Statistiques appliquées L3 2006/2007

Interrogation n°1

Exercice 1

On considère un couple de variables aléatoires réelles (X,Y) de densité conjointe :

$$f_{X,Y}(x,y) = \begin{cases} kx^3y & \text{si} \quad 0 \leq x \leq 1 \text{ et } 0 \leq y \leq 2 \\ 0 & \text{sinon} \end{cases}$$

1- Déterminer la valeur de k.

2- Déterminer les lois marginales de X et de Y. Ces variables sont-elles indépendantes?

Exercice 2

La production d'un bien est assurée par trois usines U_1, U_2 et U_3 qui fabriquent respectivement 30%, 30% et 40% du total. Les proportions des biens produits défectueux sont respectivement 4%, 3% et 2%. Quelles sont les probabilités qu'un bien choisi au hasard, et dont on constate qu'il est défectueux, provienne des usines U_1, U_2 et U_3 ?

Exercice 3

Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées selon une loi normale d'espérance m et de variance σ^2. Notons $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

1- Calculer l'espérance et la variance de $\bar{X}_n$. Quelle est la loi de $\bar{X}_n$?

2- Quel théorème permet de dire directement que $\left(\bar{X}_n\right)$ converge en probabilité vers m ? Y a-t-il convergence presque sûre de $\left(\bar{X}_n\right)$ vers m?

3- Il est possible de démontrer la convergence en probabilité de $\left(\bar{X}_n\right)$ de deux autres façons:

3-a- Montrer, en utilisant le définition de la convergence en probabilité et l'inégalité de Tchebychev, que $\left(\bar{X}_n\right)$ converge en probabilité vers m.

3-b- En utilisant le résultat de la question 1, montrer que $\left(\bar{X}_n\right)$ converge en moyenne quadratique vers m. Conclure.

3-c- Si on supposait uniquement que les X_i étaient des variables aléatoires indépendantes de même espérance m et de même variance σ^2, les théorèmes de la question 2 resteraient-ils applicables? les résultats des questions 3-a et 3-b resteraient-ils vrais? Justifier.

4- Soit Φ la fonction de répartition de la loi normale standard $N(0,1)$ et soit $\alpha > 0$.

4-a- Soit $U \rightsquigarrow N(0,1)$. Exprimer $P(|U| \leq u)$ en fonction de $\Phi(u)$ (justifier).

4-b- Démarche probabiliste : on s'intéresse à la probabilité que $\bar{X}_n$ soit au plus à une distance α de l'espérance m, ce qu'on peut écrire sous la forme $m-\alpha \leq \bar{X}_n \leq m+\alpha$. Exprimer $P\left(m-\alpha \leq \bar{X}_n \leq m+\alpha\right)$ en fonction des paramètres α, σ et n. Que se passe-t-il lorsque n tend vers l'infini?

4-c- Démarche statistique : on cherche à déterminer un intervalle centré sur l'estimateur $\bar{X}_n$ contenant m avec une certaine probabilité. Exprimer $P\left(\bar{X}_n - \alpha\frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X}_n + \alpha\frac{\sigma}{\sqrt{n}}\right)$ en fonction du paramètre α. Quelle relation doit donc satisfaire α pour que $P\left(\bar{X}_n - \alpha\frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X}_n + \alpha\frac{\sigma}{\sqrt{n}}\right) = 1-\beta$, avec $\beta \in]0,1[$?

Question-bonus (indépendante des exercices) :

On dispose d'un échantillon $X_1, X_2, ..., X_n$ de n observations indépendantes d'une variable de Bernouilli de paramètre p. En supposant que l'approximation normale de la loi binômiale est valable, déterminer l'expression d'un intervalle de confiance de niveau (proche de) 95% du paramètre p en fonction de l'estimateur $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ et de la taille de l'échantillon n. On rappelle que si $U \rightsquigarrow N(0,1)$ alors $P(U \leq 1,96) = 0.975$.