

Université d'Orléans - Licence Economie et Gestion

Statistique Mathématique

C. Hurlin. Correction Examen Novembre 2009

Exercice 1 *Maximum de Vraisemblance.*

Question préliminaire A (0.5 point) Etant donnée la définition de la variable R , le paramètre σ correspond à l'écart type des variables X et Y , donc

$$\sigma > 0 \quad (1)$$

Question préliminaire B (1 point) Etant donnée la définition de la loi de Rayleigh, si R suit une loi *Rayleigh* (1), alors :

$$R = \sqrt{X^2 + Y^2} \quad (2)$$

où X et Y sont des v.a.r. indépendantes suivant une loi normale centrée réduite $N(0, 1)$. Dès lors, $R^2 = X^2 + Y^2$, où X^2 et Y^2 suivent des lois chi-deux à un degré de liberté. Par conséquent :

$$R^2 \sim \chi^2(2) \quad (3)$$

Question 1 (2 points) On sait que les variables $\{R_1, \dots, R_N\}$ sont *i.i.d.* de même loi que R . Dès lors, on a :

$$\begin{aligned} L(R_1, \dots, R_N; \sigma) &= \prod_{i=1}^N f_R(R_i; \sigma) \\ &= \prod_{i=1}^N \frac{R_i}{\sigma^2} \exp\left(-\frac{R_i^2}{2\sigma^2}\right) \quad 0.5 \text{ point} \\ &= \prod_{i=1}^N R_i \times \frac{1}{\sigma^{2N}} \times \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N R_i^2\right) \end{aligned} \quad (4)$$

Dès lors, la log-vraisemblance associée au N -échantillon $\{R_1, \dots, R_N\}$ s'écrit :

$$\ln L(R_1, \dots, R_N; \sigma) = \sum_{i=1}^N \log(R_i) - 2N \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^N R_i^2 \quad (1.5 \text{ point}) \quad (5)$$

Question 2 (3 points) Soit $\hat{\sigma}_{MV}$ l'estimateur du maximum de vraisemblance du paramètre σ . Ce dernier vérifie :

$$\hat{\sigma}_{MV} = \text{Arg max } L(R_1, \dots, R_N; \sigma)$$

On en déduit la CN :

$$\left. \frac{\partial \ln L(R_1, \dots, R_N; \sigma)}{\partial \sigma} \right|_{\hat{\sigma}_{MV}} = -\frac{2N}{\hat{\sigma}_{MV}} + \frac{1}{\hat{\sigma}_{MV}^3} \sum_{i=1}^N R_i^2 = 0 \quad (6)$$

D'où l'on tire que :

$$2N = \frac{1}{\hat{\sigma}_{MV}^2} \sum_{i=1}^N R_i^2$$

et finalement :

$$\hat{\sigma}_{MV} = \sqrt{\frac{1}{2N} \sum_{i=1}^N R_i^2} \quad 1.5 \text{ point} \quad (7)$$

On vérifie la CS :

$$\begin{aligned} \left. \frac{\partial^2 \ln L(R_1, \dots, R_N; \sigma)}{\partial \sigma^2} \right|_{\hat{\sigma}_{MV}} &= \frac{2N}{\hat{\sigma}_{MV}^2} - \frac{3}{\hat{\sigma}_{MV}^4} \sum_{i=1}^N R_i^2 \\ &= \frac{1}{\hat{\sigma}_{MV}^4} \left(2N \hat{\sigma}_{MV}^2 - 3 \sum_{i=1}^N R_i^2 \right) \end{aligned} \quad (8)$$

Or, nous savons que

$$\hat{\sigma}_{MV}^2 = \frac{1}{2N} \sum_{i=1}^N R_i^2 \quad (9)$$

Dès lors, il vient :

$$\begin{aligned} \left. \frac{\partial^2 \ln L(R_1, \dots, R_N; \sigma)}{\partial \sigma^2} \right|_{\hat{\sigma}_{MV}} &= \frac{1}{\hat{\sigma}_{MV}^4} \left(\sum_{i=1}^N R_i^2 - 3 \sum_{i=1}^N R_i^2 \right) \\ &= -\frac{2}{\hat{\sigma}_{MV}^4} \sum_{i=1}^N R_i^2 < 0 \quad 1.5 \text{ point} \end{aligned} \quad (10)$$

Question 3 (2 points) L'estimateur du maximum de vraisemblance $\hat{\sigma}_{MV}$ est sans biais si et seulement si $E(\hat{\sigma}_{MV}) = \sigma$ ou de façon équivalente $E(\hat{\sigma}_{MV}^2) = \sigma^2$, puisque σ ne peut pas être négatif. Or, on sait que :

$$E(\hat{\sigma}_{MV}^2) = \frac{1}{2N} \sum_{i=1}^N E(R_i^2) = \frac{1}{2N} \sum_{i=1}^N E(R^2) = \frac{E(R^2)}{2} \quad 0.5 \text{ point} \quad (11)$$

