

Statistiques - Résumé

October 25, 2023

THEVENET Louis

Table des matières

1. Estimation	1
1.1. Modèle statistique, estimateurs	1
1.2. Vraisemblance	1
1.3. Inégalité de Cramér Rao	1
1.4. Maximum de vraisemblance	2
1.5. Méthode des moments (construire un estimateur)	2
1.6. Estimation Bayésienne	3
2. Tests Statistiques	4
3. Au partiel	7

1. Estimation

Qualités d'un estimateur

1.1. Modèle statistique, estimateurs

Définition 1.1.1:

On note $\hat{\theta}(X_1, \dots, X_n)$, $\hat{\theta}_n$ ou $\hat{\theta}$ l'estimateur lié aux n VA *iid* X_1, \dots, X_n elles-mêmes liées aux n observations x_1, \dots, x_n

- Biais : $b_n(\theta) = E(\hat{\theta}_n) - \theta \in \mathbb{R}^p$
- Variance : $v_n(\theta) = E\left[(\hat{\theta}_n - E(\hat{\theta}_n))^2\right]$
- Matrice de covariance : $E\left[(\hat{\theta}_n - E(\hat{\theta}_n))(\hat{\theta}_n - E(\hat{\theta}_n))^T\right]$
- Erreur quadratique moyenne (MSE) : $e_n(\theta) = E\left[(\hat{\theta}_n - \theta)^2\right] = v_n(\theta) + b_n^2(\theta)$
- un estimateur $\hat{\theta}_n$ est convergent si $\lim_{n \rightarrow +\infty} b_n(\theta) = \lim_{n \rightarrow +\infty} v_n(\theta) = 0$

1.2. Vraisemblance

Définition 1.2.1: Vraisemblance

$$L(x_1, \dots, x_n; \theta) = \begin{cases} X_i \text{ VA discrète } P[X_1 = x_1, \dots, X_n = x_n; \theta] \\ X_i \text{ VA continue } p(x_1, \dots, x_n; \theta) \end{cases}$$

1.3. Inégalité de Cramér Rao

Théorème 1.3.1:

$$\text{Var}(\hat{\theta}_n) \geq \frac{[1 + b'_n(\theta)]^2}{(-E\left[\frac{\partial^2 \ln(L(X_1, \dots, X_n; \theta))}{\partial \theta^2}\right])} = \text{BCR}(\theta)$$

- *BCR* : Borne de Cramér-Rao
- $L(X_1, \dots, X_n; \theta)$: vraisemblance
- **Hypothèses** :
 1. log-vraisemblance deux fois dérivable
 2. support de la loi indépendant de θ

BCR d'un estimateur non biaisé : $\text{BCR} = \frac{1}{E\left[\frac{\partial^2 \ln(L(X_1, \dots, X_n; \theta))}{\partial \theta^2}\right]}$

Méthode 1.3.1: Efficacité

Un estimateur est **efficace** ssi $\text{Var}(\hat{\theta}_n) = \text{BCR}$

1.4. Maximum de vraisemblance

Définition 1.4.1: Maximum de vraisemblance

$$\hat{\theta}_{\text{MV}} = \arg \max_{\theta} L(X_1, \dots, X_n; \theta)$$

Méthode 1.4.1: Recherche de $\hat{\theta}_{\text{MV}}$

- On cherche θ tel que $\frac{\partial L(X_1, \dots, X_n; \theta)}{\partial \theta} = 0$ ou $\frac{\partial \ln L(X_1, \dots, X_n; \theta)}{\partial \theta} = 0$
- Puis éventuellement tableau de variations pour vérifier ou alors on étudie $\frac{\partial^2 \ln L(X_1, \dots, X_n; \hat{\theta}_{\text{MV}})}{\partial \theta^2} < 0$

1.5. Méthode des moments (construire un estimateur)

Définition 1.5.1: Estimateur des moments

$$\hat{\theta}_{\text{Mo}} = h(\hat{m}_1, \dots, \hat{m}_q) \text{ avec } \hat{m}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

Utile car en général le paramètre à estimer θ est lié **aux premiers moments** de la loi des X_i , i.e. $\theta = h(m_1, \dots, m_q)$ avec $m_k = E[X_i^k]$, $q \geq p$

1.6. Estimation Bayésienne

Définition 1.6.1: Estimation Bayésienne

On va estimer un paramètre inconnu $\theta \in \mathbb{R}^p$ à l'aide de la vraisemblance des X_1, \dots, X_n , et une loi à priori $p(\theta)$.

