Distribution de gauss

1. Introduction

Une variable aléatoire x obéit à la distribution de Gauss (ou distribution normale), si sa densité de probabilité est :

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

L'espérance de x est μ, son écart-type est σ.

Une application importante est la somme de N nombres aléatoires, qui obéit à la distribution de Gauss lorsque N est assez grand. Le théorème de la limite centrale établit que la distribution des moyennes empiriques (calculées avec N échantillons, N très grand) d'une variable aléatoire y est une distribution de Gauss dont l'écart type est :

$$\sigma = \sqrt{\frac{var(y)}{N}} \quad (2)$$

où $var(y)$ est la variance de la variable aléatoire et N le nombre d'échantillons utilisés pour le calcul de la moyenne. μ est l'espérance de y.

2. Simulation de Monte-Carlo

Il s'agit de générer des échantillons de la variable réelle x obéissant à la distribution de Gauss. On suppose que l'on dispose d'un générateur de nombres pseudo-aléatoires délivrant des nombres (flottants) avec une distribution uniforme. On utilisera la fonction `random.uniform(a,b)`, qui génère un flottant dans l'intervalle $[a, b]$, représentant un nombre réel aléatoire de densité uniforme sur cet intervalle. Pour générer rapidement un tableau de N nombres aléatoires, on pourra aussi utiliser la fonction `numpy.random.random(N)`.

La variable aléatoire x peut être obtenue à partir de deux variables aléatoires u_1 et u_2 uniformes sur $[0, 1]$, par la transformation suivante :

$$x = \mu + \sigma\sqrt{-2\ln(u_1)}\cos(2\pi u_2) \quad (3)$$

(1) Écrire une fonction générant les nombres x, pour $\mu = 0$ et pour une valeur de σ donnée.

(2) Tester cette fonction en calculant la moyenne et l'écart-type pour un nombre N d'échantillons assez grand.

Pour vérifier que les nombres tirés obéissent bien à la distribution de Gauss, on doit calculer un histogramme des N valeurs tirées. Pour cela, il faut tout d'abord choisir un intervalle $[-x_{max}, x_{max}]$ assez large pour contenir pratiquement toutes les valeurs générées (à choisir en fonction de σ). Cet intervalle est divisé en P sous-intervalles égaux, chacun de largeur $2x_{max}/P$. L'histogramme est un tableau H à P éléments. L'élément $H[i]$ contient le nombre d'échantillons dont la valeur x se trouve dans le sous-intervalle correspondant.

L'histogramme représente une densité de probabilité. Les valeurs de H seront donc divisées d'une part par le nombre de tirages, d'autre part par la largeur des sous-intervalles.

(3) Générer l'histogramme pour un nombre de tirages assez grand. Faire une comparaison avec la densité de probabilité théorique $p(x)$ de la distribution de Gauss.