Exercices de Statistiques
Université de Lorraine

Estimation et théorie des tests

Clément Dell'Aiera

1 Généralités sur l'estimateur du maximum de vraisemblance

1. Rappeler les propriétés de l'EMV.
2. Soient X_j des variables exponentielles indépendantes de paramètre $\theta > 0$, non-observées, et T un instant de censure. Soit $\mathcal{E}^n$ l'expérience engendrée par l'observation du n-échantillon $X_j^* = \min\{T, X_j\}$. Donner une mesure qui domine le modèle et calculer sa vraisemblance.
3. Montrer que l'estimateur du maximum de vraisemblance ne dépend pas du choix de la mesure dominante.

2 Exemples de calculs de maximum de vraisemblance

Pour chaque loi, on considère un n-échantillon tiré de façon i.i.d selon cette loi. Proposer un espace des paramètres donnant un modèle identifiable. Donner une mesure dominante si possible. Calculer la vraisemblance du modèle, ainsi que la log-vraisemblance, donner les équations de vraisemblance et déterminer , s'il existe, un estimateur du maximum de vraisemblance.

1. Modèle gaussien standard, de densité par rapport à la mesure de Lebesgue
$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2\sigma^2}(x-\mu)^2 \quad ,\theta = (\mu, \sigma).$$
2. Modèle de Bernoulli
$$\mathbb{P}_\theta(X = 1) = 1 - \mathbb{P}(X = 0) = \theta.$$
3. Modèle de Laplace, où $\sigma > 0$ est connu, de densité par rapport à la mesure de Lebesgue
$$f_\theta(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\theta|}{\sigma}\right).$$

4. Modèle uniforme, de densité par rapport à la mesure de Lebesgue

$$f_\theta(x) = \frac{1}{\theta} 1_{[0,\theta]}(x).$$

5. Modèle de Cauchy, de densité par rapport à la mesure de Lebesgue

$$f_\theta(x) = \frac{1}{\pi(1-(x-\theta)^2)}.$$

6. Modèle de translation. On considère la densité

$$h(x) = \frac{e^{-\frac{|x|}{2}}}{2\sqrt{2\pi|x|}}.$$

Le modèle de translation par rapport à la densité h est le modèle dominé par la mesure de Lebesgue sur $\mathbb{R}$ de densités

$$f_\theta(x) = h(x-\theta) \quad , x \in R, \theta \in \mathbb{R}.$$

3 Méthode des moments

1. Calculer des estimateurs des moments d'ordre 1 et 2 pour l'expérience engendrée par l'observation d'un n-échantillon de variables exponentielles de paramètre $\theta > 0$. Donner l'asymptotique des ces deux estimateurs.
2. On considère le modèle de translation associé à la famille des lois de Cauchy :

$$f_\theta(x) = \frac{1}{\pi(1+(x-\theta)^2)} \quad , x \in \mathbb{R}.$$

On note g la fonction signe, qui vaut 1 si $x > 0$, -1. Trouver un estimateur pour $\theta \mapsto \mathbb{E}[g(X_1)]$ et donner ses propriétés.

4 Principe de Neyman : décision à 2 points

1. Soit f la densité d'une loi de probabilité sur $\mathbb{R}$, et $\mathcal{E}$ l'expérience statistique engendré par un n-échantillon de loi $p_\theta(x) = f(x-\theta)$. On suppose que $\Theta = \{0, \theta_0\}$ avec $\theta_0 \neq 0$. On veut tester $H_0|\theta = 0$ contre $H_1|\theta = \theta_0$.
   (a) Décrire l'expérience statistique et donner la vraisemblance du modèle.
   (b) Donner la zone de rejet du test de Neyman-Pearson de niveau α associé à H_0 et H_1.
2. L'expérimentateur observe une seule réalisation d'une v.a. X de loi de Poisson de paramètre $\theta > 0$. On veut tester $H_0|\theta = \theta_0$ contre $H_1|\theta = \theta_1$, où $\theta_0 \neq \theta_1$.
   (a) Donner la zone de rejet du test de Neyman-Pearson de niveau α associé.
   (b) Sachant que $\mathbb{P}_{\theta_0}(X > 9) = 0.032$ et $\mathbb{P}_{\theta_1}(X > 8) = 0.068$, donner une zone de rejet explicite pour $\alpha = 0.05 = 5\%$. Le test est-il optimal ?

