STK4011 and STK9011 Autumn 2016

Point estimation

Covers (most of) the following material from chapters 6 and 7:

- Section 6.2.4: pages 285-286 and 288
- Section 7.3.2: pages 334-342
- Section 7.3.3: pages 342-348

Ørnulf Borgan
Department of Mathematics
University of Oslo

Best unbiased estimators

Definition 7.3.7

An estimator W^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $\mathrm{E}_\theta W^* = \tau(\theta)$ for all θ and, for any other estimator W with $\mathrm{E}_\theta W = \tau(\theta)$, we have $\mathrm{Var}_\theta W^* \leq \mathrm{Var}_\theta W$ for all θ

We also say that W^* is a uniform minimum variance unbiased estimator (UMVUE) for $\tau(\theta)$

Information

Consider a sample $\mathbf{X} = (X_1, X_2, ..., X_n)$ with joint pdf (or pmf) $f(\mathbf{x} \mid \theta)$

We assume that expressions of the form

$$\mathrm{E}_\theta W(\mathbf{X}) = \int_{\mathcal{X}} W(\mathbf{x}) f(\mathbf{x} \mid \theta) d\mathbf{x}$$

may be differentiated with respect to θ by changing the order of differentiation and integration (summation in the case of pmf)

Then we have

$$\mathrm{E}_\theta \left(\frac{\partial}{\partial \theta} \log f(\mathbf{X} \mid \theta) \right) = 0$$

The information number or Fisher information of the sample is

$$I(\theta) = \mathrm{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(\mathbf{X} \mid \theta) \right)^2 \right] = \mathrm{Var}_\theta \left(\frac{\partial}{\partial \theta} \log f(\mathbf{X} \mid \theta) \right)$$

For the special case where $X_1, X_2, ..., X_n$ are iid with pdf (or pmf) $f(x \mid \theta)$ the information in the sample is given by $I(\theta) = nI_1(\theta)$, where $I_1(\theta)$ is the information in one observation and is given by

$$I_1(\theta) = \mathrm{E}_\theta \left[\left(\frac{\partial}{\partial \theta} \log f(X \mid \theta) \right)^2 \right] = \mathrm{Var}_\theta \left(\frac{\partial}{\partial \theta} \log f(X \mid \theta) \right)$$

If expressions of the form $\int W(\mathbf{x}) f(\mathbf{x}|\theta)d\mathbf{x}$ may be differentiated twice with respect to θ by changing the order of differentiation and integration (summation for pmf), the information in the sample may also be given as

$$I(\theta) = -\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(\mathbf{X}|\theta)\right)$$

while in the case of iid observations, the information in one observation may be given as

$$I_1(\theta) = -\mathrm{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)$$

The Cramér-Rao inequality (Theorem 7.3.9)

Consider a sample $\mathbf{X} = (X_1, X_2, ..., X_n)$ with joint pdf (or pmf) $f(\mathbf{x} \mid \theta)$ and assume that expressions of the form

$$\mathrm{E}_\theta W(\mathbf{X}) = \int_{\mathcal{X}} W(\mathbf{x}) f(\mathbf{x}|\theta)d\mathbf{x}$$

may be differentiated with respect to θ by changing the order of differentiation and integration (summation in the case of pmf). Then for any unbiased estimator $W(\mathbf{X})$ of $\tau(\theta)$, we have

$$\mathrm{Var}_\theta W(\mathbf{X}) \geq \left(\tau'(\theta)\right)^2 / I(\theta)$$

where

$$I(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\mathbf{X}|\theta)\right)^2\right]$$

Example – Binomial UMVUE (exercise 7.40)

Let $X_1, X_2, ..., X_n$ be iid Bernoulli random variables with success probability p

Here

$$\log f(x|p) = \log\left(p^x(1-p)^{1-x}\right)$$
$$= x\log p + (1-x)\log(1-p)$$

$$\frac{\partial}{\partial p}\log f(x|p) = \frac{x}{p} - \frac{1-x}{1-p}$$

$$\frac{\partial^2}{\partial p^2}\log f(x|\lambda) = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$$

