MA 3140: Statistical Inference

Dr. Sameen Naqvi
Department of Mathematics, IIT Hyderabad
Email id: sameen@math.iith.ac.in

Cramer-Frechet-Rao (CFR) Inequality

Let $\Theta \in \mathbb{R}$ be an open interval and suppose that the family $\{f_\theta : \theta \in \Theta\}$ satisfies the following regularity conditions:

(i) It has common support set S, i.e., $S = \{\boldsymbol{x} : f_\theta(\boldsymbol{x}) > 0\}$ does not depend on θ.

(ii) For $\boldsymbol{x}$ and $\theta \in \Theta$, the derivative $\frac{\partial}{\partial\theta} \log f(\boldsymbol{x}, \theta)$ exists and is finite.

(iii) For any statistic h with $E[|h(\boldsymbol{X})|] < \infty$, the operations of integration (summation) and differentiation with respect to θ can be interchanged in $E[h(\boldsymbol{X})]$, i.e.,

$$\frac{\partial}{\partial\theta} \int h(\boldsymbol{x}) f_\theta(\boldsymbol{x}) d\boldsymbol{x} = \int h(\boldsymbol{x}) \frac{\partial}{\partial\theta} f_\theta(\boldsymbol{x}) d\boldsymbol{x}$$

whenever the RHS is finite.

CFR Inequality cont'd

Let $T(\boldsymbol{X})$ be such that $Var[T(\boldsymbol{X})] < \infty$ and define $\psi(\theta) = E[T(\boldsymbol{X})]$.

If

$$I(\theta) = E\left[\frac{\partial}{\partial\theta} \log f_\theta(\boldsymbol{X})\right]^2$$

satisfies $0 < I(\theta) < \infty$, then

$$Var[T(\boldsymbol{X})] \geq \frac{[\psi'(\theta)]^2}{I(\theta)}.$$

Note: The proof is based on Cauchy-Schwartz Inequality, i.e., for any two r.v.s X and Y,

$$Var(X)\ Var(Y) \geq [Cov(X, Y)]^2.$$

Few Important Results

Claim 1:

$$E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] = 0$$

Proof:

$$\begin{aligned}
E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] &= \int \left[\frac{\partial}{\partial\theta}\log f_\theta(x)\right] f_\theta(x)dx \\
&= \int \frac{1}{f_\theta(x)}\left[\frac{\partial}{\partial\theta} f_\theta(x)\right] f_\theta(x)dx \\
&= \int \frac{\partial}{\partial\theta} f_\theta(x)dx \\
&= \frac{\partial}{\partial\theta}\int f_\theta(x)dx = \frac{\partial}{\partial\theta}1 = 0
\end{aligned}$$

Few Important Results

Claim 2: If $X_1, \dots, X_n$ are iid random variables, then

$$E\left[\frac{\partial}{\partial\theta}\log f_\theta(\boldsymbol{X})\right]^2 = \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right]^2$$

Proof:

$$\begin{aligned} E\left[\frac{\partial}{\partial\theta}\log f_\theta(\boldsymbol{X})\right]^2 &= E\left[\frac{\partial}{\partial\theta}\log \prod_{i=1}^{n} f_\theta(X_i)\right]^2 \\ &= E\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right]^2 \end{aligned}$$

Few Important Results

Proof cont'd:

$$= \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right]^2 + \sum_{i\neq j} E\left[\frac{\partial}{\partial\theta}\log f_\theta(X_i)\frac{\partial}{\partial\theta}\log f_\theta(X_j)\right]$$

$$= \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right]^2 + \sum_{i\neq j} E\left[\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right] E\left[\frac{\partial}{\partial\theta}\log f_\theta(X_j)\right]$$

$$= \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right]^2 + 0 \qquad \text{(using Claim 1)}$$

$$= \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right]^2$$

Few Important Results

Claim 3:

$$E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right]$$

Proof:

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\log f_\theta(x) &= \frac{\partial}{\partial\theta}\left[\frac{1}{f_\theta(x)}\frac{\partial}{\partial\theta}f_\theta(x)\right] \\
&= \frac{\partial}{\partial\theta}\left[\frac{\frac{\partial}{\partial\theta}f_\theta(x)}{f_\theta(x)}\right] \\
&= \frac{f_\theta(x)\frac{\partial^2}{\partial\theta^2}f_\theta(x) - \left[\frac{\partial}{\partial\theta}f_\theta(x)\right]^2}{[f_\theta(x)]^2} \\
&= \frac{1}{f_\theta(x)}\frac{\partial^2}{\partial\theta^2}f_\theta(x) - \left[\frac{\partial}{\partial\theta}\log f_\theta(x)\right]^2
\end{aligned}$$

Few Important Results

Proof cont'd: Now, taking expectation on both sides we get

$$
\begin{aligned}
E&\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right] \\
&= E\left[\frac{1}{f_\theta(x)}\frac{\partial^2}{\partial\theta^2}f_\theta(X)\right] - E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 \\
&= \int \frac{1}{f_\theta(x)}\left[\frac{\partial^2}{\partial\theta^2}f_\theta(x)\right]f_\theta(x)dx - E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 \\
&= \frac{\partial^2}{\partial\theta^2}\int f_\theta(x)dx - E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 \\
&= -E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 = -I_1(\theta)
\end{aligned}
$$

Few Important Results

Claim 4:

$$I_1(\theta) = Var\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]$$

Proof:

$$\begin{aligned}
Var\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] &= E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 - \left[E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]\right]^2 \\
&= E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 - 0 \\
&= I_1(\theta)
\end{aligned}$$

Few Important Results

Claim 5:

$$\psi'(\theta) = Cov\left[T, \frac{\partial}{\partial\theta}\log f_\theta(X)\right]$$

Proof:

$$\begin{aligned}
Cov\left[T, \frac{\partial}{\partial\theta}\log f_\theta(X)\right] &= E\left[T\frac{\partial}{\partial\theta}\log f_\theta(X)\right] - E[T]E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right] \\
&= E\left[T\frac{\partial}{\partial\theta}\log f_\theta(X)\right] \qquad \text{(using Claim 1)} \\
&= \int \left[T\frac{\partial}{\partial\theta}\log f_\theta(x)\right] f_\theta(x)dx \\
&= \int \left[T\frac{1}{f_\theta(x)}\frac{\partial}{\partial\theta}f_\theta(x)\right] f_\theta(x)dx \\
&= \int \left[T\frac{\partial}{\partial\theta}f_\theta(x)\right] dx
\end{aligned}$$

Few Important Results

Proof cont'd:

$$
\begin{aligned}
&= \frac{\partial}{\partial\theta} \int T \; f_\theta(x) dx \\
&= \frac{\partial}{\partial\theta} E[T] \\
&= \frac{\partial}{\partial\theta} \psi(\theta) = \psi^{'}(\theta)
\end{aligned}
$$

Thus,

$$
Cov\left[T, \frac{\partial}{\partial\theta} \log f_\theta(X)\right] = \psi^{'}(\theta)
$$

Note 1

The above claims also hold for $\boldsymbol{X}$, i.e.,

Claim 1': $E\left[\frac{\partial}{\partial\theta}\log f_\theta(\boldsymbol{X})\right] = 0$

Claim 3': $E\left[\frac{\partial}{\partial\theta}\log f_\theta(\boldsymbol{X})\right]^2 = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(\boldsymbol{X})\right]$

Claim 4': $I(\theta) = Var\left[\frac{\partial}{\partial\theta}\log f_\theta(\boldsymbol{X})\right]$

Claim 5': $\psi'(\theta) = Cov\left[T(\boldsymbol{X}), \frac{\partial}{\partial\theta}\log f_\theta(\boldsymbol{X})\right]$

