-\frac{t}{2}} dt = C_n \int_0^x (t/\sigma^2)^{\frac{n-3}{2}} e^{-\frac{t}{2\sigma^2}} \frac{1}{\sigma^2} dt,$$

et sa densité est donc $f_{\sigma^2}(t) = C_n (t/\sigma^2)^{\frac{n-3}{2}} e^{-\frac{t}{2\sigma^2}} \frac{1}{\sigma^2}$. Le rapport $\frac{f_{\sigma^2}}{f_{\tau^2}}$ est donc croissant lorsque $\sigma > \tau$. Le théorème 7.2 permet de conclure.

On peut ici expliciter la fonction puissance :

$$\sigma^2 \mapsto \mathbb{P}_{\sigma^2}\{Q_n > C\sigma_0^2\} = C_n \int_{\frac{C\sigma_0^2}{\sigma^2}}^{\infty} t^{\frac{n-3}{2}} e^{-\frac{t}{2}} dt$$

qui est bien une fonction croissante en σ^2 .

(ii) Hypothèse simple contre hypothèse composite

Théorème 8.2 *Le test de région critique $\{Q_n < C\sigma_0^2 \text{ ou } > C'\sigma_0^2\}$, où C et C' sont tels que $\mathbb{P}\{c \leq \chi_{n-1}^2 \leq C'\} = 1 - \alpha$ et $C^{\frac{n-1}{2}}e^{-\frac{C}{2}} = C'^{\frac{n-1}{2}}e^{-\frac{C'}{2}}$ est un test sans biais de niveau α de l'hypothèse $\sigma^2 = \sigma_0^2$ contre l'hypothèse $\sigma^2 \neq \sigma_0^2$, uniformément plus puissant parmi les tests de cette hypothèse sans biais de niveau α vérifiant que la dérivée de la fonction puissance est nulle en σ_0^2 .*

Preuve : c'est une conséquence du paragraphe 7.3. On cherche donc C et C' tels que

$$C_n \int_C^{C'} t^{\frac{n-3}{2}} e^{-\frac{t}{2}} dt = 1 - \alpha.$$

De plus, on a à étudier l'application

$$a \mapsto \mathbb{P}_a\{C\sigma_0^2 \leq Q_n \leq C'\sigma_0^2\} = C_n \int_{\frac{C\sigma_0^2}{a}}^{\frac{C'\sigma_0^2}{a}} t^{\frac{n-3}{2}} e^{-\frac{t}{2}} dt$$

qui admet pour dérivée

$$a \mapsto \left(\frac{C\sigma_0^2}{a}\right)^{\frac{n-1}{2}} e^{-\frac{C\sigma_0^2}{2a}} - \left(\frac{C'\sigma_0^2}{a}\right)^{\frac{n-1}{2}} e^{-\frac{C'\sigma_0^2}{2a}}$$

Cette fonction s'annule en $a = \sigma_0^2$ si et seulement si $C'^{\frac{n-1}{2}}e^{-\frac{C'}{2}} = C^{\frac{n-1}{2}}e^{-\frac{C}{2}}$. On peut donc appliquer le théorème du paragraphe 7.3 ce qui achève la preuve. \square

8.2 Tests sur le paramètre m .

(i) Test unilatère.

Théorème 8.3 *Le test de région critique $\{\frac{\sqrt{n}(\bar{X}_n - m_0)}{\sqrt{Q_n/(n-1)}} > C\}$, où C est tel que $\mathbb{P}\{Student_{n-1} > C\} = \alpha$, est un test de niveau α de l'hypothèse $m \leq m_0$ contre l'hypothèse $m > m_0$, uniformément plus puissant parmi les tests de cette hypothèse de niveau α . De plus, la fonction puissance $m \mapsto \mathbb{P}_m\{\frac{\sqrt{n}(\bar{X}_n - m_0)}{\sqrt{Q_n/(n-1)}} > C\}$ est croissante.*

