УДК 519.21

ВЫДЕЛЕНИЕ ТОЧЕК,ОБЛАДАЮЩИХ ПРИНЦИПОМ << Ц >>

Кнопов П.С, Пепеляева Т.В.

Институт кибернетики им. В.М.Глушкова НАН Украины,
отдел математических методов исследования операций,
ул. академика В.М. Глушкова, Киев-187, 03187
e-mail: *knopov@yahoo.com,pepelaev@yahoo.com*

Abstract In this article we investigate an easy enough algorithm for solving of a very important problem of recognition of unknown objects, which appear when we find the dependence rate between random samples of observations for the case of Gaussian random variables.

Одной из важных проблем теории распознавания является задача обнаружения цели при заданных наборах случайных наблюдений. Для решения этой задачи на практике существенным является вопрос о создании простых и в то же время надежных алгоритмов ее решения. В статье предлагается один из таких алгоритмов и оценивается вероятность ошибки принятия неверного решения.

1.Постановка задачи Пусть n,N – натуральные числа. Для каждого $j \in \{1, 2...N\}$ наблюдаются реализации наборов случайных величин

$$\begin{gathered} \Delta^{(k)}_{(j,1)} = \xi^{(k)}_{(j,1)} + \eta^{(k)}_{(j,1)}, \\ \Delta^{(k)}_{(j,2)} = \xi^{(k)}_{(j,2)} + \eta^{(k)}_{(j,2)}, k = 1, 2, ..., q \end{gathered} \tag{1}$$

Здесь $\{\xi^{(k)}_{(j,1)} | k = 1, 2, ..., q\}$ - набор независимых, одинаково распределенных $N(m, \sigma^2)$ случайных величин, не зависящих от $\{\eta^{(k)}_{(j,1)} \mid k = 1, 2, ..., q\}$, $\{\eta^{(k)}_{(j,2)} \mid k = 1, 2, ..., q\}$. Таким же является набор случайных величин $\{\xi^{(k)}_{(j,2)} \mid k = 1, 2, ..., q\}$. Наборы случайных величин $\{\eta^{(k)}_{(j,i)} \mid k = 1, 2, ..., q\}$, $i = 1, 2$ независимы, не зависят от наборов $\{\xi^{(k)}_{(j,i)} \mid k = 1, 2, ..., q\}$, $\quad i = 1, 2$, $\eta^{(k)}_{(j,i)}$ имеет распределение $N(0, \sigma^2)$.

Если случайные величины $\Delta^{(k)}_{(j,1)}$ и $\Delta^{(k)}_{(j,2)}$, $\quad k = 1, 2, ..., q$ независимы, то точка j не обладает признаком «Ц».

Если случайные величины $\Delta^{(k)}_{(j,1)}$ и $\Delta^{(k)}_{(j,2)}$, $\quad k = 1, 2, ..., q$ имеют корреляцию $\rho > 0$, то точка j обладает признаком «Ц».

Признаком «Ц» обладает ровно n из N точек.

По наблюдениям **??** требуется определить точки, обладающие признаком «Ц».

2. Рассмотрим случайную величину

$$\Delta_j = \Delta_{(j,1)} - \Delta_{(j,2)}, j = 1, 2, ..., N$$

Ясно, что $M\Delta_j = 0$, j=1,2,...,N. Случай, когда точка j обладает признаком «Ц», будет характеризоваться тем, что $D\Delta_j = 2(1-\rho)\sigma_0^2$, где $\sigma_0^2 = \sigma^2 + \sigma_1^2$. Случай, когда точка j не обладает признаком «Ц» будет характеризоваться тем, что $D\Delta_j = 2\sigma_0^2$, где $\sigma_0^2 = \sigma^2 + \sigma_1^2$.

Пусть H_0 - гипотеза, которая состоит в том, что случайная величина Δ_j имеет дисперсию $2(1-\rho)\sigma_0^2$ (точка j обладает признаком «Ц»);

H_1- гипотеза, которая состоит в том, что случайная величина Δ_j имеет дисперсию $2\sigma_0^2$ (точка j не обладает признаком «Ц»).

Для каждого $j \in 1, 2, ..., N$ будем проверять гипотезу H_0 против альтернативы H_1.

