**L10 Least square estimators**

1. Least square estimators

   (1) Linear model
   $y_1, ..., y_n$ are observations from a population system. Then data vector $y = (y_1, ..., y_n)'$ is a random vector. Model

   $$y = X\beta + \epsilon \text{ with } E(\epsilon) = 0$$

   is a linear model since by the model specification $E(y) = X\beta$ is a linear function of unknown parameter vector $\beta \in R^p$ and $E(y)$ lies in a linear space $S = \mathcal{R}(X)$. Here $X \in R^{n\times p}$ called the design matrix is known.

   (2) Least square estimators
   While $E(y) \in \mathcal{R}(X) = S$, $y$ may or may not be in $S = \mathcal{R}(X)$. By intuition one would estimate $E(y)$ by a vector in $S = \mathcal{R}(X)$ that has minimum-distance to $y$. The corresponding value of $\beta$ is called a least square estimator (LSE) of $\beta$. So $\widehat{\beta}$ is a LSE of $\beta$ if

   $$\|y - X\widehat{\beta}\|^2 \le \|y - X\beta\|^2 \text{ for all } \beta \in R^p.$$

   Estimated $y$, $\widehat{y} = X\widehat{\beta}$, is fitted value vector; $e = y - X\widehat{\beta} = y - \widehat{y}$ is residual vector; $Q(\beta) = \|y - X\beta\|^2$; $Q(\widehat{\beta}) = \|y - X\widehat{\beta}\|^2 = \|y - \widehat{y}\|^2 = \|e\|^2 = \sum_i e_i^2$ is the sum of squared errors (SSE).

   (3) Equivalent statements
   By the definitions, the following three are equivalent
   (i) $\widehat{\beta}$ is LSE of $\beta$ (ii) $\widehat{\beta}$ is a LSS to $X\beta = y$ (iii) $X\widehat{\beta} = \pi(y \mid \mathcal{R}(X))$

   (4) The collection of all least square estimators
   Let LSE($\beta$) be the collection of all LSE of $\beta$. Then

   $$\text{LSE}(\beta) = X^+y + \mathcal{N}(X)$$

   where $X^+y$ is perpendicular to $\mathcal{N}(X)$ and hence is the minimum-norm LSE

   **Ex1:** Suppose $X \in R^{n\times p}$ has full column rank. Then
   (i) $\mathcal{N}(X) = \{0\}$.
   $\supset$ is trivial. $\subset$: $\beta \in \mathcal{N}(X) \Longrightarrow X\beta = 0 \Longrightarrow \beta = X^L 0 = 0 \in \{0\}$.
   (ii) $X^+ = (X'X)^{-1}X'$
   (iii) LSE($\beta$) contains the unique $\widehat{\beta} = X^+y = (X'X)^{-1}X'y$.

2. Restricted least square estimators

   (1) LSE under $\beta \in \mathcal{D}$
   $\mathcal{D}$ is a closed convex set in $R^p$. For restricted linear model

   $$y = X\beta + \epsilon \text{ with } E(\epsilon) = 0 \text{ under } \beta \in \mathcal{D}$$

   $\widehat{\beta}$ is LSE of $\beta$ under $\beta \in \mathcal{D}$ if

   $$\widehat{\beta} \in \mathcal{D} \text{ and } \|y - X\widehat{\beta}\|^2 \le \|y - X\beta\|^2 \text{ for all } \beta \in \mathcal{D}.$$

(2) Equivalent statements
By the definitions the following three statements are equivalent
(i) $\widehat{\beta}$ is LSE of $\beta$ under $\beta \in \mathcal{D}$ (ii) $\widehat{\beta}$ is LSS to $X\beta = y$ under $\beta \in \mathcal{D}$
(iii) $X\widehat{\beta} = \pi(y \mid X\mathcal{D})$

(3) Comments
We do have formulas for $\pi(y \mid X\mathcal{D})$ when $X\mathcal{D}$ is $\mathcal{R}(\cdot), \mathcal{R}^{\perp}(\cdot), \mathcal{N}(\cdot), \mathcal{N}^{\perp}(\cdot)$ and $\mathcal{A} = x_0 + S$.

$$\pi(y|\mathcal{A}) = \pi(y \mid x_0 + S) = x_0 + \pi(y - x_0 \mid S).$$

3. LSE under equation constraint

(1) Equation restrictions

$$\begin{aligned} A\beta = 0 &\iff \beta \in \mathcal{N}(A) = \mathcal{D} \\ &\implies X\mathcal{D} = X\mathcal{N}(A) = X\mathcal{R}(I - A^+A) = \mathcal{R}[X(I - A^+A)]. \\ A\beta = b &\iff \beta \in A^+b + \mathcal{N}(A) = \mathcal{D} \\ &\implies X\mathcal{D} = XA^+b + X\mathcal{N}(A) = XA^+b + \mathcal{R}[X(I - A^+A)]. \end{aligned}$$

(2) LSE under $A\beta = 0$

The followings are equivalent.
(i) $\widehat{\beta}$ is a LSE of $\beta$ under $A\beta = 0$
(ii) $\widehat{\beta}$ is a LSS to $X\beta = y$ under $A\beta = 0$
(iii) $X\widehat{\beta} = \pi(y \mid \mathcal{R}[X(I - A^+A)]$

The collection of all LSEs of $\beta$ under $A\beta = 0$ is $[X(I - A^+A)]^+y + \mathcal{N}(X)$.