Preuve : c'est une simple conséquence du fait qu'il s'agit d'un modèle exponentiel à rapport de vraisemblance monotone, En effet, le rapport de vraisemblance est

$$\frac{f_m}{f_\mu}(x_1, \dots, x_n) = \exp\left[\frac{1}{2\sigma^2}(\mu - m)(2\bar{x}_n - m - \mu)\right]$$

qui est croissant en \bar{x}_n dès que $\mu > m$. On utilise la statistique $T_0 = \frac{\sqrt{n}(\bar{X}_n - m_0)}{\sqrt{Q_n/(n-1)}}$ qui suit, lorsque le paramètre $m = m_0$, une loi de Student à $n - 1$ degrés de liberté (cf chapitre de probabilité sur les lois associées aux vecteurs gaussiens). On peut alors appliquer le théorème 7.2

(ii) Hypothèse simple contre hypothèse composite

Théorème 8.4 *Le test de région critique $\{|\frac{\sqrt{n}(\overline{X}_n - m_0)}{\sqrt{Q_n/(n-1)}}| > C\}$, où C est tel que $\mathbb{P}\{|Student_{n-1}| > C\} = \alpha$, est un test sans biais de niveau α de l'hypothèse $m = m_0$ contre l'hypothèse $m \neq m_0$, uniformément plus puissant parmi les tests de cette hypothèse sans biais de niveau α vérifiant que la dérivée de la fonction puissance est nulle en m_0 .*

Preuve : c'est une conséquence du paragraphe 7.3 puisque l'on est toujours dans le cadre d'un modèle exponentiel à rapport de vraisemblance monotone. On montre l'existence de C tel que défini dans l'énoncé et vérifiant de plus que l'application

$$m \mapsto \mathbb{P}_m \left\{ \left| \frac{\sqrt{n}(\overline{X}_n - m_0)}{\sqrt{Q_n/(n-1)}} \right| > C \right\}$$

est minimum en m_0 . Soit $G(m) = \mathbb{P}_m \left\{ \left| \frac{\sqrt{n}(\overline{X}_n - m_0)}{\sqrt{Q_n/(n-1)}} \right| \leq C \right\}$. Les statistiques $\overline{X}_n - m_0$ et Q_n sont indépendantes et la loi du couple est $\mathcal{N}(m - m_0, \frac{\sigma^2}{n}) \otimes \chi_{n-1}^2$. On peut donc calculer $G(m)$ par une intégrale double :

$$G(m) = \int \int_{\sqrt{n-1}|x| \leq C \sqrt{v/n}} \sqrt{\frac{n}{2\pi\sigma^2}} e^{-\frac{(x-m+m_0)^2}{2\sigma^2}} v^{\frac{n-3}{2}} e^{-\frac{1}{2}v} dv dx.$$

$$G'(m) = \int \int_{\sqrt{n-1}|x| \leq C \sqrt{v/n}} \sqrt{\frac{n}{2\pi\sigma^2}} \frac{x - m + m_0}{\sigma^2} e^{-\frac{(x-m+m_0)^2}{2\sigma^2}} v^{\frac{n-3}{2}} e^{-\frac{1}{2}v} dv dx.$$

Le domaine étant symétrique, $G'(m_0) = 0$ et les hypothèses du théorème 7.3 sont vérifiées et l'on peut conclure.

En revanche, la fonction puissance n'est pas facile à expliciter.

8.3 Tests du chi-deux

- (i) indépendance
 - (ii) adéquation
- plus test d'adéquation de Kolmogorov ??

References

- [1] Ph. BARBE et M. LEDOUX : “Probabilités, de la licence à l’agrégation”, Belin, Paris, 1998.
- [2] M. COTTRELL, V. GENON-CATALOT, C. DUHAMEL, T. MEYRE : “Exercices de probabilités”, Cassini, Paris, 1999.
- [3] D. DACUNHA-CASTELLE et M. DUFLO : “Probabilités et statistique”, Masson, Paris, 1982 ; et le livre d’exercices qui va avec.
- [4] P. JAFFARD : “Initiation aux méthodes de la statistique et du calcul des probabilités”, Masson, Paris, 1978
- [5] P. JAFFARD : “Initiation aux méthodes de la statistique et du calcul des probabilités”, Masson, Paris, 1978 (réédité en 1996).
- [6] J. NEVEU : “Bases mathématiques du calcul des probabilités”, Masson, Paris, 1970.
- [7] J. NEVEU : “Cours de probabilités à l’école Polytechnique”.
- [8] G. SAPORTA : “Probabilités, analyse des données et statistique”, ed. Technip, Paris, 1990.