Statistique Inférentielle

Chapitre 1 Outils probabilistes

I Convergence et ordre

Convergence en probabilité:

$$X_n \xrightarrow{P} X \iff \forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

Convergence en loi:

$$\begin{aligned} X_n \xrightarrow{\mathcal{L}} X \quad &\iff \quad \forall f \text{ continue bornée}, \ \lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)) \\ &\iff \quad \forall t \in \mathbb{R}, \quad \mathbb{E}\left(e^{itX_n}\right) \to \mathbb{E}\left(e^{itX}\right) \end{aligned}$$

Convergence presque sûre:

$$X_n \xrightarrow{p.s.} X \iff \mathbb{P}\left(\limsup_{n\to\infty} |X_n - X| = 0\right) = 1.$$

Théorème 1.1.1.

1. $X_n \xrightarrow{p.s.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\mathcal{L}} X$
2. $X_n \xrightarrow{P} c \iff X_n \xrightarrow{\mathcal{L}} c$ si c est une constante.
3. si $X_n \xrightarrow{\mathcal{L}} X$ et $X_n - Y_n \xrightarrow{P} 0$ alors $Y_n \xrightarrow{\mathcal{L}} X$.
4. si $X_n \xrightarrow{P} X$ et $Y_n \xrightarrow{P} Y$ alors $(X_n, Y_n) \xrightarrow{P} (X, Y)$.
5. si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{P} c$ alors $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c)$

Lemme 1.1.2. (Slutsky)
Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} c$ où c est une constante alors

1. $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$.
2. $X_n Y_n \xrightarrow{\mathcal{L}} cX$.
3. $X_n / Y_n \xrightarrow{\mathcal{L}} X/c$ si $c \neq 0$.

Théorème 1.1.3. Soit g continue en tout point de C tel que $\mathbb{P}(X \in C) = 1$, alors si $X_n \xrightarrow{\mathcal{L}} X$ (resp. $\xrightarrow{P}$, $\xrightarrow{p.s.}$) alors $g(X_n) \xrightarrow{\mathcal{L}} g(X)$ (resp.$\xrightarrow{P}$, $\xrightarrow{p.s.}$).

Remarquons que ceci implique, si g est bornee, que $\mathbb{E}g(X_n) \to \mathbb{E}g(X)$ ce qui ne permet pas, par exemple d'obtenir la convergence des moments car $x \to x^k$ est non bornée.

Uniforme intégrabilité: la suite X_n est uniformément intégrable si

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{E}\left(|X_n| 1_{|X_n|>M}\right) = 0.$$

Théorème 1.1.4. (Cv des moments)
Soit g continue en tout point de C tel que $\mathbb{P}(X \in C) = 1$ et $X_n \xrightarrow{\mathcal{L}} X$. Alors $\mathbb{E}g(X_n) \to \mathbb{E}g(X) \iff (g(X_n))$ est uniformément intégrable.

Ordre en probabilité

soit (g_n) une suite de nombres positifs.

X_n est d'ordre plus petit que g_n en probabilité:

$$X_n = o_{\mathbb{P}}(g_n) \quad \text{si} \quad \frac{X_n}{g_n} \xrightarrow{P} 0.$$

X_n est au plus d'ordre g_n en probabilité:

$$X_n = O_{\mathbb{P}}(g_n) \quad \text{si} \quad \forall \varepsilon > 0, \exists M_\varepsilon, \mathbb{P}(|X_n| \geq M_\varepsilon g_n) \leq \varepsilon.$$

On dit aussi que X_n est borné en probabilité par g_n.

Lemme 1.1.5.

1. si $X_n = o_{\mathbb{P}}(f_n)$ et $Y_n = o_{\mathbb{P}}(g_n)$ alors

$$\begin{aligned} X_n Y_n &= o_{\mathbb{P}}(f_n g_n) \\ |X_n|^s &= o_{\mathbb{P}}(f_n^s) \quad \textit{pour } s > 0 \\ X_n + Y_n &= o_{\mathbb{P}}(\max(f_n, g_n)) \end{aligned}$$

2. idem pour $X_n = O_{\mathbb{P}}(f_n)$ et $Y_n = O_{\mathbb{P}}(g_n)$.
3. si $X_n = o_{\mathbb{P}}(f_n)$ et $Y_n = 0_{\mathbb{P}}(g_n)$

$$X_n Y_n = o_{\mathbb{P}}(f_n g_n)$$

Corollaire 1.1.6.

1. si $\mathbb{E}(X_n^2) = O(a_n^2)$ alors $X_n = O_{\mathbb{P}}(a_n)$.
2. si $\mathbb{E}((X_n - \mathbb{E}(X_n))^2) = O(a_n^2)$ et $\mathbb{E}(X_n) = O(a_n)$ alors $X_n = O_{\mathbb{P}}(a_n)$.

Développements limités en probabilité

Théorème 1.1.7.
si

$$X_n = a + O_{\mathbb{P}}(r_n) \quad (\text{resp.} o_{\mathbb{P}}(r_n))$$

où $r_n \to 0$ et a une constante. Si g est une fonction avec s dérivées continues en a alors

$$g(X_n) = g(a) + g'(a)(X_n - a) + \cdots + \frac{1}{(s-1)} g^{(s-1)}(a)(X_n - a)^{s-1} + O_{\mathbb{P}}(r_n^s)$$

(resp. $o_{\mathbb{P}}(r_n^s)$)

II Théorèmes Limites

Les variables considérées sont à valeurs dans $\mathbb{R}^k$.

Loi des Grands Nombres.

Théorème 1.2.1. (Loi Faible des Grands Nombres)
Soit (X_n) une suite de v.a. i.i.d. telle que $\mathbb{E}\|X_n\| < \infty$ et $\mathbb{E}X_n = \mu$, alors

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu.$$

Théorème 1.2.2. (Loi Forte des Grands Nombres)
Soit (X_n) une suite de v.a. i.i.d. telle que $\mathbb{E}\|X_n\| < \infty$ et $\mathbb{E}X_n = \mu$, alors

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s.} \mu.$$

Théorème Central Limite.

Théorème 1.2.3. (Théorème Central Limite)
Soit X_n une suite de v.a. i.i.d. de moyenne μ et de matrice de covariance (finie) Σ, alors

$$\sqrt{n}(\bar{X}_n - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)$$

où $\mathcal{N}(0, \Sigma)$ dénote le vecteur gaussien centré de matrice de covariance Σ.

