UNIVERSITÉ PARIS VI — PIERRE ET MARIE CURIE

U.F.R de Mathématiques

En vue du cours MM045 TER

Spécialité Mathématiques et Applications

présentée par

Sixin ZHANG

Estimateur du Maximum de Vraisemblance

Directeur: Lucien BIRGÉ

3 juillet 2009

Table des matières

1 Principe de la Méthode du Maximum de Vraisemblance

1.1 Le modèle statistique et l'estimateur

Dans cet article, on travaille principalement dans un modèle statistique particulier :

On se donne une suite de variables aléatoires $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ i.i.d, à valeurs dans un espace mesurable $(E, \mathcal{E})$, muni d'une famille de lois $\boldsymbol{P} = \{\mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^d\}$ sous l'hypothèse fondamentale que la vraie loi des X_i est dans cette famille.

Un modèle statistique est dit identifiable si

$$\mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2} \quad pour \quad \theta_1 \neq \theta_2.$$

Un modèle statistique est dit dominé s'il existe une mesure positive σ-finie μ sur $(E, \mathcal{E})$ telle que, pour tout $\theta \in \Theta$, il existe une densité $f_\theta(x) \in L^1(\mu)$ satisfaisant

$$f_\theta(x) = \frac{d\mathbb{P}_\theta}{d\mu}(x), \quad [\mu] - p.p.$$

Il faut dire que $f_\theta(x)$ est une fonction définie $[\mu] - p.p$ sur E, mais on peut choisir un représentant défini sur tout E qui nous convient le cas échéant, qui va être encore noté par $f_\theta(x)$.

Ayant établi un modèle statistique, on souhaite estimer la vraie loi à partir des échantillons. On se ramène souvent à estimer le vrai paramètre qui correspond à la vraie loi, i.e. à trouver un estimateur du paramètre inconnu d'un modèle statistique.

Définition 1.1 (Estimateur du paramètre inconnu)**.** Dans un modèle statistique, l'estimateur du paramètre θ est une suite de fonctions mesurables $\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{X}_n)$ à valeurs dans Θ, destinée à estimer θ.

On dit que $\hat{\theta}_n$ est consistant si pour tout $\theta \in \Theta$, on a

$$\hat{\theta}_n \xrightarrow[n \to \infty]{\mathbb{P}_\theta} \theta.$$

On dit que $\hat{\theta}_n$ est uniformément consistant si, pour tout $\delta > 0$,

$$\sup_{\theta \in \Theta} \mathbb{P}_\theta[|\hat{\theta}_n - \theta| \geqslant \delta] \xrightarrow[n \to \infty]{} 0.$$

On dit que $\hat{\theta}_n$ est asymptotiquement normal si pour tout $\theta \in \Theta$, il existe $\sigma^2(\theta) > 0$ tel que

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

Généralement, le modèle statistique doit être identifiable pour qu'un estimateur soit consistant. Par ailleurs, la méthode du maximum de vraisemblance se borne généralement au modèle statistique dominé.

Donc dans toute la suite, on va supposer que nous sommes toujours dans un MODÉLE STATISTIQUE IDENTIFIABLE ET DOMINÉ.

1.2 La méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance joue un rôle très important en Statistique. Cette méthode a été établie par R.A. Fisher entre 1912 et 1922.

Pour beaucoup de modèles courants, l'estimateur du maximum de vraisemblance est défini de façon unique, et se calcule explicitement. Sur le plan théorique on démontre, sous des hypothèses de régularité convenable, qu'il est consistant, asymptotiquement normal et théoriquement le meilleur.

Définition 1.2 (Estimateur du maximum de vraisemblance)**.** Tout estimateur $\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{X}_n)$ tel que

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} f_\theta(\boldsymbol{X}_n)$$

s'appelle estimateur du maximum de vraisemblance. La fonction de vraisemblance $f_\theta(\boldsymbol{X}_n)$ est définie par

$$f_\theta(\boldsymbol{X}_n) = \prod_{i=1}^{n} f_\theta(X_i).$$

On peut aussi définir la fonction de log-vraisemblance $L_n(\theta)$ par

$$L_n(\theta) = \log f_\theta(\boldsymbol{X}_n) = \sum_{i=1}^{n} \log f_\theta(X_i).$$

Remarque 1.3. Il faut noter que nous travaillons dans la droite achevée $\overline{\mathbb{R}} = [-\infty, +\infty]$, où on définit $\log 0 = -\infty$ et $0 \times (\pm\infty) = 0$. D'autre part, on a μ presque sûrement $f_\theta(X_i) \in [0, +\infty[$. Par suite, la fonction de log-vraisemblance est bien définie. Il faut aussi remarquer que cette définition ne rien dit sur l'existence et l'unicité de l'estimateur du maximum de vraisemblance.

2 Consistance des Estimateurs du Maximum de Vraisemblance

Dans cette partie, on va étudier la consistance des estimateurs du maximum de vraisemblance dans le cas vectoriel, i.e. $\Theta \subseteq \mathbb{R}^d$.

On va introduire deux méthodes classiques pour l'étudier. On verra que l'affinité de Hellinger et la divergence de Kullback-Leibler caractérisent bien l'écart entre l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ et le vrai paramètre θ_0. On trouvera un lien fondamental entre ces deux méthodes à la fin de cette partie.

Soient $\mathbb{P}$ et $\mathbb{Q}$ deux lois sur $(E, \mathcal{E})$ dominées par une mesure positive σ-finie μ. Posons

$$p = \frac{d\mathbb{P}}{d\mu}, \quad q = \frac{d\mathbb{Q}}{d\mu}.$$

Soit $\theta_0 \in \Theta$ qui correspond à la vraie loi. On va d'abord étudier la consistance autour du point θ_0, ensuite sur Θ.

2.1 Rapport de vraisemblance et affinité de Hellinger

Définition 2.1 (Rapport de vraisemblance)**.** On appelle rapport de vraisemblance (par rapport à θ_0) défini $\mathbb{P}_{\theta_0}$-p.s sur Θ, la quantité

$$\mathbf{Z}_n(\theta) = \frac{f_\theta(\boldsymbol{X}_n)}{f_{\theta_0}(\boldsymbol{X}_n)} = \prod_{i=1}^{n} \frac{f_\theta(X_i)}{f_{\theta_0}(X_i)}.$$

Par la définition de l'estimateur du maximum de vraisemblance, on a $\mathbf{Z}_n(\hat{\theta}_n) \geqslant \mathbf{Z}_n(\theta_0) = 1$ $\mathbb{P}_{\theta_0}$-p.s. On peut trouver des conditions suffisantes pour la consistance au point θ_0 à partir de cette inégalité. Rappelons que $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_{\theta_0}} \theta_0$ est définie par :

$$(\forall\delta > 0)(\forall\varepsilon > 0)(\exists N = N(\delta, \varepsilon) \in \mathbb{N}^*)(n \geqslant N \Rightarrow \mathbb{P}_{\theta_0}[|\hat{\theta}_n - \theta_0| \geqslant \delta] < \varepsilon).$$

Posons $\Theta_\delta = \{\theta : |\theta - \theta_0| \geqslant \delta, \theta \in \Theta\}$, on a

$$\{|\hat{\theta}_n - \theta_0| \geqslant \delta\} \subseteq \{\sup_{\theta\in\Theta_\delta} \mathbf{Z}_n(\theta) \geqslant 1\}.$$

Donc une condition suffisante pour la consistance est

$$\forall\delta > 0, \quad \mathbb{P}_{\theta_0}[\sup_{\theta\in\Theta_\delta} \mathbf{Z}_n(\theta) \geqslant 1] \xrightarrow[n\to\infty]{} 0. \tag{H_0}$$

Plus précisément,

Proposition 2.2. *Sous l'hypothèse (H_0), l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ converge (en probabilité) vers θ_0.*

Remarque 2.3. Il faut faire attention au cas $\Theta_\delta = \emptyset$, où $\sup_{\theta\in\Theta_\delta} \mathbf{Z}_n(\theta) = -\infty$, et la conclusion reste encore vraie.

On introduit maintenant l'affinité de Hellinger, qui peut nous aider à trouver des conditions plus pratiques.

Définition 2.4 (Affinité de Hellinger)**.** On appelle affinité de Hellinger entre $\mathbb{P}$ et $\mathbb{Q}$ la quantité

$$\rho(\mathbb{P},\mathbb{Q}) = \int \sqrt{p(x)q(x)}d\mu(x)$$

qui est indépendante du choix de la mesure dominante.

On a immédiatement que $0 \leqslant \rho(\mathbb{P},\mathbb{Q}) \leqslant 1$ et que $\rho(\mathbb{P},\mathbb{Q}) = 1$ s.s.i (si et seulement si) $\mathbb{P} = \mathbb{Q}$ en vertu que

$$\rho(\mathbb{P},\mathbb{Q}) = 1 - \frac{1}{2}\int \left[\sqrt{p(x)} - \sqrt{q(x)}\right]^2 d\mu(x),$$

$$\begin{aligned}\rho(\mathbb{P},\mathbb{Q}) = 1 \quad &\Leftrightarrow \quad \int \left[\sqrt{p(x)} - \sqrt{q(x)}\right]^2 d\mu(x) = 0\\ &\Leftrightarrow \quad p(x) = q(x) \quad \mu\text{-p.s} \quad \Leftrightarrow \quad \mathbb{P} = \mathbb{Q}.\end{aligned}$$

Pour simplifier l'écriture, notons

$$\xi_\theta(x) = \sqrt{f_\theta(x)}, \quad \rho(\theta,\theta_0) = \int \xi_\theta(x)\xi_{\theta_0}(x)d\mu(x).$$

Le modèle statistique étant identifiable, on a alors

$$0 \leqslant \rho(\theta,\theta_0) \leqslant 1,\ \rho(\theta,\theta_0) = 1 \ s.s.i\ \theta = \theta_0.$$

Dans la suite, on va étudier respectivement la consistance dans le cas Θ compact et le cas Θ ouvert.

