UNIVERSITY OF LONDON

Goldsmiths UNIVERSITY OF LONDON

The Cramér-Rao Theorem [1]

Math Notes | Larry Cui

April 3, 2022

The Cramér-Rao Inequality theorem provides a lower bound for the variance of an unbiased estimator of a density function parameter. Many statistics textbooks provide the theorem without giving any proof, which sometimes frustrates students like me. I therefore did some research online and referred mainly to Miller's paper to get the proof.

1 Cauchy-Schwarz Inequality Theorem and its Integral Version

Before we take on the Cramér-Rao theorem, we need to get familiar with the Cauchy-Schwarz theorem first, as it's the key in the Cramér-Rao proof.

Simply put from the geometric perspective, Cauchy-Schwarz theorem says that two sides of a triangle combined is longer than the other one.

Cauchy-Schwarz Inequality Theorem Let $\vec{u}$ and $\vec{v}$ be two vectors in $\mathbb{R}^n$, the theorem states that:

$$|u+v| \leqslant |u| + |v|$$

The Cauchy-Schwarz theorem itself is so straight forward and intuitive that no proof is needed. However, it has many variants, one of those says $|u \cdot v| \leqslant |u||v|$. We give proof below.

We know that if we square the left side of the above equation:

$$|u+v|^2 = (u+v)\cdot(u+v) = u\cdot u + v \cdot v + 2u \cdot v = |u|^2 + |v|^2 + 2u\cdot v$$

Square of the right side:

$$(|u|+|v|)^2 = |u|^2 + |v|^2 + 2|u||v|$$

From the law of cosine, we know that $u \cdot v = |u||v|\cos\theta$. As $-1 \leqslant \cos\theta \leqslant 1$, so complete the proof.

If we write out the variant in coordinates, this says

$$|u_1v_1 + u_2v_2 + \cdots + u_nv_n| \leqslant \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

Square both sides get us a more general form:

[1] This note is based on Adam Merberg and Steven J. Miller's paper: *Course Notes for Math 162: Mathematical Statistics The Cramér-Rao Inequality*.

Variant of Cauchy-Schwarz If we substitute u, v with a, b then

$$\left|\sum_{i=1}^{n} a_i b_i\right|^2 \leqslant \left(\sum_{i=1}^{n}\left|a_i^2\right|\right)\left(\sum_{i=1}^{n}\left|b_i^2\right|\right)$$

Now we give the integral variant of Cauchy-Schwarz theorem and the proof below.

Integral Variant of Cauchy-Schwarz

$$\left(\int_{-\infty}^{\infty} f(x)g(x)\,dx\right)^2 \leqslant \left(\int_{-\infty}^{\infty} f(x)^2\,dx\right)\left(\int_{-\infty}^{\infty} g(x)^2\,dx\right) \tag{1}$$

Proof: It's obvious to see that $\left|\int_{-\infty}^{\infty} f(x)g(x)\,dx\right| \leqslant \int_{-\infty}^{\infty} |f(x)g(x)|\,dx$. If we can prove that $\int_{-\infty}^{\infty} |f(x)g(x)|\,dx$ is less than or equal to the square root of the right side of the equation (1), the integral variant is proved.

Let $u = \sqrt{\int_{-\infty}^{\infty} f(x)^2\,dx}$ and $v = \sqrt{\int_{-\infty}^{\infty} g(x)^2\,dx}$, then

$$\begin{aligned}
\frac{1}{uv}\int_{-\infty}^{\infty} |f(x)g(x)|\,dx &= \int_{-\infty}^{\infty} \frac{|f(x)|}{u}\cdot\frac{|g(x)|}{v}\,dx \\
&\leqslant \int_{-\infty}^{\infty} \frac{\frac{f(x)^2}{u^2}+\frac{g(x)^2}{v^2}}{2}\,dx && \triangleright : 2ab \leqslant a^2+b^2 \\
&= \frac{1}{2}\left(\int_{-\infty}^{\infty}\frac{f(x)^2}{u^2}\,dx + \int_{-\infty}^{\infty}\frac{g(x)^2}{v^2}\,dx\right) \\
&= \frac{1}{2}\left(\frac{u^2}{u^2}+\frac{v^2}{v^2}\right) = 1
\end{aligned}$$

The above result tells us that:

$$\left|\int_{-\infty}^{\infty} f(x)g(x)\,dx\right| \leqslant \int_{-\infty}^{\infty} |f(x)g(x)|\,dx \leqslant \sqrt{\int_{-\infty}^{\infty} f(x)^2\,dx}\sqrt{\int_{-\infty}^{\infty} g(x)^2\,dx}$$

Square both sides finish the proof.

