ROCZNIKI POLSKIEGO TOWARZYSTWA MATEMATYCZNEGO
Seria III: MATEMATYKA STOSOWANA XXVI (1985)

Ministerstwo Nauki i Szkolnictwa Wyższego

WIESŁAW PASEWICZ
Szczecin

Bayesowska metoda dyskryminacyjna w przypadku specjalnej struktury macierzy kowariancji

(Praca wpłynęła do Redakcji 1.06.1983)

1. WSTĘP

Dane są dwie populacje π_1 i π_2. W populacji π_i wektor losowy $\underline{Z}_i$ ma p-wymiarowy rozkład normalny z zerowym wektorem wartości oczekiwanych i dodatnio określoną macierzą kowariancji $\underline{\Sigma}_i$ $(i = 1, 2)$. Dana jest obserwacja $\underline{z}$, która z prawdopodobieństwem a priori q_i $(q_i > 0$, $q_1 + q_2 = 1)$ pochodzi z populacji π_i $(i = 1, 2)$. Problem polega na znalezieniu bayesowskiej metody rozstrzygającej, z której populacji pochodzi obserwacja $\underline{z}$, w przypadku gdy

$$\underline{\Sigma}_i = \sigma_i^2\left[(1 - \rho_i)\,\underline{I} + \rho_i\underline{E}\right],$$

gdzie σ_i^2 jest wspólną wariancją składowych wektora losowego $\underline{Z}_i$, ρ_i jest wspólnym współczynnikiem korelacji dla wszystkich par składowych wektora losowego $\underline{Z}_i$, $\underline{I}_{p\times p}$ jest macierzą jednostkową, a $\underline{E}_{p\times p}$ jest macierzą, której wszystkie elementy są jedynkami.

2. BAYESOWSKA METODA DYSKRYMINACYJNA

Niech $f(\underline{z}|\sigma_i,\varrho_i)$ będzie funkcją gęstości wektora losowego $\underline{Z}_i$ (i = = 1,2 . W przypadku gdy parametry σ_i, ϱ_i są znane, decydujemy, że obserwacja z pochodzi z tej populacji, dla której prawdopodobieństwo a posteriori

$$(1) \qquad P(\pi_i|\underline{z}) = q_i f(\underline{z}|\sigma_i,\varrho_i) \Big/ \sum_{j=1}^{2} q_j f(\underline{z}|\sigma_i,\varrho_i), \qquad i = 1, 2,$$

jest maksymalne. Rozwiązanie to pokrywa się z rozwiązaniem uzyskanym na drodze minimalizacji ryzyka bayesowskiego (patrz np. Anderson, 1958).

W przypadku gdy σ_i, ϱ_i nie są znane, możemy posłużyć się estymatorem funkcji gęstości $f(\underline{z}|\sigma_i,\varrho_i)$, i = 1, 2. Załóżmy, że $\underline{X}_{1i}, \ldots, \underline{X}_{N_i i}$ jest próbą losową pochodzącą z populacji π_i (i = 1, 2). Oznaczamy przez $g(\sigma_i,\varrho_i)$ gęstość a priori parametrów σ_i i ϱ_i, a przez D_i wartości obliczone na podstawie N_i-elementowej próby. Bayesowskim estymatorem funkcji $f(\underline{z}|\sigma_i,\varrho_i)$ jest wartość oczekiwana funkcji gęstości wektora $\underline{Z}_i$, względem rozkładu a posteriori parametrów σ_i, ϱ_i (patrz np. Ferguson, 1967), tj.

$$(2) \qquad f(\underline{z}|D_i) = \iint P(\sigma_i,\varrho_i|D_i)\, f(\underline{z}|\sigma_i,\varrho_i)\, d\sigma_i\, d\varrho_i,$$

gdzie

$$(3) \qquad P(\sigma_i,\varrho_i|D_i) = L(\underline{X}_{1i}, \ldots, \underline{X}_{N_i i}|\sigma_i,\varrho_i)\, g(\sigma_i,\varrho_i)$$

oznacza gęstość a posteriori parametrów σ_i, ϱ_i ($i = 1, 2$), natomiast L jest funkcją wiarogodności dla próby $\underline{X}_{1i}, \ldots, \underline{X}_{N_i i}$. Zatem reguła klasyfikacyjna (1) przyjmie teraz postać

$$(4) \qquad P(\pi_i|\underline{z}) = q_i f(\underline{z}|D_i) \Big/ \sum_{j=1}^{2} q_j f(\underline{z}|D_j), \qquad i = 1, 2.$$

Dodajmy, że dla dowolnej macierzy kowariancji (nieznanej) $\underline{\Sigma}_i$ ($i = 1, 2$) oraz równych i nieznanych wartości oczekiwanych $\mu_1 = \mu_2 = \mu$ przedstawiony problem klasyfikacji jest nierozwiązany (patrz Geisser, 1964).

