

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 13 points)

Soit  $(\varepsilon_i)_{i \in \mathbf{N}}$  une suite de variables aléatoires indépendantes identiquement distribuées et centrées telles que  $0 < \text{var}(\varepsilon_0) = \sigma_\varepsilon^2 < \infty$ . 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  $\mathbb{E}[\varepsilon_0^4] = 3\sigma_\varepsilon^4$ . 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 (**0.5pts**).
3. Uniquement dans cette question, on suppose que  $\mathbb{E}[|\varepsilon_0|] < 1$ . Déterminer la limite en probabilité de  $(X_n)$ .
4. Démontrer que  $\text{cov}(X_n, X_{n+k}) = 0$  pour  $k \geq 1$ . En déduire la matrice de variance-covariance du vecteur  $(X_1, \dots, X_n)$ . Montrer pourtant que  $X_1$  et  $X_2$  ne sont pas nécessairement indépendantes (considérer par exemple le cas gaussien).
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  $\text{cov}(|X_1|, |X_2|) = 0$ , que  $\text{var}(|\varepsilon_0|) = 0$  et enfin qu'il existerait  $a > 0$  tel que  $\mathbb{P}(\varepsilon_0 = a) = \mathbb{P}(\varepsilon_0 = -a) = 1/2$ .
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  $\text{var}(|\varepsilon_1|) = 0$ . 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).
7. Si on suppose que  $\varepsilon_0$  suit la loi de Rademacher, déterminer la loi de  $X_n$  pour tout  $n \in \mathbf{N}^*$ . Montrer que pour tout  $n \in \mathbf{N}^*$  et tout  $(x_1, \dots, x_n) \in \{-1, 1\}^n$ ,

$$\mathbb{P}\left(X_1 = x_1 \cap X_2 = x_2 \cap \dots \cap X_n = x_n\right) = \mathbb{P}\left(\varepsilon_1 = x_1 \cap \varepsilon_2 = x_1 x_2 \cap \dots \cap \varepsilon_n = x_{n-1} x_n\right) \quad \text{avec } x_0 = 1.$$

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

*Proof.* 1. Pour tout  $n \geq 1$ , par indépendance, et comme  $\mathbb{E}[|\varepsilon_0|] < \infty$  du fait que  $\text{var}(\varepsilon_0) < \infty$ ,

$$\mathbb{E}[|X_n|] = \mathbb{E}\left[\prod_{k=1}^n |\varepsilon_k|\right] = \prod_{k=1}^n \mathbb{E}[|\varepsilon_k|] = (\mathbb{E}[|\varepsilon_0|])^n < \infty,$$

De même,  $\mathbb{E}[X_n] = \mathbb{E}\left[\prod_{k=1}^n \varepsilon_k\right] = \prod_{k=1}^n \mathbb{E}[\varepsilon_k] = (\mathbb{E}[\varepsilon_0])^n = 0$ .

2. Si  $\varepsilon_0 \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ , alors par intégration par parties

$$\mathbb{E}[\varepsilon_0^4] = \int_{\mathbf{R}} x^4 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \left[-x^3 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right]_{-\infty}^{\infty} + \int_{\mathbf{R}} 3x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0 + 3 \text{var}(\varepsilon_0) = 3\sigma_\varepsilon^4.$$

En utilisant l'indépendance, le fait d'être identiquement distribuée et le fait que tous ces moments existent,

$$\mathbb{E}[X_n^2] = \mathbb{E}\left[\prod_{k=1}^n \varepsilon_k^2\right] = \prod_{k=1}^n \mathbb{E}[\varepsilon_k^2] = (\sigma_\varepsilon^2)^n,$$

et

$$\mathbb{E}[X_n^4] = \mathbb{E}\left[\prod_{k=1}^n \varepsilon_k^4\right] = \prod_{k=1}^n \mathbb{E}[\varepsilon_k^4] = (3\sigma_\varepsilon^4)^n = 3^n \sigma_\varepsilon^{4n}.$$

