RAO-BLACKWELL THEOREM

1. Introduction

Let $X = (X_1, \ldots, X_n)$ be a random sample from $f_\theta$, where $\theta \in \Theta$. Let $g$ be a function of $\theta$. A estimator $W$ for $g(\theta)$ is a **uniform minimum variance unbiased estimator** (UMVUE, MVUE) if for all $\theta \in \Theta$, it is unbiased, $\mathbb{E}_\theta W = g(\theta)$, and if $Y$ is any other unbiased estimator then $\text{Var}_\theta(W) \leq \text{Var}_\theta(Y)$. Let us remark that a MVUE may not exist.

If certain regularity conditions are satisfied, we already know from the Cramer-Rao lower bound, that

$$\text{Var}_\theta(W) \geq \frac{(g'(\theta))^2}{nI(\theta)}.$$  

We also said any estimator satisfying the Cramer-Rao lower bound is efficient. However, not every MVUE is efficient, and the regularity conditions are not always satisfied, in particular, when the domain of the pdf depends on $\theta$. We will use sufficient statistics to help us find a MVUE.

One the main tools in improving estimators is the following theorem.

**Theorem 1 (Rao-Blackwell).** Let $X = (X_1, \ldots, X_n)$ be a random sample from $f_\theta$, where $\theta \in \Theta$. Let $g$ be a function of $\theta$. Let $Y$ be an unbiased estimator for $g(\theta)$. Let $T$ be a sufficient statistic for $\theta$. Set $W = \mathbb{E}_\theta(Y|T)$. Then $W$ is also an unbiased estimator, and furthermore,

$$\text{Var}_\theta(W) \leq \text{Var}_\theta(Y).$$

Sometimes $W$ is a **Rao-Blackwell estimator**. Let us remark in Theorem 1, we do not know that $W$ is a MVUE. However, it turns out that under certain conditions, $W$ is the unique MVUE. Before we prove Theorem 1, we will recall some properties of conditional expectation.

Let $X$ and $Y$ be random variables. Recall that $\mathbb{E}(X|Y)$ is also random variable, and in particular, it is a function of $Y$; that is, there exists $\phi$, such that $\phi(Y) = \mathbb{E}(X|Y)$. In earlier courses, you found $\phi$ be computing, $\phi(y) := \mathbb{E}(X|Y = y)$. Thus Theorem 1, tell us that when we are looking for MVUE, we can restrict ourselves to functions of sufficient statistics; I hope this justifies the use of the word ‘sufficient.’

An important property of conditional expectations is that

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X).$$
Recall that we define
\[\text{Var}(X|Y) := \mathbb{E}(X^2|Y) - (\mathbb{E}(X|Y))^2.\]

It follows from Jensen’s inequality, for conditional expectations, that \(\text{Var}(X|Y) \geq 0.\)

**Exercise 2.** Let \(Z = \mathbb{E}(X|Y)\) Show that
\[
\text{Var}(X) = \text{Var}(Z) + \mathbb{E}(\text{Var}(X|Y)).
\]
So that in particular, we have
\[
\text{Var}(Z) \leq \text{Var}(X).
\]

Note that it is very important that \(T\) be a sufficient statistic in Theorem 1. Consider a random sample \(X = (X_1, \ldots, X_n),\) where \(X_i \sim \text{Bern}(p),\) and \(n > 2.\) It is not hard to show that \(X_2\) is not a sufficient statistic for \(p,\) however, \(X_1\) is an unbiased estimator for \(p,\) and \(\mathbb{E}(X_1|X_2) = \mathbb{E}(X_1) = p,\) since \(X_1\) is independent of \(X_2;\) thus \(\mathbb{E}(X_1|X_2)\) is not an estimator for \(p.\)

**Proof of Theorem 1.** Note that it is immediate from the properties of conditional expectations that \(\mathbb{E}_\theta W = g(\theta),\) and that \(W\) is a function of \(X,\) since it is a function of \(T.\) It is also immediate by Exercise 2 that the claim regarding the variances holds.

However, this does not mean that \(W\) is an estimator; we need to also show that it does not depend on \(\theta.\) The fact that \(W\) does not depend on \(\theta,\) follows from the fact that the original \(Y\) is also a estimator, and hence a function of \(Y = u(X),\) for some function, that does not depend on \(\theta,\) and the fact that \(T\) is a sufficient statistic. \(\square\)

In proof of Theorem 1, to see the last claim more clearly, consider the case where all the random variables are discrete. Let \(p(x, t) = \mathbb{P}(X = x \mid T = t),\) then \(p(x, t)\) does not depend on \(\theta,\) since \(T\) is a sufficient statistic. Thus
\[
\phi(t) := \mathbb{E}(Y|T = t) = \mathbb{E}(u(X)|T = t) = \sum_x u(x)p(x, t)
\]
also does not depend on \(\theta.\) Hence \(\phi(T)\) does not depend on \(\theta,\) since \(T\) does not depend on \(\theta.\)

2. **Examples**

You will see in the following examples that one can start with very modest estimators, one’s which are not even consistent, and obtain much improved estimators by finding corresponding Rao-Blackwell estimators.
Exercise 3. Let $X = (X_1, \ldots, X_n)$ be a random sample from $f_\theta$. Suppose that the sample sum $T = X_1 + \cdots + X_n$ is a sufficient statistic for $\theta$ and that $\mathbb{E}X_1 = \theta$. Find the Rao-Blackwell estimator $\mathbb{E}(X_1|T)$. Hint: you basically did this for homework.

Exercise 4. Let $X = (X_1, \ldots, X_n)$ be a random sample, where $X_i \sim \text{Unif}(0, \theta)$. We know that $M = \max \{X_1, \ldots, X_n\}$ is a sufficient statistic for $\theta$. Note that $\mathbb{E}X_1 = \theta/2$. Find the Rao-Blackwell estimator $\mathbb{E}(X_1|T)$. Hint: you basically did this for homework.

Exercise 5. Let $X = (X_1, \ldots, X_n)$ be a random sample, where $X_i \sim \text{Poi}(\lambda)$. We know that the sample sum $T$ is a sufficient statistic for $\theta$. Here, we want to estimate $\mathbb{P}(X_1 = 0) = e^{-\lambda}$. Set $Y = 1[\mathbb{X}_1 = 0]$. Clearly, $Y$ is an unbiased estimator for $\mathbb{P}(X_1 = 0)$. Show that the Rao-Blackwell estimator satisfies

$$\mathbb{E}(Y|T) = (1 - \frac{1}{n})^T.$$

Exercise 6. Show that in Exercise 5, the Rao-Blackwell estimator is consistent.

Exercise 7. Let $X = (X_1, \ldots, X_n)$ be a random sample, where $X_i \sim N(\mu, 1)$, where $\mu \in \mathbb{R}$ is unknown. Here, we are interested in estimating $\mu^2$. In a previous exercise, you found that

$$T = (\bar{X})^2 - 1/n,$$

is an unbiased estimator for $\mu^2$ and you computed the variance of $T$. I promised you that we will be able to prove that $T$ is the MVUE, we are almost there. Compute the Fisher information for the random sample $X$, note that the variance of $T$ strictly larger than associated Cramer-Rao bound.

Exercise 8. Note that the random sample itself $X$, is trivially a sufficient statistic. What happens when you apply you find the Rao-Blackwell estimator with respect to $X$?

Exercise 9. What happens when you try to use the Rao-Blackwell theorem twice?