Détermination d'une statistique exhaustive

Loi de Poisson

La loi de Poisson est une loi de probabilité discrète qui décrit le comportement du nombre d'évènements se produisant dans un laps de temps fixé, si ces évènements se produisent avec une fréquence moyenne connue et indépendamment du temps écoulé depuis l'évènement précédent. Si le nombre moyen d'occurrences dans cet intervalle est λ (le paramètre de la loi), alors la probabilité qu'il existe exactement x occurrences ($x \in \mathbb{N}$) est

$$p_X(x;\lambda) = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \tag{1}$$

Étant donné un échantillon *i.i.d.* $(X_1, \dots, X_n)$ généré suivant une loi de Poisson de paramètre inconnu λ, $p_X(x;\lambda)$, donnez une statistique exhaustive pour λ.

Solution

La loi jointe pour l'échantillon *i.i.d.* $(x_1, \dots, x_n)$ est donnée par :

$$\begin{aligned}
p(x_1, \dots, x_n; \lambda) &= \prod_{i=1}^{n} p(x_i; \lambda) & (2)\\
&= \prod_{i=1}^{n} [e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}] & (3)\\
&= \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} \frac{1}{\prod_{i=1}^{n} x_i!} & (4)\\
&= \underbrace{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}_{u(T(x_1,\dots,x_n);\theta)} \underbrace{\frac{1}{\prod_{i=1}^{n} x_i!}}_{v(x_1,\dots,x_n)} & (5)
\end{aligned}$$

A partir de ce résultat, et d'après le théorème de factorisation (critère de Fisher-Neyman), on a : $\sum_{i=1}^{n} X_i$ est une statistique exhaustive pour λ.

Famille exponentielle

Une densité de probabilité (ou loi de probabilité dans le cas de v.a. discrètes) $f(x;\theta)$ appartient à la *famille exponentielle* si $f(x;\theta)$ peut s'écrire sous la forme

$$f(x;\theta) = \exp[\eta(\theta)T(x) + a(\theta) + b(x)] \tag{6}$$

Soit X une variable aléatoire de densité $f(x;\theta)$ appartenant à la famille exponentielle. En déduire une statistique exhaustive pour le paramètre θ pour un échantillon *i.i.d.* $(X_1, \dots, X_n)$

Solution

La densité jointe pour les observations *i.i.d.* $\mathbf{x} = (x_1, \ldots, x_n)$ est donnée par :

$$\begin{aligned}
f(\mathbf{x};\theta) &= \prod_{i=1}^{n} f(x_i;\theta) && (7)\\
&= \prod_{i=1}^{n} \exp[\eta(\theta)T(x)+a(\theta)+b(x)] && (8)\\
&= \exp\left[\sum_{i=1}^{n} (\eta(\theta)T(x_i)+a(\theta)+b(x_i))\right] && (9)\\
&= \underbrace{\exp\left[\sum_{i=1}^{n} \eta(\theta)T(x_i)+n\,a(\theta)\right]}_{u(T(x_1,\ldots,x_n);\theta)} \underbrace{\exp\left[\sum_{i=1}^{n} b(x_i)\right]}_{v(x_1,\ldots,x_n)} && (10)
\end{aligned}$$

A partir de ce résultat, et d'après le théorème de factorisation (critère de Fisher-Neyman), on a : $\sum_{i=1}^{n} T(x_i)$ est une statistique exhaustive pour θ.

Estimateurs et détermination de la Borne Inférieure de Cramér-Rao (CRLB)

Soit X une v.a. uni-variée à densité normale :

$$f(X;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}.$$

1. Calculer la CRLB pour un estimateur sans biais de l'espérance μ (variance σ^2 connue)
   La moyenne empirique $\bar{X}$ est un estimateur de l'espérance μ (démo en prochain TD),
2. Montrer qu'il est sans biais
3. En déduire qu'il est efficace
4. Calculer la CRLB pour un estimateur sans biais de la variance σ^2 (espérance μ connu)
5. La variance empirique S^2 est un estimateur biaisé de la variance σ^2 (démo en prochain TD), de variance $\frac{2\sigma^4}{n-1}$, en déduire son efficacité.

