Wiesław P a s e w i c z (Szczecin)

# ANALIZA DYSKRYMINACYJNA POPULACJI NORMALNYCH PRZY NIEZNANYCH PARAMETRACH I NIEZNANYCH PRAWDOPODOBIEŃSTWACH A PRIORI

## 1. Wstęp

Załóżmy, że w populacji $\pi_i$ obserwowany wektor losowy $\underline{Z}_i$ ma p-wymiarowy rozkład normalny z parametrami $\underline{\mu}_i, \underline{\Sigma}_i$, dla $i = 1,\ldots,k$. Dana jest obserwacja $\underline{z}$, o której wiemy, że jest obserwacją pochodzącą z jednej z $k$ populacji $\pi_1,\ldots,\pi_k$, lecz nie wiemy z której.

Niech $q_i > 0$ $(q_1 +\ldots+ q_k = 1)$ oznacza prawdopodobieństwo a priori zdarzenia, że obserwacja $\underline{z}$ pochodzi z populacji $\pi_i$, natomiast $f(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i)$ niech oznacza funkcję gęstości wektora losowego $\underline{Z}_i$, $i = 1,\ldots,k$.

W przypadku, gdy prawdopodobieństwa $q_i$ oraz funkcje $f(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i)$ są znane, to decydujemy, że obserwacja $\underline{z}$ pochodzi z tej populacji, dla której prawdopodobieństwo a posteriori

$$(1) \qquad P(\pi_i \; \underline{z}) = \frac{q_i \; f(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i)}{\sum\limits_{j=1}^{k} q_j \; f(\underline{z}|\underline{\mu}_j,\underline{\Sigma}_j)}$$

jest maksymalne.

Rozwiązanie to pokrywa się z rozwiązaniem uzyskanym na drodze minimalizacji ryzyka bayesowskiego (patrz np. Anderson [1958], rozdz.6).

Celem tej pracy jest podanie oszacowań prawdopodobieństw a posteriori $P(\pi_i|\underline{z})$ w przypadku, gdy prawdopodobieństwa $q_i$ oraz funkcje $f(\underline{z}|\underline{\pi}_i,\underline{\Sigma}_i)$ nie są znane, $i = 1,\ldots,k$.

## 2. Metoda

Niech $\underline{\bar{z}}_i$ oraz $\underline{S}_i$ oznaczają zwykłe estymatory parametrów $\underline{\pi}_i$, $\underline{\Sigma}_i$ uzyskane z próby o liczebności $N_i$, $i = 1,\ldots,k$. Posłużymy się bayesowskim estymatorem nieznanej funkcji gęstości $f(\underline{z}|\underline{\pi}_i,\underline{\Sigma}_i)$, $i = 1,\ldots,k$. Wiadomo, że gdy posługujemy się kwadratową funkcją straty, to bayesowski estymator funkcji gęstości $f(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i)$ jest równy wartości oczekiwanej tej funkcji względem rozkładu a posteriori parametrów w niej występujących. Oznaczmy ten estymator przez $h(\underline{z}|\underline{\bar{z}}_i,\underline{S}_i)$, $i = 1,\ldots,k$.

Mamy

$$(2) \qquad h(\underline{z}|\underline{\bar{z}}_i,\underline{S}_i) = \iint f(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i)\, t(\underline{\mu}_i,\underline{\Sigma}_i|\underline{\bar{z}}_i,\underline{S}_i)\, d\underline{\mu}_i\, d\underline{\Sigma}_i,$$

gdzie $t(\underline{\mu}_i,\underline{\Sigma}_i|\underline{\bar{z}}_i,\underline{S}_i)$ jest funkcją gęstości rozkładu a poste--riori parametrów $\underline{\mu}_i,\underline{\Sigma}_i$, $i = 1,\ldots,k$.

Przyjmując, że funkcja gęstości łącznego rozkładu a priori parametrów $\underline{\mu}_i,\underline{\Sigma}_i$ jest funkcją Jeffreys'a [1961] postaci

$$(3) \qquad g(\underline{\mu}_i,\underline{\Sigma}_i^{-1}) \propto |\underline{\Sigma}_i|^{\frac{p+1}{2}} \qquad (\propto \text{ oznacza proporcjonalność})$$

otrzymamy (Geisser, [1964])

$$(4) \qquad h(\underline{z}|\underline{\bar{z}}_i,\underline{S}_i) = c_i[1+N_i(N_i^2 - 1)^{-1}(\underline{z}-\underline{\bar{z}}_i)\underline{S}_i^{-1}(\underline{z}-\underline{\bar{z}}_i)]^{-\frac{N_i}{2}}$$

gdzie

$$(5)\quad c_i = [\pi^{-1}N_i(N_i+1)^{-1}]^{\frac{p}{2}}\Gamma(N_i/2)\Gamma^{-1}[(N_i-p)/2]\,|(N_i-1)\underline{S}_i|^{-\frac{1}{2}}$$

$i = 1,\ldots,k$.

