㊸ 標本分布のある性質について

青山　博次郎

§1. 緒　言

正規母集団 N (a, σ) より n 個の標本を任意抽出するとき，その標本平均 $\overline{x}$ と，標本分散 S^2 とは独立であることはよく知られている。

逆に $\overline{x}$ 及び S^2 が独立ならば，もとの分布は正規分布又は単位分布であることも証明された。[1]

しかし標本平均，標本分散の同時分布と，もとの分布との関係についてはあまり詳しい研究がないように思う。

先年「よみ書き能力調査」が行われ，その得点分布について平均と分散が抛物線的関係で結ばれていることより，もとの分布函数の型は

$y = ax^m + b$ であると推定された。[2]

これについて少しく考えてみよう。

§2. ある分布函数の型

有限なる区間 $[a, b]$ 内に於て一つの分布を考え，それが密度数 $f(x)$ をもつものとすると，分散 V，平均 M は

$$V = \int_a^b x^2 f(x)\,dx - M^2 \tag{1}$$

$$M = \int_a^b x f(x)\,dx \tag{2}$$

を満足し，また仮定により

$$\int_a^b f(x)\,dx = 1 \tag{3}$$

が成立する。

このとき V が M の二次式

$$V = AM + B - M^2 \tag{4}$$

で表わされ，従つて

$$\int_a^b x^2 f(x)\,dx = A\int_a^b x f(x)\,dx + B \tag{5}$$

が成立つ場合は如何なる場合であるかを考えよう。

(5) は (3) によつて

$$\int_a^b (x^2 - Ax - B) f(x)\,dx = 0 \tag{6}$$

となるから，(6) について調べればよい。 A，B は $f(x)$ のもつパラメータの数より少いパラメータを含む定数とする。

区間 $[a, b]$ を $[-1, 1]$ に移す変換

$$y = \frac{2x - (b+a)}{b-a} \tag{7}$$

により，(6) は

$$\int_{-1}^{1} \left\{ \frac{(b-a)^2}{4} y^2 + \frac{b^2 - a^2 - A(b-a)}{2} y + \left(\frac{(b+a)^2}{4} - \frac{A(b+a)}{2} - B \right) \right\} f_1(y) = 0 \tag{8}$$

となる。 但し $f_1(y) = f\left(\frac{(b-a)y + (b+a)}{2} \right)$ とする。

さて $f_1(y)$ を $[-1, 1]$ における *Legendre* の多項式 $P_n(y)$ を用いて（一様収斂性を仮定して）

$$f_1(y) = \sum_{n=0}^{\infty} c_n P_n(y) \tag{9}$$

で表わしたとすると，（8）は

$$\frac{(b-a)^2}{15}C_2+\frac{b^2-a^2-A(b-a)}{3}C_1+2\left(\frac{(b+a)^2}{4}+\frac{(b-a)^2}{12}-\frac{A(b+a)}{2}-B\right)C_0=0 \tag{10}$$

となる。　また（3）の条件より

$$C_0=\frac{1}{b-a} \tag{11}$$

となる。

故に（10），（11）の条件の下に

$$f(x)=C_0+C_1\frac{2x-(b+a)}{b-a}+\frac{C_2}{2}\left\{3\left(\frac{2x-(b+a)}{b-a}\right)^2-1\right\}$$

$$+C_3P_3\left(\frac{2x-(b+a)}{b-a}\right)+\cdots\cdots \tag{12}$$

なることが分る。　勿論 $f(x)\geqq 0$ なる条件を満足する様にしなければならない。

このように $C_3, C_4, \cdots\cdots$ は $f(x)\geqq 0$ を満足する限り種々の値が取り得られるから，（6）を満足する如き $f(x)$ は無数に存することが分る。

特に，$C_3=C_4=\cdots\cdots=0$ の場合は $f(x)$ は C_1, A, B のみをパラメータとして含む。

従つて A，B を一つのパラメータ α のみを含む定数とするときは $f(x)$ は

$$f(x)=A(C_1,\alpha)+B(C_1,\alpha)x+C(C_1,\alpha)x^2 \tag{13}$$

なる x の二次函数であると考えてもよい。

例．　區間 $[0,\ x_0]$ に於ての分布が密度函数

$$f(x)=ax^2+bx+c$$

をもつものとする。

このとき $\int_0^{x_0}f(x)\,dx=1$ の條件により，パラメータは

ただ二つである。

実際に分散，平均を計算すると

$$V = \frac{6}{5} x_0 M + \frac{c}{30} \bar{x}_0^3 - \frac{3}{10} x_0^2 - M^2$$

となつて，パラメータCのみを含む形にすることが出来る。

（4）によれば，母平均，母標準偏差の間には円関係が存在する。もし「よみ書き能力調査」に於て実際に標本平均，標準偏差の間に，抛物線関係 $\sigma^2 = AM + B$ が成立しているならば，標本平均標本分散の間には直線関係が存することゝなる。

この場合も上と同様にして條件（10）の代りに

$$\frac{(b-a)^3}{30} c_2 - \frac{(b-a)^4}{36} c_1^2 - \frac{A}{6}(b-a)^2 c_1 - \frac{A}{2}(b+a) - B + \frac{(b-a)^2}{12} = 0 \qquad (14)$$

