Université de Paris Ouest
Licence Sciences Economiques

Année Universitaire 2009/2010
Statistiques 3ème année

Corrigé du devoir surveillé du 3 décembre 2009

Exercice 1 (4 points). On s'intéresse au diamètre de pièces circulaires fabriquées par une machine. Du fait des imperfections de la machine, le diamètre correspond à la réalisation d'une variable aléatoire de moyenne $m = 60$ et de variance $\sigma^2 = 100$. Le fabricant souhaite vérifier que le diamètre de ces pièces est bien de taille 60. Il réalise alors un n-échantillon (X_1, X_2, $\cdots$, X_n). On s'intéresse à la moyenne empirique $\bar{X}_n$.

1. (2 points) Calculer l'espérance et la variance de $\bar{X}_n$.

Corrigé $\mathbf{E}[\bar{X}_n] = m = 60$; $\mathrm{var}(\bar{X}_n) = \sigma^2/n = 100/n$.

2. (1 point) Déterminer suivant la taille de l'échantillon la loi de $\bar{X}_n$.

Corrigé Pour des petites valeurs de n, on ne connait pas la loi de $\bar{X}_n$. Pour n grand ($n \geq 30$), on peut approcher la loi de $\bar{X}_n$ par la loi normale $\mathbf{N}(60, 100/n)$.

3. (1 point) Calculer la probabilité en tirant au hasard un échantillon de taille 100, d'observer une moyenne d'échantillon inférieure à 58.2 ?

Corrigé Pour $n = 100$, on peut donc approcher la loi de $\bar{X}_n$ par la loi $\mathbf{N}(60, 1)$ et donc la loi de $\bar{X}_n - 60$ par la loi $\mathbf{N}(0, 1)$. La probabilité cherchée est donc $\Phi(-1.8)$, où Φ est la fonction de répartition de la loi normale. Par symétrie, $\Phi(-1.8) = 1 - \Phi(1.8)$. On lit sur la table $\Phi(1.8) = 0.9641$, d'où $\Phi(-1.8) = 0.0359$.

Exercice 2 (9.5 points). Considérons un n-échantillon (X_1, X_2, $\cdots$, X_n) i.i.d de loi de densité de paramètre α, $\alpha > 0$:

$$f(x) = \frac{1}{\alpha} x^{-\frac{1}{\alpha}-1} \quad \text{si} \quad x \geq 1 \quad \text{et} \quad 0 \text{ sinon}$$

1. (1 point) Vérifier que f est une densité de probabilité.

Corrigé On vérifie que $\alpha^{-1} \int_1^\infty x^{-1/\alpha-1}\,\mathrm{d}x = 1$.

2. (1 point) Donner la fonction de vraisemblance de l'échantillon.

Corrigé La vraisemblance est le produit des vraisemblances marginales car les variables sont indépendantes, soit

$$V(\alpha, x_1, \ldots, x_n) = \alpha^{-n} \left(\prod_{i=1}^n x_i \right)^{-1/\alpha-1} .$$

3. (2.5 points) Déterminer l'estimateur du maximum de vraisemblance de α.

Corrigé Il est équivalent de maximiser la log-vraisemblance donnée par

$$L(\alpha, x_1, \ldots, x_n) = -n \log(\alpha) - \left(\frac{1}{\alpha} + 1 \right) \sum_{i=1}^n \log(x_i) .$$

La dérivée par rapport à α est

$$\frac{\partial L}{\partial \alpha}(\alpha, x_1, \ldots, x_n) = -\frac{n}{\alpha} + \frac{1}{\alpha^2}\sum_{i=1}^{n} \log(x_i) \ .$$

Elle s'annule en la valeur

$$\frac{1}{n}\sum_{i=1}^{n} \log(x_i) \ .$$

Cette valeur est bien un maximum car c'est l'unique zéro de la dérivée et la log-vraisemblance vaut $-\infty$ en 0 et $+\infty$. L'estimateur du maximum de α est donc

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} \log(X_i) \ .$$

4. (2 points) On considère la variable aléatoire $Y = \log(X)$, avec X qui suit une loi de densité $f(x)$. Calculer la fonction de répartion de Y et en déduire qu'elle est de loi exponentielle.

Corrigé Y est une variable aléatoire positive. Soit $y \geq 0$.

$$\mathbf{P}(Y \leq y) = \mathbf{P}(\log(X) \leq y) = \mathbf{P}(X \leq \mathrm{e}^y) = 1 - (\mathrm{e}^y)^{-1/\alpha} = 1 - \mathrm{e}^{-1/\alpha y} \ .$$

Y suit donc la loi exponentielle de paramètre $1/\alpha$.

