STATISTICS TUTORIAL #3

10/10/2024

Statistics Fall 2024 - TA Session 3

TA: Giacomo Caserta - giacomo.caserta@uniroma2.it
Office Hour: By appointment, office 3D-8, third floor building B

10/10/2024

Problem 1

Prove that $\overline{X}$, the mean of a random sample of size n from a distribution that is $\mathcal{N}(\theta, \sigma^2)$ $(-\infty < \theta < +\infty)$ is, for every known $\sigma^2 > 0$, an efficient estimator of θ.

Problem 2

Let $X_1, X_2, ..., X_n$ be a random sample on X that has a **Gamma** $(\alpha = 4, \beta = \theta)$ distribution with $0 < \theta < \infty$.

(a) Find the MLE of θ.

(b) Find the Fisher information $I(\theta)$.

(c) Show that the MLE of θ, which was derived in point (a), is an efficient estimator of θ.

(d) Obtain the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$.

Problem 3

If $X_1, X_2, ..., X_n$ is a random sample from a distribution with pdf:

$$f(x;\theta) = \begin{cases} \frac{3\theta^3}{(x+\theta)^4} & 0 < x < \infty \\ 0 & elsewhere \end{cases}$$

Show that $Y = 2\overline{X}$ is an unbiased estimator of θ and determine its efficiency.

PROBLEM 1

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \log f(x;\theta) \mid \theta\right]$$

GIVEN THAT THE LIKELIHOOD OF θ GIVEN X IS ALWAYS PROPORTIONAL TO THE PROBABILITY $f(X;\theta)$ THEIR LOGS WILL NECESSARILY DIFFER BY A CONSTANT THAT IS INDEPENDENT OF θ, AND THE DERIVATIVES OF THESE LOGS W.R.T. θ ARE EQUAL. THUS IN THE DEF OF FISHER INFO WE CAN USE $\log L(\theta;X)$ INSTEAD OF $\log f(X;\theta)$.

$$\log f(X;\theta) = -\log\sqrt{2\pi\sigma^2} - \frac{(X-\theta)^2}{2\sigma^2}$$

$$\frac{\partial \log f(X;\theta)}{\partial\theta} = \frac{X-\theta}{\sigma^2}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} = -\frac{1}{\sigma^2}$$

$$I(\theta) = -E\left[\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2}\right] = \frac{1}{\sigma^2}$$

CRAMER-RAO LOWER BOUND $= \dfrac{1}{(m\, I(\theta))} = \dfrac{\sigma^2}{N}$

$$VAR(\hat{\theta}) = \frac{1}{N} VAR(x) = \frac{\sigma^2}{N}$$

SINCE $\bar{X}$ IS UNBIASED FOR θ,

$VAR(\bar{X}) = \dfrac{\sigma^2}{N}$ ATTAINS THE CRLB

HENCE $\bar{X}$ IS AN EFFICIENT ESTIMATOR OF θ.

REFORMULATION FOR IID SETTINGS:

IF $X_1, X_2, \dots, X_N$ ARE I.I.D., THEN:

$$I_N(\theta) = -N E\left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X_N)\right]$$

WHILE, THE FISHER INFO FOR A SINGLE SAMPLE $I(\theta)$ IS:

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

WHICH MEANS THAT:

$$I_N(\theta) = N\, I(\theta)$$

CRAMER - RAO LOWER BOUND

IF $\hat{\theta}$ IS AN UNBIASED ESTIMATOR OF AN UNKNOWN PARAMETER θ BASED ON N INDEPENDENT OBSERVATIONS, THEN

$$VAR(\hat{\theta}) \geq \frac{1}{N\,I(\theta)}$$

WHERE $I(\theta)$ DENOTES THE FISHER INFORMATION OF ONE OBSERVATION.