Cas Θ compact dans $\mathbb{R}^d$

Du point de vue des applications, l'ensemble des paramètres est typiquement borné. On va donc étudier le cas où Θ est compact dans $\mathbb{R}^d$. Notons

$$B(\theta,\Delta) = \{t : |t-\theta| < \Delta, t \in \Theta\}$$
$$\xi_\theta^\Delta(x) = \sup_t\{\xi_t(x) : t \in \overline{B}(\theta,\Delta)\}.$$

Une condition suffisante pour que (H_0) soit réalisée est

$(\forall\delta > 0)(\forall\theta \in \Theta_\delta)(\exists\Delta_\theta = \Delta(\theta,\delta) > 0)$ tel que

$$\int \xi_\theta^{\Delta_\theta}(x)\xi_{\theta_0}(x)d\mu(x) < 1. \qquad (H_0^{\Delta_0})$$

Théorème 2.5. *Θ étant compact dans $\mathbb{R}^d$, sous l'hypothèse $(H_0^{\Delta_0})$, l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ converge (en probabilité) vers θ_0.*

Démonstration. L'ensemble $\Theta_\delta = \{\theta : |\theta - \theta_0| \geqslant \delta, \theta \in \Theta\}$ est compact puisque il est l'intersection d'un fermé et d'un compact. Par suite, il existe un nombre $T = T(\delta)$ fini de $\Omega_k = B(\theta_k, \Delta_{\theta_k})$ avec $\theta_k \in \Theta_\delta$ et $\Delta_{\theta_k} = \Delta(\theta_k, \delta)$ dans $(H_0^{\Delta_0})$ tel que

$$\Theta_\delta \subseteq \bigcup_{k=1\ldots T} \Omega_k.$$

Alors

$$\{\sup_{\theta\in\Theta_\delta} \mathbf{Z}_n(\theta) \geqslant 1\} \subseteq \bigcup_{k=1\ldots T} \{\sup_{\theta\in\overline{\Omega}_k} \mathbf{Z}_n(\theta) \geqslant 1\}.$$

L'Inégalité de Markov dit que

$$\begin{aligned}\mathbb{P}_{\theta_0}[\sup_{\theta\in\overline{\Omega}_k} \mathbf{Z}_n(\theta) \geqslant 1] \leqslant & \mathbb{E}_{\theta_0}[\sup_{\theta\in\overline{\Omega}_k} \mathbf{Z}_n^{1/2}(\theta)] \\ =& \Big(\int \xi_{\theta_k}^{\Delta_{\theta_k}}(x)\xi_{\theta_0}(x)d\mu(x)\Big)^n \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

Il s'ensuit que

$$\mathbb{P}_{\theta_0}[\sup_{\theta\in\Theta_\delta} \mathbf{Z}_n(\theta) \geqslant 1] \leqslant \sum_{k=1\ldots T} \mathbb{P}_{\theta_0}[\sup_{\theta\in\overline{\Omega}_k} \mathbf{Z}_n(\theta) \geqslant 1] \xrightarrow[n\to\infty]{} 0.$$

Donc (H_0) est bien vérifiée, d'où la conclusion. □

Posons

$$\xi_\theta^\circ(x) = \limsup_{t\to\theta} \xi_t(x) = \lim_{\Delta\to 0} \xi_\theta^\Delta(x).$$

La condition $(H_0^{\Delta_0})$ peut être décomposée en les deux conditions suivantes :

Théorème 2.6. *La condition $(H_0^{\Delta_0})$ est équivalente à la réalisation simultanée des deux conditions suivantes :*

- *Pour tous les $\theta \neq \theta_0$*

$$\int \xi_\theta^\circ(x)\xi_{\theta_0}(x)d\mu(x) < 1 \qquad (H_0^\circ)$$

- *Pour tous les $\theta \neq \theta_0$, il existe un $\tilde{\Delta}_\theta > 0$ tel que*

$$\int \xi_\theta^{\tilde{\Delta}_\theta}(x)\xi_{\theta_0}(x)d\mu(x) < +\infty. \qquad (H_I)$$

La démonstration est basée sur le théorème de la convergence monotone, on ne détaille pas la preuve ici parce que dans la partie suivante, on verra une démonstration plus générale. Maintenant, on va s'intéresser au cas où Θ est ouvert dans $\mathbb{R}^d$.

Cas Θ ouvert dans $\mathbb{R}^d$

De temps en temps, on a besoin traiter des modèles statistiques où Θ est ouvert dans $\mathbb{R}^d$. Pour avoir la consistance, on a besoin contrôler le rapport de vraisemblance $\mathbf{Z}_n(\theta)$ quand θ est loin de θ_0. On va d'abord décomposer Θ en des sous-ensembles compacts.

On admet ici une conclusion dans l'analyse classique [7] :

Lemme 2.7. *Soit Θ un ouvert dans $\mathbb{R}^d$, alors il existe une suite $\{K_j\}_{j\geqslant 1}$ de compacts dans Θ tel que*

$$\Theta = \bigcup_{j\geqslant 1} K_j, \quad K_j \subset \overset{\circ}{K}_{j+1} .$$

Une fois que la décomposition de $\Theta = \cup_{j\geqslant 1} K_j$ est choisie, on a naturellement une condition suffisante :

$$\lim_{j\to\infty} \int \sup_{\theta\in\Theta\setminus K_j} \xi_\theta(x)\cdot\xi_{\theta_0}(x) d\mu(x) < 1. \tag{H_J}$$

Théorème 2.8. *Θ étant ouvert dans $\mathbb{R}^d$, sous les hypothèses (H_J) et $(H_0^{\Delta_0})$, l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ converge (en probabilité) vers θ_0.*

Démonstration. Par (H_J), il existe un $j_0 \geqslant 1$ tel que

$$\omega_0 = \int \sup_{\theta\in\Theta\setminus K_{j_0}} \xi_\theta(x)\cdot\xi_{\theta_0}(x) d\mu(x) < 1.$$

Il s'ensuit que

$$\mathbb{P}_{\theta_0}[\sup_{\theta\in\Theta\setminus K_{j_0}} \mathbf{Z}_n(\theta) \geqslant 1] \leqslant \mathbb{E}_{\theta_0}[\sup_{\theta\in\Theta\setminus K_{j_0}} \mathbf{Z}_n^{1/2}(\theta)] = \omega_0^n \xrightarrow[n\to\infty]{} 0.$$

Or pour tout $\delta > 0$,

$$\{\sup_{\theta\in\Theta_\delta} \mathbf{Z}_n(\theta) \geqslant 1\} \subseteq \{\sup_{\theta\in\Theta\setminus K_{j_0}} \mathbf{Z}_n(\theta) \geqslant 1\} \cup \{\sup_{\theta\in\Theta_\delta\cap K_{j_0}} \mathbf{Z}_n(\theta) \geqslant 1\}.$$

$\Theta_\delta \cap K_{j_0}$ étant compact, le Théorème 2.5 implique alors que

$$\mathbb{P}_{\theta_0}[\sup_{\theta\in\Theta_\delta\cap K_{j_0}} \mathbf{Z}_n(\theta) \geqslant 1] \xrightarrow[n\to\infty]{} 0.$$

Donc (H_0) est bien vérifiée, d'où la conclusion. $\square$

2.2 Rapport de log-vraisemblance et divergence de Kullback-Leibler

Dans cette partie, on considère le rapport de log-vraisemblance qui nous permet caractériser la consistance par la loi des grands nombres. On trouvera des conclusions analogues mais plus générales.

Définition 2.9 (Rapport de log-vraisemblance)**.** On appelle rapport de log-vraisemblance (par rapport à θ_0), définie $\mathbb{P}_{\theta_0}$-p.s sur Θ, la quantité

$$\mathbf{M}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{f_\theta(X_i)}{f_{\theta_0}(X_i)} = \frac{1}{n}[L_n(\theta) - L_n(\theta_0)].$$

En raisonnant comme précédemment, $\hat{\theta}_n = \arg\max_{\theta\in\Theta} \mathbf{M}_n(\theta)$, par suite $\mathbf{M}_n(\hat{\theta}_n) \geqslant \mathbf{M}_n(\theta_0) \geqslant 0$ $\mathbb{P}_{\theta_0}$-p.s. Or $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_{\theta_0}} \theta_0$ s'écrit :

$$(\forall\delta > 0)(\forall\varepsilon > 0)(\exists N = N(\delta, \varepsilon) \in \mathbb{N}^*)(n \geqslant N \Rightarrow \mathbb{P}_{\theta_0}[|\hat{\theta}_n - \theta_0| < \delta] \geqslant 1 - \varepsilon).$$

Posons $\Theta_\delta = \{\theta : |\theta - \theta_0| \geqslant \delta, \theta \in \Theta\}$, alors

$$\{\sup_{\theta\in\Theta_\delta} \mathbf{M}_n(\theta) < 0\} \subseteq \{|\hat{\theta}_n - \theta_0| < \delta\}.$$

Donc une condition analogue est

$$\forall\delta > 0, \quad \mathbb{P}_{\theta_0}[\sup_{\theta\in\Theta_\delta} \mathbf{M}_n(\theta) < 0] \xrightarrow[n\to\infty]{} 1. \qquad (K_0)$$

Plus précisément,

Proposition 2.10. *Sous l'hypothèse (K_0), l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ converge (en probabilité) vers θ_0.*

On introduit la divergence de Kullback-Leibler pour trouver des choses plus intéressantes.

Définition 2.11 (Divergence de Kullback-Leibler)**.** On appelle divergence de Kullback-Leibler entre $\mathbb{P}$ et $\mathbb{Q}$ la quantité

$$\mathbf{K}(\mathbb{P}, \mathbb{Q}) = \begin{cases} \int \log \frac{d\mathbb{P}}{d\mathbb{Q}} d\mathbb{P} = \int \log \frac{p(x)}{q(x)} p(x) d\mu(x) & \text{si } \mathbb{P} \ll \mathbb{Q}, \\ +\infty & \text{sinon.} \end{cases}$$

On va vérifier que cette quantité est bien définie dans le cas où $\mathbb{P} \ll \mathbb{Q}$. Il suffit de monter que l'intégrale existe sur $S = \{x | p(x) > 0 \text{ et } q(x) > 0\}$ car $\mathbb{P}(S^c) = 0$. En effet, $\mathbb{P} \ll \mathbb{Q}$ et $\mathbb{Q}[q(x) = 0] = 0$ implique que $\mathbb{P}[q(x) = 0] = 0$, donc $\mathbb{P}(S^c) \leqslant \mathbb{P}[p(x)=0] + \mathbb{P}[q(x)=0] = 0$. Par suite, $\mathbb{P}$-p.s $0 < \frac{p(x)}{q(x)} < +\infty$ et $\log \frac{p(x)}{q(x)} \in \mathbb{R}$. D'autre part

$$\int_S \log \frac{p(x)}{q(x)} p(x) d\mu(x) = -\int_S \log \frac{q(x)}{p(x)} p(x) d\mu(x).$$

Comme sur $]0,+\infty[$, $\log x \leqslant x-1$ et l'égalité est vraie seulement pour $x=1$, il vient

$$\int_S \log \frac{q(x)}{p(x)} p(x) d\mu(x) \leqslant \int_S \big(q(x) - p(x)\big) d\mu(x) \leqslant 0,$$

car $\mathbb{P}(S) = 1, \mathbb{Q}(S) \leqslant 1$. Donc la divergence de Kullback-Leibler est bien définie. Il s'ensuit aussi que toujours $\mathbf{K}(\mathbb{P},\mathbb{Q}) \geqslant 0$ et que $\mathbf{K}(\mathbb{P},\mathbb{Q}) = 0$ s.s.i $\mathbb{P} = \mathbb{Q}$. Pour simplifier, on note,

$$\mathbf{K}(\theta,\theta_0) = \mathbf{K}(\mathbb{P}_{\theta_0}, \mathbb{P}_\theta) = \int \log \frac{f_{\theta_0}(x)}{f_\theta(x)} f_{\theta_0}(x) d\mu(x).$$

On a donc $\mathbf{K}(\theta,\theta_0) \geqslant 0$, et $\mathbf{K}(\theta,\theta_0) = 0$ s.s.i $\theta = \theta_0$.

