

Université d'Orléans - Licence Economie et Gestion

Statistique Mathématique

C. Hurlin. Examen Novembre 2006

Exercice 1

Question 1 (1 point) Les variables X_i sont *i.i.d.* de même loi que X où X suit une loi de Bernoulli $B(1, p)$ à chaque individu sondé une variable aléatoire X_i pour $i = 1, \dots, N$, telle que :

$$X = \begin{cases} 1 & \Pr[X = 1] = p \\ 0 & \Pr[X = 0] = 1 - p \end{cases} \quad (1)$$

ou de façon générale :

$$\Pr[X = x] = p^x (1 - p)^{1-x} \quad (2)$$

Question 2 (2 points) Les variables X_i étant indépendantes, la vraisemblance de l'échantillon s'écrit :

$$L(x_1, \dots, x_n, p) = p^{\sum x_i} (1 - p)^{N - \sum x_i} \quad (3)$$

et la log-vraisemblance :

$$\log L(x_1, \dots, x_n, p) = \log(p) \sum x_i + \log(1 - p) (N - \sum x_i) \quad (4)$$

On en déduit les équations de vraisemblance :

$$\left. \frac{\partial \log L(x_1, \dots, x_n, p)}{\partial p} \right|_{p=\hat{p}} = \frac{\sum x_i}{\hat{p}} - \frac{(N - \sum x_i)}{1 - \hat{p}} = 0 \quad (5)$$

On en déduit :

$$\hat{p} = \frac{1}{N} \sum_{i=1}^N X_i \quad (6)$$

On vérifie que \hat{p} est un maximum puisque (deuxième équation de vraisemblance) :

$$\left. \frac{\partial^2 \log L(x_1, \dots, x_n, p)}{\partial p^2} \right|_{p=\hat{p}} = -\frac{\sum x_i}{\hat{p}^2} - \frac{(N - \sum x_i)}{(1 - \hat{p})^2} < 0 \quad (7)$$

Question 3 (2 points) On montre que :

$$E(\hat{p}) = \frac{1}{N} E\left(\sum_{i=1}^N X_i\right) = \frac{1}{N} \sum_{i=1}^N E(X_i) \quad (8)$$

puisque les X_i sont *i.i.d.* de même loi que X :

$$E(\hat{p}) = \frac{1}{N} \sum_{i=1}^N E(X) = E(X) = p \quad (1 \text{ point}) \quad (9)$$

L'estimateur du MV est sans biais (1 point). De plus, on sait que :

$$Var(\hat{p}) = \frac{1}{N^2} Var\left(\sum_{i=1}^N X_i\right) \quad (10)$$

Les variables X_i étant indépendantes et de même loi que X , on a donc :

$$Var(\hat{p}) = \frac{1}{N^2} \left(\sum_{i=1}^N Var(X_i) \right) = \frac{1}{N^2} \left(\sum_{i=1}^N Var(X) \right) = \frac{Var(X)}{N} = \frac{p(1-p)}{N} \quad (11)$$

Ainsi on a :

$$E(\hat{p}) = p \quad (12)$$

$$\lim_{N \rightarrow \infty} Var(\hat{p}) = \lim_{N \rightarrow \infty} \frac{p(1-p)}{N} = 0 \quad (13)$$

Donc, on peut en conclure que l'estimateur du MV est convergent :

$$\hat{p} \xrightarrow[N \rightarrow \infty]{p} p \quad (1 \text{ point}) \quad (14)$$

Question 4 (1.5 points) Déterminons la quantité d'information de Fisher associée à l'échantillon $\{X_1, \dots, X_N\}$:

$$\begin{aligned} I_N(p) &= -E \left[\frac{\partial^2 \log L(X_1, \dots, X_n, p)}{\partial p^2} \right] \\ &= E \left[\frac{\sum X_i}{p^2} + \frac{(N - \sum X_i)}{(1-p)^2} \right] \\ &= \frac{1}{p^2} \sum E(X_i) + \frac{N - \sum E(X_i)}{(1-p)^2} \\ &= \frac{Np}{p^2} + \frac{N - Np}{(1-p)^2} \\ &= \frac{N}{p(1-p)} \end{aligned} \quad (15)$$

Par conséquent, on a :

$$Var(\hat{p}) = \frac{1}{I_N(p)} \quad (16)$$

L'estimateur du MV est efficace au sens de la borne FDCR.

Question 5 (1 point) On sait que :

$$\sqrt{N}(\hat{p} - p) \xrightarrow[N \rightarrow \infty]{d} N \left(0, \frac{1}{I_N(p)} \right) \quad (17)$$

Donc

$$\sqrt{N}(\hat{p} - p) \xrightarrow[N \rightarrow \infty]{d} N \left(0, \frac{p(1-p)}{N} \right) \quad (18)$$

Remarque : On retrouverait le même résultat par approximation de la loi Binomiale.