Nous présentons ensuite deux extensions aux cas de v.a. indépendantes non identiquement distribuées (à valeurs réelles pour simplifier).

Théorème 1.2.4. (Théorème Central Limite - Lindeberg)
Soient X_n des v.a. indépendantes centrées tels que $\mathbb{E}|X_n|^2 = \sigma_n^2$. Notons $V_n = \sum_{i=1}^n \sigma_i^2$, si

$$\lim_{n\to\infty} \frac{1}{V_n} \sum_{i=1}^n \mathbb{E}\left(X_i^2 1_{\{X_i^2 > V_n \varepsilon\}} \right) = 0 \qquad \forall \varepsilon > 0$$

alors

$$\frac{1}{\sqrt{V_n}} \sum_{i=1}^n X_i \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

Nous donnons ensuite une condition de type moment permettant de vérifier ces conditions.

Théorème 1.2.5. (Théorème Central Limite - Lyapunov)
Soient X_n des v.a. indépendantes centrées tels que $\mathbb{E}|X_n|^2 = \sigma_n^2$. Notons $V_n = \sum_{i=1}^n \sigma_i^2$, si il existe $\delta > 0$ tel que

$$\lim_n \frac{1}{V_n^{1+\delta/2}} \sum_{i=1}^n \mathbb{E}|X_i|^{2+\delta} = 0$$

alors

$$\frac{1}{\sqrt{V_n}} \sum_{i=1}^n X_i \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

La δ-méthode.

Théorème 1.2.6.
Soit $g : D_g \subset \mathbb{R}^k \to \mathbb{R}^m$ différentiable au point θ. Soient T_n des v.a. à valeurs dans D_g. Soit r_n une suite croissante vers l'infini. Si, pour une v.a. T,

$$r_n(T_n - \theta) \xrightarrow{\mathcal{L}} T$$

alors

$$r_n(g(T_n) - g(\theta)) \xrightarrow{\mathcal{L}} \nabla_\theta g(\theta) T$$

et

$$r_n(g(T_n) - g(\theta)) - \nabla_\theta g(\theta)(r_n(T_n - \theta)) \xrightarrow{P} 0$$

Chapitre 2 Distribution d'échantillonnage

I Présentation

Un "échantillon": deux sens

1. un sous-ensemble ou une suite d'individus choisis au hasard dans une population, ou encore le sous-ensemble ou la suite des valeurs d'un caractère statistique pour les individus choisis.

2. une suite de v.a. indépendantes et identiquement distribuées.

II Etude des moments empiriques

Soit une v.a. X et un n-échantillon $(X_1, ..., X_n)$ de X.

Moment empirique d'ordre r:

$$M_n^r = M_n^r(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i^r$$

moment théorique d'ordre r: $m_r(X) = \mathbb{E}(X^r)$.

Moment empirique centré d'ordre r:

$$\tilde{M}_n^r = \tilde{M}_n^r(X_1, ..., X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^r$$

moment théorique d'ordre r: $\mu_r = \mathbb{E}((X - \mathbb{E}(X))^r)$.

Remarque: on note $\bar{X}_n = M_1$ et $S_n^2 = \tilde{M}_n^r$.

Propriétés à distance finie:

1. Si $\mathbb{E}|X| < \infty$, alors $\mathbb{E}(\bar{X}_n) = \mathbb{E}(X)$.
2. Si $\mathbb{E}(X^2) < \infty$, alors $\mathbb{V}(\bar{X}_n) = \frac{\mathbb{V}(X)}{n}$ et $\mathbb{E}(S_n^2) = \frac{n-1}{n}\mathbb{V}(X)$.
3. Si $\mathbb{E}|X|^3 < \infty$, alors $\mathrm{Cov}(\bar{X}_n, S_n^2) = \frac{\mu_3}{n}(1 - \frac{1}{n})$.
4. Si $\mathbb{E}(X^4) < \infty$, alors $\mathbb{V}(S_n^2) \leq \frac{\mu_4 + \mu_2^2}{n}$.

Propriétés asymptotiques:

1. Si $\mathbb{E}|X|^r < \infty$, alors $M_n^r \xrightarrow{p.s.} \mathbb{E}X^r$.
2. Si $\mathbb{E}(X^2) < \infty$, alors $\sqrt{n}(\bar{X}_n - \mu_1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbb{V}(X))$.
3. Si $\mathbb{E}(X^4) < \infty$, alors $\sqrt{n}(S_n^2 - \mu_2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu_4 - \mu_2^2)$.

<u>Cas gaussien:</u>

soit $(X_1, ..., X_n)$ un n-échantillon de $X \sim \mathcal{N}(0,1)$.

Théorème 2.2.1. (Fisher)
pour tout échantillon $(X_1, ..., X_n)$ issu de $X \sim \mathcal{N}(0,1)$, on a

1. $\sqrt{n}\bar{X}_n \sim \mathcal{N}(0,1)$.
2. $nS_n^2 \sim \chi^2(n-1)$.
3. $\sqrt{n}\bar{X}_n$ et nS_n^2 sont indépendantes.

Conséquence: soit $(X_1, ..., X_n)$ un n-échantillon de $X \sim \mathcal{N}(m, \sigma^2)$,

$$T_n = \frac{\sqrt{n-1}(\bar{X}_n - m)}{\sqrt{S_n^2}} \sim \mathcal{T}(n-1).$$

III Echantillon ordonné

soit $(X_1, ..., X_n)$ un n-échantillon de X, de fonction de répartition F. On note $(X_1^*, ..., X_n^*)$ l'échantillon ordonné associé à $(X_1, ..., X_n)$, i.e.

"X_k^* est la $k^{ème}$ valeur de l'échantillon $(X_1, ..., X_n)$ rangé dans l'ordre croissant."

