Informatique

TP 5

Simulation

Le générateur de nombres aléatoires grand de Scilab permet de simuler, outre les lois discrètes usuelles (déjà vu), les lois continues au programme :

- `grand(m,n,'unf',a,b)` permet de simuler la loi uniforme sur $[a, b]$;
- `grand(m,n,'exp',mu)` permet de simuler la loi exponentielle d'espérance μ (i.e. de paramère $\lambda = \frac{1}{\mu}$) ;
- `grand(m,n,'nor',mu,sigma)` permet de simuler la loi normale $\mathcal{N}(\mu, \sigma^2)$ (attention, le dernier paramètre est l'écart-type et non la variance) ;
- `grand(m,n,'gam',nu,1)` permet de simuler la loi $\gamma(\nu)$.

On s'interdit dans ce TP, sauf mention expresse de l'énoncé, l'emploi de ce générateur grand. On utilisera seulement le générateur rudimentaire rand.

Partie 1 : un premier exemple, la loi exponentielle

Exercice 1 : simulation de la loi exponentielle

Soit $\lambda > 0$.

1. Soit U une variable de loi uniforme sur $[0, 1]$. Justifier que $X = -\frac{1}{\lambda}\ln(1 - U)$ suit une loi exponentielle de paramètre λ.
2. En déduire une `function y=simul_exp(lambda)` qui simule la loi exponentielle de paramètre λ.

Pour mesurer la qualité d'un algorithme de simulation, on peut tracer un histogramme représentant les valeurs $x_1, \ldots, x_n$ obtenues lors d'un grand nombre de simulations de la loi étudiée. Si `X` est le vecteur contenant ces valeurs et `S` un vecteur dont les éléments $s_1, \ldots, s_p$ sont rangés par ordre croissant, la commande `histplot(S,X)` trace des rectangles ayant pour base les classes $[s_1, s_2],]s_2, s_3], \ldots,]s_{p-1}, s_p]$ sur l'axe des abscisses et dont l'aire est proportionnelle au nombre de valeurs de X appartenant à la classe correspondante (la somme des aires des rectangles étant égale à 1). La commande `histplot(n,X)` crée l'histogramme construit sur n classes de même longueur recouvrant le plus petit intervalle contenant tous les éléments de X.

Exercice 2 : tracé de l'histogramme de la distribution statistique obtenue

Superposer sur une même figure l'histogramme des valeurs obtenues lors de n simulations de loi exponentielle de paramètre 1 avec la densité de la loi $\mathcal{E}(1)$. Faire une figure pour chacune des valeurs de $n \in \{10^2, 10^3, 10^4\}$.

Une autre méthode pour mesurer la qualité d'un algorithme de simulation est de comparer la fonction de répartition empirique à la fonction de répartition de la loi étudiée. En notant $x_1, \ldots, x_n$ les valeurs obtenues lors d'un grand nombre de simulations de la loi étudiée, la *fonction de répartition empirique* associée à cette distribution statistique est définie par

$$\overline{F}_n : y \in \mathbb{R} \longmapsto \overline{F}_n(y) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{[x_k \leqslant y]}$$

où :

$$\forall y \in \mathbb{R}, \quad \forall k \in [\![1, n]\!], \quad \mathbb{1}_{[x_k \leqslant y]} = \begin{cases} 1 & \text{si } x_k \leqslant y \\ 0 & \text{sinon} \end{cases}.$$

Pour $y \in \mathbb{R}$, $\overline{F}_n(y)$ est donc égale à la proportion de valeurs x_k qui sont inférieures ou égales à y.

Exercice 3 : tracé de la fonction de répartition empirique

Superposer sur une même figure la fonction de répartition empirique associée aux valeurs obtenues lors de n simulations de loi exponentielle de paramètre 1 avec la fonction de répartition de la loi $\mathcal{E}(1)$. Faire une figure pour chacune des valeurs de $n \in \{10^2, 10^3, 10^4\}$.

Exercice 4 : convergence de la moyenne empirique

Étant données les valeurs $x_1, \dots, x_{1000}$ obtenues lors de 1000 simulations de loi exponentielle de paramètre 1, représenter graphiquement la suite de terme général $\overline{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$, $1 \leqslant n \leqslant 1000$. Commenter.

Exercice 5 : application à la simulation de la loi géométrique

Soit X une variable aléatoire de loi exponentielle de paramètre $\lambda > 0$.

