回帰係数の推定について（I）

竹内　　清

1

回帰係数の推定に当り，2変量正規分布から無作為抽出した標本の回帰係数の分布，ならびにその性質を第1節で検討する。ついで第2節で観測誤差などの誤差が観測値に含まれる簡単な場合について，それがどのように変化するかを検討することにする。今回取扱ったのは2変量正規分布の場合であるが，多変量の場合，さらにここでの仮定が変化した場合，結果がどのように変化するかは，次回以降で検討することにする。

さて2変量 x および y は，次式で与えられるような2変量正規分布をするものと仮定する。

$$f(x,\ y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\ \sigma_y}+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}\right] \quad (1)$$

ただし，μ_x，μ_y，σ_x^2，σ_y^2，ρ は，それぞれ対応する母平均，母分散，母相関係数を表わす。すなわち

$$\mathrm{E}(x)=\mu_x,\quad \mathrm{E}(y)=\mu_y$$

$$\mathrm{E}(x-\mu_x)^2=\sigma_x^2,\quad \mathrm{E}(y-\mu_y)^2=\sigma_y^2$$

$$\frac{\mathrm{E}(x-\mu_x)(y-\mu_y)}{\sigma_x\quad \sigma_y}=\frac{\sigma_{xy}}{\sigma_x\sigma_y}=\sigma_{xy}$$

σ_{xy} は x と y の間の母共分散を表わす。無作為抽出により与えられたデータ $(x_i,\ y_i)$ $(i=1,\ 2,\cdots,\ \mathrm{n})$ に対して，y の x に対する線形回帰方程式は次式で与えられる。

$$\bar{y}_x - \bar{y} = \mathrm{r}\frac{s_y}{s_x}(x - \bar{x}) \tag{2}$$

ただし，$\bar{y}_x$ は x が与えられた場合の y の平均値，r は x と y の間の標本相関係数，$\bar{x}$ および $\bar{y}$ はそれぞれ x と y の標本平均，s_x および s_y はそれぞれ x と y の標準偏差を表わす。

上の(2)式はつぎのように書きかえられる。

$$\bar{y}_x = b_0 + b_1 x \tag{3}$$

したがって

$$\begin{aligned} b_0 &= \bar{y} - b_1\bar{x} \\ b_1 &= \mathrm{r}\frac{s_y}{s_x} = \frac{s_{xy}}{s_x^2} \end{aligned} \tag{4}$$

ここで b_1 の分布を考える。この分布は最初 В. И. Романовский によって与えられたので[(1)]，彼の展開を補足しながらみることにしよう[(2)]。

b_1 の確率密度を $f(b_1)$ とすると，これは次式で与えられる。

$$f(b_1) = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{\sqrt{\pi}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \frac{\sigma_x}{\sigma_y}\left[1 - \mathrm{r}^2 + \left(\mathrm{r} - \frac{\sigma_x}{\sigma_y}b_1\right)^2\right]^{-\frac{n}{2}} \tag{5}$$

ここで

$$\begin{aligned} \mathrm{A} &= \frac{n}{2\sigma_x^2(1-\rho^2)} \\ \mathrm{B} &= \frac{n}{2\sigma_y^2(1-\rho^2)} \\ \mathrm{C} &= \frac{n\rho}{2\sigma_x\sigma_y(1-\rho^2)} \end{aligned} \tag{6}$$

とおくと

---

(1) V. Romanovsky, "On the distribution of the regression coefficient in samples from normal population," Изв. АН СССР, 1926, вып. 6, 20, стр. 643—684.

(2) ここでは同じ著者の В. И. Романовский, Математическая Статистика, книга вторая, 1963, стр. 701—710. によっている。

$$f(b_1)=\frac{2^{n-3}\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\pi\ (n-3)\,!}\cdot\frac{(AB-C^2)^{\frac{n-1}{2}}}{B^{\frac{n-2}{2}}}\cdot(A-2Cb_1+Bb_1^2)^{-\frac{n}{2}} \tag{7}$$

また

$$2^{n-3}\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)=\sqrt{\pi}\,\Gamma(n-2) \tag{8}$$

なる関係を利用すると

$$f(b_1)=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)}\cdot\frac{(AB-C^2)^{\frac{n-1}{2}}}{B^{\frac{n-2}{2}}}\cdot(A-2Cb_1+Bb_1^2)^{-\frac{n}{2}} \tag{9}$$