Remarques:
1. Les X_k^* ne sont pas indépendants.
2. $X_1^* = \min(X_1, ..., X_n)$ et $X_n^* = \max(X_1, ..., X_n)$.
3. Si X admet une densité f alors la loi de $(X_1^*, ..., X_n^*)$ a pour densité

$$g_n(z_1, ..., z_n) = \begin{cases} n! f(z_1) ... f(z_n) & \text{si } z_1 \leq ... \leq z_n \\ 0 & \text{sinon.} \end{cases}$$

Lois marginales de l'échantillon ordonné: soit H_k la fonction de répartition de X_k^*, on a

$$\begin{aligned} H_k(x) &= \sum_{j=k}^{n} C_n^j F(x)^j (1_F(x))^{n-j} \\ &= \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} t^{k-1}(1-t)^{n-k} dt \end{aligned}$$

ce qui donne en particulier

$$H_n(x) = F^n(x) \qquad H_1(x) = 1 - (1 - F(x))^n$$

et si X a une densité f alors, X_k^* a pour densité

$$h_k(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^k (1 - F(x))^{n-k} f(x)$$

IV Fonction de répartition empirique

soit $(X_1, ..., X_n)$ un n-échantillon de X, on appelle fonction de répartition empirique:

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

Proposition 2.4.1.
pour tout échantillon $(X_1, ..., X_n)$ issu de X, on a

1. $F_n(x) \xrightarrow{p.s.} F(x)$ pour tout x.
2. $\sqrt{n}(F_n(x) - F(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, F(x)(1 - F(x)))$ pour tout x.

Théorème 2.4.2. (Glivenko-Cantelli, Kolmogorov)

1. Il existe un sous-ensemble négligeable N tel que l'on ait

$$\lim_{n\to\infty} \sup_x |F_n(x,\omega) - F(x)| = 0 \qquad \forall\omega \in \Omega\backslash N$$

2. soit $D_n = \sup_x |F_n(x,\omega) - F(x)|$, on a $\sqrt{n}D_n \xrightarrow{\mathcal{L}} Z$ où la fonction de répartition de Z est

$$\psi(x) = \sum_{k=-\infty}^{\infty} (-1)^{k-1} e^{-2k^2x^2}.$$

de plus, ni la loi de D_n ni la loi limite ψ ne sont modifiées par des transformations du type $X \to aX + b$.

Chapitre 3 Modèles et problèmes statistiques

I Modèles statistiques

Expérience aléatoire E: ensemble des résultats possibles Ω, muni d'une tribu $\mathcal{F}$.

Sur $(\Omega, \mathcal{F})$, on se donne une famille de probabilités $\mathcal{Q}$.

Observations: la v.a. $X : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A}$ et sur $(\mathcal{X}, \mathcal{A}$ on considère la famille de probabilités $\mathcal{P} = \{P = Q_X ; Q \in \mathcal{Q}\}$.

Le triplet $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ est un **modèle statistique**, et constitue la structure d'intérêt.

Exemple: $\mathcal{X} = \{0, 1\}$ et $\mathcal{P} = \{B(p), p \in [0, 1]\}$.

Modèle paramétrique:
$\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ où Θ est l'espace des paramètres. On cherche donc à identifier θ.

Modèle paramétrique dominé:
il existe une mesure positive σ-finie μ sur $\mathcal{A}$ telle que pour tout $\theta \in \Theta$, $\mathbb{P}_| theta$ admette une densite $f(\cdot, \theta)$ par rapport à μ.

Modèle paramétrique dominé homogène:
$A(\theta) = \{x \in \mathcal{X} : f(x, \theta) > 0\}$ ne dépend pas de θ.

Exemples: 1. $\mathcal{X} = \{0, 1\}$ et $\mathcal{P} = \{B(p), p \in [0, 1]\}$.
2. $\mathcal{X} = \mathbb{R}$ et $\mathcal{P} = \{\mathcal{N}(m, \sigma^2, (m, \sigma^2) \in \mathbb{R} \times \mathbb{R}_*^+\}$.

Modèle non paramétrique:

Θ n'est pas de dimension finie.

Exemple: $\Theta = \{f(\cdot); f \geq 0, ntf = 1, \int xf(x)dx = 0\}$.

Modèle semi-paramétrique:

$\Theta = \Theta_1 \times \Theta_2$ où Θ_1 est de dimension finie.

Modèle bayésien:

le paramètre d'intérêt est lui-même une variable aléatoire dont la oi reflète la connaissance *à priori* du problème.

II Problèmes statistiques

Dans la pratique, on a plusieurs observations que l'on supppose indépendantes et de même loi. On a donc le modèle d'échantillonnage $(\mathcal{X}^n, \mathcal{A}^{\otimes n}, \mathbb{P}_\theta^{\otimes n} : \theta \in \Theta)$.

<ins>Problème d'estimation statistique</ins>:
On cherche la "vraie" valeur du paramètre θ ou bien un sous-ensemble où cette "vraie valeur" a une forte probabilité de se trouver.
$\hookrightarrow$ estimation *ponctuelle*: on construit $T = \phi(X_1, ..., X_n)$ visant à approcher θ
$\hookrightarrow$ estimation *par intervalle de confiance*: on construit l'intervalle $I = I(X_1, ..., X_n)$ tel que $\mathbb{P}(\theta \in I) = \alpha$.

Problème de test statistique:
on se donne un sous-ensemble $\mathcal{P}_0 \subset \mathcal{P}$ et on s'intéresse aux deux hypothèses:

$$H_0 : \mathbb{P} \in \mathcal{P}_0 \qquad\qquad H_1 : \mathbb{P} \notin \mathcal{P}_0$$

et deux erreurs
$\hookrightarrow$ erreur de $1^{\text{ère}}$ espèce: on choisit H_1 alors que H_0 est vraie.
$\hookrightarrow$ erreur de $2^{\text{ème}}$ espèce: on choisit H_0 alors que H_1 est vraie.
et on y associe une probabilité

$$\alpha = \mathbb{P}(\text{choisir } H_1/H_0 \text{ est vraie}) \qquad\qquad \beta = \mathbb{P}(\text{choisir } H_0/H_1 \text{ est vraie})$$

On cherche alors à partir de l'échantillon observé à prendre une décision.

Chapitre 4 : Vraisemblance et exhaustivité

I Préliminaires

Définition 1 : Un modèle statistique paramétrique est donné par :

- *un espace probabilisable* $(\Omega, \mathcal{A})$*;*
- *une famille de probabilités* $\mathbb{P}_\theta$ *où* $\theta \in \Theta$ *sur* $(\Omega, \mathcal{A})$*.*

$\forall \theta \in \Theta$, X est une application mesurable de $(\Omega, \mathcal{A})$ dans $(X(\Omega), \mathcal{B})$.

