Для завершения доказательства теоремы покажем, что $\rho^n(A)\le\xi$, если $I_{\text{sum}}=N_n$. Ввиду определения числа ξ, для любой траектории $t\in L^n(A)$ и всякого индекса $i\in N_n$ найдется такая траектория $t'\ne t$, что $\gamma_i(t,t',A)\le\xi$.

Отсюда, положив $\varepsilon>\alpha>\xi$ и взяв в качестве возмущающей матрицу $B\in\Re(\varepsilon)$ с элементами

$$b_{i,j}=\begin{cases}\alpha & \text{при} \quad i\in N_n,\ j\in N(t),\\ -\alpha & \text{при} \quad i\in N_n,\ j\notin N(t),\end{cases}$$

с учетом линейности функции $\tau_i(t,t',A)$ выводим

$$\tau_i(t,t',A+B)=\tau_i(t,t',A)+\alpha\Delta_i(t,t')>\tau_i(t,t',A)+\gamma_i(t,t',A)\ \Delta_i(t,t')=0\ \forall\ i\in N_n.$$

Поэтому траектория $t\notin L^n(A+B)$. Следовательно, в силу свойства 2 получаем $\rho^n(A)\le\xi$.

Собирая все доказанное, убеждаемся в справедливости теоремы 2.

Настоящая работа выполнена при поддержке Фонда фундаментальных исследований Республики Беларусь (проект Ф95-70).

1. Ефимчик Н.Е., Подкопаев Д.П. // Вестн. Белорус. ун-та. Сер.1. 1996. №1. С.48.
2. Диниц Е.А. О решении двух задач о назначении. Исследования по дискретной оптимизации. М., 1976. С.333.
3. Подиновский В.В., Гаврилов В.М. Оптимизация по последовательно применяемым критериям. М., 1975.
4. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. М., 1982.
5. Сергиенко И.В. Математические модели и методы решения задач дискретной оптимизации. Киев, 1985.
6. Дубов Ю.А., Травкин С.И., Якимец В.Н. Многокритериальные модели формирования и выбора вариантов систем. М., 1986.
7. Емеличев В.А., Гладкий А.А., Янушкевич О.А. // Весці АН Беларусі. Сер. фіз.-мат. навук. 1996. №3. С.82.
8. Гордеев Э.Н., Леонтьев В.К. // Журн. вычисл. математики и мат. физики. 1980. Т.20. №4. С.1071.
9. Гордеев Э.Н., Леонтьев В.К. // Кибернетика. 1986. №5. С.82.
10. Sotskov Yu.N., Leontev V.K., Gordeev E.N. // Discrete Applied Mathematics. 1995. V 58. P. 169.
11. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М., 1972.

Поступила в редакцию 19.06.97.

УДК 519.2

А.Л.КОСТЕВИЧ

ОБ АСИМПТОТИЧЕСКОМ АНАЛИЗЕ ВЕРОЯТНОСТИ ОШИБКИ КЛАССИФИКАЦИИ ЦЕПЕЙ МАРКОВА

The decision rules for classification of homogeneous finite Markov chains are constructed for different levels of prior information and their performance is investigated.

Большинство методов в дискриминантном анализе разработаны для моделей наблюдения с абсолютно непрерывными условными функциями распределения, в частности для гауссовской модели наблюдения [1]. Подобные методы для моделей с дискретными распределениями и с зависимыми наблюдениями, образующими временной ряд, в частности для часто используемых на практике [1,2] цепей Маркова, исследованы мало. В данной работе рассматривается задача дискриминантного анализа однородных цепей Маркова (ОЦМ) первого порядка при различных уровнях априорной неопределенности.

1. Математическая модель. Пусть на вероятностном пространстве (Ω,F,P) наблюдаются ОЦМ $\{X_t\}$, $t=1,2,...$ из $L\ge2$ классов $\Omega_1,...,\Omega_L$ с заданными априорными вероятностями $q_1,...,q_L$, $(q_l>0)$. Класс Ω_l образуют цепи Маркова с фазовым пространством $S=\{1,...,N\}$, некоторым вектором начальных вероятностей и матрицей вероятностей одношаговых переходов:

$$\begin{aligned}&\pi^{(l)}=(\pi_{l,i}):\quad \pi_{l,i}=\mathrm{P}\{X_1=i|\Omega_l\},\\ &P^{(l)}=(p_{l,ij}):\quad p_{l,ij}=\mathrm{P}\{X_t=j|X_{t-1}=i,\Omega_l\},\end{aligned}\quad i,j\in S,\ l=\overline{1,L}. \tag{1}$$

