Foundations of Statistical Inference

J. Berestycki & D. Sejdinovic

Department of Statistics
University of Oxford

MT 2019

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

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 may be improved if we take advantage of sufficient statistics. Also when looking for MVUE we can restrict ourselves to functions of sufficient statistics.

Proof

1.

$$\begin{aligned}\widehat{\theta}_T = \mathbb{E}_X[\hat{\theta} \mid T = t] &= \int_{\mathcal{X}} \hat{\theta}(x) f(x \mid t, \theta) dx \\ &= \int_{\mathcal{X}} \hat{\theta}(x) f(x \mid t) dx\end{aligned}$$

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

3. Using the law of total variance

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

Example 12

Suppose $X_1, \ldots, X_n$ is 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 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 $T(X_1, \ldots, X_n)$ 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 $= \theta$ 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 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,

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

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 $\mathsf{Var}(\widetilde{\theta}) < \mathsf{Var}(\widehat{\theta}_T)$.

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

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

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

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

which is a contradiction.