Cramér-Rao Lower Bound

Theorem 1.1 *Let X_i be independent and identically distributed such that $\mathbf{X} \sim f_{\mathbf{X}}(\mathbf{x}, \theta)$ and let $T(\mathbf{X})$ be an estimator for θ. Then (under certain regularity conditions)*

$$\operatorname{Var}(T(\mathbf{X})) \geq \frac{[1 + b_T'(\theta)]^2}{I(\theta)}$$

where

$$\begin{aligned} I(\theta) &= \operatorname{E}\left[\left(\frac{\partial \ln(f_{\mathbf{X}}(\mathbf{x}, \theta))}{\partial \theta}\right)^2\right] \\ &= -\operatorname{E}\left[\frac{\partial^2 \ln(f_{\mathbf{X}}(\mathbf{x}, \theta))}{\partial \theta^2}\right] \end{aligned}$$

For any estimator of θ within the family of bias, $b_T(\theta)$, we now have a lower bound for the variance. A couple of special cases exist:

- For unbiased estimators we have $b_T(\theta) = 0 \Rightarrow b_T'(\theta) = 0 \Rightarrow \operatorname{Var}(T) \geq \frac{1}{I(\theta)}$
- For $T(\mathbf{X}) = c$ a constant, $b_T(\theta) = c - \theta \Rightarrow b_T'(\theta) = -1 \Rightarrow \operatorname{Var}(T) \geq 0$

Proof 1.1 *(Unidimensional) Let $V = \frac{\partial \ln(f(\mathbf{x}, \theta))}{\partial \theta}$ (assumed to exist by regularity). Then*

$$\begin{aligned} \operatorname{E}(V) &= \int_{\mathfrak{X}} \frac{1}{f(\mathbf{x}, \theta)} \frac{\partial f(\mathbf{x}, \theta)}{\partial \theta} f(\mathbf{x}, \theta) dx \quad \textit{chain rule} && (1) \\ &= \frac{\partial}{\partial \theta} \int_{\mathfrak{X}} f(\mathbf{x}, \theta) dx \quad \textit{assumes interchangeability} && (2) \\ &= 0 \quad \textit{since derivative of constant} && (3) \end{aligned}$$

Thus,

$$\begin{aligned} \operatorname{Cov}(V, T) &= \operatorname{E}(VT) - \operatorname{E}(V)\operatorname{E}(T) && (4) \\ &= \operatorname{E}(VT) - 0 \quad \textit{since } \operatorname{E}(V) = 0 && (5) \\ &= \operatorname{E}\left[T \frac{\partial(\ln(f))}{\partial \theta}\right] && (6) \\ &= \int_{\mathfrak{X}} t(\mathbf{x}) \frac{1}{f(\mathbf{x}, \theta)} \frac{\partial f(\mathbf{x}, \theta)}{\partial \theta} f(\mathbf{x}, \theta) dx && (7) \\ &= \frac{\partial}{\partial \theta} \int_{\mathfrak{X}} t(\mathbf{x}) f(\mathbf{x}, \theta) dx \quad \textit{by assumed interchangeability} && (8) \\ &= \frac{\partial}{\partial \theta} \operatorname{E}(T) && (9) \\ &= \frac{\partial}{\partial \theta} [\theta + b_T(\theta)] \quad \textit{since } b_T(\theta) = \operatorname{E}(T) - \theta && (10) \\ &= 1 + b_T'(\theta) && (11) \end{aligned}$$

We also have that

$$\begin{aligned} \operatorname{Var}(V) &= \operatorname{E}\left[(V - \operatorname{E}(V))^2\right] && (12) \\ &= \operatorname{E}\left(V^2\right) && (13) \\ &= \operatorname{E}\left[\left(\frac{\partial \ln(f)}{\partial \theta}\right)^2\right] && (14) \end{aligned}$$

Now, by the above

$$\rho^2 = \frac{(\mathrm{Cov}(V,T))^2}{\mathrm{Var}(V)\mathrm{Var}(T)} \tag{15}$$

$$\leq 1 \tag{16}$$

This implies that

$$\mathrm{Var}(T) \geq \frac{[1+b_T'(\theta)]^2}{\mathrm{E}\left[\left(\frac{\partial \ln(f)}{\partial\theta}\right)^2\right]} \tag{17}$$

Lemma 1.1 $I(\theta) = \mathrm{E}\left(V^2\right) = -\mathrm{E}\left(\frac{\partial V}{\partial\theta}\right) = -\mathrm{E}\left(\frac{\partial^2 \ln(f)}{\partial\theta^2}\right)$

Proof 1.2 *It is apparent that* $\frac{\partial V}{\partial\theta} = \frac{\partial}{\partial\theta}\left(\frac{\partial \ln(f)}{\partial\theta}\right) = \frac{\partial^2 \ln(f)}{\partial\theta^2} = \frac{\partial}{\partial\theta}\frac{f'}{f} = \frac{f''f-(f')^2}{f^2} = \frac{f''}{f} - V^2$. *However,* $\mathrm{E}\left(\frac{f''}{f}\right) = \frac{\partial^2}{\partial\theta^2}\int_{\mathfrak{X}} f(\mathbf{x},\theta)dx = 0$ *under regularity conditions. Therefore* $\mathrm{E}\left(\frac{\partial V}{\partial\theta}\right) = -\mathrm{E}(V^2) = -I(\theta)$.

