確率・統計 B
統計量の分布

若木宏文

wakaki@math.sci.hiroshima-u.ac.jp
http://home.hiroshima-u.ac.jp/ wakaki/lecture/probstatB16/index.shtml

2017.10.27

統計量の分布 (1/8)

定義 6.6

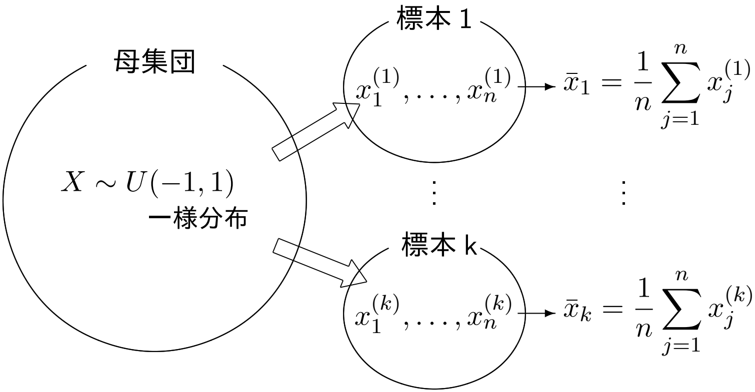
統計量

$t(x_1, \ldots, x_n)$: n 変数のボレル可測関数,
未知パラメータは含まれないものとする.
$X_1, \ldots, X_n$: ランダム標本
$\Rightarrow T = t(X_1, \ldots, X_n)$ を統計量とよぶ.

統計量の分布や, その特性値が知りたい.

統計量の分布 (2/8)

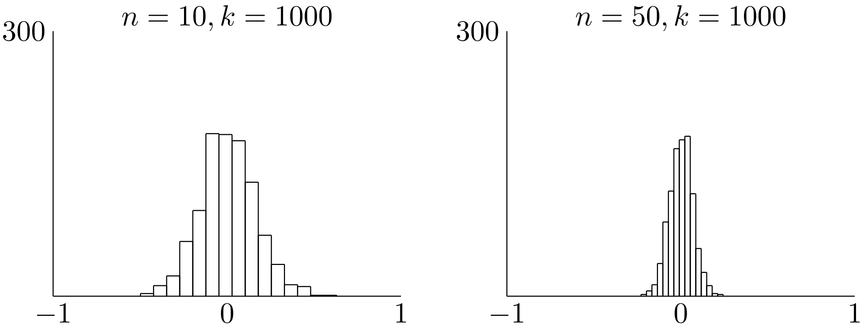
例. 統計量の分布 (数値実験)



統計量の分布 (3/8)

例. 統計量の分布 (数値実験)(つづき)

標本平均の分布

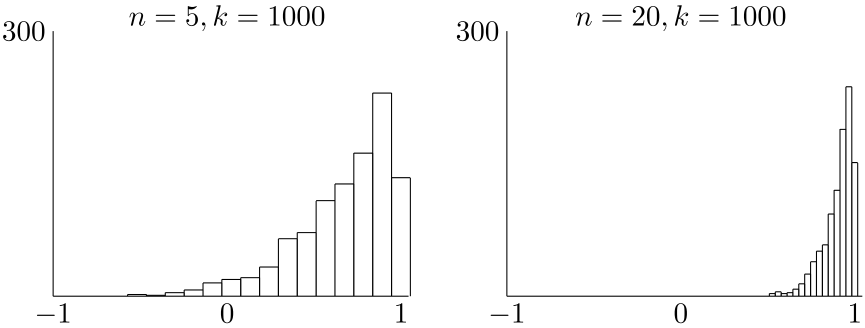


$\bar{x} = 0$ を中心にほぼ左右対称に分布している
標本数 (n) が大きくなると, 分布の「幅」が狭くなる

統計量の分布 (4/8)

例. 統計量の分布 (数値実験)(つづき)

最大値の分布 ($t = \max\{x_1, \ldots, x_n\}$)



右側に偏った分布
標本数 (n) が大きくなると, 分布の「幅」が狭くなる

統計量の分布 (5/8)

定義 6.7

標本平均, 標本分散

$X_1, \ldots, X_n$: ランダム標本

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

をそれぞれ, 標本平均, 標本分散とよぶ.

