TP 7 : Convergence en loi

I. Convergence en loi : le cas de variables discrètes

Définition 1 : si toutes les variables sont discrètes et de supports inclus dans $X(\Omega)$

Une suite (X_n) de v.a.r. discrètes converge en loi vers une v.a.r. X discrète (vers la loi de la variable X en réalité) si :

$$\forall k \in X(\Omega), \lim_{n\to+\infty} P(X_n = k) = P(X = k)$$

Exercice 1. Soit (X_n) une suite de variables aléatoires indépendantes telle que $X_n \hookrightarrow \mathcal{B}\left(n, \frac{\lambda}{n}\right)$. Nous allons vérifier que la suite (X_n) converge en loi vers une une variable $X \hookrightarrow P(\lambda)$ (de même espérance).

1. Soit X une variable aléatoire suivant une loi de Poisson de paramètre λ.
   Soit $k \in \mathbb{N}^*$. Rappeler la valeur de $p_k = P(X = k)$ et donner une relation de récurrence entre p_{k-1} et p_k.
   Compléter (sur papier) la fonction suivante qui calcule par récurrence le vecteur $P = [p_0, p_1, ..., p_{20}]$ tel que $p_k = P(X = k)$ où $X \hookrightarrow \mathcal{P}(lambda)$:

```
function P=loi_Poisson(lambda)
    P=zeros(1,21)
    P(1)=..........
    for k=1:20 do
        P(k+1)= ........
    end
endfunction
```

2. A la suite, on rajoute les instructions suivantes :

```
lambda=1
n=input('Entrer n: ')
X=grand(1,1000,....,....,....)

M=tabul(X,"i")
x=M(:,1)
f=M(:,2)/1000
bar(x,f)
bar(0:20,loi_Poisson(lambda),width=0.2, color='red')
```

   (a) Compléter (sur papier) le programme de façon que le vecteur ligne `X` contienne 1000 simulations de la loi binomaile de paramètres n et $\frac{\lambda}{n}$.
   (b) Que contient le vecteur ligne `x` ? le vecteur ligne `f` ?
   Que fait l'instruction `bar(x,f)` ?, l'instruction `bar(0:20,loi_Poisson(lambda),width=0.2, color='red')` ?

3. Ouvrir, compléter et lancer le programme*tp7exo1officiel* pour $n = 10$ puis $n = 100$, $n = 1000$. Conclure.

Exercice 2 (: ecricome 2017 (sur papier)). Soit n un entier naturel non nul.
On effectue une série illimité de tirages d'une boule avec remise dans une urne contenant n boules numérotées de 1 à n. Pour tout entier naturel k non nul, on note X_k la variable aléatoire égale au numéro de la boule obtenue au k-ième tirage.
Pour tout entier naturel k non nul, on note S_k la somme des numéros des boules obtenues lors des k premiers tirages :

$$S_k = \sum_{i=1}^{k} X_i.$$

On considère enfin la variable aléatoire T_n égale au nombre de tirages nécessaires pour que, pour la première fois, la somme des numéros des boules obtenues soit supérieure ou égale à n.

Exemple : avec $n = 10$, si les numéros obtenus aux cinq premiers tirages sont dans cet ordre $2, 4, 1, 5, 9$, alors on obtient : $S_1 = 2,\ S_2 = 6,\ S_3 = 7,\ S_4 = 12,\ S_5 = 21$ et $T_{10} = 4$.

