Exploration Numérique 2

Exercice 1

Comme $\{(X_1, Y_1)\}$ est une loi gaussienne de moyenne nulle, nous avons que $\mathbb{E}_\rho[X_1] = \mathbb{E}_\rho[Y_1] = 0$.

On rappelle que :

$$\Sigma = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, Y_1) \\ \mathrm{Cov}(X_1, Y_1) & \mathrm{Cov}(Y_1, Y_1) \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Donc $\mathrm{Cov}(X_1, X_1) = \mathrm{Var}_\rho[X_1] = \mathbb{E}_\rho[X^2] - \mathbb{E}_\rho[X_1]^2 = \mathbb{E}_\rho[X_1^2] = 1$.

Comme (X_1, Y_1) est un vecteur gaussian, X_1 et Y_1 suivent chacun une loi gaussienne. Vu leur variance, $X_1 \sim \mathcal{N}(0, 1)$ et $Y_1 \sim \mathcal{N}(0, 1)$.

Passons maintenant au calcul de $\mathbb{E}_\rho[X_1^4]$:

$$\mathbb{E}_\rho[X_1^4] = \int_{-\infty}^{+\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

On utilise le fait que :

$$I(\alpha) = \int_{-\infty}^{+\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}}$$

Donc :

$$I''(\alpha) = \int_{-\infty}^{+\infty} x^4 e^{-\alpha x^2} dx = \frac{3}{4}\sqrt{\frac{\pi}{\alpha^5}}$$

Pour $\alpha = 1/2$:

$$I''(1/2) = \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx = 3\sqrt{2\pi}$$

Donc :

$$\mathbb{E}_\rho[X_1^4] = \int_{-\infty}^{+\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} 3\sqrt{2\pi} = 3$$

Or, une fois qu'on a que $\mathbb{E}_\rho[X_1^4] = 3$, on a aussi que $\mathrm{Var}(X_1^2) = \mathbb{E}_\rho[X_1^4] - \mathbb{E}_\rho[X_1^2]^2 = 3 - 1 = 2$.

On remarque que la solution pour Y_1 est la même, ce qui nous permets d'affirmer que $\mathbb{E}_\rho[Y_1]^4 = 3$ et $\mathrm{Var}_\rho(Y_1^2) = 2$.

$\square$

Exploration Numérique 2

Exercice 2

Nous allons montrer que $Y = \rho X + \sqrt{1-\rho^2}Z$ où $Z \sim \mathcal{N}(0,1)$ et X et Z sont indépendants. Définissons la matrice :

$$A = \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix}$$

Nous avons donc que :

$$A\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X \\ -\frac{\rho}{\sqrt{1-\rho^2}}X + \frac{1}{\sqrt{1-\rho^2}}Y \end{pmatrix} := \begin{pmatrix} X \\ Z \end{pmatrix}$$

Donc $\boxed{Y = \rho X + \sqrt{1-\rho^2}Z}$ et par transformation affine d'un vecteur gaussien :

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim \mathcal{N}(0, A\Sigma A^t)$$

Ce qui implique que Z est centré. Nous avons que :

$$A\Sigma A^t = \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{\rho}{\sqrt{1-\rho^2}} \\ 0 & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Et donc :

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim \mathcal{N}(0, \mathrm{Id}_2)$$

Or, cela implique que X et Z sont indépendants et que $\boxed{Z \sim \mathcal{N}(0,1)}$, ce qu'on voulait démontrer.
Nous avons donc que :

$$\mathbb{E}_\rho[X_1^2Y_1^2] = \mathbb{E}_\rho[X_1^2(\rho^2X_1^2 + 2\rho\sqrt{1-\rho^2}X_1Z_1 + (1-\rho^2)Z_1^2)]$$

$$\mathbb{E}_\rho[X_1^2Y_1^2] = \rho^2\mathbb{E}_\rho[X_1^4] + 2\rho\sqrt{1-\rho^2}\mathbb{E}_\rho[X_1^3]\mathbb{E}_\rho[Z_1] + (1-\rho^2)\mathbb{E}_\rho[Z_1^2]$$

Comme $Z_1 \sim \mathcal{N}(0,1)$, $\mathbb{E}_\rho[Z_1^2] = 1$ et $\mathbb{E}_\rho[Z_1] = 0$. En utilisant aussi que $\mathbb{E}_\rho[X_1^4] = 3$:

$$\mathbb{E}_\rho[X_1^2Y_1^2] = 3\rho^2 + 1 - \rho^2 = 1 + 2\rho^2$$

□

Exercice 3

On développe les termes :

$$\mathrm{Cov}_\rho(X_1^2, X_1Y_1) = \mathrm{Cov}_\rho(X_1^2, X_1(\rho X_1 + \sqrt{1-\rho^2}Z_1))$$

$$\mathrm{Cov}_\rho(X_1^2, X_1Y_1) = \rho\mathrm{Cov}_\rho(X_1^2, X_1^2) + \sqrt{1-\rho^2}\mathrm{Cov}_\rho(X_1^2, X_1Z_1)$$

$$\mathrm{Cov}_\rho(X_1^2, X_1Y_1) = 2\rho + \sqrt{1-\rho^2}\mathrm{Cov}(X_1^2, X_1Z_1)$$

On observe que :

$$\mathrm{Cov}_\rho(X_1^2, X_1Z_1) = \mathbb{E}_\rho[(X_1^2 - \mathbb{E}_\rho[X_1^2])(X_1Z_1 - \mathbb{E}_\rho[X_1Z_1])]$$

Nous avons que $\forall n \in \mathbb{N}^*$, $\mathbb{E}_\rho[X_1^nZ_1] = \mathbb{E}_\rho[X_1^n]\mathbb{E}_\rho[Z_1] = 0$, donc :

$$\mathrm{Cov}_\rho(X_1^2, X_1Z_1) = \mathbb{E}_\rho[X_1^3Z_1 - X_1Z_1\mathbb{E}_\rho[X_1^2]] = \mathbb{E}_\rho[X_1^3Z_1] - \mathbb{E}_\rho[X_1^2]\mathbb{E}_\rho[X_1Z_1] = 0$$

Nous avons donc que :

$$\boxed{\mathrm{Cov}_\rho(X_1^2, X_1Y_1) = 2\rho}$$

Maintenant dévellopons l'autre terme :

$$\mathrm{Cov}_\rho(Y_1^2, X_1Y_1) = \mathrm{Cov}_\rho((\rho X_1 + \sqrt{1-\rho^2}Z_1)^2, X_1(\rho X_1 + \sqrt{1-\rho^2}Z_1))$$

$$\mathrm{Cov}_\rho(Y_1^2, X_1Y_1) = \mathrm{Cov}_\rho(\rho^2X_1^2 + (1-\rho^2)Z_1^2 + 2X_1Z_1\rho\sqrt{1-\rho^2}, \rho X_1^2 + \sqrt{1-\rho^2}X_1Z_1)$$

$$\mathrm{Cov}_\rho(Y_1^2, X_1Y_1) = \mathrm{Cov}_\rho(\rho^2X_1^2 + (1-\rho^2)Z_1^2 + 2X_1Z_1\rho\sqrt{1-\rho^2}, \rho X_1^2 + \sqrt{1-\rho^2}X_1Z_1)$$

$$\mathrm{Cov}_\rho(Y_1^2, X_1Y_1) = \rho^3\mathrm{Cov}_\rho(X_1^2, X_1^2) + \rho(1-\rho^2)\mathrm{Cov}_\rho(Z_1^2, X_1^2) + 2\rho\sqrt{1-\rho^2}\mathrm{Cov}_\rho(X_1Z_1, X_1^2) + \rho^2\sqrt{1-\rho^2}\mathrm{Cov}_\rho(X_1^2, X_1Z_1) + (1-\rho^2)^{\frac{3}{2}}\mathrm{Cov}_\rho(Z_1^2, X_1Z_1) + 2\rho(1-\rho^2)\mathrm{Cov}_\rho(X_1Z_1, X_1Z_1)$$

