9.2 Statistical Aspects of Least Squares

Exercise 2.4 (2.2): Convergence rates for consistent estimators.

For most consistent estimators of the parameters of stationary processes, the estimation error $\hat{\theta} - \theta_0$ tends to zero as $1/n$ when $n \to \infty$. For nonstationary processes, faster convergence rates may be expected. To see this, derive the variance of the least squares estimate in the model

$$Y_t = \alpha t + D_t, \ t = 1, \ldots, N$$

with D_t white noise, zero mean and variance λ^2.

Solution:

The LS estimate $\hat{\alpha}$ of α is given as the solution to the corresponding normal equations

$$\hat{\alpha} = \frac{\sum_{t=1}^{n} Y_t t}{\sum_{t=1}^{n} t^2}.$$

Thus

$$\hat{\alpha} - \alpha = \frac{\sum_{t=1}^{n} D_t t}{\sum_{t=1}^{n} t^2}$$

and

$$\mathbb{E}[\hat{\alpha} - \alpha]^2 = \mathbb{E}\frac{\sum_{t=1}^{n}\sum_{s=1}^{n} D_t t D_s s}{\left(\sum_{t=1}^{n} t^2\right)^2} = \frac{\lambda^2}{\sum_{t=1}^{n} t^2},$$

since $\mathbb{E}[D_t D_s] = \delta_{t-s}\lambda^2$. As $\sum_{t=1}^{n} t^2 = \frac{n(n+1)(2n+1)}{6}$, it follows that the variance of $\hat{\alpha}$ goes to zero as $n \to \infty$.

Exercise 2.5 (2.3)

Illustration of unbiasedness and consistency properties. Let $\{X_i\}_i$ be a sequence of i.i.d. Gaussian random variables with mean μ and variance σ. Both are unknown. Let $\{x_i\}_{i=1}^n$ be a realization of this process of length n. Consider the following estimate of μ:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{N} x_i$$

and the following two estimates of σ:

$$\begin{cases} \hat{\sigma}_1 = \frac{1}{n}\sum_{i=1}^n (x_i - \hat{\mu})^2 \\ \hat{\sigma}_2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \hat{\mu})^2 \end{cases}$$

determine the mean and the variance of the estimates $\hat{\mu}, \hat{\sigma}_1$ and $\hat{\sigma}_2$. Discuss their bias and consistency properties. Compare $\hat{\sigma}_1$ and $\hat{\sigma}_2$ in terms of their Mean Square Error (mse).

Solution:

The expected $\hat{\mu}$ is given as

$$\mathbb{E}[\hat{\mu}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \mu.$$

The variance of $\hat{\mu}$ is computed as

$$\mathbb{E}[\hat{\mu} - \mu]^2 = \mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^n\sum_{i=1}^n (X_i - \mu)(X_j - \mu)\right] = \mathbb{E}\left[\frac{1}{n^2}\sum_{i=1}^n\sum_{i=1}^n \delta_{i-j}\sigma^2\right] = \frac{\sigma^2}{n}.$$

Next note that

$$\mathbb{E}[X_i - \hat{\mu}]^2 = \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^n (X_i - X_j)\right] = \frac{n-1}{n}\sigma^2$$

The bias of $\hat{\sigma}_1$ is derived as

$$\mathbb{E}[\hat{\sigma}_1] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu})^2\right] = \frac{n-1}{n}\sigma^2$$

while the bias of $\hat{\sigma}_2$ is 0 as seen by

$$\mathbb{E}[\hat{\sigma}_2] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^n (X_i - \hat{\mu})^2\right] = \sigma^2$$

Hence $\hat{\mu}$ and $\hat{\sigma}_2$ are unbiased, while $\hat{\sigma}_1$ is 'only' asymptotically unbiased.

The variance of $\hat{\sigma}_1$ is derived as follows. First note that

$$\mathbb{E}[\hat{\sigma}_1 - \sigma^2]^2 = \sigma^4 - 2\sigma^2\mathbb{E}[\hat{\sigma}_1] + \mathbb{E}[\hat{\sigma}_1^2] = \left(\frac{2n-1}{n^2}\right)\sigma^2$$

The variance of $\hat{\sigma}_2$ is given as

$$\mathbb{E}[\hat{\sigma}_1 - \sigma^2]^2 = \sigma^4 - 2\sigma^2\mathbb{E}[\hat{\sigma}_1] + \mathbb{E}[\hat{\sigma}_1^2] = \left(\frac{2}{n-1}\right)\sigma^2$$

Hence both $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are consistent estimates, but for $n > 1$, $\hat{\sigma}_1$ gives a lower variance than $\hat{\sigma}_2$.