1. Loi de X_k :
   Pour $k \in \mathbb{N}^*$, déterminer la loi de X_k. Rappeler son espérance et sa variance.
2. Loi de S_k :
   (a) Pour $k \in \mathbb{N}^*$, déterminer $S_k(\Omega)$.
   (b) Soit $k \in [\![1, n-1]\!]$.
      i. Exprimer S_{k+1} en fonction de S_k et de X_{k+1}.
      ii. En utilisation un système complet d'événements lié à la variable aléatoire S_k, démontrer alors que :
      $$\forall i \in [\![k+1, n]\!],\ \mathrm{P}(S_{k+1} = i) = \frac{1}{n}\sum_{j=k}^{i-1} \mathrm{P}(S_k = j).$$
   (c) i. Pour $k \in \mathbb{N}^*$ et $j \in \mathbb{N}^*$, rappeler la formule du triangle de Pascal liant les nombres : $\binom{j-1}{k-1}$, $\binom{j-1}{k}$ et $\binom{j}{k}$.
      ii. En déduire que pour tout $k \in \mathbb{N}^*$ et pour tout entier naturel i supérieur ou égal à $k+1$:
      $$\sum_{j=k}^{i-1} \binom{j-1}{k-1} = \binom{i-1}{k}.$$
      iii. Pour tout entier $k \in [\![1, n]\!]$, on note $\mathcal{H}_k$ la proposition :
      $$\text{« } \forall i \in [\![k, n]\!],\ \mathrm{P}(S_k = i) = \frac{1}{n^k}\binom{i-1}{k-1} \text{ ».}$$
      Démontrer par récurrence que pour tout entier $k \in [\![1, n]\!]$, $\mathcal{H}_k$ est vraie.
3. Loi de T_n :
   (a) Déterminer $T_n(\Omega)$.
   (b) Soit $k \in [\![1, n-1]\!]$. Comparer les événements : $[T_n > k]$ et $[S_k \leq n-1]$.
   (c) En déduire que : $\forall k \in [\![0, n]\!],\ \mathrm{P}(T_n > k) = \frac{1}{n^k}\binom{n-1}{k}$.
4. convergence en loi de la suite (T_n) :
   (a) Pour tout entier naturel k non nul, démontrer que :
   $$\lim_{n \to +\infty} \mathrm{P}(T_n > k) = \frac{1}{k!}.$$
   (b) Démontrer alors que $(T_n)_{n \geq 1}$ converge en loi vers une variable aléatoire Y à valeurs dans $\mathbb{N}^*$ telle que : $\forall k \in \mathbb{N}^*,\ \mathrm{P}(Y = k) = \frac{k-1}{k!}$.
5. On rappelle qu'en langage Scilab, l'instruction `grand(1,1,'uin',1,n)` renvoie un entier aléatoire de $[\![1, n]\!]$.[1] Compléter la fonction ci-dessous, qui prend en argument le nombre n de boules contenues dans l'urne, afin qu'elle simule la variable aléatoire T_n :

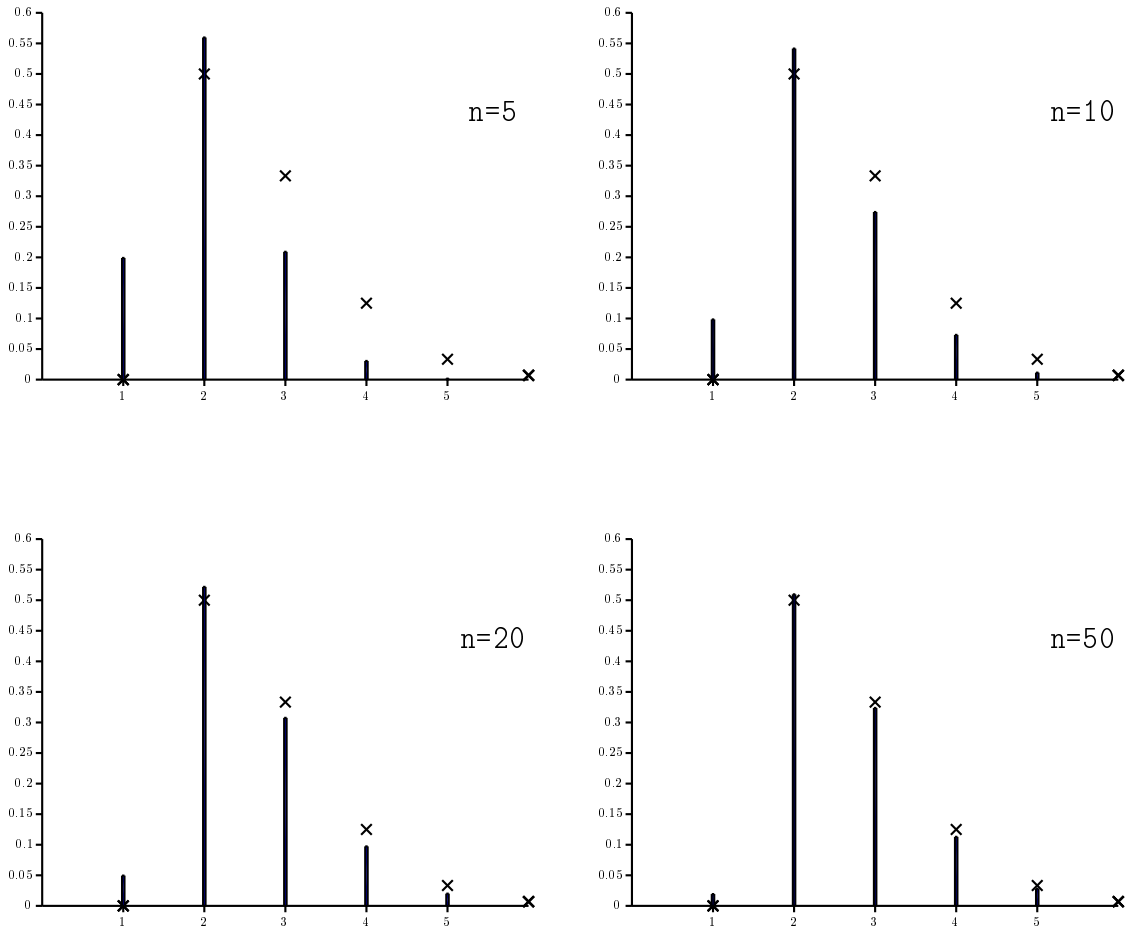
```
function y=T(n)
    S = ..........
    y = ..........
    while ..........
        tirage = grand(1,1,'uin',1,n)
        S = S + tirage
        y = ..........
    end
endfunction
```