Comme X_1 et Z_1 sont indépendants, X_1^2 et Z_1^2 le sont aussi. En plus, de manière analogue à $\mathrm{Cov}_\rho(X_1^2, X_1Z_1) = 0$, $\mathrm{Cov}_\rho(Z_1^2, X_1Z_1) = 0$:

$$\mathrm{Cov}_\rho(Y_1^2, X_1Y_1) = 2\rho^3 + 2\rho(1-\rho^2)\mathrm{Cov}_\rho(X_1Z_1, X_1Z_1)$$

On calcule $\mathrm{Cov}_\rho(X_1Z_1, X_1Z_1) = \mathrm{Var}_\rho(X_1Z_1)$:

$$\mathrm{Cov}_\rho(X_1Z_1, X_1Z_1) = \mathbb{E}_\rho[X_1^2Z_1^2] - \mathbb{E}_\theta[X_1Z_1]^2 = \mathbb{E}_\rho[X_1^2]\mathbb{E}_\rho[Z_1^2] = 1$$

Donc :

$$\mathrm{Cov}_\rho(Y_1^2, X_1Y_1) = 2\rho^3 + 2\rho(1-\rho^2) = 2\rho^3 + 2\rho - 2\rho^3$$

$$\boxed{\mathrm{Cov}_\rho(Y_1^2, X_1Y_1) = 2\rho}$$

□

Exercice 4

Nous avons que $\mathbb{E}_\rho[X_i^2] = \mathbb{E}_\rho[Y_i^2] = 1$. En plus :

$$\mathbb{E}_\theta[X_iY_i] = \mathbb{E}_\theta[X_i(\rho X_i + \sqrt{1-\rho^2}Z_i)] = \rho$$

Ce qui confirme que $\mathbb{E}_\rho[U_i] = \mathbb{E}_\rho[U_1] = \mu(\rho)$ (n-échantillon).
Nous allons d'abord calculer la matrice de variance-covariance de U_1. Nous avons déjà que :

$$\mathrm{Cov}_\rho(X_1^2, X_1^2) = 2, \quad \mathrm{Cov}_\rho(Y_1^2, Y_1^2) = 2, \quad \mathrm{Cov}_\rho(Y_1^2, X_1Y_1) = \mathrm{Cov}_\rho(X_1^2, X_1Y_1) = 2\rho$$

Nous avons aussi que :

$$\mathrm{Cov}(X_1Y_1, X_1Y_1) = \mathbb{E}_\rho[X_1^2Y_1^2] - \mathbb{E}_\rho[X_1Y_1]^2 = 1 + 2\rho^2 - \mathbb{E}_\theta[\rho X_1^2 + \sqrt{1-\rho^2}X_1Z_1]^2$$

Comme X_1 et Z_1 sont indépendants :

$$\mathrm{Cov}(X_1Y_1, X_1Y_1) = 1 + 2\rho^2 - \rho^2\mathbb{E}_\theta[X_1^2] = 1 + 2\rho^2 - \rho^2$$

$$\mathrm{Cov}(X_1Y_1, X_1Y_1) = 1 + \rho^2$$

Il suffit donc de trouver $\mathrm{Cov}_\rho(X_1^2, Y_1^2)$:

$$\mathrm{Cov}_\rho(X_1^2, Y_1^2) = \mathrm{Cov}_\rho(X_1^2, \rho^2X_1^2 + 2\rho\sqrt{1-\rho^2}X_1Z_1 + (1-\rho^2)Z_1^2)$$

$$\mathrm{Cov}_\rho(X_1^2, Y_1^2) = \rho^2 \mathrm{Cov}_\rho(X_1^2, X_1^2) + 2\rho\sqrt{1-\rho^2}\mathrm{Cov}_\rho(X_1^2, X_1 Z_1)$$

Comme montré dans l'exercice 3, $\mathrm{Cov}_\rho(X_1^2, X_1 Z_1) = 0$, donc :

$$\mathrm{Cov}_\rho(X_1^2, Y_1^2) = 2\rho^2$$

Nous avons donc tous les éléments pour la matrice de variance-covariance de U_1 :

$$\Sigma_{U_1} = \begin{pmatrix} 2 & 2\rho^2 & 2\rho \\ 2\rho^2 & 2 & 2\rho \\ 2\rho & 2\rho & 1+\rho^2 \end{pmatrix} = V(\rho)$$

Nous avons donc, par le théorème centrale limite multi-dimensionel que :

$$\sqrt{n}(\bar{U}_n - \mu) \xRightarrow{\mathbb{P}_{n,\rho}} \mathcal{N}(0, V(\rho))$$

Exercice 5

Nous avons que :

$$S_{n,X}^2 = \sum_i^n \frac{1}{n}(X_i - \bar{X}_n)^2 = \overline{X_n^2} - \overline{X_n}^2$$

$$S_{n,Y}^2 = \sum_i^n \frac{1}{n}(Y_i - \bar{Y}_n)^2 = \overline{Y_n^2} - \overline{Y_n}^2$$

$$S_{n,XY}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X_n})(Y_i - \overline{Y_n}) = \overline{X_n Y_n} - \overline{X}_n \overline{Y}_n$$

Comme $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires i.i.d., nous avons par la LFGN que :

$$\overline{X_n} \xrightarrow{\mathbb{P}_{n,\rho}-\text{proba}} \mathbb{E}_\rho[X_1] = 0$$

Donc aussi :

$$\overline{X_n} \xRightarrow{\mathbb{P}_{n,\rho}} \mathbb{E}_\rho[X_1] = 0$$

On prend la fonction continue $x \mapsto x^2$:

Exploration Numérique 2

$$\overline{X_n}^2 \xRightarrow{\mathbb{P}_{n,\rho}} 0$$

Comme cela converge à une constante :

$$\overline{X_n}^2 \xrightarrow{\mathbb{P}_{n,\rho}-\text{proba}} 0$$

De manière analogue :

$$\overline{Y_n}^2 \xrightarrow{\mathbb{P}_{n,\rho}-\text{proba}} 0$$

En plus comme $\overline{X_n} \xrightarrow{\mathbb{P}_{n,\rho}-\text{proba}} 0$ et $\overline{Y_n} \xrightarrow{\mathbb{P}_{n,\rho}-\text{proba}} 0$, par Slutksy :

$$\overline{X}_n\overline{Y}_n \xRightarrow{\mathbb{P}_{n,\rho}} 0$$

Et comme cela est une constante :

$$\overline{X}_n\overline{Y}_n \xrightarrow{\mathbb{P}_{n,\rho}-\text{proba}} 0$$

Nous avons donc que :

$$\begin{pmatrix} S^2_{n,X} \\ S^2_{n,Y} \\ S^2_{n,XY} \end{pmatrix} = \begin{pmatrix} \overline{X^2_n} \\ \overline{Y^2_n} \\ \overline{X_nY_n} \end{pmatrix} - \begin{pmatrix} \overline{X_n}^2 \\ \overline{Y_n}^2 \\ \overline{X}_n\overline{Y}_n \end{pmatrix} = \bar{U}_n - \begin{pmatrix} \overline{X_n}^2 \\ \overline{Y_n}^2 \\ \overline{X}_n\overline{Y}_n \end{pmatrix}$$

Comme chaque coordonné du vecteur $(\overline{X_n}^2, \overline{Y_n}^2, \overline{X}_n\overline{Y}_n)$ converge en loi à 0, son vecteur converge en loi au vecteur $(0,0,0)^t$. Comme cela est une constante, le vecteur converge aussi en probabilité.