Пусть $\alpha = P(H_1/H_0)$ - ошибка 1-го рода (вероятность пропуска признака «Ц»).

$\beta = P(H_0/H_1)$ - ошибка 2-го рода (вероятность ложного обнаружения признака «Ц»).

$1 - \beta$ - мощность критерия - вероятность того, что при отсутствии признака «Ц» он не будет обнаружен.

Пусть задан уровень значимости критерия α. Согласно лемме Неймана-Пирсона, критическая область K критерия наибольшей мощности заданного уровня значимости будет иметь вид

$$K = \{\vec{x} \in R^q | \frac{f_1(\vec{x})}{f_0(\vec{x})}\}, \vec{x} = (x_1, ..., x_n),$$

где f_1- плотность распределения случайной величины $\vec{\Delta}_j = (\Delta_j^{(1)}, ..., \Delta_j^{(q)})$, $j \in 1, 2, ..., N$ при гипотезе H_1; f_0- плотность распределения этой же случайной величины при гипотезе H_0.

В настоящем случае

$$f_1(\vec{x}) = \Big(\frac{1}{\sqrt{2\pi}\sqrt{2\sigma_0^2}}\Big)^q exp\Big\{ -\frac{1}{2\sigma_0^2}\sum_{i=1}^{q} x_i^2\Big\},$$

$$f_0(\vec{x}) = \Big(\frac{1}{\sqrt{2\pi}\sqrt{2(1-\rho)\sigma_0^2}}\Big)^q exp\Big\{-\frac{1}{2(1-\rho)\sigma_0^2}\sum_{i=1}^{q} x_i^2\Big\},$$

Далее,

$$\frac{f_1(\vec{x})}{f_0(\vec{x})} = (1-\rho)^{\frac{q}{2}} exp\Big\{\frac{1}{2\sigma_0^2}\Big(\frac{1}{1-\rho}-1\Big)\sum_{i=1}^{q} x_i^2\Big\} = (1-\rho)^{\frac{q}{2}} exp\Big\{\frac{\rho}{2\sigma_0^2(1-\rho)}\sum_{i=1}^{q} x_i^2\Big\}.$$

Таким образом, критическая область будет иметь вид

$$K = \Big\{\vec{x} \in R^q | \sum_{i=1}^{q} x_i^2 \geq c\Big\}.$$

Постоянная c находится по заданному уровню значимости α.

При гипотезе H_0 величина $\frac{1}{2\sigma_0^2(1-\rho)}\sum\limits_{i=1}^{q}(\Delta_j^{(i)})^2, \quad j \in 1,2,...,N$ имеет χ^2 распределение с q степенями свободы. Из уравнения

$$P\Big\{\vec{x} \in R^q | \frac{1}{2\sigma_0^2(1-\rho)}\sum_{i=1}^{q} x_i^2 \geq \frac{c}{2\sigma_0^2(1-\rho)}\Big\} = \alpha$$

по таблице квантилей χ^2 распределения находим $\frac{c}{2\sigma_0^2(1-\rho)}$ [1,с.551].

При гипотезе H_1 величина $\frac{1}{2\sigma_0^2}\sum_{i=1}^{q}(\Delta_j^{(i)})^2$ имеет χ^2 распределение с q степенями свободы. Мощность критерия

$$1-\beta = P\Big\{\frac{1}{2\sigma_0^2} \geq (1-\rho)\frac{c}{2\sigma_0^2(1-\rho)}\Big\}$$

Табл.1 $\alpha = 0.01$

q	I	II	$1-\beta$	III	$1-\beta$	IV	$1-\beta$	V	$1-\beta$	VI	$1-\beta$
5	15.086	7.543	0.2	6.344	0.3	4.758	0.5	3.017	0.7	1.508	0.85
10	23.209	1.6	0.3	9.28	0.5	6.96	0.7	5.64	0.85	2.32	0.99
15	30.578	15.28	0.4	12.23	0.7	9.17	0.85	6.1	0.98	3.5	0.99
20	37.566	18.75	0.6	15.02	0.75	11.2	0.92	7.5	0.99	3.75	0.99
25	44.314	22.15	0.65	17.7	0.85	13.3	0.96	8.86	0.99	4.4	0.99
30	50.892	25.45	0.7	20.36	0.9	15.27	0.99	10.17	0.99	5.1	0.99

I: $c/2\sigma_0^2(1-\rho)$, II: $\rho = 0.5c/2\sigma_0^2$,
III: $\rho = 0.6(1-\rho) = 0.4c/2\sigma_0^2$,
IV: $\rho = 0.7(1-\rho) = 0.3c/2\sigma_0^2$,

V: $\rho = 0.8(1-\rho) = 0.2c/2\sigma_0^2$,
VI: $\rho = 0.9(1-\rho) = 0.1c/2\sigma_0^2$.

