Lecture 38: Asymptotic properties of LSE's

We consider first the consistency of the LSE $l^\tau \hat{\beta}$ with $l \in \mathcal{R}(Z)$ for every n.

Theorem 3.11. Consider model

$$X = Z\beta + \varepsilon \tag{1}$$

under assumption A3 ($E(\varepsilon) = 0$ and $\text{Var}(\varepsilon)$ is an unknown matrix).
Suppose that $\sup_n \lambda_+[\text{Var}(\varepsilon)] < \infty$, where $\lambda_+[A]$ is the largest eigenvalue of the matrix A, and that $\lim_{n\to\infty} \lambda_+[(Z^\tau Z)^-] = 0$. Then $l^\tau \hat{\beta}$ is consistent in mse for any $l \in \mathcal{R}(Z)$.
Proof. The result follows from the fact that $l^\tau \hat{\beta}$ is unbiased and

$$\begin{aligned}\text{Var}(l^\tau \hat{\beta}) &= l^\tau (Z^\tau Z)^- Z^\tau \text{Var}(\varepsilon) Z (Z^\tau Z)^- l \\ &\le \lambda_+[\text{Var}(\varepsilon)] l^\tau (Z^\tau Z)^- l.\end{aligned}$$

Without the normality assumption on ε, the exact distribution of $l^\tau \hat{\beta}$ is very hard to obtain. The asymptotic distribution of $l^\tau \hat{\beta}$ is derived in the following result.

Theorem 3.12. Consider model (1) with assumption A3. Suppose that $0 < \inf_n \lambda_-[\text{Var}(\varepsilon)]$, where $\lambda_-[A]$ is the smallest eigenvalue of the matrix A, and that

$$\lim_{n\to\infty} \max_{1\le i\le n} Z_i^\tau (Z^\tau Z)^- Z_i = 0. \tag{2}$$

Suppose further that $n = \sum_{j=1}^k m_j$ for some integers k, m_j, $j = 1, ..., k$, with m_j's bounded by a fixed integer m, $\varepsilon = (\xi_1, ..., \xi_k)$, $\xi_j \in \mathcal{R}^{m_j}$, and ξ_j's are independent.
(i) If $\sup_i E|\varepsilon_i|^{2+\delta} < \infty$, then for any $l \in \mathcal{R}(Z)$,

$$l^\tau (\hat{\beta} - \beta) \Big/ \sqrt{\text{Var}(l^\tau \hat{\beta})} \to_d N(0, 1). \tag{3}$$

(ii) Suppose that when $m_i = m_j$, $1 \le i < j \le k$, ξ_i and ξ_j have the same distribution. Then result (3) holds for any $l \in \mathcal{R}(Z)$.
Proof. Let $l \in \mathcal{R}(Z)$. Then

$$l^\tau (Z^\tau Z)^- Z^\tau Z\beta - l^\tau \beta = 0$$

and

$$l^\tau (\hat{\beta} - \beta) = l^\tau (Z^\tau Z)^- Z^\tau \varepsilon = \sum_{j=1}^k c_{nj}^\tau \xi_j,$$

where c_{nj} is the m_j-vector whose components are $l^\tau (Z^\tau Z)^- Z_i$, $i = k_{j-1} + 1, ..., k_j$, $k_0 = 0$, and $k_j = \sum_{t=1}^j m_t$, $j = 1, ..., k$.
Note that

$$\sum_{j=1}^k \|c_{nj}\|^2 = l^\tau (Z^\tau Z)^- Z^\tau Z (Z^\tau Z)^- l = l^\tau (Z^\tau Z)^- l. \tag{4}$$

Also,

$$\begin{aligned}\max_{1\le j\le k} \|c_{nj}\|^2 &\le m \max_{1\le i\le n} [l^\tau (Z^\tau Z)^- Z_i]^2 \\ &\le m l^\tau (Z^\tau Z)^- l \max_{1\le i\le n} Z_i^\tau (Z^\tau Z)^- Z_i,\end{aligned}$$

which, together with (4) and condition (2), implies that

$$\lim_{n\to\infty}\left(\max_{1\le j\le k}\|c_{nj}\|^2 \Big/ \sum_{j=1}^{k}\|c_{nj}\|^2\right) = 0.$$

The results then follow from Corollary 1.3.