$\forall B \in \mathcal{B}$, $\mathbb{P}_\theta^X(B) = \mathbb{P}_\theta\left(X^{-1}(B)\right)$.

Définition 2 : Un n-échantillon de la loi loi $\mathbb{P}_\theta^X$ est un n-uplet de variables aléatoires mutuellement indépendantes et de même loi $\mathbb{P}_\theta^X$.

Dans la suite, on considère un n-échantillon de loi $\mathbb{P}_\theta^X$.

Si X est discrète, on suppose que la loi $\mathbb{P}_\theta^X$ est dominée par la mesure du dénombrement et a pour densité $f(x;\theta) = \mathbb{P}_\theta(X = x)$.

Si X est continue, on suppose que la loi $\mathbb{P}_\theta^X$ est dominée par la mesure de Lebesgue et a pour densité $f(x;\theta)$.

II Notion de vraisemblance

Définition 3 : Pour $(x_1, \ldots, x_n)$ *réalisation de* $(X_1, \ldots, X_n)$*, on appelle vraisemblance de l'échantillon la fonction* $V_n : \Theta \longrightarrow \mathbb{R}_+$ *telle que*

$$V_n(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta).$$

Dans le cas dicret, $V_n(\theta_0; x_1, \ldots, x_n)$ correspond à la probabilité d'observer $(x_1, \ldots, x_n)$ pour $\theta = \theta_0$.

III Notion d'exhaustivité

Définition 4 : On appelle statistique une application mesurable $T = f(X_1, \ldots, X_n)$ *fonction d'un* n*-échantillon.*

On cherche à résumer un n-échantillon de loi $\mathbb{P}_\theta^X$ par une statistique sans perdre d'informations sur θ.

Définition 5 : Une statistique $T = f(X_1, \ldots, X_n)$ est dite exhaustive pour le paramètre θ si la loi conditionnelle de $(X_1, \ldots, X_n)$ sachant $T = t$ ne dépend pas de θ.

Théorème 1 (critère de factorisation) : $T = f(X_1, \ldots, X_n)$ est une statistique exhaustive pour θ si et seulement si il existe une application $g : \mathbb{R}^n \longrightarrow \mathbb{R}_+$ et une application $h : \mathbb{R}^n \times \Theta \longrightarrow \mathbb{R}_+$ telles que :

$$V_n(\theta; x_1, \ldots, x_n) = g(x_1, \ldots, x_n) h(f(x_1, \ldots, x_n), \theta)$$

pour toutes valeurs de θ, $x_1, \ldots, x_n$.

Chapitre 5 : Information et identification

I Information

Dans le paragraphe précédent, on a utilisé sans vraiment la définir avec précision une notion d'information. On a dit qu'une statistique exhaustive pour θ conservait toute l'information sur le paramètre θ.

Soit $(\Omega, \mathcal{A}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ un modèle statistique.

Soit X une application mesurable de $(\Omega, \mathcal{A})$ dans $(X(\Omega), \mathcal{B})$ et $\mathbb{P}_\theta^X$ la probabilité image par X.

$\mathbb{P}_\theta^X$ est dominé par une mesure μ, X admet pour densité $f(x; \theta)$ par rapport à μ.

On suppose que le modèle image par X est régulier :

(H1) : Θ est un ouvert de $\mathbb{R}^p$;

(H2) : le modèle est homogène : $\{x|f(x;\theta) > 0\}$ ne dépend pas de θ ;

(H3) : pour tout $\theta \in \Theta$, la fonction $\theta \longrightarrow f$ est deux fois dérivable et les dérivées partielles premières et secondes sont continues ;

(H4) : pour tout $B \in \mathcal{B}$, la fonction $\theta \longrightarrow \int_B f(x;\theta)d\mu(x)$ est deux fois dérivable et l'on peut intervertir les opérations de dérivation et d'intégration.

Definition 1 : On appelle information de Fisher apportée par X sur θ, la matrice de variance-covariance de $\dfrac{\delta \log f}{\delta \theta}(X;\theta)$.

Le vecteur $\dfrac{\delta \log f}{\delta \theta}(X;\theta)$ est appelé vecteur du score.

$$I_X(\theta) = \mathbb{V}\left(\frac{\delta \log f}{\delta \theta}(X;\theta)\right)$$

Proposition 1 : Le vecteur du score est centré.

Corrolaire 1 : L'élément (i,j) de la matrice d'information de Fisher est tel que :

$$[I_X(\theta)]_{(i,j)} = \mathbb{E}\left(\frac{\delta \log f}{\delta \theta_i}(X;\theta)\frac{\delta \log f}{\delta \theta_j}(X;\theta)\right).$$

Théorème 1 :

$$[I_X(\theta)]_{(i,j)} = -\mathbb{E}\left(\frac{\delta^2 \log f}{\delta \theta_i \delta \theta_j}(X;\theta)j\right)$$

Soit $X_1, \ldots, X_n$ un n-échantillon de X et soient $S = f_1(X_1, \ldots, X_n)$ et $T = f_2(X_1, \ldots, X_n)$ deux statistiques.

Proposition 2 : Si S et T sont indépendantes, alors

$$I_{(S,T)}(\theta) = I_S(\theta) + I_T(\theta).$$

Corrollaire 2 :

$$I_{(X_1,\ldots,X_n)} = nI_X(\theta)$$

Théorème 2 : Pour toute statistique S, la matrice $I_S(\theta)$ est inférieure au sens de l'ordre partiel sur les matrices symétriques à la matrice $I_{(X_1,\ldots,X_n)}(\theta) = nI_X(\theta)$.

$nI_X(\theta) - I_S(\theta)$ est une matrice positive.

Théorème 3 : $nI_X(\theta) = I_S(\theta)$ si et seulement si S est une statistique exhaustive pour θ.

II Identification

Soit $(\Omega, \mathcal{A}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ un modèle statistique.

Définition 2 : Une valeur du paramètre $\theta_0 \in \Theta$ est identifiable si $\forall \theta \neq \theta_0, \quad \mathbb{P}_\theta \neq \mathbb{P}_{\theta_0}$. Le modèle est dit identifiable si toutes les valeurs du paramètre sont identifiables, ie si l'application $\theta \longrightarrow \mathbb{P}_\theta$ est injective.