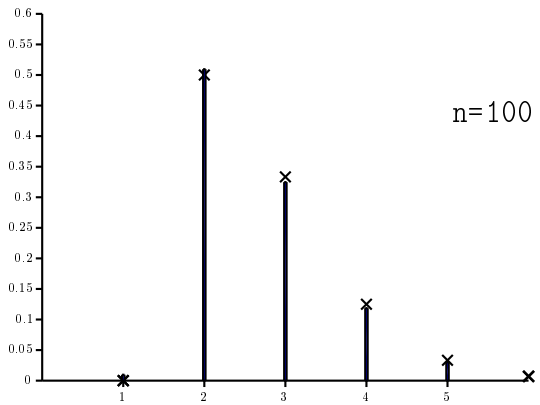
1. L'énoncé Ecricome oublie de préciser « selon la loi uniforme ».

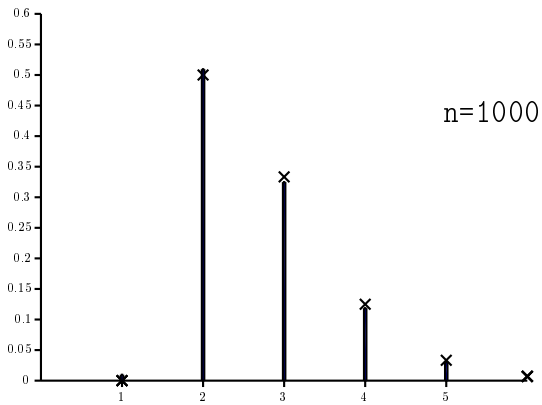
6. On suppose déclarée la fonction précédente et on écrit le script ci-dessous :

```
function y=freqT(n)
    y = zeros(1,n)
    for i=1:100000
        k = T(n)
        y(k) = y(k)+1
    end
    y = y/100000
endfunction

function y=loitheoY(n)
    y = zeros(1,n)
    for k=1:n
        y(k) = (k-1)/prod(1:k)
    end
endfunction

clf
n = input('n=?')
plot2d(loitheoY(6),style=-2)
x = freqT(n)
bar(x(1:5))
```

L'exécution de ce script pour les valeurs de n indiquées a permis d'obtenir les graphes ci-dessous :







(a) Expliquer ce que représentent les vecteurs renvoyés par les fonctions `freqT` et `loitheoY`. Comment ces vecteurs sont-ils représentés graphiquement dans chaque graphique obtenu ?

(b) Expliquer en quoi cette succession de graphiques permet d'illustrer le résultat de la question 13.

