Infarctus à l'hôpital public

Loïc Rakotoson

2020-11-13

1 Modèle paramétrique

$$f_X(x) = \frac{\alpha}{\theta^\alpha} x^{\alpha-1} 1_{0 \leq x \leq \theta} \quad \forall x \in \mathbb{R}$$

$(X_i)_{i \geq 1}$ iid de même loi que X.

1.1 Question 1

On suppose qu'on connaît θ et on cherche à estimer α.
Montrer que $Y = ln(\frac{\theta}{X}) \sim \mathcal{E}(\lambda)$.

Soit g une fonction strictement décroissante et dérivable.

$$\begin{aligned} g : \mathbb{R}_*^+ &\longrightarrow \mathbb{R} \\ x &\longmapsto y = ln(\frac{\theta}{x}) \end{aligned}$$

Alors on a $g^{-1} : y \to \theta e^{-y}$ et $\frac{\partial}{\partial y} g^{-1} = -\theta e^{-y}$.

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} \mathbb{P}[\, Y \leq y] \, = \frac{\partial}{\partial y} \mathbb{P}[\, g(X) \leq y] \\ &= \frac{\partial}{\partial y} \mathbb{P}[\, g^{-1}.g(X) \geq g^{-1}(y)] \, = \frac{\partial}{\partial y} 1 - F_X(g^{-1}(y)) \\ &= \left| \frac{\partial}{\partial y} g^{-1}(y) \right| f_X(g^{-1}(y)) = \alpha \frac{e^{-\alpha y + y}}{e^y} \\ f_Y(y) &= \alpha e^{-\alpha y} \end{aligned}$$

donc $Y \sim \mathcal{E}(\alpha)$.

$\mathbb{E}[\, Y] \, = \frac{1}{\alpha}$ et $\mathbb{V}[\, Y] \, = \frac{1}{\alpha^2}$.

On peut en déduire l'information de Fisher $I_X(\alpha)$ du modèle.

$$\mathcal{L}(x_i;\alpha) = ln\, L(x_i;\alpha) = n\, ln \frac{\alpha}{\theta^\alpha} + \sum_{i=1}^{n} ln\, x_i^{\alpha-1}$$

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha) &= \frac{n}{\alpha} - n\, ln\, \theta + \sum_{i=1}^{n} ln\, x_i \\
&= \frac{n}{\alpha} - \sum_{i=1}^{n} ln\, \theta - ln\, x_i \\
&= \frac{n}{\alpha} - \sum_{i=1}^{n} y_i \\
\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha) &= \frac{\partial}{\partial\alpha}\mathcal{L}(y_i;\alpha)
\end{aligned}$$

$$\begin{aligned}
I_X(\alpha) &= \mathbb{V}\left[\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha)\right] \\
&= n\, \mathbb{V}[Y] \\
I_X(\alpha) &= \frac{n}{\alpha^2}
\end{aligned}$$

1.2 Question 2

Estimateur de maximum de vraisemblance.

$$\begin{aligned}
&\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha) = 0 \\
\Leftrightarrow \quad & \alpha = \frac{n}{\sum_{i=1}^{n} y_i} \\
\Leftrightarrow \quad & \alpha = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} g(x_i)} \\
\tilde{\alpha}_{EMV} &= \frac{1}{\frac{1}{n}\sum_{i=1}^{n} ln \frac{\theta}{X_i}}
\end{aligned}$$

1.3 Question 3

On suppose à présent que θ et α sont inconnus.
Information de Fisher $I_X(\alpha, \theta)$ du modèle.
Comme le support de f_X dépend de θ, on ne peut pas déduire l'information de Fisher à partir de la relation de second ordre. On utilisera donc l'espérance du carrée du premier niveau des dérivée partielle.

$$\mathcal{L}(x_i;\alpha,\theta) = ln\,L(x_i;\alpha,\theta) = n\,ln\frac{\alpha}{\theta^\alpha} + \sum_{i=1}^n ln\,x_i^{\alpha-1}$$

