大標本論の数學的基礎に就いて(Ⅱ)

北川 敏男

〔Ⅲ〕多次元統計量に関する標準誤差

以上の1次元變量に関する理論は、容易にこれを殆んどそのまゝ多次元統計量に擴張し得る。簡單のため2次元變量に就いて述べる。

$k=(k_1, k_2)$（但し $k_1=1,2,\cdots,m_1$; $k_2=1,2,\cdots,m_2$）なる $m_1\times m_2$ 箇の箇所を考へる。1球をとって來てこれらの箇所に向つて投ずるとき、k箇所に納まる確率を p_k とする。n回の獨立試行中、k箇所に納つた箇數を Y_k とする。

$$\sum_k p_k = 1 \qquad \text{從って} \qquad \sum_k Y_k = n$$

とする前次の如く、原點に関する (p, q) 次の積率及び標本平均値に関する (p, q) 次の積率を定義する。

$$u'_{(p,q),n} = \frac{1}{n}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2} x_i^p y_j^q Y(i,j)$$

$$u'_{(p,q)} = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2} x_i^p y_j^q p(i,j)$$

$$u_{(p,q),n} = \frac{1}{n}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(x_i - u'_{(1,0)})^p(y_j - u'_{(0,1)})^q Y(i,j)$$

$$u_{(p,q)} = \sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(x_i - u'_{(1,0)})^p(y_j - u'_{(0,1)})^q p(i,j)$$

これらに関しては次の関係式が成立つ。

(1) $E\{Y_k\} = n p_k$

(2) $\sigma^2\{Y_k\} = n p_k(1-p_k)$

(3) $E\{Y_k Y_l\} = -n p_k p_l$

(4) $E\{(u'_{(p,q),n} - E\{u'_{(p,q),n}\})(u'_{(r,s),n} - E\{u'_{(r,s),n}\})\}$
$= n(u'_{(p+r,q+s)} - u'_{(p,q)}u'_{(r,s)})$

(5) $u_{(p,q),n} = \sum_{l=0}^{q}\sum_{r=0}^{p}\binom{p}{r}\binom{q}{l}(-1)^{r+l}u'^{\,r}_{(1,0)}u'^{\,l}_{(0,1)}u_{(p-r,q-l)}$

これらの関係式を土台とし、一般確定量列の考を用ひて、統計量の標準偏差を求めることが出來るのは、1次元の場合と

何等相違はない。茲では，標本相関係数の標準偏差を求める方法を示すに止めよう。

定理 5.

(1°) $\{u'(p,q),n\}$ $(n=1,2,3,\cdots\cdots)$ は一致推定量列であって，　$E\{u'(p,q),n\}=u'(p,q)$ にして且つ

$$u'(p,q),n=u'(p,q)+\frac{1}{\sqrt{n}}\xi(p,q)+o\left(\frac{1}{\sqrt{n}}\right)$$

と書けば

$$E\{\xi(p,q)\xi(r,s)\}=\frac{1}{n}\left(E\{u'(p,q),n-u'(p,q))(u'(r,s),n-u'(r,s))\right\}$$
$$=u'(p+r,q+s)-u'(p,q)u'(r,s)$$

(2°) $\{u(p,q),n\}$ $(n=1,2,3,\cdots)$ は $u(p,q)$ に対する一致推定量列であって，

$$E\{u(p,q),n\}=u(p,q)$$
$$u(p,q),n=u(p,q)-\frac{pu(p-1,q)}{\sqrt{n}}\xi(1,0)-\frac{qu(p,q-1)}{\sqrt{n}}\xi(0,1)+\frac{1}{\sqrt{n}}\xi(p,q)+o\left(\frac{1}{\sqrt{n}}\right)$$

但し，$\xi(1,0)$，$\xi(0,1)$，$\xi(p,q)$ は(1°)で導入したものに外ならぬものである。

(3°)
$$E\{(u(p,q),n-u(p,q))(u(r,s),n-u(r,s))\}$$
$$=\frac{1}{n}\Big\{pr\,u(p-1,q)u(r-1,s)E\{\xi^2(1,0)\}+E\{\xi_{p,q}\xi_{r,s}\}$$
$$+qs\,u(p,q-1)u(r,s-1)E\{\xi^2(0,1)\}$$
$$-pu(p-1,q)E\{\xi(1,0)\xi(r,s)\}-ru(r-1,s)E\{\xi(1,0)\xi(p,q)\}$$
$$-qu(p,q-1)E\{\xi(0,1)\xi(r,s)\}-su(r,s-1)E\{\xi(0,1)\xi(p,q)\}$$
$$+ps\,u(p-1,q)u(r,s-1)E\{\xi(1,0)\xi(0,1)\}$$
$$+qr\,u(p,q-1)u(r-1,s)E\{\xi(1,0)\xi(0,1)\}\Big\}+o\left(\frac{1}{n}\right)$$