3. PRZYPADEK, GDY PARAMETRY σ_i I ϱ_i SĄ ZNANE

Można wykazać (patrz Pasewicz, 1978), że funkcja gęstości wektora losowego $\underline{Z}_i$ ma postać

$$(5) \qquad f(\underline{z}|\sigma_i, \varrho_i) = (2\pi)^{-p/2} \left\{ \sigma_i^{\,p} (1-\varrho_i)^{(p-1)/2} \left[1 + (p-1)\varrho_i\right]^{1/2} \right\}^{-1} \times$$

$$\times \exp - \left\{ \frac{1}{2\sigma_i^{\,2}(1-\varrho_i)} \left[a - \frac{\varrho_i (a+b)}{1+(p-1)\varrho_i} \right] \right\}.$$

Zgodnie ze wzorem (1) decydujemy, że obserwacja $\underline{z}$ pochodzi z populacji π_1 wtedy, gdy

$$(6) \qquad q_1 f(\underline{z}|\sigma_1, \varrho_1) > q_2 f(\underline{z}|\sigma_2, \varrho_2).$$

Jeżeli

(7) $$q_1 f(\underline{z}|\sigma_1,\rho_1) < q_2 f(\underline{z}|\sigma_2,\rho_2),$$

to decydujemy, że obserwacja $\underline{z}$ pochodzi z populacji π_2. Jeżeli natomiast

(8) $$q_1 f(\underline{z}|\sigma_1,\rho_1) = q_2 f(\underline{z}|\sigma_2,\rho_2),$$

to decyzji o zaklasyfikowaniu obserwacji $\underline{z}$ nie podajemy albo decydujemy, że obserwacja $\underline{z}$ pochodzi z populacji π_1 albo z π_2, gdy istnieją inne przyczyny dyktujące taki wybór.

4. PRZYPADEK, GDY PARAMETRY σ_i, ρ_i NIE SĄ ZNANE

Załóżmy, że Z_{ijk} jest k-tą obserwacją, j-tej cechy z i-tej populacji, gdzie $k = 1, \ldots, N_i$, $j = 1, \ldots, p$, $i = 1, 2$.

Wprowadzimy następujące oznaczenia

(9) $$A_i = \sum_k \sum_j z_{ijk}^2 = \mathrm{tr} \sum_{k=1}^{N_i} \underline{Z}_{ki}\underline{Z}'_{ki} = pN_i S_i^2,$$

(10) $$A_i + B_i = \sum_k \Big(\sum_j z_{ijk}\Big)^2 = \mathrm{tr}\Big(\underline{E} \sum_{k=1}^{N_i} \underline{Z}_{ki}\underline{Z}'_{ki}\Big) = pN_i S_i^2 \big[1 + (p-1) r_i\big]$$

(11) $$\nu_i = N_i + 1, \qquad \bar{A}_i = A_i + a, \qquad \bar{B}_i = B_i + b,$$

gdzie

(12) $$a = \sum_j z_j^2 = \mathrm{tr}(\underline{z}\underline{z}'), \qquad a + b = \Big(\sum_j z_j\Big)^2 = \mathrm{tr}(\underline{E}\underline{z}\underline{z}').$$

Niech $\hat{f}_i(\underline{z})$ będzie bayesowskim estymatorem funkcji gęstości $f(\underline{z}|\sigma_i,\varrho_i)$ postaci (5). Prawdziwe jest następujące twierdzenie:

TWIERDZENIE. Jeżeli (Geisser i Desu, 1968) łączna gęstość rozkładu a priori parametrów σ_i, ϱ_i jest postaci

$$g(\sigma_i,\varrho_i) \propto \left\{\sigma_i(1-\varrho_i)\left[1+(p-1)\varrho_i\right]\right\}^{-1}, \tag{13}$$

to

$$\hat{f}_i(\underline{z}) = \left(\frac{p}{\pi}\right)^{p/2} \frac{\Gamma\left(\frac{\nu_i}{2}\right)\Gamma\left[\frac{(p-1)\nu_i}{2}\right]}{\Gamma\left(\frac{N_i}{2}\right)\Gamma\left[\frac{(p-1)N_i}{2}\right]} \frac{(A_i+B_i)^{N_i/2}\left[(p-1)A_i-B_i\right]^{\frac{1}{2}(p-1)N_i}}{(\bar{A}_i+\bar{B}_i)^{\nu_i/2}\left[(p-1)\bar{A}_i-\bar{B}_i\right]^{\frac{1}{2}(p-1)\nu_i}}. \tag{14}$$

D o w ó d. Istotnie. Na mocy równości (3) mamy

$$P(\sigma_i,\varrho_i|A_i,B_i) = L(A_i, B_i|\sigma_i,\varrho_i)\, g(\sigma_i,\varrho_i), \tag{15}$$

gdzie

$$L(A_i,B_i|\sigma_i,\varrho_i) = (2\pi)^{-pN_i/2}\left[\sigma_i^2(1-\varrho_i)\right]^{-\frac{1}{2}N_i(p-1)} \times \tag{16}$$

$$\times \left\{\sigma_i^2\left[1+(p-1)\varrho_i\right]\right\}^{-\frac{1}{2}N_i} \exp\left\{-\frac{1}{2p}\left(\frac{(p-1)A_i-B_i}{\sigma_i^2(1-\varrho_i)} + \frac{A_i+B_i}{\sigma_i^2\left[1+(p-1)\varrho_i\right]}\right)\right\}.$$