Cas Θ compact dans $\mathbb{R}^d$

Rappellons que $B(\theta,\Delta) = \{t : |t-\theta| < \Delta, t \in \Theta\}$. Posons

$$f_\theta^\Delta(x) = \sup_t \{f_t(x) : t \in \overline{B}(\theta,\Delta)\}.$$

Une condition suffisante pour que (K_0) soit réalisée est

$(\forall \delta > 0)(\forall \theta \in \Theta_\delta)(\exists \Delta_\theta = \Delta(\theta,\delta) > 0)$ tel que

$$\int \log \frac{f_\theta^{\Delta_\theta}(x)}{f_{\theta_0}(x)} f_{\theta_0}(x) d\mu(x) < 0. \tag{$K_0^{\Delta_0}$}$$

En effet,

Théorème 2.12. *Θ étant compact dans $\mathbb{R}^d$, sous l'hypothèse $(K_0^{\Delta_0})$, l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ converge (en probabilité) vers θ_0.*

Démonstration. Pour tout $\delta > 0$, l'ensemble Θ_δ est compact puisque il est l'intersection d'un fermé et d'un compact. Par suite, il existe un nombre fini $T = T(\delta)$ de $\Omega_k = B(\theta_k, \Delta_k)$ avec $\theta_k \in \Theta_\delta$ et $\Delta_k = \Delta(\theta_k,\delta)$ dans $(K_0^{\Delta_0})$ tel que

$$\Theta_\delta \subseteq \bigcup_{k=1..T} \Omega_k.$$

La loi des grands nombres dit que

$$\begin{aligned}
\sup_{\theta\in\Theta_\delta} \mathbf{M}_n(\theta) &\leqslant \max_{k=1..T} \sup_{\theta\in\overline{\Omega}_k} \mathbf{M}_n(\theta) \\
&= \max_{k=1..T} \frac{1}{n}\sum_{i=1}^n \sup_{\theta\in\overline{\Omega}_k} \log \frac{f_\theta(X_i)}{f_{\theta_0}(X_i)} \\
&= \max_{k=1..T} \frac{1}{n}\sum_{i=1}^n \log \frac{f_{\theta_k}^{\Delta_k}(X_i)}{f_{\theta_0}(X_i)} \xrightarrow[n\to\infty]{\mathbb{P}_{\theta_0}} \max_{k=1..T} \int \log \frac{f_{\theta_k}^{\Delta_k}(x)}{f_{\theta_0}(x)} f_{\theta_0}(x)d\mu(x) < 0.
\end{aligned}$$

Par suite, $\forall\delta > 0$,

$$\mathbb{P}_{\theta_0}[\sup_{\theta\in\Theta_\delta} \mathbf{M}_n(\theta) < 0] \xrightarrow[n\to\infty]{} 1.$$

Donc (K_0) est bien vérifiée, d'où la conclusion. □

Posons

$$f_\theta^\circ(x) = \limsup_{u\to\theta} f_u(x) = \lim_{\Delta\to 0} f_\theta^\Delta(x).$$

Le théorème suivant est analogue au Théorème 2.6 dont on a retardé la démonstration.

Théorème 2.13. *La condition $(K_0^{\Delta_0})$ est équivalente à la réalisation simultanée des deux conditions suivantes :*

- *Pour tous les $\theta \neq \theta_0$*

$$\int \log \frac{f_\theta^\circ(x)}{f_{\theta_0}(x)} f_{\theta_0}(x)d\mu(x) < 0 \qquad (K_0^\circ)$$

- *Pour tous les $\theta \neq \theta_0$, il existe un $\tilde{\Delta}_\theta > 0$ tel que*

$$\int \log \frac{f_\theta^{\tilde{\Delta}_\theta}(x)}{f_{\theta_0}(x)} f_{\theta_0}(x)d\mu(x) < +\infty. \qquad (K_I)$$

Démonstration. Il suffit de prendre $\delta = |\theta - \theta_0|$ pour que $(K_0^{\Delta_0})$ entraîne (K_0°) et (K_I), car $f_\theta^\circ(x) \leqslant f_\theta^\Delta(x)$ pour tout $\Delta > 0$.

Inversement, pour tout $\theta \neq \theta_0$, le théorème de la convergence monotone implique quand Δ est assez petit $(0 < \Delta < \tilde{\Delta}_\theta)$,

$$\lim_{\Delta\to 0} \int \log \frac{f_\theta^\Delta(x)}{f_{\theta_0}(x)} f_{\theta_0}(x)d\mu(x) = \int \log \frac{f_\theta^\circ(x)}{f_{\theta_0}(x)} f_{\theta_0}(x)d\mu(x) < 0.$$

Donc l'existence de Δ_θ dans $(K_0^{\Delta_0})$ est bien vérifiée. □

Introduisons maintenant des conditions plus simples réalisant (K_0°) et (K_I).

Proposition 2.14. *Si pour tous les* $\theta \neq \theta_0$*, il existe un ensemble* $C_\theta \in \mathcal{E}$ *avec* $\mu(C_\theta^c) = 0$ *tel que la fonction* $t \mapsto f_t(x)$ *est semi-continue supérieurement au point* θ *pour tous les* $x \in C_\theta$*, alors* (K_0°) *est réalisée.*

Démonstration. Pour tout $\theta \neq \theta_0$, on a $f_\theta^\circ(x) \leqslant f_\theta(x)$ pour les $x \in C_\theta$. Par suite,

$$\int \log \frac{f_\theta^\circ(x)}{f_{\theta_0}(x)} f_{\theta_0}(x) d\mu(x) \leqslant \int_{C_\theta} \log \frac{f_\theta(x)}{f_{\theta_0}(x)} f_{\theta_0}(x) d\mu(x) = -\mathbf{K}(\theta, \theta_0) < 0.$$

□

Proposition 2.15. *Si pour tous les* $\theta \neq \theta_0$*, il existe* $\Delta_\theta > 0$ *tel que* $f_\theta^{\Delta_\theta}(x) \in L^1(\mu)$*. Alors* (K_I) *est réalisée.*

Démonstration. Pour tous les $\theta \neq \theta_0$, il existe un $\Delta_\theta > 0$ tel que $f_\theta^{\Delta_\theta}(x) \in L^1(\mu)$. Or $\log \frac{f_\theta^{\Delta_\theta}(x)}{f_{\theta_0}(x)} f_{\theta_0}(x) \leqslant f_\theta^{\Delta_\theta}(x) - f_{\theta_0}(x)$, on a bien

$$\int \log \frac{f_\theta^{\Delta_\theta}(x)}{f_{\theta_0}(x)} f_{\theta_0}(x) d\mu(x) \leqslant \int f_\theta^{\Delta_\theta}(x) d\mu(x) - 1 < +\infty.$$

□

Jusqu'à maintenant, on a trouvé plusieurs conditions suffisantes pour que $\hat{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}_{\theta_0}} \theta_0$. Mais ce sont des conditions seulement au voisinage de θ_0. Pour avoir la consistance de l'estimateur du maximum de vraisemblance, il suffit de faire "courir" θ_0 sur Θ, d'où,

Théorème 2.16. Θ *étant compact dans* $\mathbb{R}^d$*, si pour tous les* $\theta \in \Theta$*,*

- *Il existe un ensemble* $C_\theta \in \mathcal{E}$ *avec* $\mu(C_\theta^c) = 0$ *tel que la fonction* $t \mapsto f_t(x)$ *est semi-continue supérieurement au point* θ *pour tous les* $x \in C_\theta$
- *Il existe* $\Delta_\theta > 0$ *tel que* $f_\theta^{\Delta_\theta}(x) \in L^1(\mu)$*.*

Alors l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ *est consistant.*

Il faut dire qu'à partir de conditions légèrement différentes de celles du Théorème 2.16, on peut avoir une conclusion plus forte : la consistance uniforme. En effet, en remplaçant semi-continuité supérieurement par continuité, on obtient

Théorème 2.17. Θ *étant compact dans* $\mathbb{R}^d$*, si pour tous les* $\theta \in \Theta$*,*

- *Il existe un ensemble* $C_\theta \in \mathcal{E}$ *avec* $\mu(C_\theta^c) = 0$ *tel que, la fonction* $t \mapsto f_t(x)$ *est continue au point* θ *pour tous les* $x \in C_\theta$
- *Il existe* $\Delta_\theta > 0$ *tel que* $f_\theta^{\Delta_\theta}(x) \in L^1(\mu)$*.*

Alors l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ *est uniformément consistant.*

Démonstration. On va le prouver par l'absurde. D'après le Théorème 2.16, $\hat{\theta}_n$ converge "ponctuellement", i.e. $\forall\theta\in\Theta, \forall\delta>0, \mathbb{P}_\theta[|\hat{\theta}_n-\theta|\geqslant\delta]\xrightarrow[n\to\infty]{}0$.

Pour que $\hat{\theta}_n$ soit uniformément consistant, il faut et il suffit que

$$\forall\delta>0,\quad \sup_{\theta\in\Theta}\mathbb{P}_\theta[|\hat{\theta}_n-\theta|\geqslant\delta]\xrightarrow[n\to\infty]{}0.$$

Si $\hat{\theta}_n$ n'est pas uniformément consistant, alors

$(\exists\delta_0>0)(\exists\varepsilon_0>0)(\forall k\in\mathbb{N}^*)(\exists n_k\geqslant k)(\exists\theta_k\in\Theta)$ tel que

$$\mathbb{P}_{\theta_k}[|\hat{\theta}_{n_k}-\theta_k|\geqslant\delta_0]\geqslant\varepsilon_0.$$

Θ étant compact, quand $k\to\infty$, on peut extraire une sous suite de θ_k, notée encore θ_k, qui converge vers un certain $\theta_0\in\Theta$. On a donc $\theta_k\xrightarrow[k\to\infty]{}\theta_0$ avec

$$\mathbb{P}_{\theta_k}[|\hat{\theta}_{n_k}-\theta_k|\geqslant\delta_0]\geqslant\varepsilon_0.$$

D'autre part, $\{|\hat{\theta}_{n_k}-\theta_k|\geqslant\delta_0\}\subseteq\{|\hat{\theta}_{n_k}-\theta_0|\geqslant\frac{\delta_0}{2}\}\cup\{|\theta_k-\theta_0|\geqslant\frac{\delta_0}{2}\}$ avec $\mathbb{P}_{\theta_k}[|\theta_k-\theta_0|\geqslant\frac{\delta_0}{2}]\xrightarrow[k\to\infty]{}0$ car $\theta_k\xrightarrow[k\to\infty]{}\theta_0$. Par suite, on aboutit à une contradiction si

$$\mathbb{P}_{\theta_k}[|\hat{\theta}_{n_k}-\theta_0|\geqslant\frac{\delta_0}{2}]\xrightarrow[k\to\infty]{}0. \tag{1}$$

Grâce à la convergence "ponctuelle", on a $\mathbb{P}_{\theta_0}[|\hat{\theta}_{n_k}-\theta_0|\geqslant\frac{\delta_0}{2}]\xrightarrow[k\to\infty]{}0$ et il reste à remarquer que

$$\begin{aligned}&\Big|\mathbb{P}_{\theta_k}[|\hat{\theta}_{n_k}-\theta_0|\geqslant\frac{\delta_0}{2}]-\mathbb{P}_{\theta_0}[|\hat{\theta}_{n_k}-\theta_0|\geqslant\frac{\delta_0}{2}]\Big|\\ &\leqslant\int(1_{|\hat{\theta}_{n_k}-\theta_0|\geqslant\frac{\delta_0}{2}}|f_{\theta_k}(x)-f_{\theta_0}(x)|d\mu(x)\\ &\leqslant\int|f_{\theta_k}(x)-f_{\theta_0}(x)|d\mu(x)\xrightarrow[k\to\infty]{}0.\end{aligned}$$

En effet, sur l'ensemble C_{θ_0} avec $\mu(C^c_{\theta_0})=0$, on a $f_{\theta_k}(x)\to f_{\theta_0}(x)$. De plus, il existe un $\Delta_{\theta_0}>0$ tel que $f^{\Delta_{\theta_0}}_{\theta_0}(x)\in L^1(\mu)$. Par suite, $|f_{\theta_k}(x)-f_{\theta_0}(x)|\leqslant 2f^{\Delta_{\theta_0}}_{\theta_0}(x)\in L^1(\mu)$ quand k est assez grand. On s'applique le théorème de la convergence dominée et on obtient

$$\int|f_{\theta_k}(x)-f_{\theta_0}(x)|d\mu(x)\xrightarrow[k\to\infty]{}0.$$

L'inégalité triangulaire nous dit alors

$$\mathbb{P}_{\theta_k}[|\hat{\theta}_{n_k}-\theta_0|\geqslant\frac{\delta_0}{2}]\xrightarrow[k\to\infty]{}0.$$

D'où la contradiction est bien aboutie. □

Il faut remarquer que les conclusions dans ces deux parties sont assez analogues. En effet, un lien fondamental entre l'affinité de Hellinger et la divergence de Kullback-Leibler nous explique ces phénomènes.

