DS 9

Thème : Informatique

On suppose importé dans la suite les bibliothèques `numpy`, `numpy.random`, `matplotlib.pyplot`, respectivement avec les alias `np`, `rd` et `plt`. On rappelle aussi que l'on code une matrice rectangulaire par `A=np.array([[ligne 1], ... [dernière ligne]])` et le coefficient en position $(i+1, j+1)$ s'obtient avec un décalage : `A[i,j]`.

Exercice 1 : Simulation d'une loi géométrique (tronquée)

1. En utilisant uniquement la commande `rd.random()` (qui renvoie un nombre pris au hasard dans $[0;1[$), écrire un programme qui prend en argument $p \in]0;1[$ et simule une variable aléatoire X de loi géométrique de paramètre p.

2. Soient $N \in \mathbb{N}^*$ et $p \in [0;1]$. La variable X suit une loi géométrique tronquée si $X(\Omega) = [\![1;N]\!]$ et

$$\forall k \in [\![1;N-1]\!], \qquad \mathbf{P}(X=k) = pq^{k-1} \quad \text{et} \quad \mathbf{P}(X=N) = 1-q^N.$$

Adapter le programme précédent pour écrire une fonction qui prend en arguments N, p et simule la variable X.

Exercice 2 : Nombres de Stirling de première espèce

On définit les nombres de Stirling, notés $s(n,k)$, par

$$\forall n \in \mathbb{N}^* \quad s(n,0) = s(0,n) = 0 \quad \text{et} \quad s(0,0) = 1$$

$$\forall n \in \mathbb{N}, \quad \forall k \in \mathbb{N}^*, \qquad s(n+1,k) = s(n,k-1) - ns(n,k).$$

3. Compléter le programme suivant d'arguments $N, K \in \mathbb{N}^*$ et qui renvoie la matrice $S_{N,K} = (s(n,k))_{n\in[\![0;N]\!],\ k\in[\![0;K]\!]}$.

Editeur

```python
def Stirling( ...... ):
    S=np.zeros( ...... )
    S[0,0]=1
    for n in range( ..... ):
        for k in range( ...... ):

            S[n+1,k]= .......
    return S
```

Console

```
# un test
>>> Stirling(5,5)
array([[  1.,   0.,   0.,   0.,   0.,   0.],
       [  0.,   1.,   0.,   0.,   0.,   0.],
       [  0.,  -1.,   1.,   0.,   0.,   0.],
       [  0.,   2.,  -3.,   1.,   0.,   0.],
       [  0.,  -6.,  11.,  -6.,   1.,   0.],
       [  0.,  24., -50.,  35., -10.,   1.]])
```

4. En déduire un programme qui prend en argument n et renvoie la somme

$$c_n = \sum_{k=0}^{n} (-1)^k s(n,k).$$

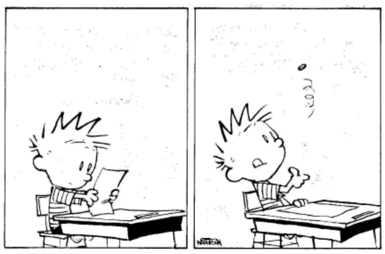
5. À l'aide de ces quelques tests, conjecturer une formule simple pour c_n.

Console

```
>>> for n in range(0,7):
    print(somme(n))
1.0
-1.0
2.0
-6.0
24.0
-120.0
720.0
```

Console

```
>>> for n in range(0,7):
    print(somme(n+1)/somme(n))
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
-7.0
```



Calvin et Hobbes

Problème A : Suite de Syracuse

Soit $N \in \mathbb{N}^*$ fixé. On définit la suite $(u_n)_n$ par la récurrence :

- $u_0 = N$;
- et pour tout entier naturel n :

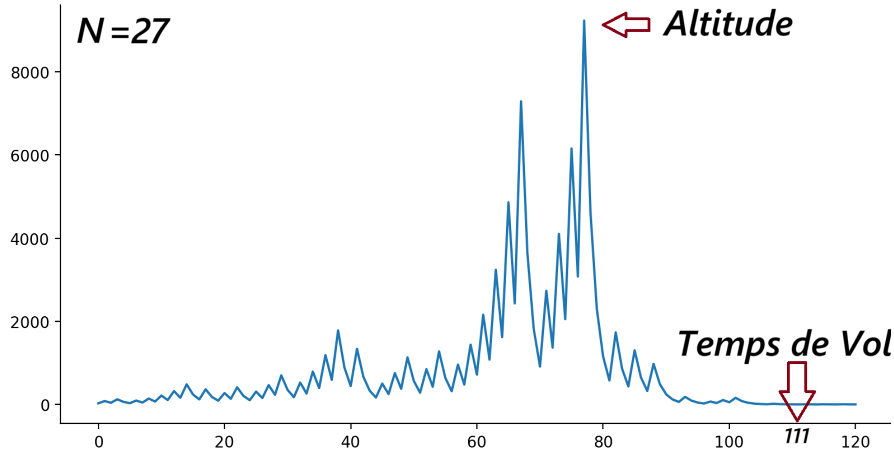
$$u_{n+1} = \begin{cases} u_n/2 & \text{si } u_n \text{ est pair,} \\ 3u_n + 1 & \text{si } u_n \text{ est impair.} \end{cases}$$

6. Compléter le programme qui prend en arguments N, n et renvoie la matrice ligne constituée des $n+1$ premiers termes de la suite de Syracuse initialisée à N : $[u_0 \quad u_1 \quad u_2 \quad \dots \quad u_n]$.

