# Estimation of the noise variance in OLS

Andersen Ang

First created: 2014. Last update : 2017-Feb

## 1 The problem

For the regression problem

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}$$

$\varepsilon \sim \mathcal{N}\left(0, \sigma^2 I\right)$, the noise is zero-mean, uncorrelated and having same variance $\sigma^2 : \mathbb{E}\left[\varepsilon_i \varepsilon_i\right] = \sigma^2 \delta_{ij}$. This is the Gaussian-Markov assumption.

We want to estimate $\beta$ for the case $n = p$ by the Ordinary Least Square Estimator $\hat{\beta}^{OLS} = \arg\min \|y - X\beta\|_2^2$ it can be shown that $\hat{\beta}^{OLS} = \left(X^T X\right)^{-1} X^T y$ and it has the distribution $\hat{\beta}^{OLS} \sim \mathcal{N}\left(\beta_0, \sigma^2 \left(X^T X\right)^{-1}\right)$ for $\varepsilon \sim \mathcal{N}\left(0, \sigma^2 I\right)$ and $\beta_0$ is the true value.

We know that the noise is zero mean , but we don't know the value of $\sigma^2$, so we want to estimate it.

## 2 Estimation of $\sigma^2$

The error $e$ (residue) can be expressed as

$$e = y - X\hat{\beta}$$

$$e = y - X\left(X^T X\right)^{-1} X^T y$$

$$e = \left(I - X\left(X^T X\right)^{-1} X^T\right) y$$

Let $G = \left(I - X\left(X^T X\right)^{-1} X^T\right)$, note that

$$G^T = \left(I - X\left(X^T X\right)^{-1} X^T\right)^T$$

$$G^T = I^T - \left(X\left(X^T X\right)^{-1} X^T\right)^T$$

By $(ABC)^T = (A[BC])^T = [BC]^T A^T = C^T B^T A^T$

$$G^T = I^T - \left(X\left(X^T X\right)^{-1} X^T\right) = G$$

And

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

$$G^2 = I - X\left(X^T X\right)^{-1} X^T - X\left(X^T X\right)^{-1} X^T + X\left(X^T X\right)^{-1} \underbrace{X^T X \left(X^T X\right)^{-1}}_{I} X^T$$

$$G^2 = I - X\left(X^TX\right)^{-1}X^T \underbrace{-X\left(X^TX\right)^{-1}X^T + X\left(X^TX\right)^{-1}X^T}_{0} = I - X\left(X^TX\right)^{-1}X^T = G$$

And

$$GX = \left(I - X\left(X^TX\right)^{-1}X^T\right)X = X - X = 0$$

So

$$G \text{ is symmetric, idempotent and orthogonal to } X$$

Then

$$e = Gy$$
$$e = G\left(X\beta + \varepsilon\right)$$

$$e = \underbrace{GX}_{0}\beta + G\varepsilon$$

$$e = \left(I - X\left(X^TX\right)^{-1}X^T\right)\varepsilon$$

Therefore

$$\mathbb{E}\left[e\right] = \mathbb{E}\left[\left(I - X\left(X^TX\right)^{-1}X^T\right)\varepsilon\right] = \left(I - X\left(X^TX\right)^{-1}X^T\right)\underbrace{\mathbb{E}\left[\varepsilon\right]}_{0} = 0$$

$$Var\left(e\right) = \mathbb{E}\left[ee^T\right] = \mathbb{E}\left[\left(G\varepsilon\right)\left(G\varepsilon\right)^T\right]$$

Since $G$ is symmetric $G = G^T$

$$Var\left(e\right) = \mathbb{E}\left[G\varepsilon\varepsilon^TG\right] = G\mathbb{E}\left[\varepsilon\varepsilon^T\right]G$$

Since $\mathbb{E}\left[\varepsilon\varepsilon^T\right] = \sigma^2 I$ and $G^2 = G$

$$Var\left(e\right) = \sigma^2 G$$

Thus the theoretical equation to estimate $\sigma^2$ is thus

$$\mathbb{E}\left[ee^T\right] = \sigma^2\left(I - X\left(X^TX\right)^{-1}X^T\right)$$

# 3 Practical way to estimate $\sigma^2$ in OLS

The above equation cannot be used directly in application since it is not "very useful".

Consider $\mathbb{E}\left[ee^T\right]$, it is a square matrix, one way to compute this value is to use the trace

$$\mathbb{E}\left[ee^T\right] = \mathbb{E}\left[\mathrm{Tr}\left(ee^T\right)\right]$$

Since trace and expectation operator can be interchanged

$$\mathbb{E}\left[ee^T\right] = \mathrm{Tr}\left(\mathbb{E}\left[ee^T\right]\right)$$

Since $\mathbb{E}\left[ee^T\right] = \sigma^2 G = \sigma^2\left(I - X\left(X^TX\right)^{-1}X^T\right)$, thus

$$\mathbb{E}\left[ee^T\right] = \mathrm{Tr}\left(\mathbb{E}\left[\sigma^2 G\right]\right)$$

$$\mathbb{E}\left[ee^T\right] = \sigma^2\mathrm{Tr}\left(G\right)$$

$$\mathbb{E}\left[ee^T\right] = \sigma^2 \mathrm{Tr}\left(I - X\left(X^T X\right)^{-1} X^T\right)$$

Since $\mathrm{Tr}(A+B) = \mathrm{Tr}\left(A\right) + \mathrm{Tr}\left(B\right)$

$$\mathbb{E}\left[ee^T\right] = \sigma^2 \left(\mathrm{Tr}I - \mathrm{Tr}\left(X\left(X^T X\right)^{-1} X^T\right)\right)$$

Since $\mathrm{Tr}\left(AB^T\right) = \mathrm{Tr}\left(B^T A\right)$

$$\mathbb{E}\left[ee^T\right] = \sigma^2 \left(\mathrm{Tr}I - \mathrm{Tr}\underbrace{\left(X^T X\left(X^T X\right)^{-1}\right)}_{I}\right)$$

Notice that the first identity matrix is $n \times n$ and the second one is $p \times p$

$$\mathbb{E}\left[ee^T\right] = \sigma^2 \left(\mathrm{Tr}I_{n\times n} - \mathrm{Tr}I_{p\times p}\right)$$

$$\mathbb{E}\left[ee^T\right] = \sigma^2 \left(n - p\right)$$

Therefore

$$\sigma^2 = \frac{\mathbb{E}\left[ee^T\right]}{n-p}$$