Si  $X_n$  était gaussienne centrée (voir  $\mathbb{E}[X_n] = 0$  obtenu à la question 1.), on devrait avoir d'après le calcul du moment d'ordre 4 d'une gaussienne centrée:

$$\mathbb{E}[X_n^4] = 3 (\mathbb{E}[X_n^2])^2 = 3 \sigma_\varepsilon^{4n}.$$

Or ici,  $\mathbb{E}[X_n^4] = 3^n \sigma_\varepsilon^{4n} \neq 3 \sigma_\varepsilon^{4n}$  si  $n \geq 2$ . Donc  $X_n$  n'est pas gaussienne pour  $n \geq 2$ .

3. Si  $\mathbb{E}[|\varepsilon_0|] < 1$ , alors  $\mathbb{E}[|X_n - 0|] = \mathbb{E}[|X_n|] = (\mathbb{E}[|\varepsilon_0|])^n \xrightarrow{n \rightarrow \infty} 0$ . Donc  $X_n$  converge dans  $\mathbb{L}^1$  vers 0, elle converge donc aussi en probabilité vers 0:  $X_n \xrightarrow[n \rightarrow +\infty]{P} 0$ .

4. D'après ce qui précède,  $\text{cov}(X_n, X_{n+k}) = \mathbb{E}[X_n X_{n+k}]$  (variables centrées). D'où

$$\text{cov}(X_n, X_{n+k}) = \mathbb{E}\left[X_n^2 \prod_{j=n+1}^{n+k} \varepsilon_j\right] = \mathbb{E}[X_n^2] \mathbb{E}\left[\prod_{j=n+1}^{n+k} \varepsilon_j\right] = \mathbb{E}[X_n^2] \cdot 0 = 0,$$

car les  $(\varepsilon_j)_{n+1 \leq j \leq n+k}$  sont indépendantes de  $X_n$  qui ne dépend que de  $(\varepsilon_j)_{1 \leq j \leq n}$  et  $(\varepsilon_i)$  forme une suite de v.a.i.i.d.

La matrice de covariance de  $(X_1, \dots, X_n)$  est donc diagonale puisque les covariances sont nulles entre les  $X_i$  d'indices différents:

$$\text{cov}((X_1, \dots, X_n)) = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\varepsilon^4 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_\varepsilon^{2n} \end{pmatrix}$$

Si  $\varepsilon_0$  a pour loi  $\mathcal{N}(0, 1)$  alors  $\text{cov}(X_1^2, X_2^2) = \mathbb{E}[\varepsilon_1^4 \varepsilon_2^2] - 1 = 3 - 1 = 2 \neq 0$  donc  $X_1^2$  et  $X_2^2$  ne sont pas indépendantes, soit  $\mathbb{E}[X_1^2 X_2^2] \neq \mathbb{E}[X_1^2] \mathbb{E}[X_2^2]$ :  $X_1$  et  $X_2$  ne sont pas indépendantes en général.

5. Si  $(X_n)$  était une suite indépendante, alors  $|X_1|$  et  $|X_2|$  seraient indépendantes, ce qui impliquerait que  $\text{cov}(|X_1|, |X_2|) = 0$ . Or  $\mathbb{E}[|X_1| |X_2|] = \mathbb{E}[|\varepsilon_1|^2] \mathbb{E}[|\varepsilon_2|]$  et  $\mathbb{E}[|X_1|] \mathbb{E}[|X_2|] = \mathbb{E}[|\varepsilon_1|] \mathbb{E}[|\varepsilon_1|] \mathbb{E}[|\varepsilon_2|]$ . Ainsi

$$\text{cov}(|X_1|, |X_2|) = \mathbb{E}[|\varepsilon_2|] (\mathbb{E}[|\varepsilon_1|^2] - \mathbb{E}[|\varepsilon_1|]^2) = \mathbb{E}[|\varepsilon_0|] \text{var}(|\varepsilon_0|),$$

puisque tous les  $\varepsilon_i$  ont même loi. D'où  $\text{var}(|\varepsilon_0|) = 0$ .

