

EXERCICE 1:

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

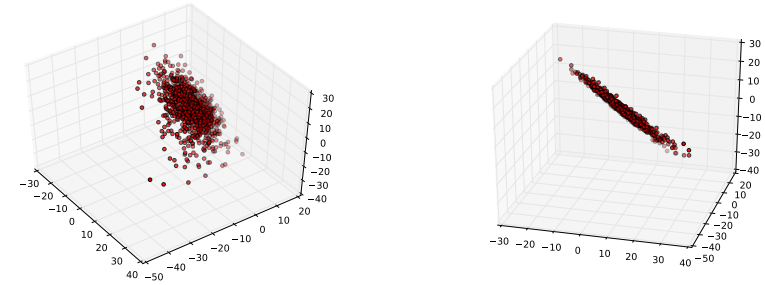
$$MX = \begin{pmatrix} Cov(X_1, U) \\ Cov(X_2, U) \\ Cov(X_3, U) \end{pmatrix} \quad \text{où } X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ et } U = \sum_{j=1}^3 x_j X_j.$$

- 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: On considère deux joueurs 1 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:

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).

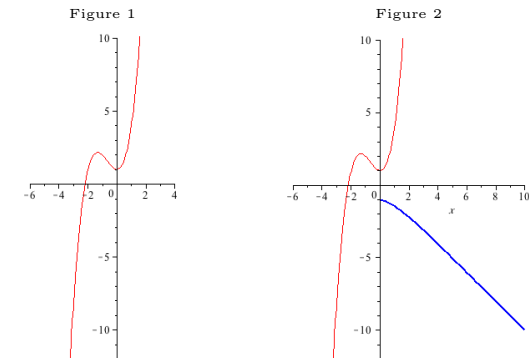
3. Expliquer comment il est possible de construire géométriquement sur cette figure le point de coordonnées $(2; r(2))$.
4. On a répété à l'aide de géogébra la construction précédente en faisant varier t de 0 à 10 avec un pas de 0.1, et on a obtenu la figure 2 ci-dessous. Que peut-on conjecturer relativement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

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 .

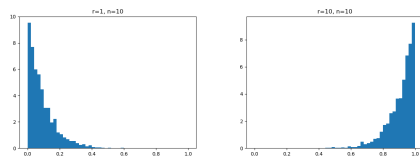
II - Classement croissant des variables uniformes :

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 .)

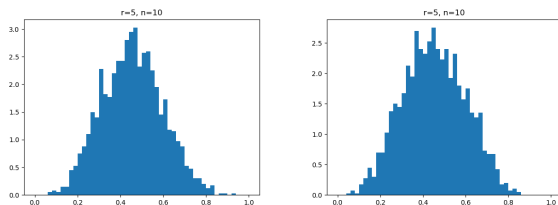
Soit $r \in \llbracket 1, n \rrbracket$. Le but de cet exercice est de savoir si on peut éventuellement se permettre d'approcher la loi de Y_r par une loi normale et peut être trouver une façon supplémentaire possible de simuler une loi normale.

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 ?

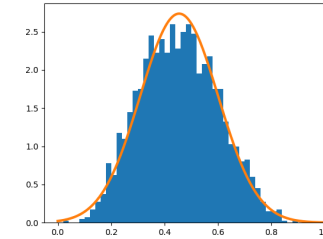


- c) Voici maintenant deux autres simulations pour le même r (ici en l'occurrence $r = 5$ pour $n = 10$.) La loi semble-t-elle coïncider avec une loi normale ?



