相関係数の起源と Galton(1889)流の相関係数簡易推定法

○椎名乾平・野村奈都 (早稲田大学教育学部) Key Words: Galton, correlation, Bravais

Simple Galtonian (1889) Estimation of Correlation Coefficient. Kenpei SHIINA, Natsu NOMURA (School of Education, Waseda Univ.)

相関係数の起源 x, y を平均からの偏差とする時, Pearson (1896) は相関係数を, 2 変量正規分布

$$z = \frac{N}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left\{\frac{x^2}{\sigma_1^2(1-\rho^2)} - \frac{2xy\rho}{\sigma_1\sigma_2(1-\rho^2)} + \frac{y^2}{\sigma_2^2(1-\rho^2)}\right\}} \quad (1)$$

の母数 ρ として定義した。しかし Pearson 自身も認めるようにその前史として Bravais や Galton の貢献がある。

Bravais の貢献 (1) と同等の誤差関数は Bravais (1846) が

$$f(x, y) = (\sqrt{ab-c^2} / \pi) e^{-(ax^2+2cxy+by^2)} \quad (2)$$

としては半世紀前に与えていた。そのためか仏語圏では ρ を Bravais-Pearson の係数と呼ぶことがある。(1)(2)の指数部分は共に 3 母数の 2 次形式であるから容易に相互変換可能である。さて, (2)の指数部分を $ax^2+2cxy+by^2=k$ と書くと,

$D = \sqrt{(b-a)^2+4c^2}$ として, 長軸 $2\sqrt{2k/(a+b+D)}$, 短軸 $2\sqrt{2k/(a+b-D)}$, x 軸と長軸の角度 $\tan\theta = (b-a+D)/(2c)$, 面積 $\pi k / \sqrt{ab-c^2}$ である楕円の方程式となる。一方(1)の指数部分は分散や相関といった統計量と対応した形となっている。

Galton の貢献 Galton(1877)は「世代が変わっても, 身長のような属性の統計的性質はなぜ一定になるのか?」という問題に注目した。親世代の身長の平均からの偏差を y , 子世代の身長の平均からの偏差を x とする。1) まず親世代の身長の変異性 (標準偏差に相当) が 1.7 なのをデータより計算した。2) 次に 中間親=(父の身長+1.08×母の身長)/2 の変異性が $1.7/\sqrt{2}=1.2$ なのを理論的に求めた。3) 相関図より, 中間親の身長区分ごとに計算した子世代の平均身長が, 親世代の約 2/3 なのを観察し ($x=2/3y$, regression に他ならない), さらに子世代の変異性が中間親の 2/3 すなわち, $1.2 \times 2/3 = .8$ となるのを理論的に導いた。4) 最後に, 中間親の身長区分ごとに計算した子世代の変異性が, 親世代の身長にかかわらず大体 1.5 で一定となるのを観察した。以上より, 子世代の身長の変異性が, 分散の和の性質より $\sqrt{0.8^2+1.5^2}=1.7$ となり, 親世代の変異性とピッタリ重なるのを見出した。以上の導出を手掛かりとして, Galton and Dickson (1886, p.63)は, 親子の身長のデータに楕円をあてはめた。すなわち

$$\frac{y^2}{(1.22)^2} + \frac{(3x-2y)^2}{9(1.50)^2} = \frac{y^2}{(1.22)^2} + \frac{(x-2/3y)^2}{(1.50)^2} \quad (3)$$

(3)を展開すれば Bravais 型の楕円表現 (現在では確率楕円と呼ばれる) となるが, 相関係数に言及はなく, 事実上(1)と(2)の中間形となっている。(3)は重要な式であるが, 唐突に結論だけ提示された感が否めない。(3)式の導出を試みると, まず傾いていない楕円の方程式 $x^2/a^2 + y^2/b^2$ に一次変換

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2} & a/b\rho \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \sqrt{1-\rho^2}x + a/b\rho y \\ y \end{pmatrix} \quad (4)$$

を施す。ここで $\sqrt{1-\rho^2}$ は楕円を x 軸に沿って「潰す」変換であり, $a/b\rho$ は, x, y の線形関係を示す項である。逆行列を用いて x, y を X, Y で表現し, $x^2/a^2 + y^2/b^2$ に代入すると,

$$\begin{aligned} x^2/a^2 + y^2/b^2 = k &\Rightarrow \frac{1}{a^2} \left(\frac{X^2}{1-\rho^2} - 2\frac{\rho}{b} \frac{XY}{1-\rho^2} + \frac{a^2}{b^2} \frac{\rho^2 Y^2}{1-\rho^2} \right) + \frac{Y^2}{b^2} \\ &= \frac{X^2}{a^2(1-\rho^2)} - 2\frac{\rho}{ab} \frac{XY}{1-\rho^2} + \frac{Y^2}{b^2(1-\rho^2)} = 2k \end{aligned}$$

図1 $\rho \cong B/A$ である

とすることができ(1)の指数部分が再現される。

Galton 流の簡易相関係数推定法 さて $a=b$ の時, 楕円は $X^2-2\rho XY+Y^2=a^2k(1-\rho^2)$ と書ける。Y 軸と平行な直線 B との接点を求めるため Y で微分し 0 と置くと, 接点の座標値は, $(a\sqrt{k}, \rho a\sqrt{k})$ となる (図 1)。尚, この点は回帰直線 (図上破線) と楕円の交点 (の推定値) である。以上より, Galton の考え方を用いて相関係数の簡易推定を行うことができる。

1. 変数 x, y を基準化し散布図を描く。
2. 図上の点がうまく収まるように (傾いた) 楕円を描く。
3. 楕円と y 軸に平行な直線との接点を求める。すると
4. **相関係数=接点の y 座標÷接点の x 座標** (図 1 の B/A)。 x, y が基準化されていないなら, 分布の幅 α, β を考慮して, **相関係数= $B/A \times \alpha/\beta$** とすればよい。この方法は, 概算値を知りたいときや, 統計学教育で意味ある方法と思われる。

Galton の楕円 (正規分布の確率楕円) の留意点

主軸 (楕円の長軸) と回帰線は $\rho=1$ の場合以外は重ならず, 回帰線の傾きは, 主軸より緩やかになる (回帰という名称の由来である)。確かに理論上はそうなるのだが, 直感的にはなかなか納得しがたい。そこで図 2 で示すような, 3D 実物模型を作製してみた ($\rho=0.3$)。この立体モデルでは視覚と触覚で主軸と回帰線の間を確認できるので, 教育上有効であろう。

図2