Вероятность того, что ни один признак «Ц» не будет пропущен, приведены в Табл.2.

n	1	5	10	15	20
P	0.99	0.95	0.9	0.86	0.81

Алгоритм: N, q- натуральные числа, $\{\Delta_j^{(S)} | j = 1, 2, ..., N; S = 1, 2, ..., q\}$ - действительные, ρ, σ_0^2- положительные, $\rho \in (0, 1), \sigma_0^2 > 2$.

1) По таблице 1 для заданного q находим $\frac{c}{2\sigma_0^2(1-\rho)}$и вычисляем c

2) Вычисляем

$$\sum_{i=1}^{q} (\Delta_j^{(S)})^2, j = 1, 2, ..., N.$$

3) Если $\sum\limits_{s=1}^{q} (\Delta_j^{(S)})^2 < C$, то j обладает признаком «Ц».

Если $\sum\limits_{s=1}^{q} (\Delta_j^{(S)})^2 \geq C$, то j не обладает признаком «Ц».

3. Укажем один способ выделения признака «Ц» при $N = 4, n = 2$. Среднее m будем предполагать неизвестным.

Положим $\bar{\Delta}_{(j,i)} = \frac{1}{q} \sum\limits_{s=1}^{q} \Delta_{(j,i)}^{(k)}, \quad j = 1, 2, 3, 4; \quad i = 1, 2.$

$$\hat{r}_j = \frac{\sum\limits_{s=1}^{q} (\Delta_{(j,1)}^{(k)} - \bar{\Delta}_{(j,1)})(\Delta_{(j,2)}^{(k)} - \bar{\Delta}_{(j,2)})}{\sqrt{\sum\limits_{s=1}^{q} (\Delta_{(j,1)}^{(k)} - \bar{\Delta}_{(j,1)})^2 \sum\limits_{s=1}^{q} (\Delta_{(j,2)}^{(k)} - \bar{\Delta}_{(j,2)})^2}}$$

$j = 1, 2, 3, 4.$

Пусть $|\hat{r}_{j_1}| > |\hat{r}_{j_2}| > |\hat{r}_{j_3}| > |\hat{r}_{j_4}|$. Тогда полагаем, что точки j_1 и j_2 обладают признаком «Ц».

Пусть, для определенности, признаком «Ц» обладают точки 1) и 4) и $P_{\text{ОШ}}$-вероятность ошибочного определения ситуации. Тогда вероятность ошибки

$$P_{\text{ош}} \leq P\{|\hat{r}_3| > |\hat{r}_1|\} + P\{|\hat{r}_3| > |\hat{r}_4|\} + P\{|\hat{r}_2| > |\hat{r}_1|\} + P\{|\hat{r}_3| > |\hat{r}_4|\}.$$

Каждая из этих вероятностей оценивается однотипно.

Рассмотрим

$$P\{|\hat{r}_3| > |\hat{r}_1|\}.$$

Величина $\hat{r}_3$ имеет плотность

$$f_q(r) = \frac{\Gamma(\frac{q-1}{2})}{\Gamma(\frac{q-2}{2})}\frac{(1-r^2)^{\frac{q-4}{2}}}{\sqrt{\pi}}, -1 \leq r \leq 1.$$

[2,с.436] причем $\hat{r}_3 = 0$, $D\hat{r}_3 = \frac{1}{q} + O\left(\frac{1}{q^{3/2}}\right)$ [2,с.393] ($\hat{r}_3 = 0$ следует из вида $f_q(r)$).