Pour cela on minimise une fonction de coût $c(\theta, \hat{\theta})$ qui représente l'erreur entre θ et $\hat{\theta}$.

Deux estimateurs principaux :

- MMSE : moyenne de la loi à posteriori

$$\hat{\theta}_{\text{MSEE}} = E[\theta \mid X_1, \dots, X_n]$$

- MAP : estimateur du maximum à posteriori de θ est définie par

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} p(\theta \mid x_1, \dots, x_n)$$

On appelle

- $p(\theta)$ loi à **priori** de θ
- $p(\theta \mid X_1, \dots, X_n)$ loi à **posteriori** de θ

Méthode 1.6.1: On peut utiliser $f(\theta \mid t_1, \dots, t_n) \propto f(t_1, \dots, t_n \mid \theta)p(\theta)$

Théorème 1.6.1: MMSE

L'estimateur MMSE minimise l'erreur quadratique moyenne (Mean Square Error, MSE)

On a

$$c(\theta, \hat{\theta}) = E[(\theta - \hat{\theta})^T (\theta - \hat{\theta})]$$

Théorème 1.6.2: MAP

L'estimateur MAP minimise la fonction de coût $E[c(\theta, \hat{\theta})]$ en moyenne (moyenne à posteriori) avec

$$c(\theta, \hat{\theta}) = \begin{cases} 1 & \text{si } \|\theta - \hat{\theta}\| > \Delta \\ 0 & \text{si } \|\theta - \hat{\theta}\| < \Delta \end{cases}$$

Avec Δ arbitrairement petit

Exemple : Estimation Bayésienne

Exemple :

- Vraisemblance : $X_i \sim \mathcal{N}(\theta, \sigma^2)$
- Loi à priori : $\theta \sim \mathcal{N}(\mu, \nu^2)$

Solution

- Loi à posteriori : $\theta \mid X_1, \dots, X_n \sim \mathcal{N}(m_p, \sigma_p^2)$
- Estimateurs : $\hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{MMSE}} = m_p = \bar{X} \left(\frac{n\nu^2}{n\mu^2 + \sigma^2} \right) + \mu \left(\frac{\sigma^2}{\sigma^2 + n\nu^2} \right)$

2. Tests Statistiques

Définition 2.1: Tests statistiques

- Observations :
 - X_1, \dots, X_n n VA i.i.d.
 - $L(X_1, \dots, X_n; \theta)$
- Hypothèses
 - Hypothèses simples : $H_0 : \theta = \theta_0, H_1 : \theta = \theta_1$
 - Hypothèse composites $H_0 : \theta = \theta_0, H_1 : \theta > \theta_0$

Construire une statistique de test $T(X_1, \dots, X_n)$ Règle de test : $\begin{cases} \text{si } T(x_1, \dots, x_n) \in \Delta: \text{rejet } H_0 \\ \text{sinon accepter } H_0 \end{cases}$

Δ : zone critique de test = zone de rejet de H_0

Définition 2.2:

- Risque de première espèce (fausse alarme) : $\alpha = \text{PFA} = P[\text{Rejeter } H_0 \mid H_0 \text{ vraie}]$
- Risque de seconde espèce (non détection) : $\beta = \text{PND} = P[\text{Rejeter } H_1 \mid H_1 \text{ vraie}]$
- Puissance du test (proba de détection) : $\pi = 1 - \beta$

Définition 2.3: p-valeur

La proba de fausse alarme la plus petite telle qu'on rejette le test, i.e. la plus petite valeur de α telle que H_0 est rejetée.

