

EXERCICE 1: [Indications] [Correction]

1. Approximation de la covariance :
 - a) Soit Z une variable aléatoire admettant une espérance. Rappeler une méthode permettant d'approcher $E[Z]$ grâce à des simulations successives de la variable Z .
 - b) En déduire une façon d'approcher $Cov(A, B)$ pour toute variable A, B admettant une covariance.
 - c) *Application* : On se donne A et B deux variables aléatoires indépendantes qui suivent toutes deux une loi normale centrée réduite. Quelle est la valeur de $Cov(A, B)$? Le vérifier de manière approximative par simulation grâce aux considérations précédentes.
(On pourra éventuellement utiliser la fonction `Gauss` de la bibliothèque `random`.)

Soient maintenant X_1, X_2, X_3 trois variables aléatoires réelles discrètes, centrées et admettant une variance. On note M leur matrice de variance-covariance, soit : $M = (m_{i,j})_{1 \leq i,j \leq 3}$, avec $m_{i,j} = Cov(X_i, X_j)$.

2. Justifier que la matrice M est diagonalisable.

3. a) Vérifier que

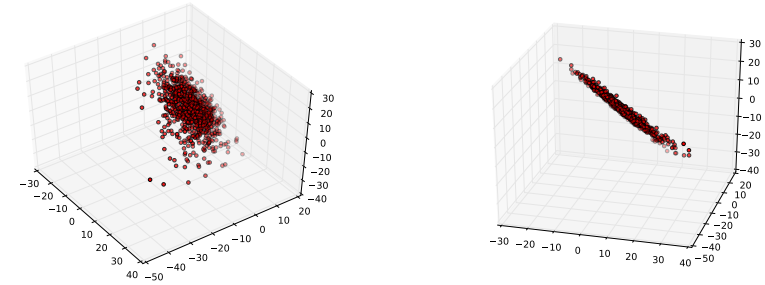
$$MX = \begin{pmatrix} Cov(X_1, U) \\ Cov(X_2, U) \\ Cov(X_3, U) \end{pmatrix} \quad \text{où } X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ et } U = \sum_{j=1}^3 x_j X_j.$$

- b) En déduire que ${}^t X M X$ s'écrit comme une variance quelquesoit le vecteur X .
 - c) En déduire que les valeurs propres de M sont positives ou nulles. *Indication* : étudier le signe de ${}^t X M X$ si X est un vecteur.
4. On suppose à partir de maintenant que

$$M = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

- a) Les variables aléatoires X_1, X_2, X_3 sont-elles mutuellement indépendantes?
- b) Déterminer les valeurs propres et espaces propres de M . (On pourra utiliser éventuellement le langage Python pour s'aider.)
- c) Déduire des calculs déjà effectués dans les questions précédentes que la variable aléatoire $X_1 + X_2 + X_3$ est de variance nulle.

5. Voici une simulation effectuée à partir d'exemples particuliers de variables X_1, X_2, X_3 sous deux angles différents.



- a) Que peut-on conjecturer sur le graphique?
- b) Démontrer et préciser la conjecture à l'aide des résultats précédents.

EXERCICE 2: [Indications] [Correction] On considère deux joueurs A et B . Pendant une partie, on lance une pièce.

- Si on obtient pile, alors B donne 1 euro à A ,
- Sinon, c'est A qui donne 1 euro à B .

On note X_n la variable aléatoire qui donne le gain total de A à l'issue de n parties indépendantes ainsi que, pour tout $n \in \mathbb{N}^*$, $u_n = P(X_{2n} = 0)$.

1. Donner la loi de X_n puis l'espérance et la variance de X_n .
2. a) Construire une fonction Python `x` d'argument `n` qui renvoie le gain aléatoire de A à l'issue de n parties indépendantes.
- b) Construire ensuite une fonction Python `u` d'argument `n` qui renvoie une approximation de u_n à l'issue de n parties indépendantes.
- c) En déduire une valeur approchée de $\lim_{n \rightarrow +\infty} \frac{u_{n+1}}{u_n}$.
3. Calculer maintenant de façon exacte

$$\lim_{n \rightarrow +\infty} \frac{u_{n+1}}{u_n}$$

Interpréter.

EXERCICE 3: [Indications] [Correction]

L'exercice comprend 3 parties. La partie 3 peut être abordée sans que la partie 2 ait été traitée.

définition d'une fonction :

Soit t un réel positif ou nul. Pour tout réel x , on pose : $P_t(x) = x^3 + tx^2 + 1$.

1. Construire une fonction **graphe** de paramètre a, b, t qui affiche le graphique de la fonction P_t sur l'intervalle $[a, b]$.
2. Démontrer que le polynôme P_t admet une unique racine réelle que l'on notera $r(t)$. On note r l'application définie sur $\mathbb{R}^+$ qui, à tout réel positif t , associe le réel $r(t)$.

ébauche de la courbe de la fonction r :

On a construit la représentation graphique de la fonction P_2 sur la figure 1 ci-dessous (en y rajoutant les axes).

3. Expliquer comment il est possible de construire géométriquement sur cette figure le point de coordonnées $(2; r(2))$.
4. On a répété à l'aide de géogébra la construction précédente en faisant varier t de 0 à 10 avec un pas de 0.1, et on a obtenu la figure 2 ci-dessous. Que peut-on conjecturer relativement :
 - a) à $r(0)$?
 - b) au signe de $r(t)$?
 - c) à la limite de la fonction r en $+\infty$?
 - d) à la branche infinie de la courbe de r en $+\infty$?

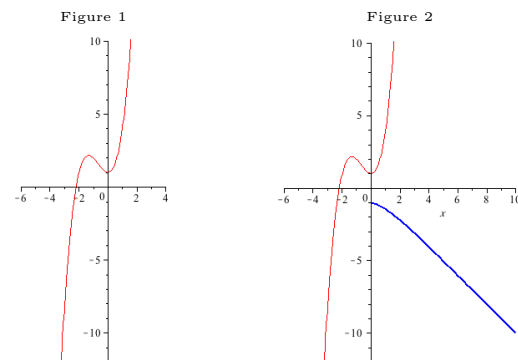
Démontrer ces conjectures.

5. Démontrer que la fonction r réalise une bijection de $\mathbb{R}^+$ vers $] -\infty, -1]$. Déterminer la fonction réciproque de la fonction r , que l'on nommera s .
6. En déduire que la fonction r est continue et dérivable sur $\mathbb{R}^+$. Dresser le tableau de variation de r .

Approche informatique :

7. Rédiger une fonction informatique qui, recevant t et un entier n , renvoie une valeur approchée de $r(t)$ à 10^{-n} près.

8. Utiliser cette fonction pour construire la courbe de r . on obtient la figure 2. Commenter la courbe obtenue.

**EXERCICE 4:** [Indications] [Correction]

I - Questions préliminaires :

1. a) Expliquer pourquoi (citer le théorème utilisé), pour une variable X , étant donné $(X_1, \dots, X_N)$ un N échantillon de X , on peut supposer que, pour N assez grand, on a

$$P(a < X < b) \simeq \frac{\text{Nombre de } i \text{ tels que } a < X_i < b}{N}$$

i.e.

- b) En déduire comment l'histogramme d'un N échantillon $(X_1, \dots, X_N)$ d'une variable X peut éventuellement donner une idée sur la loi de X .

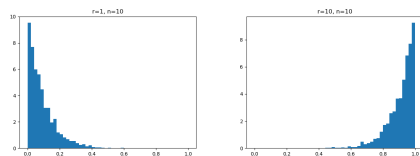
II - Classement croissant des variables uniformes :

Soit n un entier strictement supérieur à 1. $X_1, \dots, X_n$ sont n variables aléatoires réelles indépendantes de loi uniforme sur $[0; 1]$. Lorsqu'on range les X_i par ordre croissant, on obtient la suite $Y_1, Y_2, \dots, Y_n$ (Y_1 est le min des X_i et Y_n est le max des X_i .)