Note 2

Combining Claims 4' and 5', and substituting in the CFR Inequality

$$Var(T(\boldsymbol{X})) \geq \frac{[\psi'(\theta)]^2}{I(\theta)},$$

we get

$$Var(T(\boldsymbol{X})) \geq \frac{\left[Cov\left[T(\boldsymbol{X}), \frac{\partial}{\partial\theta}\log f_\theta(\boldsymbol{X})\right]\right]^2}{Var\left[\frac{\partial}{\partial\theta}\log f_\theta(\boldsymbol{X})\right]}.$$

This is true from Cauchy Schwartz Inequality -

$$Var(\boldsymbol{X})Var(\boldsymbol{Y}) \geq Cov^2(\boldsymbol{X}, \boldsymbol{Y})$$

Remark 1

- Let X and Y be independently distribution random variables with densities f_θ and g_θ, respectively. Then

$$I(\theta) = I_1(\theta) + I_2(\theta)$$

where $I_1(\theta)$, $I_2(\theta)$ and $I(\theta)$ are the information about θ contained in X, Y and (X, Y), respectively.

Proof: By definition,

$$I(\theta) = E\Big[\frac{\partial}{\partial\theta}\log f_\theta(X) + \frac{\partial}{\partial\theta}\log g_\theta(Y)\Big]^2,$$

and the result follows from the fact that the cross-product is zero (using Claim 1).

Remark 2

- Let $X_1, \ldots, X_n$ be iid random variables with common pdf (pmf) $f_\theta(x)$. Then
$$I(\theta) = nI_1(\theta).$$

Proof

$$\begin{aligned} I(\theta) = E\left[\frac{\partial}{\partial\theta}\log f_\theta(\boldsymbol{X})\right]^2 &= \sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right]^2 \\ &= nE\left[\frac{\partial}{\partial\theta}\log f_\theta(X_1)\right]^2 \\ &= nI_1(\theta) \end{aligned}$$

In this case, the inequality is:

$$Var(T(\boldsymbol{X})) \geq \frac{[\psi'(\theta)]^2}{nI_1(\theta)}.$$

Remark 3

- If $\psi(\theta) = \theta$, then

$$Var(T(\boldsymbol{X})) \geq \frac{1}{I(\theta)}.$$

- $I_1(\theta)$: Fisher information in X_1.
- $I(\theta) = nI_1(\theta)$: Fisher information in the random sample $X_1, \ldots, X_n$.
- As n gets larger, the lower bound for $Var(T(\boldsymbol{X}))$ gets smaller.

  Thus, as the Fisher information increases, the lower bound decreases, and the "best" estimator will have smaller variance.

Example 1: Revisiting Poisson Distribution

Let $X_1, \ldots, X_n$ be a random sample from Poisson (λ). Find the best UE for λ using the method of CRLB.

Solution: We know that CRLB is $\frac{[\psi^{'}(\lambda)]^2}{I(\lambda)}$.

Here, $\psi(\lambda) = \lambda$ and $\psi^{'}(\lambda) = 1$.

Now,

$$
\begin{aligned}
f_\lambda(x) &= \frac{e^{-\lambda}\lambda^x}{x!} \\
\log f_\lambda(x) &= -\lambda + x \log \lambda - \log x! \\
\frac{\partial}{\partial \lambda} \log f_\lambda(x) &= -1 + \frac{x}{\lambda} = \frac{x - \lambda}{\lambda}
\end{aligned}
$$

Example 1: Revisiting Poisson Distribution cont'd

Now,

$$E\left[\frac{\partial}{\partial\lambda}\log f_\lambda(X)\right]^2 = \frac{1}{\lambda^2}E(X-\lambda)^2 = \frac{1}{\lambda^2}\lambda = \frac{1}{\lambda}$$

So, $I(\lambda) = \frac{n}{\lambda}$ and CRLB for the variance of UE of λ is $\frac{\lambda}{n}$.

We know that $E\overline{X} = \lambda$ and $Var\overline{X} = \frac{\lambda}{n}$. Thus, we conclude that $\overline{X}$ is the best UE of λ.

