数理解析研究所講究録
第 223 巻 1974 年 1-6



# 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.　$\mathcal{P}$は、すべてのモーメントをもち、しかも、これら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．　$X_1, \cdots, X_n$ を $F \in \mathcal{P}$ からの標本とする．$F$のモーメントは $L = a_1X_1 + \cdots + a_nX_n$，（$a_j$ は任意の正の数）のモーメントできまるから、$F$は統計量$L$の分布から完全に定まる．

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

$$t = \sqrt{n}\, X_0 \Big/ \sqrt{\sum_1^n X_i^2}$$

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，$X$ の分布が $N(0, \sigma^2)$ とはいえない．　密度関数、

$$\frac{\sqrt{2}}{\pi\sigma\left[1 + \left(\frac{x}{\sigma}\right)^4\right]} \qquad \text{あるいは,} \qquad \frac{\sqrt{2}\ x^2}{\sigma^3\left[1 + \left(\frac{x}{\sigma}\right)^4\right]}$$

ともつ分布かも知れない.

以下、分布族 $\{f(x;\theta)\}$ の分布からの標本 $X_1, \cdots, X_n$ にもとづく統計量Tで、

(i) Tの分布が$\theta$に無関係、で、

(ii) Tの分布から $f(x)$ が一意的に定まる,

ようなものを求める.

A. $\theta = (\mu, \sigma^2)$, $f(x,\theta) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$ とする.
$Y_k = (X_k - \bar{X})/S$ , $k=1,\cdots,n$ とおくと, $T=(Y_1,\cdots,Y_n)$ が(i)を満足することは明らか. 適当な条件のもとで(ii)もいえる.

B. $\theta = (\theta_1, \cdots, \theta_k, \underbrace{0,\cdots,0}_{s-k})$ , $f(x;\theta) = f(x-\theta)$
$x = (x_1, \cdots, x_s)$ が与えられたとき. $x^* = (x_1, \cdots, x_k, 0, \cdots, 0)$ とおく. 統計量 $T = (x_1 - x_3^*, x_2 - x_3^*)$ がてきとうな条件下で(i), (ii)を満足する.

証明. (i)は明らか. $f(x)$ の ch.f. を $\varphi(t)$ とする.
$(x_1 - x_3^*) - (x_2 - x_3^*) = x_1 - x_2$ だから、 Tの分布から $|\varphi(t)|^2$ が一意に定まる. Tの ch.f. を $\psi(s,t)$ とすると,

$$\psi(s,t)=E\left(e^{i(x_1-x_3^*)s'+i(x_2-x_3^*)t'}\right)$$

$$=\varphi(s)\,\varphi(t)\,\overline{\varphi(s^*+t^*)}\ .$$

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

とおくと,

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

homogenious equation

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

は　$t=0$ とおいて　$A(s)=A(s^*)$　となるから.

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

これより,

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

かくして, $s=0$ の近傍で, $\varphi(s)$ は $\psi(s,t)$ から.
factor $e^{i\Sigma c_j s_j}$ を除いて一意的に定まる.

C.

Aにおいて, $n\geq 9$ とする.　統計量

$$T=\left(\log\left|\frac{Y_3-Y_1}{Y_9-Y_7}\right|,\ \log\left|\frac{Y_2-Y_1}{Y_8-Y_7}\right|,\ \log\left|\frac{Y_6-Y_4}{Y_9-Y_7}\right|,\ \log\left|\frac{Y_5-Y_4}{Y_8-Y_7}\right|,\ \mathrm{sign}(Y_3-Y_1),\mathrm{sign}(Y_2-Y_1),\ \mathrm{sign}(Y_6-Y_4),\mathrm{sign}(Y_5-Y_4)\right)$$

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. $(Z_{11}, Z_{12})$, $(Z_{21}, Z_{22})$, $(Z_{31}, Z_{32})$ とこの分布からの標本として,

$$V_j = (\log|Z_{j1}|, \log|Z_{j2}|, \mathrm{sign}\, Z_{j1}, \mathrm{sign}\, Z_{j2}), \quad j=1,2,3$$

を作ると, $V_1, V_2, V_3$ は, 4変量の 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, $(V_1 - V_3^*,\ V_2 - V_3^*)$ $(= T)$ の分布から $h$ が, したがって, $g$ が, したがって $f$ が一意に定まる.

D. $Y_k$, $k=1,\cdots,n$ とAのように作ると, $(Y_1,\cdots,Y_n)$ は, $Y_1+\cdots+Y_n=0$, $Y_1^2+\cdots+Y_n^2=1$ 上の分布になる. 正規分布なら $(Y_1,\cdots Y_n)$ の分布は一様分布に, 逆に $(Y_1,\cdots Y_n)$ の分布が一様なら $X$ の分布は正規分布である. さらに $(Y_1,\cdots,Y_n)$ の分布が連続な密度関数 $g$ をもつとする.

(つぎのことがいえる:) $n \geq 6$ として,

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

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

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

ただし、 $$y_1+y_2+y_3 = y_4+y_5+y_6,$$
$$y_1^2+y_2^2+y_3^2 = y_4^2+y_5^2+y_6^2$$
$$(y_1, y_2, y_3) \neq (y_4, y_5, y_6),$$

をみたす3点, $z_1, z_2, z_3$ における $g$の値が等しいなら、
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,

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

と書ける. 上の条件から,

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

いま $f(t_0) > 0$ とすると, $t_0$ の近くの $t$, $0$ に近い $s$ にたいして,

$$f(t+sy_1) \cdot f(t+sy_2) \cdot f(t+sy_3) = f(t+sy_4) \cdot f(t+sy_5) \cdot f(t+sy_6)$$

が成りたつことになる.

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

とおくと,

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

これを解いて,

$$\xi(t) = \text{二次の多項式}.$$

が得られる.

一般に、 $\xi_j$ が連続なら.

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

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、 $X_1, \cdots, X_n$ が同じ分布に 従うという仮定なしになりたつ。

以上, Linnik, Zinger らによる仕事の紹介である.