Foundations of Statistical Inference

Julien Berestycki

Department of Statistics
University of Oxford

MT 2016

Lecture 4 - Consequences of the Cramér-Rao Lower Bound.
Searching for a MVUE. Rao-Blackwell Theorem.

Previously on SB2a

Theorem

Theorem 2 : Cramér-Rao inequality (and bound). If $\widehat{\theta}$ is an unbiased estimator of θ, then subject to certain regularity conditions on $f(x;\theta)$, we have

$$Var(\widehat{\theta}) \geq I_\theta^{-1}.$$

where I_θ, the Fisher information, is given by $I_\theta = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ell(\theta)\right]$

Corollary (1)

There exists an unbiased estimator $\widehat{\theta}$ which attains the CR lower bound (under regularity conditions) if and only if

$$S(x,\theta) = \frac{\partial \ell}{\partial \theta} = I_\theta(\widehat{\theta} - \theta)$$

and then X belongs to an exponential family.

Previously on SB2a

Corollary (3)

Suppose $\widetilde{\theta}(X)$ is an unbiased estimator that attains the CRLB, and so is a MVUE. Suppose that the MLE $\widehat{\theta}$ is a solution to $\partial\ell/\partial\theta = 0$ (so, not on boundary). Then $\widetilde{\theta} = \widehat{\theta}$.

Efficiency

Definition

The (Bahadur) **efficiency** of an estimator $\widetilde{\theta}$ is defined as a comparison of the variance of $\widetilde{\theta}$ with the CR bound I_θ^{-1}. That is

$$\text{Efficiency of } \widetilde{\theta} = \frac{I_\theta^{-1}}{\text{Var}[\widetilde{\theta}]} = \frac{1}{I_\theta \text{Var}[\widetilde{\theta}]}$$

The **asymptotic efficiency** is the limit as $n \to \infty$.

There are similar definitions for the relative efficiency of two estimators.

Asymptotic normality of MLE

Revision from Part A Statistics As the sample size $n \to \infty$, the MLE

$$\widehat{\theta} \approx N(\theta, I_\theta^{-1}).$$

This is a powerful and general result. Assuming the usual regularity conditions hold then it tells us that the MLE has the following properties

1. it is asymptotically unbiased
2. it is asymptotically efficient i.e. it attains the CRLB asymptotically.
3. it has a normal distribution asymptotically.

Extensions to the Cramér-Rao inequality

CRLB for biased θ

If $\widehat{\theta}$ is an estimator with bias $b(\theta) = \text{bias}(\widehat{\theta})$, then

$$\text{Var}[\widehat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta}\right)^2 I_\theta^{-1}$$

Proof Begin with $\mathbb{E}_\theta(\widehat{\theta}(X)) = \theta + b(\theta)$ (in 1.) and $\mathbb{E}_\theta(\widehat{g}(X)) = g(\theta)$ (in 2.). Differentiate both sides and proceed as above to find $\text{Cov}[U, V] = (1 + \partial b/\partial\theta)$ (in 1.) and $\text{Cov}[U, V] = \partial g/\partial\theta$ (in 2., with $U = \widehat{g}$). The bound is against $\text{Cov}[U, V]^2$ which leads to the results above.

Extensions to the Cramér-Rao inequality

Suppose that the regularity conditions hold and that $0 < I_\theta < \infty$. Set $\gamma = g(\theta)$ for some differentiable function g with $g' \neq 0$ and let T be a **regular** unbiased estimator of γ

$$\int T(x) \frac{\partial}{\partial \theta} L(\theta, x) dx = \frac{\partial}{\partial \theta} \int T(x) L(\theta, x) dx$$

CRLB again

$$V(T) \geq \frac{(g'(\theta))^2}{I_\theta} \quad \forall \theta$$

with equality if and only if

$$\frac{\partial \ell}{\partial \theta} = I_\theta (T(x) - g(\theta))/g'(\theta) \quad \forall \theta.$$

