[Review] Random Variables:

Let X_1, X_2, $\cdots$, X_n be n random variavles, which are mutually independently and identically distributed.

mutually independent $\Longrightarrow$ $f(x_i, x_j) = f_i(x_i)f_j(x_j)$ for $i \neq j$.

$f(x_i, x_j)$ denotes a joint distribution of X_i and X_j.

$f_i(x)$ indicates a marginal distribution of X_i.

identical $\Longrightarrow$ $f_i(x) = f_j(x)$ for $i \neq j$.

[End of Review]

[Review] Mean and Variance:

Let X and Y be random variables (continuous type), which are independently distributed.

Definition and Formulas:

- $\mathrm{E}(g(X)) = \displaystyle\int g(x)f(x)\mathrm{d}x$ for a function $g(\cdot)$ and a density function $f(\cdot)$.
- $\mathrm{V}(X) = \mathrm{E}((X-\mu)^2) = \displaystyle\int (x-\mu)^2 f(x)\mathrm{d}x$ for $\mu = \mathrm{E}(X)$.
- $\mathrm{E}(aX+b) = a\mathrm{E}(X)+b$ and $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$.
- $\mathrm{E}(X \pm Y) = \mathrm{E}(X) \pm \mathrm{E}(Y)$ and $\mathrm{V}(X \pm Y) = \mathrm{V}(X) + \mathrm{V}(Y)$.

[End of Review]

Mean and Variance of $\hat{\beta}_2$: $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2, but they are not necessarily normal.

Remember that we do not need normality assumption to obtain mean and variance but the normality assumption is required to test a hypothesis.

From (16), the expectation of $\hat{\beta}_2$ is derived as follows:

$$\mathrm{E}(\hat{\beta}_2) = \mathrm{E}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \beta_2 + \mathrm{E}(\sum_{i=1}^{n} \omega_i u_i) = \beta_2 + \sum_{i=1}^{n} \omega_i \mathrm{E}(u_i) = \beta_2. \qquad (17)$$

It is shown from (17) that the ordinary least squares estimator $\hat{\beta}_2$ is an **unbiased estimator (不偏推定量)** of β_2.

From (16), the variance of $\hat{\beta}_2$ is computed as:

$$\begin{aligned}
\mathrm{V}(\hat{\beta}_2) &= \mathrm{V}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \mathrm{V}(\sum_{i=1}^{n} \omega_i u_i) = \sum_{i=1}^{n} \mathrm{V}(\omega_i u_i) = \sum_{i=1}^{n} \omega_i^2 \mathrm{V}(u_i) \\
&= \sigma^2 \sum_{i=1}^{n} \omega_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.
\end{aligned} \tag{18}$$

The third equality holds because $u_1, u_2, \cdots, u_n$ are mutually independent.

The last equality comes from (15).

Thus, $\mathrm{E}(\hat{\beta}_2)$ and $\mathrm{V}(\hat{\beta}_2)$ are given by (17) and (18).

Gauss-Markov Theorem (ガウス・マルコフ定理): $\hat{\beta}_2$ has minimum variance within a class of the linear unbiased estimators.

$\longrightarrow$ **best linear unbiased estimator (BLUE, 最良線型不偏推定量)**

(Proof is omitted.)

Distribution of $\hat{\beta}_2$: We discuss the small sample properties of $\hat{\beta}_2$.

In order to obtain the distribution of $\hat{\beta}_2$ in small sample, the distribution of the error term has to be assumed.

Therefore, the extra assumption is that $u_i \sim N(0, \sigma^2)$.

Writing (16), again, $\hat{\beta}_2$ is represented as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i.$$

First, we obtain the distribution of the second term in the above equation.

It is well known that sum of normal random variables results in a normal distribution.

Therefore, $\sum_{i=1}^{n} \omega_i u_i$ is distributed as:

$$\sum_{i=1}^{n} \omega_i u_i \sim N(0,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2).$$

Therefore, $\hat{\beta}_2$ is distributed as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} \omega_i u_i \sim N(\beta_2,\ \sigma^2 \sum_{i=1}^{n} \omega_i^2),$$

or equivalently,

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma \sqrt{\sum_{i=1}^{n} \omega_i^2}} = \frac{\hat{\beta}_2 - \beta_2}{\sigma / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \sim N(0, 1),$$

for any n.

Moreover, replacing σ^2 by its estimator $s^2 = \dfrac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)^2$, it is known that we have:

$$\frac{\hat{\beta}_2 - \beta_2}{s / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \sim t(n-2),$$

where $t(n-2)$ denotes t distribution with $n-2$ degrees of freedom.

Thus, under normality assumption on the error term u_i, the $t(n-2)$ distribution is used for the confidence interval and the testing hypothesis in small sample.

Or, taking the square on both sides,

$$\left(\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\right)^2 \sim F(1, n-2).$$

[Review] Confidence Interval (信頼区間，区間推定)):

Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Then, we can obtain: $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$, where $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})^2$.

