

統計量の分布から母集団分布がどこまでわかるか

統数研 清水 良一

例1. P は、すべてのモーメントともち、しかも、二点のモーメントによって一意的に定まるような分布の族とする。

X_1, \dots, X_n を $F \in P$ からの標本とする。 F のモーメントは $L = a_1 X_1 + \dots + a_n X_n$, (a_j は任意の正の数) のモーメントで定まるから、 F は統計量 L の分布から完全に定まる。

例2. X_0, X_1, \dots, X_n が $N(0, \sigma^2)$ からの標本なら、

$$t = \sqrt{n} \frac{X_0}{\sqrt{\sum_{i=1}^n X_i^2}}$$

は d.f. n の t 分布に従う。しかし、逆に、統計量 t の分布が t 分布であるからといって、 X_0 の分布が $N(0, \sigma^2)$ とはいえない。密度関数、

$$\frac{\sqrt{2}}{\pi \sigma [1 + (\frac{x}{\sigma})^2]} \quad \text{あるいは} \quad \frac{\sqrt{2} x^2}{\sigma^3 [1 + (\frac{x}{\sigma})^2]}$$

とも一つ分布がも知れない。

以下、分布族 $\{f(x; \theta)\}$ の分布から標本 X_1, \dots, X_n にまとづく統計量 T で、

- (i) T の分布が θ に無関係で、
- (ii) T の分布から $f(x)$ が一意的に定まる、
それをもと求めよ。

A. $\theta = (\mu, \sigma^2)$, $f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} f\left(\frac{x-\mu}{\sigma}\right)$ とする。

$Y_k = (X_k - \bar{X})/S$, $k=1, \dots, n$ とおき、 $T = (Y_1, \dots, Y_n)$ が(i)を満足することは明らか。適当な条件のもとで(ii)も成る。

B. $\theta = (\theta_1, \dots, \theta_n, \underbrace{0, \dots, 0}_{s-k})$, $f(x; \theta) = f(x - \theta)$
 $x = (x_1, \dots, x_s)$ が s で n より少ないとす。 $x^* = (x_1, \dots, x_n, 0, \dots, 0)$
 とおき。統計量 $T = (x_1 - x_3^*, x_2 - x_3^*)$ が(i), (ii)を満足する。

証明。(i)は明らか。 $f(x) \rightarrow \text{ch.f. } \varphi(t) \rightarrow 3$ 。

$(x_1 - x_3^*) - (x_2 - x_3^*) = x_1 - x_2$ だから、 T の分布
 から $|\varphi(t)|^2$ が一意に定まる。 $T \rightarrow \text{ch.f. } \psi(s, t)$ すと、

$$\begin{aligned}\psi(s, t) &= E(e^{i(x_1 - x_3^*) s' + i(x_2 - x_3^*) t'}) \\ &= \varphi(s) \varphi(t) \overline{\varphi(s^* + t^*)}.\end{aligned}$$

ここで $A(t) = \arg \varphi(t)$, $S(s, t) = \arg \psi(s, t)$

とおこう,

$$S(s, t) = A(s) + A(t) - A(s^* + t^*).$$

homogeneous equation

$$A(s) + A(t) - A(s^* + t^*) = 0$$

if $t = 0$ とおうと $A(s) = A(s^*)$ とおこう.

$$A(s^*) + A(t^*) = A(s^* + t^*).$$

これが、

$$A(s) = A(s^*) = \sum_{j=1}^k c_j s_j.$$

$s < 0$, $s = 0$ の近傍で, $\varphi(s)$ は $\psi(s, t)$ にさ.

factor $e^{i \sum c_j s_j}$ を除いて一意性を定めよ.

C.