Under the conditions of Theorem 3.12, $\mathrm{Var}(\varepsilon)$ is a diagonal block matrix with $\mathrm{Var}(\xi_j)$ as the jth diagonal block, which includes the case of independent ε_i's as a special case.

Exercise 80 shows that condition (2) is almost a necessary condition for the consistency of the LSE.

The following lemma tells us how to check condition (2).

Lemma 3.3. The following are sufficient conditions for (2).
(a) $\lambda_+[(Z^\tau Z)^-] \to 0$ and $Z_n^\tau (Z^\tau Z)^- Z_n \to 0$, as $n\to\infty$.
(b) There is an increasing sequence $\{a_n\}$ such that $a_n \to \infty$, $a_n/a_{n+1} \to 1$, and $Z^\tau Z/a_n$ converges to a positive definite matrix.
Proof. (a) Since $Z^\tau Z$ depends on n, we denote $(Z^\tau Z)^-$ by A_n.
Let i_n be the integer such that $h_{i_n} = \max_{1\le i\le n} h_i$.
If $\lim_{n\to\infty} i_n = \infty$, then

$$\lim_{n\to\infty} h_{i_n} = \lim_{n\to\infty} Z_{i_n}^\tau A_n Z_{i_n} \le \lim_{n\to\infty} Z_{i_n}^\tau A_{i_n} Z_{i_n} = 0,$$

where the inequality follows from $i_n \le n$ and, thus, $A_{i_n} - A_n$ is nonnegative definite.
If $i_n \le c$ for all n, then

$$\lim_{n\to\infty} h_{i_n} = \lim_{n\to\infty} Z_{i_n}^\tau A_n Z_{i_n} \le \lim_{n\to\infty} \lambda_n \max_{1\le i\le c}\|Z_i\|^2 = 0.$$

Therefore, for any subsequence $\{j_n\}\subset\{i_n\}$ with $\lim_{n\to\infty} j_n = a \in (0,\infty]$, $\lim_{n\to\infty} h_{j_n} = 0$. This shows that $\lim_{n\to\infty} h_{i_n} = 0$.
(b) Omitted.

If $n^{-1}\sum_{i=1}^n t_i^2 \to c$ and $n^{-1}\sum_{i=1}^n t_i \to d$ in the simple linear regression model (Example 3.12), where c is positive and $c > d^2$, then condition (b) in Lemma 3.3 is satisfied with $a_n = n$ and, therefore, Theorem 3.12 applies.

In the one-way ANOVA model (Example 3.13),

$$\max_{1\le i\le n} Z_i^\tau (Z^\tau Z)^- Z_i = \lambda_+[(Z^\tau Z)^-] = \max_{1\le j\le m} n_j^{-1}.$$

Hence conditions related to Z in Theorem 3.12 are satisfied if and only if $\min_j n_j \to \infty$. Some similar conclusions can be drawn in the two-way ANOVA model (Example 3.14).

Functions of unbiased estimators

If the parameter to be estimated is $\vartheta = g(\theta)$ with a vector-valued parameter θ and U_n is a vector of unbiased estimators of components of θ, then $T_n = g(U_n)$ is often asymptotically unbiased for ϑ.
Assume that g is differentiable and $c_n(U_n - \theta) \to_d Y$. Then

$$\text{amse}_{T_n}(P) = E\{[\nabla g(\theta)]^\tau Y\}^2/c_n^2$$

(Theorem 2.6). Hence, T_n has a good performance in terms of amse if U_n is optimal in terms of mse (such as the UMVUE or BLUE).

Example 3.22. Consider a polynomial regression of order p:

$$X_i = \beta^\tau Z_i + \varepsilon_i, \qquad i = 1, ..., n,$$

where $\beta = (\beta_0, \beta_1, ..., \beta_{p-1})$, $Z_i = (1, t_i, ..., t_i^{p-1})$, and ε_i's are i.i.d. with mean 0 and variance $\sigma^2 > 0$.
Suppose that the parameter to be estimated is $t_\beta \in \mathcal{T} \subset \mathcal{R}$ such that

$$\sum_{j=0}^{p-1} \beta_j t_\beta^j = \max_{t \in \mathcal{T}} \sum_{j=0}^{p-1} \beta_j t^j.$$

Note that $t_\beta = g(\beta)$ for some function g.
Let $\hat{\beta}$ be the LSE of β.
Then the estimator $\hat{t}_\beta = g(\hat{\beta})$ is asymptotically unbiased and its amse can be derived under some conditions.