Классы $\{\Omega_l\}$ различаются параметрами $\{\pi^{(l)},P^{(l)}\}$, причем эти параметры в приложениях часто неизвестны. Будем также рассматривать стационарное распределение вероятностей $\pi^{(l)*}=(\pi^*_{l,i})$ и матрицу вероятностей пар состояний $\Pi^{(l)}=(P_{l,ij})$ цепей Маркова из класса Ω_l:

$$P_{l,ij}=\mathrm{P}\{X_{t-1}=i,X_t=j|\Omega_l\},\quad i,j\in S,\ l=\overline{1,L}. \tag{2}$$

Требуется построить решающее правило (РП) для классификации наблюдаемой в течение T_0 единиц времени цепи Маркова $X^{(0)}=\{x_{0,1},\ldots,x_{0,T_0}\}$ в L классов: $d=d(X^{(0)})$, $d\in\{1,\ldots,L\}$, оптимальное в смысле вероятности ошибочной классификации:

$$r=\mathrm{P}\{d\neq\nu\}\to\min_{d(\cdot)}, \tag{3}$$

где $\nu\in\{1,\ldots,L\}$ — дискретная случайная величина с распределением вероятностей $\mathrm{P}(\nu=l)=q_l$, указывающая истинный ненаблюдаемый номер класса регистрируемой ОЦМ.

2. Статистические оценки параметров ОЦМ. Часто параметры классов (1), (2) неизвестны, и регистрируется классифицированная обучающая выборка следующего вида:

$$X=\{X^{(1)},\ldots,X^{(L)}\},\quad X^{(l)}=\{x_{l,1},\ldots,x_{l,T_l}\},\quad l=\overline{1,L},$$

где $X^{(l)}$ — наблюдаемая в течение T_l единиц времени реализация ОЦМ из класса Ω_l. В таком случае при построении РП вместо неизвестных параметров (1), (2) будем пользоваться их оценками [2], построенными по обучающей выборке X:

$$\hat{p}_{l,ij}=\frac{n_{l,ij}}{n_{l,i\cdot}},\quad \hat{P}_{l,ij}=\frac{n_{l,ij}}{T_l},\quad n_{l,ij}=\sum_{t=2}^{T_l}\delta_{i,x_{l,t-1}}\delta_{j,x_{l,t}},\quad i,j=\overline{1,N},\quad l=\overline{0,L}, \tag{4}$$

где $\delta_{i,j}$ — символ Кронекера.

Отметим два вспомогательных результата [2].

Лемма 1. *Пусть* $\theta_l=(p_{l,11},\ldots,p_{l,NN})'$ — *N^2-вектор параметров (1) цепи Маркова,* $\hat\theta_l=(\hat p_{l,11},\ldots,\hat p_{l,NN})'$ — *N^2-вектор оценок (4). Тогда при $T_l\to\infty$ выполняется свойство асимптотической нормальности оценок* $(l=\overline{0,L})$:

$$L\left\{\sqrt{T_l}\left(\hat\theta_l-\theta_l\right)\right\}\to N_{N^2}\left(0,\Sigma(\theta_l)\right),\quad \Sigma(\theta_l)=\left(\sigma_{ijst}(\theta_l)\right),$$

$$\sigma_{ijst}(\theta_l)=1/\pi^*_{l,i}\left(\delta_{is}\left(\delta_{jt}p_{l,ij}-p_{l,ij}p_{l,st}\right)\right),\quad i,j,s,t=\overline{1,N}.$$

Лемма 2. *Пусть* $\xi_l=(P_{l,11},\ldots,P_{l,NN})'$ — *N^2-вектор параметров (2) цепи Маркова,* $\hat\xi_l=(\hat P_{l,11},\ldots,\hat P_{l,NN})'$ — *N^2-вектор оценок (4). Тогда при $T_l\to\infty$ выполняется свойство асимптотической нормальности оценок* $(l=\overline{0,L})$:

$$L\left\{\sqrt{T_l}\left(\hat\xi_l-\xi_l\right)\right\}\to N_{N^2}\left(0,\Sigma(\xi_l)\right),\quad \Sigma(\xi_l)=\left(\sigma_{ijst}(\xi_l)\right),$$

$$\sigma_{ijst}(\xi_l)=\pi^*_{l,i}P_{l,ij}\left(\delta_{is}\delta_{jt}-\pi^*_{l,s}p_{l,st}\right)+p_{l,ij}p_{l,st}\left(\pi^*_{l,i}c_{l,js}+\pi^*_{l,s}c_{l,ti}\right),\quad i,j,s,t=\overline{1,N},$$

