ECE 2 - Année 2020-2021
Lycée français de Vienne
Mathématiques - F. Gaunard
http://frederic.gaunard.com

T.P. n°4

Variables aléatoires à densité
Méthode d'inversion
Histogrammes

☞ *Nouvelles commandes:* `cdfnor()`, `histplot()`

1 Loi uniforme, commande `rand()` et histogrammes

☞ On rappelle que les instructions `rand(1,N)` ou `rand(1:N)` permettent de créer un vecteur ligne dont les N composantes sont des réalisations (indépendantes) de la loi appelée avec la fonction `rand()`, ce qu'on appelle aussi un $N-$échantillon de la loi appelée. (On peut, en permutant les deux arguments créer aussi un vecteur colonne.)

Sans modification préalable, chaque appel de la fonction `rand()` renvoie un nombre aléatoire généré selon la loi uniforme $\mathcal{U}([0;1[)$.

☞ Sinon, l'instruction `grand(n, m,'unf', a,b)` renvoie une matrice de taille $n \times m$ de réels représentant des simulations de la loi uniforme sur l'intervalle réel $[a;b[$.

Exercice 1.

(1) À l'aide de la fonction `rand()`, écrire une fonction `y=unif(a,b)` simulant une loi uniforme sur un intervalle $[a;b]$.
(2) L'instruction `rand('normal')` indique que dorénavant, chaque appel de la fonction `rand()` simulera une loi normale centrée-réduite. Écrire alors une fonction `y=LaplaceGauss(mu,sigma)` simulant une variable aléatoire de loi de Laplace-Gauss de paramètres μ et σ^2.

☞ L'instruction `grand(m,n,'nor',mu,sigma)` simule une loi normale $\mathcal{N}(\mu,\sigma^2)$ (et crée une matrice à m lignes et n colonnes) et `grand(m,n,'exp', 1/lambda)` une loi exponentielle $\mathcal{E}(\lambda)$.

Histogrammes et commande `histplot()`

Les histogrammes sont des outils de représentation graphique des données qui sont particulièrement fréquents en statistiques. Dans l'exemple classique d'utilisation, on dispose d'un jeu de données fourni sous la forme d'un vecteur. Analyser cet ensemble de données consiste à en effectuer une répartition selon certaines classes *i.e.* comptabiliser les données comprises entre deux bornes a et b, celles entre b et c, celles entre c et d . . .

☞ La commande `histplot(x, U)` permet de représenter l'histogramme de répartitions des valeurs du vecteurs `U` selon les classes choisies par l'argument `x`.

Le premier argument `x` peut être:

- un nombre de classes:

  exemple: `histplot(10, U)`

  Dans ce cas la première classe commence avec la valeur du plus petit élément de `U` et la dernière classe se termine avec la valeur du plus grand;

- un un vecteur d'**éléments croissants**. Deux éléments successifs du vecteur définissent une classe

  exemple: `histplot([0:0.1:1], U)`

Dans les deux cas, la hauteur de chaque barre b est obtenue par

$$\frac{N(b)}{N(U) \times L(b)}$$

où

- $N(b)$ est le nombre d'éléments de la barre b;
- $N(U)$ est le nombre d'éléments du vecteur U;
- $L(b)$ est la largeur de la barre b.

En particulier, la hauteur de la barre ne correspond *a priori* pas au nombre d'éléments dans la classe correspondante. En attribuant la valeur `%f` à l'option `normalization`, on peut alors lire en ordonnée le nombre d'éléments de chaque classe. Pour cela, on écrira donc par exemple

```
histplot([0:0.1:1], U, normalization=%f)
```

Exercice 2.

(1) Générer un échantillon de taille 50 de la loi uniforme sur $[0;1[$ que l'on notera U.
(2) Représenter l'histogramme des valeurs de U en 10 classes.
(3) Représenter l'histogramme des valeurs de U en classes de largeurs 0.1, commençant à 0 et se terminant à 1.
(4) Ajouter une option pour lire le nombre de valeurs par classe avec la hauteur de chaque barre. Combien y a-t-il de valeurs comprises entre 0.3 et 0.4?

Exercice 3.