Example 3.23. In the study of the reliability of a system component, we assume that

$$X_{ij} = \boldsymbol{\theta}_i^\tau z(t_j) + \varepsilon_{ij}, \quad i = 1, ..., k, \ j = 1, ..., m.$$

Here X_{ij} is the measurement of the ith sample component at time t_j;
$z(t)$ is a q-vector whose components are known functions of the time t;
$\boldsymbol{\theta}_i$'s are unobservable random q-vectors that are i.i.d. from $N_q(\theta, \Sigma)$, where θ and Σ are unknown;
ε_{ij}'s are i.i.d. measurement errors with mean zero and variance σ^2;
$\boldsymbol{\theta}_i$'s and ε_{ij}'s are independent.
As a function of t, $\boldsymbol{\theta}^\tau z(t)$ is the degradation curve for a particular component and $\theta^\tau z(t)$ is the mean degradation curve.
Suppose that a component will fail to work if $\boldsymbol{\theta}^\tau z(t) < \eta$, a given critical value.
Assume that $\boldsymbol{\theta}^\tau z(t)$ is always a decreasing function of t.
Then the reliability function of a component is

$$R(t) = P(\boldsymbol{\theta}^\tau z(t) > \eta) = \Phi\left(\frac{\theta^\tau z(t) - \eta}{s(t)}\right),$$

where $s(t) = \sqrt{[z(t)]^\tau \Sigma z(t)}$ and Φ is the standard normal distribution function.
For a fixed t, estimators of $R(t)$ can be obtained by estimating θ and Σ, since Φ is a known function.
It can be shown (exercise) that the BLUE of θ is the LSE

$$\hat{\theta} = (Z^\tau Z)^{-1} Z^\tau \bar{X},$$

where Z is the $m \times q$ matrix whose jth row is the vector $z(t_j)$, $X_i = (X_{i1}, ..., X_{im})$, and $\bar{X}$ is the sample mean of X_i's.
The estimation of Σ is more difficult.
It can be shown (exercise) that a consistent (as $k \to \infty$) estimator of Σ is

$$\hat{\Sigma} = \frac{1}{k} \sum_{i=1}^{k} (Z^\tau Z)^{-1} Z^\tau (X_i - \bar{X})(X_i - \bar{X})^\tau Z (Z^\tau Z)^{-1} - \hat{\sigma}^2 (Z^\tau Z)^{-1},$$

where

$$\hat{\sigma}^2 = \frac{1}{k(m-q)} \sum_{i=1}^{k} [X_i^\tau X_i - X_i^\tau Z (Z^\tau Z)^{-1} Z^\tau X_i].$$

Hence an estimator of $R(t)$ is

$$\hat{R}(t) = \Phi\left(\frac{\hat{\theta}^\tau z(t) - \eta}{\hat{s}(t)}\right),$$

where

$$\hat{s}(t) = \sqrt{[z(t)]^\tau \hat{\Sigma} z(t)}.$$

$Y_{i1} = X_i^\tau Z (Z^\tau Z)^{-1} z(t)$
$Y_{i2} = [X_i^\tau Z (Z^\tau Z)^{-1} z(t)]^2$
$Y_{i3} = [X_i^\tau X_i - X_i^\tau Z (Z^\tau Z)^{-1} Z^\tau X_i]/(m-q)$
$Y_i = (Y_{i1}, Y_{i2}, Y_{i3})$ It is apparent that $\hat{R}(t)$ can be written as $g(\bar{Y})$ for a function

$$g(y_1, y_2, y_3) = \Phi\left(\frac{y_1 - \eta}{\sqrt{y_2 - y_1^2 - y_3 [z(t)]^\tau (Z^\tau Z)^{-1} z(t)}}\right).$$

Suppose that ε_{ij} has a finite fourth moment, which implies the existence of $\text{Var}(Y_i)$.
The amse of $\hat{R}(t)$ can be derived (exercise).