где $c_{l,ij}=\sum_{k=0}^{\infty}\left(p^{(k)}_{l,ij}-\pi^*_{l,j}\right)<\infty$, $p^{(k)}_{l,ij}=\left(\left(P^{(l)}\right)^k\right)_{ij}$ — *вероятность перехода ОЦМ за k шагов из состояния i в состояние j при условии $\nu=l$.*

3. Решающее правило. Рассмотрим сначала случай известных параметров (1) классов. В [3] построено байесовское решающее правило (БРП), оптимальное по критерию (3), которое для случая двух классов (L=2) можно записать в виде:

$$d = d_0\left(X^{(0)}\right) = \mathbf{1}\left(\Lambda\left(X^{(0)}\right) - h\right) + 1, \tag{5}$$

$$\Lambda = \Lambda\left(X^{(0)}\right) = \sum_{i,j=1}^{N} \hat{P}_{0,ij} \ln \frac{p_{2,ij}}{p_{1,ij}}, \quad h = \frac{1}{T_0} \ln \frac{q_1 \pi_{1,x_{0,1}}}{q_2 \pi_{2,x_{0,1}}}, \tag{6}$$

где $\mathbf{1}(\cdot)$ — функция единичного скачка, $\hat{P}_{0,ij}$ определены в (4).

Примем обозначения:

$$a_l = (-1)^l \sum_{i,j=1}^{N} P_{l,ij} \ln \frac{p_{2,ij}}{p_{1,ij}}, \quad \sigma_l^2 = \sum_{i,j,s,t=1}^{N} \ln \frac{p_{2,ij}}{p_{1,ij}} \ln \frac{p_{2,st}}{p_{1,st}} \sigma_{ijst}(\xi_l), \quad l \in \{1,2\}, \tag{7}$$

где a_l — информационное расстояние Кульбака [2]. Отметим, что при сближении параметров классов $P^{(1)}$, $P^{(2)}$ величина $a_l/\sigma_l \to 0$, (l=1,2).

Теорема 1. *Если имеет место асимптотика растущей длительности наблюдения и сближающихся классов:*

$$T_0 \to \infty, \quad \sqrt{T_0}\frac{a_l}{\sigma_l} \to \delta_l > 0, \quad l \in \{1,2\},$$

то вероятность ошибочной классификации (3) БРП (5), (6) имеет предельное значение:

$$r_0 \to q_1 \Phi(-\delta_1) + q_2 \Phi(-\delta_2), \tag{8}$$

где $\Phi(\cdot)$ — функция распределения стандартного нормального закона.

Доказательство. Согласно (6) и утверждению леммы 2, Λ является линейной комбинацией асимптотически нормальных случайных величин, поэтому, в силу теоремы Андерсона [4] о функциональном преобразовании последовательности асимптотически нормальных случайных величин, условное распределение Λ при условии $\nu = l$ также является асимптотически нормальным:

$$L\left\{\sqrt{T_0}(\Lambda - a_l) \mid \nu = l\right\} \to N\left(0, \sigma_l^2\right), \quad T_0 \to \infty.$$

Далее доказательство аналогично последней части доказательства теоремы 2.■

Следствие. Если классы Ω_1, Ω_2 равновероятны ($q_1 = q_2$); $\sigma_1 = \sigma_2 = \sigma$, $a_2 = -a_1$ *и при $T_0 \to \infty$:*

$$\sqrt{T_0}\frac{J_K(2{:}1)}{\sigma} \to \delta, \quad \text{то} \quad r_0 \to \Phi\left(-\frac{\delta}{2}\right), \quad J_K(2{:}1) = \sum_{i,j=1}^{N}\left(P_{2,ij} - P_{1,ij}\right)\ln\frac{p_{2,ij}}{p_{1,ij}} > 0,$$

где $J_K(2{:}1)$ — информационное расхождение Кульбака [2].

Этот случай, например, всегда выполняется при N=2 и $P^{(l)} = P^{(l)\prime}$.

