EXAMEN STATISTIQUE - 1TR

Mardi 8 Décembre 2015

Partiel sans document (Une feuille A4 recto-verso autorisée)

Exercice 1 : Estimation (11 points)

On considère n variables aléatoires $X_1, ..., X_n$ indépendantes suivant la même loi continue de densité

$$p(x_i; r, \alpha) = \begin{cases} \frac{r\alpha^r}{(\alpha+x_i)^{r+1}} & \text{si } x_i > 0 \\ 0 \text{ sinon} \end{cases}$$

avec $\alpha > 0$ et $r > 2$. On admettra qu'une telle loi est de moyenne $E[X_i] = \frac{\alpha}{r-1}$ et de variance $\text{Var}[X_i] = \frac{2\alpha^2}{(r-1)(r-2)}$. On suppose que le paramètre α est connu et on cherche à étudier des estimateurs du paramètre r.

1. Montrer que la vraisemblance associée à l'échantillon $(X_1, ..., X_n)$ admet un unique maximum global pour une valeur de r que l'on déterminera. En déduire l'estimateur du maximum de vraisemblance de r noté $\hat{r}_{\text{MV}}$ en fonction de n, α et de l'échantillon $(X_1, ..., X_n)$.
2. En utilisant l'expression de $E[X_i]$, déterminer un estimateur des moments du paramètre r noté $\hat{r}_{\text{Mo}}$.
3. Montrer que si X_i possède la densité $p(x_i; r, \alpha)$ définie ci-dessus, alors $V_i = \ln\left(\frac{\alpha+X_i}{\alpha}\right)$ suit une loi exponentielle de densité
$$g(v_i; r) = \begin{cases} re^{-rv_i} & \text{si } v_i > 0 \\ 0 \text{ sinon} \end{cases}$$
En utilisant les tables de lois, déterminer la fonction caractéristique de V_i et montrer que $V = \sum_{i=1}^n V_i$ suit une loi gamma $\Gamma(r, n)$. En déduire que $\hat{r}_{\text{MV}}$ suit une loi inverse gamma $\text{IG}(\theta, \nu)$ de paramètres $\theta = nr$ et $\nu = n$.
4. En utilisant le fait que $\hat{r}_{\text{MV}}$ suit une loi inverse gamma $\text{IG}(\theta, \nu)$ de paramètres $\theta = nr$ et $\nu = n$, montrer que $\hat{r}_{\text{MV}}$ est un estimateur biaisé du paramètre r et en déduire un estimateur non biaisé de r noté $\hat{r}^*$. Déterminer la variance de $\hat{r}^*$ et en déduire que cet estimateur est convergent.
5. Déterminer la borne de Cramer-Rao d'un estimateur non biaisé de r. L'estimateur $\hat{r}^*$ est-il l'estimateur efficace du paramètre r ?
6. On suppose qu'on dispose d'une information a priori sur le paramètre r résumée dans une loi gamma $\Gamma(a, b)$. Montrer que la loi a posteriori de $r|x_1, ..., x_n$ est également une loi gamma dont on précisera les paramètres. Quel est l'estimateur du maximum a posteriori du paramètre r noté $\hat{r}_{\text{MAP}}$?

Exercice 2: Tests Statistiques (9 points)

Comme dans l'exercice précédent, on considère n variables aléatoires $X_1, ..., X_n$ indépendantes suivant la même loi continue de densité

$$p(x_i; r, \alpha) = \begin{cases} \frac{r\alpha^r}{(\alpha+x_i)^{r+1}} & \text{si } x_i > 0 \\ 0 \text{ sinon} \end{cases}$$

avec $\alpha > 0$ et $r > 2$. Dans un premier temps, on cherche à effectuer le test d'hypothèses simples

$$\begin{aligned} H_0 &: r = r_0 \\ H_1 &: r = r_1 \end{aligned}$$

avec $r_1 > r_0 > 2$.

1. Montrer que le test de Neyman Pearson conduit à la statistique de test

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

   et indiquer la région critique de ce test.

