Statistiques appliquées L3
Groupe 10 2006/2007

Correction Interrogation n°1

**Exercice** 1
**1-** $Cov(X,Y) = E[(X - E(X)(Y - E(Y)))] = E(XY) - E(X)E(Y)$ or $E(XY) = E(X)E(Y)$ car $X$ et $Y$ sont indépendantes. D'où $Cov(X,Y) = 0$
**2-a**
$n_1 = 40 \geq 15$ et $n_1p = 4 < 15$ donc la loi binomiale $B(40\ ; 0,1)$ est proche de de la Poisson $P(n_1p)$. On peut alors utiliser l'approximation par la loi de Poisson pour la loi de $X$.
$n_2 = 200 \geq 15$ et $n_2p = 20 > 15$ donc la loi binomiale $B(200\ ; 0,1)$ est proche de de la normale $N(n_1p\ ;\ n_1p(1-p))$. Donc on peut utiliser l'approximation par la loi Normale pour la loi de $Y$.

**2-b** $Z$ est la somme de 2 v.a.r indépendantes Binomiales donc $Z$ est binomiale $Z \rightsquigarrow B(n_1+n_2, p)$.

**2-c** $E(Z) = E(X+Y) = E(X) + E(Y)$ or $X \rightsquigarrow B(n_1, p)$ donc $E(X) = n_1p$ (cours). De même $E(Y) = n_2p$ et donc $E(Z) = (n_1+n_2)p$.
$V(Z) = V(X+Y) = V(X) + V(Y) + 2Cov(X,Y) = V(X) + V(Y)$ or $X \rightsquigarrow B(n_1, p)$ donc $V(X) = n_1p(1-p)$(cours). De même $V(X) = n_2p(1-p)$
$V(Z) = (n_1+n_2)p(1-p)$

**Exercice** 2 (Ex 12 du $TD1$)
**1-** On sait que $k$ doit vérifier :

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = 1$$

On doit respecter la condition $0 \leq x \leq y$ pour intégrer, d'où :

$$k\int_0^{+\infty}\left(\int_x^{+\infty} e^{-y}dy\right)dx = k\int_0^{+\infty}\left[-e^{-y}\right]_x^{+\infty}dx = k\int_0^{+\infty} e^{-x}dx = k\left[-e^{-x}\right]_0^{+\infty}dx = 1$$

Donc $k = 1$. Et la densité conjointe se réecrit :

$$f_{X,Y}(x,y) = \begin{cases} e^{-y} & \text{si} \quad 0 \leq x \leq y \\ 0 & \text{sinon} \end{cases}$$

**2-** Lois marginales :
- De $X$ :

$$f_X(x,y) = \begin{cases} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy & \text{si} \quad x > 0 \\ 0 & \text{sinon} \end{cases}$$

Sachant que $x < y$,

$$f_X(x,y) = \int_x^{+\infty} e^{-y}dy = \left[-e^{-y}\right]_x^{+\infty} = e^{-x}$$

D'où, la loi marginale de $X$:

$$f_X(x,y) = \begin{cases} e^{-x} & \text{si} \quad x > 0 \\ 0 & \text{sinon} \end{cases}$$

C'est la densité de la loi exponentielle de paramètre 1.

- De $Y$ :

$$f_Y(x,y) = \begin{cases} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx & \text{si} \quad y > 0 \\ 0 & \text{sinon} \end{cases}$$

Or $0 < x < y$,

$$f_y(x,y) = \int_0^y e^{-y}dx = \left[-e^{-y}\right]_0^y = ye^{-y}$$

D'où la loi marginale de $Y$ :

$$f_Y(x,y) = \begin{cases} y\ e^{-y} & \text{si} \quad y > 0 \\ 0 & \text{sinon} \end{cases}$$

Calculons $f_X(x,y) \times f_Y(x,y) = e^{-x} \times y\ e^{-y} = y\ e^{-(x+y)} \neq f_{X,Y}(x,y)$ d'où $X$ et $Y$ ne sont pas indépendantes.

**3-** Lois conditionnelles:

- De $X/Y = y$ :

$$f_{X/Y=y}(x,y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_Y(x,y)} = \frac{e^{-y}}{ye^{-y}} = \frac{1}{y} & \text{si} \quad 0 \le x \le y \\ 0 & \text{sinon} \end{cases}$$

- De $Y/X = x$ :

$$f_{Y/X=x}(x,y) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x,y)} = \frac{e^{-y}}{e^{-x}} = e^{x-y} & \text{si} \quad 0 \le x \le y \\ 0 & \text{sinon} \end{cases}$$

**Exercice** 3 (12 **points) :**

1- Par linéarité de l'espérance :

$$E\left(\bar{X}_n\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{nm}{n} = m$$

