Course Statistical Inference SIF5084

Faglig kontakt under eksamen:

Bo Lindqvist 735 93532

Hjelpemidler: (C):

K.Rottman: Mathematische Formelsamlung,

Statistiske tabeller og formler, Tapir,

Godkjent kalkulator med tomt minne,

Selvskrevet gult titteark p A4-ark utdelt av farlrer,

Engelsk-Norsk ordbok.

1. A sample $X_1, ..., X_n$ is taken from a gamma distribution with parameters θ and $1/\theta$:

$$X_1, ..., X_n \sim gamma\left(\theta, \frac{1}{\theta}\right)$$

i. e. pdf of X_i is

$$f(x;\theta) = \frac{\theta^\theta}{\Gamma(\theta)} x^{\theta-1} e^{-\theta x} I_{\{x>0\}}, \quad \theta > 0.$$

Find a one-dimensional sufficient statistic for θ.

Solution. The likelihood function is

$$L(\theta; X_1, ..., X_n) = \theta^{n\theta}[\Gamma(\theta)]^{-n}(X_1 \cdot ... \cdot X_n)^{\theta-1} e^{-\theta \sum X_i} =$$

$$= \theta^{n\theta}[\Gamma(\theta)]^{-n}(X_1 \cdot ... \cdot X_n e^{-\sum X_i})^{\theta}(X_1 \cdot ... \cdot X_n)^{-1}.$$

Put

$$T(X_1, ..., X_n) = \prod X_i \cdot e^{-\sum X_i},$$

$$g(T, \theta) = \theta^{n\theta}[\Gamma(\theta)]^{-n}[T(X_1, ..., X_n)]^{\theta},$$

and

$$h(X_1, ..., X_n) = (X_1 \cdot ... \cdot X_n)^{-1}.$$

Then

$$L(\theta; X_1, ..., X_n) = g(T(X_1, ..., X_n), \theta)h(X_1, ..., X_n)$$

and hence, due to the factorization theorem, $T(X_1, ..., X_n)$ is a (univariate) sufficient statistic.

2. Let $X_1, ..., X_n$ be a sample taken from a normal distribution with zero mean and unknown variance θ^2:

$$X_1, ..., X_n \sim N(0, \theta^2)$$

a) Find the (expected) Fisher information.

b) Consider the following estimator of θ^2:

$$T_n = \frac{2}{n}X_1^2 + \frac{n-2}{n(n-1)}\sum_{i=2}^{n} X_i^2.$$

Is this estimator unbiased?

c) Is T_n consistent?

d) Is the estimator T_n efficient? (We call an unbiased estimator efficient if its variance coincides with the lower bound of the Cramer-Rao inequality).

e) Find MLE (maximum likelihood estimator) of θ^2. Is it unbiased? Consistent? Efficient?

f) Give an example of a biased (but consistent!) estimator whose variance is less then $\frac{2\theta^4}{n}$ for all n. Find its bias and MSE (mean squared error). Compare the latter with MSE of the MLE.

Solution.

a) Denote the Fisher information of the sample and that of one observation by $I(\theta)$ and $I_0(\theta)$ respectively. Then $I(\theta) = nI_0(\theta)$ and

$$I_0(\theta) = E\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2 = E\left(\frac{X^2}{\theta^3} - \frac{1}{\theta}\right)^2 = \frac{1}{\theta^6}EX^4 - \frac{1}{\theta^2}$$

To find EX^4 we can use mgf: $EX^n = {M_X}^{(n)}(0)$.

