Lecture 37: Unbiased Estimator for the QoIs

We will continue with the linear model

\[ Y = Xq_0 + \varepsilon, \quad X \in \mathbb{R}^{n \times p}, \quad n \gg p. \]

where \( \mathbb{E}(\varepsilon) = 0, \quad \text{Var}(\varepsilon) = \sigma_0^2 I. \)

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

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

Preliminary work: Expected value and variance.

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

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

\[ \mathbb{E}(Z) = \int_{\mathbb{R}} x \, p(x) \, dx. \]

From this definition we see that \( \mathbb{E} \) is linear.

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 \( \mathbb{E}(d) = d \): 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,

\[
\text{Var}(z) = \mathbb{E}((z - \mathbb{E}(z))(z - \mathbb{E}(z)))
= \mathbb{E}(z^2) - \mathbb{E}(z)^2
\]

If \( z_1 \) and \( z_2 \) are scalar random variables,

\[
\text{Cov}(z_1, z_2) = \mathbb{E}((z_1 - \mathbb{E}(z_1))(z_2 - \mathbb{E}(z_2)))
= \mathbb{E}(z_1 z_2) - \mathbb{E}(z_1)\mathbb{E}(z_2).
\]

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

\[
\text{Var}(\mathbf{z}) = \begin{bmatrix}
\text{Var}(z_1) & \text{Cov}(z_1, z_2) & \cdots & \text{Cov}(z_1, z_n) \\
\text{Cov}(z_2, z_1) & \text{Var}(z_2) & \cdots & \text{Cov}(z_2, z_n) \\
\vdots & \vdots & \ddots & \vdots \\
\text{Cov}(z_n, z_1) & \cdots & \text{Cov}(z_n, z_{n-1}) & \text{Var}(z_n)
\end{bmatrix}
\]

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

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

\[
\text{Var}(\mathbf{z}) = \mathbb{E}((\mathbf{z} - \mathbb{E}(\mathbf{z}))(\mathbf{z} - \mathbb{E}(\mathbf{z}))^T),
\]
Where the expected value of a vector is taken component-wise:

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

Now \( \text{Var}(z) = \mathbb{E} \left( (z - \mathbb{E}(z))(z - \mathbb{E}(z))^\top \right) \)

\[
= \mathbb{E}(zz^\top - \mathbb{E}(z)z^\top - z \mathbb{E}(z)^\top + \mathbb{E}(z)\mathbb{E}(z)^\top)
= \mathbb{E}(zz^\top) - \mathbb{E}(z)\mathbb{E}(z)^\top.
\]

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

\[
\left( \text{Var}(z) \right)_{j,k} = \mathbb{E}(z_j z_k) - \left( \mathbb{E}(z) \mathbb{E}(z)^\top \right)_{j,k}
= \mathbb{E}(z_j z_k) - \mathbb{E}(z_j) \mathbb{E}(z_k).
\]

\[
= \text{Cov}(z_j, z_k)
\]

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

**Unbiased estimator**

Our first task is to show that the least squares solution \( \hat{\beta} = (X^\top X)^{-1}X^\top Y \) is an unbiased estimator. Thus we compute
\[ E(\hat{\beta}) = E\left( (X^T X)^{-1} X^T y \right) \]
\[ = E\left( (X^T X)^{-1} X^T (Xq_0 + \epsilon) \right) \quad \text{since } y = Xq_0 + \epsilon \]
\[ = E\left( \frac{(X^T X)^{-1} X^T X q_0 + (X^T X)^{-1} X^T \epsilon}{I} \right) \]
\[ = E(q_0) + E((X^T X)^{-1} X^T \epsilon) \quad \uparrow \text{ } q_0 \text{ is constant} \quad \uparrow \text{ linearity of expected value} \]
\[ = q_0 + (X^T X)^{-1} X^T \underbrace{E(\epsilon)}_{=0 \text{ by assumption}} \]
\[ = q_0. \]

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