Properties of Estimators II

Bruce M. Boghosian

Department of Mathematics

Tufts University

Outline

Properties studied to date

- **Unbiasedness:** Suppose that $\vec{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ is a random sample from the continuous PDF $f_Y(y; \theta)$, where θ is an unknown parameter. An estimator $\hat{\theta}(\vec{Y})$ is said to be *unbiased* for θ if $E(\hat{\theta}) = \theta$ for all θ. A similar definition holds for discrete random variables.
- **Efficiency:** Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for parameter θ. If $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$, we say that $\hat{\theta}_1$ is *more efficient* than $\hat{\theta}_2$.
- **Relative efficiency:** The *relative efficiency* of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is $\frac{\mathrm{Var}(\hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1)}$.
- **Absolute efficiency:** The Cramér-Rao bound gives a theoretical minimum variance for any unbiased estimator.

The Cramér-Rao bound

Bruce M. Boghosian

- Let $f_Y(y;\theta)$ be a continuous PDF with continuous first and second derivatives
- Suppose that $\{y \mid f_Y(y) \neq 0\}$ does not depend on θ
- We are given n samples $\vec{Y} = \{Y_1, Y_2, \ldots, Y_n\}$
- Let $\hat{\theta}(\vec{Y})$ be an unbiased estimator of θ
- Then

$$\mathrm{Var}(\hat{\theta}) \geq \left\{ n\, E\left[\left(\frac{\partial \ln f_Y(Y;\theta)}{\partial \theta} \right)^2 \right] \right\}^{-1} = \left\{ -n\, E\left[\frac{\partial^2 \ln f_Y(Y;\theta)}{\partial \theta^2} \right] \right\}^{-1}$$

- This gives us a lower bound on the efficiency of any unbiased estimator.
- The *absolute efficiency* of an unbiased estimator $\hat{\theta}$ is the ratio of the Cramér-Rao lower bound to the variance of $\hat{\theta}$

The Bernoulli and binomial distributions

- We may suppose that $p_X(k;p) = p^k(1-p)^{1-k}$ where $k \in \{0,1\}$
- Flip coin n times, and define $X = X_1 + X_2 + \cdots + X_n$ where $X_j \in \{0,1\}$.
- Define the unbiased estimator $\hat{p} = X/n$
- The variance of the result is

$$\mathrm{Var}(\hat{p}) = \mathrm{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\mathrm{Var}(X) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Cramér-Rao bound for binomial dist. – second-derivative method

Bruce M. Boghosian

- To calculate the Cramér-Rao bound, note

$$\ln p_{X_j}(X_j;p) = X_j \ln p + (1-X_j)\ln(1-p)$$

- Taking derivatives,

$$\frac{\partial \ln p_{X_j}(X_j;p)}{\partial p} = \frac{X_j}{p} - \frac{1-X_j}{1-p}$$

$$\frac{\partial^2 \ln p_{X_j}(X_j;p)}{\partial p^2} = -\frac{X_j}{p^2} - \frac{1-X_j}{(1-p)^2}$$

- Taking the expectation value

$$\left\{-n\,E\left[\frac{\partial^2 \ln p_{X_j}(X_j;p)}{\partial p^2}\right]\right\}^{-1} = \left\{n\left(\frac{p}{p^2}+\frac{1-p}{(1-p)^2}\right)\right\}^{-1} = \frac{p(1-p)}{n}$$

- So $\mathrm{Var}(\hat{p})$ achieves the Cramér-Rao bound. It is maximally efficient.

Cramér-Rao bound for binomial dist. – first-derivative method

- Recall that the first derivative was

$$\frac{\partial \ln p_{X_j}(X_j;p)}{\partial p} = \frac{X_j}{p} - \frac{1-X_j}{1-p}$$

- Squaring the first derivative yields

$$\left(\frac{\partial \ln p_{X_j}(X_j;p)}{\partial p}\right)^2 = \frac{X_j^2}{p^2} - 2\frac{X_j}{p}\frac{1-X_j}{1-p} + \frac{(1-X_j)^2}{(1-p)^2} = \frac{(X_j-p)^2}{p^2(1-p)^2}$$

- Taking the expectation value then yields

$$\left\{n\, E\left[\frac{(X_j-p)^2}{p^2(1-p)^2}\right]\right\}^{-1} = \left\{n\left(\frac{p(1-p)}{p^2(1-p)^2}\right)\right\}^{-1} = \frac{p(1-p)}{n}$$

- So $\mathrm{Var}(\hat{p})$ achieves the Cramér-Rao bound. It is maximally efficient.

