

ELEMENTS D'ANALYSE MULTIVARIEE ET SIMULATIONS - EXERCICES

Exercice 1

Soient X , Y et Z , 3 variables aléatoires réelles indépendantes gaussiennes centrées réduites (loi $N(0, 1)$). On définit les variables aléatoires $U = X+Y+Z$ et $V = X-Y$.

Quelle est la loi du couple (U, V) ? Les variables aléatoires U et V sont elles indépendantes ?

Exercice 2

Soit $\mathbf{Z} \sim N_p(0, I_p)$ un vecteur aléatoire normal de dimension p et $\mathbf{Y} = \mathbf{AZ} + \mathbf{c}$ une transformation linéaire de \mathbf{Z} , où \mathbf{A} est une matrice de dimension $\dim(\mathbf{A}) = k \times p$ et $\mathbf{c} \in \mathbb{R}^k$.
Montrez que $\mathbf{Y} \sim N_k(\mathbf{c}, \mathbf{AA}^t)$

Exercice 3

Soient 2 populations P_1 et P_2 bidimensionnelles normalement distribuées selon les lois suivantes :

$$P_1 \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \right) \quad \text{et} \quad P_2 \sim N_2 \left(\begin{pmatrix} 2 \\ 2 \end{pmatrix}; \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

Tracer sommairement les ellipses d'isodensité correspondant à un niveau de probabilité 0.95.

On utilisera le fait que $P(\chi_2^2 < 6) = 0.95$

Exercice 4

On suppose que l'on dispose de n observations d'une loi normale $N(\mu; \sigma^2)$.

- 1) Donner les expressions de l'intervalle de confiance à 95% pour la moyenne et de l'intervalle de tolérance pour une valeur future dans le cas où l'écart type est connu.
- 2) Reprendre la question précédente en supposant que l'écart type est inconnu.

Exercice 5

On suppose un échantillon bivarié de moyenne $\mathbf{g} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ et de variance $V = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$.

Donner les expressions des ellipsoïdes de confiance et de tolérance de probabilité 0,975 pour un échantillon de taille $n = 50$.

On utilisera le fait que $P(F_{2,48} < 4) = 0,975$.

Exercice 6 *Données infarctus.*

Tester l'égalité des moyennes des 2 groupes en utilisant les résultats fournis par SAS.

Exercice 7 *Simulation*

- 1- Calculer la période de la méthode de Lehmer lorsque $m = 7$, $r_0 = 3$ et $a = 4$.
- 2- A partir d'un générateur de nombres pseudo-aléatoires, on obtient une variable aléatoire $R \square U[0;1]$, comment obtenir une variable aléatoire $S \square Cauchy$ en utilisant la méthode de l'anamorphose.

Rappel : La densité de probabilité d'une loi de Cauchy est : $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$

- 3- On cherche à obtenir une variable aléatoire suivant une distribution de Poisson. Comment obtenir cette variable à partir de variables uniformes ?

Indices : On passera par une somme de lois exponentielles.

Exercice 8 *Le bootstrap*

On simule B échantillons bootstrap à partir d'un échantillon de distribution inconnue.

- 1- On veut créer un intervalle de confiance bootstrap basé sur les percentiles. Quelles bornes de la distribution bootstrap doit-on utiliser pour un intervalle de confiance à 90 % ?
- 2- Soit un échantillon (X_1, \dots, X_n) de données, θ un paramètre d'intérêt de distribution inconnue et T une statistique pour estimer ce paramètre. Décrire le processus afin d'estimer le biais, la variance, l'intervalle de confiance à 95% et la distribution d'échantillonnage de T par la méthode du bootstrap.