Remarquons que, comme $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires i.i.d, $\frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}(0, \frac{1}{n})$. En plus, $\sqrt{n}\overline{X_n} \sim \mathcal{N}(0,1)$, et donc $n\overline{X_n}^2 \sim \chi^2(1)$. Donc $\sqrt{n}\overline{X_n}^2 \xRightarrow{\mathbb{P}_{n,\rho}} 0$.

Par le même argument, $\sqrt{n}\overline{Y_n}^2 \xRightarrow{\mathbb{P}_{n,\rho}} 0$.

Remarquons en plus que :

$$\sqrt{n}|\overline{X}_n\overline{Y}_n| \leq \sqrt{n}\frac{1}{2}(\overline{X}^2_n + \overline{Y}^2_n) \xRightarrow{\mathbb{P}_{n,\rho}} 0$$

Donc $\sqrt{n}\overline{X}_n\overline{Y}_n \xRightarrow{\mathbb{P}_{n,\rho}} 0$, et le vecteur :

$$\sqrt{n}\begin{pmatrix}\overline{X_n}^2\\ \overline{Y_n}^2\\ \overline{X}_n\overline{Y}_n\end{pmatrix} \xRightarrow{\mathbb{P}_{n,\rho}} \begin{pmatrix}0\\0\\0\end{pmatrix}$$

En plus, comme cela converge à une constante :

$$\sqrt{n}\begin{pmatrix}\overline{X_n}^2\\ \overline{Y_n}^2\\ \overline{X}_n\overline{Y}_n\end{pmatrix} \xrightarrow{\mathbb{P}_{n,\rho}-\text{proba}} \begin{pmatrix}0\\0\\0\end{pmatrix}$$

Nous pouvons donc appliquer Slutsky :

$$\sqrt{n}(\bar{U}_n - \mu) \xRightarrow{\mathbb{P}_{n,\rho}} \mathcal{N}(0, V(\rho))$$

$$\sqrt{n}([S^2_{n,X}{}^2, S^2_{n,Y}, S^2_{n,XY}] + [\overline{X_n}^2, \overline{Y_n}^2, \overline{X}_n\overline{Y}_n] - \mu) \xRightarrow{\mathbb{P}_{n,\rho}} \mathcal{N}(0, V(\rho))$$

$$\sqrt{n}([S^2_{n,X}{}^2, S^2_{n,Y}, S^2_{n,XY}] - \mu) \xRightarrow{\mathbb{P}_{n,\rho}} \mathcal{N}(0, V(\rho))$$

$\square$

Exercice 6

On définit ϕ: $x, y, z \mapsto \frac{z}{\sqrt{xy}}$ sur $\mathbb{R}^3_{>0}$, une fonction $\mathcal{C}^1$. On calcule sa différentielle en $x, y, z \in \mathbb{R}^3_{>0}$:

$$d\phi(x,y,z) : v \in \mathbb{R}^3 \mapsto \begin{bmatrix}-\frac{yz}{2(xy)^{3/2}} & -\frac{xz}{2(xy)^{3/2}} & \frac{1}{(xy)^{1/2}}\end{bmatrix} \cdot v \in \mathbb{R}$$

On voit que:

$$\phi(S^2_{n,X}, S^2_{n,Y}, S_{n,XY}) = R_n$$
$$\phi(\mu) = \rho$$
$$d\phi(\mu) : v \mapsto \begin{bmatrix}-\frac{\rho}{2} & -\frac{\rho}{2} & 1\end{bmatrix} \cdot v$$

Alors, comme $\sqrt{n}\left([S^2_{n,X}, S^2_{n,Y}, S^2_{n,XY}] - \mu\right) \xRightarrow{\mathbb{P}_{n,\rho}} \mathcal{N}(0, V(\rho))$, par la methode Delta:

$$\sqrt{n}\left(\phi(S^2_{n,X}, S^2_{n,Y}, S^2_{n,XY}) - \phi(\mu)\right) = \sqrt{n}\left(R_n - \rho\right) \xRightarrow{\mathbb{P}_{n,\rho}} d\phi(\mu)\left(\mathcal{N}(0, V(\rho))\right)$$

Et on voit que:

$$d\phi(\mu)\left(\mathcal{N}(0,V(\rho))\right)=\mathcal{N}\left(0,\begin{bmatrix}-\frac{\rho}{2} & -\frac{\rho}{2} & 1\end{bmatrix}V(\rho)\begin{bmatrix}-\frac{\rho}{2} & -\frac{\rho}{2} & 1\end{bmatrix}^T\right)$$

Et que:

$$\begin{bmatrix}-\frac{\rho}{2} & -\frac{\rho}{2} & 1\end{bmatrix}V(\rho)\begin{bmatrix}-\frac{\rho}{2} & -\frac{\rho}{2} & 1\end{bmatrix}^T=\begin{bmatrix}-\frac{\rho}{2} & -\frac{\rho}{2} & 1\end{bmatrix}\begin{pmatrix}2 & 2\rho^2 & 2\rho\\ 2\rho^2 & 2 & 2\rho\\ 2\rho & 2\rho & 1+\rho^2\end{pmatrix}\begin{bmatrix}-\frac{\rho}{2} & -\frac{\rho}{2} & 1\end{bmatrix}^T=(1-\rho^2)^2$$

Donc:

$$\sqrt{n}\left(R_n-\rho\right)\xRightarrow{\mathbb{P}_{n,\rho}}\mathcal{N}(0,(1-\rho^2)^2)$$

Exercice 7

On a que :

$$\frac{1}{1-\rho^2}=\frac{1}{(1+\rho)(1-\rho)}=\frac{A}{1+\rho}+\frac{B}{1-\rho}=\frac{(B+A)+\rho(B-A)}{(1+\rho)(1-\rho)}$$

Cela implique que $A+B=1$ est $A=B$, donc $A=B=1/2$. Nous avons donc que :

$$g(\rho)=\int\frac{1}{1-\rho^2}d\rho=\int\left[\frac{1/2}{1-\rho}+\frac{1/2}{1+\rho}\right]d\rho=\frac{1}{2}\left[\int\frac{d\rho}{1-\rho}+\int\frac{d\rho}{1+\rho}\right]$$

Faisons les changements de variables $u=1-\rho$ et $v=1+\rho$:

$$g(\rho)=\frac{1}{2}\left[-\int\frac{du}{u}+\int\frac{dv}{v}\right]=\frac{1}{2}(-\log(1-\rho)+\log(1+\rho))=\frac{1}{2}\log\frac{1+\rho}{1-\rho}+C$$

Où est une constante que nous pouvons définir comme étant $C=0$.