1. Déterminer la loi de la variable $\mathrm{Y} = \lfloor \mathrm{X} \rfloor$.

2. En déduire une `function simul_geometrique(p)` de simulation de la loi géométrique de paramètre $p \in \,]0,1[$.

Partie 2 : méthode d'inversion

Soit X une variable aléatoire dont la fonction de répartition $\mathrm{F_X}$ est continue et strictement croissante sur $\mathbb{R}$, sauf peut-être en dehors d'un intervalle ouvert I où elle peut être constante (égale à 0 ou à 1). Elle réalise donc une bijection de I sur $]0,1[$. Soit $\mathrm{Q_X} : \,]0,1[\, \longrightarrow \mathbb{R}$ la fonction réciproque, appelée *fonction quantile*. Soit U une variable de loi uniforme sur $]0,1[$. Par stricte croissance de $\mathrm{F_X}$, on a :

$$\forall x \in \mathbb{R}, \quad \mathbb{P}\big(\mathrm{Q_X}(\mathrm{U}) \leqslant x\big) = \mathbb{P}\big(\mathrm{U} \leqslant \mathrm{F_X}(x)\big) = \mathrm{F_X}(x)$$

et la variable $\mathrm{Q_X}(\mathrm{U})$ a donc même loi que X. La méthode d'inversion consiste à simuler la variable U puis à lui appliquer la fonction $\mathrm{Q_X}$ pour simuler X.

Exercice 6 : simulation de la loi de Cauchy

La loi de Cauchy a pour densité la fonction

$$f : x \in \mathbb{R} \longmapsto \frac{1}{\pi}\frac{1}{1+x^2}.$$

1. Écrire une `function y=simul_Cauchy()` qui simule la loi de Cauchy en utilisant la méthode d'inversion.

2. Implémenter et exécuter le code suivant :

Listing 1 : code à analyser

```
n=10000;   // nombre de simulations
N=10;  // nombre de trajectoires
U=rand(N,n);
X=tan(%pi*(U-1/2));
E=cumsum(X,'c')./cumsum(ones(N,n),'c');
plot2d(1:n,E',rect=[0 -10 n 10])
```

Que met-on ainsi en évidence ?

Partie 3 : simulation de la loi normale

Scilab propose la fonction `cdfnor` pour calculer les valeurs de la fonction de répartition Φ_{m,σ^2} de la loi normale $\mathcal{N}(m, \sigma^2)$ et de son inverse :

- l'instruction `cdfnor('PQ',x,m,sigma)` renvoie le réel $p = \Phi_{m,\sigma^2}(x)$;
- l'instruction `cdfnor('X',m,sigma,p,q)`, où le dernier paramètre q doit être égal à $1 - p$, renvoie le réel x tel que $\Phi_{m,\sigma^2}(x) = p$.

Exercice 7 : simulation avec `grand`

Superposer sur une même figure l'histogramme représentant la distribution statistique observée lors de 1000 simulations de loi $\mathcal{N}(1, 2^2)$ par la commande `grand` avec la densité de cette loi.

Exercice 8 : méthode des 12 uniformes

Le théorème limite central (cf. chapitre convergences et approximations en probabilités) énonce qu'étant donnée une suite (X_n) de variables aléatoires mutuellement indépendantes suivant une même loi admettant une espérance m et une variance σ^2, la variable aléatoire centrée réduite associée à $\mathrm{S}_n = \mathrm{X}_1 + \cdots + \mathrm{X}_n$:

$$\mathrm{S}_n^* = \frac{\mathrm{S}_n - nm}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}(\overline{\mathrm{X}}_n - m)$$

converge en loi vers la loi normale $\mathcal{N}(0, 1)$.
On considérera que si $\mathrm{X}_1, \ldots, \mathrm{X}_{12}$ sont des variables aléatoires indépendantes de loi uniforme sur $[0, 1]$, alors S_{12}^* suit approximativement une loi $\mathcal{N}(0, 1)$.

1. En déduire une `function y=simul_normale12(N,m,sigma2)` qui renvoie N simulations de la loi $\mathcal{N}(m, \sigma^2)$.
2. Superposer sur une même figure la représentation graphique de la fonction de répartition empirique de 1000 simulations de loi $\mathcal{N}(1, 2^2)$ avec la fonction de répartition de cette loi.

Exercice 9 : méthode de Box-Müller

On montre que si U et V sont deux variables aléatoires indépendantes de loi uniforme sur $[0, 1]$, alors la variable aléatoire $\mathrm{T} = \sqrt{-2\ln \mathrm{U}}\cos(2\pi \mathrm{V})$ suit la loi $\mathcal{N}(0, 1)$.

1. Écrire une `function y=simul_normaleBM(m,sigma2)` qui renvoie N simulations de la loi $\mathcal{N}(m, \sigma^2)$.
2. Superposer sur une même figure la représentation graphique de la fonction de répartition empirique de 1000 simulations de loi $\mathcal{N}(1, 2^2)$ avec la fonction de répartition de cette loi.

Exercice 10 : théorème de stabilité

Vérifier expérimentalement le théorème de stabilité pour les lois normales en utilisant `grand`.