Bruce M. Boghosian

Properties of Estimators

Unbiasedness

Bruce M. Boghosian

School of Arts and Sciences

Department of Mathematics

Tufts University

Outline

Ambiguity associated with estimators

- For the example of the uniform distribution,

$$f_X(x) = \left\{ \begin{array}{ll} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{array} \right\}$$

  - the MLE estimator was $\hat{a}(\vec{X}) = \min_j X_j$
  - the MM estimator was $\hat{a}(\vec{X}) = M_1 - \sqrt{3}\sqrt{M_2 - M_1^2}$, for sample moments

$$M_1 = \frac{1}{n}\sum_j^n X_j \quad \text{and} \quad M_2 = \frac{1}{n}\sum_j^n X_j^2$$

- Which one is "right"?
- There is no single answer to that question. We must instead identify desirable properties of estimators, and see which estimators have which of those properties.

Estimators are themselves random variables!

- Estimators such as $\hat{a}(\vec{X})$, are functions of r.v.s.
- Hence estimators are themselves random variables.
- They presumably have probability density functions, though it is not always easy to figure out what they are.
- For estimating the sample mean $\hat{\mu}(\vec{X})$, we were able to use the CLT to study its distribution
- For other estimators, that approach may not be possible
- Because they have density functions, however, we know certain things about them:
  - They have means. We may speak of $E(\hat{a}(\vec{X}))$ & $E(\hat{\mu}(\vec{X}))$.
  - They have variances. We may speak of $\sigma^2_{\hat{a}}$ & $\sigma^2_{\hat{\mu}}$

Desirable properties of estimators

- **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 $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$, 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 $\frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}$.

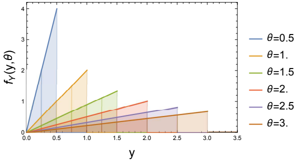
Tufts UNIVERSITY

Unbiasedness Example 1

Bruce M. Boghosian

- Consider the one-parameter probability density function

$$f_Y(y;\theta) = \begin{cases} \frac{2y}{\theta^2} & \text{if } 0 \le y \le \theta \\ 0 & \text{otherwise} \end{cases}$$



- Normalization: $\int_0^\theta dy\ \frac{2y}{\theta^2} = 1$
- Mean: $\mu = \int_0^\theta dy\ \frac{2y}{\theta^2} y = \frac{2}{3}\theta$
- Mean square: $E(Y^2) = \int_0^\theta dy\ \frac{2y}{\theta^2} y^2 = \frac{1}{2}\theta^2$
- Variance: $\text{Var}(Y) = \int_0^\theta dy\ \frac{2y}{\theta^2}(y-\mu)^2 = \frac{1}{18}\theta^2$
- Standard deviation: $\sigma_Y = \sqrt{\text{Var}(Y)} = \frac{1}{3\sqrt{2}}\theta$

Method of moments

- Let $M_1 = \frac{1}{n}\sum_{j=1}^{n} Y_j$ be the average.
- Set the theoretical mean equal to the average: $\frac{2}{3}\theta_e = M_1$
- Hence $\theta_e = \frac{3}{2}M_1$
- MM estimator is then

$$\hat{\theta}_{\text{mm}}(\vec{Y}) = \frac{3}{2n}\sum_{j=1}^{n} Y_j$$

- MM estimator is unbiased

$$E(\hat{\theta}_{\text{mm}}(\vec{Y})) = \frac{3}{2n}\sum_{j=1}^{n} E(Y_j) = \frac{3}{2n}\sum_{j=1}^{n}\frac{2}{3}\theta = \frac{3}{2n}n\left(\frac{2}{3}\theta\right) = \theta.$$

Maximum likelihood estimation

- If $\max_j y_j > \theta$, the likelihood will be zero
- So suppose that $\theta > \max_j y_j$
- Likelihood is $L(\vec{y};\theta) = \prod_{j=1}^n \left(\frac{2y_j}{\theta^2}\right)$
- This clearly increases as θ decreases, so the MLE estimator is

$$\hat{\theta}_{\text{mle}}(\vec{y}) = \max_j y_j$$

- To calculate $E(\hat{\theta}_{\text{mle}})$, we need $f_{Y_{\max}}(y)$, but we can calculate this using what we know about order statistics.

Calculation of CDF

- Note $F_Y(y) = 0$ for $y \leq 0$, and $F_Y(y) = 1$ for $y \geq \theta$
- For $0 < y < \theta$: $F_Y(y) = \int_0^y dz\ f_Y(z) = \int_0^y dz\ \frac{2z}{\theta^2} = \frac{y^2}{\theta^2}$
- Hence the CDF is

$$F_Y(y) = \begin{cases} 0 & \text{if } y \leq 0 \\ \frac{y^2}{\theta^2} & \text{if } 0 \leq y \leq \theta \\ 0 & \text{if } y \geq \theta \end{cases}$$

- From our theorem on order statistics

$$f_{Y_{\max}}(y) = \begin{cases} n\left(\frac{y^2}{\theta^2}\right)^{n-1} \frac{2y}{\theta^2} & \text{if } 0 \leq y \leq \theta \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 2n\frac{y^{2n-1}}{\theta^{2n}} & \text{if } 0 \leq y \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

- Note that this is normalized

$$\int_0^\theta dy\ f_{Y_{\max}}(y) = \frac{2n}{\theta^{2n}} \frac{\theta^{2n}}{2n} = 1.$$

Bias of maximum likelihood estimation

Bruce M. Boghosian

- Expectation value of the MLE estimator is then

$$E(\hat{\theta}_{\text{mle}}) = \int_0^\theta dy\ f_{Y_{\max}}(y)y = \int_0^\theta dy\ 2n\frac{y^{2n-1}}{\theta^{2n}}y = \frac{2n}{\theta^{2n}}\frac{\theta^{2n+1}}{2n+1} = \frac{2n}{2n+1}\theta.$$

- The *MLE estimator is biased* since

$$E(\hat{\theta}_{\text{mle}}(\vec{y})) = \frac{2n}{2n+1}\theta \neq \theta$$

- It is *asymptotically unbiased*, since $\lim_{n\to\infty} E(\hat{\theta}_{\text{mle}}(\vec{y})) = \theta$.