Rappelons que $\mathbb{P}$ et $\mathbb{Q}$ sont deux lois sur $(E, \mathcal{E})$ dominées par une mesure positive σ-finie μ avec

$$p = \frac{d\mathbb{P}}{d\mu}, \quad q = \frac{d\mathbb{Q}}{d\mu}.$$

Lemme 2.18.

$$-\mathbf{K}(\mathbb{P}, \mathbb{Q}) \leqslant 2 \log \rho(\mathbb{P}, \mathbb{Q}) \leqslant 2[\rho(\mathbb{P}, \mathbb{Q}) - 1].$$

Démonstration. Pour avoir la première inégalité, on applique l'Inégalité de Jensen à la fonction $-\frac{1}{2}\mathbf{K}(\mathbb{P}, \mathbb{Q})$ sachant que log est concave. Pour la deuxième inégalité, il suffit s'appliquer l'inégalité $\log x \leqslant x - 1$ sur $[0, +\infty[$, d'où la conclusion. □

On peut alors déduire les conclusions suivantes par le même raisonnement,

Proposition 2.19.

$$\begin{aligned} (H_0^{\Delta_0}) &\Rightarrow (K_0^{\Delta_0}) \\ (H_0^{\circ}) &\Rightarrow (K_0^{\circ}) \\ (H_I) &\Rightarrow (K_I). \end{aligned}$$

Cette proposition nous dit que les conditions déduites par la divergence de Kullback-Leibler sont plus générales que celles par l'affinité de Hellinger.

3 Modèle Régulier et Normalité Asymptotique

Dans cette partie, on va étudier la normalité asymptotique des estimateurs du maximum de vraisemblance dans le cas où Θ est un ouvert de $\mathbb{R}$. Les estimateurs par la méthode des moments généralement possèdent cette propriété grâce au théorème limite centrale. R. Fisher a trouvé que dans certains modèles réguliers, la normalité asymptotique pour les estimateurs du maximum de vraisemblance reste vraie, de plus $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} N(0, \mathbf{I}^{-1}(\theta))$, où $\mathbf{I}(\theta)$ s'appelle l'information de Fisher, qui est une quantité très importante dans Statistique.

3.1 Modèle régulier

On a bien étudié la consistance pour les estimateurs du maximum de vraisemblance. Rappelons que $\xi_\theta(x) = \sqrt{f_\theta(x)}$, qui peut être considérée comme une fonction de Θ dans $L^2(\mu)$. On va s'intéresser plutôt à cet opérateur car $L^2(\mu)$ est un espace de Hilbert.

Nous allons d'abord introduire le produit scalaire et la distance dans $L^2(\mu)$. Rappelons que l'affinité de Hellinger est définie par

$$\rho(\mathbb{P}, \mathbb{Q}) = \int \sqrt{p(x)q(x)}d\mu(x).$$

Cette quantité peut naturellement être interprétée comme un produit scalaire dans $L^2(\mu)$. Soit $\xi, \eta \in L^2(\mu)$, on définit

$$\langle \xi, \eta \rangle_{\mathrm{L}^2(\mu)} = \int \xi \eta d\mu$$

$$\|\xi\|_{\mathrm{L}^2(\mu)} = \sqrt{\langle \xi, \xi \rangle_{\mathrm{L}^2(\mu)}} = \sqrt{\int \xi^2 d\mu}.$$

Il faut dire que si le produit scalaire et la distance de ξ, η sont indépendants du choix de la mesure dominante μ, on les note simplement $\langle \xi, \eta \rangle$ et $\|\xi\|$. (Par exemple : quand ξ et η correspondent aux racines des densités de deux probabilités dans un modèle statistique dominé).

Par suite, la distance entre $\mathbb{P}$ et $\mathbb{Q}$ dans $L^2(\mu)$ est caractérisée par :

Définition 3.1 (Distance de Hellinger)**.** On appelle distance de Hellinger entre $\mathbb{P}$ et $\mathbb{Q}$ la quantité

$$\mathbf{h}(\mathbb{P}, \mathbb{Q}) = \frac{1}{\sqrt{2}} \|\sqrt{p} - \sqrt{q}\| = \left(\frac{1}{2} \int \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 d\mu(x) \right)^{1/2}.$$

Par définition, on a $\mathbf{h}^2(\mathbb{P},\mathbb{Q}) = 1 - \rho(\mathbb{P},\mathbb{Q})$. On peut alors vérifier facilement que $\mathbf{h}(\mathbb{P},\mathbb{Q})$ satisfait les trois conditions classiques pour être une distance. D'ailleurs, $0 \leqslant \mathbf{h}(\mathbb{P},\mathbb{Q}) \leqslant 1$ et $\mathbf{h}(\mathbb{P},\mathbb{Q}) = 0$ s.s.i $\mathbb{P} = \mathbb{Q}$.

Notons $\mathbf{h}(\theta+u,\theta) = \mathbf{h}(\mathbb{P}_{\theta+u},\mathbb{P}_\theta)$ où $\theta, \theta+u \in \Theta$. Alors

$$0 \leqslant \mathbf{h}(\theta+u,\theta) \leqslant 1, \text{ et } \mathbf{h}(\theta+u,\theta) = 0 \text{ s.s.i } u = 0.$$

On introduit maintenant le modèle régulier qu'on va étudier :

Définition 3.2 (Modèle régulier). Un modèle statistique est dit régulier si

1. Θ est un intervalle ouvert de $\mathbb{R}$.
2. $\xi_\theta(x)$ est absoluement continue sur Θ pour $[\mu]$-p.p x.
3. L'application $\xi'_\theta(x) : \Theta \times E \to \mathbb{R}$ est mesurable.
4. L'opérateur $\xi'_\theta : \Theta \to L^2(\mu)$ est continu.

L'absolue continuité de $\xi_\theta(x)$ implique que la dérivée $\xi'_\theta(x)$ par rapport à θ existe presque sûrement, mais on peut définir la dérivée partout en supposant que $\xi'_\theta(x) = 0$ sur l'ensemble non-dérivable qui est de mesure nulle. Il faut aussi dire que la régularité ne dépend pas de notre choix de μ, qui implique que généralement notre conclusions ne dépendent pas de notre choix de μ.

L'analyse classique nous dit les choses suivantes [6] :

Lemme 3.3. *Soit ξ_θ absolument continue sur Θ, de dérivée ξ'_θ, on a*

1. $f'_\theta = 2\xi_\theta\xi'_\theta$.
2. $\xi_\theta(x) = 0 \Rightarrow \xi'_\theta(x) = 0$.
3. *Si $\theta, \theta+u \in \Theta$, alors pour $[\mu]$-p.p x*

$$\xi_{\theta+u}(x) - \xi_\theta(x) = \int_\theta^{\theta+u} \xi'_t(x)dt = u \cdot \int_0^1 \xi'_{\theta+\lambda u}(x)d\lambda.$$

Remarque 3.4. Il faut faire attention au cas $u < 0$, où par définition

$$\int_\theta^{\theta+u} \xi'_t(x)dt = -\int_{\theta+u}^{\theta} \xi'_t(x)dt.$$

Dans un modèle régulier, la continuité de l'opérateur $\xi'_\theta : \Theta \mapsto L^2(\mu)$ implique que $\xi'_\theta \in L^2(\mu)$, on peut alors définir :

Définition 3.5 (Information de Fisher). On appelle Information de Fisher la quantité

$$\mathbf{I}(\theta) = 4\int_{\{x:f_\theta(x)>0\}} [\xi'_\theta(x)]^2 d\mu(x).$$

Dans le modèle régulier, l'Information de Fisher peut aussi être définie par $\mathbf{I}(\theta) = 4\left\|\xi'_\theta\right\|^2$, puisque $(f_\theta(x) = 0) \Rightarrow (\xi_\theta(x) = 0) \Rightarrow (\xi'_\theta(x) = 0)$.

Quelques propriétés dans le modèle régulier sont données par la proposition suivante :

Proposition 3.6. *Dans le modèle régulier, on a*

1. $\mathbf{I}(\theta)$ *est continue sur* Θ.
2. *L'application* $\xi_\theta : \Theta \to L^2(\mu)$ *est continue, et*
$$\int |\xi_{\theta+u}(x) - \xi_\theta(x) - \xi'_\theta(x) \cdot u|^2 d\mu(x) = o(u^2), \quad u \to 0. \tag{2}$$
3. *L'affinité de Hellinger* $\rho : \Theta^2 \to \mathbb{R}$ *est continue.*
La distance de Hellinger $\mathbf{h} : \Theta^2 \to \mathbb{R}$ *est continue.*
4. *Si* $\theta \in \Theta$ *et* $u \neq 0$, *alors*
$$\lim_{u\to 0} \frac{\mathbf{h}^2(\theta+u,\theta)}{u^2} = \frac{\mathbf{I}(\theta)}{8}. \tag{3}$$
5. *Si* $\theta, \theta+u \in \Theta$ *et* $u \neq 0$, *alors*
$$\frac{\mathbf{h}^2(\theta+u,\theta)}{u^2} \leqslant \sup_{\lambda\in[0,1]} \frac{\mathbf{I}(\theta+\lambda u)}{8}.$$
6. J *étant compact dans* Θ, *alors pour tout* $\theta \in J$ *t.q..* $\mathbf{I}(\theta) > 0$,
$$g_J(\theta) = \inf_{\substack{u\in J-\theta \\ u\neq 0}} \frac{\mathbf{h}^2(\theta+u,\theta)}{u^2} > 0.$$

Démonstration. En vue du lemme 3.3 et du modèle régulier, on a

1. Par la continuité de ξ'_θ dans $L^2(\mu)$,
$$\begin{aligned} |\mathbf{I}(\theta+u) - \mathbf{I}(\theta)| &= 4\left| \left\|\xi'_{\theta+u}\right\|^2 - \left\|\xi'_\theta\right\|^2 \right| \\ &\leqslant 4\big(\left\|\xi'_{\theta+u}\right\| + \left\|\xi'_\theta\right\| \big) \cdot \big(\left\|\xi'_{\theta+u} - \xi'_\theta\right\| \big) \xrightarrow[u\to 0]{} 0. \end{aligned}$$
2. La continuité de ξ_θ dans $L^2(\mu)$ vient directement de la continuité de ξ'_θ dans $L^2(\mu)$ et du fait que $\xi_{\theta+u} - \xi_\theta = \int_\theta^{\theta+u} \xi'_t dt$. Il reste à remarquer que :
$$\begin{aligned} \int |\xi_{\theta+u} - \xi_\theta - \xi'_\theta \cdot u|^2 d\mu &= \int \left(\int_\theta^{\theta+u} (\xi'_t - \xi'_\theta) dt \right)^2 d\mu \\ &\leqslant \int \left(u \int_\theta^{\theta+u} (\xi'_t - \xi'_\theta)^2 dt \right) d\mu \\ &= u \int_\theta^{\theta+u} \left\|\xi'_t - \xi'_\theta\right\|^2 dt \\ &= o(u^2). \end{aligned}$$