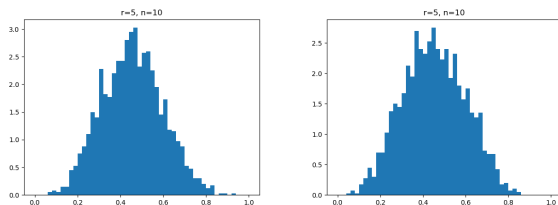
Soit $r \in \llbracket 1, n \rrbracket$. Le but de cet exercice est de savoir si on peut éventuellement se permettre d'approcher la loi de Y_r par une loi normale et peut être trouver une façon supplémentaire possible de simuler une loi normale.

2. Créer une fonction de paramètre n permettant de simuler un vecteur $(Y_1, \dots, Y_n)$.

3. a) Construire une fonction `histogramme_Unif_Croissant` de paramètres r, n qui affiche un histogramme (avec par exemple 50 classes) des valeurs de Y_r obtenues sur 2000 simulations de l'expérience.
Pour ce faire, on dispose par exemple de la fonction `hist` de la bibliothèque `matplotlib.pyplot`, qui, en demandant par exemple `hist(Y, density=True, bins=50)` rend l'histogramme des fréquences (option `density=True`) de la liste des valeurs Y découpées en 50 classes (option `bins=50`)
- b) Voici quelques observations d'histogrammes obtenues à l'aide de la fonction Python précédente pour $r = 1$ (à gauche) et $r = n$ (à droite) que vous pourrez confirmer à l'aide de votre fonction précédente. La loi semble-t-elle coïncider avec une loi normale ?

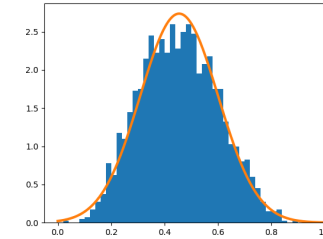


- c) Voici maintenant deux autres simulations pour le même r (ici en l'occurrence $r = 5$ pour $n = 10$.) La loi semble-t-elle coïncider avec une loi normale ?



4. Supposons que Y_r suit une loi proche d'une loi normale $\mathcal{N}(\mu_r, \sigma_r^2)$. Exprimer μ_r et σ_r^2 par-rapport à Y_r et donner une approximation de ces paramètres grâce aux données obtenues par simulation. (Sur quel principe vous appuyez-vous pour votre affirmation ?)

Ci-dessous, pour information, un histogramme pour $r = 5$ et $n = 10$ superposé avec la densité d'une variable qui suit $\mathcal{N}(\mu_r, \sigma_r^2)$ avec les paramètres trouvés dans la question précédente. (Il n'est pas demandé de réaliser cette superposition.)



III - Loi de Y_r :

Voyons maintenant si on peut confirmer les phénomènes observés. Commençons par déterminer précisément la loi de Y_r .

5. Soit x un réel, on note N_x le nombre de variables X_i vérifiant $X_i \leq x$.
- Donner la loi de N_x .
 - Comparer pour tout r de $\{1, \dots, n\}$ les événements $\{N_x \geq r\}$ et $\{Y_r \leq x\}$.
 - Montrer que la fonction de répartition de Y_r est donnée par la formule suivante :

$$F_r(t) = \begin{cases} 0 & \text{si } t < 0 \\ \sum_{k=r}^n \binom{n}{k} t^k (1-t)^{n-k} & \text{si } t \in [0; 1] \\ 1 & \text{si } t > 1 \end{cases}$$

- d) Soit $t \in [0, 1]$. Montrer que

$$F_{n+1-r}(t) = 1 - F_r(1-t).$$

6. a) La variable Y_r admet-elle une densité ?
b) Montrer que $\mathbb{E}[Y_r] = \frac{r}{n+1}$.
c) Calculer la variance de Y_r .

7. Montrer que :

- a) $\mathbb{E}[Y_{n+1-r}] = 1 - \mathbb{E}[Y_r]$
- b) $\mathbb{V}(Y_r) = \mathbb{V}(Y_{n+1-r})$.

IV - Évaluation de la corrélation avec la loi normale :

On souhaite dans cette question établir un lien entre la loi de Y_r et la loi normale $\mathcal{N}(\mu_r, \sigma_r^2)$. On rappelle pour cela :

- qu'on a noté F_r la fonction de répartition de Y_r .
- Qu'on a calculé explicitement μ_r et σ_r^2 (à retrouver dans l'énoncé.)

Notons maintenant Φ_r la fonction de répartition d'une variable suivant $\mathcal{N}(\mu_r, \sigma_r^2)$.

8. Montrer que

$$\Phi_{n+1-r}(t) = 1 - \Phi_r(1 - t)$$

On rappelle que deux variables ont même loi ssi elles ont même fonction de répartition. Pour savoir si on peut assimiler la loi de Y_r avec une loi normale, on souhaite donc voir dans quelle mesure on a (ou on n'a pas) $F_r \simeq \Phi_r$ et avec quelle précision. Pour ce faire, nous allons étudier le nuage de points

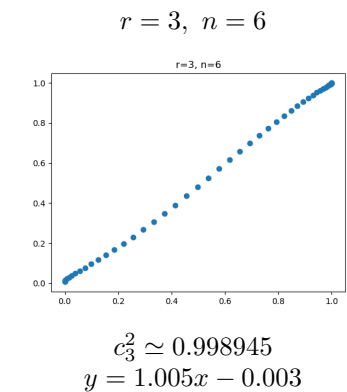
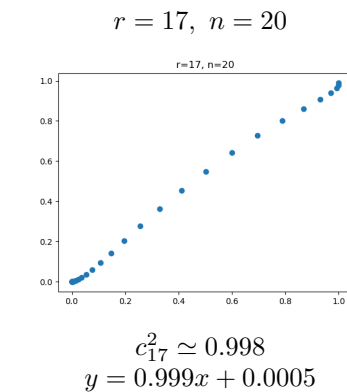
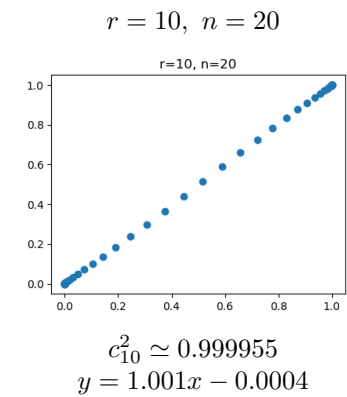
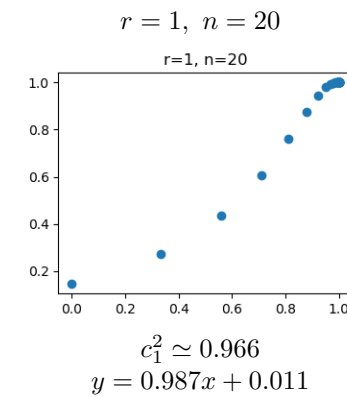
$$\{(X_i^{(r)}, Y_i^{(r)})\}_{i=0, \dots, M} = \{(F_r(x_i), \Phi_r(x_i))\}_{i=0, \dots, M}$$

où $x_i = \frac{i}{M}$ pour lequel nous allons calculer le coefficient de corrélation (et éventuellement observer la droite d'ajustement si elle est pertinente.)