Funkcja $h(\underline{z}|\bar{\underline{z}}_i,\underline{S}_i)$ dana wzorem (4) jest funkcją gęstości p-wymiarowego rozkładu t (Cornish, [1954]).

Zajmiemy się teraz nieznanym prawdopodobieństwem a priori $q_i$, $i = 1,\ldots,k$.

Oznaczmy przez $A_i$ zdarzenie, że losowo wybrana obserwacja z $N$ elementowej próby pobranej z populacji $\pi_1,\ldots,\pi_k$ jest obserwacją pochodzącą z populacji $\pi_i$, $i = 1,\ldots,k$.

Załóżmy, że w $N$-elementowej próbie wykluczające się zdarzenia $A_1,A_2,\ldots,A_k$ wystąpiły odpowiednio $N_1,N_2,\ldots,N_k$ razy. Załóżmy dalej, że nieznane prawdopodobieństwo $q_i$ jest estymowane przez $N_i/N$ gdzie $N = \sum\limits_i N_i$ oraz $P(A_i) = q_i$.

Zatem prawdopodobieństwo, że zdarzenie $A_i$ wystąpi $N_i$ razy wynosi $q_i^{N_i}$ i łączne prawdopodobieństwo, że zdarzenie $A_1$ wystąpi $N_1$ razy, $A_2$ wystąpi $N_2$ razy, itd., $A_k$ wystąpi $N_k$ razy wynosi $q_1^{N_1}\cdot q_2^{N_2}\cdot\ldots\cdot q_k^{N_k}$.

Ponieważ liczba takich zdarzeń jest równa $\frac{N!}{N_1!N_2!\ldots N_k!}$, to łączny rozkład wektora losowego $N_1,N_2,\ldots,N_{k-1}$ jest $(k-1)$ - wymiarowym rozkładem wielomianowym, którego funkcja masy jest postaci:

$$(6)\quad p(N_1,N_2,\ldots,N_{k-1}|q_1,q_2,\ldots,q_{k-1}) = \frac{N!}{\prod\limits_{j=1}^{k} N_j!}\prod_{j=1}^{k} q_j^{N_j},$$

gdzie $N_k = N - N_1 - \ldots - N_{k-1}$, oraz $q_k = 1 - q_1 - \ldots - q_{k-1}$.

Stąd funkcję wiarygodności możemy zapisać następująco:

$$L(q_1,q_2,\ldots,q_{k-1}) \propto \prod_{j=1}^{k} q_j^{N_j} . \tag{7}$$

Można dowieść (patrz np. Wilks, [1962]), że zachodzi następująca równość:

$$\int_S x_1^{\nu_1-1} x_2^{\nu_2-1} \ldots x_k^{\nu_k-1} (1-x_1-\ldots-x_k)^{\nu_{k+1}-1} dx_1 dx_2 \ldots dx_k = \tag{8}$$

$$= \frac{\prod_{j=1}^{k+1} \Gamma(\nu_j)}{\Gamma\left(\sum_{j=1}^{k+1} \nu_j\right)} ,$$

gdzie $S$ jest simplexem postaci:

$$(x_1,x_2,\ldots,x_k):\ x_i \geq 0,\quad i = 1,2,\ldots,k,\quad \sum_{i=1}^{k} x_i \leq 1$$

oraz $\nu_i > 0$, dla $i = 1,2,\ldots,k$.

Całka występująca w (8) nazywa się całką Dirichleta i jest ona uogólnieniem funkcji beta.

W związku z tym funkcja gęstości postaci:

$$f(x_1,x_2,\ldots,x_k) = \tag{9}$$

$$= \begin{cases} \dfrac{\Gamma(\sum_j \nu_j)}{\prod_j \Gamma(\nu_j)} x_1^{\nu_1-1} \ldots x_k^{\nu_k-1} (1-x_1-\ldots-x_k)^{\nu_{k+1}-1}, & \\ & \text{dla } (x_1,\ldots,x_k) \in S \\ 0 & \text{dla } (x_1,\ldots,x_k) \notin S \end{cases}$$

jest nazywana funkcją gęstości k-wymiarowego rozkładu Dirichleta. Załóżmy teraz, że funkcja gęstości prawdopodobieństwa a priori $q$ łącznego rozkładu zmiennych losowych $q_1,\ldots,q_{k-1}$ jest funkcją gęstości rozkładu Dirichleta postaci:

$$(10)\qquad g(q_1,q_2,\ldots,q_{k-1}) \propto \prod_{j=1}^{k} q_j^{\alpha_j} \qquad (\alpha_j > 0,\ j=1,2,\ldots,k)$$

przy czym równość (10) wystąpi wtedy, gdy jej prawą stronę pomnożymy przez $\Gamma[\sum_j(\alpha_j+1)]/\prod_j\Gamma(\alpha_j+1)$.