3. En vertu que $\mathbf{h}^2(s,t) = 1-\rho(s,t)$ pour tout $(s,t) \in \Theta^2$, il suffit de prouver la continuité de l'affinité de Hellinger. Soit $(s_n, t_n) \in \Theta^2$ tel que $s_n \to s, t_n \to t$, alors grâce à la continuité de ξ_θ dans $L^2(\mu)$, on a

$$\begin{aligned}\left|\rho(s_n,t_n) - \rho(s,t)\right| &\leqslant \int \left|\xi_{s_n}\xi_{t_n} - \xi_s\xi_t\right| d\mu \\ &\leqslant \int |\xi_{s_n}| \cdot |\xi_{t_n} - \xi_t| + |\xi_t| \cdot |\xi_{s_n} - \xi_s| d\mu \\ &\leqslant \|\xi_{s_n}\| \, \|\xi_{t_n} - \xi_t\| + \|\xi_{s_n} - \xi_s\| \, \|\xi_t\| \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

4. Il suffit d'utiliser (2) pour obtenir que

$$\frac{\mathbf{h}^2(\theta+u,\theta)}{u^2} = \frac{1}{2} \cdot \frac{\|\xi_{\theta+u} - \xi_\theta\|^2}{u^2} \xrightarrow[u\to 0]{} \frac{1}{2} \cdot \left\|\xi'_\theta\right\|^2 = \frac{\mathbf{I}(\theta)}{8}.$$

5. L'intervalle $[\theta, \theta+u] \subseteq \Theta$ car $\theta, \theta+u \in \Theta$. On s'applique le Théorème de Fubini et l'Inégalité de Cauchy-Schwarz comme précédemment et on obtient

$$\begin{aligned}\mathbf{h}^2(\theta+u,\theta) &= \frac{1}{2}\int \left(\xi_{\theta+u} - \xi_\theta\right)^2 d\mu = \frac{1}{2}\int \left(\int_\theta^{\theta+u} \xi'_t dt\right)^2 d\mu \\ &\leqslant \frac{u}{2}\int_\theta^{\theta+u} \left\|\xi'_t\right\|^2 dt \leqslant u^2 \cdot \sup_{\lambda\in[0,1]} \frac{\mathbf{I}(\theta+\lambda u)}{8}.\end{aligned}$$

6. Par l'absurde. Si $g_J(\theta) = 0$, alors il existe une suite $u_n \in J - \theta$ et $u_n \neq 0$ tel que

$$\frac{\mathbf{h}^2(\theta+u_n,\theta)}{u_n^2} \xrightarrow[n\to\infty]{} 0.$$

J est compact, alors $J - \theta$ l'est aussi. On peut donc extraire une sous-suite (on la note encore u_n) de u_n qui converge vers un $u_0 \in J - \theta$.

• Si $u_0 \neq 0$, la continuité de $\mathbf{h}$ abouti la contradiction suivante

$$\frac{\mathbf{h}^2(\theta+u_n,\theta)}{u_n^2} \xrightarrow[n\to\infty]{} \frac{\mathbf{h}^2(\theta+u_0,\theta)}{u_0^2} > 0.$$

• Si $u_0 = 0$, d'après (3), on a aussi contradiction car

$$\frac{\mathbf{h}^2(\theta+u_n,\theta)}{u_n^2} \xrightarrow[n\to\infty]{} \frac{\mathbf{I}(\theta)}{8} > 0.$$

D'où la conclusion.

□

3.2 Locale asymptotique normalité

L'idée de la locale asymptotique normalité (LAN) est d'approximer une famille de loi générale par des lois gaussiennes. Elle a été introduite par A. Wald et bien développée par L. LeCam et J. Hajek. Comme on a vu dans les parties précédentes, le rapport de vraisemblance joue un rôle très important pour étudier les estimateurs, cette quantité peut être représentée sous une forme particulière sous condition LAN. On va aussi montrer que le modèle régulier qu'on vient d'introduire possède la propriété LAN, mais on va d'abord se placer dans un cadre plus général.

Soit Θ un ouvert de $\mathbb{R}$, on peut alors définir la dérivabilité de l'opérateur $\xi_\theta : \Theta \to L^2(\mu)$. En général, la dérivée d'un opérateur entre deux espaces de Banach X et Y est définie par une forme linéaire continue entre X et Y, i.e. un élément dans $\mathcal{L}(X,Y)$. Ici $\Theta \subseteq \mathbb{R}$, en vertu que $\mathcal{L}(\mathbb{R},Y)$ est isométrique à Y, on peut ainsi dire que ξ_θ est dérivable dans $L^2(\mu)$ s'il existe une fonction (dérivé) $\xi'_\theta : \Theta \to L^2(\mu)$, t.q.

$$\left\|\xi_{\theta+u} - \xi_\theta - \xi'_\theta \cdot u\right\|^2 = o(u^2), \quad \text{quand } u \to 0.$$

Il faut remarquer que la dérivabilité d'un opérateur implique immédiatement la continuité de cet opérateur. D'ailleurs,

Proposition 3.7. *Si ξ_θ est dérivable dans $L^2(\mu)$ de dérivé ξ'_θ, alors (quand $u \to 0$)*

$$\|\xi_{\theta+u} - \xi_\theta\|^2 = \left\|\xi'_\theta \cdot u\right\|^2 + o(u^2) \tag{4}$$

$$\int \left|(\xi_{\theta+u} - \xi_\theta)^2 - (\xi'_\theta \cdot u)^2\right| d\mu = o(u^2). \tag{5}$$

Démonstration. La dérivabilité de ξ_θ dans $L^2(\mu)$ dit que $\|\xi_{\theta+u} - \xi_\theta - \xi'_\theta \cdot u\| = o(u)$. Alors

$$\frac{\|\xi_{\theta+u} - \xi_\theta\|}{u} \xrightarrow[u\to 0]{} \left\|\xi'_\theta\right\|,$$

$$\frac{\|\xi_{\theta+u} - \xi_\theta\|^2}{u^2} \xrightarrow[u\to 0]{} \left\|\xi'_\theta\right\|^2.$$

Par suite (4) est bien vérifié. Pour (5), il suffit d'utiliser l'Inégalité de Cauchy-Schwarz :

$$\int \left|(\xi_{\theta+u} - \xi_\theta)^2 - (\xi'_\theta \cdot u)^2\right| d\mu \leqslant \left\|\xi_{\theta+u} - \xi_\theta + \xi'_\theta \cdot u\right\| \cdot \left\|\xi_{\theta+u} - \xi_\theta - \xi'_\theta \cdot u\right\| = o(u^2).$$

$\square$

Corollaire 3.8. *Si ξ_θ est dérivable dans $L^2(\mu)$ et $u \in \mathbb{R}$ fixé, alors (quand $n \to \infty$)*

$$\left\|\xi_{\theta+\frac{u}{\sqrt{n}}} - \xi_\theta\right\|^2 = \left\|\xi'_\theta \cdot \frac{u}{\sqrt{n}}\right\|^2 + o(\frac{1}{n}) \tag{6}$$

$$\int \left|(\xi_{\theta+\frac{u}{\sqrt{n}}} - \xi_\theta)^2 - (\xi'_\theta \cdot \frac{u}{\sqrt{n}})^2\right| d\mu = o(\frac{1}{n}) \tag{7}$$

$$\int \xi'_\theta \cdot \xi_\theta d\mu = 0. \tag{8}$$

Démonstration. Il reste à prouver (8). Prenons $u = 1$, on a alors

$$\sqrt{n}(\xi_{\theta+\frac{1}{\sqrt{n}}} - \xi_\theta) \xrightarrow[n\to\infty]{L^2(\mu)} \xi'_\theta, \quad \xi_{\theta+\frac{1}{\sqrt{n}}} \xrightarrow[n\to\infty]{L^2(\mu)} \xi_\theta.$$

Par la continuité du produit scalaire $\langle \cdot, \cdot \rangle$, on a

$$\begin{aligned}\left\langle \xi'_\theta, 2\xi_\theta \right\rangle &= \lim_{n\to\infty} \left\langle \sqrt{n}(\xi_{\theta+\frac{1}{\sqrt{n}}} - \xi_\theta), (\xi_{\theta+\frac{1}{\sqrt{n}}} + \xi_\theta) \right\rangle \\ &= \lim_{n\to\infty} \sqrt{n} \int (f_{\theta+\frac{1}{\sqrt{n}}} - f_\theta) d\mu = \lim_{n\to\infty} 0 = 0.\end{aligned}$$

D'où la conclusion. □

On introduit maintenant la définition de LAN :

Définition 3.9 (Locale asymptotique normalité)**.** Un modèle statistique est dit localement asymptotiquement normal (LAN) au point $\theta \in \Theta$ s'il existe une fonction non-dégénérée $\varphi(n) = \varphi(n, \theta)$ t.q. pour tout $u \in \mathbb{R}$, asymptotiquement $\theta + \varphi(n)u \in \Theta$ et

$$\prod_{i=1}^n \frac{f_{\theta+\varphi(n)u}}{f_\theta}(X_i) = \exp\left\{u\Delta_{n,\theta} - \frac{1}{2}u^2 + \psi_{n,\theta}(u)\right\},$$

où

$$\psi_{n,\theta}(u) \xrightarrow[n\to\infty]{\mathbb{P}_\theta} 0, \quad \Delta_{n,\theta} \xrightarrow[n\to\infty]{\mathcal{L}} N(0,1).$$

Il y a beaucoup de modèles statistiques classiques qui satisfont cette propriété. En effet, Le modèle régulier précédent est LAN aux points θ où $\mathbf{I}(\theta) > 0$ avec $\varphi(n, \theta) = [n\mathbf{I}(\theta)]^{-1/2}$:

Théorème 3.10. *Dans le modèle régulier, la fonction de rapport de vraisemblance au point θ t.q. $\mathbf{I}(\theta) > 0$ s'écrit sous la forme suivante :*

$$\prod_{i=1}^n \frac{f_{\theta+u/\sqrt{n\mathbf{I}(\theta)}}}{f_\theta}(X_i) = \exp\left\{u\Delta_{n,\theta} - \frac{1}{2}u^2 + \psi_{n,\theta}(u)\right\},$$

où

$$\psi_{n,\theta}(u) \xrightarrow[n\to\infty]{\mathbb{P}_\theta} 0, \quad \Delta_{n,\theta} \xrightarrow[n\to\infty]{\mathcal{L}} N(0,1).$$

