

Feuille de préparation pour l'interro 1 du 10 décembre

Exercice 1:

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées selon la loi normale centrée réduite. Pour tout $n \in \mathbb{N}^*$, on définit $S_n = \sum_{i=1}^n X_i$.

1. Pour $n \in \mathbb{N}^*$, déterminer la loi du vecteur $Y_n = \begin{pmatrix} S_1 \\ \vdots \\ S_n \end{pmatrix}$.
2. Soit $n \in \mathbb{N}^*$. Pour tout $j \in \{1, \dots, n\}$, déterminer la loi du vecteur $Z_j = \begin{pmatrix} S_j \\ S_n \end{pmatrix}$
 - (a) à partir de la loi des X_1, \dots, X_n .
 - (b) à partir de la loi de Y_n .
3. Soit $n \geq 2$. Pour tout $j \in \{1, \dots, n-1\}$, on considère la variable aléatoire $A_j = S_j - \frac{j}{n}S_n$. Les variables A_j et S_n sont-elles indépendantes?

Exercice 2:

Soient X_1, \dots, X_n des variables aléatoires représentant les relevés au cours du temps d'un sondeur écho radar définies par:

$$\begin{cases} X_1 = b_1 + U_1 \\ X_j = b_j + \theta U_{j-1} + U_j, \forall j \in \{2, \dots, n\} \end{cases}$$

où b_1, \dots, b_n, θ sont des réels positifs et U_1, \dots, U_n sont des variables aléatoires indépendantes et identiquement distribuées selon la loi $\mathcal{N}(0, \sigma^2)$.

1. Montrer que $X = (X_1, \dots, X_n)'$ est un vecteur gaussien.
2. Déterminer la loi de X .

Exercice 3:

Soit un couple (X, Y) de densité $f(x, y) = \frac{1}{4\pi} \exp \left[-\frac{1}{2}T(x, y) \right]$ avec

$$T(x, y) = \frac{1}{2}x^2 + y^2 - xy + 4x - 7y + \frac{25}{2}.$$

On considère la variable aléatoire $A = 2Y - X$.

1. Les variables aléatoires A et X sont-elles indépendantes?
2. Quelle est la loi de A ?

Exercice 4

Afin de les sélectionner, un directeur d'agence fait passer un test d'aptitude à 27 candidats dont 15 hommes et 12 femmes. Il note le temps (en minutes) nécessaire à chacun des candidats pour répondre au test. On note X_1, \dots, X_{12} les variables donnant le temps des 12 candidates, et Y_1, \dots, Y_{15} les variables donnant le temps des 15 candidats. Les 27 réponses sont modélisées par :

$$X_i \sim \mathcal{N}(m_F, \sigma_F^2) ; Y_i \sim \mathcal{N}(m_H, \sigma_H^2) ; \text{ les 27 variables sont indépendantes.}$$

Les résultats observés sont :

$$\sum_{i=1}^{12} x_i = 99, 1 ; \sum_{i=1}^{12} x_i^2 = 836, 15 ; \sum_{i=1}^{15} y_i = 134, 3 ; \sum_{i=1}^{15} y_i^2 = 1217, 7.$$

1. Construire un test d'égalité des variances des réponses des hommes et des femmes au niveau 5% et donner la conclusion de ce test, ainsi que son niveau observé (p-value).
2. On suppose maintenant que les variances sont égales : $\sigma_F^2 = \sigma_H^2 = \sigma^2$.

- (a) On note $f(u_1, \dots, u_{12}, v_1, \dots, v_{15})$ la densité de probabilité de $(X_1, \dots, X_{12}, Y_1, \dots, Y_{15})$ au point $(u_1, \dots, u_{12}, v_1, \dots, v_{15})$. Cette densité dépend des paramètres m_F, m_H, σ^2 . On appelle vraisemblance de $(X_1, \dots, X_{12}, Y_1, \dots, Y_{15})$ la variable définie par :

$$V(X_1, \dots, X_{12}, Y_1, \dots, Y_{15}, m_F, m_H, \sigma^2) = f(X_1, \dots, X_{12}, Y_1, \dots, Y_{15}).$$

Calculer l'estimateur du maximum de vraisemblance $(\hat{m}_F, \hat{m}_H, \hat{\sigma}^2)$ défini par :

$$V(X_1, \dots, X_{12}, Y_1, \dots, Y_{15}, \hat{m}_F, \hat{m}_H, \hat{\sigma}^2) = \max_{(m_F, m_H, \sigma^2)} V(X_1, \dots, X_{12}, Y_1, \dots, Y_{15}, m_F, m_H, \sigma^2)$$

Indication : plutôt que de maximiser $V(X_1, \dots, X_{12}, Y_1, \dots, Y_{15}, m_F, m_H, \sigma^2)$, maximiser la log-vraisemblance définie par $\ln [V(X_1, \dots, X_{12}, Y_1, \dots, Y_{15}, m_F, m_H, \sigma^2)]$.

- (b) Rappeler la loi de $\hat{\sigma}^2$ et en déduire un estimateur sans biais de σ^2 , noté S^2 .

Montrer que $\mathbb{E}[(\hat{\sigma}^2 - \sigma^2)^2] < \mathbb{E}[(S^2 - \sigma^2)^2]$.

Donner un intervalle de confiance pour σ de niveau de confiance 95%.

- (c) Si la performance des candidats des deux sexes lors du test n'est évaluée que par le temps nécessaire pour y répondre, peut-on affirmer qu'il y a une différence entre la performance moyenne des candidats et celle des candidates ?

On répondra à l'aide d'un test statistique et on calculera le niveau observé de ce test (p-value).

Exercice 5

Théorème de Fisher-Cochran, version décentrée : Soit $Y \sim \mathcal{N}_n(m, \sigma^2 I_d)$ un vecteur gaussien de \mathbb{R}^n et soient E_1, \dots, E_p des sous-espaces vectoriels orthogonaux de \mathbb{R}^n , de dimension respective r_1, \dots, r_p . On note P_{E_i} la matrice de projection orthogonale de \mathbb{R}^n sur E_i . Les vecteurs $P_{E_i} Y$ sont mutuellement indépendants de loi respective $\mathcal{N}_n(P_{E_i} m, \sigma^2 P_{E_i})$ et les variables $\left\| \frac{P_{E_i} Y}{\sigma} \right\|^2$ sont mutuellement indépendantes de loi respective $\chi^2 \left(r_i, \left\| \frac{P_{E_i} m}{\sigma} \right\|^2 \right)$.

Théorème de Fisher : Soit Y_1, \dots, Y_n un n -échantillon de loi $\mathcal{N}(m, \sigma^2)$. On note $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ et $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$. La loi du couple $(\bar{Y}, \hat{\sigma}^2)$ est donnée par : $\bar{Y} \sim \mathcal{N}(m, \frac{\sigma^2}{n})$, $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-1)$, les deux variables sont indépendantes.

1. Démontrer le théorème de Fisher-Cochran, version décentrée.
2. En utilisant le théorème de Fisher-Cochran, démontrer le théorème de Fisher.