Problèmes inverses — Estimation par maximum de vraisemblance

Corrigé des exercices 6, 7 et 8

Exercice 6

1. On vérifie facilement que

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2.$$

On peut donc appliquer la loi des grands nombres. On en déduit que

$$\hat{\sigma}_n^2 \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbb{E}_\theta(X_1^2) - (\mathbb{E}_\theta X_1)^2 = \theta.$$

2. On a

$$\sqrt{n}(\hat{\sigma}_n^2 - \theta) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \mathbb{E}_\theta(X_1^2)\right) - \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}_\theta X_1\right)(\bar{X}_n + \mathbb{E}_\theta X_1). \tag{1}$$

Les deux termes relèvent du TCL. Pour étudier leur combinaison linéaire, il est nécessaire d'écrire un TCL vectoriel :

$$\sqrt{n}\left[\left(\frac{1}{n}\sum_{i=1}^n X_i\,,\,\frac{1}{n}\sum_{i=1}^n X_i^2\right) - \left(\mathbb{E}_\theta X_1, \mathbb{E}_\theta(X_1^2)\right)\right] \xrightarrow[n\to+\infty]{\text{en loi sous } \mathbb{P}_\theta} (G_1, G_2),$$

où $(G_1, G_2) \sim \mathcal{N}_2(0, \Sigma)$, où Σ est la matrice de variance-covariance du vecteur (X_1, X_1^2), c'est-à-dire

$$\Sigma = \begin{pmatrix} \theta & \mathrm{Cov}_\theta(X_1, X_1^2) \\ \mathrm{Cov}_\theta(X_1, X_1^2) & \mathrm{Var}_\theta(X_1^2) \end{pmatrix}.$$

Par ailleurs, la loi des grands nombres permet de traiter le terme $\bar{X}_n + \mathbb{E}_\theta X_1$, qui converge presque sûrement vers $2\mathbb{E}_\theta X_1$. En inserrant les résultats précédents dans l'équation (1) et en utilisant le lemme de Slutsky, on obtient

$$\sqrt{n}(\hat{\sigma}_n^2 - \theta) \xrightarrow[n\to+\infty]{\text{en loi sous } \mathbb{P}_\theta} G_1 - 2G_2\mathbb{E}_\theta X_1 \sim \mathcal{N}(0, \alpha^2),$$

où

$$\alpha^2 = \mathrm{Var}_\theta(G_1 - 2G_2\mathbb{E}_\theta X_1) = \theta - 4\mathbb{E}_\theta X_1 \mathrm{Cov}_\theta(X_1, X_1^2) + 4\left(\mathbb{E}_\theta X_1\right)^2 \mathrm{Var}_\theta(X_1^2).$$

Donc la vitesse de l'estimateur $\hat{\sigma}_n^2$ est $\sqrt{n}$ et sa loi limite est $\mathcal{N}(0, \alpha^2)$.

Exercice 7

1. On a $\mathbb{E}_\theta X_1 = \int_{\mathbb{R}} x\mathcal{B}(\theta)(dx) = \theta$. La méthode des moments consiste à remplacer la loi de X_1, $\mathcal{B}(\theta)$, par la mesure empirique des observations, $\hat{\mathbb{Q}}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. On obtient l'estimateur

$$\hat{\theta}(X_1, \ldots, X_n) = \int_{\mathbb{R}} x\hat{\mathbb{Q}}_n(dx) = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}_n.$$

2. Pour tout $(x_1, \ldots, x_n) \in \{0,1\}^n$, on a

$$L_n(x_1, \ldots, x_n; \theta) = \mathbb{P}_\theta((X_1, \ldots, X_n) = (x_1, \ldots, x_n)) = \prod_{i=1}^n (\theta \mathbb{1}_{x_i=1} + (1-\theta)\mathbb{1}_{x_i=0}) = \theta^{n\bar{x}_n}(1-\theta)^{n(1-\bar{x}_n)},$$

où $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$. On cherche le maximum de cette fonction. On calcule

$$\frac{d}{d\theta}(\ln L_n(x_1, \ldots, x_n; \theta)) = \frac{n\bar{x}_n}{\theta} - \frac{n(1-\bar{x}_n)}{1-\theta} = \frac{n\bar{x}_n - n\theta}{\theta(1-\theta)}.$$

Cette dérivée s'annule si et seulement si $\theta = \bar{x}_n$, elle est strictement négative pour $\theta > \bar{x}_n$ et strictement positive si $\theta < \bar{x}_n$. Donc $\theta \mapsto L_n(x_1, \ldots, x_n; \theta)$ atteint son maximum pour $\theta = \bar{x}_n = \hat{\theta}(x_1, \ldots, x_n)$.