上の(9)式の証明は，Романовский によって与えられているが，それを補足しながら説明するとつぎのようになる。

$\xi_{11}=s_x^2$，$\xi_{12}=s_{xy}$，$\xi_{22}=s_y^2$

とおくと，ξ_{11}，ξ_{12}，ξ_{22} の同時度数関数は次式で与えられる。

$$f(\xi_{11},\xi_{12},\xi_{22})'=\frac{2^{n-3}}{\pi(n-3)\,!}(AB-C^2)^{\frac{n-1}{2}}(\xi_{11}\xi_{22}-\xi_{12}^2)^{\frac{n-4}{2}}\times e^{-A\xi_{11}-B\xi_{22}+2C\xi_{12}} \tag{10}$$

ところで

$$b_1=\frac{\xi_{12}}{\xi_{11}}$$

であるから，$\xi_{12}=b_1\xi_{11}$．これから $b_1 d\xi_{11}=d\xi_{12}$。

また

$$\xi_{11}\xi_{22}-\xi_{12}^2=\xi_{11}\xi_{22}-b_1^2\xi_{11}^2=\xi_{11}(\xi_{22}-b_1^2\xi_{11})$$

これらの関係を(10)式に代入し，ξ_{11}，ξ_{22} について変域全体にわたって積分すると，b_1 の周辺度数が求められる。すなわち

$$f(b_1)=L\int_0^\infty\int_0^\infty\xi_{11}(\xi_{11}\xi_{22}-b_1^2\xi_{12})^{\frac{n-4}{2}}\times \times e^{-A\xi_{11}-B\xi_{22}+2Cb_1\xi_{12}}d\xi_{11}d\xi_{22} \tag{11}$$

ただし

$$L=\frac{2^{n-3}}{\pi(n-3)\,!}(AB-C^2)^{\frac{n-1}{2}}\ 。$$

ここで

$$t=\xi_{22}-\xi_{11}b_1^2$$

と変換すると，$dt=d\xi_{22}$ であるから

$$f(b_1)=L\int_0^\infty\int_0^\infty \xi_{11}^{\frac{n-2}{2}}t^{\frac{n-4}{2}}e^{-(A-2Cb_1+Bb_1^2)\xi_{11}-Bt}d\xi_{11}dt$$

$$=L\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n}{2}\right)B^{-\frac{n-2}{2}}(A-2Cb_1+Bb_1^2)^{-\frac{n}{2}} \tag{12}$$

これから(9)式が導かれる。

以上の展開のうち(12)式の結果はつぎのようにに導かれる。

(12)式において

$$\int_0^\infty\int_0^\infty \xi_{11}^{\frac{n-2}{2}}t^{\frac{n-4}{2}}e^{-(A-2Cb_1+Bb_1^2)\xi_{11}-Bt}d\xi_{11}dt$$

$$=\int_0^\infty t^{\frac{n-4}{2}}e^{-Bt}dt\int_0^\infty \xi_{11}^{\frac{n-2}{2}}e^{-(A-2Cb_1+Bb_1^2)\xi_{11}}d\xi_{11} \tag{13}$$

ところで χ^2 の密度関係数 $f(\chi^2)$ は次式で与えられる。

$$f(\chi^2)=\frac{(\chi^2)^{\frac{n}{2}-1}e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \tag{14}$$

nは自由度で

$$\Gamma\left(\frac{n}{2}\right)=\int_0^\infty x^{\frac{n}{2}}e^{-x}dx\ ,\qquad x>0 \tag{15}$$

したがって

$$\int_0^\infty t^{n-4}e^{-Bt}dt$$

において，$Bt=z$ とおくと $dz=Bdt$ であるから

$$\int_0^\infty \left(\frac{z}{B}\right)^{\frac{n-4}{2}} e^{-z} B^{-1} dz = \int_0^\infty B^{-\frac{n-4}{2}-1} z^{\frac{n-2}{2}-1} e^{-z} dz$$

$$= B^{-\frac{n-2}{2}} \Gamma\left(\frac{n-2}{2}\right) \tag{16}$$

また

$$\int_0^\infty \xi_{11}^{\frac{n-2}{2}} e^{-(A-2Cb_1+Bb_1^2)\xi_{11}} d\xi_{11}$$

