MSc. Data Science & Artificial Intelligence

Statistical Inference Theory

Dr. Christine MALOT

Final project

Author: Joris LIMONIER

joris.limonier@gmail.com

Due: February 24, 2022

Contents

1 Exercise

Let $X_1, \ldots, X_n$ be Independent and Identically Distributed (iid) random variables with density:

$$f_\theta(x) = (k+1)\theta^{-k-1}x^k \mathbf{1}_{[0,\theta]}(x) \tag{1}$$

where $k \in \mathbb{N}_{\geq 0}$ and $\theta > 0$ is unknown.
Let $m := n(k+1)$.

1.1 Question 1

Let $\mathcal{L}$ be the likelihood function. Let $x_1, \ldots, x_n$ be a sample drawn from $X_1, \ldots, X_n$.
We want to find $\hat{\theta}$ maximizing $\mathcal{L}$, that is:

$$\hat{\theta} = \arg\max_{\theta \in \mathbb{R}_{>0}} \mathcal{L}(\theta) \tag{2}$$

and we have:

$$\begin{aligned}
\mathcal{L}(\theta) &= \prod_{i=1}^{n} f_\theta(x_i) \\
&= \prod_{i=1}^{n} (k+1)\theta^{-k-1}x_i^k \mathbf{1}_{[0,\theta]}(x_i) \\
&= (k+1)^n \theta^{-n(k+1)} \prod_{i=1}^{n} x_i^k \mathbf{1}_{[0,\theta]}(x_i)
\end{aligned}$$

Since we have that:

$$\exists\, 1 \leq i \leq n,\ \theta < x_i \implies \mathcal{L}(\theta) = 0$$

We can enforce:

$$\theta \geq x_{\max} := \max\{x_1, \ldots, x_n\} \tag{3}$$

which yields:

$$\hat{\theta} = \arg\max_{\theta \in \mathbb{R}_{\geq x_{\max}}} (k+1)^n \theta^{-n(k+1)} \prod_{i=1}^{n} x_i^k$$

Now, since $\mathcal{L}$ is decreasing (with respect to θ), maximizing it implies choosing θ as small as possible. By (3) however, we have a lower bound on θ. This results in:

$$\hat{\theta} = x_{\max}$$

1.2 Question 2

The bias of $\hat{\theta}$ is defined as:

$$\text{bias}(\hat{\theta}) := \mathbb{E}[\hat{\theta}] - \theta$$

Let $X_{\max} := \max\{X_1, \ldots, X_n\}$. We start by computing $\mathbb{P}(\hat{\theta} \le t)$:

$$\begin{aligned}
\mathbb{P}(\hat{\theta} \le t) &= \mathbb{P}(X_{\max} \le t) \\
&= \mathbb{P}\left(X_1 \le t, \ldots X_n \le t\right) \\
&= \prod_{i=1}^{n} \mathbb{P}\left(X_i \le t\right) && \textit{(independence)} \\
&= \mathbb{P}\left(X_1 \le t\right)^n && \textit{(identical distribution)}
\end{aligned}$$

We compute the CDF of X_1:

$$\begin{aligned}
\mathbb{P}\left(X_1 \le t\right) &= \int_{-\infty}^{t} f_\theta(s) ds \\
&= \int_{-\infty}^{t} (k+1)\theta^{-k-1} s^k \mathbf{1}_{[0,\theta]}(s) ds
\end{aligned}$$

The indicator function tells us that:

$$\mathbb{P}\left(X_1 \le t\right) = \begin{cases} 0 & t < 0 \\ 1 & t > \theta \end{cases}$$

So we can now focus on the case $t \in [0, \theta]$:

$$\begin{aligned}
\mathbb{P}\left(X_1 \le t\right) &= (k+1)\theta^{-k-1} \int_0^t s^k ds \\
&= (k+1)\theta^{-k-1} \left[\frac{s^{k+1}}{k+1}\right]_0^t \\
&= \theta^{-k-1} \left[s^{k+1}\right]_0^t \\
&= \left[\frac{t}{\theta}\right]^{k+1}
\end{aligned}$$

So to recap, the CDF of X_1 is:

$$\mathbb{P}\left(X_1 \le t\right) = \begin{cases} 0 & t < 0 \\ \left[\frac{t}{\theta}\right]^{k+1} & t \in [0, \theta] \\ 1 & t > \theta \end{cases}$$

and therefore, the CDF of $\hat{\theta} = X_{\max}$ is given by:

$$
\begin{aligned}
\mathbb{P}(\hat{\theta} \leq t) &= \mathbb{P}\left(X_1 \leq t\right)^n \\
&= \begin{cases} 0 & t < 0 \\ \left[\frac{t}{\theta}\right]^{n(k+1)} & t \in [0, \theta] \\ 1 & t > \theta \end{cases} \\
&= \begin{cases} 0 & t < 0 \\ \left[\frac{t}{\theta}\right]^{m} & t \in [0, \theta] \\ 1 & t > \theta \end{cases}
\end{aligned}
$$

