Distribution of regression coefficients (cont.)

Back to our problem: \( Y = X\beta + \epsilon \) where \( \epsilon_i \) are iid \( N(0, \sigma^2) \). We have
\[
Y \sim N(X\beta, \sigma^2 I).
\]

Therefore,
\[
\hat{\beta} = (X^T X)^{-1} X^T Y \sim N(\beta, \sigma^2 (X^T X)^{-1}).
\]

In particular,
\[
E(\hat{\beta}) = \beta.
\]

Thus, \( \hat{\beta} \) is unbiased.
Statistical consistency of least squares

- We saw that $E(\hat{\beta}) = \beta$.
- What happens as the sample size $n$ goes to infinity? We expect $\hat{\beta} = \hat{\beta}(n) \rightarrow \beta$.

A sequence of estimators $\{\theta_n\}_{n=1}^{\infty}$ of a parameter $\theta$ is said to be consistent if $\theta_n \rightarrow \theta$ in probability ($\theta_n \overset{P}{\rightarrow} \theta$) as $n \rightarrow \infty$.

(Recall: $\theta_n \overset{P}{\rightarrow} \theta$ if for every $\epsilon > 0$,
$$\lim_{n \rightarrow \infty} P(|\theta_n - \theta| \geq \epsilon) = 0.$$)

In order to prove that $\hat{\beta}_n$ (estimator with $n$ samples) is consistent, we will make some assumptions on the data generating model.

(Without any assumptions, nothing prevents the observations to be all the same for example...)

Observations: $y = (y_i) \in \mathbb{R}^n$, $X = (x_{ij}) \in \mathbb{R}^{n \times p}$. Let $x_i = (x_{i1}, \ldots, x_{in}) \in \mathbb{R}^p$ ($i = 1, \ldots, n$).

We will assume:
- $(x_i)_{i=1}^{n}$ are iid random vectors.
- $y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i$ where $\epsilon_i$ are iid $N(0, \sigma^2)$.
- The error $\epsilon_i$ is independent of $x_i$.
- $E x_{ij}^2 < \infty$ (finite second moment).
- $Q = E(x_i x_i^T) \in \mathbb{R}^{p \times p}$ is invertible.

Under these assumptions, we have the following theorem.

Theorem: Let $\hat{\beta}_n = (X^T X)^{-1} X^T y$. Then, under the above assumptions, we have $\hat{\beta}_n \overset{P}{\rightarrow} \beta$.

Proof of the theorem

We have
$$\hat{\beta} = (X^T X)^{-1} X^T y = \left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i^T\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} x_i y_i\right).$$

Using Cauchy–Schwarz,
$$E(|x_{ij} x_{ik}|) \leq (E(x_{ij}^2) E(x_{ik}^2))^{1/2} < \infty.$$

In a similar way, we prove that $E(|x_{ij} y_i|) < \infty$.

By the weak law of large numbers, we obtain
$$\frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \overset{P}{\rightarrow} E(x_i x_i^T) = Q,$$
$$\frac{1}{n} \sum_{i=1}^{n} x_i y_i \overset{P}{\rightarrow} E(x_i y_i).$$
Proof of the theorem (cont.)

Using the continuous mapping theorem, we obtain

\[ \hat{\beta}_n \xrightarrow{p} E(x_i x_i^T)^{-1} E(x_i y_i). \]

(Define \( g : \mathbb{R}^{p \times p} \times \mathbb{R}^p \to \mathbb{R}^p \) by \( g(A, b) = A^{-1} b \).)

Recall: \( y_i = x_i^T \beta + \epsilon_i \). So

\[ x_i y_i = x_i x_i^T \beta + x_i \epsilon_i. \]

Taking expectations,

\[ E(x_i y_i) = E(x_i x_i^T) \beta + E(x_i \epsilon_i). \]

Note that \( E(x_i \epsilon_i) = 0 \) since \( x_i \) and \( \epsilon_i \) are independent by assumption.

We conclude that

\[ \beta = E(x_i x_i^T)^{-1} E(x_i y_i) \]

and so \( \hat{\beta}_n \xrightarrow{p} \beta \). \[ \square \]