STATISTIQUE MATHÉMATIQUE - CORRIÉ DE L'EXAMEN.

ENSAI - Mastères spécialisés - 2025/2026 -

ENSAI École nationale de la statistique et de l'analyse de l'information

Le barème initial a été modifié. Le I est maintenant sur 18, le II est sur 8.

I. 18 points

1°. $\boxed{\text{0.5 point}}$

$$\theta = \frac{1-p}{p}.$$

Interprétation : θ est le rapport $\mathbb{P}[\text{échec}]/\mathbb{P}[\text{succès}]$. C'est donc le nombre moyen d'échecs avant que n'apparaisse le premier succès. $\boxed{\text{0.5 point}}$

2°. Avec $p = \frac{1}{1+\theta}$ et $1-p = \frac{\theta}{1+\theta}$, on a immédiatement $\boxed{\text{1 point}}$

$$\mathbb{E}[X] = \frac{1}{p} = 1+\theta \text{ et } \mathbb{V}(X) = \frac{1-p}{p^2} = \theta(1+\theta).$$

3°. Par la méthode des moments, on a $\widehat{\theta}_1 = \overline{X} - 1$ $\boxed{\text{1 point}}$ et par suite, $\mathbb{E}[\widehat{\theta}_1] = \theta \Rightarrow$ le biais est nul $\boxed{\text{0.5 point}}$. Par ailleurs $\boxed{\text{1 point}}$

$$\mathbb{E}\left[(\widehat{\theta}_1 - \theta)^2\right] = \mathbb{V}(\overline{X}) = \frac{\theta(1+\theta)}{n}.$$

4°. Consistance, convergence en moyenne quadratique, normalité asymptotique.

D'après la LFGN, licite car les variables sont i.i.d. et possèdent un moment d'ordre 1, $\overline{X} \xrightarrow{p.s.} 1+\theta$ $\boxed{\text{1 point}}$.
D'après le théorème de l'application continue ($g(x) = x-1$ est continue) $\widehat{\theta}_1 \xrightarrow{p.s.} \theta$ $\boxed{\text{1 point}}$.

Comme $\widehat{\theta}_1$ est sans biais et $\mathbb{V}(\widehat{\theta}_1) = \theta(1+\theta)/n \to 0$, on a la convergence en moyenne quadratique. $\boxed{\text{1 point}}$.

D'après le TLC (licite, car les variables ont une variance finie) $\boxed{\text{1 point}}$,

$$\sqrt{n}(\widehat{\theta}_1 - \theta) = \sqrt{n}(\overline{X} - (1+\theta)) \rightsquigarrow \mathcal{N}\big(0,\ \theta(1+\theta)\big).$$

5°. Vraisemblance et EMV. $\boxed{\text{1 point}}$

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^{x_i - 1} = \theta^{S_n - n}(1+\theta)^{-S_n}.$$

Log-vraisemblance : $\ell(\theta) = (S_n - n)\log\theta - S_n \log(1+\theta)$.
Le score est donc $\ell'(\theta) = \frac{S_n - n}{\theta} - \frac{S_n}{1+\theta}$.
L'équation $\ell'(\theta) = 0$ donne

$$\widehat{\theta}_2 = \frac{S_n}{n} - 1 = \overline{X} - 1.$$

Il s'agit bien d'un maximum car l'expression change également de signe en ce point. $\boxed{\text{1 point}}$

On constate que $\widehat{\theta}_2 = \widehat{\theta}_1$.

6°. Modèle exponentiel et régularité.

La famille $\{(1-p)^{x-1}p : x \geq 1\}$ (ou sa reparamétrisation en θ) est une famille exponentielle unidimensionnelle à support indépendant du paramètre $\boxed{\text{1 point}}$. elle est régulière (paramètre intérieur, log-vraisemblance C^2, information finie $\boxed{\text{0.5 point}}$). La vraisemblance s'écrit

$$L(\theta) = \exp\left(s_n \ln\left(\frac{\theta}{1+\theta}\right) - n\ln\theta\right),$$

ce qui prouve que le modèle est exponentiel, avec fonction de partition donnée par $\lambda(\theta) = \ln(\theta/(1+\theta))$.

