Now, in order to obtain the properties of $\hat{\beta}$ such as mean, variance, distribution and so on, (19) is rewritten as follows:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u. \qquad (20)\end{aligned}$$

Taking the expectation on both sides of (20), we have the following:

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}(\beta + (X'X)^{-1}X'u) = \beta + (X'X)^{-1}X'\mathrm{E}(u) = \beta,$$

because of $\mathrm{E}(u) = 0$ by the assumption of the error term u_i.

Thus, unbiasedness of $\hat{\beta}$ is shown.

The variance of $\hat{\beta}$ is obtained as:

$$\begin{aligned}\mathrm{V}(\hat{\beta}) &= \mathrm{E}((\hat{\beta} - \beta)(\hat{\beta} - \beta)') = \mathrm{E}\Big((X'X)^{-1}X'u((X'X)^{-1}X'u)'\Big) \\ &= \mathrm{E}((X'X)^{-1}X'uu'X(X'X)^{-1}) = (X'X)^{-1}X'\mathrm{E}(uu')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}.\end{aligned}$$

The first equality is the definition of variance in the case of vector.

In the fifth equality, $\mathrm{E}(uu') = \sigma^2 I_n$ is used, which implies that $\mathrm{E}(u_i^2) = \sigma^2$ for all i and $\mathrm{E}(u_i u_j) = 0$ for $i \neq j$.

Remember that $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2.

Under normality assumption on the error term u, it is known that the distribution of $\hat{\beta}$ is given by:

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}).$$

Proof:

θ_u: $n \times 1$, $\quad u$: $n \times 1$, $\quad \theta_\beta$: $k \times 1$, $\quad \hat{\beta}$: $k \times 1$

The moment-generating function of u, i.e., $\phi_u(\theta_u)$, is:

$$\phi_u(\theta_u) = \mathrm{E}\Big(\exp(\theta_u' u)\Big) = \exp\Big(\frac{\sigma^2}{2}\theta_u'\theta_u\Big),$$

which is $N(0, \sigma^2 I_n)$.

The moment-generating function of $\hat{\beta}$, i.e., $\phi_\beta(\theta_\beta)$, is:

$$\begin{aligned}\phi_\beta(\theta_\beta) &= \mathrm{E}\Big(\exp(\theta_\beta'\hat{\beta})\Big) = \mathrm{E}\Big(\exp(\theta_\beta'\beta + \theta_\beta'(X'X)^{-1}X'u)\Big) \\ &= \exp(\theta_\beta'\beta)\mathrm{E}\Big(\theta_\beta'(X'X)^{-1}X'u\Big) = \exp(\theta_\beta'\beta)\phi_u\Big(\theta_\beta'(X'X)^{-1}X'\Big) \\ &= \exp(\theta_\beta'\beta)\exp\Big(\frac{\sigma^2}{2}\theta_\beta'(X'X)^{-1}\theta_\beta\Big) = \exp\Big(\theta_\beta'\beta + \frac{\sigma^2}{2}\theta_\beta'(X'X)^{-1}\theta_\beta\Big),\end{aligned}$$

which is equivalent to the normal distribution with mean β and variance $\sigma^2(X'X)^{-1}$.

Note that $\quad \theta_u = X(X'X)^{-1}\theta_\beta$. QED

Taking the jth element of $\hat{\beta}$, its distribution is given by:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 a_{jj}), \qquad \text{i.e.,} \qquad \frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{a_{jj}}} \sim N(0,1),$$

where a_{jj} denotes the jth diagonal element of $(X'X)^{-1}$.

Replacing σ^2 by its estimator s^2, we have the following t distribution:

$$\frac{\hat{\beta}_j - \beta_j}{s\sqrt{a_{jj}}} \sim t(n-k),$$

where $t(n-k)$ denotes the t distribution with $n-k$ degrees of freedom.

s^2 is taken as follows:

$$s^2 = \frac{1}{n-k}\sum_{i=1}^n e_i^2 = \frac{1}{n-k}e'e = \frac{1}{n-k}(y - X\hat{\beta})'(y - X\hat{\beta}),$$

which leads to an unbiased estimator of σ^2.

Proof:

Substitute $y = X\beta + u$ and $\hat{\beta} = \beta + (X'X)^{-1}X'u$ into $e = y - X\hat{\beta}$.

$$\begin{aligned}e &= y - X\hat{\beta} = X\beta + u - X(\beta + (X'X)^{-1}X'u) \\ &= u - X(X'X)^{-1}X'u = (I_n - X(X'X)^{-1}X')u\end{aligned}$$

$I_n - X(X'X)^{-1}X'$ is idempotent and symmetric, because we have:

$$(I_n - X(X'X)^{-1}X')(I_n - X(X'X)^{-1}X') = I_n - X(X'X)^{-1}X,'$$
$$(I_n - X(X'X)^{-1}X')' = I_n - X(X'X)^{-1}X'.$$

s^2 is rewritten as follows:

$$\begin{aligned} s^2 &= \frac{1}{n-k}e'e = \frac{1}{n-k}((I_n - X(X'X)^{-1}X')u)'(I_n - X(X'X)^{-1}X')u \\ &= \frac{1}{n-k}u'(I_n - X(X'X)^{-1}X')'(I_n - X(X'X)^{-1}X')u \\ &= \frac{1}{n-k}u'(I_n - X(X'X)^{-1}X')u \end{aligned}$$

