Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées

IS Math314 **Année 2004-2005**

Examen, 20 mai 2005, durée 3 heures.

- Ce sujet comporte 3 pages.
- Documents autorisés : polycopié du cours IPE, polycopié Simulation, notes manuscrites du cours d'IS.
- Calculatrices autorisées.

Ex 1. *Estimation du paramètre de position de la loi log-normale (8 points)*

On rappelle qu'une variable aléatoire Y suit la loi gaussienne $\mathfrak{N}(m,\sigma)$, de paramètres $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$, si elle admet pour densité la fonction

$$g : \mathbb{R} \to \mathbb{R}, \quad y \longmapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{(y-m)^2}{2\sigma^2}\Big).$$

On sait qu'alors Y admet des moments de tout ordre et que

$$\mathbf{E}Y = m, \quad \operatorname{Var} Y = \sigma^2.$$

Ces affirmations préliminaires sur les lois gaussiennes, n'ont pas à être redémontrées dans les copies.

1) Exprimez $\mathbf{E}(Y^2)$ en fonction des paramètres m et σ.

2) On dit qu'une variable aléatoire X suit la loi log-normale de paramètres m et σ, avec $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}_+^*$, si elle admet pour densité la fonction

$$f : \mathbb{R} \to \mathbb{R}, \quad x \longmapsto \frac{1}{x\sigma\sqrt{2\pi}} \exp\Big(-\frac{(\ln x-m)^2}{2\sigma^2}\Big)\mathbf{1}_{]0,+\infty[}(x).$$

La fonction f est donc nulle sur $\mathbb{R}_-$ et strictement positive sur $\mathbb{R}_+^*$. Vérifiez à l'aide d'un changement de variable que f est bien une densité de probabilité sur $\mathbb{R}$.

3) Vérifiez de même que la variable aléatoire $\ln X$ a des moments de tout ordre et que

$$\mathbf{E}(\ln X) = m, \quad \operatorname{Var}(\ln X) = \sigma^2.$$

4) On se propose dans la suite de l'exercice d'étudier l'estimation du paramètre m. Pour cela on dispose d'un échantillon *observé* $X_1(\omega), \dots, X_n(\omega)$ avec n grand. On considère donc que les variables aléatoires X_i utilisées ici sont indépendantes et de même loi que X et ceci, quelles que soient les valeurs prises par les paramètres m et σ. En utilisant la question précédente, proposez en justifiant votre réponse, un estimateur T_n de m qui soit *fortement consistant*, *i.e.* presque-sûrement convergent vers m lorsque n tend vers $+\infty$.

5) Expliquez brièvement comment construire à partir de la valeur observée $T_n(\omega)$ un intervalle de confiance au niveau $1-\varepsilon$ pour m, en distinguant les deux cas σ connu, puis σ inconnu. Indiquez quels sont les théorèmes limites utilisés pour cela.

6) On suppose que σ est connu et m inconnu. Déterminez l'estimateur de m par la méthode du maximum de vraisemblance.

Les deux exercices suivants sont liés. On pourra traiter l'exercice 3 en admettant les résultats de l'exercice 2.

Ex 2. *Estimation de quantiles (7 points)*

Soit F une fonction de répartition continue et strictement croissante sur $\mathbb{R}$. On sait qu'elle réalise alors une bijection de $\mathbb{R}$ sur $]0,1[$ et que son inverse (au sens classique) F^{-1} est *continue* sur $]0,1[$ et strictement croissante sur cet intervalle. Pour tout $u \in]0,1[$, on définit le quantile associé q par

$$q := F^{-1}(u).$$

On se propose d'estimer q pour *u fixé* à partir d'une suite de variables aléatoires $(X_k)_{k\geq 1}$, définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$, indépendantes et de même loi de fonction de répartition F.

On rappelle que la fonction de répartition empirique F_n construite sur l'échantillon $X_1, \dots, X_n$ est définie par

$$\forall \omega \in \Omega, \quad t \mapsto F_n(\omega, t) := \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{]-\infty, X_k(\omega)]}(t)$$

et que pour tout $\omega \in \Omega$, $F_n(\omega, .)$ est une fonction de répartition donc en particulier *croissante* au sens large sur $\mathbb{R}$ et *continue à droite* en tout point de $\mathbb{R}$. Pour t fixé dans $\mathbb{R}$, l'application partielle $F_n(., t)$, notée plus simplement $F_n(t)$ et définie par $F_n(t) : \Omega \to [0,1]$, $\omega \mapsto F_n(\omega, t)$ est une variable aléatoire.