Editeur

```
def syracuse(n,N):
    U= ...
    U[0]= ...
    for i in range(0,n):
        if  ...   :
            U[i+1]=  ...
        else :
            U[i+1]=  ...
    return U
```

On suppose maintenant que, pour tout entier $N \geqslant 1$, il existe toujours un entier n tel que $u_n = 1$. On définit alors V_N, le temps de vol de N comme le plus petit indice n tel que $u_n = 1$ où $(u_k)_{k\in\mathbb{N}}$ désigne la suite de Syracuse initialisée avec N. On définit aussi l'altitude A_N de N comme la plus grande valeur obtenue par la suite (u_n).
Ci-dessous, on a représenté les 120 premiers termes de la suite initialisée avec N = 27.

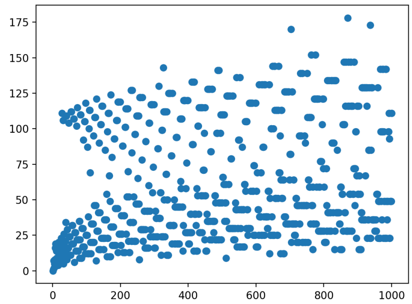


7. Écrire une fonction d'entête `def Vol(N):` renvoyant le temps de vol V_N.

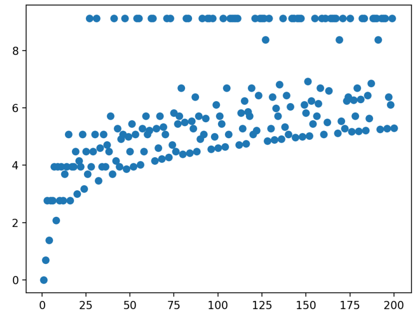
8. Faire de même avec l'altitude A_N.

9. Que font les programmes suivants ?

Console
```
plt.clf()
N=1000
V=np.zeros(N)
for i in range(1,N):
    V[i-1]=vol(i)
plt.plot(np.arange(1,N+1),V,'o')
plt.show()
```



Console
```
N=200
M=np.ones(N)
for i in range(1,N+1):
    V[i-1]=vol(i)
    M[i-1]=Altitude(i)
plt.plot(np.arange(1,N+1),np.log(M),'o')
plt.show()
```



Problème B : Étude d'un minimum

10. Soit Z, une variable aléatoire. On suppose que l'on dispose d'une commande `simuZ()` qui permet de simuler la variable Z. Donner un programme `Frep` qui prend en argument un réel x et donne une approximation de $F_Z(x) = \mathbf{P}(Z \leqslant x)$.

11. Dans la suite, on fixe $n \in \mathbb{N}^*$, et on pose

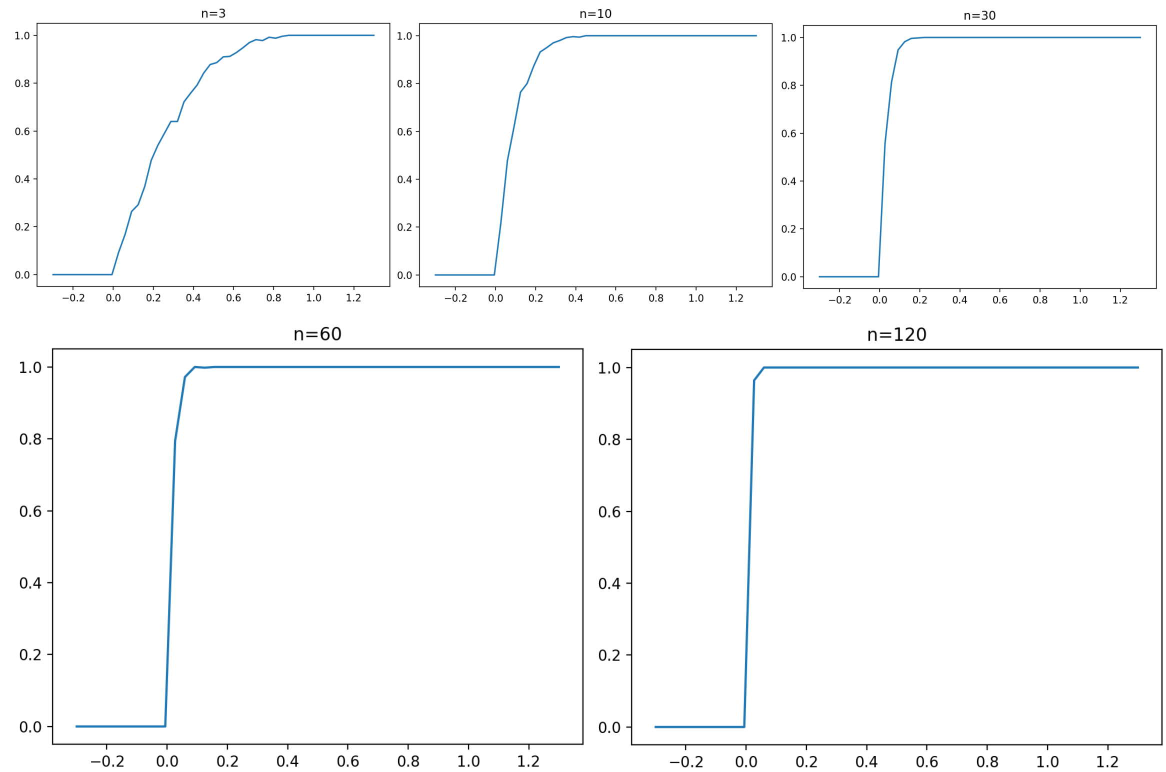
$$Z_n = \min(X_1, X_2, \ldots, X_n)$$

où les variables $X_1, X_2, \ldots, X_n$ sont indépendantes et de loi uniforme sur $[0;1]$.
Donner un programme qui prend en argument n et simule Z_n.