2. On admet que T_n suit une loi gamma $\Gamma(r_j, n)$ sous l'hypothèse H_j avec $j \in \{0, 1\}$, et on suppose qu'on sait calculer la fonction Q_n définie par

$$Q_n(x) = \frac{1}{\Gamma(n)} \int_0^x u^{n-1} e^{-u}\, du$$

   ainsi que son inverse $Q_n^{-1}(x)$ pour toutes les valeurs de n et de x. Déterminer la valeur du seuil K_α du test de Neyman Pearson en fonction de r_0, α et de Q_n^{-1}.

3. Déterminer la puissance du test en fonction de K_α, r_1 et Q_n. En déduire les courbes COR du test étudié dans cet exercice et tracer la forme de ces courbes pour différentes valeurs du couple (r_0, r_1).

4. Lorsqu'on ne peut pas calculer les fonctions Q_n et Q_n^{-1}, on peut approcher la loi de T_n par une loi normale d'après le théorème de la limite centrale. Calculer le seuil K_α et la puissance du test lorsqu'on utilise cette loi approchée en fonction de n, r_0, r_1 et de la fonction de répartition de la loi normale $\mathcal{N}(0, 1)$. En déduire l'expression des courbes COR résultant de ces approximations.

Partie I

1) Vraisemblance de l'échantillon (2pts)

$$L(x_1,\dots,x_n;r)=\prod_{i=1}^{n}\frac{r\alpha^r}{(\alpha+x_i)^{r+1}}=\frac{(r\alpha^r)^n}{\prod_{i=1}^{n}(\alpha+x_i)^{r+1}}$$

donc $\ln L(x_1,\dots,x_n;r)=n\ln r+nr\ln\alpha-\sum_{i=1}^{n}(r+1)\ln(\alpha+x_i)$

$$\frac{\partial \ln L}{\partial r}\geq 0 \iff \frac{n}{r}+r\ln\alpha-\sum_{i=1}^{n}\ln(\alpha+x_i)\geq 0$$

$$\iff r\leq\frac{n}{\sum_{i=1}^{n}\ln\left(\frac{\alpha+x_i}{\alpha}\right)}$$

Rq = puisque $x_i>0$, on a $\frac{\alpha+x_i}{\alpha}>1$ et donc $\ln\left(\frac{\alpha+x_i}{\alpha}\right)>0$

D'où le tableau de variations

r	0		$\frac{n}{\sum_{i=1}^{n}\ln\left(\frac{\alpha+x_i}{\alpha}\right)}$		$+\infty$
$\frac{\partial \ln L}{\partial r}$		$+$	0	$-$	
$\ln L$		$\nearrow$		$\searrow$	

1pt

On en déduit que $\frac{n}{\sum_{i=1}^{n}\ln\left(\frac{\alpha+x_i}{\alpha}\right)}$ est un maximum global unique de L

d'où

$$\boxed{\hat{r}_{MV}=\frac{n}{\sum_{i=1}^{n}\ln\left(\frac{\alpha+X_i}{\alpha}\right)}}$$

1pt

2) Méthode des moments (1pt)

$$E[X_i]=\frac{\alpha}{r-1}\iff r=1+\frac{\alpha}{E[X_i]}$$

d'où

$$\boxed{\hat{r}_{MO}=1+\frac{\alpha}{\frac{1}{n}\sum_{i=1}^{n}X_i}}$$

1pt

3) $V_i = \ln\left(\frac{\alpha + X_i}{\alpha}\right) \iff X_i = \alpha\,(e^{V_i} - 1)$ (3 points) 2

Le changement de variables est donc bijectif de $\mathbb{R}^+$ dans D, où D est défini par

$$X_i > 0 \iff \alpha(e^{V_i} - 1) > 0 \iff V_i > 0$$

c'est-à-dire $\boxed{D = \mathbb{R}^+}$. De plus, on a $\frac{dX_i}{dV_i} = \alpha e^{V_i}$, d'où la densité de V_i en appliquant le théorème du changement de variables

$$\pi(v_i) = \frac{r\,\alpha^r}{(\alpha e^{v_i})^{r+1}}\,\left|\alpha e^{v_i}\right|\,\mathbb{1}_D(v_i)$$