Example 2: Binomial Distribution

Let $X \sim Bin(n, p)$, n is known and $0 \leq p \leq 1$. Find the best UE for p using the method of CRLB.

Solution: We know that CRLB is $\frac{[\psi^{'}(p)]^2}{I(p)}$.

Here, $\psi(p) = p$ and $\psi^{'}(p) = 1$.

Now,

$$f_p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\log f_p(x) = \log \binom{n}{x} + x \log p + (n-x)\log(1-p)$$

$$\frac{\partial}{\partial p} \log f_p(x) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}$$

Example 2: Binomial Distribution cont'd

Now,

$$E\left[\frac{\partial}{\partial p}\log f_p(X)\right]^2 = E\left[\frac{X-np}{p(1-p)}\right]^2 = \frac{1}{p^2(1-p)^2}E(X-np)^2$$
$$= \frac{1}{p^2(1-p)^2}np(1-p)$$
$$= \frac{n}{p(1-p)}$$

So, CRLB for the variance of UE of p is $\frac{p(1-p)}{n}$.

We know that $E\left(\frac{X}{n}\right) = p$ and $Var\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}$. Thus, we conclude that $\frac{X}{n}$ is the best UE of p.

Example 3: Geometric Distribution

Let $X \sim Geometric(\theta)$.

(i) X: Number of trials until the first success.

The pmf is

$$P(X = x) = \theta(1-\theta)^{x-1}, \; x = 1, 2, \ldots$$

with $E(X) = \frac{1}{\theta}$ and $Var(X) = \frac{1-\theta}{\theta^2}$.

(ii) X: Number of failures preceeding the first success.

The pmf is

$$P(X = x) = \theta(1-\theta)^{x}, \; x = 0, 1, 2, \ldots$$

with $E(X) = \frac{1-\theta}{\theta}$ and $Var(X) = \frac{1-\theta}{\theta^2}$.

Example 3: Geometric Distribution cont'd

Suppose we are interested to estimate θ, i.e., $\psi(\theta) = \theta$.

Now,

$$
\begin{aligned}
f_\theta(x) &= \theta(1-\theta)^x,\ x = 0, 1, 2, \ldots,\ 0 < \theta < 1 \\
\log f_\theta(x) &= \log\theta + x\log(1-\theta) \\
\frac{\partial}{\partial\theta}\log f_\theta(x) &= \frac{1}{\theta} - \frac{x}{1-\theta} \\
E\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 &= E\left[\frac{1}{\theta} - \frac{X}{1-\theta}\right]^2 = \frac{1}{\theta^2(1-\theta)}.
\end{aligned}
$$

So, $I(\theta) = \frac{1}{\theta^2(1-\theta)}$, and CRLB for variance of an UE of θ is $\theta^2(1-\theta)$.

Example 3: Geometric Distribution cont'd

Note that here we are interested in θ which is not the mean. In fact, it is interpreted as $P(X = 0) = \theta$.

Define an estimator for θ as

$$T(X) = \begin{cases} 1, & \text{if } X = 0 \\ 0, & \text{if } X \neq 0 \end{cases}$$

Then, $E[T(X)] = \theta$, $E[T^2(X)] = \theta$ and $Var[T(X)] = \theta - \theta^2 = \theta(1-\theta) > \theta^2(1-\theta)$.

Hence, CRLB is not attained.

Example 3: Geometric Distribution cont'd

Another approach:

$$
\begin{aligned}
& E[T(X)] = \theta, \forall\ 0 < \theta < 1 \\
\implies & \sum_{x=0}^{\infty} T(x)\theta(1-\theta)^x = \theta \\
\implies & T(0)\theta + T(1)\theta(1-\theta) + T(2)\theta(1-\theta)^2 \ldots = \theta \\
\implies & T(0) + T(1)(1-\theta) + T(2)(1-\theta)^2 + \ldots = 1
\end{aligned}
$$

Solving this, we get $T(0) = 1,\ T(1) = T(2) = \ldots = 0$.

Since $T(X)$ is the only UE of θ, it is also the UMVUE.

Thanks for your patience!