$$p(x) = \inf\{\alpha \in]0, 1[\mid x \in \mathcal{R}_\alpha\}$$

Théorème 2.1: Théorème de Neyman-Pearson

Test paramétrique à hypothèses simples : $H_0 : \theta = \theta_0, H_1 : \theta = \theta_1$

Rejet de H_0 si :

Cas X_i continues	Cas X_i discètes
$\frac{L(x_1, \dots, x_n \mid H_1)}{L(x_1, \dots, x_n \mid H_0)}$ $= \frac{p(x_1, \dots, x_n \mid \theta_1)}{p(x_1, \dots, x_n \mid \theta_0)} > S_\alpha$	$\frac{L(x_1, \dots, x_n \mid H_1)}{L(x_1, \dots, x_n \mid H_0)}$ $= \frac{P[X_1 = x_1, \dots, X_n = x_n \mid \theta_1]}{P[X_1 = x_1, \dots, X_n = x_n \mid \theta_0]} > S_\alpha$

Exemple : $X_i \sim \mathcal{P}(\lambda)$

Règle du test : si $T = \sum X_i > S_\alpha$ alors rejet H_0

\leadsto zone critique du test : $\Delta =]S_\alpha, +\infty[$

Détermination du seuil :

$$\alpha = P[\text{rejeter } H_0 \mid H_0 \text{ vraie}] = P[T = \sum X_i > S_\alpha \mid ?]$$

où ? représente la loi de T sous H_0 (on aurait $\lambda = \lambda_0$)

Loi de T ?

avec $T = \sum X_i$, $X_i \sim \mathcal{P}(\lambda)$ iid

On calcule la F.C. :

$$\begin{aligned}\Phi_T(u) &= E[e^{iT u}] \\ &= \prod_{i=1}^n E[e^{ju \Sigma X_i}] \\ &= e^{n\lambda(e^{ju} - 1)}\end{aligned}$$

Ainsi

$$T \sim \mathcal{P}(n\lambda)$$

Donc :

- Sous H_0 : $T \sim \mathcal{P}(n\lambda_0)$
- Sous H_{10} : $T \sim \mathcal{P}(n\lambda_1)$

Puis

$$\begin{aligned}\alpha &= P[T > S_\alpha \mid T \sim \mathcal{P}(n\lambda_0)] \\ &= 1 - P[T < S_\alpha \mid T \sim \mathcal{P}(n\lambda_0)] \\ &= 1 - \sum_{k=0}^{\text{Plan}(S_\alpha)} \left(\frac{(n\lambda_0)^k}{k!} \right) e^{-n\lambda_0}\end{aligned}$$

Et avec $\alpha = 5\%$

$$95\%? + \begin{cases} k=0: e^{-\lambda_0 n} \\ k=1: \frac{n\lambda_0}{1} e^{-\lambda_0 n} \\ k=2: \frac{(n\lambda_0)^2}{2!} e^{-n\lambda_0} \end{cases}$$

Généralement on utilise un calculateur

Théorème 2.2: Test du rapport de vraisemblance généralisé (Neyman Pearson pour hypothèses composites)

Très compliqué à la main, on le fera sûrement pas en TD (ni partiel en théorie)

Test paramétrique à hypothèses composites : $H_0 : \theta \in \Theta_0$, $H_1 : \theta \in \Theta_1$

Rejet de H_0 si :

$$\frac{L(x_1, \dots, x_n \mid \hat{\theta}_1^{\text{MV}})}{L(x_1, \dots, x_n \mid \hat{\theta}_0^{\text{MV}})} > S_\alpha$$

où $\hat{\theta}_0^{\text{MV}}$ et $\hat{\theta}_1^{\text{MV}}$ sont les estimateurs du maximum de vraisemblance de θ sous les hypothèses H_0 et H_1

Remarque : $L(x_1, \dots, x_n \mid \hat{\theta}_i^{\text{MV}}) = \sup_{\theta \in \Theta_i} L(x_1, \dots, x_n \mid \theta)$

3. Au partiel

- Faire un test de Neymann Pearson, construire une statistique etc