

REPRESENTATION OF STATISTICS OF DISCRIMINANT ANALYSIS AND ASYMPTOTIC EXPANSIONS WHEN THE DIMENSIONS OF THE SPACE ARE COMPARABLE WITH THE SAMPLE SIZES

A. D. Deev

1970

SovietRxiv

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

UDC 519.281

MATHEMATICS

A. D. DEEV

REPRESENTATION OF STATISTICS OF DISCRIMINANT ANALYSIS AND ASYMPTOTIC EXPANSIONS WHEN THE DIMENSIONS OF THE SPACE ARE COMPARABLE WITH THE SAMPLE SIZES

(Presented by Academician A. N. Kolmogorov on 20 IV 1970)

1. As is well known, when classifying an observation X into one of several (S) populations, information about which is given by a preliminary sample, one should distinguish three kinds of errors: a) the error $P(H_i | H_j)$, unknown to the investigator, which is obtained for completely specified distributions because of the statistical nature of the problem (the error in assigning X to the i -th population when in fact X belongs to the j -th ($i \neq j$)); b) the error due to the one-time construction of the classifier from a fixed sample, i.e., the conditional error $P(H_i | H_j, \{X_\alpha^{(k)}\})$, $\alpha = 1, \dots, N_k$, $k = 1, \dots, S$, with i, j fixed. The conditional error is a random variable whose distribution it is desirable to estimate, and its first characteristic is its average over all possible samples of fixed size $N = (N_1, N_2, \dots, N_s)$; c) $P_{N;ij} = MP(H_i | H_j, \{X_\alpha^{(k)}\})$. Clearly, for any linear functional of the errors $P(H_i | H_j)$, $L = \sum C_{ij} P(H_i | H_j)$, the minimum will be less than the minimum $L_N = \sum C_{ij} MP(H_i | H_j, \{X_\alpha^{(k)}\})$.

The study of distributions connected with discriminant functions is a difficult problem, very far from a satisfactory solution. In the present note a representation is proposed for certain classification statistics in terms of simple one-dimensional random variables (normal and χ^2), which makes it possible to obtain asymptotic expansions useful for practical purposes.

2. We shall consider the problem of classifying an observation X into one of two normal populations $\pi_i \sim N(\mu_i, \Sigma)$ ($i = 1, 2$) with common covariance matrix Σ . By sufficiency considerations the sample information $\{X_\alpha^{(i)}\}$, $\alpha = 1, \dots, N_i$, is reduced to $\mathfrak{M} = \{\bar{X}^{(1)}, \bar{X}^{(2)}, A\}$, where

$$\bar{X}^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} X_{\alpha}^{(i)}$$

are estimates of the population means μ_i , and

$$S = \frac{1}{f} A = \frac{1}{N_1 + N_2 - 2} A = \frac{1}{N_1 + N_2 - 2} \sum_{i=1}^2 \sum_{\alpha=1}^{N_i} (X_{\alpha}^{(i)} - \bar{X}^{(i)}) (X_{\alpha}^{(i)} - \bar{X}^{(i)})'$$

is an unbiased estimate of the common covariance matrix Σ . Let $W = (X - \frac{1}{2}\bar{X}^{(1)} - \frac{1}{2}\bar{X}^{(2)})' S^{-1} (\bar{X}^{(2)} - \bar{X}^{(1)})$ be the so-called Anderson statistic, usually used for classifying X according to the rule: accept $X \in \pi_1$ if $W < C$ and $X \in \pi_2$ if $W > C$, where C is the classification threshold chosen by the investigator; usually $C = 0$, as we shall also assume, although this is inessential.

The distribution of W is complicated and has been studied by many authors (¹⁻³). In particular, one should note the work (⁴), in which for the first time a representation of W in terms of simple statistics was given, useful for simulation, numerical analysis, obtaining asymptotic expansions, etc.

In the parametric space there exists only one parameter—the so-called Mahalanobis distance $\rho^2 = (\mu_2 - \mu_1)' \Sigma^{-1} (\mu_2 - \mu_1)$, and the errors are functions of ρ ; for example, they can be made equal: $P(H_1 | H_2) = P(H_2 | H_1) = \Phi(-\rho/2)$, where $\Phi(\cdot)$ is the distribution function of a standard normal variate $N(0, 1)$.