$\boxed{1\,\text{pt}}$

$$\boxed{\pi(v_i) = r\,e^{-r v_i}\,\mathbb{1}_{\mathbb{R}^+}(v_i)}$$

On reconnait la densité d'une loi exponentielle, c'est-à-dire d'une loi gamma $\Gamma(r, 1)$. La fonction caractéristique de V_i est donnée dans les tables de lois

$$\phi_{V_i}(t) = E\left[e^{iV_i t}\right] = \boxed{\frac{1}{1 - \frac{it}{r}}}$$

Puisque les va X_i sont indépendants, les va V_i le sont aussi, d'où

$$\phi_V(t) = E\left[e^{itV}\right] = E\left[e^{it\sum_{k=1}^n V_k}\right] = E\left[\prod_{k=1}^n e^{itV_k}\right]$$

$$\underset{\text{ind des va } V_k}{=} \prod_{k=1}^n E\left[e^{itV_k}\right] = \frac{1}{\left(1 - \frac{it}{r}\right)^n}$$

On reconnait la fonction caractéristique d'une loi gamma $\Gamma(r, n)$

Donc $\boxed{V \sim \Gamma(r, n)}$ $\boxed{1\,\text{pt}}$

Pour avoir la loi de $\hat{r}_{MV}$, il suffit de remarquer que

$$\hat{r}_{MV} = \frac{n}{\sum_{k=1}^n \ln\left(\frac{\alpha + X_k}{\alpha}\right)} = \frac{n}{\sum_{k=1}^n V_k} = \frac{n}{V}$$

Pour avoir la loi de $\hat{r}_{MV}$, il suffit de faire le changement de variables $W = \frac{n}{V}$ qui est bijectif de $\mathbb{R}^+$ dans $\mathbb{R}^+$. De plus

$$W = \frac{n}{V} \Leftrightarrow V = \frac{n}{W} \Rightarrow \frac{dV}{dW} = \frac{-n}{W^2} \Rightarrow \left|\frac{dV}{dW}\right| = \frac{n}{W^2}.$$

On en déduit la densité de W

$$h(w) = \frac{r^n}{\Gamma(n)} e^{-\frac{rn}{w}} \left(\frac{n}{w}\right)^{n-1} \frac{n}{w^2} \mathbb{1}_{\mathbb{R}^+}(w)$$

Soit
$$\boxed{h(w) = \frac{(nr)^n}{\Gamma(n)} \frac{1}{w^{n+1}} e^{-\frac{nr}{w}} \mathbb{1}_{\mathbb{R}^+}(w)}$$

On reconnaît la densité d'une loi inverse gamma de paramètres nr et n, c'est à dire $\boxed{\hat{r}_{MV} = \frac{n}{V} \sim IG(nr, n)}$ $\boxed{1\,pt}$

4) En utilisant les tables de lois, on obtient les résultats suivants (2.5 points)

$$E[\hat{r}_{MV}] = \frac{nr}{n-1}$$

$$Var[\hat{r}_{MV}] = \frac{n^2 r^2}{(n-1)^2(n-2)}$$

L'estimateur $\hat{r}_{MV}$ est donc biaisé, mais on peut construire l'estimateur

$$\boxed{\hat{r}^* = \frac{n-1}{n}\hat{r}_{MV} = \frac{n-1}{\sum_{i=1}^{n} \ln\left(\frac{\alpha + x_i}{\alpha}\right)}}$$ $\boxed{1\,pt}$

qui vérifie $E[\hat{r}^*] = \frac{n-1}{n} E[\hat{r}_{MV}] = r$ et donc est non-biaisé.

De manière immédiate, on obtient

$$Var(\hat{r}^*) = \left(\frac{n-1}{n}\right)^2 Var(\hat{r}_{MV}) = \boxed{\frac{r^2}{n-2}}$$ $\boxed{1\,pt}$

$\left.\begin{array}{l} Var[\hat{r}^*] \xrightarrow[n\to+\infty]{} 0 \\ \hat{r}^* \text{ non biaisé} \end{array}\right| \Rightarrow \hat{r}^*$ convergent (estimateur convergent deux) $\boxed{0{,}5\,pt}$

5) Borne de Cramér-Rao d'un estimateur non-biaisé de r (1.5pt)