Construction of unbiased version of MLE estimator

- We can construct an unbiased version of the MLE estimator by defining

$$\hat{\theta}_3(\vec{y}) := \frac{2n+1}{2n}\hat{\theta}_{\text{mle}} := \frac{2n+1}{2n}\max_j y_j.$$

- We can see that $\hat{\theta}_3$ is unbiased since

$$E(\hat{\theta}_3) = \frac{2n+1}{2n}\frac{2n}{2n+1}\theta = \theta,$$

- There is no problem with creating unbiased estimators in this way.
- Note that $\hat{\theta}_3$ is not the MLE estimator, but it is arguably preferable to it.

Example 2: Unbiasedness requirement for a linear estimator

- Suppose $X_1, X_2, \ldots, X_n$ have PDF $f_X(x, \theta)$, with theoretical mean $E(X) = \theta$.
- Hence $E(X_j) = \theta$ for $j = 1, 2, \ldots, n$
- Suppose we construct the estimator $\hat{\theta}(\vec{X}) = \sum_{j=1}^{n} a_j X_j$.
- $\therefore E\left(\hat{\theta}(\vec{X})\right) = \sum_{j=1}^{n} a_j E(X_j) = \sum_{j=1}^{n} a_j \mu = \left(\sum_{j=1}^{n} a_j\right) \mu$
- So $\hat{\theta}$ is unbiased iff $\sum_{j=1}^{n} a_j = 1$

Example 3: The variance of the normal distribution

- Recall the normal distribution with theoretical mean μ and variance $v = \sigma^2$

$$f_X(x) = \frac{1}{\sqrt{2\pi v}} \exp\left[-\frac{(x-\mu)^2}{2v}\right]$$

- Recall that the MLE and MM estimators were

$$\hat{\mu}(\vec{X}) = \overline{X} := \frac{1}{n}\sum_{j}^{n} X_j$$

$$\hat{v}(\vec{X}) = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$$

- Clearly, the estimator $\hat{\mu}$ is unbiased from the previous example $\sum_{j=1}^{n}\frac{1}{n} = 1$
- What about the estimator $\hat{v}$?

Two useful lemmas

Bruce M. Boghosian

- Lemma 1

$$\begin{aligned} E(X_j\,\overline{X}) =& E\left(X_j \frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} E\left(X_j X_k\right) \\ =& \frac{1}{n}\sum_{k\neq j}^{n} E\left(X_j\right) E\left(X_k\right) + \frac{1}{n} E\left(X_j^2\right) \\ =& \frac{n-1}{n}\mu^2 + \frac{1}{n}(\mu^2 + v) = \mu^2 + \frac{1}{n} v \end{aligned}$$

- Lemma 2

$$\begin{aligned} E(\overline{X}^2) =& E\left(\frac{1}{n}\sum_{j=1}^{n} X_j\,\overline{X}\right) = \frac{1}{n}\sum_{j=1}^{n} E\left(X_j\,\overline{X}\right) \\ =& \frac{1}{n}\sum_{j=1}^{n}\left(\mu^2 + \frac{1}{n} v\right) = \mu^2 + \frac{1}{n} v \end{aligned}$$

Expectation of the variance estimator

Bruce M. Boghosian

- Using our lemmas, we have

$$
\begin{aligned}
E(\hat{v}) =& E\left(\frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2\right) = E\left(\frac{1}{n}\sum_{j=1}^{n}\left(X_j^2 - 2X_j\overline{X} + \overline{X}^2\right)\right)\\
=& \frac{1}{n}\sum_{j=1}^{n}\left[E\left(X_j^2\right) - 2E\left(X_j\overline{X}\right) + E\left(\overline{X}^2\right)\right]\\
=& \frac{1}{n}\sum_{j=1}^{n}\left[\left(\mu^2 + v\right) - 2\left(\mu^2 + \frac{1}{n}v\right) + \left(\mu^2 + \frac{1}{n}v\right)\right]\\
=& \frac{n-1}{n}v
\end{aligned}
$$

- Since $E(\hat{v}) \neq v$, the *variance estimator is biased*, but *asymptotically unbiased*.

Constructing an unbiased variance estimator

- We have found $E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$
- We construct the *sample variance* estimator,

$$\boxed{\hat{S}^2(\vec{X}) := \frac{n}{n-1}\hat{\sigma}^2(\vec{X}) = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2}$$

- There is an associated *sample standard deviation* estimator

$$\boxed{\hat{S}(\vec{X}) := \sqrt{\frac{n}{n-1}}\,\hat{\sigma}(\vec{X}) = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2}}$$

- This is used in interval estimation, instead of the estimated standard deviation.

Summary

Bruce M. Boghosian

- We have identified some desirable properties of estimators.
- We defined and described the property of *unbiasedness*.
- We presented three examples illustrating this property.