Owing to the conditional normality of W with respect to $\mathfrak{M}$,

$$P(W > 0 | \mathfrak{M}, H_1) = \Phi \left(\frac{(\mu_1 - \frac{1}{2}\bar{X}^{(1)} - \frac{1}{2}\bar{X}^{(2)})' S^{-1} (\bar{X}^{(2)} - \bar{X}^{(1)})}{[(\bar{X}^{(2)} - \bar{X}^{(1)})' S^{-1} \Sigma S^{-1} (\bar{X}^{(2)} - \bar{X}^{(1)})]^{1/2}} \right),$$

$$P(W < 0 | \mathfrak{M}, H_2) = \Phi \left(\frac{(\frac{1}{2}\bar{X}^{(1)} + \frac{1}{2}\bar{X}^{(2)} - \mu_2)' S^{-1} (\bar{X}^{(2)} - \bar{X}^{(1)})}{[(\bar{X}^{(2)} - \bar{X}^{(1)})' S^{-1} \Sigma S^{-1} (\bar{X}^{(2)} - \bar{X}^{(1)})]^{1/2}} \right).$$

Let us denote the arguments by ξ_1 and ξ_2 . We express ξ_1 and ξ_2 through six simple random variables. By invariance considerations, we shall take the matrix Σ to be the identity matrix I_p . Put $Z_1 = f_1^{-1/2} (\bar{X}^{(2)} - \bar{X}^{(1)})$ and $Z_2 = (N_1 \bar{X}^{(1)} + N_2 \bar{X}^{(2)}) / (N_1 + N_2)^{1/2}$, $f_1 = (N_1 + N_2) / N_1 N_2$; then Z_1 and Z_2 are independent and have distributions $Z_1 \sim N_1 (f_1^{-1/2} \Delta \mu, I_p)$, $Z_2 \sim N((N_1 \mu_1 + N_2 \mu_2) / (N_1 +$

$N_2)^{1/2}, I_p)$ ($\Delta\mu = \mu_2 - \mu_1$). In this notation ξ_1 and ξ_2 are represented in the following form:

$$\xi_1 = \left[\frac{1}{(N_1 + N_2)^{1/2}} (MZ_2 - Z_2)' A^{-1} Z_1 - \frac{N_2}{N_1 + N_2} \Delta\mu' A^{-1} Z_1 + \frac{N_2 - N_1}{2(N_1 + N_2)} f_1^{1/2} Z_1' A^{-1} Z_1 \right] / (Z_1' A^{-2} Z_1)^{1/2},$$

$$\xi_2 = \left[-\frac{1}{(N_1 + N_2)^{1/2}} (MZ_2 - Z_2)' A^{-1} Z_1 - \frac{N_1}{N_1 + N_2} \Delta\mu' A^{-1} Z_1 - \frac{N_2 - N_1}{2(N_1 + N_2)} f_1^{1/2} Z_1' A^{-1} Z_1 \right] / (Z_1' A^{-2} Z_1)^{1/2}.$$

The quantity

$$\nu_2 = \frac{(Z_2 - MZ_2)' A^{-1} Z_1}{(Z_1' A^{-2} Z_1)^{1/2}}$$

has a standard normal distribution and does not depend on A and Z_1 ; as for the joint distribution of $\beta_1 = \Delta\mu' A^{-1} Z_1$, $\beta_2 = Z_1' A^{-1} Z_1$, and $\beta_3 = Z_1' A^{-2} Z_1$, the following representation is valid:

Theorem 1. If $Z \sim N(\Delta\mu, I_p)$ and $A \sim W(I_p, f)$, where Z and A are independent, and $W(I_p, f)$ denotes the Wishart distribution in p -dimensional space with f degrees of freedom, then the statistics $(\beta_1, \beta_2, \beta_3)$ are expressed in terms of 5 independent random variables:

$$\beta_1 = (\chi_{f-p+1}^2)^{-1} |\Delta\mu| [\nu_1 + |\Delta\mu| + R \sin \theta \sqrt{\chi_{p-1}^2}],$$

$$\beta_2 = (\chi_{f-p+1}^2)^{-1} [(\nu_1 + |\Delta\mu|)^2 + \chi_{p-1}^2],$$

$$\beta_3 = (\chi_{f-p+1}^2)^{-2} [(\nu_1 + |\Delta\mu|)^2 + \chi_{p-1}^2] (1 + R^2),$$

$\nu_1 \sim N(0, 1)$; χ_{f-p+1}^2 and χ_{p-1}^2 are χ^2 -variates; R^2 has a beta distribution of the second kind (see (6)) with

$$p(R^2) = \frac{\Gamma((f+1)/2)}{\Gamma((f-p+2)/2)\Gamma((p-1)/2)} \frac{(R^2)^{(p-3)/2}}{(1+R^2)^{(f+1)/2}}$$

and

$$p(\theta) = \frac{\Gamma((p-1)/2)}{\Gamma(1/2)\Gamma((p-2)/2)} \cos^{p-1} \theta.$$

We note that the beta distribution of the second kind is the ratio of independent χ^2 -variates: $R^2 = \chi_{p-1}^2 / \chi_{f-p+2}^2$.