On peut affaiblir la notion précédente à une notion locale.

Définition : Une valeur du paramètre $\theta_0 \in \Theta$ est localement identifiable s'il existe un voisinage $\mathcal{V}_0$ de θ_0 tel que $\forall \theta \neq \theta_0, \quad \theta \in \mathcal{V}_0, \quad \mathbb{P}_\theta \neq \mathbb{P}_{\theta_0}$. Un modèle est dit localement identifiable si toutes les valeurs du paramètre sont localement identifiables.

Théorème 3 : Si la matrice d'information de Fisher $I_X(\theta)$ est régulière, le modèle image par X est localement identifiable.

Chapitre 6 : Familles exponentielles

I Définition

Soit $(\Omega, \mathcal{A}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ un modèle statistique.

Définition : On dit que $\{\mathbb{P}_\theta; \theta \in \Theta\}$ *appartient à la famille exponentielle si,* $\forall \theta \in \Theta$, $\mathbb{P}_\theta$ *est dominée par une mesure* μ *et s'il existe une famille de densités par rapport à* μ *du type :*

$$f(x;\theta) = C(\theta)h(x)\exp\left(\sum_{i=1}^{r} \alpha_i(\theta)T_j(x)\right)$$

où $C(\cdot)$ *et* $h(\cdot)$ *sont des fonction,s positives. La statistique* $T = (T_1, \ldots, T_r)$ *est appelée statistique canonique.*

Nombreuses propriétés...

II Liens avec la notion d'exhaustivité

D'après le théorème de factorisation, la statistique canonique d'une famille exponentielle est une statistique exhaustive.

Soit X une application mesurable de $(\Omega, \mathcal{A})$ dans $(X(\Omega), \mathcal{B})$ et $X_1, \ldots, X_n$ un n-échantillon de loi X.

Soient S_1 et S_2 deux statistiques telles que $S_1 = h(S_2)$. Si S_1 est exhaustive pour θ alors S_2 l'est aussi.

Définition : Une statistique S est exhaustive minimale pour θ, si elle est exhaustive et si pour toute autre statistique exhaustive T pour θ, il existe une application h telle que $S = h(T)$.

Lemme 1 : Deux statistiques exhaustives minimales pour θ sont en liaison bijective.

Lemme 2 : Si S est exhaustive minimale pour θ, toute statistique exhaustive pour θ de la forme $T = h(S)$ est également minimale.

Ouvert de $\mathbb{R}$: l'ensemble vide ou intervalle ouvert $]a_i, b_i[$

Ouvert de $\mathbb{R}^n$: produit de n ensembles ouvert de $\mathbb{R}$

On appelle intérieur d'un sous-ensemble A de $\mathbb{R}^n$ la réunion, éventuellement vide de tous les ouverts contenus dans A. C'est le plus grand ouverts contenus dans A.

Théorème 1 : Si $\{\mathbb{P}_\theta^X ; \theta \in \Theta\}$ appartient à la famille exponentielle et si l'intérieur de $\alpha(\Theta)$ ($\alpha(\theta) = (\alpha_1(\theta), \ldots, \alpha_r(\theta))$) est non vide, alors la statistique $T = \left(\sum_{i=1}^n T_1(X_i), \ldots, \sum_{i=1}^n T_r(X_i)\right)$ est exhaustive minimale pour θ.

II Liens avec la notion de complétude

Définition : Le modèle $(\Omega, \mathcal{A}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ *est complet si*

$$\int f(x) d\mathbb{P}_\theta(x) = 0, \quad \forall \theta \in \Theta \quad \Longrightarrow f = 0 \ \mathbb{P}_\theta \text{ presque partout } \forall \theta \in \Theta.$$

Une statistique S *est complète si le modèle image par* S *est complet.*

Définition : La statistique S *est libre pour* θ *si sa loi ne dépend pas de* θ*.*

Proposition 1 : Si T est une statistique exhaustive et complète pour θ et si S est une statistique libre pour θ, alors T et S sont indépendantes.

Théorème 2 : Si $\{\mathbb{P}_\theta^X ; \theta \in \Theta\}$ appartient à la famille exponentielle et si l'intérieur de $\alpha(\Theta)$ ($\alpha(\theta) = (\alpha_1(\theta), \ldots, \alpha_r(\theta))$) est non vide, alors la statistique $T = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_r(X_i)\right)$ est complète.

III Liens avec la notion d'identification

Proposition 2 : Soit un modèle exponentiel sous forme naturelle, de densités

$$\exp\left[\sum_{j=1}^{r} \lambda_j T_j(x) - A(\lambda)\right] g(x)$$

$\lambda \in \Lambda$, **ouvert de** $\mathbb{R}^r$**.**

Une condition nécessaire et suffisante pour que le modèle soit identifiable est que $I_X(\lambda)$ **soit régulière.**

Chapitre 7 : Estimation ponctuelle

Soit $(\Omega, \mathcal{A}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ un modèle statistique.

On observe un n-échantillon $X_1, \dots, X_n$ de loi $\mathbb{P}_\theta^X$ et $X(\Omega) = \mathbb{R}$.

Un estimateur de θ (ou de $g(\theta)$), paramètre inconnu, est une statistique $T_n = f(X_1, \dots, X_n)$ proche de θ (ou de $g(\theta)$) dans un sens qu'il convient de définir.

I Qualités d'un estimateur

Définition 1 : $T_n = f(X_1, \ldots, X_n)$ *est un estimateur non biaisé de* $g(\theta)$ *si et seulement si* $\mathbb{E}_\theta[T_n] = g(\theta)$, $\forall \theta \in \Theta$.

Il faut bien comprendre que T_n est une statistique, ie une quantité aléatoire. La réalisation de T_n, notée t_n est une estimation de $g(\theta)$.

Définition 2 : $T_n = f(X_1, \ldots, X_n)$ *est un estimateur faiblement (resp. fortement) convergent de* $g(\theta)$ *si et seulement si* T_n *converge en probabilité (resp. presque sûrement) vers* $g(\theta)$, $\forall \theta \in \Theta$.