That is,

$$P\Big(-t_{\alpha/2}(n-1) < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1)\Big) = 1-\alpha$$

i.e.,

$$P\Big(\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\Big) = 1-\alpha.$$

Note that $t_{\alpha/2}(n-1)$ is obtained from the t distribution table, given α and $n-1$.

Then, replacing $\overline{X}$ by $\overline{x}$, we obtain the $100(1-\alpha)\%$ confidence interval of μ as follows:

$$(\overline{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ \overline{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}).$$

[End of Review]

In the case of OLS,

$$P\Big(-t_{\alpha/2}(n-2) < \frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} < t_{\alpha/2}(n-2)\Big) = 1 - \alpha,$$

where $t_{\alpha/2}(n-2)$ denotes $100 \times \alpha/2\%$ point from the $t(n-2)$ distribution.

Rewriting,

$$P\Big(\hat{\beta}_2 - t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}} < \beta_2 < \hat{\beta}_2 + t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\Big) = 1 - \alpha.$$

Replacing $\hat{\beta}_2$ and s^2 by observed data, the $100(1-\alpha)\%$ confidence interval of β_2 is given by:

$$\Big(\hat{\beta}_2 - t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}},\ \hat{\beta}_2 + t_{\alpha/2}(n-2)\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}\Big).$$

[Review] Testing the Hypothesis (仮説検定):

Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2.

Then, we obtain: $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$, where $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$, which is known as the unbiased estimator of σ^2.

- The null hypothesis $H_0: \ \mu = \mu_0$, where μ_0 is a fixed number.
- The alternative hypothesis $H_1: \ \mu \neq \mu_0$

Under the null hypothesis, we have the disribution: $\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$.

Replacing $\overline{X}$ and S^2 by $\overline{x}$ and s^2, compare $\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}}$ and $t(n-1)$.

H_0 is rejected when $\left|\dfrac{\overline{x}-\mu_0}{s/\sqrt{n}}\right| > t_{\alpha/2}(n-1)$.

$t_{\alpha/2}(n-1)$ is obtained from the significance level α and the degrees of freedom $n-1$.

[End of Review]

In the case of OLS, the hypotheses are as follows:

- The null hypothesis $H_0: \ \beta_2 = \beta_2^*$
- The alternative hypothesis $H_1: \ \beta_2 \neq \beta_2^*$

Under H_0,

$$\frac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}} \sim t(n-2).$$

Replacing $\hat{\beta}_2$ and s^2 by the observed data, compare $\dfrac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}$ and $t(n-2)$.

H_0 is rejected at significance level α when $\left|\dfrac{\hat{\beta}_2 - \beta_2^*}{s/\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}\right| > t_{\alpha/2}(n-1)$.

(*) $\hat{\beta}_2$ = Coefficient, $\quad \dfrac{s}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}$ = Standard Error,

s = Standard Error of Regression

3　多重回帰

n 組のデータ $(Y_i, X_{1i}, X_{2i}, \cdots, X_{ki})$, $i = 1, 2, \cdots, n$ を用いて，k 変数の多重回帰モデルを考える。

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,$$

ただし，X_{ji} は j 番目の説明変数の第 i 番目の観測値を表す。u_i は誤差項 (または，攪乱項) で，同じ仮定を用いる (すなわち，$u_1, u_2, \cdots, u_n$ は互いに独立に，平均ゼロ，分散 σ^2 の正規分布に従う)。

$\beta_1, \beta_2, \cdots, \beta_k$ は推定されるべきパラメータである。

すべての i について，$X_{1i} = 1$ とすれば，β_1 は定数項として表される。

次のような関数 $S(\beta_1, \beta_2, \cdots, \beta_k)$ を定義する。

$$S(\beta_1, \beta_2, \cdots, \beta_k) = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (Y_i - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2$$

このとき,

$$\min_{\beta_1,\beta_2,\cdots,\beta_k} S(\beta_1, \beta_2, \cdots, \beta_k)$$

となるような $\beta_1, \beta_2, \cdots, \beta_k$ を求める。$\Longrightarrow$ 最小自乗法

このときの解を $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ とする。

最小化のためには,

$$\frac{\partial S(\beta_1, \beta_2, \cdots, \beta_k)}{\partial \beta_1} = 0, \quad \frac{\partial S(\beta_1, \beta_2, \cdots, \beta_k)}{\partial \beta_2} = 0, \quad \cdots, \quad \frac{\partial S(\beta_1, \beta_2, \cdots, \beta_k)}{\partial \beta_k} = 0$$

を満たす $\beta_1, \beta_2, \cdots, \beta_k$ が $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ となる。