Question 6 (1 point) Estimation ponctuelle :

$$\hat{p} = \frac{1}{N} \sum_{i=1}^N X_i = \frac{98}{200} = 0.49 \quad (19)$$

Question 7 (1 point) Si $N > 30$, on peut approximer la loi de la fréquence empirique $\hat{p} = \frac{\sum_{i=1}^N X_i}{N}$ par une loi normale :

$$\hat{p} \rightsquigarrow N\left(0, \frac{p(1-p)}{N}\right) \quad (20)$$

On en déduit l'intervalle de confiance au seuil de $\alpha\%$ suivant la formule :

$$\Pr\left[\hat{p} - \Phi^{-1}(1 - \alpha/2) \sqrt{\frac{p(1-p)}{N}}; \hat{p} + \Phi^{-1}(1 - \alpha/2) \sqrt{\frac{p(1-p)}{N}}\right] = 1 - \alpha \quad (21)$$

où $\Phi(\cdot)$ désigne la fonction de répartition de la loi normale centrée réduite. On a donc

$$\alpha = 5\% \quad IC_5 = [0.42; 0.56] \quad (22)$$

$$\alpha = 2\% \quad IC_5 = [0.41; 0.57] \quad (23)$$

Le deuxième intervalle est moins précis puisque le niveau de risque est plus élevé.

Question 8 (1.5 point) Le candidat est élu (en prévision) si la probabilité que l'on vote pour lui est supérieure à 0.5. On cherche N tel que :

$$\Pr[\hat{p} > 0.5] = 0.95 \quad (24)$$

$$\iff \Pr[\hat{p} \leq 0.5] = 0.05 \quad (25)$$

$$\iff \Pr\left[\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} \leq \frac{0.5 - p}{\sqrt{\frac{p(1-p)}{N}}}\right] = 0.05 \quad (26)$$

$$\iff \Phi\left(\frac{0.5 - p}{\sqrt{\frac{p(1-p)}{N}}}\right) = 0.05 \quad (27)$$

$$\iff \frac{0.5 - p}{\sqrt{\frac{p(1-p)}{N}}} = \Phi^{-1}(0.05) \quad (28)$$

D'où l'on tire que pour $p = 0.52$; on a $N = 1689$

Exercice 2 *Vrai - Faux Barème : 5 points.*

Toute réponse non justifiée sera considérée comme fautive.

Question 1 FAUX. La loi normale étant symétrique, son moment d'ordre 3 centré (skewness) est nul.

$$\mu_3 = E[(X - E(X))^3] = 0 \quad (29)$$

Rappelons que la fonction génératrice de moments centrés de la loi normale est :

$$\mu(r) = E[(X - m)^r] = \begin{cases} \sigma^r 2^{\frac{r}{2}} \frac{\Gamma(\frac{r+1}{2})}{\Gamma(\frac{1}{2})} & r \text{ pair} \\ 0 & r \text{ impair} \end{cases} \quad (30)$$

Question 2 (0.5 point) FAUX, car pour cela il faut que les variables X et Y soient indépendantes pour que la variable Z suivent une loi de Student.

Question 3 (0.5 point) FAUX Soit $\{X_1, X_2, \dots, X_N\}$ un N -échantillon de variables aléatoires *i.i.d.* de même loi que X , où X suit une loi du chi-deux à K degrés de liberté. On sait que

$$E(X) = K \quad \text{Var}(X) = 2K$$

Donc si les variables X_i sont indépendantes et de même loi que, alors :

$$E\left(\sum_{i=1}^N X_i\right) = N K \quad \text{Var}\left(\sum_{i=1}^N X_i\right) = 2 N K \quad (31)$$

Question 4 (0.5 point) FAUX La loi faible des grands nombres implique que

$$\bar{X}_N \xrightarrow[N \rightarrow \infty]{p} \mu \quad (32)$$

où p désigne la convergence en probabilité. Mais la convergence en probabilité n'implique pas la convergence en moyenne quadratique.

Question 5 (1.5 point) VRAI. Calculons tout d'abord l'espérance de la v.a.r. S_N .

$$E(S_N) = \frac{1}{N} E\left[\sum_{i=1}^N (X_i - \bar{X}_N)^2\right] \quad (33)$$

Considérons la décomposition suivante :

$$\begin{aligned} \sum_{i=1}^N (X_i - \bar{X}_N)^2 &= \sum_{i=1}^N (X_i - m + m - \bar{X}_N)^2 \\ &= \sum_{i=1}^N [(X_i - m) - (\bar{X}_N - m)]^2 \\ &= \sum_{i=1}^N (X_i - m)^2 + N (\bar{X}_N - m)^2 - 2 (\bar{X}_N - m) \sum_{i=1}^N (X_i - m) \end{aligned}$$

Or, on rappelle que $\bar{X}_N = \sum_{i=1}^N X_i / N$. Dès lors,

$$\begin{aligned} 2 (\bar{X}_N - m) \sum_{i=1}^N (X_i - m) &= 2 (\bar{X}_N - m) \left(\sum_{i=1}^N X_i - Nm \right) \\ &= 2 (\bar{X}_N - m) (N \bar{X}_N - Nm) \\ &= 2N (\bar{X}_N - m)^2 \end{aligned}$$