Lemma 1.2 *For* X_i *iid* $f_X(x,\theta)$, $I_{X_1\cdot X_2\cdots X_n}(\theta) = nI_{X_i}(\theta)$.

Example 1.1 *Let* $X_i \overset{iid}{\sim} N\left(\mu_0,\sigma^2\right)$ *so that* $\theta=\sigma^2$. *Then*

$$f_X(x,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu_0)^2\right) \tag{18}$$

so that

$$\ln(f) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\theta) - \frac{1}{2\theta}(x-\mu_0)^2 \tag{19}$$

Thus

$$\frac{\partial \ln(f)}{\partial\theta} = -\frac{1}{2\theta} + \frac{(x-\mu_0)^2}{2\theta^2} \tag{20}$$

$$= \frac{1}{2\theta^2}\left(-\theta + (x-\mu_0)^2\right) \tag{21}$$

Taking the expectation gives

$$\mathrm{E}\left(\left(\frac{\partial \ln(f)}{\partial\theta}\right)^2\right) = \mathrm{E}\left[\frac{1}{4\theta^4}\left(\theta^2 - 2\theta(x-\mu_0)^2 + (x-\mu_0)^4\right)\right] \tag{22}$$

$$= \frac{1}{4\theta^4}\left[\theta^2 - 2\theta^2 + \mathrm{E}\left((x-\mu_0)^4\right)\right] \tag{23}$$

$$= \frac{1}{4\theta^2}\left[-1 + \mathrm{E}\left(\left(\frac{x-\mu_0}{\sqrt{\theta}}\right)^4\right)\right] \tag{24}$$

Noting that $Y = \frac{x-\mu_0}{\sqrt{\theta}} \sim N(0,1)$ *we can use its mgf to find the expected value. That is, since* $M_Y(s) = e^{s^2}/2$ *then* $M^{(4)}(s)|_{s=0} = 3$. *Thus*

$$\mathrm{E}\left(\left(\frac{\partial \ln(f)}{\partial\theta}\right)^2\right) = \frac{1}{4\theta^2}[-1+3] \tag{25}$$

$$= \frac{1}{2\theta^2} \tag{26}$$

$$= \frac{1}{2\sigma^4} \tag{27}$$

$$= I_{X_i}(\theta) \tag{28}$$

Using this information provides

$$\begin{aligned} CRLB &= \frac{1}{I_n(\theta)} \quad (29)\\ &= \frac{1}{n/(2\sigma^4)} \quad (30)\\ &= \frac{2\sigma^4}{n} \quad (31) \end{aligned}$$

Now, $\mathrm{E}\left[(X-\mu_0)^2\right] = \sigma^2$ *and* $\mathrm{E}\left[\left((X-\mu_0)^2\right)^2\right] = 3\sigma^4$ *by mgf. So,* $\mathrm{Var}\left((X-\mu_0)^2\right) = 3\sigma^4 - (\sigma^2)^2 = 2\sigma^4$. *Hence,*

$$\begin{aligned} \mathrm{Var}\left(\frac{\sum(X_i-\mu_0)^2}{n}\right) &= \frac{2\sigma^4}{n} \quad (32)\\ &= \frac{1}{I_n(\theta)} \quad (33)\\ &= CRLB \quad (34) \end{aligned}$$

From this we deduce that $T(\mathbf{X}) = \frac{\sum(X_i-\mu_0)^2}{n}$ *is efficient. As an aside, note that*

$$\begin{aligned} \mathrm{E}(T) &= \frac{\sum \mathrm{E}\left[(X_i - \mathrm{E}(X_i))^2\right]}{n} \quad (35)\\ &= \frac{n}{n}\sigma^2 \quad (36)\\ &= \sigma^2 \quad (37) \end{aligned}$$

which implies that T *is unbiased for* $\theta = \sigma^2$ *when* μ_0 *is known.*

When is the CRLB attained?

Theorem 1.2 *Let* $\mathbf{X} \sim f(\mathbf{x}, \theta)$ *and* $T(\mathbf{X})$ *be an estimator of* θ. *Then* $T(\mathbf{X})$ *attains the CRLB iff* $f(\mathbf{x}, \theta) = h(\mathbf{x})c(\theta)e^{T(\mathbf{x})\omega(\theta)}$.

One more piece of good news.

Theorem 1.3 *Under regularity conditions, the mle is asymptotically normal and attains the CRLB.*