

Contrôle Continu

Rappels

- Il sera tenu compte du soin apporté à la présentation et à la rédaction.
- Si vous n'arrivez pas à démontrer un résultat, vous pouvez l'admettre pour la suite de l'exercice.

EXERCICE 1 (6 points) Soit $\theta > 0$ et X une variable aléatoire de densité (par rapport à la mesure de Lebesgue)

$$f_{\theta}(x) = \theta e^{-\theta(x-1)} \mathbf{1}_{\{x>1\}}.$$

On considère un échantillon i.i.d. $X_1, \dots, X_n$ de même loi que X .

1. Calculer et représenter la fonction de répartition F_{θ} de X . Déterminer la médiane de X .

Correction. Un calcul élémentaire donne

$$F_{\theta}(x) = \left(1 - e^{-\theta(x-1)}\right) \mathbf{1}_{\{x>1\}},$$

bijection de $]1, \infty[$ vers $]0, 1[$. On en déduit la médiane $x_{1/2} = 1 + \frac{\log 2}{\theta}$.

2. En déduire un estimateur $\hat{\theta}_n$ de θ . Montrer sa consistance et sa normalité asymptotique.

Correction. Puisque $f_{\theta}(x_{1/2}) > 0$, on sait que la médiane empirique $x_{1/2}(n) = X_{(\lceil \frac{n}{2} \rceil)}$ est un estimateur fortement consistant et asymptotiquement normal de $x_{1/2}$, avec

$$\sqrt{n} (x_{1/2}(n) - x_{1/2}) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \theta^{-2}).$$

Comme $x_{1/2} = 1 + \frac{\log 2}{\theta} = \varphi(\theta)$ avec φ C^1 -difféomorphisme de $]0, \infty[$ vers $]1, \infty[$, l'estimateur $\hat{\theta}_n := \varphi^{-1}(x_{1/2}(n)) = \frac{\log 2}{x_{1/2}(n) - 1}$ est un estimateur fortement consistant et asymptotiquement normal de θ , avec

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, (\theta / \log 2)^2).$$

3. Quelle est la loi de la variable $T := X - 1$? En déduire $\mathbb{E}_{\theta}[X]$ et $\text{Var}_{\theta}(X)$.

Correction. Par translation, on sait que T a pour densité

$$f_T(t) = f_{\theta}(x+1) = \theta e^{-\theta t} \mathbf{1}_{\{t>0\}},$$

donc $T \sim \mathcal{E}(\theta)$, d'où

$$\mathbb{E}_{\theta}[X] = 1 + \mathbb{E}[T] = 1 + \frac{1}{\theta} \quad \text{et} \quad \text{Var}_{\theta}(X) = \text{Var}(T) = \frac{1}{\theta^2}.$$

4. Le modèle $(f_\theta)_{\theta>0}$ est-il régulier ? Si oui, calculer l'information de Fisher $I(\theta) = I_1(\theta)$ pour une observation.

Correction. Pour tout réel x , la fonction $\theta \mapsto f_\theta(x)$ est C^1 donc les deux premiers points de la définition d'un modèle régulier sont vérifiés. De plus, avec les notations du cours et puisque $X > 1$ p.s. si $X \sim f_\theta$,

$$\ell_\theta(X) = \log f_\theta(x) = \log \theta - (X-1)\theta \quad \text{donc} \quad \ell'_\theta(X)^2 = \left(\frac{1}{\theta} - (X-1)\right)^2 = \left(X - \left(1 + \frac{1}{\theta}\right)\right)^2,$$

donc, d'après les calculs de $\mathbb{E}_\theta[X]$ et $\text{Var}_\theta(X)$ ci-dessus,

$$\mathbb{E}_\theta [\ell'_\theta(X)^2] = \text{Var}_\theta(X) = \frac{1}{\theta^2}.$$

La fonction $\theta \mapsto \theta^{-2}$ étant continue sur $]0, \infty[$, le modèle est régulier, d'information de Fisher $I(\theta) = \theta^{-2}$.

5. Déterminer l'estimateur du maximum de vraisemblance $\tilde{\theta}_n$.

Correction. Puisque $X_i > 1$ p.s. si $X_i \sim f_\theta$, on peut se passer des indicatrices dans la vraisemblance définie pour tout $t > 0$ par

$$L_n(t) = \prod_{i=1}^n f_t(X_i) = \prod_{i=1}^n t e^{-t(X_i-1)} \mathbf{1}_{\{X_i>1\}} = \prod_{i=1}^n t e^{-t(X_i-1)}.$$

Ainsi la log-vraisemblance s'écrit-elle

$$\ell_n(t) = n \log t - n(\bar{X}_n - 1)t,$$

donc $\ell_n(t) \geq 0$ ssi $t \leq 1/(\bar{X}_n - 1)$ et l'EMV est $\tilde{\theta}_n := 1/(\bar{X}_n - 1)$.