統計量の分布 (6/8)

P_n : ランダム標本 $X_1, \ldots, X_n$ の経験分布
X^* : P_n に従う確率変数

$\Rightarrow$

$$\mathsf{E}(X^*) = \sum_{i=1}^{n} X_i \mathsf{P}(X^* = X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

$$\mathsf{Var}(X^*) = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \mathsf{P}(X^* = X_i) = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

標本平均は, 経験分布の平均

標本分散は, 経験分布の分散の $\dfrac{n}{n-1}$ 倍

統計量の分布 (7/8)

定理 6.2

$X_1, \ldots, X_n$ をランダム標本とし

$$\mathrm{E}(X_i) = \mu, \quad \mathrm{Var}(X_i) = \sigma^2, \quad i = 1, \ldots, n$$

とする．このとき次が成り立つ．

(1) $\mathrm{E}(\bar{X}_n) = \mu$.

(2) $\mathrm{Var}(\bar{X}_n) = \frac{1}{n}\sigma^2$.

(3) $\mathrm{E}(S_n^2) = \sigma^2$.

(証明 は (3) のみ板書で)

メモ用紙

メモ用紙

統計量の分布 (8/8)

定理 6.3

定理 6.2 と同じ仮定のもとで次が成り立つ.

(1) $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$.

(2) $S_n^2 \xrightarrow{p} \sigma^2$.

(3) $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1)$.

(証明は板書で)

メモ用紙

メモ用紙

標本平均, 標本分散の分布 (1/15)

母集団分布が正規分布の場合
標本平均と標本分散の関数の分布を導出する.
平均や分散に関する「区間推定」や「検定」で必要となる.

定理 6.8

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ とする. このとき, 次が成り立つ.

(1) $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

(2) $(n-1)S_n^2/\sigma^2 \sim \chi^2_{n-1}$.

(3) $\bar{X}_n$ と S_n は互いに独立である.

(証明には, 定義といくつかの定理が必要)

標本平均, 標本分散の分布 (2/15)

定義 6.1

カイ 2 乗分布 確率変数 X の確率密度関数が

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\frac{n}{2})2^{n/2}} e^{-x/2} x^{n/2-1}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

で与えられるとき, X は自由度 n のカイ **2** 乗分布に従うといい, $X \sim \chi_n^2$ とかく.

標本平均, 標本分散の分布 (3/15)

$\boldsymbol{X} = (X_1, \ldots, X_n)'$ の確率密度関数が

$$f(x_1, \ldots, x_n) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}$$

と表せるとき, n 次元正規分布に従うといい,
$\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$ と書く. ここで,

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}.$$

パラメータ $\boldsymbol{\mu}$ は任意の定数, すなわち, $-\infty < \mu_i < \infty$ $(i = 1, \ldots, n)$ で, Σ は正定値行列である.

標本平均, 標本分散の分布 (4/15)

命題 1

$\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$, $Z_1, \ldots, Z_n \overset{i.i.d.}{\sim} N(0,1)$ とする.

(1) $\boldsymbol{Z} \sim N_n(\boldsymbol{0}, I_n)$
ただし, I_n は n 次単位行列.

(2) A : n 次正則行列, $\boldsymbol{b}$: n 次元ベクトル
$A\boldsymbol{Z} + \boldsymbol{b} \sim N_n(\boldsymbol{b}, AA')$

(3) $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$
$\Rightarrow \quad \mathsf{E}(\boldsymbol{X}) = \boldsymbol{\mu}, \mathsf{Var}(\boldsymbol{X}) = \Sigma,$
$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) := \mathsf{E}(e^{i\boldsymbol{t}'\boldsymbol{X}}) = \exp(i\boldsymbol{t}'\boldsymbol{\mu} - \boldsymbol{t}'\Sigma\boldsymbol{t}/2)$

(4) B : $m \times n$ 行列, $\mathrm{rank} B = m$, $\boldsymbol{b}$: m 次元ベクトル,
$\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma), \quad \boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$
$\Rightarrow \quad \boldsymbol{Y} \sim N_m(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B')$

標本平均, 標本分散の分布 (5/15)

命題 1 の証明 (1)

$N(0,1)$ の確率密度関数は $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ であるから,

$$
\begin{aligned}
f_{\boldsymbol{Z}}(z_1, \ldots, z_n) &= \prod_{j=1}^{n} \left\{ \frac{1}{\sqrt{2\pi}} e^{-z_j^2/2} \right\} \\
&= (2\pi)^{-n/2} \exp\left\{ -\frac{1}{2}(z_1^2 + \cdots + z_n^2) \right\}
\end{aligned}
$$

となり, $N_n(\mathbf{0}, I_n)$ の確率密度関数と一致する.

