M1 MAS
Statistics

Correction of Exam
November 2024, 12th
Statistics

Exercise 1 Consider a n-sized sample $(X_1, \dots, X_n)$ of iid r.v. such that $X_i \sim \mathcal{N}(\theta, \theta^2)$, with $\theta > 0$ unknown.

We study the two estimators:

$$\hat{\theta}_n = \bar{X}_n \quad , \quad \text{and} \quad \tilde{\theta}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}.$$

1- Are these estimators consistent?

- From the Strong Law of Large Numbers (SLLN), since X_i are iid with $\mathbb{E}[X_i] = \theta < +\infty$, we have $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n \to +\infty]{p.s.} \mathbb{E}[X_i] = \theta$

Hence $\hat{\theta}_n \xrightarrow[n\to+\infty]{p.s.} \theta \Rightarrow \hat{\theta}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \theta \Rightarrow \hat{\theta}_n$ is a consistent estimator of θ.

- $\tilde{\theta}_n$ is the square root of the second-order empirical moment:

$$\tilde{\theta}_n^2 = \frac{1}{n-1}\sum(X_i - \bar{X}_n)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - 2\bar{X}_n \underbrace{\sum_{i=1}^{n} X_i}_{= n\bar{X}_n} + n\bar{X}_n^2\right)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - 2n\bar{X}_n^2 + n\bar{X}_n^2\right] = \frac{1}{n-1}\sum X_i^2 - \frac{n}{n-1}(\bar{X}_n)^2$$

By the SLLN, we have: $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}[X_i^2] = Var(X_i) + (\mathbb{E}[X_i])^2 = 2\theta^2$

$$\bar{X}_n \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}[X_i] = \theta$$

Hence, $\tilde{\theta}_n \xrightarrow[n\to+\infty]{p.s.} \sqrt{2\theta^2 - \theta^2} = \theta$ by continuity. $\tilde{\theta}_n$ is thus consistent.

② - Compute the Fisher information of X_i :

$$I_1(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln L(\theta;X)\right] \text{ with}$$

$$\ln L(\theta;x) = \ln f_X(x;\theta) = \ln\left(\frac{1}{\sqrt{2\pi\theta^2}}e^{-\frac{(x-\theta)^2}{2\theta^2}}\right)$$

$$= -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\theta^2) - \frac{(x-\theta)^2}{2\theta^2} = \ell(\theta;x)$$

$$\rightarrow \frac{\partial}{\partial\theta}\ell(\theta;x) = -\frac{1}{\cancel{2}}\cancel{2\theta}\frac{1}{\theta^{\cancel{2}}} - \left(\frac{-2(x-\theta)\times 2\theta^2 - 4\theta(x-\theta)^2}{4\theta^4}\right)$$

$$= -\frac{1}{\theta} - \frac{1}{4\theta^4}\left(-4\theta^2x + \cancel{4\theta^3} - 4\theta x^2 + 8\theta^2x - \cancel{4\theta^3}\right)$$

$$= -\frac{1}{\theta} - \frac{1}{4\theta^4}\left(4\theta^2x - 4\theta x^2\right) = -\frac{1}{\theta} - \frac{x}{\theta^2} + \frac{x^2}{\theta^3}$$

$$\rightarrow \frac{\partial^2}{\partial\theta^2}\ell(\theta;x) = \frac{\partial}{\partial\theta}\left(-\frac{1}{\theta} - \frac{x}{\theta^2} + \frac{x^2}{\theta^3}\right) = +\frac{1}{\theta^2} - x\times -2\theta^{-3} + x^2\times(-3)\theta^{-4}$$

$$= \frac{1}{\theta^2} + \frac{2x}{\theta^3} - \frac{3x^2}{\theta^4}$$

Thus $I_1(\theta) = -\mathbb{E}\left[\frac{1}{\theta^2} + \frac{2X}{\theta^3} - \frac{3X^2}{\theta^4}\right] = -\left[\frac{1}{\theta^2} + \frac{2}{\theta^3}\mathbb{E}(X) - \frac{3}{\theta^4}\overbrace{\mathbb{E}[X^2]}^{=Var(X)+(\mathbb{E}[X])^2}\right]$

$$= -\left[\frac{1}{\theta^2} + \frac{2}{\theta^3}\times\theta - \frac{3}{\theta^4}(\theta^2+\theta^2)\right] = -\left(\frac{1}{\theta^2} + \frac{2}{\theta^2} - \frac{6}{\theta^2}\right) = \frac{3}{\theta^2}$$

③ - Is $\hat{\theta}_n$ an efficient estimator?