The Cauchy-Schwarz inequality

- Note that

$$\begin{aligned} 0 \leq & \left(\vec{X}z + \vec{Y}\right) \cdot \left(\vec{X}z + \vec{Y}\right) \\ =& \vec{X} \cdot \vec{X} z^2 + 2\vec{X} \cdot \vec{Y} z + \vec{Y} \cdot \vec{Y} \\ =& \|X\|^2 z^2 + 2\vec{X} \cdot \vec{Y} z + \|\vec{Y}\|^2. \end{aligned}$$

- Because the above quadratic in z has at most one real root, its discriminant must be less than or equal to zero, so

$$4\left(\vec{X} \cdot \vec{Y}\right)^2 - 4\|X\|^2\|\vec{Y}\|^2 \leq 0.$$

- From this follows the *Cauchy-Schwarz inequality*,

$$\boxed{\left|\vec{X} \cdot \vec{Y}\right| \leq \|X\|\|Y\|}$$

Standard deviation as a vector norm

Bruce M. Boghosian

- Given random data $\{X_j\}_{j=1}^n$ and $\{Y_j\}_{j=1}^n$ with means

$$\mu_X = \frac{1}{n}\sum_{j=1}^n X_j \qquad \mu_Y = \frac{1}{n}\sum_{j=1}^n Y_j$$

- Define the *deviations from the means*

$$\vec{X} := \{X_j - \mu_X\}_{j=1}^n \qquad \vec{Y} := \{Y_j - \mu_Y\}_{j=1}^n$$

- The standard deviations are then

$$\sigma_X = \sqrt{\frac{1}{n}\sum_{j=1}^n (X_j - \mu_X)^2} = \sqrt{\frac{1}{n}\vec{X}\cdot\vec{X}} = \frac{1}{\sqrt{n}}\|\vec{X}\|$$

$$\sigma_Y = \sqrt{\frac{1}{n}\sum_{j=1}^n (Y_j - \mu_Y)^2} = \sqrt{\frac{1}{n}\vec{Y}\cdot\vec{Y}} = \frac{1}{\sqrt{n}}\|\vec{Y}\|$$

Covariance and correlation as inner products

- The *covariance* between X and Y is then

$$\mathrm{Cov}(X,Y) = \frac{1}{n}\sum_{j=1}^{n}(X_j-\mu_X)(Y_j-\mu_Y) = \frac{1}{n}\vec{X}\cdot\vec{Y}$$

- The *Pearson correlation coefficient* between X and Y is then

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{\frac{1}{n}\vec{X}\cdot\vec{Y}}{\frac{1}{\sqrt{n}}\|\vec{X}\|\frac{1}{\sqrt{n}}\|\vec{Y}\|} = \frac{\vec{X}\cdot\vec{Y}}{\|\vec{X}\|\;\|\vec{Y}\|}$$

- Note that, by the Cauchy-Schwarz inequality, we have

$$|\rho_{X,Y}| = \frac{\left|\vec{X}\cdot\vec{Y}\right|}{\|\vec{X}\|\;\|\vec{Y}\|} \le 1,$$

so $\rho(X,Y)\in[-1,+1]$, or equivalently $|\mathrm{Cov}(X,Y)|^2 \le \mathrm{Var}(X)\,\mathrm{Var}(Y)$