証明は1次元の場合と全く平行である。(3°)式の右辺の $E\{\ \}$ はすべて(1°)に依り，その値を陽示し得ることに注意されたい。

定理 6. 母集団相関係数を ρ，大きさ n の標本に関する標本相関係数を r_n とする。即ち

$$\rho=\frac{u(1,1)}{\sqrt{u(2,0)\,u(0,2)}}$$

$$r_n = \frac{u_{(1,1),n}}{\sqrt{u_{(2,0),n}\ u_{(0,2),n}}}$$

然る時には，$u_{(1,0)} = u_{(0,1)} = 0$ のとき

$$\frac{\sigma^2_{r_n}}{\rho^2} = \frac{1}{n}\left\{ \frac{u_{(2,2)}}{u^2_{(1,1)}} + \frac{1}{4}\frac{u_{(4,0)}}{u^2_{(2,0)}} + \frac{1}{4}\frac{u_{(0,4)}}{u^2_{(0,2)}} + \frac{1}{2}\frac{u_{(2,2)}}{u_{(2,0)}u_{(0,2)}} - \frac{u_{(3,1)}}{u_{(1,1)}u_{(2,0)}} - \frac{u_{(1,3)}}{u_{(1,1)}\ u_{(0,2)}} \right\} + o\left(\frac{1}{n}\right)$$

證明：定理５に依り $\{u_{(2,0),n}\}$, $\{u_{(1,1),n}\}$, $\{u_{(0,2),n}\}$ は一致統計量列であって

$$u_{(2,0),n} = u_{(2,0)} + \frac{\zeta_{(2,0)}}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)$$
$$u_{(1,1),n} = u_{(1,1)} + \frac{\zeta_{(1,1)}}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)$$
$$u_{(0,2),n} = u_{(0,2)} + \frac{\zeta_{(0,2)}}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)$$

と書かれるから

$$\frac{1}{\sqrt{u_{(2,0),n}}} = \frac{1}{\sqrt{u_{(2,0)}}}\left(1 - \frac{\zeta_{(2,0)}}{2u_{(2,0)}\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\right)$$
$$\frac{1}{\sqrt{u_{(0,2),n}}} = \frac{1}{\sqrt{u_{(0,2)}}}\left(1 - \frac{\zeta_{(0,2)}}{2u_{(0,2)}\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\right)$$

従って

$$r_n = \frac{u_{(1,1),n}}{\sqrt{u_{(2,0),n}\ u_{(0,2),n}}}$$

に以上の式を代入すれば

$$\frac{\sqrt{u_{(2,0)}u_{(0,2)}}}{u_{(1,1)}}(r_n - \rho) = -\frac{\zeta_{(2,0)}}{2u_{(2,0)}\sqrt{n}} - \frac{\zeta_{(0,2)}}{2u_{(0,2)}\sqrt{n}} + \frac{\zeta_{(1,1)}}{u_{(1,1)}\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)$$

両辺を自乗し，平均値を求めると

$$左辺 = \frac{1}{\rho^2}E\{(r_n-\rho)^2\} = \frac{\sigma^2_{r_n}}{\rho^2}$$

$$右辺 = \frac{1}{n}\left\{\frac{E\{\zeta^2_{(1,1)}\}}{u^2(1,1)} + \frac{E\{\zeta^2_{(2,0)}\}}{4u^2(2,0)} + \frac{E\{\zeta^2_{(0,2)}\}}{4u^2(0,2)} + \frac{1}{2}\frac{E\{\zeta_{(2,0)}\zeta_{(0,2)}\}}{u(2,0)\ u(0,2)} - \frac{E\{\zeta_{(2,0)}\zeta_{(1,1)}\}}{u(2,0)\ u(1,1)} - \frac{E\{\zeta_{(1,1)}\zeta_{(0,2)}\}}{u(0,2)\ u(1,1)}\right\} + o\left(\frac{1}{n}\right)$$