9. Pourquoi l'étude de ce nuage peut nous permettre de répondre à la question posée ?

10. Ci-dessous quelques exemples de nuage pour $M = 50$ points, auquel on a rajouté :

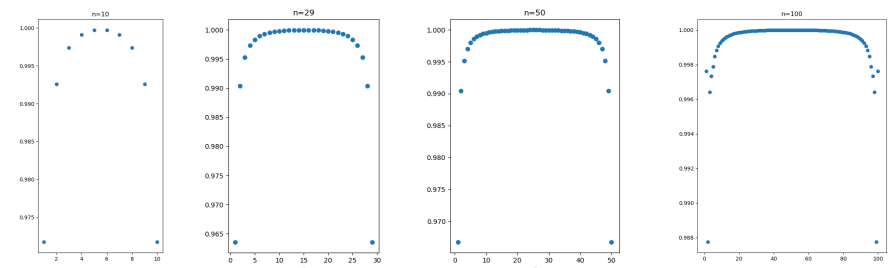
- le coefficient de corrélation au carré du nuage (noté c_r^2 et appelé "coefficient de détermination")
- la droite d'ajustement du nuage



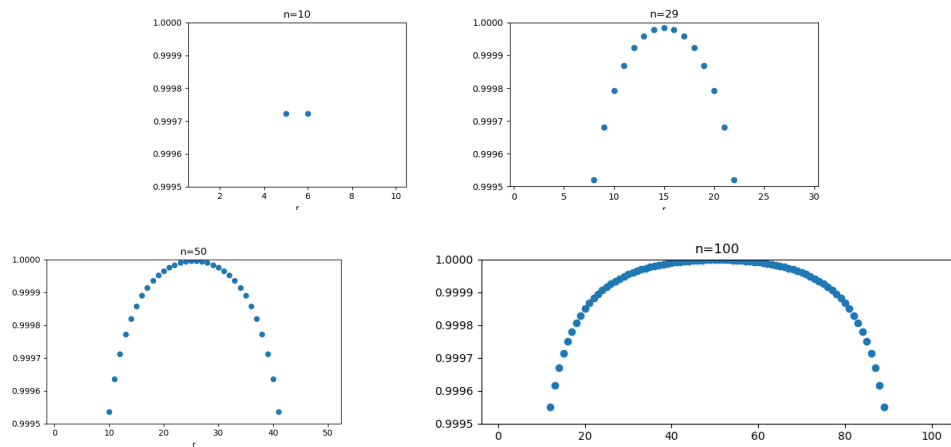
a) Quelles informations peut-on tirer de ces données ?

b) Notons $\overline{X^{(r)}}$ la moyenne des $X_i^{(r)}$ et $\overline{Y^{(r)}}$ la moyenne des $Y_i^{(r)}$. Montrer que

11. On appelle "coefficient de détermination" le carré du coefficient de corrélation (i.e. ici c_r^2). Voici quelques graphiques des coefficients de déterminations (r en abscisse et c_r^2 en ordonnée.)



Pour plus de précision, on a fait un zoom sur la partie dont les ordonnées étaient supérieures à 0,999.



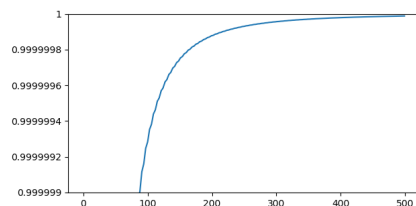
- Montrer maintenant que $c_r = c_{n+1-r}$
- Expliquer alors le caractère symétrique des courbes.
- Conjecturer le sens de variation des courbes en observant celles qui sont dessinées.

On s'intéresse maintenant au(x) point(s) maximum(s) de ces courbes pour savoir lequel des F_r a la meilleure corrélation avec Φ_r .

- Combien semble-t-il y avoir de points où la courbe est maximale pour $n = 10$, $n = 29$? (puis pour n pair, n impair?)
 - Soit $n \in \mathbb{N}^*$ fixé. En admettant le sens de variation établi précédemment de manière graphique, prouver à l'aide des calculs faits ce qui a été avancé dans la question précédente et montrer que :
 - si n est impair, le maximum des c_r^2 est atteint au point (r, c_r^2) avec $r = \frac{n+1}{2}$
 - si n est pair, le maximum est atteint aux points (j, c_j^2) avec $j = \frac{n}{2}$ et $j = \frac{n}{2} + 1$.

On note finalement $d_{max}^{(n)}$ cette valeur maximale des c_r^2 .

- Voici un dernier graphique qui donne maintenant $d_{max}^{(n)}$ en fonction de n en abscisse. De l'interprétation graphique que vous en ferez, en déduire une façon de simuler une variable qui suit approximativement une loi normale $\mathcal{N}(0, 1)$. Critiquez ce résultat par-rapport à vos connaissances et au résultat obtenu.



Exercice 1

1. a) D'après la loi faible des grands nombres, on sait que, si $(Z_1, \dots, Z_N)$ est un N échantillon de la variable Z avec N grand, on a

$$\mathbb{E}[Z] \simeq \frac{Z_1 + \dots + Z_N}{N}$$

- b) On sait que, d'après la formule de HK,

$$\text{Cov}(A, B) = \mathbb{E}(AB) - \mathbb{E}(A)\mathbb{E}(B)$$

D'où, pour N grand,

$$\text{Cov}(A, B) \simeq \frac{A_1 B_1 + \dots + A_N B_N}{N} - \frac{A_1 + \dots + A_N}{N} \cdot \frac{B_1 + \dots + B_N}{N}$$

- c) le programme suivant permet de répondre à la question :

```

1 import random as rd
2
3 def Cov():
4     N=1000
5     # on va simuler A et B N fois :
6     A=[rd.gauss(0,1) for _ in range(N)]
7     B=[rd.gauss(0,1) for _ in range(N)]
8     # E(A):
9     EA=sum(A)/N
10    # E(A):
11    EB=sum(B)/N
12    # E(AB):
13    produit=[A[i]*B[i] for i in range(N)]
14    EAB=sum(produit)/N
15    # résultat :
16    return(EAB-EA*EB)
```

On trouve par exemple : -0.017 qui est en effet proche de 0.

2. Comme la covariance est symétrique, il en résulte que la matrice est symétrique elle aussi.

De plus, une covariance est un nombre réel. La matrice est donc symétrique réelle, donc diagonalisable par le théorème spectral.

3. a) Soit $X = (x_1, x_2, x_3)$. Alors

$$\begin{aligned}
 MX &= \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \text{Cov}(X_1, X_3) \\ \text{Cov}(X_2, X_1) & \dots & \dots \\ \text{Cov}(X_3, X_1) & \dots & \dots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\
 &= \begin{pmatrix} \sum_{j=1}^3 x_j \text{Cov}(X_1, X_j) \\ \sum_{j=1}^3 x_j \text{Cov}(X_2, X_j) \\ \sum_{j=1}^3 x_j \text{Cov}(X_3, X_j) \end{pmatrix} \\
 &= \begin{pmatrix} \text{Cov}(X_1, \sum_{j=1}^3 x_j X_j) \\ \text{Cov}(X_2, \sum_{j=1}^3 x_j X_j) \\ \text{Cov}(X_3, \sum_{j=1}^3 x_j X_j) \end{pmatrix} \\
 &= \begin{pmatrix} \text{Cov}(X_1, U) \\ \text{Cov}(X_2, U) \\ \text{Cov}(X_3, U) \end{pmatrix} \quad \text{avec } U = \sum_{j=1}^3 x_j X_j
 \end{aligned}$$

- b) Grâce à la question précédente, on a

$$\begin{aligned}
 {}^t X M X &= \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} \text{Cov}(X_1, U) \\ \text{Cov}(X_2, U) \\ \text{Cov}(X_3, U) \end{pmatrix} \\
 &= \sum_{i=1}^3 x_i \text{Cov}(X_i, U) \\
 &= \text{Cov}\left(\sum_{i=1}^3 x_i X_i, U\right) \\
 &= \text{Cov}(U, U) = V(U) \geq 0
 \end{aligned}$$

- c) • Remarquons que pour X quelconque, on sait d'après la question précédente que

$${}^tXMX = V(U)$$

Or toute variance est toujours positive.

- Supposons maintenant que X soit vecteur propre non nul de valeur propre λ . Alors, comme $U = \sum_{j=1}^3 x_j X_j$,

$$0 \leq V\left(\sum_{i=1}^3 x_i X_i\right) = {}^tX \underbrace{MX}_{=\lambda X} = \lambda {}^tXX = \lambda \underbrace{\|X\|^2}_{\geq 0}$$

Ceci donne au final

$$0 \leq \lambda$$

4. a) Les valeurs propres sont donc bien positives ou nulles.
Comme $Cov(X_i, X_j) = -1 \neq 0$, les variables aléatoires (X_i) ne sont pas indépendantes.

