Analysis of Variance and Design of Experiments

General Linear Hypothesis and Analysis of Variance

:::

Lecture 5

Least squares and Maximum Likelihood Estimation of Parameters

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Slides can be downloaded from http://home.iitk.ac.in/~shalab/sp

Regression model for the general linear hypothesis:

Let $Y_1, Y_2, ..., Y_n$ **be a sequence of *n* independent random variables associated with responses. Then we can write it as**

$$Y_i = \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, \; i = 1, 2, ..., n; \; j = 1, 2, ..., p$$

where $\beta_1, \beta_2, ..., \beta_p$ **are the unknown parameters and** x_{ij}**'s are the known values of independent covariates** $X_1, X_2, ..., X_p$, ε_i**'s are i.i.d. random error component with**

$$E(\varepsilon_i) = 0 \;, \quad Var(\varepsilon_i) = \sigma^2, \quad Cov(\varepsilon_i, \varepsilon_j) = 0 (i \neq j).$$

$\hat{\beta} = (X'X)^{-1} X'y$ **(LSE) is the best linear unbiased estimator of** β **in the sense of having minimum variance in the class of linear and unbiased estimator.**

Least squares estimate of β:

If $L'\beta$ is a linear parametric function where $L = (\ell_1, \ell_2, \ldots, \ell_p)'$ is a non-null vector, then the least-squares estimate of $L'\beta$ is $L'\hat{\beta}$.

A question arises that what are the conditions under which a linear parametric function $L'\beta$ admits a unique least-squares estimate in the general case.

The concept of estimable function is needed to find such conditions.

Estimable functions:

A linear function $\lambda'\beta$ of the parameters with known λ is said to be an estimable parametric function (or estimable) if there exists a linear function *L'Y* of *Y* such that

$$E(L'Y) = \lambda'\beta$$

for all $\beta \in R^p$.

Note that not all parametric functions are estimable.

Following results will be useful in understanding further topics.

Estimable functions:

Theorem 1: **A linear parametric function $L'\beta$ admits a unique least squares estimate if and only if $L'\beta$ is estimable.**

Theorem 2 (Gauss Markoff theorem):

If the linear parametric function $L'\beta$ is estimable then the linear estimator $L'\hat{\beta}$ where $\hat{\beta}$ is a solution of $X'X\hat{\beta} = X'Y$ is the best linear unbiased estimator of $L'\beta$ in the sense of having minimum variance in the class of all linear and unbiased estimators of $L'\beta$.

Estimable functions:

Theorem 3: If the linear parametric functions

$$\phi_1 = l_1^{'}\beta,\ \phi_2 = l_2^{'}\beta, \ldots, \phi_k = l_k^{'}\beta$$

are estimable, then any linear combination of $\phi_1, \phi_2, \ldots, \phi_k$ is also estimable.

Estimable functions:

Theorem 4: **All linear parametric functions in β are estimable if and only if X has full rank.**

If X is not of full rank, then some linear parametric functions do not admit the unbiased linear estimators and nothing can be inferred about them.

The linear parametric functions which are not estimable are said to be confounded.

Estimable functions:

Theorem 5: Let $L_1'\beta$ and $L_2'\beta$ be two estimable parametric functions and let $L_1'\hat{\beta}$ and $L_2'\hat{\beta}$ be their least squares estimators. Then

$$Var(L_1'\hat{\beta}) = \sigma^2 \; L_1'(X'X)^{-1}L_1$$

$$Cov(L_1'\hat{\beta}, L_2'\hat{\beta}) = \sigma^2 \; L_1'(X'X)^{-1}L_2$$

assuming that *X* is a full rank matrix.

If not, the generalized inverse of *X'X* can be used in place of the unique inverse.

Estimator of σ^2 based on least squares estimation:

Consider an estimator of σ^2 as,

$$\hat{\sigma}^2 = \frac{1}{n-p}(y - X\hat{\beta})'(y - X\hat{\beta})$$

$$= \frac{1}{n-p}[y - X(X'X)^{-1}X'y]'[y - X(X'X)^{-1}X'y]$$

$$= \frac{1}{n-p}y'[I - X(X'X)^{-1}X'][I - X(X'X)^{-1}X']y$$

$$= \frac{1}{n-p}y'[I - X(X'X)^{-1}X']y$$

where the hat matrix $[I - X(X'X)^{-1}X']$ is an idempotent matrix with its trace as

Estimator of σ^2 based on least squares estimation:

$$
\begin{aligned}
tr[I - X(X'X)^{-1}X'] &= trI - trX(X'X)^{-1}X' \\
&= n - tr(X'X)^{-1}X'X \\
&\qquad (\text{using the result } tr(AB) = tr(BA)) \\
&= n - tr\, I_p \\
&= n - p.
\end{aligned}
$$

Using the result $E(y'Ay) = tr(A)\Sigma$ **with** $Cov(y) = \Sigma$ **, we have**

$$
\begin{aligned}
E(\hat{\sigma}^2) &= \frac{\sigma^2}{n-p} tr[I - X(X'X)^{-1}X'] \\
&= \sigma^2
\end{aligned}
$$

and so $\hat{\sigma}^2$ **is an unbiased estimator of** σ^2**.**

Maximum likelihood estimation:

The least-squares method does not use any distribution of the random variables in case of the estimation of parameters.

We need the distributional assumption in case of least squares only while constructing the tests for hypothesis and the confidence intervals.

For maximum likelihood estimation, we need the distributional assumption from the beginning.