式の $E\{\ \}$ は定理5に依り容易に得られる。即ち ζ の定義から

$$E\{\zeta(p,q)\zeta(r,s)\} = nE\{(u(p,q),n - u(p,q))(u(r,s),n - u(r,s))\} + o(1)$$

假定 $u(0,1) = u(1,0) = 0$ に注意して

$$\frac{\sigma^2_{r_n}}{\rho^2} = \frac{1}{n}\left\{\frac{u(2,2) - u^2(1,1)}{u^2(1,1)} + \frac{u(4,0) - u^2(2,0)}{4u^2(2,0)} + \frac{u(0,4) - u^2(0,2)}{4u^2(0,2)} + \frac{u(2,2) - u(2,0)\,u(0,2)}{2u(2,0)\ u(0,2)} - \frac{u(3,1) - u(2,0)\,u(1,1)}{u(2,0)\ u(1,1)} - \frac{u(1,3) - u(0,2)\,u(1,1)}{u(0,2)\ u(1,1)}\right\} + o\left(\frac{1}{n}\right)$$

これより直ちに求むる式を得る。

§4 変量の変換

4.1 変量の変換　統計量の変換に関する補題1(§2参照)は，いろいろな方面に於いて利用さるべきものであって，所謂変量変換法と大標本論との関係がこれによって窺ひ得るのである。茲では2，3の例に依ってその一班を窺ふことに止める。

〔I〕 Poisson の分布　Poisson 分布に関する変量変換には次の2通りの場合(1°及2°)がある。

定理 7.　(1°) $\{X_n\}$ $(n=1, 2, 3, \cdots)$ は相互に独立な統計変量であって各 X_n は同一平均値 m なる Poisson 分布に従ふものとする。然る時には

$$S_n = (X_1 + X_2 + \cdots\cdots + X_n)/n$$

と置けば、S_n 従って $\sqrt{S_n}$ は一致推定量列であって

$$E\{\sqrt{S_n}\} = \sqrt{m} + o(1),\quad \sigma^2\{\sqrt{S_n}\} = \frac{1}{4} + o\left(\frac{1}{n}\right)$$

(2) X は平均値 m なる Poisson 分布に従ふものとする。然る時には m が充分に大になるとき

$$E\{\sqrt{X}\} = \sqrt{m} + o(1),\quad \sigma^2\{\sqrt{X}\} = \frac{1}{4} + o(1)$$

証明。(1) に就いて。中心極限定理に依り任意の与へられた α, β に対して

$$Pr.\left\{m - \frac{\alpha\sigma}{\sqrt{n}} \leqq S_n \leqq m + \frac{\beta\sigma}{\sqrt{n}}\right\}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\alpha}^{\beta} e^{-t^2/2}\,dt + o(1)$$

但し、各 X_k は Poisson 分布に従ふから $\sigma^2 = m$ である。

補題 1 (§2) に依り

$$Pr\left\{\sqrt{m - \frac{\alpha\sigma}{\sqrt{n}}} \leqq \sqrt{S_n} \leqq \sqrt{m + \frac{\beta\sigma}{\sqrt{n}}}\right\}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\alpha}^{\beta} e^{-t^2/2}\,dt + o(1)$$

然るに $\sigma^2 = m$ に注意すれば

$$\sqrt{m \pm \frac{\alpha\sigma}{\sqrt{n}}} = \sqrt{m} \pm \frac{1}{2}\,\frac{\alpha\sigma}{\sqrt{m}\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)$$

$$= \sqrt{m} \pm \frac{\alpha}{2\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)$$

これに依り (1) を得る

(2) に就いて。

$$Y = \frac{X - m}{\sigma},\quad y = \frac{x - m}{\sigma}.$$

と置き

$$Pr.\{X = x\} = \frac{m^x e^{-m}}{x!} = \frac{\sigma^{2\sigma^2 + 2\sigma y}}{(\sigma^2 + \sigma y)!}e^{-\sigma^2} = h(y)$$