Величина $\hat{r}_1$ имеет плотность

$$g_q(r) = \frac{q-2}{\pi}(1-\rho_1^2)^{\frac{q-1}{2}}(1-r^2)^{\frac{q-4}{2}}\int\limits_0^1 \frac{x^{q-2}}{(1-\rho_1 rx)^{q-1}}\frac{dx}{\sqrt{1-x^2}},$$

причем

$$\hat{r}_1 = \rho_1 + O\Big(\frac{1}{q}\Big), D\hat{r}_1 = \frac{(1-\rho_1^2)^2}{q} + O\Big(\frac{1}{q^{3/2}}\Big)$$

[2,с.393].

Вероятность $P\{|\hat{r}_3| > |\hat{r}_1|\}$ оценим следующим образом.

$$P\{|\hat{r}_3 > \hat{r}_1|\} \leq P\{|\hat{r}_3 \geq \varepsilon\} + P\{|\hat{r}_1 < \varepsilon\}.$$

Далее,

$$P\{|\hat{r}_3 \geq \varepsilon\} \leq \frac{1}{\varepsilon^2}\Big[\frac{1}{q}O\Big(\frac{1}{q^{3/2}}\Big)\Big] \leq \frac{1}{\varepsilon^2 q} + \frac{c}{\varepsilon^2 q^{3/2}}.$$

Здесь и далее буквой будем обозначать различные константы (зависящие, быть может, только от ρ).

Вероятность $P\{|\hat{r}_1 \geq \varepsilon\}$ оценим двумя способами.

Пусть ε отделено от $|\rho|$, т.е. $\varepsilon \leq |\rho| - \varepsilon_0$, ε_0 - фиксировано, от q не зависит. Для достаточно больших q $|Er| > \varepsilon$ для этих q имеем:

$$\begin{gathered} P\{|\hat{r}_1 \geq \varepsilon\} \leq P\{|\hat{r}_1 - E\hat{r}_1| \geq |E\hat{r}_1| - \varepsilon\} \leq \\ \leq \frac{1}{\left[|\rho| - \varepsilon + O\left(\frac{1}{q}\right)\right]^2} \left[\frac{(1-\rho^2)^2}{q} + O\left(\frac{1}{q^{3/2}}\right)\right] \leq \\ \leq \frac{1}{(|\rho|-\varepsilon)^2}\left(1 + \frac{c}{q}\right)\left[\frac{(1-\rho^2)^2}{q} + \frac{c}{q^{3/2}}\right] \leq \\ \leq \frac{1}{(|\rho|-\varepsilon)^2}\frac{(1-\rho^2)^2}{q} + \frac{c}{q^{3/2}}. \end{gathered} \tag{2}$$

Эта оценка не учитывает малость оцениваемой вероятности при малых ε. Для построения другой оценки воспользуемся видом плотность $g_q(r)$. Имеем при $|r| < \varepsilon$, ε отделено от $|\rho|$:

$$g_q(r) \leq \begin{cases} \frac{q-2}{\pi}(1-\rho^2)^{\frac{q-1}{2}} \int\limits_0^1 \frac{x^{q-2}}{(1-|\rho_1|\varepsilon)^{q-1}} \frac{dx}{\sqrt{1-x^2}}, & \rho_1 r > 0, \\ \frac{q-2}{\pi}(1-\rho^2)^{\frac{q-1}{2}} \int\limits_0^1 \frac{x^{q-2}dx}{\sqrt{1-x^2}}, & \rho_1 r < 0. \end{cases}$$

Интеграл

$$\int\limits_0^1 \frac{x^{q-2}dx}{\sqrt{1-x^2}} = |x^2 = t| = \frac{1}{2}\int\limits_0^1 (1-t)^{-\frac{1}{2}} t^{\frac{q-3}{2}} dt = \frac{1}{2} B\left(\frac{1}{2}, \frac{q-1}{2}\right) = \frac{1}{2}\frac{sqrt\pi\Gamma(\frac{q-1}{2})}{\Gamma(\frac{q}{2})} =$$

$$= \frac{\sqrt{\pi}}{2}\frac{\Gamma(\frac{q-1}{2})}{\Gamma(\frac{q}{2})}.$$

Для $\Gamma(x)$ справедлива формула Стирлинга

$$\Gamma(x) = cx^{x-\frac{1}{2}}e^{-x}e^{\frac{\Theta(x)}{12x}}, \qquad 0 \leq \Theta(x) \leq 1$$