**Proof**

$$\begin{aligned} & \widehat{\beta} \text{ is a LSE of } \beta \text{ under } A\beta = 0 \\ \iff & X\widehat{\beta} = \pi(y \mid \mathcal{R}[X(I - A^+A)]) = X[X(I - A^+A)]^+y \\ \iff & N\left\{\widehat{\beta} - [X(I - A^+A)]^+y\right\} = 0 \iff \widehat{\beta} - [X(I - A^+A)]^+y \in \mathcal{N}(X) \\ \iff & \widehat{\beta} \in [X(I - A^+A)]^+y + \mathcal{N}(X). \end{aligned}$$

(3) LSE under consistent $A\beta = b$

The followings are equivalent
(i) $\widehat{\beta}$ is a LSE of $\beta$ under $A\beta = b$
(ii) $\widehat{\beta}$ is a LSS to $X\beta = y$ under $A\beta = b$
(iii) $X\widehat{\beta} = \pi(y \mid XA^+b + \mathcal{R}[X(I - A^+A)])$.

The collection of all LSEs of $\beta$ under $A\beta = b$ is $A^+b + [X(I - A^+A)]^+(y - XA^+b) + \mathcal{N}(X)$

**Proof**

$$\begin{aligned} & \widehat{\beta} \text{ is a LSE of } beta \text{ under } A\beta = b \\ \iff & X\widehat{\beta} = \pi(y \mid XA^+b + \mathcal{R}[X(I - A^+A)]) \\ \iff & X\widehat{\beta} = XA^+b + \pi(y - XA^+b \mid \mathcal{R}[X(I - A^+A)]) \\ \iff & X\widehat{\beta} = XA^+b + X[X(I - A^+A)]^+(y - XA^+b) \\ \iff & \widehat{\beta} \in A^+b + [X(I - A^+A)]^+(y - XA^+b) + \mathcal{N}(X). \end{aligned}$$

**L11 Kronecker product and vectorization**

1. Kronecker product

   (1) Defintion

   For $A = (a_{ij})_{m\times n}$ and $B \in R^{p\times q}$, $A \otimes B \overset{def}{=\!=} \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix} \in R^{mp\times nq}$.

   (2) Associative property: $(A \otimes B) \otimes C = A \otimes (B \otimes C)$. So $A \otimes B \otimes C$.

   (3) Distributive property:
   $(A_1 + A_2) \otimes B = (A_1 \otimes B) + (A_2 \otimes B)$ and $A \otimes (B_1 + B_2) = (A \otimes B_1) + (A \otimes B_2)$.

   (4) Scalar multiplication: $\alpha A = \alpha \otimes A,\ (\alpha A) \otimes (\beta B) = \alpha\beta(A \otimes B) = (\beta A) \otimes (\alpha B)$.

   (5) $(A \otimes B)' = A' \otimes B'$. Recall: $(AB)' = B'A'$.

   **Ex1:** $\otimes$ is not commutative: $A \otimes B \neq B \otimes A$. For example,

   $$A = (1, 2),\ B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \Rightarrow A \otimes B = \begin{pmatrix} 1 & 2 & 2 & 4 \\ 3 & 4 & 6 & 8 \end{pmatrix},\ B \otimes A = \begin{pmatrix} 1 & 2 & 2 & 4 \\ 3 & 6 & 4 & 8 \end{pmatrix}.$$

   **Ex2:** From (3) and (4),

   $$(aA + bB) \otimes (cC + dD) = ac(A \otimes C) + ad(A \otimes D) + bc(B \otimes C) + bd(B \otimes D).$$

2. Matrix multiplication and Kronecker product

   (1) $(A_1A_2) \otimes (B_1B_2) = (A_1 \otimes B_1)(A_2 \otimes B_2)$

   (2) If $A \in R^{m\times m}$ and $B \in R^{n\times n}$ are non-singular, so is $A \otimes B$ and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.
   **Pf:** $(A \otimes B)(A^{-1} \otimes B^{-1}) = I_m \otimes I_n = I_{mn}$. Recall: $(AB)^{-1} = B^{-1}A^{-1}$.

   (3) $A^- \otimes B^- \subset (A \otimes B)^-$
   **Pf:** $(A \otimes B)(A^- \otimes B^-)(A \otimes B) = (AA^-A) \otimes (BB^-B) = A \otimes B$.

   (4) $(A \otimes B)^+ = A^+ \otimes B^+$.
   **Pf:** The result can be proven by showing the four conditions. For example,
   (iii) $(A \otimes B)(A^+ \otimes B^+) = (AA^+) \otimes (BB^+)$ is symmetric.