Définition 3 : $T_n = f(X_1, \ldots, X_n)$ *est dite asymptotiquement normale s'il existe pour tout* $n \geq 1$ *deux suites réelles* $(a_n)_{n\geq 1}$ *et* $(b_n)_{n\geq 1}$ *où* $a_n > 0$ *telles que* $\left(\dfrac{T_n - b_n}{a_n}\right)$ *converge en loi vers une loi normale centrée réduite.*

Les suites $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ peuvent dépendre du paramètre inconnu θ.

Si T_n est un estimateur fortement convergent de $g(\theta)$, on pourra alors prendre $b_n = g(\theta)$ et s'il existe $(a_n)_{n\geq 1}$ telle que $\left(\dfrac{T_n - b(\theta)}{a_n}\right)$ converge en loi vers une loi normale centrée réduite, alors on dira que T_n est un estimateur asymptotiquement normal de $g(\theta)$.

Comment comparer deux estimateurs ?

On cherche des estimateurs optimaux relativement à un critère donné. On considère souvent l'erreur quadratique moyenne. Si T_n est un estimateur de $g(\theta)$, alors

$$EQM(T_n) = \mathbb{E}_\theta[(T_n - g(\theta))^2] = \mathbb{V}_\theta(T_n) + (\mathbb{E}_\theta(T_n) - g(\theta))^2.$$

Dans la classe des estimateurs non biaisés, ce critère est égal à la variance.

Soit $\mathcal{S}_n$ la classe des estimateurs non biaisés.

Définition 4 : On dit que $T_n \in \mathcal{S}_n$ est uniformément optimal dans la classe $\mathcal{S}_n$ si $\mathbb{V}_\theta(T_n) \leq \mathbb{V}_\theta(Q_n)$, $\forall Q_n \in \mathcal{S}_n$ et $\forall \theta \in \Theta$.

On dit plus précisement que T_n est un estimateur non biaisé de variance uniformément minimale (UMVUE).

Définition 5 : On dit que $T_n \in \mathcal{S}_n$ est localement optimal dans la classe $\mathcal{S}_n$ pour la valeur θ_0 du paramètre θ si $\mathbb{V}_{\theta_0}(T_n) \leq \mathbb{V}_{\theta_0}(Q_n)$, $\forall Q_n \in \mathcal{S}_n$.

On suppose que les hypothèses (H_1)-(H_4) sont vérifiées, de plus on suppose que :

H_5 : $0 < I_X(\theta) < \infty$;
H_6 : T_n estimateur non biaisé de $g(\theta)$ admet un densité régulière.

Théorème 1 : Soit $T_n \in \mathcal{S}_n$. Si (H_1)-(H_6) vérifiées alors :
i) $\theta \longrightarrow g(\theta)$ est dérivable, $\forall \theta \in \Theta$;
ii) borne FDCR :

$$\mathbb{V}_\theta(T_n) \geq \frac{(g'(\theta))^2}{I_n(\theta)}.$$

Définition 6 : T_n *estimateur non biaisé de* $g(\theta)$ *sera dit efficace si et seulement si* $\mathbb{V}_\theta(T_n) = \frac{(g'(\theta))^2}{I_n(\theta)}$.

Définition 7 : T_n *estimateur non biaisé de* $g(\theta)$ *sera dit asymptotiquement efficace si et seulement si* $\mathbb{E}_\theta(T_n)$ *tend vers* $g(\theta)$ *et* $\left(\mathbb{V}_\theta(T_n) \times \frac{I_n(\theta)}{(g'(\theta))^2}\right)$

tend 1 *lorsque* n *tend vers l'infini,* $\forall \theta \in \Theta$.

II Estimation non biaisée optimale

Théorème 2 (Théorème de Rao-Blackwell) : on considère une statistique exhaustive W pour θ et un estimateur T_n non biaisé de $g(\theta)$, alors la variable aléatoire $R_n = \mathbb{E}_\theta[T_n|W]$ vérifie les propriétés suivantes :
i) R_n **est un estimateur non biaisé de** $g(\theta)$
ii) $\mathbb{V}_\theta(R_n) \leq \mathbb{V}\theta(T_n)$**,** $\forall\theta \in \Theta$**.**

Théorème 3 (Théorème de Lehman-Scheffé) : Soit W une statistique exhaustive et complète pour θ. On pose $T_0 = \phi_0(W)$. Si T_0 est un estimateur sans biais de $g(\theta)$ alors T_0 est l'unique estimateur non biasié de variance uniformément minimale pour $g(\theta)$.

III Méthodes de construction d'estimateurs

La méthode du maximum de vraisemblance consiste, après avoir observé $(x_1, \ldots, x_n)$, à choisir l'estimation t de θ qui maximise la fonction $\theta \longrightarrow V_n(\theta; x_1, \ldots, x_n)$.

En général, une telle estimation t est fonction de $(x_1, \ldots, x_n)$: $t = \phi(x_1, \ldots, x_n)$. L'estimateur correspondant sera noté $\widehat{\theta}_n = \phi(X_1, \ldots, X_n)$. On dit que $\widehat{\theta}_n$ est l'estimateur du maximum de vraisemblance θ (EMV).

Sous certaines hypothèses, $\widehat{\theta}_n$ converge presque sûrement vers θ et $\sqrt{n}(\widehat{\theta}_n - \theta)$ converge en loi vers une loi normale centrée de variance $1/I_X(\theta)$.

Chapitre 8 : Estimation par intervalles

Soit $(\Omega, \mathcal{A}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ un modèle statistique.

On observe un n-échantillon $X_1, \ldots, X_n$ de loi $\mathbb{P}_\theta^X$ et $X(\Omega) = \mathbb{R}$.

Nous allons introduire la notion d'estimation ensembliste pour un paramètre inconnu.

Supposons que $\theta \in \mathbb{R}$.

Soient deux statistiques $A_n = f_1(X_1, \ldots, X_n$ et $B_n = f_2(X_1, \ldots, X_n)$ telles que $A_n \leq B_n$. Soit $\alpha \in [0, 1]$.

Définition 1 : On dira que $[A_n, B_n]$ est un intervalle de confiance de niveau $1 - \alpha$ pour θ si

$$\mathbb{P}_\theta(\theta \in [A_n, B_n]) = 1 - \alpha.$$

On doit comprendre un intervalle de confiance de niveau $1-\alpha$ comme un intervalle aléatoire qui a une probabilité $1-\alpha$ de contenir θ et non comme un intervalle fixé auquel θ aléatoire appartient avec une probabilité $1-\alpha$.

