Properties of the least squares estimates

2019-01-18

Warmup

Let a and b be scalar constants, and X be a scalar random variable.

Fill in the blanks

$$\mathsf{E}\left(aX + b\right) = \underline{\qquad\qquad}$$
$$\mathsf{Var}\left(aX + b\right) = \underline{\qquad\qquad}$$

Goal

Recall that the least squares estimates are:

$$\hat{\beta}_{p\times 1} = \left(X^T X\right)^{-1} X^T y$$

Our goal today is to learn about the statistical properties of these estimates, in particular their expectation and variance.

Random Vectors

$\hat{\beta}$ is a vector-valued random variable, so we first need to cover a little more background.

Let $U_1, \dots, U_n$ be scalar random variables. Then the vector

$$\mathbf{U} = \left(U_1, \dots, U_n\right)^T$$

is a vector valued random variable, a.k.a. a **random vector**.

The expectation of $\mathbf{U}$ is the vector of expectations,

$$\mathsf{E}\left(\mathbf{U}\right) = \left(\mathsf{E}\left(U_1\right), \dots, \mathsf{E}\left(U_n\right)\right)^T$$

And the variance-covariance matrix of $\mathbf{U}$ is

$$\mathsf{Var}\left(\mathbf{U}\right) = \mathsf{Cov}\left(\mathbf{U}\right) = \begin{pmatrix} \mathsf{Var}\left(U_1\right) & \mathsf{Cov}\left(U_1, U_2\right) & \cdots & \mathsf{Cov}\left(U_1, U_n\right) \\ \mathsf{Cov}\left(U_2, U_1\right) & \mathsf{Var}\left(U_2\right) & \cdots & \mathsf{Cov}\left(U_2, U_n\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}\left(U_n, U_1\right) & \mathsf{Cov}\left(U_n, U_2\right) & \cdots & \mathsf{Var}\left(U_n\right) \end{pmatrix}_{n\times n}$$

For example, the errors in multiple linear regression, ϵ_i, $i = 1, \dots, n$ are independent with mean 0 and variance σ^2.

Then,

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \mathsf{E}\left(\epsilon\right) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}, \quad \mathsf{Var}\left(\epsilon\right) = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix} = \sigma^2 I_n$$

Properties of Expectation and Variance for random vectors

Let:

- $\mathbf{U}_{n\times 1}$ be a random vector
- $A_{m\times n}$ be a constant matrix
- $b_{m\times 1}$ be a constant vector

Then:

$$\mathsf{E}\left(A\mathbf{U}+b\right)=A\mathsf{E}\left(\mathbf{U}\right)+b$$
$$\mathsf{Var}\left(A\mathbf{U}+b\right)=A\mathsf{Var}\left(\mathbf{U}\right)A^T$$

These are the vector analogs of the properties you wrote down in the warmup.

Find E (y) **and Var** (y) where $y_{n\times 1}$ satisfies the multiple linear regression equation:

$$y=X\beta+\epsilon$$

and $\mathsf{E}\left(\epsilon\right)=\mathbf{0}$, $\mathsf{Var}\left(\epsilon\right)=\sigma^2 I$

Expectation of the least squares estimates

Assume the regression set up (with the usual dimensions):

$$y = X\beta + \epsilon$$

where X is fixed with rank p, $\mathsf{E}(\epsilon) = 0$, and $\mathsf{Var}(\epsilon) = \sigma^2 I_n$.

Fill in the blanks to show the least squares estimates are unbiased

$$\mathsf{E}\left(\hat{\beta}\right) = \mathsf{E}\left(\left(X^T X\right)^{-1} X^T y\right)$$

$$= \mathsf{E}\left(\left(X^T X\right)^{-1} X^T \left(\qquad \right)\right) \quad \text{plug in the regression equation for } y$$

$$= \mathsf{E}\left(\qquad + \qquad \right) \quad \text{expand}$$

$$= \mathsf{E}\left(\quad + \qquad \right) \quad \text{simplify term on the left } A^{-1}A = I$$

$$= \quad + \qquad \mathsf{E}\left(\ \right) \quad \text{property of expectation}$$

$$= \qquad \text{regression assumptions}$$

$$= \beta$$

Variance-covariance matrix of the least square estimates

Fill in the blanks to find the variance covariance matrix of least squares estimates

$$\text{Var}\left(\hat{\beta}\right) = \text{Var}\left(\left(X^T X\right)^{-1} X^T y\right)$$

$$= \text{Var}\left(\left(X^T X\right)^{-1} X^T X\beta + \left(X^T X\right)^{-1} X^T \epsilon\right) \quad \text{plug in reg. eqn. and expand}$$

$$= 0 + \left(\qquad\qquad\right)\text{Var}\left(\epsilon\right)\left(\qquad\qquad\right)^T \quad \text{property of Var}$$

$$= \left(\qquad\qquad\right)\quad\left(\qquad\qquad\right)^T \quad \text{regression assumption}$$

$$= \quad\left(\qquad\qquad\right)\left(\qquad\qquad\right)^T \quad \text{move scalar to front}$$

$$= \sigma^2\left(\qquad\qquad\right) \qquad\qquad \text{distribute transpose}$$

$$= \sigma^2 \qquad\qquad \text{since } A^{-1}A = I$$

$$= \sigma^2\left(X^T X\right)^{-1}$$

We can pull out the variance of a particular parameter estimate, say $\hat{\beta}_i$, from the diagonal of the matrix:

$$\text{Var}\left(\beta_i\right) = \sigma^2 (X^T X)^{-1}_{i+1 i+1}$$

where A_{ij} indicates the element in the i'th row and j'th column of the matrix A.

Why $i + 1$?

The off diagonal terms tell us about the covariance between parameter estimates.

Estimating σ

To make use of the variance-covariance results we need to be able to estimate σ^2.

An unbiased estimate is:

$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2 = \frac{||e||^2}{n-p}$$

The denominator $n - p$ is known as the model **degrees of freedom**.

Standard errors of particular parameters

The standard error of a particular parameter is then the squareroot of the variance replacing σ^2 with its estimate:

$$\text{SE}\left(\beta_i\right) = \hat{\sigma}^2 \sqrt{(X^T X)^{-1}_{i+1 i+1}}$$

Gauss Markov Theorem

You might wonder if we can find estimates with better properties. The Guass-Markov theorem says the least squares estimates are BLUE (**B**est **L**inear **U**nbiased **E**stimator).

> Of all linear, unbiased estimates, the least squares estimates have the smallest variance.

Of course if you are willing to have a non-linear estimate and/or, a biased estimate you might be able to find an estimate with smaller variance.

Proof, see Section 2.8 in Faraday

Summary

For the linear regression model

$$Y = X\beta + \epsilon$$

where $\mathsf{E}(\epsilon) = 0_{n\times 1}$, $\mathsf{Var}(\epsilon) = \sigma^2 I_n$, and the matrix $X_{n\times p}$ is fixed with rank p.

The least squares estimates are

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Furthermore, the least squares estimates are BLUE, and

$$\mathsf{E}\left(\hat{\beta}\right) = \beta, \qquad \mathsf{Var}\left(\hat{\beta}\right) = \sigma^2 (X^T X)^{-1}$$

$$\mathsf{E}\left(\hat{\sigma}^2\right) = \mathsf{E}\left(\frac{1}{n-p}\sum_{i=1}^{n} e_i^2\right) = \sigma^2$$

We have not used any Normality assumptions to show these properties.