UPS TOULOUSE III **AGRÉGATION EXTERNE 2005**
TP2

SIMULATION DE VARIABLES ALÉATOIRES

1 Méthode par Inversion.

Soit X une variable aléatoire réelle de fonction de répartition F. On appelle inverse généralisée de F, la fonction F^{-1} définie pour tout $y \in]0,1]$ par

$$F^{-1}(y) = \inf\{x \in \mathbb{R} \ / F(x) \geq y\}.$$

La méthode de simulation par inversion repose sur le lemme suivant.

Lemme 1. Si U est une variable aléatoire de loi uniforme $\mathcal{U}([0,1])$, alors $F^{-1}(U)$ a même loi que X. De plus, si F est continue sur $\mathbb{R}$, alors $F(X)$ suit la loi uniforme $\mathcal{U}([0,1])$.

Exercice 1. Utiliser le code Matlab suivant pour engendrer N réalisations de variables aléatoires indépendantes et de loi exponentielle de paramètre $\lambda > 0$ ou bien de loi de Cauchy de paramètre $c > 0$.
clear
$N = input$('Entrez la taille de l'échantillon N : ') ;
$\lambda = input$('Préciser la valeur du paramètre λ : ') ;
$c = input$('Préciser la valeur du paramètre c : ') ;
$X = rand(N,1); \ Y = -\log(X)/\lambda; \ Z = c * \tan(\pi * (X - 0.5));$

Tracer les moyennes empiriques successives de Y et vérifier la loi des grands nombres pour Y. Augmenter le nombre N de réalisations pour affiner la précision. Comparer vos résultats de simulations avec le générateur *rexpweib* de Matlab. Effectuer le même exercice pour la loi normale $\mathcal{N}(0,1)$. Que se passe-t-il sur Z ? Quelle est la loi de la moyenne empirique associée à Z ? Conclure.

Exercice 2. Engendrer N réalisations de variables aléatoires indépendantes de loi de Pareto dont la densité est : $\dfrac{ab^a}{x^{a+1}}\mathbb{1}_{[b,+\infty[}(x)$, $b > 0$, de loi triangulaire :$\dfrac{2}{a}(1 - \dfrac{x}{a})\mathbb{1}_{[0,a]}(x)$, $a > 0$, de loi de Weibull : $\mathbb{P}(X > x) = e^{-(\frac{x}{b})^a}, x > 0, a > 0, b > 0$.

Lemme 2. Soit $x_1, x_2, \cdots, x_n$ des nombres réels tous différents et soit $p_1, p_2, \cdots, p_n$ des nombres réels positifs tels que $\sum_{i=1}^n p_i = 1$. On pose $s_0 = 0$ et pour tout $1 \leq k \leq n$, $s_k = \sum_{i=1}^k p_i$. Soit U est une variable aléatoire de loi uniforme $\mathcal{U}([0,1])$ et

$$X = \sum_{k=1}^n x_k \mathbb{1}_{(s_{k-1} \leq U \leq s_k)}.$$

Alors, X est une variable aléatoire de loi discrète $P = p_1\delta_{x_1} + p_2\delta_{x_2} + \cdots + p_n\delta_{x_n}$.

Exercice 3. Utiliser la fonction Matlab *probadis.m* afin d'engendrer une réalisation aléatoire d'une loi discrète à support fini.

```
function realis = probadis(n, x, p)
% Générateur aléatoire d'une loi discrète à support fini.
% n ≥ 1 est le cardinal du support de la loi.
% p est un vecteur de n nombres réels positifs tel que sum(p) = 1.
% x est un vecteur de n nombres réels donnant le support de la loi.
% Les réalisations successives sont indépendantes.
r = rand; a = 0; b = p(1);
 for i = 1 : n − 1,
        if ((r >= a) & (r < b))
              realis = x(i);
              return;
        end
        a = b; b = b + p(i + 1);
 end
 realis = x(n);
 return;
```

Créer un code Matlab permettant d'engendrer un vecteur aléatoire X contenant N réalisations indépendantes et de même loi Binomiale $\mathcal{B}(n, p)$ où les valeurs $N, n \geq 1$ et $0 < p < 1$ sont affectées par l'utilisateur. Pour N assez grand, vérifier la loi des grands nombres sur les moyennes empiriques successives de X. Comparer vos résultats de simulations avec le générateur *rbinom* de Matlab.

