Note on Cramér-Rao Inequality
March 12, 2023

- (Cramér-Rao Inequality, general case) Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be a random vector, with the joint density $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}}(y_1, \ldots, y_n)$. Let $\hat{\theta}$ be an estimator of θ. Then the *Cramér-Rao inequality* is given by

$$\operatorname{Var}(\hat{\theta}) \geq \frac{\left(\frac{\partial}{\partial\theta}E(\hat{\theta})\right)^2}{E\left(\left(\frac{\partial}{\partial\theta}\ln f_{\mathbf{Y}}(\mathbf{Y})\right)^2\right)}$$

  - If the components of $\mathbf{Y} = (Y_1, \ldots, Y_n)$ consist of discrete random variables, we may replace $f_{\mathbf{Y}}(\mathbf{y})$ by $p_{\mathbf{Y}}(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y})$.
  - The right-hand side is called the *Cramér-Rao Lower Bound.*
  - The denominator of the right-hand side

$$I(\theta) = E\left(\left(\frac{\partial}{\partial\theta}\ln f_{\mathbf{Y}}(\mathbf{Y})\right)^2\right)$$

    is called the *Fisher Information.*

- (Cramér-Rao Inequality, iid case) Now, if $Y_1, \ldots, Y_n$ are iid random variables, each with density $f_Y(y)$. Let $\hat{\theta}$ be an estimator of θ. Then the Cramér-Rao inequality becomes

$$\operatorname{Var}(\hat{\theta}) \geq \frac{\left(\frac{\partial}{\partial\theta}E(\hat{\theta})\right)^2}{nE\left(\left(\frac{\partial}{\partial\theta}\ln f_Y(Y)\right)^2\right)}$$

  - If the Y's are discrete, we may replace $f_Y(y)$ by $p_Y(y) = P(Y = y)$.
  - Note now that the Fisher Information becomes

$$I(\theta) = nE\left(\left(\frac{\partial}{\partial\theta}\ln f_Y(Y)\right)^2\right)$$

  - This form has an advantage over the general case in that we deal with a univariate density $f_Y(y)$ instead of the joint density.

- (Cramér-Rao Inequality, iid case, exponential family) If $Y_1, \ldots, Y_n$ are iid random variables, each with pdf $f_Y(y)$ that belongs to an exponential family. Let $\hat{\theta}$ be an estimator of θ. Then the Cramér-Rao inequality is now

$$\operatorname{Var}(\hat{\theta}) \geq \frac{\left(\frac{\partial}{\partial\theta}E(\hat{\theta})\right)^2}{-nE\left(\frac{\partial^2}{\partial\theta^2}\ln f_Y(Y)\right)}$$

- Again, we may replace $f_Y(y)$ by $p_Y(y) = P(Y = y)$ if the Y's are discrete.
- The Fisher Information is now

$$I(\theta) = -nE\left(\frac{\partial^2}{\partial\theta^2}\ln f_Y(Y)\right)$$

- This form has a major advantage in that it is always easier to deal with the second derivative over squaring then taking the expectation.

- (Cramér-Rao Inequality, unbiased case) In all previous cases, if $\hat{\theta}$ is unbiased for θ, then since $E(\hat{\theta}) = \theta$, it follows that

$$\left(\frac{\partial}{\partial\theta}E(\hat{\theta})\right)^2 = \left(\frac{\partial}{\partial\theta}\theta\right)^2 = 1^2 = 1$$

and hence

$$\text{Var}(\hat{\theta}) \geq \frac{\left(\frac{\partial}{\partial\theta}E(\hat{\theta})\right)^2}{I(\theta)} = \frac{1}{I(\theta)} = [I(\theta)]^{-1}$$

- **Efficiency** If $\text{Var}(\hat{\theta})$ attains the Cramér-Rao Lower Bound, then $\hat{\theta}$ is called the *efficient estimator* of θ.

