Université René Descartes
U.F.R. de Mathématiques et Informatique

Cours de Statistique Fondamentale

MST2
1999-2000

C. Graffigne

Contents

Chapter 1

Décision statistique

Exemples :

1. Durée de vie moyenne d'une ampoule —– Estimation.
2. Efficacité d'un médicament —– Test.
3. Encadrement de la durée de vie d'une ampoule —– Intervalle de confiance.

Mais on peut aussi vouloir utiliser des décisions plus complexes, par exemple classer k marques d'un même produit.

1.1 La décision statistique

Un statisticien observe un phénomène dont la loi dépend d'un paramètre $\theta \in \Theta$, afin d'entreprendre une action$a \in A$ dont le coût $C(\theta, a)$, supposé positif, représente la "perte" du statisticien.

Avant de se décider, la statisticien observe un vecteur aléatoire X de loi P_θ. Sa décision, fonction du résultat de l'expérience, sera une fonction δ mesurable de $\mathbb{R}^n$ dans A, ensemble des actions supposé muni d'une tribu $\mathcal{T}$.

Utiliser la décision δ signifie que si $X = x$, $\delta(x)$ est la décision prise. Le coût $C(\theta, \delta)$ est alors une variable aléatoire. Si le statisticien choisit la règle de décision (ou la stratégie) δ, son risque moyen (ou coût moyen) sera, lorsque le paramètre est θ :

$$R(\theta, \delta) = \int_{\mathbb{R}^n} C(\theta, \delta(x)) dP_\theta(x) = E_\theta(C(\theta, \delta(X)))$$

1.1.1 Le cas de l'estimation

Pour estimer θ , l'ensemble des actions est Θ . Plus généralement, si g est une fonction de Θ dans U ouvert de $\mathbb{R}^k$, estimer $g(\theta)$ consiste à prendre $A = U$ et à attribuer une valeur a à $g(\theta)$.

Une décision est ici appelée estimateur, son coût sera le plus souvent égal à $C(\theta, a) = ||g(\theta) - a||^2$, ou, pour un estimateur $T(X)$, $C(g(\theta), T(X)) = J(g(\theta) - T(X))$ où J est une fonction convexe, nulle en 0.

1.1.2 Les tests

Soit (Θ_0, Θ_1) une partition de Θ , supposons que l'on veuille décider du sous-ensembles dans lequel se trouve θ . C'est un problème de choix entre deux hypothèses : $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$. On prend alors $A = \{0, 1\}$, le coût de l'erreur $C(\theta, i)$ est nul pour $\theta \in \Theta_i$ et égal à 1 pour $\theta \notin \Theta_i$. On peut aussi prendre plus généralement $A = [0, 1]$ afin de permettre des décisions aléatoires et des coûts variables.

En général $\delta = \mathbf{1}_D$ où $D \in \mathcal{A} = \mathcal{P}(\Theta)$, ensemble des parties de Θ , est la région de rejet du test. Le risque du test est alors :

$$R(\theta,\delta) = C(\theta,0)P(D^c) + C(\theta,1)P(D) = \begin{cases} P(D) & \text{si } \theta \in \Theta_0 \\ P(D^c) & \text{si } \theta \in \Theta_1 \end{cases}$$

$R(\theta,\delta)$ est donc égal au niveau ou au risque de seconde espèce suivant les cas.

Remarque : Si $\mathcal{A} = \{a_1, \ldots, a_k\}$, on parle de problème à décision multiple.

1.1.3 Régions de confiance

La décision correspondante est un fonction δ prenant ses valeurs dans $A = \mathcal{P}(\Theta)$, ensemble des parties de Θ . le coût de la décision δ, si le paramètre est θ , est $C(\theta,\delta) = \mathbf{1}_{\delta^c}(\theta) = \mathbf{1}_{\theta\notin\delta}$. Le risque associé est alors :

$$R(\theta,\delta) = \int_{\mathbb{R}^n} \mathbf{1}_{\theta\notin\delta(x)} dP_\theta(x) = P_\theta(\theta \notin \delta) = P_\theta(\{x|\theta \notin \delta(x)\})$$

1.1.4 Décision randomisée

Définition 1.1 : *Une loi de probabilité δ sur l'ensemble des fonctions mesurables est une règle de décision randomisée. On définit $R(\theta,\delta) = E(R(\theta,Z))$ où $Z \sim \delta$ et Z est à valeurs dans D.*

1.2 Décisions minimax et bayésienne

1.2.1 Comparaison de décisions

Une décision δ est dite **aussi bonne que** la décision δ' si pour tout θ , $R(\theta,\delta)$ est inférieur ou égal à $R(\theta,\delta')$, elle est **meilleure** que δ' si l'inégalité est stricte pour au moins un θ .

Une décision δ est **optimale** si elle est aussi bonne que toutes les autres, elle est **admissible** s'il n'existe aucune décision meilleure.

Remarque : $\forall\theta \in \Theta, R(\theta,\delta) = R(\theta,\delta') \not\Longrightarrow \delta = \delta'$.

1.2.2 Stratégie du minimax

On veut choisir δ à partir de la quantité $R(\theta,\delta)$. La stratégie du minimax consiste à minimiser en δ

$$R(\delta) = \sup_\theta R(\theta,\delta).$$

Remarque : Tous les θ sont traités de la même façon.

Proposition 1.1 : *Une décision admissible et de risque constant est minimax.*

Dém : par l'absurde. Si δ est admissible et de risque constant, supposons que δ ne soit pas minimax, alors

$$\exists\delta', sup_{\theta\in\Theta}R(\theta,\delta') < sup_{\theta\in\Theta}R(\theta,\delta) = R(\delta) \Longrightarrow \forall\theta \in \Theta, R(\theta,\delta') < R(\delta) = R(\theta,\delta)$$

donc δ ne serait pas admissible. □

1.2.3 Stratégie bayésienne

Lors d'une approche bayésienne, on considère que θ est une fonction aléatoire $\tilde{\theta}$ de loi ν (appelée loi a priori).

Nous observons $\tilde{X}$, vecteur aléatoire, et supposons que $(\tilde{\theta}, \tilde{X})$ suit une loi telle que :

$$\begin{cases} \tilde{\theta} \text{ suit } \nu, \theta \text{ à valeurs dans } \Theta \\ \mathcal{L}(\tilde{X}|\tilde{\theta} = \theta) = P_\theta \end{cases}$$

Le risque est alors défini de la manière suivante :

$$R(\theta, \delta) = E[C(\theta, \delta(\tilde{X}))|\tilde{\theta} = \theta]$$

On appelle risque bayésien le nombre

$$r_\nu(\delta) = E[R(\tilde{\theta}, \delta)] = \int_\Theta R(\theta, \delta) d\nu(\theta) = \int_\Theta \left(\int_\Omega C(\theta, \delta(x)) dP_\theta(x) \right) d\nu(\theta)$$

Définition 1.2 : *δ est une décision bayésienne si elle minimise $r_\nu(.)$.*

Remarques :

- Après inversion du conditionnement :

$$r_\nu(\delta) = \int_\Omega \left(\int_\Theta C(\theta, \delta(x)) d\nu_{\tilde{\theta}|\tilde{X}=x}(\theta) \right) dP_{\tilde{X}}(x)$$

- Si δ minimise $\int_\Theta C(\theta, \delta(x)) d\nu_{\tilde{\theta}|\tilde{X}=x}(\theta)$ pour tout x, alors a fortiori, $\delta(X)$ est bayésien.

- si $C(x, y) = (x - y)^2$ et $g : \Theta \longrightarrow \mathbb{R}^k$ alors $T = h(X)$ est bayésien ssi $T = \int_\Theta g(\theta) d\nu_{\tilde{\theta}|\tilde{X}=x}(\theta)$. Si le modèle est dominé :
$$T = \int_\Theta g(\theta) \frac{f_\theta(X)}{\int_\Theta f_{\theta'}(X) d\nu(\theta')} d\nu(\theta)$$

1.2.4 Construction d'estimateurs admissibles

Soit T un estimateur bayésien de $g(\theta)$. T est admissible si l'une des 3 conditions suivantes est vérifiée :

1. $\Theta \in \mathbb{R}^p$, ν charge tous les ouverts de Θ et $\theta \longrightarrow R(\theta, T')$ est continue pour tous T' estimateur de $g(\theta)$.
2. Θ est dénombrable et ν charge tous les points de Θ .
3. T est unique en tant qu'estimateur bayésien P_θ p.s., pour tout θ dans Θ (c.a.d. que si T' minimise $r_\nu(.)$, alors $P_\theta(\{T' = T\}) = 0, \forall \theta \in \Theta$).

Dém : 1. Par l'absurde.

Si T n'est pas admissible ; $\exists T'$ tq $\forall \theta, R(\theta, T') \leq R(\theta, T)$ et $\exists \theta_0 \in \Theta$ tq $R(\theta_0, T') < R(\theta_0, T)$.

On suppose $\theta \underset{n \to +\infty}{\longrightarrow} R(\theta, T')$ continue, donc :

$$\exists V \in \mathcal{V}(\theta_0), R(\theta, T) - R(\theta, T') > \frac{R(\theta_0, T) - R(\theta_0, T')}{2} > 0$$

Or ν charge V donc $\nu(V) > 0$ et quand on intègre :

$$\begin{matrix} \int_V [R(\theta, T) - R(\theta, T')] \, d\nu(\theta) > 0 \\ \int_\theta -V [R(\theta, T) - R(\theta, T')] \, d\nu(\theta) \geq 0 \end{matrix} \Longrightarrow \int_\Theta [R(\theta, T) - R(\theta, T')] \, d\nu(\theta) > 0 \Longrightarrow r_\nu(T) > r_\nu(T')$$

ce qui contredit l'hypothèse T bayésien. $\square$

Chapter 2

Modèles statistiques

2.1 Formulation

2.1.1 Exemples :

Inspection

Considérons une livraison de produits manufacturés formée de N objets. Un certain nombre de ces objets sont défectueux. Soit $N\theta$ ce nombre inconnu, on voudrait avoir plus d'information sur la valeur de θ .

On tire un échantillon de n objets, sans remise, et on les vérifie. L'observation est donc, pour chaque objet i, une variable binaire indiquant qu'il est défectueux ($X_i = 1$) ou non ($X_i = 0$), et on construit X la variable aléatoire valant le nombre d'objets défectueux : $X = \sum_{i=1}^{n} X_i$. Alors,

$$P(X = k) = \frac{C_{N\theta}^{k} C_{N-N\theta}^{n-k}}{C_N^n} \text{ où } \max(n - N(1-\theta), 0) \leq k \leq \min(N\theta, n)$$

c'est-à-dire $X \sim \mathcal{H}(N\theta, N, n)$, loi hypergéométrique. Nous avons donc a considérer une famille de lois indexée par θ .

Mesure :

Un expérimentateur réalise n mesures indépendantes d'une constante physique. Soient $X_1, X_2, \ldots X_n$ ces n mesures.

On peut supposer que $X_i = \mu + \epsilon_i$, $1 \leq i \leq n$, où $\epsilon = (\epsilon_1, \epsilon_2, \ldots \epsilon_n)$ est le vecteur d'erreurs.

Plusieurs types d'hypothèses sont envisageables :

1. ϵ est indépendant de μ
2. les ϵ_i sont indépendantes
3. les ϵ_i sont iid
4. la loi des ϵ_i est continue (à densité), et symétrique par rapport à 0
5. les ϵ_i sont iid de loi $\mathcal{N}(0, \sigma^2)$, σ^2 inconnu
6. les ϵ_i sont iid de loi $\mathcal{N}(0, \sigma^2)$, σ^2 connu

Là encore, on a à faire à une famille de lois, plus ou moins précisément définies. Dans les deux derniers cas, la famille est paramétrée par $\theta = (\mu, \sigma^2)$ ou par $\theta = \mu$.

Remarques :

- On ne sait vraiment traiter le deuxième exemple que dans les cas 5. ou 6. mais ces hypothèses peuvent être **fausses**, alors les conclusions risquent d'être fausses elles aussi.
- toutes les échelles de "certitudes" sur le modèle choisi sont possibles.
- On va pour le moment admettre que le modèle est correct.

2.1.2 Modèles paramétrés ou non

Soient un espace de réalisations Ω, et $X = (X_1, \ldots, X_n)$ un vecteur aléatoire représentant les observations (ou données). On considère la loi de X, et on suppose qu'elle appartient à une famille de lois $\mathcal{P}$ sur $\mathbb{R}^n$. Les paramètres intervenant dans cette famille peuvent-être ceux qui nous intéressent ou des paramètres parasites (ex. σ^2 dans l'exemple précédent, 5.).

En fait $\mathcal{P}$ peut être paramétrée ou non paramétrée, si l'on considère les cas 5., 6. ou 4. de l'exemple précédent.

Si on admet que $\mathcal{P}$ est paramétrée par θ (contenant à la fois les paramètres parasites et les autres) et $\theta \in \Theta$ espace des paramètres. On note $\mathcal{P}\{P_\theta, \theta \in \Theta\}$.

Définition 2.1 : *Un paramètre θ est identifiable dans Θ si $\theta_1 \neq \theta_2 \Longrightarrow P_{\theta_1} \neq P_{\theta_2}$, sinon il n'est pas identifiable.*

θ n'est pas identifiable si $\exists \theta_1, \theta_2 \in \Theta$ tels que $P_{\theta_1} = P_{\theta_2}$, on ne peut pas alors espérer "retrouver" θ_1.

Exemple : Soient X_1, X_2 indépendantes, de lois $\mathcal{N}(\alpha_1+\nu, 1)$ et $\mathcal{N}(\alpha_2+\nu, 1)$, où $\alpha_1, \alpha_2, \nu \in \mathbb{R}$. $\theta = (\alpha_1, \alpha_2, \nu)$ n'est pas identifiable, mais pour ν fixé, $\theta = (\alpha_1, \alpha_2)$ est identifiable.

Définition 2.2 : *Les modèles pour lesquels l'une des hypothèses suivantes est vérifiée :*

- *Toutes les P_θ ont des densités $f_\theta(x)$*
- *Toutes les P_θ sont discrètes et il existe un ensemble $\{x_1, x_2, \ldots, x_n, \ldots\}$ indépendant de θ tel que* $\sum_{i \in \mathbb{N}} P_\theta(x_i) = 1$.

sont appelés réguliers.

Pour toutes les opérations classiques en statistique : estimation, test, intervalle de confiance, on utilise des "statistiques", c'est-à-dire des fonctions intermédiaires.

2.2 Statistiques - Exhaustivité

Définition 2.3 : *Une statistique est une fonction aléatoire $T(X)$.*

Définition 2.4 : *Une statistique $T(X)$ est exhaustive (suffisante) pour un paramètre θ ssi la loi de $(X|T(X) = t)$ ne dépend pas de θ .*

Exemple : Un machine fabrique n objets, chaque objet est défectueux avec la probabilité $1 - \theta$, correct avec la probabilité θ . On construit des variables aléatoires X_i prenant la valeur 0 si l'objet est défectueux, et la valeur 1 sinon. $X = (X_1, X_2, \ldots, X_n)$ est l'observation. Alors :

$$P(X_1 = x_1, \ldots, X_n = x_n) = \theta^s (1-\theta)^{n-s}, \text{ où } s = \sum_{i=1}^{n} x_i$$

$$\text{et } P(X = x|S = s) = \frac{P(X = x, S = s)}{P(S = s)} = \frac{P(X = s)}{P(S = s)} = \frac{1}{C_n^x} \text{ où } S = \sum_{i=1}^{n} X_i$$

La statistique $S = \sum_{i=1}^{n} X_i$ est exhaustive.

Définition 2.5 : *Deux statistiques T_1, T_2 sont équivalentes si*

$$T_1(x) = T_1(y) \Longleftrightarrow T_2(x) = T_2(y).$$

Théorème 2.1 : *Pour un modèle régulier, $T : \mathbb{R}^n \longrightarrow \mathbb{I}$ est exhaustive ssi $\exists g : I \times \Theta \longrightarrow \mathbb{R}$ et $h : \mathbb{R}^n \longrightarrow \mathbb{R}$ tels que $f_\theta(x) = g(T(x), \theta)h(x)$ ou $P_\theta(x) = g(T(x), \theta)h(x)$, $\forall x \in \mathbb{R}^n$ et $\theta \in \Theta$.*

Dém : Nous allons faire la démonstration de ce théorème dans le cas discret. La démonstration se fait de manière similaire dans le cas continu. Calculons la loi conditionnelle :

$$P_\theta(X = x | T(X) = t) = \frac{P_\theta(X = x, T(X) = t)}{P_\theta(T(X) = t)} = \begin{cases} 0 & \text{si } T(x) \neq t \\ \frac{g(t,\theta)h(x)}{P_\theta(T(X)=t)} & \text{si } T(x) = t \end{cases}$$

Or, $P_\theta(T(X) = t) = \sum_{x/T(x)=t} P_\theta(X = x) = \sum_{x/T(x)=t} g(t, \theta)h(x) = g(t, \theta) \sum_{x/T(x)=t} h(x)$ et donc

$$P_\theta(X = x | T(X) = t) = \begin{cases} 0 & \text{si } T(x) \neq t \\ \frac{h(x)}{\sum_{x/T(x)=t} h(x))} & \text{si } T(x) = t \end{cases}$$

qui est bien indépendant de θ , on a bien vérifié que si la loi se factorise sous la forme annoncée dans le théorème, $T(X)$ est exhaustive.

Réciproquement, si $T(X)$ est exhaustive, posons $g(t, \theta) = P_\theta(T(x) = t)$ et $h(x) = P_\theta(X = x | T(X) = T(x))$, on a bien alors l'égalité recherchée. □

Exemples :

- $X_1, \ldots, X_n$ intervalles de temps entre les arrivées successives de n clients, alors :

$$\begin{aligned} P_\theta(x_1, \ldots, x_n) &= \theta^n exp[-\theta \textstyle\sum_{i=1}^{n} x_i] \text{ si } \forall i, x_i > 0 \\ &= 0 \text{ sinon} \end{aligned}$$

  Soit $T(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$, $g(t, \theta) = \theta^n e^{-\theta t}$ si $t > 0$ et $\theta > 0$ et $h(x_1, \ldots, x_n) = 1$ si $\forall i$, $x_1 > 0$, 0 sinon. Alors $P_\theta(x_1, \ldots, x_n) = g(t, \theta).h(x_1, \ldots, x_n)$.

- Considérons une population de taille θ , d'individus numérotés. On tire n individus et on observe leurs rangs, soit $T(X) = \max(X_1, \ldots, X_n)$. Alors $P(x_1, \ldots, x_n) = \theta^{-n}$ si $x_i \in \{1, 2, \ldots, \theta\}$ donc $P(x_1, \ldots, x_n) = \theta^{-n}\mathbf{1}_{\{\max x_i \leq \theta\}}$

- Soit $X_1, \ldots, X_n$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, μ et σ^2 inconnus. Soit $\theta = (\mu, \sigma^2)$. Alors

$$\begin{aligned} f_\theta(x_1, \ldots, x_n) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[\frac{-n\mu^2}{2\sigma^2}\right].\exp \frac{-1}{2\sigma^2}\left[\sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i\right] \end{aligned}$$

  Les statistiques $T_1(X) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ et $T_2(X) = (\frac{1}{n}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2)$ sont exhaustives

2.3 Familles exponentielles

2.3.1 Familles à 1 paramètre

Définition 2.6 : *$\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ est une famille exponentielle à 1 paramètre s'il existe des fonctions $c(.)$ et $d(.)$ de Θ dans $\mathbb{R}$, des fonctions $T(.)$ et $S(.)$ de $\mathbb{R}^n$ dans $\mathbb{R}$ et $A \subset \mathbb{R}^n$ tels que :*

$$\begin{aligned} & f_\theta(x) = \exp[c(\theta)T(x) + d(\theta) + S(x)]\mathbf{1}_A(x) \\ \textit{ou} \quad & P_\theta(x) = \exp[c(\theta)T(x) + d(\theta) + S(x)]\mathbf{1}_A(x) \end{aligned}$$

Remarques :

- Il n'y a pas d'unicité, a priori.
- A ne dépend pas de θ .
- $T(X)$ est exhaustive pour θ : $g(T(x),\theta) = \exp[c(\theta)T(x) + d(\theta)]$ et $h(x) = \exp[S(x)]\mathbf{1}_A(x)$.

Exemples :

1. $f_\mu(x) = \frac{1}{\sqrt{2\pi}.\sigma_0}\exp[\frac{-1}{2\sigma_0^2}(x-\mu)^2]$, σ_0 fixé
   $c(\mu) = \frac{\mu}{\sigma_0^2}$, $d(\mu) = \frac{-\mu^2}{2\sigma_0^2}$, $T(x) = x$ et $S(x) = -\left(\frac{x^2}{2\sigma_0^2} + \ln(\sqrt{2\pi}.\sigma_0)\right)$.
2. $f_\sigma^2(x) = \frac{1}{\sqrt{2\pi}.\sigma}\exp[\frac{-1}{2\sigma^2}(x-\mu_0)^2]$, μ_0 fixé
   $c(\sigma^2) = \frac{-1}{2\sigma^2}$, $T(x) = (x-\mu_0)^2$ et $d(\sigma) = -\ln(\sqrt{2\pi}.\sigma)$.
3. $P_\theta(x) = C_n^x\theta^x(1-\theta)^{n-x}\mathbf{1}_A(x)$, $A = \{0,1,\ldots,n\}$
   $P_\theta(x) = \exp\left[x\ln(\frac{\theta}{1-\theta}) + n\ln(1-\theta) + \ln C_n^x\right]\mathbf{1}_A(x)$ et $c(\theta) = \ln(\frac{\theta}{1-\theta})$, $d(\theta) = n\ln(1-th)$, $T(x) = x$ et $S(x) = \ln C_n^x$.

Théorème 2.2 : *Soit $\{P_\theta\}$ une famille exponentielle discrète à un paramètre, définie par c, d, S, T et A, et soit $\{Q_\theta\}$ la famille de distributions de $T(X)$. Alors, $\{Q_\theta\}$ est une famille exponentielle discrète à un paramètre, de probabilités $Q_\theta(t) = \exp[c(\theta)t + d(\theta) + S^*(t)]\mathbf{1}_{A^*}(t)$.*

Dém : Par définition :

$$\begin{aligned} Q_\theta(t) = P_\theta(T(X)=t) = \textstyle\sum_{x/T(x)=t} P_\theta(x) &= \textstyle\sum_{x/T(x)=t}\exp[c(\theta)T(x) + d(\theta) + S(x)]\mathbf{1}_A(x) \\ &= \exp[c(\theta)t + d(\theta)].\textstyle\sum_{x/T(x)=t}\exp[S(x)]\mathbf{1}_{A^*}(t) \end{aligned}$$

où A^* est l'image de A par T. Il suffit alors de poser $S^*(t) = \ln(\sum_{x/T(x)=t}\exp[S(x)])$. □

Remarque : Même résultat pour une famille de lois exponentielles à densité.

Posons $\eta = c(\theta)$, alors dans le cas continu, la famille de loi devient

$$f_\theta(x) = \exp[\eta T(x) + d_0(\eta) + S(x)]\mathbf{1}_A(x)$$

$$\text{avec } d(\eta) = -\ln\int_A exp[\eta T(x) + S(x)]dx$$

Dans le cas discret :

$$P_\theta(x) = \exp[\eta T(x) + d_0(\eta) + S(x)]\mathbf{1}_A(x)$$

$$\text{avec } d(\eta) = -\ln\sum_{x\in A} exp[\eta T(x) + S(x)]$$

Ces expressions sont directement obtenues en identifiant les constants de normalisation dans les deux cas.

Soit $H = \{\eta = c(\theta) |\ \theta \in \Theta$ et $d_0(\eta)$ fini$\}$ alors H est un intervalle de $\mathbb{R}$ ou $\mathbb{R}$ tout entier, et la famille de lois paramétrée par $\eta \in H$ contient celle paramétrée par $\theta \in \Theta$. On appelle cette paramétrisation de la famille exponentielle sa **forme canonique**, et H est appelé l'espace canonique des paramètres.

Théorème 2.3 : *Si η appartient à l'intérieur de H, la fonction génératrice de $T(X)$ existe et est donnée par : $\psi(s) = \exp[d_0(\eta) - d_0(s+\eta)]$, pour s voisin de 0. De plus, $E(T(X)) = -d_0'(\eta)$ et $Var(T(X)) = -d_0''(\eta)$.*

Dém : La démonstration de ce théorème s'obtient directement par le calcul :

$$\psi(s) = E(e^{sT(X)}) = \sum_{x\in A} e^{sT(x)} \exp[\eta T(x) + d_0(\eta) + S(x)]$$

$$= \sum_{x\in A} \exp[(\eta+s)T(x) + d_0(\eta) + S(x)] = \exp[d_0(\eta) - d_0(\eta+s)]$$

et les calculs de $E(T(X))$ et $Var(T(X))$ s'en déduisent. □

2.3.2 Familles à k paramètres

Définition 2.7 : *$\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ est une famille exponentielle à k paramètres s'il existe des fonctions $c_1, c_2, \ldots, c_k$ et d de Θ dans $\mathbb{R}$, des fonctions $T_1, T_2, \ldots, T_k$ et S de $\mathbb{R}^n$ dans $\mathbb{R}$ et $A \subset \mathbb{R}^n$ tels que :*

$$f_\theta(x) = \exp[\textstyle\sum_{i=1}^k c_i(\theta)T_i(x) + d(\theta) + S(x)]\mathbf{1}_A(x)$$
$$\text{ou} \quad P_\theta(x) = \exp[\textstyle\sum_{i=1}^k c_i(\theta)T_i(x) + d(\theta) + S(x)]\mathbf{1}_A(x)$$

Exemple : Soit $P_\theta = \mathcal{N}(\mu, \sigma^2)$, $\Theta = \{(\mu, \sigma^2)/\mu \in \mathbb{R}, \sigma \in \mathbb{R}^{+*}\}$.

$$f_\theta(x) = \exp\left[\frac{\mu}{\sigma^2}x - \frac{x^2}{2\sigma^2} - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \ln(2\pi\sigma^2)\right)\right] (x-\mu)^2]$$

$c_1(\theta) = \frac{\mu}{\sigma^2}$, $c_2(\theta) = \frac{-1}{2\sigma^2}$, $T_1(x) = x$, $T_2(x) = x^2$, $d(\theta) = \frac{-1}{2}\left(\frac{\mu^2}{\sigma^2} + \ln(2\pi\sigma^2)\right)$, $S(x) = 0$ et $A = \mathbb{R}$.

Remarques :

- Il n'y a pas d'unicité de la décomposition
- $T(X) = (T_1(X), \ldots, T_k(X))$ est une statistique exhaustive.