2 Cramér-Rao Theorem

Inequality Description Cramér-Rao Inequality provides with a lower bound on the variance of an unbiased estimator for a parameter. Let $f(y)$ be a probability density function of a population with parameter θ. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables of a sample size n from that population. Let $\hat{\theta}(Y_1, Y_2, \ldots, Y_n)$ be an unbiased estimator for θ. We have:

$$\operatorname{Var}(\hat{\theta}) \geqslant \left\{nE\left[\left(\frac{\partial \ln f(y)}{\partial\theta}\right)^2\right]\right\}^{-1} = \left\{-nE\left[\frac{\partial^2 \ln f(y)}{\partial\theta^2}\right]\right\}^{-1} \tag{2}$$

Proof: In above inequality, E denotes the expected value with respect to the pdf $f(y)$. Let's re-write $\hat{\theta}(Y_1, Y_2, \ldots, Y_n)$ as $\hat{\theta}(\vec{y})$. When we say $\hat{\theta}$ is unbiased, we are actually saying that $E(\hat{\theta}) = \theta$. So we have the following equation:

$$0 = E(\hat{\theta} - \theta) = \int (\hat{\theta}(\vec{y}) - \theta) f(y)\, dy$$

If we differentiate both sides with respect to θ,

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \int (\hat{\theta}(\vec{y}) - \theta) f(y)\, dy \\ &= \int \frac{\partial}{\partial \theta}\Big(\hat{\theta}(\vec{y}) - \theta\Big) \cdot f(y)\, dy + \int (\hat{\theta}(\vec{y}) - \theta) \cdot \frac{\partial}{\partial \theta} f(y)\, dy \\ &= \int -1 \cdot f(y)\, dy + \int (\hat{\theta}(\vec{y}) - \theta) \cdot \frac{\partial}{\partial \theta} f(y)\, dy && \triangleright : \frac{\partial}{\partial \theta}\hat{\theta}(\vec{y}) = 0 \\ 1 &= \int (\hat{\theta}(\vec{y}) - \theta) \cdot \frac{\partial}{\partial \theta} f(y)\, dy \end{aligned}$$

Now we need to re-write the integral. We know that in $f(y)$ and dy, y is dummy variable so can be replaced by any other variable notations. And we also know that as $\int f(y)\, dy = 1$, we can multiply integral with this pdf integral as many as we want. The right side of the above equation can be re-written as:

$$\begin{aligned} \int (\hat{\theta}(\vec{y}) - \theta) \cdot \frac{\partial}{\partial \theta} f(y)\, dy &= \int \cdots \int (\hat{\theta}(\vec{y}) - \theta) \cdot \frac{\partial}{\partial \theta} f(y_1) f(y_2) \ldots f(y_n)\, dy_1\, dy_2 \ldots dy_n \\ &= \int \cdots \int (\hat{\theta}(\vec{y}) - \theta) \cdot \\ &\quad \left(\frac{\partial f(y_1)}{\partial \theta} f(y_2) \ldots f(y_n) + \frac{\partial f(y_2)}{\partial \theta} f(y_1) \ldots f(y_n) + \cdots + \frac{\partial f(y_n)}{\partial \theta} f(y_1) \ldots f(y_{n-1}) \right) \\ &\quad dy_1\, dy_2 \ldots dy_n \end{aligned}$$

By chain differentiation, we have $\dfrac{\partial \ln g}{\partial \theta} = \dfrac{1}{g}\dfrac{\partial g}{\partial \theta}$, so $\dfrac{\partial f(y_1)}{\partial \theta} = \dfrac{\partial \ln f(y_1)}{\partial \theta} f(y_1)$, and so on. This transformation allows us combine terms together to simplify the equation. We also use $\phi(\vec{y})$ to denote $f(y_1) f(y_2) \ldots f(y_n)$, and $d\vec{y}$ to denote $dy_1\, dy_2 \ldots dy_n$.