- b) • Avec le langage Python :

Avec le programme :

```
1 import numpy as np
2
3 M=np.matrix([
4 [2,-1,-1],
5 [-1,2,-1],
6 [-1,-1,2]
7 ])
8
9 [Vap,Vep]=np.linalg.eig(M)
10 print(Vap)
11 print(Vep)
```

Le résultat pour les valeurs propres :

```
[ 3.00000000e+00 -3.77809194e-16  3.00000000e+00]
```

semble laisser penser que les valeurs propres sont 0,3

Les résultats pour les vecteurs propres :

```
[[ 0.81649658 -0.57735027  0.30959441]
 [-0.40824829 -0.57735027 -0.80910101]
 [-0.40824829 -0.57735027  0.49950661]]
```

semblent quant à eux laisser penser que les espaces propres (à lire en colonnes!!) sont :

$$E_3 = Vect\left(\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 \\ -8 \\ 5 \end{pmatrix}\right), E_0 = Vect\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right),$$

après vérification avec

```
1 # pour vap 0
2 V0=np.matrix([[1],[1],[1]])
3 print(M*V0) # on trouve bien M*V0=0*V0
4
5 # pour vap 3
6 V1=np.matrix([[2],[-1],[-1]])
7 print(M*V1) # on trouve bien M*V1=3*V1
8
9 V2=np.matrix([[3],[-8],[5]])
10 print(M*V2) # on trouve bien M*V2=3*V2
```

cela fonctionne, avec $\dim E_0 + \dim E_3 \geq 3$. Il n'y a donc pas d'autre valeur propre et on peut dire que

$$Sp(M) = \{0, 3\}, \quad E_3 = Vect \left(\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 3 \\ -8 \\ 5 \end{pmatrix} \right), E_0 = Vect \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right),$$

- Avec calcul :

En utilisant le fait que λ est valeur propre ssi $rg(M - \lambda Id) < 3$, on obtient que les valeurs propres de M sont 3 et 0, les sous-espaces propres étant

$$E_0 = Vect \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right) \quad E_3 = Vect \left(\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}; \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right)$$

c) D'après la question 2b) $V(X_1 + X_2 + X_3) = (1 \ 1 \ 1) M \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} =$

$$(1 \ 1 \ 1) \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0.$$

5. a) On dirait bien que les variables sont sur un plan. Ceci signifierait qu'il existe a, b, c, d tels que

$$aX_1 = bX_2 + cX_3 + d = 0$$

- b) On a vu que

$$V(X_1 + X_2 + X_3) = 0.$$

Or, on rappelle qu'une variable de variance nulle est une variable constante avec probabilité de 1. Ainsi, il existe $c \in \mathbb{R}$ tel que, avec probabilité 1 :

$$X_1 + X_2 + X_3 = c.$$

Ainsi, avec une probabilité 1 d'avoir raison, on peut dire que (X_1, X_2, X_3) est sur un plan.

En remarque, on pourrait même observer que ce plan passe par 0. En effet, les variables étant centrées par hypothèse, par linéarité de l'espérance, on a que

$$\mathbb{E}[X_1 + X_2 + X_3] = 0.$$

Ceci signifie que, comme $X_1 + X_2 + X_3 = c$, on a

$$\mathbb{E}[c] = 0$$

Or, comme toute constante, $\mathbb{E}[c] = c$ et donc

$$c = 0$$

D'où, avec probabilité 1 :

$$\boxed{X_1 + X_2 + X_3 = 0}$$

Exercice 2

1. On note S_n le nombre de "faces" obtenus au cours des n lancers, alors

$$S_n \hookrightarrow \mathcal{B}(n, \frac{1}{2})$$

Le gain X_n de A correspond à

$$X_n = \text{nb de piles} - \text{nombre de faces} = (n - S_n) - S_n$$

i.e.

$$X_n = n - 2S_n$$

Ainsi,

$$\text{Supp}(X_n) = \{n - 2k \mid k \in \{0, \dots, n\}\}$$

avec

$$P(X_n = n - 2k) = P(S_n = k) = \binom{n}{k} \frac{1}{2^n}$$

et

$$E(X_n) = n - 2E(S_n) = n - 2n \frac{1}{2}$$

i.e.

$$E(X_n) = 0$$

pas très étonnant !

3. $u_n = P(X_{2n} = 0) = P(2n - 2S_{2n} = 0) = P(S_{2n} = n) = \binom{2n}{n} \frac{1}{4^n}$
d'où

$$\frac{u_{n+1}}{u_n} = \frac{\binom{2n+2}{n+1} \frac{1}{4^{n+1}}}{\binom{2n}{n} \frac{1}{4^n}} = \frac{2n+1}{2(n+1)}$$

$$\lim_{n \rightarrow +\infty} \frac{u_{n+1}}{u_n} = 1$$

Si A est ruiné à un rang pair, asymptotiquement, il a toutes les chances de resté ruiné par la suite. (Avec au moins la satisfaction de ne pas avoir de dettes.)

Exercice 3

- 1.

```
1 import matplotlib.pyplot as plt
2
3 def polynome(t,x):
4     return(x**3+t*x**2+1)
5
6 def graphe(a,b,t):
7     plt.figure('P_t')
8     plt.clf()
9     # Dessin de la courbe
10    X=np.linspace(a,b,(b-a)*100)
11    Y=[polynome(t,x) for x in X]
12    plt.plot(X,Y)
13    plt.show()
```

2. On dérive le polynôme t (par-rapport à x bien entendu, car t est un indice fixé) :

$$P'_t(x) = 3x^2 + 2xt = x(3x + 2t)$$

On peut faire un tableau de signe de P'_t facilement :

x	$-\infty$	$-\frac{2t}{3}$	0	$+\infty$
$P'_t(x)$		+	0	-

On observe sur le tableau de variation de P_t (tableau que l'on prendra soin de dessiner pour une meilleure observation) que $P_t(0) = 1$ et

$$P_t(-\frac{2t}{3}) \geq P_t(0) > 0.$$

Grâce aux sens de variations, comme P_t est continue, par le théorème de la bijection, on observera qu'il existe une unique racine sur $\mathbb{R}$, cette racine étant plus précisément comprise entre $-\infty$ et $-\frac{2t}{3}$ ou plus largement entre $-\infty$ et 0 .

3. $r(2)$ est l'abscisse x correspondant à $P_2(x) = 0$. Pour la remettre en ordonnée, il suffit par exemple de dessiner le cercle de centre l'origine du repère et de rayon x et de garder l'ordonnée obtenue par intersection avec l'axe des ordonnées.
4. a) $r(0) = -1$:

En effet, $r(0)$ est la racine de P_0 par définition. Or, $P_0(x) = x^3 + 1$. Sa seule racine réelle est bien $r(0) = -1$.

b) r est négative :

Ceci a été démontré dans la première question où on a stipulé (après étude du tableau de variation) que la seule racine possible était comprise entre $-\infty$ et $-\frac{2t}{3} \leq 0$

c) Manifestement, $\lim_{t \rightarrow +\infty} r(t) = -\infty$:

On a vu dans la première question que $r(t) \leq -\frac{2t}{3}$. Par théorème de comparaison des limites, si on fait tendre t vers $+\infty$, on obtient le résultat.

d) On dirait bien que la branche infinie de la courbe en $+\infty$ est

$$y = -x$$

En effet, $r(t)$ est racine de P_t , c'est-à-dire que

$$P_t(r(t)) = 0$$

ce qui se traduit par

$$r(t)^3 + tr(t)^2 + 1 = 0$$

d'où, en divisant par $tr(t)^2$:

$$\frac{r(t)}{t} = -1 - \frac{1}{t \cdot r(t)^2}$$

Or, le membre de droite tend vers -1 (oui, rappelons que d'après la question précédente, on a $\lim_{+\infty} r = -\infty$) D'où

$$\lim_{t \rightarrow +\infty} \frac{r(t)}{t} = -1$$

On obtient donc une direction asymptotique

$$y = -x$$

Mais est-ce une asymptote ? Il faut étudier $\lim_{t \rightarrow +\infty} r(t) + t$, mais d'après l'égalité précédente,

$$r(t) + t = -\frac{1}{r^2(t)} \xrightarrow{t \rightarrow +\infty} 0$$

On est donc maintenant certain que r admet une asymptote

$$y = -x$$

5. Méthode 1 :

Il s'agit de trouver le sens de variation de r (mais ceci n'est pas trop dans l'idée de l'exercice car c'est ce qui est demandé dans la question suivante :)

On sait d'après les questions précédentes que le début du tableau de variation est le suivant :

t	0	$+\infty$
$r(t)$	-1	$-\infty$

Mais nous n'en savons pas plus. Si on réussit à démontrer que r est strictement décroissante, alors on aura bien une bijection de $[0, +\infty[$ dans $] -\infty, 0]$.