2.4 Modèle bayésien

Considérons une famille $\{P_\theta, \theta \in \Theta\}$, et introduisons une variable aléatoire $\tilde{\theta}$ à valeurs dans Θ et de loin π. π représente l'information dont on dispose sur la variable θ avant l'expérience, elle est appelée la loi à priori. P_θ est alors la loi conditionnelle de X sachant $\tilde{\theta} = \theta$. Si les deux variables X et θ sont discrètes ou à densité, la loi jointe de (θ, X) est donc $P_\theta(x)\pi(\theta)$. On notera en particulier $P(x|\theta)$ au lieu de $P_\theta(x)$, et par les formules classiques d'inversion :

$$\pi(\theta|x) = \frac{\pi(\theta)P(x|\theta)}{\sum_{t\in\Theta}\pi(t)P(x|t)} \text{ dans le cas discret}$$

$$\pi(\theta|x) = \frac{\pi(\theta)P(x|\theta)}{\int_\Theta \pi(t)P(x|t)dt} \text{ dans le cas continu}$$

π et P sont alors des lois discrètes ou des densités suivant les cas.

Définition 2.8 : *Si $\Theta \subset \mathbb{R}^k$, les lois exponentielles*

$$P(x|\theta) = \exp\left[\sum_{j=1}^k c_j(\theta)T_j(x) + S(x) + d(\theta)\right]\mathbf{1}_A(x)$$

$$\text{et } \pi(\theta|t_1, t_2, \ldots, t_k) = \exp\left[\sum_{j=1}^k c_j(\theta)t_j + d(\theta) - \ln K(t_1, t_2, \ldots, t_k)\right]$$

$$où\ (t_1, t_2, \ldots, t_k) \in \Omega = \{(t_1, t_2, \ldots, t_k)/K(t_1, t_2, \ldots, t_k) < +\infty\}$$

sont dites conjuguées. $K(t_1, t_2, \ldots, t_k)$ *est la constante de normalisation.*

Chapter 3

Méthodes d'estimation

En règle générale, après avoir construit un modèle statistique $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, on cherche à estimer θ ou une fonction $q(\theta)$. Pour estimer $q(\theta)$, on commence par choisir une statistique qui devrait être "proche" de θ . De nombreux estimateurs sont a priori utilisables, nous allons, dans ce chapitre, présenter quelques estimateurs classiquement utilisés. On note en général $\hat{\theta}$ un estimateur de θ.

3.1 Principe de substitution

Nous supposons que $X_1, \ldots, X_n$ est un échantillon d'une distribution connue ou partiellement connue.

3.1.1 Substitution des fréquences

Supposons que chaque X_i suive une loi multinomiale dont les valeurs $v_1, \ldots, v_k$ sont connues, mais dont les probabilités d'apparition $p_1, \ldots, p_k$ sont inconnues.

On construit alors les variables aléatoires N_i égales au nombre d'indices j tels que $X_j = v_i$. La méthode de substitution des fréquences consiste alors à estimer chaque p_i par $\hat{p_i} = N_i/n$, et plus généralement, une fonction $q(p_i, \ldots, p_k)$ est estimée par $\hat{q}(p_i, \ldots, p_k) = q(\hat{p_i}, \ldots, \hat{p_k}) = q(N_1/n, \ldots, N_k/n)$.

Remarque : Il n'y a pas en général unicité d'un tel estimateur.

Exemple : Supposons que l'on échantillonne une population génétiquement stable, afin d'observer un gène simple ayant deux allèles. Supposons que les trois génotypes soient identifiables, nous allons alors considérer trois types d'individus dont les probabilités d'apparition sont données par$p_1 = \theta^2$, $p_2 = 2\theta(1-\theta)$ et $p_3 = (1-\theta)^2$ où θ est un paramètre réel compris entre 0 et 1. Nous avons alors plusieurs estimés possibles de θ : $\sqrt{N_1/n}$ ou $1 - \sqrt{N_3/n}$

3.1.2 Méthode des moments

Notons $m_1(\theta), \ldots, m_r(\theta)$ les r premiers moments de la loi considérée : $m_j(\theta) = E_\theta(X_1^j)$. Et construisons les moments empiriques : $\hat{m}_j = \frac{1}{n}\sum_{i=1}^n X_i^j$, $j \geq 1$, qui sont des estimateurs des moments $m_j(\theta)$. De même que précédemment, un estimateur de $q(m_1(\theta), \ldots, m_r(\theta))$ sera obtenu par $q(\hat{m}_1(\theta), \ldots, \hat{m}_r(\theta))$

Exemples :

- Si $q(\theta) = Var(X) = m_2(\theta) - m_1^2(\theta)$, on obtient la variance empirique $\hat{\sigma}^2 = \hat{m}_2 - \hat{m}_1^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2 = \sum_{i=1}^n (X_i - \bar{X})^2$.
- Si $X_1, \ldots, X_n$ est un $\mathcal{N}(\mu, \sigma^2)$ n-échantillon, les estimateurs obtenus par la méthode des moments pour μ et σ^2 sont $\bar{X}$ et $\hat{\sigma}^2$.
- Estimation de la taille d'une population : $P_\theta(x_i) = 1/\theta$ et $E_\theta(X_i) = \sum_{x_i=1}^\theta x_i/\theta = \theta(\theta+1)/2\theta = (\theta+1)/2$. L'estimateur de θ que l'on obtient est alors $\hat{\theta} = 2\bar{X} - 1$, ce qui est clairement un "mauvais" estimateurs, en particulier si $\max_i X_i > 2\bar{X} - 1$, puisque alors $\theta \geq \max_i X_i$.

Remarque : Il n'y a pas non plus d'unicité de tels estimateurs, par exemple si la loi considérée est une loi de poisson de paramètre θ , θ est à la fois l'espérance et la variance de X_i ce qui permet donc de construire au moins deux estimateurs.

3.2 Moindres carrés

3.2.1 Modèles de régression linéaires ou non

Supposons que l'on observe $Y_1, \ldots, Y_n$ et que

$$Y_i = g_i(\theta_1, \ldots, \theta_r) + \epsilon_i, \quad 1 \leq i \leq n$$

où les fonctions g_i sont des fonctions connues, et les réels $\theta_1, \ldots, \theta_n$ sont des paramètres inconnus, variants dans $\Omega \subset \mathbb{R}^r$. Supposons de plus que les ϵ_i vérifient les hypothèses suivantes :

$$E(\epsilon_i) = 0, \quad 1 \leq i \leq n$$

$$Var(\epsilon_i) = \sigma^2 > 0, \quad 1 \leq i \leq n$$

$$Cov(\epsilon_i, \epsilon_j) = 0, \quad 1 \leq i < j \leq n$$

Ces hypothèses seront en particulier vérifiées si les ϵ_i sont iid. d'espérance 0 et de variance σ^2. Un cas particulier souvent utilisé est celui ou les ϵ_i sont un n-échantillon de loi $\mathcal{N}(0, \sigma^2)$.

Exemple : Le modèle de régression linéaire : $Y_i = \theta_1 + \theta_2 x_i + \epsilon_i$ ou encore $\theta = (\theta_1, \theta_2)$ et $g(\theta_1, \theta_2) = \theta_1 + \theta_2 x_i$. Soit, par exemple, x la quantité de phosphore répandue dans un champ, et Y la quantité de phosphore retrouvée dans un plan de maïs après 38 jours. 9 échantillons de terrains ont été utilisé pour obtenir :

x_i	1	4	5	9	11	13	23	23	28
y_i	64	71	54	81	76	93	77	95	109

L'estimateur des moindres carrés consiste à construire $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_r)$ de manière à ce qu'il réalise le minimum de

$$\sum_{i=1}^{n} (y_i - E(Y_i))^2 = \sum_{i=1}^{n} (y_i - g(\theta_1, \ldots, \theta_n))^2$$

dans Ω. L'estimateur aux moindres carré d'un fonction $q(\theta)$ est alors $q(\hat{\theta})$.

Si les fonctions g_i sont toutes différentiables et à valeurs dans un fermé, alors $\hat{\theta}$ est toujours défini. **Si Ω est un ouvert**, $\hat{\theta}$ doit vérifier :

$$\frac{\partial}{\partial \theta_j} \sum_{i=1}^{n} [y_i - g_i(\theta)]^2 = 0$$

Ces équations sont appelées les équations normales, elles peuvent encore s'écrire :

$$\sum_{i=1}^{n} [y_i - g_i(\theta)] \frac{\partial}{\partial \theta_j} g_i(\theta) = 0, \quad j = 1, \ldots, r$$

Si les fonctions $g_i(\theta)$ sont linéaires, ces équations forment un système linéaire. Dans ce cas, nous parleront de moindres carrés linéaires et si de plus les ϵ_i sont iid de loi $\mathcal{N}(0, \sigma^2)$, nous parlerons de modèle linéaire.

Exemples :

- Modèle de mesure : $Y_i = \theta_1 + \epsilon_i$, $r = 1$, alors $\frac{\partial}{\partial \theta_1} g_1(\theta_1) = 1$ et les équations normales deviennent :

$$\sum_{i=1}^{n} (y_i - \theta_1) = 0$$

  dont la solution est $\hat{\theta}_1 = (1/n) \sum_{i=1}^{n} y_i = \bar{y}$. On vérifie qu'il s'agit bien d'un minimum.

- Modèle de régression : les équations normales sont

$$\begin{cases} \sum_{i=1}^n (y_i - \theta_1 - \theta_2 x_i) = 0 \\ \sum_{i=1}^n x_i(y_i - \theta_1 - \theta_2 x_i) = 0 \end{cases}$$

  que l'on résout en

$$\hat{\theta}_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{et} \quad \hat{\theta}_1 = \bar{y} - \hat{\theta}_2 \bar{x}$$

  On vérifie qu'il s'agit bien d'un minimum.

  La droite $y = \hat{\theta}_1 + \hat{\theta}_2 x$ est la droite de régression. C'est la droite qui minimise la somme des carrés des distances "verticales" des points (x_i, y_i) à la droite.

3.2.2 Moindres carrés pondérés

Supposons cette fois que les ϵ_i aient des variances différentes : $Var(\epsilon_i) = w_i\sigma^2$, où σ^2 est inconnu et les w_i sont connus. Le modèle précédent n'est plus directement applicable, mais si on pose $Z_i = Y_i/\sqrt{w_i} = (g_i(\theta)/\sqrt{w_i}) + (\epsilon_i/\sqrt{w_i})$, $1 \le i \le n$, alors $Var(\epsilon_i/\sqrt{w_i}) = w_i\sigma^2/w_i = \sigma^2$. Donc en posant $g_i^*(\theta) = g_i(\theta)/\sqrt{w_i}$ et $\epsilon_i^* = \epsilon_i/\sqrt{w_i}$, on retombe sur un modèle similaire au précédent :

$$Z_i = g_i^*(\theta) + \epsilon_i^*, \quad 1 \le i \le n$$

L'estimateur des moindres carrés pondéré est la valeur $\hat{\theta}$ qui minimise

$$\sum_{i=1}^n [z_i - g_i^*(\theta)]^2 = \sum_{i=1}^n [y_i - g_i(\theta)]^2/w_i.$$

On obtient des équations normales comme précédemment.

3.3 Maximum de vraisemblance

Soit $p_\theta(x)$ une loi discrète ou une densité, $\theta \in \Theta \subset \mathbb{R}^k$. Lorsque l'on considère cette fonction comme une fonction de θ , on l'appelle la vraisemblance et on la note $L(\theta, x)$. On peut considérer que $L(\theta, x)$ est une mesure de la vraisemblance que θ ait produit x

La méthode du maximum de vraisemblance consiste à rechercher la valeur de $\hat{\theta}(x)$ qui a le plus vraisemblablement produit la donnée x. C'est-à-dire de chercher $\hat{\theta}(x)$ qui satisfait :

$$L(\hat{\theta}(x), x) = p_{\hat{\theta}}(x)(x) = max\{L(\theta, x), \theta \in \Theta\} = max\{p_\theta(x), \theta \in \Theta\}$$

Si un tel estimateur existe, on estime une fonction $q(\theta)$ par $q(\hat{\theta}(x)$.

Remarques :

- Un estimateur du maximum de vraisemblance n'existe pas nécessairement, et n'est pas nécessairement unique s'il existe.
- Si la statistique $T(X)$ est exhaustive, l'estimateur ne dépend que de $T(X)$.

3.3.1 Familles à un paramètre

Exemples :

- Distribution Gaussienne avec variance connue : $X \sim \mathcal{N}(\mu, \sigma^2)$, où σ^2 est connue. Alors

$$L(\theta, x) = \frac{1}{\sigma}\phi(\frac{\theta - x}{\sigma}) \text{ où } \phi \text{ est la densité gaussienne centrée reduite}$$

  Le maximum de vraisemblance est donc obtenu pour $\hat{\theta}(x) = x$.

- Soit $X_1, \ldots, X_n$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, σ^2 connu.

$$\begin{aligned} L(\mu, (x_1, \ldots, x_n)) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[\frac{-n\mu^2}{2\sigma^2}\right] . \exp\frac{-1}{2\sigma^2}\left[\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i\right] \end{aligned}$$

$L(\mu, x)$ est maximal en μ ssi $\phi(\mu) = \frac{n\mu^2}{2\sigma^2} - \frac{\mu}{\sigma^2}\sum_{i=1}^n x_i + \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2$ est minimum en μ.

$$\phi'(\mu) = \frac{n\mu}{\sigma^2} - \frac{1}{\sigma^2}\sum_{i=1}^n x_i = 0 \Longleftrightarrow \mu = \frac{1}{n}\sum_{i=1}^n x_i$$

On vérifie que $\mu = \bar{x}$ correspond bien à un minimum et qu'il appartient à $\Theta = \mathbb{R}$. $\hat{\mu} = \bar{X}$ est l'estimateur du maximum de vraisemblance.

- Estimation de la taille d'une population : $L(\theta, x) = \theta^{-n}\prod_{i=1}^n \mathbf{1}_{\{1,\ldots,\theta\}}(x_i)$, l'estimateur du maximum de vraisemblance est alors $\hat{\theta} = \max(x_1, \ldots, x_n)$.

Posons

$$\mathcal{L}(\theta, x) = \ln L(\theta, x) = \ln p_\theta(x)$$

et supposons que Θ est un ouvert et que $\mathcal{L}$ est différentiable en θ pour x fixé. Alors, si $\hat{\theta}(x)$ existe, il doit satisfaire l'équation de vraisemblance

$$\frac{\partial}{\partial\theta}\mathcal{L}(\theta, x) = 0$$

En effet, puisque la fonction ln est croissante, $\hat{\theta}$ maximise $L(\theta, x)$ ssi il maximise $\mathcal{L}(\theta, x)$, et en dérivant, on obtient l'équation précédente.

Pour n variables aléatoires indépendantes de densités $f_i(\theta, x)$, alors

$$\mathcal{L}(\theta, x) = \ln\prod_{i=1}^n f_i(\theta, x_i) = \sum_{i=1}^n \ln f_i(\theta, x_i)$$

et nous obtenons l'équation de vraisemblance suivante :

$$\sum_{i=1}^n \frac{\partial}{\partial\theta}\ln f_i(\theta, x_i) = 0$$

Exemples :

- Considérons une population constituée de trois types d'individus labellés 1, 2 et 3 associés aux probabilités d'apparition

$$P_\theta(1) = \theta^2, \quad P_\theta(2) = 2\theta(1-\theta), \quad P_\theta(3) = (1-\theta)^2 \text{ où } 0 < \theta < 1$$

Si on observe 3 individus et qu'on obtienne $x_1 = 1$, $x_2 = 2$ et $x_3 = 1$, alors

$$L(\theta, x) = P_\theta(1)P_\theta(2)P_\theta(3) = 2\theta^5(1-\theta)$$

et l'équation de vraisemblance est

$$\frac{\partial}{\partial\theta}\mathcal{L}(\theta, x) = \frac{5}{\theta} - \frac{1}{1-\theta} = 0$$

qui a une solution unique $\hat{\theta} = \frac{5}{6}$, puisque

$$\frac{\partial^2}{\partial\theta^2}\mathcal{L}(\theta, x) = \frac{-5}{\theta^2} - \frac{1}{(1-\theta)^2} < 0$$

pour tout $\theta \in]0, 1[$, donc la fonction $\mathcal{L}(\theta, x)$ est concave en θ .

- Soit X le nombre de clients arrivant à un guichet en n heures. Si l'on admet que les temps d'arrivée suivent un processus de Poisson, alors X suit une loin de Poisson de paramètre $n\lambda$, où λ représente le nombres d'arrivées en une heure.

$$L(\lambda, x) = \frac{(\lambda n)^x e^{-\lambda n}}{x!}, \quad k = 0, 1, \ldots$$

et l'équation de vraisemblance est donc

$$\frac{\partial}{\partial \lambda} \mathcal{L}(\lambda, x) = -n + \frac{x}{\lambda} = 0$$

L'estimateur de maximum de vraisemblance est donc $\hat{\lambda} = x/n$ si x est strictement positif, il n'est pas défini si $x = 0$.

Théorème 3.1 : *Soit*

$$p_\theta(x) = \exp[c(\theta)T(x) + d(\theta) + S(x)]\mathbf{1}_A(x), \quad \theta \in \Theta$$

et soit C l'intérieur de l'ensemble des valeurs prises par c. Supposons que c soit injective. Si l'équation $E_\theta(T(X)) = T(x)$ a un solution $\hat{\theta}(x)$ pour laquelle $c(\hat{\theta}(x)) \in C$, alors $\hat{\theta}(x)$ est le seul estimateur du maximum de vraisemblance du paramètre θ .

Dém : Supposons tout d'abord que $c(\theta) = \theta$ (c'est-à-dire que la famille est sous sa forme naturelle), alors

$$\frac{\partial}{\partial \theta} \mathcal{L}(\theta, x) = T(x) + d'(\theta) \text{ et } \frac{\partial^2}{\partial \theta^2} \mathcal{L}(\theta, x) = d''(\theta)$$

Or $-d'(\theta) = E_\theta(T(X))$ et $-d''(\theta) = Var_\theta(T(X))$, donc $\mathcal{L}$ est concave en θ , cqfd.

Dans la cas général

$$\mathcal{L}(\theta, x) = c(\theta)T(x) + d(\theta) + S(X) = c(\theta)T(x) + d_0(c(\theta)) + S(X)$$

du fait que c est injective. Donc si $c(\hat{\theta})$ est tel que $E_{c(\hat{\theta})}(T(X)) = T(x)$, et puisque c est injective, $\hat{\theta}$ est bien l'estimateur du maximum de vraisemblance. □

Exemples :

- Soit $X_1, \ldots, X_n$ un n-échantillon de loi gaussienne $\mathcal{N}(\theta, \sigma^2)$, σ^2 connu. $T(X) = \sum_{i=1}^n X_i$, $c(\theta) = \theta/\sigma^2$ est bijective et $E_\theta(T(X)) = n\theta$ qui conduit à $\sum_{i=1}^n x_i = n\theta$ et donc à $\hat{\theta} = \bar{x}$ (si $\Theta = \mathbb{R}$).
- Considérons une population constituée de trois types d'individus labellés 1, 2 et 3 associés aux probabilités d'apparition

$$P_\theta(1) = \theta^2, \quad P_\theta(2) = 2\theta(1-\theta), \quad P_\theta(3) = (1-\theta)^2 \text{ où } 0 < \theta < 1$$

On observe n individus et on considère n_1, n_2 et n_3, les nombres d'individus dans chaque classe, alors

$$L(\theta, x) = 2^{n_2} \theta^{2n_1+n_2} (1-th)^{2n_3+n_2} = 2^{n_2} \left(\frac{\theta}{1-\theta}\right)^{2n_1+n_2} (1-th)^{2n}$$

d'où $c(\theta) = \ln(\theta/(1-\theta))$, $T(x) = 2n_1 + n_2$ et $E_\theta(T(X)) = E_\theta(2N_1 + N_2) = 2n\theta^2 + 2n\theta(1-\theta) = 2n\theta$. On obtient donc $\hat{\theta} = (2n_1 + n_2)/2n$.

3.3.2 Familles à k paramètres

Si $\theta = (\theta_1, \ldots, \theta_k)$ est un vecteur de paramètres variant dans $\Theta \subset \mathbb{R}^k$. L'approche reste la même que pour un paramètre réel. Si $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ existe, il doit vérifier le système d'équations suivant :

$$\frac{\partial}{\partial \theta} \mathcal{L}(\theta, x) = 0, \quad i = 1, \ldots, k$$

pour tous les x tels que $\mathcal{L}$ admet des dérivées partielles en θ .

Théorème 3.2 : *Soit*

$$p_\theta(x) = \exp[\sum_{i=1}^{k} c_i(\theta)T_i(x) + d(\theta) + S(x)]\mathbf{1}_A(x), \quad \theta \in \Theta$$

et soit C l'intérieur de l'ensemble des valeurs prises par $c = (c_1, \ldots, c_k)$. On suppose que la fonction c est injective. Si les équations $E_\theta(T_i(X)) = T_i(x)$, pour $i = 1, \ldots, k$, ont une solution $\hat{\theta}(x) = (\hat{\theta}_1(x), \ldots, \hat{\theta}_k(x))$ pour laquelle $c(\hat{\theta}(x)) = (c_1(\hat{\theta}(x)), \ldots, c_k(\hat{\theta}(x))) \in C$, alors $\hat{\theta}(x)$ est le seul estimateur du maximum de vraisemblance du paramètre θ .

Exemple : Loi normale avec les deux paramètres inconnus : Soit $X_1, \ldots, X_n$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$ et $\theta = (\mu, \sigma^2)$, alors $c_1(\theta) = \mu/\sigma^2$, $T_1(x) = \sum_{i=1}^n x_i$, $c_2(\theta) = -1/(2\sigma^2)$, $T_2(x) = \sum_{i=1}^n x_i^2$. D'où $E_\theta(T_1(X)) = n\mu$ et $E_\theta(T_2(X)) = n(\sigma^2 + \mu^2)$. Pour $n > 2$,

$$\begin{cases} n\hat{\mu} = \sum_{i=1}^n x_i \\ n(\hat{\sigma}^2 + \hat{\mu}^2) = \sum_{i=1}^n x_i^2 \end{cases} \Longrightarrow \begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = (1/n)\sum_{i=1}^n (x_i - \bar{x})^2 \end{cases}$$

$\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$ est l'estimateur du maximum de vraisemblance.

3.3.3 Liens avec les autres méthodes

- Un estimateur des moindres carrés peut-être considéré comme un estimateur du maximum de vraisemblance : Supposons que les ϵ_i sont indépendants et suivent une gaussienne $\mathcal{N}(0, \sigma^2)$. Alors

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

  et donc $L(\theta, x)$ est maximal ssi $\sum_{i=1}^n (y_i - g_i(\theta))^2$ est minimal.

- D'un point ce vue bayésien : Si Θ est fini et que l'on choisit une loi a priori uniforme sur Θ , alors l'estimateur du maximum de vraisemblance maximise la loi a posteriori; si Θ est un intervalle, et si la loi a priori est uniforme sur Θ , alors l'estimateur de vraisemblance maximise la densité a posteriori.

Chapter 4

Comparaison d'estimateurs - Optimalité

4.1 Critères d'estimation

Lors du chapitre sur la théorie de la décision, nous avons déjà vu des critères d'estimation du type $R(\theta,T) = E|T(X) - q(\theta)|$ ou $R(\theta,T) = E(T(X) - q(\theta))^2$. Ce dernier est le risque quadratique que nous utiliserons pour la suite.

Remarque : Si $T(X) \sim \mathcal{N}(q(\theta), \sigma_T^2)$, alors $E|T(X) - q(\theta)| = \sqrt{\frac{2}{\pi}}\sqrt{R(\theta,T)}$

Définition 4.1 : *Le biais de l'estimateur T est $b(\theta,T) = E(T(X)) - q(\theta)$. On dit que T est sans biais si $b(\theta,T) = 0$, qu'il est biaisé sinon.*

Remarque : $R(\theta,T) = Var(T(X)) + b^2(\theta,T)$.

Exemple : Soit $X_1,\dots,X_n$ un n-échantillon de loi $\mathcal{N}(\mu,\sigma^2)$, alors $\hat{\mu} = \bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ donc $b(\mu,\bar{X}) = 0$ et $R(\mu,\bar{X}) = Var(\bar{X}) = \sigma^2/n$. De même $n\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-1}$ et donc

$$b(\sigma^2,\hat{\sigma}^2) = \frac{\sigma^2}{n}E(n\hat{\sigma}^2/\sigma^2) - \sigma^2 = ((n-1)/n - 1)\sigma^2 = -\sigma^2/n$$

$$\text{et } R(\sigma^2,\hat{\sigma}^2) = \frac{\sigma^4}{n^2}Var\left(\frac{n\hat{\sigma}^2}{\sigma^2}\right) + \frac{\sigma^4}{n^2} = \frac{\sigma^4}{n^2}[2(n-1)+1] = \frac{\sigma^4(2n-1)}{n^2}$$

Par contre, si $s^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, alors $b(s^2,\sigma^2) = 0$.

Définition 4.2 : *Un estimateur T sans biais de θ est dit UVMB (Uniformément de Variance Minimum parmi les estimateurs sans Biais) si*

$$Var(T(X)) \le Var(S(X)), \forall S \text{ estimateur sans biais de } \theta$$

(c.a.d. $R(\theta,T) \le R(\theta,S)$.)

Remarques :

- Un estimateur non biaisé n'existe pas forcément.
- Un estimateur UVMB peut être non admissible.
- $\hat{\theta}$ non biaisé pour θ n'implique pas que $q(\hat{\theta})$ soit non biaisé pour $q(\theta)$.
- Il existe d'après la théorie de la décision plusieurs critères d'estimation : $\int R(\theta,T)\pi(\theta)d\theta$ pour l'estimateur bayesien et $\max_\theta R(\theta,T)$ pour l'estimateur minimax.

4.2 Estimateurs sans biais uniformément de variance minimale

Théorème 4.1 (de Rao-Blackwell) : *Si $T(X)$ est exhaustive pour θ et S est un estimateur sans biais de $q(\theta)$ tel que $E_\theta(|S(X)|) < +\infty, \forall\theta \in \Theta$. Alors $T^*(X) = E(S(X)|T(X))$ est telle que*

$$\forall\theta \in \Theta, E_\theta[T^*(X) - q(\theta)]^2 \leq E_\theta[S(X) - q(\theta)]^2.$$

Si $Var(S(X)) < +\infty$ il y a inégalité stricte sauf si $T^*(X) = S(X)$ P_θ p.s.