A はおうと, $n \geq 2$ とおう. 統計量

$$T = \left(\log \left| \frac{Y_2 - Y_1}{Y_3 - Y_1} \right|, \log \left| \frac{Y_2 - Y_1}{Y_3 - Y_2} \right|, \log \left| \frac{Y_3 - Y_2}{Y_3 - Y_1} \right|, \log \left| \frac{Y_3 - Y_2}{Y_3 - Y_4} \right|, \text{sign}(Y_3 - Y_1), \text{sign}(Y_2 - Y_1), \right. \\ \left. \text{sign}(Y_3 - Y_4), \text{sign}(Y_3 - Y_5) \right)$$

が条件(i), (ii) を満たす。

証明. f は (Z_1, Z_2) , $Z_1 = X_3 - X_1$, $Z_2 = X_2 - X_1$ の分布から定まる. この分布は scale family $\{\sigma^{-2} g(x/\sigma)\}$. に属する. (Z_{11}, Z_{12}) , (Z_{21}, Z_{22}) , (Z_{31}, Z_{32}) とこの 3 つ分布からうの標本として,

$D_j = (\log|Z_{j1}|, \log|Z_{j2}|, \text{sign } Z_{j1}, \text{sign } Z_{j2})$, $j=1, 2, 3$ と IF とし, D_1, D_2, D_3 は, 4 变量, location family $\{h(x-\theta); \theta = (\log \sigma, \log \sigma, 0, 0)\}$ からの標本である.

よって, B から, $(D_1 - D_3^*, D_2 - D_3^*) (= T)$ の分布から t が, ($T = \text{triv. } \mathcal{C}$, g が: $|T - t|$ が \mathcal{C} の f で一意に定まる).

D. Y_k , $k=1, \dots, n$ $\in A$ のすなは IF とし, (Y_1, \dots, Y_n) は, $Y_1 + \dots + Y_n = 0$, $Y_1^2 + \dots + Y_n^2 = 1$ 上の分布である.

正規分布から (Y_1, \dots, Y_n) の分布は一様分布, 逆に (Y_1, \dots, Y_n) の分布が一様なら X の分布の正規分布である. さらには (Y_1, \dots, Y_n) の分布が連続な密度関数 g をもつとする.

(つきのことばへとる:) $n \geq 6$ とする.

$$Z_1 = (y_1, y_2, y_3, y_4, y_5, y_6, \dots, y_n),$$

$$Z_2 = (y_1, y_2, y_3, y_1, y_2, y_3, \dots, y_n),$$

$$Z_3 = (y_4, y_5, y_6, y_4, y_5, y_6, \dots, y_n),$$

$$\text{E. } y_1 + y_2 + y_3 = y_4 + y_5 + y_6,$$

$$y_1^2 + y_2^2 + y_3^2 = y_4^2 + y_5^2 + y_6^2$$

$$(y_1, y_2, y_3) \neq (y_4, y_5, y_6),$$

故此 $\tau_2 \neq 3$ 等， Z_1, Z_2, Z_3 为 $5:1+3$ 之值亦等 $1:1:2:4$ 。

X の分布は正規分布である。

証明. X の分布の確率密度を f とするとき、

$$g(y_1, \dots, y_n) = C \cdot \int_{-\infty}^{\infty} dt \int_0^{\infty} s^{n-2} \cdot f(t+sy_1) \cdots f(t+sy_n) ds$$

と書ける。上, 条件から,

$$\int_{-\infty}^{\infty} dt \int_0^{\infty} s^{n-2} [f(t+sy_1) \cdot f(t+sy_2) \cdot f(t+sy_3) \\ - f(t+sy_4) \cdot f(t+sy_5) \cdot f(t+sy_6)] ds = 0.$$

八、若 $f(t_0) > 0$ 且 $f' \neq 0$, 则 t_0 的附近 $\exists t$, 使 $f(t) < 0$

"S" is being.

$$f(t+s y_1) \cdot f(t+s y_2) \cdot f(t+s y_3) \\ = f(t+s y_4) \cdot f(t+s y_5) \cdot f(t+s y_6)$$

が成り立つことになる。

$$\xi(t) = \log f(t_0 + t)$$

とおくと、

$$\sum_1^3 \xi(t + s y_j) - \sum_4^6 \xi(t + s_j) = 0.$$

6

これを解いて、

$\xi(t) = \text{二次, 多項式}.$

が得られる。

一般に、 ξ_j が連続たり。

$$\sum_{j=1}^n \xi_j(t + s y_j) = 0$$

から $\xi_j(t)$ が多項式であることが導かれ。こゝで述べた結果は、 X_1, \dots, X_n が同じ分布 π 従うという仮定下に得られた。

以上、Linnik, Zinger らによる仕事の紹介である。