Exercice 4. Pour N, N_1 et $n \geq 1$ avec $N_1, n \leq N$, la loi hypergéométrique $\mathcal{H}(N, N_1, n)$ est donnée, pour tout $k \in [0; n]$ avec $0 \leq k \leq n$, par $B_p(X = k) = C_{N_1}^k C_{N-N_1}^{n-k} / C_N^n$. Créer un code Matlab permettant d'engendrer un vecteur aléatoire Z contenant M réalisations i.i.d. de loi $\mathcal{H}(N, N_1, n)$ où les valeurs $M, N \geq 1$ et $N_1, n \leq N$ sont affectées par l'utilisateur.

a) Si N tend vers l'infini et le rapport N_1/N tend vers p avec $0 < p < 1$, montrer que X converge en loi vers la loi Binomiale $\mathcal{B}(n, p)$. Pour M, N assez grand et $N_1 = pN$ avec $0 < p < 1$, tracer l'histogramme de Z et comparer le à la loi $\mathcal{B}(n, p)$.

b) Si N, N_1 et n tendent vers l'infini et le produit nN_1/N tend vers $\lambda > 0$, montrer que X converge en loi vers la loi de Poisson $\mathcal{P}(\lambda)$. Pour M, N assez grand, $N_1 = \lambda\sqrt{N}$ et $n = \sqrt{N}$, tracer l'histogramme de Z et comparer le à la loi $\mathcal{P}(\lambda)$.

Exercice 5. Soit (X_n) une suite de variables aléatoires indépendantes et de même loi exponentielle $\mathcal{E}(\lambda)$ avec $\lambda > 0$. Si $S_n = \sum_{k=1}^n X_k$, $N_0 = 0$ et pour tout $t > 0$, $N_t = \sum_{n=1}^\infty \mathbb{I}_{(S_n \leq t)}$, (N_t) est un processus de Poisson d'intensité λ. Montrer que, pour tout $t > 0$, N_t suit la loi de Poisson $\mathcal{P}(\lambda t)$. En déduire un code Matlab permettant d'engendrer un vecteur aléatoire Y contenant N réalisations indépendantes et de même loi $\mathcal{P}(\lambda)$ où les valeurs $N \geq 1$ et $\lambda > 0$ sont affectées par l'utilisateur. Pour N assez grand, vérifier la loi des grands nombres sur les moyennes empiriques successives de Y. Comparer vos résultats de simulations avec le générateur *rpoiss* de Matlab.

2 Méthode par Troncature.

La méthode de simulation par troncature repose sur le lemme suivant.

Lemme 3. Si X est une variable aléatoire à valeurs dans $\mathbb{Z}$, de fonction de répartition F et soit G la fonction de répartition d'une variable aléatoire continue Y telle que, pour tout $n \in \mathbb{Z}$, $G(n + 1) = F(n)$. Si U est une variable aléatoire de loi uniforme $\mathcal{U}([0, 1])$, alors la partie entière $[G^{-1}(U)]$ a même loi que X.

Exercice 6. Utiliser le code Matlab suivant pour engendrer N réalisations de variables aléatoires indépendantes et de loi uniforme discrète sur $\{1, 2, \cdots, a\}$ avec $a \in \mathbb{N}^*$ ou bien de loi géométrique de paramètre $p = 1 - \exp(-\lambda)$ avec $\lambda > 0$.

```
clear
N = input('Taille de l'échantillon N : ') ; a = input('Valeur du paramètre a : ') ;
if a ~= round(a)
        disp('La valeur du paramètre a doit etre un entier positif!!')
        break
 end
λ = input('Préciser la valeur du paramètre λ : ') ;
X = rand(N, 1); Y = fix(1 + a * X); Z = fix(- log(X)/λ);
```

Tracer les moyennes empiriques successives de Y et vérifier la loi des grands nombres pour Y. Augmenter le nombre N de réalisations pour affiner la précision. Effectuer le même exercice pour la loi géométrique associée à Z. Comparer vos résultats de simulations avec le générateur *rgeom* de Matlab.

3 Simulations de Gaussiennes.

L'algorithme de Box-Muller de simulation de gaussiennes repose sur le lemme suivant.

Lemme 4. Soit (X, Y) un couple aléatoire de $\mathbb{R}^2$. Alors, (X, Y) suit la loi normale $\mathcal{N}(0, I_2)$ si et seulement si $X = r \cos \theta$ et $Y = r \sin \theta$ où r et θ sont deux variables aléatoires indépendantes avec r^2 de loi exponentielle $\mathcal{E}(1/2)$ et θ de loi uniforme $\mathcal{U}([0, 2\pi])$.

Il découle de ce lemme que, si U et V sont deux variables aléatoires indépendantes de loi uniforme $\mathcal{U}([0, 1])$, alors

$$X = \sqrt{-2 \log U} \cos(2\pi V) \qquad \text{et} \qquad Y = \sqrt{-2 \log U} \sin(2\pi V)$$

sont indépendantes et de loi normale $\mathcal{N}(0, 1)$.