PROBLEM 2

$$X_i \sim \text{Gamma}(\underset{\alpha}{4}, \underset{\beta}{\theta}) \quad (\text{I.I.D.})$$

$$f(x|\alpha,\beta) = \begin{cases} \dfrac{x^{\alpha-1} e^{-x/\theta}}{\Gamma(\alpha)\,\theta^{\alpha}} & \text{IF } x>0 \\ 0 & \text{OTHERWISE} \end{cases}$$

$$f(x|4,\theta) = \begin{cases} \dfrac{x^{3} e^{-x/\theta}}{\Gamma(4)\,\theta^{4}} & \text{IF } x>0 \\ 0 & \text{OTHERWISE} \end{cases}$$

WE WRITE THE LOG-LIKELIHOOD FUNCTION STARTING FROM THE LIKELIHOOD FUNCTION:

$$L(\alpha, \beta \mid X) =$$

$$\prod_{i=1}^{N} \frac{x_i^3 \, e^{-x_i/\theta}}{\Gamma(4)\,\theta^4}$$

$$= \prod_{i=1}^{N} \frac{x_i^3 \, e^{-x_i/\theta}}{\Gamma(4)\,\theta^4}$$

$$= \frac{1}{\Gamma(4)\,\theta^4} \prod_{i=1}^{N} x_i^3 \, e^{-x_i/\theta}$$

$$= \frac{1}{\Gamma(4)\,\theta^4} \left(\prod_{i=1}^{N} x_i^3 \right) \left(\prod_{i=1}^{N} e^{-x_i/\theta} \right)$$

= TAKE THE LOG:

$$\ell(\theta) = \sum_i \left[-\log \Gamma(4) - 4 \log \theta - 3 \log(x_i) - \frac{x_i}{\theta} \right]$$

$$\ell'(\theta) = \sum_i \left[-\frac{4}{\theta} + \frac{x_i}{\theta^2} \right] =$$

$$= n\left(-4\theta + \bar{x}\right) / \theta^2$$

SET $\ell'(\theta) = 0$

$$\theta = \frac{\bar{x}}{4}$$

THEN $\ell''\left(\frac{\bar{x}}{4}\right) < 0$, THUS, THE MLE OF

θ IS $\hat{\theta} = \frac{\bar{x}}{4}$

2B) FIND $I(\theta)$

THE LOG-PDF WRITES AS FOLLOWS:

$$\log f(x;\theta) = \text{CONSTANT} - 4\log\theta + 3\log(x) - \frac{x}{\theta}$$

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = -\frac{4}{\theta} + \frac{x}{\theta^2}$$

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = \frac{4}{\theta^2} - \frac{2x}{\theta^3}$$

RECALL THAT, WE WROTE THE PDF AS:

$$f(x;\alpha;\beta) = \frac{1}{\Gamma(\alpha)\,\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}$$

$\hookrightarrow$ THE MEAN IS: $E(X) = \alpha\beta$

IN THIS CASE $\alpha = 4$, $\beta = \theta$

THE MEAN BECOMES $E(x) = 4\theta$

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

$$= E\left[\frac{2x}{\theta^3} - \frac{4}{\theta^2}\right]$$

$$= \frac{E[2x]}{\theta^3} - \frac{4}{\theta^2}$$

$$= \frac{2E[X]}{\theta^3} - \frac{4}{\theta^2}$$

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

$$= \frac{8}{\theta^2} - \frac{4}{\theta^2} = \frac{4}{\theta^2}$$

$$VAR(X) = \alpha \beta^2 \longrightarrow 4\theta^2$$

$$VAR(\hat{\theta}) = VAR\left(\frac{\bar{X}}{4}\right) =$$

$$= \frac{1}{16}\left[n \sigma^2 \right] =$$

$$= \frac{1}{16N}\left[N\, 4\, \theta^2 \right] =$$

$$= \frac{1}{4N} N \theta^2 = \frac{\theta^2}{4N}$$

SINCE $VAR(\hat{\theta})$ AND CRLB ARE THE SAME, $\hat{\theta}$ IS AN EFFICIENT ESTIMATOR FOR θ.

D) OBTAIN THE ASYMPTOTIC DISTRIBUTION OF

$$\sqrt{N}(\hat{\theta}-\theta)$$

$\boxed{\text{TH}}$ ASSUME $X_1, X_2, \ldots, X_N$ ARE IID WITH PDF $f(x;\theta)$ SUCH THAT THE REGULARITY CONDITIONS ARE SATISFIED.
SUPPOSE FURTHER THAT $I(\theta)$ SATISFIES $0 < I(\theta) < \infty$, THEN ANY CONSISTENT SEQUENCE OF SOLUTIONS OF THE MLE EQUATIONS SATISFIES:

$$\sqrt{N}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

$$\sqrt{N}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \frac{\theta^2}{4}\right)$$