- Some Examples (Exercise 9.8, page 448 of WMS)
  - Suppose that $Y_1, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$ random variables. Each Y has the pdf

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \quad -\infty < y < \infty$$

We have already seen that $Y \sim N(\mu, \sigma^2)$ belongs to an exponential family. To show that $\hat{\mu} = \overline{Y}$ is an efficient estimator of μ, we first note that $E(\hat{\mu}) = E(\overline{Y}) = \mu$ so that $\hat{\mu}$ is an unbiased estimator of μ. Now, $\text{Var}(\hat{\mu}) \geq [I(\mu)]^{-1}$, and the Fisher Information is

$$\begin{aligned}
I(\mu) &= -nE\left(\frac{\partial^2}{\partial\mu^2}\ln f_Y(Y)\right) = -nE\left(\frac{\partial^2}{\partial\mu^2}\ln\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(Y-\mu)^2}{2\sigma^2}\right)\right]\right)\\
&= -nE\left(\frac{\partial^2}{\partial\mu^2}\left[\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \left(\frac{(Y-\mu)^2}{2\sigma^2}\right)\right]\right) = -nE\left(\frac{\partial}{\partial\mu}\left(\frac{2(Y-\mu)}{2\sigma^2}\right)\right)\\
&= -nE\left(\frac{-1}{\sigma^2}\right) = \frac{n}{\sigma^2}
\end{aligned}$$

and hence the Cramér-Rao Lower Bound is

$$[I(\mu)]^{-1} = \left[\frac{n}{\sigma^2}\right]^{-1} = \frac{\sigma^2}{n}$$

Since

$$\text{Var}(\hat{\mu}) = \text{Var}(\overline{Y}) = \frac{\sigma^2}{n}$$

we have that

$$\text{Var}(\hat{\mu}) = [I(\mu)]^{-1}$$

so that $\text{Var}(\hat{\mu})$ attains the Cramér-Rao Lower Bound; therefore, $\hat{\mu} = \overline{Y}$ is the efficient estimator of μ.

– Suppose that $Y_1, \ldots, Y_n$ are iid Poisson(λ) random variables, where each Y has the probability function

$$p_Y(y) = P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, \ldots$$

We have seen that $Y \sim$ Poisson(λ) belongs to an exponential family. We will show that $\hat{\lambda} = \overline{Y}$ is an efficient estimator of λ. Again, $E(\hat{\lambda}) = E(\overline{Y}) = \lambda$ so that $\hat{\lambda}$ is an unbiased estimator of λ. Now, the Fisher Information will be given by

$$\begin{aligned}
I(\lambda) &= -nE\left(\frac{\partial^2}{\partial\lambda^2}\ln p_Y(Y)\right) = -nE\left(\frac{\partial^2}{\partial\lambda^2}\ln\left[\frac{e^{-\lambda}\lambda^Y}{Y!}\right]\right) \\
&= -nE\left(\frac{\partial^2}{\partial\lambda^2}\left(-\lambda + Y\ln\lambda - \ln(Y!)\right)\right) = -nE\left(\frac{\partial}{\partial\lambda}\left(-1 + \frac{Y}{\lambda}\right)\right) \\
&= -nE\left(\frac{-Y}{\lambda^2}\right) = \frac{n}{\lambda^2}E(Y) = \frac{n\lambda}{\lambda^2} = \frac{n}{\lambda}
\end{aligned}$$

and hence the Cramér-Rao Lower Bound is

$$[I(\lambda)]^{-1} = \left[\frac{n}{\lambda}\right]^{-1} = \frac{\lambda}{n}$$

but since

$$\mathrm{Var}(\hat{\lambda}) = \mathrm{Var}(\overline{Y}) = \frac{\lambda}{n}$$

we have that

$$\mathrm{Var}(\hat{\lambda}) = [I(\lambda)]^{-1}$$

so that $\mathrm{Var}(\hat{\lambda})$ attains the Cramér-Rao Lower Bound; therefore, $\hat{\lambda} = \overline{Y}$ is the efficient estimator of λ.