TP 4 : Simulation de lois usuelles

Rappels : L'instruction `rand()` renvoit un nombre choisi au hasard dans le segment $[0,1]$ (i.e. suivant une loi uniforme sur $[0;1]$), l'instruction `rand(m,n)` renvoit une matrice de taille $m \times n$ dont tous les coefficients sont des nombres choisis au hasard dans le segment $[0,1]$.

Théorie : la loi forte des grands nombres
Soient X_1, X_2, ..., X_n n variables aléatoires indépendantes et identiquement distribuées (de même loi) admettant une espérance m.
On pose

$$\overline{X_n} = \frac{X_1 + X_2 + ...X_n}{n}$$

alors

$$\boxed{\lim_{n \to +\infty} \overline{X_n} = m}$$

Autrement dit, la moyenne empirique de n simulations de la même loi converge vers la moyenne théorique (l'espérance).

En particulier, ce théorème permet d'affirmer que si on fait n simulations $(X_1, X_2, ...X_n)$ de la même loi discrète, alors :

$$\boxed{\text{la fréquence } f_n \text{ d'apparition d'une valeur } k \text{ converge vers } P(X_1 = k) \text{ lorsque } n \text{ tend vers } +\infty}$$

En effet, si on pose $Y_i = 1$ si $X_i = k$ et 0 sinon alors $Y_i \hookrightarrow B(P(X_1 = k))$ et
$f_n = \frac{Y_1+Y_2+...+Y_n}{n} \to E(Y_1) = P(X_1 = k)$ d'après la loi forte des grands nombres.

1) Les lois uniformes

a) Loi uniforme sur $[\![n_1; n_2]\!]$

Rappels : Si X est un nombre ENTIER choisi au hasard parmi les entiers n_1, $n_1 + 1$,$n_1 + 2$, ...n_2 alors X suit une loi uniforme sur $[\![n_1; n_2]\!]$.
Nous allons vérifier sur un exemple que l'instruction `grand(m,n,'uin',n1,n2)` simule "correctement" une matrice de taille $m \times n$ dont chaque coefficient suit une loi uniforme sur $[\![n_1; n_2]\!]$.

Remarque
Il est possible de simuler la même matrice de loi en remplaçant l'instruction `grand(m,n,'uin',n1,n2)` par l'instruction `floor((n2-n1)*rand(m,n)+n1)`.

Dans la suite on prend l'exemple de $n_1 = -2$ et $n_2 = 5$.

Exercice 1.

1. Créer un vecteur (ligne) $(U_1, U_2, ..., U_{10000})$ contenant un échantillon de 10000 simulations de la loi uniforme sur $[\![-2; 5]\!]$.
   Afficher les 100 premières valeurs et vérifier que les valeurs obtenues sont cohérentes.
2. **Convergence de la moyenne empirique :**
   (a) On considère l'instruction `--> M=cumsum(U)./[1:10000]`. Quelle est la suite (M_n) dont le vecteur `M` calculer les 100 premières valeurs ?
   (b) Remplir la suite d'instructions suivantes qui calcule autrement le vecteur M ; en utilisant la commande `mean` au lieu de la commande `cumsum` :

```
M=zeros(1,10000)
for n=1:10000 do
    M(n)=**********
end
```

(c) Expliquer et rajouter les instructions suivantes :

```
plot(1:10000, M, '+')
e=(-2+5)/2
plot2d([0 10000], [e e])
```

Exécute plusieurs fois le programme. La loi forte des grands nombres est-elle vérifiée ?

3. **Convergence des fréquences :** A l'aide de la fonction `tabul`, afficher le tableau de contingence (modalités-effectifs) de l'échantillon U. Commenter.
Créer un vecteur colonne f contenant la liste des fréquences des différentes valeurs prises par $(U_1, U_2, ..., U_{10000})$ puis représenter dans une nouvelle fenêtre le diagramme en bâton des fréquences à l'aide de l'instruction `bar`.
Commenter.

b) Loi uniforme sur $[a; b]$

Rappels : Si X est un nombre REEL choisi au hasard dans le segment $[a; b]$ alors X suit une loi uniforme sur $[a, b]$.
Nous allons vérifier sur un exemple que l'instruction `grand(m,n,'unf',a,b)` simule une matrice de taille $m \times n$ dont chaque coefficient suit une loi uniforme sur $[a; b]$.

Remarque

Il est possible de simuler la même matrice de loi en remplaçant l'instruction `grand(m,n,'unf',a,b)` par l'instruction `(b-a)*rand(m,n)+a`.

Dans la suite on prend l'exemple de $a = -2$ et $b = 5$.

Exercice 2.