Ainsi  $|\varepsilon_0| = a$  p.s., avec  $a > 0$  puisque  $\text{var}(\varepsilon_0) > 0$ , ce qui implique que  $\varepsilon_0$  peut être égal à  $a$  ou à  $-a$ . Et comme  $\mathbb{E}[\varepsilon_0] = 0$ ,

$$\mathbb{P}(\varepsilon_0 = a) = \mathbb{P}(\varepsilon_0 = -a) = \frac{1}{2}.$$

6. Si  $(X_n)$  est identiquement distribuée, alors  $\mathbb{E}[|X_n|] = \mathbb{E}[|X_1|]$  pour tout  $n \in \mathbf{N}^*$ . Comme  $\mathbb{E}[|X_n|] = (\mathbb{E}[|\varepsilon_0|])^n$ , on a donc nécessairement  $\mathbb{E}[|\varepsilon_0|] = 1$ .

De plus,  $\text{var}(|X_2|) = \text{var}(|X_1|)$ , soit  $\mathbb{E}[\varepsilon_0^2]^2 - \mathbb{E}[|\varepsilon_0|]^4 = \mathbb{E}[\varepsilon_0^2] - \mathbb{E}[|\varepsilon_0|]^2$  ce qui impose  $\text{var}(|\varepsilon_0|) (\mathbb{E}[\varepsilon_0^2] + \mathbb{E}[|\varepsilon_0|]^2 - 1) = 0$ , d'où  $\text{var}(|\varepsilon_0|) = 0$ . On en déduit que  $|\varepsilon_0| = 1$  p.s., et comme  $\mathbb{E}[\varepsilon_0] = 0$ ,

$$\mathbb{P}(\varepsilon_0 = 1) = \mathbb{P}(\varepsilon_0 = -1) = \frac{1}{2}.$$

7. Si  $\varepsilon_0$  suit la loi de Rademacher, alors  $X_n$  est aussi à valeurs dans  $\{-1, 1\}$  comme produit de 1 et de  $-1$ . De plus, par récurrence,  $\mathbb{P}(X_1 = 1) = \mathbb{P}(\varepsilon_1 = 1) = 1/2$ , et si  $\mathbb{P}(X_n = 1) = 1/2$  alors, en utilisant l'indépendance,

$$\begin{aligned} \mathbb{P}(X_{n+1} = 1) &= \mathbb{P}(\varepsilon_{n+1} X_n = 1) = \mathbb{P}(\varepsilon_{n+1} X_n = 1 \cap X_n = 1) + \mathbb{P}(\varepsilon_{n+1} X_n = 1 \cap X_n = -1) \\ &= \mathbb{P}(\varepsilon_{n+1} = 1) \mathbb{P}(X_n = 1) + \mathbb{P}(\varepsilon_{n+1} = -1) \mathbb{P}(X_n = -1) = 1/4 + 1/4 = 1/2. \end{aligned}$$

On en déduit que  $X_n$  suit une loi de Rademacher pour tout  $n \in \mathbf{N}^*$ .

Pour  $(x_1, \dots, x_n) \in \{-1, 1\}^n$ , on a  $\varepsilon_k = X_{k-1} \varepsilon_k$ , d'où  $\varepsilon_k = x_{k-1} \varepsilon_k$  puisqu'alors  $X_k = x_{k-1} x_{k-1} \varepsilon_k = x_k$  du fait que  $x_{k-1}^2 = 1$ . Ainsi,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(\varepsilon_1 = x_1, \varepsilon_2 = x_1 x_2, \dots, \varepsilon_n = x_{n-1} x_n).$$

Par indépendance des  $\varepsilon_i$ , on obtient que pour tout  $n \in \mathbf{N}^*$