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

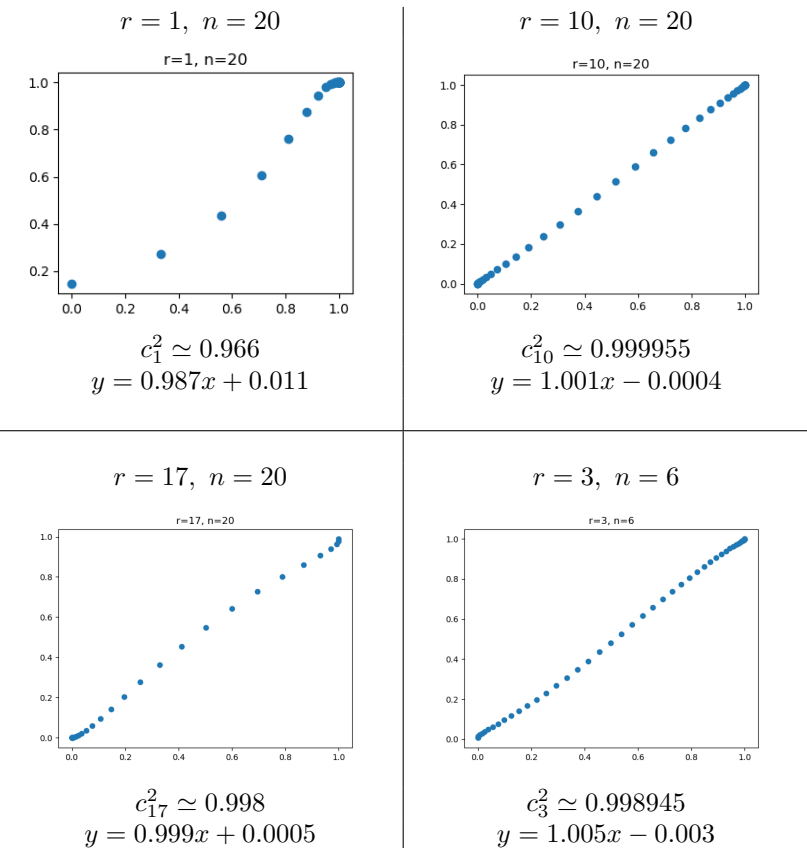
$$\{(X_i^{(r)}, Y_i^{(r)})\}_{i=0, \dots, M} = \{(F_r(x_i), \Phi_r(x_i))\}_{i=0, \dots, M}$$

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 ?

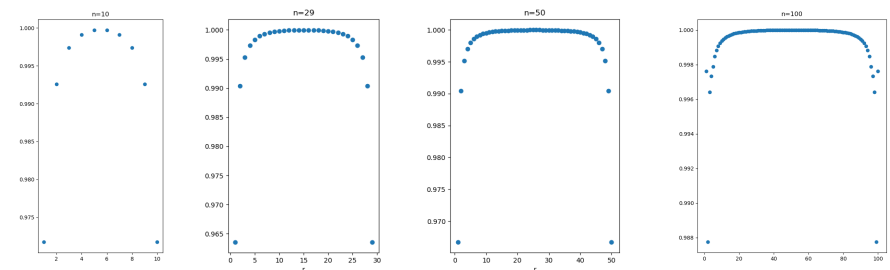
10. Ci-dessous quelques exemples de nuage pour $M = 50$ points, auquel on a rajouté :

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

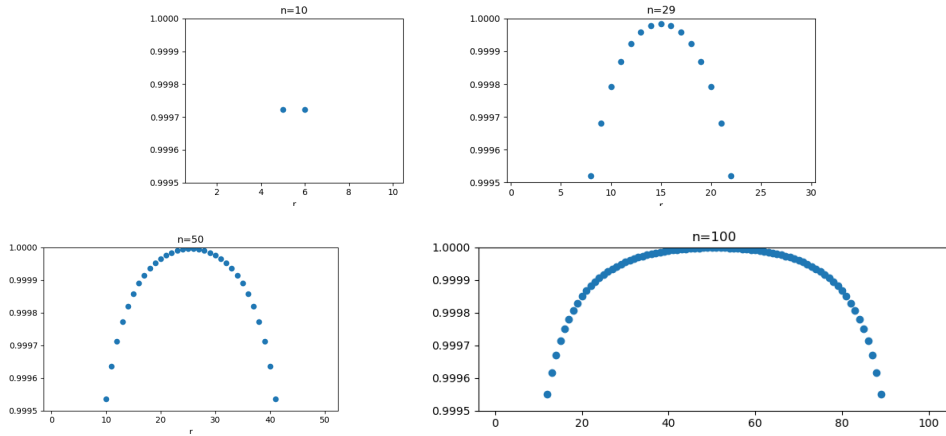


- a) Quelles informations peut-on tirer de ces données ?
- b) Notons $\overline{X^{(r)}}$ la moyenne des $X_i^{(r)}$ et $\overline{Y^{(r)}}$ la moyenne des $Y_i^{(r)}$. Montrer que

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.