6. L'estimateur $\tilde{\theta}_n$ est-il asymptotiquement efficace ?

Correction. En considérant le C^1 -difféomorphisme $\psi(\theta) := \mathbb{E}_\theta[X] = 1 + 1/\theta$, on note que $\tilde{\theta}_n = \psi^{-1}(\bar{X}_n)$. Or le TCL assure que

$$\sqrt{n} (\bar{X}_n - \psi(\theta)) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \theta^{-2}),$$

donc la méthode Delta permet d'en déduire que

$$\sqrt{n} (\tilde{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \theta^2) = \mathcal{N}(0, 1/I(\theta)),$$

et $\tilde{\theta}_n$ est asymptotiquement efficace.

EXERCICE 2 (4 points) Soit $0 < \theta < 1$ et $X \sim \mathcal{B}(\sqrt{\theta})$, loi de Bernoulli de paramètre $\sqrt{\theta}$.

1. Rappeler ce que valent $\mathbb{E}_\theta[X]$ et $\text{Var}_\theta(X)$. En déduire un estimateur $\hat{\theta}_n$ de θ et préciser sa normalité asymptotique.

Correction. On sait que $\mathbb{E}_\theta[X] = \sqrt{\theta}$ et $\text{Var}_\theta(X) = \sqrt{\theta}(1 - \sqrt{\theta})$. On note que $\psi(\theta) := \mathbb{E}_\theta[X] = \sqrt{\theta}$ établit un C^1 -difféomorphisme de $]0, 1[$ vers $]0, 1[$, donc on considère $\hat{\theta}_n := \psi^{-1}(\bar{X}_n) = (\bar{X}_n)^2$. Par le TCL, on sait que

$$\sqrt{n} (\bar{X}_n - \psi(\theta)) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \sqrt{\theta}(1 - \sqrt{\theta})),$$

donc la méthode Delta assure que

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \theta^2) = \mathcal{N}(0, 4\theta^{3/2}(1 - \sqrt{\theta})),$$

2. Soit $0 < \alpha < 1$. Donner un intervalle de confiance de niveau asymptotique $(1 - \alpha)$ pour θ .

Correction. Notons $\sigma^2(\theta) := 4\theta^{3/2}(1 - \sqrt{\theta})$. Clairement la fonction $\theta \mapsto \sigma(\theta)$ est continue sur $]0, 1[$. L'estimateur $\hat{\theta}_n$ est un estimateur consistant de θ (puisque'il est asymptotiquement normal), donc le Théorème de Continuité assure que $\sigma(\hat{\theta}_n)$ est un estimateur consistant de $\sigma(\theta)$. Le Théorème de Slutsky permet alors de conclure que

$$\sqrt{n} \frac{\hat{\theta}_n - \theta}{\sigma(\hat{\theta}_n)} \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

En notant $\Phi^{-1}(1 - \alpha/2)$ le quantile d'ordre $(1 - \alpha/2)$ de la gaussienne standard, il s'ensuit qu'un intervalle de confiance de niveau asymptotique $(1 - \alpha)$ pour θ est

$$I_n := \left[\hat{\theta}_n - \frac{\Phi^{-1}(1 - \alpha/2)\sigma(\hat{\theta}_n)}{\sqrt{n}} ; \hat{\theta}_n + \frac{\Phi^{-1}(1 - \alpha/2)\sigma(\hat{\theta}_n)}{\sqrt{n}} \right].$$

3. Donner un test de niveau asymptotique α pour décider entre $H_0 : \theta = 1/2$ et $H_1 : \theta \neq 1/2$.

Correction. Deux possibilités :

- (a) On applique directement le résultat précédent : en notant $\mathbf{X} := (X_1, \dots, X_n)$, on sait qu'on définit un test de niveau asymptotique α en considérant

$$T_n(\mathbf{X}) = \mathbf{1}_{I_n \cap \{1/2\} = \emptyset},$$

c'est-à-dire que l'on rejette H_0 ssi

$$\left| \hat{\theta}_n - \frac{1}{2} \right| > \frac{\Phi^{-1}(1 - \alpha/2)\sigma(\hat{\theta}_n)}{\sqrt{n}}.$$