On a $\frac{\partial \ln L}{\partial r} = \frac{n}{r} + n \ln \alpha - \sum_{i=1}^{n} \ln(\alpha + x_i)$ (voir question 1)

Donc $\frac{\partial^2 \ln L}{\partial r^2} = -\frac{n}{r^2} \Rightarrow E\left[-\frac{\partial^2 \ln L}{\partial r^2}\right] = \frac{n}{r^2}$

d'où la borne de Cramér-Rao

$$\boxed{BCR = \frac{r^2}{n}}$$

1pt

$Var(\hat{r}^*) = \frac{r^2}{n-2} > \frac{r^2}{n} = BCR$ donc $\hat{r}^*$ n'est pas l'estimateur efficace de r

0.5pt

6) La loi a posteriori de $r | X_1, \ldots, X_n$ vérifie (2points)

$$p(r | x_1, \ldots, x_n) \underset{\text{proportionnel à}}{\propto} \underbrace{p(x_1, \ldots, x_n ; r)}_{\text{vraisemblance}} \underbrace{p(r)}_{\text{loi a priori de } r}$$

soit

$$p(r | x_1, \ldots, x_n) \propto \frac{r^n \alpha^{nr}}{\left[\prod_{i=1}^{n} (\alpha + x_i)\right]^{r+1}} e^{-ar} r^{b-1} \mathbb{1}_{\mathbb{R}^+}(r)$$

$$\propto r^{n+b-1} \exp\left[-r\left(a - \cancel{b}n \ln \alpha + \sum_{i=1}^{n} \ln(\alpha + x_i)\right)\right]$$

On reconnait une loi gamma de paramètres $a + \sum_{i=1}^{n} \ln\left(\frac{\alpha + x_i}{\alpha}\right)$ et $n+b$

soit

$$\boxed{r | x_1, \ldots, x_n \sim \Gamma\left(a + \sum_{i=1}^{n} \ln\left(\frac{\alpha + x_i}{\alpha}\right), n+b\right)}$$

1pt

L'estimateur du maximum a posteriori de r s'obtient en maximisant $\ln p(r | x_1, \ldots, x_n)$ par rapport à r. On obtient

$$\boxed{\hat{r}_{MAP} = \frac{1 + \frac{b-1}{n}}{\frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{\alpha + x_i}{\alpha}\right) + \frac{a}{n}}}$$

1pt

On remarquera que $\dfrac{\hat{r}_{MAP}}{\hat{r}_{MV}} \xrightarrow[n \to +\infty]{} 1$

L'estimateur MAP se comporte comme l'estimateur du maximum de vraisemblance lorsque $n \to +\infty$, ce qui est un résultat très classique

$$\boxed{\text{Partie II = Test Statistique}}$$

1) Test de Neyman Pearson (1 point)

Il est défini par Rejet de H_0 si $\dfrac{p(x_1, \dots, x_n \mid H_1)}{p(x_1, \dots, x_n \mid H_0)} > \text{Seuil}$

Soit

Rejet de H_0 si $\dfrac{\prod_{i=1}^{n} \left(\frac{\alpha + x_i}{\alpha}\right)^{r_0}}{\prod_{i=1}^{n} \left(\frac{\alpha + x_i}{\alpha}\right)^{r_1}} > S_\alpha$

i.e., si $\sum_{i=1}^{n} (r_0 - r_1) \ln\left(\frac{\alpha + x_i}{\alpha}\right) > \ln S_\alpha$

Puisque $r_0 - r_1 < 0$, on déduit

$$\boxed{\text{Rejet de } H_0 \text{ si } \sum_{i=1}^{n} \ln\left(\frac{\alpha + x_i}{\alpha}\right) < k_\alpha}$$ 1pt

2) Détermination du seuil k_α (1 point)

$\alpha = P[\text{Rejeter } H_0 \mid H_0 \text{ vraie}] = P[T_n < k_\alpha \mid H_0 \text{ vraie}]$

Soit $\alpha = P[T_n < k_\alpha \mid T_n \sim \Gamma(r_0, n)] = \displaystyle\int_{-\infty}^{k_\alpha} \frac{r_0^n}{\Gamma(n)} e^{-r_0 x} x^{n-1} dx$