- $\hat{\theta}_n$ is unbiased, and $Var(\hat{\theta}_n) = Var(\bar{X}_n) = \frac{Var(X_i)}{n} = \frac{\theta^2}{n}$
- $\frac{1}{I_n(\theta)} = \frac{1}{nI_1(\theta)} = \frac{\theta^2}{3n} < \frac{\theta^2}{n} = Var(\hat{\theta}_n) \Rightarrow \hat{\theta}_n$ does not reach the Cramer-Rao bound $\Rightarrow$ it is not efficient.

Exercise 2 Consider the sample $(X_1, \dots, X_n)$ of iid random variables, where $X_i \sim \mathcal{N}(0,1)$.

1 - $\forall i = 1, \dots, n$, we know that $X_i^2 \sim \chi_1^2$, and that $\sum_{i=1}^n X_i \sim \chi_n^2$. (Chi-square distribution with n degrees of freedom).

2 - $$\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) = \sum_{i=1}^n X_i^2 - 2\bar{X}_n \sum_{i=1}^n X_i + n\bar{X}_n^2$$
$$= \sum_{i=1}^n X_i^2 - 2n\bar{X}_n^2 + n\bar{X}_n^2 = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2.$$

Now, consider the random vector $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ and $A \in \mathcal{M}_{n\times n}(\mathbb{R})$, such that $AA^t = A^tA = I_n$, where the last column of A is $\begin{pmatrix} \frac{1}{\sqrt{n}} \\ \vdots \\ \frac{1}{\sqrt{n}} \end{pmatrix}$

Let $Z = A^t X$.

3 - What is the distribution of X?

$\forall i$, $X_i \sim \mathcal{N}(0,1)$. Moreover, X_i are iid $\Bigg\}$ $\Rightarrow$ X is a Gaussian vector, i.e. $X \sim \mathcal{N}_n(\mu, \Gamma)$

where $\mu = \mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ and $\Gamma = Var(X) = \begin{pmatrix} Var(X_1) & \cdots & Cov(X_n, X_1) \\ \vdots & \ddots & \\ Cov(X_n, X_1) & \cdots & Var(X_n) \end{pmatrix} = I_n$

Finally, $X \sim \mathcal{N}_n\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}\right)$.

4 - Remind that $Z = A^t X$. $\Rightarrow$ • $\mathbb{E}[Z] = \mathbb{E}[A^t X] = A^t \mathbb{E}[X] = A^t \underset{\text{column vector of } 0}{0_n} = 0_n$

• $Var(Z) = \mathbb{E}[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])^t] = \mathbb{E}[(A^tX - \underbrace{\mathbb{E}[A^tX]}_{=0})(A^tX - \underbrace{\mathbb{E}[A^tX]}_{=0})^t]$

$= \mathbb{E}[A^tX(A^tX)^t] = \mathbb{E}[A^tXX^t(A^t)^t] = \mathbb{E}[A^tXX^tA] = A^t\mathbb{E}[XX^t]A$

$= A^t \, \mathbb{E}\left[(X-\underbrace{\mathbb{E}(X)}_{=0})(X-\underbrace{\mathbb{E}(X)}_{=0})^t\right] A = A^t \, Var(X) A = A^t I_n A = A^t A = I_n$.

- $Z = A^t X$: each component Z_i of the random vector Z is a linear combination of independent random variables $X_1, \dots, X_n$, each $X_i \sim \mathcal{N}(0,1)$.
Hence Z_i is Gaussian, and Z is also Gaussian: $Z \sim \mathcal{N}_n(0_n, I_n)$.