$$\begin{aligned} 1 &= \int \cdots \int (\hat{\theta}(\vec{y}) - \theta) \left[\phi(\vec{y}) \sum_{i=1}^{n} \frac{\partial \ln f(y_i)}{\partial \theta} \right] d\vec{y} \\ &= \int \cdots \int \Big[(\hat{\theta}(\vec{y}) - \theta) \cdot \phi(\vec{y})^{1/2} \Big] \left[\phi(\vec{y})^{1/2} \sum_{i=1}^{n} \frac{\partial \ln f(y_i)}{\partial \theta} \right] d\vec{y} && \triangleright : \phi(\vec{y}) = (\phi(\vec{y})^{1/2})^2 \\ 1 &= \left(\int \cdots \int \Big[(\hat{\theta}(\vec{y}) - \theta) \cdot \phi(\vec{y})^{1/2} \Big] \left[\phi(\vec{y})^{1/2} \sum_{i=1}^{n} \frac{\partial \ln f(y_i)}{\partial \theta} \right] d\vec{y} \right)^2 && \triangleright : \text{square both sides} \\ &\leqslant \int \cdots \int (\hat{\theta}(\vec{y}) - \theta)^2 \cdot \phi(\vec{y})\, d\vec{y} \cdot \int \cdots \int \left(\sum_{i=1}^{n} \frac{\partial \ln f(y_i)}{\partial \theta} \right)^2 \phi(\vec{y})\, d\vec{y} && \triangleright : \text{Cauchy-Schwarz kicks in} \end{aligned}$$

As we said above, y is but dummy variable, so $\int \cdots \int (\hat{\theta}(\vec{y}) - \theta)^2 \cdot \phi(\vec{y})\, d\vec{y} = \int (\hat{\theta}(\vec{y}) - \theta)^2 \cdot$

$f(y)\,dy = \mathrm{Var}\,(\hat{\theta})$. The inequality becomes:

$$1 \leqslant \mathrm{Var}\,(\hat{\theta}) \cdot \int \cdots \int \left(\sum_{i=1}^{n} \frac{\partial \ln f(y_i)}{\partial \theta} \right)^2 \phi(\vec{y})\, d\vec{y} \tag{3}$$

Now let's turn attention to the second part of the inequality:

$$\begin{aligned} \int \cdots \int \left(\sum_{i=1}^{n} \frac{\partial \ln f(y_i)}{\partial \theta} \right)^2 \phi(\vec{y})\, d\vec{y} &= \int \cdots \int \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \ln f(y_i)}{\partial \theta} \frac{\partial \ln f(y_j)}{\partial \theta} \phi(\vec{y})\, d\vec{y} \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \int \cdots \int \frac{\partial \ln f(y_i)}{\partial \theta} \frac{\partial \ln f(y_j)}{\partial \theta} \phi(\vec{y})\, d\vec{y} \\ &= I_1 + I_2 \end{aligned}$$

where

$$\begin{aligned} I_1 &= \int \cdots \int \sum_{i=1}^{n} \left(\frac{\partial \ln f(y_i)}{\partial \theta} \right)^2 \phi(\vec{y})\, d\vec{y} \\ I_2 &= \sum_{\substack{1 \leqslant i,j \leqslant n \\ i \neq j}}^{n} \int \cdots \int \frac{\partial \ln f(y_i)}{\partial \theta} \frac{\partial \ln f(y_j)}{\partial \theta} \phi(\vec{y})\, d\vec{y} \end{aligned}$$

We tackle these two terms one by one:

$$\begin{aligned} I_1 &= \int \cdots \int \sum_{i=1}^{n} \left(\frac{\partial \ln f(y_i)}{\partial \theta} \right)^2 \phi(\vec{y})\, d\vec{y} \\ &= \sum_{i=1}^{n} \int \left(\frac{\partial \ln f(y_i)}{\partial \theta} \right)^2 f(y_i)\, dy_i \cdot \int \cdots \int \prod_{\substack{k=1 \\ k \neq i}}^{n} f(y_k)\, dy_k \\ &= \sum_{i=1}^{n} E\left[\left(\frac{\partial \ln f(y_i)}{\partial \theta} \right)^2 \right] \cdot 1^{n-1} && \triangleright : \text{dummy variable for } y_k \\ &= nE\left[\left(\frac{\partial \ln f(y_i)}{\partial \theta} \right)^2 \right] && \triangleright : \text{expected value is called Fisher Information} \end{aligned}$$

