

Corrigé : Feuille de travaux dirigés 3

Solution Exercice 1 On a $X_i \stackrel{i.i.d.}{\sim} \mathcal{Pois}(\lambda)$. Le modèle pour une observation est $\mathcal{P} = \{\mathcal{Pois}(\lambda), \lambda > 0\}$. Le paramètre est λ et l'espace des paramètres est $\Lambda =]0, +\infty[$.

1. On a

$$\begin{aligned} \mathbb{E}_\lambda(X_1) &= \sum_{k \geq 0} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda \sum_{k \geq 1} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda \sum_{k \geq 0} \frac{\lambda^k}{(k)!} e^{-\lambda} \\ &= \lambda. \end{aligned}$$

Donc en considérant l'observation $X = (X_1, \dots, X_n)$ et en posant

$$S(X) = \frac{1}{n} \sum_{i=1}^n X_i$$

On a (par linéarité de l'espérance)

$$\mathbb{E}_\lambda(S(X)) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\lambda(X_i) = \lambda.$$

Le biais de S en tant qu'estimateur de λ , est donc

$$\text{biais}(S, \lambda) = \mathbb{E}_\lambda(S(X)) - \lambda = 0,$$

de sorte que S est un estimateur non biaisé de λ .

2. On est dans le cadre de l'estimation d'une fonction du paramètre, $\theta = g(\lambda) = e^{-\lambda}$.

(a) Un estimateur est une fonction des données, donc

$$S_1(X) = e^{-S(X)}$$

est un estimateur de θ , qui paraît « raisonnable » puisque S estime sans biais λ . Cependant S_1 est biaisé. En effet, puisque la fonction g est strictement convexe, et puisque la variable aléatoire $Y = S(X) : \Omega \rightarrow \mathbb{R}^+$ n'est pas constante, on a

$$g(\mathbb{E}(Y)) < \mathbb{E}(g(Y)),$$

c'est-à-dire, puisque $\mathbb{E}(Y) = \mathbb{E}(S(X)) = \lambda$ et $g(Y) = e^{-S(X)} = S_1$,

$$e^{-\lambda} < \mathbb{E}(S_1),$$

de sorte que $\text{biais}(S_1, \lambda) = \mathbb{E}(S_1(X)) - g(\lambda) > 0$ et S_1 est biaisé.

- (b) Un estimateur non biaisé de $\theta = e^{-\lambda}$ est une statistique S_2 telle que $\mathbb{E}_\lambda(S_2(X)) = e^{-\lambda}$. On remarque que $e^{-\lambda} = \mathbb{P}_\lambda(X_1 = 0) = \mathbb{E}[\mathbb{1}_{\{0\}}(X_1)]$, où $\mathbb{1}$ est la fonction

$$\text{indicateur : } \mathbb{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, en posant

$$S_2(X) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{0\}}(X_i)$$

on a bien $\mathbb{E}_\lambda(S_2(X)) = \frac{1}{n} \sum_{i=1}^n \mathbb{P}_\lambda(X_i = 0) = e^{-\lambda}$.

3. Efficacité des estimateurs : rappelons qu'un estimateur $S(X)$ non-biaisé d'une quantité $g(\lambda) \in \mathbb{R}$ est appelé *efficace* lorsque $\text{Var}_\lambda(S_2) = g'(\lambda)^2 / I(\lambda)$, où $I(\lambda)$ est l'information de Fisher relative à l'observation X , lorsque $X \sim P_\lambda$.

Dans le cas où $X = (X_1, \dots, X_n)$ est un échantillon i.i.d., avec $X_i \sim P_\lambda$, le vecteur X est distribué selon la loi produit $P_\lambda^{\otimes n}$ et l'information de Fisher pour le n -échantillon est $I(\lambda) = nI_1(\lambda)$, où $I_1(\lambda)$ est l'information de Fisher pour une seule observation $X_1 \sim P_\lambda$.

- (a) S_1 est biaisé donc il ne peut pas être efficace.
 (b) S est non biaisé. Il reste à calculer sa variance et l'information de Fisher $I_1(\lambda)$. Un calcul similaire à celui de la question 1. donne $\text{Var}_\lambda(X_1) = \lambda$, d'où

$$\text{Var}_\lambda(S(X)) = \frac{1}{n^2} \times \text{Var}_\lambda\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \times \sum_{i=1}^n \text{Var}_\lambda(X_i) = \frac{\lambda}{n}$$

où l'on a utilisé l'indépendance des X_i pour que la variance de la somme soit la somme des variances.

