Fiche 4 : Loi normale

Définition de la loi normale

La loi normale est l'une des lois de probabilité les plus adaptées pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires. Également appelée loi gaussienne, loi de Gauss ou loi de Laplace-Gauss, c'est une loi de probabilité continue qui dépend de deux paramètres : son espérance μ et son écart type σ. La densité de probabilité de la loi normale est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Lorsqu'une variable aléatoire X suit la loi normale, elle est dite gaussienne ou normale et il est habituel d'utiliser la notation avec la variance σ^2 : $X \sim \mathcal{N}(\mu, \sigma^2)$. L'objectif de ce TD est de se familiariser avec des propriétés importantes de la loi normale et de voir comment les utiliser pour générer une variable aléatoire suivant cette loi.

Question 1.1 : De la même façon que R vous permet de générer des nombres aléatoires suivant la loi uniforme à l'aide de la fonction `runif`, il est possible de générer des nombres suivant la loi normale à l'aide de la fonction `rnorm`. R fournit également les fonctions de densité (`dnorm`), de repartition (`pnorm`), de quantile (`qnorm`).

- Afficher la loi de réparition (i.e., la fonction de densité également connue sur le nom de *probability density function*).
- Afficher la fonction de répartition (également connue sous le nom *cumulative distribution function*).
- Afficher l'inverse de la fonction de répartition (également connue sous le nom *quantile function*).

Question 1.2 : Considérons une variable $X \sim \mathcal{N}(0, 1)$. Calculez avec R les probabilités suivantes : $\mathbb{P}(X \in [-1, 1])$, $\mathbb{P}(X \in [-2, 2])$, $\mathbb{P}(X \in [-3, 3])$.

Propriétés importantes de la loi normale

Linéarité Si $X \sim \mathcal{N}(0, 1)$ alors $m + \sigma X \sim \mathcal{N}(m, \sigma^2)$

Convolution Si $X \sim \mathcal{N}(m, \sigma^2)$, $Y \sim \mathcal{N}(m', \sigma'^2)$ et X et Y sont indépendantes alors $X + Y \sim \mathcal{N}(m + m', \sigma^2 + \sigma'^2)$.

La linéarité s'obtient assez simplement en effectuant un changement de variable. La convolution est bien moins évidente. En effet, s'il est évident que $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = m + m'$ et que Var$(X + Y) =$ Var$(X) +$ Var$(Y) = \sigma + \sigma'^2$ (car X et Y sont indépendantes), le fait que $X + Y$ suive une loi normale est loin d'être évident... La somme de deux lois uniformes n'est pas uniforme par exemple.

Question 1.3 : Illustrez la linéarité et la convolution en générant des échantillons, en traçant leurs histogrammes et en superposant les lois correspondant.

Génération

Bien sûr, en R, la fonction rnorm vous permet de construire des échantillons suivant la loi normale mais comment faire si on ne dispose que d'un générateur de nombres uniformes ?
Commençons par la méthode "historique". Considérons des variables U_i suivant chacune une loi uniforme sur $[0, 1]$. On s'intéresse à la variable X définie par $X = U_1 + U_2 + \ldots + U_{12} - 6$.
Question 1.4 : Générez des échantillons de X et traçez l'histogramme correspondant. Comparez-le à la densité de la loi $\mathcal{N}(0, 1)$.
On pourrait bien sûr utiliser l'inverse de la fonction de répartition mais elle n'a pas une tête sympathique. Il y a moyen de la discrétiser mais le fait que la loi soit non bornée complique forcément un peu les choses au bord. On va donc regarder une autre méthode, la méthode de Box-Müller, qui utilise des propriétés fortes de la loi normale.

Question 1.5 : Tracez l'histogramme 2D correspondant à des échantillons de $U(-1, 1) \times U(-1, 1)$.

Question 1.6 : Comparez l'histogramme 2D précédent à celui d'échantillons de $\mathcal{N}(0, 1) \times \mathcal{N}(0, 1)$. Que remarquez-vous ?

Question 1.7 : Considérons $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 1)$. Tracez l'histogramme d'échantillons de $X^2 + Y^2$. À quelle loi cet histogramme vous-fait-il penser ?

Question 1.8 : Déduisez des observations précédentes une méthode pour générer deux variables indépendantes (X, Y) de loi normales à partir de deux variables indépendantes U_1, U_2 de loi uniforme sur $[0, 1]$. Testez votre algorithme.