において

$$A-2Cb_1+Bb_1{}^2=w$$

とおくと

$$\frac{1}{(A-2Cb_1+Bb_1{}^2)} dw = d\xi_{11}$$

であるから，上式は

$$\int_0^\infty w^{\frac{n}{2}-1} \frac{1}{(A-2Cb_1+Bb_1{}^2)^{-\frac{n}{2}+1}} \cdot \frac{1}{(A-2Cb_1+Bb_1{}^2)^{-1}} e^{-w} dw$$

$$=(A-2Cb_1+Bb_1{}^2)^{-\frac{n}{2}} \Gamma\left(\frac{n}{2}\right) \tag{17}$$

さて以上のようにして，b_1 の密度関係 $f(b_1)$ が求められたが，これを利用して b_1 に関する種々の性質が導かれる。

b_1 の期待値 $E(b_1)$ と分散 $\sigma_{b_1}^2$ はつぎのようにして求められるであろう。直接積分を行なうとすれば

$$E(b_1)=\int_\Omega b_1 f(b_1) db_1 \tag{18}$$

Ω は b_1 の変域を表わす。この変域はつぎのように定められるであろう。

$$b_1 = r\frac{s_y}{s_x}$$

であるから，

$$-1 \leq r \leq 1$$

$s_x \geq 0$

$s_y \geq 0$

より，b_1 の変域は $-s_y/s_x$ から s_y/s_x となるであろう。

またb₁の分散は

$$\sigma^2_{b_1} = \int_\Omega b_1{}^2 f(b_1) \mathrm{d} b_1 - \mathrm{E}^2(b_1) \tag{19}$$

から求められるであろう。

また積率母関数を用いてもそれぞれの結果は求められる。この場合は

$$\mathrm{M}(\theta) = \int_{-\infty}^{\infty} \mathrm{e}^{\theta b_1} f(b_1) \mathrm{d} b_1 \tag{20}$$

とおいて右辺の積分を適当に行なった上で，θ に関して適当に i 回微分して $\theta=0$ とおけば所要の結果が求められるであろう。

Романовский は，$\xi_{11}=s_x{}^2$，$\xi_{22}=s_y{}^2$，$\xi_{12}=s_{xy}$ として積率母関数を用いて b_1 の性質を導いている。

$$\mathrm{M}(\theta) = \int_\Omega \mathrm{e}^{\alpha s_x{}^2+\beta s_y{}^2+\gamma s_{xy}} f(b_1) \mathrm{d} b_1$$

$$= (\mathrm{AB}-\mathrm{C}_2)^{\frac{\mathrm{n}-1}{2}} [(\mathrm{A}-\alpha)(\mathrm{B}-\beta)-(\mathrm{C}+\tfrac{1}{2}+\gamma)^2]^{-\frac{\mathrm{n}-1}{2}} \tag{21}$$

この式から，$s_x{}^2$，$s_y{}^2$，s_{xy} の積率が求められる。結局これからつぎの結果が導かれる。b_1 の積率として，偶数次の場合，ν を正の整数として

$$\mathrm{B}_{1,2\nu} = \frac{(2\nu)\,!\,(1-\rho^2)^\nu}{2^\nu}\left(\frac{\sigma_y}{\sigma_x}\right)^{2\nu} \times$$

$$\times \sum_{\rho=0}^{\nu} \frac{2^\rho(\rho/\sqrt{1-\rho^2})^{2\rho}}{(2\rho)\,!\,(\nu-\rho)\,!\,(\mathrm{n}-3)(\mathrm{n}-5)\cdots(\mathrm{n}-2\nu+2\rho-1)} \tag{22}$$

b_1 の奇数次の積率の場合には

$$\mathrm{B}_{1,2\nu+1} = \frac{(2\nu+1)\,!\,(1-\rho^2)^{\frac{2\nu+1}{2}}}{2^\nu}\left(\frac{\sigma_y}{\sigma_x}\right)^{2\nu+1} \times$$

$$\times \Sigma \frac{2^\rho\,(\rho^2/(1-\rho^2))^{2\rho}}{(2\rho+1)\,!\,(\mathrm{k}\rho)\,!\,(\mathrm{n}-3)(\mathrm{n}-5)\cdots(\mathrm{n}-2\nu+2\rho-1)} \tag{23}$$

これから b_1 の期待値と分散および標準偏差は，結局つぎのように求められる。

$$E(b_1)=B_{1,1}=\rho\frac{\sigma_y}{\sigma_x}=\beta_1 \tag{24}$$

$$\sigma^2_{b_1}=E(b_1^2)-E^2(b_1)=\frac{\sigma_y^2}{\sigma_x^2}\frac{(1-\rho^2)}{(n-3)} \tag{25}$$

$$\sigma_{b_1}=\frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-\rho^2}{n-3}} \tag{26}$$