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha,\theta) &= \frac{n}{\alpha} - \sum_{i=1}^n y_i \\
\frac{\partial}{\partial\theta}\mathcal{L}(x_i;\alpha,\theta) &= -\frac{n\alpha}{\theta}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha,\theta)\right)^2\right] &= \frac{n^2}{\alpha^2} + \mathbb{E}\left[\left(\sum_{i=1}^n y_i\right)^2\right] - 2\frac{n}{\alpha}\mathbb{E}\left[\sum_{i=1}^n y_i\right] \\
&= \frac{n^2}{\alpha^2} + \mathbb{V}\left[\sum_{i=1}^n y_i\right] + \left(\mathbb{E}\left[\sum_{i=1}^n y_i\right]\right)^2 - 2\frac{n}{\alpha}\mathbb{E}\left[\sum_{i=1}^n y_i\right] \\
&= \frac{n^2}{\alpha^2} + n\mathbb{V}[Y] + n^2\mathbb{E}^2[Y] - 2\frac{n^2}{\alpha}\mathbb{E}[Y] \\
\mathbb{E}\left[\left(\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha,\theta)\right)^2\right] &= \frac{n}{\alpha^2}
\end{aligned}$$

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\mathcal{L}(x_i;\alpha,\theta)\right)^2\right] = \frac{n^2\alpha^2}{\theta^2}$$

$$\mathbb{E}\left[\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha,\theta)\,.\,\frac{\partial}{\partial\theta}\mathcal{L}(x_i;\alpha,\theta)\right] = \frac{n^2\alpha}{\theta}\mathbb{E}[Y] - \frac{n^2\alpha}{\theta\alpha} = 0$$

$$\begin{aligned}
I_X(\alpha,\theta) &= \mathbb{E}\left[\nabla_{\alpha,\theta}\mathcal{L}(x_i;\alpha,\theta)^t.\nabla_{\alpha,\theta}\mathcal{L}(x_i;\alpha,\theta)\right] \\
&= \begin{pmatrix} \mathbb{E}\left[\left(\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha,\theta)\right)^2\right] & \mathbb{E}\left[\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha,\theta)\,.\,\frac{\partial}{\partial\theta}\mathcal{L}(x_i;\alpha,\theta)\right] \\ \mathbb{E}\left[\frac{\partial}{\partial\theta}\mathcal{L}(x_i;\alpha,\theta)\,.\,\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha,\theta)\right] & \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\mathcal{L}(x_i;\alpha,\theta)\right)^2\right] \end{pmatrix} \\
I_X(\alpha,\theta) &= \begin{pmatrix} n\alpha^{-2} & 0 \\ 0 & n^2\alpha^2\theta^{-2} \end{pmatrix}
\end{aligned}$$

1.4 Question 4

Estimateurs.

$\frac{\partial^2}{\partial^2\theta}\mathcal{L}(x_i;\alpha,\theta) > 0$ donc $\mathcal{L}(x_i;\alpha,\theta)$ est convexe.
$\mathcal{L}(x_i;\alpha,\theta)$ est maximale pour $\hat{\theta}_{EMV} = max_{j=1}^{n} X_j$ car $\frac{\alpha}{\theta^\alpha}$ est décroissante.
$\hat{\alpha}_{EMV} = \tilde{\alpha}_{EMV}$ avec $\theta : \hat{\theta}_{EMV}$ car $\frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha,\theta) = \frac{\partial}{\partial\alpha}\mathcal{L}(x_i;\alpha)$. Ainsi:

$$\hat{\theta}_{EMV} = max_{j=1}^{n} X_j \qquad \hat{\alpha}_{EMV} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} ln\frac{\hat{\theta}_{EMV}}{X_i}}$$

2 Modèle bayesien

```r
n <- length(X)
```

2.1 Question 5

Loi de Jeffreys.

$$\sqrt{\det(I_X(\alpha,\theta))} = \frac{n\sqrt{n}}{\theta}$$
$$\pi(\alpha,\theta) \propto \frac{1}{\theta}$$

$$L(x|\alpha,\theta)\pi(\alpha,\theta) \propto \alpha^n \theta^{-n\alpha-1} \prod_{i=1}^{n} x_i^{\alpha-1} 1_{0\leq x_i \leq \theta}$$

Comme on a l'indicateur $1_{0\leq x_i\leq\theta}$, $\theta \in [max_j X_j, +\infty[$ et $\alpha > 0$.

