CHAPTER 2 - POINT ESTIMATION

§2.1 Bias and variance considerations

What criteria could we use to determine whether an estimator $T(\mathbf{X})$ is a "good" estimator of parameter θ? Previous units have introduced you to:

Bias

$$\text{Bias}(T(\mathbf{X})) = \mathsf{E}(T(\mathbf{X})) - \theta$$

Mean Square Error

$$\begin{aligned} \text{MSE}(T(\mathbf{X})) &= \mathsf{E}((T(\mathbf{X}) - \theta)^2) \\ &= \text{Bias}(T(\mathbf{X}))^2 + \text{Var}(T(\mathbf{X})) \end{aligned}$$

Consistency If $T_n(\mathbf{X})$ is the estimator based on $X_1, \ldots, X_n$, then $T_n(\mathbf{X})$ is consistent for θ if for every $\epsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}(|T_n(\mathbf{X}) - \theta| < \epsilon) = 1$$

Nobody wants an estimator which is inconsistent.

Example of a consistent estimator

Suppose $X_1, \ldots, X_n$ iid $U(0, \theta)$ and $T_n = \max_{i=1,\ldots,n} X_i$ (the MLE of θ as we saw in the previous chapter). The CDF of T_n can be derived

$$\begin{aligned} \mathsf{P}(T_n < t) &= \mathsf{P}(X_1 < t \cap X_2 < t \cap \ldots \cap X_n < t) \\ &= \prod_{i=1}^{n} \mathsf{P}(X_i < t) \\ &= \left(\frac{t}{\theta}\right)^n, \ 0 < t < \theta \end{aligned}$$

Hence we can calculate that T_n has expectation $n\theta/(n+1)$, i.e. it is a biased estimator of θ. But is it consistent?

$$\begin{aligned} \mathsf{P}(|T_n - \theta| < \epsilon) &= \mathsf{P}(\theta - T_n < \epsilon) \\ &= \mathsf{P}(T_n > \theta - \epsilon) \\ &= 1 - \left(\frac{\theta - \epsilon}{\theta}\right)^n, \ 0 < \theta - \epsilon < \theta \end{aligned}$$

As $n \to \infty$, $\mathsf{P}(|T_n - \theta| < \epsilon) \to 1$, so although this estimator is biased, it is consistent.

More generally, any estimator $T_n(\mathbf{X})$ which has a MSE which tends to zero as $n \to \infty$ will be consistent:

$$\begin{aligned} \mathsf{P}(|T_n - \theta| > \epsilon) &= \mathsf{P}((T_n - \theta)^2 > \epsilon^2) \\ &\leq \frac{\mathsf{E}((T_n - \theta)^2)}{\epsilon^2} \text{ by Chebychev's inequality.} \end{aligned}$$

In this chapter, we will concentrate on unbiased estimators, in which case, minimising the MSE is equivalent to minimising the variance of the estimator. If we can see that the resulting variance tends to zero as $n \to \infty$, then the estimator we have is consistent.

§2.2 The Rao-Blackwell Theorem

We start with a theorem which allows us to take any unbiased estimator of a function of parameter θ, $\phi(\theta)$, even an inconsistent estimator, and to generate an improved unbiased estimator in the sense that its variance is smaller (or at least, no greater).

Theorem 2.1

The *Rao-Blackwell Theorem* states that if T is an unbiased estimator of $\phi(\theta)$ and S is a sufficient statistic for the underlying distribution $f_\theta(x)$, then setting $\tilde{T} = \mathsf{E}(T|S)$, we have that $\tilde{T}$ is a statistic with

$$\begin{aligned} \mathsf{E}(\tilde{T}) &= \phi(\theta) \\ \text{Var}(\tilde{T}) &\leq \text{Var}(T) \end{aligned}$$

Proof: For notational simplicity, we will assume continuous quantities. There are three parts of the proof:

- **$\mathsf{E}(\tilde{T})$?**

$$\begin{aligned} \tilde{T} &= \mathsf{E}(T|S=s) \\ &= \int_t tf(t|s)dt \text{ ie a function of } s \\ \mathsf{E}(\tilde{T}) &= \int_s \left\{ \int_t tf(t|s)dt \right\} f(s)ds \\ &= \int_s \int_t tf(t,s)dtds \\ &= \int_t \int_s tf(t,s)dsdt \\ &= \int_t t \left\{ \int_s f(t,s)ds \right\} dt \\ &= \int_t tf(t)dt \\ &= \mathsf{E}(T) \end{aligned}$$

So, since T is unbiased for $\phi(\theta)$, so is $\tilde{T}$.