On se place maintenant dans un cadre plus général où LAN reste encore vraie : Soient Θ un ouvert de $\mathbb{R}$, l'opérateur $\xi_\theta : \Theta \to L^2(\mu)$ et il existe une fonction (dérivé) $\xi'_\theta : \Theta \to L^2(\mu)$ t.q. pour tout $u \in \mathbb{R}$ fixé on a (quand $n \to \infty$)

$$\int_{\{f_\theta>0\}} |\xi_{\theta+\frac{u}{\sqrt{n}}} - \xi_\theta - \xi'_\theta \cdot \frac{u}{\sqrt{n}}|^2 d\mu = o(\frac{1}{n}). \tag{H_d}$$

$$\int_{\{f_\theta=0\}} f_{\theta+\frac{u}{\sqrt{n}}} d\mu = o(\frac{1}{n}). \tag{H_s}$$

Toutes les conclusions de la Proposition 3.7 et du Corollaire 3.8 restent encore vraies si on se restreint à $S_\theta = \{x : f_\theta(x) > 0\}$. Plus précisément,

Corollaire 3.11. *Sous (H_d) et (H_s) et $u \in \mathbb{R}$ fixé, on a (quand $n \to \infty$)*

$$\int_{S_\theta} (\xi_{\theta+\frac{u}{\sqrt{n}}} - \xi_\theta)^2 d\mu = \int_{S_\theta} (\xi'_\theta \cdot \frac{u}{\sqrt{n}})^2 d\mu + o(\frac{1}{n}). \tag{9}$$

$$\int_{S_\theta} \left|(\xi_{\theta+\frac{u}{\sqrt{n}}} - \xi_\theta)^2 - (\xi'_\theta \cdot \frac{u}{\sqrt{n}})^2\right| d\mu = o(\frac{1}{n}). \tag{10}$$

$$\mathbb{E}_\theta\left[\frac{\xi'_\theta}{\xi_\theta}(X_i)\right] = \int_{S_\theta} \xi'_\theta \cdot \xi_\theta d\mu = 0. \tag{11}$$

Posons

$$\eta_i(u) = \frac{\xi_{\theta+u/\sqrt{n}}}{\xi_\theta}(X_i) - 1$$

Cette quantité est bien définie sur S_θ, de plus :

Lemme 3.12. *Sous (H_d) et (H_s), $u \in \mathbb{R}$, $\varepsilon > 0$, on a (quand $n \to \infty$)*

$$\mathbb{E}_\theta\big[\eta_i^2(u)\big] - \frac{\mathbf{I}(\theta)}{4n}u^2 = o(\frac{1}{n}), \tag{12}$$

$$\mathbb{E}_\theta\left|\eta_i^2(u) - \big(\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\big)^2\right| = o(\frac{1}{n}), \tag{13}$$

$$\mathbb{E}_\theta\big[\eta_i(u)\big] + \frac{\mathbf{I}(\theta)}{8n}u^2 = o(\frac{1}{n}), \tag{14}$$

$$\mathbb{P}_\theta\big[|\eta_i(u)| > \varepsilon\big] = o(\frac{1}{n}). \tag{15}$$

Démonstration. (12) vient directement de (H_d) et (9), tandis que (13) résulte de (H_d) et (10). Pour avoir (14), il suffit de remarquer que, sous (H_s)

$$\begin{aligned}\mathbb{E}_\theta\big[\eta_i^2(u)\big] &= \int_{S_\theta} |\xi_{\theta+\frac{u}{\sqrt{n}}} - \xi_\theta|^2 d\mu \\ &= \int |\xi_{\theta+\frac{u}{\sqrt{n}}} - \xi_\theta|^2 d\mu + o(\frac{1}{n}) = -2\mathbb{E}_\theta\big[\eta_i(u)\big] + o(\frac{1}{n}).\end{aligned}$$

Pour (15), il suffit de traiter le cas $u \neq 0$:

$$\begin{aligned}\mathbb{P}_\theta\big[|\eta_i(u)| > \varepsilon\big] &\leqslant \mathbb{P}_\theta\Big[\big|\eta_i(u) - \frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\big| > \varepsilon/2\Big] \\ &\quad + \mathbb{P}_\theta\Big[\big|\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\big| > \varepsilon/2\Big] \\ &\leqslant (\frac{2}{\varepsilon})^2\mathbb{E}_\theta\Big[\big|\eta_i(u) - \frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\big|^2\Big] \\ &\quad + \mathbb{P}_\theta\Big[\big|\frac{\xi'_\theta}{\xi_\theta}(X_i)\big|^2 > (\frac{\varepsilon\sqrt{n}}{2u})^2\Big] \\ &= o(\frac{1}{n}) + o(\frac{1}{n}) = o(\frac{1}{n}).\end{aligned}$$

En effet, (H_d) dit directement que

$$\mathbb{E}_\theta\Big[\big|\eta_i(u) - \frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\big|^2\Big] = \int_{\{f_\theta>0\}} |\xi_{\theta+\frac{u}{\sqrt{n}}} - \xi_\theta - \xi'_\theta \cdot \frac{u}{\sqrt{n}}|^2 d\mu = o(\frac{1}{n}).$$

D'autre part, $\mathbb{E}_\theta[|\frac{\xi'_\theta}{\xi_\theta}(X_i)|^2] = \frac{\mathbf{I}(\theta)}{4} < \infty$ entraîne

$$\begin{aligned}\mathbb{P}_\theta\Big[\big|\frac{\xi'_\theta}{\xi_\theta}(X_i)\big|^2 > (\frac{\varepsilon\sqrt{n}}{2u})^2\Big] &= \mathbb{E}_\theta\Big[1_{\{|\frac{\xi'_\theta}{\xi_\theta}(X_i)|^2 > \frac{n\varepsilon^2}{4u^2}\}}\Big] \\ &\leqslant \mathbb{E}_\theta\Big[1_{\{|\frac{\xi'_\theta}{\xi_\theta}(X_i)|^2 > \frac{n\varepsilon^2}{4u^2}\}} \cdot \big|\frac{\xi'_\theta}{\xi_\theta}(X_i)\big|^2\Big] \cdot \frac{4u^2}{n\varepsilon^2} = o(\frac{1}{n}).\end{aligned}$$

D'où la conclusion. □

On arrive maintenant à une conclusion fondamentale pour LAN :

Théorème 3.13. *Si (H_d) et (H_s) sont réalisées, $u \in \mathbb{R}$, alors*

$$\sum_{i=1}^n \log\frac{\xi_{\theta+u/\sqrt{n}}}{\xi_\theta}(X_i) = \frac{u}{\sqrt{n}}\sum_{i=1}^n \frac{\xi'_\theta}{\xi_\theta}(X_i) - \frac{u^2}{4}\mathbf{I}(\theta) + \Gamma_{n,\theta}(u), \quad \Gamma_{n,\theta}(u) \xrightarrow[n\to\infty]{\mathbb{P}_\theta} 0.$$

Démonstration.

$$\sum_{i=1}^{n} \log \frac{\xi_{\theta+u/\sqrt{n}}}{\xi_\theta}(X_i) = \sum_{i=1}^{n} \log\big(1+\eta_i(u)\big).$$

Par la formule de Taylor, $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3(1+\alpha x)^3}, \alpha \in]0,1[$, par suite

$$\sum_{i=1}^{n} \log\big(1+\eta_i(u)\big) = \sum_{i=1}^{n} \bigg(\eta_i(u) - \frac{1}{2}\cdot \eta_i^2(u) + \frac{1}{3(1+\alpha_i(u)\eta_i(u))^3}\cdot \eta_i^3(u)\bigg).$$

Soit $\varepsilon > 0$ assez petit, on peut alors avoir $\frac{1}{3(1+\alpha_i(u)\eta_i(u))^3} < \frac{1}{2}$ sur les événements $A_n(u) = \{\max_{1\leqslant i\leqslant n} |\eta_i(u)| \leqslant \varepsilon\}$. Par suite, il suffit de prouver

$$\mathbb{P}_\theta\big[A_n^c(u)\big] \xrightarrow[n\to\infty]{} 0, \tag{16}$$

$$\mathbb{P}_\theta\bigg[\bigg|\sum_{i=1}^{n}\eta_i(u) - \sum_{i=1}^{n}\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i) + \frac{\mathbf{I}(\theta)}{8}u^2\bigg| > \varepsilon\bigg] \xrightarrow[n\to\infty]{} 0, \tag{17}$$

$$\mathbb{P}_\theta\bigg[\bigg|\bigg(\sum_{i=1}^{n}\eta_i^2(u)\bigg) - \frac{\mathbf{I}(\theta)}{4}u^2\bigg| > \varepsilon\bigg] \xrightarrow[n\to\infty]{} 0, \tag{18}$$

$$\mathbb{P}_\theta\bigg[\sum_{i=1}^{n}\big|\eta_i^3(u)\big| > \varepsilon\bigg] \xrightarrow[n\to\infty]{} 0. \tag{19}$$

Pour (16), grâce à (15), on a

$$\mathbb{P}_\theta\big[A_n^c(u)\big] \leqslant \sum_{i=1}^{n}\mathbb{P}_\theta\big[|\eta_i(u)| > \varepsilon\big] = n\cdot o(\frac{1}{n}) \xrightarrow[n\to\infty]{} 0.$$

En vue de (13), on a

$$\begin{aligned}&\mathbb{P}_\theta\bigg[\bigg|\sum_{i=1}^{n}\eta_i^2(u) - \sum_{i=1}^{n}\big(\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\big)^2\bigg| > \varepsilon\bigg]\\ &\leqslant \frac{n}{\varepsilon}\cdot\mathbb{E}_\theta\bigg|\eta_i^2(u) - \big(\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\big)^2\bigg| = O(n)\cdot o(\frac{1}{n}) \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

Par suite, (18) vient directement du fait que (loi des grands nombres)

$$\sum_{i=1}^{n}\big(\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\big)^2 = \frac{u^2}{n}\sum_{i=1}^{n}\big(\frac{\xi'_\theta}{\xi_\theta}(X_i)\big)^2 \xrightarrow[n\to\infty]{\mathbb{P}_\theta} \frac{\mathbf{I}(\theta)}{4}u^2.$$

Pour avoir (17), grâce à (14) et (11), il suffit de regarder

$$\begin{aligned}
&\mathbb{P}_\theta\Big[\Big|\sum_{i=1}^n\Big(\eta_i(u)-\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)-\mathbb{E}_\theta\big[\eta_i(u)\big]\Big)\Big|>\varepsilon\Big]\\
&\quad\leqslant\frac{1}{\varepsilon^2}\mathbb{E}_\theta\Big[\Big|\sum_{i=1}^n\Big(\eta_i(u)-\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)-\mathbb{E}_\theta\big[\eta_i(u)\big]\Big)\Big|^2\Big]=\frac{n}{\varepsilon^2}\mathbb{E}_\theta\Big[\Big|\eta_i(u)-\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)-\mathbb{E}_\theta\big[\eta_i(u)\big]\Big|^2\Big]\\
&\quad=\frac{n}{\varepsilon^2}\mathbb{V}_\theta\Big[\eta_i(u)-\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\Big]=\frac{n}{\varepsilon^2}\Big(\mathbb{E}_\theta\Big[\Big|\eta_i(u)-\frac{u}{\sqrt{n}}\frac{\xi'_\theta}{\xi_\theta}(X_i)\Big|^2\Big]-\mathbb{E}^2_\theta\big[\eta_i(u)\big]\Big)\\
&\quad=O(n)\cdot\big(o(\frac{1}{n})+o(\frac{1}{n})\big)=o(1)\xrightarrow[n\to\infty]{}0.
\end{aligned}$$