$V(\bar{X}_n) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right)$ or les $X_i$ sont indépendantes donc $V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i)$ d'où :

$$V(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Puisque $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ est une combinaison linéaire de variables indépendantes suivant une loi normale alors $\bar{X}_n$ suit également une loi normale. Les paramètres de cette loi sont $E\left(\bar{X}_n\right) = m$ et $V(\bar{X}_n) = \frac{\sigma^2}{n}$ :

$$\bar{X}_n \rightsquigarrow N\left(m, \frac{\sigma^2}{n}\right)$$

**2-** Les $X_i$ étant des variables indépendantes admettant la même espérance et la même variance, la loi faible des grands nombres permet d'affirmer directement que $\left(\bar{X}_n\right)$ converge en probabilité vers l'espérance commune des $X_i$ c'est-à-dire $m$. Par ailleurs les $X_i$ suivent la même loi donc la loi forte des grands nombres est également applicable: $\left(\bar{X}_n\right)$ converge presque sûrement vers $m$.

**3-** L'application de l'inégalité de Tchebychev à $\bar{X}_n$ donne, pour tout $\varepsilon > 0$ fixé,

$$P(\left|\bar{X}_n - E(\bar{X}_n)\right| \geq \varepsilon) \leq \frac{V(\bar{X}_n)}{\varepsilon^2}$$

donc sachant que $E\left(\bar{X}_n\right) = m$ et $V(\bar{X}_n) = \frac{\sigma^2}{n}$, on obtient:

$$P(\left|\bar{X}_n - m\right| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \underset{n \longrightarrow +\infty}{\longrightarrow} 0$$

d'où $P(\left|\bar{X}_n - m\right| \geq \varepsilon) \underset{n \longrightarrow +\infty}{\longrightarrow} 0$ pour tout $\varepsilon > 0$, ce qui, justement veut dire que $\left(\bar{X}_n\right)$ converge en probabilité vers $m$.

**4-** Le meilleur estimateur de $m$ est la moyenne empirique $\bar{X}_n$. Et $\bar{X}_n \rightsquigarrow N\left(m, \frac{\sigma^2}{n}\right)$ d'après (1)
**5-** On a:

$$P(|U| \leq u) = P(-u \leq U \leq u) = P\left(U \leq u\right) - P(U < -u) = F(u) - F(-u)$$

or la loi normale standard est symétrique par rapport à 0 (au sens où sa densité est symétrique par rapport à 0) donc $F(-u) = 1 - F(u)$. Conclusion :

$$P(|U| \leq u) = 2F(u) - 1$$

**6-** $\sigma^2 = 9$ et $\overline{x}_n = 20$
**6-a** Intervalle Bilatérale de confiance pour $m$ au niveau de confiance 95%
On sait que :

$$\bar{X}_n \rightsquigarrow N\left(m, \frac{\sigma^2}{n}\right)$$

Soit

$$U = \sqrt{n}\frac{\bar{X}_n - m}{\sigma} \rightsquigarrow N\left(0, 1\right)$$

On sait que :

$$P\left(|U| < u\right) = 0,95 = 2F\left(u\right) - 1$$

D'où:

$$F\left(u\right) = 0,975 \text{ et donc } u = 1,96$$

Ainsi :

$$\begin{aligned}
P\left(|U| < 1,96\right) &= P\left(-1,96 < U < 1,96\right) = 0,95 \\
&= P\left(-1,96 < \sqrt{n}\frac{\bar{X}_n - m}{\sigma} < 1,96\right) = 0,95 \\
&= P\left(\bar{X}_n - 1,96\frac{\sigma}{\sqrt{n}} < m < \bar{X}_n + 1,96\frac{\sigma}{\sqrt{n}}\right) = 0,95 \\
&= P\left(20 - 1,96\frac{3}{\sqrt{25}} < m < 20 + 1,96\frac{3}{\sqrt{25}}\right) = 0,95
\end{aligned}$$

D'où $m$ à 95% de chance de se trouver entre $18,824$ et $21,176$

**6-b** On vient de montrer que l'IC de $m$ à 95% est $\left[\bar{X}_n - 1,96\frac{\sigma}{\sqrt{n}} \quad ; \quad \bar{X}_n + 1,96\frac{\sigma}{\sqrt{n}}\right]$
La longueur de l'intervalle($d$) de confiance est la distance qui sépare sa borne supérieure de sa borne inférieure. On cherche le nombre d'observations ($n'$) tel que $d < 2$

$$\begin{aligned} d &= \left(\bar{X}_n + 1,96\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X}_n - 1,96\frac{\sigma}{\sqrt{n}}\right) \\ &= 2 \times 1,96\frac{\sigma}{\sqrt{n}} < 2 \end{aligned}$$

D'où

$$\begin{aligned} 1,96\frac{\sigma}{\sqrt{n}} &< 1 \\ 1,96 \quad \sigma &< \sqrt{n} \\ (1,96)^2 \quad \sigma^2 &< n \end{aligned}$$

Et

$$n > 34,5744 \text{ soit } n > 35$$

Il faut au moins 35 observations pour que la longueur de l'intervalle de confiance de $m$à 95% soit inférieure à 2