12. À l'aide du code suivant et des différents tests, que pouvez-vous conjecturer sur la convergence en loi de la suite $(Z_n)_{n\in\mathbb{N}^*}$?

Editeur

```
n= ... # valeur à définir à chaque test
x=np.linspace(-0.3,1.3,50)
y=np.zeros(50)
for i in range(50):
    y[i]=Frep(x[i])    # la fonction Frep est construite avec Z=Zn
plt.plot(x,y,'-')
plt.show()
```

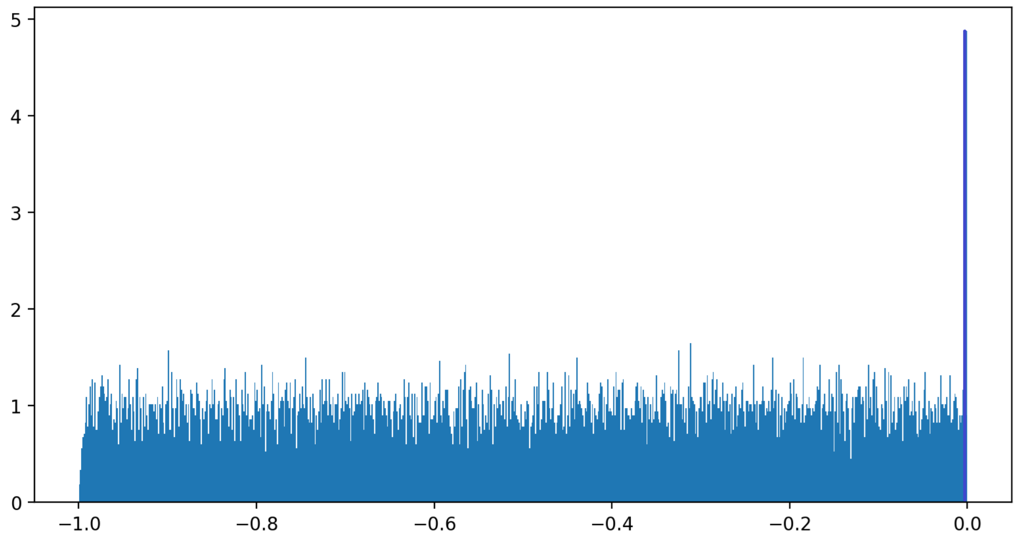


13. Prouver votre conjecture.

14. On définit maintenant la variable $T_n = Z_n - X_n$. Justifier que $\mathbf{P}(T_n = 0) = 1/n$. Est-ce que T_n est une variable aléatoire à densité ?

15. Écrire une fonction python d'argument n qui renvoie une simulation de T_n.

16. La figure A présente un histogramme de 750 rectangles donnant la répartition de 20000 réalisations de la variable aléatoire T_{200}. La figure B est un zoom de la partie droite.

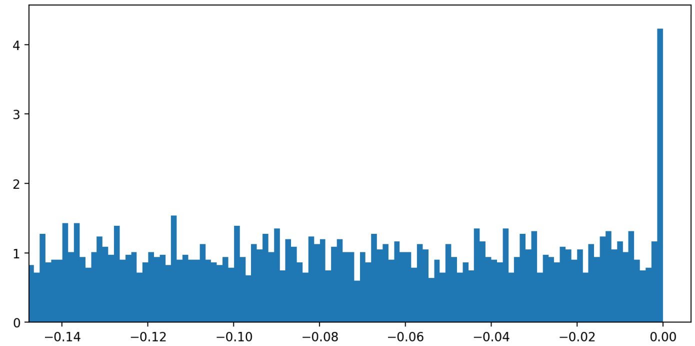


- FIGURE A -

Console

```python
plt.clf()
m=20000
Ech=np.zeros(m)
for i in range(m):
    Ech[i]=simuTn(200)

plt.clf()
plt.hist(Ech,750,density=True)
plt.show()
```

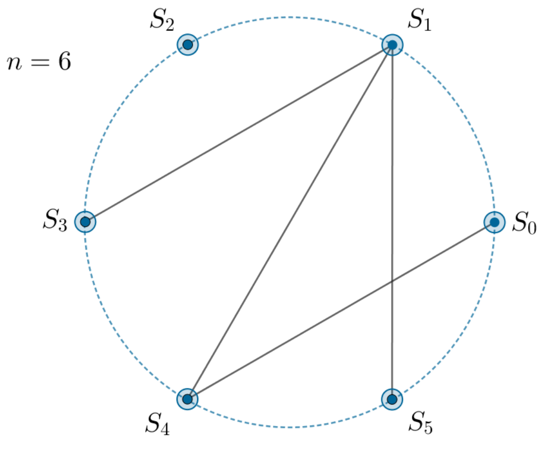


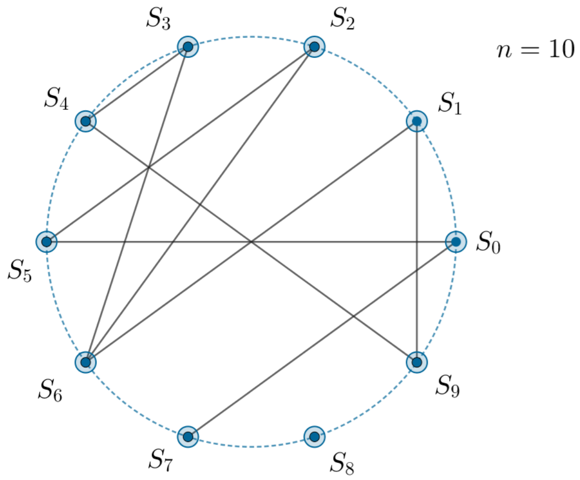
- FIGURE B -

a) Est-ce que la variable T_n vous semble discrète ? Justifier.
b) Le rectangle le plus à droite de la figure B est-t-il cohérent avec la question 14 ?