すなわち, $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ は,

$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki}) X_{1i} = 0,$$
$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki}) X_{2i} = 0,$$

$$\vdots$$

$$\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})X_{ki} = 0,$$

を満たす。

さらに,

$$\sum_{i=1}^{n} X_{1i}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}^2 + \widehat{\beta}_2 \sum_{i=1}^{n} X_{1i}X_{2i} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{1i}X_{ki},$$

$$\sum_{i=1}^{n} X_{2i}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{2i} + \widehat{\beta}_2 \sum_{i=1}^{n} X_{2i}^2 + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{2i}X_{ki},$$

$$\vdots$$

$$\sum_{i=1}^{n} X_{ki}Y_i = \widehat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{ki} + \widehat{\beta}_2 \sum_{i=1}^{n} X_{2i}X_{ki} + \cdots + \widehat{\beta}_k \sum_{i=1}^{n} X_{ki}^2,$$

行列表示によって，

$$\begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix} \begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix},$$

が得られ，$\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ についてまとめると，

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \vdots \\ \widehat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki} \\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{pmatrix},$$

を解くことになる。$\Longrightarrow$ コンピュータによって計算

3.1　推定量の性質

$\beta_1, \beta_2, \cdots, \beta_k$ の最小二乗推定量は $\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ とする。

誤差項 (または，攪乱項) u_i の分散 σ^2 の推定量 s^2 は,

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n}\widehat{u}_i^2 = \frac{1}{n-k}\sum_{i=1}^{n}(Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki})^2$$

として表される。

このとき,

$$\mathrm{E}(\widehat{\beta}_j) = \beta_j, \qquad \mathrm{E}(s^2) = \sigma^2,$$

を証明することが出来る。(証明略)

分布について：　$\widehat{\beta}_1, \widehat{\beta}_2, \cdots, \widehat{\beta}_k$ の分散は以下のように表される。

$$\mathrm{V}\begin{pmatrix}\widehat{\beta}_1\\ \widehat{\beta}_2\\ \vdots\\ \widehat{\beta}_k\end{pmatrix} = \begin{pmatrix}\mathrm{V}(\widehat{\beta}_1) & \mathrm{Cov}(\widehat{\beta}_1,\widehat{\beta}_2) & \cdots & \mathrm{Cov}(\widehat{\beta}_1,\widehat{\beta}_k)\\ \mathrm{Cov}(\widehat{\beta}_2,\widehat{\beta}_1) & \mathrm{V}(\widehat{\beta}_2) & \cdots & \mathrm{Cov}(\widehat{\beta}_2,\widehat{\beta}_k)\\ \vdots & \vdots & \ddots & \vdots\\ \mathrm{Cov}(\widehat{\beta}_k,\widehat{\beta}_1) & \mathrm{Cov}(\widehat{\beta}_k,\widehat{\beta}_2) & \cdots & \mathrm{V}(\widehat{\beta}_k)\end{pmatrix}$$

$$= \sigma^2 \begin{pmatrix}\sum X_{1i}^2 & \sum X_{1i}X_{2i} & \cdots & \sum X_{1i}X_{ki}\\ \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \cdots & \sum X_{2i}X_{ki}\\ \vdots & \vdots & \ddots & \vdots\\ \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \cdots & \sum X_{ki}^2\end{pmatrix}^{-1}$$

$\widehat{\beta}_j$ の分散 (すなわち，上の逆行列の j 番目の対角要素) を，

$$\mathrm{V}(\widehat{\beta}_j) = \sigma^2_{\widehat{\beta}_j},$$

として，その推定量を $s^2_{\widehat{\beta}_j}$ とする。

このとき,

$$\widehat{\beta}_j \sim N(\beta_j, \sigma^2_{\widehat{\beta}_j}),$$

となり，標準化すると,

$$\frac{\widehat{\beta}_j - \beta_j}{\sigma_{\widehat{\beta}_j}} \sim N(0, 1),$$

が得られる。さらに,

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k),$$

となり (証明略)，しかも，$\widehat{\beta}_j$ と s^2 の独立性から (証明略),

$$\frac{\widehat{\beta}_j - \beta_j}{s_{\widehat{\beta}_j}} \sim t(n-k)$$

となる。

よって，通常の区間推定や仮説検定を行うことが出来る。

決定係数について：　また，決定係数 R^2 についても同様に表される。

$$R^2 = \frac{\sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2} = 1 - \frac{\sum_{i=1}^n \widehat{u}_i^2}{\sum_{i=1}^n (Y_i - \overline{Y})^2}$$

ただし，$\widehat{Y}_i = \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} + \cdots + \widehat{\beta}_k X_{ki}$，$Y_i = \widehat{Y}_i + \widehat{u}_i$ である。

R^2 は，説明変数を増やすことによって，必ず大きくなる。なぜなら，説明変数が増えることによって，$\sum_{i=1}^n \widehat{u}_i^2$ が必ず減少するからである。