II. Convergence en loi : cas général

Définition 2

De manière générale, un suite (X_n) converge en loi vers une variable X (vers la loi de la variable X en réalité) si :

$$\forall x \in \mathbb{R}, \lim_{n \to +\infty} F_{X_n}(x) = F_X(x)$$

Autrement dit ; la fonction de répartition de X_n doit converger en tout point vers celle de X.

Exercice 3. Soit (U_n) une suite de variables aléatoires indépendantes suivant toutes la loi uniforme sur $[0;1]$.
On note $M_n = max(U_1, U_2, ...U_n)$ et $X_n = n(1 - M_n)$.
Nous allons étudier la convergence en loi de la suite de variables aléatoires (X_n).

1. Compléter (sur papier) la fonction suivante qui effectue une simulation x de la variable X_n .

```
function x=simulation_X(n)
   u=grand(1,n,'unf',0,1)
   m=................
   x=................
endfunction
```

2. Compléter (sur papier) le programme suivant qui effectue 1000 simulations de la variable X_n (à l'aide de 1000 appels à la fonction `simulation_X`, rangées dans un vecteur ligne `s`, et afficher l'histogramme des fréquences obtenu (on découpera l'intervalle des valeurs en 100 classes de même longueur).

```
n=input('entrer n:')
s=zeros(1,1000)
for k=1:1000 do
    s(k)= ..........
end
histplot(100 ,s)
```

3. On rappelle que la densité de probabilité d'une loi exponentielle de paramètre 1 est $f(x) = 0$ si $x < 0$ et $f(x) = exp(-x)$ si $x \geq 0$.
Compléter et lancer le programme *TP7exo3officiel* pour $n = 2$, $n = 5$, $n = 10$ et $n = 100$ et émettre une conjecture.

4. Nous allons vérifier la convergence en loi en terme de convergence de la fonction de répartition comme demandé dans la définition :

   (a) **Point cours : fonction de répartition empirique**
   La fonction de répartition d'une variable aléatoire X_n est définie par $F_{X_n}(t) = P(X_n \leq t)$.

Pour obtenir une valeur approchée de cette fonction de répartition, on calcule la fréquence, $F_{\text{emp}}(t)$, de l'évènement $(X_n \leq t)$ sur la liste s des simulations de X_n. La fonction ainsi définie est appelée **fonction de répartition empirique de** X_n.

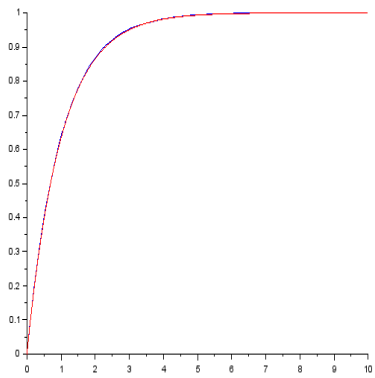
Que fait l'instruction suivante ?

```
F_emp=sum(s <=t)/length(s)
```

(b) Compléter (sur papier) le programme suivant qui :
- Construit un vecteur $x = [x_1, x_2, ..., x_{1000}]$ qui divise le segment $[0; 10]$ de l'axe des abscisses en 1000 valeurs équi-réparties ;
- Construit un vecteur $F = [F_1, F_2, ..., F_{1000}]$ tel que pour tout k, F_k est une estimation de $F_{X_n}(x_k) = P(X_n \leq x_k)$;
- Trace la courbe de la fonction de répartition empirique ainsi créée, ainsi que la courbe de la fonction de répartition d'une loi exponentielle de paramètre 1.

```
scf(1)
x=linspace(0,10,1000)
for k=1:1000 do
    F_emp(k)=............
end
plot2d(x,F_emp)

function y=F(x)
    if x<0 then
        y=0
    else
        y=1-exp(-x)
    end
endfunction
fplot2d(x,F,style=5)
```

On obtient le graphique suivant :



Commenter.

Exercice 4 (: edhec 2017 (sur papier)).

1. Soit V une variable aléatoire suivant une loi exponentielle de paramètre 1, c'est-à-dire admettant pour fonction de répartition la fonction F définie sur $\mathbb{R}$ par $F(x) = \begin{cases} 0 & \text{si} x < 0 \\ 1 - e^{-x} & \text{si } x \geq 0 \end{cases}$
Déterminer la fonction de répartition de $W = -\ln(V)$ (*on dit que W suit une loi de Grumbel*).