□

Exercice 8

On veut appliquer de nouveau la méthode Delta pour la fonction $g\colon\rho\mapsto\frac{1}{2}\log\frac{1+\rho}{1-\rho}$ définit auparavant. On voit que g est continue et strictement croissante sur $]-1,1[$ et donc admet une inverse $g^{-1}:\rho\mapsto\tanh(\rho)$[1]. Aussi, par la définition de g, on voit que $g'\colon\rho\mapsto\frac{1}{1-\rho^2}$. Alors, comme $\sqrt{n}\left(R_n-\rho\right)\xRightarrow{\mathbb{P}_{n,\rho}}\mathcal{N}(0,(1-\rho^2)^2)$, par la méthode Delta:

[1]On voit que $x=\frac{1}{2}\log\frac{1+\rho}{1-\rho}$ implique que $e^{2x}=\frac{1+\rho}{1-\rho}$. Alors $\rho=\frac{e^{2x}-1}{e^{2x}+1}=\tanh(x)$

$$\sqrt{n}\left(\frac{1}{2}\log\frac{1+R_n}{1-R_n}-\frac{1}{2}\log\frac{1+\rho}{1-\rho}\right)=\sqrt{n}\left(g(R_n)-g(\rho)\right)\xrightarrow{\mathbb{P}_{n,\rho}} Z\sim\frac{1}{1-\rho^2}\mathcal{N}(0,(1-\rho^2)^2)=\mathcal{N}(0,1)$$

Exercice 9

On calcule la probabilité asymptotique:

$$\mathbb{P}_{n,\rho}\left(-z_{1-\alpha/2}\leq\sqrt{n}\left(W_n-g(\rho)\right)\leq z_{1-\alpha/2}\right)\xrightarrow{n\to\infty}1-\alpha$$

Ou z_β dénomme le quantile β d'une gaussiene centrée réduite. Alors, comme g est inversible:

$$\mathbb{P}_{n,\rho}\left(g^{-1}(W_n-z_{1-\alpha/2}/\sqrt{n})\leq\rho\leq g^{-1}\left(W_n+z_{1-\alpha/2}/\sqrt{n}\right)\right)\xrightarrow{n\to\infty}1-\alpha$$

On en déduit l'intervalle de confiance asymptotique a probabilité de couverture $1-\alpha$ pour ρ:

$$\mathcal{I}^n_{1-\alpha}(\rho)=\left[\frac{\exp\{2(W_n-z_{1-\alpha/2}/\sqrt{n})\}-1}{\exp\{2(W_n-z_{1-\alpha/2}/\sqrt{n})\}+1},\frac{\exp\{2(W_n+z_{1-\alpha/2}/\sqrt{n})\}-1}{\exp\{2(W_n+z_{1-\alpha/2}/\sqrt{n})\}+1}\right]$$

Exercice 10

On veut appliquer un developpement limite de $g^{-1}=\tanh$ au tour de W_n. En notant que $(g^{-1})'=\tanh'=\mathrm{sech}^2$, alors $\forall h\in\mathbb{R}$:

$$g^{-1}(W_n+h)=\underbrace{g^{-1}(W_n)}_{R_n}+\mathrm{sech}^2(W_n)h+o(|h|^2)$$

Comme $\mathrm{sech}^2(W_n)=1-R_n^2$ [2], alors on a que:

$$g^{-1}(W_n\pm z_{1-\alpha/2}/\sqrt{n})=R_n\pm\frac{z_{1-\alpha/2}}{\sqrt{n}}(1-R_n^2)+o\left(\frac{1}{n}\right)$$

Dans le comportement asymptotique on peut négliger la contribution du terme $o\left(\frac{1}{n}\right)$, ce qu'entraîne:

$$\begin{aligned}\mathcal{I}^n_{1-\alpha}(\rho)&=\left[g^{-1}(W_n-z_{1-\alpha/2}/\sqrt{n}),g^{-1}(W_n+z_{1-\alpha/2}/\sqrt{n})\right]\\&\overset{n\to\infty}{\sim}\left[R_n-\frac{z_{1-\alpha/2}(1-R_n^2)}{\sqrt{n}},R_n+\frac{z_{1-\alpha/2}(1-R_n^2)}{\sqrt{n}}\right]\end{aligned}$$

Qui est aussi un intervalle de confiance asymptotique de probabilité de couverture $1-\alpha$.

[2] On calcule $\mathrm{sech}^2(W_n)=4/(e^{W_n}+e^{-W_n})^2=4/(1+e^{2W_n}+e^{-2W_n})=4/(1+\frac{1+R_n}{1-R_n}+\frac{1-R_n}{1+R_n})=1-R_n^2$

Pratique

In [ ]:

```
import numpy as np

import matplotlib.pyplot as plt

import scipy.stats as stt
from scipy.stats import shapiro
from scipy.stats import probplot
```

In [ ]:

```
N = 1000
ns = [100, 200, 300]
rho = 1/3
alpha = .05
```

Dans les plots qui suivent :

Quantité 1 - $\sqrt{n}(R_n - \rho)$

Quantité 2 - $\sqrt{n}\left(\frac{1}{2}log(\frac{1+R_n}{1-R_n}) - \frac{1}{2}log(\frac{1+\rho}{1-\rho})]\right)$

Les intervalles de confiance sont montrés pour une instance.

In [ ]:

```
def run_n(n, do_plot=True):
    sigma = np.matrix([[1, rho], [rho, 1]])
    samples = np.random.multivariate_normal([0,0], cov=sigma, size=(N, n))

    X = samples[:,:,0]
    Y = samples[:,:,1]

    X_bar = np.mean(X, axis=1).reshape(-1, 1)
    Y_bar = np.mean(Y, axis=1).reshape(-1, 1)

    SnX2 = np.sum((X - X_bar) ** 2, axis=1) / n
    SnY2 = np.sum((Y - Y_bar) ** 2, axis=1) / n

    SnXY2 = np.sum((X - X_bar) * (Y - Y_bar), axis=1) / n

    R = SnXY2 / np.sqrt(SnX2 * SnY2)

    devR = np.sqrt(n) * (R - rho)
    devW = np.sqrt(n) * (np.log((1 + R) / (1 - R)) / 2 - np.log((1 + rho) / (1 - rho))
/ 2)

    # Plots de devR
    if do_plot:
      plt.figure()
      ppR = probplot(devR, dist='norm', plot=plt)
      plt.grid()
      plt.title(f"Plot q-q ({n=}) de la Quantite 1")

      plt.figure()
      plt.hist(devR, density=True)
      plt.title(f"Histogramme ({n=}) de la Quantite 1")

      plt.figure()
      plt.boxplot(devR)
      plt.title(f"Boxplot ({n=}) de la Quantite 1")

      # Plots de devW
      plt.figure()
      ppR = probplot(devW, dist='norm', plot=plt)
      plt.grid()
      plt.title(f"Plot q-q ({n=}) de la Quantite 2")

      plt.figure()
      plt.hist(devW, density=True)
      plt.title(f"Histogramme ({n=}) de la Quantite 2")

      plt.figure()
      plt.boxplot(devW)
      plt.title(f"Boxplot ({n=}) de la Quantite 2");

    # Test Shapiro-Wilks
    _, pR = shapiro(devR)
    _, pW = shapiro(devW)

    if do_plot:
      print(f"Pour {n=}:")
      print(f"  Quantité 1 - p-valeur={pR:.2e}")
      print(f"  Quantité 2 - p-valeur={pW:.2e}")
```

```
    # Intervalles de confiance
    z = stt.norm.ppf(1 - alpha / 2)

    R0 = R[0]
    IR = [R0 - z * (1 - R0 ** 2) / np.sqrt(n), R0 + z * (1 - R0 ** 2) / np.sqrt(n)]

    W0 = np.log((1 + R0) / (1 - R0)) / 2
    IW = [(np.exp(2 * (W0 - z / np.sqrt(n))) - 1) / (np.exp(2 * (W0 - z / np.sqrt(n)))
+ 1),
          (np.exp(2 * (W0 + z / np.sqrt(n))) - 1) / (np.exp(2 * (W0 + z / np.sqrt(n)))
+ 1)]

    if do_plot:
      print(f"  Intervalle (1) : {IR}")
      print(f"  Intervalle (2) : {IW}")

    return pR, pW
```