Donc (17) est bien vérifiée. Il reste à prouver (19), qui vient directement de (16) et (18), en effet

$$\begin{aligned}
\mathbb{P}_\theta\Big[\sum_{i=1}^n\big|\eta_i^3(u)\big|>\varepsilon\Big]\leqslant&\mathbb{P}_\theta\Big[\max_{1\leqslant i\leqslant n}|\eta_i(u)|>\frac{\varepsilon}{1+\frac{\mathbf{I}(\theta)}{4}u^2}\Big]\\
&+\mathbb{P}_\theta\Big[\sum_{i=1}^n\big|\eta_i^2(u)\big|>1+\frac{\mathbf{I}(\theta)}{4}u^2\Big]\xrightarrow[n\to\infty]{}0.
\end{aligned}$$

□

Rappelons que $\mathbf{I}(\theta)=4\left\|\xi'_\theta\right\|^2$ et que (H_d) et (H_s) sont satisfaites dans le modèle régulier. On a alors immédiatement une conclusion équivalente à celle dans Théorème 3.10 :

Corollaire 3.14. *Dans le modèle régulier, la fonction de rapport de log-vraisemblance au point θ où $\mathbf{I}(\theta)>0$ s'écrit sous la forme suivante :*

$$\sum_{i=1}^n\log\frac{\xi_{\theta+u/\sqrt{n}}}{\xi_\theta}(X_i)=\frac{u}{\sqrt{n}}\sum_{i=1}^n\frac{\xi'_\theta}{\xi_\theta}(X_i)-u^2\left\|\xi'_\theta\right\|^2+\Gamma_{n,\theta}(u)$$

où

$$\Gamma_{n,\theta}(u)\xrightarrow[n\to\infty]{\mathbb{P}_\theta}0,\qquad\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\xi'_\theta}{\xi_\theta}(X_i)\xrightarrow[n\to\infty]{\mathcal{L}}N(0,\left\|\xi'_\theta\right\|^2).$$

Dans le modèle régulier, on peut renforcer notre conclusion. On admet ici une version uniforme de LAN [6], qui nous servira dans la partie suivante :

Corollaire 3.15. *Dans le modèle régulier, pour tout $T>0$, la fonction de rapport de log-vraisemblance au point θ où $\mathbf{I}(\theta)>0$ s'écrit sous la forme suivante :*

$$\sum_{i=1}^n\log\frac{\xi_{\theta+u/\sqrt{n}}}{\xi_\theta}(X_i)=\frac{u}{\sqrt{n}}\sum_{i=1}^n\frac{\xi'_\theta}{\xi_\theta}(X_i)-u^2\left\|\xi'_\theta\right\|^2+\Gamma_{n,\theta}(u),$$

où

$$\sup_{|u|\leqslant T}|\Gamma_{n,\theta}(u)|\xrightarrow[n\to\infty]{\mathbb{P}_\theta}0,\qquad\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\xi'_\theta}{\xi_\theta}(X_i)\xrightarrow[n\to\infty]{\mathcal{L}}N(0,\left\|\xi'_\theta\right\|^2).$$

3.3 Consistance et normalité asymptotique

On a bien étudié LAN dans le cadre général et dans le modèle régulier. Dans cette partie, on va étudier la relation entre la consistance et la normalité asymptotique pour les estimateurs du maximum de vraisemblance. On trouvera les conditions pour que l'existence (au sens asymptotique) et la consistance des estimateurs du maximum de vraisemblance entraînent la normalité asymptotique dans le modèle régulier.

On va s'intéresser à un point $\theta \in \Theta$ qui satisfait $\mathbf{I}(\theta) > 0$. Dans toute la suite, on va fixer un tel θ comme le vrai paramètre. Le rapport de vraisemblance défini sur $\Theta - \theta$ s'écrit :

$$\mathbf{Z}_n(u) = \prod_{i=1}^{n} \frac{f_{\theta+u}(X_i)}{f_\theta(X_i)} = \Big(\frac{\zeta_n(\theta+u)}{\zeta_n(\theta)}\Big)^2, \quad \text{où} \quad \zeta_n(\theta) = \prod_{i=1}^{n} \xi_\theta(X_i).$$

Posons $p_n(u) = \mathbf{Z}_n^{3/4}(u)$, alors

Lemme 3.16. *Dans le modèle régulier, on a*

$$\mathbb{E}_\theta\big[\mathbf{Z}_n^{1/2}(u)\big] = \rho^n(\theta+u,\theta), \quad \mathbb{E}_\theta\big[p_n(u)\big] \leqslant \rho^{n/2}(\theta+u,\theta),$$
$$p_n'(u) = \frac{3}{2}\frac{\zeta_n'(\theta+u)}{\zeta_n^{1/2}(\theta)}\mathbf{Z}_n^{1/4}(u), \quad \mathbb{E}_\theta\big[|p_n'(u)|\big] \leqslant \frac{3}{4}\sqrt{n\mathbf{I}(\theta+u)\rho^n(\theta+u,\theta)}.$$

Démonstration. Par la définition de l'affinité de Hellinger, vue que les variables aléatoires $X_1, \ldots, X_n$ sont i.i.d, on a immédiatement que

$$\mathbb{E}_\theta\big[\mathbf{Z}_n(u)\big] \leqslant 1, \quad \mathbb{E}_\theta\big[\mathbf{Z}_n^{1/2}(u)\big] = \rho^n(\theta+u,\theta).$$

Par l'Inégalité de Cauchy-Schwarz,

$$\mathbb{E}_\theta\big[p_n(u)\big] \leqslant \Big(\mathbb{E}_\theta\big[\mathbf{Z}_n(u)\big]\cdot \mathbb{E}_\theta\big[\mathbf{Z}_n^{1/2}(u)\big]\Big)^{1/2} \leqslant \rho^{n/2}(\theta+u,\theta),$$

$$p_n'(u) = \frac{\partial}{\partial u}\mathbf{Z}_n^{3/4}(u) = \frac{\partial}{\partial u}\Big(\frac{\zeta_n(\theta+u)}{\zeta_n(\theta)}\Big)^{3/2} = \frac{3}{2}\Big(\frac{\zeta_n(\theta+u)}{\zeta_n(\theta)}\Big)^{1/2}\frac{\zeta_n'(\theta+u)}{\zeta_n(\theta)} = \frac{3}{2}\frac{\zeta_n'(\theta+u)}{\zeta_n^{1/2}(\theta)}\mathbf{Z}_n^{1/4}(u),$$

$$\begin{aligned}
\mathbb{E}_\theta\big[|p_n'(u)|\big] &\leqslant \frac{3}{2}\mathbb{E}_\theta\Big[\Big|\frac{3}{2}\frac{\zeta_n'(\theta+u)}{\zeta_n^{1/2}(\theta)}\mathbf{Z}_n^{1/4}(u)\Big|\Big] \\
&\leqslant \frac{3}{2}\Big(\mathbb{E}_\theta\Big[\frac{\big|\zeta_n'(\theta+u)\big|^2}{\zeta_n(\theta)}\Big]\cdot \mathbb{E}_\theta\big[\mathbf{Z}_n^{1/2}(u)\big]\Big)^{1/2} \\
&= \frac{3}{4}\sqrt{n\mathbf{I}(\theta+u)\rho^n(\theta+u,\theta)}.
\end{aligned}$$

□

On a déjà étudié la consistance des estimateurs du maximum de vraisemblance (où l'on suppose leur existence). Dans le modèle régulier, dès qu'ils sont consistants, ils convergent à la vitesse au plus $n^{-1/2}$:

Théorème 3.17. *Dans le modèle régulier, si l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ converge (en probabilité) vers θ avec $\mathbf{I}(\theta) > 0$, alors pour tout $T > 0$,*

$$\limsup_{n\to\infty} \mathbb{P}_\theta\big[|\hat{\theta}_n - \theta| \geqslant T/\sqrt{n}\big] \leqslant 2(1+3\sqrt{\pi})\exp\big[-T^2\mathbf{I}(\theta)/16\big]. \tag{20}$$

Démonstration. D'après la proposition 3.6, $\mathbf{I}(t)$ est continue sur Θ. Θ étant un intervalle ouvert avec $\mathbf{I}(\theta) > 0$, on peut alors choisir un $\varepsilon > 0$ assez petit tel que $\mathbf{I}(t) > 0$ sur $J_\varepsilon = [\theta-\varepsilon, \theta+\varepsilon]$; on peut aussi trouver un $g_\varepsilon > 0$ grâce à la même proposition telle que $g_\varepsilon u^2 \leqslant \mathbf{h}^2(\theta+u,\theta)$ pour les $u \in [-\varepsilon, \varepsilon]$.

Pour $T > 0$ fixé, on va étudier la borne asymptotique de $\mathbb{P}_\theta\big[|\hat{\theta}_n - \theta| \geqslant T/\sqrt{n}\big]$. Posons $\delta_n = T/\sqrt{n}$, quand n est assez grand pour que $\delta_n < \varepsilon$, on a

$$\mathbb{P}_\theta\big[|\hat{\theta}_n - \theta| \geqslant \delta_n\big] \leqslant \mathbb{P}_\theta\big[|\hat{\theta}_n - \theta| > \varepsilon\big] + \mathbb{P}_\theta\big[|\hat{\theta}_n - \theta| \in [\delta_n, \varepsilon]\big].$$

Comme $\hat{\theta}_n$ est consistant en θ, $\mathbb{P}_\theta\big[|\hat{\theta}_n - \theta| > \varepsilon\big] \xrightarrow[n\to\infty]{} 0$. D'autre part,

$$\mathbb{P}_\theta\big[|\hat{\theta}_n - \theta| \in [\delta_n, \varepsilon]\big] \leqslant \mathbb{P}_\theta\Big[\sup_{\delta_n \leqslant |u| \leqslant \varepsilon} p_n(u) \geqslant 1\Big] \leqslant \mathbb{E}_\theta\Big[\sup_{\delta_n \leqslant |u| \leqslant \varepsilon} p_n(u)\Big] \leqslant H_+(\delta_n) + H_-(\delta_n),$$

où

$$\begin{aligned} H_+(\delta_n) &= \mathbb{E}_\theta\Big[\sup_{\delta_n \leqslant u \leqslant \varepsilon} p_n(u)\Big] \qquad\qquad H_-(\delta_n) = \mathbb{E}_\theta\Big[\sup_{-\varepsilon \leqslant u \leqslant -\delta_n} p_n(u)\Big] \\ &= \mathbb{E}_\theta\Big[\sup_{\delta_n \leqslant u \leqslant \varepsilon} \Big[p_n(\delta_n) + \int_{\delta_n}^{u} p_n'(t)dt\Big]\Big] \\ &\leqslant \mathbb{E}_\theta\big[p_n(\delta_n)\big] + \int_{\delta_n}^{\varepsilon} \mathbb{E}_\theta\big[|p_n'(t)|\big]dt \\ &\leqslant \rho^{n/2}(\theta+\delta_n, \theta) + \int_{\delta_n}^{\varepsilon} \frac{3}{4}\sqrt{n\mathbf{I}(\theta+t)\rho^n(\theta+t,\theta)}dt. \end{aligned}$$

En vertu que $1 - x \leq \exp(-x)$, on a

$$\rho(\theta+u,\theta) = 1 - \mathbf{h}^2(\theta+u,\theta) \leqslant \exp[-\mathbf{h}^2(\theta+u,\theta)],$$