- **Var$(\tilde{T})$?**

$$\begin{aligned} \text{Var}(T) &= \mathsf{E}((T-\phi(\theta))^2) \\ &= \mathsf{E}((T-\tilde{T}+\tilde{T}-\phi(\theta))^2) \\ &= \mathsf{E}((T-\tilde{T})^2) + 2\mathsf{E}((T-\tilde{T})(\tilde{T}-\phi(\theta))) + \mathsf{E}((\tilde{T}-\phi(\theta))^2) \\ &= \mathsf{E}((T-\tilde{T})^2) + 2\mathsf{E}((T-\tilde{T})(\tilde{T}-\phi(\theta))) + \text{Var}(\tilde{T}) \\ \text{now } \mathsf{E}((T-\tilde{T})(\tilde{T}-\phi(\theta))) &= \int_s \int_t (t-\mathsf{E}(T|s))(\mathsf{E}(T|s)-\phi(\theta))f(t,s)dtds \\ &= \int_s \int_t (\mathsf{E}(T|s)-\phi(\theta))(t-\mathsf{E}(T|s))f(t|s)f(s)dtds \\ &= \int_s (\mathsf{E}(T|s)-\phi(\theta)) \left[\int_t (t-\mathsf{E}(T|s))f(t|s)dt \right] f(s)ds \\ \text{and } \int_t (t-\mathsf{E}(T|s))f(t|s)dt &= \int_t tf(t|s)dt - \int_t \mathsf{E}(T|s)f(t|s)dt \end{aligned}$$

$$
\begin{aligned}
&= \mathsf{E}(T|s) - \mathsf{E}(T|s) \\
&= 0 \\
\text{so } \mathsf{E}((T-\tilde{T})(\tilde{T}-\phi(\theta))) &= 0
\end{aligned}
$$

Since $\mathsf{E}((T-\tilde{T})^2) \geq 0$ and $\mathsf{E}((T-\tilde{T})(\tilde{T}-\phi(\theta))) = 0$, this shows that $\text{Var}(T) \geq \text{Var}(\tilde{T})$.

- **Is $\tilde{T}$ a statistic?** Since S is a sufficient statistic, the conditional distribution of $\mathbf{X}|S$ does not depend on θ. Since $T = T(\mathbf{X})$ is an estimator of $\phi(\theta)$ it is a function of $\mathbf{X}$ but not of θ. As a result the distribution of $T(\mathbf{X})|S$ cannot be a function of θ and so $\tilde{T} = \mathsf{E}(T(\mathbf{X})|S)$ is a function of $\mathbf{X}$ via S but not of θ. Hence $\tilde{T}$ is a statistic.

Examples of the Rao-Blackwell theorem

Suppose $X_1, \ldots, X_n$ are iid $Pois(\mu)$ random variables. We can use the Factorisation theorem to see that $\sum_{i=1}^n X_i$ is a sufficient statistic for μ. Consider two different estimation problems (so requiring two different choices of initial unbiased estimator):

Estimating μ We need an estimator whose expectation is μ. If $X \sim Pois(\mu)$, $\mathsf{E}(X) = \mu$, and so a simple unbiased estimator of μ is $T(\mathbf{X}) = X_1$. Notice that although this estimator is unbiased, it is a very poor estimator since its variance is independent of n and it is not consistent. Applying the Rao-Blackwell theorem, $\tilde{T} = \mathsf{E}(X_1|\sum_{i=1}^n X_i = s)$. To calculate this, we first find the conditional distribution of $(X_1|\sum_{i=1}^n X_i = s)$:

$$
\begin{aligned}
\mathsf{P}(X_1 = x_1|\sum_{i=1}^n X_i = s) &= \frac{\mathsf{P}(X_1 = x_1 \cap \sum_{i=1}^n X_i = s)}{\mathsf{P}(\sum_{i=1}^n X_i = s)} \\
&= \frac{\mathsf{P}(X_1 = x_1 \cap \sum_{i=2}^n X_i = s - x_1)}{\mathsf{P}(\sum_{i=1}^n X_i = s)} \\
&= \frac{\mathsf{P}(X_1 = x_1)\mathsf{P}(\sum_{i=2}^n X_i = s - x_1)}{\mathsf{P}(\sum_{i=1}^n X_i = s)}
\end{aligned}
$$

As we know that the sum of m iid $Pois(\mu)$ random variables is a $Pois(m\mu)$,

$$
\begin{aligned}
\mathsf{P}(X_1 = x_1|\sum_{i=1}^n X_i = s) &= \frac{[\mu^{x_1}e^{-\mu}/x_1!][((n-1)\mu)^{s-x_1}e^{-(n-1)\mu}/(s-x_1)!]}{(n\mu)^s e^{-n\mu}/s!} \\
&= \frac{s!}{x_1!(s-x_1)!}\frac{(n-1)^{s-x_1}}{n^s} \\
&= \frac{s!}{x_1!(s-x_1)!}\left(\frac{1}{n}\right)^{x_1}\left(1-\frac{1}{n}\right)^{s-x_1}
\end{aligned}
$$

Then $\tilde{T} = \sum_{x_1} x_1 \mathsf{P}(X_1 = x_1|\sum_{i=1}^n X_i = s)$ using this conditional distribution. However we can save ourselves the work of calculating the summation by noticing that the conditional distribution is the pmf of a $Bin(s, 1/n)$ distribution. As a result:

$$
\begin{aligned}
\tilde{T}(\mathbf{X}) &= \frac{s}{n} \\
&= \frac{\sum_{i=1}^n X_i}{n}
\end{aligned}
$$

In this case, properties of expectation and variance allow us to verify that $\tilde{T}$ is unbiased for μ with $\text{Var}(\tilde{T}) = \text{Var}(T)/n$. $\tilde{T}$ is a consistent estimator of μ.

Estimating $\mathsf{P}(X=0)$ To find an unbiased estimator of $p_0 = \mathsf{P}(X = 0)$, we could use the indicator function of any event which happens with probability p_0, for example:

$$\begin{aligned} T &= I_{[X_1=0]} \\ \mathsf{E}(T) &= \mathsf{P}(I_{[X_1=0]} = 1) \\ &= p_0 \end{aligned}$$

Then applying the Rao-Blackwell theorem

$$\begin{aligned} \tilde{T} &= \mathsf{E}(I_{[X_1=0]} | \sum_{i=1}^{n} X_i = s) \\ &= \mathsf{P}(I_{[X_1=0]} = 1 | \sum_{i=1}^{n} X_i = s) \\ &= \mathsf{P}(X_1 = 0 | \sum_{i=1}^{n} X_i = s) \end{aligned}$$

But we have already found this conditional probability when we found the complete conditional pmf in the first part of this example. So

$$\tilde{T}(\mathbf{X}) = \left(1 - \frac{1}{n}\right)^{\sum_{i=1}^{n} X_i}$$

§2.3 Cramér-Rao lower bound and efficiency

In the previous section, we learned that conditioning an unbiased estimator on a sufficient statistic helps in finding a lower variance unbiased estimator. This raises the question of how low a variance could be achieved by an unbiased estimator.