以上の結果から，$n\to\infty$ のとき

$$t=\frac{b_1-\beta_1}{\sigma_{b_1}}$$

は N (0,1) に近づくことが分るであろう。また $n\to\infty$ のとき

$$\frac{s_y}{s_x}\sqrt{\frac{1-r^2}{n}}$$

は

$$\frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-\rho^2}{n}}$$

に近づく。したがって，$n\to\infty$ のとき

$$t=\frac{b_1-\beta^1}{\frac{s^y}{s_x}\sqrt{\frac{1-r^2}{n}}}$$

は N (0,1) に近づくことが分かるであろう。

以上の展開では 2 変量正規分布をする母集団から無作為に n 組の (x_i, y_i) ($i=1,2,\cdots,n$) を抽出したとき，母回帰係数 β_1 の推定値として標本回帰係数 b_1 を使った場合の問題を考察したわけである。次節では，これまでの仮定を変えた場合，すなわち，観測誤差を導入した場合，その結果はどのように変化するかを検討するものである。

2

前節の展開において x と y とに観測誤差があるときどうなるかを検討するのがここでの問題である。つぎのモデルを仮定する。

$$X=x+u \tag{27}$$

$$Y=y+v \tag{28}$$

ただし，X,Y は観測値で，x は X の真の値で相互に独立に $N(\mu_x, \sigma_x)$，u は観測誤差で相互に独立に $N(0, \sigma_u)$，x と u とは独立，また y は Y の真の値で相互に独立に $N(\mu_y, \sigma_y)$，v は観測誤差で相互に独立に $N(0, \sigma_v)$，y と v は独立。したがって X および Y も確率変数でそれぞれ $N(\mu_x, \sqrt{\sigma_x^2+\sigma_u^2})$，$N(\mu_y, \sqrt{\sigma_y^2+\sigma_v^2})$ となる。[(3)]

ここでは X について考えてみよう。Y についても同様の類推ができる。上のモデルで x と u とは独立であるが，x と X は独立でないことに注意。すなわち

$$
\begin{aligned}
&E\{X-E(X)\}\{x-E(x)\}\\
&=E\{x+u-E(x+u)\}\{x-\mu_x\}\\
&=E(x-\mu_x+u)(x-\mu_x)\\
&=E\{(x-\mu_x)^2+u(x-\mu_x)\}\\
&=E(x-\mu_x)^2+E(xu)-\mu_x E(u)
\end{aligned}
$$

ところで，$E(xu)=0$，$E(u)=0$ であるから，上式は結局 x の分散 σ_x^2 となり0ではない。

X と x の同時密度関数 $f(X, x)$ は次式で与えられる。最初にたてた仮定から

$$
f(X,\ x)=\frac{1}{2\pi\sigma_X\sigma_x\sqrt{1-\rho^2}}\exp\Bigg[-\frac{1}{2(1-\rho^2)}\times \\
\times\left\{\left(\frac{X-\mu_X}{\sigma_X}\right)^2-\frac{2\rho(X-\mu_X)(X-\mu_x)}{\sigma_X\,\sigma_x}+\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\}\Bigg] \tag{29}
$$

(3) 平均 μ_x で標準偏差 σ_x の正規分布を $N(\mu_x, \sigma_x)$ で表わす。他も同様にして示される。

ただし ρ は x と X の間の母相関係数で $\rho=\frac{\sigma_x\mathrm{X}}{\sigma_x\sigma_\mathrm{X}}$ 。

X の周辺度数 $f_1(\mathrm{X})$ は上式から

$$f_1(\mathrm{X})=\int_{-\infty}^{\infty}f(\mathrm{X},\ x)\,\mathrm{d}x$$

で与えられる。これは結局次式となる

$$f_1(\mathrm{X})=\frac{1}{\sqrt{2\pi}\,\sigma_\mathrm{X}}\exp\left\{-\tfrac{1}{2}\left(\frac{\mathrm{X}-\mu_\mathrm{X}}{\sigma_\mathrm{X}}\right)^2\right\} \qquad (30)$$

