数理解析研究所講究録
第 231 巻 1975 年 113-120

統計的判別法について

静岡大　工　多賀保志
西　晃央
阪　大　養　石井恵一

§1. 1次元の場合

1次2次のモーメントが指定された2個の1次元分布[平均 μ_1, μ_2 $(\mu_1 \neq \mu_2)$, 分散 σ_1^2, σ_2^2] に従う母集団 Π_1, Π_2 から等確率で得られる観測値に基づく誤判別の確率に関して次の結果がある。

__定理1__ [Becker, Chernoff]

$$\sup_{F \in \mathcal{F}} \inf_{\phi \in \Phi} e(\phi, F) = \left[2(1+S^2)\right]^{-1}$$

ここで、 $S = |\mu_1 - \mu_2| / (\sigma_1 + \sigma_2)$, $F = (F_1, F_2)$, $\mathcal{F} = (\mathcal{F}_1, \mathcal{F}_2)$

$\mathcal{F}_i = \mathcal{F}(\mu_i, \sigma_i^2)$ は平均 μ_i 分散 σ_i^2 を持つすべての確率分布からなる分布族である。$\phi \equiv \phi(x) = (\phi_1(x), \phi_2(x))$, $\phi_i(x) \geq 0$, $\phi_1(x) + \phi_2(x) = 1$

x を観測したとき $\phi_i(x)$ の確率で Π_i からの観測値と判定する。

$$e(\phi, F) = \frac{1}{2}\left\{\int(1-\phi_1)\,dF_1 + \int(1-\phi_2)\,dF_2\right\}$$

である。

更に4次のモーメント τ_1, τ_2 を指定した場合には次の結果が

ある．

<u>定理2</u>［西，多賀］　$i=1$ or 2 に対して $|\xi^*|<\sqrt{\nu_i}$ 且 $\psi_i(\xi^*)>0$ 又は $|\xi^*|>\sqrt{\tau_i/\nu_i}$ 且 $\psi_i(\xi^*)>0$ が成立することが

$$\sup_{F\in\mathfrak{F}(\mu,\nu,\tau)}\ \inf_{\phi\in\Phi} e(\phi,F) < \sup_{F\in\mathfrak{F}(\mu,\nu)}\ \inf_{\phi\in\Phi} e(\phi,F)$$

となるための十分条件である．

ここで，$\xi^*=(\sigma_1\mu_2+\sigma_2\mu_1)/(\sigma_1+\sigma_2)$，$\psi_i(\xi)=\sigma_i^2\xi^4-(2\nu_i\sigma_i^2+\rho_i^2)\xi^2+2\mu_i\rho_i^2\xi+(\sigma_i^2\nu_i^2-\rho_i^2\mu_i^2)$，$\nu_i=\sigma_i^2+\mu_i^2$，$\rho_i^2=\tau_i-\nu_i^2$ である．

<u>注意</u>　$\nu_1=\nu_2$，$\tau_1=\tau_2$ の場合は定理2の条件は必要十分条件である．

§2. 多次元の場合．

§1 での μ_i, σ_i^2 の代りに平均ベクトル μ_i，分散共分散行列 Σ_i $(i=1,2)$ が指定された場合に数理計画法の観点からの Chernoff の結果の拡張として次の定理が得られている．

<u>定理3</u>［石井，多賀］　π_i を観測値が Π_i から得られる事前分布とする．$\pi_1\leq\pi_2$ と仮定してさしつかえない．

<u>(i).</u>　$1\leq\pi_2/\pi_1<1+(\mu_1-\mu_2)'\Sigma_2^{-1}(\mu_1-\mu_2)$ の場合．

$$\max_{F\in\mathfrak{F}}\ \inf_{\phi\in\Phi} e(\phi,F)=\min_{\phi\in\Phi}\ \sup_{F\in\mathfrak{F}} e(\phi,F)=1/t.$$

ここで，$x'(\mu_1-\mu_2)/\sqrt{x'\Sigma_2 x}\geq\sqrt{\pi_2/\pi_1-1}$ なる領域内の任意の x に対して，t に関する方程式

$$\sqrt{x'\Sigma_1 x}\sqrt{\pi_1 t-1}+\sqrt{x'\Sigma_2 x}\sqrt{\pi_2 t-1}-x'(\mu_1-\mu_2)=0$$