Theorem 2.2

The *Cramér-Rao Lower Bound* states that if $X_1, \ldots, X_n$ have a joint distribution $f_\theta(\mathbf{x})$ and $T(\mathbf{X})$ is an unbiased estimator of $\phi(\theta)$, then provided

1. $S_\theta = \{\mathbf{x} : f_\theta(\mathbf{x}) > 0\}$ does not depend on θ, and

2. The operations of integration with respect to $\mathbf{x}$ and differentiation with respect to θ can be interchanged:

$$\text{Var}(T(\mathbf{X})) \geq \frac{(\phi'(\theta))^2}{\mathsf{E}\left[\left(\frac{d \ln f_\theta(\mathbf{x})}{d\theta}\right)^2\right]}$$

Proof: Let $U = \frac{d \ln f_\theta(\mathbf{x})}{d\theta}$ then we know that

$$\begin{aligned} \int_x f_\theta(\mathbf{x}) d\mathbf{x} &= 1 \text{ as pdf} && (1) \\ \text{and } \int_x \mathbf{T}(\mathbf{x}) f_\theta(\mathbf{x}) d\mathbf{x} &= \phi(\theta) \text{ as unbiased} && (2) \end{aligned}$$

$$\begin{aligned}
(1) \Rightarrow \quad \frac{d}{d\theta}\int_x f_\theta(\mathbf{x})d\mathbf{x} &= 0 \\
\Rightarrow \quad \int_x \frac{df_\theta(\mathbf{x})}{d\theta}d\mathbf{x} &= 0 \\
\Rightarrow \quad \int_x \frac{1}{f_\theta(\mathbf{x})}\frac{df_\theta(\mathbf{x})}{d\theta}f_\theta(\mathbf{x})d\mathbf{x} &= 0 \\
\Rightarrow \quad \int_x \frac{d\ln f_\theta(\mathbf{x})}{d\theta}f_\theta(\mathbf{x})d\mathbf{x} &= 0 \\
\Rightarrow \quad \mathsf{E}(U) &= 0 \qquad (3) \\
(2) \Rightarrow \quad \frac{d}{d\theta}\int_x T(\mathbf{x})f_\theta(\mathbf{x})d\mathbf{x} &= \phi'(\theta) \\
\Rightarrow \quad \int_x T(\mathbf{x})\frac{df_\theta(\mathbf{x})}{d\theta}d\mathbf{x} &= \phi'(\theta) \\
\Rightarrow \quad \int_x T(\mathbf{x})\frac{d\ln f_\theta(\mathbf{x})}{d\theta}f_\theta(\mathbf{x})d\mathbf{x} &= \phi'(\theta) \\
\Rightarrow \quad \mathsf{E}(TU) &= \phi'(\theta) \qquad (4) \\
\text{Now } \mathrm{Cov}(UT) &= \mathsf{E}\left[(U-\mathsf{E}(U))(T-\mathsf{E}(T))\right] \\
&= \mathsf{E}\left[U(T-\phi(\theta))\right] \text{ by (3) and bias} \\
&= \mathsf{E}(UT) - \phi(\theta)\mathsf{E}(U) \\
&= \phi'(\theta) \text{ by (3) and (4)} \\
\text{The Cauchy-Schwartz inequality} \Rightarrow \quad (\mathrm{Cov}(UT))^2 &\leq \mathrm{Var}(U)\mathrm{Var}(T) \\
\Rightarrow \quad \mathrm{Var}(T) &\geq \frac{(\phi'(\theta))^2}{\mathrm{Var}(U)} \\
\text{But } \mathrm{Var}(U) &= \mathsf{E}(U^2) \\
&= \mathsf{E}\left[\left(\frac{d\ln f_\theta(\mathbf{x})}{d\theta}\right)^2\right]
\end{aligned}$$

- From the proof of the CRLB, we already have that

$$\begin{aligned}
\int_x \frac{d\ln f_\theta(\mathbf{x})}{d\theta}f_\theta(\mathbf{x})d\mathbf{x} &= 0 \\
\Rightarrow \quad \frac{d}{d\theta}\int_x \frac{d\ln f_\theta(\mathbf{x})}{d\theta}f_\theta(\mathbf{x})d\mathbf{x} &= 0 \\
\Rightarrow \quad \int_x \left[\frac{d^2\ln f_\theta(\mathbf{x})}{d\theta^2}f_\theta(\mathbf{x}) + \frac{d\ln f_\theta(\mathbf{x})}{d\theta}\frac{df_\theta(\mathbf{x})}{d\theta}\right]d\mathbf{x} &= 0 \\
\Rightarrow \quad \int_x \left[\frac{d^2\ln f_\theta(\mathbf{x})}{d\theta^2}f_\theta(\mathbf{x}) + \frac{d\ln f_\theta(\mathbf{x})}{d\theta}\frac{d\ln f_\theta(\mathbf{x})}{d\theta}f_\theta(\mathbf{x})\right]d\mathbf{x} &= 0 \\
\text{i.e. } \mathsf{E}\left[\frac{d^2\ln f_\theta(\mathbf{x})}{d\theta^2}\right] + \mathsf{E}\left[\left(\frac{d\ln f_\theta(\mathbf{x})}{d\theta}\right)^2\right] &= 0
\end{aligned}$$