Sachant que :

$$\begin{aligned} E(R_i^2) &= V(R_i) + E(R_i)^2 \\ &= \left(\frac{4 - \pi}{2} \right) \sigma^2 - \frac{\pi}{2} \sigma^2 \\ &= 2\sigma^2 \quad 1 \text{ point} \end{aligned} \quad (12)$$

Dès lors, on a :

$$E(\hat{\sigma}_{MV}^2) = \sigma^2 \quad (13)$$

L'estimateur $\hat{\sigma}_{MV}$ est sans biais puisque σ est strictement positif, donc cette égalité implique (0.5 point) :

$$E(\hat{\sigma}_{MV}) = \sigma \quad (14)$$

Question 4 (1 point) On admet que $V(R^2) = 4\sigma^4$. Dès lors, puisque les R_i sont *i.i.d.* de même loi que R , il vient :

$$V(\hat{\sigma}_{MV}^2) = \frac{1}{4N^2} \sum_{i=1}^N V(R_i^2) = \frac{1}{4N^2} \sum_{i=1}^N V(R^2) = \frac{V(R^2)}{4N} \quad 0.5 \text{ point} \quad (15)$$

D'où l'on tire que :

$$V(\hat{\sigma}_{MV}^2) = \frac{\sigma^4}{N}$$

Question 5 (1 point) Par conséquent :

$$\lim V(\hat{\sigma}_{MV}^2) = \lim \frac{\sigma^4}{N} = 0 \quad \text{lorsque } N \rightarrow \infty \quad 0.5 \text{ point} \quad (16)$$

Sachant que par ailleurs $E(\hat{\sigma}_{MV}^2) = \sigma^2$, on en déduit que $\hat{\sigma}_{MV}^2$ est un estimateur convergent de σ^2 :

$$\hat{\sigma}_{MV}^2 \xrightarrow[N \rightarrow \infty]{p} \sigma^2 \quad 0.5 \text{ point} \quad (17)$$

D'après le théorème de Slutsky (0.5 point), soit $\{X_T\}_{T=1}^\infty$ une suite de $(n, 1)$ vecteurs admettant une limite en probabilité définie par c , et soit $g(\cdot)$ une fonction continue en c , satisfaisant $g: R^n \rightarrow R^n$, et ne dépendant pas de T , alors :

$$g(X_T) \xrightarrow[T \rightarrow \infty]{p} g(c) \quad (18)$$

Ici en posant $g(x) = \sqrt{x}$, il vient immédiatement que :

$$\hat{\sigma}_{MV} \xrightarrow[N \rightarrow \infty]{p} \sigma \quad (19)$$

L'estimateur $\hat{\sigma}_{MV}$ est convergent en probabilité.

Question 6 (1.5 points) On admet que :

$$\frac{\partial^2 \ln L(R_1, \dots, R_N; \sigma)}{\partial \sigma^4} = \frac{N}{\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^N R_i^2 \quad (20)$$

Dès la borne la quantité d'information Fisher vaut :

$$\begin{aligned} I_N(\sigma^2) &= E \left[-\frac{\partial^2 \ln L(R_1, \dots, R_N; \sigma)}{\partial \sigma^4} \right] \\ &= -\frac{N}{\sigma^4} + \frac{1}{\sigma^6} \sum_{i=1}^N E(R_i^2) \\ &= -\frac{N}{\sigma^4} + \frac{2N\sigma^2}{\sigma^6} \\ &= \frac{N}{\sigma^4} \quad 1 \text{ point} \end{aligned} \quad (21)$$

Dès lors,

$$V(\hat{\sigma}_{MV}^2) = \frac{1}{I_N(\sigma^2)} = \frac{\sigma^4}{N} \quad 0.5 \text{ point}$$

L'estimateur du maximum de vraisemblance $\hat{\sigma}_{MV}^2$ est efficace au sens de la borne FDCR

Question 7 (1 point) La loi asymptotique de l'estimateur $\hat{\sigma}_{MV}^2$ est la suivante :

$$\sqrt{N} (\hat{\sigma}_{MV}^2 - \sigma^2) \xrightarrow{L} N \left(0, \frac{\sigma^4}{N} \right) \quad (22)$$

=====

Exercice 2 *Vrai Faux.*

Toute réponse non justifiée sera considérée comme fausse. *Attention : certaines propositions contenues dans les énoncés peuvent être incomplètes.*

Question 1 (0.5 point) FAUX. Sous les conditions de régularités vues en cours, l'estimateur du MV \hat{v} est asymptotiquement distribué selon une loi normale

Question 2 (0.5 point) FAUX La fonction de densité d'une v.a.r. X distribuée selon une loi Student à v degrés de liberté est en effet leptokurtique, mais son degré de kurtosis décroît avec v . En effet, lorsque v tend vers l'infini, la distribution de Student converge vers la distribution normale, qui est une distribution mesokurtique.

