TP 9 : Simulation de lois à densité

1. La méthode d'inversion (de la fonction de répartition)

La méthode d'inversion prouve que l'on peut simuler toutes les variables aléatoires à densité dont on connait explicitement l'expression de la fonction de répartition à partir d'une uniforme sur $[0;1]$.
En effet, les lois à densité sont toutes fonction d'une uniforme, dès lors que l'on émet quelques hypothèses peu restrictives, la plus importante étant de connaitre l'expression de F.
On remarque donc directement que la méthode ne permettra pas de simuler une loi normale.
On rappelle au passage que l'on simule une loi normale par approximation à l'aide du théorème central limite.

a) Etude théorique de la méthode

Exercice 1. Soit X une variable aléatoire à densité dont une densité f est continue sur $\mathbb{R}$, strictement positive sur $\mathbb{R}$ sauf en un nombre fini de points.
On note F la fonction de répartition de X.

1. Démontrer que F réalise une bijection de $\mathbb{R}$ dans un ensemble que l'on explicitera.
   Dresser le tableau de variations de F^{-1}.
2. Soit $U \hookrightarrow \mathcal{U}([0;1])$. Justifier l'existence et déterminer la loi de $Z = F^{-1}(U)$.
3. En déduire une instruction en Scilab simulant une variable de même loi que X utilisant uniquement `rand()` (on supposera que la fonction F^{-1} était déjà prédéfinie dans Scilab).

Théorème 1 : méthode d'inversion générale

Soit X une variable aléatoire de support $X(\Omega) =]a;b[$, dont la fonction de répartition F est une bijection de $]a;b[$ sur $]0;1[$.
Si $U \hookrightarrow \mathcal{U}([0;1])$, alors $F^{-1}(U)$ suit la même loi que X.

b) Application à la simulation d'une exponentielle

Exercice 2. Soit X une variable aléatoire suivant une loi exponentielle de paramètre λ.

1. Vérifier que X suit bien les hypothèses du théorème 1.
2. Déterminer l'expression de la réciproque F^{-1} de la restriction à $X(\Omega)$ de sa fonction de répartition F.
3. Ecrire une `function X=inversion_exp(m,n,lambda)` qui renvoit une matrice de taille $m \times n$ dont chaque coefficient est la simulation d'une exponentielle de paramètre lambda par méthode d'inversion.
4. Ecrire une suite d'instructions qui prend en entrées n et λ, puis simule n variables aléatoires suivant une loi exponentielle de paramètre λ et trace, sur le même graphique, l'histogramme des fréquences des simulations obtenues sur le segment $[0;10]$, avec 50 classes de même largeur, ainsi que la courbe de la densité de X restreint à $[0;10]$.
   Exécuter le programme plusieurs fois, en gardant λ fixé ($\lambda = 0.5$ par exemple), et en augmentant n : $n = 100$ puis $n = 1000$ puis $n = 100000$.
5. L'instruction `grand(m,n,'exp',1/lambda)` renvoit une matrice de taille $m \times n$ dont chaque coefficient est la simulation d'une exponentielle de paramètre lambda.
   Comparer les résultats obtenus par méthode d'inversion et les résultats obtenus par `grand`.
6. Comparer, à l'aide de l'instruction `timer()`, les temps requis pour faire les calculs (avec $n = 100000$) par les deux méthodes. Commenter.

c) Autres exemples

Exercice 3. Soit X une variable aléatoire de fonction de répartition définie par $F(x) = \begin{cases} 0 & \text{si } x < 1 \\ 1 - \frac{1}{x^2} & \text{si } x \geq 1 \end{cases}$.

1. Vérifier rapidement que F vérifie les conditions du théorème.
2. Déterminer la réciproque de la restriction de F à $X(\Omega)$.
3. Utiliser la méthode d'inversion pour tracer l'histogramme de la loi empirique de X. On choisira la subdivision de classes $x = [0 : 0.1 : 10]$.
4. Tracer la densité (théorique) de X sur le même graphique.
5. Conclure.

Exercice 4. Soit X une variable aléatoire de fonction de répartition définie par $F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 2(1 - e^{-x}) & \text{si } x \in [0; \ln(2)] \\ 1 & \text{si } x \geq \ln(2) \end{cases}$.

1. Vérifier rapidement que F vérifie les conditions du théorème.
2. Déterminer la réciproque de la restriction de F à $X(\Omega)$.
3. Utiliser la méthode d'inversion pour tracer l'histogramme de la loi empirique de X. On choisira la subdivision de classes $x = [-1 : 0.01 : \ln(2) + 1]$.
4. Tracer la densité (théorique) de X sur le même graphique.
5. Conclure.

2. La fonction cdfnor

On rappelle que la fonction notée Φ désigne la fonction de répartition d'une loi normale centrée réduite. On ne connait pas l'expression de cette fonction donc elle est tabulée. On rappelle à ce propos quelques valeurs du tableau de Φ :
$\Phi(1,96) \simeq 0,975$, $\Phi(0,7) \simeq 0,758$.