Calculons l'information de Fisher pour une observation. Ici, le modèle est dominé par la mesure discrète sur \mathbb{N} , $\mu = \sum_{i \in \mathbb{N}} \delta_i$, et la densité par rapport à cette mesure de référence est $p_\lambda(x) = \mathbb{P}_\lambda(X = x) = \lambda^x / x! e^{-\lambda}$.

$$\begin{aligned} I_1(\lambda) &= \mathbb{E}_\lambda \left[\left(\frac{\partial \log p_\lambda(X_1)}{\partial \lambda} \right)^2 \right] \\ &= \mathbb{E}_\lambda \left[\left(\frac{\partial (X_1 \log(\lambda) - \lambda - \log(X_1!))}{\partial \lambda} \right)^2 \right] \\ &= \mathbb{E}_\lambda \left[(X_1/\lambda - 1)^2 \right] \\ &= \frac{1}{\lambda^2} \mathbb{E} \left[(X_1 - \lambda)^2 \right] \\ &= \frac{1}{\lambda^2} \text{Var}_\lambda(X_1) \\ &= 1/\lambda \end{aligned}$$

D'où, pour n observations, $I(\lambda) = nI_1(\lambda) = n/\lambda$.

On peut conclure : l'estimateur S de λ est non biaisé et vérifie :

$$\forall \lambda > 0, \text{Var}_\lambda(S(X)) = \frac{\lambda}{n} = \frac{1}{I(\lambda)}.$$

Donc S atteint la borne de Cramér-Rao (car ici $g_0(\lambda) = \lambda$, de sorte que $g'(\lambda) = 1$), c'est-à-dire S est efficace.

- (c) Puisque S_2 est un estimateur de $g(\lambda)$, non biaisé, S_2 est efficace si et seulement si S_2 atteint la borne de Cramér-Rao, qui vaut

$$g'(\lambda)^2/I(\lambda) = \lambda e^{-2\lambda}/n.$$

On remarque que les variables aléatoires $Z_i = \mathbb{1}_{\{0\}}(X_i)$ sont des variables de Bernoulli de paramètre $p = \mathbb{P}(X_i = 0) = e^{-\lambda}$. Ainsi, nS_2 suit une loi binomiale de paramètres $(n, p = e^{-\lambda})$, de sorte que $\text{Var}_\lambda(S_2) = e^{-\lambda}(1 - e^{-\lambda})/n$. D'après le cours (on admet que le modèle est régulier), on a pour tout λ

$$\text{Var}_\lambda(S_2) = e^{-\lambda}(1 - e^{-\lambda})/n \geq g'(\lambda)^2/I(\lambda) = \lambda e^{-2\lambda}/n$$

De plus, l'inégalité est stricte pour au moins une valeur de λ (prendre $\lambda = 1$), donc S_2 n'est pas efficace.

Solution Exercice 2 [paramètre de translation]

Remarque préliminaire : le modèle n'est pas régulier car l'ensemble des points x tels que $p_\theta(x) > 0$ dépend de θ (c'est ensemble est $[\theta, +\infty[$).

1. l'espérance de X_1 vaut

$$\mathbb{E}_\theta(X_1) = \int_\theta^{+\infty} x e^{\theta-x} dx \stackrel{\text{calcul simple}}{=} \theta + 1.$$

En prenant $\hat{\theta}_n(X) = \frac{1}{n} \sum (X_i) - 1$, on a bien par linéarité de l'espérance

$$\mathbb{E}_\theta(\hat{\theta}_n(X)) = \frac{1}{n} \sum_i \mathbb{E}(X_i) - 1 = \theta.$$

Ainsi $\hat{\theta}_n$ est non biaisé.

2. Risque quadratique : Pour $\theta > 0$, il est donnée par

$$R(\theta, \hat{\theta}_n) = \mathbb{E} [(\hat{\theta}_n(X) - \theta)^2]$$

et puisque $\mathbb{E}(\hat{\theta}_n(X)) = \theta$, on a $R(\theta, \hat{\theta}_n) = \text{Var}_\theta(\hat{\theta}_n(X))$. De plus, pour une observation $X_1 \sim P_\theta$,

$$\begin{aligned} \text{Var}_\theta(X_1) &= \int_\theta^\infty (x - \theta - 1)^2 e^{\theta-x} dx \\ &= \int_0^\infty (t - 1)^2 e^{-t} dt \\ &= \left[-(t - 1)^2 e^{-t} \right]_0^\infty + 2 \underbrace{\int_0^\infty (t - 1) e^{-t} dt}_0 \\ &= 1. \end{aligned}$$

