Доклады Академии наук СССР
1970. Том 195, № 4

УДК 519.281

МАТЕМАТИКА

А. Д. ДЕЕВ

ПРЕДСТАВЛЕНИЕ СТАТИСТИК ДИСКРИМИНАНТНОГО АНАЛИЗА И АСИМПТОТИЧЕСКИЕ РАЗЛОЖЕНИЯ ПРИ РАЗМЕРНОСТЯХ ПРОСТРАНСТВА, СРАВНИМЫХ С ОБЪЕМОМ ВЫБОРОК

(Представлено академиком А. Н. Колмогоровым 20 IV 1970)

1. Как известно, при классификации наблюдения X в одну из нескольких (S) популяций, информации о которых задаются предварительной выборкой, следует различать три вида ошибок: а) неизвестная исследователю ошибка $P(H_i|H_j)$, которая получается при полностью определенных распределениях из-за статистической природы задачи (речь идет об ошибке отнесения X к i-й популяции, когда на самом деле X принадлежит j-й $(i \neq j)$; б) ошибка, обусловленная однократностью построения классификатора по фиксированной выборке, т. е. условная ошибка $P(H_i|H_j, \{X_\alpha^{(k)}\})$, $\alpha = 1, \ldots, N_k$, $k = 1, \ldots, S$, i, j фиксированы. Условная ошибка является случайной величиной, распределение которой желательно оценить и первой характеристикой является ее усреднение по всевозможным выборкам фиксированного объема $N = (N_1, N_2, \ldots, N_s)$; в) $P_{N, ij} = MP(H_i|H_j, \{X_\alpha^{(k)}\})$. Очевидно, что для любого линейного функционала от ошибок $P(H_i|H_j)$ $L = \Sigma C_{ij} P(H_i|H_j)$ минимум будет меньше минимума $L_N = \Sigma C_{ij} MP(H_i|H_j, \{X_\alpha^{(k)}\})$.

Изучение распределений, связанных с дискриминантными функциями, является сложной задачей, весьма далекой от удовлетворительного решения. В настоящей заметке предлагается представление некоторых статистик классификации через простые одномерные случайные величины (нормальные и χ^2), которое позволяет получить асимптотические разложения, полезные для практических целей.

2. Мы будем рассматривать задачу классификации наблюдения X в одну из двух нормальных популяций $\pi_i \sim N(\mu_i, \Sigma)$ $(i = 1, 2)$ с общей ковариационной матрицей Σ. По соображениям достаточности выборочная информация $\{X_\alpha^{(i)}\}$ $\alpha = 1, \ldots, N_i$ редуцируется в $\mathfrak{M} = \{\overline{X}^{(1)}, \overline{X}^{(2)}, A\}$, где $\overline{X}^{(i)} = \frac{1}{N_i} \sum_{\alpha=1}^{N_i} X_\alpha^{(i)}$ — оценки средних популяций μ_i, а $S = \frac{1}{f} A =$ $= \frac{1}{N_1 + N_2 - 2} A = \frac{1}{N_1 + N_2 - 2} \sum_{i=1}^{2} \sum_{\alpha=1}^{N_i} (X_\alpha^{(i)} - \overline{X}^{(i)})(X_\alpha^{(i)} - \overline{X}^{(i)})'$ — несмещенная оценка общей ковариационной матрицы Σ. Пусть $W = (X - {}^1/_2\overline{X}^{(1)} - {}^1/_2\overline{X}^{(2)\prime} S^{-1}(\overline{X}^{(2)} - \overline{X}^{(1)})$ — так называемая статистика Андерсона, обычно используемая для классификации X согласно правилу: принимаем $X \in \pi_1$ при $W < C$ и $X \in \pi_2$ при $W > C$, C — порог классификации, выбираемый исследователем; обычно $C = 0$, что мы и будем полагать, хотя это несущественно.

Распределение W сложно и исследовалось многими авторами ([1–3]). Особенно следует отметить работу ([4]), в которой впервые дано представление W через простые статистики, полезное для моделирования, численного анализа, получения асимптотических разложений и т. п.

В параметрическом пространстве существует всего один параметр — так называемое расстояние Махаланобиса $\rho^2 = (\mu_2 - \mu_1)' \Sigma^{-1} (\mu_2 - \mu_1)$, и ошибки являются функциями от ρ, например, могут быть сделаны равными: $P(H_1|H_2) = P(H_2|H_1) = \Phi(-\rho/2)$, где $\Phi(\cdot)$ — функция распределения стандартной нормальной величины $N(0, 1)$.