つぎに X が与えられた場合の x の条件付度数関数 $f(x\mid\mathrm{X})$ は次式で与えられる。

$$f(x\mid\mathrm{X})=\frac{f(x,\ \mathrm{X})}{f_1(\mathrm{X})}$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_x\sqrt{1-\rho^2}}\exp\left[-\tfrac{1}{2}\left\{\frac{x_\mathrm{X}-\mu_x-\rho\frac{\sigma_x}{\sigma_\mathrm{X}}(\mathrm{X}-\mu_\mathrm{X})}{\sigma_x\sqrt{1-\rho^2}}\right]^2 \qquad (31)$$

そこで，X が与えられた場合の x の期待値 $\mathrm{E}(x\mid\mathrm{X})$ は次式で与えられる。

$$\mathrm{E}(x\mid\mathrm{X})=\int_{-\infty}^{\infty}xf(x\mid\mathrm{X})\mathrm{d}x$$

$$\int_{-\infty}^{\infty}x\frac{1}{\sqrt{2\pi}\,\sigma_x\sqrt{1-\rho^2}}\exp\left[-\tfrac{1}{2}\left\{\frac{x_\mathrm{X}-\mu_x-\rho\frac{\sigma_x}{\sigma_\mathrm{X}}(\mathrm{X}-\mu_x)}{\sigma_x\sqrt{1-\rho^2}}\right]^2\mathrm{d}x$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_x\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}x\ \exp\left[-\tfrac{1}{2}\left\{\frac{x_\mathrm{X}-\mu_x-\rho\frac{\sigma_x}{\sigma_\mathrm{X}}(\mathrm{X}-\mu_x)}{\sigma_x\sqrt{1-\rho_2}}\right\}\right]^2\mathrm{d}x \qquad (32)$$

上式において

$$\frac{x_\mathrm{X}-\mu_x-\rho\frac{\sigma_x}{\sigma_\mathrm{X}}(\mathrm{X}-\mu_x)}{\sigma_x\sqrt{1-\rho^2}}=t$$

とおくと

$$\mathrm{d}x=\sigma_x\sqrt{1-\rho^2}\ \mathrm{d}t$$

したがって

$$\frac{1}{\sqrt{2\pi}\,\sigma_x\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\left\{\sigma_x\sqrt{1-\rho^2}\,t+\mu_x+\rho\frac{\sigma_x}{\sigma_X}(X-\mu_x)\right\}\sigma_x\sqrt{1-\rho^2}\,e^{-\frac{t^2}{2}}dt$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left\{\sigma_x\sqrt{1-\rho^2}\,t+\mu_x+\rho\frac{\sigma_x}{\sigma_X}(X-\mu_x)\right\}e^{-\frac{t^2}{2}}dt$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sigma_x\sqrt{1-\rho^2}\,te^{-\frac{t^2}{2}}dt+\frac{\left\{\mu_x+\rho\frac{\sigma_x}{\sigma_X}(X-\mu_x)\right\}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt$$

上式右辺第1項は0となり，第2項は，結局 $\mu_x+\rho\frac{\sigma_x}{\sigma_X}(X-\mu_x)$ となる。したがって

$$E(x\mid X)=\mu_x+\rho\frac{\sigma_x}{\sigma_X}(X-\mu_x) \tag{33}$$

上式に ρ, σ_x, σ_X の値をそれぞれ代入することにより，結局次式がえられる。

$$E(x\mid X)=\frac{\sigma_u{}^2\mu_x+\sigma_x{}^2X}{\sigma_x{}^2+\sigma_u{}^2} \tag{34}$$

上式から分るように，観測値 X が与えられたとき，真の値 x の平均値は，μ_x と X の加重平均値である。すなわち

$$\frac{\sigma_u{}^2}{\sigma_x{}^2+\sigma_{u2}}\mu_x+\frac{\sigma_x{}^2}{\sigma_x{}^2+\sigma_u{}^2}X \tag{35}$$

から，X の分散中に占める x と u のそれぞれの分散の相対的割合が分かり，かつ x の平均 μ_x が分かれば観測値 X が与えられた場合の x の平均値が分かることになる。したがって最初に仮定したモデルが成り立つ場合，すなわち観測誤差がある場合，観測値 X そのものだけを使って分析するのは危険である。

Y についても同様にして

$$E(y\mid Y)=\frac{\sigma_v{}^2\mu_y+\sigma_y{}^2Y}{\sigma_y{}^2+\sigma_v{}^2} \tag{36}$$