Ainsi, par indépendance des X_i ,

$$\text{Var}_\theta(\hat{\theta}_n(X)) = \frac{1}{n} \text{Var}_\theta(X_1) = \frac{1}{n}.$$

3. On considère $\tilde{\theta}_n(X) = \min_{i=1}^n X_i$.

Remarque : l'idée de choisir $\tilde{\theta}_n$ vient du fait que θ est la borne inférieure des $\{x : p_\theta(x) > 0\}$. Ainsi, pour $x > \theta$ on a $\mathbb{P}_\theta(X_1 \leq x) > 0$ alors que pour $x < \theta$, $\mathbb{P}_\theta(X \leq x) = 0$. Dans ce sens, θ est la plus petite valeur « possible » pour les X_i . Pour calculer la loi de $\tilde{\theta}_n(X)$, il est dans ce cas plus facile de calculer sa fonction de répartition que sa densité (à supposer que cette dernière existe). En effet, si on appelle \tilde{F}_θ cette fonction de répartition,

$$\tilde{F}_\theta(x) = \mathbb{P}_\theta(\tilde{\theta}_n(X) \leq x),$$

on a

$$\begin{aligned} 1 - \tilde{F}_\theta(x) &= \mathbb{P}_\theta(\tilde{\theta}_n(X) > x) \\ &= \mathbb{P}_\theta[\cap_{i=1}^n X_i > x] \\ &= \prod_{i=1}^n \mathbb{P}_\theta(X_i > x) \quad (\text{indépendance}) \\ &= \mathbb{P}_\theta(X_1 > x)^n. \end{aligned}$$

Reste à calculer cette dernière quantité en intégrant la densité de X_1 , ce qui donne

$$\forall x \geq \theta, \mathbb{P}_\theta(X_1 > x) = \int_x^\infty e^{\theta-t} dt = e^{\theta-x},$$

et pour $x \leq \theta$, $\mathbb{P}_\theta(X_1 > x) = 1$. Finalement $\mathbb{P}_\theta(X_1 > x) = e^{-\max(x-\theta, 0)}$. D'où

$$\tilde{F}_\theta(x) = 1 - e^{-n \max(x-\theta, 0)}.$$

4. Puisque $\tilde{F}_\theta(\theta) = 0$, on a, avec probabilité 1, $\tilde{\theta}_n(X) \geq \theta$. Ainsi, la variable aléatoire $Z = \tilde{\theta}_n(X) - \theta$ est presque sûrement positive. Comme elle n'est pas constante, son espérance est strictement positive. D'où : $\tilde{\theta}_n$ est biaisé.
5. Le risque quadratique de $\tilde{\theta}_n$ est $R(\theta, \tilde{\theta}_n) = \mathbb{E}[(\theta - \tilde{\theta}_n(X))^2]$. On utilise le fait que pour une variable aléatoire $Y : \Omega \rightarrow \mathbb{R}^+$, $\mathbb{E}(Y) = \int_{t=0}^+ \mathbb{P}(Y > t) dt$. On pose $Y = (\tilde{\theta}_n - \theta)^2$. Alors, pour $t \geq 0$, $\mathbb{P}_\theta(Y > t) = \mathbb{P}_\theta(\tilde{\theta}_n - \theta > \sqrt{t})$ car avec probabilité

1, $\tilde{\theta}_n - \theta \geq 0$. Ainsi,

$$\begin{aligned}
 R(\theta, \tilde{\theta}_n) &= \mathbb{E}(Y) \\
 &= \int_{t=0}^{\infty} \mathbb{P}_{\theta}(Y > t) dt \\
 &= \int_{t=0}^{\infty} \mathbb{P}_{\theta}(\tilde{\theta}_n - \theta > \sqrt{t}) dt \\
 &= \int_{t=0}^{\infty} \mathbb{P}_{\theta}(\tilde{\theta}_n - \theta > \sqrt{t}) dt \\
 &= \int_{t=0}^{\infty} e^{-n\sqrt{t}} dt \quad (\text{cf question précédente}) \\
 &= \int_{u=0}^{\infty} e^{-nu} 2u du \\
 &= \frac{2}{n^2} \int_{r=0}^{\infty} e^{-r} r dr \\
 &= \frac{2}{n^2}
 \end{aligned}$$

6. Lorsque n est plus grand que 2, on a

$$\forall \theta, \quad R(\theta, \tilde{\theta}_n) = 2/n^2 < 1/n = R(\theta, \hat{\theta}_n).$$

Le risque du deuxième estimateur (qui est pourtant biaisé) est uniformément plus faible que celui du premier estimateur. $\tilde{\theta}_n$ est donc préférable à $\hat{\theta}_n$.