Pour estimer q, on introduit la suite de variables aléatoires Q_n définies par

$$Q_n = F_n^{-1}(u) := \inf\{t \in \mathbb{R};\ F_n(t) \geq u\}.$$

1) Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires et Y une variable aléatoire, définies sur le même $(\Omega, \mathcal{F}, P)$. On suppose que Y_n converge presque-sûrement vers Y quand n tend vers $+\infty$ et que $g : \mathbb{R} \to \mathbb{R}$ est une fonction continue sur $\mathbb{R}$. Expliquez pourquoi $g(Y_n)$ converge p.s. vers $g(Y)$. Justifiez de même la convergence p.s. de $g(Y_n)$ vers $g(c)$ lorsque Y_n converge p.s. vers la constante c et g est une fonction borélienne [1] $\mathbb{R} \to \mathbb{R}$, continue au point c.

2) Justifiez les inégalités entre variables aléatoires (*i.e.* vraies pour tout $\omega \in \Omega$) :

$$\forall n \geq 1, \quad F_n(Q_n) \geq u, \quad \text{et} \quad F_n\Big(Q_n - \frac{1}{n}\Big) < u.$$

1. On est ainsi assurés que $g(Y_n)$ est une v.a.

3) On définit la variable aléatoire positive Δ_n par

$$\forall \omega \in \Omega, \quad \Delta_n(\omega) := \|F_n(\omega, .) - F\|_\infty = \sup_{t\in\mathbb{R}} |F_n(\omega, t) - F(t)|.$$

Montrez que pour tout $n \geq 1$,

$$F(Q_n) \geq u - \Delta_n \quad \text{et} \quad F\Big(Q_n - \frac{1}{n}\Big) < u + \Delta_n$$

et en déduire l'encadrement (vrai pour tout $n \geq 1$ et tout $\omega \in \Omega$) :

$$F^{-1}(u - \Delta_n) \leq Q_n \leq \frac{1}{n} + F^{-1}(u + \Delta_n).$$

4) Expliquez pourquoi $F^{-1}(u - \Delta_n)$ converge presque-sûrement vers $q = F^{-1}(u)$ lorsque n tend vers l'infini.

5) Montrez que Q_n converge p.s. vers q lorsque n tend vers l'infini.

Ex 3. *Estimation des paramètres d'une loi de Cauchy (5 points)*

La loi de Cauchy $\mathrm{Cau}(a, b)$ de paramètres $a \in \mathbb{R}$ et $b \in \mathbb{R}_+^*$ a pour densité

$$f_{a,b} : \mathbb{R} \to \mathbb{R}, \quad x \longmapsto \frac{1}{\pi b \Big(1 + \big(\frac{x-a}{b}\big)^2\Big)}.$$

1) Vérifiez que la fonction de répartition $F_{a,b}$ de la loi $\mathrm{Cau}(a, b)$ est donnée par

$$\forall t \in \mathbb{R}, \quad F_{a,b}(t) = \frac{1}{\pi} \arctan\left(\frac{t-a}{b}\right) + \frac{1}{2}.$$

En déduire que si Y suit la loi $\mathrm{Cau}(0, 1)$, pour tout réel a et tout réel $b > 0$, la v.a. $X := a + bY$ suit la loi $\mathrm{Cau}(a, b)$.

2) On se propose d'estimer a et b au vu d'un échantillon observé $X_1(\omega), \dots, X_n(\omega)$ avec n grand. On considère donc que les variables aléatoires X_i utilisées ici sont indépendantes et de même loi $\mathrm{Cau}(a, b)$, les paramètres a et b étant inconnus. Comme a est un paramètre de position, on envisage son estimation par la moyenne empirique

$$M_n := \frac{1}{n} \sum_{k=1}^{n} X_k.$$

Cet estimateur est-il fortement consistant ?

3) Expliquez brièvement pourquoi $F_{a,b}$ est continue sur $\mathbb{R}$ et strictement croissante sur $\mathbb{R}$. Elle admet donc un inverse au sens classique $F_{a,b}^{-1} :]0, 1[\to \mathbb{R}$. On définit alors les *quartiles* q_1, q_2, q_3 de la loi $\mathrm{Cau}(a, b)$ par

$$q_i := F_{a,b}^{-1}\Big(\frac{i}{4}\Big), \quad \text{pour } i = 1, 2, 3.$$

Le deuxième quartile q_2 est la *médiane* et $q_3 - q_1$ est appelé *écart inter-quartiles*. Ce paramètre donne une idée de la dispersion de la loi autour de sa médiane. Calculez les quartiles et l'écart inter-quartiles de la loi $\mathrm{Cau}(a, b)$ en fonction des paramètres a et b.

4) En utilisant le résultat de l'exercice 2, proposez des estimateurs fortement consistants des paramètres a et b.