MSc. Data Science & Artificial Intelligence

Refreshers: Basic algebra for data analysis

Dr. Marco CORNELI

Midterm

Author: Joris LIMONIER

joris.limonier@hotmail.fr

Date: November 26, 2021

Contents

1 Exercise 1

1.1 Question (a)

Unbiasedness First, we prove that $\hat{F}_N(u)$ is unbiased, that is $\mathbb{E}\left[\hat{F}_N(u)\right] = F_N(u)$.

$$\begin{aligned}
\mathbb{E}\left[\hat{F}_N(u)\right] &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{]-\infty,u]}(X_i)\right] && \textit{linearity of the expected value} \\
&= \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[\mathbf{1}_{]-\infty,u]}(X_i)\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} \mathbb{P}(X_i \leq u) && X_i \textit{ are i.i.d.} \\
&= \frac{1}{N} N\mathbb{P}(X_1 \leq u) \\
&= \mathbb{P}(X_1 \leq u) \\
&= F(u)
\end{aligned}$$

Consistency Consistency of $\hat{F}_N(u)$ comes from the law of large numbers: the empirical mean converges in probability for the true mean.

1.2 Question (b)

Let $p_{\bar{u}} := \mathbb{P}(X_i \geq \bar{u})$. We have:

$$\begin{aligned}
p_{\bar{u}} &:= \mathbb{P}(X_i \geq \bar{u}) \\
&= 1 - \mathbb{P}(X_i \leq \bar{u}) + \mathbb{P}(X_i = \bar{u}) && F \textit{ continuous} \\
&= 1 - \mathbb{P}(X_i \leq \bar{u}) \\
&= 1 - F(\bar{u})
\end{aligned}$$

Now we plug in our estimator for $F(\bar{u})$:

$$\hat{p}_{\bar{u}} = 1 - \hat{F}_N(\bar{u})$$

1.3 Question (c)

Let $\alpha = 0.025$, therefore $z_{1-\alpha/2} \approx 1.96$.
Let

$$\mu_{p_{\bar{u}}} := \frac{1}{N} \sum_{i=1}^{N} \hat{p}_{\bar{u}}$$

and

$$\sigma_{p_{\bar{u}}} := \sqrt{\sum_{i=1}^{N} \left(\hat{F}_N(\bar{u}) - \mu_{p_{\bar{u}}} \right)^2}$$

Then by the CLT, we have that:

$$\mathbb{P}\left(z_{\alpha/2} < \frac{\hat{p}_{\bar{u}} - \mu_{p_{\bar{u}}}}{\sigma_{p_{\bar{u}}}/\sqrt{n-1}} < z_{1-\alpha/2} \right) \approx 1 - \alpha$$

1.4 Question (d)

1. Let *nb_boot* be the number of bootstrap replicates you want to perform
2. *store_boot* = [ ]
3. for b in 1, ..., *nb_boot*:
   (a) $s :=$ sample N times with replacement from $\{X_1, \ldots, X_N\}$
   (b) Compute $\hat{F}_N^*(\bar{u})$ on s
   (c) Compute $\hat{p}_{\bar{u}}^* = 1 - \hat{F}_N^*(\bar{u})$ from the $\hat{F}_N^*(\bar{u})$ that was just computed
   (d) Append $\hat{p}_{\bar{u}}^*$ to *store_boot*
4. Define α as desired.
5. Define $\hat{p}_{\bar{u},\alpha/2}^*$ as the $\frac{\alpha}{2}$-th fractile over *store_boot*
6. Define $\hat{p}_{\bar{u},1-\alpha/2}^*$ as the $\frac{1-\alpha}{2}$-th fractile over *store_boot*
7. Define $\hat{p}_{\bar{u}} = 1 - \hat{F}_N(\bar{u})$.
8. The confidence interval is given by

$$\left[2\hat{p}_{\bar{u}} - \hat{p}_{\bar{u},1-\alpha/2}^*, 2\hat{p}_{\bar{u}} - \hat{p}_{\bar{u},\alpha/2}^* \right]$$