PROBLEM 3

$$f_\theta(x) = \frac{3\theta^3}{(x+\theta)^4} \qquad x>0;\ \theta>0$$

$$\log f_\theta(x) = \log(3\theta^3) - \log(x+\theta)^4$$

$$= 3\log(\theta) - 4\log(x+\theta)$$

$$\frac{\partial^2}{\partial\theta^2}\left[3\log(\theta) - 4\log(x+\theta)\right]$$

$$\frac{\partial}{\partial\theta}\left[\frac{3}{\theta} - \frac{4}{x+\theta}\right]$$

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

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

WE NEED TO APPLY THE LAW OF THE UNCONCIOUS STATISTICIAN:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

$$g(x) = \frac{4}{(x+\theta)^2}$$

$$f_\theta(x) = \frac{3\theta^3}{(x+\theta)^4}$$

$$E[g(X)] = \int_0^{\infty} \frac{4}{(x+\theta)^2} \frac{3\theta^3}{(x+\theta)^4}\, dx$$

I NEED TO COMPUTE:

$$E\left[\frac{3}{\theta^2} - \frac{4}{(X+\theta)^2}\right]$$

$$= \frac{3}{\theta^2} - E[g(x)] =$$

$$= \frac{3}{\theta^2} - \int_0^{+\infty} \frac{4}{(x+\theta)^2} \frac{3\theta^3}{(x+\theta)^4} dx$$

$$= \frac{3}{\theta^2} - 4 \int_0^{+\infty} \frac{3\theta^3}{(x+\theta)^6} dx$$

$$u = x + \theta$$

$$du = dx$$

When $x = 0$, $u = \theta$

When $x \longrightarrow \infty$, $u \to \infty$

$$\int_\theta^\infty \frac{3\theta^3}{u^6} du$$

$$= 3\theta^3 \int_\theta^\infty \frac{1}{u^6} du$$

$$\int u^{-6}\,du = -\frac{1}{5u^5}$$

$$\Rightarrow 3\theta^3 \left[-\frac{1}{5u^5}\right]_{\theta}^{\infty}$$

$$\left[-\frac{1}{5u^5}\right]_{\infty} = 0$$

$$\left[-\frac{1}{5u^5}\right]_{\theta} = -\frac{1}{5\theta^5}$$

BACK TO THE INTEGRAL:

$$= 3\theta^3\left(0-\left(-\frac{1}{5\theta^5}\right)\right) = 3\theta^3\,\frac{1}{5\theta^5} = \frac{3}{5\theta^2}$$

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

$$= \frac{3}{\theta^2} - 4\,\frac{3}{5\theta^2} = \frac{3}{\theta^2} - \frac{12}{5\theta^2}$$

$$= \frac{15-12}{5\theta^2} = \frac{3}{5\theta^2}$$

FOR $X_1, \dots, X_N \overset{IID}{\sim} f_\theta(x)$, LET

$$\hat{\theta} = 2\bar{X}$$

THEN $E[\hat{\theta}] = E[2\bar{X}]$

$$= 2E[\bar{X}] = 2\frac{1}{N}[N\mu] =$$

$$\overset{!}{=} 2E[X_1]$$

$$VAR[\hat{\theta}] = 2^2 \frac{VAR(X_1)}{N}$$

$$E[X_1] = \int_0^\infty x \frac{3\theta^3}{(x+\theta)^4} dx$$

$$= \int_0^\infty [(x+\theta) - \theta] \frac{3\theta^3}{(x+\theta)^4} dx$$

$$= \boxed{\int_0^{+\infty} \frac{3\theta^3}{(x+\theta)^3} dx} - \theta$$

LET US SOLVE THE INTEGRAL

$$u = x + \theta$$

$$du = dx$$

THE LIMITS BECOME $[\theta, +\infty]$

$$\int_{\theta}^{+\infty} \frac{3\theta^3}{u^3}\, du - \theta$$

$$= 3\theta^3 \int_{\theta}^{+\infty} u^{-3}\, du - \theta$$

$$= 3\theta^3 \left[- \frac{u^{-2}}{2} \right]_{\theta}^{\infty} - \theta$$

$$= \frac{3}{2}\theta^3 \left[- \frac{1}{u^2} \right]_{\theta}^{\infty} - \theta$$