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{k=1}^n \mathbb{P}(\varepsilon_k = x_{k-1} x_k) = \left(\frac{1}{2}\right)^n = \prod_{k=1}^n \mathbb{P}(X_k = x_k),$$

donc  $(X_n)$  est une suite de variables aléatoires i.i.d. □

## Exercice 2 (Sur 15 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.

Sur  $(\Omega, \mathcal{A}, \mathbb{P})$ , on définit  $X$  comme une variable de Bernoulli de paramètre  $p \in [0, 1]$ ,  $U_1$  et  $U_2$  deux variables aléatoires continues 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  $Z(\Omega)$  (**0.5 pts**) ? En déduire que  $\mathbb{E}[|Z|^k] < \infty$  pour tout  $k \in \mathbf{N}$  (**0.5 pts**).
2. Calculer  $\mathbb{E}[Z]$  (**0.5 pts**) et montrer que  $\text{var}(Z) = \frac{1}{12} s^2 (4 - 3p^2)$  (**1 pt**).
3. Déterminer  $\mathbb{P}(Z \leq z)$  pour  $z \in \mathbf{R}$  (on pourra utiliser la formule des probabilités totales) (**1 pt**) et en déduire que  $Z$  est une variable aléatoire continue (**0.5 pts**) de densité  $f_Z$  telle que :

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

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  $p$  (**1 pt**). En déduire que l'estimateur du maximum de vraisemblance  $p_n^*$  de  $p$  sur  $[0, 1]$  est unique et vérifie

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

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

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

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  $0 < \varepsilon \leq s/2$ ,

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

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  $\frac{1-p}{s}$  (**1 pt**).

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

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

(b) Montrer que :

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

(c) Montrer que  $\sqrt{n} \mathbb{P}(s - n^{-3/4} < Z \leq s) \rightarrow 0$ .

(d) En déduire que

$$\mathbb{E}[|\sqrt{n} (\hat{p}_n - \tilde{p}_n)|] \rightarrow 0.$$

(e) Enfin, en déduire que

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

9. Donner à partir de  $(Z_1, \dots, Z_n)$  un intervalle de confiance à 95% pour  $p$  lorsque  $p$  et  $s$  sont inconnus (**0.5 pts**).

*Proof.* 1. On a  $X(\Omega) = \{0, 1\}$ ,  $U_1(\Omega) = [0, s]$  et  $U_2(\Omega) = [0, 2s]$ . Par conséquent  $Z(\Omega) = [0, 2s]$ .

$Z$  est une variable aléatoire bornée,  $|Z| \leq 2s$  : donc  $\mathbb{E}[|Z|^k] \leq (2s)^k < \infty$ .

2.  $\mathbb{E}[Z] = \mathbb{E}[X U_1] + \mathbb{E}[(1-X) U_2] = \mathbb{E}[X] \mathbb{E}[U_1] + \mathbb{E}[(1-X)] \mathbb{E}[U_2]$  puisque  $X, U_1$  et  $U_2$  sont supposées indépendantes. Par conséquent  $\mathbb{E}[Z] = p \frac{s}{2} + (1-p) s = (2-p) \frac{s}{2}$ .  
 $\mathbb{E}[Z^2] = \mathbb{E}[X^2 U_1^2] + \mathbb{E}[(1-X)^2 U_2^2] + 2 \mathbb{E}[X(1-X) U_1 U_2] = \mathbb{E}[X U_1^2] + \mathbb{E}[(1-X) U_2^2]$  puisque  $X(1-X) = 0$ . Par indépendance, on obtient  $\mathbb{E}[Z^2] = p \mathbb{E}[U_1^2] + (1-p) \mathbb{E}[U_2^2] = p \frac{1}{3} s^2 + (1-p) \frac{4}{3} s^2 = \frac{1}{3} s^2 (4 - 3p)$ . Finalement,  $\text{var}(Z) = \frac{1}{3} s^2 (4 - 3p) - (2-p)^2 \frac{s^2}{4} = \frac{1}{12} s^2 (4 - 3p^2)$ .