Proof of the Cramér-Rao bound I

Bruce M. Boghosian

- Suppose we have random variable X with a one-parameter PDF $f(x;\theta)$.
- We have an estimator $\hat{t}(X)$ whose expectation is $\psi(\theta)$
- Estimators are random variables, so let's give this one a name

$$T = \hat{t}(X)$$

- The expectation value of T is a function of the parameter θ,

$$E(T) = \int dx\ f(x;\theta)\hat{t}(x) = \psi(\theta).$$

- We want to show that there is a lower bound on

$$\mathrm{Var}(T) = E(T^2) - (E(T))^2 = E(T^2) - [\psi(\theta)]^2$$

Proof of the Cramér-Rao bound II

Bruce M. Boghosian

- Define another random variable

$$V = \frac{\partial}{\partial\theta} \ln f(X;\theta) = \frac{1}{f(X;\theta)} \frac{\partial}{\partial\theta} f(X;\theta)$$

- Note that this has zero mean

$$\begin{aligned} E(V) &= \int dx \; f(x;\theta) V \\ &= \int dx \; f(x;\theta) \frac{1}{f(X;\theta)} \frac{\partial}{\partial\theta} f(X;\theta) \\ &= \int dx \; \frac{\partial}{\partial\theta} f(X;\theta) \\ &= \frac{\partial}{\partial\theta} \int dx \; f(X;\theta) \\ &= 0. \end{aligned}$$

Proof of the Cramér-Rao bound III

- Now consider the covariance of V and T,

$$
\begin{aligned}
\mathrm{Cov}(V,T) =& E\left[(T-\psi(\theta))\left(\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right)\right] \\
=& E\left[T\left(\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right)\right] \\
=& \int dx\; f(X;\theta)\hat{t}(x)\left(\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right) \\
=& \frac{\partial}{\partial\theta}\int dx\; f(X;\theta)\hat{t}(x) \\
=& \psi'(\theta)
\end{aligned}
$$

Proof of the Cramér-Rao bound IV

- To recap, starting with random variable X,
  - We constructed estimator $T = \hat{t}(X)$ with expectation $E(T) = \psi(\theta)$
  - We defined $V = \frac{\partial}{\partial\theta} \ln f(X;\theta)$ with expectation $E(V) = 0$
  - We found $\text{Cov}(V, T) = \psi'(\theta)$
- Now, by the Cauchy-Schwarz inequality, we have

$$\text{Var}(T)\,\text{Var}(V) \geq |\text{Cov}(V,T)|^2 = |\psi'(\theta)|^2,$$

- and from this it follows that

$$\text{Var}(T) \geq \frac{|\psi'(\theta)|^2}{\text{Var}(V)} = \frac{|\psi'(\theta)|^2}{E\left[n\left(\frac{\partial}{\partial\theta}\ln f(X_j;\theta)\right)^2\right]}$$

Proof of the Cramér-Rao bound V

- We have found that

$$\mathrm{Var}(T) \geq \frac{|\psi'(\theta)|^2}{\mathrm{Var}(V)} = \frac{|\psi'(\theta)|^2}{E\left[n\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)^2\right]}$$

- In the event that the estimator $\hat{t}$ is for θ itself, and is unbiased so that $E(T) = \psi(\theta) = \theta$, the above result becomes

$$\mathrm{Var}(T) \geq \frac{1}{E\left[n\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)^2\right]}$$

- This gives us the first-derivative form of the theorem.

Proof of the Cramér-Rao bound VI

Bruce M. Boghosian

- Finally, note that

$$
\begin{aligned}
E\left[\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)^2\right] &= \int dx\ f(x;\theta)\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)\\
&= \int dx\ f(x;\theta)\frac{1}{f(x;\theta)}\left(\frac{\partial f(X_i;\theta)}{\partial\theta}\right)\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)\\
&= \int dx\ \frac{\partial f(X_i;\theta)}{\partial\theta}\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)\\
&= -\int dx\ f(X_i;\theta)\left(\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)\right)\\
&= E\left[-\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)\right].
\end{aligned}
$$

- This gives us the second-derivative form of the theorem.

Summary

- We have reviewed the Cramér-Rao bound with an example.
- We have learned about both the first- and second-derivative forms of CR bound.
- We learned about and proved the Cauchy-Schwarz inequality.
- We proved both forms of the CR bound.