R^2 を基準にすると，被説明変数にとって意味のない変数でも，説明変数が多いほど，よりよいモデルということになる。この点を改善するために，自由度修正済み決定係数 $\overline{R}^2$ を用いる。

$$\overline{R}^2 = 1 - \frac{\sum_{i=1}^n \widehat{u}_i^2/(n-k)}{\sum_{i=1}^n (Y_i - \overline{Y})^2/(n-1)},$$

$\sum_{i=1}^n \widehat{u}_i^2/(n-k)$ は u_i の分散 σ^2 の不偏推定量であり，$\sum_{i=1}^n (Y_i - \overline{Y})^2/(n-1)$ は Y_i の分散の不偏推定量である。

R^2 と $\overline{R}^2$ との関係は，

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k},$$

となる。さらに，

$$\frac{1-\overline{R}^2}{1-R^2} = \frac{n-1}{n-k} \geq 1,$$

という関係から，$\overline{R}^2 \leq R^2$ という結果を得る。($k = 1$ のときのみに，等号が成り立つ。)

数値例：　今までと同じ数値例で，$\overline{R}^2$ を計算する。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$	$\widehat{u}_i$
1	6	10	60	100	6.8	−0.8
2	9	12	108	144	8.1	0.9
3	10	14	140	196	9.4	0.6
4	10	16	160	256	10.7	−0.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$
	35	52	468	696	35	0
平均	$\overline{Y}$	$\overline{X}$				
	8.75	13				

まず R^2 は,

$$R^2 = 1 - \frac{\sum \widehat{u}_i^2}{\sum Y_i^2 - n\overline{Y}^2} = 1 - \frac{(-0.8)^2 + 0.9^2 + 0.6^2 + (-0.7)^2}{35 - 4 \times 8.75^2} = 1 - \frac{2.30}{10.75} = 0.786$$

となり，$\overline{R}^2$ は，

$$\overline{R}^2 = 1 - \frac{\sum \widehat{u}_i^2/(n-k)}{(\sum Y_i^2 - n\overline{Y}^2)/(n-1)} = 1 - \frac{2.30/(4-2)}{10.75/(4-1)} = 0.679$$

となる。

注意： R^2 や $\overline{R}^2$ を比較する場合，被説明変数が同じことが必要である。被説明変数が異なる場合 (例えば，被説明変数を上昇率とするかそのままの値を用いるかによって，被説明変数が異なる)，誤差項 u_i の標準誤差で比較すべきである (標準誤差の小さいモデルを採用する)。$\Longrightarrow$ 関数型の選択

4　系列相関：*DW* について

4.1　*DW* について

最小自乗法の仮定の一つに,「攪乱項 $u_1, u_2, \cdots, u_n$ はそれぞれ独立に分布する」というものがあった。ダービン・ワトソン比 (*DW*) とは，誤差項の系列相関，すなわち，u_i と u_{i-1} との間の相関の有無を検定するために考案された。

$\Longrightarrow$ 時系列データのときのみ有効

$u_1, u_2, \cdots, u_n$ の系列について，それぞれの符号が，+ + + - - - - + + - - - + + のように，プラスが連続で続いた後で，マイナスが連続で続くというような場合, $u_1, u_2, \cdots, u_n$ は正の系列相関があると言う。また，+ - + - + - + - + のように交互にプラス，マイナスになる場合，$u_1, u_2, \cdots, u_n$ 負の系列相関があると言う。

特徴： $u_1, u_2, \cdots, u_i$ から u_{i+1} の符号が予想できる。$\Longrightarrow$「$u_1, u_2, \cdots, u_n$ はそれぞれ独立に分布する」という仮定に反する。

すなわち，ダービン・ワトソン比とは，回帰式が

$$Y_i = \alpha + \beta X_i + u_i,$$

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときに，H_0：$\rho = 0, H_1$：$\rho \neq 0$ の検定である。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

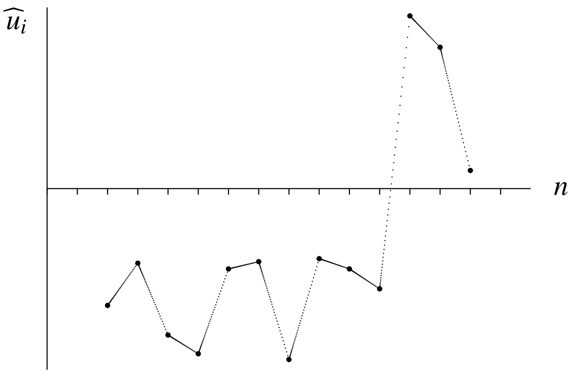

図 4： 正の系列相関

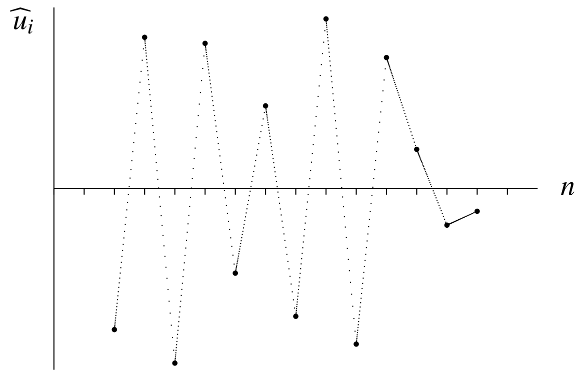

図 5： 負の系列相関

ダービン・ワトソン比の定義は次の通りである。

$$DW=\frac{\sum_{i=2}^{n}(\widehat{u_i}-\widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2}$$