Solution

1. On a

$$\begin{aligned}
\ln f(X;\mu) &= \ln \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} && (11)\\
&= \ln\left[\frac{1}{\sigma\sqrt{2\,\pi}}\right] - \frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2. && (12)
\end{aligned}$$

et

$$\begin{aligned}
\frac{\partial \ln f(X;\mu)}{\partial\mu} &= \frac{1}{\sigma^2}\left(\frac{X-\mu}{\sigma}\right) && (13)\\
\left(\frac{\partial \ln f(X;\mu)}{\partial\mu}\right)^2 &= \frac{1}{\sigma^4}\left(\frac{X-\mu}{\sigma}\right)^2 && (14)
\end{aligned}$$

Ainsi

$$\mathbb{E}\left[\left(\frac{\partial \ln f(X;\mu)}{\partial\mu}\right)^2\right] = \frac{1}{\sigma^4}\mathbb{E}\left[(X-\mu)^2\right] = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2} \qquad (15)$$

La CRLB pour l'estimateur de μ est donc

$$\mathrm{CRLB} = \frac{1}{n\mathbb{E}\left[\left(\frac{\partial \ln f(X;\mu)}{\partial\mu}\right)^2\right]} = \frac{\sigma^2}{n} \qquad (16)$$

2. on a $\mathbb{E}[\bar{X}] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu$. L'estimateur $\bar{X}$ est donc sans biais.
3. Ensuite, on a $\mathrm{Var}(\bar{X}) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$

   Puisque la variance de $\bar{X}$ est $\frac{\sigma^2}{n}$, il est donc à variance minimale pour μ quand la v.a. X est a densité normale $\mathcal{N}(\mu, \sigma^2)$
4. Soit $\theta = \sigma^2$, on a

$$\begin{aligned} \ln f(X;\theta) &= \ln \frac{1}{\sqrt{\theta 2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\frac{(X-\mu)^2}{\theta}} && (17) \\ &= -\frac{1}{2}\ln 2\pi\theta - \frac{1}{2}\frac{(X-\mu)^2}{\theta}. && (18) \end{aligned}$$

et

$$\begin{aligned} \frac{\partial \ln f(X;\theta)}{\partial\theta} &= -\frac{1}{2\theta} + \frac{(X-\mu)^2}{2\theta^2} && (19) \\ \frac{\partial^2 \ln f(X;\theta)}{\partial^2\theta} &= \frac{1}{2\theta^2} - \frac{(X-\mu)^2}{\theta^3} && (20) \\ & && (21) \end{aligned}$$

Ainsi

$$\mathbb{E}\left[\frac{\partial^2 \ln f(X;\theta)}{\partial^2\theta}\right] = \frac{1}{2\theta^2} - \frac{\mathbb{E}\left[(X-\mu)^2\right]}{\theta^3} = \frac{1}{2\theta^2} - \frac{\theta}{\theta^3} = -\frac{1}{2\theta^2} = -\frac{1}{2\sigma^4} \quad (22)$$

La CRLB pour l'estimateur de σ^2 est donc donnée par [1]

$$\mathrm{CRLB} = -\frac{1}{n\mathbb{E}\left[\frac{\partial^2 \ln f(X;\mu)}{\partial^2\theta}\right]} = \frac{2\sigma^4}{n} \quad (23)$$

5. On a $\mathrm{Var}(S^2) = \frac{2\sigma^4}{n}$ donc l'efficacité de S^2 est donnée par :

$$e(S^2) = \frac{\mathrm{CRLB}}{\mathrm{Var}(S^2)} = \frac{2\sigma^4/n}{2\sigma^4/(n-1)} = \frac{n-1}{n}$$

S^2 n'est donc pas un estimateur efficace pour σ^2 (car $e(S^2) < 1$). Cependant, S^2 est *asymptotiquement* efficace car

$$\lim_{n\to+\infty} e(S^2) = 1$$

1. ⚠ Remarque : ici on a utilisé la deuxième forme de la CRLB