- From the result above and the fact that the independence means that the joint distribution is the product of the marginals

$$I_n(\theta) = -\mathsf{E}\left[\frac{d^2\ln f_\theta(\mathbf{x})}{d\theta^2}\right]$$

$$
\begin{aligned}
&= -\mathsf{E}\left[\frac{d^2 \ln \prod_{i=1}^n f_\theta(x_i)}{d\theta^2}\right] \\
&= -\mathsf{E}\left[\frac{d^2 \sum_{i=1}^n \ln f_\theta(x_i)}{d\theta^2}\right] \\
&= -\mathsf{E}\left[\sum_{i=1}^n \frac{d^2 \ln f_\theta(x_i)}{d\theta^2}\right] \\
&= \sum_{i=1}^n -\mathsf{E}\left[\frac{d^2 \ln f_\theta(x_i)}{d\theta^2}\right] \\
&= nI_1(\theta)
\end{aligned}
$$

Before seeing some examples of the theorem in action, we make one further definition:

Definition 2.1

The *Efficiency* of an unbiased estimator T is defined to be the ratio of the Cramér-Rao Lower Bound to the variance of T.

Examples of the Cramér-Rao Lower Bound and efficiency

Normal Suppose $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$ random variables with σ^2 known.

$$
\begin{aligned}
f_\mu(\mathbf{x}) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2) \\
\ln(f_\mu(\mathbf{x})) &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2 \\
\frac{d\ln(f_\mu(\mathbf{x}))}{d\mu} &= \frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu) \\
\frac{d^2\ln(f_\mu(\mathbf{x}))}{d\mu^2} &= \frac{1}{\sigma^2}\sum_{i=1}^n -1 \\
\text{So } I_n(\mu) &= -\mathsf{E}(-n/\sigma^2) \\
&= n/\sigma^2
\end{aligned}
$$

So, if the goal is to estimate μ, then the CRLB is σ^2/n. In this Normal case, we know that $\bar{\mathbf{X}}$ is an unbiased estimator of μ which has a variance of σ^2/n. In other words, $\bar{\mathbf{X}}$ has an efficiency of 1.

Exponential Suppose $X_1, \ldots, X_n$ iid $Exp(\mu)$ random variables where μ is the mean parameter.

$$
\begin{aligned}
f_\mu(\mathbf{x}) &= \mu^{-n}\exp(-\frac{\sum_{i=1}^n x_i}{\mu}) \\
\ln(f_\mu(\mathbf{x})) &= -n\ln\mu - \frac{\sum_{i=1}^n x_i}{\mu} \\
\frac{d\ln(f_\mu(\mathbf{x}))}{d\mu} &= \frac{-n}{\mu} + \frac{\sum_{i=1}^n x_i}{\mu^2}
\end{aligned}
$$

$$\begin{aligned}\frac{d^2\ln(f_\mu(\mathbf{x}))}{d\mu^2} &= \frac{n}{\mu^2} - \frac{2\sum_{i=1}^n x_i}{\mu^3} \\ \text{So } I_n(\mu) &= -\frac{n}{\mu^2} + \frac{2\sum_{i=1}^n \mathsf{E}(X_i)}{\mu^3} \\ &= n/\mu^2\end{aligned}$$