DW は近似的に，次のように表される。

$$\begin{aligned}DW&=\frac{\sum_{i=2}^{n}(\widehat{u_i}-\widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2}=\frac{\sum_{i=2}^{n}\widehat{u}_i^2-2\sum_{i=2}^{n}\widehat{u_i}\widehat{u}_{i-1}+\sum_{i=2}^{n}\widehat{u}_{i-1}^2}{\sum_{i=1}^{n}\widehat{u}_i^2}\\&=\frac{2\sum_{i=1}^{n}\widehat{u}_i^2-(\widehat{u}_1^2+\widehat{u}_n^2)}{\sum_{i=1}^{n}\widehat{u}_i^2}-2\frac{\sum_{i=2}^{n}\widehat{u_i}\widehat{u}_{i-1}}{\sum_{i=1}^{n}\widehat{u}_i^2}\approx 2(1-\widehat{\rho}),\end{aligned}$$

以下の 2 つの近似が用いられる。

$$\begin{aligned}&\frac{\widehat{u}_1^2+\widehat{u}_n^2}{\sum_{i=1}^{n}\widehat{u}_i^2}\approx 0,\\&\frac{\sum_{i=2}^{n}\widehat{u_i}\widehat{u}_{i-1}}{\sum_{i=1}^{n}\widehat{u}_i^2}=\frac{\sum_{i=2}^{n}\widehat{u_i}\widehat{u}_{i-1}}{\sum_{i=2}^{n}\widehat{u}_{i-1}^2+\widehat{u}_n^2}\approx\frac{\sum_{i=2}^{n}\widehat{u_i}\widehat{u}_{i-1}}{\sum_{i=2}^{n}\widehat{u}_{i-1}^2}=\widehat{\rho},\end{aligned}$$

すなわち，$\widehat{\rho}$ は $\widehat{u_i}$ と $\widehat{u}_{i-1}$ の回帰係数である。$u_i=\rho u_{i-1}+\epsilon_i$ において，u_i, u_{i-1} の代わりに $\widehat{u_i},\widehat{u}_{i-1}$ に置き換えて，ρ の推定値 $\widehat{\rho}$ を求める。

1. DW の値が 2 前後のとき，系列相関なし ($\widehat{\rho} = 0$ のとき，$DW \approx 2$)。

2. DW が 2 より十分に小さいとき，正の系列相関と判定される。

3. DW が 2 より十分に大きいとき，負の系列相関と判定される。

正確な判定には，データ数 n とパラメータ数 k に依存する。表 1 を参照せよ。

k' は定数項を除くパラメータ数を表すものとする。

See `http://www.stanford.edu/~clint/bench/dwcrit.htm` for the DW table.

Table 1: ダービン・ワトソン統計量の 5 % 点の上限と下限

	$k'=1$		$k'=2$		$k'=3$		$k'=4$		$k'=5$		$k'=6$		$k'=7$		$k'=8$		$k'=9$		$k'=10$		k'
n	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl	du	dl
6	0.610	1.400	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
7	0.700	1.356	0.467	1.896	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
8	0.763	1.332	0.559	1.777	0.367	2.287	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588	—	—	—	—	—	—	—	—	—	—	—	—	—
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822	—	—	—	—	—	—	—	—	—	—	—
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.315	2.645	0.203	3.004	—	—	—	—	—	—	—	—	—
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.380	2.506	0.268	2.832	0.171	3.149	—	—	—	—	—	—	—
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.444	2.390	0.328	2.692	0.230	2.985	0.147	3.266	—	—	—	—	—
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360	—	—	—
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.471	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438	—
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304	0.09
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184	0.13
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.668	0.321	2.873	0.244	3.073	0.17
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974	0.22
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.691	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885	0.26
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.731	2.124	0.637	2.290	0.546	2.461	0.461	2.633	0.380	2.806	0.30
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.735	0.34
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670	0.39
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.750	2.174	0.666	2.318	0.584	2.464	0.506	2.613	0.43
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.013	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560	0.47
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513	0.50
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470	0.54
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.959	0.874	2.071	0.798	2.188	0.723	2.309	0.649	2.431	0.57

$$DW = \frac{\sum_{i=2}^{n}(\widehat{u_i} - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2} \approx 2(1-\widehat{\rho}) \longrightarrow 2(1-\rho)$$