Вследствие условной нормальности W относительно $\mathfrak{M}$

$$P(W > 0 \mid \mathfrak{M}, H_1) = \Phi\left(\frac{(\mu_1 - {}^1/_2\overline{X}^{(1)} - {}^1/_2\overline{X}^{(2)})' S^{-1} (\overline{X}^{(2)} - \overline{X}^{(1)})}{[(\overline{X}^{(2)} - \overline{X}^{(1)})' S^{-1} \Sigma S^{-1} (\overline{X}^{(2)} - \overline{X}^{(1)})]^{1/2}}\right),$$

$$P(W < 0 \mid \mathfrak{M}, H_2) = \Phi\left(\frac{({}^1/_2\overline{X}^{(1)} + {}^1/_2\overline{X}^{(2)} - \mu_2)' S^{-1} (\overline{X}^{(2)} - \overline{X}^{(1)})}{[(\overline{X}^{(2)} - \overline{X}^{(1)})' S^{-1} \Sigma S^{-1} (\overline{X}^{(2)} - \overline{X}^{(1)})]^{1/2}}\right).$$

Обозначим через ξ_1 и ξ_2 аргументы. Выразим ξ_1 и ξ_2 через шесть простых случайных величин. По соображениям инвариантности матрицу Σ будем считать единичной I_p. Обозначим $Z_1 = f_1^{-1/2}(\overline{X}^{(2)} - \overline{X}^{(1)})$ и $Z_2 = (N_1\overline{X}^{(1)} + N_2\overline{X}^{(2)})/(N_1 + N_2)^{1/2}$, $f_1 = (N_1 + N_2)/N_1N_2$; тогда Z_1 и Z_2 независимы и имеют распределения $Z_1 \sim N_1(f_1^{-1/2}\Delta\mu, I_p)$, $Z_2 \sim N((N_1\mu_1 + N_2\mu_2)/(N_1 + N_2)^{1/2}, I_p)$ $(\Delta\mu = \mu_2 - \mu_1)$. В указанных обозначениях ξ_1 и ξ_2 представятся в следующем виде:

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

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

Величина $\nu_2 = \dfrac{(Z_2 - MZ_2)' A^{-1}Z_1}{(Z_1' A^{-2} Z_1)^{1/2}}$ имеет стандартное нормальное распределение и не зависит от A и Z_1; что касается совместного распределения $\beta_1 = \Delta\mu' A^{-1}Z_1$, $\beta_2 = Z_1' A^{-1} Z_1$ и $\beta_3 = Z_1' A^{-2} Z_1$, то справедливо следующее представление:

Теорема 1. *Если $Z \sim N(\Delta\mu, I_p)$ и $A \sim W(I_p, f)$, где Z и A независимы, а $W(I_p, f)$ обозначает распределение Уишарта в p-мерном пространстве с f степенями свободы, то статистики $(\beta_1, \beta_2, \beta_3)$ выражаются через 5 независимых случайных величин:*

$$\beta_1 = (\chi^2_{f-p+1})^{-1} |\Delta\mu| [\nu_1 + |\Delta\mu| + R \sin\theta \sqrt{\chi^2_{p-1}}],$$
$$\beta_2 = (\chi^2_{f-p+1})^{-1} [(\nu_1 + |\Delta\mu|)^2 + \chi^2_{p-1}],$$
$$\beta_3 = (\chi^2_{f-p+1})^{-2} [(\nu_1 + |\Delta\mu|)^2 + \chi^2_{p-1}](1 + R^2),$$

$\nu_1 \sim N(0, 1)$; χ^2_{f-p+1} и χ^2_{p-1} — χ^2-величины; R^2 имеет B-распределение II рода (см. ([6])) с

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

и $p(\theta) = \dfrac{\Gamma((p-1)/2)}{\Gamma(1/2)\,\Gamma((p-2)/2)} \cos^{p-1}\theta$. Заметим, что B-распределение II рода есть отношение независимых χ^2-величин: $R^2 = \chi^2_{p-1}/\chi^2_{f-p+2}$.

Доказательство основано на стандартной технике случайного ортогонального преобразования, не зависящего от A (см. ([4])), и лемме о распределении первой строки матрицы A^{-1}.

Лемма. *Пусть $A = \{a_{ij}\} \sim W(I_p, f)$ и $A^{-1} = \{a^{ij}\}$, тогда a^{11} и $g' = \left(\dfrac{a^{12}}{a^{11}}, \dfrac{a^{13}}{a^{11}}, \ldots, \dfrac{a^{1p}}{a^{11}}\right)$ независимы и имеют распределения $a^{11} \sim \dfrac{1}{\chi^2_{f-p+1}}$,*