Maximum likelihood estimation:

Suppose $y_1, y_2, \ldots, y_n$ **are independently and identically distributed following a normal distribution with**

- **mean:** $E(y_i) = \sum_{j=1}^{p} \beta_j x_{ij}$ **and**
- **variance:** $Var(y_i) = \sigma^2$ **(*i* =1, 2,…, *n*).**

Then the likelihood function of $y_1, y_2, \ldots, y_n$ **is**

$$L(y \mid \beta, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right]$$

where $y = (y_1, y_2, \ldots, y_n)'$.

Maximum likelihood estimation:

Then

$$L = \ln L(y|\beta, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

Differentiating the log-likelihood with respect to β **and** σ^2,

we have

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow X'X\tilde{\beta} = X'y$$

$$\frac{\partial L}{\partial \sigma^2} = 0 \Rightarrow \tilde{\sigma}^2 = \frac{1}{n}(y - X\tilde{\beta})'(y - X\tilde{\beta})$$

Maximum likelihood estimation:

Assuming the full rank of *X*, the normal equations are solved and the maximum likelihood estimators are obtained as

$$\tilde{\beta} = (X'X)^{-1}X'y$$

$$\tilde{\sigma}^2 = \frac{1}{n}(y - X\tilde{\beta})'(y - X\tilde{\beta})$$

$$= \frac{1}{n}y'\left[I - X(X'X)^{-1}X'\right]y.$$

The second-order differentiation conditions can be checked and they are satisfied for $\hat{\beta}$ and $\tilde{\sigma}^2$ to be the maximum likelihood estimators.

Maximum likelihood estimation:

Note that in the maximum likelihood estimator $\tilde{\beta}$ is the same as the least-squares estimator $\tilde{\beta}$ and

- **$\tilde{\beta}$ is an unbiased estimator of β, i.e., $E(\tilde{\beta}) = \beta$ unlike the least-squares estimator but**

- **$\tilde{\sigma}^2$ is not an unbiased estimator of σ^2, i.e.,**

$$E(\tilde{\sigma}^2) = \frac{n-p}{n}\sigma^2 \neq \sigma^2$$

like the least-squares estimator.

Now we use the following theorems for developing the test of hypothesis.

Maximum likelihood estimation:

Theorem 6: Let $Y = (Y_1, Y_2, ..., Y_n)'$ follow a multivariate normal distribution $N(\mu, \Sigma)$ with mean vector μ and positive definite covariance matrix Σ

Then $Y'AY$ follows a non-central chi-square distribution with *p* degrees of freedom and non-centrality parameter $\mu' A \mu$, i.e., $\chi^2(p, \mu' A\mu)$ if and only if ΣA is an idempotent matrix of rank *p*.

Maximum likelihood estimation:

Theorem 7: **Let** $Y = (Y_1, Y_2, ..., Y_n)'$ **follow a multivariate normal distribution** $N(\mu, \Sigma)$ **with mean vector** μ **and positive definite covariance matrix** Σ.

Let $Y'A_1Y \sim \chi^2(p_1, \mu' A_1 \mu)$ **and** $Y'A_2Y \sim \chi^2(p_2, \mu' A_2 \mu)$.

Then $Y'A_1Y$ **and** $Y'A_2Y$ **are independently distributed if** $A_1 \Sigma A_2 = 0.$

Maximum likelihood estimation:

Theorem 8: **Let** $Y=(Y_1,Y_2,...,Y_n)'$ **follow a multivariate normal distribution** $N(\mu,\sigma^2 I)$.

Then the maximum likelihood (or least squares) estimator $L'\hat{\beta}$ **of estimable linear parametric function is independently distributed of** $\hat{\sigma}^2$;

$$L'\hat{\beta} \sim N\left[L'\beta, L'(X'X)^{-1}L\right]$$

and

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p)$$

where $rank(X) = p.$

Maximum likelihood estimation:

Proof: **Consider** $\hat{\beta} = (X'X)^{-1}X'Y$, **then**

$$E(L'\hat{\beta}) = L'(X'X)^{-1}X'E(Y)$$
$$= L'(X'X)^{-1}X'X\beta$$
$$= L'\beta$$

$$Var(L'\hat{\beta}) = L'Var(\hat{\beta})L$$
$$= L'E(\hat{\beta}-\beta)'(\hat{\beta}-\beta)L$$
$$= \sigma^2 L'(X'X)^{-1}L.$$

Since $\hat{\beta}$ **is a linear function of *y* and** $L'\hat{\beta}$ **is a linear function of** $\hat{\beta}$**, so** $L'\hat{\beta}$ **follows a normal distribution** $N\left[L'\beta, \underline{\sigma^2 L'(X'X)^{-1}L}\right]$

Maximum likelihood estimation:

Proof: **Let** $A = I - X(X'X)^{-1}X'$ **and** $B = L'(X'X)^{-1}X'$, **then**

$$L'\hat{\beta} = L'(X'X)^{-1}X'Y = BY$$

and

$$n\hat{\sigma}^2 = (Y - X\beta)'\left[I - X(X'X)^{-1}X'\right](Y - X\beta) = Y'AY.$$

So, using Theorem 6 with *rank*(*A*) = *n* - *p*,

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p).$$

Also $$BA = L'(X'X)^{-1}X' - L'(X'X)^{-1}X'X(X'X)^{-1}X'$$

$$= 0.$$

So using Theorem 7, $Y'AY$ **and** BY **are independently distributed.**