Statistiques appliquées - L3 d'Economie
Interrogation écrite N°2 - Groupe 9 - Corrigé

Marc SANGNIER - marc.sangnier@ens-cachan.fr

22 décembre 2007

Exercice 1

Soit une variable aléatoire réelle continue X de densité $f(x) = \frac{2}{\theta}\left(1 - \frac{x}{\theta}\right)$ si $0 \leq x \leq \theta$, $f(x) = 0$ sinon. On dispose d'un échantillon de n observations x_i de la variable aléatoire X.

Question 1

$$E(X) = \int_{-\infty}^{+\infty} x * f(x)dx = \int_0^\theta x\frac{2}{\theta}\left(1 - \frac{x}{\theta}\right)dx = \frac{2}{\theta}\int_0^\theta x - \frac{x^2}{\theta}dx$$

$$\Longleftrightarrow E(X) = \frac{2}{\theta}\left[\frac{x^2}{2} - \frac{x^3}{3\theta}\right]_0^\theta = \frac{2}{\theta}\left(\frac{\theta^2}{2} - \frac{\theta^3}{3\theta}\right) = \frac{2}{\theta}\left(\frac{3\theta^2 - 2\theta^2}{3 * 2}\right) = \frac{\theta}{3}$$

Question 2

L'estimation par la méthode des moments consiste en la résolution d'équations issues de l'association des moments théoriques d'une variable aléatoire avec les moments empiriques des observations de cette variable.

Question 3

Pour estimer le paramètre θ par la méthode des moments, on associe le moment empirique d'ordre 1 (la moyenne empirique) à l'espérance de la variable X.

$$E(X) = \bar{X} \Longleftrightarrow \frac{\theta}{3} = \frac{1}{n}\sum_{i=1}^n x_i \Longleftrightarrow \hat{\theta} = 3 * \frac{1}{n}\sum_{i=1}^n x_i$$

Exercice 2

Soit une variable aléatoire réelle continue X qui suit une loi Normale d'espérance m et de variance σ^2. Sa densité s'écrit :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

On dispose d'un échantillon de n observations x_i de la variable aléatoire X. On suppose que ces observations sont indépendantes.

Question 1

La loi Normale est utile pour modèliser des valeurs dont il n'y a aucune raison de penser a priori qu'il y en a davantage au dessus de la moyenne qu'en dessous de la moyenne. Ceci vient de la symètrie de la loi Normale par rapport à sa moyenne.

Question 2

La fonction de vraisemblance de l'échantillon $\{x_1; x_2; ...; x_n\}$ est :

$$L(x_1; x_2; ...; x_n; m; \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x_i-m)^2}{2\sigma^2}\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^n exp\left(-\frac{(x_i-m)^2}{2\sigma^2}\right)$$

Question 3

La fonction de log-vraisemblance de cet échantillon est :

$$log(L(x_1; x_2; ...; x_n; m; \sigma^2)) = ln\left[\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} exp\left(-\frac{(x_i - m)^2}{2\sigma^2}\right)\right] = n * ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \sum_{i=1}^{n} \frac{(x_i - m)^2}{2\sigma^2}$$

$$\iff log(L(x_1; x_2; ...; x_n; m; \sigma^2)) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$$

Question 4

Estimons la variance σ^2 par la méthode du maximum de vraisemblance.

$$\frac{\partial log\left[L(m; \sigma^2; x_1; x_2; ...; x_n)\right]}{\partial \sigma^2} = 0$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - m)^2 = 0 \iff \frac{1}{2\sigma^2}\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m)^2 = \frac{1}{2\sigma^2}n \iff \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m)^2 = n$$

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$$

Question 5

$\hat{\sigma^2}$ dépend du paramètre inconnu m.

Question 6

Estimons l'espérance m par la méthode du maximum de vraisemblance.

$$\frac{\partial log\left[L(m; \sigma^2; x_1; x_2; ...; x_n)\right]}{\partial m} = 0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - m) = 0 \iff \sum_{i=1}^{n}(x_i - m) = 0 \iff n * m = \sum_{i=1}^{n} x_i \iff \hat{m} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Question 7

L'estimateur $\hat{m}$ suit une loi Normale d'espérance m et de variance $\frac{\sigma^2}{n}$.

Question 8

Un estimateur est dit convergent si il est sans biais et si sa variance tend vers 0 lorsque n tend vers l'infini.

$$E(\hat{m}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}n * m = m$$

$\hat{m}$ est donc sans biais.

$$V(\hat{m}) = \frac{\sigma^2}{n}$$

On voit immédiatement que lorsque n tend vers l'infini, $V(\hat{m})$ tend vers 0.
$\hat{m}$ est donc convergent.

