

TP2

Objectif

On se propose de simuler des variables aléatoires à l'aide de la méthodes de Box-Muller ainsi que des vecteurs gaussiens indépendants et non-indépendants.

Partie I : Algorithmes de Box-Muller

La méthode de Box-Muller est une technique populaire pour générer des échantillons aléatoires à partir d'une distribution normale centrée réduite. La méthode repose sur la transformation de deux variables aléatoires uniformes indépendantes en deux variables aléatoires normales indépendantes (Voir Fig.1).

Plus précisément, la méthode de Box-Muller consiste en les étapes suivantes :

1. Générer deux variables aléatoires uniformes indépendantes U_1 et U_2 , chacune étant dans l'intervalle $[0,1]$.
2. Calculer Z_1 et Z_2 comme suit :

$$\begin{aligned}Z_1 &= \sqrt{-2 \ln(U_1)} \cos(2\pi U_2) \\Z_2 &= \sqrt{-2 \ln(U_1)} \sin(2\pi U_2)\end{aligned}$$

Les variables Z_1 et Z_2 sont i.i.d. et suivent la loi normale centrée réduite.

Exercice

1. Écrire une fonction qui retourne n réalisations de la loi normale $\mathcal{N}(0,1)$ par la méthode de Box-Muller.
2. Valider l'algorithme à l'aide d'un outil graphique (Q-Q plot et Fonction de répartition empirique).

Partie II : Simulation de vecteurs gaussiens

Un générateur de la loi $\mathcal{N}(0, 1)$ permet de simuler un vecteur aléatoire $X = (X_1, \dots, X_d)$ de \mathbb{R}^d de loi normale multivariée $\mathcal{N}(\mu, \sigma^2)$.

1^{er} cas : Matrice de covariance Σ diagonale.

Il suffit de générer d variables X_i pour $i = 1, \dots, d$, indépendantes de lois normale respectives $\mathcal{N}(\mu_i, \Sigma_{ii})$. Pour ce faire, il suffit de simuler Z_1, \dots, Z_d des variables aléatoires i.i.d. suivant la loi $\mathcal{N}(0, 1)$ et d'utiliser la transformation $X_i = \mu_i + Z_i \sqrt{\Sigma_{ii}}$.

2^{eme} cas : Matrice de covariance Σ non diagonale.

La méthode précédente n'est pas envisageable. On utilise alors la décomposition de Cholesky de Σ .

Exercice

Soit $X = (X_1, X_2)$ de loi normale multivariée, et soient :

$$\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix} \text{ et } \Sigma_2 = \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}$$

1. Écrire une fonction qui permet de simuler une suite de vecteurs $(X^{(n)})_{n \geq 1} = (X_{1,n}, X_{2,n})_{n \geq 1}$ qui suivent la loi $\mathcal{N}(\mu, \Sigma_1)$.
Utiliser la fonction créer dans la première partie et Voir Indications.
2. Simuler une suite de vecteurs $(X^{(n)})_{n \geq 1} = (X_{1,n}, X_{2,n})_{n \geq 1}$ qui suivent la loi $\mathcal{N}(\mu, \Sigma_2)$.
Utiliser la fonction prédéfinie *rnorm*.
3. On suppose que $X = (X_1, X_2)$ de loi $\mathcal{N}(\mu, \Sigma_2)$.
 - (a) Quelle est la loi de $X_1 + X_2$?
 - (b) Comparer la densité de $X_1 + X_2$ avec l'histogramme de simulation.

Indications

- *rnorm*(n, μ, σ) permet de simuler directement n réalisations i.i.d. de la loi de normale $\mathcal{N}(\mu, \sigma^2)$ (par défaut $\mu = 0$ et $\sigma = 1$).
- *dnorm*, *pnorm* et *qnorm* correspondent respectivement à la densité, à la fonction de répartition et au quantile d'une loi normale.
- *quantile*($x, probs$) retourne les quantiles d'un échantillon x pour un vecteur de probabilités *probs*.
- *qqplot* permet de tracer un diagramme Quantile-Quantile.
- La fonction *chol* de R permet d'obtenir la décomposition de Cholesky d'une matrice symétrique définie-positive.

Soit $\mathbf{Z} = (Z_1, Z_2) \sim \mathcal{N}(\mathbf{0}_2, \mathbf{I}_2)$, $\rho \in [-1, 1]$ une constante. En posant,

$$X_1 = \mu_1 + \sigma_1 Z_1 \quad \text{et} \quad X_2 = \mu_2 + \sigma_2 \left(\rho Z_1 + \sqrt{1 - \rho^2} Z_2 \right),$$

on a

$$\mathbf{X} = (X_1, X_2) \sim \mathcal{N} \left\{ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \right\}.$$

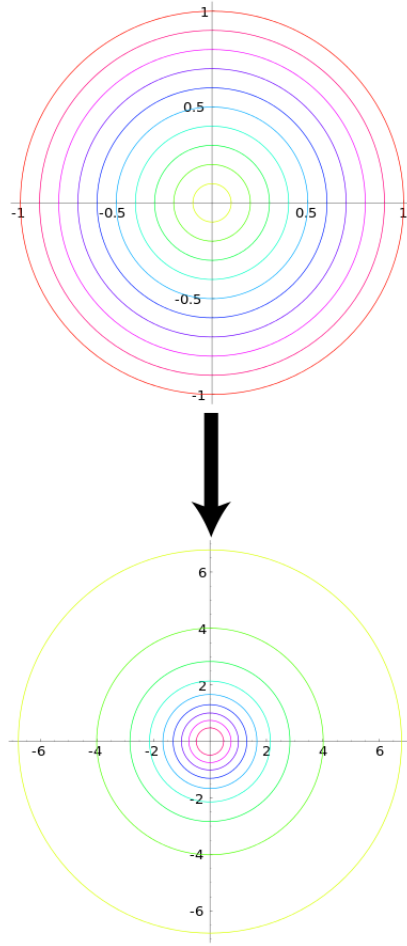


FIGURE 1 – Représentation graphique de la transformation : les cercles de départ, répartis uniformément autour de l'origine, deviennent un nouvel ensemble de cercles centrés, dont la répartition est proche de l'origine puis s'étirole rapidement. Les plus grands cercles de départ correspondent aux plus petits cercles d'arrivée, et vice-versa.