3. Tout d'abord,  $F_Z(z) = \mathbb{P}(Z \leq z) = 0$  si  $z < 0$  et  $F_Z(z) = \mathbb{P}(Z \leq z) = 1$  si  $z > 2s$  puisque  $Z(\Omega) = [0, 2s]$ .  
Si  $z \in [0, 2s]$ , en utilisant la formule des probabilités totales,

$$\begin{aligned}\mathbb{P}(Z \leq z) &= \mathbb{P}(Z \leq z \cap X = 1) + \mathbb{P}(Z \leq z \cap X = 0) \\ &= \mathbb{P}(U_1 \leq z \cap X = 1) + \mathbb{P}(U_2 \leq z \cap X = 0) \\ &= p \mathbb{P}(U_1 \leq z) + (1-p) \mathbb{P}(U_2 \leq z) \quad (\text{par indépendance}) \\ &= p \frac{1}{s} \min(z, s) + (1-p) \frac{z}{2s}.\end{aligned}$$

Par conséquent  $\mathbb{P}(Z \leq z) = (1+p) \frac{z}{2s}$  si  $0 \leq z \leq s$  et  $\mathbb{P}(Z \leq z) = p + (1-p) \frac{z}{2s}$  si  $s < z \leq 2s$ . Finalement,  $F_Z$  est une fonction continue (facile à vérifier pour  $z = 0, s$  et  $2s$ ), et est dérivable pour  $z \in \mathbf{R} \setminus \{0, s, 2s\}$  :  $Z$  est une variable aléatoire continue de densité :

$$f_Z(z) = (1+p) \frac{1}{2s} \mathbb{I}_{0 \leq z \leq s} + (1-p) \frac{1}{2s} \mathbb{I}_{s < z \leq 2s} = \left( \frac{1}{2s} (p+1) \mathbb{I}_{0 \leq z \leq s} + (1-p) \mathbb{I}_{s < z \leq 2s} \right) \mathbb{I}_{0 \leq z \leq 2s}.$$

4. La vraisemblance de  $p$  à partir de  $(Z_1, \dots, Z_n)$  est :

$$\begin{aligned}L_p(z_1, \dots, z_n) &= \prod_{i=1}^n \frac{1}{2s} (p+1) \mathbb{I}_{0 \leq z_i \leq s} (1-p) \mathbb{I}_{s < z_i \leq 2s} \quad \text{pour } (z_1, \dots, z_n) \in [0, 2s]^n \\ &= \frac{1}{(2s)^n} (p+1)^{\sum_{i=1}^n \mathbb{I}_{0 \leq z_i \leq s}} (1-p)^{\sum_{i=1}^n \mathbb{I}_{s < z_i \leq 2s}}.\end{aligned}$$

Par conséquent, la log-vraisemblance de  $p$  à partir de  $(Z_1, \dots, Z_n)$  est :

$$\hat{\ell}_n(p) = -n \ln(2s) + \ln(1+p) \sum_{i=1}^n \mathbb{I}_{0 \leq Z_i \leq s} + \ln(1-p) \sum_{i=1}^n \mathbb{I}_{s < Z_i \leq 2s}$$

La dérivée première de  $p \mapsto \hat{\ell}_n(p)$  s'annule pour  $p$  tel que

$$(1-p) \sum_{i=1}^n \mathbb{I}_{0 \leq Z_i \leq s} = (1+p) \sum_{i=1}^n \mathbb{I}_{s < Z_i \leq 2s} \implies p = \hat{p}_n = \frac{1}{n} \sum_{i=1}^n (2 \mathbb{I}_{0 \leq Z_i \leq s} - 1).$$

De plus  $\hat{\ell}_n''(p) = -\frac{1}{(1+p)^2} \sum_{i=1}^n \mathbb{I}_{0 \leq Z_i \leq s} - \frac{1}{(1-p)^2} \sum_{i=1}^n \mathbb{I}_{s < Z_i \leq 2s} < 0$  pour tout  $p \in [0, 1)$  (fonction concave), donc  $\hat{p}_n$  est bien l'unique maximum de  $\hat{\ell}_n(p)$ . Mais comme il est possible d'avoir  $\hat{p}_n < 0$ , dans ce cas le maximum est atteint en 0. D'où l'expression de  $p_n^*$ .