CRLB for p dimensional parameter

Suppose that $\theta = (\theta_1, \ldots, \theta_p)$ and $\hat{\theta}$ is **unbiased** so that

$$\forall i = 1, \ldots p, \quad E[\hat{\theta}_i] = \theta_i.$$

Definition

Fisher information matrix

$$I_{\theta^*} = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell|_{\theta=\theta^*}\right]_{i,j=1,\ldots,p} = E\left[\frac{\partial\ell}{\partial\theta_i}\frac{\partial\ell}{\partial\theta_j}|_{\theta=\theta^*}\right]_{i,j=1,\ldots,p}$$

For matrices A, B $A \succeq B$ means that $B - A$ is positive semi definite i.e.

$$x^t(A - B)x \geq 0, \quad \forall x \in \mathbb{R}^p.$$

CRLB for p dimensional parameter

Recall that $V(\hat{\theta})$ is the $p \times p$ covariance matrix of $\hat{\theta}(X)$

Theorem (CRLB)

$$V(\hat{\theta}) \succeq I_\theta^{-1} \quad \textit{under regularity conditions.}$$

and in particular $Var(\widehat{\theta}_i) \geq [I^{-1}]_{ii},\ i = 1, \ldots, p.$
$\hat{\theta}(X)$ *attains the bound if and only if*

$$\frac{\partial \log f(x;\theta)}{\partial\theta} = I_\theta(\hat{\theta} - \theta).$$

Exercise: verify that we have already proved $\mathrm{Var}(\widehat{\theta}_i) \geq [I_{ii}]^{-1}$. Note that $[I^{-1}]_{ii} \geq [I_{ii}]^{-1}$ (GJJ) so bound above is stronger.
Exercise For an Exponential family in canonical form,

$$I_{ij} = -\frac{\partial^2}{\partial\phi_i\partial\phi_j}nD(\phi).$$

Rao-Blackwell

The CRLB may not be achievable but will still wish to search for an MVUE. Sufficiency plays an important role in the search for a MVUE.

Theorem (Rao-Blackwell Theorem (RBT) (GJJ 2.5.2))

Let $X_1, \ldots, X_n$ be a random sample of observations from $f(x;\theta)$.
Suppose that T is a sufficient statistic for θ and that $\widehat{\theta}$ is any unbiased estimator for θ.
Define a new estimator $\widehat{\theta}_T = \mathbb{E}[\hat{\theta} \mid T]$. Then

1. *$\widehat{\theta}_T$ is a function of T alone;*
2. *$\mathbb{E}[\widehat{\theta}_T] = \theta$;*
3. *$Var(\widehat{\theta}_T) \leq Var(\widehat{\theta})$.*

Comment This says that estimators maybe be improved if we take advsantage of sufficient statistics. Also when looking for MVUE we can restrict ourselves to functions of sufficient statistics.

Proof

1.

$$\widehat{\theta}_T = \mathbb{E}_X[\hat{\theta} \mid T = t] = \int_\chi \hat{\theta}(x) f(x \mid t, \theta) dx = \int_\chi \hat{\theta}(x) f(x \mid t) dx$$

2. $\mathbb{E}[\widehat{\theta}_T] = \mathbb{E}_T[\mathbb{E}[\hat{\theta} \mid T]] = \mathbb{E}[\widehat{\theta}] = \theta$ (by law of total expectation)

3. Using the law of total variance

$$\begin{aligned} \mathrm{Var}(\widehat{\theta}) &= \mathrm{Var}(\mathbb{E}[\hat{\theta} \mid T]) + \mathbb{E}_T[\mathrm{Var}(\hat{\theta} \mid T)] \\ &= \mathrm{Var}(\widehat{\theta}_T) + \mathbb{E}_T[\mathrm{Var}(\hat{\theta} \mid T)] \\ \Rightarrow \mathrm{Var}(\widehat{\theta}) &\geq \mathrm{Var}(\widehat{\theta}_T) \end{aligned}$$

Example 12

Suppose $X_1, \dots, X_n$ be a random sample from Bernoulli(θ).

It is easy to see that $\widehat{\theta} = X_1$ is unbiased for θ. Also, we have seen before that $T = \sum_{i=1}^N X_i$ is sufficient for θ.