Question 3 (0.5 point) VRAI Soit $\{X_1, X_2, \dots, X_N\}$ un N -échantillon de variables aléatoires *i.i.d.* de même loi que X , où $X \sim N(m, \sigma^2)$, alors on sait que S_N^2 , la variance empirique corrigée, vérifie :

$$S_N^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X}_N)^2 \quad (23)$$

avec :

$$\left(\frac{1}{N-1} \right) \frac{S_N^2}{\sigma^2} \sim \chi^2(N-1) \quad (24)$$

Dès lors, on sait que :

$$V \left[(N-1) \frac{S_N^2}{\sigma^2} \right] = 2(N-1) \quad (25)$$

$$\iff \frac{(N-1)^2}{\sigma^4} V(S_N^2) = 2(N-1) \quad (26)$$

D'où l'on tire que:

$$V(S_N^2) = \frac{2\sigma^4}{(N-1)} \quad (27)$$

Question 4 (0.5 point) FAUX. S'il existe un estimateur efficace alors celui-ci correspond à l'estimateur du maximum de vraisemblance. On peut donc concevoir un estimateur du maximum de vraisemblance non efficace, voir même biaisé (cf. exercices TD).

Question 5 (0.5 point) VRAI. D'après le théorème central limite, lorsque N tend vers l'infini, on sait que :

$$\sqrt{N} (\bar{X}_N - E(X)) \xrightarrow[N \rightarrow \infty]{L} N(0, Var(X)) \quad (28)$$

Si X suit une loi du chi-deux à K degrés de libertés, on a $E(X) = K$ et $Var(X) = 2K$, dès lors :

$$\sqrt{N} (\bar{X}_N - K) \xrightarrow[N \rightarrow \infty]{L} N(0, 2K) \quad (29)$$

Question 6 (0.5 point) FAUX. Si X et Y sont deux v.a.r. indépendantes distribuées selon une loi normale $N(0, 1)$, alors $X + Y$ suit une loi normale $N(0, 2)$. Dès lors, on a :

$$\Pr(X + Y \leq 1.8124) = 0.90 \quad (30)$$

Question 7 (0.5 point) VRAI. Si X et Y sont deux v.a.r. indépendantes telles que X admette une distribution normale $N(0, 1)$ et Y une distribution du chi-deux à 9 degrés de liberté alors la variable $X/\sqrt{Y/9} = 3X/\sqrt{Y}$ suit une loi de Student à 9 degrés de liberté :

$$\Pr\left(3\frac{X}{\sqrt{Y}} \leq 1.3830\right) = 0.90$$

Question 8 (0.5 point) VRAI. En effet, si X et Y sont deux variables aléatoires indépendantes où X suit une loi du chi-deux à v_1 degrés de liberté et Y suit une loi du chi-deux à v_2 degrés de liberté, alors la variable Z vérifie :

$$Z = \frac{X/v_1}{Y/v_2} = \frac{v_2 X}{v_1 Y} \sim F(v_1, v_2) \quad (31)$$

où $F(v_1, v_2)$ désigne la loi de Fisher-Snedecor à v_1 et v_2 degrés de liberté.

=====

Exercice 3 *Intervalle de confiance. Barème : 4 points*

On veut estimer la moyenne m d'une variable aléatoire X suivant une loi normale de variance connue $\sigma^2 = 6.25$ à l'aide d'un échantillon de taille $n = 10$ valeurs indépendantes. La moyenne \bar{X} de l'échantillon observé est égale à 4.3. Pour plus de détails.cf. cours.

Question 1 (2 points) On sait que $\Phi^{-1}(1 - 0.15/2) = 1.4395$. Dès lors, on a :

$$\begin{aligned} IC &= \left[4.3 \pm \Phi^{-1}(1 - 0.15/2) \sqrt{\frac{6.25}{100}} \right] \\ &= [3.1620; 5.4380] \end{aligned}$$

Question 2 (2 points) On sait que $F^{-1}(1 - 0.15/2; 9) = 1.5737$, dès lors on a :

$$\begin{aligned} IC &= \left[4.3 \pm F^{-1}(1 - 0.15/2; 9) \sqrt{\frac{6.25}{100}} \right] \\ &= [3.0559; 5.5441] \end{aligned}$$