Problème C : Graphes aléatoires d'Erdős et Rényi

Soient $n \in \mathbb{N} \setminus \{0;1\}$ et $p \in]0;1[$.
On place n points régulièrement sur un cercle numérotés de S_0 à S_{n-1}. Pour chaque paire de points $\{S_i;S_j\}$ (pour $i \neq j$), on décide de relier (ou non) les deux points par un segment [1] avec probabilité p.





On peut représenter la situation à l'aide d'une matrice $A = (a_{i,j}) \in \mathscr{M}_n(\mathbb{R})$, dite d'adjacence, définie par

$$\forall\, i,\, j \in [\![1;n]\!], \qquad a_{i,j} = \begin{cases} 1 & \text{si les points } S_{i-1} \text{ et } S_{j-1} \text{ sont reliés} \\ 0 & \text{sinon.} \end{cases}$$

On pourra noter que la matrice A est symétrique avec une diagonale nulle.

1. Dans le vocabulaire des graphes, on parle plutôt de sommets pour les points, et d'arêtes au lieu des segments.

17. Écrire un programme d'arguments n et p qui renvoie la matrice d'adjacence d'un graphe construit aléatoirement par la règle précédente.

18. Un point est dit isolé s'il n'est relié à aucun autre point. Que dire de la colonne $j+1$ de la matrice d'adjacence si le point S_j est isolé ? En déduire un programme d'argument une matrice d'adjacence A renvoyant le nombre de points isolés du graphe associé.
On pourra utiliser la commande `np.sum(A,0)` *qui renvoie une matrice ligne donnant la somme sur chaque colonne.*

19. Pour tout entier $n \geqslant 2$, on pose alors N_n, la variable qui renvoie le nombre de points isolés d'un graphe aléatoire avec les paramètres n et

$$p = \frac{\ln(n)}{n} - \frac{1}{n}.$$

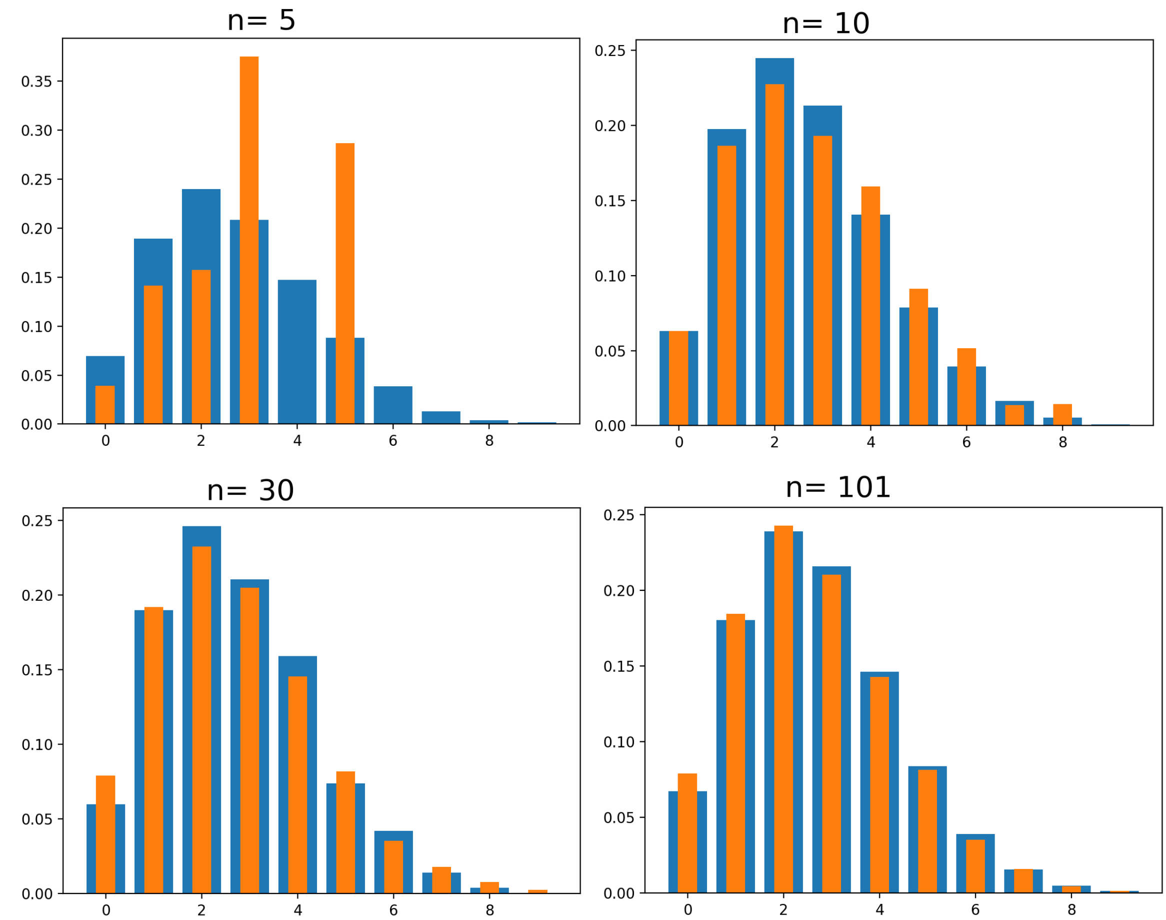
Écrire une fonction python `Nn` qui prend en argument n et renvoie une simulation de la variable N_n.

