Deuxième année
2002-2003

Théorie des tests

Corrigé des travaux dirigés n°2

Guillaume Lacôte
Bureau **E03**
✉ Guillaume.Lacote@ensae.fr
☞ http://ensae.no-ip.com//SE207/

Dernière mise-à-jour: 20030402-10h07

Corrigé de l'exercice 1

(Voir correction manuscrite)

Corrigé de l'exercice 2

☞ Q1 (a) On calcule d'abord la densité, qui vaut 0 pour $x < 1$ et

$$f_{1,\gamma}(x) = \gamma\, x^{-(\gamma+1)} = \frac{\gamma}{x^{\gamma+1}}$$

pour $x \geq 1$.
D'où $\ln f_{1,\gamma}(x) = \ln\gamma - (\gamma+1)\ln x$. Une application directe du lemme de Neyman-Pearson fait donc intervenir la statistique $\sum_{i=1}^{n} \ln X_i$. (Remarquons que cette statistique est suffisante pour γ.)

(b) Il s'agit d'abord d'effectuer des intégrations par parties. On obtient :

$$E_{\gamma_0}(\ln X_1) = \gamma_0 \int_1^{+\infty} \ln x\; x^{-(\gamma_0+1)}\,dx = \frac{1}{\gamma_0}$$

et

$$\mathbb{V}(\gamma_0)(\ln X_1) = \gamma_0 \int_1^{+\infty} \left(\ln x - \frac{1}{\gamma_0}\right)^2 x^{-(\gamma_0+1)}\,dx = \frac{1}{\gamma_0^2}$$

Par le théorème central limite (que l'on peut utiliser ici en raison de l'existence du moment d'ordre 2), on en déduit que :

$$\frac{\gamma_0}{\sqrt{n}} \sum_{i=1}^{n} \left(\ln X_i - \frac{1}{\gamma_0}\right)$$

converge en distribution vers la loi normale $\mathcal{N}(0,1)$, de fonction de répartition (fr) notée Φ. Ainsi,

$$\begin{aligned}
P\left\{\frac{1}{n}\sum_{i=1}^{n} \ln X_i \;>\; c_{\alpha,n}\right\} &= P\left\{\sum_{i=1}^{n}\left(\ln X_i - \frac{1}{\gamma_0}\right) \;>\; n\left(c_{\alpha,n} - \frac{1}{\gamma_0}\right)\right\} \\
&= P\left\{\frac{\gamma_0}{\sqrt{n}}\sum_{i=1}^{n}\left(\ln X_i - \frac{1}{\gamma_0}\right) \;>\; \gamma_0\sqrt{n}\left(c_{\alpha,n} - \frac{1}{\gamma_0}\right)\right\} \\
&\longrightarrow \lim_{n\to\infty}\left[1 - \Phi\left(\gamma_0\sqrt{n}\left(c_{\alpha,n} - \frac{1}{\gamma_0}\right)\right)\right] \\
&= \alpha
\end{aligned}$$

d'où

$$\lim_{n\to\infty} \gamma_0\sqrt{n}\left(c_{\alpha,n} - \frac{1}{\gamma_0}\right) = u_\alpha$$

avec $1 - \Phi(u_\alpha) = \alpha$. Finalement, l'approximation (asymptotique) cherchée est :

$$c_{\alpha,n} \;\approx\; \frac{1}{\gamma_0} + \frac{u_\alpha}{\gamma_0\sqrt{n}}$$

(c) On a :

$$\begin{aligned} P\{\ln X_1 \le y\} &= P\{X_1 \le \exp y\} \\ &= 1 - \left(\frac{\exp y}{\exp(\ln a)}\right)^{-\gamma_0} \\ &= 1 - \exp(-\gamma_0 (y - \ln a)) \end{aligned}$$

Puisque $a = 1$, il s'agit d'une loi exponentielle de paramètre $\lambda = \gamma_0$.
Par conséquent, la va $\sum_{i=1}^n \ln X_i$, somme de n va iid de loi $\mathrm{Exp}(\gamma_0^{-1})$, est une loi Gamma de moyenne n/γ_0 et de variance n/γ_0^2. Avec une table des lois Gamma, on pourrait donc déterminer la valeur critique $c_{\alpha,n}$ exactement (et le test étudié peut alors être considéré comme "à distance finie").

