Feuille de TP n° 2-Fonctions de répartitions, programmation et simulation de variables aléatoires

1 Fonction de répartition et inverse généralisé

1.1 Fonction de répartition

Soit X une variable aléatoire réelle, on note F_X sa fonction de répartition définie pour tout $x \in \mathbb{R}$ par $F_X(x) = \mathbb{P}(\{X \leq x\})$. Sous Scilab on fait appel à ces fonction avec la commande `cdf*` où `*` désigne la loi utilisée.

▷ `help cdf`

▷ `[F,G]=cdfbin("PQ",zeros(1,10),[1:10],18/38*ones(1,10),1-18/38*ones(1,10)), G(G>=0.95)=[];n=length(G)+1`

Que représente n ?

On peut définir l'inverse généralisée de F_X en posant pour tout $y \in]0,1[$, $F^{-}(y) = \inf\{x \in \mathbb{R} \ ; \ F(x) \geq y\}$.

▷ `[x]=cdfnor("X",0,1,0.975,1-0.975)`

En déduire la valeur de α pour laquelle la probabilité qu'une variable aléatoire normale centrée réduite apartienne à $[-\alpha, \alpha]$ soit égale à 95%.

1.2 Fonction de répartition empirique

Soit $X^n(\omega) = (X_1(\omega), \cdots, X_n(\omega))$ un n-échantillon d'une variable aléatoire X. La fonction de répartition empirique de cette échantillon est donnée pour tout $x \in \mathbb{R}$ par

$$F_{X^n(\omega)}(x) = \frac{Card\{i \in \{1, \cdots, n\} \ ; \ X_i(\omega) \leq x\}}{n}.$$

Sous Scilab on trace des fonctions constantes par morceaux à l'aide de la commande `plot2d2`.

2 Programmation

2.1 Boucles et instructions conditionnelles

▷ `help for`

▷ `I=[1:10000]; S=0;`
`for k=I, S=S+(2*(rand()<1/2)-1)*1/k; end; S`

▷ `help while`

▷ `S=0; T=0;`
```
while abs(S)<=1; S=S+grand(1,'nor',0,1/100); T=T+1; end; T
```
▷ `help if, help then, help else, help elseif`

▷ `U=rand(10000,2); k=0`
```
for i=[1:10000];
if U(i,1)^2+U(i,2)^2<=1 then k=k+1;
end;
end;
pi=4*k/10000
```
Que remarquez vous ?

2.2 Fonctions

On peut créer ses propres fonctions sous Scilab. Pour cela on commence par créer un fichier *nomdufichier.sci* dans lequel on écrira une ou plusieurs fonctions. Chaque fonction a la syntaxe suivante :

$$\begin{array}{c} \text{function } [y_1, \cdots, y_m] = \text{nomdelafunction}(x_1, \cdots, x_n) \\ \vdots \\ \textit{instructions} \\ \vdots \\ \text{endfunction} \end{array}$$

où les y_i sont les arguments de sortie et les x_i les arguments d'entrée. Dans les deux cas la liste peut être vide. On execute ensuite le fichier à l'aide des commandes

```
getf("\nomdurépertoiredesauvegarde\nomdufichier.sci")
```

ou

```
exec("\nomdurépertoiredesauvegarde\nomdufichier.sci")
```

Les fonctions sont des variables à part entière de Scilab, on peut aussi construire des fonctions de manière récursive.

2.3 Exercices

► Tracer les fonctions de répartitions des lois $\mathcal{N}(0,1)$, $\mathcal{E}(1)$, $\mathcal{B}(\frac{1}{2},10)$, et $\mathcal{P}(1)$.

► Ecrire une fonction `frempirique(X)` qui à partir d'un échantillon X trace sa fonction de répartition empirique.

► Tracer les fonctions de répartitions empiriques d'un échantillon de taille 100 et 1000 d'une loi $\mathcal{N}(0,1)$ et les superposer sur le même graphique avec la véritable fonction de répartion. Que remarquer vous ? Illustrer plus précisément ce résultat.

▶ Ecrire une fonction `TCL(n,m)` qui à partir d'un m-échantillons de la variable aléatoire

$$\sqrt{n}(\frac{X_1 + \cdots , X_n}{n} - 0),$$

où les X_i sont indépendantes et de loi uniforme sur $\{-1, 1\}$, tracer son histogramme et sa fonction de répartition empirique (sur des fenêtres distinctes) superposés à la densité et à la fonction de répartition d'une loi normale centrée réduite. Qu'en pensez vous ?

▶ Au casino on joue à la roulette et on décide tant qu'on a pas gagné de parier sur la couleur rouge, avec probabilité $\frac{18}{38}$, en doublant la mise à chaque fois. On commence par miser 1 et on dispose d'un capital de 1000. Soit τ le temps de ruine et C le capital maximum atteind pendant la partie. Ecrire une fonction `Tempsdejeu(n)` donnant un histogramme et la fonction de répartition empirique d'un n-échantillon de τ ainsi que sa moyenne empirique. Ecrire de la même façon une fonction `Capitalmax(n)`.

