



Deuxième année  
2002-2003

Théorie des tests  
*Séance de travaux dirigés n°2*

Guillaume Lacôte  
Bureau **E03**

✉ [Guillaume.Lacote@ensae.fr](mailto:Guillaume.Lacote@ensae.fr)

☞ <http://ensae.no-ip.com//SE207/>

Exercice 1
------------

- ☞ Q1 Dans l'exercice 1 du TD 9 du cours d'Estimation et tests, on a testé l'adéquation à une loi de Poisson du nombre d'appels par seconde dans un central téléphonique. Les données étaient les suivantes :

Nombre d'appels par seconde	Effectifs observés
0	6
1	15
2	40
3	42
4	37
5	30
6	10
7	9
8	5
9	3
10	2
11	1

La mise en oeuvre du test du  $\chi^2$  nécessitant un nombre fini de classes d'effectifs suffisants, on a regroupé les modalités de la manière suivante :

$$\begin{cases} J_k = \{k\} \text{ pour } k = 0, 1, \dots, K-1 \\ J_K = \{K, K+1, \dots\}, \end{cases}$$

avec  $K = 8$ . On note  $n$  le nombre total d'observations et  $n_k$  les effectifs des  $K+1$  classes.

- ☞ Q2 Ecrire la statistique  $\mathcal{T}_n^2$  du test de l'adéquation à la famille des lois de Poisson (le paramètre étant inconnu).
- ☞ Q3 On note  $\mu_0$  la vraie loi des observations et on pose comme dans le cours

$$\gamma_0 = 1 - \frac{1}{I_0} \sum_{k=0}^K \left( \int_{J_k} s_0 d\mu_0 \right)^2 \mu_0^{-1}(J_k)$$

- (a) Expliciter, dans le cas où  $\mu_0$  est la loi de Poisson de paramètre  $\lambda$ , les quantités

$$q_k = \mu_0(J_k), \quad s_0(k), \quad \int_{J_k} s_0 d\mu_0, \quad I_0$$

en fonction de  $k, \lambda$  et  $p_k(\lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$ .

(b) Montrer que  $\gamma_0$  s'écrit aussi

$$\gamma_0 = \lambda \left( \sum_{k \geq K} p_k \right) \left[ \frac{\sum_{k \geq K} p_k s_0(k)^2}{\sum_{k \geq K} p_k} - \left( \frac{\sum_{k \geq K} p_k s_0(k)}{\sum_{k \geq K} p_k} \right)^2 \right].$$

(c) Interpréter le terme entre crochets comme une variance. En déduire que  $\gamma_0 > 0$ .

☞ Q4 (a) Rappeler l'expression de la distribution limite sous l'hypothèse nulle de  $\mathcal{T}_n^2$  en fonction de  $\gamma_0$ .

Quelle approximation le corrigé du TD 9 faisait-il implicitement ? Quelle conséquence l'erreur commise a-t-elle sur le niveau du test ?

(b) Estimer  $\gamma_0$  à l'aide d'une calculatrice programmable ou d'un logiciel de calcul numérique (on utilisera les valeurs  $K = 8$  et  $\lambda = \hat{\lambda}_{\text{emv}} = 3,7$ ).

L'approximation faite était-elle justifiée ?

---

## Corrigé de l'exercice 1

(Voir correction manuscrite)

★  
★ ★

## Exercice 2

On appelle loi de Pareto de paramètres  $a > 0$  et  $\gamma > 0$ , et on note  $\text{Pareto}(a, \gamma)$ , la mesure de probabilité sur l'ensemble  $\mathbf{R}$  des réels dont la fonction de répartition  $F_{a,\gamma}$  est définie par  $F_{a,\gamma}(x) = 0$  pour  $x < a$  et

$$F_{a,\gamma}(x) = 1 - \left( \frac{x}{a} \right)^{-\gamma} \quad \text{pour } x \geq a$$

☞ Q1 On suppose que  $a$  est connu. On peut alors se ramener au cas  $a = 1$ , ce que l'on fera ici. Soit  $X_1, \dots, X_n$  un échantillon de variables aléatoires indépendantes et de même loi (iid)  $\text{Pareto}(1, \gamma)$ .

- (a) Soient  $\gamma_1 > 0$  et  $\gamma_0 > 0$  deux valeurs distinctes du paramètre  $\gamma$ . On souhaite tester l'hypothèse simple  $H_0 = [\gamma = \gamma_0]$  contre l'alternative simple  $H_1 = [\gamma = \gamma_1]$ .  
Montrer que le test le plus puissant de niveau  $0 < \alpha < 1$  de  $H_0$  contre  $H_1$  a une région de rejet de la forme  $n^{-1} \sum_{i=1}^n \ln X_i > c_{\alpha, n}$  ou de la forme  $n^{-1} \sum_{i=1}^n \ln X_i < c_{\alpha, n}$ .  
Discuter.
- (b) Déterminer, sous  $H_0$ , l'espérance et la variance de la variable aléatoire  $\ln X_1$ . Dédire du théorème central limite une approximation de  $c_{\alpha, n}$  utilisable lorsque  $n$  est assez grand.
- (c) Quelle est, sous  $H_0$ , la loi de  $\ln X_1$ ?  
De quelle forme est la loi de  $\sum_{i=1}^n \ln X_i$ ?
- (d) Toujours dans le cas  $a = 1$  connu, déterminer le test uniformément le plus puissant (UPP) de niveau  $\alpha$  de  $H_0 : [\gamma \leq \gamma_0]$  contre  $H_1 = [\gamma > \gamma_0]$ .