Dém : $E(T^*) = E[E(S|X)] = E(S) \Longrightarrow b(\theta, T^*) = b(\theta, S) = 0$. On veut donc démontrer que

$$E[E(S|T) - ES]^2 \leq E(S - ES)^2.$$

Or, $E(S-ES)^2 = E\left[E\left((S-ES)^2|T\right)\right]$ et on va calculer $E\left((S-ES)^2|T=t\right)$

$$E\left((S-ES)^2|T=t\right) = E\left((S - E(S|T=t) + E(S|T=t) - ES)^2\,|T=t\right)$$

$$= E\left((S-E(S|T))^2|T=t\right) + E\left((E(S|T)-ES)^2|T=t\right) + 2E\left((S-E(S|T))(E(S|T)-ES)|T=t\right)$$

$$= E\left((S-E(S|T))^2|T=t\right) + E\left((E(S|T)-ES)^2|T=t\right)$$

$$+2E\left(S.E(S|T=t) - E(S|T=t)^2 + ES.E(S|T=t) - S.ES\ |T=t\right)$$

$$= E\left((S-E(S|T))^2|T=t\right) + E\left((E(S|T)-ES)^2|T=t\right)$$

$$+2E(S|T=t)E\left(S|T=t\right) - 2E(S|T=t)^2 + 2ES.E(S|T=t) - 2ES.E(S|T=t)$$

$$= E\left((S-E(S|T))^2|T=t\right) + E\left((E(S|T)-ES)^2|T=t\right)$$

$$\geq E\left((E(S|T)-ES)^2|T=t\right)$$

Si $Var(S(X)) < +\infty$, il ne peut y avoir égalité que si $E\left((S-E(S|T))^2|T=t\right) = 0$ c'est-à-dire si $S = E(S|T) = T^*$ P_θ *p.s.* □

Définition 4.3 : *Une statistique T définie sur le modèle statistique $(\Omega, \mathcal{A}, P_\theta, \theta \in \Theta)$ est complète si toute f fonction à valeurs réelles telle que $f(T)$ est intégrable et $E_\theta(f(T)) = 0$, $\forall\theta \in \Theta$, est nulle P_θ p.s.$\forall\theta \in \Theta$.*

Exemple : Soit $X_1, \ldots, X_n$ un n-échantillon de loi $\mathcal{P}(\theta)$. Alors $T(X) = \sum_{i=1}^n X_i$ est exhaustive et $T \sim \mathcal{P}(n\theta)$. On vérifie que T est complète : soit g telle que $g(T)$ est intégrable et $E_\theta(g(T)) = \sum_{i=1}^n g(i)\frac{e^{-n\theta}(n\theta)^i}{i!} = 0\ \forall\theta \in \Theta \Longrightarrow \forall i,\ \frac{g(i)n^i}{i!} = 0 \Longrightarrow \forall i, g(i) = 0$.

Théorème 4.2 (de Lehmann-Scheffe) : *Si $T(X)$ est exhaustive complète et si $S(X)$ est un estimateur sans biais de $q(\theta)$, alors $T^*(X) = E(S(X)|T(X))$ est UVMB pour $q(\theta)$. Si $Var_\theta(S(X)) < +\infty$, $\forall\theta \in \Theta$, $T^*(X)$ est le seul UVMB de $q(\theta)$ P_θ p.s.*

Dém : $b(q(\theta), T^*) = b(q(\theta), S)$ donc T^* est non biaisé. D'après le théorème de Rao-Blackwell, $E(T^*(X) - q(\theta))^2 \leq E(S(X) - q(\theta))^2$ c'est-à-dire $Var(T^*(X)) \leq Var(S(X))$.

Il reste à vérifier que T^* est indépendant de la statistique S choisie : Soient S_1 et S_2 deux estimateurs non biaisés de $q(\theta)$. Puisque T est exhaustive, on peut écrire $S_1 = g_1(T)$ et $S_2 = g_2(T)$ et puisque ces estimateurs sont non biaisés : $E_\theta(g_1(T) - g_2(T)) = 0$ ce qui implique, puisque T est complète, que $g_1(T) = g_2(T)$ P_θ *p.s.* et donc que $E(S_1(X)|T(X)) = E(S_2(X)|T(X))$. L'unicité se déduit alors du cas d'unicité du théorème de Rao-Blackwell. □

Théorème 4.3 : *Soit $(P_\theta,\ \theta \in \Theta)$ une famille exponentielle à k paramètres telle que $c = (c_1, \ldots, c_k)$ soit défini sur une ensemble d'intérieur non vide.*
Alors $T(X) = (T_1(X), \ldots, T_k(X))$ est complète et exhaustive.

Dém : Exhaustivité déjà démontrée. Le reste provient de l'unicité de la transformée de Laplace et est admis. □

Exemples :

1. Soient $X_1, \ldots, X_n$ de loi $\mathcal{N}(\mu, \sigma^2)$ et $\theta = (\mu, \sigma^2)$. La statistique $T(X) = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ est exhaustive et complète. Donc $\bar{X}$ est UVMB si σ^2 est inconnu. De même $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ est UVMB si μ est inconnu, par contre ce n'est plus vrai si μ est connu.

2. Soient $X_1, \ldots, X_n$ de loi $\mathcal{U}(0, \theta)$, $\theta \in \mathbb{R}^{+*}$. Alors $L(x, \theta) = \frac{1}{\theta^n}$ si $max(x_1, \ldots, x_n) < \theta$ et 0 sinon, et donc $M = max(X_1, \ldots, X_n)$ est exhaustive et complète : $P(M < t) = P(X_1 < t)^n$ et donc la densité est $m(t) = n\theta^{-n}t^{n-1}$ si $0 < t < \theta$ ce qui donne $E_\theta(g(M)) = n\theta^{-n} \int_0^\theta g(t)t^{n-1}dt = 0\ \forall \theta \in \Theta$ et donc $g(t) = 0,\ \forall t \in]0, \theta[$. Cependant, M est biaisé : $E_\theta(M) = n\theta/(n+1)$ mais $(n+1)M/n$ ne l'est pas et est donc UVMB.

4.3 Inégalité de Cramer-Rao – Dimension 1

Soit $(P_\theta,\ \theta \in \Theta)$ une famille régulière à densité, et Θ un ouvert de $\mathbb{R}$. On peut donner les mêmes résultats sur des familles de probabilités discrètes en utilisant $P_\theta(x)$ au lieu de $f_\theta(x)$. On note :

(I) $A = \{x / f_\theta(x) > 0\}$ est indépendant de θ et $\forall x \in A, \forall \theta \in \Theta, \frac{\partial}{\partial \theta} \ln f_\theta(x)$ existe et est fini.

(II) si T est une statistique telle que $E_\theta(|T|) < \infty, \forall \theta \in \Theta$, alors $\frac{\partial}{\partial \theta} \int T(x) f_\theta(x) dx = \int T(x) \frac{\partial}{\partial \theta} f_\theta(x) dx$.

Proposition 4.1 : *Si $f_\theta(x) = exp[c(\theta)T(x) + d(\theta) + S(x)]\mathbf{1}_A(x)$ est une famille exponentielle à un paramètre pour laquelle c et d sont dérivables et c a une dérivée continue non nulle sur Θ alors* **(I)** *et* **(II)** *sont vérifiées.*

Définition 4.4 : $I(\theta) = E\left[\frac{\partial}{\partial \theta} \ln f_\theta(X)\right]^2$ *est l'information de Fisher.*

Remarques :

- $0 \leq I(\theta) \leq +\infty$.
- $I(\theta) = \int_{\mathbb{R}} \left(\frac{f'_\theta(x)}{f_\theta(x)}\right)^2 f_\theta(x) dx = \int_{\mathbb{R}} \frac{f'^2_\theta(x)}{f_\theta(x)} dx$.

Théorème 4.4 (Inégalité de Cramer-Rao) : *Soit $T(X)$ une statistique telle que $Var_\theta(T(X)) < +\infty\ \forall \theta \in \Theta$, et soit $\psi(\theta) = E_\theta(T(X))$. On suppose que* **(I)** *et* **(II)** *sont vérifiées et que $0 < I(\theta) < +\infty$, alors ψ est dérivable et*

$$\forall \theta \in \Theta,\ Var_\theta(T(X)) \geq \frac{\psi'(\theta)^2}{I(\theta)}.$$

Dém :

$$\int f_\theta(x) dx = 1 \Longrightarrow \int \frac{\partial}{\partial \theta} f_\theta(x) dx = 0 \Longrightarrow \int \frac{\partial}{\partial \theta} \ln f_\theta(x) f_\theta(x) dx = 0$$

Par ailleurs

$$\psi(\theta) = \int T(x) f_\theta(x) dx \Longrightarrow \psi'(\theta) = \int T(x) \frac{\partial}{\partial \theta} f_\theta(x) dx = \int T(x) \frac{\partial}{\partial \theta} \ln f_\theta(x) f_\theta(x) dx$$

$$\Longrightarrow \psi'(\theta) = Cov(\frac{\partial}{\partial \theta} \ln f_\theta(X), T(X))$$

$$\Longrightarrow |\psi'(\theta)| \leq \sqrt{Var(T(X)).Var(\frac{\partial}{\partial \theta} \ln f_\theta(X))} = \sqrt{Var(T(X)).I(\theta)}$$

□

Corollaire 1 : *Sous les hypothèses du théorème, et si T est non biaisé pour θ, alors*

$$\forall \theta \in \Theta,\ Var(T(X)) \geq \frac{1}{I(\theta)}$$

Corollaire 2 : *Si $X_1, \ldots, X_n$ est une n-échantillon de densité f_θ et sous les hypothèses du théorème :*

$$\forall \theta \in \Theta,\ Var(T(X)) \geq \frac{\psi'(\theta)^2}{nI_1(\theta)}$$

où $I_1(\theta)$ est l'information de Fisher associé à la variable X_1.

Dém :

$$I(\theta) = Var(\frac{\partial}{\partial\theta} \ln f_\theta^{(n)}(X)) = Var(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f_\theta(X_i)) = \sum_{i=1}^{n} Var(\frac{\partial}{\partial\theta} \ln f_\theta(X_i)) = nI_1(\theta)$$

□

Remarques :

- On peut vérifier que les hypothèses **(I)** et **(II)** sont nécessaires pour le corollaire 2 en considérant une loi exponentielle de densité $f(x) = (2x/\theta^2)\mathbf{1}_{[0,\theta]}(x)$. On a alors $I(\theta) = 4n^2/\theta^2$ et $I_1(\theta) = 4/\theta^2$.
- $1/I(\theta)$ est appelé la borne inférieure de Cramer-Rao.
- Pour un estimateur biaisé de $\psi(\theta)$ et sous les hypothèses du théorème, on obtient : $E_\theta(T(X) - \psi(\theta))^2 = Var(T(X)) + b^2(\psi(\theta), T)$ et donc

$$E_\theta(T(X) - \psi(\theta))^2 \geq b^2(\psi(\theta), T) + \frac{\left(\frac{\partial}{\partial\theta} b(\psi(\theta), T) + \psi'(\theta)\right)^2}{I(\theta)}$$

- Dans le cas d'une famille exponentielle à un paramètre écrite sous forme normale, on a $E_\theta(T(X)) = -d'(\theta)$ et $Var_\theta(T(X)) = -d''(\theta)$. L'inégalité de Cramer-Rao devient alors une égalité.

Définition 4.5 : *Un estimateur T sans biais, de carré intégrable de $\psi(\theta)$, tel que*

$$\frac{\partial}{\partial\theta} \int T(x) f_\theta(x) dx = \int T(x) \frac{\partial}{\partial\theta} f_\theta(x) dx \ \text{ et } Var_\theta(T(X)) = \psi'^2(\theta)/I(\theta)$$

est dit efficace.

Théorème 4.5 : *Supposons que la famille $(P_\theta, \theta \in \Theta)$ satisfasse aux hypothèses* **(I)** *et* **(II)** *et qu'un estimateur sans biais T^* de $\psi(\theta)$ atteigne la borne inférieure de l'inégalité de Cramer-Rao pour tout $\theta \in \Theta$. Alors (P_θ) est une famille exponentielle à un paramètre ayant une densité ou une probabilité discrète de la forme :*

$$P_\theta(x) = exp[c(\theta)T^*(x) + d(\theta) + S(x)]\mathbf{1}_A(x).$$

De plus, T^* est UVMB pour $\psi(\theta)$.

Réciproquement, si (P_θ) est une famille exponentielle à un paramètre pour laquelle c a une dérivée continue non nulle sur Θ , alors $T(X)$ atteint la borne minimale de l'inégalité de Cramer-Rao et est UVMB pour $E_\theta(T(X))$.

Dém : Admis pour l'essentiel. □

4.4 Inégalité de Cramer-Rao – Dimension p

Soit $(P_\theta,\ \theta \in \Theta)$ une famille régulière à densité, et Θ un ouvert de $\mathbb{R}^p$. On fait des hypothèses similaires à celles faites en dimension 1 :

(I) $A = \{x / f_\theta(x) > 0\}$ est indépendant de θ et $\forall x \in A, \forall \theta \in \Theta, \nabla_\theta \ln f_\theta(x)$ existe et est fini.

(II) si $T = (T_1, \ldots, T_k)$ est une statistique telle que $\forall i, \forall \theta \in \Theta,\ E_\theta(|T_i|) < \infty$, alors $\nabla_\theta \int T(x) f_\theta(x) dx = \int T(x) \nabla_\theta f_\theta(x) dx$.

Définition 4.6 : *On appelle matrice d'information de Fisher la matrice $I(\theta) = (I_{i,j}(\theta))_{1 \leq i,j \leq p}$ où :*

$$I_{i,j}(\theta) = E_\theta\left(\frac{\partial}{\partial \theta_i} \ln f_\theta(x) \frac{\partial}{\partial \theta_j} \ln f_\theta(x)\right)$$

si elle existe et qu'elle est inversible.

Remarque : Cette matrice, lorsqu'elle existe, est symétrique positive. C'est la matrice de covariance de $\nabla_\theta \ln f_\theta(x)$.

Théorème 4.6 : *Soit $T(X)$ une statistique vectorielle de carré intégrable pour tout $\theta \in \Theta$, et soit $\psi(\theta) = E_\theta(T(X)) \in \mathbb{R}^k$. On suppose que* **(I)** *et* **(II)** *sont vérifiées et que la matrice d'information de Fisher existe, alors ψ est différentiable et la matrice de covariance $\Gamma_\theta(T(X))$ vérifie*

$$\forall \theta \in \Theta,\ \Gamma_\theta(T(X)) \geq {}^t\nabla_\theta \psi(\theta).I^{-1}(\theta).\nabla_\theta \psi(\theta)$$

Remarque : Il s'agit ici d'inégalités sur des matrices symétriques positives : $M \geq N$ ssi $M - N$ est positive.

4.5 Théorie des grands échantillons

On considère $(X_n, n \geq 1)$ une suite d'observations indépendantes, et, pour tout $n \geq 1$, soit $T_n = T(X_1, \ldots, X_n)$ un estimateur de $q(\theta)$ basé sur le n-échantillon $(X_1, \ldots, X_n)$, tous les T_n étant obtenus de manière similaires.

4.5.1 Consistance

Définition 4.7 : *On dit que la suite T_n est fortement (resp. faiblement) consistante si $T_n \xrightarrow[n \to +\infty]{p.s.} q(\theta)$ (resp. $T_n \xrightarrow[n \to +\infty]{P} q(\theta)$).*

Remarque : Pour la convergence presque sûre : $P(\lim_n T_n = q(\theta)) = 1$, et pour la convergence en probabilité : $\forall \epsilon > 0, \lim_n P(|T_n - q(\theta)| \geq \epsilon) = 0$.

Théorème 4.7 : *Si T_n est obtenu par substitution de fréquence, dans le cas multinomial :*

$$T_n = h(\frac{N_1}{n}, \ldots, \frac{N_k}{n}) \text{ où } q(\theta) = h(p_1(\theta), \ldots, p_k(\theta))$$

et si h est continue, alors T_n est fortement consistant.

Si T_n est obtenu par substitution des moments :

$$T_n = g(\hat{m}_1, \ldots, \hat{m}_k) \text{ où } q(\theta) = g(m_1, \ldots, m_k)$$

si tous les moments m_i, $i \leq 2k$, existent et si g est continue, alors T_n est fortement consistant.

Dém : Par la loi forte des grands nombres. □

Remarque : On peut montrer que les estimateurs UVMB sont faiblement consistants, et que les estimateurs du maximum de vraisemblance le sont généralement.

4.5.2 Normalité asymptotique

Définition 4.8 : *Soient $\mu_n(\theta) = E_\theta(T_n(X))$ et $\sigma_n^2(\theta) = Var_\theta(T_n(X))$, un estimateur T est dit asymptotiquement non biaisé pour $q(\theta)$ si $(\mu_n - q(\theta))/\sigma_n(\theta)$ tend vers 0 quand n tend vers $+\infty$.*

Remarque : Si $\sigma_n(\theta) \xrightarrow[n\to+\infty]{} 0$ et $\mu_n(\theta) \xrightarrow[n\to+\infty]{} q(\theta)$, il est facile de montrer que T_n est faiblement consistant en utilisant l'inégalité de Chebychev :

$$P(|T_n(X) - \mu_n| \geq 2\epsilon) \leq (\sigma_n^2(\theta)/(4\epsilon^2))$$

et en choisissant n tel que $|q(\theta) - \mu_n| < \epsilon$.

Théorème 4.8 : *Si T_n est obtenu par substitution de fréquence, dans le cas multinomial :*

$$T_n = h(\frac{N_1}{n}, \ldots, \frac{N_k}{n}) \text{ où } q(\theta) = h(p_1(\theta), \ldots, p_k(\theta))$$

et si h admet des dérivées partielles continues, alors

$$\sqrt{n}(T_n - h(p_1, \ldots, p_k)) \xrightarrow[n\to+\infty]{\mathcal{L}(P_\theta)} \mathcal{N}(0, \sigma_h^2)$$

$$\text{où } \sigma_h^2 = \sum_{i=1}^{k} p_i \left[\frac{\partial}{\partial p_i} h(p)\right]^2 - \left[\sum_{i=1}^{k} p_i \frac{\partial}{\partial p_i} h(p)\right]^2$$

Si T_n est obtenu par substitution des moments :

$$T_n = g(\hat{m}_1, \ldots, \hat{m}_k) \text{ où } q(\theta) = g(m_1, \ldots, m_k)$$

si tous les moments m_i, $i \leq 2k$, existent et si g admet des dérivées partielles continues, alors

$$\sqrt{n}(T_n - g(m_1, \ldots, m_k)) \xrightarrow[n\to+\infty]{\mathcal{L}(P_\theta)} \mathcal{N}(0, \tau_g^2)$$

$$\text{où } \tau_g^2 = \sum_{i=1}^{2k} b_i m_i - \left[\sum_{i=1}^{k} m_i \frac{\partial}{\partial m_i} g(m)\right]^2 = Var\left(\sum_{i=1}^{k} \frac{\partial}{\partial m_i} g(m) X_1^i\right)$$

$$\text{et } b_i = \sum_{j+l=i} \frac{\partial}{\partial m_j} g(m) \frac{\partial}{\partial m_l} g(m)$$

Définition 4.9 : *Une suite d' estimateur T_n est dit consistante asymptotiquement normale pour $q(\theta)$ si $\sqrt{n}(T_n - q(\theta))$ est asymptotiquement normal (converge en loi vers une loi normale).*

Définition 4.10 : *Une suite d'estimateurs T_n est asymptotiquement efficace pour l'estimation de $q(\theta)$ si l'information de Fisher est définie, et si*

$$\sqrt{n}(T_n - q(\theta)) \xrightarrow[n\to+\infty]{\mathcal{L}(P_\theta)} \mathcal{N}\left(0, \frac{q'^2(\theta)}{I_1(\theta)}\right)$$

Remarque : Définition dans le cas vectoriel similaire, pour la forme matricielle de la borne de Cramer-Rao.

4.6 Estimateurs sans biais et estimateurs du maximum de vraisemblance

- Dans les cas de familles exponentielles de statistiques $T_i(X)$, les deux approches "maximum de vraisemblance" et "sans biais" conduisent à $T_i(X)$.
- Si le nombre d'échantillons est grand, UVMB et "maximum de vraisemblance" donnent la même réponse.
- Le maximum de vraisemblance existe plus souvent et sont plus facilement calculables.
- Quelquefois, le maximum de vraisemblance se comporte vraiment mal, pour tout n.
- Ni le maximum de vraisemblance, ni les UVMB ne se comportent de manière satisfaisante systématiquement.

Chapter 5

Intervalles de confiance - Tests

5.1 Intervalle de confiance

5.1.1 Définitions

Soit $T(X)$ un estimateur de $q(\theta)$. Quelles est la précision avec laquelle $T(X)$ estime $q(\theta)$?, est-il possible d'affirmer $q(\theta) \in [T(X) - a, T(X) + a]$?, ou plus exactement $P(q(\theta) \in [T(X) - a, T(X) + a]) \approx 1$ que l'on doit plutôt écrire $P(T(X) \in [q(\theta) - a, q(\theta) + a])$ puisque $T(X)$ est la variable aléatoire.

Exemple : Soient $X_1, \ldots, X_n$ n mesures de θ obtenues avec un instrument de mesure dont la précision est connue : les X_i sont iid, de loi $\mathcal{N}(\theta, \sigma^2)$, et σ^2 est connu. $\bar{X}$ est un estimateur de θ, et $\bar{X} \sim \mathcal{N}(\theta, \sigma^2/n)$. Donc $\sqrt{n}(\bar{X} - \theta)/\sigma) \sim \mathcal{N}(0,1)$ et :

$$P(|\bar{X} - \theta| > a) = \phi\left(-\frac{a\sqrt{n}}{\sigma}\right) + \left(1 - \phi\left(\frac{a\sqrt{n}}{\sigma}\right)\right) = 2\phi\left(-\frac{a\sqrt{n}}{\sigma}\right)$$

où ϕ est la fonction de répartition de la loi $\mathcal{N}(0,1)$.

On peut donc choisir a de manière que $P(\bar{X} - a \leq \theta \leq \bar{X} + a) = 1 - P(|\bar{X} - \theta| > a) = 1 - \alpha$ où α est fixé dans $[0,1]$, par exemple $\alpha = 5\%, 1\%$.

Le problème consiste alors à trouver deux statistiques T^- et T^+ telles que $P(T^- \leq T^+) = 1 - \alpha$.

Remarque : Dans le cas discret, de telles statistiques peuvent ne pas exister, il faut alors prévoir une définition plus large du niveau.

Définition 5.1 : *Un intervalle $[T^-, T^+]$ formé de deux statistiques T^- et T^+ telles que $T^- \leq T^+$ est un intervalle de confiance au niveau $1 - \alpha$ pour $q(\theta)$ si*

$$\forall \theta \in \Theta, P_\theta(T^- \leq q(\theta) \leq T^+) \geq 1 - \alpha$$

Le coefficient de sécurité (ou niveau de confiance) du test est la borne inférieure des probabilités $P_\theta(T^- \leq q(\theta) \leq T^+)$ pour $\theta \in \Theta$.

Remarques :

- En général, $\alpha = 5\%, 1\%, 0.1\%$.
- Dans le cas de lois symétriques pour $T(X)$, on considère le plus souvent des intervalles de confiance de la forme $[T(X) - a, T(X) + a]$.
- En général, il n'y a pas d'intervalle de confiance de seuil fixé et de longueur uniformément (en θ) minimale.

Définition 5.2 : *Un intervalle de confiance est dit sans biais si*

$$\forall \theta, \theta' \in \Theta, P_\theta(T^- \leq q(\theta) \leq T^+) \geq P_\theta(T^- \leq q(\theta') \leq T^+)$$

C'est à dire que l'intervalle considéré a au moins autant de chance de contenir la vraie valeur de $q(\theta)$ que toute autre valeurs $q(\theta')$.

Définition 5.3 : *Une statistique $T^-(X)$ est appelée borne inférieure de confiance au niveau $1-\alpha$ pour $q(\theta)$ si $\forall \theta \in \Theta, P(T^-(X) \leq q(\theta)) \geq 1-\alpha$. De même, une statistique $T^+(X)$ est appelée borne supérieure de confiance au niveau $1-\alpha$ pour $q(\theta)$ si $\forall \theta \in \Theta, P(q(\theta) \leq T^+(X)) \geq 1-\alpha$*

5.1.2 Exemples

1) Intervalle de confiance pour μ, lorsque $X_1, \ldots, X_n$ est un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, σ^2 connu.
$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ et donc on trouve dans les tables une valeur de a telle que

$$P(-a \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq a) = 1-\alpha \text{ et } P(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < -a) = P(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > a) = \alpha/2$$

$[\bar{X} - \frac{a\sigma}{\sqrt{n}}, \bar{X} + \frac{a\sigma}{\sqrt{n}}]$ est un intervalle de confiance au niveau $1-\alpha$ pour μ.
$\bar{X} - \frac{a\sigma}{\sqrt{n}}$ est la borne inférieur de confiance au niveau $1-\alpha/2$.
$\bar{X} + \frac{a\sigma}{\sqrt{n}}$ est la borne supérieur de confiance au niveau $1-\alpha/2$.

2) Intervalle de confiance pour μ, lorsque $X_1, \ldots, X_n$ est un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, σ^2 inconnu.
Il faut cette fois commencer par estimer σ^2, par exemple par

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

On sait alors que $\sqrt{n}(\bar{X}-\mu)$ et $(n-1)S^2/\sigma^2$ sont indépendantes et que $\sqrt{n}(\bar{X}-\mu) \sim \mathcal{N}(0, \sigma^2)$ et $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$. Donc $\dfrac{\sqrt{n}(\bar{X}-\mu)}{S} \sim T_{n-1}$ loi de Student à $n-1$ degrés de liberté.

On trouve dans les tables une valeur de t (par approximation avec la loi normale si n est trop grand) telle que

$$P(-t \leq \frac{\sqrt{n}(\bar{X}-\mu)}{S} \leq t) = 1-\alpha \text{ et } P(\frac{\sqrt{n}(\bar{X}-\mu)}{S} < -t) = P(\frac{\sqrt{n}(\bar{X}-\mu)}{S} > t) = \alpha/2$$

$[\bar{X} - \frac{tS}{\sqrt{n}}, \bar{X} + \frac{tS}{\sqrt{n}}]$ est un intervalle de confiance au niveau $1-\alpha$ pour μ.
$\bar{X} - \frac{tS}{\sqrt{n}}$ est la borne inférieur de confiance au niveau $1-\alpha/2$.
$\bar{X} + \frac{tS}{\sqrt{n}}$ est la borne supérieur de confiance au niveau $1-\alpha/2$.

3) Intervalle de confiance pour σ^2, lorsque $X_1, \ldots, X_n$ est un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, μ connu.
$\tilde{\sigma}^2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2}$ est à la fois l'estimateur du maximum de vraisemblance et l'estimateur UVMB de σ^2. De plus $n\tilde{\sigma}^2/\sigma^2 \sim \chi^2_n$. Soient α_1, α_2 tels que $\alpha_1 + \alpha_2 = \alpha$, on choisi a et b tels que :

$$P(\frac{n\tilde{\sigma}^2}{\sigma^2} \geq a) = 1-\alpha_2$$

$$P(\frac{n\tilde{\sigma}^2}{\sigma^2} \leq b) = 1-\alpha_1$$

$$\text{d'où } P(a \leq \frac{n\tilde{\sigma}^2}{\sigma^2} \leq b) = 1-\alpha$$

$[\frac{n\tilde{\sigma}^2}{b}, \frac{n\tilde{\sigma}^2}{a}]$ est un intervalle de confiance au niveau $1-\alpha$ pour σ^2.
On peut montrer qu'il existe un seul couple (α_1, α_2) qui minimise uniformément en σ^2 la longueur de ces intervalles. $\alpha_1 = \alpha_2 = \alpha/2$ sont proches de ces valeurs optimales.

4) Intervalle de confiance pour θ, lorsque $X \sim \mathcal{B}(n,\theta)$. Lorsque n est assez petit, les tables permettent, en fonction de la valeur observée x, d'obtenir directement un intervalle de confiance numérique, généralement à un niveau de sécurité de 95%.