は唯一の実根 $t=t(x)$ を持つ。また $\max t(x)=t(b)\equiv t_0$ なる b は正数倍を除いて一意に存在する。更に $e(\phi,F)$ の鞍点を具体的に構成できる。

(ii). $\pi_2/\pi_1 \geqq 1+(\mu_1-\mu_2)'\Sigma_2^{-1}(\mu_1-\mu_2)$ の場合．

$$\sup_{F\in\mathcal{F}}\inf_{\phi\in\Phi} e(\phi,F)=\min_{\phi\in\Phi}\sup_{F\in\mathcal{F}} e(\phi,F)=\pi_1$$

$\sup_{F\in\mathcal{F}} e(\phi,F)$ は $\phi_1^*(x)\equiv 0$ ，$\phi_2^*(x)\equiv 1$ で最小化できるが、$\inf_{\phi\in\Phi} e(\phi,F)$ を最大化する F^* は必ずしも存在しない．

次に判別方法を線型に限定した場合を考える。Φ_0 を線型判別法の全体，$\Phi^b\subset\Phi_0$ は $\phi_1(x)=\chi_{\{x\mid b'x\geqq c\}}(x)$ $(-\infty<c<\infty)$ なる線型判別法の全体とする．ここでベクトル b は定理3 (i) で導入されたものである．

定理4［石井，多賀］

(i). $1\leqq \pi_2/\pi_1 < 1+(\mu_1-\mu_2)'\Sigma_2^{-1}(\mu_1-\mu_2)$ の場合．

$$\sup_{F\in\mathcal{F}}\inf_{\phi\in\Phi^b} e(\phi,F)=\sup_{F\in\mathcal{F}}\inf_{\phi\in\Phi_0} e(\phi,F)=\sup_{F\in\mathcal{F}}\inf_{\phi\in\Phi} e(\phi,F)=1/t_0$$

しかし、$\inf_{\phi\in\Phi_0}\sup_{F\in\mathcal{F}} e(\phi,F)$ は一般には $\inf_{\phi\in\Phi}\sup_{F\in\mathcal{F}} e(\phi,F)$ より大きい．

(ii). $\pi_2/\pi_1 \geqq 1+(\mu_1-\mu_2)'\Sigma_2^{-1}(\mu_1-\mu_2)$ の場合．

$$\sup_{F\in\mathcal{F}}\inf_{\phi\in\Phi^b} e(\phi,F)=\inf_{\phi\in\Phi^b}\sup_{F\in\mathcal{F}} e(\phi,F)=\pi_1$$

§3. 1次元変量の多次元化

2N次までのモーメントが指定された1次元確率変数Xを考える。

$$E\{(X, X^2, \cdots X^N)\} = (\mu_1, \mu_2, \cdots, \mu_N)$$

$$E\{(X-\mu_1, X^2-\mu_2, \cdots X^N-\mu_N)'(X_1-\mu_1, X^2-\mu_2, \cdots, X^N-\mu_N)\}$$

$$= (\mu_{l+m} - \mu_l \cdot \mu_m)_{l, m=1, \cdots, N}$$

さて、平均ベクトル $(\mu_1^{(i)}, \cdots\cdots \mu_N^{(i)})$

　　分散共分散行列 $(\mu_{l+m}^{(i)} - \mu_l^{(i)} \cdot \mu_m^{(i)})_{l, m=1, \cdots, N}$ 　　$(i=1, 2)$

を持つN次元分布を $\mathcal{G}^{(N)} = (\mathcal{G}_1^{(N)}, \mathcal{G}_2^{(N)})$, $\mu_l (l=1, 2, \cdots 2N)$ を l 次モーメントとして持つ1次元分布を $\mathcal{F}^{(N)} = (\mathcal{F}_1^{(N)}, \mathcal{F}_2^{(N)})$ とおく. $\pi_1 = \pi_2 = 1/2$ の場合を考えると定理3 (i) から

$$\sup_{F \in \mathcal{F}^{(N)}} \inf_{\phi \in \Phi} e(\phi, F) \leq \max_{G \in \mathcal{G}^{(N)}} \inf_{\phi \in \Phi} e(\phi, G) = 1/t_0 \equiv 1/t_{0(N)}$$

がいえる. 左辺の評価として $1/t_{0(N)}$ のNに関しての変動を調べる. $1/t_{0(N)}$ がNに関して減少することは容易にわかる.