前節の議論では，$u=v=0$ と考えたが，本節のモデルが成立する場合，それでは回帰係数の推定値はどのような性質をもつかをここで検討することにする。前節と同様に

$$\overline{Y}_X = b_0' + b_1' X \tag{37}$$

$$b_0' = \overline{Y} - b_1' \overline{X} \tag{38}$$

$$b_1' = r\frac{s_Y}{s_X} = \frac{s_{XY}}{s_X^2} \tag{39}$$

ここで b_1' の分布を考える。

前節の結果を利用すると

$$f(b_1') = \frac{(1-\rho^2)^{\frac{n-1}{2}}}{\sqrt{\pi}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \frac{\sigma_X}{\sigma_Y}\left[1 - r^2 + (r - \frac{\sigma_X}{\sigma_Y} b_1')^2\right]^{-\frac{n}{2}} \tag{40}$$

ところで上式において，

$$\sigma_X = \sqrt{\sigma_x^2 + \sigma_u^2}$$

$$\sigma_Y = \sqrt{\sigma_y^2 + \sigma_v^2}$$

ここで

$$A = \frac{n}{2(\sigma_x^2 + \sigma_u^2)(1-\rho^2)}$$

$$B = \frac{n}{2(\sigma_y^2 + \sigma_v^2)(1-\rho^2)}$$

$$C = \frac{n\rho}{2\sqrt{\sigma_x^2 + \sigma_u^2}\sqrt{\sigma_y^2 + \sigma_v^2}\,(1-\rho^2)}$$

とおくと

$$f(b_1') = \frac{2^{n-3}\,\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}{\pi\,(n-3)\,!} \cdot \frac{(AB - C^2)^{\frac{n-1}{2}}}{B^{\frac{n-2}{2}}} \cdot (A - 2Cb_1 + Bb_1^{2\prime})^{-\frac{n}{2}} \tag{41}$$

また　$$2^{n-3}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right) = \sqrt{\pi}\,\Gamma\,(n-2)$$

なる関係を利用すると

$$f(b_1') = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)} \cdot \frac{(AB - C^2)^{\frac{n-1}{2}}}{B^{\frac{n-2}{2}}} \cdot (A - 2Cb_1 + Bb_1^{2\prime})^{-\frac{n}{2}} \tag{42}$$

この b_1' の度数関数を用いて b_1' についての種々の性質が求められるであろう。

まず b_1' の期待値を求めよう。前節の結果を利用すると

$$\begin{aligned} E(b_1') &= \rho\sigma_Y/\sigma_X \\ &= \frac{\sigma_{XY}}{\sigma_X^2} \end{aligned} \tag{43}$$

ところで σ_{XY} は

$$\begin{aligned} \sigma_{XY} &= E\{X-E(X)\}\{Y-E(Y)\} \\ &= E(x+u-\mu_x)(y+v-\mu_y) \\ &= E(x-\mu_x)(y-\mu_y)+E(uy)+E(xv)+E(uv) \\ &= \sigma_{xy}+E(uy)+E(xv)+E(uv) \end{aligned}$$

したがって

$$E(b_1') = \frac{\sigma_{xy}+E(uy)+E(xv)+E(uv)}{\sigma_x^2+\sigma_u^2} \tag{44}$$

ところで前節において

$$E(b_1) = \rho\frac{\sigma_y}{\sigma_x} = \frac{\sigma_{xy}}{\sigma_x^2} = \beta_1$$

で b_1 は不偏推定量である。そこで $E(b_1')$ が不偏であるためには

$$\frac{\sigma_{xy}+E(uy)+E(vx)+E(uv)}{\sigma_x^2+\sigma_u^2} = \frac{\sigma_{xy}}{\sigma_x^2} \tag{45}$$

なる条件が満たされなければならない。

$$\frac{\sigma_u^2}{\sigma_x^2} = k \tag{46}$$

ただし k は定数，とおけば

$$\sigma_u^2 = k\sigma_x^2$$

したがって上式は

$$\frac{\sigma_{xy}+E(uy)+E(vx)+E(uv)}{(1+k)\sigma_x^2} = \frac{\sigma_{xy}}{\sigma_x^2}$$

$$\frac{\sigma_{xy}+E(uy)+E(vx)+E(uv)}{\sigma_x^2} = \frac{(1+k)\sigma_{xy}}{\sigma_x^2}$$