Comme on n'a pas de formule pour r , on procède de la façon suivante :

Soient $u, v \geq 0$ tels que $u < v$. Montrons que $r(u) > r(v)$. (On note ici ru et rv respectivement pour ces deux quantités.)

On sait que

$$P_u(ru) = 0 = P_v(rv)$$

et que

$$ru, rv \in] -\infty, -\frac{2t}{3}]$$

Supposons que $ru < rv$. Comme P_u est croissante sur $] -\infty, -\frac{2t}{3}]$ (cf première question), on a

$$0 = P_u(ru) < P_u(rv)$$

d'où, comme $0 = P_v(rv)$, on a

$$P_v(rv) < P_u(rv)$$

Mais pour tout x , on a

$$P_u(x) > P_v(x) \Leftrightarrow x^3 + ux^2 + 1 > x^3 + vx^2 + 1 \Leftrightarrow ux^2 > vx^2 \Leftrightarrow u > v$$

D'où finalement $u > v$, ce qui est contradictoire. L'hypothèse de départ est donc mauvaise. On a donc

$$r_u > r_v$$

et r est donc décroissante, d'où le tableau de variation :

t	0	$+\infty$
$r(t)$	-1	$-\infty$

Ainsi, r est bijective de $\mathbb{R}^+$ vers $] -\infty, -1]$. On note s sa réciproque. Cherchons sa formule : On sait que

$$\begin{aligned} t = s(y) &\Leftrightarrow y = r(t) \\ &\Leftrightarrow y^3 + ty^2 + 1 = 0 \\ &\Leftrightarrow t = \frac{-1 - y^3}{y^2} \text{ car } y \neq 0 \text{ sinon } 0 = 1 \text{ dans l'égalité ci-dessus} \end{aligned}$$

D'où

$$s(y) = \frac{-1 - y^3}{y^2}$$

Méthode 2 : (la plus indiquée au vue de l'énoncé)

Montrons que r est injective sur son intervalle de définition $[0, +\infty[$. Soit $y = r(t)$. Montrons que t est unique :

$$\begin{aligned} y = r(t) &\Leftrightarrow y^3 + ty^2 + 1 = 0 \\ &\Leftrightarrow t = \frac{-1 - y^3}{y^2} \text{ car } y \neq 0 \text{ sinon } 0 = 1 \text{ dans l'égalité ci-dessus} \end{aligned}$$

Il n'y a donc qu'une seule possibilité pour t .

De plus, on a

$$\begin{aligned} t \geq 0 &\Leftrightarrow \frac{-1 - y^3}{y^2} \geq 0 \\ &\Leftrightarrow -1 - y^3 \geq 0 \text{ (car } y^2 > 0) \\ &\Leftrightarrow -1 \geq y^3 \\ &\Leftrightarrow -1 \geq y \\ &\Leftrightarrow -1 \geq r(t) \end{aligned}$$

On en déduit que

$$r : [0, +\infty[\rightarrow] -\infty, -1]$$

et que de plus r est surjective sur ces intervalles (car si $y \in] -\infty, -1]$, on a $t = \frac{-1 - y^3}{y^2} \geq 0$ et $r(t) = y$.)

(Sinon, on peut également étudier l'application $y \mapsto \frac{-1-y^3}{y^2}$ pour le constater.)
Autrement dit, $rr : [0, +\infty[\rightarrow]-\infty, -1]$ est injective et surjective, donc bijective, et la formule de sa réciproque est

$$s(y) = \frac{-1-y^3}{y^2}$$

6. la fonction obtenue pour s est continue, dérivable de dérivée strictement négative. (à calculer) Il en va donc de même pour r et elle est également décroissante.

7. On fait une dichotomie sur $[a, 0]$ (ou $[a, -\frac{2t}{3}]$) où $P_t(a) < 0$. Cherchons donc d'abord un tel a .

Observons par exemple que

$$P_t(-2t) = (-2t)^3 + t(-2t)^2 + 1 = -8t^3 + 4t^3 + 1 = -4t^3 + 1$$

Ainsi, si $t > 1$, $P_t(-2t) < -3 < 0 = P_t(r(t))$, donc, par croissance de P_t , on a

$$-2t \leq r(t)$$

On peut donc prendre

$$a = -2t$$

Si $t \in [0, 1]$,

$$P_t(x) = x^3 + tx^2 + 1 \leq x^3 + x^2 + 1$$

Avec $x = -2$,

$$P_t(-2) \leq -8 + 5 < 0$$

Donc

$$-2 < r(t)$$

on peut donc prendre

$$a = -2.$$

Pour résumer, si on veut ne faire qu'un seul cas, on peut prendre directement $a = \min(-2t, -2)$.

```

1 def Dichotomie(t,n):
2     """Recherche la racine d'une fonction à 10^(-n
3         ) près."""
4     gauche=min(-2,-2*t)
5     droite=0
6     while droite-gauche>=10**(-n):
7         milieu=(gauche+droite)/2
8         if milieu**3+t*milieu**2+1>0:
9             droite=milieu
10        else:
11            gauche=milieu
12    return(milieu)

```

8.

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 def Courbe(n):
5     plt.clf()
6     T=np.linspace(0,10,101)
7     Y=[Dichotomie(t,n) for t in T]
8     plt.plot(T,Y)
9
10 n=30
11 taquet = 0.999
12 correlations(n,taquet)

```

Exercice 4

1. a) Soit S_i la variable de Bernoulli de succès $a < X_i < b$. Dans ce cas, on a

$$S_1 + \dots + S_N = \text{Nombre de } i \text{ tels que } a < X_i < b$$

D'après la loi faible des grands nombres, on sait que

$$\mathbb{E}[S_1] \simeq \frac{S_1 + \dots + S_N}{N}$$

En parallèle, on sait également que

$$\mathbb{E}[S_1] = P(\text{"succès"}) = P(a < X_1 < b) = P(a < X < b)$$

CQFD

- b) Construire l'histogramme du N échantillon consiste en abscisse à découper les résultats en classes délimitées par des nombres $a_1 < \dots < a_n$ où $a_1 < \min X_i$ et $a_n > \max X_i$ et à dessiner un rectangle de base $[a_i, a_{i+1}[$ d'aire la fréquence d'obtention de la classe. On obtient donc, d'après la question précédente, pour chaque classe $[a_i, a_{i+1}[$, un rectangle dont l'aire vaut environ

$$P(a_i < X < a_{i+1})$$

Ce qui correspond en parallèle à l'aire sous la courbe de la densité de X si X est une variable à densité.

2.