Korzystając z (7) i (10) otrzymamy gęstość a posteriori łącznego rozkładz zmiennych $q_1,\ldots,q_{k-1}$ :

$$(11)\qquad P(q_1,q_2,\ldots,q_{k-1}|N_1,N_2,\ldots,N_{k-1}) \propto \prod_{j=1}^{k} q_j^{\alpha_j+N_j},$$

przy czym równość w (11) zachodzi wtedy gdy prawą stronę (11) pomnożymy przez $\Gamma[\sum_j (N_j+\alpha_j+1)]/\prod_j \Gamma(N_j+\alpha_j+1)$.

Ponieważ prawdopodobieństwo a posteriori zdarzenia że obserwacja $\underline{z}$ pochodzi z populacji $\pi_i$ zależy od $q_i$, to w celu znalezienia bezwarunkowego prawdopodobieństwa a posteriori $P(\pi_i|\underline{z})$ musimy zmodyfikować gęstość (11) przez uzależnienie jej od $\underline{z}$.

Lemat.

Niech $A,B,C$ będą zdarzeniami spełniającymi następujące warunki:

(i) $A \cap C = A$

(ii) $P(B \cap C) = P(B)\,P(C)$.

Zachodzi następujący związek

(12) $P(A|\,B\cap C) \propto P(A\cap C)\ P(B|\,A)$.

Dowód.

Ponieważ zachodzi równość

$P(A \cap B \cap C) = P(B|A \cap C)\; P(A \cap C)$, to na mocy (i) mamy

$P(A \cap B \cap C) = P(B|A)\; P(A \cap C)$.

Dzieląc obie strony ostatniej równości przez $P(B \cap C)$ i korzystając z (ii) mamy

$$\frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(B|A)\; P(A \cap C)}{P(B)\; P(C)} .$$

Ostatecznie więc

$$P(A|B \cap C) = \frac{1}{P(B)} P(B|A)\; P(A|C) \propto P(B|A)\; P(A|C), \quad \text{c.n.d.}$$

Z (12) wynika, że

$$(13) \quad P(q_1,q_2,\ldots,q_{k-1}|\underline{z},N_1,N_2,\ldots,N_{k-1}) =$$

$$P(q_1,q_2,\ldots,q_{k-1}|N_1,N_2,\ldots,N_{k-1})\; h(\underline{z}|q_1,q_2,\ldots,q_{k-1})\; \frac{1}{h(\underline{z})},$$

gdzie: $h(\underline{z}|q_1,q_2,\ldots,q_{k-1}) = \sum_{j=1}^{k} q_j\; h(\underline{z}|\underline{\mu}_j,\underline{\Sigma}_j)$,

$$h(\underline{z}) = \sum_{j=1}^{k} (N_j + \alpha_j+1)\; h(\underline{z}|\underline{\mu}_j,\underline{\Sigma}_j) / \sum_{j=1}^{k} (N_j+\alpha_j+1).$$

Zatem:

$$(14) \quad P(q_1,q_2,\ldots,q_{k-1}|\underline{z},N_1,N_2,\ldots,N_{k-1}) \propto \prod_j q_j^{N_j+\alpha_j} \sum_j q_j\; h(\underline{z}|\underline{\mu}_j,\underline{\Sigma}_j)$$

oraz

$$(15) \quad \int\ldots\int P(q_1,\ldots,q_{k-1}|\underline{z},N_1,\ldots,N_{k-1})\, P(\pi_i|\underline{z},\underline{q})\, dq_1\ldots dq_{k-1} =$$

$$P(\pi_i|\underline{z}) \propto \int\ldots\int \prod_j q_j^{N_j+\alpha_j} \sum_j q_j h(\underline{z}|\underline{\mu}_j,\underline{\Sigma}_j)\, P(\pi_i|\underline{z},q)\, dq_1\ldots dq_{k-1},$$

gdzie $\quad P(\pi_i|\underline{z},q) = q_i h(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i) / \sum_j q_j h(\underline{z}|\underline{\mu}_j,\underline{\Sigma}_j).$

