Corrigé : Feuille de travaux dirigés 3

Solution Exercice 1 On a $X_i \overset{i.i.d.}{\sim} \mathcal{P}oiss(\lambda)$. Le modèle pour une observation est $\mathcal{P} = \{\mathcal{P}oiss(\lambda), \lambda > 0\}$. 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 $X = (X_1, \ldots, X_n)$ 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 parait « 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 $Y = S(X) : \Omega \to \mathbb{R}^+$ 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 $e^{-\lambda} = \mathbb{P}_\lambda(X_1 = 0) = \mathbb{E}\Big[\mathbb{1}_{\{0\}}(X_1)\Big]$, où $\mathbb{1}$ est la fonction indicatrice : $\mathbb{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon.} \end{cases}$

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 $\mathbb{V}\mathrm{ar}_\lambda(S_2) = g'(\lambda)^2/I(\lambda)$, où $I(\lambda)$ est l'information de Fisher relative à l'observation X, lorsque $X \sim \mathrm{P}_\lambda$.

Dans le cas où $X = (X_1, \dots, X_n)$ est un échantillon i.i.d. , avec $X_i \sim \mathrm{P}_\lambda$, le vecteur X est distribué selon la loi produit $\mathrm{P}_\lambda^{\otimes n}$ 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 $X_1 \sim \mathrm{P}_\lambda$.

(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 $\mathbb{V}\mathrm{ar}_\lambda(X_1) = \lambda$, d'où

$$\mathbb{V}\mathrm{ar}_\lambda(S(X)) = \frac{1}{n^2} \times \mathbb{V}\mathrm{ar}_\lambda(\sum_{i=1}^{n} X_i) = \frac{1}{n^2} \times \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}_\lambda(X_i) = \frac{\lambda}{n}$$

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 $p_\lambda(x) = \mathbb{P}_\lambda(X = x) = \lambda^x/x!e^{-\lambda}$.

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

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 :

$$\forall \lambda > 0, \mathbb{V}\mathrm{ar}_\lambda(S(X)) = \frac{\lambda}{n} = \frac{1}{I(\lambda)}.$$

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 $\mathbb{V}\mathrm{ar}_\lambda(S_2) = e^{-\lambda}(1 - e^{-\lambda})/n$. D'après le cours (on admet que le modèle est régulier), on a pour tout λ

$$\mathbb{V}\mathrm{ar}_\lambda(S_2) = e^{-\lambda}(1 - e^{-\lambda})/n \geq g'(\lambda)^2/I(\lambda) = \lambda e^{-2\lambda}/n$$

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 $X = (X_1, \ldots, X_n)$ un éhantillon i.i.d. où $X_i \sim \mathcal{B}er(\theta)$ ($\theta \in]0, 1[$).

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

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

— 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

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

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

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

2. On cherche maintenant un estimateur de la variance des X_i, $g(\theta) = \mathbb{V}\mathrm{ar}_\theta(X_1) = \theta(1-\theta)$. 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 :

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

On cherche un estimateur non biaisé de la forme

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

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

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

de sorte que $\widehat{v}$ est non biaisé.

Solution Exercice 3 1. On obtient le MV est annulant la dérivée partielle en θ de la log vraisemblance (qui est bien différentiable et strictement concave sur $]0,+\infty[$, en tant que fonction de θ) :

$$\frac{\partial}{\partial\theta}(\log p_\theta^{\otimes n}(\mathbf{x})) = n\left(\frac{1}{\theta} + \log u\right) - \sum \log X_i = \frac{n}{\theta} - \sum \log(X_i/u),$$

d'où le résultat :

$$\widehat{\theta}_{MV}(X) = \frac{n}{\sum_{i=1}^n \log\left(\frac{X_i}{u}\right)}.$$

2. Risque quadratique
Pour $y > 0$,
$$\mathbb{P}_\theta(\log(X_i/u) > y) = \mathbb{P}_\theta(X_i > ue^y) = 1 - F_\theta(ue^y) = (u/(ue^y))^\theta = e^{-\theta y}$$
donc $Y_i := \log(X_i/u) \sim \mathcal{E}(\theta)$.
3. D'après le point 2, $V_i := \frac{1}{n}\log(X_i/u) \sim \mathcal{E}(n\theta)$: on l'obtient en calculant $\mathbb{P}_\theta(V_i > y)$ pour $y > 0$. Ainsi, par le point 2. donné dans l'énoncé, $\frac{1}{n}\sum_{i=1}^n \log(X_i/u) \sim \mathcal{G}amma(n, n\theta)$, d'où le résultat.
4. Par définition le biais de $\widehat{\theta}_{MV}$ est
$$\mathbb{E}_\theta(\theta_{MV}(X)) - \theta = \frac{n\theta}{n-1} - \theta = \frac{\theta}{n-1}$$
où la deuxième égalité vient de l'expression de l'espérance d'un inverse Gamma donnée dans l'énoncé. Le risque quadratique de $\widehat{\theta}_{MV}$ est donc, d'après l'expression de la variance d'un inverse gamma,
$$R(\theta, \widehat{\theta}_{MV}) = \text{biais}^2(\theta) + \text{var}_\theta(\widehat{\theta}_{MV}(X)) = \frac{\theta^2}{(n-1)^2} + \frac{n^2\theta^2}{(n-1)^2(n-2)} = \frac{\theta^2}{(n-1)^2}(1 + \frac{n^2}{n-2}).$$
5. pour une observation,
$$I_1(\theta) = \mathbb{E}_\theta[(\frac{-\partial}{\partial\theta}\log p_\theta(X_1))^2] = \mathbb{E}_\theta[(\frac{1}{\theta} - \log(X_1/u))^2]$$
$$= \mathbb{V}\text{ar}_\theta(Y)$$
où $Y \sim \exp(\lambda)$ d'où
$$I_1(\theta) = \frac{1}{\lambda^2}$$
puis
$$I_n(\theta) = \frac{n}{\lambda^2}.$$
6. Puisque $\widehat{\theta}_c = \frac{n-1}{n}\widehat{\theta}$, on a directement
$$\mathbb{V}\text{ar}_\theta(\widehat{\theta}_c(X)) = \frac{(n-1)^2}{n^2}\mathbb{V}\text{ar}_\theta(\widehat{\theta}_{MV}) = \frac{\theta^2}{n-2}$$
7. L'estimateur $\widehat{\theta}_{MV}$ est biaisé : il ne peut pas être efficace. L'estimateur $\widehat{\theta}_c$ est non-biasé mais sa variance $\frac{\theta^2}{n-2}$ est strictement supérieure à la borne de Cramér-Rao $\frac{1}{I_n(\theta)} = \frac{\theta^2}{n}$. Il n'est donc pas non plus efficace.
8. En comparant $R(\theta, \widehat{\theta}_{MV})$ et $R(\theta, \widehat{\theta}_c) = \mathbb{V}\text{ar}_\theta(\widehat{\theta}_c(X))$, on constate que le premier est supérieur au second, pour tout θ, dès que $n \geq 4/3$. Or on a supposé $n > 2$ donc on choisit $\widehat{\theta}_c$.