```

1 def Simul_Uniforme_Croissant(n):
2     """Simule n variables aléatoires Xi suivant
        une loi uniforme et les range par ordre
        croissant."""
3     X=[rd.random() for i in range(n)]
4     # On va classer ces nombres par comparaison 2
        à 2:
5     for i in range(n): # pour chaque i fixé, on
        met en premier le min de X[i:] où les i-1
        premiers termes sont déjà classés.
6         for j in range(i+1,n):
7             if X[j]<X[i]:
8                 [X[i],X[j]]=X[j],X[i]]
9     return(X)

```

3. a)

```

1 def histogramme_Unif_Croissant(r,n):
2     plt.clf()
3     """Dessine un histogramme de Yr en découpant l'
        univers en p classes et en réalisant N fois
        l'expérience."""
4     N=2000
5     nb_classes=50
6     if r>n:
7         print('impossible, r doit être <=n')
8     else:
9         Yr=[simul_Uniforme_Croissant(n)[r-1] for i
            in range(N)]
10        plt.hist(Yr,density=True,bins=np.linspace
            (0,1,nb_classes+1))
11        plt.title('r='+str(r)+' , n='+str(n))
12        plt.show()

```

- b) Notons en premier lieu que quoi qu'il en soit, on a $\text{Supp}(Y_r) = [0, 1]$. Ainsi, il est impossible que l'histogramme se prolonge à gauche de 0 et à droite de 1. Pour que la loi coïncide avec une loi normale, il faudrait que la courbe soit "en cloche", comme la densité d'une loi normale. Or ici, on voit bien qu'aucune des deux courbes ne vérifie cette propriété. Il semblerait donc que ces variables ne suivent pas du tout une loi normale.

- c) L'allure de l'histogramme est moins flagrante, la courbe est plus en triangle qu'en cloche, mais ça se rapproche quand même d'une loi normale. Peut être que cette loi est proche d'une loi normale.
4. Les paramètres de la loi normale étant son espérance et sa variance, on choisirait comme paramètre l'espérance et la variance de Y_r . On ne les connaît pas ici, mais on sait grâce à la loi faible des grands nombres que la moyenne empirique est un estimateur sans biais de l'espérance et que la variance empirique est un estimateur de la variance. On peut donc utiliser ces données.
5. a) N_x compte le nombre de succès dans la suite des n événements indépendants $X_i \leq x$ de probabilité :

$$\begin{cases} x & \text{si } x \in [0, 1] \\ 0 & \text{si } x < 0 \\ 1 & \text{si } x \geq 1 \end{cases}$$

Ainsi,

$$N_x \begin{cases} \hookrightarrow \mathcal{B}(n, x) & \text{si } x \in [0, 1] \\ \text{est constante égale à } 0 & \text{si } x < 0 \\ \text{est constante égale à } n & \text{si } x > 1 \end{cases}$$

b) On a

$$\{N_x \geq r\} = \{Y_r \leq x\}$$

- c) • Fonction de répartition F_r de Y_r :

On a que $\text{Supp}(Y_r) = [0; 1]$, d'où

$$F_r(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{si } t > 1 \end{cases}$$

Si $t \in [0; 1]$, on a

$$\begin{aligned} F_r(t) &= P(Y_r \leq t) \\ &= P(N_t \geq r) \\ &= \sum_{k=r}^n \binom{n}{k} t^k (1-t)^{n-k} \end{aligned}$$

D'où

$$F_r(t) = \begin{cases} 0 & \text{si } t < 0 \\ \sum_{k=r}^n \binom{n}{k} t^k (1-t)^{n-k} & \text{si } t \in [0; 1] \\ 1 & \text{si } t > 1 \end{cases}$$

Remarquons qu'on retrouve la loi du max

$$F_n(t) = \begin{cases} 0 & \text{si } t < 0 \\ t^n & \text{si } t \in [0; 1] \\ 1 & \text{si } t > 1 \end{cases}$$

et du min :

$$F_1(t) = \begin{cases} 0 & \text{si } t < 0 \\ \sum_{k=1}^n \binom{n}{k} t^k (1-t)^{n-k} & \text{si } t \in [0; 1] \\ 1 & \text{si } t > 1 \end{cases} = \begin{cases} 0 & \text{si } t < 0 \\ 1 - (1-t)^n & \text{si } t \in [0; 1] \\ 1 & \text{si } t > 1 \end{cases}$$

d) Soit $t \in [0, 1]$. On a

$$\begin{aligned} F_{n+1-r}(t) &= P(Y_{n+1-r} \leq t) \\ &= P(N_t \geq n+1-r) \\ &= 1 - P(N_t < n+1-r) \\ &= 1 - P(N_t \leq n-r) \\ &= \sum_{k=0}^{n-r} \binom{n}{k} t^k (1-t)^{n-k} \end{aligned}$$

d'un autre côté,

$$\begin{aligned} 1 - F_r(1-t) &= 1 - P(Y_r \leq 1-t) \\ &= 1 - P(N_{1-t} \geq r) \\ &= 1 - \sum_{j=r}^n \binom{n}{j} (1-t)^j t^{n-j} \\ &= 1 - \sum_{k=n-r}^0 \binom{n}{n-k} (1-t)^{n-k} t^k \quad \text{en posant } k = n-j \\ &= 1 - \sum_{k=0}^{n-r} \binom{n}{k} t^k (1-t)^{n-k} \quad \text{car } \binom{n}{k} = \binom{n}{n-k} \\ &= F_{n+1-r}(t) \end{aligned}$$

i.e. au final

$$\boxed{F_{n+1-r}(t) = 1 - F_r(1-t)}.$$

6. a) On observe que

$$\lim_{t \rightarrow 0^-} F_r(t) = 0 = F_r(0)$$

et

$$\lim_{t \rightarrow 1^-} F_r(t) = 1 = \binom{n}{n} 1^n = F_r(1)$$

la fonction F est donc continue sur $\mathbb{R}$. Elle est également clairement $\mathcal{C}^1$ au moins sur $\mathbb{R} - \{0, 1\}$. En conclusion,

$$\boxed{Y_r \text{ est à densité}}$$

b) On note f une densité de Y_r . Comme Y_r est à support dans $[0; 1]$, on a

$$\mathbb{E}(Y_r) = \int_0^1 x f(x) dx = [x F_{Y_r}(x)]_0^1 - \int_0^1 F_{Y_r}(x) dx = 1 - \int_0^1 F_{Y_r}(x) dx$$

• Calculons $\int_0^1 F_{Y_r}(x) dx$:

$$\int_0^1 F_{Y_r}(x) dx = \sum_{k=r}^n \underbrace{\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx}_{I_{k,n}}$$

Soit $k \in \{1, \dots, n\}$, on a

$$\begin{aligned} I_{k,n} &= \int_0^1 \binom{n}{k} \underbrace{t^k}_u \underbrace{(1-t)^{n-k}}_{v'} dx \\ &= \left[-t^k \frac{(1-t)^{n-k+1}}{n-k+1} \right]_0^1 + \int_0^1 \binom{n}{k} \frac{k}{n-k+1} x^{k-1} (1-x)^{n-k+1} dx \\ &= \int_0^1 \underbrace{\binom{n}{k} \frac{k}{n-k+1}}_{\binom{n}{k-1}} t^{k-1} (1-x)^{n-k+1} dx \quad (\text{car } n-k+1 \neq 0) \\ &= I_{k-1,n} \end{aligned}$$

$$\begin{aligned} \text{Ainsi, } I_{k,n} = I_{0,n} &= \int_0^1 \binom{n}{0} t^0 (1-t)^n dx \\ &= \int_0^1 (1-t)^n dx = \left[\frac{-(1-t)^{n+1}}{n+1} \right]_0^1 \\ &= \frac{1}{n+1} \end{aligned}$$