20. Les codes et simulations suivants permettent de conjecturer une convergence en loi. Laquelle ?

Editeur

```
n= ... # on teste avec 5, 10, 30, 101
m=10000
Ech=np.zeros(m)
Poi=rd.poisson(np.exp(1),m)
for i in range(m):
    Ech[i]=  Nn(n)

plt.clf()
classe=np.linspace(0,10,11)-0.5
plt.hist(Poi,classe,rwidth=0.8,density=True)
plt.hist(Ech,classe,rwidth=0.4,density=True)
plt.show()
```



– FIN –

DS 9 - solution

Exercice 1

1. Si X désigne le rang du premier succès dans une infinité d'expériences de Bernoulli de paramètre p, on sait que

$$X \hookrightarrow \mathscr{G}(p).$$

On en déduit l'idée du code :

```
def geo(p):
    n=1
    while rd.random()>p:
        n+=1
    return n
```

2. On rajoute une condition pour renvoyer N si on a déjà $N-1$ échecs :

```
def geoT(p,N):
    n=1
    while rd.random()>p and n<N:
        n+=1
    return n
```

Exercice 2

Un exercice qui présente une variante de la construction du triangle de Pascal avec python.

3.

```
def Stirling(N,K):
    S=np.zeros([N+1,K+1])
    S[0,0]=1
    for n in range(N):
        for k in range(K+1):
            S[n+1,k]=S[n,k-1]-n*S[n,k]
    return S
```

4.

```
def somme(n):
    c=0
    S=Stirling(n,n)
    e=1
    for k in range(n+1):
        e=-e
        c+=e*S[n,k]
    return c
```

5. On conjecture que pour tout $n \in \mathbb{N}$

$$c_n = (-1)^n\, n!$$

Problème A

6.

```
def syracuse(n,N):
    U=np.zeros(n+1)
    U[0]=N
    for i in range(0,n):
        if 2*int(U[i]/2)==U[i]:
            U[i+1]=U[i]/2
        else :
            U[i+1]=3*U[i]+1
    return U
```

7.

```
def vol(N):
u=N
n=0
while u!=1:
    n+=1
    if 2*int(u/2)==u:
        u=u/2
    else :
        u=3*u+1
return n
```

8.

```
def Altitude(N):
u=N
M=u
while u!=1:
    if 2*int(u/2)==u:
        u=u/2
    else :
        u=3*u+1
        M=max(M,u)
return M
```

9. Le premier affiche le temps de vol en fonction de $N \in [\![1;1000]\!]$. Le second affiche le logarithme de l'attitude en fonction de $N \in [\![1;200]\!]$.

Problème B

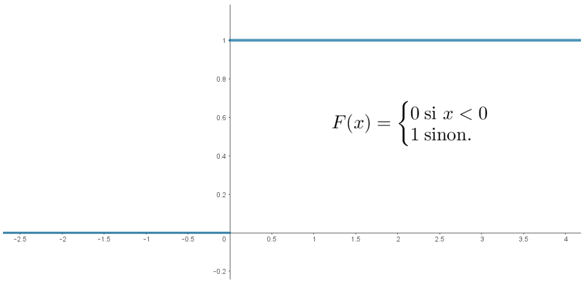
10.

```python
def Frep(x):
    c=0
    for i in range(500):
        c+=(simuZ(n)<x)
    return c/500
```

11.

```python
def simuZ(n):
    return np.min(rd.random(n))
```

12. Le programme affiche une approximation de la courbe de la fonction de répartition de Z_n sur $[-0.3; 1.3]$. On constate une convergence vers la fonction de répartition F donnée par :



Cette fonction de répartition est celle d'une variable constante égale à 0. On conjecture donc une convergence en loi vers une variable constante égale à 0.

13. Soit $x \in \mathbb{R}$.

→ Si $x < 0$, on constate que

$$F_{Z_n}(x) = 0 \underset{n\to\infty}{\longrightarrow} 0 = F(x).$$

→ Si $x > 0$, on a par indépendance et l'égalité en loi des variables X_i

$$\begin{aligned} 1 - F_{Z_n}(x) &= \mathbf{P}(Z_n > x) \\ &= \mathbf{P}\big([X_1 > x] \cap \ldots \cap [X_n > x]\big) \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i > x) \\ &= \prod_{i=1}^{n} \big(1 - F_{X_i}(x)\big) \\ &= \big(1 - F_{X_1}(x)\big)^n \underset{n\to\infty}{\longrightarrow} 0 \end{aligned}$$

car $1 - F_{X_1}(x) \in [0; 1[$. On en déduit

$$F_{Z_n}(x) = 0 \underset{n\to\infty}{\longrightarrow} 1 = F(x).$$

→ Il est inutile d'étudier la convergence pour $x = 0$ car 0 n'est pas un point de continuité de F.

Finalement pour tout $x \in \mathbb{R}^*$,

$$F_{Z_n}(x) = 0 \underset{n\to\infty}{\longrightarrow} F(x).$$

C'est la définition de la convergence en loi. La conjecture est prouvée.

14. Par symétrie du problème, pour tout indice $i \in [\![1; n]\!]$

$$\mathbf{P}(Z_n = X_i) = \mathbf{P}(Z_n = X_n) \qquad (\bullet)$$

Pour détailler davantage, on peut justifier en disant que les variables $Z_n - X_i$ et $Z_n - X_n$ ont même loi.