Dalej mamy

$$P(\pi_i|\underline{z}) = \frac{\Gamma[\sum_j (N_j+\alpha_j+1)]}{\prod_j \Gamma(N_j+\alpha_j+1)} \frac{h(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i)}{h(\underline{z})} \int\ldots\int \prod_j q_j^{N_j+\alpha_j} q_i dq_1\ldots dq_{k-1} =$$

$$= \frac{\Gamma[\sum_j (N_j+\alpha_j+1)]}{\prod_j \Gamma(N_j+\alpha_j+1)} \frac{h(\underline{z}|\underline{\mu}_j,\underline{\Sigma}_j)}{h(\underline{z})} \frac{\Gamma(N_i+\alpha_i+2) \prod_{\substack{j=1 \\ j\neq i}}^{k} \Gamma(\alpha_j+N_j+1)}{\Gamma[\sum_{j=1}^{k} (\alpha_j+N_j+1)+1]} =$$

$$= \frac{\Gamma[\sum_j (N_j+\alpha_j+1)] h(\underline{z}|\mu_i,\Sigma_i)}{\prod_j \Gamma(N_j+\alpha_j+1)\, h(\underline{z})} \frac{(\alpha_i+N_i+1)\Gamma(N_i+\alpha_i+1) \prod_{j\neq i} \Gamma(\alpha_j+N_j+1)}{[1+\Sigma_j(\alpha_j+N_j+1)]} =$$

$$= \frac{h(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i)}{h(\underline{z})} \frac{\alpha_i + N_i + 1}{\Sigma_j \ (\alpha_j+N_j+1)} .$$

Tak więc

$$(16) \qquad P(\pi_i|\underline{z}) = \frac{(N_i + \alpha_i + 1)\ h(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i)}{\sum_j (N_j+\alpha_j+1)h(\underline{z}|\underline{\mu}_j,\underline{\Sigma}_j)}, \quad i=1,2,\ldots,k.$$

Równość (16) przedstawia poszukiwane prawdopodobieństwo a posteriori zaklasyfikowania obserwacji $\underline{z}$ do populacji $\pi_i$, gdy nieznane są prawdopodobieństwa a priori $q_i$, że obserwacja $\underline{z}$ pochodzi z populacji $\pi_i$, oraz gdy nieznane są parametry $\underline{\mu}_i$, $\underline{\Sigma}_i$, $i = 1,2,\ldots,k$.

<u>Uwaga.</u>

Wielkość $\alpha_i$ należy interpretować jako częściową informację o wartości $q_i$, jeżeli $q_i$ nie jest znane, $i = 1,\ldots,k$.

Jeżeli $\alpha_1 = \ldots = \alpha_k = \alpha$ oraz $N_1 = \ldots = N_k = M$, to równość (16) przyjmuje postać

$$P(\pi_i|\underline{z}) = \frac{h(\underline{z}|\underline{\mu}_i,\underline{\Sigma}_i)}{\sum_j h(\underline{z}|\underline{\mu}_j,\underline{\Sigma}_j)}, \qquad i = 1,\ldots,k.$$

Zauważmy, że przypadek ten odpowiada sytuacji, kiedy $q_i$ nie są znane i kiedy zakłada się, że wszystkie $q_i$ są sobie równe tzn. $q_1 = q_2 = \ldots = q_k = \frac{1}{k}$.

## Literatura cytowana

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

Cornish, E.A., 1954, The multivariate t - distribution associated with a set of normal sample deviates, Australian Journal of Physics, 7, str.531 - 542.

Jeffreys, H., 1961, Theory of Probability, 3rd ed., Oxford University Press, Oxford.

Geisser, S., 1963, Posterior Odds for Multivariate Normal Classifications, Journal of the Royal Statistical Society, str. 69 - 76.

Wilks, S.S., 1962, Mathematical Statistics, Wiley, New York.

## DISCRIMINANT ANALYSIS OF NORMAL POPULATIONS WITH UNKNOWN PARAMETERS AND UNKNOWN PRIOR PROBABILITIES


### Summary

Let us assume that the observed random vector $\underline{Z}_i=(z_1,\ldots,z_p)'$ from population $\pi_i$ has a p-dimensional normal distribution with a mean vector $\underline{\mu}_i$ and a positive definite covariance matrix $\underline{\Sigma}_i$, $i = 1,2,\ldots,k$. A multivariate observation $\underline{z}$ is known and it belongs to one of $k$ multivariate normal populations $\pi_i$, but it is not known to which.

If the prior probability that $\underline{z}$ belongs to $\pi_i$ is $q_i$, where $\Sigma q_i = 1$, and the parameters of $\pi_i$ are known, it is relatively easy to compute the likelihood that $\underline{z}$ belongs to $\pi_i$ and then to combine this, via Bayes's theorem, with the prior probability in order to obtain the posterior probability.

In the present paper we consider the discrimination between $\pi_i$'s when the $q_i$'s and the parameters of $\pi_i$'s are a priori unknown.