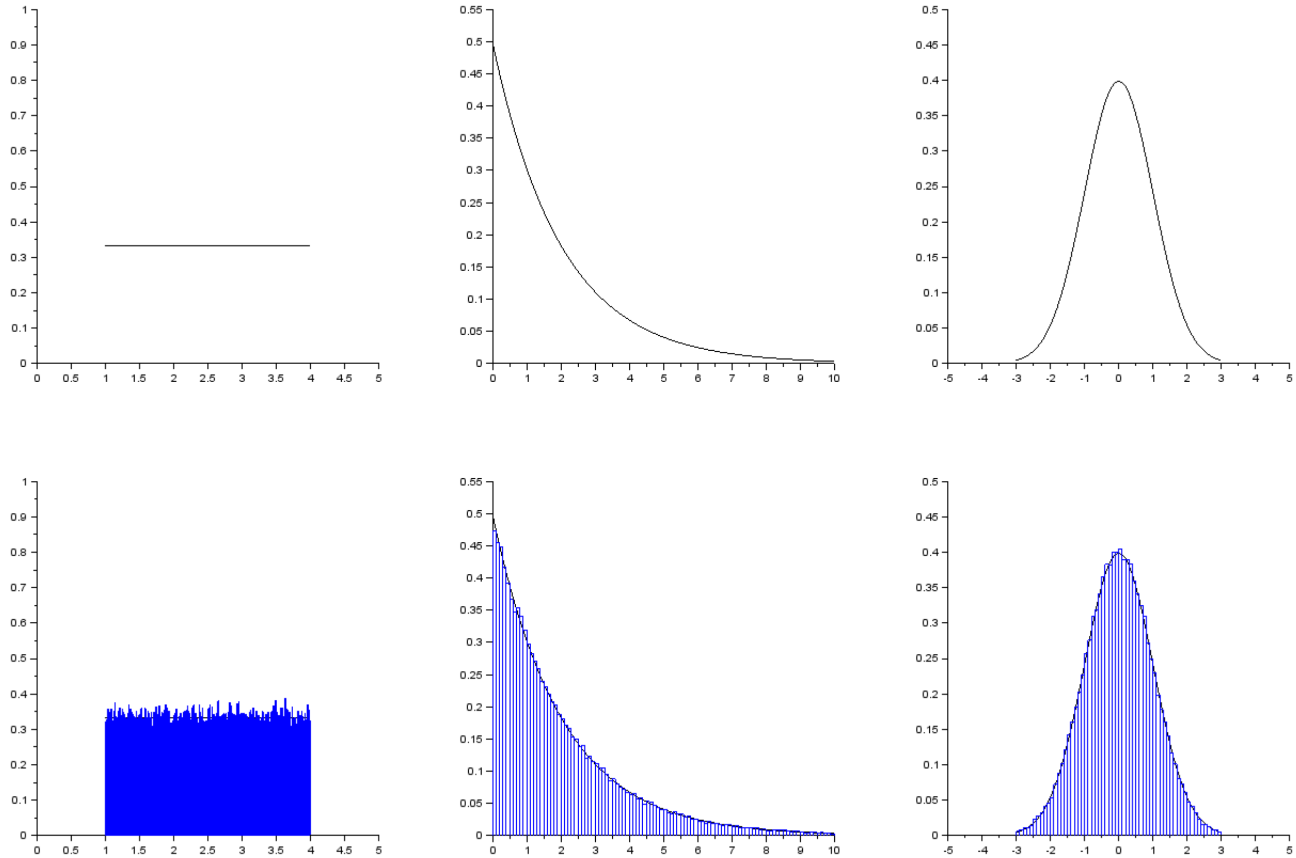
(1) Expliquer les trois scripts suivants

```
//script 1

x=1:0.01:4;
y=ones(1,301)./3;
plot2d(x,y)
N=10000;
t=grand(1,N,"unf",1,4);
histplot(x,t)
```

```
//script 2

lambda=1/2;
x=0:0.1:10;
y=lambda*exp(-lambda*x)
plot2d(x,y)
N=100000;
t=grand(1,N,"exp",1/
lambda);
histplot(200,t)
```

```
//script 3

x=-3*s:0.1:3*s;
y=exp(-x^2/2)/sqrt(2*%pi
)
plot2d(x,y)

N=100000;
mu=0;
s=1;
t=grand(1,N,"nor",mu,s);
histplot(x, t)
```

(2) Commenter alors les graphiques obtenus, reproduits ci-dessous.