Hence the information in one observation is

$$I_1(p) = \mathrm{E}_p\left(\frac{X}{p^2} + \frac{1-X}{(1-p)^2}\right) = \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p(1-p)}$$

and the information in the sample is $I(p) = n/\left[p(1-p)\right]$

By the Cramér-Rao inequality, we have for any unbiased estimator W for p that

$$\mathrm{Var}_p W \geq \frac{1}{I(p)} = \frac{p(1-p)}{n}$$

Now $\bar{X}$ is an unbiased estimator for p and

$$\mathrm{Var}_p\, \bar{X} = \frac{p(1-p)}{n}$$

Hence $\bar{X}$ is an UMVUE for p

Example 7.3.13 – uniform $(0,\theta)$

Let $X_1, X_2, ..., X_n$ be iid and $\text{uniform}(0,\theta)$

For this example the conditions of the Cramér-Rao inequality are not fulfilled

Example 7.3.14 – normal variance bound

Let $X_1, X_2, ..., X_n$ be iid $n(\mu,\sigma^2)$ random variables, where both parameters are unknown

Here S^2 is an unbiased estimator for σ^2, but it does not attain the Cramér-Rao lower bound

Proof of the Cramér-Rao inequality

The proof of the Cramér-Rao inequality is an application of the following well-known result on correlation:

For random variables Z and Y we have

$$-1 \leq \text{corr}(Z,Y) \leq 1$$

with $|\text{corr}(Z,Y)| = 1$ if and only if there exist constants $a \neq 0$ and b such that $Y = aZ + b$

From this result it follows that

$$\text{Var}\, Z \geq \frac{[\text{Cov}(Z,Y)]^2}{\text{Var} Y} \qquad (*)$$

with equality if and only if there exist constants $a \neq 0$ and b such that $Y = aZ + b$

To prove the Cramér-Rao inequality we use (*) with

$$Z = W(\mathbf{X}) \quad \text{and} \quad Y = \frac{\partial}{\partial\theta} \log f(\mathbf{X}|\theta)$$

We know that

$$\text{Var}_\theta\left(\frac{\partial}{\partial\theta} \log f(\mathbf{X}|\theta)\right) = \text{E}_\theta\left[\left(\frac{\partial}{\partial\theta} \log f(\mathbf{X}|\theta)\right)^2\right]$$

So we only need to prove that

$$\tau'(\theta) = \frac{d}{d\theta}\text{E}_\theta W(\mathbf{X}) = \text{Cov}_\theta\left(W(\mathbf{X}),\, \frac{\partial}{\partial\theta} \log f(\mathbf{X}|\theta)\right)$$

Note that we have equality if and only if $Y = aZ + b$

This gives the corollary on the next slide

Corollary 7.3.15

Consider a sample $\mathbf{X} = (X_1, X_2, ..., X_n)$ with joint pdf (or pmf) $f(\mathbf{x} \mid \theta)$ and assume that the conditions of the Cramér-Rao inequality are fulfilled

Then an unbiased estimator $W(\mathbf{X})$ of $\tau(\theta)$ attains the Cramér-Rao lower bound if and only if there exist a function $a(\theta)$ such that

$$a(\theta)\{W(\mathbf{X}) - \tau(\theta)\} = \frac{\partial}{\partial\theta} \log f(\mathbf{X}|\theta)$$

Example 7.3.16 – normal variance bound

Let $X_1, X_2, ..., X_n$ be iid $n(\mu,\sigma^2)$

If μ is known, one may attain the Cramér-Rao lower bound when estimating σ^2, otherwise not

Sufficiency and unbiased estimators

We will now see how sufficiency may help us to find the best unbiased estimators

We remember that a statistic $T = T(\mathbf{X})$ is a sufficient statistic for θ if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ

We will make use of the following results on conditional means and variances:

$$\mathrm{E}X = \mathrm{E}\left[\mathrm{E}(X \mid Y)\right] \qquad \text{(Theorem 4.4.3)}$$

$$\mathrm{Var}X = \mathrm{Var}\left[\mathrm{E}(X \mid Y)\right] + \mathrm{E}\left[\mathrm{Var}(X \mid Y)\right] \qquad \text{(Thm 4.4.7)}$$

Theorem 7.3.17 (Rao-Blackwell)

Let $W = W(\mathbf{X})$ be any unbiased estimator of $\tau(\theta)$ and let $T = T(\mathbf{X})$ be sufficient statistic for θ

Then $\phi(T) = \mathrm{E}(W \mid T)$ is an unbiased estimator of $\tau(\theta)$ and $\mathrm{Var}_\theta\, \phi(T) \le \mathrm{Var}_\theta W$ for all θ