Отсюда при $x \to \infty$

$$\frac{\Gamma(x)}{\Gamma(x+\frac{1}{2})} \sim \frac{x^{x-\frac{1}{2}}e^{-x}}{(x+\frac{1}{2})^x e^{-x-\frac{1}{2}}} \sim \sqrt{e}\frac{1}{\sqrt{x}(1+\frac{1}{2x})^x} \sim \frac{1}{\sqrt{x}}, \quad x \to +\infty.$$

Поэтому

$$\int\limits_0^1 \frac{x^{q-2}dx}{\sqrt{1-x^2}} \sim \frac{\sqrt{\pi}}{2}\frac{1}{\sqrt{\frac{q}{2}}} = \frac{\sqrt{\pi}}{\sqrt{2}\cdot\sqrt{q}}, \quad q \to \infty;$$

$$g_q(r) \leq \begin{cases} c\sqrt{q}\frac{(1-\rho^2)^{\frac{q-1}{2}}}{(1-|\rho_1|\varepsilon)^{q-1}}, & \rho_1 r > 0, \quad |r| < \varepsilon, \\ c\sqrt{q}(1-\rho^2)^{\frac{q-1}{2}}, & \rho_1 r < 0, \quad |r| < \varepsilon. \end{cases}$$

Тогда

$$P\{|\hat{r}_1 < \varepsilon\} \leq c\sqrt{q}(1-\rho^2)^{\frac{q-1}{2}}\varepsilon\Big[1+\tfrac{1}{(1-|\rho_1|\varepsilon)^{n-1}}\Big] \leq c\sqrt{q}\varepsilon(1-\rho^2)^{\frac{q-1}{2}}/(1-|\rho_1|\varepsilon)^{q-1}. \tag{3}$$

Этой оценкой имеет смысл пользоваться, если

$$\frac{1-\rho^2}{(1-|\rho|\varepsilon)^2} < 1,$$

или

$$1-|\rho|\varepsilon > \sqrt{1-\rho^2}; \quad \varepsilon < \frac{1-\sqrt{1-\rho^2}}{|\rho|} = \frac{|\rho|}{1+\sqrt{1-\rho^2}}.$$

Построим с помощью оценок (2) и (3) две оценки для искомой вероятности. Итак, в первом случае

$$P\{|\hat{r}_3 > \hat{r}_1|\} \leq \frac{1}{\varepsilon^2 q} + \frac{1}{(|\rho|-\varepsilon)^2}\frac{(1-\rho^2)^2}{q} + \frac{c}{\varepsilon^2 q^{3/2}}.$$

Минимизируя по ε выражение

$$\frac{1}{\varepsilon^2} + \frac{(1-\rho^2)^2}{(|\rho|-\varepsilon)^2},$$

получаем для критической точки этой функции от ε уравнение

$$\frac{1}{\varepsilon^3} = \frac{(1-\rho^2)^2}{(|\rho|-\varepsilon)^3}; \quad \frac{|\rho|-\varepsilon}{\varepsilon} = (1-\rho^2)^{2/3},$$

Откуда

$$\varepsilon = \frac{|\rho|}{1+(1-\rho^2)^{2/3}}.$$

Тогда при этом

$$\frac{1}{\varepsilon^2} + \frac{(1-\rho^2)^2}{(|\rho|-\varepsilon)^2} = \frac{[1+(1-\rho^2)^{2/3}]^3}{\rho^2},$$

Действительно,

$$\frac{1}{\varepsilon^2} + \frac{\varepsilon^2(1-\rho^2)^2}{(|\rho|-\varepsilon)^{4/3}} = \frac{1}{\varepsilon^2}[1+(1-\rho^2)^{2/3}] = \frac{[1+(1-\rho^2)^{2/3}]^3}{\rho^2}.$$

Итак,

$$P\{|\hat{r}_3 > \hat{r}_1|\} \leq \frac{1}{\varepsilon^2}[1+(1-\rho^2)^{2/3}] = \frac{[1+(1-\rho^2)^{2/3}]^3}{\rho^2}\frac{1}{q} + \frac{c}{q^{3/2}}. \quad (4)$$