7°. Lorsque θ varie dans, $]0,\infty[$, $\lambda(\theta)$ varie dans $\mathbb{R}$, qui est un ouvert. Par suite, S_n est exhaustive et complète pour θ (théorème de factorisation). $\boxed{\text{1 point}}$

Comme $\widehat{\theta}_2 = \overline{X} - 1$ est sans biais et fonction de S_n, d'après le théorème de Lehmann-Scheffé, c'est l'estimateur VUMSB de θ. $\boxed{\text{1 point}}$

8°. L'information de Fisher existe car le modèle est régulier. En repartant de l'expression du score, $\boxed{\text{1 point}}$

$$\mathbb{I}_n(\theta) = \mathbb{V}(l'(\theta)) = \frac{n}{\theta(1+\theta)}.$$

Or $\mathbb{V}(\widehat{\theta}_2) = \frac{\theta(1+\theta)}{n} = \mathbb{I}_n(\theta)^{-1}$ et $\widehat{\theta}_2$ est sans biais : $\widehat{\theta}_2$ est efficace. $\boxed{\text{1 point}}$

9°. De $\pi(\theta) \propto \theta^{a-1}(1+\theta)^{-(a+b)}$ et $L(\theta) \propto \theta^{S_n - n}(1+\theta)^{-S_n}$, on déduit par application de la formule de Bayes que la loi *a posteriori* est une loi bêta de type II de paramètres $\mathscr{B}'\big(a + S_n - n,\ b+n\big)$. $\boxed{\text{1 point}}$

La moyenne *a posteriori* (si $b+n > 1$) est donc

$$\widehat{\theta}_3 = \mathbb{E}[\theta \mid X] = \frac{a + S_n - n}{b+n-1},$$

fonction de S_n. Son biais est :

$$\mathbb{E}_\theta[\widehat{\theta}_3] - \theta = \frac{a+n\theta}{b+n-1} - \theta = \frac{a - \theta(b-1)}{b+n-1} \xrightarrow[n\to\infty]{} 0,$$

L'estimateur est donc biaisé mais asymptotiquement sans biais. $\boxed{\text{1 point}}$

II. 8 points

1°.
$X - \theta \sim \mathcal{N}(0,1)$. En posant $Z = X - \theta$ et en utilisant les propriétés de symétrie de la fonction de répartition de la loi normale,

$$\mathbb{P}(X > 0) = \mathbb{P}(Z > -\theta) = 1 - \phi(-\theta) = 1 - (1 - \phi(\theta)) = \phi(\theta).$$

$\boxed{\text{1 point}}$
2°. Pour $p = \phi(\theta)$, par la règle de la dérivée de l'inverse donne $\boxed{\text{0.5 point}}$

$$(\phi^{-1})'(p) = \frac{1}{\phi'(\phi^{-1}(p))} = \frac{1}{\phi'(\theta)} = \frac{1}{f(\theta)}.$$

3°. Par indépendance et identique distribution des X_i, $Y_i = \mathbb{1}_{[X_i>0]}$ suit une loi de Bernoulli de paramètre $p = \phi(\theta)$ et $n\overline{Y}_n = \sum_{i=1}^n Y_i$ suit donc une loi binomiale de paramètres (n, p). $\boxed{\text{1 point}}$

Par suite, on a immédiatement $\mathbb{E}[\overline{Y}_n] = p = \phi(\theta)$ $\boxed{\text{0.5 point}}$

4°. Par la méthode des moments, on résoud $\overline{Y}_n = \phi(\theta)$, ce qui donne $\boxed{\text{1 point}}$

$$\widehat{\theta}_n = \phi^{-1}(\overline{Y}_n).$$

5°. En appliquant la LFGN (les variables en jeu sont i.i.d. et possèdent un moment d'ordre 1), $\overline{Y}_n \xrightarrow{p.s.} p = \phi(\theta)$. Par continuité de ϕ^{-1}, $\widehat{\theta}_n = \phi^{-1}(\overline{Y}_n) \xrightarrow{p.s.} \phi^{-1}(p) = \theta$ ce qui prouve la consistance forte. $\boxed{\text{1 point}}$

6°. En appliquant le TLC pour la séquence de Bernoulli,

$$\sqrt{n}\,(\overline{Y}_n - p) \rightsquigarrow \mathcal{N}\big(0,\ p(1-p)\big), \qquad p = \phi(\theta),$$

avec $p = \phi(\theta)$.

La variance limite est $\sigma^2 = \phi(\theta)\big(1-\phi(\theta)\big)$. $\boxed{\text{1 point}}$

7°. En utilisant la méthode delta (ϕ et ϕ^{-1} sont de classe C^∞ sur leur ensemble de définition, donc dérivables) avec $g(u) = \phi^{-1}(u)$ et $g'(p) = 1/f(\theta)$ $\boxed{\text{1 point}}$, on a :

$$\sqrt{n}\,(\widehat{\theta}_n - \theta) = \sqrt{n}\big(g(\overline{Y}_n) - g(p)\big) \rightsquigarrow \mathcal{N}\Big(0,\ g'(p)^2\,p(1-p)\Big)$$

La variance limite est donc $\boxed{\text{1 point}}$

$$\overline{\sigma}^2 = \frac{\phi(\theta)\,[1-\phi(\theta)]}{f(\theta)^2} = \frac{\phi(\theta)\,\big(1-\phi(\theta)\big)}{f(\theta)^2}$$