1. **Estimating the mean** μ The CRLB for estimating μ is simply $1/I_n(\mu) = \mu^2/n$. One known unbiased estimator of μ is $\bar{\mathbf{X}}$ and, as with the previous example, we can see that this has efficiency 1 since the variance of an $Exp(\mu)$ is μ^2.
2. **Estimating the rate** $1/\mu$ In this case, $\phi(\mu) = \mu^{-1}$, so $\phi'(\mu) = -\mu^{-2}$. The CRLB is thus $1/(n\mu^2)$. We do have an unbiased estimator of $1/\mu$ from Problem Sheet 2; to find its efficiency we need to find its variance:

$$\begin{aligned}T(\mathbf{X}) &= \frac{n-1}{\sum_{i=1}^n X_i} \\ \text{where } \mathsf{E}(T) &= \frac{1}{\mu} \text{ and } \sum_{i=1}^n X_i \sim Gam(1/\mu, n) \\ \mathsf{E}\left(\frac{1}{\sum_{i=1}^n X_i}\right)^2 &= \int_0^\infty \frac{1}{y^2}\frac{(1/\mu)^n}{\Gamma(n)} y^{n-1}e^{-y}dy \\ &= \frac{(1/\mu)^2\Gamma(n-2)}{\Gamma(n)}\int_0^\infty \frac{1}{y^2}\frac{(1/\mu)^{n-2}}{\Gamma(n-2)} y^{n-3}e^{-y}dy \\ &= \frac{(1/\mu)^2\Gamma(n-2)}{\Gamma(n)} \\ &= \frac{(n-3)!}{(n-1)!\mu^2} \\ \text{and thus } Var(T) &= (n-1)^2\mathsf{E}\left(\frac{1}{\sum_{i=1}^n X_i}\right)^2 - \frac{1}{\mu^2} \\ &= \frac{1}{(n-2)\mu^2}\end{aligned}$$

So, this estimator does not attain the CRLB; its efficiency is $\frac{n-2}{n}$.

§2.4 Minimum variance unbiased estimators

How do we know whether or not there will be an unbiased estimator which attains the Cramér-Rao Lower Bound?

Theorem 2.3

An unbiased estimator $T(\mathbf{x})$ of $\phi(\theta)$ attains the Cramér-Rao Lower Bound if and only if we can factorise:

$$\frac{d\ln f_\theta(\mathbf{x})}{d\theta} = a(\theta)(T(\mathbf{x}) - \phi(\theta))$$

for some function $a(\theta)$.

Proof: Considering the proof of the Cramér-Rao Lower Bound, equality is achieved if and only if

$$\begin{aligned}(\text{Cov}(U,T))^2 &= \text{Var}(U)\text{Var}(T) \\ \text{where } U &= \frac{d \ln f_\theta(\mathbf{x})}{d\theta}\end{aligned}$$

This occurs only when U is a linear function of T, ie

$$U(\mathbf{x}, \theta) = a(\theta)T(\mathbf{x}) + b(\theta)$$

for some functions of θ, $a(\theta)$ and $b(\theta)$. Since $\mathsf{E}(U(\mathbf{X}, \theta)) = 0$ (from the proof of the Cramér-Rao Lower Bound again) and $\mathsf{E}(T(\mathbf{X})) = \phi(\theta)$ (since it is unbiased),

$$\begin{aligned}\mathsf{E}(U(\mathbf{X}, \theta)) &= a(\theta)\mathsf{E}(T(\mathbf{X})) + b(\theta) \\ \Rightarrow\ 0 &= a(\theta)\phi(\theta) + b(\theta) \\ \text{i.e. } b(\theta) &= -a(\theta)\phi(\theta) \\ \text{and so } \frac{d \ln f_\theta(\mathbf{x})}{d\theta} &= a(\theta)(T(\mathbf{x}) - \phi(\theta))\end{aligned}$$

Examples

We will return to the three examples of the previous section illustrating the CRLB

Normal

$$\begin{aligned}\frac{d \ln(f_\mu(\mathbf{x}))}{d\mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) \\ &= \frac{n}{\sigma^2}(\bar{\mathbf{x}} - \mu)\end{aligned}$$

This confirms that $\bar{\mathbf{X}}$ is an unbiased estimator of μ which attains the CRLB.