The proof is based on the standard technique of a random orthogonal transformation independent of A (see (4)), and on the lemma concerning the distribution of the first row of the matrix A^{-1} .

Lemma. Let $A = \{a_{ij}\} \sim W(I_p, f)$ and $A^{-1} = \{a^{ij}\}$. Then a^{11} and $g' = \left(\frac{a^{12}}{a^{11}}, \frac{a^{13}}{a^{11}}, \dots, \frac{a^{1p}}{a^{11}}\right)$ are independent and have distributions $a^{11} \sim 1/\chi_{f-p+1}^2$,

$$p(g) = \frac{\Gamma((f+1)/2)}{\pi^{(p-1)/2} \Gamma((f-p+2)/2)} (1 + g'g)^{-(f+1)/2}.$$

The proof of the lemma is based on the rule for multiplying block-partitioned matrices and on Theorem 4.3.2 of [5]. Namely, if

$$A = \begin{pmatrix} a_{11} & a'_{(1)} \\ a_{(1)} & A_{22} \end{pmatrix} \quad \text{and} \quad A^{-1} = \begin{pmatrix} a^{11} & a^{(1)'} \\ a^{(1)} & A^{22} \end{pmatrix},$$

then

$$a^{(1)'} = -a^{11} a'_{(1)} A_{22}^{-1}$$

(the quantity $a'_{(1)} A_{22}^{-1}$ is the row of coefficients of the formal sample regression of the first coordinate on the remaining ones), whence by Theorem 4.3.2 of [5] a^{11} and $g' = a^{(1)'} / a^{11}$ are independent. The distribution of a^{11} has been indicated; the conditional distribution of g' (when the coordinates except the first are fixed) is normal with zero mean and covariance matrix A_{22}^{-1} , where A_{22} has distribution $W(I_{p-1}, f)$. Integrating the joint density of g' and A_{22} over the set of positive definite A_{22} (see [5], Ch. 7), we obtain $p(g)$, which completes the proof of the lemma.

In order not to encumber the exposition, we give the final form of the representation of the conditional errors only for $N_1 = N_2 = N$:

$$P\{W > 0 \mid \mathfrak{M}, H_1\} = \Phi \left(\frac{\nu_2}{(2N)^{1/2}} - \frac{1}{2} - \frac{\rho(\nu_1 + \sqrt{N/2} \rho + R \sin \theta \sqrt{\chi_{p-1}^2})}{(1 + R^2)^{1/2} [(\nu_1 + \sqrt{N/2} \rho)^2 + \chi_{p-1}^2]^{1/2}} \right),$$

$$P\{W < 0 \mid \mathfrak{M}, H_2\} = \Phi \left(-\frac{\nu_2}{(2N)^{1/2}} - \frac{1}{2} - \frac{\rho(\nu_1 + \sqrt{N/2} \rho + R \sin \theta \sqrt{\chi_{p-1}^2})}{(1 + R^2)^{1/2} [(\nu_1 + \sqrt{N/2} \rho)^2 + \chi_{p-1}^2]^{1/2}} \right).$$

A. N. Kolmogorov proposed studying the behavior of the distribution of classifiers when $p/N_i \rightarrow \lambda_i$ ($p \rightarrow \infty$, $N_i \rightarrow \infty$) and gave a first approximation to the classification errors by means of a rule based on two statistics:

$$\Delta_i = (X - \bar{X}^{(1)})' S^{-1} (\bar{X}^{(2)} - \bar{X}^{(1)}), \quad \Delta_1 + \Delta_2 = r^2 = (\bar{X}^{(2)} - \bar{X}^{(1)})' S^{-1} (\bar{X}^{(2)} - \bar{X}^{(1)}),$$

and allowing, generally speaking, a rejection region for classification. The consideration of this problem is the subject of a separate paper; here we give the principal and first term of the expansion of the distribution of W in the indicated asymptotics. A similar expansion in the usual asymptotics (p fixed, $N_i \rightarrow \infty$) was obtained in [7], but already for moderate values of p the corrections are comparable with the principal term (see Table 2, p. 1292 of [7]); therefore it seems advisable to us to expand for $p/N_i \rightarrow \lambda_i = \text{const}$. The expansion is obtained by synthesizing the ideas of preliminary averaging with respect to the normal measure and representation through simple statistics, followed by application of Laplace's method to the conditional characteristic function.