Exercice 8

1. Le fait que $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 \mathrm{Id})$ signifie que les ε_i sont i.i.d. et centrés, donc l'aléa du modèle (provenant uniquement des ε_i, puisque les R_{ij} sont connus) est supposé indépendant des régresseurs.

Soit le modèle $Y = R\theta + \varepsilon$, où $\theta \in \mathbb{R}^k$ et $\varepsilon \sim \mathcal{N}_n((m, \ldots, m), \sigma^2 \mathrm{Id})$. On se ramène au modèle de X en rajoutant un régresseur $R_{k+1} = (1, \ldots, 1)^T$ et une coordonnée supplémentaire à θ qui vaut m.

2. Dans ce modèle, la matrice R est connue, le paramètre θ est à estimer, et X est le vecteur des observations. Le vecteur X est un vecteur gaussien, dont les moyenne et variance sont faciles à calculer. On obtient le modèle statistique

$$\left(\mathbb{R}^n, \left\{\mathcal{N}_n(R\theta, \sigma^2 \mathrm{Id})\right\}_{\theta \in \mathbb{R}^k}\right).$$

3. a) Montrons d'abord que $\Pi_E = R(R^T R)^{-1} R^T$. Pour cette application linéaire, si $x \in E$, alors il existe $y \in \mathbb{R}^k$ tel que $x = Ry$ et

$$\Pi_E x = \Pi_E R y = R(R^T R)^{-1} R^T R y = Ry = x.$$

De plus, pour tout $x \in \mathbb{R}^n$, $x - \Pi_E x$ est orthogonal à E, car pour tout $y \in \mathbb{R}^k$,

$$\langle Ry, x - \Pi_E x \rangle = y^T R^T x - y^T R^T R (R^T R)^{-1} R^T x = y^T R^T x - y^T R^T x = 0.$$

Ceci suffit à caractériser l'application $R(R^T R)^{-1} R$ comme la projection orthogonale sur E.

Ceci nous conduit à définir $\hat{\theta} = (R^T R)^{-1} R^T X$ de façon à ce que $R\hat{\theta} = \Pi_E X$. C'est un estimateur sans biais de θ car

$$\mathbb{E}_\theta \hat{\theta} = (R^T R)^{-1} R^T \, \mathbb{E}_\theta X = (R^T R)^{-1} R^T R \theta = \theta.$$

b) En projetant la relation $X = R\theta + \varepsilon$, on a $\Pi_E X = R\theta + \Pi_E \varepsilon$, donc $X - \Pi_E X = \varepsilon - \Pi_E \varepsilon = \Pi_{E^\perp} \varepsilon$. Soit $\{u_1, \ldots, u_n\}$ une base orthonormée de $\mathbb{R}^n$ telle que $\{u_1, \ldots, u_{n-k}\}$ est une base orthonormée de $E^\perp$. Soit P la matrice de changement de base correspondante. Alors

$$P\varepsilon = (\varepsilon'_1, \ldots, \varepsilon'_n) \sim \mathcal{N}_n(0, \sigma^2 P^T \mathrm{Id} P) = \mathcal{N}_n(0, \sigma^2 \mathrm{Id}),$$

ce qui signifie que les ε'_i sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$.

On en déduit que

$$\|\Pi_{E^\perp}\varepsilon\|^2 = (\varepsilon'_1)^2 + \ldots + (\varepsilon'_{n-k})^2 = \sigma^2 Z,$$

où Z suit la loi du χ^2 à $n-k$ degrés de liberté. En particulier, $\mathbb{E}_\theta\|\Pi_{E^\perp}\varepsilon\|^2 = \sigma^2(n-k)$ et donc $\hat{\sigma}^2 = \frac{\|X-\Pi_E X\|^2}{n-k}$ est un estimateur sans biais de σ^2.

c) La densité du vecteur gaussien X en $(x_1, \ldots, x_n) \in \mathbb{R}^n$ est

$$L_n(x_1, \ldots, x_n; \theta, \sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|x - R\theta\|^2}{2\sigma^2}\right).$$

À $\sigma > 0$ fixé, cette densité est maximale en θ tel que $R\theta$ soit égal à la propjection orthogonale de x sur E, c'est-à-dire pour $\theta = \hat{\theta}$. Pour $\theta = \hat{\theta}$ fixé, la fonction définie par

$$\psi(\sigma) = \ln L_n(x_1, \ldots, x_n; \theta, \sigma) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{\|x - R\hat{\theta}\|^2}{2\sigma^2}$$

satisfait $\psi'(\sigma) = -\frac{n}{\sigma} + \frac{\|x-R\theta\|^2}{\sigma^3}$ et est donc maximale en $\sigma = \frac{\|x-R\hat{\theta}\|}{\sqrt{n}} = \frac{\|x-\Pi_E x\|}{\sqrt{n}} = \sqrt{\frac{n-k}{n}}\,\hat{\sigma}$.