$-1 < \rho < 1$ なので (証明略)，近似的に $0 \le DW \le 4$ となる。

- $0 \le DW \le dl \qquad \longrightarrow\ u_i$ に正の系列相関
- $dl \le DW \le du \qquad \longrightarrow\ u_i$ に正の系列相関と判定できない
- $du \le DW \le 4 - du \qquad \longrightarrow\ u_i$ に系列相関なし
- $4 - du \le DW \le 4 - dl \qquad \longrightarrow\ u_i$ に負の系列相関と判定できない
- $4 - dl \le DW \le 4 \qquad \longrightarrow\ u_i$ に負の系列相関

数値例：　今までと同じ数値例で，DW を計算する。

i	Y_i	X_i	X_iY_i	X_i^2	$\widehat{Y}_i$	$\widehat{u}_i$
1	6	10	60	100	6.8	−0.8
2	9	12	108	144	8.1	0.9
3	10	14	140	196	9.4	0.6
4	10	16	160	256	10.7	−0.7
合計	$\sum Y_i$	$\sum X_i$	$\sum X_iY_i$	$\sum X_i^2$	$\sum \widehat{Y}_i$	$\sum \widehat{u}_i$
	35	52	468	696	35	0
平均	$\overline{Y}$	$\overline{X}$				
	8.75	13				

$$
\begin{aligned}
DW &= \frac{\sum_{i=2}^{n}(\widehat{u}_i - \widehat{u}_{i-1})^2}{\sum_{i=1}^{n}\widehat{u}_i^2} \\
&= \frac{(-0.8-0.9)^2+(0.9-0.6)^2+(0.6-(-0.7))^2}{(-0.8)^2+0.9^2+0.6^2+(-0.7)^2} = \frac{4.67}{2.30} = 2.03
\end{aligned}
$$

推定結果の表記方法：　回帰モデル：

$$Y_i = \alpha + \beta X_i + u_i,$$

の推定の結果，$\widehat{\alpha} = 0.3, \widehat{\beta} = 0.65$, $s_{\widehat{\alpha}} = \sqrt{10.0005} = 3.163$, $s_{\widehat{\beta}} = \sqrt{0.0575} = 0.240$, $\dfrac{\widehat{\alpha}}{s_{\widehat{\alpha}}} = 0.095$, $\dfrac{\widehat{\beta}}{s_{\widehat{\beta}}} = 2.708$, $s^2 = 1.15$ (すなわち，$s = 1.07$), $R^2 = 0.786$, $\overline{R}^2 = 0.679$, $DW = 2.03$ を得た。

これらをまとめて，

$$Y_i = \underset{(0.095)}{0.3} + \underset{(2.708)}{0.65}\, X_i,$$

$$R^2 = 0.786, \quad \overline{R}^2 = 0.679, \quad s = 1.07, \quad DW = 2.03,$$

ただし，係数の推定値の下の括弧内は t 値を表すものとする。

または,

$$Y_i = \underset{(3.163)}{0.3} + \underset{(0.240)}{0.65} X_i,$$

$$R^2 = 0.786, \quad \overline{R}^2 = 0.679, \quad s = 1.07, \quad DW = 2.03,$$

ただし，係数の推定値の下の括弧内は標準誤差を表すものとする。

のように書く。$s = \sqrt{1.15} = 1.07$ に注意。

4.2　系列相関のもとで回帰式の推定

回帰式が

$$Y_i = \alpha + \beta X_i + u_i,$$

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときの推定を考える。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

u_i を消去すると，

$$(Y_i - \rho Y_{i-1}) = \alpha(1-\rho) + \beta(X_i - \rho X_{i-1}) + \epsilon_i,$$

となり，

$Y_i^* = (Y_i - \rho Y_{i-1}),\ X_i^* = (X_i - \rho X_{i-1})$

を新たな変数として，

$$Y_i^* = \alpha' + \beta X_i^* + \epsilon_i,$$

に最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とするなので，最小二乗法を適用が可能となる。ただし，$\alpha' = \alpha(1-\rho)$ の関係が成り立つことに注意。

より一般的に，回帰式が

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,$$

$$u_i = \rho u_{i-1} + \epsilon_i,$$

のときの推定を考える。ただし，$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とする。

u_i を消去すると，

$$(Y_i - \rho Y_{i-1}) = \beta_1(X_{1i} - \rho X_{1,i-1}) + \beta_2(X_{1i} - \rho X_{2,i-1}) + \cdots + \beta_k(X_{1i} - \rho X_{k,i-1}) + \epsilon_i,$$

となり，

$Y_i^* = (Y_i - \rho Y_{i-1}),\ X_{1i}^* = (X_{1i} - \rho X_{1,i-1}),\ X_{2i}^* = (X_{2i} - \rho X_{2,i-1}),\ \cdots,\ X_{ki}^* = (X_{ki} - \rho X_{k,i-1})$ を新たな変数として，