De plus, on a

$$\bigcup_{i=1}^{n} [Z_n = X_i] = \Omega$$

D'où $$\mathbf{P}\Big(\bigcup_{i=1}^{n} [Z_n = X_i]\Big) = 1.$$

Les événements $\big([Z_n = X_i]\big)_{i \in [\![1;n]\!]}$ ne sont pas deux à deux incompatibles mais, on a toutefois

$$\forall\, i, j \in [\![1; n]\!], \quad i \neq j \quad \Rightarrow \quad \mathbf{P}\big([Z_n = X_i] \cap [Z_n = X_j]\big) = 0.$$

En effet, on a une première inclusion

$$[Z_n = X_i] \cap [Z_n = X_j] \subset [X_i = X_j]$$

puis

$$\mathbf{P}\big([Z_n = X_i] \cap [Z_n = X_j]\big) \leqslant \mathbf{P}\big([X_i = X_j]\big) = 0.$$

Car par le théorème de sommation, on vérifie que $X_i - X_j$ est une variable à densité, donc

$$[X_i = X_j] = [X_i - X_j = 0]$$

est un événement négligeable.

Justifions maintenant que

$$\sum_{i=1}^{n} \mathbf{P}\big([Z_n = X_i]\big) = \mathbf{P}\Big(\bigcup_{i=1}^{n} [Z_n = X_i]\Big) = 1.$$

Pour cela, on introduit les événements :
B : il existe deux indices $i \neq j$ tels que $[X_i = X_j]$ et pour tout $i \in [\![1; n]\!]$

$$\tilde{A_i} = [Z_n = X_i] \cap \overline{B}.$$

Ainsi, on vérifie que

$$\mathbf{P}(B) = 0;$$

$$\forall\, i \in [\![1; n]\!], \qquad \mathbf{P}(\tilde{A_i}) = \mathbf{P}([Z_n = X_i])$$

et B et A_i, $i \in [\![1; n]\!]$ forment un système complet d'événements. Finalement

$$\mathbf{P}(B) + \sum_{i=1}^{n} \mathbf{P}(\tilde{A_i}) = 1.$$

C'est-à-dire

$$\sum_{i=1}^{n} \mathbf{P}\big([Z_n = X_i]\big) = 1.$$

Et en reprenant la relation $(\bullet)$

$$n\mathbf{P}\big([Z_n = X_n]\big) = 1.$$

D'où le résultat.

- Comme $\mathbf{P}(T_n = 0) \neq 0$, la variable T_n n'est pas à densité.

15.

```
def simuTn(n):
    m=np.min(rd.random(n-1))
    xn=rd.random()
    if m<xn :
        return m-xn
    else :
        return 0
```

16.a) Non, la variable ne semble pas discrète car toutes les valeurs comprises entre -1 et 0 semblent possibles. L'ensemble des valeurs prises par la variable n'est pas fini ni dénombrable, elle n'est pas discrète.

16.b) Oui, car on constate que la probabilité que T_n prenne la valeur 0 n'est pas négligeable. On a précisément $\mathbf{P}(T_{200} = 0) = 1/200$. Donc sur les 20000 réalisations, on s'attend à avoir autour d'une centaine de cas où T_n prend la valeur 0.
L'aire du dernier rectangle correspond pratiquement à la probabilité $\mathbf{P}(T_{200})$. Comme la base du rectangle est de longueur $\ell = 1/750$, la hauteur est donnée par

$$h = \frac{\text{Aire}}{\ell} \simeq \frac{750}{200} \simeq 4.$$

Problème C

17.

```
def matrice(n,p):
    A=np.zeros([n,n])
    for i in range(n):
        for j in range(i+1,n):
            A[i,j]=(rd.random()<p)
            A[j,i]=A[i,j]
    return A
```

18.

```
def Isole(A):
C=np.sum(A,1)==0
return np.sum(C)
```

19.

```
def Nn(n):
p=np.log(n)/n-1/n
return Isole(matrice(n,p))
```

20. Le programme renvoie, à n fixé, deux histogrammes :
→ L'un obtenu à partir de 10 000 réalisations d'une variable loi de Poisson de paramètre e ;
→ L'autre obtenu à partir de 10 000 réalisations de N_n.

On constate que plus n est grand, plus les histogrammes se rapprochent. Sachant que l'aire d'un rectangle au niveau de l'entier i est proportionnelle à la probabilité que la variable prenne la valeur i, on s'attend à

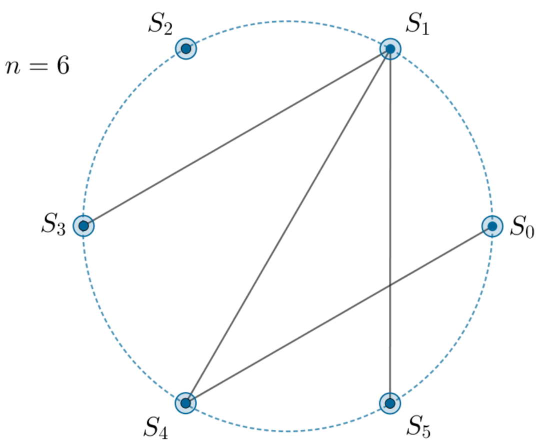
$$\forall i \in \mathbb{N}, \qquad \mathbf{P}(N_n = i) \underset{n\to\infty}{\longrightarrow} \mathbf{P}(X = i)$$