$(\because\ \boldsymbol{z}' I_n^{-1} \boldsymbol{z} = \sum_{j=1}^{n} z_j^2)$

標本平均, 標本分散の分布 (6/15)

定理 3.3

$\boldsymbol{X} = (X_1, \dots, X_n)$: 連続型確率ベクトル
$f(x_1, \dots, x_n)$: $\boldsymbol{X}$ の確率密度関数
$y_j(\boldsymbol{x}) = y_j(x_1, \dots, x_n)$ $((j = 1, 2, \dots, n))$:
 C^1-級関数で $\boldsymbol{y} = (y_1(\boldsymbol{x}), \dots, y_n(\boldsymbol{x}))$ と $\boldsymbol{x}$ が 1 対 1 対応
$J = \left(\frac{\partial x_j}{\partial y_i}\right)_{i,j=1,\cdot,n}$: 逆変換のヤコビ行列

$\det(J) \neq 0$ ならば, $\boldsymbol{Y} = (y_1(\boldsymbol{X}), \dots, y_n(\boldsymbol{X}))$ の確率密度関数は

$$f_{\boldsymbol{Y}}(y_1, \dots, y_n) = f(x_1(\boldsymbol{y}), \dots, x_n(\boldsymbol{y}))|\det(J)|$$

標本平均, 標本分散の分布 (7/15)

命題 1 の証明 (2)

定理 3.3 を用いると $\boldsymbol{X} = A\boldsymbol{Z} + \boldsymbol{b}$ の確率密度関数は

$$
\begin{aligned}
&(2\pi)^{-n/2} \exp\left[-\frac{1}{2}\{A^{-1}(\boldsymbol{x}-\boldsymbol{b})\}'\{A^{-1}(\boldsymbol{x}-\boldsymbol{b})\}\right] |\det(A^{-1})| \\
=&(2\pi)^{-n/2}|AA'|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{b})'(AA')^{-1}(\boldsymbol{x}-\boldsymbol{b})\right\}
\end{aligned}
$$

となり, $N_n(\boldsymbol{b}, AA')$ の確率密度関数と一致する.

標本平均, 標本分散の分布 (8/15)

命題 1 の証明 (3)

Σ は正定値行列であるから, $\Sigma = AA'$ となる正則行列 A が存在する. このとき, (2) より $A\boldsymbol{Z} + \boldsymbol{\mu} \sim N_n(\boldsymbol{\mu}, \Sigma)$ となる.

$N(0,1)$ の特性関数は, $e^{-t^2/2}$ であるから, $\boldsymbol{Z}$ の特性関数は

$$\begin{aligned} &\mathsf{E}[\exp(it_1 Z_1 + \cdots + it_n Z_n)] \\ &= \prod_{j=1}^{n} \mathsf{E}[e^{it_j Z_j}] = \exp\left(-\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\right) \ (\boldsymbol{t} = (t_1, \ldots, t_n)') \end{aligned}$$

したがって $\boldsymbol{X}$ の特性関数は

$$\begin{aligned} \mathsf{E}[\exp(i\boldsymbol{t}'\boldsymbol{X})] &= \mathsf{E}[\exp\{i\boldsymbol{t}'(A\boldsymbol{Z} + \boldsymbol{\mu})\}] = e^{i\boldsymbol{t}'\boldsymbol{\mu}}\mathsf{E}[\exp\{i(A'\boldsymbol{t})'\boldsymbol{Z}\}] \\ &= e^{i\boldsymbol{t}'\boldsymbol{\mu}} \exp\left\{-\frac{1}{2}\boldsymbol{t}'(AA')\boldsymbol{t}\right\} = \exp(i\boldsymbol{t}'\boldsymbol{\mu} - \boldsymbol{t}'\Sigma\boldsymbol{t}/2) \end{aligned}$$

標本平均, 標本分散の分布 (9/15)

命題 1 の証明 (4)