   **Ex3:** Extensions
   $(A_1 \cdots A_k) \otimes (B_1 \cdots B_k) \otimes \cdots \otimes (C_1 \cdots C_k) = (A_1 \otimes B_1 \otimes \cdots \otimes C_1) \cdots (A_k \otimes B_k \otimes \cdots \otimes C_k)$
   $(A_1 \otimes \cdots \otimes A_k)^{-1} = A_1^{-1} \otimes \cdots \otimes A_k^{-1}$ $\qquad$ $(A_1 \otimes \cdots \otimes A_k)^+ = A_1^+ \otimes \cdots \otimes A_k^+$
   $A_1^- \otimes \cdots \otimes A_k^- \subset (A_1 \otimes \cdots \otimes A_k)^-$.

   **Comment:** For (1), the LHS is always the RHS. But the RHS may not be written as the LHS. For example, $(A_{1\times 2} \otimes B_{1\times 3})(C_{3\times 1} \otimes D_{2\times 1}) \neq (A_{1\times 2}C_{3\times 1}) \otimes (B_{1\times 3}D_{2\times 1})$.

3. Vectorization

   (1) Definition

   For $A = (A_1, ..., A_n) \in R^{m\times n}$, $\text{vec}(A) = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix} \in R^{mn}$ defines a 1-1 mapping between $R^{m\times n}$ and $R^{mn}$.

(2) Properties of the transformation vec$(\cdot)$.

(i) vec$(\cdot)$ is a linear transformation
The transformation preserves linear combinations, i.e., for $C = \alpha A + \beta B$, after the transformation $\text{vec}(C) = \alpha\text{vec}(A) + \beta\text{vec}(B)$, i.e.,

$$\text{vec}(\alpha A + \beta B) = \alpha\,\text{vec}(A) + \beta\,\text{vec}(B).$$

(ii) vec$(\cdot)$ preserves inner products
For $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{m\times n}$, the Frobenius inner product

$$\langle A,\, B\rangle = \text{tr}(B'A) = \sum_{i,j} a_{ij}b_{ij}.$$

After the transformation, $\langle\text{vec}(A),\,\text{vec}(B)\rangle = [\text{vec}(B)]'[\text{vec}(A)] = \sum_{ij} a_{ij}b_{ij}$.
Thus $\langle\text{vec}(A),\,\text{vec}(B)\rangle = \langle A,\, B\rangle$.

(3) While generally $A \otimes B \neq B \otimes A$, for vectors $x \in R^m$ and $y \in R^n$,

$$x \otimes y' = y' \otimes x = xy' = \begin{pmatrix} x_1y_1 & \cdots & x_1y_n \\ \vdots & \ddots & \vdots \\ x_my_1 & \cdots & x_my_n \end{pmatrix}.$$

So $\text{vec}(x \otimes y') = \text{vec}(y' \otimes x) = \text{vec}(xy') = y \otimes x$.

(4) $\text{vec}(AXB) = (B' \otimes A)\,\text{vec}(X)$.

**Pf:** Suppose $X \in R^{p\times q}$ and $I_q = (e_1, ..., e_q)$.

$$\begin{aligned}
\text{vec}(AXB) &= \text{vec}[A(X_1, .., X_q)(e_1, .., e_q)'B] = \text{vec}\left[(AX_1, ..., AX_q)\begin{pmatrix} e_1'B \\ \vdots \\ e_q'B \end{pmatrix}\right] \\
&= \text{vec}[(AX_1)(B'e_1)' + \cdots + (AX_q)(B'e_q)'] \\
&= \text{vec}[(AX_1)(B'e_1)'] + \cdots + \text{vec}[(AX_q)(B'e_q)'] \\
&= (B'e_1) \otimes (AX_1) + \cdots + (B'e_q) \otimes (AX_q) \\
&= (B' \otimes A)(e_1 \otimes X_1) + \cdots + (B' \otimes A)(e_q \otimes X_q) \\
&= (B' \otimes A)\,\text{vec}(X_1e_1') + \cdots + (B' \otimes A)\,\text{vec}(X_1e_q') \\
&= (B' \otimes A)\,\text{vec}(X_1e_1' + \cdots + X_qe_q') \\
&= (B' \otimes A)\text{vec}(XI') = (B' \otimes A)\,\text{vec}(X).
\end{aligned}$$

**Comment:** $Y = AXB$ is a linear transformation from $R^{p\times q}$ to $R^{m\times n}$.
By 1-1 mapping, this transformation has image as a linear transformation from $R^{pq}$ to $R^{mn}$. This transformation is $\text{vec}(Y) = (B' \otimes A)\,\text{vec}(X)$.

**Ex4:** 36 (1) on p70. Show that $\text{tr}(ABC) = [\text{vec}(A')]' \cdot (I \otimes B) \cdot \text{vec}(C)$.

$$\begin{aligned}
\text{tr}(ABC) &= \text{tr}[(A')'BC] = \langle BC,\, A'\rangle = \langle\text{vec}(BC),\,\text{vec}(A')\rangle = [\text{vec}(A')]'[\text{vec}(BC)] \\
&= [\text{vec}(A')]'[\text{vec}(BCI)] = [\text{vec}(A')]' \cdot (I \otimes B) \cdot \text{vec}(C)
\end{aligned}$$