Lecture 37: Unbiased Estimator for the QoIs

We will continue with the linear model

$$Y = X q_0 + \varepsilon, \qquad X \in \mathbb{R}^{n \times p}, \quad n \geq p.$$

Where $\mathbb{E}(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \sigma_0^2 I$. $\qquad q_0 =$ Quantities of Interest ("QoIs")

LAST TIME we saw that the value $\hat{q}$ that minimizes the norm of the mismatch, $\|Y - Xq\|$, over all $q \in \mathbb{R}^p$ is given by $\hat{q} = (X^T X)^{-1} X^T Y$.

IN THIS LECTURE we shall show that $\hat{q}$ is an unbiased estimator for q_0: this means that

$$\mathbb{E}(\hat{q}) = q_0.$$

<u>Preliminary work</u>: Expected value and variance.

Let z be a random variable taking values in $\Omega \subseteq \mathbb{R}$ according to the probability distribution p. (Note that $\int_\Omega p(x)\,dx = 1.$)

The expected value of z, $\mathbb{E}(z)$, then is given by

$$\mathbb{E}(z) = \int_\Omega x\, p(x)\, dx.$$

From this definition we see that $\mathbb{E}$ is <u>linear</u>. If z_1 and z_2 are two random variables with the same distribution p and $\alpha \in \mathbb{R}$ is a constant, then $\mathbb{E}(\alpha z_1 + z_2) = \alpha\,\mathbb{E}(z_1) + \mathbb{E}(z_2)$.

Also note that $\mathbb{E}(\alpha) = \alpha$: the expected value of a constant is that constant.

Now recall the definitions of variance and covariance from the last lecture: If z is a scalar random variable,

$$\mathrm{Var}(z) = \mathbb{E}\big((z - \mathbb{E}(z))(z - \mathbb{E}(z))\big)$$
$$= \mathbb{E}(z^2) - \mathbb{E}(z)^2$$

If z_1 and z_2 are scalar random variables,

$$\mathrm{Cov}(z_1, z_2) = \mathbb{E}\big((z_1 - \mathbb{E}(z_1))(z_2 - \mathbb{E}(z_2))\big)$$
$$= \mathbb{E}(z_1 z_2) - \mathbb{E}(z_1)\,\mathbb{E}(z_2).$$

We will need the Variance of a vector of random variables, $z \in \mathbb{R}^n$, defined as

$$\mathrm{Var}(z) = \begin{bmatrix} \mathrm{Var}(z_1) & \mathrm{Cov}(z_1, z_2) & \cdots & \mathrm{Cov}(z_1, z_n) \\ \mathrm{Cov}(z_2, z_1) & \mathrm{Var}(z_2) & \cdots & \mathrm{Cov}(z_2, z_n) \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}(z_n, z_1) & \cdots & & \mathrm{Var}(z_n) \end{bmatrix}$$

NOTE THAT THIS IS WHAT WE MEAN BY $\mathrm{Var}(\varepsilon) = \sigma_0^2 I$ AT THE START OF THIS LECTURE.

Note that $\mathrm{Var}(z) \in \mathbb{R}^{n \times n}$ is a symmetric matrix.

Since $\mathrm{Cov}(z_j, z_k) = \mathbb{E}(z_j z_k) - \mathbb{E}(z_j)\,\mathbb{E}(z_k)$, we can compactly write

$$\mathrm{Var}(z) = \mathbb{E}\big((z - \mathbb{E}(z))(z - \mathbb{E}(z))^T\big),$$

where the expected value of a vector is taken componentwise:

$$\mathbb{E}(z) = \mathbb{E}\left(\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}\right) = \begin{bmatrix} \mathbb{E}(z_1) \\ \vdots \\ \mathbb{E}(z_n) \end{bmatrix}.$$

Now
$$\begin{aligned} \mathrm{Var}(z) &= \mathbb{E}\left((z-\mathbb{E}(z))(z-\mathbb{E}(z))^T\right) \\ &= \mathbb{E}\left(zz^T - \mathbb{E}(z)z^T - z\,\mathbb{E}(z)^T + \mathbb{E}(z)\mathbb{E}(z)^T\right) \\ \left(\substack{\text{linearity of} \\ \text{expected} \\ \text{value}}\right) &= \mathbb{E}(zz^T) - \mathbb{E}(z)\mathbb{E}(z)^T - \mathbb{E}(z)\mathbb{E}(z)^T + \mathbb{E}(z)\mathbb{E}(z)^T \\ &= \mathbb{E}(zz^T) - \mathbb{E}(z)\mathbb{E}(z)^T \end{aligned}$$

By this formulation, the (j,k) entry of $\mathrm{Var}(z)$ is

$$\begin{aligned} (\mathrm{Var}(z))_{j,k} &= \mathbb{E}(zz^T)_{j,k} - (\mathbb{E}(z)\mathbb{E}(z)^T)_{j,k} \\ &= \mathbb{E}(z_j z_k) - \mathbb{E}(z_j)\mathbb{E}(z_k) \\ &= \mathrm{Cov}(z_j, z_k), \end{aligned}$$

consistent with the original definition of $\mathrm{Var}(z)$.

Unbiased estimator

Our first task is to show that the least squares solution $\hat{q} = (X^T X)^{-1} X^T Y$ is an unbiased estimator. Thus we compute

$$\mathbb{E}(\hat{q}) = \mathbb{E}((X^TX)^{-1}X^TY)$$

$$= \mathbb{E}((X^TX)^{-1}X^T(Xq_0+\varepsilon)) \qquad \text{since } Y = Xq_0+\varepsilon$$

$$= \mathbb{E}(\underbrace{(X^TX)^{-1}X^TX}_{I}q_0 + (X^TX)^{-1}X^T\varepsilon)$$

$$= \mathbb{E}(q_0) + \mathbb{E}((X^TX)^{-1}X^T\varepsilon)$$

$\updownarrow$ q_0 is constant $\qquad \updownarrow$ linearity of expected value

$$= q_0 + (X^TX)^{-1}X^T\underbrace{\mathbb{E}(\varepsilon)}_{=0 \text{ by assumption}}$$

$$= q_0.$$

Hence $\mathbb{E}(\hat{q}) = q_0$, so $\hat{q}$ gives an unbiased estimator for q_0.