が得られるだけで其他は全く同様である。

次にこのような関係の生ずるのはいかなる分布の型に由来するかを考えてみることにする。

§ 3. 標本平均，標本分散の分布ともとの分布函数との関係．

標本平均 $\bar{x}$ と，標本分散 S^2 の同時分布函数は分らないので，$\bar{x}$ と S^2 に線形関係が存するときもとの分布の型の特長をしらべてみることにする。

一般に与えられた母集団の大きさを N とし，これより n 個の標本を任意抽出し，標本平均，標本分散，母平均，母分散をそれぞれ，$\bar{x}$，S^2，m，σ^2 とおくと

$$\left.\begin{aligned} E(\bar{x}) &= m \\ V(\bar{x}) &= \frac{N-n}{N-1}\frac{\sigma^2}{n} \\ E(S^2) &= \frac{N}{N-1}\frac{n-1}{n}\sigma^2 \end{aligned}\right\} \qquad (15)$$

$$V(S^2) = \frac{N(N-n)(n-1)\sigma^4}{(N-1)^2(N-2)(N-3)n^3}\left\{(Nn-N-n-1)(N-1)\alpha_4 - (N^2 n - 3N^2 + 6N - 3n - 3)\right\}, \quad \alpha_4 \text{ は尖度}$$

が成立つ。

このとき $cov(\bar{x}, S^2)$ の期待値は $\frac{N}{N-1}\frac{n-1}{n}\rho(\bar{x}, S^2)\sigma_{\bar{x}}\sigma_{S^2}$

であるから

$$\rho(\bar{x}, S^2) = \frac{N-1}{N}\frac{n}{n-1}\frac{1}{\sigma_{\bar{x}}\sigma_{S^2}} E\left\{(S^2 - \frac{N}{N-1}\frac{n-1}{n}\sigma^2)(\bar{x} - m)\right\} \quad (16)$$

右辺の期待値を計算すると

$$E\left\{(S^2 - \frac{N}{N-1}\frac{n-1}{n}\sigma^2)(\bar{x}-m)\right\} = \frac{n-1}{n^2}\alpha_3\sigma^3, \quad \alpha_3 \text{ は歪度} \quad (17)$$

故に（15），（16），（17）より

$$\rho(\bar{x}, S^2) = \frac{(N-1)^2 n}{N(N-n)}\sqrt{\frac{(N-1)(N-2)(N-3)}{N(n-1)}}\frac{\alpha_3}{\sqrt{(Nn-N-n-1)(N-1)\alpha_4 - (N^2 n - 3N^2 + 6N - 3n - 3)}} \quad (18)$$

無限母集団のときは $N \to \infty$ とおいて

$$\rho(\bar{x}, S^2) = \frac{n\alpha_3}{(n-1)\sqrt{\alpha_4 - 3 + \frac{2n}{n-1}}} \quad (19)$$

これより $\alpha_3 = 0$ のとき，即ち対称な分布函数（勿論対称でないものもあるが）ならば $\rho(\bar{x}, S^2) = 0$ となる。

正規分布なら勿論 $\rho(\bar{x}, S^2) = 0$ である。

かくて回帰が線形のときは $\alpha_3 \gtreqless 0$ に従って回帰直線は右上り，水平，右下りとなる。逆にこれよりもとの分布函数の歪度が推定される。

例 1. 母集団として 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 なる 10 個の数を記入したそれぞれ 19, 9, 5, 4, 4, 3, 2, 2, 1, 1 枚のカードを作り，これより 10 枚のカードを任意抽出（毎回元に戻して）することを50回くり返えした。

このとき実際に $\bar{x}$ と S^2 の標本相関数は 0.6883 であつて，$\rho = 0$ なる仮説は棄却される。

回帰係数の直線性の検定より直線回帰があると考えられるが実際，その回帰直線は右上りであることが示された。（$\alpha_3 > 0$ の場合）

例 2. 母集団として 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 なる11個の数を記入したそれぞれ 1, 2, 4, 7, 7, 8, 7 7, 4, 2 1 枚のカードを作り，これより前例と同様に 10 枚のカードを任意抽出することを50回くり返えした。

$\bar{x}$ と S^2 との標本相関係数は 0.129 であり，$\rho = 0$ の仮説は捨てられない。

実際に二つの回帰直線を作つてみるとほゞ直交していることが示される。（$\alpha_3 = 0$ の場合）

例 3. 「よみ書き能力調査」における「学歴なし」，「小学中退」の人の能力分布は $\alpha_3 > 0$ で，「小学卒」「高小卒以上」「全員」の場合の能力分布は $\alpha_3 < 0$ である。

このとき $\bar{x}$, S^2 の回帰直線は上述の結果とよく一致している。

（註） 1) R. C. Geary: The distribution of "Student's" ratio for non-normal samples. Journ. Royal Statist. Soc., Supplement 3 (1936) 及び T. Kawata and H. Sakamoto: On the characterisation of the normal population by the independance of the sample mean and the sample variance, Journ Math. Soc. Japan vol.1

No. 2 (1949)

2) 林 知己夫 他 : リテラシイ調査にあらわれた分布の型など, 統数研講究録 vol. 5 No. 6. 1949

3) Kendall : Advanced Mathematical Statistics vol. 1. P. 284.