(5) _ We have

$$\sum_{i=1}^n Z_i^2 = Z^t Z = (A^t X)^t (A^t X) = X^t (A^t)^t (A^t X) = X^t A A^t X = X^t X = \sum_{i=1}^n X_i^2$$

(6) _

$$\begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} = A^t X = \begin{pmatrix} a_{11} & \cdots a_{1n-1} & \frac{1}{\sqrt{n}} \\ a_{21} & \cdots & \vdots \\ \vdots & & \vdots \\ a_{n1} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix}^t \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & & \vdots \\ a_{1n-1} & \cdots & \vdots \\ \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

Therefore $Z_n = \frac{1}{\sqrt{n}} X_1 + \frac{1}{\sqrt{n}} X_2 + \dots + \frac{1}{\sqrt{n}} X_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$.

(7) _ We know that:

$$\to \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n X_i^2 - n(\bar{X}_n)^2 \overset{(5)\text{ et }(6)}{=} \sum_{i=1}^n Z_i^2 - \frac{n}{n^2}\left(\sum X_i\right)^2 = \sum_{i=1}^n Z_i - \underbrace{\frac{\left(\sum X_i\right)^2}{n}}_{= Z_n^2} = \sum_{i=1}^{n-1} Z_i^2$$

$$\to \sum_{i=1}^n X_i^2 \sim \chi^2_n, \quad Z_i \sim \mathcal{N}(0,1), \; Z_i \text{ iid}$$

Hence $Z_i^2 \sim \chi^2_1 \Rightarrow \sum_{i=1}^{n-1} Z_i^2 \sim \chi^2_{n-1}$.

Exercise 3 Consider the continuous random variable X with density:

$$f_\theta(x) = c\, e^{-\theta x + 1} \, \mathbb{1}_{x \geq -1}, \text{ with } \begin{cases} \theta > 0 \text{ unknown} \\ c \text{ a constant.} \end{cases}$$

1. We know that f_θ is a density: $\Rightarrow$ it is continuous
$\Rightarrow$ it is positive
$\Rightarrow$ $\int f_\theta = 1$.

$$\int_{\mathcal{X}} \left(c\, e^{-\theta x+1}\right) \mathbb{1}_{x \geq -1}\, dx = 1$$

$$\Updownarrow$$

$$\int_{-1}^{+\infty} c\, e^{-\theta x+1}\, dx = 1 \Leftrightarrow c e \left[- \frac{e^{-\theta x}}{\theta} \right]_{-1}^{+\infty} = 1$$

Hence $c e\left(0 + \frac{e^{\theta}}{\theta}\right) = 1 \Leftrightarrow \frac{c}{\theta} e^{\theta+1} = 1 \Leftrightarrow c = \theta e^{-\theta-1}$.

2. • $\mathbb{E}[X] = \int_{\mathcal{X}} x f_\theta(x)\, dx = \int_{-\infty}^{+\infty} x\, \theta e^{-\theta-1} e^{-\theta x+1} \mathbb{1}_{x \geq -1}\, dx$

$$= \int_{-1}^{+\infty} \theta x\, e^{-\theta(x+1)}\, dx = \theta \int_{-1}^{+\infty} \underbrace{x}_{u} \underbrace{e^{-\theta(x+1)}}_{v'}\, dx = \cancel{\theta}\left[\left[\frac{x e^{-\theta(x+1)}}{\cancel{\theta}} \right]_{-1}^{+\infty} - \int_{-1}^{+\infty} - \frac{e^{-\theta(x+1)}}{\cancel{\theta}}\, dx \right]$$

$$= \left[-x e^{-\theta(x+1)} \right]_{-1}^{+\infty} + \int_{-1}^{+\infty} e^{-\theta(x+1)}\, dx = \left(0 + (-1) e^{0}\right) + \left[- \frac{e^{-\theta(x+1)}}{\theta} \right]_{-1}^{+\infty}$$

$$= -1 + \left(-0 + \frac{1}{\theta}\right)$$

$$= \frac{1}{\theta} - 1.$$

Using the SLLN, we have: since $\begin{cases} X_i \text{ iid} \\ \mathbb{E}[X_i] < \infty \end{cases}$.

