Partiel de Statistique

Mercredi 16 janvier 2008, 8h00-10h00

Partiel sans documents (une feuille manuscrite A4 est autorisée)

RAPPELS LOI GAMMA

LOI	Densité de probabilité	moyenne	variance	fonction caractéristique.
Gamma $\Gamma(\theta, \nu)$	$f(x) = \frac{\theta^\nu}{\Gamma(\nu)} e^{-\theta x} x^{\nu-1}$ $\theta > 0,\ \nu > 0$ $x \geq 0$	$\frac{\nu}{\theta}$	$\frac{\nu}{\theta^2}$	$\frac{1}{\left(1 - i\frac{t}{\theta}\right)^\nu}$

RAPPELS FONCTION GAMMA

$$\Gamma(n) = \int_0^\infty u^{n-1} e^{-u} du = (n-1)!$$

Exercice 1 : Estimation

On considère une variable aléatoire X de loi de Pareto (ce que l'on notera dans la suite $X \sim P(a)$) de densité

$$h(x) = \frac{a}{x^{a+1}} \mathbb{I}_{[1,\infty[}(x),$$

où $a > 0$ est un paramètre inconnu et où $\mathbb{I}_{[1,\infty[}(x)$ est la fonction indicatrice sur le domaine $[1, \infty[$ ($\mathbb{I}_{[1,\infty[}(x) = 1$ si $x \in [1, \infty[$ et $\mathbb{I}_D(x) = 0$ sinon). Dans cette partie, on dispose de n observations $x_1, ..., x_n$ associées à cette loi de Pareto et on cherche à estimer le paramètre a.

1) Montrer que l'estimateur du maximum de vraisemblance de a s'écrit (en n'oubliant pas de faire un tableau de variation)

$$\widehat{a}_{MV} = \frac{n}{\sum_{i=1}^n \ln X_i}.$$

- Montrer que la variable aléatoire $Z = \ln X$ suit une loi gamma dont on déterminera les paramètres.
- Déterminer la fonction caractéristique de $U = \sum_{i=1}^n Z_i = \sum_{i=1}^n \ln X_i$ lorsque $(X_1, ..., X_n)$ est un échantillon de loi de Pareto $P(a)$. En déduire que U suit une loi gamma de paramètres a et n, i.e. $U \sim \Gamma(a, n)$. Déterminer la moyenne et la variance de la variable aléatoire $\frac{1}{U}$ notées $E\left[\frac{1}{U}\right]$ et $V\left[\frac{1}{U}\right]$.
- Déterminer le biais de l'estimateur $\widehat{a}_{MV}$. En déduire un estimateur non biaisé de a noté $\widetilde{a}$ et déterminer la variance de cet estimateur.
- L'estimateur $\widetilde{a}$ est-il l'estimateur efficace de a ?

2) On suppose maintenant qu'on dispose d'une information a priori sur le paramètre a résumée dans la densité a priori

$$g(a) = \lambda e^{-\lambda a} \mathbb{I}_{[0,\infty[}(a).$$

- Montrer que la densité a posteriori du paramètre a notée $f(a|x_1, ..., x_n)$ est la densité d'une loi gamma dont on déterminera les paramètres.
- Déterminer l'estimateur MAP du paramètre a et étudier son comportement lorsque n est "grand".
- Déterminer l'estimateur MMSE du paramètre a.

Exercice 2 : Tests Statistiques

On dispose toujours de n observations $x_1, ..., x_n$ associées à une loi de Pareto $P(a)$ et on considère le test d'hypothèses simples

$$\begin{aligned} H_0 &: a = a_0 \\ H_1 &: a = a_1 \text{ avec } a_1 > a_0 \end{aligned}$$

1) Déterminer la statistique du test de Neyman-Pearson $U(X_1, ..., X_n)$ (notée U pour simplifier) et la zone de rejet de H_0 issue de ce test. On suppose dans ce qui suit que U suit une loi gamma $\Gamma(a_0, n)$ sous l'hypothèse H_0 et une loi gamma $\Gamma(a_1, n)$ sous l'hypothèse H_1.

- Déterminer l'équation intégrale permettant de déterminer le seuil du test de Neyman-Pearson noté S_α en fonction de α.
- Déterminer l'équation intégrale permettant de déterminer la puissance du test de Neyman-Pearson en fonction de S_α.

On suppose dans ce qui suit que $n = 1$ pour simplifier les calculs.

- Montrer que la première équation intégrale donnée ci-dessus permet de déterminer S_α en fonction de α.
- Montrer que la deuxième équation intégrale donnée ci-dessus permet de déterminer la puissance π_α du test de Neyman-Pearson en fonction de α et vérifier que

$$\lim_{\alpha \to 0} \pi_\alpha = 0 \text{ et } \lim_{\alpha \to 1} \pi_\alpha = 1$$

- Représenter les courbes caractéristiques opérationnelles du récepteur pour $(a_0, a_1) = (1, 2)$ et pour $(a_0, a_1) = (1, 3)$. Commentaires.

2) On désire faire un test de Kolmogorov pour tester si les observations $x_1, ..., x_n$ sont issues d'une loi de Pareto de paramètre $a_0 = 1$, c'est-à-dire

$$\begin{aligned} H_0 &: X_i \sim P(1) \\ H_1 &: \text{non } H_0 \end{aligned}$$

- Déterminer et tracer la fonction de répartition d'une loi de Pareto $P(1)$ notée $F_0(x)$.
- Tracer la fonction de répartition $\widehat{F}(x)$ des observations (pour simplifier les calculs, les observations sont entières mais ce serait très peu probable en pratique !)

$$x_1 = 2, x_2 = 3, x_3 = 4, x_4 = 5, x_5 = 6.$$

- Calculer

$$\sup_{x \in \mathbb{R}} \left| F_0(x) - \widehat{F}(x) \right|$$

  On pourra recopier et compléter le tableau suivant

x_i	2.00	3.00	4.00	5.00	6.00
$F(x_i)$					
E_i^+					
E_i^-					
$Max(E_i^+, E_i^-)$					

- Expliquer comment on peut conclure.