と置けば Stirling の公式に依り

$$h(y) = \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\left\{1 - \frac{1}{\sigma}\left(\frac{y}{2} - \frac{y^3}{6}\right) - \frac{1}{\sigma^2}\left(\frac{1}{12} - \frac{3y^2}{8} + \frac{y^4}{6} - \frac{y^6}{72}\right)\right.$$

$$\left. + \cdots\cdots\right\}$$

となる事が知られて居る。依って y/σ が余り大でない限り、正規分布

し Poisson 分布に対する良い近似となり、μ^3/σ が小さい程，その近似は良好となる。この事実に注意すれば (1°) と同様に論證し

〔II〕出現頻度（2項分布の場合）

定理 8。$\{X_n\}$ $(n=1,2,3,\cdots)$ は相互に独立な統計変量であって，各 X_n は 0 或は 1 になり、1 になる確率を p，0 になる確率を $q(=1-p)$ とする。

然る時，$S_n=(X_1+X_2+\cdots+X_n)/n$ と置けば，S_n 従って $\sin^{-1}\sqrt{S_n}$ は一致推定量列であって

$$E\{\sin^{-1}\sqrt{S_n}\}=\sin^{-1}p+o(1)$$
$$\sigma^2\{\sin^{-1}\sqrt{S_n}\}=\frac{1}{4}+o\left(\frac{1}{n}\right)$$

となる。

證明。$\sigma\{X_n\}=\sqrt{pq}$，$\sigma\{S_n\}=\sqrt{pq/n}$ に注意すれば

$$Pr.\left\{p-\alpha\sqrt{\frac{pq}{n}}\leqq S_n\leqq p+\beta\sqrt{\frac{pq}{n}}\right\}$$
$$=Pr.\left\{\sin^{-1}\sqrt{p-\alpha\sqrt{\frac{pq}{n}}}\leqq\sin^{-1}\sqrt{S_n}\leqq\sin^{-1}\sqrt{p+\beta\sqrt{\frac{pq}{n}}}\right\}$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\alpha}^{\beta}e^{-t^2/2}dt+o(1)$$

然るに $\gamma=\alpha$ 又は β として

$$\sin^{-1}\sqrt{p\pm\gamma\sqrt{\frac{pq}{n}}}=\sin^{-1}p\pm\frac{d}{dp}(\sin^{-1}\sqrt{p})\gamma\sqrt{\frac{pq}{n}}+o\left(\frac{1}{\sqrt{n}}\right)$$

となる。然るに

$$\frac{d}{dp}(\sin^{-1}\sqrt{p})=\frac{1}{2\sqrt{p}\sqrt{1-p}}=\frac{1}{2\sqrt{pq}}$$

これを前式に代入すれば求める結果を得る。

§4.2. 統計量の変換　今茲に N 箇の箇所（地区）があって，その第 k 番箇所に対しては，確率変数 X_k と或る未知常数 m_k とが対応し，次の如き性質をもつとする。即ち X_k は負ならざる整数のみをとるものと，X_k が或る $x\ (\geqq 0)$ になる確率は，

$$Pr.\{X_k=x\}=f(x;m_k)\quad(x=0,1,2,\cdots)$$

に依って与へられるものとする。

以下 $\{m_k\}$ $(k=1,2,\cdots,N)$ は s 箇の常数 $\theta_1,\theta_2,\cdots,\theta_s$ の既知函数とし

$$m_k=\varphi_k(\theta_1,\theta_2,\cdots,\theta_s)\quad(k=1,2,3,\cdots;N)$$

とする。

問題は各 k について才 k 番箇所では実測値 x_k なるとき、これら N 箇の $\{x_k\}$ の値から、$\theta_1,\theta_2,\cdots\theta_s$ を推定する問題である。最尤法に依りこの解を求めるに

$$L=\log\prod_{k=1}^{N}f(x_k;m_k)$$

と置いて

$$\frac{\partial L}{\partial\theta_j}=\sum_{k=1}^{N}\frac{1}{f(x_k;m_k)}\cdot\frac{\partial f(x_k;m_k)}{\partial m_k}\frac{\partial m_k}{\partial\theta_j}=0$$

$$(j=1,2,\cdots;s)$$

を解いて L を最大ならしむべき $(\theta_1,\theta_2,\cdots,\theta_s)$ の値を求める。

然るに、この方程式が解き難いものであると実際上不便である。又変量分析法の場合の如き際には、各箇所に対応する確率変数の分散は相等しいのが望ましい。このため

$$x=g(y),\qquad m=g(\alpha)$$

なる変換を行ひ、函数 g を適当にとって、上述の目的に適合させるようとする。即ち

$$\frac{\partial L}{\partial\theta_j}=\sum_{k=1}^{N}\frac{1}{f(g(y_k),g(\alpha_k))}\left(\frac{\partial f}{\partial m}\right)_{g(y_k),g(\alpha_k)}\cdot g'(\alpha_k)\frac{\partial\alpha_k}{\partial\theta_j}=0$$