La relation $\mathbb{P}_\theta(\theta \in [A_n, B_n] = 1 - \alpha$ peut se décomposer d'une infinité de manières.
À tout couple $(\alpha_1, \alpha_2) \in \mathbb{R}_+$ tel que $\alpha_1 + \alpha_2 = \alpha$, $\mathbb{P}_\theta(\theta \in]-\infty, A_n[) = \alpha_1$ et $\mathbb{P}_\theta(\theta \in]B_n, \infty[) = \alpha_2$, on peut associer un intervalle de confiance de niveau $1-\alpha$. Souvent, on répartit les risque de manière symétrique, soit $\alpha_1 = \alpha_2 = \alpha/2$.

On a considéré le cas d'un intervalle de confiance bilatéral, il peut arriver que la recherche d'un intervalle unilatéral s'avère plus pertinente. Il s'agira d'un intervalle de la forme $[C_n, \infty[$ ou $]-\infty, C_n]$ où $C_n = f_3(X_1, \ldots, X_n,)$.

Définition 2 : On dira que $[A_n, B_n]$ est un intervalle de confiance de niveau asymptotiquement égal à $1-\alpha$ pour θ si $\mathbb{P}_\theta(\theta \in [A_n, B_n])$ tend vers $1-\alpha$ lorsque n tend vers l'infini.

Définition 3 : Une fonction pivotale pour θ est une variable aléatoire $T_n = \phi(X_1, \ldots, X_n, \theta)$ dont la loi est connue et indépendante de θ.

On peut construire un intervalle de confiance à partir d'une fonction pivotale.

Définition 4 : Une fonction asymptotiquement pivotale pour θ est une variable aléatoire $T_n = \phi(X_1, \ldots, X_n, \theta)$ qui converge en loi vers une loi connue et indépendante de θ.

Supposons $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$.

Soit $1 - \alpha_1 = \mathbb{P}_\theta(\theta_1 \in [A_n, B_n])$ et $1 - \alpha_2 = \mathbb{P}_\theta(\theta_2 \in [C_n, D_n])$.

La région de confiance définie par $\Gamma_n = [A_n, B_n] \times [C_n, D_n]$ pour θ est de niveau au moins égal à $1 - \alpha_1 - \alpha_2$.

Chapitre 9 : Tests d'hypothèses

Soit $(\Omega, \mathcal{A}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ un modèle statistique.

On observe un n-échantillon $X_1, \ldots, X_n$ de loi $\mathbb{P}_\theta^X$ et $X(\Omega) = \mathbb{R}$.

On souhaite vérifier si une hypothèse portant sur θ est vraie ou non.

I Exemple intruductif

Des relevés effectués pendant de nombreuses années ont permis d'établir que le niveau naturel des pluies dans la Beauce en millimètres par an suit une loi normale de moyenne 600 et de variance 50^2.

Des entrepreneurs, surnommés faiseurs de pluie, prétendaient pouvoir augmenter de 50 mm le niveau moyen de pluie, ceci par insémination des nuages au moyen d'iodure d'argent.

Leur procédé fut mis à l'essai entre 1951 et 1959 et on releva les hauteurs de pluies suivantes :

Année	1951	1952	1953	1954	1955	1956	1957	1958	1959
mm	510	614	780	512	501	534	603	788	650

Que pouvait-on en conclure ?

Deux hypothèses s'affrontaient ou bien l'insémination était sans effet, ou bien elle augmentait réellement le niveau moyen de pluie de 50 mm.

Soit X la variable aléatoire normale modélisant le niveau annuel des pluies et μ sa moyenne. Ces hypothèses pouvaient se formaliser comme suit :

$$\begin{cases} H_0 : \mu = 600mm \\ H_1 : \mu = 650mm \end{cases}.$$

Comment décider ?

Nous disposions de 9 observations $(x_1, x_2, \ldots, x_9)$, réalisation du 9-échantillon $(X_1, X_2, \ldots, X_9)$.

Puisqu'il s'agit de tester la valeur μ, il est naturel de s'intéresser à $\overline{X_9} = \frac{1}{9}\sum_{i=1}^{9} X_i$ qui est un estimateur sans biais et convergent de μ.

$\overline{X_9}$ est la variable de décision. C'est-à-dire la fonction des observations qui nous va nous permettre de trancher entre les 2 hypothèses.

Si H_0 est vraie : $\overline{X_9} \sim \mathcal{N}\left(600, 50^2/9\right)$.

Si H_1 est vraie : $\overline{X_9} \sim \mathcal{N}\left(650, 50^2/9\right)$.

Si $\overline{X_9}$ est trop grand, c'est-à-dire supérieure à un seuil k que l'on doit déterminer, on optera pour H_1.

Si $\overline{X_9} \leq k$, on conservera H_0.

L'ensemble des événements $\left\{\overline{X_9} > k\right\}$ région de rejet de H_0.

Comment déterminer le seuil k ?

Procédé onéreux : les agriculterus ne souhaitaient pas abandonner l'hypothèse H_0 sans en fixer le risque d'erreur.

Ils acceptaient un risque maximal de rejeter H_0 en se trompant de 5%.

On peut alors calculer le seuil k :

$$\mathbb{P}(\text{rejeter} \quad H_0 \quad \text{à tort}) =$$

$$\mathbb{P}\left(\overline{X_9} > k / \overline{X_9} \sim \mathrm{N}\left(600, \frac{50^2}{9}\right)\right) = 0.05$$

Grâce aux tables des fractiles de la loi normale, on trouve : $k = 627.5$.

La règle de décision est donc la suivante :

- si $\overline{X_9} > 627.5mm$, on rejette l'hypothèse H_0 ;
- si $\overline{X_9} \leq 627.5mm$, on conserve l'hypothèse H_0.

Les données relevées indiquent que $\overline{x_9} = 610.2mm$.

La conclusion est donc de conserver H_0.

L'insémination était sans effet notable sur le niveau des pluies : les valeurs observées pouvaient donc être dues au hasard en l'absence de toute influence de l'iodure d'argent.

Rien ne dit que conserver H_0 mette à l'abri de se tromper : en effet, les faiseurs de pluies ont peut-être raison, mais on ne s'en est pas aperçu.