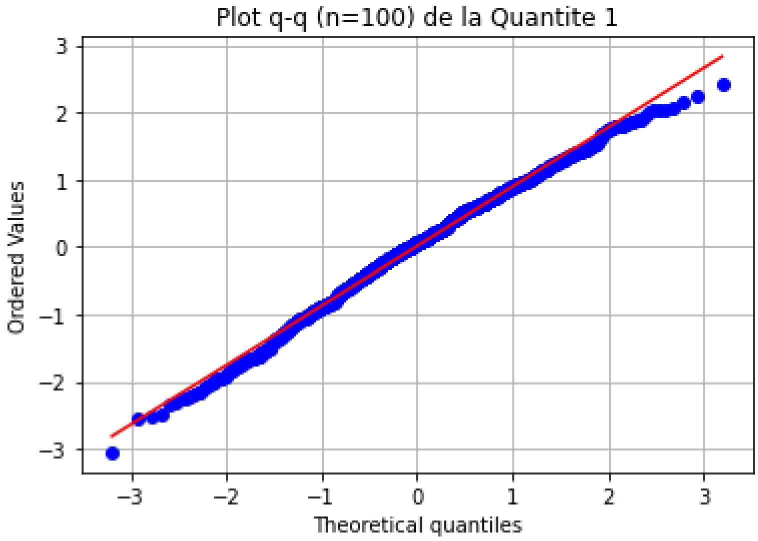
In [ ]:

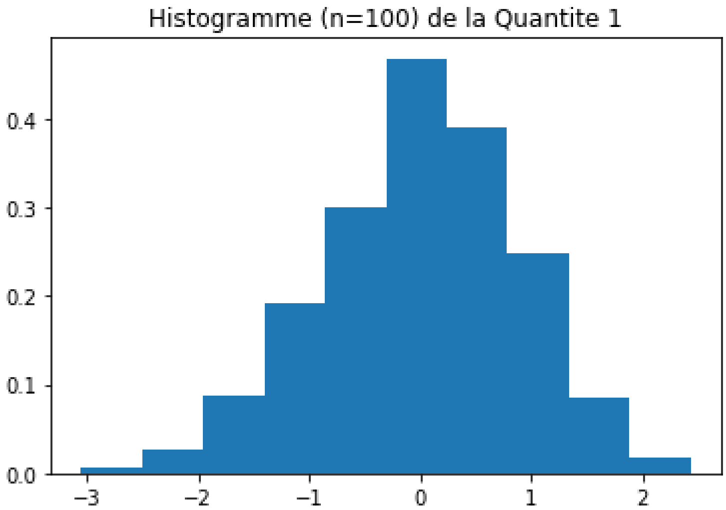
```
# n = 100
run_n(n=100)
```

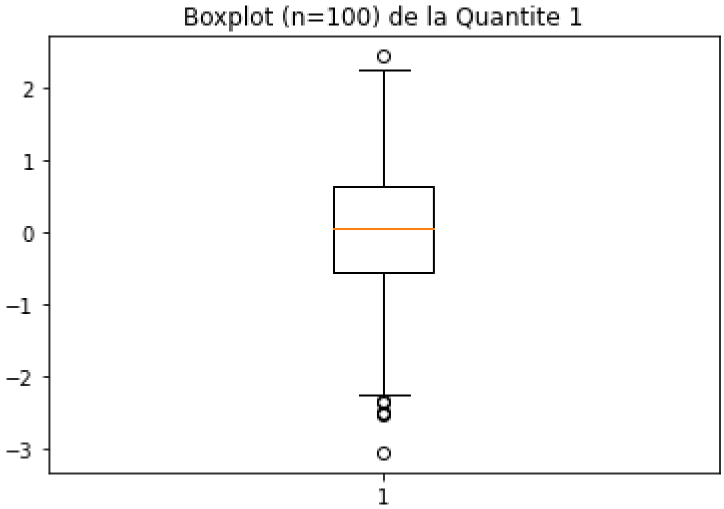
```
Pour n=100:
  Quantité 1 - p-valeur=1.42e-03
  Quantité 2 - p-valeur=5.23e-01
  Intervalle (1) : [0.025026123750093526, 0.39936839560671455]
  Intervalle (2) : [0.019472229864091573, 0.38971799762722314]
```

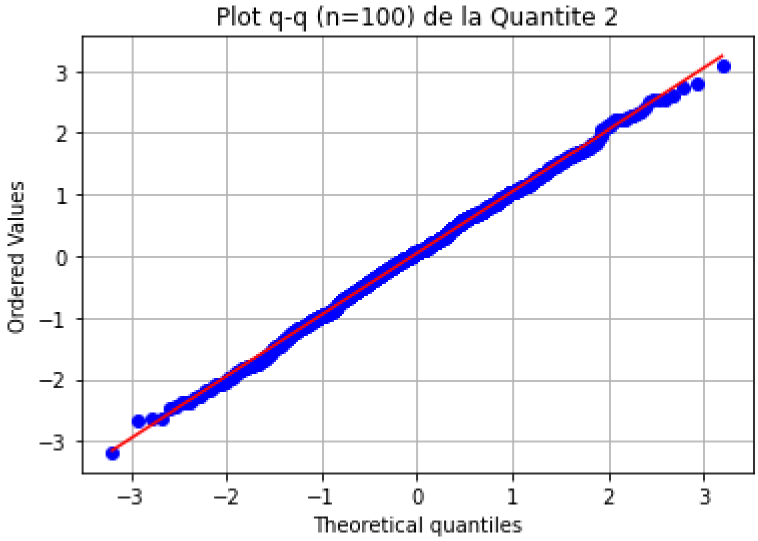
Out[ ]:

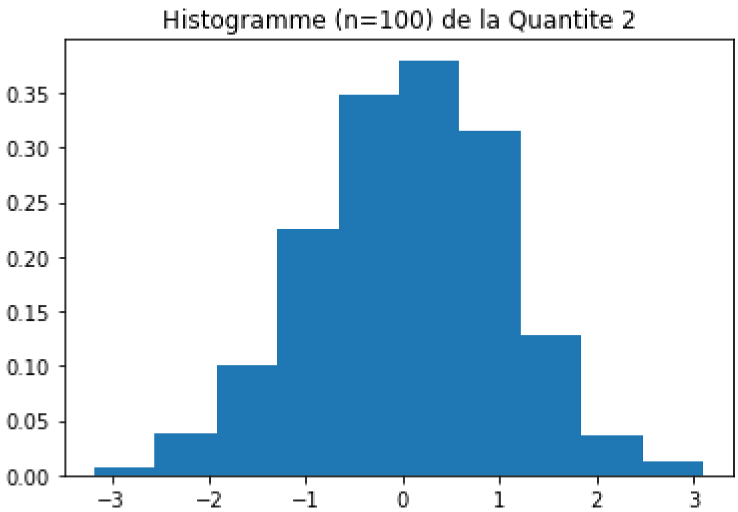
```
(0.0014160366263240576, 0.5233244895935059)
```