(d) La justification consiste à s'assurer que le test est à rapport de vraisemblance monotone (RVM) en la statistique

$$T_n = \sum_{i=1}^n \ln X_i$$

ce qui se prouve facilement. Le test UPP cherché a donc encore, d'après le cours, la même région de rejet :

$$\frac{1}{n}\sum_{i=1}^n \ln X_i > c_{\alpha,n}$$

☞ Q2 (a) i. On a $F_0^{-1}(t) = (1-t)^{-1/\gamma_0}$ pour $0 \le t < 1$.

ii. Comme la densité vaut 0 pour pour $x < 1$ et

$$f_{1,\gamma}(x) = \gamma\, x^{-(\gamma+1)} = \frac{\gamma}{x^{\gamma+1}}$$

pour $x \ge 1$, on a

$$\ln f_{1,\gamma}(x) = \ln\gamma - (\gamma+1)\ln x$$

Donc la log-vraisemblance est

$$\ell_n = n\,\ln\gamma - (\gamma+1)\left(\sum_{i=1}^n \ln X_i\right)$$

L'équation de la vraisemblance s'écrit par conséquent

$$\frac{n}{\gamma} = \left(\sum_{i=1}^n \ln X_i\right)$$

ce dont on tire

$$\hat\gamma_n = \frac{n}{\left(\sum_{i=1}^n \ln X_i\right)}$$

iii. Comme $\ln f_{1,\gamma}(x) = \ln\gamma - (\gamma+1)\ln x$, il vient

$$s_0(x) = \frac{1}{\gamma_0} - \ln x$$

pour $x \geq 1$. Les calculs précédents conduisent alors à

$$I_0 = \frac{1}{\gamma_0^2}$$

iv. D'après le cours, la fonction $\phi_0(x)$ est égale à

$$I_0^{-1}\, s_0(x) = \gamma_0^2 \left(\frac{1}{\gamma_0} - \ln x \right)$$

pour $x \geq 1$.

v. D'après le cours, c'est la loi normale de moyenne 0 et de variance γ_0^2.

(b) D'après le cours, il faut d'abord calculer les fonctions g_0 et h_0, toutes deux définies sur $[0,1]$. Or

$$g_0(t) = \left.\frac{\partial F_{1,\gamma}}{\partial\gamma}\right|_{\gamma=\gamma_0} \left(F_0^{-1}(t)\right)$$

et par ailleurs

$$\frac{\partial F_{1,\gamma}}{\partial\gamma}(x) = \ln x\; x^{-\gamma}$$

pour $x \geq 1$. Par suite,

$$g_0(t) = -\frac{1}{\gamma_0}(1-t)\,\ln(1-t)$$

D'autre part, $h_0 = s_0 \circ F_0^{-1}$, soit

$$h_0(t) = \frac{1}{\gamma_0}\left(1 + \ln(1-t)\right).$$

Donc

$$\int_0^u h_0(t)\,dt = \frac{1}{\gamma_0}\left(u + \int_0^u \ln(1-t)\,dt\right).$$

Par le changement de variables $s = 1-t$ puis par un calcul direct de primitive, on obtient $\int_0^u h_0(t)\,dt = g_0(u)$

On peut alors appliquer le résultat du cours, et on trouve

$$r_{\hat{B}}(u_1, u_2) = u_1 \wedge u_2 - u_1u_2 - (1-u_1)(1-u_2)\ln(1-u_1)\ln(1-u_2)$$

qui ne dépend plus de γ_0.

(c) Un calcul détaillé (à faire !) conduit à la même formule. Cela s'explique par le résultat de la question 1.3. Les valeurs critiques se déduisent donc des valeurs analogues pour tester l'adéquation du modèle exponentiel. Elles sont tabulées dans le livre de D'Agostino & Stephens, par exemple. On les utilise sur l'échantillon transformé défini par :

$$V_i = 1 \;-\; X_i^{-\hat{\gamma}_n}$$

dans le cas $a = 1$. Ou bien, ce qui revient au même, on teste directement l'exponentialité sur l'échantillon des $\ln X_i$.

(d) Ou bien on utilise l'estimateur du maximum de vraisemblance, ou bien l'estimateur de γ basé sur les données groupées, ce qui conduit à une équation simple. On utilise les résultats du cours pour les lois limites.

☞ Q3 (a) On procède comme pour l'exponentielle

$$F(x) = 1 - \exp -\frac{x-\mu}{\theta}$$

On estime le paramètre a par $X_{(1,n)}$.

(b) On montre, à partir des indications, que si $\rho > 1$, il existe un entier aléatoire $n_0(\omega)$, fini p.s., tel que

$$0 \leq U_{(1,n)} \leq cte \left(\frac{(\ln n)^\rho}{n} \right)$$

pour tout $n \geq n_0(\omega)$.

(c) On en déduit que l'estimateur $X_{(1,n)}$ de a est superefficace. On applique donc les mêmes tables sur les

$$V_i = 1 \ - \ \left(\frac{X_i}{X_{(1,n)}} \right)^{-\hat{\gamma}_n}$$

pour $X_i \neq X_{(1,n)}$.

Corrigé de l'exercice 3

(Voir correction manuscrite)