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i$$

最小二乗法を適用する。$\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ は互いに独立とするなので，最小二乗法を適用が可能となる。

$\boldsymbol{\rho}$ の求め方について (その 1):　DW は近似的に $DW \approx 2(1 - \widehat{\rho})$ と表されるので，DW から ρ の推定値 $\widehat{\rho}$ を逆算して，

$Y_i^* = (Y_i - \widehat{\rho} Y_{i-1}),\ X_{1i}^* = (X_{1i} - \widehat{\rho} X_{1,i-1}),\ X_{2i}^* = (X_{2i} - \widehat{\rho} X_{2,i-1}),\ \cdots,\ X_{ki}^* = (X_{ki} - \widehat{\rho} X_{k,i-1})$ を新たな変数として，

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i,$$

に最小二乗法を適用する。

$\boldsymbol{\rho}$ の求め方について (その 2):　収束計算によって求める。 $\longrightarrow$ コクラン・オーカット法

1. $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,\quad i = 1, 2, \cdots, n$

   を最小二乗法で推定する。 $\longrightarrow$ $\widehat{\beta}_1, \cdots, \widehat{\beta}_k, \widehat{u}_i$ を得る。

2. $\widehat{u}_i = \rho \widehat{u}_{i-1} + \epsilon_i,\quad i = 2, 3, \cdots, n$

   を最小二乗法で推定する。 $\longrightarrow$ $\widehat{\rho}$ を得る。

3. $\rho^{(m-1)} = \widehat{\rho}$ とおく。

4. $Y_i^* = (Y_i - \rho^{(m-1)} Y_{i-1}),\ X_{1i}^* = (X_{1i} - \rho^{(m-1)} X_{1,i-1}),\ X_{2i}^* = (X_{2i} - \rho^{(m-1)} X_{2,i-1}),\ \cdots,$
$X_{ki}^* = (X_{ki} - \rho^{(m-1)} X_{k,i-1})$ を計算する。

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \cdots + \beta_k X_{ki}^* + \epsilon_i, \quad i = 2, 3, \cdots, n$$

を最小二乗法で推定する。 $\longrightarrow \widehat{\beta}_1, \cdots, \widehat{\beta}_k$ を得る。

5. $\widehat{u}_i = Y_i - \widehat{\beta}_1 X_{1i} - \widehat{\beta}_2 X_{2i} - \cdots - \widehat{\beta}_k X_{ki}, \quad i = 1, 2, \cdots, n$

を計算する。

6. ステップ 2 に戻り，$m = 1, 2, \cdots$ について繰り返す。

収束先を $\beta_1, \beta_2, \cdots, \beta_k, \rho$ の推定値とする。

5　不均一分散 (不等分散)

回帰式が

$$Y_i = \alpha + \beta X_i + u_i$$

の場合を考える。X_i が外生変数，Y_i は内生変数，u_i は互いに独立な同一の分布を持つ攪乱項 (最小二乗法に必要な仮定) とする。「独立な同一の分布」の意味は「攪乱項 $u_1, u_2, \cdots, u_n$ はそれぞれ独立に平均ゼロ，分散 σ^2 の分布する」である。分散が時点に依存する場合，代表的には，分散が他の変数 (例えば，z_i) に依存する場合，すなわち，u_i の平均はゼロ，分散は $\sigma_*^2 z_i^2$ の場合は，最小二乗法の仮定に反する。そのため，単純には，$Y_i = \alpha + \beta X_i + u_i$ に最小二乗法を適用できない。以下のような修正が必要となる。

$$\frac{Y_i}{z_i} = \alpha\frac{1}{z_i} + \beta\frac{X_i}{z_i} + \frac{u_i}{z_i} = \alpha\frac{1}{z_i} + \beta\frac{X_i}{z_i} + u_i^*$$

このとき，新たな攪乱項 u_i^* は平均ゼロ，分散 σ_*^2 の分布となる (すなわち,「同

一の」分布)。

$$\mathrm{E}(u_i^*) = \mathrm{E}\left(\frac{u_i}{z_i}\right) = \left(\frac{1}{z_i}\right)\mathrm{E}(u_i) = 0$$

u_i の仮定 $\mathrm{E}(u_i) = 0$ が使われている。

$$\mathrm{V}(u_i^*) = \mathrm{V}\left(\frac{u_i}{z_i}\right) = \left(\frac{1}{z_i}\right)^2 \mathrm{V}(u_i) = \sigma_*^2$$

u_i の仮定 $\mathrm{V}(u_i) = \sigma_*^2 z_i^2$ が最後に使われている。

よって，$\dfrac{Y_i}{z_i}, \dfrac{1}{z_i}, \dfrac{X_i}{z_i}$ を新たな変数として，最小二乗法を適用することができる。