On pose s, t et $S_{\alpha,\theta}$ tel que:

$$\begin{aligned}
s &= \sum_{i=1}^{n} ln\, x_i \\
t &= max_j\, x_j \\
S_{\alpha,\theta} &= \int_{\alpha,\theta} L(x|\alpha,\theta)\pi(\alpha,\theta)\, d(\alpha,\theta)
\end{aligned}$$

$$\begin{aligned}
S_\theta &= \int_{t}^{+\infty} L(x|\alpha,\theta)\pi(\alpha,\theta)\, d\theta \\
&= \alpha^n \prod_{i=1}^{n} x_i^{\alpha-1} \int_{t}^{+\infty} \theta^{-n\alpha-1}\, d\theta \\
&= \alpha^{n-1} \prod_{i=1}^{n} x_i^{\alpha-1} \frac{1}{nt^{n\alpha}} \\
S_\theta &= \alpha^{n-1} e^{(\alpha-1)s} \frac{1}{nt^{n\alpha}}
\end{aligned}$$

$$\begin{aligned}
S_{\alpha,\theta} &= \int_{0}^{+\infty} S_\theta\, d\alpha \\
&= \frac{1}{n} e^{-s} \int_{0}^{+\infty} \frac{1}{t^{n\alpha}} \alpha^{n-1} e^{\alpha s} d\alpha \\
S_{\alpha,\theta} &= \frac{1}{n} e^{-s}\, \Gamma(n)\, (-s)^{-n} \left(1 - \frac{n\, ln\, t}{s}\right)^{-n}
\end{aligned}$$

$$\begin{aligned}
S_{\alpha,\theta}^{-1} &= n\, e^{\sum_{i=1}^{n} ln\, x_i} \left(\sum_{i=1}^{n} ln\, x_i^{-1}\right)^n \frac{\left[-\left(n\, ln\, max_j\, x_j - \sum_{i=1}^{n} ln\, x_i\right) \left(\sum_{i=1}^{n} ln\, x_i\right)^{-1}\right]^n}{\Gamma(n)} \\
&= n \prod_{i=1}^{n} x_i \left(\sum_{i=1}^{n} ln\, x_i^{-1}\right)^n \frac{\left(\sum_{i=1}^{n} ln \frac{max_j\, x_j}{x_i}\right)^n \left(\sum_{i=1}^{n} ln\, x_i^{-1}\right)^{-n}}{\Gamma(n)} \\
S_{\alpha,\theta}^{-1} &= n \prod_{i=1}^{n} x_i \frac{\left(\sum_{i=1}^{n} ln \frac{max_j\, x_j}{x_i}\right)^n}{\Gamma(n)}
\end{aligned}$$

Alors on obtient:

$$\begin{aligned}
\pi(\alpha,\theta|x) &= L(x|\alpha,\theta)\, \pi(\alpha,\theta)\, S_{\alpha,\theta}^{-1} \\
&= \frac{\alpha^n}{\theta^{n\alpha+1}} \frac{\left(\prod_{i=1}^{n} x_i\right)^\alpha}{\prod_{i=1}^{n} x_i} n \prod_{i=1}^{n} x_i \frac{\left(\sum_{i=1}^{n} ln \frac{max_j\, x_j}{x_i}\right)^n}{\Gamma(n)} 1_{\theta \geq max_j\, x_j} \\
\pi(\alpha,\theta|x) &= \frac{\alpha^n}{\theta^{n\alpha+1}}\, n \left(\prod_{i=1}^{n} x_i\right)^\alpha \frac{\left(\sum_{i=1}^{n} ln \frac{max_j\, x_j}{x_i}\right)^n}{\Gamma(n)} 1_{\theta \geq max_j\, x_j}
\end{aligned}$$