5) Intervalle de confiance pour θ, lorsque $Y \sim \mathcal{B}(n,\theta)$, lorsque n est assez grand. On peut alors approximer la loi de Y par $\mathcal{N}(n\theta, n\theta(1-\theta))$, d'où, pour $X = Y/n$:

$$\frac{\sqrt{n}(X-\theta)}{\sqrt{\theta(1-\theta)}} \sim \mathcal{N}(0,1)$$

et on trouve dans les tables la valeur de a telle que

$$P\left(\left|\frac{\sqrt{n}(X-\theta)}{\sqrt{\theta(1-\theta)}}\right| < a\right) = 1-\alpha$$

Pour obtenir l'intervalle de confiance, il faut résoudre l'inégalité $(X-\theta)^2 \leq a^2\theta(1-\theta)/n$ en θ.

$$(X-\theta)^2 \leq a^2\theta(1-\theta)/n \Longleftrightarrow (1+a^2/n)\theta^2 - (2X + a^2/n)\theta + X^2 \leq 0$$

or $\Delta = (2X^2+a^2/n)^2 - 4X^2(1+a^2/n) = a^2/n(a^2/n+4X(1-X)) > 0$, l'équation admet donc deux solutions distinctes, et l'intervalle de confiance est défini par ces deux solutions :

$$\left[\frac{nX + a^2/2 - a\sqrt{a^2/4 + nX - nX^2}}{n+a^2}, \frac{nX + a^2/2 + a\sqrt{a^2/4 + nX - nX^2}}{n+a^2}\right]$$

La longueur de cet intervalle est $L = \dfrac{2a\sqrt{a^2/4 + nX - nX^2}}{n+a^2}$. On peut vérifier que $L \geq a/\sqrt{n+a^2}$, car $a^2/4 + nX - nX^2 \leq (a^2+n)/4$, et ainsi calculer la valeur minimale de n permettant d'obtenir des intervalles de longueur inférieure à une constante donnée.

6) Intervalle de confiance pour θ, lorsque $X \sim \mathcal{P}(\theta)$. Certaines tables permettent, en fonction de la valeur observée x, d'obtenir directement un intervalle de confiance numérique, généralement à un niveau de sécurité de 95%.

7) Intervalle de confiance pour θ, lorsque $X_1, \ldots, X_n$ est un échantillon de loi $\mathcal{P}(\theta)$. En supposant n assez grand, $\bar{X}$ suit approximativement une loi $\mathcal{N}(\theta, \theta/n)$ et donc on choisi a tel que

$$P(-a \leq \frac{\bar{X}-\theta}{\sqrt{\theta/n}} \leq a) = 1-\alpha$$

Pour obtenir l'intervalle de confiance, il faut résoudre l'inégalité $(\bar{X}-\theta)^2 \leq a^2\theta/n$ en θ.

$$(\bar{X}-\theta)^2 \leq a^2\theta/n \Longleftrightarrow \theta^2 - (2\bar{X} + a^2/n)\theta + X^2 \leq 0$$

or $\Delta = (2\bar{X}^2 + a^2/n)^2 - 4\bar{X}^2 = (a^2/n)(a^2/n + 4\bar{X}) > 0$, l'équation admet donc deux solutions distinctes, et l'intervalle de confiance est défini par ces deux solutions :

$$\left[\bar{X} + a^2/(2n) - \sqrt{a^4/(4n^2) + a^2\bar{X}/n}, \bar{X} + a^2/(2n) + \sqrt{a^4/(4n^2) + a^2\bar{X}/n}\right]$$

5.1.3 Régions de confiance

Soit $q(\theta) = (q_1(\theta), \ldots, q_r(\theta))$, on construit alors deux statistiques vectorielles $T^-(X) = (T_1^-(X), \ldots, T_r^-(X))$ et $T^+(X) = (T_1^+(X), \ldots, T_r^+(X))$ puis $I(X) = \{q(\theta)/\forall j, T_j^-(X) \leq q_j(\theta) \leq T_j^+(X)\}$ est une région de confiance au niveau $1-\alpha$ $P(q(\theta) \in I(X)) \geq 1-\alpha$.

Remarque : Si les couples $(T_1^-, T_1^+), \ldots, (T_r^-, T_r^+)$ sont indépendantes, le pavé $[T_1^-, T_1^+] \times \ldots \times [T_r^-, T_r^+]$ est une région de confiance au niveau $\prod_{i=1}^r (1-\alpha_i)$.

Exemple : Pour un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, soient $I_1(X) = [\bar{X} - tS/\sqrt{n}, \bar{X} + tS/\sqrt{n}]$ et $I_2(X) = [(n-1)S^2/b, (n-1)S^2/a]$ des intervalles de confiance au niveau $1-\alpha/2$ pour μ et σ^2, alors $I_1(X) \times I_2(X)$ est une région de confiance pour (μ, σ^2) avec un niveau approximatif de $1-\alpha$.

5.2 Tests

5.2.1 Définitions

Le but est de répondre à des questions du type : un médicament est-il efficace ?, un objet manufacturé est-il trop souvent défectueux ?

Plus précisément, on observe $X \sim P_\theta$, $\theta \in \Theta$, et on veut déterminer parmi deux hypothèses $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta_1$ celle qui est vraie. On va alors tester $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$ et décider d'accepter ou de rejeter H_0, le rejet de H_0 impliquant l'acceptation de H_1.

Remarques :

- La formulation d'un test n'est pas symétrique en H_0 et H_1.
- Il n'est pas nécessaire que $\Theta = \Theta_1 \cup \Theta_2$, mais il faut que $\Theta_1 \cap \Theta_2 = \emptyset$.
- Les hypothèses H_0 et H_1 peuvent être composites.

Exemple : Soit θ la probabilité qu'un patient prenant un médicament donné guérisse, et soit S le nombre de guérisons observées parmi n patients prenant ce médicament. Notons θ_0 la probabilité de guérison spontanée.

$S \sim \mathcal{B}(n, \theta)$, $\Theta = [0,1]$, on veut tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta > \theta_0$. Si S est grand par rapport à n, on a envie de dire que $\theta > \theta_0$...

Définition 5.4 : *Un test de l'hypothèse Θ_0 contre l'hypothèse Θ_1 de niveau α est la donnée d'un ensemble aléatoire R qui dépend uniquement de X tel que*

$$\sup_{\theta \in \Theta_0} P_\theta(X \in R(X)) \leq \alpha$$

$R(X)$ s'appelle le domaine de rejet, si $X \in R(X)$, on rejette Θ_0, on l'accepte sinon.

Lien avec la théorie de la décision : Un test est une décision $d : (\Omega, \mathcal{A}) \longrightarrow \{0,1\}$ mesurable telle que

$$\begin{aligned} d(\omega) &= 1 \text{ si } \omega \in R \\ &= 0 \text{ si } \omega \in R^c \end{aligned}$$

où $\Omega = R \cup R^c$, R^c et $R \cap R^c = \emptyset$: désigne le complémentaire de R dans Ω.

La fonction de perte est : $L(\theta, d) = \mathbf{1}_{\Theta_0} d + \mathbf{1}_{\Theta_1}(1-d)$, d'où le risque :

$$R(\theta, d) = \underbrace{\mathbf{1}_{\Theta_0} P_\theta(R)}_{\substack{\text{erreur de} \\ \text{1-ere espèce}}} + \underbrace{\mathbf{1}_{\Theta_1}(1 - P_\theta(R))}_{\substack{\text{erreur de} \\ \text{2-ème espèce}}}$$

La fonction de puissance du test est $\beta(\theta) = P_\theta(R), \theta \in \Theta$, on considérera aussi la puissance du test $\beta_0 = \inf_{\theta \in \Theta_1} P_\theta(R)$.

Définition 5.5 : *Un test est dit sans biais si $\beta(\theta) \geq \alpha, \forall \theta \in \Theta_1$, α étant le niveau du test.*

Problème : Parmi les tests de niveau α, on s'intéresse plus particulièrement aux tests qui maximise $\beta(\theta), \theta \in \Theta_1$.

Définition 5.6 : *Un test est dit unilatère si $R(X)$ est du type $\{T(X) > a\}$, il est bilatère si $R^c(X)$ est du type $\{a < T(X) < b\}$.*

5.2.2 Exemples

1) Soit $X_1, \ldots, X_n$ un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, σ^2 connu, on veut tester $H_0 : \mu = 0$ contre $H_1 : \mu = 1$.

$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ va être utilisé comme statistique du test, on construit alors $R(X) = \{\bar{X} > a\}$ tel que pour $\mu = 0, P_\mu(R(X)) = \alpha$, α étant un niveau fixé.

$$P_{\mu=0}(R(X)) = P_{\mu=0}(\bar{X} > a) = P_{\mu=0}\left(\frac{\bar{X}}{\sigma/\sqrt{n}} > \frac{a}{\sigma/\sqrt{n}}\right) = P\left(U > \frac{a}{\sigma/\sqrt{n}}\right)$$

où $U \sim \mathcal{N}(0,1)$, $\sqrt{n}a/\sigma = u$ est alors lu dans la table en fonction de α d'après la condition $P(U > \sqrt{n}a/\sigma) = \alpha$, et on remplace : $a = \sigma u/\sqrt{n}$ et $R(X) = \{\bar{X} > \sigma u/\sqrt{n}\}$.

Pour un $\bar{x}$ observé, on accepte H_0 si $\bar{x} \leq \sigma u/\sqrt{n}$, on rejette H_0 sinon. On peut aussi calculer la puissance β_0 de ce test : $\beta_0 = \inf_{\mu > 0} P_\mu(R)$ et en faisant tendre μ vers 0, on obtient $\beta_0 = \alpha$.

2) Soit $X_1, \ldots, X_n$ un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, σ^2 connu, on veut tester $H_0 : \mu \leq 0$ contre $H_1 : \mu > 0$.

$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ va être utilisé comme statistique du test, on construit alors $R(X) = \{\bar{X} > a\}$ tel que $\forall \mu \leq 0, P_\mu(R(X)) \leq \alpha$, α étant un niveau fixé.

$$\forall \mu \leq 0, P_\mu(R(X)) = P_\mu(\bar{X} > a) = P_\mu\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{a - \mu}{\sigma/\sqrt{n}}\right)$$

$$\text{or } \forall \mu \leq 0, P_\mu\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{a - \mu}{\sigma/\sqrt{n}}\right) \leq P\left(U > \frac{a}{\sigma/\sqrt{n}}\right)$$

où $U \sim \mathcal{N}(0,1)$, $\sqrt{n}a/\sigma = u$ est alors lu dans la table en fonction de α d'après la condition $P(U > \sqrt{n}a/\sigma) = \alpha$, et on remplace : $a = \sigma u/\sqrt{n}$ et $R(X) = \{\bar{X} > \sigma u/\sqrt{n}\}$.

Pour un $\bar{x}$ observé, on accepte H_0 si $\bar{x} \leq \sigma u/\sqrt{n}$, on rejette H_0 sinon. On peut aussi calculer la puissance β_0 de ce test : $\beta_0 = \inf_{\mu > 0} P_\mu(R)$ et en faisant tendre μ vers 0, on obtient $\beta_0 = \alpha$.

3) Soit $X_1, \ldots, X_n$ un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, σ^2 connu, on veut tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$.

$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ va de nouveau être utilisé comme statistique du test, on construit $R^c(X) = \{\mu_0 - a \leq \bar{X} \leq \mu_0 + a\}$ tel que $P_{\mu_0}(R^c(X)) = 1 - \alpha$, α étant un niveau fixé.

$$P_{\mu_0}(R^c(X)) = P_{\mu_0}(\mu_0 - a \leq \bar{X} \leq \mu_0 + a) = P_{\mu_0}\left(\frac{-a}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq \frac{a}{\sigma/\sqrt{n}}\right)$$

$$= P\left(\frac{-a}{\sigma/\sqrt{n}} \leq U \leq \frac{a}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

$\sqrt{n}a/\sigma = u$ est lu dans la table en fonction de α, et on remplace : $a = \sigma u/\sqrt{n}$ et $R^c(X) = \{-\sigma u/\sqrt{n} \leq \bar{X} \leq \sigma u/\sqrt{n}\}$.

Pour un $\bar{x}$ observé, on accepte H_0 si $-\sigma u/\sqrt{n} \leq \bar{x} \leq \sigma u/\sqrt{n}$, on rejette H_0 sinon. On peut aussi calculer la puissance β_0 de ce test : $\beta_0 = \inf_{\mu \neq \mu_0} P_\mu(R)$ et en faisant tendre μ vers μ_0, on obtient $\beta_0 = \alpha$.

Remarque : L'intervalle d'acceptation que l'on obtient ainsi est exactement l'intervalle de confiance que nous avons construit précédemment.

5.2.3 Test uniformément le plus puissant (optimal)

On peut étendre la définition d'un test précédemment en considérant une décision $\phi : (\Omega, \mathcal{A}) \longrightarrow [0,1]$, où $\phi(\omega)$ représente la probabilité de rejeter H_0 lorsque l'on observe ω. ϕ est appelée la fonction critique. La fonction de perte est alors : $L(\theta, d) = \mathbf{1}_{\Theta_0}\phi + \mathbf{1}_{\Theta_1}(1-\phi)$, et le risque : $R(\theta, d) = \mathbf{1}_{\Theta_0}E_\theta(\phi) + \mathbf{1}_{\Theta_1}(1 - E_\theta(\phi))$. La fonction de puissance du test est donc $\beta(\theta) = E_\theta(\phi), \theta \in \Theta$.

Définition 5.7 : *Si ϕ et ϕ' sont les fonctions critiques de deux tests de niveau α, on dit que ϕ est plus puissant que ϕ' si*

$$\forall \theta \in \Theta_1, \beta(\theta) = E_\theta(\phi) \geq E_\theta(\phi') = \beta'(\theta).$$

Un test est dit uniformément le plus puissant (UPP) parmi tous les tests de niveau inférieur à α si

$$\forall \theta \in \Theta_1, \forall \phi', \beta(\theta) = E_\theta(\phi) \geq E_\theta(\phi') = \beta'(\theta).$$

On dit aussi qu'un tel test est optimal.

Définition 5.8 : *Un test est dit simple si Θ_0 est réduit à un point.*

Définition 5.9 : *Une statistique de test $T(X)$ est dite optimale si les tests UPP peuvent s'écrire en fonction de T.*

Théorème 5.1 : *Soit $(\Omega, \mathcal{A}, P_\theta, \theta \in \Theta)$ un modèle statistique où $\Theta = \{\theta_0, \theta_1\}$. On suppose que P_{θ_0} et P_{θ_1} admettent pour densités f_{θ_0} et f_{θ_1} par rapport à une mesure μ, et on teste $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$, alors :*

1. *$\forall \phi_\lambda$ définis par :*

$$\begin{array}{lll} \phi_\lambda(\omega) & = 1 & si \quad f_{\theta_1}(\omega) > \lambda f_{\theta_0}(\omega) \\ & = 0 & si \quad f_{\theta_1}(\omega) < \lambda f_{\theta_0}(\omega) \\ & = \gamma(\omega) & si \quad f_{\theta_1}(\omega) = \lambda f_{\theta_0}(\omega) \end{array}$$

   où $\gamma(\omega) \in [0,1], \lambda \geq 0$.

   ϕ_λ est la fonction critique d'un test uniformément le plus puissant à son niveau pour H_0 contre H_1.

   En particulier, si $\lambda = +\infty$, et avec la convention $0 \times +\infty = 0$, le test devient :

$$\begin{array}{ll} \phi_\lambda(\omega) & = 1 \; si \quad f_{\theta_0}(\omega) = 0 \\ & = 0 \; si \quad f_{\theta_0}(\omega) > 0 \end{array}$$

   et il est le plus puissant de niveau 0.

2. *$\forall \alpha \in [0,1], \exists \alpha \geq 0$ et γ constant tel que ϕ_λ avec $\gamma(\omega) = \gamma$ soit de niveau α.*

3. *Si ϕ' est le test uniformément le plus puissant de seuil α pour tester H_0 contre H_1, il est de la forme ϕ_λ.*

Remarque : Ces tests sont appelés des tests de Neyman Pearson (NP).

Dém :

1. Soient ϕ_λ et ϕ' tel que $0 \leq \phi'(\omega) \leq 1, \forall \omega \in \Omega$ et $E_{\theta_0}(\phi') \leq E_{\theta_0}(\phi_\lambda)$

$$\Longleftrightarrow \beta_{\phi'}(\theta_0) \leq \beta_{\phi_\lambda}(\theta_0) \text{ (niveau).}$$

$$\text{or } \left\{ \begin{array}{l} f_{\theta_1}(\omega) > \lambda f_{\theta_0}(\omega) \Longrightarrow \phi_\lambda(\omega) = 1 \Longrightarrow \phi_\lambda(\omega) - \phi'(\omega) \geq 0 \\ f_{\theta_1}(\omega) < \lambda f_{\theta_0}(\omega) \Longrightarrow \phi_\lambda(\omega) = 0 \Longrightarrow \phi_\lambda(\omega) - \phi'(\omega) \leq 0 \end{array} \right.$$

$$\text{D'où } \int_\Omega (\phi_\lambda(\omega) - \phi'(\omega))(f_{\theta_1}(\omega) - \lambda f_{\theta_0}(\omega)) d\mu(\omega) \geq 0$$

$$\text{et } E_{\theta_1}(\phi_\lambda) - E_{\theta_1}(\phi') \geq \lambda(E_{\theta_0}(\phi_\lambda) - E_{\theta_0}(\phi')) \geq 0 \text{ car } E_{\theta_0}(\phi') \leq E_{\theta_0}(\phi_\lambda)$$

Finalement, $E_{\theta_1}(\phi') \leq E_{\theta_1}(\phi_\lambda)$, cqfd.

En particulier, si ϕ_λ et ϕ' sont de niveau 0, $\phi'(\omega) = 0$ sur $\{\omega / f_0(\omega) > 0\}$.

2. Si $\alpha = 0$, alors $\lambda = +\infty$.

   Si $0 < \alpha \leq 1$, $E_{\theta_0}(\phi_\lambda) \begin{array}{l} = P_{\theta_0}(f_{\theta_1} > \lambda f_{\theta_0}) + \gamma P_{\theta_0}(f_{\theta_1} = \lambda f_{\theta_0}) \\ = P_{\theta_0}(f_{\theta_1}/f_{\theta_0} > \lambda) + \gamma P_{\theta_0}(f_{\theta_1} = \lambda f_{\theta_0}) \end{array}$

   Par définition, si $f_{\theta_0}(\omega) \neq 0, f_{\theta_1}(\omega)/f_{\theta_0}(\omega)$ est défini, sinon, on pose $f_{\theta_1}(\omega)/f_{\theta_0}(\omega) = +\infty$

   La fonction $u : \lambda \longrightarrow P_{\theta_0}(f_{\theta_1}/f_{\theta_0} > \lambda)$ est décroissante de 1 à 0, et continue à droite, il y a donc deux cas

   - $\exists \lambda_0 / P_{\theta_0}(f_{\theta_1}/f_{\theta_0} > \lambda) = \alpha$, alors$\gamma = 0$.
   - sinon $\exists \lambda_0$ tel que $u(\lambda_0^-) \geq \alpha \geq u(\lambda_0)$ et on pose alors

   $$\gamma = (\alpha - u(\lambda_0))/(u(\lambda_0^-) - u(\lambda_0)).$$

   Remarque : λ_0 est unique sauf sur un ensemble de mesure nulle pour chaque loi.

3. Soit ϕ' un test uniformément le plus puissant au niveau α. D'après **2.**, il existe une fonction ϕ_λ de même niveau et d'après **1.**, ϕ_λ est le plus puissant, donc

   $$\left\{ \begin{array}{ll} E_{\theta_0}(\phi_\lambda) = E_{\theta_0}(\phi') & \text{(niveau)} \\ E_{\theta_1}(\phi_\lambda) = E_{\theta_1}(\phi') & \text{(puissance)} \end{array} \right.$$

   $\Longrightarrow \int_\Omega (\phi_\lambda(\omega) - \phi'(\omega))(f_{\theta_1}(\omega) - \lambda f_{\theta_0}(\omega)) d\mu(\omega) = 0$
   $\Longrightarrow \phi_\lambda \equiv \phi'$ sur $\{\omega / f_{\theta_1}(\omega) \neq \lambda f_{\theta_0}(\omega)\}, \mu$ *pp*.
   $\Longrightarrow \phi_\lambda \equiv \phi', \mu$ *pp*.

   si $E_{\theta_0}(\phi') < \alpha$ et $E_{\theta_1}(\phi') < 1$, on peut augmenter la puissance ou diminuer le niveau de ce test, donc $E_{\theta_0}(\phi') = \alpha$ ou $E_{\theta_1}(\phi') = 1$.

□

Corollaire 1 : *Soit β la puissance du test de niveau α le plus puissant de H_0 contre H_1 ($0 < \alpha < 1$) alors $\alpha < \beta$ sauf si $f_0 = f_1$ auquel cas $\alpha = \beta$.*

Dém : Le test défini par $\phi(\omega) \equiv \alpha$ a pour puissance $E_{\theta_1}(\phi) = \int_\Omega \alpha f_1(\omega) d\mu(\omega) = \alpha$, donc la puissance β est supérieure ou égale à α et si $\alpha = \beta < 1$, le test défini par ϕ est le plus puissant d'où $f_0(\omega) = f_1(\omega), \mu$ *pp* et $P_{\theta_0} = P_{\theta_1}$ d'après la définition des tests NP. □

Définition 5.10 : *Une famille de densité $f_\theta, \theta \in \Theta$ intervalle de $\mathbb{R}$, a un rapport de vraisemblance monotone s'il existe une statistique réelle $T(X)$ telle que pour tout couple (θ', θ''), le rapport $f_{\theta''}/f_{\theta'}$ soit, sur son ensemble de définition, une fonction strictement monotone de T.*

Remarque : On peut supposer que le rapport de vraisemblance est une fonction strictement croissante de T, sinon, on échange t en $-T$, et étendre sa définition à $\mathbb{R}$ en posant $f_{\theta''}/f_{\theta'} = +\infty$ si $f_{\theta''} \neq 0$ et $f_{\theta'} = 0$.

Théorème 5.2 : *Soit $(\Omega, \mathcal{A}, P_\theta, \theta \in \Theta)$ un modèle statistique où Θ est un intervalle de $\mathbb{R}$. On suppose que, pour tout θ, P_θ admet une densité f_θ par rapport à une mesure μ, et que le rapport de vraisemblance est monotone en T, alors $\forall \theta_0 \in \Theta$, pour tester $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$, il existe, $\forall \alpha \in [0, 1]$, un test UPP de niveau α de la forme :*

$$\begin{array}{llll} \phi(\omega) & = 1 & si & T(\omega) > C \\ & = 0 & si & T(\omega) < C \\ & = \gamma & si & T(\omega) = C \end{array}$$

où C et γ sont des constantes déterminées par α.

Dém : On commence par construire Φ_λ UPP pour $\theta = \theta_0$ contre $\theta = \theta_1$, θ_1 fixé $> \theta_0$.

$$\begin{aligned}\Phi_\lambda(\omega) &= \mathbf{1}_{f_1 > \lambda f_0} + \gamma \mathbf{1}_{f_1 = \lambda f_0} \\ &= \mathbf{1}_{f_1/f_0 > \lambda} + \gamma \mathbf{1}_{f_1/f_0 = \lambda} \\ &= \mathbf{1}_{T > c} + \gamma \mathbf{1}_{T = c}\end{aligned}$$

car f_1/f_0 est une fonction croissante de T.

Le même test (γ et c fixés) peut être utilisé pour $\theta = \theta'$ contre $\theta = \theta''$ avec $\theta' < \theta''$. Le niveau de ce test est alors $E_{\theta'}(\Phi)$ et sa puissance est $E_{\theta''}(\Phi)$. C'est un test UPP à ce niveau et on a montré que $E_{\theta'}(\Phi) \leq E_{\theta''}(\Phi)$. D'où, $\forall \theta', \theta''$ tels que $\theta' < \theta''$, alors $E_{\theta'}(\Phi) \leq E_{\theta''}(\Phi)$.

Le test de $\theta \leq \theta_0$ contre $\theta > \theta_0$ défini par Φ a alors pour niveau : $\sup_{\theta \leq \theta_0} E_\theta(\Phi) = E_{\theta_0}(\Phi)$ et pour puissance : $\inf_{\theta > \theta_0} E_\theta(\Phi)$.

Soit maintenant ψ un autre test de $\theta \leq \theta_0$ contre $\theta > \theta_1$, avec $\theta_1 > \theta_0$, au niveau $\alpha = E_{\theta_0}(\Phi)$, on veut montrer que $E_{\theta_1}(\psi) \leq E_{\theta_1}(\Phi)$.

Or, ψ est aussi un test de $\theta = \theta_0$ contre $\theta = \theta_1$, et dans ce cas, son niveau est toujours inférieur ou égal à α car il est inférieur à $\sup_{\theta \leq \theta_0} E_\theta(\psi) = \alpha$.

Or Φ est UPP pour ce test et donc $E_{\theta_1}(\psi) \leq E_{\theta_1}(\Phi)$, $\forall \theta_1 > \theta$, cqfd. $\square$

Remarque : Pour les tests bilatères, il n'y a pas de résultats généraux sur les tests UPP, et on va voir que même pour les familles exponentielles, parfois il n'existe pas de test UPP.

5.2.4 Exemples

1) Soit $X_1, \ldots, X_n$ un échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, σ^2 connu, on veut tester $H_0 : \mu = 0$ contre $H_1 : \mu = v$.

Le rapport de vraisemblance est

$$L(x_1, \ldots, x_n, 0, v) = \exp \frac{1}{2\sigma^2}\left[\sum_{i=1}^n x_i^2 - \sum_{i=1}^n (x_i - v)^2\right] = \exp\left(\frac{v}{\sigma^2}\left[\sum_{i=1}^n x_i\right] - \frac{nv^2}{2\sigma^2}\right).$$

$\phi_\lambda(\omega) = \mathbf{1}_{f_1/f_0 > \lambda}(\omega)$ est la fonction critique d'un test le plus puissant pour son niveau, donc en fait, pour tout fonction F strictement croissante, la fonction $\psi_\lambda(\omega) = \mathbf{1}_{F(f_1/f_0) > \lambda}(\omega)$ est aussi la fonction critique d'un test le plus puissant pour son niveau.

Choisissons ici : $F(f_1/f_0) = \sqrt{n}\bar{X}/\sigma = (\sigma/v\sqrt{n})\left[\ln(f_1/f_0) + (nv^2/2\sigma^2)\right]$, la région de rejet est alors : $\{\sqrt{n}\bar{X}/\sigma > \lambda\}$ qui correspond à un test UPP.

2) Soit $X_1, \ldots, X_n$ un échantillon de loi $P_\theta(x) = 1/\theta$ si $x \in \{1, \ldots, \theta\}$, et 0 sinon, $\theta \in \mathbb{N}^*$. On veut tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$.