Take the expectation of $u'(I_n - X(X'X)^{-1}X')u$ and note that $\mathrm{tr}(a) = a$ for a scalar a.

$$\begin{aligned} \mathrm{E}(s^2) &= \frac{1}{n-k}\mathrm{E}\Big(\mathrm{tr}\big(u'(I_n - X(X'X)^{-1}X')u\big)\Big) = \frac{1}{n-k}\mathrm{E}\Big(\mathrm{tr}\big((I_n - X(X'X)^{-1}X')uu'\big)\Big) \\ &= \frac{1}{n-k}\mathrm{tr}\Big((I_n - X(X'X)^{-1}X')\mathrm{E}(uu')\Big) = \frac{1}{n-k}\sigma^2\mathrm{tr}\Big((I_n - X(X'X)^{-1}X')I_n\Big) \\ &= \frac{1}{n-k}\sigma^2\mathrm{tr}(I_n - X(X'X)^{-1}X') = \frac{1}{n-k}\sigma^2(\mathrm{tr}(I_n) - \mathrm{tr}(X(X'X)^{-1}X')) \\ &= \frac{1}{n-k}\sigma^2(\mathrm{tr}(I_n) - \mathrm{tr}((X'X)^{-1}X'X)) = \frac{1}{n-k}\sigma^2(\mathrm{tr}(I_n) - \mathrm{tr}(I_k)) \\ &= \frac{1}{n-k}\sigma^2(n-k) = \sigma^2 \end{aligned}$$

$\longrightarrow$ s^2 is an unbiased estimator of σ^2.

Note that we do not need normality assumption for unbiasedness of s^2.

Under normality assumption for u, the distribution of s^2 is:

$$\frac{(n-k)s^2}{\sigma^2} = \frac{u'(I_n - X(X'X)^{-1}X')u}{\sigma^2} \sim \chi^2(\mathrm{tr}(I_n - X(X'X)^{-1}X'))$$

Note that $\mathrm{tr}(I_n - X(X'X)^{-1}X') = n - k$, because

$$\mathrm{tr}(I_n) = n$$
$$\mathrm{tr}(X(X'X)^{-1}X') = \mathrm{tr}((X'X)^{-1}X'X) = \mathrm{tr}(I_k) = k$$

Trace (トレース):

1. A: $n \times n$, $\mathrm{tr}(A) = \sum_{i=1}^n a_{ii}$, where a_{ij} denotes an element in the ith row and the jth column of a matrix A.
2. a: scalar (1×1), $\mathrm{tr}(a) = a$
3. A: $n \times k$, B: $k \times n$, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$
4. $\mathrm{tr}(X(X'X)^{-1}X') = \mathrm{tr}((X'X)^{-1}X'X) = \mathrm{tr}(I_k) = k$
5. When X is a vector of random variables, $\mathrm{E}(\mathrm{tr}(X)) = \mathrm{tr}(\mathrm{E}(X))$

Asymptotic Normality (without normality assumption on u): Using the central limit theorem, without normality assumption we can show that as $n \longrightarrow \infty$, under the condition of $\frac{1}{n}X'X \longrightarrow M$ we have the following result:

$$\frac{\hat{\beta}_j - \beta_j}{s\sqrt{a_{jj}}} \longrightarrow N(0,1),$$

where M denotes a $k \times k$ constant matrix.

Thus, we can construct the confidence interval and the testing procedure, using the t distribution under the normality assumption or the normal distribution without the normality assumption.

4 Properties of OLSE

1. Properties of $\hat{\beta}$: **BLUE (best linear unbiased estimator，最良線形不偏推定量)**, i.e., minimum variance within the class of linear unbiased estimators **(Gauss-Markov theorem，ガウス・マルコフの定理)**

   Proof:

   Consider another linear unbiased estimator, which is denoted by $\widetilde{\beta} = Cy$.

   $$\widetilde{\beta} = Cy = C(X\beta + u) = CX\beta + Cu,$$

   where C is a $k \times n$ matrix.

Taking the expectation of $\widetilde{\beta}$, we obatin:

$$\mathrm{E}(\widetilde{\beta}) = CX\beta + C\mathrm{E}(u) = CX\beta$$

Because we have assumed that $\widetilde{\beta} = Cy$ is unbiased, $\mathrm{E}(\widetilde{\beta}) = \beta$ holds.

That is, we need the condition: $CX = I_k$.

Next, we obtain the variance of $\widetilde{\beta} = Cy$.

$$\widetilde{\beta} = C(X\beta + u) = \beta + Cu.$$

Therefore, we have:

$$\mathrm{V}(\widetilde{\beta}) = \mathrm{E}(\widetilde{\beta} - \beta)(\widetilde{\beta} - \beta)' = \mathrm{E}(Cuu'C') = \sigma^2 CC'$$