但し $x_k=g(y_k)$, $m_k=g(\alpha_k)$ とする。

変量分析法に用ひられる変換は、以上の考へに依るものである。

例1. $f(x,m)=\dfrac{n!}{x!(n-x)!}m^x(1-m)^{n-x}$ とすれば

$$\frac{1}{f(x;m)}\cdot\frac{\partial f(x;m)}{\partial m}=\frac{x}{m}-\frac{n-x}{1-m}=n\frac{\frac{x}{n}-m}{m(1-m)}$$

従って $\partial L/\partial\theta_j=0$ $(j=1,2,\cdots,s)$ は

$$n\sum_{k=1}^{N}\frac{\frac{x_k}{n}-m_k}{m_k(1-m_k)}\frac{\partial m_k}{\partial\theta_j}=0\quad(j=1,2,\cdots,s)$$

上述の目的に適合させるには、$x_k/n=Z$ とおき、添次 k を

取り去り

$$\frac{z-m}{m(1-m)}g'(\alpha) = \frac{g(y)-g(\alpha)}{g(\alpha)(1-g(\alpha))}g'(\alpha)$$
$$= C(y-\alpha)+o(|y-\alpha|)$$

となる様函数 g を決定すればよい，但し C は常数である。任意の y 及び α に対して成立つ故，

$$\frac{g'(\alpha)}{\sqrt{g(\alpha)(1-g(\alpha))}} = C_1 \qquad (C_1=\sqrt{C})$$

この微分方程式を解いて

$$\sin^{-1}\sqrt{g(\alpha)} = C_1\alpha + C_2$$

を得る。C 及び C' は常数である。

例2. Poisson 分布の場合。

$$f(x;m) = \frac{m^x e^{-m}}{x!}$$

然る時

$$\frac{1}{f(x;m)}\frac{\partial f(x;m)}{\partial m} = \frac{x-m}{m}$$

従って $\partial L/\partial\theta_j = 0\ (j=1,2,\cdots,s)$ は

$$\sum_{k=1}^{N}\frac{x_k-m_k}{m_k}\frac{\partial m_k}{\partial\theta_j} = 0 \quad (j=1,2,\cdots s)$$

依って求むる変換 g を得るには

$$\sum_{k=1}^{N}\frac{g(y_k)-g(\alpha_k)}{g(\alpha_k)}g'(\alpha_k)\frac{\partial\alpha_k}{\partial\theta_j} = 0 \ (j=1,2,\cdots s)$$

に於いて

$$\frac{g(y)-g(\alpha)}{g(\alpha)}g'(\alpha) = C(y-\alpha)+o(|y-\alpha|)$$

なる様函数 g を決定すればよい. 即ち

$$\frac{g'(\alpha)^2}{g(\alpha)} = C_1$$

となる故

$$g(\alpha) = (C_1\alpha + C_2)^2$$

となる。

§5. 変量分析法への應用

§5.1. 変量分析法に於ける平均差の使用

変量分析法に於いては従来，平方和，乗積和が最も根本的な役割を演じ，全理論は実に二次形式論と密接不可分な関係にあるのである。併しその根本の要求に於ては，要するに，変動を変動因に分析することである。変動を表現するのに，敢て分散乃至相関量に依ることを必要としない以上，変量分析法に於ても，平方和，乗積和のみに由らねばならぬとは限らないわけであらう。二次形式論の直交変換等に関する美麗なる理論は，変量分析法に於ても根本的な役割を果す。平方和，乗積和を使用しないと云ふ事になれば，直交変換等も利用出来ないことになり，従って，深い理論は不可能になりはすまいかとの懸念があらう。此の点の懸念を解くまで深いが，とにかく次の方法に依れば，少くも簡単な場合には，役に立つこともあらうかと思ふ。