Le rapport de vraisemblance :

$$\begin{aligned}L(x_1, \ldots, x_n, \theta_0, \theta_1) &= \infty \text{ si} \quad \theta_0 < \max(x_1, \ldots, x_n) \leq \theta_1 \\ &= \left(\tfrac{\theta_0}{\theta_1}\right)^n \text{ si} \quad 1 < \max(x_1, \ldots, x_n) \leq \theta_0\end{aligned}$$

L est monotone en $M_n = \max(X_1, \ldots, X_n)$. la région de rejet : $R = \{M_n > m\}$ correspond à un test UPP de niveau $P_\theta(M_n \geq m) = 1 - \left(\frac{j-1}{\theta_0}\right)^n$.

5.2.5 Lemme de Neyman-Pearson

Théorème 5.3 (Lemme de Neyman-Pearson) : *Soit δ_k une fonction critique : $\delta_k = \mathbf{1}_{f_{\theta_1}/f_{\theta_0} \geq k}$ pour $\theta = \theta_0$ contre $\theta = \theta_1$. Soit Φ la fonction critique d'un autre test tel que $E_{\theta_0}(\Phi) \leq E_{\theta_0}(\delta_k)$. Alors, $E_{\theta_1}(\Phi) \leq E_{\theta_1}(\delta_k)$, c'est-à-dire $\beta_\Phi(\theta_1) \leq \beta_{\delta_k}(\theta_1)$.*

Dém : $0 \leq \Phi \leq 1$, on étudie le cas "à densité".

- si$f_{\theta_1}/f_{\theta_0} < k \implies \delta_k = 0 \implies \Phi(x)(f_{\theta_1}(x) - kf_{\theta_0}(x)) \leq 0$
- si$f_{\theta_1}/f_{\theta_0} \geq k \implies \delta_k = 1 \implies \Phi(x)(f_{\theta_1}(x) - kf_{\theta_0}(x)) \leq (f_{\theta_1}(x) - kf_{\theta_0}(x))$

$$\begin{array}{lll} \text{D'où} & \Phi(x)(f_{\theta_1}(x)-kf_{\theta_0}(x)) & \leq \delta_k(x)(f_{\theta_1}(x)-kf_{\theta_0}(x)) \\ & \int \Phi(x)(f_{\theta_1}(x)-kf_{\theta_0}(x)) & \leq \int \delta_k(x)(f_{\theta_1}(x)-kf_{\theta_0}(x)) \\ & E_{\theta_1}(\Phi(X))-E_{\theta_1}(\delta_k(X)) & \leq k\left[E_{\theta_0}(\Phi(X))-E_{\theta_0}(\delta_k(X))\right] \end{array}$$

Or $k \geq 0$ et $E_{\theta_0}(\Phi(X)) \leq E_{\theta_0}(\delta_k(X))$ d'où $E_{\theta_1}(\Phi(X)) \leq E_{\theta_1}(\delta_k(X))$. □

Remarque : On peut généraliser aux familles ayant un rapport de vraisemblance monotone et aux tests $\theta \leq \theta_0$ contre $\theta > \theta_0$.

Théorème 5.4 : *Un test UPP est sans biais (quelles que soient les hypothèses).*
Dém : Soit Φ ce test, de niveau α, et $\Phi' \equiv \alpha$. Φ' est un autre test de niveau égal à α. Puisque Φ est UPP : $\forall \theta \in \Theta_1$, $E_\theta(\Phi) \geq E_\theta(\Phi') = \alpha$. □

5.2.6 Familles exponentielles

Théorème 5.5 : *Si X suit une loi exponentielle à un paramètre de densité ou de loi discrète*

$$f_\theta(x) = \exp\left[c(\theta)T(X) + d(\theta) + S(X)\right]\mathbf{1}_A(x)$$

où $\theta \in \mathbb{R}$, et c est une fonction strictement croissante en θ et $T(X)$ est non dégénérée, alors : $\forall \theta_0$

a) ⋆ *$T(X)$ est une statistique de test optimale pour $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$.*

⋆ *Il existe un test UPP de niveau α défini par :*

$$\begin{array}{llll} \phi(\omega) & = 1 & si & T(\omega) > C \\ & = 0 & si & T(\omega) < C \\ & = \gamma & si & T(\omega) = C \end{array}$$

C et γ sont des constantes déterminées par le niveau.

⋆ *La puissance du test UPP de niveau α est croissante en θ.*

b) ⋆ *$T(X)$ est une statistique de test optimale pour $H_0 : \theta \leq \theta_1$ ou $\theta \geq \theta_2$ contre $H_1 : \theta_1 < \theta < \theta_2$, où $\theta_1 < \theta_2$.*

⋆ *Il existe un test UPP de niveau α défini par :*

$$\begin{array}{llll} \phi(\omega) & = 1 & si & C_1 < T(\omega) < C_2 \\ & = 0 & si & T(\omega) < C_1 \text{ ou } T(\omega) > C_2 \\ & = \gamma_i & si & T(\omega) = C_i \end{array}$$

C_1, C_2, γ_1 et γ_2 sont des constantes déterminées par le niveau et la puissance : $E_{\theta_1}(\phi) = E_{\theta_0}(\phi) = \alpha$.

Exemples :

1) Soit $X_1, \ldots, X_n$ un échantillon de loi $\Gamma(p, 1/\theta)$, on veut tester :

1. $H_0 : \theta \leq \theta_0$ contre $H_0 : \theta > \theta_0$
2. $H_0 : \theta \geq \theta_0$ contre $H_0 : \theta < \theta_0$

1. $f(\theta, x) = \exp\left[(-1/\theta)\sum_{i=1}^n x_i + (p-1)\sum_{i=1}^n \ln x_i - np\ln\theta - n\ln\Gamma(p)\right]$. D'où $c(\theta) = (-1/\theta)$ et $T(X) = \sum_{i=1}^n X_i$. $c(\theta)$ est bien croissante en θ, et on recherche donc un test de domaine de rejet $R(X) = \{T(X) \geq c\}$ ou c est choisi de manière que $P(R(X)) = \alpha$, or $(1/\theta)\sum_{i=1}^n X_i \sim \Gamma(np, 1)$ d'où c.
2. Poser $\eta = 1/\theta$ et résoudre en η avec $T(X) = -\sum_{i=1}^n X_i$.

2) Soit $S \sim \mathcal{B}(n,\theta)$, on veut tester : $H_0 : \theta \leq \theta_0$ contre $H_0 : \theta > \theta_0$. Soit $\delta_k(S) = 1$ si $S \geq b$ et 0 sinon. Le rapport de vraisemblance est

$$L(s,\theta_0,\theta_1) = \frac{\theta_1^s(1-\theta_1)^{n-s}}{\theta_0^s(1-\theta_0)^{n-s}} = \rho^s \left(\frac{1-\theta_1}{(1-\theta_0}\right)^n \text{ où } \rho = \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}$$

$$s \geq k \Longleftrightarrow L(s,\theta_0,\theta_1) \geq \rho^k \left(\frac{1-\theta_1}{(1-\theta_0}\right)^n$$

donc δ_k est un test de NP et il est UPP.

On peut aussi considérer que l'on a une loi exponentielle : $P(X=k) = C_n^k \theta^k (1-\theta)^{n-k} = exp[k \ln \frac{\theta}{1-\theta} + n\ln(1-\theta) + \ln C_n^k]$. Donc $c(\theta) = \ln \frac{\theta}{1-\theta}$ et $c'(\theta) = \frac{1}{\theta(1-\theta)} > 0$, donc c est croissante strictement. De plus $T(X) = X$ est non dégénérée. D'où $T(X)$ est optimale et UPP, on construit $\Phi(\omega) = \mathbf{1}_{C_1 < T(\omega) < C_2} + \sum_{i=1,2} \gamma_i \mathbf{1}_{T(\omega)=C_i}$.

Pour un n-échantillon, même chose avec $T(X) = \sum X_i$.

Théorème 5.6 : *Soit $(\Omega, \mathcal{A}, P_\theta, \theta \in \Theta)$, Θ intervalle de $\mathbb{R}$. On suppose que ce modèle est exponentiel à un paramètre, et sous forme canonique :*

$$\exp[\theta.T(x) + S(x) + d(\theta)]$$

Soient $\theta_1 < \theta_2$ dans Θ, et

$$\begin{aligned} \phi(\omega) &= 1 \quad si \quad T(\omega) < C_1 \text{ ou } T(\omega) > C_2 \\ &= 0 \quad si \quad T(\omega) \in]C_1, C_2[\\ &= \gamma_i \quad si \quad T(\omega) = C_i \end{aligned}$$

a) *telle que $E_{\theta_1}(\Phi) = E_{\theta_2}(\Phi) = \alpha$. Alors Φ est associé à un test UPP parmi des tests sans biais, de niveau α, de $H_0 : \theta_1 \leq \theta \leq \theta_2$ contre $H_0 : \theta < \theta_1$ ou $\theta > \theta_2$.*

De plus, $\forall \alpha$, $0 < \alpha < 1$, $\exists \Phi$ de cette forme telle que $E_{\theta_1}(\Phi) = E_{\theta_2}(\Phi)$.

b) *telle que $E_{\theta_0}(\Phi) = \alpha$ et $\left(\frac{\partial}{\partial\theta} E_\theta(\Phi)\right)_{\theta=\theta_0} = 0$. Alors Φ est associé à un test UPP parmi des tests sans biais, de niveau α, de $H_0 : \theta = \theta_0$ contre $H_0 : \theta \neq \theta_0$.*

De plus, $\forall \alpha$, $0 < \alpha < 1$, $\exists \Phi$ de cette forme telle que $E_{\theta_0}(\Phi) = \alpha$ et $\left(\frac{\partial}{\partial\theta} E_\theta(\Phi)\right)_{\theta=\theta_0} = 0$.

5.2.7 Exemples

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

a) μ inconnu et σ^2 connu. On veut tester $H_0 : \mu \leq \mu_0$ contre $H_1 : \mu > \mu_0$

$$\begin{aligned} f_\theta(x_1,\ldots,x_n) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right] \\ &= \exp\left[\frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 - \frac{n}{2} \ln(2\pi\sigma^2)\right] \end{aligned}$$

$c(\mu) = \mu/\sigma^2$ est croissante sur $\mathbb{R}$ et $T(X) = \sum_{i=1}^n X_i$ non dégénérée. On construit donc $\phi_\lambda(\omega) = \mathbf{1}_{T>\lambda}(\omega)$, associé à $R = \{T(X) > \lambda\}$, où λ est défini par : $E_{\mu_0}(\phi_\lambda) = \alpha$.

$T(X) = \sum_{i=1}^n X_i \sim \mathcal{N}(n\mu, n\sigma^2)$, donc

$$E_{\mu_0}(\phi_\lambda) = P_{\mu_0}(T(X) > \lambda) = P_{\mu_0}\left(\frac{T(X) - n\mu_0}{\sqrt{n}\sigma} > \frac{\lambda - n\mu_0}{\sqrt{n}\sigma}\right) = \alpha$$

$\frac{\lambda - n\mu_0}{\sqrt{n}\sigma}$ est lu dans la table $\mathcal{N}(0,1)$, donc $\lambda = \sqrt{n}\sigma u + n\mu_0$.

b) μ inconnu et σ^2 connu. On veut tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$

Pour les mêmes raisons que précédemment, on construit

$$\phi_\lambda(\omega) = \mathbf{1}_{T>c_2 \text{ ou } T<c_1}(\omega) + \gamma_1 \mathbf{1}_{T=c_1}(\omega) + \gamma_2 \mathbf{1}_{T=c_2}(\omega).$$

La loi étant continue, $\gamma_1 = \gamma_2 = 0$, et le test est associé à $R = \{T(X) > c_2\} \cup \{T(X) < c_1\}$, où λ est défini par : $E_{\mu_0}(\phi_\lambda) = \alpha$ et $\left(\frac{\partial}{\partial\theta} E_\mu(\Phi)\right)_{\mu=\mu_0} = 0$.

2) Soit $X_1, \ldots, X_n$ un n-échantillon de loi de Cauchy $f_\Delta(x) = \frac{1}{1+(x-\Delta)^2}\frac{1}{\pi}$.

$$f_\Delta(x_1, \ldots, x_n) = \frac{1}{\pi^n}\prod_{i=1}^{n}\frac{1}{1+(x_i-\Delta)^2}$$

On veut tester $\mathcal{H} : (H_0 : \Delta = \Delta_0$ contre $H_1 : \Delta > \Delta_0)$. Pour cela, nous allons considérer les deux tests d'hypothèses :

$\alpha)$ $H_0 : \Delta = \Delta_0$ contre $H_1 : \Delta = \Delta_1$
$\beta)$ $H_0 : \Delta = \Delta_0$ contre $H_1 : \Delta = \Delta_2$

où $\Delta_0 < \Delta_1 < \Delta_2$.

Test UPP pour $\alpha)$: $\phi_1(\omega) = \mathbf{1}_{f_1/f_0>\lambda_1}(\omega) + \gamma_1\mathbf{1}_{f_1/f_0=\lambda_1}(\omega)$.

Test UPP pour $\beta)$: $\phi_2(\omega) = \mathbf{1}_{f_2/f_0>\lambda_2}(\omega) + \gamma_2\mathbf{1}_{f_2/f_0=\lambda_2}(\omega)$.

Un test UPP pour $\mathcal{H}$ doit aussi être UPP pour $\alpha)$ et $\beta)$. Donc sa fonction critique doit s'écrire sous les formes ϕ_1 et ϕ_2. Ceci n'est possible que si f_2/f_1 est une fonction croissante de f_1/f_0.

Pour $n = 1$: $L(x, 0, \Delta_2) = \frac{f_2}{f_0}(x) = \frac{1}{1+(x-\Delta_2)^2}$, et $L(x, 0, \Delta_1) = \frac{f_1}{f_0}(x) = \frac{1}{1+(x-\Delta_1)^2}$.

$$\text{Donc } L(x, 0, \Delta_2) < L(x', 0, \Delta_2) \iff (x-\Delta_2)^2 > (x-\Delta_2)^2 \iff (x-x')(\frac{x+x'}{2} - \Delta_2) > 0$$

$$\text{et } L(x, 0, \Delta_1) < L(x', 0, \Delta_1) \iff (x-\Delta_1)^2 > (x-\Delta_1)^2 \iff (x-x')(\frac{x+x'}{2} - \Delta_1) > 0$$

Donc les fonctions $L(x, 0, \Delta_2)$ et $L(x, 0, \Delta_1)$ ne sont pas de même monotonie et il n'existe pas de test UPP pour $\mathcal{H}$.

5.2.8 Tests du rapport de vraisemblance

Soit $(\Omega, \mathcal{A}, P_\theta, \theta \in \Theta)$ un modèle régulier, on note $(f_\theta(x))_\theta \in \Theta$ la famille des densités (ou lois discrètes)., et on suppose $\Theta = \Theta_1 \cup \Theta_2$ et $\Theta_1 \cap \Theta_2 =$. On veut tester $\mathcal{H} : (H_0 : \theta \in \Theta_1$ contre $H_1 : \theta \in \Theta_2)$.

Définition 5.11 : *Le test de domaine de rejet $\{L(X) > \lambda\}$ où*

$$L(X) = \frac{\sup_{\theta\in\Theta_2} f_\theta(X)}{\sup_{\theta\in\Theta_1} f_\theta(X)}$$

est appelé un test du rapport de vraisemblance.

Remarques :

- Si F est une fonction strictement croissante, $\{L(X) > \lambda\}$ est équivalent à $\{F(L(X)) > \lambda\}$. (par exemple, $F = \ln$.)
- Pas de résultat générique sur la qualité de ces tests.
- Il existe des propriétés asymptotiques pour ces tests. (quand la taille de l'échantillon tend vers $+\infty$.)
- Si $\sup_{\theta\in\Theta} f_\theta(x)$ est atteint pour $\theta \in \Theta_2$, on peut écrire

  $$L(X) = \frac{\sup_{\theta\in\Theta} f_\theta(X)}{\sup_{\theta\in\Theta_1} f_\theta(X)}$$

  qui est plus facile à calculer. En particulier, le numérateur est alors atteint pour l'estimateur du maximum de vraisemblance quand il existe.

Exemples :

1. Soit $X_1, \ldots, X_n$ un n-échantillon de loi $\mathcal{N}(\mu, \sigma^2)$, μ et σ^2 inconnus. On veut tester $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$.

   On considère donc $\Theta = \mathbb{R} \times \mathbb{R}^{+*}$, $\Theta_1 = \{(\mu_0, \sigma^2), \sigma^2 \in \mathbb{R}^{+*}\}$ et $\Theta_2 = \{(\mu, \sigma^2), \sigma^2 \in \mathbb{R}^{+*}, \mu \in \mathbb{R} - \{\mu_0\}\}$

   $M_0 = \sup_{\theta \in \Theta_1} f_\theta(x)$ est atteint pour $\mu = \mu_0$ et pour l'estimateur du maximum de vraisemblance de σ^2 à μ_0 connu : $\hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2$. Donc

   $$M_0 = \exp(-\frac{1}{2n} \ln(2\pi) - \frac{1}{2n} \ln(\hat{\sigma}_0^2) - \frac{n}{2}).$$

   $M = \sup_{\theta \in \Theta} f_\theta(x)$ est atteint pour l'estimateur du maximum de vraisemblance de (μ, σ^2) : $(\hat{\mu}, \hat{\sigma}^2) = (\bar{X}, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2)$. Donc

   $$M = \exp\left[-\frac{1}{2n} \ln(2\pi) - \frac{1}{2n} \ln(\hat{\sigma}^2) - -\frac{n}{2} \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right].$$

   D'où $M/M_0 = \exp(-\frac{1}{2n}[\ln(\hat{\sigma}^2) - \ln(\hat{\sigma}_0^2)]$, et le test du rapport de vraisemblance est obtenu par $\{M/M_0 > \lambda\} = \{\hat{\sigma}_0^2/\hat{\sigma}^2 > \lambda'\}$.

   Or $\hat{\sigma}_0^2/\hat{\sigma}^2 = 1 + (\bar{X} - \mu_0)^2/\hat{\sigma}^2$, et donc $\{\hat{\sigma}_0^2/\hat{\sigma}^2 > \lambda'\} = \{\left|\frac{\bar{X}-\mu_0}{\hat{\sigma}}\right| > \lambda''\}$. Sous H_0, la loi de $\sqrt{n}\frac{\bar{X}-\mu_0}{s}$ est une loi de Student à $n-1$ degrés de liberté, et on obtient les domaines de rejet suivants :

   pour $(H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0) : R = \{\left|\sqrt{n}\frac{\bar{X}-\mu_0}{s}\right| > c\}$

   pour $(H_0 : \mu \leq \mu_0$ contre $H_1 : \mu > \mu_0) : R = \{\sqrt{n}\frac{\bar{X}-\mu_0}{s} > c'\}$

   pour $(H_0 : \mu \geq \mu_0$ contre $H_1 : \mu < \mu_0) : R = \{\sqrt{n}\frac{\bar{X}-\mu_0}{s} < c''\}$

2. Soit $X_1, \ldots, X_{n_1}$ un n_1-échantillon de loi $\mathcal{N}(\mu_1, \sigma^2)$, $Y_1, \ldots, Y_{n_2}$ un n_2-échantillon de loi $\mathcal{N}(\mu_2, \sigma^2)$ μ_1, μ_2 et σ^2 inconnus. On pose $n = n_1 + n_2$ et on veut tester $H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$.

   On considère donc $\Theta_1 = \{(\mu_1, \mu_2, \sigma^2), \sigma^2 \in \mathbb{R}^{+*}, \mu_1 \in \mathbb{R}, \mu_2 = \mu_1\}$ et $\Theta_2 = \{(\mu_1, \mu_2, \sigma^2),\ \sigma^2 \in \mathbb{R}^{+*}, \mu_1, \mu_2 \in \mathbb{R}\}$

   $$\ln f_\theta(x, y) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_2)^2$$

   Après calculs, on obtient :

   $$T = \sqrt{\frac{n_1 n_2}{n}} \frac{\bar{Y} - \bar{X}}{S} \text{ où } S^2 = \frac{1}{n-2} \left[\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2\right]$$

   Sous H_0, la loi de T est une loi de Student à $n-2$ degrés de liberté car $\bar{X}/\sigma$, $\bar{Y}/\sigma$, $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$ et $\frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \bar{Y})^2$ sont indépendantes de lois $\mathcal{N}(\mu_1/\sigma, 1/n_1)$, $\mathcal{N}(\mu_2/\sigma, 1/n_2)$, $\chi^2_{n_1-1}$ et $\chi^2_{n_2-1}$.

   On obtient les domaines de rejet suivants :

   pour $(H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2) : R = \{|T| > c\}$

   pour $(H_0 : \mu_1 \leq \mu_2$ contre $H_1 : \mu_1 > \mu_2) : R = \{T < c'\}$

   pour $(H_0 : \mu_1 \geq \mu_2$ contre $H_1 : \mu_1 < \mu_2) : R = \{T > c''\}$

3. Soit $(X_1, Y_1), \ldots, (X_n, Y_n)$ un n-échantillon de loi $\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. On veut tester $H_0 : \rho = 0$ contre $H_1 : \rho \neq 0$.

   $$\ln f_\theta(x, y) = -n \ln(2\pi\sigma_1^2\sigma_2^2\sqrt{1-\rho^2})$$
   $$-\frac{1}{2(1-\rho^2)} \left[\sum_{i=1}^n (x_i - \mu_1)^2/\sigma_1^2 + \sum_{i=1}^n (x_i - \mu_1)(y_i - \mu_2)/\sigma_1^2\sigma_2^2 \sum_{i=1}^n (y_i - \mu_2)^2/\sigma_2^2\right]$$

Les estimateurs du maximum de vraisemblance sont :

$$\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad \hat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \qquad \hat{\rho} = \sum_{i=1}^{n}(X_i - \mu_1)(Y_i - \mu_2)/\hat{\sigma}_1^2\hat{\sigma}_2^2$$

Après calculs, on obtient le logarithme du rapport de vraisemblance : $\ln \lambda(x, y) = \frac{-n}{2}\ln(1 - \hat{\rho}^2)$, qui est une fonction croissante de $\hat{\rho}^2$, le domaine de rejet pour le test bilatère est donc $R = \{|\hat{\rho}| > c\}$.

On montre que, sous H_0, $T = \sqrt{n-2}\hat{\rho}/\sqrt{1-\hat{\rho}^2} \sim \mathcal{T}_{n-2}$.

Chapter 6

Modèles linéaires

6.1 Introduction, définitions

Définition 6.1 : *$Y = (Y_1, \ldots, Y_n)$ suit un modèle linéaire (généralisé) si*

$$Y = X\beta + \epsilon$$

où X est une matrice (non aléatoire) de taille (n, p), β est un vecteur (non aléatoire) dans $\mathbb{R}^p$, Y et ϵ sont des vecteurs aléatoires à valeurs dans $\mathbb{R}^n$. ϵ est supposé centré, de matrice de covariance $\sigma^2 I_n$.

Remarque : Y est une variable aléatoire à valeur dans $\mathbb{R}^n$, et $E(Y) = X\beta$, $Var(Y) = \sigma^2 I_n$.

Définition 6.2 : *Un modèle linéaire est gaussien si $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$.*

Exemples :

1. Performances d'un traitement sur une population.
   m_k sujets reçoivent le traitement k. Y_{kl} est le rendement de la parcelle l ayant subi le traitement k.
   $Y_{kl} = \beta_k + \epsilon_{kl}$ ——> **Modèle d'analyse de la variance à un facteur**.

$$Y = (Y_{11}, Y_{12}, \ldots, Y_{1n_1}, Y_{21}, Y_{12}, \ldots, Y_{2n_2}, \ldots, Y_{pn_p})'$$

$$\epsilon = (\epsilon_{11}, \epsilon_{12}, \ldots, \epsilon_{1n_1}, \epsilon_{21}, \epsilon_{12}, \ldots, \epsilon_{2n_2}, \ldots, \epsilon_{pn_p})'$$

$$X = \begin{pmatrix} \mathbb{1}_{n_1} & & & \\ & \mathbb{1}_{n_2} & (0) & \\ & (0) & \ddots & \\ & & & \mathbb{1}_{n_p} \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

   où $\mathbb{1}_{n_k}$ est un vecteur colonne formé de 1, de taille n_k, et les ϵ_{kl} sont iid $\mathcal{N}(0, \sigma^2)$, $1 \leq k \leq p$, $1 \leq l \leq n_k$.

2. $Y_j = \beta_0 + \beta_1 x_{j1} + \ldots + \beta_p x_{jp} + \epsilon_j$ ——> **Modèle de régression linéaire à p facteurs**.

$$Y = (Y_1, \ldots, Y_n)' \qquad \epsilon = (\epsilon_1, \ldots, \epsilon_n)'$$

$$X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & & & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

3. $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ où $1 \leq i \leq p$, $1 \leq j \leq n_i$, $n_1 + \ldots + n_p = n$

$$Y = (Y_{11}, Y_{12}, \ldots, Y_{1n_1}, Y_{21}, Y_{12}, \ldots, Y_{2n_2}, \ldots, Y_{pn_p})'$$

$$X = \begin{pmatrix} & \mathbb{1}_{n_1} & & & \\ \mathbb{1}_n & & \mathbb{1}_{n_2} & (0) & \\ & & (0) & \ddots & \\ & & & & \mathbb{1}_{n_p} \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}$$

6.2 Estimation

On considère un modèle linéaire gaussien $Y = X\beta + \epsilon = \xi + \epsilon$ en posant $E(Y) = X\beta = \xi$.

6.2.1 Forme canonique

Il s'agit de faire un changement de base orthogonal. Notons $(v_1, \ldots, v_n)$ la nouvelle base et $Z_i = Y'v_i$, $\eta_i = \xi' v_i$. Alors $Y = \sum_{i=1}^n Z_i v_i$. Remarquons que si M est la matrice de changement de base, elle est supposée orthogonale : $M^{-1} = M'$.

Soit E l'espace vectoriel image de la matrice X (sous entendu de l'endomorphisme associé à la matrice X). On note r le rang de X, c'est à dire la dimension de E, et on choisit $(v_1, \ldots, v_r)$ une base orthonormée de E. Ensuite, on complète cette base par $(v_{r+1}, \ldots, v_n)$ de manière que $(v_1, \ldots, v_n)$ soit orthonormée.

Or $\xi = X\beta \in E$, donc $\begin{cases} \xi = \sum_{i=1}^r a_i v_i \text{ avec } a_i = \xi.v_i, i = 1, \ldots, r \\ \xi \perp v_i, i = r+1, \ldots, n \text{ c'est à dire } \xi.v_i = 0, i = r+1, \ldots, n \end{cases}$

Proposition 6.1 : *Les Z_i sont indépendantes et de loi $\mathcal{N}(\eta_i, \sigma^2)$.*

Dém : Ceci provient du fait que le changement de base est orthogonal et que les Y_i sont indépendantes. □

Proposition 6.2 : $(Z_1, \ldots, Z_r, \sum_{i=1}^n Z_i^2)$ *est une statistique exhaustive complète pour* $(\eta_1, \ldots, \eta_r, \sigma^2)$.