$\sigma_x^2>0$ であるから，両辺に σ_x^2 をかけ

$$\sigma_{xy}+\mathrm{E}(uy)+\mathrm{E}(vx)+\mathrm{E}(uv)=(1+k)\sigma_{xy}$$

これから

$$\mathrm{E}(uy)+\mathrm{E}(vx)+\mathrm{E}(uv)=k\sigma_{xy} \tag{47}$$

上の(47)式の条件が満たされる場合，b_1' は不偏推定量となる。

ところでこのような場合，u と y とは独立，また v と x とも独立，また u と v とも独立と仮定すれば，b_1' の期待値はつぎのようになる。

$$\mathrm{E}(b_1')=\frac{\sigma_{xy}}{\sigma_x^2+\sigma_u^2} \tag{48}$$

ところで，観測誤差がある場合，$\sigma_u^2\neq 0$ であるから

$$\mathrm{E}(b_1')=\frac{\sigma_{xy}}{\sigma_x^2+\sigma_u^2}<\mathrm{E}(b_1)=\frac{\sigma_{xy}}{\sigma_x^2} \tag{49}$$

したがって b_1' は偏りのある推定値となる。偏りの程度は

$$\frac{\sigma_u^2}{\sigma_x^2}=k$$

とすれば（ただし k は定数）

$$\sigma_u^2=k\sigma_x^2$$

であるから，これを上式に代入し

$$\mathrm{E}(b_1')=\frac{1}{(1+k)}\,\frac{\sigma_{xy}}{\sigma_x^2}=\frac{1}{1+k}\beta_1 \tag{50}$$

$k>0$ であるから

$$\frac{1}{1+k}<1$$

したがって，b_1' は小さい方へ偏りをもつことになる。したがって b_1' に（$1+k$）を乗じたものを用いると

$$\mathrm{E}\{(1+k)b_1'\}=\beta_1 \tag{51}$$

となることが分かるであろう。

つぎに b_1' の分散を求めよう。前節の展開から

$$\sigma^2{}_{b_1'}=\mathrm{E}(b_1'^2)-\mathrm{E}^2(b_1')$$

$$=\frac{\sigma_{\mathrm{Y}}{}^2}{\sigma_{\mathrm{X}}{}^2}\frac{1-\rho^2}{\mathrm{n}-3} \tag{52}$$

上式の $\sigma_{\mathrm{Y}}{}^2$, $\sigma_{\mathrm{X}}{}^2$, ρ^2 を真の値と誤差で表わすと

$$\frac{\sigma_{\mathrm{Y}}{}^2}{\sigma_{\mathrm{X}}{}^2}\frac{(1-\rho^2)}{(\mathrm{n}-3)}=\frac{\sigma_x{}^2\sigma_y{}^2+\sigma_x{}^2\sigma_v{}^2+\sigma_y{}^2\sigma_u{}^2+\sigma_u{}^2\sigma_v{}^2-\{\sigma_{xy}+\mathrm{E}(uy)+\mathrm{E}(xv)+\mathrm{E}(uv)\}^2}{(\sigma_x{}^2+\sigma_u{}^2)^2\ (\mathrm{n}-3)} \tag{53}$$

もし $\mathrm{E}(uy)=\mathrm{E}(xv)=\mathrm{E}(uv)=0$，すなわち，それぞれの変数は独立であると仮定すれば

$$\sigma^2{}_{b_1'}=\frac{\sigma_{\mathrm{Y}}{}^2(1-\rho^2)}{\sigma_{\mathrm{X}}{}^2(\mathrm{n}-3)}=\frac{\sigma_x{}^2\sigma_y{}^2+\sigma_x{}^2\sigma_v{}^2+\sigma_y{}^2\sigma_u{}^2+\sigma_u{}^2\sigma_v{}^2-\sigma^2{}_{xy}}{(\sigma_x{}^2+\sigma_u)^2\ (\mathrm{n}-3)} \tag{54}$$

ところで

$$\sigma^2{}_{b_1'}=\frac{\sigma_y{}^2}{\sigma_x{}^2}\frac{(1-\rho^2)}{(\mathrm{n}-3)}=\frac{\sigma_x{}^2\sigma_y{}^2-\sigma_{xy}{}^2}{(\sigma_x{}^2)^2(\mathrm{n}-3)} \tag{55}$$

前と同様に

$$\frac{\sigma_u{}^2}{\sigma_x{}^2}=k\quad,\qquad\frac{\sigma_v{}^2}{\sigma_y{}^2}=l\quad,\qquad(k,\ l\ \text{は定数})$$

と仮定すれば

$$\sigma^2{}_{b_1'}=\frac{(1+k)(1+l)\sigma_x{}^2\sigma_y{}^2-\sigma_{xy}{}^2}{(1+k)^2\ \sigma_x{}^4(\mathrm{n}-3)} \tag{56}$$

