

## Troisième Année Licence M.I.A.S.H.S. 2025 – 2026

## Statistique 2

Examen terminal, Mai 2026

*Examen de 2h00. Tout document ou calculatrice est interdit.***Exercice 1 (Sur 17 points)**

Soit  $(\varepsilon_i)_{i \in \mathbf{N}}$  une suite de variables aléatoires indépendantes identiquement distribuées, centrées, telles que  $\mathbb{E}[\varepsilon_0^2] < \infty$  et  $0 < \sigma_\varepsilon^2 = \text{var}(\varepsilon_0)$ . Pour tout  $n \in \mathbf{N}^*$ , on définit la variable aléatoire  $X_n$  telle que

$$X_n = \prod_{k=1}^n \varepsilon_k.$$

1. Pour tout  $n \in \mathbf{N}^*$ , démontrer que  $\mathbb{E}[|X_n|] < \infty$  puis que  $\mathbb{E}[X_n] = 0$  (**1pt**)?
2. Uniquement dans cette question, on suppose que  $\varepsilon_0$  suit la loi  $\mathcal{N}(0, \sigma_\varepsilon^2)$ . Démontrer que  $\mathbb{E}[\varepsilon_0^4] = 3\sigma_\varepsilon^4$  (**1pt**). Après avoir calculé  $\mathbb{E}[X_n^2]$  et  $\mathbb{E}[X_n^4]$ , déduire que pour tout  $n \geq 2$ ,  $X_n$  ne peut pas suivre une loi gaussienne (**2pts**).
3. Uniquement dans cette question, on suppose que  $\mathbb{E}[|\varepsilon_0|] < 1$ . Déterminer la limite en probabilité de  $(X_n)$  (**1pt**).
4. Démontrer que  $\text{cov}(X_n, X_{n+k}) = 0$  pour  $k \geq 1$  (**0.5pts**). En déduire la matrice de variance-covariance du vecteur  $(X_1, \dots, X_n)$  (**0.5pts**). Montrer pourtant que  $X_1$  et  $X_2$  ne sont pas nécessairement indépendantes (considérer par exemple le cas où  $\varepsilon_0$  suit une loi gaussienne) (**2pts**).
5. On voudrait déterminer une condition nécessaire sur la loi de  $\varepsilon_0$  pour que la suite  $(X_n)_{n \in \mathbf{N}^*}$  soit une suite de v.a. indépendantes. Démontrer que si cela était vrai, alors cela impliquerait que  $\text{cov}(|X_1|, |X_2|) = 0$  (**0.5pts**), que  $\text{var}(|\varepsilon_0|) = 0$  (**0.5pts**) et enfin qu'il existerait  $a > 0$  tel que  $\mathbb{P}(\varepsilon_0 = a) = \mathbb{P}(\varepsilon_0 = -a) = 1/2$  (**1pt**).
6. On voudrait déterminer une condition nécessaire sur la loi de  $\varepsilon_0$  pour que la suite  $(X_n)_{n \in \mathbf{N}^*}$  soit une suite de v.a. identiquement distribuées. Montrer que ceci ne serait possible que si  $\mathbb{E}[|\varepsilon_0|] = 1$  et  $\text{var}(|\varepsilon_0|) = 0$  (**1.5pts**). En déduire qu'alors  $\mathbb{P}(\varepsilon_0 = 1) = \mathbb{P}(\varepsilon_0 = -1) = 1/2$  (on dit que  $\varepsilon_0$  suit la loi de Rademacher) (**1pt**).
7. Si on suppose que  $\varepsilon_0$  suit la loi de Rademacher, déterminer la loi de  $X_n$  pour tout  $n \in \mathbf{N}^*$  (**1.5pts**). Montrer que pour tout  $n \geq 2$  et tout  $(x_1, \dots, x_n) \in \{-1, 1\}^n$ ,

$$\mathbb{P}(X_1 = x_1 \cap X_2 = x_2 \cap \dots \cap X_n = x_n) = \mathbb{P}(\varepsilon_1 = x_1 \cap \varepsilon_2 = x_1 x_2 \cap \dots \cap \varepsilon_n = x_{n-1} x_n) \quad (\mathbf{2pts}).$$

En déduire alors que  $(X_n)$  est une suite de v.a.i.i.d. (**1pt**).

**Exercice 2 (Sur 12 points)**

Dans ce qui suit, on note  $\mathbb{I}_{x \in A}$  la fonction de  $x$  qui vaut 1 si  $x \in A$  et 0 sinon.