Il y avait deux manières de se tromper :

• croire les faiseurs de pluie alors qu'ils n'étaient pour rien dans le résultat obtenu ;

• ne pas croire les faiseurs de pluie alors que leur méthode est bonne

Le premier risque noté α est fixé par les agriculteurs : $\alpha = 0.05$. Le deuxième risque noté β n'est pas maîtrisé.

Pour cette expérience,

$$\beta = \mathbb{P}(\text{rejeter} \quad H_1 \quad \text{à tort}) =$$

$$\mathbb{P}\left(\overline{X_9} \leq 627.5 / \overline{X_9} \sim \mathrm{N}\left(650, \frac{50^2}{9}\right)\right) = 0.088.$$

La probabilité de ne pas croire les faiseurs de pluie alors que leur méthode est bonne est de 0.088.

Les deux hypothèses ne jouent pas des rôles symétriques.

II Notions générales

Un test est un mécanisme qui permet de trancher entre deux hypothèses au vu des résultats d'un échantillon.

Soient Θ_0 et Θ_1 deux sous-ensembles formant une partition Θ : $\Theta = \Theta_0 \cup \Theta_1$ et $\Theta_0 \cap \Theta_1 = \emptyset$.

Hypothèse nulle : $H_0 : \theta \in \Theta_0$; Hypothèse alternative : $H_1 : \theta \in \Theta_1$

Hypothèses simples : $\theta = \theta_0$

Hypothèses composites : $\theta > \theta_0$, $\theta = (\theta_1, \theta_2)$ et $\theta_1 = a, \theta_2 \in \mathbb{R}$...

Définition 1 : On appelle région critique un sous-ensemble aléatoire W *de* $\mathbb{R}^n$ *associé à la règle de décision suivante : si* $(X_1, \ldots, X_n) \in W$, H_0 *est rejetée ; sinon* H_0 *est accpetée.*

4 cas possibles schématisés avec les probabilités correspondantes :

	Vérité	
Décision	H_0	H_1
H_0	$1 - \alpha$	β
H_1	α	$1 - \beta$

α et β sont les probabilités d'erreur de première et deuxième espèce.

Définition 2 : On appelle risque de première espèce d'un test de région critique W, la fonction α_W suivante :

$$\alpha_W(\theta) = \mathbb{P}_\theta^X(W), \quad \forall \theta \in \Theta_0.$$

Définition 3 : On appelle risque de deuxième espèce d'un test de région critique W, la fonction β_W suivante :

$$\beta_W(\theta) = 1 - \mathbb{P}_\theta^X(W), \quad \forall \theta \in \Theta_1.$$

Définition 4 : On appelle niveau d'un test de région critique W, la quantité α^* suivante :

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

Définition 5 : On appelle puissance d'un test de région critique W, *la fonction* ρ_W *suivante :*

$$\rho_W(\theta) = \mathbb{P}_\theta^X(W), \quad \forall \theta \in \Theta_1.$$

Définition 6 : Un test de région critique W *est dit sans biais si,* $\forall \theta \in \Theta_1$, $\rho_W(\theta) \geq \alpha^*$.

Définition 7 : Un test de région critique W *est dit convergent si,* $\forall \theta \in \Theta_1$, $\rho_W(\theta)$ *tend vers 1 lorsque* n *tend vers l'infini.*

β varie en sens contraire de α.

Dans la pratique des tests d'hypothèses, il est de règle de se fixer α^* comme donné ce qui fait que l'hypothèse nulle H_0 joue un rôle prééminent.

Les valeurs courantes pour α^* sont par exemple $0.01, 0.05, 0.1$.

Définition 8 : Un test de région critique W est de risque de première espèce asymptotiquement égal à α_W si, $\forall \theta \in \Theta_0$, $\mathbb{P}_\theta^X(W)$ tend vers $\alpha_W(\theta)$ lorsque n tend vers l'infini.

Définition 9 : Un test de région critique W est de risque de deuxième espèce asymptotiquement égal à β_W si, $\forall \theta \in \Theta_1$, $1 - \mathbb{P}_\theta^X(W)$ tend vers $\beta_W(\theta)$ lorsque n tend vers l'infini.

Définition 10 : Un test de région critique W est niveau asymptotiquement égal à α^* si $\sup_{\theta \in \Theta_0} \alpha(\theta)$ tend vers α^* lorsque n tend vers l'infini.

III Tests optimaux

On dit que le test de région critique W est meilleur que le test de région critique W' si :

$$\forall \theta \in \Theta_0, \quad \alpha_W(\theta) \leq \alpha_{W'}(\theta) \quad ;$$
$$\forall \theta \in \Theta_1, \quad \beta_W(\theta) \leq \beta_{W'}(\theta).$$

Un test est dit optimal s'il est meilleur que tous les autres.

En général, il n'existe pas de test optimal. on adopte alors l'optique Neyman qui consiste à fixer le niveau et à chercher le test de puissance maximale, test UPP.

Théorème 1 (Lemme de Neyman-Pearson) : On suppose que $\Theta_0 = \{\theta_0\}$ et $\Theta_1 = \{\theta_1\}$ (test entre deux hypothèses simples). Le test basé sur la région critique réalisée suivante :

$$w_{k_\alpha} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \,|V_n(\theta_0; x_1, \ldots, x_n) < k_\alpha V_n(\theta_1; x_1, \ldots, x_n)\}$$

est le test UPP de niveau (de risque de première espèce) α.
La constante k_α est déterminée par l'équation :

$$\alpha = \mathbb{P}_{\theta_0}(W_{k_\alpha}).$$

III Test du rapport des vraisemblances

Définition 11 : Le test du rapport des vraisemblances de $H_0 : \theta \in \Theta_0$ *contre* $H_1 : \theta \in \Theta_1$ *est le test associé à la région critique réalisée :*

$$w_{k_\alpha} = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n \left| \sup_{\theta \in \Theta_0} V_n(\theta; x_1, \ldots, x_n) < k_\alpha \sup_{\theta \in \Theta_1} V_n(\theta; x_1, \ldots, x_n) \right. \right\}.$$

La constante k_α *est déterminée par l'équation :*

$$\alpha = \mathbb{P}_{\theta_0}(W_{k_\alpha}).$$

Ce test est utilisé lorsque au moins l'une des deux hypothèses est composite, c'est notamment le cas en présence d'un paramètre de nuisance.