

Cases - Study for Exam 1

Ines Pancorbo

Case 1: Find and discuss an example of improving an estimator using Rao-Blackwell.

The Rao-Blackwell Theorem states that if W is any unbiased estimator of $\tau(\theta)$, and T is any sufficient statistic for θ then $\mathbb{E}(W|T)$ is a uniformly better unbiased estimator of $\tau(\theta)$, i.e., $\mathbb{E}(\mathbb{E}(W|T)) = \tau(\theta)$ and $\text{Var}(\mathbb{E}(W|T)) \leq \text{Var}(W)$ for all θ .

Let $\mu > 0$ be a parameter. Consider X_1, X_2, \dots, X_n i.i.d Poisson(μ).

Consider the unbiased estimator of μ , X_1 (note: $\mathbb{E}(X_1) = \mu$). We can improve this estimator using the Rao-Blackwell Theorem by finding a sufficient statistic, T , for μ and then considering the statistic $\mathbb{E}(X_1|T)$.

The poisson distribution belongs to the family of exponential distributions. To show this notice,

$$f(x|\mu) = \frac{e^{-\mu}\mu^x}{x!}\mathbf{1}(x)_{\{0,1,2,\dots\}} = e^{-\mu}\left(\frac{1}{x!}\mathbf{1}(x)_{\{0,1,2,\dots\}}\right)\mu^x = e^{-\mu}\left(\frac{1}{x!}\mathbf{1}(x)_{\{0,1,2,\dots\}}\right)e^{\ln(\mu)x}$$

And so the density is given by,

$$h(x) = \frac{1}{x!}\mathbf{1}(x)_{\{0,1,2,\dots\}}, \quad c(\mu) = e^{-\mu}, \quad \text{and} \quad e^{w_1(\mu)t_1(x)} = e^{\ln(\mu)x}$$

This implies, by Theorem 6.2.10 of Casella and Berger, that $\sum_{i=1}^n t_1(X_i) = \sum_{i=1}^n X_i$ is a sufficient statistic for μ .

So let $T = \sum_{i=1}^n X_i$.

Now let us consider,

$$\mathbb{E}(X_1|T) = \mathbb{E}(X_1|\sum_{i=1}^n X_i = t)$$

and find a simpler expression for this statistic.

Notice that

$$\mathbb{E}\left(\sum_{i=1}^n X_i \mid \sum_{i=1}^n X_i = t\right) = t$$

and that

$$t = \mathbb{E}\left(\sum_{i=1}^n X_i \mid \sum_{i=1}^n X_i = t\right) = \sum_{i=1}^n \mathbb{E}(X_i \mid \sum_{i=1}^n X_i = t)$$

by linearity properties of expectation.

Further, because X_1, X_2, \dots, X_n are iid,

$$\sum_{i=1}^n \mathbb{E}(X_i \mid \sum_{i=1}^n X_i = t) = t \implies \mathbb{E}(X_i \mid \sum_{i=1}^n X_i = t) = \frac{t}{n}.$$

And so $\mathbb{E}(X_1|T) = \frac{T}{n} = \bar{X}$. So \bar{X} is a uniformly better unbiased estimator of μ than X_1 , in the sense that both are unbiased,

$$\mathbb{E}(\bar{X}) = \mu = \mathbb{E}(X_1)$$

but,

$$\text{Var}(\bar{X}) = \frac{\mu}{n} \leq \mu = \text{Var}(X_1).$$

Just as an aside note: This example can be a bit circular if we choose a sample size of 1, i.e., we let $n = 1$. If so, $T = X_1$ and by Rao-Blackwell we will just get back the same estimator, X_1 .

Case 2: Find and discuss an example of a sufficient statistic that is not complete

Let $\theta > 0$ be a parameter. Let X_1, X_2, \dots, X_n i.i.d Uniform $[-\theta, 2\theta]$. So the density of X_i is,

$$f_{X_i}(x_i) = \frac{1}{3\theta} \mathbf{1}(x_i)_{[-\theta, 2\theta]}$$

This is a case in which the statistic $(X_{(1)}, X_{(n)})$ although sufficient for θ is not complete.