Dans la suite, on désigne par n un entier naturel non nul et par $X_1, \ldots, X_n$ des variables aléatoires définies sur le même espace probabilisé, indépendantes et suivant la même loi que V, c'est à dire la loi $\mathcal{E}(1)$.

On considère la variable aléatoire Y_n définie par $Y_n = \max(X_1, X_2, \ldots, X_n)$, c'est à dire que pour tout ω de Ω, on a : $Y_n(\omega) = \max(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$. On admet que Y_n est une variable

aléatoire à densité.

2. Montrer que la fonction de répartition F_{Y_n} de Y_n est définie par :

$$F_{Y_n}(x) = \begin{cases} 0 & \text{si } x < 0 \\ (1-e^{-x})^n & \text{si } x \geq 0 \end{cases}$$

3. On pose $Z_n = Y_n - \ln(n)$.
   (a) On rappelle que `grand(1,n,'exp',1)` simule n variables aléatoires indépendantes et suivant toutes la loi exponentielle de paramètre 1.
   Compléter la déclaration de fonction `Scilab` suivante afin qu'elle simule la variable aléatoire Z_n.

```
function Z=f(n)
x=grand(1,n,'exp',1)
Z=...
endfunction
```

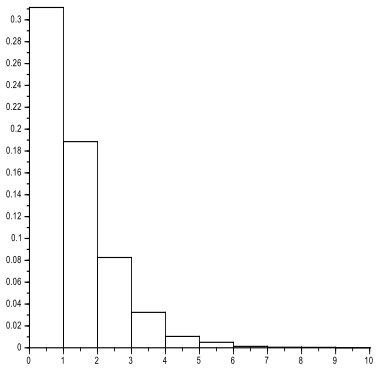
   (b) Voici deux scripts :
   - script (1)

```
V=grand(1,10000,'exp',1)
W=-log(V)
s=linspace(0,10,11)
histplot(s,W)
```

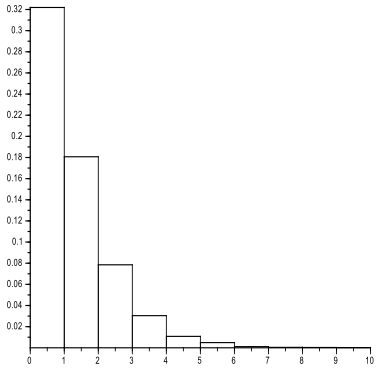
   - script (2)

```
n=input('entrez la valeur de n : ')
Z=[]  // la matrice-ligne Z est vide
for k=1:10000
   Z=[Z,f(n)]
end
s=linspace(0,10,11)
histplot(s,Z)
```

   Chacun des scripts simule 10000 variables indépendantes, regroupe les valeurs renvoyées en 10 classes qui sont les intervalles $[0,1]$, $]1,2]$, $]2,3]$, ..., $]9,10]$ et trace l'histogramme correspondant (la largeur de chaque rectangle est égale à 1 et leur hauteur est proportionnelle à l'effectif de chaque classe).
   Le script (1) dans lequel les variables aléatoires suivent la loi de Grumbel (loi suivie par W), renvoie l'histogramme (1) ci-dessous, alors que le script (2) dans lequel les variables aléatoires suivent la même loi que Z_n, renvoie l'histogramme (2) ci-dessous, pour lequel on a choisi $n = 1000$.



Histogramme (1)



Histogramme(2) pour $n = 1000$

Quelle conjecture peut-on émettre quant au comportement de la suite des variables aléatoires (Z_n).

4. On note F_{Z_n} la fonction de répartition de Z_n.
   (a) Justifier que, pour tout réel x, on a : $F_{Z_n}(x) = F_{Y_n}(x + \ln(n))$.
   (b) Déterminer explicitement $F_{Z_n}(x)$.
   (c) Montrer que, pour tout réel x, on a : $\lim\limits_{n \to +\infty} n \ln\left(1 - \dfrac{e^{-x}}{n}\right) = -e^{-x}$.
   (d) Démontrer le résultat conjecturé à la question 2)b.