On fait le changement de variables $u = r_0 x$ et on obtient

$$\alpha = \int_{-\infty}^{r_0 k_\alpha} \frac{\cancel{r_0^n}}{\Gamma(n)} e^{-u} \left(\frac{u}{\cancel{r_0}}\right)^{n-1} \frac{du}{\cancel{r_0}} = Q_n(r_0 k_\alpha)$$

d'où $\alpha = Q_n(r_0 k_\alpha)$ et donc $\boxed{k_\alpha = \frac{1}{r_0} Q_n^{-1}(\alpha)}$ 1pt

3) Puissance du test et courbes COR (2 points)

La puissance du test est $\Pi = P[\text{Rejeter } H_0 \mid H_1 \text{ vraie}] = 1-\beta$

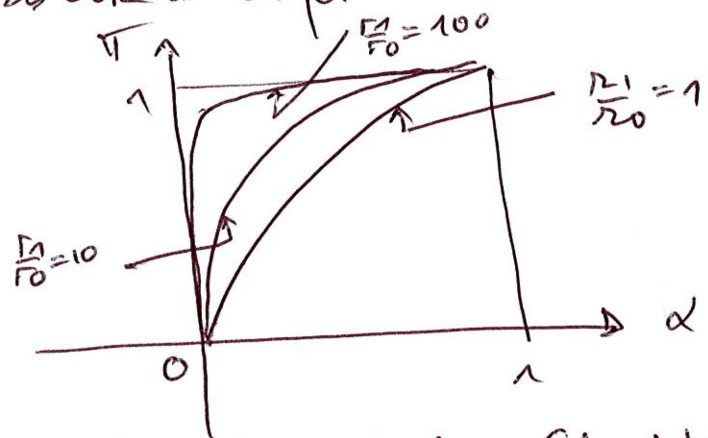
donc $\Pi = P[T_n < k_\alpha \mid T_n \sim \Gamma(r_1, n)]$

$$\boxed{\Pi = Q_n(r_1 k_\alpha)}$$

En remplaçant k_α par son expression, on obtient

$$\boxed{\Pi = Q_n\left[\frac{r_1}{r_0} Q_n^{-1}(\alpha)\right]}$$ 1pt

On remarque que Π est une fonction croissante de $\frac{r_1}{r_0}$. On a donc des courbes COR de la forme



1pt

4) Théorème de la limite centrale (4 points)

Puisque T_n est la somme de n variables aléatoires indépendantes $V_i = \ln\left(\frac{\alpha + X_i}{\alpha}\right)$ et de même loi, on peut appliquer le théorème de la limite centrale

$$T_n \underset{n \text{ grand}}{\sim} \mathcal{N}\left(nE[V_i],\ n\operatorname{Var}V_i\right)$$

Puisque T_n suit une loi gamma $\Gamma(r_j, n)$ sous l'hypothèse H_j, on a

$$E[T_n] = \frac{n}{r_j} \quad \text{et} \quad \operatorname{Var}[T_n] = \frac{n}{r_j^2}$$

Soit

$$\boxed{T_n \underset{\#}{\sim} \mathcal{N}\left(\frac{n}{r_j}, \frac{n}{r_j^2}\right) \text{ sous l'hypothèse } H_j}$$ 1pt

On en déduit

$$\alpha = P\left[T_n < K_\alpha \mid r = r_0\right) \sim P\left[\frac{T_n - n/r_0}{\underbrace{\sqrt{n/r_0^2}}_{U_n}} < \frac{K_\alpha - n/r_0}{\sqrt{n/r_0^2}} \;\middle|\; U_n \sim N(0,1)\right]$$

soit

$$\alpha = F_{N(0,1)}\left[\frac{K_\alpha - n/r_0}{\sqrt{n/r_0^2}}\right] = \boxed{F_{N(0,1)}\left[\frac{r_0 K_\alpha - n}{\sqrt{n}}\right]}$$

$\boxed{1\,pt}$

On en déduit

$$\frac{r_0 K_\alpha}{\sqrt{n}} - \sqrt{n} = F^{-1}_{N(0,1)}(\alpha)$$

ou

$$\boxed{K_\alpha = \frac{\sqrt{n}}{r_0}\left[\sqrt{n} + F^{-1}_{N(0,1)}(\alpha)\right]} \quad (*)$$