By differentiating the CDF of $\hat{\theta} = X_{\max}$, we can obtain its PDF:

$$
\begin{aligned}
\mathbb{P}(\hat{\theta} = t) &= \frac{d}{dt}\mathbb{P}(\hat{\theta} \leq t) \\
&= \begin{cases} \frac{d}{dt}0 & t < 0 \\ \frac{d}{dt}\left[\frac{t}{\theta}\right]^{m} & t \in [0, \theta] \\ \frac{d}{dt}1 & t > \theta \end{cases} \\
&= \begin{cases} 0 & t < 0 \\ m\frac{t^{m-1}}{\theta^m} & t \in [0, \theta] \\ 0 & t > \theta \end{cases} \\
&= m\frac{t^{m-1}}{\theta^m}\mathbf{1}_{[0,\theta]}(t)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[\hat{\theta}] &= \mathbb{E}[X_{\max}] \\
&= \int_{-\infty}^{\infty} t\mathbb{P}(X_{\max} = t)dt \\
&= \int_{-\infty}^{\infty} tm\frac{t^{m-1}}{\theta^m}\mathbf{1}_{[0,\theta]}(t)dt \\
&= \int_{0}^{\theta} tm\frac{t^{m-1}}{\theta^m}dt \\
&= \frac{m}{\theta^m}\int_{0}^{\theta} t^m dt \\
&= \frac{m}{\theta^m}\left[\frac{t^{m+1}}{m+1}\right]_0^{\theta} \\
&= \frac{m}{\theta^m}\frac{\theta^{m+1}}{m+1} \\
&= \frac{m}{m+1}\theta
\end{aligned}
$$

Now, from the definition of the bias:

$$\begin{aligned}
\text{bias}(\hat{\theta}) &= \mathbb{E}[\hat{\theta}] - \theta \\
&= \frac{m}{m+1}\theta - \theta \\
&= \frac{-1}{m+1}\theta
\end{aligned}$$

1.3 Question 3

We want to find λ_1 and λ_2 such that $\lambda_1\hat{\theta}$ is unbiased and $\lambda_2\hat{\theta}$ has the smallest quadratic error.

1.3.1 Unbiased

$$\begin{aligned}
&\text{bias}(\lambda_1\hat{\theta}) = 0 \\
\implies &\mathbb{E}[\lambda_1\hat{\theta}] - \theta = 0 \\
\implies &\lambda_1\mathbb{E}[\hat{\theta}] - \theta = 0 \\
\implies &\lambda_1\frac{m}{m+1}\theta - \theta = 0 \\
\implies &\lambda_1 = \theta\frac{m+1}{m\theta} \\
\implies &\lambda_1 = \frac{m+1}{m}
\end{aligned}$$

1.3.2 Smallest quadratic error

We define the quadratic error (mean squared error) is defined as follows:

$$MSE(\lambda_2\hat{\theta}) := \mathbb{E}\left[\left(\lambda_2\hat{\theta} - \theta\right)^2\right]$$

we want to minimize it:

$$\begin{aligned}
MSE(\lambda_2\hat{\theta}) &:= \mathbb{E}\left[\left(\lambda_2\hat{\theta} - \theta\right)^2\right] \\
&= \mathbb{E}\left[\hat{\theta}^2\right]\lambda_2^2 - 2\theta\mathbb{E}\left[\hat{\theta}\right]\lambda_2 + \theta^2 \\
&= \mathbb{E}\left[\hat{\theta}^2\right]\lambda_2^2 - \frac{2m\theta^2}{m+1}\lambda_2 + \theta^2
\end{aligned}$$

We compute $\mathbb{E}\left[\hat{\theta}^2\right]$:

$$\begin{aligned}
\mathbb{E}\left[\hat{\theta}^2\right] &= \int_{-\infty}^{\infty} t^2 \mathbb{P}(\underbrace{\hat{\theta}}_{=X_{\max}} = t)dt \\
&= \int_{-\infty}^{\infty} t^2 m \frac{t^{m-1}}{\theta^m} \mathbf{1}_{[0,\theta]}(t)dt \\
&= \int_0^{\theta} t^2 m \frac{t^{m-1}}{\theta^m} dt \\
&= \frac{m}{\theta^m} \int_0^{\theta} t^{m+1} dt \\
&= \frac{m}{\theta^m(m+2)} \left[t^{m+2}\right]_0^{\theta} \\
&= \frac{m\theta^2}{m+2}
\end{aligned}$$

Therefore, the quadratic error becomes:

$$\begin{aligned}
MSE(\lambda_2\hat{\theta}) &= \mathbb{E}\left[\hat{\theta}^2\right]\lambda_2^2 - \frac{2m\theta^2}{m+1}\lambda_2 + \theta^2 \\
&= \frac{m\theta^2}{m+2}\lambda_2^2 - \frac{2m\theta^2}{m+1}\lambda_2 + \theta^2
\end{aligned}$$

We have a parabola, with the coefficient of the term of degree 2 that is positive, so we know that the minimum will occur when the derivative becomes 0. We differentiate and equate to 0:

$$\begin{aligned}
& \frac{\partial}{\partial \lambda_2} MSE(\lambda_2\hat{\theta}) = 0 \\
\implies & \frac{\partial}{\partial \lambda_2}\left[\frac{m\theta^2}{m+2}\lambda_2^2 - \frac{2m\theta^2}{m+1}\lambda_2 + \theta^2\right] = 0 \\
\implies & \frac{2m\theta^2}{m+2}\lambda_2 - \frac{2m\theta^2}{m+1} = 0 \\
\implies & \frac{1}{m+2}\lambda_2 = \frac{1}{m+1} \\
\implies & \lambda_2 = \frac{m+2}{m+1}
\end{aligned}$$

We can evaluate the MSE at the minimum:

$$\begin{aligned}
MSE(\lambda_2\hat{\theta}) &= \frac{m\theta^2}{m+2}\left(\frac{m+2}{m+1}\right)^2 - \frac{2m\theta^2}{m+1}\left(\frac{m+2}{m+1}\right) + \theta^2 \\
&= \frac{m\theta^2(m+2)}{(m+1)^2} - \frac{2m\theta^2(m+2)}{(m+1)^2} + \theta^2 \\
&= \theta^2 - \frac{m\theta^2(m+2)}{(m+1)^2} \\
&= \theta^2\left[1 - \frac{m(m+2)}{(m+1)^2}\right]
\end{aligned}$$