Dém : Ceci provient du fait que l'on a une famille exponentielle. □

Définition 6.3 : *Si r est minimal ($r = rg(X)$), cette nouvelle écriture du modèle : $Z_i = \eta_i + \epsilon_i'$ est une forme canonique.*

6.2.2 Estimation des moyennes

Z_i est l'estimateur UVMB pour η_i, et en général, $\sum_{i=1}^r d_i Z_i$ est l'estimateur UVMB pour $\sum_{i=1}^r d_i \eta_i$. Donc $\hat{\xi}_j = \sum_{i=1}^r Z_i v_{ji}$ puisque $\xi_j = \sum_{i=1}^r \eta_i v_{ji}$

Remarques :

- C'est aussi l'estimateur du maximum de vraisemblance.
- $\hat{\xi}$ est déterminé en fonction de Z, mais Z n'est pas connu explicitement.
- Comment passer de ξ à β ($\xi = X\beta$) ?

Théorème 6.1 :

1. *$\hat{\xi}$ est l'unique projection orthogonale de Y sur E : $\hat{\xi} = \sum_{i=1}^p \hat{\beta}_i x_i = X\hat{\beta}$ où x_i est la i-ème colonne de X.*

2. *Si $p = r$, alors $\beta = (\beta_1, \ldots, \beta_p)'$ est identifiable. L'estimateur des moindres carrés est unique et*

$$\hat{\beta} = (X'X)^{-1}X'Y$$

3. *De plus, toujours si $p = r$, $\hat{\beta} = (X'X)^{-1}X'\hat{\xi}$ et $\beta = (X'X)^{-1}X'\xi$*

Dém :

1. $\forall t \in E$, $Y - t = \sum_{i=1}^{r}(Z_i - t'v_i)v_i + \sum_{i=r+1}^{n} Z_i v_i$ d'où

$$||Y-t||^2 = \sum_{i=1}^{n}(Y_i - t_i)^2 = \sum_{i=1}^{r}(Z_i - t'v_i)^2 + \sum_{i=1}^{r} Z_i^2$$

   $||Y - t||^2$ est minimale pour $t \in E$ si $\forall i \in \{1, \ldots, r\}$, $Z_i = t'v_i \Longleftrightarrow t = \hat{\xi}$.

2. $p = r$, l'estimateur des moindres carrés est caractérisé par un système d'équations normales :

$$\frac{\partial}{\partial \beta_k}\sum_{i=1}^{n}(Y_i - \sum_{i=1}^{p} x_{ij}\beta_j)^2 = 0, \forall k \in \{1, \ldots, p\}$$

$$-2\sum_{i=1}^{n}(Y_i - \sum_{i=1}^{p} x_{ij}\beta_j)x_{ik} = 0, \forall k \in \{1, \ldots, p\}$$

$$\sum_{i=1}^{n} Y_i x_{ik} = \sum_{j=1}^{p}\beta_j\left(\sum_{i=1}^{n} x_{ij}x_{ik}\right), \forall k \in \{1, \ldots, p\}$$

   c'est à dire $X'Y = X'X\beta$.

   Remarque : $X'X$ est de taille (p, p), elle est régulière (inversible) si $r = p$.

   L'estimateur des moindres carrés est donc $\hat{\beta} = (X'X)^{-1}X'Y$.

3. $\hat{\xi} = X\hat{\beta} \in E \Longrightarrow (X'X)^{-1}X'(X\hat{\beta}) = \hat{\beta}$.

   $\xi = X\beta \in E \Longrightarrow (X'X)^{-1}X'(X\beta) = \beta$.

□

Remarques :

- si $r = p$, $E\hat{\beta} = (X'X)^{-1}X'E(Y) = \beta$, $\hat{\beta}$ est donc sans biais pour β.

  $Var(\hat{\beta}) = E[(\hat{\beta} - E\hat{\beta})(\hat{\beta} - E\hat{\beta})]$, or $\hat{\beta} - E\hat{\beta} = (X'X)^{-1}X'(Y - X\beta)$, donc

$$\begin{aligned} Var(\hat{\beta}) &= E[(X'X)^{-1}X'\epsilon((X'X)^{-1}X')'] \\ &= E[(X'X)^{-1}X'\epsilon\epsilon'X((X'X)^{-1})'] = \sigma^2(X'X)^{-1} \end{aligned}$$

  puisque $\epsilon = Y - X\beta$.

- La relation $X'Y = (X'X)\hat{\beta}$ reste vraie, même si $r \neq p$. Elle a alors une infinité de solutions qui conduisent toutes au même $\hat{\xi}$.

Définition 6.4 : *Lorsque $r = p$, on dit que le modèle est régulier. Sinon, on dit que le modèle est singulier.*

Remarque : La taille de $X'X$ est (p, p), en général, p est petit, alors que par contre n est grand.

Proposition 6.3 : *Si ψ est une forme linéaire (application linéaire à valeurs dans $\mathbb{R}$) et si $p = r$, $\psi(\hat{\xi})$ est UVMB pour $\phi(\xi)$.*

Dém : $\left.\begin{array}{l} \beta = (X'X)^{-1}X'\xi \\ \hat{\beta} = (X'X)^{-1}X'\hat{\xi} \end{array}\right\} \Longrightarrow$ $\hat{\beta}$ est UVMB pour β et donc les combinaisons linéaires de β ont la même propriété. □

Proposition 6.4 : *Pour un modèle régulier*

- $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1})$
- $[Cov(\hat{\beta}_j, \hat{\beta}_k)]_{j,k} = \sigma^2(X'X)^{-1}$
- $[Cov(\hat{\xi}_j, \hat{\xi}_k)]_{j,k} = \sigma^2 P$ *où* $P = X(X'X)^{-1}X'$
- $\hat{\epsilon} = Y - \hat{\xi}$ *est centrée*

Exemples :

1. $Y_i = \beta_1 + \beta_2 x_i + \epsilon_i$, où les ϵ_i sont iid de loi $\mathcal{N}(0, \sigma^2)$, $i = 1, \dots n$.

   L'estimateur des moindres carrés est

$$\left\{\begin{array}{l} \hat{\beta}_2 = \dfrac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{x} \end{array}\right.$$

   Par ailleurs,

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\text{donc } X' = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \quad \text{et} \quad X'X = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

   $X'X$ est de rang 2 si $\exists i, j$ tels que $x_i \neq x_j$. On se place dans ce cas.

   On considère $v_1 = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})'$ et $v_2 = (\tau_1, \dots, \tau_n)'$ où

$$\tau_i = \frac{x_i - \bar{x}}{[\sum_{i=1}^n (x_i - \bar{x})^2]^{1/2}}$$

   On vérifie que (v_1, v_2) est une base orthonormée de E, et on obtient :

$$\hat{\xi} = \sum_{i=1}^2 (Y'.v_j)v_j = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i\right) v_1 + \left(\frac{1}{[\sum(x_i - \bar{x})^2]^{1/2}} \sum_{i=1}^n (x_i - \bar{x})Y_i\right) v_2$$

$$\text{D'où } \hat{\xi}_k = \bar{Y} + (x_k - \bar{x})\frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum(x_i - \bar{x})^2}$$

   On retrouve ainsi les estimateurs précédents en utilisant les relations :

$$\left\{\begin{array}{l} \xi_1 = \beta_1 + \beta_2 x_1 \\ \xi_2 = \beta_1 + \beta_2 x_2 \end{array}\right. \Longrightarrow \left\{\begin{array}{l} \beta_2 = \frac{\xi_2 - \xi_1}{x_2 - x_1} \\ \beta_1 = \xi_1 - \beta_2 x_1 \end{array}\right.$$

   où x_1 et x_2 sont supposés distincts.

2. $Y_{ij} = \beta_i + \epsilon_{ij}$

$$X = \begin{pmatrix} \mathbb{1}_{n_1} & & & \\ & \mathbb{1}_{n_2} & (0) & \\ & (0) & \ddots & \\ & & & \mathbb{1}_{n_p} \end{pmatrix} \quad \text{et} \quad X'X = \begin{pmatrix} n_1 & & & \\ & n_2 & (0) & \\ & (0) & \ddots & \\ & & & n_p \end{pmatrix}$$

$$\text{D'où } \hat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} 1/n_1 & & & \\ & 1/n_2 & (0) & \\ & (0) & \ddots & \\ & & & 1/n_p \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n_1} Y_{1j} \\ \sum_{j=1}^{n_2} Y_{2j} \\ \vdots \\ \sum_{j=1}^{n_p} Y_{pj} \end{pmatrix} = \begin{pmatrix} Y_{1.} \\ Y_{2.} \\ \vdots \\ Y_{p.} \end{pmatrix}$$

où $Y_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$.

6.2.3 Théorème de Gauss-Markov

Théorème 6.2 S : *oit ψ une application linéaire de $\mathbb{R}^p$ dans $\mathbb{R}^q$ et $Y = X\beta + \epsilon$ un modèle linéaire où, pour tout i et tout j, $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2 < +\infty$ et $Cov(\epsilon_i, \epsilon_j) = 0$. L'estimateur des moindres carrés $\psi(\hat{\beta})$ est un estimateur non biaisé pour $\psi(\beta)$, uniformément de variance minimale parmi les estimateurs non biaisés <u>linéaires</u> en Y.*

Dém : On peut écrire $\psi(U) = A.U$ où $U \in \mathbb{R}^p$ et $A \in \mathcal{M}(q,p)$. Soit $T(Y)$ un autre estimateur non biaisé de $\psi(\beta)$ linéaire en Y, il s'écrit alors $T(Y) = T.Y$. On veut montrer que $Var(TY) \geq Var(A\hat{\beta})$. Remarquons qu'il s'agit ici de matrices de covariance et que l'inégalité signifie en fait que la matrice $Var(TY) - Var(A\hat{\beta})$ est définie positive.

Or $Var(TY) = T(VarY)T' = \sigma^2 TT'$. Soit $P = X(XX')^{-1}X'$, on peut vérifier facilement que $P + (I - P)(I - P) = I$, et donc on peut écrire :

$$Var(TY) = \sigma^2 TT' = \sigma^2 TIT' = \sigma^2 T(P + (I-P)(I-P))T'$$

$$= \sigma^2 TPT' + \sigma^2 T(I-P)(I-P)T'$$

$$= \sigma^2 (TX)(X'X)^{-1}(TX)' + \sigma^2 (T - TP)(T' - PT')$$

On a supposé que TY et $A\hat{\beta}$ sont deux estimateurs sans biais de $A\beta$, donc $E(TY) = TE(Y) = TX\beta$ est aussi égal à $E(A\hat{\beta}) = A\beta$ et ce pour tout β, ce qui implique $TX = A$. On remplace alors dans l'équation précédente :

$$Var(TY) = \sigma^2 A(X'X)^{-1}A' + \sigma^2[T - TX(X'X)^{-1}X'][T' - X(X'X)^{-1}X'T']$$

$$= Var(A\hat{\beta}) + \sigma^2[T - A(X'X)^{-1}X'][T' - X(X'X)^{-1}A'] = Var(A\hat{\beta}) + Var(TY - A\hat{\beta})$$

$$\begin{array}{lll} \text{en effet} & Var(TY - A\hat{\beta}) & = E[(TY - A\hat{\beta})(TY - A\hat{\beta})'] \\ & & = E[(TY - A(X'X)^{-1}X'Y)(TY - A(X'X)^{-1}X'Y)'] \\ & & = E[(T - A(X'X)^{-1}X')YY'(T - A(X'X)^{-1}X')'] \\ & & = (T - A(X'X)^{-1}X')E[YY'](T - A(X'X)^{-1}X')' \\ & & = \sigma^2(T - A(X'X)^{-1}X')(T - A(X'X)^{-1}X')' \end{array}$$

□

6.2.4 Estimation de la variance σ^2

Proposition 6.5 : $S^2 = \frac{1}{n-r}\sum_{i=r+1}^{n} Z_i^2$ *est un estimateur sans biais UVMB pour σ^2 et $(n-r)S^2/\sigma^2 \sim \chi^2_{n-r}$. De plus, si le modèle est régulier, $\hat{\beta}$ et S^2 sont indépendants.*

Dém : Provient directement du fait que les variables $Z_{r+1}, Z_{r+2}, \ldots, Z_n$ sont iid $\mathcal{N}(0, \sigma^2)$. De plus, $S^2 = \frac{1}{n-r}\sum_{i=1}^n Z_i^2 - \sum_{i=1}^r Z_i^2$ ne dépend que de la statistique $(Z_1, \ldots, Z_r, \sum_{i=1}^n Z_i^2)$ qui est exhaustive et complète, donc S^2 est UVMB. □

Le vecteur $(Y_1 - \hat{\xi}_1, Y_2 - \hat{\xi}_2, \ldots, Y_n - \hat{\xi}_n)$ est appelé le vecteur des résidus et $Y - \hat{\xi} = \sum_{i=r+1}^n Z_i v_i$, donc $S^2 = \frac{1}{n-r}||Y - \hat{\xi}||^2 = \frac{1}{n-r}\sum_{i=1}^n (Y_i - \hat{\xi}_i)^2$. $||Y - \hat{\xi}||^2 = \sum_{i=1}^n (Y_i - \hat{\xi}_i)^2$ est appelé le résidu.

Remarque : Dans le cas non gaussien, S^2 est toujours sans biais pour σ^2, mais on ne sait pas s'il est UVMB.

6.2.5 Intervalles de confiance

Remarque : $\hat{\xi}$ et $Y - \hat{\xi}$ sont indépendants car $\hat{\xi} = \sum_{i=1}^r Z_i v_i$ et $Y - \hat{\xi} = \sum_{i=r+1}^n Z_i v_i$, les Z_i sont des gaussiennes indépendantes et la base $(v_1, \ldots, v_n)$ est orthogonale.

Proposition 6.6 : *Tout estimateur $\psi(\hat{\xi}) = \sum_{i=1}^n w_i \hat{\xi}_i$ est tel que $\psi(\hat{\xi}) \sim \mathcal{N}(\psi(\xi), \sigma_\psi^2)$ où on pose $\psi(\hat{\xi}) = \sum_{i=1}^n a_i Y_i$ et $\sigma_\psi^2 = Var(\psi(\hat{\xi})) = \sigma^2 \sum_{i=1}^n \sigma_i^2$.*

Proposition 6.7 : *Si $p = r$, $\hat{\beta}$ et S^2 sont indépendants et $\sum_{j=1}^r d_j \hat{\beta}_j \sim \mathcal{N}(\sum_{j=1}^r d_j \beta_j, \sigma_d^2)$ où $\sigma_d^2 = \sigma^2 \sum_{i=1}^n a_i^2$ lorsque l'on pose $\sum_{j=1}^r d_j \beta_j = \sum_{i=1}^n a_i Y_i$.*

Ces deux propositions permettent de construire des intervalles de confiance pour $\psi(\hat{\xi})$ dans le cas général et pour $\sum_{j=1}^r d_j \hat{\beta}_j$ dans le cas régulier.

Exemple : $Y_i = \beta_1 + \beta_2 x_i + \epsilon_i$, où les ϵ_i sont iid de loi $\mathcal{N}(0, \sigma^2)$, $i = 1, \ldots n$. On a déjà démontré que :

$$\begin{cases} \hat{\beta}_2 = \dfrac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{x} \end{cases}$$

et on obtient donc

$$\begin{cases} Var(\hat{\beta}_2) = \dfrac{\sigma^2 \sum_{i=1}^n (x_i - \bar{x})^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} = \dfrac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ Var(\hat{\beta}_1) = Var(\bar{Y}) - \bar{x}^2 Var(\hat{\beta}_2) - 2\bar{x} Cov(\bar{Y}, \hat{\beta}_2) = \sigma^2 \left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \end{cases}$$

et les intervalles de confiance :

$$\text{pour } \beta_2 : \left[\hat{\beta}_2 - \frac{st}{\sqrt{\sum (x_i - \bar{x})^2}}, \hat{\beta}_2 + \frac{st}{\sqrt{\sum (x_i - \bar{x})^2}} \right]$$

$$\text{pour } \beta_1 : \left[\hat{\beta}_1 - \frac{st}{\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}}, \hat{\beta}_1 + \frac{st}{\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}} \right]$$

6.3 Tests sur les modèles linéaires

Beaucoup de tests sont de la forme : $H_0 : \xi \in E_0$ contre $H_1 : \xi \in E - E_0$ où E_0 est une sous-espace vectoriel de E.

Dans le cas gaussien, on peut exprimer facilement le test du rapport de vraisemblance pour ces hypothèses :

$$f_\theta(y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\frac{1}{2\sigma^2}||y - \xi||^2)$$

$(\hat{\xi},\hat{\sigma}^2)$ l'estimateur du maximum de vraisemblance dans $E \times \mathbb{R}^{+*}$ est obtenu par :

$$\begin{cases} \hat{\xi} = \text{ projection orthogonale de } Y \text{ sur } E \\ \hat{\sigma}^2 = \frac{1}{n}\|y-\hat{\xi}\|^2 \end{cases}$$

$(\hat{\xi}_0,\hat{\sigma}_0^2)$ l'estimateur du maximum de vraisemblance dans $E_0 \times \mathbb{R}^{+*}$ est obtenu par :

$$\begin{cases} \hat{\xi}_0 = \text{ projection orthogonale de } Y \text{ sur } E_0 \\ \hat{\sigma}_0^2 = \frac{1}{n}\|y-\hat{\xi}_0\|^2 \end{cases}$$

Le rapport de vraisemblance s'écrit

$$L(y) = \ln\left(\frac{\sup_{E\times\mathbb{R}^{+*}} f_{\xi,\sigma}(y)}{\sup_{E_0\times\mathbb{R}^{+*}} f_{\xi,\sigma}(y)}\right) = \left(\frac{\|y-\hat{\xi}_0\|^2}{\|y-\hat{\xi}\|^2}\right)^{n/2}$$

En fait, on considère la statistique

$$T_n = \frac{n-r}{r-q}\;\frac{\|y-\hat{\xi}_0\|^2 - \|y-\hat{\xi}\|^2}{\|y-\hat{\xi}\|^2}$$

qui est une fonction croissante de $L(y)$. T_n est appelé la statistique de Fisher.

Théorème 6.3 : *Sous H_0 ($\xi \in E_0$), $T_n \sim F(r-q,n-r)$. En général, T_n suit une loi de Fisher non centrée de coefficient de non centralité $\delta^2 = \|\hat{\xi}-\hat{\xi}_0\|^2/\sigma^2$ à $r-q$, $n-r$ degrés de liberté.*

Dém : On reprend la forme canonique : $Z' = Y'A$ où les Z_i sont iid de loi $\mathcal{N}(\mu_i,\sigma^2)$. On note $q = dimE_0$ et $r = dimE$. Puisque $E_0 \subset E$, on peut choisir la base $(v_1,\ldots,v_n)$ de manière que $(v_1,\ldots,v_q)$ soit une base de E_0 et que $(v_1,\ldots,v_r)$ soit une base de E. Ainsi $\|Y-\hat{\xi}\|^2 = \sum_{i=r+1}^n Z_i^2$ et $\|Y-\hat{\xi}_0\|^2 = \sum_{i=q+1}^n Z_i^2$, et donc

$$T_n = \frac{n-r}{r-q}\;\frac{\sum_{i=q+1}^n Z_i^2 - \sum_{i=r+1}^n Z_i^2}{\sum_{i=r+1}^n Z_i^2} = \frac{n-r}{r-q}\;\frac{\sum_{i=q+1}^r Z_i^2}{\sum_{i=r+1}^n Z_i^2}$$

$$\begin{cases} \frac{1}{\sigma^2}\sum_{i=q+1}^r Z_i^2 \sim \chi^2_{r-q} \\ \frac{1}{\sigma^2}\sum_{i=r+1}^n Z_i^2 \sim \chi^2_{n-r} \end{cases} \quad \text{et elles sont indépendantes}$$

T_n suit donc une loi de Fisher à $(r-q,n-r)$ degrés de liberté. □

6.3.1 Hypothèses testables

Définition 6.5 : *$g(\beta) \in \mathbb{R}^k$ est linéairement estimable s'il existe un estimateur linéaire en Y sans biais pour $g(\beta)$.*

Proposition 6.8 : *q est linéairement estimable ssi $g(\beta) = AX\beta$ où $A \in \mathcal{M}(k,n)$. De plus APY est l'estimateur sans biais UVMB parmi les estimateurs linéaires ($P = X(X'X)^{-1}X'$).*

Dém : "$\Longrightarrow$" On suppose g linéairement estimable, c'est à dire : $\exists T/T = AY$ et $E(T) = E(AY) = g(\beta) = A.E(Y) = AX\beta$ donc $g(\beta) = AX\beta$.

"$\Longleftarrow$" On suppose que $g(\beta) = AX\beta$ et on construit $T = AY$. $PY = X(X'X)^{-1}X'Y$ est un estimateur sans biais pour $X\beta$ donc APY est sans biais pour $AX\beta$ et on conclut en appliquant le théorème de Gauss-Markov au modèle $Y = \xi + \epsilon$. □

Remarque : Toute reparamétrisation régulière du modèle donne les mêmes estimateurs : Soit $Y = X_1\gamma_1 + \epsilon$ avec $X_1\gamma_1 = X\beta$ et X_1 injective, alors $AX\beta = AX_1\gamma_1$ et $\hat{\gamma}_1 = (X_1'X_1)^{-1}X_1'Y$ d'où $\hat{g} = AX_1\hat{\gamma}_1 = AX_1(X_1'X_1)^{-1}X_1'Y = APY$.

Définition 6.6 : *Un hypothèse est testable si elle équivaut à une hypothèse linéaire sur une fonction estimable de β.*

6.3.2 Application à la régression linéaire

Soit $Y_i = \beta_1 + \beta_2 x_i + \epsilon_i$, où les ϵ_i sont iid de loi $\mathcal{N}(0, \sigma^2)$, $i = 1, \dots n$.

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$X'X$ est de rang 2 si $\exists i, j$ tels que $x_i \neq x_j$. On se place dans ce cas.

- On veut tester : $H_0 : \beta_2 = 0$ contre $H_1 : \beta_2 \neq 0$.

  Sous H_0 : $\xi_i = \beta_1$, $i = 1, \dots n$ et $\hat{\beta}_1 = \bar{Y}$ est l'estimateur des moindres carrés.

  Sous H_1 : $\begin{cases} \hat{\beta}_2 = \dfrac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{x} \end{cases}$ et $\hat{\xi}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i = \hat{Y} + \hat{\beta}_2 (x_i - \bar{x})$

  On obtient alors

$$T_n = \frac{n-2}{2-1} \frac{||\hat{\xi} - \hat{\xi}_0||^2}{||Y - \hat{\xi}||^2} = (n-2) \frac{\sum_{i=1}^n (\bar{Y} + \hat{\beta}_2 (x_i - \bar{x}) - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y} - \hat{\beta}_2 (x_i - \bar{x}))^2}$$

$$= (n-2) \frac{\hat{\beta}_2^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (Y_i - \bar{Y} - \hat{\beta}_2 (x_i - \bar{x}))^2}$$

  Sous H_0, $T_n \sim \mathcal{F}(1, n-2)$. Le domaine de rejet est alors de la forme $R = \{T_n \geq f\}$ où f est lu dans la table $\mathcal{F}(1, n-2)$.

  Remarque : On peut écrire $T_n = V_n^2$ avec $V_n = \frac{\sqrt{n-2} Z_2}{\sqrt{\sum_{i=3}^n Z_i^2}}$. En effet, $\hat{\xi} - \hat{\xi}_0 = \sum_{i=q+1}^r Z_i v_i = Z_2 v_2$ et donc $||\hat{\xi} - \hat{\xi}_0||^2 = Z_2^2$. De plus, $||Y - \hat{\xi}||^2 = \sum_{i=r+1}^n Z_i^2$.

  Sous H_0 : $Z_2 \sim \mathcal{N}(0, \sigma^2)$ et $V_n \sim \mathcal{T}_{n-2}$.

  Sous H_1 : $Z_2 \sim \mathcal{N}(\eta_2, \sigma^2)$ et $V_n \sim \mathcal{T}_{n-2}$ non centrée.

  Le domaine de rejet est donc de la forme $\{|V_n| \geq t\}$.

- On veut tester : $H_0 : \beta_2 \leq 0$ contre $H_1 : \beta_2 > 0$. Alors $R = \{U_n \geq t\}$, en effet $U_n = \frac{Z_2/\sigma}{\sqrt{S^2}}$, or $Z_2/\sigma \sim \mathcal{N}(\eta_2, \sigma^2)$ où $\eta_2 = \xi' v_2$ et $\eta_2 \neq 0$ si $\xi \notin E - E_0$. Donc

$$\eta_2 = \xi' v_2 = \sum_{i=1}^n \xi_i \tau_i = \sum_{i=1}^n (\beta_1 + \beta_2 x_i) \frac{x_i - \bar{x}}{\sqrt{\sum (x_i - \bar{x})^2}} = \beta_2 \sqrt{\sum (x_i - \bar{x})^2}$$

6.3.3 Application aux modèles d'analyse de la variance

Modèle à un facteur $Y_{kl} = \beta_k + \epsilon_{kl}$ où les ϵ_{kl} sont iid $\mathcal{N}(0, \sigma^2)$, $1 \leq k \leq p$, $1 \leq l \leq n_k$.

$$X = \begin{pmatrix} \mathbb{1}_{n_1} & & & \\ & \mathbb{1}_{n_2} & (0) & \\ & (0) & \ddots & \\ & & & \mathbb{1}_{n_p} \end{pmatrix} \quad \text{et} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

On veut tester : $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p$ contre $H_1 : \exists i, j \in \{1, \dots, p\}, \beta_i \neq \beta_j$.

Sous H_0, $Y_{kl} = \beta + \epsilon_{kl}$ donc $\hat{\xi}_0 = (Y_{..}, \dots, Y_{..})'$.

Sous H_1, $\hat{\xi} = \begin{pmatrix} Y_{1.}\mathbb{1}_{n_1} \\ Y_{2.}\mathbb{1}_{n_2} \\ \vdots \\ Y_{p.}\mathbb{1}_{n_p} \end{pmatrix}$.

On obtient $||\hat{\xi} - \hat{\xi}_0||^2 = \sum_{i=1}^p n_i (Y_{i.} - Y_{..})^2$ et

$$T_n = \frac{n-p}{p-1} \frac{\sum_{i=1}^p n_i (Y_{i.} - Y_{..})^2}{\sum_{k=1}^p \sum l = 1^{n_k} (Y_{kl} - Y_{k.})^2}$$

Sous H_0, $T_n \sim \mathcal{F}(p-1, n-p)$, donc le domaine de rejet est $R = \{T_n \geq f\}$ où f est lu dans la table $\mathcal{F}(p-1, n-p)$.

Sous H_1, $T_n \sim \mathcal{F}(p-1, n-p)$ non centrée, coefficient de non centralité $\delta^2 = \frac{1}{\sigma^2} \sum_{k=1}^p n_k (\beta_k - \bar{\beta})^2$ où $\bar{\beta} = \frac{1}{n} \sum_{i=1}^p n_i \beta_i$.