I_2 has a total of $n^2 - n$ integral terms, each of which takes the same form. We look into one term first:

$$\begin{aligned} \int \cdots \int \frac{\partial \ln f(y_i)}{\partial \theta} \frac{\partial \ln f(y_j)}{\partial \theta} \phi(\vec{y})\, d\vec{y} &= \int \frac{\partial \ln f(y_i)}{\partial \theta} f(y_i)\, dy_i \int \frac{\partial \ln f(y_j)}{\partial \theta} f(y_j)\, dy_j \int \cdots \int \prod_{\substack{k=1 \\ k \neq i,j}}^{n} f(y_k)\, dy_k \\ &= E\left[\frac{\partial \ln f(y_i)}{\partial \theta} \right] \cdot E\left[\frac{\partial \ln f(y_j)}{\partial \theta} \right] \cdot 1^{n-2} \qquad \triangleright : \text{dummy variable for } y_k \end{aligned}$$

But we know that $E\left[\frac{\partial \ln f(y)}{\partial \theta}\right] = 0$ as shown below, so $I_2 = 0$ since each of its terms is 0.

$$\begin{aligned} 1 &= \int f(y)\, dy \qquad \triangleright : \text{differentiate both sides} \\ 0 &= \int \frac{\partial f(y)}{\partial \theta}\, dy = \int \frac{1}{f(y)} \frac{\partial f(y)}{\partial \theta} f(y)\, dy = \int \frac{\partial \ln f(y)}{\partial \theta} f(y)\, dy = E\left[\frac{\partial \ln f(y)}{\partial \theta} \right] \end{aligned}$$

Combine I_1 and I_2 and put them back into equation (3), we complete the proof of the first term on the theorem. But what does the second term $-nE\left[\frac{\partial^2 \ln f(y)}{\partial\theta^2}\right]$ mean exactly? We are going to move around the forms of the second derivative to see what we can get of the expected value from there.

$$
\begin{aligned}
E\left[\frac{\partial^2 \ln f(y)}{\partial\theta^2}\right] &= E\left[\frac{\partial}{\partial\theta}\left(\frac{1}{f(y)}\cdot\frac{\partial f(y)}{\partial\theta}\right)\right] \\
&= E\left[-\frac{1}{f(y)^2}\frac{\partial f(y)}{\partial\theta}\frac{\partial f(y)}{\partial\theta} + \frac{1}{f(y)}\frac{\partial}{\partial\theta}\frac{\partial f(y)}{\partial\theta}\right] \qquad \triangleright : \frac{\partial \ln g}{\partial\theta} = \frac{1}{g}\frac{\partial g}{\partial\theta} \\
&= -E\left[\frac{\partial \ln f(y)}{\partial\theta}\frac{\partial \ln f(y)}{\partial\theta}\right] + E\left[\frac{1}{f(y)}\frac{\partial}{\partial\theta}\frac{\partial f(y)}{\partial\theta}\right]
\end{aligned}
$$

We know that the first term on the right side of the equation is exactly $-I_1/n$, then we see where it'll get us from the second term,

$$
\begin{aligned}
E\left[\frac{1}{f(y)}\frac{\partial}{\partial\theta}\frac{\partial f(y)}{\partial\theta}\right] &= \int \frac{1}{f(y)}\frac{\partial}{\partial\theta}\frac{\partial f(y)}{\partial\theta}\cdot f(y)\,dy \\
&= \frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\int f(y)\,dy \\
&= \frac{\partial}{\partial\theta}\frac{\partial}{\partial\theta}\cdot 1 \\
&= 0
\end{aligned}
$$

Combine together, $-nE\left[\frac{\partial^2 \ln f(y)}{\partial\theta^2}\right]$ is exactly equal to $nE\left[\left(\frac{\partial \ln f(y)}{\partial\theta}\right)^2\right]$. Second term of the theorem proved.