De la même manière

$$\boxed{\Pi = F_{N(0,1)}\left[\frac{r_1 K_\alpha - n}{\sqrt{n}}\right]} \qquad \boxed{1\,pt}$$

d'où, en remplaçant K_α par son expression $(*)$

$$\boxed{\Pi = F_{N(0,1)}\left[\frac{r_1}{r_0}\sqrt{n} - \sqrt{n} + \frac{r_1}{r_0} F^{-1}_{N(0,1)}(\alpha)\right]} \qquad \boxed{1\,pt}$$

On remarquera qu'à nouveau Π ne dépend de r_0 et de r_1 que via le rapport $\frac{r_1}{r_0}$.

BARÈME

8

Partie I (13 points)

1) $\hat{r}_{MV} = \dfrac{n}{\sum_{i=1}^{n} \ln\left(\frac{\alpha + X_i}{\alpha}\right)}$ 1pt

max de vraisemblance

$$\begin{array}{c|ccccc} r & 0 & & \frac{n}{\sum \ln\left(\frac{\alpha+X_i}{\alpha}\right)} & & +\infty \\ \hline L & & \nearrow & & \searrow & \end{array}$$ 1pt

2

2) $\hat{r}_{MO} = 1 + \dfrac{\alpha}{\bar{x}}$ 1pt

1

3) $\pi(v_i) = r e^{-r v_i} \mathbb{1}_{\mathbb{R}^+}(v_i)$ 1pt

$V = \sum_{i=1}^{n} v_i \sim \Gamma(r, n)$ 1pt

$\hat{r}_{MV} = \dfrac{n}{V} \sim IG(n, nr)$ 1pt

3

4) $\hat{r}^* = \dfrac{n-1}{n} \hat{r}_{MV}$ 0.5

$E[\hat{r}_{MV}] = \dfrac{nr}{n-1}$ 0.5

$var(\hat{r}^*) = \dfrac{r^2}{n-2}$ 1pt

$\hat{r}^*$ convergent 0.5

2,5

5) $BCR = \dfrac{r^2}{n}$ 1

$\hat{r}^*$ non efficace 0.5

1,5

· $r \mid x_1, \dots, x_n \sim \Gamma\left(a + \sum_{i=1}^{n} \ln\left(\frac{\alpha + x_i}{\alpha}\right), b+n\right)$ 1pt

$\hat{r}_{MAP} = \dfrac{1 + \frac{b-1}{n}}{\frac{1}{n}\sum_{k=1}^{n} \ln\left(\frac{\alpha + X_k}{\alpha}\right) + \frac{a}{n}}$ 1pt

$\dfrac{\hat{r}_{MAP}}{\hat{r}_{MV}} \xrightarrow[n \to \infty]{} 1$ 1pt

3

Partie II (8 points)

1) Rejet de H_0 si $T_n < k_\alpha$

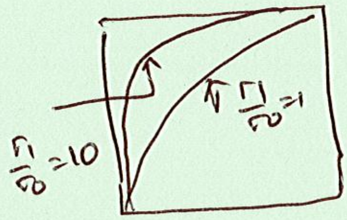
$T_n = \sum_{k=1}^{n} \ln\left(\frac{\alpha + X_i}{\alpha}\right)$

1

2) $\alpha = Q_n(r_0 k_\alpha)$

1

3) $\pi = Q_n\left[\frac{r_1}{r_0} Q_n^{-1}(\alpha)\right]$



2

4) $T_J \sim N\left(\frac{n}{r_J}, \frac{n}{r_J^2}\right)$ 1pt

$\alpha = F_{N(0,1)}\left[\dfrac{k_\alpha - \frac{n}{r_0}}{\sqrt{\frac{n}{r_0^2}}}\right]$ 1pt

$\pi = F\left[\dfrac{k_\alpha - \frac{n}{r_1}}{\sqrt{\frac{n}{r_1^2}}}\right]$ 1pt

$\pi = F\left[\frac{r_1}{r_0} F^{-1}(\alpha) + \frac{r_1}{r_0}\sqrt{n} - \sqrt{n}\right]$

1pt

4