☞ Q2 Soit  $X_1, \dots, X_n$  un échantillon de variables aléatoires iid.

- (a) On suppose encore  $a = 1$  connu. On note  $\gamma_0$  la vraie valeur de  $\gamma$  sous  $H_0$ , et on note  $F_0 = F_{1, \gamma_0}$ .
- Calculer la fonction inverse  $F_0^{-1}(\cdot)$ .
  - Déterminer l'estimateur du maximum de vraisemblance  $\hat{\gamma}_n$  basé sur l'échantillon.
  - Déterminer la fonction de score  $s_0(x) = s(x; \gamma_0)$  et l'information de Fisher  $I_0 = I(\gamma_0)$ .
  - En déduire une expression de la forme

$$n^{1/2} (\hat{\gamma}_n - \gamma_0) = n^{-1/2} \sum_{i=1}^n \phi_0(X_i) + o_P(1)$$

sous  $H_0$ , en précisant la fonction  $\phi_0$

- Quelle est la limite en distribution de  $n^{1/2}(\hat{\gamma}_n - \gamma_0)$  sous  $H_0$ , quand  $n \rightarrow \infty$ ?
- (b) Dédire des questions précédentes l'expression de la fonction de covariance du processus gaussien centré  $\hat{\mathbf{B}}(\cdot)$ , limite en distribution sous  $H_0$ , quand  $n \rightarrow \infty$ , de la suite de processus

$$\hat{\mathbf{B}}_n(t) = n^{\frac{1}{2}} [(\mathbf{F}_n - F_{1, \hat{\gamma}_n}) \circ F_0^{-1}](t), \quad t \in [0, 1]$$

où  $\mathbf{F}_n(\cdot)$  désigne la fonction de répartition empirique.

- Comparer le résultat obtenu à celui que l'on obtient lorsque l'on teste l'adéquation du modèle exponentiel, et en déduire les valeurs critiques, pour  $\alpha = 0.10, 0.05, 0.025$  et  $0.01$ , des tests de Kolmogorov-Smirnov et de Cramér-von Mises pour tester l'adéquation du modèle Pareto  $(1, \gamma)$ .
- Mettre en application du test d'adéquation du chi-deux pour tester l'adéquation de Pareto  $(1, \gamma)$ , avec  $\gamma$  inconnu.

☞ Q3 On se propose de tester l'adéquation de Pareto  $(a, \gamma)$ , lorsque  $a$  et  $\gamma$  sont tous deux inconnus.

- (a) Soit  $U_1, \dots, U_n$  un échantillon de va iid uniformes sur  $[0, 1]$ , et soit  $U_{(1, n)}$  la plus petite observation de cet échantillon. Montrer que  $nU_{(1, n)}$  converge en distribution vers la loi exponentielle de moyenne 1 quand  $n \rightarrow \infty$ .

- (b) Montrer que si  $(A_j)_{j \geq 1}$  est une suite quelconque d'événements et si  $\sum_{j=1}^{+\infty} P(A_j)$  converge, alors presque sûrement seuls un nombre fini de  $A_j$  ont lieu.
- (c) Proposer un test de l'adéquation de Pareto( $a, \gamma$ ).

## 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  $\gamma$ .)

(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  $\Phi$ . Ainsi,

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

d'où

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

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 :

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

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  $\text{Exp}(\gamma_0^{-1})$ , est une loi Gamma de moyenne  $n/\gamma_0$  et de variance  $n/\gamma_0^2$ . 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  $0 \leq t < 1$ .

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  $x \geq 1$ , 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  $\gamma_0^2$ .

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

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

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

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

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

$$r_B(u_1, u_2) = u_1 \wedge u_2 - u_1 u_2 - (1 - u_1)(1 - u_2) \ln(1 - u_1) \ln(1 - u_2)$$

qui ne dépend plus de  $\gamma_0$ .

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

★  
★ ★

### Exercice 3

On rappelle la définition de l'inverse généralisée d'une fonction de répartition  $F$  :

$$F^{\text{inv}}(u) = \inf\{x \in \mathbb{R}, F(x) \geq u\}$$

- ☞ Q1 Soit  $X$  une variable aléatoire suivant la loi de Poisson de paramètre  $\lambda > 0$ . Déterminer et représenter graphiquement sa fonction de répartition  $F_\lambda$  et l'inverse généralisée  $F_\lambda^{\text{inv}}$  de  $F_\lambda$ .



☞ Q2 (a) Pour toute f.d.r.  $F$ , vérifier l'équivalence :

$$u \leq F(x) \iff F^{\text{inv}}(u) \leq x$$

(b) Soit  $U$  une v.a. uniforme sur  $[0, 1]$ . On pose :

$$Y = F_{\lambda}^{\text{inv}}(U)$$

Montrer que  $Y$  suit la même loi que  $X$ .

(c) Expliquer par quel algorithme on construit  $Y$  à partir de  $U$ .

☞ Q3 Déterminer la f.d.r.  $G_{\lambda}$  de la v.a.  $Z = F_{\lambda}(Y)$ .

Cette v.a. est-elle uniforme sur  $[0, 1]$ ? La f.d.r.  $G_{\lambda}$  dépend-elle de  $\lambda$ ?

☞ Q4 Peut-on utiliser les tests de Kolmogorov-Smirnov ou de Cramer-von Mises pour tester l'adéquation au modèle de Poisson?

---

## Corrigé de l'exercice 3

(Voir correction manuscrite)

★  
★ ★