2.2 Question 6

Comme $ln(x)$ est strictement croissante, on a la relation:
$(\hat{\alpha},\hat{\theta})_{MAP} = argmax_{\alpha,\theta}\pi(\alpha,\theta|x) = argmax_{\alpha,\theta}\, ln\, \pi(\alpha,\theta|x)$.

$$\begin{aligned} ln\,\pi(\alpha,\theta|x) &= ln\,n + \alpha\sum_{i=1}^{n} ln\,x_i + n\,ln\left(\sum_{i=1}^{n} ln\frac{max\,x}{x_i}\right) \\ &+ n\,ln\,\alpha - ln\,\Gamma n - n\alpha\,ln\,\theta - ln\,\theta \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial\alpha} ln\,\pi(\alpha,\theta|x) &= \sum_{i=1}^{n} ln\,x_i + \frac{n}{\alpha} - n\,ln\,\theta \\ \frac{\partial}{\partial\theta} ln\,\pi(\alpha,\theta|x) &= -\frac{n\alpha+1}{\theta} \\ \nabla_{\alpha,\theta} ln\,\pi(\alpha,\theta|x) = 0_{\mathbb{R}^2} &\Leftrightarrow \begin{cases} \alpha & = -\frac{1}{n} \\ \theta & = e^{-n+\frac{1}{n}\sum_{i=1}^{n} ln\,x_i} \end{cases} \end{aligned}$$

Or on devrait avoir $\alpha > 0$ et $\theta \geq max_j\,x_j$ donc:

$$\begin{cases} \sum_{i=1}^{n} ln\,x_i + \frac{n}{\alpha} - n\,ln\,\theta & = 0 \\ \alpha & > 0 \\ \theta & \geq max_j\,x_j \end{cases} \Leftrightarrow \begin{cases} \alpha & = \frac{n}{\sum_{i=1}^{n} ln\frac{\theta}{x_i}} \\ \theta & = max_j\,x_j \end{cases}$$

$$(\hat{\alpha},\hat{\theta})_{MAP} = \left(\frac{n}{\sum_{i=1}^{n} ln\frac{max_j\,x_j}{x_i}}, max_j\,x_j\right)$$

Valeurs pour le jeu de données `infarctus`.

```
t <- max(X)
a <- n/(sum(log(t/X)))
hat <- c(a,t)
hat
```

```
## [1]  2.370016 99.810744
```

2.3 Question 7

$$\pi(\alpha,\theta) \propto \frac{\alpha^3 e^{-2\alpha}}{\theta}$$

$$L(x|\alpha,\theta)\pi(\alpha,\theta) \propto \alpha^{n+3}\theta^{-n\alpha-1}e^{-2\alpha}\prod_{i=1}^{n} x_i^{\alpha-1} 1_{0\leq x_i\leq\theta}$$

Comme à la question 5, on garde les variables s et t.

$$\begin{aligned}
S_\theta &= \alpha^{n+3} e^{-2\alpha} \prod_{i=1}^{n} x_i^{\alpha-1} \int_t^{+\infty} \theta^{-n\alpha-1}\, d\theta \\
&= \alpha^{n+2} e^{-2\alpha} \prod_{i=1}^{n} x_i^{\alpha-1} \frac{1}{nt^{n\alpha}} \\
S_\theta &= \alpha^{n+2} e^{(\alpha-1)s-2\alpha} \frac{1}{nt^{n\alpha}}
\end{aligned}$$

$$\begin{aligned}
S_{\alpha,\theta} &= \int_0^{+\infty} S_\theta \, d\alpha \\
&= \frac{1}{n} e^{-s} \int_0^{+\infty} \frac{1}{t^{n\alpha}} \alpha^{n+2} e^{(s-2)\alpha}\, d\alpha \\
S_{\alpha,\theta} &= \frac{1}{n} e^{-s}\, \Gamma(n+3)\, (2-s)^{n+3} \left(1 + \frac{n\, ln\, t}{2-s}\right)^{n+3}
\end{aligned}$$