- (b) Sous $H_0 : \theta = 1/2$, on a

$$\sqrt{n} \left(\hat{\theta}_n - \frac{1}{2} \right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \sqrt{2} - 1),$$

donc

$$J_n := \left[\hat{\theta}_n - \frac{\Phi^{-1}(1 - \alpha/2)(\sqrt{2} - 1)^{1/2}}{\sqrt{n}} ; \hat{\theta}_n + \frac{\Phi^{-1}(1 - \alpha/2)(\sqrt{2} - 1)^{1/2}}{\sqrt{n}} \right]$$

est un intervalle de confiance de niveau asymptotique $(1 - \alpha)$ pour $1/2$, au sens où

$$\mathbb{P}_{1/2} \left(\frac{1}{2} \in J_n \right) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha.$$

Par conséquent, le test rejetant H_0 ssi

$$\left| \hat{\theta}_n - \frac{1}{2} \right| > \frac{\Phi^{-1}(1 - \alpha/2)(\sqrt{2} - 1)^{1/2}}{\sqrt{n}}$$

est de niveau asymptotique α .

4. Déterminer la p-valeur $\alpha_0(\mathbf{x})$ pour une réalisation $\mathbf{x} = (x_1, \dots, x_n)$.

Correction. Notons $\hat{\theta}_n(\mathbf{x})$ la réalisation de l'estimateur. Il y a deux possibilités, en fonction du test proposé en question précédente :

(a) Puisque

$$\alpha_0(\mathbf{x}) = \inf \left\{ \alpha \in [0, 1], \left| \hat{\theta}_n(\mathbf{x}) - \frac{1}{2} \right| > \frac{\Phi^{-1}(1 - \alpha/2) \sigma(\hat{\theta}_n(\mathbf{x}))}{\sqrt{n}} \right\},$$

il vient

$$\alpha_0(\mathbf{x}) = 2 \left(1 - \Phi \left(\frac{\sqrt{n} \left| \hat{\theta}_n(\mathbf{x}) - \frac{1}{2} \right|}{\sigma(\hat{\theta}_n(\mathbf{x}))} \right) \right).$$

(b) Puisque

$$\alpha_0(\mathbf{x}) = \inf \left\{ \alpha \in [0, 1], \left| \hat{\theta}_n(\mathbf{x}) - \frac{1}{2} \right| > \frac{\Phi^{-1}(1 - \alpha/2)(\sqrt{2} - 1)^{1/2}}{\sqrt{n}} \right\},$$

on obtient cette fois

$$\alpha_0(\mathbf{x}) = 2 \left(1 - \Phi \left(\frac{\sqrt{n} \left| \hat{\theta}_n(\mathbf{x}) - \frac{1}{2} \right|}{(\sqrt{2} - 1)^{1/2}} \right) \right).$$

5. (Bonus) Le modèle $(P_\theta)_{0 < \theta < 1} = (\mathcal{B}(\sqrt{\theta}))_{0 < \theta < 1}$ est-il régulier ? Si oui, calculer l'information de Fisher $I(\theta) = I_1(\theta)$ pour une observation. L'estimateur $\hat{\theta}_n$ est-il asymptotiquement efficace ?

Correction. Pour tout $\theta \in]0, 1[$, la densité de la loi $\mathcal{B}(\sqrt{\theta})$ par rapport à la mesure de comptage sur $\{0, 1\}$ s'écrit

$$f_\theta(x) = (\sqrt{\theta})^x (1 - \sqrt{\theta})^{1-x},$$

donc pour tout $x \in \{0, 1\}$, la fonction $\theta \mapsto f_\theta(x)$ est C^1 sur $]0, 1[$, ce qui assure que les deux premiers points de la définition d'un modèle régulier sont vérifiés. Pour le dernier point, en notant $X \sim \mathcal{B}(\sqrt{\theta})$, on a

$$\ell_\theta(X) = \frac{1}{2} X \log \theta + (1 - X) \log(1 - \sqrt{\theta}),$$

donc le score est

$$\ell'_\theta(X) = \frac{X - \sqrt{\theta}}{2\theta(1 - \sqrt{\theta})}.$$

Puisque $\mathbb{E}_\theta[X] = \sqrt{\theta}$ et $\text{Var}_\theta(X) = \sqrt{\theta}(1 - \sqrt{\theta})$, il s'ensuit que le moment d'ordre 2 du score est

$$\mathbb{E}_\theta [\ell'_\theta(X)^2] = \frac{\text{Var}_\theta(X)}{4\theta^2(1 - \sqrt{\theta})^2} = \frac{1}{4\theta^{3/2}(1 - \sqrt{\theta})},$$

qui est une fonction de θ continue sur $]0, 1[$, donc le modèle est régulier et l'information de Fisher pour une donnée est $I_1(\theta) = (4\theta^{3/2}(1 - \sqrt{\theta}))^{-1}$. En particulier, l'estimateur $\hat{\theta}_n$ est asymptotiquement efficace.