On a alors

$$\begin{aligned} \rho^{n/2}(\theta+\delta_n,\theta) &\leqslant \exp\Big[-\frac{n}{2}\cdot\mathbf{h}^2(\theta+\delta_n,\theta)\Big] \\ &= \exp\Big[-\frac{T}{2}\cdot\frac{\mathbf{h}^2(\theta+\delta_n,\theta)}{\delta_n^2}\Big] \xrightarrow[n\to\infty]{} \exp\big[-T^2\mathbf{I}(\theta)/16\big]. \end{aligned}$$

Posons $\mathbf{I}_\varepsilon = \sup_{|u|\leqslant\varepsilon} \mathbf{I}(\theta+u)$. Vue que $g_\varepsilon u^2 \leqslant \mathbf{h}^2(\theta+u,\theta)$ pour les $u \in [-\varepsilon,\varepsilon]$, on a [1],

$$\begin{aligned}\int_{\delta_n}^{\varepsilon} \frac{3}{4}\sqrt{n\mathbf{I}(\theta+t)\rho^n(\theta+t,\theta)}dt &\leqslant \frac{3\sqrt{\mathbf{I}_\varepsilon}}{4}\sqrt{n}\int_{\delta_n}^{\varepsilon}\exp[-\frac{n}{2}\cdot\mathbf{h}^2(\theta+t,\theta)]dt \\ &\leqslant \frac{3\sqrt{\mathbf{I}_\varepsilon}}{4}\sqrt{n}\int_{\delta_n}^{\varepsilon}\exp[-\frac{n}{2}\cdot g_\varepsilon t^2]dt = \frac{3\sqrt{\mathbf{I}_\varepsilon}}{4\sqrt{g_\varepsilon}}\int_{\sqrt{g_\varepsilon}T}^{\sqrt{g_\varepsilon n}\varepsilon}\exp[-\frac{x^2}{2}]dx \\ &\leqslant \frac{3\sqrt{\mathbf{I}_\varepsilon}}{4\sqrt{g_\varepsilon}}\int_{\sqrt{g_\varepsilon}T}^{\infty}\exp[-\frac{x^2}{2}]dx \leqslant \frac{3\sqrt{\mathbf{I}_\varepsilon}}{4\sqrt{g_\varepsilon}}\sqrt{2\pi}\exp[-\frac{(T\sqrt{g_\varepsilon})^2}{2}].\end{aligned}$$

On en déduit alors que

$$\limsup_{n\to\infty}\int_{\delta_n}^{\varepsilon}\frac{3}{4}\sqrt{n\mathbf{I}(\theta+t)\rho^n(\theta+t,\theta)}dt \leqslant \frac{3\sqrt{\mathbf{I}_\varepsilon}}{4\sqrt{g_\varepsilon}}\sqrt{2\pi}\exp[-\frac{(T\sqrt{g_\varepsilon})^2}{2}].$$

Même raisonnement pour $H_-(\delta_n)$, on obtient

$$\limsup_{n\to\infty} H_+(\delta_n) \leqslant \exp\big[-T^2\mathbf{I}(\theta)/16\big] + \frac{3\sqrt{\mathbf{I}_\varepsilon}}{4\sqrt{g_\varepsilon}}\sqrt{2\pi}\exp[-\frac{(T\sqrt{g_\varepsilon})^2}{2}]$$

$$\limsup_{n\to\infty} H_-(\delta_n) \leqslant \exp\big[-T^2\mathbf{I}(\theta)/16\big] + \frac{3\sqrt{\mathbf{I}_\varepsilon}}{4\sqrt{g_\varepsilon}}\sqrt{2\pi}\exp[-\frac{(T\sqrt{g_\varepsilon})^2}{2}].$$

Pour conclure, on fait ε tendre vers 0, vue que

$$\mathbf{I}_\varepsilon = \sup_{|u|\leqslant\varepsilon}\mathbf{I}(\theta+u) \xrightarrow[\varepsilon\to 0]{} \mathbf{I}(\theta)$$

$$g_\varepsilon = \inf_{\substack{|u|\leqslant\varepsilon \\ u\neq 0}}\frac{\mathbf{h}^2(\theta+u,\theta)}{u^2} \xrightarrow[\varepsilon\to 0]{} \frac{\mathbf{I}(\theta)}{8}.$$

On a alors

$$\exp\big[-T^2\mathbf{I}(\theta)/16\big] + \frac{3\sqrt{\mathbf{I}_\varepsilon}}{4\sqrt{g_\varepsilon}}\sqrt{2\pi}\exp[-\frac{(T\sqrt{g_\varepsilon})^2}{2}] \xrightarrow[\varepsilon\to 0]{} (1+3\sqrt{\pi})\exp\big[-T^2\mathbf{I}(\theta)/16\big]$$

D'où la conclusion. □

1. $\forall z > 0$,

$$\int_z^\infty e^{-u^2/2}du < \sqrt{2\pi}\cdot e^{-z^2/2}.$$

D'après le Corollaire 3.15, le modèle régulier possède la propriété LAN uniforme aux points θ où $\mathbf{I}(\theta) > 0$, qui implique que pour tout $T > 0$,

$$\sum_{i=1}^{n} \log \frac{\xi_{\theta+u/\sqrt{n}}}{\xi_\theta}(X_i) = \frac{u}{\sqrt{n}} \sum_{i=1}^{n} \frac{\xi'_\theta}{\xi_\theta}(X_i) - u^2 \left\|\xi'_\theta\right\|^2 + \Gamma_{n,\theta}(u), \tag{21}$$

où

$$\sup_{|u|\leqslant T} |\Gamma_{n,\theta}(u)| \xrightarrow[n\to\infty]{\mathbb{P}_\theta} 0, \qquad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\xi'_\theta}{\xi_\theta}(X_i) \xrightarrow[n\to\infty]{\mathcal{L}} N(0, \left\|\xi'_\theta\right\|^2).$$

On arrive à la conclusion qui entraîne la normalité asymptotique :

Théorème 3.18. *Dans le modèle régulier, si l'estimateur du maximum de vraisemblance $\hat\theta_n$ converge (en probabilité) vers θ avec $\mathbf{I}(\theta) > 0$, alors il est asymptotiquement normal en θ satisfaisant $\sqrt{n}(\hat\theta_n - \theta) \xrightarrow[n\to\infty]{\mathcal{L}} N(0, \mathbf{I}^{-1}(\theta))$.*

Démonstration. Posons

$$F_n(u) = \sum_{i=1}^{n} \log \frac{\xi_{\theta+u/\sqrt{n}}}{\xi_\theta}(X_i), \quad \hat u_n = \underset{u\in\sqrt{n}(\Theta-\theta)}{\arg\max}\ F_n(u),$$

$$G_n(u) = \frac{u}{\sqrt{n}} \sum_{i=1}^{n} \frac{\xi'_\theta}{\xi_\theta}(X_i) - u^2 \left\|\xi'_\theta\right\|^2, \quad \tilde u_n = \underset{u\in\sqrt{n}(\Theta-\theta)}{\arg\max}\ G_n(u).$$

Pour tout $\varepsilon > 0$ et $T > 0$, on regard les événements suivants :

$$A_n = \{\sup_{|u|\leqslant T} |F_n(u) - G_n(u)| \leqslant \varepsilon\}, \quad B_n = \{|\tilde u_n| \leqslant T\}, \quad C_n = \{|\hat u_n| \leqslant T\}.$$

Or

$$\hat u_n = (\hat\theta_n - \theta)\sqrt{n}, \quad \tilde u_n = \frac{1}{2\left\|\xi'_\theta\right\|^2} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\xi'_\theta}{\xi_\theta}(X_i)$$

D'après (20) et (21),

$$\sup_{|u|\leqslant T} \Gamma_{n,\theta}(u) \xrightarrow[n\to\infty]{\mathbb{P}_\theta} 0 \quad \tilde u_n \xrightarrow[n\to\infty]{\mathcal{L}} N(0, \mathbf{I}^{-1}(\theta)),$$

$$\limsup_{n\to\infty} \mathbb{P}_\theta\big[|\hat u_n| \geqslant T\big] \leqslant 2(1 + 3\sqrt{\pi}) \exp\big[-T^2\mathbf{I}(\theta)/16\big].$$

Alors quand n et T sont assez grands, on peut avoir

$$\mathbb{P}_\theta[A_n^c] \leqslant \varepsilon, \quad \mathbb{P}_\theta[B_n^c] \leqslant \varepsilon, \quad \mathbb{P}_\theta[C_n^c] \leqslant \varepsilon.$$

Il reste à considérer l'événement $A_n \cap B_n \cap C_n$, sur lequel on a

$$\left|\hat u_n - \tilde u_n\right| \leqslant \sqrt{\frac{8\varepsilon}{\mathbf{I}(\theta)}}. \tag{22}$$

En effet, sur $A_n \cap B_n \cap C_n$ et pour tout u t.q. $|u| \leqslant T$, on a

$$G_n(u) - \varepsilon \leqslant F_n(u) \leqslant G_n(u) + \varepsilon.$$

Par suite

$$G_n(\tilde{u}_n) - \varepsilon \leqslant F_n(\tilde{u}_n) \leqslant F_n(\hat{u}_n) \leqslant G_n(\hat{u}_n) + \varepsilon,$$
$$\Longrightarrow G_n(\hat{u}_n) - G_n(\tilde{u}_n) + 2\varepsilon \geqslant 0.$$

Alors (22) est bien vérifiée [2], et on peut donc dire

$$\hat{u}_n - \tilde{u}_n \xrightarrow[n\to\infty]{\mathbb{P}_\theta} 0.$$

Par suite

$$\sqrt{n}(\hat{\theta}_n - \theta) = \hat{u}_n \xrightarrow[n\to\infty]{\mathcal{L}} N(0, \mathbf{I}^{-1}(\theta)).$$

D'où la conclusion. □

2. En général, pour une fonction $g(x) = -ax^2 + bx$ avec $a > 0$, on a

$$\tilde{x} = \arg\max_{x\in\mathbb{R}} g(x) = \frac{b}{2a}, \quad g(\tilde{x}) = \frac{b^2}{4a}.$$

Si $g(x) - g(\tilde{x}) + 2\varepsilon \geqslant 0$ alors

$$|x - \tilde{x}| \leqslant \frac{1}{2}|x_1 - x_2| = \sqrt{\frac{2\varepsilon}{a}},$$

où x_1, x_2 sont les deux racines d'équation $g(x) - g(\tilde{x}) + 2\varepsilon = 0$ et il suffit de remarquer que $|x_1 - x_2| = \sqrt{(x_1 + x_2)^2 - 4x_1x_2}$ pour conclure.

Références

[1] I.A. IBRAGIMOV, Statistical Estimation, Springer-Verlag, 1981.

[2] A.A. BOROVKOV, Statistique Mathématique, Moscou, 1987.

[3] A.W. VAN DER VAART, Asymtotic Statistics, Cambridge University Press, 1998.

[4] L. LECAM, On the assumptions used to prove asymptotic normality of maximum likelihood estimates, The Annals of Mathematical Statistics, 1970.

[5] J. HAJEK, Local asymptotic minimax and admissibility in estimation, Proc. 6th Berkeley Symp. on Math. Stat and Prob, 1972.

[6] L. BIRGÉ, Statistique Mathématique, Université Paris VI, 2009.

[7] N. LERNER, Lecture Notes on Real Analysis, Université Paris VI, 2009.