$$\begin{aligned}
S_{\alpha,\theta}^{-1} &= n\, e^s\, (2-s)^{n+3} \frac{(2-s+n\, ln\, t)^{n+3} (2-s)^{-n-3}}{\Gamma(n+3)} \\
&= n \prod_{i=1}^{n} x_i \frac{(2 - \sum_{i=1}^{n} ln\, x_i - \sum_{i=1}^{n} ln\, max_j\, x_j)^{n+3}}{\Gamma(n+3)} \\
S_{\alpha,\theta}^{-1} &= n \prod_{i=1}^{n} x_i \frac{\left(2 + \sum_{i=1}^{n} ln \frac{max_j\, x_j}{x_i}\right)^{n+3}}{\Gamma(n+3)}
\end{aligned}$$

Ainsi, on obtient:

$$\begin{aligned}
\pi(\alpha,\theta|x) &= L(x|\alpha,\theta)\, \pi(\alpha,\theta)\, S_{\alpha,\theta}^{-1} \\
&= \frac{\alpha^{n+3} e^{-2\alpha}}{\theta^{n\alpha+1}} \frac{(\prod_{i=1}^{n} x_i)^\alpha}{\prod_{i=1}^{n} x_i} n \prod_{i=1}^{n} x_i \, \frac{\left(2 + \sum_{i=1}^{n} ln \frac{max_j\, x_j}{x_i}\right)^{n}}{\Gamma(n+3)} 1_{\theta \geq max_j\, x_j} \\
\pi(\alpha,\theta|x) &= \frac{\alpha^{n+3} e^{-2\alpha}}{\theta^{n\alpha+1}}\, n \left(\prod_{i=1}^{n} x_i\right)^\alpha \frac{\left(2 + \sum_{i=1}^{n} ln \frac{max_j\, x_j}{x_i}\right)^{n+3}}{\Gamma(n+3)} 1_{\theta \geq max_j\, x_j}
\end{aligned}$$

2.4 Question 8

Comme à la question 6, on évaluera donc uniquement la dérivée partielle par rapport à α et on prendra la plus petite valeur que θ peut prendre pour avoir le dénominateur minimal:

$(\tilde{\alpha}, \tilde{\theta})_{MAP} = argmax_{\alpha,\theta} \pi(\alpha,\theta|x) = argmax_{\alpha,\theta}\, ln\, \pi(\alpha,\theta|x)$.

$$ln\,\pi(\alpha,\theta|x) = ln\,n + \alpha\sum_{i=1}^{n} ln\,x_i + (n+3)\,ln\left(2+\sum_{i=1}^{n} ln\frac{max\,x}{x_i}\right) + (n+3)\,ln\,\alpha - 2\alpha - n\alpha\,ln\,\theta - ln\,\theta$$

$$\frac{\partial}{\partial\alpha} ln\,\pi(\alpha,\theta|x) = \sum_{i=1}^{n} ln\,x_i + \frac{n+3}{\alpha} - n\,ln\,\theta\, - 2$$

$$\begin{cases} \sum_{i=1}^{n} ln\,x_i + \frac{n+3}{\alpha} - n\,ln\,\theta\, - 2 & = 0 \\ \alpha & > 0 \\ \theta & \geq max_j\,x_j \end{cases} \Leftrightarrow \begin{cases} \alpha & = \frac{n+3}{2+\sum_{i=1}^{n} ln\frac{\theta}{x_i}} \\ \theta & = max_j\,x_j \end{cases}$$

$$(\tilde{\alpha},\tilde{\theta})_{MAP} = \left(\frac{n+3}{2+\sum_{i=1}^{n} ln\frac{max_j\,x_j}{x_i}}, max_j\,x_j\right)$$

```
a <- (n+3)/(2+sum(log(t/X)))
tilde <- c(a,t)
tilde
```

```
## [1]  2.361845 99.810744
```

2.5 Question 9

La machine étant incapable de calculer/représenter $\Gamma(n+3)$ avec $n =$, ~~on choisira un échantillon aléatoire parmi nos individus~~.

2.6 Question 10