We have

$$M_X(t) = e^{\theta^2 t^2/2}$$

$$M''_X(t) = e^{\theta^2 t^2/2}(\theta^4 t^2 + \theta^2)$$
$$M'''_X(t) = e^{\theta^2 t^2/2}(\theta^6 t^3 + 3\theta^4 t)$$
$$M_X^{(4)}(t) = e^{\theta^2 t^2/2}(\theta^2 t)(\theta^6 t^3 + 3\theta^4 t) + e^{\theta^2 t^2/2}(3\theta^6 t^2 + 3\theta^4)$$

Hence $EX^4 = M^{(4)}(0) = 3\theta^4$

and

$$I_0(\theta) = \frac{2}{\theta^2}, \quad I(\theta) = \frac{2n}{\theta^2}$$

b)

$$ET_n = \frac{2}{n}EX_1^2 + \frac{(n-2)}{n(n-1)}\sum_{i=2}^{n} EX_i^2 = \theta^2\left(\frac{2}{n} + \frac{(n-2)}{n(n-1)}(n-1)\right) = \theta^2$$

i.e. T_n is unbiased.

c) It is consistent: $\frac{2}{n}X_1^2 \xrightarrow{P} 0,$

$$\frac{n-2}{n} \longrightarrow 1, \qquad \frac{1}{n-1}\sum_{i=2}^{n} X_i^2 \xrightarrow{P} EX^2 = \theta^2$$

d)

$$VarT_n = \frac{4}{n^2}Var(X_1^2) + \frac{(n-2)^2}{n^2(n-1)^2}\sum_{i=2}^{n} Var(X_i^2) =$$
$$= Var(X^2)\left[\frac{4}{n^2} + \frac{(n-2)^2(n-1)}{n^2(n-1)^2}\right] = Var(X^2)\frac{1}{n-1} = \frac{2\theta^4}{n-1}$$

since

$$Var(X^2) = EX^4 - (EX^2)^2 = 3\theta^4 - \theta^4 = 2\theta^4$$

(EX^4 was obtained in part (a)).

The Cramer-Rao lower bound is (use part (a))

$$\frac{\left[\frac{d}{d\theta}(\theta^2)\right]^2}{I(\theta)} = \frac{4\theta^2}{2n/\theta^2} = \frac{2\theta^4}{n} < \frac{2\theta^4}{n-1} = Var(T_n).$$

Hence T_n is not efficient

e)

$$L(\theta; X_1, ..., X_n) = (2\pi)^{-n/2}\theta^{-n}e^{-\frac{1}{2\theta^2}\sum X_i^2}.$$
$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^3}\sum X_i^2.$$

So, the MLE is $T_{MLE} = \frac{1}{n}\sum X_i^2$.

$$ET_{MLE} = \frac{1}{n}\sum EX_i^2 = \theta^2$$

(unbiased);

due to the Law of Large Numbers $\frac{1}{n}\sum X_i^2 \xrightarrow{P} EX^2 = \theta^2$

(consistent).

$$Var(T_{MLE}) = \frac{1}{n^2}\sum Var(X_i^2) = \frac{1}{n}Var(X^2) =$$

$$= \frac{1}{n}(EX^4 - (EX^2)^2) = \frac{1}{n}(3\theta^4 - \theta^4) = \frac{2\theta^4}{n}$$

which coincides with the Cramer-Rao lower bound (see part (d)), therefore T_{MLE} is efficient.

f) Any estimator of the form

$$V_n = c_n T_{MLE} = \frac{c_n}{n}\sum X_i^2,$$

where $0 < c_n < 1$ and $c_n \to 1$ as $n \to \infty$, satisfies these conditions. It is biased:

$$EV_n = c_n ET_{MLE} = c_n\theta^2 < \theta^2;$$

consistent (evidently, since $c_n \to 1$ and $T_{MLE} \xrightarrow{P} \theta^2$); and

$$Var(V_n) = c_n^2 Var(T_{MLE}) = c_n^2\frac{2\theta^4}{n} < \frac{2\theta^4}{n}.$$

Bias of V_n is

$$b(V_n) = c_n\theta^2 - \theta^2 = (c_n - 1)\theta^2;$$

MSE is

$$MSE(V_n) = [b(V_n)]^2 + Var(V_n) = [(c_n - 1)\theta^2]^2 + c_n^2\frac{2\theta^4}{n} =$$

$$= (c_n - 1)^2\theta^4 + c_n^2\frac{2\theta^4}{n} = [\frac{n(c_n - 1)^2}{2} + c_n^2]\frac{2\theta^4}{n}.$$

For example, if $c_n = \frac{n-1}{n}$, then

$$\frac{n(c_n - 1)^2}{2} + c_n^2 = \frac{2n^2 - 3n + 2}{2n^2} < \frac{2n^2}{2n^2} = 1$$

and $MSE(V_n) < MSE(T_{MLE})$.