2 La fonction `cdfnor()`

Concernant la **fonction de répartition** de la loi normale $\mathcal{N}(\mu, \sigma^2)$, la fonction `cdfnor()` permet de "tout" calculer. Mais il faut être vigilant quant aux arguments qu'elle prend. Plus précisément:

- `[P,Q]=cdfnor("PQ",x,mu,sigma)` correspond à la formule

$$P = \Phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \mathrm{d}t, \qquad Q = 1 - P.$$

- `X=cdfnor("X",mu, sigma, P, Q)` permet de calculer $\Phi^{-1}_{\mu,\sigma^2}(P)$ en connaissant μ, σ et P mais en précisant $Q = 1 - P$.
- `M=cdfnor("Mean", sigma, P, Q, x)` sert à calculer l'espérance d'une variable aléatoire $Z \hookrightarrow \mathcal{N}(\mu, \sigma^2)$ connaissant σ et $P(Z \leq \mathtt{x}) = P$.
- `S=cdfnor("Std", P, Q, x, mu)` sert à résoudre le problème analogue avec μ connu mais pas σ (attention à l'ordre des arguments!).

Exercice 4. (D'après **ECRICOME 2009**)

Sur une ligne de bus, une enquête a permis de révéler que le retard (ou l'avance) sur l'horaire officiel du bus à une station donnée, peut être représenté(e) par une variable aléatoire réelle, notée X, exprimée en minutes, qui suit une loi normale $\mathcal{N}(m, \sigma^2)$.

On admet de plus que la probabilité que le retard soit inférieur à 7 minutes est égale à $p = 0.8413$ et que l'espérance de X est de 5 minutes.

(1) Déterminer la valeur de σ.
(2) Quelle est la probabilité que le retard soit supérieur à 9 minutes ?

Exercice 5. Completer le script ci-dessous afin qu'il permette de comparer la fonction de répartition d'une variable aléatoire $X \hookrightarrow \mathcal{N}(0,1)$ et la fonction de répartition empirique à partir d'un échantillon de taille 1000 sur l'intervalle $[-2;2]$ avec 50 points.

```
function y=repartition(x)
    y=..................................
endfunction

X=-5:0.1:5;
Y=feval(X, repartition)
plot2d(X,Y, style=5, leg="répartition_théorique")


U=grand(1, 1000, 'nor',0,1);

function y=rep_emp(x)
    y=...................................
endfunction

Z=feval(X, rep_emp)
plot2d(X,Z, style=2, leg="répartition_empirique")
```

3 Méthode d'inversion

Exercice 6. (Extrait de **EML 2015**) On considère une v.a. U telle que $U \hookrightarrow \mathcal{U}([0;1])$ et on introduit la v.a. V définie par
$$V = -\frac{1}{\lambda}\ln(1-U),$$
où λ est un réel strictement positif.

(1) Montrer que $V \hookrightarrow \mathcal{E}(\lambda)$.
(2) En déduire l'écriture d'une fonction `SciLab`, notée `y=loi_expo(lambda)` qui, prenant en argument un paramètre `lambda` renvoie une simulation de la loi exponentielle $\mathcal{E}$(`lambda`).

Exercice 7. (Extrait de **EDHEC 2019**) Dans cet exercice θ désigne un réel de $]0;1/2[$. On considère la fonction
$$f(x) = \begin{cases} \frac{1}{\theta x^{1+1/\theta}}, & \text{si } x \geq 1 \\ 0, & \text{sinon.} \end{cases}$$

(1) Montrer que f peut être considérée comme une densité de probabilité. On note X une variable aléatoire de densité f.
(2) Préciser la fonction de répartition de X.
(3) Montrer que X admet espérance et variance et les calculer.
(4) On pose $Y = \ln(X)$. Montrer que Y suit une loi exponentielle dont on précisera le paramètre.
(5) On rappelle qu'en `SciLab`, la commande `grand(1,1, "exp", 1/lambda)` simule une variable aléatoire suivant la loi exponentielle de paramètre λ. Écrire des commandes `SciLab` utilisant `grand()` et permettant de simuler X.

Exercice 8. (Extrait de **HEC 2015**) On considère une v.a. $X \hookrightarrow \mathcal{E}(\lambda)$ (où $\lambda > 0$) et F la fonction définie sur $\mathbb{R}$ par
$$F(x) = \exp\left(e^{-\lambda x}\right).$$
On *peut montrer* que F est la fonction de répartition d'une v.a. T (loi de Gumbel de paramètre λ). On peut aussi montrer que
$$Z = -\frac{1}{\lambda}\ln(\lambda X)$$
suit la même loi que T.

(1) On considère $\lambda = 1$. Écrire en `SciLab` les commandes qui permettent de simuler la loi de T.
(2) Montrer que Z et T ont la même loi.