Let us show that it is sufficient for θ using the Factorization Theorem:

$$f(\underline{x}|\theta) = \prod_{i=1}^n \frac{1}{3\theta} \mathbf{1}(x_i)_{[-\theta, 2\theta]} = \frac{1}{(3\theta)^n} \prod_{i=1}^n \mathbf{1}(x_i)_{[-\theta, 2\theta]} = \frac{1}{(3\theta)^n} \mathbf{1}(x_{(1)})_{[-\theta, \infty)} \mathbf{1}(x_{(n)})_{(-\infty, 2\theta]}$$

And so by letting,

$$h(x) = 1, \text{ and } g((X_{(1)}, X_{(n)})|\theta) = \frac{1}{(3\theta)^n} \mathbf{1}(x_{(1)})_{[-\theta, \infty)} \mathbf{1}(x_{(n)})_{(-\infty, 2\theta]}$$

we have that $(X_{(1)}, X_{(n)})$ is a sufficient statistic for θ .

Now let us show $(X_{(1)}, X_{(n)})$ is not complete. We can show this by considering an interesting statistic given our random sample: $T = \max\left(-X_{(1)}, \frac{X_{(n)}}{2}\right)$.

T is sufficient for θ as well by the Factorization Theorem,

$$f(\underline{x}|\theta) = \prod_{i=1}^n \frac{1}{3\theta} \mathbf{1}(x_i)_{[-\theta, 2\theta]} = \frac{1}{(3\theta)^n} \prod_{i=1}^n \mathbf{1}(x_i)_{[-\theta, 2\theta]} = \frac{1}{(3\theta)^n} \mathbf{1}(\theta)_{[\max(-x_{(1)}, \frac{x_{(n)}}{2}), \infty)}$$

since

$$-\theta \leq x_i \leq 2\theta \implies -\theta \leq x_{(1)} \leq \dots \leq x_{(n)} \leq 2\theta \implies \theta \geq -x_{(1)}, \theta \geq \frac{x_{(n)}}{2}$$

Let us find the distribution of T .

For $0 \leq t \leq \theta$, we have

$$P(T \leq t) = P(-t \leq X_{(1)}, X_{(n)} \leq 2t) = P(-t \leq X_1, X_2, \dots, X_n \leq 2t) = (P(-t < X_1 < 2t))^n = \left(\frac{t}{\theta}\right)^n$$

So T has density,

$$f_T(t) = \frac{nt^{n-1}}{\theta^n} \mathbf{1}(t)_{[0, \theta]}$$

This means T is distributed as $Y_{(n)}$ where Y_1, \dots, Y_n are i.i.d Uniform $[0, \theta]$ and studying the properties of T as an estimator of θ reduces to studying the properties of $Y_{(n)}$. And we have shown in class that $Y_{(n)}$ is a complete statistic for θ . Let us show it again:

Let $g(Y_{(n)})$ be such that $E(g(Y_{(n)})) = 0$ for every $\theta > 0$ and note that $E(g(Y_{(n)})) = n\theta^{-n}G(\theta)$, with

$$G(\theta) = \int_0^\theta t^{n-1} g(t) dt.$$

G is differentiable almost everywhere and $G'(\theta) = \theta^{n-1}g(\theta)$. If $G(\theta) = 0$ for every $\theta > 0$ then $G' = 0$ and hence, $g = 0$. So $Y_{(n)}$ is complete for θ . This means that T is complete for θ .

Further, since T is sufficient and complete, it is minimally sufficient. Now noting that T is a function of $(X_{(1)}, X_{(n)})$ it is clear that $(X_{(1)}, X_{(n)})$ is not minimally sufficient (by definition of minimally sufficient statistic). And consequently, $(X_{(1)}, X_{(n)})$ cannot be complete.

There are other ways to show $(X_{(1)}, X_{(n)})$ is not complete. However, doing it the way done above, by considering the statistic T , we get a bonus: We can notice that T is not only complete and sufficient for θ , but also the MLE for θ and by Lehmann-Scheffe, $\frac{n+1}{n}T$ is the UMVUE of θ .