$\boldsymbol{Y}$ の特性関数は (3) を用いると

$$\begin{aligned}\mathsf{E}[\exp\{i\boldsymbol{t}'(B\boldsymbol{X}+\boldsymbol{b})\}] &= e^{i\boldsymbol{t}'\boldsymbol{b}}\mathsf{E}[\exp\{i(B'\boldsymbol{t})'\boldsymbol{X}\}] \\ &= e^{i\boldsymbol{t}'\boldsymbol{b}}\exp(i(B'\boldsymbol{t})'\boldsymbol{\mu}-(B'\boldsymbol{t})'\Sigma(B'\boldsymbol{t})/2) \\ &= \exp(i\boldsymbol{t}'(B\boldsymbol{\mu}+\boldsymbol{b})-\boldsymbol{t}'B\Sigma B'\boldsymbol{t}/2)\end{aligned}$$

これは, $N_m(B\boldsymbol{\mu}+\boldsymbol{b}, B\Sigma B')$ の特性関数と一致する.

標本平均, 標本分散の分布 (10/15)

定理 6.7

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(0,1)$ とし, $V = X_1^2 + \cdots + X_n^2$ とする. このとき, $V \sim \chi_n^2$ である.

(証明)

X と Y を独立で, $X \sim \chi_m^2, Y \sim \chi_n^2$ とする. このとき

$Z = X + Y$ の分布関数は $z > 0$ に対して

$\mathcal{D} = \{(x,y) \,|\, x \geq 0, y \geq 0, x + y \leq z\}$ とおくと

$$\begin{aligned}
F_Z(z) &= \mathsf{P}(X + Y \leq z) \\
&= \iint_{\mathcal{D}} \frac{1}{\Gamma(m/2)2^m} e^{-x/2} x^{m/2-1} \frac{1}{\Gamma(n/2)2^n} e^{-y/2} y^{n/2-1} dxdy, \\
&= \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{m+n}} \int_0^z \left\{ \int_0^{z-y} e^{-(x+y)/2} x^{m/2-1} y^{n/2-1} dx \right\} dy
\end{aligned}$$

標本平均, 標本分散の分布 (11/15)

定理 6.7 の証明（続き）

Z の確率密度関数は

$$\frac{d}{dz}F_Z(z) = \frac{1}{\Gamma(m/2)\Gamma(n/2)2^{m+n}}\left\{\int_0^{z-z} e^{-(x+z)/2}x^{m/2-1}z^{n/2-1}dx\right\}$$

$$+\frac{1}{\Gamma(m/2)\Gamma(n/2)2^{m+n}}\int_0^z\left\{e^{-((z-y)+y)/2}(z-y)^{m/2-1}y^{n/2-1}\right\}dy$$

$$=\frac{e^{-z/2}z^{(m+n)/2-2}}{\Gamma(m/2)\Gamma(n/2)2^{m+n}}\int_0^z\left(1-\frac{y}{z}\right)^{m/2-1}\left(\frac{y}{z}\right)^{n/2-1}dy$$

$$\left(\frac{y}{z}=u \text{ と変数変換}, dy\to zdu\right)$$

$$=\frac{e^{-z/2}z^{(m+n)/2-1}}{\Gamma(m/2)\Gamma(n/2)2^{m+n}}\int_0^1(1-u)^{m/2-1}u^{n/2-1}du$$

$$=C\ e^{z/2}z^{(m+n)/2-1}\ (C \text{ は定数})$$

標本平均, 標本分散の分布 (12/15)

定理 6.7 の証明（続き）

$X_1 \sim N(0,1)$ のとき, $V = X_1^2$ の分布関数は $v > 0$ に対して

$$F_V(v) = \mathsf{P}(X_1^2 \leq v) = \mathsf{P}(-\sqrt{v} \leq X_1 \leq \sqrt{v})$$
$$= \frac{1}{2\pi} \int_{-\sqrt{v}}^{\sqrt{v}} e^{-x^2/2} dx = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\sqrt{v}} e^{-x^2/2} dx$$

確率密度関数は $\frac{d}{dv} F_V(v) = \frac{\sqrt{2}}{\sqrt{\pi}} \frac{1}{2\sqrt{v}} e^{-v/2}$ となり,
χ_1^2 の確率密度関数と一致する.
n に関する帰納法より, $(X_1^2 + \cdots + X_{n-1}^2) + X_n^2 \sim \chi_n^2$ である.

標本平均, 標本分散の分布 (13/15)

定理 6.8 (再掲)

$X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$ とする. このとき, 次が成り立つ.

(1) $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

(2) $(n-1)S_n^2/\sigma^2 \sim \chi^2_{n-1}$.