5. (1.5 point) En utilisant les résultats de la question précédente dire si l'estimateur obtenu est sans biais ? convergent ?

Corrigé L'estimateur du maximum de vraisemblance est donc la moyenne empirique d'un échantillon de la loi exponentielle de paramètre $1/\alpha$. On a donc

$$\mathbf{E}[\hat{\alpha}] = \mathbf{E}[\bar{Y}_n] = \mathbf{E}[Y] = \alpha \ .$$

L'estimateur $\hat{\alpha}$ est donc sans biais. Il est aussi convergent par la loi des grands nombres.

6. (1 point) Calculer l'information relative à α contenue dans l'échantillon.

Corrigé L'information relative à α contenue dans l'échantillon est

$$I(\alpha) = -\mathbf{E}\left[\frac{\partial^2 L}{\partial \alpha^2}\right] = -\frac{n}{\alpha^2} - \frac{2n}{\alpha^3}\mathbf{E}[\bar{Y}_n] = -\frac{n}{\alpha^2} - \frac{2n}{\alpha^2} = \frac{n}{\alpha^2} \ .$$

7. (0.5 point) En déduire si l'estimateur précédent est efficace ?

Corrigé On a $\mathrm{var}(\hat{\alpha}) = \mathrm{var}(\bar{Y}_n) = \alpha^2/n = I^{-1}(\alpha)$, donc $\hat{\alpha}$ est efficace.

Exercice 3. '[7.5 points] Un état souhaite connaître la proportion p de ses électeurs en accord avec un nouveau texte de loi. Pour cela, il décide d'organiser un sondage sur un n-échantillon $(X_1, X_2, \ldots, X_n)$ et sur 100 personnes interrogées, 80 pensent voter oui.

1. (2 point) Quelle est la loi des variables X_i ? Sont-elles indépendantes ? Donner leur espérance et leur variance.

Corrigé Les variables X_i sont des variables de Bernoulli de même paramètre p que l'on peut considérer comme indépendantes car on a interrogé un nombre d'individus n très petit devant la population totale. On a $\mathbf{E}[X_i] = p$ et $\mathrm{var}(X_i) = p(1-p)$.

2. (1 point) Proposer un estimateur sans biais de p.

Corrigé La moyenne empirique $\bar{X}_n$, définie pr $\bar{X}_n = (X_1 + \cdots + X_n)/n$, est un estimateur sans biais de p car $\mathbf{E}[\bar{X}_n] = p$.

3. (2.5 points) On décide maintenant de prendre comme estimateur de p la moyenne empirique $\bar{X}_n$. Par quelle loi peut-on approximer celle de $\bar{X}_n$? En déduire un intervalle de confiance pour p au niveau de confiance 95%.

Corrigé Pour n grand, on peut approcher la loi de $\bar{X}_n$ par la loi normale. Plus précisément, $\sqrt{n}(\bar{X}_n - p)$ suit approximativement la loi $\mathbf{N}(0, p(1-p))$. On peut de plus remplacer $p(1-p)$ par l'estimateur empirique de la variance qui a l'expression $\bar{X}_n(1-\bar{X}_n)$ pour la loi de Bernoulli. Un intervalle de confiance de niveau 95% est donc

$$I_1 = \left[\bar{X}_n - 1.96 \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} \;,\; \bar{X}_n + 1.96 \frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} \right] .$$

Application numérique. La demi-longueur de l'intervalle de confiance est $1.96\sqrt{.8 * .2}/10 = 0.0784$, d'où

$$I_1 = [0.7216, 0.8784] .$$

4. (2 points) Montrer que pour $p \in [0,1]$ on a $p(1-p) \leq \frac{1}{4}$ en déduire un nouvel intervalle de confiance pour p au niveau de confiance 95%.

Corrigé La fonction $p \to p(1-p)$ est maximale pour $p = 1/2$ (par symétrie) et donc $p(1-p) \leq 1/2$ pour tout $p \in [0,1]$. On en déduit donc un nouvel intervalle de niveau au moins 95% :

$$\left[\bar{X}_n - \frac{1.96}{2\sqrt{n}} \;,\; \bar{X}_n + \frac{1.96}{2\sqrt{n}} \right] .$$

Application numérique. La demi-longueur de l'intervalle de confiance est maintenant $1.96/(2 * 10) = 0.098$, d'où

$$I_2 = [0.702, 0.898] .$$

Le second intervalle est plus large que le premier mais il est en fait de niveau supérieur à 95%.