Exponential

$$\begin{aligned}\frac{d \ln(f_\mu(\mathbf{x}))}{d\mu} &= \frac{-n}{\mu} + \frac{\sum_{i=1}^n x_i}{\mu^2} \\ &= \frac{n}{\mu^2}(\bar{\mathbf{x}} - \mu) \\ &= n(\frac{\sum_{i=1}^n x_i}{n\mu^2} - \frac{1}{\mu})\end{aligned}$$

This confirms two different points. Firstly that $\bar{\mathbf{X}}$ is an unbiased estimator of μ which attains the CRLB. Second, no unbiased estimator of $1/\mu$ exists which attains the corresponding CRLB.

Theorem 2.3 gives us a relatively simple way to find fully efficient unbiased estimators, if they exist. We can also say more about any estimator which attains the CRLB:

Theorem 2.4

If there is a statistic attaining the CRLB, then that statistic is sufficient.
Proof: If the CRLB is attained, then

$$\begin{aligned}
\frac{d\ln f_\theta(\mathbf{x})}{d\theta} &= a(\theta)(T(\mathbf{x}) - \phi(\theta)) \\
\ln f_\theta(\mathbf{x}) &= a^*(\theta)T(\mathbf{x}) + b^*(\theta) + c^*(\mathbf{x}) \\
\text{where } \frac{da^*(\theta)}{d\theta} &= a(\theta), \\
\frac{db^*(\theta)}{d\theta} &= -a(\theta)\phi(\theta), \\
\text{and } \frac{dc^*(\mathbf{x})}{d\theta} &= 0. \\
\text{Exponentiating } f_\theta(\mathbf{x}) &= \exp(a^*(\theta)T(\mathbf{x}) + b^*(\theta) + c^*(\mathbf{x})) \\
&= \exp(a^*(\theta)T(\mathbf{x}) + b^*(\theta)) \exp(c^*(\mathbf{x}))
\end{aligned}$$

and so by the Factorisation theorem we can see that $T(\mathbf{X})$ is a sufficient statistic.

What do we do if Theorem 2.3 shows us that no fully efficient unbiased estimator exists for a particular estimation problem? First we will need one rather technical definition:

Definition 2.2

A sufficient statistic S for θ in the family of distributions indexed by $\theta \in \Theta$ is *complete* if

$$\mathsf{E}(g(S)) = 0 \ \ \forall\theta \in \Theta \ \ \Rightarrow \ \ g(s) = 0 \ \text{ (almost everywhere).}$$

Theorem 2.5

Regardless of the attainment of the CRLB, provided the sufficient statistic S is complete, the Minimum Variance Unbiased Estimator of $\phi(\theta)$ is $\tilde{T} = \mathsf{E}(T|S)$ where T is any unbiased estimator of $\phi(\theta)$. That is, we should apply the Rao-Blackwell theorem.
Proof: Suppose, for contradiction, that this is not true, that is there is some other estimator T' such that $\mathsf{E}(T') = \phi(\theta)$ and

$$\text{Var}(T') < \text{Var}(\tilde{T}) \tag{5}$$

Apply the Rao-Blackwell theorem to T' giving $\tilde{T}' = \mathsf{E}(T'|S = s)$, then $\mathsf{E}(\tilde{T}') = \phi(\theta)$ and

$$\text{Var}(\tilde{T}') \leq \text{Var}(T') < \text{Var}(\tilde{T}) \tag{6}$$

Both $\tilde{T}$ and $\tilde{T}'$ are functions of the sufficient statistic since both are the result of an application of the Rao-Blackwell theorem. Let $g(S) = \tilde{T} - \tilde{T}'$, then

$$\begin{aligned}
\mathsf{E}(g(S)) &= \mathsf{E}(\tilde{T}) - \mathsf{E}(\tilde{T}') \\
&= \phi(\theta) - \phi(\theta) \\
&= 0 \ \forall\theta.
\end{aligned}$$

Since S is a complete sufficient statistic, this implies that $g(S) = 0$, i.e. that $\tilde{T} = \tilde{T}'$. But this contradicts Equation (6) and therefore our assumption that we can find an unbiased estimator with strictly smaller variance than $\tilde{T}$ is false.