(3) $\bar{X}_n$ と S_n は互いに独立である.

標本平均, 標本分散の分布 (14/15)

定理 6.8 の証明

$\boldsymbol{X} = (X_1, \ldots, X_n)' \Rightarrow \boldsymbol{X} \sim N_n(\mu \mathbf{1}_n, \sigma^2 I_n) \quad (\mathbf{1}_n = (1, \ldots, 1)')$

H：第 1 行が $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$ である n 次の直交行列

$\boldsymbol{Y} = (Y_1, \ldots, Y_n)' = H\boldsymbol{X}$ と定義すると命題 1(4) より

$$\boldsymbol{Y} \sim N_n(\mu H \mathbf{1}_n, \sigma^2 I_n)$$

となる.

$$\mathrm{E}(\boldsymbol{Y}) = \mu \begin{pmatrix} 1/\sqrt{n} & \cdots & 1/\sqrt{n} \\ * & \cdots & * \\ \vdots & & \vdots \\ * & \cdots & * \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sqrt{n}\mu \\ 0 \\ \vdots \\ 0 \end{pmatrix} =: \boldsymbol{\eta}$$

標本平均, 標本分散の分布 (15/15)

定理 6.8 の証明（続き）

$Y_1, \ldots, Y_N$ の確率密度関数は

$$(2\pi)^{-n/2}|\sigma^2 I_n|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\eta})'\left(\sigma^2 I_n\right)^{-1}(\boldsymbol{y}-\eta)\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-(y_1-\sqrt{n}\mu)^2/(2\sigma^2)} \prod_{j=2}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-y_j^2/(2\sigma^2)}$$

したがって

$Y_1, \ldots, Y_n$ は独立, $Y_1 \sim N(\sqrt{n}\mu, \sigma^2)$, $Y_2, \ldots, Y_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$

$\bar{X}_n = \frac{1}{\sqrt{n}} Y_1 \sim N(\mu, \frac{\sigma^2}{n})$, $\frac{Y_j}{\sigma} \sim N(0,1)$.

$$(n-1)S_n^2 = \sum_{j=1}^{n}(X_j - \bar{X}_n)^2 = \sum_{j=1}^{n} X_j^2 - n\bar{X}_n^2 = \sum_{j=1}^{n} Y_j^2 - Y_1^2$$

したがって $\frac{(n-1)S_n^2}{\sigma^2} = \sum_{j=2}^{n} \frac{Y_j^2}{\sigma^2} \sim \chi^2_{n-1}$ で $\bar{X}_n$ と独立.

他の統計量の分布 (1/4)

定義 7.10

(t-分布) 確率変数 T の確率密度関数が

$$f_T(t) = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{1}{2}n\right)}\left(1+\frac{1}{n}t^2\right)^{-(n+1)/2}$$

で与えられるとき，X は自由度 n の t-分布に従うといい，$T \sim t_n$ とかく．

定理 7.9

$X \sim N(0,1)$, $Y \sim \chi^2_n$, X と Y は独立とする．このとき

$$T = \frac{X}{\sqrt{Y/n}} \sim t_n$$

(証明は省略，テキスト 125 ページ参照)

他の統計量の分布 (2/4)

注 7.5

正規母集団 $N(\mu, \sigma^2)$ からの大きさ n の標本 $X_1, \ldots, X_n$ に基づく標本平均, 標本分散を $\bar{X}_n$, S_n^2 とする. このとき

$$T = t(X_1, \ldots, X_n) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}^2$$

他の統計量の分布 (3/4)

定義 7.11

F-分布確率変数 V の確率密度関数が

$$f_V(v) = \begin{cases} \dfrac{\Gamma\left(\frac{1}{2}(m+n)\right)}{\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)} \left(\dfrac{m}{n}\right)^{m/2} v^{m/2-1}\left(1+\dfrac{m}{n}v\right)^{-(m+n)/2}, & v > 0 \\ 0, & v \leq 0 \end{cases}$$

で与えられるとき, V は自由度 m, n の F-分布に従うといい,
$V \sim F_{m,n}$ とかく.

他の統計量の分布 (4/4)

定理 7.10

$X \sim \chi_m^2$, $Y \sim \chi_n^2$, X と Y は独立とする. このとき

$$V = \frac{X/m}{Y/n}$$

は自由度 m, n の F-分布に従う.

(証明は省略, テキスト 127 ページ参照)