$$\text{Finalement, } \int_0^1 F_{Y_r}(x) dx = \sum_{k=r}^n I_{k,n} = \sum_{k=r}^n \frac{1}{n+1} = \frac{n-r+1}{n+1}$$

• Conclusion :

$$\boxed{\mu_r = E(Y_r) = 1 - \frac{n-r+1}{n+1} = \frac{r}{n+1}}$$

c) On a

$$E(Y_i^2) = \int_0^1 x^2 f(x) dx = [x^2 F_{Y_i}(x)]_0^1 - 2 \int_0^1 x F_{Y_i}(x) dx = 1 - 2 \int_0^1 x F_{Y_i}(x) dx$$

• calculons $\int_0^1 x F_{Y_i}(x) dx$:

$$\begin{aligned} \int_0^1 x F_{Y_i}(x) dx &= \sum_{k=i}^n \underbrace{\int_0^1 \binom{n}{k} x^{k+1} (1-x)^{n-k} dx}_{= \frac{\binom{n}{k}}{\binom{n+1}{k+1}} I_{k+1, n+1}} \\ &= \frac{\binom{n}{k}}{\binom{n+1}{k+1}} I_{k+1, n+1} \end{aligned}$$

Or, on sait déjà que

$$I_{k, n} = \frac{1}{n+1}$$

et de plus,

$$\frac{\binom{n}{k}}{\binom{n+1}{k+1}} = \frac{k+1}{n+1}$$

d'où

$$\begin{aligned} \int_0^1 x F_{Y_i}(x) dx &= \sum_{k=i}^n \frac{k+1}{n+1} I_{k+1, n+1} \\ &= \frac{1}{(n+1)(n+2)} \sum_{k=i}^n (k+1) \\ &= \frac{1}{(n+1)(n+2)} \sum_{k=i+1}^{n+1} k \\ &= \frac{1}{(n+1)(n+2)} \left(\frac{(n+1)(n+2)}{2} - \frac{i(i+1)}{2} \right) \\ &= \frac{1}{2} - \frac{i(i+1)}{2(n+1)(n+2)} \end{aligned}$$

et donc

$$E(Y_i^2) = 1 - 2 \left(\frac{1}{2} - \frac{i(i+1)}{2(n+1)(n+2)} \right) = \frac{i(i+1)}{(n+1)(n+2)}$$

Au final,

$$V(Y_r) = E(Y_r^2) - E(Y_r)^2 = \frac{r(r+1)}{(n+1)(n+2)} - \left(\frac{r}{n+1} \right)^2 = \frac{r}{n+1} \left(\frac{r+1}{n+2} - \frac{r}{n+1} \right)$$

i.e.

$$\boxed{V(Y_r) = \frac{r}{(n+1)^2} \left(\frac{n-r+1}{n+2} \right)}$$

7. a) On a, d'après la question précédente (valeur de l'espérance)

$$\begin{aligned} \mathbb{E}[Y_{n+1-r}] &= \frac{n+1-r}{n+1} \\ &= \frac{n+1}{n+1} - \frac{r}{n+1} \\ &= 1 - \mathbb{E}[Y_r] \end{aligned}$$

i.e.

$$\boxed{\mathbb{E}[Y_{n+1-r}] = 1 - \mathbb{E}[Y_r]}$$

b) on a, d'après la question précédente (valeur de la variance)

$$\begin{aligned} \mathbb{V}[Y_{n+1-r}] &= \frac{n+1-r}{(n+1)^2} \left(\frac{n-(n+1-r)+1}{n+2} \right) \\ &= \frac{n+1-r}{(n+1)^2} \left(\frac{r}{n+2} \right) \\ &= \mathbb{V}[Y_r] \end{aligned}$$

8. Posons T_r une variable telle que

$$T_r \hookrightarrow \mathcal{N}(\mu_r, \sigma_r^2).$$

On a alors, d'après les questions 7a et 7b),

$$T_{n+1-r} \hookrightarrow \mathcal{N}(\underbrace{\mu_{n+1-r}}_{=1-\mu_r}, \underbrace{\sigma_{n+1-r}^2}_{=\sigma_r^2}).$$

D'où

$$T_{n+1-r} \hookrightarrow \mathcal{N}(1 - \mu_r, \sigma_r^2).$$

Ainsi,

$$1 - T_{n+1-r} \hookrightarrow \mathcal{N}(\mu_r, \sigma_r^2)$$

ce qui est la même loi que T_r .

Puis

$$\begin{aligned} \Phi_{n+1-r}(t) &= P(T_{n+1-r} \leq t) \\ &= P(1 - T_{n+1-r} \leq 1 - t) \\ &= P(T_r \leq 1 - t) \text{ car } T_r \text{ et } 1 - T_{n+1-r} \text{ ont la même loi} \\ &= \Phi_r(1 - t) \end{aligned}$$

9. En établissant une relation entre les abscisses et les ordonnées, on établit une relation entre F_r et Φ_r . Ainsi, si on a une droite d'ajustement fiable de type $y = ax + b$, cela signifie que

$$F_r(x) = a\Phi_r(x) + b$$

Autrement dit, $F_r = \Phi_r$ signifie qu'on aurait $a = 1$ et $b = 0$. Dans les autres cas, cela signifie qu'on n'aura pas $F_r \simeq \Phi_r$.

10. a) On constate que la droite d'ajustement est plus ou moins proche de la droite

$$y = x$$

(Mis à part pour le cas $r = 1$ ou on est un peu plus loin.)

ce qui signifierait, si le coefficient de corrélation est bon, que $F_r \simeq \Phi_r$, autrement dit, que la loi de Y_r peut être considérée comme étant la loi normale de paramètre μ_r et σ_r .

Or, les coefficients ne sont très bons ici que dans le cas où $r = 10, n = 20$ (ou encore dans le cas où $r = 3, n = 6$). Ceci confirme les histogrammes obtenus au début, où on constate que la loi de Y_r est plus ou moins proche d'une normale suivant les valeurs de r . Il semblerait pour l'instant que si r est proche de 1 ou n , l'approximation n'est pas bonne du tout.

- b) • Montrons que $\overline{X^{(n+1-r)}} = 1 - \overline{X^{(r)}}$:

Par définition de la moyenne, on a

$$\begin{aligned} \overline{X^{(n+1-r)}} &= \frac{1}{N+1} \sum_{i=0}^N X_i^{(n+1-r)} \\ &= \frac{1}{N+1} \sum_{i=0}^N F_{n+1-r} \left(\frac{i}{N} \right) \quad \text{car } X_i^{(j)} = F_j \left(\frac{i}{N} \right) \\ &= \frac{1}{N+1} \sum_{i=0}^N \left(1 - F_r \left(1 - \frac{i}{N} \right) \right) \\ &\quad \text{car } t = \frac{i}{N} \in [0, 1] \text{ et } F_{n+1-r}(t) = 1 - F_r(1-t) \\ &= \frac{1}{N+1} \sum_{i=0}^N \left(1 - F_r \left(\frac{N-i}{N} \right) \right) \\ &= \frac{1}{N+1} \sum_{j=N}^0 \left(1 - F_r \left(\frac{j}{N} \right) \right) \quad \text{avec le changement d'indice } j = N-i \\ &= \frac{N+1}{N+1} - \underbrace{\frac{1}{N+1} \sum_{j=0}^N F_r \left(\frac{j}{N} \right)}_{=\overline{X^{(r)}}} \\ &= 1 - \overline{X^{(r)}} \end{aligned}$$

D'où

$$\boxed{\overline{X^{(n+1-r)}} = 1 - \overline{X^{(r)}}}$$

- Montrons que $\overline{Y^{(n+1-r)}} = 1 - \overline{Y^{(r)}}$:

La propriété sur $Y_i^{(j)} = \Phi_j \left(\frac{i}{N} \right)$ est exactement la même que celle de $X_i^{(j)}$, c'est-à-dire

$$\Phi_{n+1-r}(t) = 1 - \Phi_r(1-t) \quad \forall t \in [0, 1]$$

Le calcul fait pour X est donc exactement le même pour Y et la propriété finale en résulte également :

$$\boxed{\overline{Y^{(n+1-r)}} = 1 - \overline{Y^{(r)}}}$$

11. a) On sait que par définition, on a

$$c_j = \frac{Cov(X^{(j)}, Y^{(j)})}{\mathbb{V}(X^{(j)})\mathbb{V}(Y^{(j)})}$$

Mais

$$Cov(X^{(j)}, Y^{(j)}) = \frac{1}{N+1} \sum_{i=0}^N (X_i^{(j)} - \overline{X^{(j)}})(Y_i^{(j)} - \overline{Y^{(j)}})$$