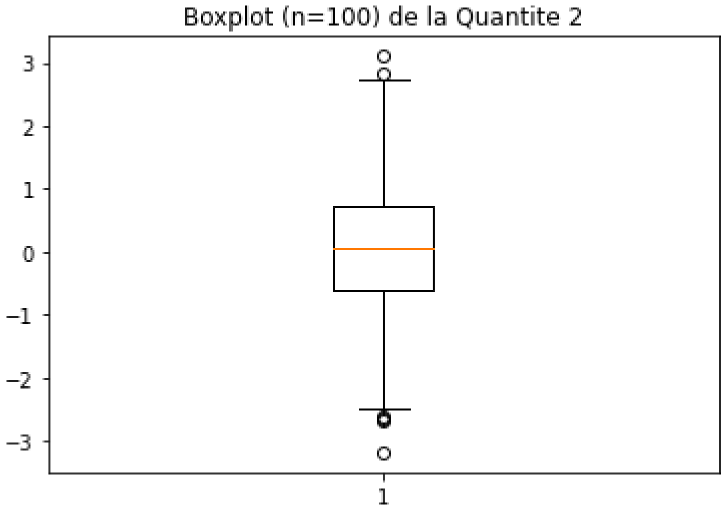





Boxplot (n=100) de la Quantite 1
2
1
0
−1
−2
−3
1


Plot q-q (n=100) de la Quantite 2
Ordered Values
3
2
1
0
−1
−2
−3
Theoretical quantiles


Histogramme (n=100) de la Quantite 2
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−3
−2
−1
0
1
2
3


Boxplot (n=100) de la Quantite 2
3
2
1
0
−1
−2
−3
1

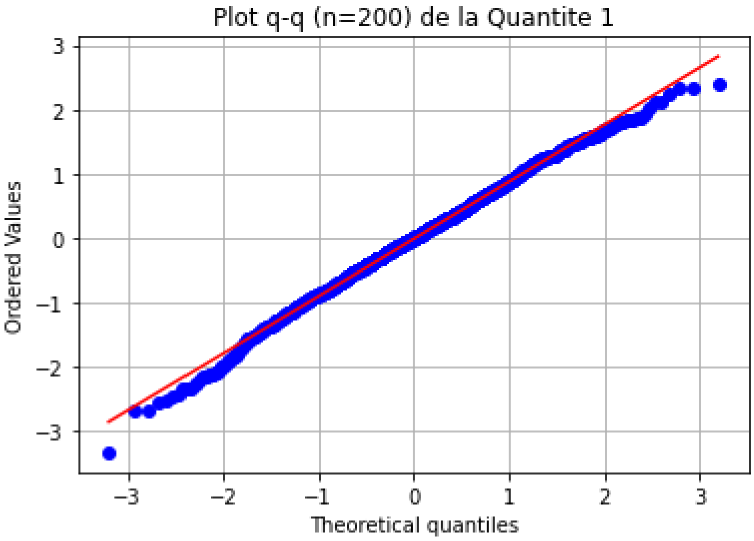
In [ ]:

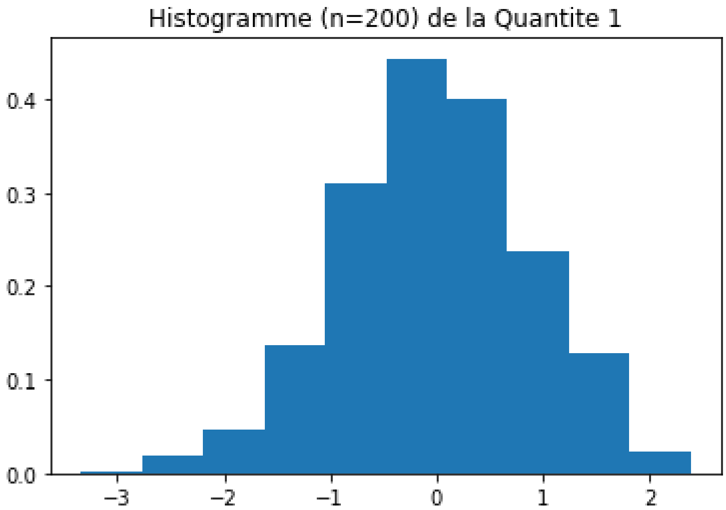
```
# n = 200
run_n(n=200)
```

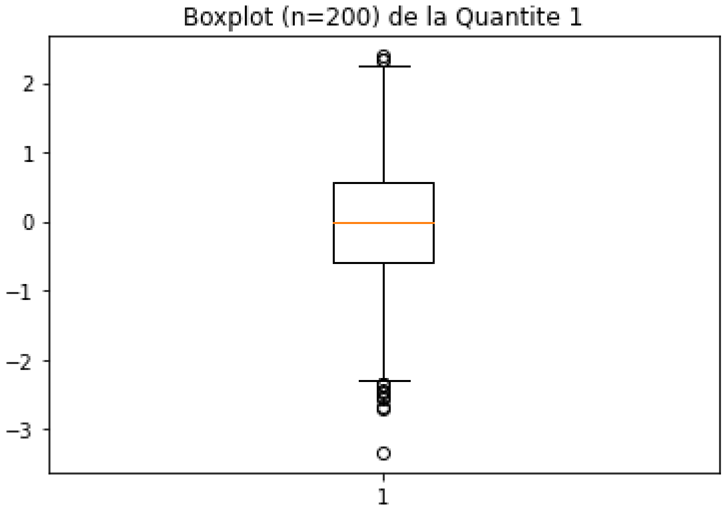
```
Pour n=200:
  Quantité 1 - p-valeur=2.87e-02
  Quantité 2 - p-valeur=4.93e-01
  Intervalle (1) : [0.20160373403971046, 0.44941540113123335]
  Intervalle (2) : [0.1966130882713951, 0.4433460287433841]
```

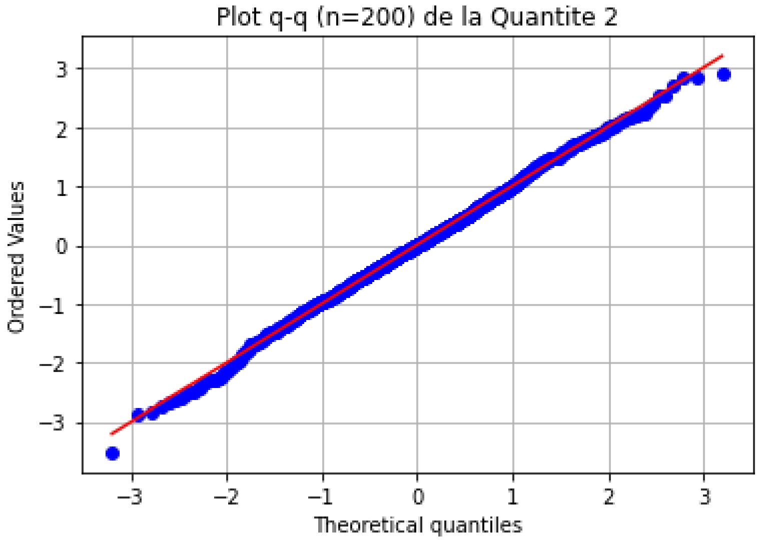
Out[ ]:

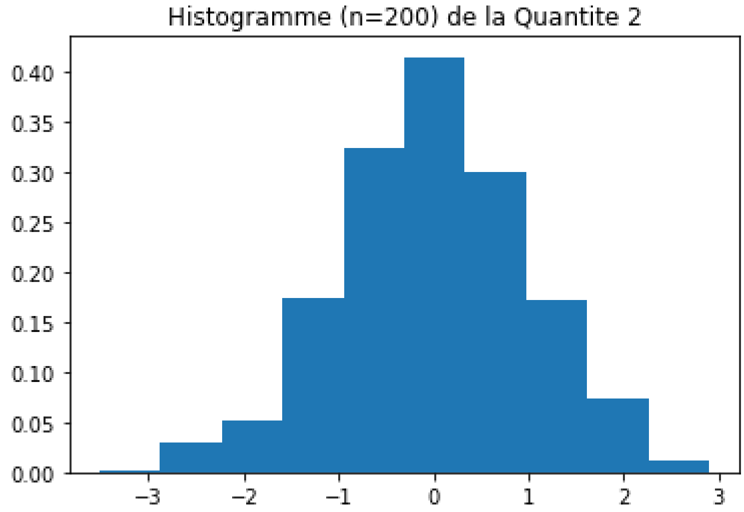
```
(0.028659336268901825, 0.49285581707954407)
```

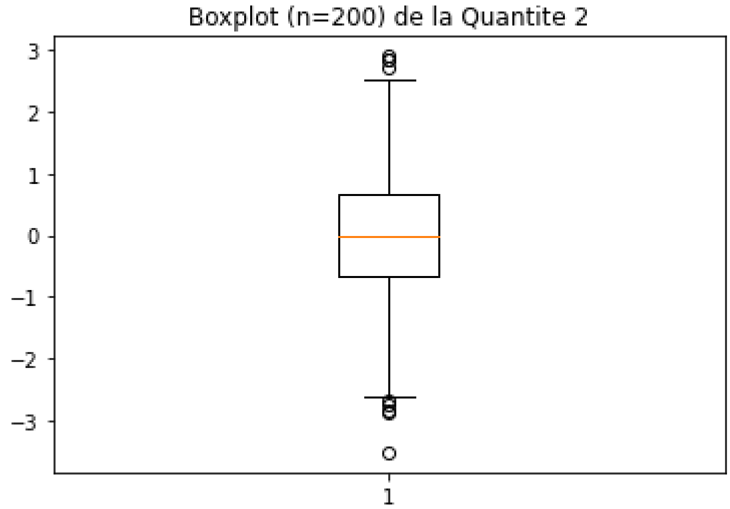





Boxplot (n=200) de la Quantite 1
2
1
0
−1
−2
−3
1


Plot q-q (n=200) de la Quantite 2
Ordered Values
Theoretical quantiles
3
2
1
0
−1
−2
−3


Histogramme (n=200) de la Quantite 2
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−3
−2
−1
0
1
2
3


Boxplot (n=200) de la Quantite 2
3
2
1
0
−1
−2
−3
1

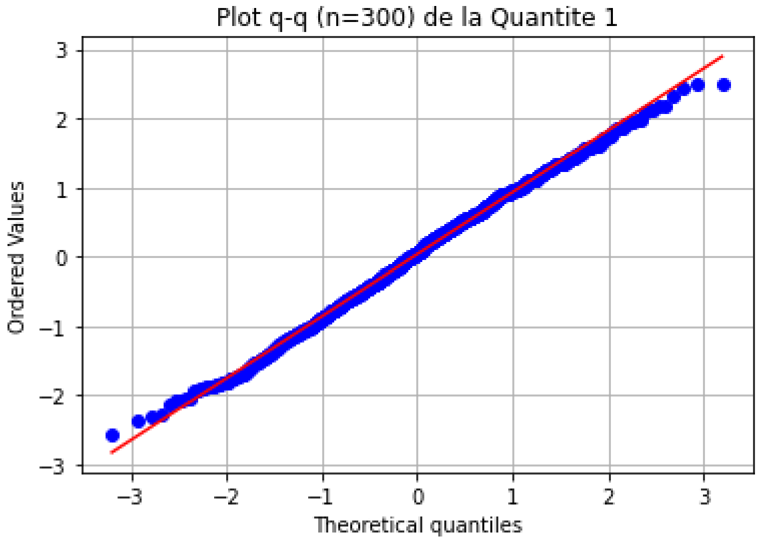
In [ ]:

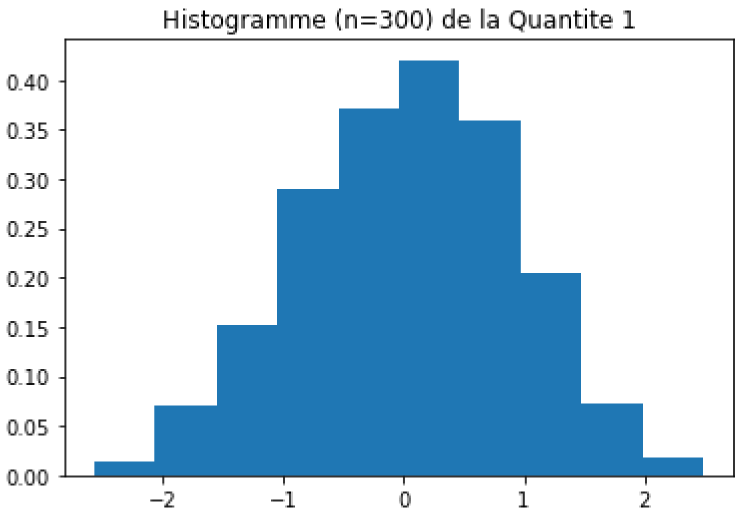
```
# n = 300
run_n(n=300)
```

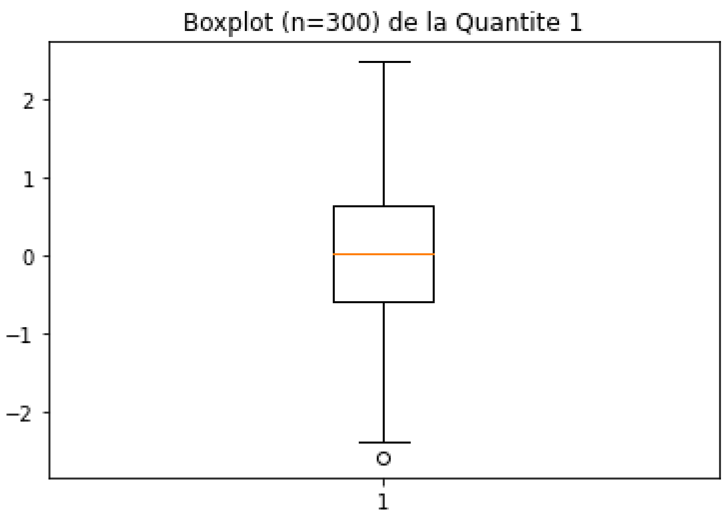
```
Pour n=300:
  Quantité 1 - p-valeur=6.18e-02
  Quantité 2 - p-valeur=1.78e-01
  Intervalle (1) : [0.35553421489559534, 0.5367995631296639]
  Intervalle (2) : [0.351141877563713, 0.5320948152280667]
```

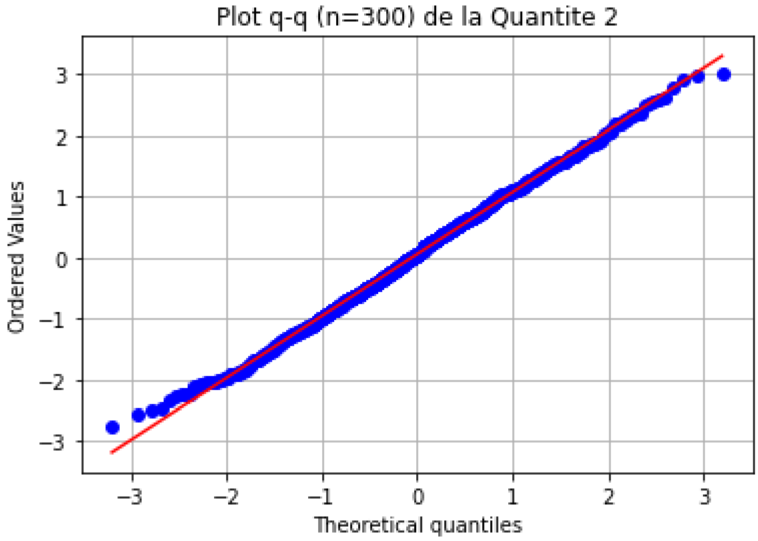
Out[ ]:

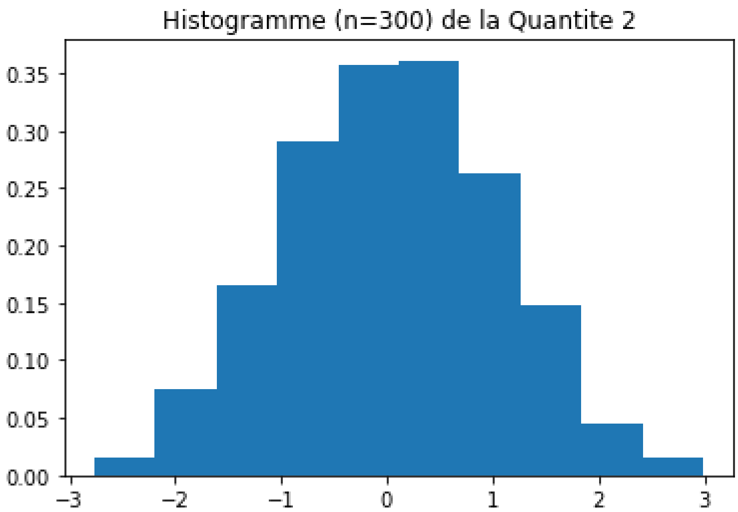
```
(0.061803679913282394, 0.17752140760421753)
```

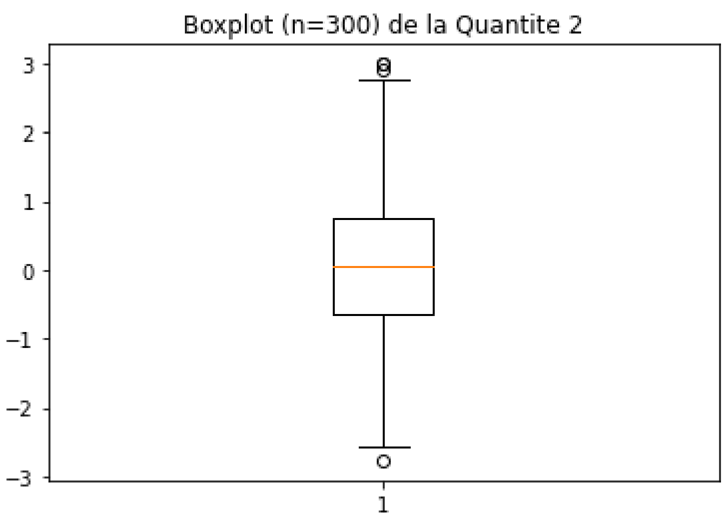







Plot q-q (n=300) de la Quantite 2
Ordered Values
Theoretical quantiles
3
2
1
0
−1
−2
−3


Histogramme (n=300) de la Quantite 2
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−3
−2
−1
0
1
2
3


Boxplot (n=300) de la Quantite 2
3
2
1
0
−1
−2
−3
1

In [ ]:

```
n = 300
N_ps = 1000

pR = np.zeros(N_ps)
pW = np.zeros(N_ps)
for i in range(N_ps):
    pR[i], pW[i] = run_n(n, do_plot=False)
```

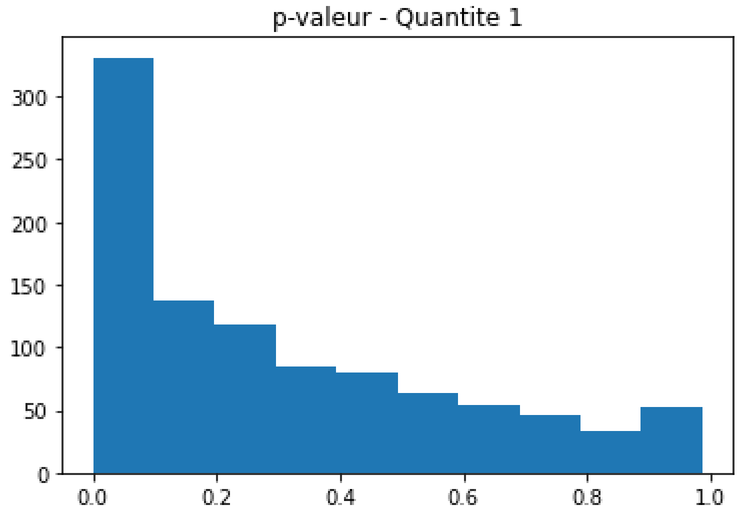
In [ ]:

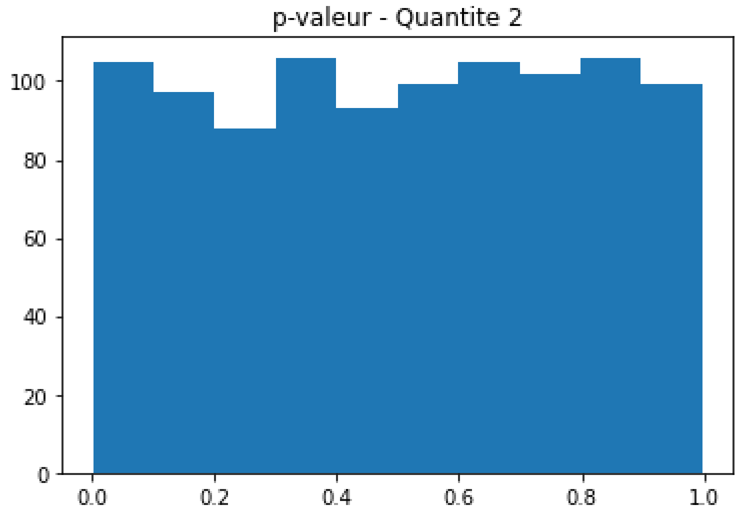
```
plt.figure()
plt.hist(pR, bins=10)
plt.title("p-valeur - Quantite 1")

plt.figure()
plt.hist(pW, bins=10)
plt.title("p-valeur - Quantite 2")
```

Out[ ]:

```
Text(0.5, 1.0, 'p-valeur - Quantite 2')
```





Conclusions

La forme du graphe des p-valeurs des tests de Shapiro-Wilks pour la quantité 1 nous montre que l'hypothèse nulle peut être rejetée car plusieurs tests obtiennent des p-valeurs petites. En revanche les tests pour la quantité 2 ne nous permettent pas de rejeter l’hypothèse nulle car les p-valeurs sont distribués presque uniformément entre zéro et un. La même approche est utilisée dans le polycopié (Figure I-3.4). Alors, construire un intervalle de confiance en utilisant la quantité 1 entraîne une approximation plus grossière qu’entrainerait la construction à partir de la quantité 2.

Notons aussi que l’intervalle donné par la deuxième formule est proche de l’intervalle de la première formule et l’approche de plus en plus quand n augmente. Ce comportement est attendu car la deuxième formule est asymptotiquement égale à la première.