Case 4: Find and discuss an example where the maximum likelihood estimator (MLE) does not exist

Let $\theta > 0$ be a parameter. Let X_1, X_2, \dots, X_n i.i.d Uniform(0, θ). So the density of X_i is,

$$f_{X_i}(x_i) = \frac{1}{\theta} \mathbf{1}_{(x_i)_{(0, \theta)}}$$

Note,

$$0 < x_i < \theta \implies 0 < x_{(1)} \leq \dots \leq x_{(n)} < \theta \implies \theta > x_{(n)}$$

Let us try and find the MLE. Consider the likelihood function,

$$L(\theta|\underline{x}) = \left(\prod_{i=1}^n \frac{1}{\theta} \right) \mathbf{1}_{(\theta)_{(x_{(n)}, \infty)}} = \frac{1}{\theta^n} \mathbf{1}_{(\theta)_{(x_{(n)}, \infty)}}$$

Clearly, $L(\theta|\underline{x})$ is a strictly monotone decreasing function of θ defined on $(x_{(n)}, \infty)$. Clearly, the supremum of $L(\theta|\underline{x})$ is at $\theta = x_{(n)}$. Since $L(\theta|\underline{x})$ is not defined at $\theta = x_{(n)}$, it does not achieve a maximum value on $(x_{(n)}, \infty)$. Consequently, the MLE for θ does not exist for a random sample of uniform (continuous) random variables X_i , $i = 1, \dots, n$, with support $(0, \theta)$.

Takeaways from this example: According to Casella and Berger, maximum likelihood estimation consists in estimating parameters of a probability distribution by maximizing the likelihood function so that under this model the data is most probable. The point in the parameter space that maximizes the likelihood function is called the MLE.

The idea of maximum likelihood is intuitive but can have a clear shortcoming as illustrated with the above example: The likelihood function must have a maximum. In the example given above, we were dealing with a strictly monotone decreasing continuous function defined on an open interval, and so, because of the non-compact support, it did not achieve its supremum (Extreme Value Theorem).

Case 7: Find and discuss an example where the estimator is consistent but not unbiased

Consider X_1, X_2, \dots, X_n i.i.d $N(\mu, \sigma^2)$, where both parameters are unknown, and consider estimation of σ^2 .

Consider the MLE for σ^2 . From example 0.0.4 of the notes, we know it is,

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

We can show that $\hat{\sigma}_n^2$ is biased for σ^2 . Consider the sample variance i.e.,

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Notice that $\hat{\sigma}_n^2 = \frac{n-1}{n} S_n^2$ and we know S_n^2 is unbiased for σ^2 (showed in problem 5 of HW 2).

Then we have that,

$$\mathbb{E}(\hat{\sigma}_n^2) = \mathbb{E}\left(\frac{n-1}{n} S_n^2\right) = \frac{n-1}{n} \sigma^2.$$

So $\mathbb{E}(\hat{\sigma}_n^2) \neq \sigma^2$ and so is a biased estimator of σ^2 .

Now, let us show that $\hat{\sigma}_n^2$ is consistent. Let us find the bias of $\hat{\sigma}_n^2$,

$$\text{Bias}(\hat{\sigma}_n^2) = \mathbb{E}(\hat{\sigma}_n^2) - \sigma^2 = \frac{n-1}{n} \sigma^2 - \sigma^2 = \sigma^2 \left(1 - \frac{1}{n} - 1\right) = -\frac{\sigma^2}{n}.$$

Now note,

$$\lim_{n \rightarrow \infty} \text{Bias}(\hat{\sigma}_n^2) = \lim_{n \rightarrow \infty} \left(-\frac{\sigma^2}{n}\right) = 0.$$

Let us find the variance of $\hat{\sigma}_n^2$.

First note that $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$ as given in the lecture notes. This implies,

$$\text{Var}\left(\frac{(n-1)S_n^2}{\sigma^2}\right) = 2(n-1) \implies \text{Var}(S_n^2) = \frac{2\sigma^4}{n-1}$$

$$\text{Var}(\hat{\sigma}_n^2) = \text{Var}\left(\frac{n-1}{n} S_n^2\right) = \frac{(n-1)^2}{n^2} \frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^4}{n^2}$$

And so,

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\sigma}_n^2) = \lim_{n \rightarrow \infty} \left(\frac{2(n-1)\sigma^4}{n^2}\right) = 0.$$

And so, since $\text{Bias}(\hat{\sigma}_n^2) \rightarrow 0$, $\text{Var}(\hat{\sigma}_n^2) \rightarrow 0$ as $n \rightarrow \infty$ we have that $\hat{\sigma}_n^2$ is consistent for σ^2 . Therefore, we have shown that $\hat{\sigma}_n^2$ is consistent but not unbiased.