3. Let $X_1, ..., X_n$ be a sample taken from a $(\theta, 1)$ normal distribution:

$$X_1, ..., X_n \sim N(\theta, 1).$$

a) For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ find the acceptance region of the significance level α likelihood ratio test.

b) Find the $(1 - \alpha)$ confidence interval that results from inverting the likelihood ratio test of part (a).

Solution.

a) The acceptance region: $\lambda(X) \geq c$ where

$$\lambda(X) = \frac{L(\theta_0; X)}{\sup L(\theta; X)} = \frac{L(\theta_0; X)}{L(\hat{\theta}_{MLE}; X)},$$

and c is found from the condition

$$P_{\theta_0}(\lambda(X) \geq c) = 1 - \alpha$$

$$\hat{\theta}_{MLE} = \frac{1}{n} \sum X_i = \bar{X}$$

$$\lambda(X) = e^{-\frac{n}{2}(\bar{X} - \theta_0)^2}$$

So, the acceptance region

$$\theta_0 - \frac{1}{\sqrt{n}} l_{1-\frac{\alpha}{2}} \leq \bar{X} \leq \theta_0 + \frac{1}{\sqrt{n}} l_{1-\frac{\alpha}{2}}$$

where l_δ - δ-quantile of the standard normal distribution.

b) Inverting the test of part (a) we obtain the following $(1 - \alpha)$ confidence interval:

$$\left[\bar{X} - \frac{1}{\sqrt{n}} l_{1-\frac{\alpha}{2}}, \bar{X} + \frac{1}{\sqrt{n}} l_{1-\frac{\alpha}{2}} \right].$$

4. Observations $Y_1, ..., Y_n$ are described by the relationship

$$Y_i = \theta \cdot e^{x_i^2}(1 + x_i^2) + \varepsilon_i$$

where $x_1, ..., x_n$ are fixed constants and $\varepsilon_1, ..., \varepsilon_n$ are iid $N(0, \sigma^2)$.

a) Find LSE (least squares estimator) of θ.

b) Find MLE of θ.

Solution.

a) $\sum(Y_i - \theta \cdot e^{x_i^2}(1+x_i^2))^2 \longrightarrow \min$

$$\frac{\partial}{\partial\theta} = -2\sum(Y_i - \theta \cdot e^{x_i^2}(1+x_i^2))e^{x_i^2}(1+x_i^2) = 0$$

$$\sum Y_i e^{x_i^2}(1+x_i^2) = \theta \sum \left[e^{x_i^2}(1+x_i^2)\right]^2$$

$$\hat{\theta}_{LSE} = \frac{\sum Y_i e^{x_i^2}(1+x_i^2)}{\sum \left[e^{x_i^2}(1+x_i^2)\right]^2}$$

b) The likelihood function

$$L(\theta; Y) = (2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\{-\frac{1}{2\sigma^2}\sum(Y_i - \theta e^{x_i^2}(1+x_i^2))^2\}$$

$$\ln L = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum(Y_i - \theta e^{x_i^2}(1+x_i^2))^2$$

we have the same minimization problem as in part (a) and hence the same result:

$$\hat{\theta}_{MLE} = \hat{\theta}_{LSE} = \frac{\sum Y_i e^{x_i^2}(1+x_i^2)}{\sum \left[e^{x_i^2}(1+x_i^2)\right]^2}.$$