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{a.s.} \mathbb{E}[X_i] = \frac{1}{\theta} - 1$. The method of moments leads to define the estimator $\hat{\theta}_n$ such that $\bar{X}_n = \frac{1}{\hat{\theta}_n} - 1 \Leftrightarrow \hat{\theta}_n = \frac{1}{1 + \bar{X}_n}$

3. Let $g(x) = \frac{1}{1+x}$. It is clear that $g(x)$ is a continuous function on $]-1; +\infty[$.
Then by continuity, since $\bar{X}_n \xrightarrow[n\to+\infty]{a.s.} \frac{1}{\theta} - 1$ then $g(\bar{X}_n) \xrightarrow[n\to+\infty]{a.s.} g\left(\frac{1}{\theta} - 1\right)$

i.e. $\hat{\theta}_n = g(\bar{X}_n) \xrightarrow[n\to+\infty]{a.s.} \frac{1}{\cancel{1} + \frac{1}{\theta} \cancel{-1}} = \theta \Rightarrow \hat{\theta}_n$ is thus strongly consistent.

(4) - We seek an exhaustive / complete / sufficient statistic for θ:
Let us look at the likelihood function: $L(\theta;x)$.

$$L(\theta;x) = f_{(X_1,\dots,X_n)}((x_1,\dots,x_n);\theta) \overset{X_i \text{ iid}}{=} \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \theta e^{-1-\theta} e^{-\theta x_i+1} \mathbb{1}_{x_i \geq 1}$$

Hence $L(\theta;x) = \theta^n e^{-n\theta} e^{-\theta \sum_{i=1}^n x_i} \mathbb{1}_{\min(x_i)\geq -1} = (\theta e^{-\theta})^n e^{-\theta\sum_{i=1}^n x_i} \mathbb{1}_{\min(x_i)\geq -1}$

By the factorization theorem of Fisher-Neyman:
t is a sufficient statistic $\Leftrightarrow$ there exist 2 measurable functions f and g such that: $\forall x \in]-1;+\infty[$, $\forall \theta > 0$,

$$L(\theta;x) = g(t(x);\theta)h(x)$$

Here $\begin{cases} h(x) = \mathbb{1}_{\min_i(x_i)\geq -1} \\ g(t(x);\theta) = (\theta e^{-\theta})^n e^{-\theta\sum_{i=1}^n x_i} \end{cases}$

We deduce that $T(X) = \sum_{i=1}^n X_i$ is a sufficient statistic.

(5) - It is given that the density function of $T = \sum_{i=1}^n X_i$ is

$$f_T(t) = \frac{\theta^n (t+n)^{n-1}}{(n-1)!} e^{-\theta(t+n)} \mathbb{1}_{t\geq -n}.$$

Denote $T_n = \hat{\theta}_n = (1+\bar{X}_n)^{-1}$, what is the density h of $\hat{\theta}_n$?

We have $T_n = \dfrac{1}{1+\frac{1}{n}\sum_{i=1}^n X_i} = \dfrac{1}{1+\frac{T}{n}}$

$$F_{T_n}(t) = \mathbb{P}(T_n \leq t) = \mathbb{P}\left(\frac{1}{1+\frac{T}{n}} \leq t\right) = \mathbb{P}\left(1+\frac{T}{n} \geq \frac{1}{t}\right) = \mathbb{P}\left(\frac{T}{n} \geq \frac{1}{t}-1\right)$$

$$= \mathbb{P}\left(T \geq \frac{n(1-t)}{t}\right) = 1 - \mathbb{P}\left(T \leq \frac{n(1-t)}{t}\right) = 1 - F_T\left(\frac{n(1-t)}{t}\right)$$

$$\Rightarrow f_{T_n}(t) = \frac{d}{dt}F_{T_n}(t) = \frac{d}{dt}\left(1 - F_T\left(\frac{n(1-t)}{t}\right)\right) = -\left(\frac{-nt-n(1-t)}{t^2}\right) f_T\left(\frac{n(1-t)}{t}\right)$$

(4)

Therefore $f_{T_n}(t) = \frac{\cancel{nt} + n - \cancel{nt}}{t^2} \frac{\theta^n \left(\frac{n(1-t)}{t} + n\right)^{n-1}}{(n-1)!} e^{-\theta\left(\frac{n(1-t)}{t}+n\right)} \mathbb{1}_{\frac{\cancel{n}(1-t)}{t} \geq -\cancel{n}}$.