Exercice 9. (D'après **HEC 2017**) Pour $a, b > 0$, on introduit la fonction $f_{a,b}$ définie sur $\mathbb{R}$ par
$$f_{a,b}(x) = \begin{cases} (a+bx)\exp\left(-ax - \frac{b}{2}x^2\right), & \text{si } x \geq 0 \\ 0, & \text{si } x < 0. \end{cases}$$

(1) Montrer que $f_{a,b}$ est une densité de probabilité et préciser la fonction de répartition de la variable aléatoire correspondante.

On dira qu'une variable aléatoire X ayant pour densité $f_{a,b}$ suit une loi exponentielle linéaire de paramètres a et b, ce qu'on notera $X \hookrightarrow \mathcal{E}_\ell(a,b)$.

(2) Soit $Y \hookrightarrow \mathcal{E}(1)$. On pose
$$X = \frac{-a + \sqrt{a^2 + 2bY}}{b}.$$
  (a) Montrer que $X \hookrightarrow \mathcal{E}_\ell(a,b)$.
  (b) En déduire l'écriture d'une fonction `y=grandlinexp(a,b,n)` permettant de générer un vecteur colonne de taille n dont chaque composante suit une loi exponentielle linéaire de paramètres a et b.

4 Symétrisation

Exercice 10. Soit Y une variable aléatoire suivant la loi exponentielle de paramètre $\lambda > 0$ et soit $\mathcal{E}$ une variable aléatoire suivant une loi uniforme sur $\{-1, 1\}$ et indépendante de Y. On pose $X = \mathcal{E}Y$.

(1) En utilisant un s.c.e associé à $\mathcal{E}$, montrer que la fonction de répartition de X est donnée par :
$$F_X(x) = \begin{cases} \dfrac{e^{\lambda x}}{2} & \text{si } x < 0 \\ 1 - \dfrac{e^{-\lambda x}}{2} & \text{si } x \geq 0 \end{cases}$$
(2) En déduire que X est une variable aléatoire à densité et qu'une densité de X est $f_X(x) = (\lambda/2)e^{-\lambda|x|}$.
(3) Quelle est la loi suivie par la variable $Z = -(1/\lambda)\ln(1-U)$ lorsque U suit la loi uniforme sur $]0;1[$?
(4) Compléter le programme suivant pour qu'il simule une variable aléatoire X suivant la loi de Laplace. On rappelle que $X = \mathcal{E}Y$ est défini en préambule.

```
function X=Laplace(lambda)
    if .......... then
        eps=-1
    else
        eps=1
    end
    U=rand() ;
    Y=............
    X=............
endfunction
```

Autre(s) Exercice(s)

Soit a un réel strictement positif.

(1) Pour tout entier n supérieur ou égal à 2, on pose :

$$I_n(a) = \int_a^{+\infty} \frac{1}{t^n} \mathrm{d}t$$

Montrer que l'intégrale $I_n(a)$ converge et vaut $\dfrac{1}{(n-1)\,a^{n-1}}$.

(2) Soit f la fonction définie sur $\mathbb{R}$ par :

$$f(t) = \begin{cases} 0, & \text{si } t < a \\ \dfrac{3\,a^3}{t^4}, & \text{si } t \geq a \end{cases}$$

(a) Démontrer que f est bien une densité de probabilté. Soit X une variable aléatoire admettant f pour densité.
(b) Donner la fonction de répartition de X.
(c) Démontrer que X admet une espérance et calculer cette espérance.
(d) Démontrer que X admet une variance et que celle-ci vaut $\dfrac{3\,a^2}{4}$.

(3) Soit U une variable aléatoire suivant la loi uniforme sur $]0,1]$. On pose : $Y = \dfrac{a}{U^{\frac{1}{3}}}$.

(a) Déterminer $Y(\Omega)$.
(b) Déterminer la fonction de répartition de Y et vérifier que Y et X suivent la même loi.
(c) Écrire une fonction en langage `SciLab` d'en-tête `function Y = simulX(a, m, n)` prenant en argument un réel a strictement positif et deux entiers naturels m et n non nuls, qui renvoie une matrice à m lignes et n colonnes dont chaque coefficient est un réel choisi de façon aléatoire et indépendante en suivant la loi de X.

(4) (a) Calculer $P(X > 2a)$.
(b) Calculer $P_{X>2a}(X > 6a)$.
(c) On suppose que la fonction `SciLab` ci-avant a été programmée correctement. Compléter le script ci-dessous afin qu'il renvoie une valeur permettant de vérifier le résultat de la question précédente.

```
a=10
N=10000
s1=0
s2=0
X = simulX(a, 1, N)
for k=1:N
    if .................... then
        s1=s1+1
        if X(k) > 6*a  then
            .........................
        end
    end
end
if s1>0 then
    disp(.......................)
end
```