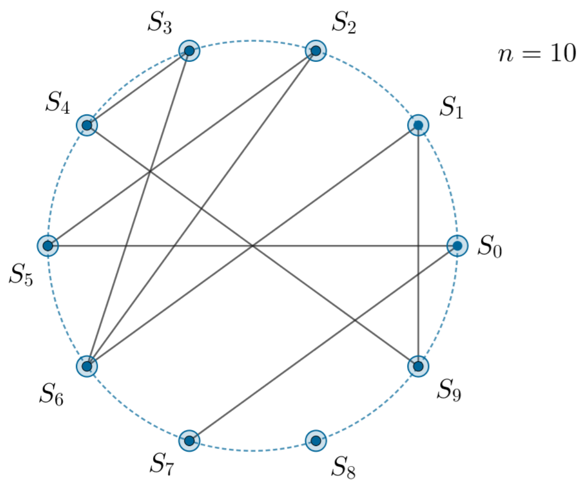
où $X \hookrightarrow \mathscr{P}(e)$. On s'attend donc à une convergence en loi de la suite $(N_n)_n$ vers une variable X de loi de Poisson de paramètre e.

Complément au DS 9

THÈME : GRAPHES ALÉATOIRES D'ERDŐS ET RÉNYI
facultatif

Soient $n \in \mathbb{N} \setminus \{0;1\}$ et $p \in]0;1[$.
On place n points régulièrement sur un cercle numérotés de S_1 à S_n. Pour chaque paire de points $\{S_i;S_j\}$ (pour $i \neq j$), on décide de relier (ou non) les deux points par un segment avec probabilité p.





On parle alors de graphes aléatoires , les points sont les sommets du graphe et les segments sont les arêtes du graphe. De plus, on dit que deux sommets sont voisins s'ils sont reliés par une arête.

1. Quel est le nombre maximal d'arêtes d'un graphe contenant n sommets ?

Pour tous $i, j \in [\![1;n]\!]$, on pose

$$T_{i,j} = \begin{cases} 1 & \text{s'il existe une arête entre les sommets } S_i \text{ et } S_j \\ 0 & \text{sinon.} \end{cases}$$

Par construction, les variables $T_{i,j}$ suivent une loi de Bernoulli de paramètre p et sont mutuellement indépendantes.

2. Soit $k \in [\![1;n]\!]$. On considère la variable aléatoire D_k égale au nombre de voisins du sommets k.
- **a)** Exprimer D_k à l'aide des variables $T_{i,j}$.
- **b)** Reconnaître la loi de D_k.

3. Pour $k \neq \ell$, est-ce que les variables D_k et D_ℓ sont indépendantes ?

4. On dit que le sommet k est isolé si $D_k = 0$ est réalisé. On note Z_n la variable aléatoire comptant les sommets isolés d'un graphe aléatoire G à n sommets et X_k la variable aléatoire de Bernoulli qui vaut 1 si le sommet k est isolé et 0 sinon.
- **a)** Montrer que : $Z_n = \sum\limits_{k=1}^{n} X_k$. En déduire que $\mathbf{E}(Z_n) = n(1-p)^{n-1}$.
- **b)** Montrer que : $Z_n^2 = \sum\limits_{k=1}^{n} X_k + 2 \sum\limits_{1 \leqslant i < j \leqslant n} X_i X_j$.
- **c)** Justifier que $\mathbf{P}\left(\left[X_i = 1\right] \cap \left[X_j = 1\right]\right) = (1-p)^{2n-3}$. En déduire que $\mathbf{E}\left(Z_n^2\right) = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}$.

- On suppose désormais que $p = p_n = c\frac{\ln(n)}{n}$, avec $c > 0, c \neq 1$.

5. Montrer que $\left(1-p_n\right)^{n-1} \underset{n \to +\infty}{\sim} \left(1-p_n\right)^n$ puis que $\left(1-p_n\right)^{n-1} \underset{n \to +\infty}{\sim} n^{-c}$.

6. On suppose que $c > 1$. À l'aide de l'inégalité de Markov, déduire la limite de $\mathbf{P}([Z_n \geqslant 1])$ puis de $\mathbf{P}([Z_n = 0])$, lorsque $n \to +\infty$.

7. En remarquant que $[Z_n = 0] \subset [|Z_n - \mathbf{E}(Z_n)| \geqslant \mathbf{E}(Z_n)]$, montrer que

$$\mathbf{P}([Z_n = 0]) \leqslant \frac{\mathbf{V}(Z_n)}{\mathbf{E}(Z_n)^2}.$$

8. En déduire que si $c < 1$, $\mathbf{P}([Z_n = 0]) \to 0$ quand $n \to +\infty$.

– FIN –

DS 9 complément- solution

Ce sujet est adapté du sujet 0 maths appliquées HEC-ESSEC.

1. Le nombre maximal d'arêtes correspond au nombre de paires de sommets soit

$$\binom{n}{2} = \frac{n(n-1)}{2}.$$

2. On a

$$D_k = \sum_{i=1}^{k-1} T_{i,k} + \sum_{i=k+1}^{n} T_{k,i}.$$

La variable D_k est la somme de $(n-1)$ variables de Bernoulli indépendantes de paramètre p d'où D_k suit la loi binomiale de paramètres $n-1$ et p.