§4. 多変量化による簡便法とexact boundの比較

4次までのモーメント(非退化)が指定されている場合を考える. F_i の4次までのモーメント $\mu_1^{(i)}, \mu_2^{(i)}, \mu_3^{(i)}, \mu_4^{(i)}$ $(i=1, 2)$ が指定されているものとし、これらをモーメントとして持つあらゆる F_i の全体を $\mathcal{F}_i$ $(i=1, 2)$ とする. $\mathcal{F}_1$, $\mathcal{F}_2$ の prior probability を π_1, π_2 $(\pi_1 + \pi_2 = 1)$ とし、分類方式 $\phi = (\phi_1, \phi_2)$ を用いるとき、両母

集団の分布 F_1, F_2 に対する誤判別の確率は $e(\phi, F) = \pi_1 \int \phi_2(x) dF_1(x) + \pi_2 \int \phi_1(x) dF_2(x)$ である。これに対してミニマックス定理：

$$\inf_{\phi} \sup_{F} e(\phi, F) = \sup_{F} \inf_{\phi} e(\phi, F) \quad (= v \text{ とおく})$$

は常に成り立つが、鞍点 (ϕ^*, F^*) は存在するとは限らない。しかし、F_1^*, F_2^* の一方あるいは両方の4次のモーメントが指定された μ_4 より小さい場合を許せば「任意の ϕ と $F_1 \in \mathcal{F}_1$, $F_2 \in \mathcal{F}_2$ に対し

$$e(\phi^*, F) \leqq e(\phi^*, F^*) \leqq e(\phi, F^*)$$」

となる ϕ^* と F^* の組が常に存在する。4次のモーメントに"欠損"が起るのは、実数空間が compact でないために $e(\phi^*, F)$ の sup に近づく F の列に対応する4次のモーメントが"流出部分"をもつことがあるからである。以下の F_i^* はこのような意味で用いるので、必ずしも $F_i^* \in \mathcal{F}_i$ でないことに注意されたい。

v の値を求めるには ϕ^* と $F^* = (F_1^*, F_2^*)$ を求めればよいが、F^* は次のような諸種の場合に分類される。

①. F_1^*, F_2^* ともに3点分布でそのうち1点が F_1^* と F_2^* に共通

①′. ①で F_1^* と F_2^* の一方または両方が(4次のモーメントに)欠損のある2点分布(無限遠点を $x^4 dF_i$ の support に含めれば①とみなせる).

②. ともに4点分布で、そのうち2点が共通

②′. ②で F_1^*, F_2^* の一方のみが欠損(をもちうる)3点分布

②″. ②で F_1^*, F_2^* の一方のみが欠損のある2点分布

③. F_1^*, F_2^* ともに欠損のある2点分布で、2点とも共通

上で②'と③の場合は、$v = \min(\pi_1, \pi_2)$ となり、たとえば、$\pi_1 \leqq \pi_2$ のときは、ϕ^*として $\phi_1^*(x) \equiv 0, \phi_2^*(x) \equiv 1$ をとればよい。

上記の各場合の結果を実際に求める方法を詳述するのは省略し、①(①')と②(②')の場合にその求め方と多変量化による簡便法でどのくらいlossがあるかを例示する。

①の場合

$$M_i = \begin{pmatrix} 1 & \mu_1^{(i)} & \mu_2^{(i)} \\ \mu_1^{(i)} & \mu_2^{(i)} & \mu_3^{(i)} \\ \mu_2^{(i)} & \mu_3^{(i)} & \mu_4^{(i)} \end{pmatrix} \qquad (\text{Hankel matrix}) \qquad (i=1,2)$$

$$K_i(x) = \frac{1}{\pi_i} (1 \ x \ x^2)\, M_i^{-1} \begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$$

とおく。もし方程式 $K_1(x) = K_2(x)$ が実根 x_0 をもち、$K_1'(x_0) \cdot K_2'(x_0) < 0$ が成り立つならば、求める v の値は

$$v = 1/K_1(x_0) \quad (= 1/K_2(x_0))$$

である。(このような x_0 が存在しないときは①以外の場合になる)。なお、x_0 の1点分布のHankel matrixをM_0とすると$1/v$は、$\frac{1}{\pi_i} M_i^{-1} M_0$ $(i=1,2)$ の0でない唯一の共通固有値である。