5. Considérons la suite  $(Y_n)$  telle que  $Y_i = 2 \mathbb{I}_{0 \leq Z_i \leq s} - 1$ . Puisque  $(Z_i)$  sont iidrv, il en est de même pour  $(Y_i)$ . De plus,  $\mathbb{E}[Y_i] = 2 \mathbb{P}(0 \leq Z \leq s) - 1 = 2(p+1)/2 - 1 = p$ , donc  $\mathbb{E}[\hat{p}_n] = p$  pour tout  $p \in [0, 1)$  : c'est un estimateur sans biais de  $p$ .

De plus,  $\text{var}(Y_i) = 4 \text{var}(\mathbb{I}_{0 \leq Z_i \leq s}) = 4 \frac{(p+1)}{2} \left(1 - \frac{(p+1)}{2}\right) = 1 - p^2$ . Toutes les hypothèses du théorème central limite sont vérifiées, et on a donc :

$$\sqrt{n} (\bar{Y}_n - \mathbb{E}[Y_1]) = \sqrt{n} (\hat{p}_n - p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1 - p^2).$$

6. Pour  $0 < \varepsilon < s/2$ ,

$$\begin{aligned}\mathbb{P}(\hat{s}_n \leq s - \varepsilon) &= \mathbb{P}\left(\max_{1 \leq i \leq n} Z_i \leq 2s - 2\varepsilon\right) \\ &= \mathbb{P}(Z_1 \leq 2s - 2\varepsilon \cap Z_2 \leq 2s - 2\varepsilon \cap \dots \cap Z_n \leq 2s - 2\varepsilon) \\ &= \mathbb{P}(Z \leq 2s - 2\varepsilon)^n \quad (\text{par indépendance et identité de loi}) \\ &= \left(p + (1-p) \frac{2s - 2\varepsilon}{2s}\right)^n \quad (\text{d'après (3) et puisque } 2s - 2\varepsilon \geq s) \\ &= \left(1 - (1-p) \frac{\varepsilon}{s}\right)^n.\end{aligned}$$

D'après ce qui précède, pour tout  $\varepsilon > 0$ , comme  $\hat{s}_n \leq s$ ,

$$\mathbb{P}(|\hat{s}_n - s| \geq \varepsilon) = \mathbb{P}(\hat{s}_n \leq s - \varepsilon) = \left(1 - (1-p) \frac{\varepsilon}{s}\right)^n \xrightarrow[n \rightarrow \infty]{} 0.$$

7. D'après ce qui précède, pour  $x \geq 0$ ,

$$\mathbb{P}(n(s - \hat{s}_n) \leq x) = \mathbb{P}(\hat{s}_n \geq s - \frac{x}{n}) = 1 - \left(1 - (1-p) \frac{x}{sn}\right)^n \xrightarrow[n \rightarrow \infty]{} 1 - \exp\left(-\frac{(1-p)}{s} x\right),$$

qui est la fonction de répartition d'une loi exponentielle de paramètre  $\frac{(1-p)}{s}$ .

8. (a) Puisque  $\hat{s}_n \leq s$ , on a

$$\sqrt{n} (\hat{p}_n - \tilde{p}_n) = 2 \sqrt{n} \sum_{i=1}^n \mathbb{I}_{0 < Z_i \leq s} - \mathbb{I}_{0 < Z_i \leq \hat{s}_n} = 2 \sqrt{n} \sum_{i=1}^n \mathbb{I}_{0 < Z_i \leq \hat{s}_n} + \mathbb{I}_{\hat{s}_n < Z_i \leq s} - \mathbb{I}_{0 < Z_i \leq \hat{s}_n} = 2 \sqrt{n} \sum_{i=1}^n \mathbb{I}_{\hat{s}_n < Z_i \leq s}.$$