2 Exercise 2

2.1 Question (a)

$$
\begin{aligned}
\mathbb{E}\left[X_i\right] &= \int_0^{+\infty} x f_\theta(x) dx \\
&= \int_0^{+\infty} x\theta e^{-\theta x} dx \\
& \hspace{20em} \textit{(by parts)} \\
&= \left[-xe^{-\theta x}\right]_0^{+\infty} - \int_0^{+\infty} -e^{-\theta x} dx \\
&= [0-0] + \frac{1}{-\theta}\left[e^{-\theta x}\right]_0^{+\infty} \\
&= \frac{1}{-\theta}[0-1] \\
&= \frac{1}{\theta}
\end{aligned}
$$

Therefore we choose $\hat{\theta}_{MM} := 1/\bar{X}$, with $\bar{X} := \frac{1}{N}\sum_{i=1}^N X_i$.

2.2 Question (b)

Recall that the likelihood function is given by:

$$
\mathcal{L}(\theta) := \prod_{i=1}^N f_\theta\left(X_i\right)
$$

therefore in our case, we have $\forall x \geq 0$:

$$
\begin{aligned}
\mathcal{L}(\theta) &:= \prod_{i=1}^N f_\theta\left(X_i\right) \\
&= \prod_{i=1}^N \theta e^{-\theta X_i}
\end{aligned}
$$

Now since $\log : \mathbb{R}_{>0} \to \mathbb{R}$ is an increasing function, maximizing the likelihood or its logarithm are equivalent tasks. We have:

$$\begin{aligned}
\log \mathcal{L}(\theta) &= \log \left[\prod_{i=1}^{N} \theta e^{-\theta X_i} \right] \\
&= \sum_{i=1}^{N} \log \left[\theta e^{-\theta X_i} \right] \\
&= [N \log \theta] - \theta \sum_{i=1}^{N} X_i
\end{aligned}$$

which we differentiate with respect to θ and set to 0, in order to find the maximum likelihood estimator:

$$\begin{aligned}
&\frac{\partial}{\partial \theta} \log \mathcal{L}(\theta) \Big|_{\theta=\theta_{ML}} = 0 \\
\implies &\frac{\partial}{\partial \theta} \left[[N \log \theta] - \theta \sum_{i=1}^{N} X_i \right] \Big|_{\theta=\theta_{ML}} = 0 \\
\implies &\frac{N}{\theta_{ML}} - \sum_{i=1}^{N} X_i = 0 \\
\implies &\theta_{ML} = \frac{N}{\sum_{i=1}^{N} X_i} \\
\implies &\theta_{ML} = 1/\bar{X}
\end{aligned}$$

2.3 Question (c)

We know that we have the following:

$$I_N(\theta) = N \cdot I(\theta) \tag{1}$$

with

$$I(\theta) := -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X) \right] \tag{2}$$

We compute the components one after the other:

$$\begin{aligned}\frac{\partial^2}{\partial\theta^2}\log f_\theta(X) &= \frac{\partial}{\partial\theta}\left[(\log\theta) - \theta X\right] \\ &= \frac{1}{\theta} - X\end{aligned}$$

therefore:

$$\begin{aligned}I(\theta) &= -\mathbb{E}\left[\frac{1}{\theta} - X\right] \\ &= -\int_0^{+\infty}\left[\frac{1}{\theta} - x\right] f_\theta(x)dx \\ &= -\int_0^{+\infty}\left[\frac{1}{\theta} - x\right]\theta e^{-\theta x}dx \\ &= -\int_0^{+\infty}\left[1 - \theta x\right] e^{-\theta x}dx \\ & \qquad\qquad\qquad\qquad\qquad\qquad \textit{(by parts)} \\ &= \left[(1-\theta x)\frac{-1}{\theta}e^{-\theta x}\right]_0^{+\infty} - \int_0^{+\infty} -\theta\frac{-1}{\theta}e^{-\theta x}dx \\ &= \left[(x - \frac{1}{\theta})e^{-\theta x}\right]_0^{+\infty} - \int_0^{+\infty} e^{-\theta x}dx \\ &= [0-0] - \frac{-1}{\theta}\left[e^{-\theta x}\right]_0^{+\infty} \\ &= -\frac{1}{\theta}[0-1] \\ &= \frac{1}{\theta}\end{aligned}$$

thus $I_N(\theta) = \frac{N}{\theta}$.