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité,  $X$  comme une variable de Bernoulli de paramètre  $p \in [0, 1]$  définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$ ,  $U_1$  et  $U_2$  deux variables aléatoires continues définies sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , de lois uniformes sur  $[0, s]$  et  $[0, 2s]$  respectivement (avec  $s > 0$ ), les trois variables  $X$ ,  $U_1$  et  $U_2$  étant indépendantes. Enfin, on définit

$$Z = X U_1 + (1 - X) U_2.$$

1. Quel est  $Z(\Omega)$  (**0.5pts**) ? En déduire que  $\mathbb{E}[|Z|^k] < \infty$  pour tout  $k \in \mathbf{N}$  (**0.5pts**).
2. Calculer  $\mathbb{E}[Z]$  (**0.5pts**) et montrer que  $\text{var}(Z) = \frac{1}{12} s^2 (4 - 3p^2)$  (**1pt**).
3. Déterminer  $\mathbb{P}(Z \leq z)$  pour  $z \in \mathbf{R}$  (on pourra utiliser la formule des probabilités totales) (**1pt**) et en déduire que  $Z$  est une variable aléatoire continue (**0.5pts**) de densité  $f_Z$  telle que :

$$f_Z(z) = \frac{1}{2s} (p+1) \mathbb{I}_{0 \leq z \leq s} (1-p) \mathbb{I}_{s < z \leq 2s} \mathbb{I}_{0 \leq z \leq 2s} \quad \text{pour } z \in \mathbf{R} \quad (\mathbf{0.5pts}).$$

4. On suppose maintenant que  $(Z_1, \dots, Z_n)$  est un échantillon observé de variables aléatoires indépendantes et identiquement distribuées, de même loi que  $Z$ . On suppose que  $s$  est connu, mais que  $p$  est inconnu. Déterminer la log-vraisemblance de  $(Z_1, \dots, Z_n)$  en fonction de  $p$  (**1pt**). En déduire que l'estimateur du maximum de vraisemblance  $p_n^*$  de  $p$  sur  $[0, 1]$  est unique et vérifie

$$p_n^* = \max(0, \hat{p}_n) \quad \text{avec} \quad \hat{p}_n = \frac{1}{n} \sum_{i=1}^n (2 \mathbb{I}_{0 \leq Z_i \leq s} - 1) \quad (\mathbf{2pts}).$$

5. Montrer que  $\hat{p}_n$  est sans biais (**0.5pts**) et que pour tout  $0 \leq p < 1$ ,

$$\sqrt{n} (\hat{p}_n - p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1 - p^2) \quad (\mathbf{0.5pts}).$$

6. On suppose maintenant que  $s$  est inconnu. On pose  $\hat{s}_n = \frac{1}{2} \max_{1 \leq i \leq n} Z_i$ . Montrer que pour tout  $0 < \varepsilon \leq s/2$ ,

$$\mathbb{P}(\hat{s}_n \leq s - \varepsilon) = \left(1 - (1-p) \frac{\varepsilon}{s}\right)^n \quad (\mathbf{1pt}).$$

En déduire que  $\hat{s}_n \xrightarrow[n \rightarrow +\infty]{\mathcal{P}} s$  lorsque  $p \in [0, 1)$  (**0.5 pts**).

7. Montrer que pour  $p \in [0, 1)$ ,  $n(s - \hat{s}_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{E}\left(\frac{1-p}{s}\right)$ , loi exponentielle de paramètre  $\frac{1-p}{s}$  (**1pt**).

8. **Questions bonus** On pose  $\tilde{p}_n = \frac{1}{n} \sum_{i=1}^n (2 \mathbb{I}_{0 \leq Z_i \leq \hat{s}_n} - 1)$ .

(a) Montrer que  $\sqrt{n} (\hat{p}_n - \tilde{p}_n) = 2 \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{I}_{\hat{s}_n < Z_i \leq s}$  (**1pt**).

(b) Montrer que  $\mathbb{P}(\hat{s}_n < Z \leq s) \leq \mathbb{P}(\hat{s}_n < Z \leq s \cap \hat{s}_n \geq s - n^{-3/4}) + \mathbb{P}(\hat{s}_n \leq s - n^{-3/4})$  (**1pt**).

(c) Montrer que  $\sqrt{n} \mathbb{P}(s - n^{-3/4} < Z \leq s) \xrightarrow[n \rightarrow \infty]{} 0$  (**2pts**).

(d) En déduire que  $\mathbb{E}[|\sqrt{n} (\hat{p}_n - \tilde{p}_n)|] \xrightarrow[n \rightarrow \infty]{} 0$  (**1pt**).

(e) Enfin, en déduire que

$$\sqrt{n} (\tilde{p}_n - p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1 - p^2) \quad (\mathbf{1pt}). \quad (1)$$

9. En utilisant (1), donner un intervalle de confiance à 95% pour  $p$  lorsque  $p$  et  $s$  sont inconnus (**1pt**).