したがって $k\to0$, $l\to0$ であれば，すなわち，$\sigma_u{}^2\to0$, $\sigma_v{}^2\to0$ であれば

$$\sigma^2{}_{b_1'}\fallingdotseq\sigma^2{}_{b_1}$$

となる。$\sigma^2{}_{b_1'}$ と $\sigma^2{}_{b_1}$ のどちらが大きいかは

$$\sigma^2{}_{b_1'}-\sigma^2{}_{b_1} \tag{57}$$

または

$$\sigma^2{}_{b_1'}/\sigma^2{}_{b_1} \tag{58}$$

を吟味するとよいであろう。

$$\sigma^2{}_{b_1'}-\sigma^2{}_{b_1}=\frac{(1+k)\sigma_x{}^2\sigma_y{}^2(l-k)+k\sigma_{xy}{}^2(2+k)}{(1+k)^2\sigma_x{}^4(\mathrm{n}-3)} \tag{59}$$

$\mathrm{n}>3$ なるかぎり上式の分母は正であるから，$\sigma^2{}_{b_1'}$ と $\sigma^2{}_{b_1}$ の大小の判定は

分子の符号による。分子が

$$l(1+k)\sigma_x^2\sigma_y^2+k\sigma_{xy}^2(2+k)-k(1+k)\sigma_x^2\sigma_y^2>0 \tag{60}$$

なるかぎり，$\sigma^2_{b_1'}$ の方が $\sigma^2_{b_1}$ より大となる。

さてnが大きい場合

$$\frac{s_Y}{s_X}\sqrt{\frac{1-r^2}{n}}$$

は漸近的に

$$\frac{\sigma_Y}{\sigma_X}\sqrt{\frac{1-\rho^2}{n}}$$

となるであろう。s_X, s_Y, および s_{XY} は $n\to\infty$ のとき，それぞれ漸近的に σ_x, σ_y, およびよ σ_{xy} となるであろうか。これをみるにはスルツキーの定理[(4)]を利用するとよい。

s_X^2 についてみよう。

$$s_X^2=s_x^2+s_u^2 \tag{61}$$

であるから，s_X^2 の漸近線をみるには，s_x^2 と s_u^2 の漸近線をみるとよい。

$$\lim_{n\to\infty} s_x^2=\sigma_x^2 \tag{62}$$

$$\lim_{n\to\infty} s_u^2=\sigma_u^2 \tag{63}$$

であるから

$$\lim_{n\to\infty} s_X^2=\sigma_x^2+\sigma_u^2 \tag{64}$$

同様にして

---

(4) E. E. Слуцкий, "С стохастических асимптотах и пределах (1925)", E. E. Слуцкий, Избранные Труды, Академия Наук СССР, 1960, стр. 25～98.
この書物は，経済学でスルッキー方程式として有名なロシヤの数学者 E. E. Слуцкий の確率論および数理統計学の論文集である。上掲の論文は，1925年にイタリーの統計学の雑誌 *Metron* に独文で発表されたものである。すなわち，
E. Slutsky, "Über Stochastische Asymptoten und Grenzwerte", *Metron*, V, 3, XII, 1925, pp. 3—89.
上記の論文集では，ロシヤ語以外の論文はすべて露訳されて再録してある。

$$\lim_{n\to\infty} s_Y{}^2 = \sigma_y{}^2 + \sigma_v{}^2 \tag{65}$$

$$\lim_{n\to\infty} s_{XY} = \sigma_{xy} \tag{66}$$

ただし，u と y, x と v, u と v は独立と仮定。

かくして β_1 の推定量として $b_1' = \dfrac{s_{XY}}{s_X{}^2} = r\dfrac{s_Y}{s_X}$ を用いると，

$$\lim_{n\to\infty} E(b_1') = \frac{\sigma_{xy}}{\sigma_x{}^2 + \sigma_u{}^2} \tag{67}$$

で

$$\lim_{n\to\infty} s_u{}^2 = \sigma_u{}^2 = 0 \tag{68}$$

でないかぎり，b_1' は不偏，一致推定量とはならない。観測誤差そのものは，一般に標本の大きさを大きくしたからといってちらばりはそのまま小さくはならないであろう。これは標本の大きさとは別個の性質をもった問題である。

次回にはこれらの問題を含めて検討することにしよう。