Dokonując podstawienia

$$(17)\qquad \begin{cases} [1+(p-1)\varrho_i]\sigma_i^2 = \alpha_i, \\ (1-\varrho_i)\sigma_i^2 = \theta_i, \end{cases}$$

równość (15) można zapisać w postaci

$$(18)\qquad P(\alpha_i,\theta_i|\,A_i,B_i) \propto \alpha_i^{-\frac{1}{2}N_i-1}\,\theta_i^{-\frac{1}{2}(p-1)N_i-1}\,\times$$

$$\times\,\exp\left\{-\frac{1}{2p}\left[\frac{A_i+B_i}{\alpha_i}+\frac{(p-1)A_i-B_i}{\theta_i}\right]\right\}.$$

Ponieważ powyższa funkcja jest proporcjonalna (zarówno ze względu na zmienną α_i, jak i ze względu na θ_i) do gęstości uogólnionego rozkładu gamma, więc

$$(19)\qquad P(\alpha_i,\theta_i|\,A_i,B_i) = (2p)^{-pN_i/2}\,\frac{(A_i+B_i)^{\frac{1}{2}N_i}}{\alpha_i^{\frac{1}{2}N_i+1}\,\Gamma\left(\frac{N_i}{2}\right)}\,\times$$

$$\times\,\frac{[(p-1)A_i-B_i]^{\frac{1}{2}(p-1)N_i}}{\theta_i^{\frac{1}{2}(p-1)N_i+1}\,\Gamma\left[\frac{1}{2}(p-1)N_i\right]}\,\exp\left\{-\frac{1}{2p}\left[\frac{A_i+B_i}{\alpha_i}+\frac{(p-1)A_i-B_i}{\theta_i}\right]\right\}.$$

Funkcję $\hat{f}_i(\underline{z})$ obliczymy jako wartość oczekiwaną funkcji $f(\underline{z}|\sigma_i,\varrho_i) = f(\underline{z}|\alpha_i,\theta_i)$ względem rozkładu a posteriori (19), tzn.

$$(20) \qquad \hat{f}_i(\underline{z}) = \int_0^\infty \int_0^\infty f(\underline{z}|\alpha_i, \theta_i)\, P(\alpha_i, \theta_i | A_i, B_i)\, d\alpha_i\, d\theta_i .$$

Wykorzystując niezależność zmiennych α_i i θ_i (prawa strona równości (19) jest iloczynem funkcji gęstości zmiennych losowych α_i i θ_i) oraz dwukrotnie uwzględniając gęstość uogólnionego rozkładu gamma w prawej stronie równości (20), otrzymamy tezę twierdzenia.

Zaklasyfikowanie obserwacji $\underline{z}$ opieramy teraz na związkach (6), (7), (8), gdzie w miejsce funkcji $f(\underline{z}|\sigma_i, \varrho_i)$ podstawiamy funkcję $\hat{f}_i(\underline{z})$. Jeżeli założymy, że $N_1 = N_2 = N$, to

$$(21) \qquad \ln \frac{\hat{f}_1(\underline{z})}{\hat{f}_2(\underline{z})} = N \ln \frac{A_1+B_1}{A_2+B_2} + N(p-1) \ln \frac{(p-1)A_1-B_1}{(p-1)A_2-B_2} +$$

$$+ (N+1) \ln \frac{\bar{A}_2+\bar{B}_2}{\bar{A}_1+\bar{B}_1} + (N+1)(p-1) \ln \frac{(p-1)\bar{A}_2-\bar{B}_2}{(p-1)\bar{A}_1-\bar{B}_1}$$

i obserwację $\underline{z}$ zaklasyfikujemy do populacji π_1 wtedy, gdy

$$(22) \qquad \ln \frac{\hat{f}_1(\underline{z})}{\hat{f}_2(\underline{z})} > c,$$

gdzie $c = \ln \frac{q_2}{q_1}$.

Jeśli nierówność (22) nie będzie zachodziła, to podejmujemy decyzję podobną do decyzji opartej na związku (7) albo (8).

LITERATURA CYTOWANA

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, New York 1958.

[2] T.S. Ferguson, Mathematical Statistics - A Decision Theoretic Approach, Academic Press, New York, London 1967.

[3] S. Geisser, Posterior odds for multivariate normal classifications, J.R. Statist. Soc. B 26 (1964), 69-76.

[4] S. Geisser, M.M. Desu, Predictive zero-mean uniform discrimination, Biometrika 55 (1968), 519-524.

[5] W. Pasewicz, Klasyfikacja wielowymiarowych obserwacji w przypadku równych wektorów średnich, Ósme Colloqium Metodologiczne z Agro-Biometrii, PAN 1978.