On construit ce que l'on appelle un tableau d'analyse de la variance (ANOVA) :

	somme des carrés	d^o de liberté	SC/dl	F
inter-classe	$SC_B = \sum_{k=1}^p n_k (Y_{k.} - Y..)^2$	$p-1$	$MS_B = \dfrac{SC_B}{p-1}$	$\dfrac{MS_B}{MS_W}$
intra-classe	$SC_W = \sum_{k=1}^p \sum_{l=1}^{n_k} (Y_{kl} - Yk.)^2$	$n-p$	$MS_W = \dfrac{SC_W}{n-p}$	
total	$SC_T = SC_B + SC_W$ $= \sum_{k=1}^p \sum_{l=1}^{n_k} (Y_{kl} - Y..)^2$	$n-1$		

Exemple : Sur 3 populations socio-économiques, on mesure le taux de cholestérol :

I	403	311	269	336	259					
II	312	222	302	420	420	386	353	210	286	290
III	403	244	353	235	319	260				

On obtient :

	somme des carrés	d^o de liberté	SC/dl	F
inter-classe	1202.5	2	601.2	0.126
intra-classe	85750.5	18	4763.9	
total	86953	20		

Dans la table $\mathcal{F}(2, 18)$ à 95%, on lit $f = 3.55$. Puisque $0.126 < 3.55$, on accepte H_0 et on conclut que le taux de cholestérol ne dépend pas de la classe socio-économique.

Modèle à deux facteurs $Y_{ijk} = \beta_{ij} + \epsilon_{ijk}$ où les ϵ_{ijk} sont iid $\mathcal{N}(0, \sigma^2)$, $1 \leq k \leq c$, $1 \leq j \leq b$, $1 \leq i \leq p$. On pose $n = bpc$.

$$X = \begin{pmatrix} \mathbb{1}_c & & & \\ & \mathbb{1}_c & (0) & \\ & (0) & \ddots & \\ & & & \mathbb{1}_c \end{pmatrix}$$

On écrit :

$$\beta_{ij} = \beta_{..} + (\beta_{i.} - \beta_{..}) + (\beta_{.j} - \beta_{..}) + (\beta_{ij} - \beta_{i.} - \beta_{.j} + \beta_{..})$$

$$\beta_{ij} = \underbrace{\beta_{..}}_{\mu} + \underbrace{(\beta_{i.} - \beta_{..})}_{\alpha_i} + \underbrace{(\beta_{.j} - \beta_{..})}_{\lambda_j} + \underbrace{(\beta_{ij} - \beta_{i.} - \beta_{.j} + \beta_{..})}_{\gamma_{ij}}$$

où α_i représente l'influence du premier facteur et β_j l'influence du deuxième facteur.

On aura un modèle additif si $\beta_{ij} = \beta + \alpha_i + \lambda_j$ où $\sum_{i=1}^p \alpha_i = 0$ et $\sum_{j=1}^b \lambda_j = 0$.

Remarque : le modèle $Y_{ijk} = \beta + \alpha_i + \lambda_j + \epsilon_{ijk}$ n'est pas identifiable sans les hypothèses $\sum_{i=1}^p \alpha_i = 0$ et $\sum_{j=1}^b \lambda_j = 0$. Cependant il l'est avec ces hypothèses. Par ailleurs, le modèle initial $Y_{ijk} = \beta_{ij} + \epsilon_{ijk}$ est lui aussi identifiable, et il est plus facile de calculer ainsi les estimateurs aux moindres carrés. On obtient, $\hat{\beta}_{ij} = Y_{ij.}$ et donc $\hat{\mu} = \frac{1}{bc}\sum_{i=1}^p \sum_{j=1}^b \hat{\beta}_{ij} = Y_{...}$, $\hat{\alpha}_i = \frac{1}{b}\sum_{j=1}^b \hat{\beta}_{ij} = Y_{i..}$ $\hat{\lambda}_j = \frac{1}{p}\sum_{i=1}^p \hat{\beta}_{ij} = Y_{.j.}$.

Si on veut tester, pour le modèle additif, $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$ contre $H_1 : \exists i \in \{1,\ldots,p\}, \alpha_i \neq 0$.

Sous H_0, $Y_{ijk} = \beta + \lambda_j + \epsilon_{ijk}$ avec la condition $\sum_{j=1}^b \lambda_j = 0$, ce qui est un modèle d'analyse de la variance à 1 facteur et à pc observations. Donc $\hat{\xi}_0 = \begin{pmatrix} Y_{.1.}\mathbb{1}_{pc} \\ Y_{.2.}\mathbb{1}_{pc} \\ \vdots \\ Y_{.b.}\mathbb{1}_{pc} \end{pmatrix}$.

On obtient

$$T_n = \frac{n-(p+b-1)}{(p+b-1)-b} \frac{\sum_{i,j,k}(Y_{i..} - Y_{...})^2}{\sum_{i,j,k}(Y_{ijk} - Y_{i..} - Y_{.j.} + Y_{...})^2}$$

$$= \frac{n-p-b+1)}{p-1} \frac{bc\sum_i (Y_{i..} - Y_{...})^2}{\sum_{i,j,k}(Y_{ijk} - Y_{i..} - Y_{.j.} + Y_{...})^2}$$

Sous H_0, $T_n \sim \mathcal{F}(p-1, n-p-b+1)$, donc le domaine de rejet est $R = \{T_n \geq f\}$ où f est lu dans la table $\mathcal{F}(p-1, n-p-b+1)$.

On construit le tableau d'analyse de la variance (ANOVA) :

	somme des carrés	d^o de liberté	SC/dl	F
1er facteur	$SC_1 = bc\sum_i (Y_{i..} - Y\ldots)^2$	$p-1$	$MS_1 = \frac{SC_1}{p-1}$	$\frac{MS_1}{MS_R}$
2ème facteur	$SC_2 = pc\sum_j (Y_{.j.} - Y\ldots)^2$	$b-1$	$MS_2 = \frac{SC_2}{b-1}$	$\frac{MS_2}{MS_R}$
Résidu	$SC_R = \sum_{i,j,k}(Y_{ijk} - Y_{i..}$ $- Y_{.j.} + Y_{...})^2$	$n-p-b+1$	$MS_R = \frac{SC_R}{n-p-b+1}$	
total	$SC_T = \sum_{i,j,k}(Y_{ijk} - Y\ldots)^2$	$n-1$		

Exemple :

	somme des carrés	d^o de liberté	SC/dl	F
1er facteur	66	3	22	1.0
2ème facteur	407	3	135.7	6.3
Résidu	193.9	9	21.5	
total	666.9	15		

Dans la table $\mathcal{F}(3,9)$, on trouve : $f = 3.86$ à 95%, $f = 5.07$ à 97.5% et $f = 6.99$ à 99%. Et on conclut à la non influence du premier facteur à 95% (et donc aussi à 97.5% et 99%) et à l'influence du deuxième facteur à 99%, mais pas à 97.5% (et donc aussi 95%).

Pour le modèle général $Y_{ijk} = \mu + \alpha_i + \lambda_j + \gamma_{ij} + \epsilon_{ijk}$ avec les conditions $\sum_{i=1}^p \alpha_i = 0$, $\sum_{j=1}^b \lambda_j = 0$ et $\sum_{i,j} \gamma_{ij} = 0$, considérons $H_0 : \gamma_{ij} = 0, \forall i,j$ contre $H_1 : \exists i,j, \gamma_{ij} \neq 0$.

On obtient alors

$$T_n = \frac{n-pb}{(p-1)(b-1)} \frac{c\sum_{i,j}(Y_{ij.} - Y_{i..} - Y_{.j.} + Y_{...})^2}{\sum_{i,j,k}(Y_{ijk} - Y_{ij.})^2}$$

Sous H_0, $T_n \sim \mathcal{F}((b-1)(p-1), n-pb)$, donc le domaine de rejet est $R = \{T_n \geq f\}$ où f est lu dans la table $\mathcal{F}((b-1)(p-1), n-pb)$. On construit ainsi un autre tableau d'analyse de la variance auquel on rajoute une ligne correspondant aux interactions.

Remarque : Si $c = 1$, $n = pb$ et ce test n'est pas utilisable.

6.4 Intervalles de confiance simultanés

Exemple : $Y_{ij} = \beta_i + \epsilon_{ij}$, $i = 1, \ldots, 4$. On sait obtenir des intervalles de confiance indépendamment sur β_1, β_2, β_3 et β_4. La question est : peut-on en déduire des relations du type : $\beta_1 < \beta_2$ ou $\beta_3 > \beta_1$? Le problème est que ces intervalles de confiance n'ayant pas été calculés simultanément, $P(\beta_i \in A_i) = 1 - \alpha$ mais $P(\beta_1 \in A_1, \ldots, \beta_4 \in A_4)$ est certainement plus petit que $1 - \alpha$ (dans le cas d'indépendance, pour $\alpha = 5\%$, on obtient environ 81%).

On veut donc construire une région de confiance $A = \prod_{k=1}^{4} A_k$ avec $P(\beta_k \in A_k, k = 1, \ldots, 4) = 1 - \alpha$.

Une première possibilité consiste à utiliser l'inégalité de Bonferoni :

$$P(\cap_{k=1}^{K} A_k) \geq 1 - \sum_{k=1}^{K} P(A_k^c)$$

et donc

$$P(\beta_k \in A_k, k = 1, \ldots, K) \geq 1 - \sum_{k=1}^{K} P(\beta_k \in A_k^c) = 1 - K\alpha = 1 - \alpha' \Longrightarrow \alpha = \alpha'/K$$

Cependant cette minoration est grossière et conduit à des intervalles de confiance très larges.

6.4.1 Méthode de Tuckey

$Y_{ij} = \beta_i + \epsilon_{ij}$, $1 \leq i \leq p$, $1 \leq j \leq c$.

On va considérer deux problèmes :

Problème I : trouver $\{I_i, 1 \leq i \leq p\}$ tel que $P(\beta_i \in I_i, 1 \leq i \leq p) \geq 1 - \alpha$.

Problème II : trouver $\{I_{ij}, 1 \leq i < j \leq p\}$ tel que $P(\beta_j - \beta_i \in I_{ij}, 1 \leq i < j \leq p) \geq 1 - \alpha$.

(I) Il est naturel de rechercher une constante k telle que

$$P = P\left(Y_{i.} - k\frac{S}{\sqrt{c}} \leq \beta_i \leq Y_{i.} + k\frac{S}{\sqrt{c}}, 1 \leq i \leq p\right) = 1 - \alpha$$

$$\text{où } S^2 = \frac{1}{(c-1)p}\sum_{i=1}^{p}\sum_{j=1}^{c}(Y_{ij} - Y_{i.})^2$$

$$\text{alors } P = P\left(-k \leq \frac{\sqrt{c}(Y_{i.} - \beta_i)}{S} \leq k, 1 \leq i \leq p\right) = P\left(\max_{1 \leq i \leq p}\left|\frac{\sqrt{c}(Y_{i.} - \beta_i)}{S}\right| \leq k\right)$$

On a vu que les variables $V_i = \sqrt{c}(Y_{i.} - \beta_i)/\sigma$ sont iid $\mathcal{N}(0, 1)$ et sont indépendantes de $W = (c-1)pS^2/\sigma^2 \sim \chi^2_{(c-1)p}$.

Définition : $\max_{1 \leq i \leq p} |V_i|/W$ suit une loi de student du module maximum de paramètre $(p, (c-1)p)$.

Il existe des tables pour ces lois, et on trouve donc k dans ces tables.

(II) On commence de la même manière :

$$P = P\left(Y_{i.} - Y_{j.} - d\frac{S}{\sqrt{c}} \leq \beta_i - \beta_j \leq Y_{i.} - Y_{j.} + d\frac{S}{\sqrt{c}}, 1 \leq i < j \leq p\right) = 1 - \alpha$$

$$= P\left(\max_{1 \leq i < j \leq p}\left|\frac{V_i - V_j}{S/\sigma}\right| \leq d\right)$$

Or $\max_{1 \leq i < j \leq p}\left|\frac{V_i - V_j}{S/\sigma}\right| = (\max_i V_i - \min_i V_i)/(S/\sigma)$.

Définition 6.7 : *$(\max_i V_i - \min_i V_i)/(S/\sigma)$ suit une loi du rang de student de paramètre $(p, (c-1)p)$.* Il existe des tables pour ces lois, et on trouve donc d dans ces tables.

6.4.2 Méthode de Scheffé

Supposons que, dans l'exemple précédent, on veuille comparer $(\beta_1+\beta_2)/2$ et $(\beta_3+\beta_4)/2$. On considère donc $\phi=\sum_{j=1}^p a_j\beta_j$ où $\sum_{j=1}^p a_j=0$. ϕ est appelé un contraste.

On peut encore écrire $\phi=\sum_{j=1}^p \omega_j\alpha_j$ où $\alpha_j=\beta_j-\beta.$ car $\phi=\sum_{j=1}^p a_j(\alpha_j+\beta.)=\sum_{j=1}^p a_j\alpha_j+\beta.\sum_{j=1}^p a_j=\sum_{j=1}^p a_j\alpha_j$.

On recherche un intervalle de confiance simultané pour tous les w_i. On commence par rechercher un intervalle de confiance pour $\sum_{j=1}^p \omega_j\alpha_j$. Soient $\hat\alpha_i=Y_{i.}-Y_{..}$ et $\hat\phi=\sum_{j=1}^p \omega_j\hat\alpha_j$, on vérifie alors que $I_i=[\hat\phi-KST(w)/\sqrt{c},\hat\phi+KST(w)/\sqrt{c}]$ où $T^2(w)=\sum_{j=1}^p(w_i-\bar w)^2$ est un intervalle de confiance pour ϕ.

$$P(\hat\phi-KST(w)/\sqrt{c}\leq\phi\leq\hat\phi+KST(w)/\sqrt{c},\forall w\in\mathbb{R}^p)$$

$$=P(\max_{w\in\mathbb{R}^p}\frac{\sqrt{c}|\hat\phi-\phi|}{ST(w)}\leq K)\leq P(\frac{c\sum(\hat\alpha_i-\alpha_i)^2}{(p-1)S^2}\leq\frac{K^2}{p-1})$$

en effet, $|\hat\phi-\phi|=|\sum w_j(\hat\alpha_i-\alpha_i)|\leq\sqrt{\sum(w_j-\bar w)^2}\sqrt{\sum(\hat\alpha_j-\alpha_j)^2}$. De plus, il y a égalité dans l'inégalité précédente ssi $(w_j)_j$ est colinéaire à $(\hat\alpha_j-\alpha_j)_j$.

Pour conclure, on remarque que $\sum_j(\hat\alpha_j-\alpha_j)^2=\sum_j(Y_{j.}-Y_{..}-\beta_j+\beta.)^2$ donc

$$\frac{c\sum(\hat\alpha_i-\alpha_i)^2}{(p-1)S^2}\sim\mathcal{F}(p-1,(c-1)p)$$

d'où l'intervalle de confiance annoncé où $K=\sqrt{(p-1)f}$ et f est lu dans la table $\mathcal{F}(p-1,(c-1)p)$.

Interprétation géométrique :

$$\sum_{j=1}^p\omega_j\alpha_j\in I_w,\forall w\in\mathbb{R}^p\Longleftrightarrow\frac{c}{p-1}\frac{\sum(\hat\alpha_i-\alpha_i)^2}{S^2}\leq f\Longleftrightarrow\sum(\hat\alpha_i-\alpha_i)^2\leq\frac{f(p-1)S^2}{c}$$

Il s'agit donc d'une sphère dans $\mathbb{R}^p$. Avec la condition $\sum_{j=1}^p\alpha_j=0$, on obtient une hypersphère, c'est à dire l'intersection d'une sphère et d'un hyperplan.

Remarque : La méthode de Tuckey donne en général des intervalles de confiance plus grands que celle de Scheffé pour des combinaisons linéaires complexes des β_i, mais des intervalles plus petits pour celles qui sont simples ($\beta_i-\beta_j$ par exemple).

Chapter 7

Analyse de donnés discrètes

7.1 Test d'ajustement, hypothèses simples

Exemples :

1. On dispose d'un générateur de nombres aléatoires qui donne des nombres entiers compris entre 0 et 9 avec une probabilité uniforme : $p_i = 1/10, i \in \{0, \ldots, 9\}$.

   On veut tester ce générateur : on observe $X_1, \ldots, X_n$ sorties successives et on veut tester $H_0 : p_i = 1/10, \forall i$ contre $H_1 : (p_0, \ldots, p_9) \neq (1/10, \ldots, 1/10)$.

2. Vérification d'une théorie sur la répartition sur des croisements d'animaux ou de plantes par exemple.

Dans les deux cas, $X_1, \ldots, X_n$ est un n-échantillon, prenant ses valeurs dans un ensemble discret connu $\{v_1, \ldots, v_k\}$ et dont les probabilités d'occurrence $\theta_1, \ldots, \theta_k$ sont inconnues, et on veut tester : $H_0 : \theta_i = \theta_{i0}, \forall i \in \{1, \ldots, k\}$ contre $H_1 : \theta_i \neq \theta_{i0}$ pour au moins un i.

On note N_i le nombre d'occurrences de v_i dans le n-uplet $(x_1, \ldots, x_n)$ et on obtient :

$$P(X_1 = x_1, \ldots, X_n = x_n) = \theta_1^{n_1} \ldots \theta_k^{n_k}$$

On a déjà vu en particulier que $(N_1, \ldots, N_k)$ est une statistique exhaustive pour $\theta = (\theta_1, \ldots, \theta_k)$, de plus $N \sim \mathcal{M}(n, \theta_1, \ldots, \theta_k)$ et $\Theta = \{\theta \mid \forall i, \theta_i \geq 0 et \sum_{i=1}^k \theta_i = 1\}$.

On calcule le rapport de vraisemblance :

$$P_\theta(n_1, \ldots, n_k) = \frac{n!}{n_1! \ldots n_k!} \theta_1^{n_1} \ldots \theta_k^{n_k}$$

$$\text{et } \sup_{\theta \in \Theta} P_\theta(n_1, \ldots, n_k) = P_{\hat{\theta}}(n_1, \ldots, n_k)$$

où $\hat{\theta}$ est l'estimateur du maximum de vraisemblance. En remarquant que $\theta_k = 1 - \sum_{i=1}^{k-1} \theta_i$, on dérive en $\theta_1, \ldots, \theta_{k-1}$ et on obtient les équations normales :

$$\sum_{i=1}^k \frac{\partial}{\partial \theta_i} \log \theta_j^{n_j} = \sum_{i=1}^k \frac{n_j}{\theta_l} \frac{\partial \theta_j}{\partial \theta_i} = 0, \forall i \in \{1, \ldots, k-1\}$$

d'où $\hat{\theta}_k = \hat{\theta}_i = n_k/n_i, j = 1, \ldots, k-1$ et $\hat{\theta}_i = n_i/n, i = 1, \ldots, k$.

Le rapport de vraisemblance s'écrit alors $L(n_1, \ldots, n_k) = P_{\hat{\theta}}(n)/P_{\theta_0}(n)$ et

$$\log L(n) = \log(P_{\hat{\theta}}(n)) - \log(P_{\theta_0}(n)) = \sum_{i=1}^k n_i[\log \hat{\theta}_i - \log \theta_{i0}] = \sum_{i=1}^k n_i[\log(n_i/n) - \log \theta_{i0}].$$

Le domaine de rejet associé au test du rapport de vraisemblance est donc $R = \{\sum_{i=1}^k n_i[\log(n_i/n) - \log\theta_{i0}] > u\}$ cependant la loi de cette statistique n'est pas simple.

En fait, on utilise le plus souvent la statistique du χ^2 :

$$X^2 = \sum_{i=1}^k \frac{(N_i - n\theta_i)^2}{n\theta_{i0}}$$

qui a l'intérêt d'avoir une loi qui peut être approximée par un χ^2_{k-1}.

On peut vérifier que, sous H_0 et quand n est grand, X^2 est proche de $2\log L$, en effet :

$$\log\hat{\theta}_i - \log\theta_{i0} \approx \frac{(\hat{\theta}_i - \theta_{i0})}{\theta_{i0}} - \frac{(\hat{\theta}_i - \theta_{i0})^2}{2\theta_{i0}^2} \approx \frac{N_i - n\theta_{i0}}{n\theta_{i0}} - \frac{(N_i - n\theta_{i0})^2}{n^2\theta_{i0}^2}$$

$$\text{donc } 2\sum_{i=1}^k N_i(\log\hat{\theta}_i - \log\theta_{i0}) \approx 2\sum_{i=1}^k \left(\frac{(N_i - n\theta_{i0})^2}{n\theta_{i0}} - \frac{(N_i - n\theta_{i0})^3}{2n^2\theta_{i0}^2}\right) + 2\sum_{i=1}^k \left((N_i - n\theta_{i0}) - \frac{(N_i - n\theta_{i0})^2}{2n\theta_{i0}}\right)$$

en remplaçant N_i par $N_i - n\theta_{i0} + n\theta_{i0}$, $\sum_{i=1}^k N_i$ par $\sum_{i=1}^k n\theta_{i0}$ et en regroupant les termes deux à deux, on obtient alors :

$$\sum_{i=1}^k \frac{(N_i - n\theta_{i0})^2}{n\theta_{i0}} - \sum_{i=1}^k \frac{(N_i - n\theta_{i0})^3}{n^2\theta_{i0}^2}$$

$$\text{Or } \left|\sum_{i=1}^k \frac{(N_i - n\theta_{i0})^3}{n^2\theta_{i0}^2}\right| \leq \sum_{i=1}^k \frac{(N_i - n\theta_{i0})^2}{n\theta_{i0}} \times \max_i \left|\frac{\hat{\theta}_i - \theta_{i0}}{\theta_{i0}}\right|$$

et $|\hat{\theta}_i - \theta_{i0}| \longrightarrow 0$ en probabilité sous H_0 par la loi forte des grands nombres et donc $2\log L \approx X^2$.

Il reste a étudier la loi de X^2 ce qui découle de la proposition suivante.

Proposition 7.1 : *Soit $N = (N_1, \ldots, N_k)$ de loi $\mathcal{M}(n, \theta)$, la suite des v.a. $\left(\frac{N_1 - n\theta_1}{\sqrt{n\theta_1}}, \ldots, \frac{N_k - n\theta_k}{\sqrt{n\theta_k}}\right)$ converge en loi, quand n tend vers $+\infty$, vers $\mathcal{N}(0, \Gamma)$, où $\Gamma = Id - (\sqrt{\theta_1}, \ldots, \sqrt{\theta_k})'(\sqrt{\theta_1}, \ldots, \sqrt{\theta_k})$.*

Admettons cette proposition, on en déduit que, pour $k = 2$, $X^2 = \frac{(N_1 - n\theta_{10})^2}{n\theta_{10}(1 - \theta_{10})}$ or $\frac{N_1 - n\theta_{10}}{\sqrt{n\theta_{10}(1 - \theta_{10})}}$ tend en loi vers $\mathcal{N}(0, 1)$ et donc X^2 tend en loi vers χ^2_1. De même pour k quelconque, par application de la même proposition, X^2 tend en loi vers un χ^2_{k-1}.

Le domaine de rejet du test est alors de la forme $\{X^2 > x\}$ et x est approximé par la loi χ^2_{k-1} (pour n grand).

Remarque : On peut se poser le problème d'un ajustement dans le cas continu, on peut alors faire des regroupements en classes et utiliser un test d'ajustement du χ^2 mais il est clair que l'on perd de l'information par ce regroupement. On verra un autre test plus tard (Kolmogorov) ainsi que comment faire ces regroupements.

7.2 Ajustement à une famille de lois. Tables de contingence

On considère $N = (N_1, \ldots, N_k)$ où $N \sim \mathcal{M}(n, \theta)$ et on veut tester : $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$, avec $\Theta = \{\theta \mid \theta_i \geq 0, \forall i \text{ et } \sum_{i=1}^k \theta_i = 1\}$.

Exemples :

1. Modèle génétique : Soit θ_i la probabilité d'apparition d'un certain phénotype lors d'un croisement. Par exemple, sur 4 phénotypes, on fait l'hypothèse :

$$\begin{cases} \theta_1 = (2 + \eta)/4 \\ \theta_2 = \theta_3 = (1 - \eta)/4 \quad \text{où } \eta \text{ est un paramètre appartenant à } [0, 1] \\ \theta_4 = \eta/4 \end{cases}$$

   et donc $\Theta_0 = \{(2 + \eta)/4, (1 - \eta)/4, (1 - \eta)/4, \eta/4), \eta \in [0, 1]\}$.

2. Quand on a supposé dans les chapitres précédents que nos observations suivaient des lois $\mathcal{N}(\mu, \sigma^2)$ ou $\mathcal{E}(\lambda)$ par exemple, il faudrait pouvoir tester ces hypothèses. Par exemple, pour $\mathcal{E}(\lambda)$, en faisant des regroupements, on peut construire $\theta_j = e^{-a_{j-1}\lambda} - e^{-a_j\lambda}$, $1 \leq j \leq k$, $a_0 = 0$, $a_k = +\infty$ et $\Theta_0 = \{(1 - e^{-a_1\lambda}, 1 - e^{-a_2\lambda}, \ldots, 1 - e^{-a_{k-1}\lambda}), \lambda > 0\}$

3. Beaucoup de problèmes se posent en termes binaires : un individu est ou n'est pas atteint par une maladie, est ou n'est pas fumeur, est home ou femme, ... On peut vouloir savoir si de telles caractéristiques sont liées ou non. On notera $\bar{A}$ l'évènement contraire de A.

   On construit ainsi 4 classes : AB, $A\bar{B}$, $\bar{A}B$, $\bar{A}\bar{B}$ et on construit $\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}$ les probabilités correspondantes. L'indépendance des deux évènements A et B peut alors être caractérisée par : $\theta_{ij} = (\theta_{i1} + \theta_{i2})(\theta_{1j} + \theta_{2j})$ pour $i \neq j$.

   On construit alors une table de contingence d'ordre 2 :

	B	$\bar{B}$
A	N_{11}	N_{12}
$\bar{A}$	N_{21}	N_{22}

   Donc $N = (N_{11}, N_{12}, N_{21}, N_{22}) \sim \mathcal{M}(n, \theta_{11}, \theta_{12}, \theta_{21}, \theta_{22})$ et on teste : $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \notin \Theta_0$, avec $\Theta_0 = \{(\eta_1\eta_2, \eta_1(1-\eta_2), (1-\eta_1)\eta_2, (1-\eta_1)(1-\eta_2)) \mid \eta_1 \in [0,1], \eta_2 \in [0,1]\}$.