• **Fonction `cdfnor` avec l'option `'PQ'` :**
L'instruction `cdfnor('PQ',X, Mean, Std)` ("cdf" est l'abréviation de "cumulative distribution fonction" i.e. fonction de répartition) renvoit, si on lui donne `X`, `Mean` et `Std`, la valeur de $F(X)$ où F est la fonction de répartition d'une loi normale d'espérance `Mean` et de variance `Std^2` :
Quelle valeur devrait renvoyer l'instruction `->cdfnor('PQ', 1.96,0,1)` ? Le vérifier.
De même demander la valeur de $\Phi(0,7)$ puis la valeur en $x = 3,4$ de la fonction de répartition d'une normale d'espérance 2 et de variance 4.
Retrouver la valeur obtenue par le calcul.
Remarque : L'option s'appelle `'PQ'` car on peut forcer la fonction à renvoyer 2 valeurs : la première, notée P est la valeur renvoyée ci-dessus ($P(X \leq x)$) , la seconde Q vaut $1 - P = 1 - P(X \leq x) = P(X \geq x)$.
Exemple : taper `-> [P,Q]=cdfnor('PQ',1.96,0,1)`. On obtient $P(X \geq 1,96) = 0,025$.

• **Fonction `cdfnor` avec l'option `'X'` :**
L'instruction `cdfnor('X',Mean,Std,P,Q)` renvoit, si on lui donne `Mean`, `Std`, `P` et `Q=1-P`, la valeur de $X = F^{-1}(P)$ où F est la fonction de répartition d'une loi normale d'espérance `Mean` et de variance `Std^2` :
Vérifier à l'aide de Scilab que $\Phi^{-1}(0,975) \simeq 1,96$.
Soit $X \hookrightarrow \mathcal{N}(0,1)$. A l'aide de Scilab, déterminer $x \in \mathbb{R}$ tel que $P(X \leq x) = 0,5$. Commenter.
Soit $X \hookrightarrow \mathcal{N}(1,9)$. A l'aide de Siclab, déterminer $x \in \mathbb{R}$ tel que $P(X \leq x) = 0,75$ puis $y \in \mathbb{R}$ tel que $P(X \geq y) = 0,05$.

• De même, l'instruction `cdfnor('Mean', Std,P,Q,X)` renvoit l'espérance de la loi normale si on lui donne l'écart type `Std` ainsi que les valeurs `P`, `Q` et `X` indiquant que pour cette loi normale on doit avoir $F(X) = P$ et $1 - F(X) = Q$ où F est sa fonction de répartition.
Enfin, l'instruction `cdfnor('Std',P,Q,X, Mean)` renvoit l'écart-type de la loi normale si on lui donne l'espérance `Mean` ainsi que les valeurs `P`, `Q` et `X` indiquant que pour cette loi normale on doit avoir $F(X) = P$ et $1 - F(X) = Q$ où F est sa fonction de répartition.
Exemples : soit $X \hookrightarrow \mathcal{N}(0, \sigma^2)$. On sait que $P(X \geq 1) = 0.25$. Déterminer sa variance.

3. Fonction de répartition empirique

Rappels sur les tests :
Taper l'instruction `->U=[ 0 9 8 3 4 5 6 8 7 1 2 8 9 ]` puis l'instruction `U <=5`.

Que fait cette instruction ?
Taper `sum(U <=5)`.
Comment expliquer le résultat ?

Algorithme de calcul de la fonction de répartition empirique :
Soit U une suite de réalisations d'une variable aléatoire X.
On peut estimer $F(t) = P(X \leq t)$ par la proportions de réalisations $U(i)$ dont la valeur est inférieure à t (fréquence des valeurs inférieures à t). Cette valeur $F_{emp}(t)$ sera appelée la valeur de la **fonction de répartition empirique** en t.

1. Donner l'instruction qui calcule cette fréquence.
2. En déduire une `function []=repartition_empirique(U, a,b,N)` qui :
   - divise l'intervalle $[a, b]$ en N subdivisions de même longueur et mémorise la suite des abscisses x_1, x_2, ...x_N obtenues dans un vecteur x.
   - Crée un vecteur $Femp$ mémorisant les valeurs $F_{emp}(x_1)$,$F_{emp}(x_2)$,...$F_{emp}(x_n)$ obtenues pour la simulation U ;
   - Trace le graphe de la fonction de répartition empirique obtenue en rouge.

Exercice 5.
1. Ecrire un programme qui demande n en entrée, effectue n simulations de variables suivant la loi normale centrée réduite et trace, sur le segment $[-5; 5]$ et avec 50 points, le graphe de la fonction de répartition empirique obtenue à partir de cet échantillon. Appliquer ce programme pour $n = 100$.
2. Rajouter dans la `function []=fct_rep_emp(U, a,b,N)` le tracé du graphe de la fonction de répartition (théorique) d'une normale centrée réduite, à l'aide de la fonction `cdfnor` avec l'option 'PQ'. Comparer.
3. Recommencer avec $n = 1000$, $n = 100000$.
4. Modifier pour comparer la fonction de répartition (théorique) d'une variable X suivant la loi exponentielle de paramètre 0.5 et la fonction de répartition empirique obtenue à partir d'un n-échantillon de même loi que X. On tracera les fonctions sur $[0; 10]$ avec 50 points.
   Exécuter le programme pour $n = 100$, puis $n = 1000$, $n = 100000$.
5. Modifier pour comparer la fonction de répartition (théorique) d'une variable X suivant la loi uniforme sur $[0, 1]$ et la fonction de répartition empirique obtenue à partir d'un n-échantillon de même loi que X. On tracera les fonctions sur $[-2; 3]$ avec 50 points.
   Exécuter le programme pour $n = 100$, puis $n = 1000$, $n = 100000$.