不均一分散の検定について

$$\widehat{u_i^2} = \gamma z_i + \epsilon_i$$

を推定し，γ の推定値 $\widehat{\gamma}$ の有意性の検定を行う (通常の t 検定)。

z_i は回帰式に含まれる変数でもよい。例えば，u_i の平均はゼロ，分散は $\sigma_*^2 X_i^2$ の

場合，各変数を X_i で割って，

$$\frac{Y_i}{X_i} = \alpha\frac{1}{X_i} + \beta + \frac{u_i}{X_i} = \alpha\frac{1}{X_i} + \beta + u_i^*$$

を推定すればよい。β は定数項として推定されるが，意味は限界係数 (すなわち，傾き) と同じなので注意すること。

6　推定量の求め方

6.1　最小二乗法

・n 個のデータ (実現値)： $x_1, x_2, \cdots, x_n$

・背後に対応する確率変数を仮定： $X_1, X_2, \cdots, X_n$

・$\mathrm{E}(X_i) = \mu$，$\mathrm{V}(X_i) = \sigma^2$ を仮定

母数 (μ, σ^2) を推定する。

観測データ $x_1, x_2, \cdots, x_n$ をもとにして，μ の最小二乗推定値を求める。

$$\min_{\mu} \sum_{i=1}^{n} (x_i - \mu)^2$$

μ の解を $\widehat{\mu}$ とすると，

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

となり，$\widehat{\mu} \equiv \overline{x}$ を得る。

すなわち，

$$\frac{\mathrm{d} \sum_{i=1}^{n} (x_i - \mu)^2}{\mathrm{d}\mu} = 0$$

を μ について解く。

μ の最小二乗推定量はデータ x_i を対応する確率変数 X_i で置き換えて，

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

となり，$\widehat{\mu} \equiv \overline{X}$ を得る ($\widehat{\mu}$ について，推定値と推定量は同じ記号を使っている)。

以上を回帰分析に応用すると，

$$\min_{\alpha,\beta} \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

を解くことになる。

すなわち，

$$\frac{\partial \sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2}{\partial \alpha} = 0$$
$$\frac{\partial \sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2}{\partial \beta} = 0$$

の連立方程式を α, β について解く。

6.2　最尤法

n 個の確率変数 $X_1, X_2, \cdots, X_n$ は互いに独立で，同じ確率分布 $f(x) \equiv f(x;\theta)$ とする。ただし，θ は母数で，例えば，$\theta = (\mu, \sigma^2)$ である。

$X_1, X_2, \cdots, X_n$ の結合分布は，互いに独立なので，

$$f(x_1, x_2, \cdots, x_n; \theta) \equiv \prod_{i=1}^{n} f(x_i; \theta)$$

と表される。

観測データ $x_1, x_2, \cdots, x_n$ を与えたもとで，$\prod_{i=1}^{n} f(x_i; \theta)$ は θ の関数として表される。すなわち，

$$l(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

となる。

$l(\theta)$ を尤度関数と呼ぶ。

$$\max_{\theta}\ l(\theta)$$

となる θ を最尤推定値 $\widehat{\theta} = \widehat{\theta}(x_1, x_2, \cdots, x_n)$ と呼ぶ。

データ $x_1, x_2, \cdots, x_n$ を確率変数 $X_1, X_2, \cdots, X_n$ で置き換えて，$\widehat{\theta} = \widehat{\theta}(X_1, X_2, \cdots, X_n)$ を最尤推定量と呼ぶ。

$$\max_{\theta} \ l(\theta)$$

と

$$\max_{\theta} \ \log l(\theta)$$

の θ の解はともに同じものであることに注意。$\log l(\theta)$ を対数尤度関数と呼ぶ。

最尤推定量の性質： n が大きいとき，

$$\widehat{\theta} \sim N(\theta, \sigma_\theta^2)$$

ただし，

$$\sigma_\theta^2 = \frac{1}{\sum_{i=1}^n \mathrm{E}\Big[\Big(\dfrac{\mathrm{d}\log f(X_i;\theta)}{\mathrm{d}\theta}\Big)^2\Big]}$$

$$= -\frac{1}{\sum_{i=1}^{n} \mathrm{E}\Big[\dfrac{\mathrm{d}^2 \log f(X_i;\theta)}{\mathrm{d}\theta^2}\Big]}$$

θ がベクトル $(k \times 1)$ の場合，n が大きいとき，

$$\widehat{\theta} \sim N(\theta, \Sigma_\theta)$$

ただし，

$$\begin{aligned}
\Sigma_\theta &= \Big(\sum_{i=1}^{n} \mathrm{E}\Big[\Big(\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big)\Big(\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big)'\Big]\Big)^{-1} \\
&= -\Big(\sum_{i=1}^{n} \mathrm{E}\Big[\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'}\Big]\Big)^{-1}
\end{aligned}$$

例 1： 正規母集団 $\mathrm{N}(\mu, \sigma^2)$ からの標本値 $x_1, x_2, \cdots, x_n$ を用いて，

(1) σ^2 が既知のとき，μ の最尤推定値と最尤推定量