T.P. I : Vecteurs gaussiens

1 Simulation de la loi normale par la méthode de Box-Muller.

Exercice 1.

Soient X et Y deux variables indépendantes de loi normales centrées réduites.

1. Déterminer la loi du couple (X, Y). Puis montrer que la densité du couple (R^2, Θ), où $R^2 = X^2 + Y^2$ et $\Theta = \tan^{-1}(Y/X)$, est donnée par

$$f(d, \theta) = \frac{1}{4\pi} e^{-d/2}, \quad 0 < d < \infty, 0 < \theta < 2\pi$$

2. Montrer que les variables aléatoires R^2 et Θ sont indépendantes de lois respectivement exponentielle $\mathcal{E}(1/2)$ et uniforme sur $]0, 2\pi[$.

3. Réciproquement, en déduire que si U et V sont deux variables aléatoires indépendantes de loi uniformes sur $]0, 1[$ alors les variables aléatoires

$$X = \sqrt{-2\ln U} \cos(2\pi V) \quad Y = \sqrt{-2\ln U} \sin(2\pi V)$$

sont indépendantes de loi $\mathcal{N}(0, 1)$.

En déduire une méthode pour simuler un échantillon de la loi normale centrée réduite puis d'une loi normale quelconque. Vous vérifierez par deux méthodes vos résultats.

Exercice 2.

Soient U_1 et U_2 deux variables indépendantes de loi uniforme sur $[0, 1]$. En utilisant les résultats précédents montrer que

$$X = \sqrt{-\frac{2\ln S}{S}} V_1 \quad Y = \sqrt{-\frac{2\ln S}{S}} V_2$$

avec $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$ et $S = V_1^2 + V_2^2$, sont indépendantes de loi $\mathcal{N}(0, 1)$.

En déduire une méthode pour simuler un échantillon de la loi normale centrée réduite puis d'une loi normale quelconque. Vous vérifierez par deux méthodes vos résultats.

2 Simulation de lois normales multidimensionnelles.

Exercice 3.

Si $X_1, \ldots, X_k$ sont des variables aléatoires indépendantes de loi $\mathcal{N}(0, 1)$ alors le vecteur $(X_1, \ldots, X_k)$ suit la loi normale dans $\mathbb{R}^k$, d'espérrance nulle et de matrice de covariance l'identité. On en déduit la simulation d'une loi normale d-dimensionnelle suivante.

Soit μ un vecteur de $\mathbb{R}^k$ et Σ est une matrice de taille k symétrique semi-définie positive. Soient $(X_1, \dots, X_k)$ un vecteur aléatoire de loi $\mathcal{N}(0, I)$. Soit A une matrice carrée d'ordre k telle que $AA^t = \Sigma$. Alors le vecteur $Y = AX + \mu$ est un vecteur gaussien de moyenne μ et de matrice de covariance Σ.

D'un point de vue pratique, lorsque Σ est définie positive, la décomposition de Cholesky donne une matrice triangulaire inférieure A qui vérifie $AA^t = \Sigma$.

Dans le cas où $k = 2$ cette décomposition est explicite. En effet soit $Y = (Y_1, Y_2)$ un vecteur gaussien d'espérance $\mu = (\mu_1, \mu_2)$ et de matrice de covariance

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$$

Dans cette notation $\rho = \mathrm{Corr}(Y_1, Y_2) = \mathrm{Cov}(Y_1, Y_2)/(\sigma_1\sigma_2)$ représente la corrélation entre Y_1 et Y_2. Si on pose

$$A = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}$$

alors on vérifie aisément que $AA^t = \Sigma$.

Utiliser cette méthode pour simuler $n = 10000$ réalisation de Y suivant une loi $\mathcal{N}(0, \Sigma)$ où

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Faire varier la valeur de ρ ($\rho = -0.75,\ 0.75,\ -1,\ 1,\ 0$) et représenter les points obtenus. Commenter les différrences

Vérifier que la matrice A proposée est bien la décomposition de Cholesky de Σ en utilisant la commande *chol*.

Exercice 4.

On suppose que (Y_1, Y_2) est un couple gaussien centré et de matrice de covariance :

$$\Sigma = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$$

On construit les variables suivante :

$$X_1 = 2Y_1 - Y_2,$$
$$X_2 = Y_2 - Y_1.$$

1. Montrer que X_1 et X_2 sont indépendantes et de même loi $\mathcal{N}(0, 1)$.

2. Simuler n couples indépendants et de même loi que (Y_1, Y_2). Vérifier votre simulation.

3. On suppose que Y_1 et Y_2 représente la résultat d'une expérience. On suppose que l'expérience est faussée si le premier résultat est inférieur à 0.4 et le deuxième est supérieur à 3. Calculer à l'aide d'une simulation la probabilité que l'expérience soit faussée.

Déterminer la valeur théorique de cette probabilité.

4. Après la mise en place d'un autre procédé, les lois des variables Y_1 et Y_2 ont été un peu modifiée : (Y_1, Y_2) est maintenant un couple gaussien de moyenne $(-0.4, 2)$ et de matrice de covariance Σ. Simuler n couples gaussiens suivant cette loi. Ce procédé est-il avantageux ?

3 Loi du chi-deux et illustration du théorème de Cochran.

Exercice 5.

1. Pour $m = 2$, puis $m = 4$, simuler m échantillons de taille $n = 10000$ de la loi normale $\mathcal{N}(0, 1)$. Ces échantillons seront placés dans une matrice X à n lignes et m colonnes. Calculer l'échantillon Y obtenu en calculant la somme des carrées de chaque ligne de cette matrice.

Représenter un histogramme des valeurs de Y. Superposer sur le même graphique la densité de la loi du chi-deux à m degrés de liberté (*dchisq*).

Représenter la fonction de répartition empirique de Y. Superposer sur le même graphique la fonction de répartition de la loi du chi- deux à m degrés de liberté (*pchisq*).

2. Calculer la moyenne empiriques M de chacune des lignes de X, retrancher M à chacune des colonnes de X et calculer l'échantillon Z obtenu en prenant la somme des carrées de chaque ligne.

Représenter un histogramme des valeurs de Z. Superposer sur le même graphique la densité de la loi du chi-deux à $m-1$ degrés de liberté. Représenter la fonction de répartition empirique de Z. Superposer sur le même graphique la fonction de répartition de la loi du chi-deux à $m - 1$ degrés de liberté.

Justifier que le vecteur Z de la deuxième ligne peut également être obtenu par la commande $z = y - m * M^2$.