- (b) D'après la formule de la probabilité totale, on a :

$$\begin{aligned}\mathbb{P}(\hat{s}_n < Z \leq s) &= \mathbb{P}(\hat{s}_n < Z \leq s \cap \hat{s}_n \geq s - n^{-3/4}) + \mathbb{P}(\hat{s}_n < Z \leq s \cap \hat{s}_n \leq s - n^{-3/4}) \\ &\leq \mathbb{P}(s - n^{-3/4} < Z \leq s \cap \hat{s}_n \geq s - n^{-3/4}) + \mathbb{P}(\hat{s}_n \leq s - n^{-3/4}) \\ &\leq \mathbb{P}(s - n^{-3/4} < Z \leq s) + \mathbb{P}(\hat{s}_n \leq s - n^{-3/4}),\end{aligned}$$

puisque  $\mathbb{P}(A \cap B) \leq P(A)$  pour tous événements  $A$  et  $B$ .

- (c) Maintenant, pour  $n$  suffisamment grand,  $\sqrt{n} \mathbb{P}(s - n^{-3/4} < Z \leq s) = \sqrt{n} (F_Z(s) - F_Z(s - n^{-3/4})) = \frac{1}{2} (p+1) n^{-1/4} \xrightarrow[n \rightarrow \infty]{} 0$ .

(d) We have  $\mathbb{E}[|\sqrt{n}(\hat{p}_n - \tilde{p}_n)|] = \mathbb{E}[\sqrt{n}(\hat{p}_n - \tilde{p}_n)] = 2\sqrt{n}\mathbb{P}(\hat{s}_n < Z \leq s)$ .

Nous avons vu que  $\sqrt{n}\mathbb{P}(s - n^{-3/4} < Z \leq s) \xrightarrow{n \rightarrow \infty} 0$ .

On a également

$$\sqrt{n}\mathbb{P}(\hat{s}_n \leq s - n^{-3/4}) = \sqrt{n}\left(1 - (1-p)\frac{n^{-3/4}}{s}\right)^n = \sqrt{n}\exp\left(n \log\left(1 - (1-p)\frac{n^{-3/4}}{s}\right)\right) \sim \sqrt{n}\exp\left(-\frac{(1-p)}{s}n^{1/4}\right) \xrightarrow{n \rightarrow \infty} 0.$$

Donc  $\mathbb{E}[|\sqrt{n}(\hat{p}_n - \tilde{p}_n)|] = 2\sqrt{n}\mathbb{P}(\hat{s}_n < Z \leq s) \xrightarrow{n \rightarrow \infty} 0$ .

(e) Puisque la convergence  $\mathbb{L}^1$  implique la convergence en probabilités,  $\mathbb{E}[|\sqrt{n}(\hat{p}_n - \tilde{p}_n)|] \xrightarrow{n \rightarrow \infty} 0$  implique  $\sqrt{n}(\hat{p}_n - \tilde{p}_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{P}} 0$ .

Puisque  $\sqrt{n}(\tilde{p}_n - p) = \sqrt{n}(\tilde{p}_n - \hat{p}_n) + \sqrt{n}(\hat{p}_n - p)$  et  $\sqrt{n}(\tilde{p}_n - \hat{p}_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{P}} 0$  and  $\sqrt{n}(\hat{p}_n - p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1 - p^2)$ , alors

$\sqrt{n}(\tilde{p}_n - p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1 - p^2)$  à partir du Lemme de Slutsky.

9. A partir du Lemme de Slutsky, on a  $\sqrt{n}(1 - \tilde{p}_n^2)^{-1/2}(\tilde{p}_n - p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$ . Donc un intervalle de confiance à 95%-confiance de  $p$  est:

$$\left[\tilde{p} - \frac{q_{0.975}}{\sqrt{n}}(1 - \tilde{p}_n^2)^{1/2}, \tilde{p} + \frac{q_{0.975}}{\sqrt{n}}(1 - \tilde{p}_n^2)^{1/2}\right],$$

avec  $q_{0.975}$  le quantile de la distribution gaussienne standard d'ordre 0.975.

□