Ainsi,

$$\begin{aligned} Cov(X^{(n+1-r)}, Y^{(n+1-r)}) &= \frac{1}{N+1} \sum_{i=0}^N (X_i^{(n+1-r)} - \overline{X^{(n+1-r)}})(Y_i^{(n+1-r)} - \overline{Y^{(n+1-r)}}) \\ &= \frac{1}{N+1} \sum_{i=0}^N (X_i^{(n+1-r)} - (1 - \overline{X^{(r)}}))(Y_i^{(n+1-r)} - (1 - \overline{Y^{(r)}})) \\ &\quad \text{car } \overline{X^{(n+1-r)}} = 1 - \overline{X^{(r)}} \text{ et } \overline{Y^{(n+1-r)}} = 1 - \overline{Y^{(r)}} \\ &= \frac{1}{N+1} \sum_{i=0}^N \left(\underbrace{F_{n+1-r} \left(\frac{i}{N} \right) - 1 + \overline{X^{(r)}}}_{= F_r \left(1 - \frac{i}{N} \right)} \right) \left(\underbrace{\Phi_{n+1-r} \left(\frac{i}{N} \right) - 1 + \overline{Y^{(r)}}}_{= \Phi_r \left(1 - \frac{i}{N} \right)} \right) \\ &= \frac{1}{N+1} \sum_{i=0}^N \left(F_r \left(\frac{N-i}{N} \right) - \overline{X^{(r)}} \right) \left(\Phi_r \left(\frac{N-i}{N} \right) - \overline{Y^{(r)}} \right) \\ &= \frac{1}{N+1} \sum_{j=0}^N \left(F_r \left(\frac{j}{N} \right) - \overline{X^{(r)}} \right) \left(\Phi_r \left(\frac{j}{N} \right) - \overline{Y^{(r)}} \right) \\ &\quad \text{(avec le changement d'indice } j = N - i) \\ &= Cov(X^{(r)}, Y^{(r)}) \quad \text{(par définition de } Cov(X^{(r)}, Y^{(r)}).) \end{aligned}$$

le même procédé permet de montrer que

$$\mathbb{V}(X^{(n+1-r)}) = \mathbb{V}(X^{(r)})$$

ainsi que

$$\mathbb{V}(Y^{(n+1-r)}) = \mathbb{V}(Y^{(r)})$$

puis de conclure que

$$c_r = c_{n+1-r}$$

b) Les courbes sont les points (r, c_r) , avec $r = 0, \dots, n$. Comme on a $c_r = c_{n+1-r}$, ceci signifie que pour les abscisses r et $n+1-r$, les ordonnées sont les mêmes. La courbe est donc symétrique par-rapport à l'axe d'abscisse x qui vérifie $x = n+1-x$, c'est-à-dire

$$x = \frac{n+1}{2}$$

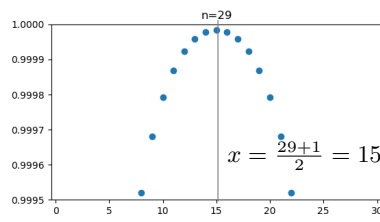
12. b) Les courbes semblent d'abord croissantes jusqu'à un certain point, puis décroissantes (à cause de la symétrie déjà expliquée précédemment.) On dirait bien qu'il y a deux points pour $n = 10$ et un seul pour $n = 29$. Il se pourrait qu'il y a deux points pour n pair et un seul pour n impair. Ceci s'explique par le fait que, si on admet que les courbes sont bien strictement croissantes puis strictement décroissantes, on sait qu'elles sont symétriques par-rapport à l'axe $x = \frac{n+1}{2}$. Ainsi :

- Si n est impair :

Alors $n+1$ est pair et $\frac{n+1}{2}$ est entier. Aussi, le maximum sera atteint au point d'intersection avec l'axe de symétrie. Il n'y aura donc qu'un **seul maximum**, et ceci pour

$$r = \frac{n+1}{2}$$

par exemple pour $n = 29$:



Un seul point maximum

- Si n est pair :

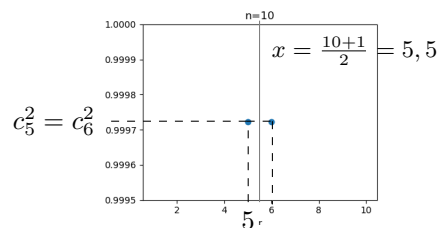
Alors $n+1$ est impair alors $\frac{n+1}{2}$ n'est pas entier. En revanche, les deux entiers qui l'entourent sont $r = \lfloor \frac{n+1}{2} \rfloor = \frac{n}{2}$ et $\lfloor \frac{n+1}{2} \rfloor + 1 = \frac{n}{2} + 1$. De plus, on sait que $c_r = c_{n-r+1}$ donc

$$c_{\frac{n}{2}} = c_{n-\frac{n}{2}+1} = c_{\frac{n}{2}+1}$$

Ainsi, par positivité,

$$c_{\frac{n}{2}}^2 = c_{\frac{n}{2}+1}^2$$

Les deux points sont donc deux points maximaux.



13. La courbe étant croissante, on voit que plus n augment, plus $d_{max}^{(n)}$ augmente également. De plus, $d_{max}^{(n)}$ tend (graphiquement) vers 1 (on observe du moins que pour $n \geq 100$, le coefficient de détermination est déjà supérieur à 0,999999, ce qui est déjà une excellente valeur pour la fiabilité des nuages.

En conclusion, on peut admettre graphiquement que pour n grand et $r = \lfloor \frac{n+1}{2} \rfloor$, on a $F_r \simeq \Phi_r$ et d'une façon relativement précise, i.e. :

$$Y_r \underset{\sim}{\hookrightarrow} N(\mu_r, \sigma^2)$$

SI on centre et qu'on réduit :

$$\frac{Y_r - \mu_r}{\sigma_r} \underset{\sim}{\hookrightarrow} \mathcal{N}(0, 1)$$

On rappelle que

$$\mu_r = \frac{r}{n+1}$$

et

$$\sigma_r^2 = \frac{r}{(n+1)^2} \left(\frac{n-r+1}{n+2} \right)$$

Le plus simple est donc de prendre par exemple un n impair et grand : (par exemple $n = 201$ ou plus) puis on peut dire que

$$r = \frac{n+1}{2}$$

et donc

$$\mu_r = \frac{r}{n+1} = \frac{n}{2(n+1)} = \frac{1}{2}$$

$$\sigma_r^2 = \frac{r}{(n+1)^2} \left(\frac{n-r+1}{n+2} \right) = \frac{1}{2(n+1)} \left(\frac{n+2}{2(n+2)} \right) = \frac{1}{4(n+2)}$$

Autrement dit, pour n impair grand et $r = \frac{n+1}{2}$, on a

$$\frac{(Y_r - \frac{1}{2})\sqrt{n+2}}{2} \underset{\sim}{\hookrightarrow} \mathcal{N}(0, 1)$$

Comme on sait engendrer Y_r , (cf qst ??), on sait donc simuler la variable ci-dessus.

Commentaires : Connaissant le théorème central limite, on sait qu'il existe au moins une autre façon de construire une variable de loi normale (à peu près) en prenant une moyenne empirique sur un n échantillon d'une variable T : $\overline{T}_n = \frac{T_1 + \dots + T_n}{n}$ avec n grand car

$$\overline{T}_n \underset{\sim}{\hookrightarrow} \mathcal{N}\left(\mathbb{E}[T], \frac{\mathbb{V}(T)}{n}\right)$$

On aurait donc pu prendre directement $T_i = X_i$ et dire que

$$\overline{X}_n = \frac{\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]}{\frac{\sigma(X_1)}{\sqrt{n}}} \underset{\sim}{\hookrightarrow} \mathcal{N}(0, 1)$$

Or,

$$\mathbb{E}[X_1] = \frac{1}{2}$$

et

$$\mathbb{V}[X_1] = \frac{1}{12}$$

D'où

$$\overline{X}_n = 2\sqrt{3n} \left(\frac{X_1 + \dots + X_n}{n} - \frac{1}{2} \right) \underset{\sim}{\hookrightarrow} \mathcal{N}(0, 1)$$

En revanche, on ne sait pas quelle est la méthode la plus précise des deux.