The Rao-Blackwell theorem shows that when looking UMVUEs, we may restrict our attention to estimators that are functions of a sufficient statistic

Theorem 7.3.19

If W is a best unbiased estimator of $\tau(\theta)$, then W is unique

If W is an unbiased estimator for $\tau(\theta)$ and we have another estimator U that satisfies $\mathrm{E}_\theta U = 0$ for all θ (i.e. U is an unbiased estimator of 0), then $\phi_a = W + aU$ is an unbiased estimator of $\tau(\theta)$

Note that

$$\mathrm{Var}_\theta\, \phi_a = \mathrm{Var}_\theta W + a^2\, \mathrm{Var}_\theta U + 2a\, \mathrm{Cov}_\theta(W, U)$$

If for some $\theta = \theta_0$ we have $\mathrm{Cov}_{\theta_0}(W, U) \ne 0$, then we may find a value of a such that

$$\mathrm{Var}_{\theta_0}\, \phi_a < \mathrm{Var}_{\theta_0} W$$

It follows that an UMVUE has to be uncorrelated with all unbiased estimators of 0

The converse is also true:

Theorem 7.3.20

Let W be an unbiased estimator of $\tau(\theta)$, then W is UMVUE if and only is it is uncorrelated with all unbiased estimators of 0

This results makes it interesting to identify situations where there exists no unbiased estimators of 0 (except 0 itself)

Definition 6.2.21

Let $f(t \mid \theta)$ be the family of pdfs or pmfs of a statistic $T = T(\mathbf{X})$. The family of distributions is complete if $\mathrm{E}_\theta\, g(T) = 0$ for all θ implies that $P_\theta(g(T) = 0) = 1$ for all θ . We also say that T is a complete statistic

Examples 6.2.23 & 7.3.22 – uniform (0,θ)

Let $X_1, X_2, ..., X_n$ be iid and $\text{uniform}(0,\theta)$

Then $T = \max X_i$ is a sufficient statistic

So when looking for UMVUE for θ we may restrict attention to estimators based on $T = \max X_i$

We may show that $T = \max X_i$ is a complete statistic

It follows that there exist no unbiased estimators of 0 based on $T = \max X_i$ (except 0 itself)

Thus $\hat{\theta} = \frac{n+1}{n} \max X_i$ is an unbiased estimator of θ that is uncorrelated with all estimators of 0

It follows that $\hat{\theta}$ is the unique UMVUE for θ

The argument on the previous slide holds quite generally, and gives the following result:

Theorem 7.3.23 (Lehmann-Scheffé)

Let $T = T(\mathbf{X})$ be a complete sufficient statistic for θ and let $\phi(T)$ be an estimator based only on T with $\mathrm{E}_\theta \phi(T) = \tau(\theta)$. Then $\phi(T)$ is the unique UMVUE of $\tau(\theta)$

Theorem 6.2.25

Let $X_1, X_2, ..., X_n$ be iid observations from an exponential family with pdf or pmf of the form

$$f(x \mid \boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x) \right)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$ and the parameter space Θ contain an open set in $\mathcal{R}^k$. Then

$$T(\mathbf{X}) = \left(\sum_{j=1}^{n} t_1(X_j), ..., \sum_{j=1}^{n} t_k(X_j) \right)$$

is a complete sufficient statistic

Example – UMVUE for the normal distribution

Let $X_1, X_2, ..., X_n$ be iid $n(\mu, \sigma^2)$ random variables, where both parameters are unknown

The normal pdf may be written

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2} \right) \exp\left(\frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2 \right)$$

Hence

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} X_i \, , \, \sum_{j=1}^{n} X_i^2 \right)$$

is a complete sufficient statistics

It follows that $\bar{X}$ and S^2 are UMVUE for μ and σ^2

If we combine the result of Lehmann-Scheffé
(thm 7.3.23) with the Rao-Blackwell construction
(thm 7.3.17), we obtain the following result
(cf. example 7.3.24):

Let $T = T(\mathbf{X})$ be a complete sufficient statistic for θ
and let W be an unbiased estimator of $\tau(\theta)$.
Then $\phi(T) = \mathrm{E}(W \mid T)$ is the unique UMVUE of $\tau(\theta)$.