Modifier le programme précédent pour répondre aux questions suivantes :

1. Créer un vecteur (ligne) $(U_1, U_2, ..., U_{10000})$ contenant un échantillon de 10000 simulations de la loi uniforme sur $[-2; 5]$.
Afficher les 100 premières valeurs et vérifier que les valeurs obtenues sont cohérentes.
2. **Convergence de la moyenne empirique :**
   (a) **Création de la suite de moyennes** $\left(M_n = \frac{U_1+..+U_n}{n}\right)$ **:**Créer un vecteur (ligne) M de taille 1×10000 dont le coefficient $M(n)$ contient $\left(\frac{U_1+..+U_n}{n}\right)$.

   (b) Représenter graphiquement la suite (M_n) et tracer sur le même graphique la fonction constante égale à l'espérance de la loi.
   Relancer plusieurs fois le programme : la loi forte des grands nombres est-elle vérifiée ?
3. **Convergence des fréquences :**
   (a) Peut-t-on construire le diagramme en bâton des fréquences comme précédemment ? Pourquoi ?
   Construire l'histogramme des fréquences correspondant aux classes $[-2; -1,9]$, $]-1,9; -1,8]$, ...$]4,9; 5]$.
   L'instruction `grand(m,n,'unf',a,b)` vous paraît-elle simuler des lois uniformes continues de manière satisfaisante ?
   Donner l'allure puis l'expression de la fonction f telle que la probabilité d'obtenir un nombre entre x_1 et x_2 est l'aire sous la courbe de f entre x_1 et x_2.

   (b) On rappelle que la densité de probabilité d'une loi uniforme sur $[a; b]$ est définie par $f(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{1}{b-a} & \text{si } x \in [a; b] \\ 0 & \text{si } x > b \end{cases}$. Comparer cette expression avec l'expression obtenue à la question précédente.
   Définir la fonction f puis tracer sur le même graphique la fonction f sur l'intervalle $[-3, 6]$.
   Commenter.

2) La loi géométrique

Exercice 3.

1. On rappelle que si X est un variable aléatoire donnant le rang d'apparition du premier succès au cours d'une suite indéfinie d'épreuves de Bernoulli indépendantes et de même paramètre p, alors X suit une loi géométrique de paramètre p.
   En utilisant uniquement cette propriété, écrire une `function x=geo(p)` qui effectue une simulation x d'une variable aléatoire X suivant une loi géométrique de paramètre p.
   On n'utilisera non pas la fonction `grand` *mais uniquement la fonction* `rand`.
2. A l'aide de la fonction `geo`, simuler 100 variables aléatoires géométriques de paramètre 1/3 et tracer le diagramme en bâton des fréquences obtenues.
3. **Comparaison avec la loi théorique** : Compléter la fonction suivante qui renvoit le vecteur `L=[L1,L2,...,L20]` contenant la loi d'une variable X géométrique de paramètre p jusqu'à $n = 20$; c'est-à-dire que `Li` contient la valeur de $P(X = i) = p(1-p)^{i-1}$:

```
function L=loigeometrique(p)
    L=zeros(1,20)
    L(1)=*************
    for i=2:20 do
        L(i)=***********
    end
endfunction
```

4. Tracer le diagramme en bâton de la loi théorique d'une géométrique. Comparer avec le diagramme des fréquences obtenu précédemment.
   Relancer le programme pour 1000 simulations puis pour 10000 simulations. Commenter.

Dorénavant, nous simulerons des variables suivant des lois géométriques par l'instruction `grand(m,n,'geom',p)` qui simule une matrice de taille $m \times n$ dont chaque coefficient suit une loi géométrique de paramètre p.

3) La loi binomiale

L'instruction `grand(m,n,'bin',k,p)` simule une matrice de taille $m \times n$ dont chaque coefficient est la simulation d'une variable aléatoire suivant la loi binomial de taille k et de paramètre p.
L'instruction `binomial(p,n)` renvoit le vecteur L de la loi de probabilité d'une variable X binomiale de taille n et de paramètre p, c'est-à-dire que la première case `L(1)` contient $P(X = 0)$, que la seconde case `L(2)` contient $P(X = 1)$,etc...`L(n)` contient $P(X = n - 1)$.
Attention à ne pas confondre ces 2 fonctions ! l'une simule une variable aléatoire (renvoit UN nombre aléatoire), l'autre donne sa loi (fixe).

Exercice 4.
Un secrétaire envoie un courrier nécessitant un réponse à 20 contacts.
On admet que les contacts répondent à tout message de façon indépendante et avec la probabilité $p = 1/4$.
On note X le nombre de réponses reçues.
La secrétaire envoie alors un message de rappel aux $20 - X$ contacts qui n'ont pas répondu. On note Y le nombre de réponses à ce second message et $Z = X + Y$ le nombre total de réponses.