Exercice 7. Utiliser le code Matlab suivant pour engendrer N réalisations de variables aléatoires indépendantes et de loi normale $\mathcal{N}(m, \sigma^2)$ où la moyenne $m \in \mathbb{R}$ et la variance $\sigma^2 > 0$ sont affectées par l'utilisateur. Tracer également l'histogramme associé.

```
clear
N = input('Taille de l'échantillon N : ') ; m = input('Valeur de la moyenne m : ') ;
σ^2 = input('Préciser la valeur de la variance σ^2 : ') ;
X = m * ones(N, 1) + sqrt(σ^2) * sqrt(-2 * log(rand(N, 1))). * cos(2 * pi * rand(N, 1)) ;
clf
[E, C] = histo(X, sqrt(N), 0, 1) ;
hold on
title('Simulation d'une loi gaussienne') ; xlabel('Valeurs') ;  ylabel('Effectifs') ;
plot(C, dnorm(C, m, σ),'r-') ; legend('Empirique','Théorique') ;
hold off
```

Un second algorithme de simulations de gaussiennes provient de la méthode d'acceptation/rejet. Elle permet d'éviter l'appel à la fonction cosinus.

Lemme 5. Soit (X, Y) un couple de variables aléatoires de loi uniforme sur le disque unité $\mathcal{D} = \{(x, y) \in \mathbb{R}^2 \,/ x^2 + y^2 < 1\}$. Soit (r, θ) le couple de coordonnées polaires associé à (X, Y), $X = r \cos \theta$ et $Y = r \sin \theta$. On pose $R = 2\sqrt{-\log r}/r$. Alors, (RX, RY) suit la loi normale $\mathcal{N}(0, I_2)$.

Exercice 8. Reprendre l'exercice 5 avec le code Matlab suivant.
clear
$N = input$('Entrez la taille de l'échantillon N : ') ; $m = input$('Précisez la valeur de la moyenne m : ') ;
$\sigma^2 = input$('Précisez la valeur de la variance σ^2 : ') ;
$max = round(3 * N/2)$;
$X = 2 * rand(max, 1) - ones(max, 1);\ Y = 2 * rand(max, 1) - ones(max, 1)$;
$S = X.\hat{}2 + Y.\hat{}2;\ X = X(find(S < 1));\ Y = Y(find(S < 1))$;
$r = sqrt(X.\hat{}2 + Y.\hat{}2);\ R = 2 * sqrt(-log(r))./r$;
$Z = R(1 : N). * X(1 : N);\ T = m * ones(N, 1) + sqrt(\sigma^2) * Z$;

Lemme 6. Soit $m \in \mathbb{R}^d$ et Γ une matrice réelle, carrée d'ordre d, symétrique et définie positive. Soit A une matrice carrée d'ordre d telle que $AA^t = \Gamma$ et soit X un vecteur aléatoire gaussien $\mathcal{N}(0, I_d)$. Alors, $Y = AX + m$ suit la loi normale $\mathcal{N}_d(m, \Gamma)$.

Exercice 9. Créer un code Matlab permettant d'engendrer N réalisations d'un vecteur aléatoire de loi $\mathcal{N}_d(m, \Gamma)$ où la dimension d, la moyenne m et la matrice de covariance Γ sont affectées par l'utilisateur. Pour calculer la racine carrée A de Γ, on utilisera la commande *sqrtm* ou bien la décomposition de Cholesky *chol*.

Exercice 10. Utiliser le code Matlab suivant afin de simuler une trajectoire brownienne.
clear; *clf*
$N = input$('Entrez la taille de l'échantillon N : ') ; $a = 0;\ b = 1;\ h = (b - a)/N$;
$plot([a : h : b], cumsum(sqrt(h) * randn(N + 1, 1)))$;*hold on*
title('Simulation d'une trajectoire brownienne') ;
xlabel('Temps t') ; *ylabel*('Valeurs de Bt') ;*grid*; *hold off*

Exercice 11. Une autre application de la méthode du rejet.
En utilisant la méthode du rejet simuler une variable gaussienne à partir d'une variable de Laplace de densité :$\frac{1}{2}e^{-|x|}$.
Plus généralement trouver le meilleur choix dans la famille des lois de densité : $\frac{\lambda}{2}e^{-\lambda|x|}, \lambda > 0$.

Exercice 12. Utiliser la méthode de l'échantillonnage préférentiel à partir de la loi de Laplace pour calculer les moments d'une loi $\mathcal{N}(0, 1)$. Expliciter la variance de votre estimateur.

4 Représentations graphiques.

% Illustration de la LGN pour la loi exponentielle.
$clear; n = 1000;\ lambda = 0.5$; $X = -log(rand(n, 1))/lambda$;

$figure$;	Création d'une nouvelle fenêtre graphique.
$plot(cumsum(X)'./[1 : length(X)],' b')$	Trace les moyennes empiriques successives de X.
$title('Loi\ des\ Grands\ Nombres')$	Titre de la figure.
$xlabel('Nombre\ de\ realisations')$	Titre des abscisses.
$ylabel('Moyennes\ empiriques')$	Titre des ordonnées.
hold on	Garde la fenêtre graphique.
$plot(1/lambda * ones(n, 1),' r --');$	Trace la limite théorique.
$legend('Empirique',' Theorique');$	Légende.

Exercice 18. Ajouter à vos codes Matlab les représentations graphiques rencontrées ci-dessus.