Solution Exercice 3 $X = (X_1, \dots, X_n)$ un échantillon i.i.d. où $X_i \sim \mathcal{Ber}(\theta)$ ($\theta \in]0, 1[$).

- On considère $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ comme estimateur de θ . Pour que \bar{X} soit efficace, il faut et il suffit que (i) \bar{X} soit non biaisé et (ii) la variance de \bar{X} atteigne la borne de Cramer-Rao pour le modèle de Bernoulli avec n observations.
 - \bar{X} est sans biais car $\mathbb{E}(\bar{X}) = \frac{1}{n} \sum \mathbb{E}(X_i) = \mathbb{E}(X_1) = \theta$.
 - La variance de \bar{X} est

$$\forall \theta \in]0, 1[, \text{Var}_{\theta}(\bar{X}) \stackrel{\text{independance}}{=} \frac{1}{n^2} \sum_i \text{Var}(X_i) = \frac{\theta(1-\theta)}{n}.$$

- Calculons l'information de Fisher pour une observation : Tout d'abord, la mesure de référence est $\mu = \delta_0 + \delta_1$ et la densité de la loi P_{θ} par rapport à μ est

$$\forall x \in \{0, 1\}, p_{\theta}(x) = \theta^x (1-\theta)^{1-x}.$$

(en effet, on vérifie que l'on a $p_{\theta}(1) = \theta = \mathbb{P}_{\theta}(X_1 = 1)$ et $p_{\theta}(0) = 1 - \theta = \mathbb{P}_{\theta}(X_1 = 0)$. Ainsi

$$\forall x \in \{0, 1\}, \forall \theta \in]0, 1[, \frac{\partial \log p_{\theta}(x)}{\partial \theta} = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)}$$

D'où

$$\begin{aligned} I_1(\theta) &= \mathbb{E}_\theta \left(\frac{\partial \log p_\theta(X_1)}{\partial \theta} \right)^2 \\ &= \mathbb{E}_\theta \left[\left(\frac{X_1 - \theta}{\theta(1-\theta)} \right)^2 \right] \\ &= \frac{1}{\theta^2(1-\theta)^2} \text{Var}_\theta(X_1) \\ &= \frac{1}{\theta(1-\theta)}. \end{aligned}$$

D'où, pour n observations, $I(\theta) = nI_1(\theta) = \frac{n}{\theta(1-\theta)}$.

Conclusion : \bar{X} est non biaisé et $\text{Var}_\theta(X) = 1/I(\theta)$, pour tout $\theta \in]0, 1[$ donc \bar{X} est efficace.

2. On cherche maintenant un estimateur de la variance des X_i , $g(\theta) = \text{Var}_\theta(X_1) = \theta(1-\theta)$. Puisque \bar{X} est un estimateur non biaisé de θ , l'estimateur $S(X) = \bar{X}(1-\bar{X})$ est un candidat naturel pour estimer $g(\theta)$. Pourtant il est biaisé :

$$\begin{aligned} \mathbb{E}_\theta[S(X)] &= \mathbb{E}_\theta(\bar{X}) - \mathbb{E}_\theta(\bar{X}^2) = \theta - \frac{1}{n^2} \left(\sum_{i=1}^n \mathbb{E}(X_i^2) + 2 \sum_{i<j} \mathbb{E}(X_i X_j) \right) \\ &= \theta - \frac{1}{n^2} \left(n\theta + 2 \frac{n(n-1)}{2} \theta^2 \right) \quad \text{car par indépendance } \mathbb{E}(X_i X_j) = \mathbb{E}(X_i) \mathbb{E}(X_j) = \theta^2 \\ &= \theta - \frac{1}{n} \left(\theta + (n-1)\theta^2 \right) \\ &= \left(1 - \frac{1}{n} \right) \theta(1-\theta) < \theta(1-\theta) \end{aligned}$$

On cherche un estimateur non biaisé de la forme

$$\hat{v} = \eta S.$$

C'est-à-dire, on cherche à "débiaiser" S en le multipliant par une constante. D'après ce qui précède, si l'on prend $\eta = \frac{1}{1-\frac{1}{n}} = \frac{n}{n-1}$, on a

$$\mathbb{E}\hat{v}(X) = \eta \mathbb{E}S(X) = \theta(1-\theta),$$

de sorte que \hat{v} est non biaisé.