$$= \frac{3}{2}\theta^3 \left(- 0 - \frac{1}{\theta^2} \right) - \theta$$

$$\frac{3}{2}\theta^3 \frac{1}{\theta^2} - \theta$$

$$= \frac{3}{2}\theta - \theta = \frac{1}{2}\theta$$

$$E[X_1^2] = \int_0^{+\infty} x^2 \frac{3\theta^3}{(x+\theta)^4} dx$$

$$= 3\theta^3 \int_0^{+\infty} \frac{x^2}{(x+\theta)^4} dx$$

$$u = \frac{x}{\theta}$$

$$x = u\theta$$

$$dx = \theta\, du$$

WHEN $x = 0$, $u = 0$

WHEN $x = +\infty$, $u = +\infty$

$$= 3\theta^3 \int_0^{+\infty} \frac{(\theta u)^2}{(\theta u + \theta)^4} \, du$$

$$= 3\theta^3 \int_0^{+\infty} \frac{\theta^2 u^2}{(\theta(u+1))^4} \, du$$

$$= 3\theta^3 \int_0^{+\infty} \frac{\theta^2 u^2}{\theta^4 (u+1)^4} \, du$$

$$= 3\theta^3 \int_0^{+\infty} \frac{1}{\theta^2} \frac{u^2}{(u+1)^4} \, du$$

$$= 3\theta \int_0^{+\infty} \frac{u^2}{(u+1)^4} \, du$$

WE SOLVE THIS USING INTEGRATION BY PARTS:

$$\int v\,dw = v\,w - \int w\,dv$$

$$v = u^2 \longrightarrow dv = 2\,u\,du$$

$$dw = \frac{1}{(u+1)^4}\,du \longrightarrow w = -\frac{1}{3}\,\frac{}{(u+1)^3}$$

$$= \left[u^2\left(-\frac{1}{3}\,\frac{}{(u+1)^3}\right)\right]_0^{+\infty} + \int_0^{+\infty} 2u\,\frac{1}{3(u+1)^3}\,du$$

AT $u = 0 \longrightarrow 0$

AT $u = +\infty \longrightarrow 0$

WE ARE LEFT WITH:

$$\frac{2}{3}\int_0^{+\infty} \frac{u}{(u+1)^3}\,du$$

$$v = u+1$$
$$dv = du$$

THE LIMITS BE $v \in [1, \infty]$

$$\int_1^\infty \frac{v-1}{v^3}\, dv$$

$$= \int_1^{+\infty} \left(\frac{1}{v^2} - \frac{1}{v^3} \right) dv$$

$$= \int_1^{+\infty} \frac{1}{v^2}\, dv = \left[-\frac{1}{v} \right]_1^{+\infty} = 1$$

$$= \int_1^{+\infty} \frac{1}{v^3}\, dv = \left[-\frac{1}{2v^2} \right]_1^{+\infty} = \frac{1}{2}$$

SUMMING UP THE TWO PARTS:

$$\int_1^{+\infty} \left(\frac{1}{v^2} - \frac{1}{v^3} \right) dv = 1 - \frac{1}{2} = \frac{1}{2}$$

$$\frac{2}{3} \left(1 - \frac{1}{2} \right) = \frac{1}{3}$$

SUMMING UP:

$$3\theta^2 \int_0^{+\infty} \frac{u^2}{(u+1)^4 du} = 3\theta^2 \frac{1}{3} = \theta^2$$

$$VAR(\hat{\theta}) = \frac{4}{N} \left[E(X_1^2) - E(X_1)^2 \right] = \frac{4}{m} \frac{3\theta^2}{4} = \frac{3\theta^2}{N}$$

IS THE ESTIMATOR OF θ, NAMELY $\hat{\theta}$ UNBIASED?

YES!

$$E(X_1) = \frac{\theta}{2} \Longrightarrow E(\hat{\theta}) = 2E(X_1)$$

$$= 2\frac{\theta}{2} = \theta$$

NOW LET US ASSESS THE EFFICIENCY OF $\hat{\theta}$

$$\text{EFF}(\hat{\theta}) = \frac{[m\, I(\theta)]^{-1}}{\text{VAR}(\hat{\theta})} = \frac{5\theta^2/3m}{3\theta^2/m} = \frac{5}{9} < 1$$