Question 9

Construisons analytiquement un intervalle de confiance à 95% pour la vraie valeur de m.

$$1-0,05 = P\left(\hat{m}-a < m < \hat{m}+a\right) = P\left(-a < m-\hat{m} < a\right) = P(-a < \hat{m}-m < a) = P\left(\frac{-a}{\sqrt{\frac{\sigma^2}{n}}} < \frac{\hat{m}-m}{\sqrt{\frac{\sigma^2}{n}}} < \frac{a}{\sqrt{\frac{\sigma^2}{n}}}\right)$$

Or, $E(\hat{m}) = m$ et $V(\hat{m}) = \sigma^2$. $\frac{\hat{m}-m}{\sqrt{\frac{\sigma^2}{n}}}$ suit donc une loi Normale centrée-réduite. Notons $F(.)$ la fonction de répartition de la loi Normale centrée-réduite.

$$1-0,05 = F\left(\frac{a}{\sqrt{\frac{\sigma^2}{n}}}\right) - F\left(\frac{-a}{\sqrt{\frac{\sigma^2}{n}}}\right) = F\left(\frac{a}{\sqrt{\frac{\sigma^2}{n}}}\right) - 1 + F\left(\frac{a}{\sqrt{\frac{\sigma^2}{n}}}\right) = 2F\left(\frac{a}{\sqrt{\frac{\sigma^2}{n}}}\right) - 1$$

$$F\left(\frac{a}{\sqrt{\frac{\sigma^2}{n}}}\right) = \frac{2-0,05}{2} = 0,975$$

L'intervalle de confiance à 95% pour m est donc :

$$[\hat{m}-a; \hat{m}+a]$$

Avec a tel que $F\left(\frac{a}{\sqrt{\frac{\sigma^2}{n}}}\right) = 0,975$

Question 10

On dispose d'une réalisation de l'échantillon telle que $\hat{m} = 100$ et $\sqrt{\frac{\sigma^2}{n}} = 10$. On sait par ailleurs que pour une loi Normale centrée-réduite de fonction de répartition $F(.)$: $F(u) = 0,975 \Longleftrightarrow u = 1,96$. Donnez l'expression numérique de l'intervalle de confiance à 95% pour la vraie valeur de m.

On en déduit :

$$\frac{a}{\sqrt{\frac{\sigma^2}{n}}} = 1,96 \Longleftrightarrow a = 1,96 * \sqrt{\frac{\sigma^2}{n}} \Longleftrightarrow a = 1,96 * 10 = 19,6$$

D'où l'intervalle de confianceà 95% pour m :

$$[100-19,6; 100+19,6] = [80,4; 119,6]$$

Exercice 3

Soit la fonction de production Cobb-Douglas suivante :

$$Y = AK^{\alpha}L^{\beta} \tag{1}$$

Dans l'équation (1), Y représente la quantité de biens produite, A est un réel non nul, K le capital (nombre de machines) et L le travail (nombre d'ouvriers).

Question 1

Utilisons une transformation logarithmique :

$$y = a + \alpha k + \beta l$$

Avec $log(A) = a$, $log(Y) = y$, $log(L) = l$ et $log(K) = k$.

Question 2

Le modèle économètrique permettrait d'estimer les paramètres a, α et β par la méthodes des moindres carrés est le suivant :

$$y_i = a + \alpha k_i + \beta l_i + \varepsilon_i \tag{2}$$

Dans l'équation (2), y_i représente le logarithme de la quantité de biens produits par l'entreprise i, k_i le logarithme du nombre de machines utilisées par l'entreprise i, l_i le logarithme du nombre de travailleurs employés par l'entrepise i et ε_i rassemble l'ensemble des facteurs non obersvés ou non mesurés propres à l'entrprise i qui peuvent avoir un impact sur la quantité de bien qu'elle produit.

Question 3

On suppose que l'estimation du modèle économètrique nous a permis d'obtenir l'expression suivante, où $\hat{a}$, $\hat{\alpha}$ et $\hat{\beta}$ sont les estimateurs de a, α et β :

$$y = \hat{a} + \hat{\alpha}k + \hat{\beta}l \tag{3}$$

La différentielle totale de l'équation (3) est :

$$dy = \hat{\alpha}dk + \hat{\beta}dl \Longleftrightarrow d(log(Y)) = \hat{\alpha}d(log(K)) + \hat{\beta}d(log(L)) \Longleftrightarrow \frac{dY}{Y} = \hat{\alpha}\frac{dK}{K} + \hat{\beta}\frac{dL}{L}$$

Avec Y la quantité de biens produite, K le nombre de machines utilisées et L le nombre de travailleurs employés.

Question 4

$\hat{\alpha}$ s'interprète de la façon suivante : si le nombre de machines utilisées augmente de 1%, alors la quantité de biens produite augmente de $\hat{\alpha}$%; si le nombre de machines utilisées augmente de u%, alors la quantité de biens produite augmente de $u * \hat{\alpha}$%.

$\hat{\beta}$ s'interprète de façon similaire.