5 Neyman-Pearson : familles à rapport de vraisemblance monotone

1. Soit $\mathcal{E}$ l'expérience statistique engendrée par un n-échantillon de loi normale $\mathcal{N}(\theta, \sigma^2)$, où σ^2 est connu, et $\theta \in \Theta = \mathbb{R}$. On souhaite tester $H_0 | \theta = \theta_0$ contre $H_1 | \theta = \theta_1$, où $\theta_0 < \theta_1$.
   (a) Décrire le modèle ainsi que la vraisemblance. On choisira la mesure de Lebesgue comme mesure dominante.
   (b) Calculer le rapport de vraisemblance
   $$\frac{f(\theta_1, Z)}{f(\theta_0, Z)}.$$
   (c) Donner la zone de rejet pour le test de Neyman-Pearson associé.
2. Pour la même expérience statistique, on a un test optimal (uniformément plus puissant) de H_0 contre H_1 donné par la région de rejet
   $$\mathcal{R} = \{\overline{X}_n > c\}$$
   où c est solution de $\mathbb{P}_{\theta_0}(\overline{X}_n > c) = \alpha$.
   (a) Calculer explicitement la valeur de la constante $c = c(\theta_0, \alpha)$.
   (b) Calculer la puissance de ce test.

6 Exercice

L'expérimentateur observe 2 échantillons indépendants $X_1, ..., X_n$ et $Y_1, ..., Y_m$ de tailles distinctes $n \neq m$, de lois respectives $\mathcal{N}(\mu_1, \sigma_1^2)$ et $\mathcal{N}(\mu_2, \sigma_2^2)$. Il souhaite tester
$$H_0 \ : \ \mu_1 = \mu_2 \quad \text{contre} \quad H_1 : \ \mu_1 \neq \mu_2.$$
Si $s_{n,1}^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \overline{X}_n)^2$ et $s_{m,2}^2 = \frac{1}{m} \sum_{j=1}^m (Y_j - \overline{Y}_m)^2$, construire un test basé sur la statistique
$$T_{n,m} = \frac{\overline{X}_n - \overline{Y}_m}{\sqrt{s_{n,1}^2 + s_{m,2}^2}}$$
et étudier sa consistance.

7 Neyman-Pearson : loi exponentielle

On observe un n-échantillon $\underline{x} = (X_1, ..., X_n)$ de variables iid de loi exponentielle de paramètre $\lambda > 0$, de densité
$$x \mapsto \lambda \exp(-\lambda x) 1_{x \leq 0}.$$

1. Rappeler l'espérance et la variance (les calculer si besoin) d'une loi exponentielle de paramètre λ. On rappelle que $2\lambda \sum_{j=1}^n X_j$ suit alors une loi du χ^2 à $2n$ degrés de liberté.
2. Ecrire le modèle statistique engendré par l'observation $\underline{x}$.

3. Calculer l'estimateur du maximum de vraisemblance $\hat{\lambda}^{MV}$ de λ.
4. Montrer que $\hat{\lambda}^{MV}$ est asymptotiquement normal, et calculer sa variance limite.
5. Soient $0 < \lambda_0 < \lambda_1$. Construire un test d'hypothèse de

$$H_0 : \lambda = \lambda_0 \text{ contre } H_1 : \lambda = \lambda_1$$

de niveau α et uniformément plus puissant. Expliciter le choix du seuil définissant la région critique. Montrer que le test est consistant, i.e. que l'erreur de seconde espèce du test tend vers 0 lorsque $n \to \infty$.