Let

$$G_i(t) = P \left\{ W < \frac{f}{f-p+1} (-1)^i \left[\frac{\rho^2}{2} + \frac{(N_2 - N_1)(p-1)}{2N_1 N_2} \right] + tD \mid H_i \right\},$$

where

$$D^2 = \frac{f^2(f+1)}{(f-p+1)^2(f-p+2)} \frac{N_1 + N_2 + 1}{N_1 + N_2} \left(\rho^2 + \frac{(p-1)(N_1 + N_2)}{N_1 N_2} \right).$$

We note that

$$G_1(t; \rho^2, N_1, N_2, p) \equiv 1 - G_2(-t; \rho^2, N_2, N_1, p).$$

Theorem 2. For $p \geq 2$,

$$G_2(t) = \Phi(t) + \frac{1}{p-1} \sum_{s=1}^4 a_s^1 \Phi^{(s)}(t) + O\left(\frac{1}{p^2}\right); \quad \Phi^{(s)}(t) = \frac{d^s \Phi(t)}{dt^s};$$

$$a_1^1 = -\frac{2\rho^2\gamma + (1+2\gamma)(\lambda_1 - \lambda_2)}{2\omega^{1/2}(\rho^2 + \lambda_1 + \lambda_2)^{1/2}};$$

$$a_2^1 = \frac{1}{2\omega(\rho^2 + \lambda_1 + \lambda_2)} \left\{ \frac{\rho^2(1+\gamma)\lambda_1^2}{\lambda_1 + \lambda_2} + \frac{(1+\gamma)^2(\lambda_1 - \lambda_2)^2}{2} + \frac{\gamma\rho^4}{2} \right\} + \frac{\gamma^2}{\omega} + 3\gamma + \frac{1}{2} \frac{\lambda_1 + \lambda_2}{\rho^2 + \lambda_1 + \lambda_2};$$

$$a_3^1 = -\frac{\lambda_1 + \lambda_2}{2\omega^{1/2}(\rho^2 + \lambda_1 + \lambda_2)^{3/2}} \left\{ \frac{2\lambda_1\rho^2}{\lambda_1 + \lambda_2} + \lambda_1 - \lambda_2 \right\} - \frac{\gamma(\rho^2 + \lambda_1 - \lambda_2)}{\omega^{1/2}(\rho^2 + \lambda_1 + \lambda_2)^{1/2}};$$

$$a_4^1 = \gamma + \frac{1}{4} \frac{\gamma^2}{\omega} + \frac{1}{4} \frac{(\lambda_1 + \lambda_2)(2\rho^2 + \lambda_1 + \lambda_2)}{(\rho^2 + \lambda_1 + \lambda_2)^2};$$

$$\lambda_i = (p-1)/N_i; \quad \gamma = \lambda_1\lambda_2/(\lambda_1 + \lambda_2 - \lambda_1\lambda_2);$$

$$\omega = \gamma + 1 = (\lambda_1 + \lambda_2)/(\lambda_1 + \lambda_2 - \lambda_1\lambda_2).$$

Analogous expansions have been obtained for other statistics of discriminant analysis.

In conclusion, the author considers it his pleasant duty to express his profound gratitude to A. N. Kolmogorov for posing the problem and to Yu. N. Blagoveshchenskii for supervising the research.

Received
16 IV 1970

REFERENCES

- ¹ A. Wald, *Ann. Math. Stat.*, **15**, 1, 145 (1944).
- ² T. W. Anderson, *Psichometrika*, **16**, No. 1, 31 (1951).
- ³ R. Sitgreaves, *Ann. Math. Stat.*, **23**, 263 (1952).
- ⁴ A. N. Bowker, in: *Contributions to Probability and Statistics*, Stanford Univ. Press, 1960, p. 442.
- ⁵ T. Anderson, *An Introduction to Multivariate Statistical Analysis*, Moscow, 1963.
- ⁶ M. Kendall, A. Stuart, *The Theory of Distributions*, 1, Moscow, 1967.
- ⁷ M. Okamoto, *Ann. Math. Stat.*, **34**, 4, 1286 (1963).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.