Оценка (3) дает

$$P\{|\hat{r}_3 > \hat{r}_1|\} \leq \frac{1}{\varepsilon^2 q} + \frac{1}{\varepsilon^2 q^{3/2}} + c\sqrt{n}\varepsilon\Big(\frac{\sqrt{1-\rho^2}}{1-|\rho|\varepsilon}\Big)^{n-1}.$$

Выберем ε так, чтобы

$$\frac{\sqrt{1-\rho^2}}{1-|\rho|\varepsilon} = 1 - \frac{1}{\sqrt{q}}.$$

Тогда

$$|\rho|\varepsilon = -\frac{\sqrt{1-\rho^2}}{1-\frac{1}{\sqrt{q}}} + 1;$$

$$\varepsilon = \frac{|\rho|}{1+\sqrt{1-\rho^2}} + O\big(\frac{1}{\sqrt{q}}\big).$$

При этом

$$\frac{1}{\varepsilon^2 q} = \Big[\frac{|\rho|}{1+\sqrt{1-\rho^2}}\Big]^{-2}\frac{1}{q} + O\big(\frac{1}{q^{3/2}}\big);$$

$$c\sqrt{n}\Big(\frac{\sqrt{1-\rho^2}}{1-|\rho|\varepsilon}\Big)^{q-1} = O\big(\frac{1}{q^{3/2}}\big), \quad q \to \infty.$$

Итак, получена оценка

$$P\{|\hat{r}_3 > \hat{r}_1|\} \leq \frac{(1+\sqrt{1-\rho^2})^2}{\rho^2}\frac{1}{q} + \frac{c}{q^{3/2}} \quad (5)$$

Объединяя оценки (4) и (5), получаем

$$P\{|\hat{r}_3 > \hat{r}_1|\} \leq \frac{A_\rho}{q} + \frac{C_\rho}{q^{3/2}} \stackrel{df}{=} F_\rho(q), \quad (6)$$

где

$$A_\rho \stackrel{df}{=} min\Big\{\frac{(1+\sqrt{1-\rho^2})^2}{\rho^2}, \frac{[1+(1-\rho^2)^{2/3}]^3}{\rho^2}\Big\} \quad (7)$$

$C_\rho < \infty$.

Сравним оценки (4), (5). Составим неравенство

$$\Big[\frac{1+\sqrt{1-\rho^2}}{|\rho|}\Big]^2 < \frac{[1+(1-\rho^2)^{2/3}]^3}{\rho^2}$$

$(1+\gamma^{1/2})^2 < [1+\gamma^{2/3}]^3, \gamma = 1-\rho^2$.

При малых γ левая часть больше (т.к. $\gamma^{1/2} > \gamma^{2/3}$); при γ , близких к единице, неравенство справедливо.

Итак, при достаточно малых $|\rho|$ выгоднее пользоваться оценкой (5); при достаточно больших $|\rho|$ - оценкой (4). В результате мы получили неравенство вида

$$P\{|\hat{r}_3 > \hat{r}_1|\} \leq F_\rho(q).$$

Тогда искомая вероятность ошибки

$$P_{\text{ош}} \leq 4[F_\rho(q) + F_\rho].$$

Замечание.Оценки, подобные (3) и соответственно (4), опираются на асимптотику первых двух моментов выборочного коэффициента корреляции, можно строить без предположения о нормальности рассматриваемых величин. Достаточно предположить существование четвертых моментов; нужно также предполагать известными не только корреляции ρ_j , но и центральные (смешанные) моменты со второго по четвертый включительно. Схема рассуждений, основанная на неравенстве Чебышева и асимптотических формулах для первых двух моментов выборочного коэффициента корреляции, остается без изменений; последние формулы даны в монографии Крамера на с.393 [2].

Заключение

Результаты статьи могут быть использованы при решении конкретных задач распознавания. Простота алгоритма и оценка вероятности принятия неверных решений делает его весьма удобным и легко реализуемым на практике.

Список литературы

1. КендаллМ. Дж„ Стьюарт А. Теория распределений. М.: «Наука», 1966.
2. Крамер Г. Математические методы статистики. М.: «Мир», 1975.