1. Quelle est la loi de X ? Quelle est la loi de Y sachant $X = i$?
2. Compléter la fonction suivante qui simule 1 fois cette expérience et la valeur z prise par la variable Z :

```
function z=reponses()
    x=grand(1,1,****, *****,****)
    y=grand(1,1,****,****,****)
    z=x+y
endfunction
```

3. Visualiser la loi empirique de Z à l'aide d'un 1000-échantillon U de la loi de Z.

4. Construire sur le même graphique (taille des bâtons et couleur différentes), le diagramme en bâton de la distribution théorique d'une loi binomiale de taille 20 et de paramètre 7/16. *On utilisera l'instruction* `L=binomial(p,n)` qui construit le vecteur `L=[L1,L2,...,L21]` où `Li` contient la valeur de $P(X = i - 1)$.
   Emettre une conjecture.
5. En lisant le diagramme, donner une valeur approchée des probabilités $P(X = 0)$, $P(X = 1)$, $P(X = 2)$.

Exercice 5.
Ouvrir le fichier Scinote intitulé *TP4exo5officiel* et faire une copie du fichier dans votre dossier personnel. Le fichier vous propose le programme suivant :

```
clear
clf()
function X=inconnue(n,p)
    for i=1:n do
        U(i)=grand(1,1,'bin',1,p)
    end
    X=sum(U)
endfunction
n=10, p=1/3
for i=1:1000 do
    X(i)=inconnue(10,1/3)
end
M=tabul(X,'i')
x=M(:,1)
f=M(:,2)/1000
subplot(1,2,1)
bar(x,f,width=0.2,color='b')
legend('diagramme des fréquences de X')
subplot(1,2,2)
bar(0:10,binomial(1/3,10),width=0.2,color='r')
legend('loi d''une binomiale de taille 10 et de paramètre 1/3')
```

Lisez attentivement le programme et exécutez-le.
Quel théorème du cours est illustré par ce programme ?

4) La loi de Poisson

L'instruction `grand(m,n,'poi',lambda)` simule une matrice de taille $m \times n$ dont chaque coefficient est la simulation d'une variable aléatoire suivant la loi de Poisson de paramètre lambda.

Exercice 6.
On admet que le nombre N de têtards issus des oeufs pondus en mars et avril d'une année suit une loi de Poisson de paramètre lambda $= 20$.
Ces têtards sont soumis à des prédateurs nombreux et voraces, et on admet que chacun d'entre eux a une probabilité $p = 0,05$ de parvenir à son développement complet, et qu'ils se développent de façon indépendante.
On note X le nombre de têtards qui parviennent à leur développement complet et se transforment donc en une grenouille verte d'environ 2cm de long.

1. Quelle est la loi de X sachant $N = k$?
2. Ecrire une `function x=grenouilles()` qui simule cette expérience aléatoire et renvoit la valeur x prise par la variable X.
3. Mettre dans un vecteur U 1000 simulations de la loi de X et tracer le diagramme en bâton des fréquences cumulées obtenues.
4. En Scilab, `cdf` est l'abréviation de "cumulated distribution fonction" qui signifie fonction de répartition.
   Ainsi, l'instruction `cdfpoi("PQ",x,lambda)` renvoit la valeur de $F(x)$ où F est la fonction de répartition de la loi de Poisson de paramètre lambda.
   A la suite du programme, taper les instructions suivantes :

```
lambda=20
p=0.05
for i=1:10 do
   F(i)=cdfpoi("PQ", i-1,lambda*p)
end
bar(0:9, F, width=0.1,color='r')
```

Emettre une conjecture.

5. En lisant le diagramme, donner une valeur approchée des probabilités $P(X = 0)$, $P(X = 1)$, $P(X = 2)$.

Exercice 7.
Ouvrir le fichier Scinote intitulé *TP4exo7officiel* et faire une copie du fichier dans votre dossier personnel. Le fichier vous propose le programme suivant :

```
clear
clf()
function X=inconnue()
    for i=1:10 do
        U(i)=grand(1,1,'poi',i)
    end
    X=sum(U)
endfunction
for i=1:1000 do
    X(i)=inconnue()
end
M=tabul(X,'i')
x=M(:,1)
f=M(:,2)/1000
subplot(1,2,1)
bar(x,cumsum(f),width=0.2,color='b')
a=get("current_axes") ; //pour régler les options des axes
a.data_bounds([1 2]) = [0 100];
a.data_bounds([3 4]) = [0 1.2];
legend('diagramme des fréquences cumulées de X')
subplot(1,2,2)
for i=1:101 do
   FR(i)=cdfpoi("PQ",i-1,55)
end
bar(0:100,FR,width=0.2,color='r')
a=get("current_axes") ; //pour régler les options des axes
a.data_bounds([1 2]) = [0 100];
a.data_bounds([3 4]) = [0 1.2];
legend('fonction de répartition d''une Poisson de paramètre 55')
```

Lisez attentivement le programme et exécutez-le.
Quel nouveau théorème du cours est illustré par ce programme ?