Properties of Estimators (1)

STA 211: The Mathematics of Regression

Yue Jiang

February 21, 2023

The following material was used by Yue Jiang during a live lecture.

Without the accompanying oral comments, the text is incomplete as a record of the presentation.

An important disclaimer

This is not a mathematical statistics class. There are semester-long (and multiple semester-long) courses on probability, and so what we cover in just two lectures scarcely touches on even the basics.

However, familiarity with some of these concepts, such as probability distributions and certain aspects of them (e.g., expectation, variance, etc.) are needed to more fully grasp linear models. As such, we will be presenting a very abridged treatment of some of the fundamentals needed to proceed.

Review: The linear model in matrix form

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

We've constructed the **ordinal least squares** estimate for $\boldsymbol{\beta}$ (in two ways now!), given by $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. But what properties does this estimator have?

Bias

The **bias** of an estimator is the difference between its expected value and the *true* value of the parameter it is estimating.

Suppose we're estimating a parameter θ using some estimator $\hat{\beta}$. Then the bias of this estimator is:

$$Bias(\hat{\theta}) = E_\theta(\hat{\theta} - \theta)$$

An estimator is said to be **unbiased** if $E_\theta(\hat{\theta}) = \theta$ (that is, if the bias is 0).

The sample mean is unbiased for the population mean

Suppose X is a random variable with some well-defined expectation $E(X) < \infty$. Consider an i.i.d. sample $X_1, \cdots, X_n$ from this random variable where each X_i thus has expectation $E(X_i) = E(X)$. Then the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is unbiased for the population mean:

$$
\begin{aligned}
E(\frac{1}{n}\sum_{i=1}^{n} X_i) &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X) \\
&= \frac{1}{n} nE(X) \\
&= E(X)
\end{aligned}
$$

Two more unbiased estimators for the population mean

In that same scenario (i.i.d. sample $X_1, \cdots, X_n$ where each X_i has expectation $E(X_i) = E(X)$), another unbiased estimator for $E(X)$ could simply be the first value X_1.

- What do you think of this estimator?

Now suppose you have another i.i.d. sample $Y_1, \cdots, Y_n$ where $Y \sim N(0, 1000000)$. Consider an estimator for $E(X)$ given by $X_1 + \bar{Y}$.

- Show that this estimator is unbiased.
- What do you think of this estimator?

Two estimators for variance

Suppose X is a random variable with some well-defined expectation $E(X) < \infty$ and well-defined variance $Var(X) \equiv \sigma^2 < \infty$. Consider two popular estimators for σ^2:

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

where $\bar{X}$ is the sample mean. It turns out that $\tilde{\sigma}^2$ is biased for σ^2 whereas s^2 is unbiased for σ^2 (explaining the $n-1$ term in the denominator of the sample variance!).

- Can you support this using a simulation?
- Is $\sqrt{\tilde{\sigma}^2}$ (un)biased for σ? How about $\sqrt{s^2}$? (Try out some more simulations!)

The OLS estimator is unbiased for $\widehat{\boldsymbol{\beta}}$

We saw last time that the OLS estimator was unbiased for $\boldsymbol{\beta}$ (even if we didn't have the terminology!) if the errors were assumed to have mean $\mathbf{0}$ and the true model was linear with $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$:

$$
\begin{aligned}
E(\widehat{\boldsymbol{\beta}}) &= E((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}) \\
&= E((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})) \\
&= E((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}) + E((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\epsilon}) \\
&= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T E(\boldsymbol{\epsilon}) \\
&= \boldsymbol{\beta}
\end{aligned}
$$

(note that this is vector-valued!)

Back to HW 4

Let's consider the random variable given by the difference between some estimator $\hat{\theta}$ and what it's trying to estimate, θ:

$$Var(\hat{\theta} - \theta) = E((\hat{\theta} - \theta)^2) - (E(\hat{\theta} - \theta))^2$$

Notice also that $Var(\hat{\theta} - \theta) = Var(\hat{\theta})$. Thus:

$$\begin{aligned} Var(\hat{\theta}) &= E((\hat{\theta} - \theta)^2) - (E(\hat{\theta} - \theta))^2 \\ E((\hat{\theta} - \theta)^2) &= Var(\hat{\theta}) + (E(\hat{\theta} - \theta))^2 \end{aligned}$$

- What does this mean in plain English?

The mean squared error

$$\underbrace{E((\hat{\theta}-\theta)^2)}_{MSE} = \underbrace{Var(\hat{\theta})}_{Variance} + \underbrace{(E(\hat{\theta}-\theta))^2}_{Bias^2}$$

The mean squared error

$$\underbrace{E((\hat{\theta} - \theta)^2)}_{MSE} = \underbrace{Var(\hat{\theta})}_{Variance} + \underbrace{(E(\hat{\theta} - \theta))^2}_{Bias^2}$$

Consider the three unbiased estimators we had for the population mean on slides 5 and 6: $\bar{X}$, X_1, and $X_1 + \bar{Y}$

- Which one has the lowest/highest MSE? (assume $Var(X) = \sigma^2$)
- What *are* the MSEs for each of these estimators?
- Can you think of a situation in which you might want a *biased* estimate of some parameter θ?

In comparing these three unbiased estimators, the one that has the lowest variance is said to be the most **efficient**.

Aside

As a brief aside, at some point you might have this thought:

If we derived the OLS estimator by minimizing "some" MSE, and we've shown that it's unbiased for $\boldsymbol{\beta}$*, then does this have any implications for its variance? Does it have the smallest possible variance among* **all** *unbiased estimators?*

- What's your reaction to this line of thinking? (More on this two lectures from now...)

Homework 5: Due Feb. 21

1. Suppose an i.i.d. random sample comes from a distribution that is symmetric about its expectation (say, $N(0, 1)$). Use numerical simulations to suggest that the sample mean and sample median are both unbiased estimates for the population mean (in this case 0), but that the sample mean is more efficient. Use simulations to further estimate the relative efficiency of these two estimators (the ratio of their variances).
2. Suppose $X_1, \cdots, X_{16}$ is an i.i.d. random sample of size 16, each of which has density $f(x) = \lambda e^{-\lambda x}$, where $x > 0$ and for some $\lambda > 0$. Consider two estimators of λ: $\hat{\lambda} = \bar{X}$ and $\tilde{\lambda} = 2$. What is the bias of each of these estimators? For what values of λ, if any, is the MSE of $\tilde{\lambda}$ lower than the MSE of $\hat{\lambda}$?
3. Suppose an i.i.d. random sample comes from $N(\mu, \sigma^2)$. Demonstrate that the MSE of $\tilde{\sigma}^2$ is lower than the MSE of s^2.