On se place dans les cas où $\Theta_0 = \{(\theta_1(\eta), \ldots, \theta_k(\eta) \mid \eta \in E\}$ avec $E \subset \mathbb{R}^r$ et $r < k-1$ (paramétrage). On calcule alors le rapport de vraisemblance :

$$\sup_{\theta \in \Theta} P_\theta(n_1, \ldots, n_k) = P_{\hat{\theta}}(n_1, \ldots, n_k) \quad \text{où} \quad \hat{\theta} = (\frac{n_1}{n}, \ldots, \frac{n_k}{n})$$

$$\sup_{\theta \in \Theta_0} P_\Theta(n_1, \ldots, n_k) = \sup_{\eta \in E} P_{\theta(\eta)}(n_1, \ldots, n_k) = P_{\theta(\hat{\eta})}(n_1, \ldots, n_k)$$

$$\text{et donc } \log L(n_1, \ldots, n_k) = \sum_{i=1}^{k} n_i \left[\log \frac{n_i}{n} - \log \theta_i(\hat{\eta})\right]$$

Si $\eta \longrightarrow \theta(\eta)$ admet des dérivées partielles, $\hat{\eta}$ vérifie :

$$\frac{\partial}{\partial \eta_j} \log P_{\theta(\eta)}(n_1, \ldots, n_k) = 0, 1 \leq j \leq r$$

$$\frac{\partial}{\partial \eta_j} \left(\sum_{i=1}^{k} n_i \log \theta_i(\eta)\right) = 0$$

$$\sum_{i=1}^{k} \frac{n_i}{\theta_i(\eta)} \frac{\partial}{\partial \eta_j} \theta_i(\eta) = 0, 1 \leq j \leq r$$

En posant $\theta_i(\hat{\eta}) = \hat{\theta}_i + r_n$, où r_n tend en loi vers 0 quand n tend vers $+\infty$ sous H_0.

On peut vérifier : $2 \log L(N_1, \ldots, N_k) \approx \displaystyle\sum_{i=1}^{k} \frac{(N_i - n - th_i(\hat{\eta}))^2}{n\theta_i(\hat{\eta})} = X^2$ qui est la statistique du χ^2.

Sauf dans des cas particuliers, il est en général difficile de calculer les lois de $2 \log L$ ou de X^2, donc les domaines de rejet. Sous certaines conditions, on peut montrer que, sous H_0, $X^2 \approx \chi^2_{k-r-1}$.

Exemples : Reprenons les trois exemples précédents :

1. L'équation normale $\dfrac{n_i}{\theta_i(\eta)} \dfrac{\partial}{\partial \eta_j} \theta_i(\eta) = 0$ avec $j = 1$ devient :

$$\frac{n_1}{(2+\eta)/4} \frac{1}{4} + \frac{n_2}{(1-\eta)/4} \frac{-1}{4} + \frac{n_3}{(1-\eta)/4} \frac{-1}{4} + \frac{n_4}{(\eta/4)} \frac{1}{4} = 0$$

$$\Longrightarrow \frac{n_1}{2+\eta} - \frac{n_2+n_3}{1-\eta} + \frac{n_4}{\eta} = 0 \Longrightarrow -n\eta^2 + \eta(n_1 - 2(n_2+n_3) - n_4) + 2n_4 = 0$$

Le discriminent est $\Delta = (n_1 - 2(n_2+n_3) - n_4)^2 + 8nn_4 > 0$ et on obtient deux racines :

$$\eta_1 = \frac{(n_1 - 2(n_2+n_3) - n_4) - \sqrt{\Delta}}{2n} \leq 0 \quad \text{et} \quad \eta_2 = \frac{(n_1 - 2(n_2+n_3) - n_4) + \sqrt{\Delta}}{2n} \geq 0$$

car $\sqrt{\Delta} > |n_1 - 2(n_2+n_3) - n_4|$. Il y a donc une seule racine positive qui correspond bien a un maximum et donc $\hat{\eta} = \eta_2$. Il suffit alors de remplacer pour avoir $\log L$ et, sous H_0, comme précédemment $X^2 \approx \chi_{4-1-1} = \chi^2_2$.

2. On obtient cette fois :

$$\sum_{j=1}^{k} n_j \frac{a_{j-1}e^{-a_{j-1}\lambda} - a_j e^{-a_j\lambda}}{e^{-a_{j-1}\lambda} - e^{-a_j\lambda}} = 0$$

qu'on ne peut résoudre analytiquement.

3. On obtient :

$$\begin{cases} \dfrac{n_{11}+n_{12}}{\hat{\eta}_1} = \dfrac{n_{21}+n_{22}}{1-\hat{\eta}_1} \\ \dfrac{n_{11}+n_{21}}{\hat{\eta}_2} = \dfrac{n_{12}+n_{22}}{1-\hat{\eta}_2} \end{cases} \quad \text{d'où} \quad \begin{cases} \hat{\eta}_1 = \dfrac{n_{11}+n_{12}}{n} \\ \hat{\eta}_2 = \dfrac{n_{12}+n_{21}}{n} \end{cases}$$

et $X^2 = n \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(N_{ij} - R_iC_j/n)^2}{R_iC_j}$ où $R_i = N_{i1} + N_{i2}$ et $C_j = N_{1j} + N_{2j}$. Sous H_0, $k = 4$, $r = 2$ et $X^2 \approx \chi^2_1$.

On peut généraliser à des tables pour lesquelles chaque couple d'observations est formé des variables X et Y. X pouvant prendre p valeurs et Y, b valeurs.

$$\text{Ainsi} \quad X^2 = n \sum_{i=1}^{p}\sum_{j=1}^{b} \frac{(N_{ij} - \frac{R_iC_j}{n})^2}{R_iC_j} \quad \text{où} \quad R_i = \sum_{j=1}^{b} N_{ij} \quad \text{et} \quad C_j = \sum_{i=1}^{p} N_{ij}$$

$$\text{ou encore} \quad X^2 = n \sum_{i=1}^{p}\sum_{j=1}^{b} \frac{(\frac{N_{ij}}{n} - N_i.N.j)^2}{N_i.N.j} \quad \text{et} \quad X^2 \sim \chi^2_{(p-1)(b-1)} \quad \text{sous } H_0.$$

7.3 Applications

1. Essayer de déterminer si un groupe d'animaux expérimentaux répond de manière homogène à une maladie donnée. On observe 10 échantillons de 5 souris :

échant. i	1	2	3	4	5	6	7	8	9	10
Nb de décés (sur 5)	3	1	5	5	3	2	4	2	3	5

Alors $N_i \sim \mathcal{B}(c_i, \theta_i)$, $c_i = 5$, $\forall i$. De plus les N_i sont supposés indépendants et on veut tester H_0 : $\forall i, \theta_i = \theta$ contre $H_1 : \exists i, j$ tq $\theta_i \neq \theta_j$.

On considère $X^2 = \sum_{i=1}^{p} \frac{(N_i - c_i\hat{\theta}_.)^2}{c_i\hat{\theta}_.(1-\hat{\theta}_.)}$ où $n = c_1 + \ldots c_p$ et $\hat{\theta}_. = \frac{1}{n}\sum_{i=1}^{p} N_i$. Si n est grand et sous H_0, $X^2 \approx \chi^2_{p-1}$.

Ici, $X^2 = \sum_{i=1}^{p} \frac{(N_i - N_.)^2}{c\hat{\theta}_.(1-\hat{\theta}_.)}$ car $\forall i$, $c_i = c$ et $c\hat{\theta}_. = N_.$. On observe $n_. = 3.3$ et $x^2 = 16.1$ et on accepte H_0.

2. Soit $\theta(x)$ la proportion d'individus qui utilisent les transports en commun plutôt que leur véhicule quand la distance pour rejoindre un transport en commun est réduite de $x$km. On choisit $x_1, \ldots, x_p$ et, pour chaque x_i, on observe c_i individus. On peut alors supposer que $N_i \sim \mathcal{B}(c_i, \theta(x_i))$.

   Par commodité, on considère (modèle "Logit") que $\log \dfrac{\theta(x)}{1-\theta(x)} = \beta_1 + \beta_2 x \Longleftrightarrow \theta(x) = \dfrac{1}{1+e^{-(\beta_1+\beta_2 x)}}$ avec $\beta_2 > 0$ de manière que la fonction θ soit croissante.

   Soit $Y_i = \log \dfrac{\hat{\theta}_i}{1-\hat{\theta}_i)}$, alors $Y_i = \beta_1 + \beta_2 x + \epsilon_i$. On suppose que les ϵ_i sont indépendants, de loi $\mathcal{N}(0, \sigma_i^2/c_i)$. C'est un modèle de régression, mais les variances varient. On pose $w_i = \sigma_i/\sqrt{c_i}$ et on obtient $Y_i/w_i = \beta_1/w_i + \beta_2 x/w_i + \epsilon'_i$ où cette fois, les ϵ'_i sont indépendants, de loi $\mathcal{N}(0,1)$.

   On estime w_i par le maximum de vraisemblance : $\hat{w}_i = \dfrac{1}{\sqrt{c_i \hat{\theta}_i(1-\hat{\theta}_i)}}$, puis β_1 et β_2 par les moindres carrés : $\hat{\beta}_2 = \dfrac{\sum_{i=1}^p s_i Y_i (x_i - \tilde{x})}{\sum_{i=1}^p s_i (x_i - \tilde{x})^2}$ et $\beta_1 = \tilde{Y} - \hat{\beta}_2 \tilde{x}$ où $\tilde{Y} = \dfrac{\sum_{i=1}^p s_i Y_i}{\sum_{i=1}^p s_i}$, $\tilde{x} = \dfrac{\sum_{i=1}^p s_i x_i}{\sum_{i=1}^p s_i}$ et $s_i = 1/\hat{w}_i$.

   On peur aussi rechercher des intervalles de confiance de la même manière.

Chapter 8

Modèles non paramétrés

8.1 Méthode du rang pour comparer deux échantillons (Wilcoxon)

Soient $X_1, \dots, X_{n_1}$ et $Y_1, \dots, Y_{n_2}$ deux échantillons de lois continues F et G. On s'intéresse à des tests de type $H_0 : F = G$ contre $H_1 : Y$ est "plus grand" que X.

On ordonne simultanément les deux échantillons et on construit $R_1, \dots, R_n$ avec $n = n_1 + n_2$ de la manière suivante :

$$\begin{array}{ll} R_1 = \text{rang de } X_1 & R_{n_1+1} = \text{rang de } Y_1 \\ R_2 = \text{rang de } X_2 & R_{n_1+2} = \text{rang de } Y_2 \\ \vdots & \vdots \\ R_{n_1} = \text{rang de } X_{n_1} & R_{n_1+n_2} = \text{rang de } Y_{n_2} \end{array}$$

On construit $W_X = \sum_{i=1}^{n_1} R_i$ et $W_Y = \sum_{i=n_1+1}^{n_1+n_2} R_i$. Sous H_0, R suit une loi uniforme sur Σ_n (ensemble des permutations de $\{1, \dots, n\}$, qui est de cardinal $n!$) :

$$P(R_1 = r_1, \dots, R_n = r_n) = 1/n!.$$

Remarques :

- Sous H_0, les lois de R, W_X et W_Y sont indépendantes, donc les lois de W_X et W_Y peuvent être (sont) tabulées.
- $P(R_{n_1+1} = s_1, \dots, R_{n_1+n_2} = s_{n_2}) = \dfrac{n_1! n_2!}{n!}$, en effet, chaque n-uplet a pour probabilité $1/n!$ et il y a $n_1!$ possibilités de choix l'ordre dans lequel sont obtenus les X_i et $n_2!$ possibilités de choix de choix pour les Y_i.
- On notera $S_1, \dots, S_{n_2}$ les rangs ordonnés des Y_i.

Exemple : Entretien d'un ventilateur suivant deux plans : PI et PII. On observe : $X_1, \dots, X_5$ l'espérance de vie avant panne lorsque le plan PI est utilisée et $Y_1, \dots, Y_5$ l'espérance de vie avant panne pour PII.

x_i	7	26	10	8	29
y_i	3	150	40	34	32

Alors $w_Y = 1 + 7 + 8 + 9 + 10 = 35$ et on peut calculer $P(W_Y \geq 35)$:

$P(W_Y = 40) = P(S = (6, 7, 8, 9, 10)) = 1/252$
$P(W_Y = 39) = P(S = (5, 7, 8, 9, 10)) = 1/252$
$P(W_Y = 38) = P(S = (4, 7, 8, 9, 10)) + P(S = (5, 6, 8, 9, 10)) = 2/252$
... etc

On obtient $P(W \geq 35) = 19/252 = 0.0754$, on accepte H_0 à 5% et on rejette H_0 à 10%.

Théorème 8.1 : *Supposons que F et G soient continues et que $0 < P(X < Y) < 1$ et soit U le nombre de couples (X_i, Y_j) tels que $X_i < Y_j$, alors*

$$P\left(\frac{U - E(U)}{\sqrt{Var(U)}} \leq t\right) \longrightarrow \Phi(t) \quad quand \quad min(n_1, n_2) \longrightarrow +\infty$$

avec $E(U) = n_1 n_2/2$ et $Var(U) = n_1 n_2 (n+1)/12$ sous l'hypothèse H_0.

Remarque : L'approximation est utilisable en pratique dès que n_1 et n_2 sont supérieurs ou égaux à 8.

On peut aussi construire des intervalles de confiance dans le cas gaussien : Soient $X_1, \ldots, X_{n_1}$ et $Y_1, \ldots, Y_{n_2}$ deux échantillons de lois $\mathcal{N}(\mu, \sigma^2)$ et $N(\mu + \Delta, \sigma^2)$. On va commencer par tester : $H_0 : \Delta = \Delta_0$ contre $H_1 : \Delta \neq \Delta_0$.

On construit un test de Wilcoxon bilatère, le domaine de rejet est donc $|U(X, Y - \Delta_0) - n_1 n_2/2| \geq c$ d'où $P(k_\alpha \leq U(X, Y - \Delta) \leq n_1 n_2 - k_\alpha) = 1 - \alpha$ et $I = \{\Delta \mid k_\alpha \leq U(X, Y - \Delta) \leq n_1 n_2 - k_\alpha\}$ est un intervalle de confiance au niveau $1 - \alpha$.

Soient $D_1 < D_2 < \ldots < D_{n_1 n_2}$ les différences $Y_j - Y_i$ rangées dans l'ordre croissant, alors $I = [D_{k_\alpha}, D_{n_1 n_2 - k_\alpha + 1}]$. En effet, $I = \{\Delta \mid U(X, Y - \Delta) \leq n_1 n_2 - k_\alpha\} \cap \{\Delta \mid U(X, Y - \Delta) \leq k_\alpha - 1\}^c$ or $\{\Delta \mid U(X, Y - \Delta) \leq k\} = [D_{n_1 n_2 - k}; +\infty[$ puisque $U(X, Y - \Delta)$ est le nombre de couples (i, j) tels que $(Y_j - \Delta) - X_i > 0$ c'est à dire $Y_j - X_i > \Delta$ et que donc $U(X, Y - \Delta) \leq k \Longleftrightarrow \Delta \geq D_{n_1 n_2 - k}$.

Exemple :

x_i	80	85	60	75	58	73
y_i	84	100	81	90	76	91

On a $P(U \geq 5) = 0.987$ d'où on déduit $P(5 \leq U \leq 31) = 0.974$ donc $I = [D_5; D_{32}]$ est un intervalle de confiance à 95.4%, on obtient $[1, 31]$:

	76	81	84	90	91	100
58	18	23	26	32	33	42
60	16	21	24	30	31	40
73	3	8	11	17	18	27
75	1	6	9	15	16	25
80	-4	1	4	10	11	20
85	-9	-4	-1	5	6	15

8.2 Test du signe et test signe et rang

On considère n couples indépendants (X_i, Y_i), $i = 1, \ldots, n$ et $Z_i = Y_i - X_i$, on note F la loi de Z_i, que l'on suppose continue. On veut tester :

H_0 : la loi de Z_i est symétrique en 0, c'est à dire $F(t) = 1 - F(-t)$, $\forall t$

contre H_1 : Z_i est "plus grand" que $-Z_i$, c'est à dire $P(Y_i - X_i > t) \geq P(X_i - Y_i > t)$.

8.2.1 Test du signe

On définit S le nombre de $Z_i = Y_i - X_i$ qui sont positifs ou nuls. Alors, $S \sim \mathcal{B}(n, 1 - F(0))$ et sous H_0, $F(0) = 1/2$ alors que sous H_1, $F(0) < 1/2$.

Remarque : si $n \geq 10$, on approxime la loi binomiale par la loi gaussienne $\mathcal{N}(n/2, n/4)$ (sous H_0).

8.2.2 Test signe et rang

On considère maintenant R_i le rang de $|Z_i|$ dans la suite des $|Z_1|, \ldots, |Z_n|$ rangés dans l'ordre croissant, multiplié par le signe de Z_i.

Exemple : Soient les valeurs observées :

$y_i - z_i$	1.1	-2.0	-3.1	-1.3
r_i	1	-3	-4	-2

On pose $W = \sum_{i=1}^{s} T_i = \frac{1}{2}\sum_{i=1}^{n} R_i + n(n+1)/4$ où $T_1 < \ldots < T_s$ sont les R_i positifs ordonnés dans l'ordre croissant et donc s est le nombre de R_i positifs. On considère de plus $I_k = 1$ si $Z_k > 0$ et $I_k = 0$ sinon.

Théorème 8.2 : *Sous H_0 : F est continue, symétrique en 0*

1. $I_1, \ldots, I_n$ *sont iid de loi* $\mathcal{B}(1, 1/2)$.
2. $(|R_1|, \ldots, |R_n|)$ *et* $(I_1, \ldots, I_n)$ *sont indépendantes.*
3. $P(|R_1| = |r_1|, \ldots, |R_n| = |r_n|) = 1/n!, \forall(|r_1|, \ldots, |r_n|)$.
4. $I_1, \ldots, I_n$ *sont iid de loi* $\mathcal{B}(1, 1/2)$.
5. $P(R_1 = r_1, \ldots, R_n = r_n) = 1/2^n n!$.
6. $P(T_1 = t_1, \ldots, T_s = t_s, S = s) = 1/2^n$ *où* $S = \sum_{j=1}^{n} I_j$.

Dém :

1. évident sous H_0.
2. Remarquons tout d'abord que $P(|Z_1| \leq t, I_1 = 1) = P(0 < Z_1 \leq t) = \frac{1}{2}P(|Z_1| \leq t)$ sous H_0. Ensuite
$$P(|Z_1| \leq t_1, \ldots, |Z_n| \leq t_n, I_1 = 1, \ldots, I_n = 1) = P(0 < Z_1 \leq t_1, \ldots, 0 < Z_n \leq t_n)$$
$$= \prod_{i=1}^{n} P(0 < Z_i \leq t_i) = \prod_{i=1}^{n} \frac{1}{2} P(|Z_i| \leq t_i) = \frac{1}{2^n} \prod_{i=1}^{n} P(|Z_i| \leq t_i)$$
donc $(|Z_1|, \ldots, |Z_n|)$ et $(I_1, \ldots, I_n)$ sont indépendantes et par suite $(|R_1|, \ldots, |R_n|)$ et $(I_1, \ldots, I_n)$ sont indépendantes.
3. Démontré dans la section précédente.
4. $P(R_1 = r_1, \ldots, R_n = r_n) = P(|R_1| = |r_1|, \ldots, |R_n| = |r_n|, I_1 = \epsilon_1, \ldots, I_n = \epsilon_n) = \frac{1}{2^n}\frac{1}{n!}$ où ϵ_i est le signe de r_i.
5. $P(T_1 = t_1, \ldots, T_s = t_s, S = s) = \sum_{r \in A} P(R_1 = r_1, \ldots, R_n = r_n)$ où $A = \{(r_1, \ldots, r_n)$ avec exactement s valeurs positives et prenant les valeurs $(t_1, \ldots, t_s)\}$. Il suffit alors de montrer que $card(A) = n!$ or il faut choisir s positions parmi n ce qui donne C_n^s choix, il y a ensuite $s!$ choix de l'ordre des t_i puis $(n-s)!$ choix de l'ordre des $(n-s)$ valeurs de r_i restant. Ce qui donne bien $card(A) = C_n^s s!(n-s)! = n!$.

Exemple : Considérons les observations suivantes :

	1	2	3	4	5	6	7
x_i	17.2	21.6	19.5	19.1	22.0	18.7	20.3
y_i	18.3	20.8	20.9	21.2	22.7	18.6	21.9
z_i	1.1	-0.8	1.4	2.1	0.7	-0.1	1.6
rangs signés	4	-3	5	7	2	-1	6

d'où $w = 2 + 4 + 5 + 6 + 7 = 24$. On calcule $P(W \geq 24) = P(W \leq 28 - 24) = 0.0547$ et on accepte H_0 à 5% alors que l'on rejette à 10%.

Sur les mêmes observations, on peut utiliser un test du signe : $s = 5$ et $P(S \geq 5) = 0.2266$, donc on accepte H_0 à 5% et à 10%; ou un test de student : $t = 2.32$, et $P(T_n \geq 2.32) < 0.05$ qui conduit à rejeter H_0 à 5% et à 10%.

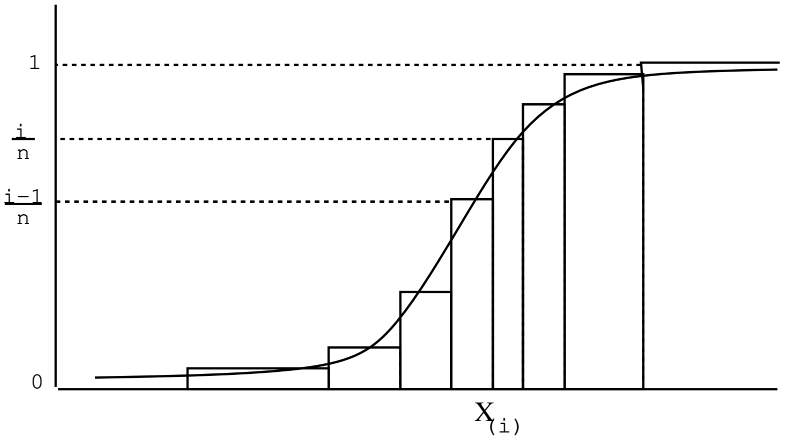

Figure 8.1: Fonction de répartition F_0 et fonction de répartition empirique

8.3 Ajustement

8.3.1 Test de Kolmogorov

Considérons dans un premier temps le problème qui consiste à estimer une fonction de répartition $F(x)$. Soit $\hat{F}_n(x) = \frac{\text{nb de } X_i < x}{n}$ la fonction de répartition empirique, alors $n\hat{F}_n(x) \sim \mathcal{B}(n,p)$ où $p = F(x)$. Donc $\hat{F}_n(x)$ est sans biais pour $F(x)$ et de variance $\frac{F(x)(1-F(x))}{n}$, de plus $\hat{F}_n(x)$ converge en probabilité vers $F(x)$ quand n tend vers $+\infty$.

On sait aussi que si les observations sont $X_1 = x_1, \ldots, X_n = x_n$ et si X a pour loi $\hat{F}_n$, alors $P(X = x_i) = 1/n$, $i = 1, \ldots, n$. De plus, le théorème de Glivenko-Cantelli permet d'obtenir que $\sup_x |\hat{F}_n(x) - F(x)|$ converge en probabilité vers 0.

On veut tester $H_o : F = F_0$ contre $H_1 : F \neq F_0$, et pour cela, on considère la statistique $D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_0(x)|$.

Théorème 8.3 : *Sous H_0, la loi de D_n est la même pour toute F_0.*

Dém : On va construire le cas F_0 strictement croissante, on peut alors considérer F_0^{-1}.

$$D_n = \sup_{0<u<1} |\hat{F}_n(F_0^{-1}(u)) - F_0(F_0^{-1}(u))| = \sup_{0<u<1} |\hat{F}_n(F_0^{-1}(u)) - u|$$

or $\hat{F}_n(F_0^{-1}(u)) = \frac{\text{nb de } X_i \leq F_0^{-1}(u)}{n} = \frac{\text{nb de } F_0(X_i) \leq u}{n}$.

Soit $U_i = F_0(X_i)$, $U_1, \ldots, U_n$ est un n-échantillon de loi $\mathcal{U}(0,1)$ puisque $P(F_0(X_i) \leq u) = P(X_i \leq F_0^{-1}(u)) = F_0(F_0^{-1}(u)) = u$, pour $0 < u < 1$. Donc $D_n = \sup_{0<u<1} |\hat{F}_n^*(u) - u|$ où $\hat{F}_n^*$ est la distribution empirique de l'échantillom de loi uniforme $U_1, \ldots, U_n$ et est donc indépendante de F_0.

Exemple : On veut vérifier si les observations suivantes peuvent être un n-échantillon de loi $\mathcal{N}(\mu_0, \sigma_0^2)$ avec $\mu_0 = 370.6$ et $\sigma_0^2 = 29.1$. On observe :

356.4	362.5	394.7	356.0	387.6	305.1	385.1	383.2	346.6	314.2
394.8	370.7	370.8	434.2	365.2	377.1	365.9	384.4	297.4	404.3
412.0	349.1	344.5							

Alors $D_n = \max_{1 \leq i \leq n} max \left(\frac{i}{n} - F_0(X_{(i)}), F_0(X_{(i)}) - \frac{i-1}{n} \right)$ où les $X_{(i)}$ sont les X_i ordonnés dans l'ordre croissant (statistique d'ordre). Cette nouvelle écriture de D_n est valable pour les lois à densité et provient du fait que l'on recherche la distance maximale entre une suite de segments horizontaux et une fonction croissante.

Le maximum est ici atteint pour $i = 3$ et alors $D_n = 0.104$. Avec $n = 23$, on obtient un domaine de rejet $R = \{D_n \geq 0.275\}$ et donc on accepte H_0.

Remarque : Malheureusement, cette technique s'adapte mal aux cas où μ et σ^2 sont inconnus. On peut penser à les estimer par $\bar{X}$ et S^2 mais, dans ce cas, on n'a plus la même loi :

$$D_n^* = \sup_{x\in\mathbb{R}} \left| \hat{F}_n(x) - \Phi\left(\frac{x-\bar{X}}{S}\right) \right|$$

8.3.2 Etude de la forme d'une distribution

On va cette fois construire un intervalle de confiance pour F, en utilisant le test de Kolmogorov :

$$P(D_n \leq c) = 1-\alpha \Longrightarrow P(-c \leq \hat{F}_n(x) - F(x) \leq c) = 1-\alpha$$

$$\Longleftrightarrow P(\hat{F}_n(x) - c \leq F(x) \leq \hat{F}_n(x) + c) = 1-\alpha$$

On peut aussi estimer $\Phi^{-1}(F(x))$ par $\Phi^{-1}(\hat{F}_n(x))$ et transformer la région de confiance ou "bande de confiance" obtenue précédemment : $[\Phi^{-1}(\hat{F}_n(x)-c);\Phi^{-1}(\hat{F}_n(x)+c)]$; on obtient ainsi deux lignes polygonales encadrant $\hat{F}_n(x)$: $(X_{(i)}, \Phi^{-1}(\frac{i}{n}-c))$ et $(X_{(i)}, \Phi^{-1}(\frac{i}{n}+c))$.

Un autre possibilité consiste à estimer f de la manière suivante :

$$\hat{f}_n(x) = \frac{\hat{F}_n(x+h_n) - \hat{F}_n(x-h_n)}{2h_n} = \frac{\text{nb de } X_i \in [x-h_n; x+h_n]}{2nh_n}$$

On peut montrer que alors $\hat{f}_n(x)$ converge en probabilité vers $f(x)$ lorsque h_n tend vers 0 et nh_n tend vers $+\infty$ en remarquant que $2nh_n\hat{f}_n(x) \sim \mathcal{B}(n, F_n(x+h_n) - F_n(x-h_n))$.

Remarque : Un problème important pour cette approche est le choix de h_n.