We can use RBT to construct an estimator with smaller variance

$$\begin{aligned}
\mathbb{E}[X_1 \mid T = t] = P(X_1 = 1 \mid T = t) &= \frac{P(X_1 = 1, \sum_{i=1}^N X_i = t)}{P(\sum_{i=1}^N X_i = t)} \\
&= \frac{P(X_1 = 1, \sum_{i=2}^N X_i = t-1)}{{}^N C_t \theta^t (1-\theta)^{N-t}} \\
&= \frac{\theta . {}^{N-1} C_{t-1} \theta^{t-1} (1-\theta)^{N-t}}{{}^N C_t \theta^t (1-\theta)^{n-t}} \\
&= \frac{t}{N}
\end{aligned}$$

Corollary (4)

If an MVUE $\widehat{\theta}$ for θ exists, then there is a function $\widehat{\theta}_T$ of the sufficient statistic T for θ which is an MVUE.

Proof If $\widehat{\theta}$ is a MVUE and T is minimal sufficient then by RBT we can construct $\widehat{\theta}_T$. Which implies $\widehat{\theta}_T$ is a function of T alone, is unbiased and variance no larger than $\widehat{\theta}$. Hence is also a MVUE.

Comment This says that we can restrict our search for a MVUE to those based on minimal sufficient statistics.

Completeness

Definition (Complete Sufficient Statistics)

Let $T(X_1, \dots, X_n)$ be a sufficient statistic for θ. The statistic T is **complete** if, whenever $h(T)$ is a function of T for which $\mathbb{E}[h(T)] = 0$ for all θ, then $h(T) \equiv 0$ almost everywhere.

Complete = θ can be estimated on the basis of T: the distributions corresponding to different values of the parameters are distinct.

Lemma (4)

Suppose T is a complete sufficient statistic for θ, and $g(T)$ unbiased for θ, so $\mathbb{E}[g(T)] = \theta$. Then $g(T)$ is the unique function of T which is an unbiased estimator of θ.

Proof If there were two such unbiased estimators $g_1(T), g_2(T)$, then $\mathbb{E}[g_1(T) - g_2(T)] = \theta - \theta = 0$ for all θ, so $g_1(T) = g_2(T)$ almost everywhere.

Question If we have an unbiased estimator what are the sufficient conditions for it to be MVUE?

Lemma (5)

If an MVUE for θ exists and T is a complete and minimal sufficient statistic for θ, and suppose $h = h(T)$ is unbiased for θ, then $h(T)$ is a MVUE.

This Lemma combines the results of Corollary 4 and Lemma 4.

Proof If an MVUE exists then there is a function of T which is an MVUE, by the RB Corollary 4. But $h(T)$ is the only function of T which is unbiased for θ (from Lemma 4). So h must be the function of T which an MVUE.

Question Finally, how can we construct a MVUE?

Theorem (Lehmann-Scheffé Theorem)

Let T be a complete sufficient statistic for θ, and let $\widehat{\theta}$ be an unbiased estimator for θ, then the unbiased estimator $\widehat{\theta}_T = \mathbb{E}[\hat{\theta} \mid T]$ has the smallest variance among all unbiased estimators of θ. That is,

$$Var(\widehat{\theta}_T) \quad \leq \quad Var(\widetilde{\theta})$$

for all unbiased estimators $\widetilde{\theta}$.

Comment This theorem says that if we can find any unbiased estimator and a complete sufficient statistic T then we can construct a MVUE.

Proof

Suppose $\widetilde{\theta}$ exists with $\text{Var}(\widetilde{\theta}) < \text{Var}(\widehat{\theta}_T)$.

Then by RBT we can construct $\widetilde{\theta}_T = \mathbb{E}[\widetilde{\theta} \mid T]$ such that

$$\text{Var}(\widetilde{\theta}_T) \leq \text{Var}(\widetilde{\theta}) < \text{Var}(\widehat{\theta}_T)$$

But $\widetilde{\theta}_T$ and $\widehat{\theta}_T$ are both unbiased and T is complete, so by Lemma 4 we have $\widetilde{\theta}_T = \widehat{\theta}_T$ and

$$\text{Var}(\widetilde{\theta}_T) = \text{Var}(\widehat{\theta}_T)$$

which is a contradiction.