On obtient ainsi

$$\begin{aligned} \sum_{i=1}^N (X_i - \bar{X}_N)^2 &= \sum_{i=1}^N (X_i - m)^2 + N (\bar{X}_N - m)^2 - 2N (\bar{X}_N - m)^2 \\ &= \sum_{i=1}^N (X_i - m)^2 - N (\bar{X}_N - m)^2 \end{aligned} \quad (34)$$

Par conséquent, on obtient en espérance :

$$E(S_N) = \frac{1}{n} E \left[\sum_{i=1}^n (X_i - m)^2 \right] - \frac{1}{n} E \left[n (\bar{X}_n - m)^2 \right] \quad (35)$$

$$= \frac{1}{n} E \left[\sum_{i=1}^n (X_i - m)^2 \right] - E \left[(\bar{X}_n - m)^2 \right] \quad (36)$$

On reconnaît dans le terme de gauche la définition de la variance de la variable X_i . Etant donnée que les variables X_i sont indépendantes et de même loi que X , on a :

$$E \left[\sum_{i=1}^N (X_i - m)^2 \right] = \sum_{i=1}^N E \left[(X_i - m)^2 \right] = \sum_{i=1}^N E \left[(X - m)^2 \right] = \sum_{i=1}^N \text{var}(X) = N \sigma^2 \quad (37)$$

De la même façon, on sait que :

$$E \left[(\bar{X}_n - m)^2 \right] = E \left[(\bar{X}_N - E(\bar{X}_N))^2 \right] = \text{var}(\bar{X}_N) = \frac{\sigma^2}{N} \quad (38)$$

Ainsi, on obtient finalement la propriété suivante :

$$E(S_N) = \sigma^2 - \frac{\sigma^2}{N} = \left(\frac{N-1}{N} \right) \sigma^2 \quad (39)$$

Donc S_N est un estimateur biaisé de la variance σ^2 .

Autre façon de montrer ce résultat : On admet le résultat de cours selon lequel :

$$\left(\frac{N-1}{\sigma^2} \right) S_N^c \rightsquigarrow \chi^2(N-1) \quad (40)$$

où $S_N^c = (N-1)^{-1} \sum_{i=1}^N (X_i - \bar{X}_N)^2$ est la variance corrigée. On sait que S_N^c est un estimateur sans biais puisque :

$$\left(\frac{N-1}{\sigma^2} \right) E(S_N^c) = N-1 \iff E(S_N^c) = \sigma^2 \quad (41)$$

Or on sait en outre que :

$$S_N = \left(\frac{N-1}{N} \right) S_N^c \quad (42)$$

Par conséquent, on retrouve :

$$E(S_N) = \left(\frac{N-1}{N} \right) E(S_N^c) = \left(\frac{N-1}{N} \right) \sigma^2 \quad (43)$$

Question 6 (1 point) VRAI Construisons tout d'abord l'estimateur du MV. La vraisemblance de l'échantillon de v.a.r. *i.i.id.* $\{X_1, X_2, \dots, X_N\}$ s'écrit :

$$L(X_1, X_2, \dots, X_N, \lambda) = \prod_{i=1}^N \Pr[X = X_i] = e^{-\lambda N} \frac{\lambda^{\sum X_i}}{\prod_{i=1}^N X_i!} \quad (44)$$

On en déduit la log-vraisemblance de l'échantillon :

$$\log L(X_1, X_2, \dots, X_N, \lambda) = -N\lambda + \log(\lambda) \left(\sum_{i=1}^N X_i \right) - \log \left(\prod_{i=1}^N X_i! \right) \quad (45)$$

La première équation de vraisemblance est :

$$\left. \frac{\partial \log L(X_1, \dots, X_n, \lambda)}{\partial \lambda} \right|_{\lambda=\hat{\lambda}} = -N + \frac{1}{\hat{\lambda}} \left(\sum_{i=1}^N X_i \right) = 0 \quad (46)$$

On en déduit un candidat pour l'estimateur du MV :

$$\hat{\lambda} = \bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i \quad (47)$$

La seconde équation de vraisemblance permet de vérifier qu'il s'agit bien d'un maximum puisque :

$$\left. \frac{\partial^2 \log L(X_1, \dots, X_n, \lambda)}{\partial \lambda^2} \right|_{\lambda=\hat{\lambda}} = -\frac{1}{\hat{\lambda}^2} \left(\sum_{i=1}^N X_i \right) = -\frac{N}{\bar{X}_N} < 0 \quad (48)$$

Question 7 (1.5 point) VRAI La vraisemblance s'écrit :

$$L(X_1, X_2, \dots, X_N, \theta) = \begin{cases} \left(\frac{1}{20-\theta} \right)^N & \text{si } X_i \in [\theta, 20] \text{ pour } i = 1, \dots, N \\ 0 & \text{sinon} \end{cases} \quad (49)$$

Donc il faut que $\hat{\theta}$ soit plus petit (ou égal) à la plus petite des valeurs X_i pour que la vraisemblance soit nul et dans ce cas il faut que θ soit le plus grand possible pour la vraisemblance soit maximale. Donc, la vraisemblance est maximale pour :

$$\hat{\theta} = \inf_{\{i\}} X_i \quad (50)$$