▶ Ecrire une fonction `Evénementrare(m,n)` qui trace l'histogramme et la fonction de répartition empirique d'un m-échantillon de loi $\mathcal{B}(n, \frac{1}{n})$. Comparer avec une loi $\mathcal{E}(1)$. En fait la loi $\mathcal{E}(\lambda)$ "approxime" assez bien la loi $\mathcal{B}(n, \frac{\lambda}{n})$ pour n assez grand (on verra plus tard qu'il y a convergence en loi). On appel cela la loi des événements rares. Pourquoi ?

▶ Ecrire une fonction `pont(n)` qui trace, étant donné un n-échantillon $(U_1(\omega), \cdots , U_n(\omega))$ de loi uniforme sur $[0, 1]$, la fonction aléatoire $t \in [0, 1] \mapsto X_n^\omega(t)$. Où

$$X_n^\omega(t) = \frac{1}{\sqrt{(n)}} \sum_{k=1}^{n} (\mathbb{1}_{U_k(\omega)\leq t} - t).$$

Comparer la loi de $X_n(t)$ et la loi $\mathcal{N}(0, t(1 - t))$.

3 Simulation de variables aléatoires

3.1 Lois discrètes

Soit $\mu = p_1\delta_1 + \cdots + p_n\delta_n$ une loi de probabilité sur $\{1, \cdots , n\}$. Soit U une variable aléatoire uniforme sur $[0, 1]$. Alors la variable aléatoire

$$X = \mathbb{1}_{\{U<p_1\}} + 2\mathbb{1}_{\{p_1<U<p_1+p_2\}} + \cdots + n\mathbb{1}_{\{p_1+\cdots+p_{n-1}<U<1\}}$$

est une variable aléatoire de loi μ. Comment simuler un jet de dé ?

3.2 Inversion de la fonction de répartition

Soit X une variable aléatoire réelle, F_X sa fonction de répartition et F_X^- sa pseudo inverse. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Alors $Y = F_X^-(U)$ est une variable aléatoire de même loi que X.

3.3 Méthode du rejet

3.3.1 Lois uniformes

On veut ici simuler des variables aléatoires uniformes sur des ensembles bornés B de $\mathbb{R}^d$. Supposons que $B \subset E$ et que l'on sait simuler une loi uniforme sur E. Soit (E_n) une suite de variables aléatoires indépendantes uniformes sur E. Soit τ le plus petit entier n tel que $E_n \in B$. Alors E_τ est une variable aléatoire uniforme sur B indépendante de τ.

3.3.2 A partir de la densité

Soit X une variable aléatoire réelle de densité f. On suppose qu'il existe Y une variable aléatoire réelle de densité g qu'on sait simuler et telle qu'il existe $c \geq 0$ tel que $f \leq c.g$. Soit (U_n) une suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$ et (W_n) une suite de variables aléatoires indépendantes de même loi que Y. On pose $Z_n = cU_n g(W_n)$ et τ le plus petit entier n tel que $f(w_n) \leq Z_n$. Alors W_τ a même loi que X et est indépendante de τ.

3.4 Exercices

► Ecrire une fonction `discr(p,n)` qui simule un n-échantillon de loi $p = (p_1, \cdots, p_m)$ sur $\{1, \cdots, n\}$. Comment simuler une loi discrette sur un autre ensemble que $\{1, \cdots, n\}$? Ecrire une fonction qui simule un n-échantillon d'une loi uniforme sur $\{-1, 0, 1\}$. Comment simuler une loi de probabilité sur $\mathbb{N}^*$? Ecrire une fonction qui simule un n-échantillon de la loi

$$\mu = \sum_{k \in \mathbb{N}^*} \frac{6}{\pi^2} \frac{1}{k^2} \delta_k.$$

► Simuler une loi $\mathcal{E}(1)$ à partir d'une loi uniforme. Comment simuler une loi géométrique à partir d'une loi exponentielle (utiliser la partie entière) ? Simuler une loi de Cauchy de densité $x \mapsto \frac{1}{\pi^2(1+x^2)}$ à partir d'une loi uniforme.

► Comment simuler une variable aléatoire de densité $x \mapsto \frac{1}{2}\exp(-|x|)$? En déduire une manière de simuler une loi normale centrée réduite à partir d'une loi exponentielle de paramètre 1.

► Ecrire une fonction `Unifboule(n,d)` qui donne un n-échantillon de loi uniforme sur la boule euclidienne de $\mathbb{R}^d$. Ecrire ensuite une fonction donnant une approximation du volume $V(d)$ de cette boule et tracer la fonction $d \mapsto V(d)$.