3. Non, les variables ne sont pas indépendantes. Par exemple,

$$[D_k = 0] \cap [D_j = n-1] = \varnothing$$

est de probabilité nulle alors que

$$\mathbf{P}(D_k = 0)\mathbf{P}(D_j = n-1) \neq 0.$$

4.a) Pour tout indice k,

$$\mathbf{E}(X_k) = \mathbf{P}([X_k = 1]) = \mathbf{P}([D_k = 0])$$

qui vaut $(1-p)^{n-1}$ d'après la loi de D_k, d'où par linéarité de l'espérance,

$$\mathbf{E}(Z_n) = n(1-p)^{n-1}.$$

4.b) On a

$$\begin{aligned} {Z_n}^2 &= \left(\sum_{k=1}^{n} X_k\right)^2 \\ &= \sum_{(i,j)\in[\![1,n]\!]^2} X_i X_j \\ &= \sum_{1\leqslant i=j\leqslant n} X_i X_j + \sum_{1\leqslant i<j\leqslant n} X_i X_j + \sum_{1\leqslant j<i\leqslant n} X_i X_j. \end{aligned}$$

Or comme toute variable de Bernoulli $X_i^2 = X_i$ et on a aussi

$$\sum_{1\leqslant j<i\leqslant n} X_i X_j = \sum_{1\leqslant i<j\leqslant n} X_i X_j$$

d'où on obtient bien que :

$$Z_n^2 = \sum_{k=1}^{n} X_k + 2\sum_{1\leqslant i<j\leqslant n} X_i X_j$$

4.c) On a pour $i<j$,

$$[X_i = 1] \cap \left[X_j = 1\right]$$

est réalisé si et seulement si les événements $\left[T_{k,i} = 0\right]$ pour $k<i$, $\left[T_{i,k} = 0\right]$ pour $k>i$, $[T_{k,j} = 0]$ pour $k<j$ et $k\neq i$, $\left[T_{j,k} = 0\right]$ pour $k>j$ sont réalisés. Or ces événements sont au nombre de $(n-1)+(n-2) = 2n-3$, indépendants et de probabilité p. D'où

$$\mathbf{P}\left([X_i = 1] \cap [X_j = 1]\right) = (1-p)^{2n-3}.$$

De plus par linéarité de l'espérance :

$$\mathbf{E}\left(Z_n^2\right) = n(1-p)^{n-1} + 2\sum_{1\leqslant i<j\leqslant n} \mathbf{E}\left(X_i X_j\right).$$

Or

$$\begin{aligned} \mathbf{E}\left(X_i X_j\right) &= \mathbf{P}\left(\left[X_i X_j = 1\right]\right) \\ &= \mathbf{P}\left([X_i = 1] \cap \left[X_j = 1\right]\right) \\ &= (1-p)^{2n-3} \end{aligned}$$

et il y a $\dfrac{n(n-1)}{2}$ couples (i,j) tels que $1\leqslant i<j\leqslant n$, alors on a bien :

$$\mathbf{E}\left(Z_n^2\right) = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}.$$

5. On remarque que $(1-p_n)^{n-1} \sim (1-p_n)^n$ car $1-p_n \to 1$ quand $n\to+\infty$. De plus

$$(1-p_n)^n = \exp\left(n\ln\left(1-c\frac{\ln(n)}{n}\right)\right).$$

Or,

$$n\ln\left(1-c\frac{\ln(n)}{n}\right) = n\left(-c\frac{\ln(n)}{n} - c^2\frac{\ln(n)^2}{2n^2} + o\left(\frac{\ln(n)^2}{n^2}\right)\right),$$

d'où

$$n\ln\left(1-c\frac{\ln(n)}{n}\right) = -c\ln(n) + o(1)$$

donc

$$(1-p_n)^n \sim n^{-c}.$$

6. Si $c>1$, $\mathbf{E}(Z_n) \to 0$ car $\mathbf{E}(Z_n) \sim n^{1-c}$, quand $n\to+\infty$. D'après l'inégalité de Markov,

$$\mathbf{P}([Z_n \geqslant 1]) \leqslant \mathbf{E}(Z_n),$$

d'où

$$\mathbf{P}([Z_n \geqslant 1]) \underset{n\to\infty}{\longrightarrow} 0$$

et ainsi

$$\mathbf{P}([Z_n = 0]) \underset{n\to\infty}{\longrightarrow} 1.$$

7. Appliquons l'inégalité de Bienaymé-Tchebychev à Z_n pour $\varepsilon = \mathbf{E}(Z_n)$:

$$\mathbf{P}([|Z_n - \mathbf{E}(Z_n)| \geqslant \mathbf{E}(Z_n)]) \leqslant \frac{\mathbf{V}(Z_n)}{\mathbf{E}(Z_n)^2}$$

Or si $Z_n = 0$ alors

$$|Z_n - \mathbf{E}(Z_n)| \geqslant \mathbf{E}(Z_n),$$

donc

$$\mathbf{P}([Z_n = 0]) \leqslant \mathbf{P}([|Z_n - \mathbf{E}(Z_n)| \geqslant \mathbf{E}(Z_n)])$$

et par transitivité,

$$\mathbf{P}([Z_n = 0]) \leqslant \frac{\mathbf{V}(Z_n)}{\mathbf{E}(Z_n)^2}.$$

D'après les questions précédentes,

$$\frac{\mathbf{V}(Z_n)}{\mathbf{E}(Z_n)^2} = \frac{1}{n\left(1-p_n\right)^{n-1}} + \frac{n-1}{n\left(1-p_n\right)} - 1.$$

Or $\frac{n-1}{n(1-p_n)} \underset{n\to\infty}{\longrightarrow} 1$ et si $c < 1$,

$$\frac{1}{n\left(1-p_n\right)^{n-1}} \underset{n\to\infty}{\longrightarrow} 0,$$

d'où

$$\mathbf{P}([Z_n = 0]) \underset{n\to\infty}{\longrightarrow} 0.$$