[例] $\mu_1^{(1)} = 1 \quad \mu_2^{(1)} = 2 \quad \mu_3^{(1)} = 4 \quad \mu_4^{(1)} > 8$

$\mu_1^{(2)} = -1, \quad \mu_2^{(2)} = 2 \quad \mu_3^{(2)} = -4 \quad \mu_4^{(2)} > 8$ $\qquad \pi_1 = \pi_2 = 1/2$

$$M_1 = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 2 & 4 & \mu_4^{(1)} \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 2 & -4 \\ 2 & -4 & \mu_4^{(2)} \end{pmatrix}$$

$$K_1(x) - K_2(x) = \frac{1}{2}(1 \ x \ x^2)(M_1^{-1} - M_2^{-1})\begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}$$

$$= -\frac{1}{2}x\left\{4 + \left(\frac{1}{\mu_4^{(2)}-8} - \frac{1}{\mu_4^{(1)}-8}\right)(x+x^3) + \left(\frac{1}{\mu_4^{(2)}-8} + \frac{1}{\mu_4^{(1)}-8}\right)x^2\right\}$$

根 $x_0=0$ について $K_1(0)=K_2(0)=4$, $K_1'(0)=-4$, $K_2'(0)=4$ だから.

$K_1'(0)K_2'(0)=-16<0$. 故に求める v は $v=1/4$ である.

一方, $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} X \\ X^2 \end{pmatrix}$ として2変量化すると,

$$\mu_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \ \Sigma_1 = \begin{pmatrix} 1 & 2 \\ 2 & \mu_4^{(1)}-4 \end{pmatrix}, \ \mu_2 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \ \Sigma_2 = \begin{pmatrix} 1 & -2 \\ -2 & \mu_4^{(2)}-4 \end{pmatrix}$$

となり, 定理3(i)に代入して t_0 の値を求めると, $1/t_0$ の値として $1/4$ を得て, この場合は多変量化による損失はない.

② (②′) の例

$$\mu_1^{(1)}=0 \quad \mu_2^{(1)}=\sigma_1^2 \quad \mu_3^{(1)}=0 \quad \mu_4^{(1)}=3\sigma_1^4 \qquad \pi_1=\pi_2=1/2$$

$$\mu_1^{(2)}=0 \quad \mu_2^{(2)}=\sigma_2^2 \quad \mu_3^{(2)}=0 \quad \mu_4^{(2)}=3\sigma_2^4$$

$$M_i = \begin{pmatrix} 1 & 0 & \sigma_i^2 \\ 0 & \sigma_i^2 & 0 \\ \sigma_i^2 & 0 & 3\sigma_i^4 \end{pmatrix} \qquad (i=1,2)$$

方法は省略して結果だけを書くと, この場合は F_2^* が欠損のある3点分布で②′の場合になり, $\sigma_1=\lambda\sigma_2$ とすると,

$$v = \frac{1}{6\lambda^2}\left\{(1+\lambda^2) + \sqrt{\lambda^4+2\lambda^2-2}\right\}$$

を得る. たとえば $\lambda=2$ のときは $v=(5+\sqrt{22})/24 \fallingdotseq 0.40375$.

一方、多変量化による簡便法では、0.42372… になるから、簡便法による評価の相対誤差は約5% である。

より高次のモーメントまで指定されている場合も理論的には同様であるが、$\tilde{F}^*$の型による場合分けが一層複雑になり、vの値を求める計算も非常に厄介になる。従って、多変量化によって2次までのモーメントの場合に置き換える簡便法が有効であろう。その場合の評価誤差は一般的には計算できないが、一つの目安として上の例のような場合について試算した。

文献

[1] Chernoff, H. (1971), "A bound on the classification error for discriminating between populations with specified means and variances", Studi di probabilitā, statistica e ricerca operativa in onore di Giuseppe Pompilj.

[2] 西晃央, 多賀保志(1974), "判別誤確率の上限について" 日本数学会年会

[3] Isii,K., Taga,Y. (1974), "Mathematical programming approach to a minimax theorem of statistical discrimination applicable to pattern recognition". IFIP Conference of Optimization Techniques in Novosibirsk.