We get

$$f_{T_n}(t) = \frac{n}{t^2} \frac{\theta^n \left(\frac{n}{t}\right)^{n-1}}{(n-1)!} e^{-\theta \frac{n}{t}} \mathbb{1}_{\frac{1-t}{t} \geq -1}$$

$$= \frac{1}{t} \times \cancel{\left(\frac{n}{t}\right)} \left(\frac{n}{t}\right)^{n-\cancel{1}} \frac{\theta^n}{(n-1)!} e^{-\frac{n\theta}{t}} \mathbb{1}_{\frac{1}{t} \geq 0}$$

$$= \frac{1}{t} \left(\frac{\theta n}{t}\right)^n \frac{e^{-n\theta/t}}{(n-1)!} \mathbb{1}_{\frac{1}{t} \geq 0} = \mathbb{1}_{t \geq 0}$$

(6) - We indeed have found the same density function:

$$h(z) = \frac{(n\theta)^n}{(n-1)!} z^{-n-1} e^{-n\theta/z} \mathbb{1}_{z \geq 0}.$$

Let us compute $\mathbb{E}[T_n]$: by definition,

$$\mathbb{E}[T_n] = \int_{-\infty}^{+\infty} z\, h(z)\, dz = \int_{-\infty}^{+\infty} z \frac{(n\theta)^n}{(n-1)!} z^{-n-1} e^{-n\theta/z} \mathbb{1}_{z \geq 0}\, dz$$

$$= \int_0^{+\infty} \frac{(n\theta)^n}{(n-1)!} z^{-n} e^{-n\theta/z} dz \overset{z = 1/t}{=} \int_{+\infty}^{0} \frac{(n\theta)^n}{(n-1)!} t^n e^{-n\theta t} \times -\frac{1}{t^2} dt$$

$$= +\int_0^{+\infty} \frac{(n\theta)^n}{(n-1)!} t^{n-2} e^{-n\theta t} dt = \frac{(n\theta)^n}{(n-1)!} \int_0^{+\infty} \underbrace{t^{n-2}}_{u} \underbrace{e^{-n\theta t}}_{v'} dt$$

$$= \frac{(n\theta)^n}{(n-1)!} \left[\left[-t^{n-2} \frac{e^{-n\theta t}}{n\theta} \right]_0^{+\infty} - \int_0^{+\infty} -\frac{e^{-n\theta t}}{n\theta} (n-2) t^{n-3} dt \right]$$

$$= \frac{(n\theta)^n}{(n-1)!} \left[(0-0) + \underbrace{\frac{n-2}{n\theta} \int_0^{+\infty} t^{n-3} e^{-n\theta t} dt}_{\text{almost the same integral as previously!}} \right]$$

thus, iterating the same integration by parts leads to:

$$\mathbb{E}[T_n] = \frac{(n\theta)^n}{(n-1)!} \frac{(n-2)!}{(n\theta)^{n-2}} \int_0^{+\infty} e^{-n\theta t} dt = \frac{(n\theta)^n}{(n-1)!} \frac{(n-2)!}{(n\theta)^{n-2}} \underbrace{\left[- \frac{e^{-n\theta t}}{n\theta} \right]_0^{+\infty}}_{= 1/n\theta}$$

$$= \frac{(n\theta)^2}{(n-1)} \times \frac{1}{\cancel{n\theta}} = \frac{n\theta}{n-1}.$$

7 - If $T_n = \hat{\theta}_n$ were an estimator of θ, T_n would be biased since

$\mathbb{E}[T_n] = \frac{n}{n-1}\theta \neq \theta$. However, it is asymptotically unbiased,

since $\lim_{n \to +\infty} \frac{n}{n-1}\theta = \theta$.

An unbiased estimator would be $T_n' = \frac{n-1}{n} T_n$.