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

Доказательство леммы основано на правиле умножения матриц, разбитых на блоки, и теореме 4.3.2 ([5]). Именно, если $A = \begin{pmatrix} a_{11} & a_{(1)} \\ a_{(1)} & A_{22} \end{pmatrix}$ и $A^{-1} = \begin{pmatrix} a^{11} & a^{(1)'} \\ a^{(1)} & A^{22} \end{pmatrix}$, то $a^{(1)'} = -a^{11} a_{(1)}' A_{22}^{-1}$ (величина $a_{(1)}' A_{22}^{-1}$ есть строка коэффициентов формальной выборочной регрессии первой координаты на остальные), отсюда по теореме 4.3.2 ([5]) a^{11} и $g' = a^{(1)'} / a^{11}$ независимы. Распределение a^{11} указано, условное распределение g' (при фиксировании координат кроме первой) нормально с нулевым средним и ковариационной матрицей A_{22}^{-1}, причем A_{22} имеет распределение $W(I_{p-1}, f)$. Интегрируя совместную плотность g' и A_{22} по множеству положительно определенных A_{22} (см. ([5]), гл. 7), получаем $p(g)$, что и завершает доказательство леммы.

Чтобы не загромождать изложение, приведем конечный вид представления условных ошибок лишь для $N_1 = N_2 = N$

$$P\{W > 0 \mid \mathfrak{M}, H_1\} = \Phi\left(\frac{v_2}{(2N)^{1/2}} - \frac{1}{2} \frac{\rho (v_1 + \sqrt{N/2}\,\rho + R \sin\theta \sqrt{\chi^2_{p-1}})}{(1 + R^2)^{1/2} [(v_1 + \sqrt{N/2}\,\rho)^2 + \chi^2_{p-1}]^{1/2}} \right),$$

$$P\{W < 0 \mid \mathfrak{M}, H_2\} = \Phi\left(-\frac{v_2}{(2N)^{1/2}} - \frac{1}{2} \frac{\rho (v_1 + \sqrt{N/2}\,\rho + R \sin\theta \sqrt{\chi^2_{p-1}})}{(1 + R^2)^{1/2} [(v_1 + \sqrt{N/2}\,\rho)^2 + \chi^2_{p-1}]^{1/2}} \right).$$

А. Н. Колмогоров предложил изучить поведение распределения классификаторов при $p/N_i \to \lambda_i$ ($p \to \infty$, $N_i \to \infty$) и дал первое приближение ошибок классификации с помощью правила, основанного на двух статистиках: $\Delta_i = (X - \overline{X}^{(i)})' S^{-1} (\overline{X}^{(2)} - \overline{X}^{(1)})$; $\Delta_1 + \Delta_2 = r^2 = (\overline{X}^{(2)} - \overline{X}^{(1)})' S^{-1} (\overline{X}^{(2)} - \overline{X}^{(1)})$ и допускающего, вообще говоря, зону отказа от классификации. Рассмотрение этой задачи составляет предмет отдельной статьи, мы же дадим здесь главный и первый член разложения распределения W в указанной асимптотике. Подобное разложение в обычной асимптотике (p — фиксировано, $N_i \to \infty$) получено в ([7]), но уже при умеренных значениях p-поправки сравнимы с главным членом (см. табл. 2, стр. 1292 ([7])), и поэтому нам кажется целесообразным разложение при $p/N_i \to \lambda_i = \text{const}$. Разложение получено синтезированием идей предварительного усреднения по нормальной мере и представления через простые статистики с последующим применением метода Лапласа к условной характеристической функции.

Пусть

$$G_i(t) = P\left\{W < \frac{f}{f-p+1}(-1)^i \left[\frac{\rho^2}{2} + \frac{(N_2 - N_1)(p-1)}{2N_1N_2}\right] + tD \mid H_i\right\},$$

где

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

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

Теорема 2. *При* $p \geqslant 2$

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

$$a_1^1 = -\frac{2\rho^2\gamma + (1+2\gamma)(\lambda_1 - \lambda_2)}{2\omega^{1/2}(\rho^2 + \lambda_1 + \lambda_2)^{1/2}}; \quad a_2^1 = \frac{1}{2\omega(\rho^2 + \lambda_1 + \lambda_2)} \left\{ \frac{\rho^2 (1+\gamma) \lambda_1^2}{\lambda_1 + \lambda_2} + \right.$$

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

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

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

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

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

Аналогичные разложения получены для других статистик дискриминантного анализа.

В заключение автор считает своим приятным долгом выразить горячую признательность А. Н. Колмогорову за постановку задачи и Ю. Н. Благовещенскому за руководство исследованием.

Поступило
16 IV 1970

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] A. Wald, Ann. Math. Stat., 15, 1, 145 (1944). [2] T. W. Anderson, Psichometrika, 16, № 1, 31 (1951). [3] R. Sitgreaves, Ann. Math. Stat., 23, 263 (1952). [4] A. H. Bowker, In: Contributions to Probability and Statistics, Stauford Univ. Press, 1960, p. 142. [5] Т. Андерсон, Введение в многомерный статистический анализ, М., 1963. [6] М. Кендалл, А. Стьюарт, Теория распределений, 1, М., 1967. [7] M. Okamoto, Ann. Math. Stat., 34, 4, 1286 (1963).