Exploration Numérique 2

25 septembre 2022

Soit $\{(X_i, Y_i)\}_{i=1}^n$ une suite de n-échantillons d'une loi gaussienne bivariée de moyenne nulle et de matrice de covariance

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

où $\rho \in (-1, 1)$. Le coefficient de corrélation empirique R_n est donné par

$$R_n = \frac{S_{n,XY}^2}{S_{n,X} S_{n,Y}} \text{ où}$$

$$S_{n,XY}^2 = \frac{1}{n} \sum_{i=1}^n \left(X_i - \bar{X}_n\right) \left(Y_i - \bar{Y}_n\right) \quad S_{n,X}^2 = \frac{1}{n} \sum_{i=1}^n \left(X_i - \bar{X}_n\right)^2 \quad S_{n,Y}^2 = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \bar{Y}_n\right)^2$$

1. Montrer que $\mathbb{E}_\rho[X_1^4] = \mathbb{E}_\rho[Y_1^4] = 3$ et $\mathrm{Var}_\rho(X^2) = \mathrm{Var}_\rho(Y^2) = 2$.
2. Montrer que, sous $\mathbb{P}_\rho$, $Y = \rho X + \sqrt{1-\rho^2} Z$ où Z et X sont indépendants. En déduire
$$\mathbb{E}_\rho[X_1^2 Y_1^2] = 1 + 2\rho^2$$
3. Montrer que
$$\mathrm{Cov}_\rho(X_1^2, X_1 Y_1) = \mathrm{Cov}_\rho(Y_1^2, X_1 Y_1) = 2\rho.$$

Pour $i \in \mathbb{N}$, nous notons $U_i = [X_i^2, Y_i^2, X_i Y_i]^T$, $\mu(\rho) = \mathbb{E}_\rho[U_1]$.

4. Montrer que pour tout $\rho \in]-1, 1[$, $\mu(\rho) = [1, 1, \rho]^T$ et
$$\sqrt{n}\big(\bar{U}_n - \mu\big) \overset{\mathbb{P}_{n,\varrho}}{\Longrightarrow} \mathrm{N}\left(0, V(\rho)\right)$$
où
$$V = \begin{pmatrix} 2 & 2\rho^2 & 2\rho \\ 2\rho^2 & 2 & 2\rho \\ 2\rho & 2\rho & 1+\rho^2 \end{pmatrix}$$
5. En déduire que
$$\sqrt{n}\big([S_{n,X}^2, S_{n,Y}^2, S_{n,XY}^2 - \mu\big) \overset{\mathbb{P}_{n,\varrho}}{\Longrightarrow} \mathrm{N}\left(0, V(\rho)\right)$$
6. Montrer [en utilisant la forme multivariée de la Delta-méthode] que, pour tout $\rho \in]-1, 1[$,
$$\sqrt{n}(R_n - \rho) \overset{\mathbb{P}_{n,\varrho}}{\Longrightarrow} \mathrm{N}\left(0, (1-\rho^2)^2\right)$$

Ppur construire un intervalle de confiance asymptotique, nous utilisons une technique de stabilisation de la variance

7. Montrer que

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

8. En déduire que, pour tout $\rho \in \left]-1, 1\right[$,

$$\sqrt{n}\left[\frac{1}{2}\log\frac{1+R_n}{1-R_n} - \frac{1}{2}\log\frac{1+\rho}{1-\rho}\right] \overset{\mathbb{P}_{n,\rho}}{\Longrightarrow} Z \sim N(0,1).$$

9. Soit $\alpha \in \left]0, 1\right[$. Posons $W_n = g(R_n)$. Montrer que

$$\left(\frac{\exp\left\{2\left(W_n - z_{(1-\alpha/2)}/\sqrt{n}\right)\right\} - 1}{\exp\left\{2\left(W_n - z_{(1-\alpha/2)}/\sqrt{n}\right)\right\} + 1}, \frac{\exp\left\{2\left(W_n + z_{(1-\alpha/2)}/\sqrt{n}\right)\right\} - 1}{\exp\left\{2\left(W_n + z_{(1-\alpha/2)}/\sqrt{n}\right)\right\} + 1}\right) \tag{1}$$

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

On peut bien entendu construire un intervalle de confiance asymptotique par Studentization.

10. Montrer que

$$\left(R_n - \frac{z_{(1-\alpha/2)}\left(1-R_n^2\right)}{\sqrt{n}}, \quad R_n + \frac{z_{(1-\alpha/2)}\left(1-R_n^2\right)}{\sqrt{n}}\right) \tag{2}$$

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

En statistique, le test de Shapiro–Wilk teste l'hypothèse nulle selon laquelle un échantillon est issu d'une population normalement distribuée ; voir [Shapiro and Wilk, 1972]/ Pour le test, vous utiliserez `scipy.stats.shapiro()` Vous utiliserez aussi `scipy.stats.probplot()` pour les $q-q$-plots.

Nous considérons des tailles d'échantillons $n = 100$, $n = 200$, $n = 300$. Nous répétons pour chaque taille d'échantillons les expériences $N = 1000$ pour $\rho = 1/3$.

1. Visualisez les histogrammes, les boxplots, les quantile-quantile plot de

$$\sqrt{n}(R_n - \rho) \quad \text{et} \quad \sqrt{n}\left[\frac{1}{2}\log\frac{1+R_n}{1-R_n} - \frac{1}{2}\log\frac{1+\rho}{1-\rho}\right]$$

   et calculer les p-valeurs du test de Shapiro-Wilks. Que conclure ?

2. Calculer les intervalles de confiance asymptotique par les formules (1) et (2). Qu'observe-t-on ?

Références

[Shapiro and Wilk, 1972] Shapiro, S. S. and Wilk, M. (1972). An analysis of variance test for the exponential distribution (complete samples). *Technometrics*, 14(2) :355–370.