При неизвестных параметрах классов (1) будем пользоваться подстановочным РП (ПРП) $d = d_1(X^{(0)}, X)$, получаемым из БРП (5), (6) заменой неизвестных параметров (1) их оценками (4). Для удобства анализа запишем ПРП в виде (5) со следующими дискриминантной функцией и пороговым значением:

$$\hat{\Lambda} = \hat{\Lambda}\left(X^{(0)}, X\right) = \sum_{i,j=1}^{N} \hat{P}_{0,ij} \ln \frac{\hat{p}_{2,ij}}{\hat{p}_{1,ij}}, \quad h = \frac{1}{T_0} \ln \frac{q_1}{q_2}. \tag{9}$$

Определим $N^2 \times N^2$-матрицы:

$$\Sigma(\eta_l) = \left(\sigma_{ijst}(\eta_l)\right): \quad \Sigma(\eta_l) = \mathrm{diag}\left\{\Sigma_1(\eta_l), \ldots, \Sigma_N(\eta_l)\right\}, \quad i,j,s,t = \overline{1,N},$$

$$\Sigma_i(\eta_l) = 1/\pi_{l,i}^{*}\left(\mathrm{diag}\left\{1/p_{l,i1}, \ldots, 1/p_{l,iN}\right\} - \mathbf{1}_{N \times N}\right), \quad i = \overline{1,N}, l \in \{1,2\}, \tag{10}$$

где $\mathbf{1}_{N \times N}$ — $N \times N$-матрица, состоящая из единиц.

Теорема 2. *Если имеет место асимптотика растущих длительностей наблюдения и сближающихся классов:*

$$T_0, T_l \to \infty, T_l / T_0 \to \lambda_l \quad (0 < \lambda_l < \infty), \quad \sqrt{T_0}\frac{a_l}{\tilde{\sigma}_l} \to \tilde{\delta}_l > 0, \quad l \in \{1,2\},$$

$$\tilde{\sigma}_l^2 = \sigma_l^2 + \sum_{i,j,s,t=1}^{N} P_{l,ij} P_{l,st} \left(\frac{1}{\lambda_1}\sigma_{ijst}(\eta_1) + \frac{1}{\lambda_2}\sigma_{ijst}(\eta_2) \right) \tag{11}$$

и $p_{l,ij} > 0$, $i,j = \overline{1,N}$, *то вероятность ошибочной классификации (3) ПРП (5), (9) имеет предельное значение:*

$$r \to r_* = q_1 \Phi(-\tilde{\delta}_1) + q_2 \Phi(-\tilde{\delta}_2), \tag{12}$$

где $\{a_l, \sigma_l\}$ *определены в (7),* $\{\sigma_{ijst}(\eta_l)\}$ *задаются в (10).*

Доказательство. Согласно (5), (9), условные вероятности ошибочной классификации можно записать в виде:

$$\begin{aligned} r_1 &= \mathrm{P}\{d \neq \nu | \nu = 1\} = 1 - \mathrm{P}\left\{\hat{\Lambda} < 1/T_0 \ln q_1 | \nu = 1\right\}, \\ r_2 &= \mathrm{P}\{d \neq \nu | \nu = 2\} = \mathrm{P}\left\{\hat{\Lambda} < 1/T_0 \ln q_1 | \nu = 2\right\}. \end{aligned} \tag{13}$$

Введем следующие N^2-векторы: $\hat{\eta}_l = g(\hat{\theta}_l) = \left(g_{11}(\hat{\theta}_l), \ldots, g_{NN}(\hat{\theta}_l)\right)$, $\eta_l = g(\theta_l)$, где $g_{ij}(\hat{\theta}_l) = \ln \hat{p}_{l,ij}$ и $N^2 \times N^2$ – матрицы производных:

$$D^{(l)} = \left. \left(\frac{\partial g_{ij}}{\partial \hat{p}_{l,st}} \right) \right|_{\hat{\theta}_l = \theta_l} : \quad D^{(l)} = \mathrm{diag}\left\{B_1^{(l)}, \ldots, B_N^{(l)}\right\}, \quad i,j,s,t = \overline{1,N},$$

$$B_i^{(l)} = \mathrm{diag}\left\{1/p_{l,i1}, \ldots, 1/p_{l,iN}\right\}, \quad i = \overline{1,N}.$$

Тогда, согласно лемме 1 и теореме Андерсона [4] о функциональном преобразовании последовательности асимптотически нормальных случайных величин, при $T_0 \to \infty$ имеем:

$$L\left\{\sqrt{T_0}\left(\hat{\eta}_l - \eta_l\right)\right\} \to N_{N^2}\left(0, \frac{1}{\lambda_l}\Sigma(\eta_l)\right), \tag{14}$$

где $\Sigma(\eta_l) = D^{(l)}\Sigma(\theta_l)D^{(l)\prime}$ и имеет вид (10).

Введем составной вектор параметров $\hat{\varphi}_l = \left(\hat{\xi}_l' \vdots \hat{\eta}'\right)'$, где $\hat{\eta} = \hat{\eta}_2 - \hat{\eta}_1$. Из (9) нетрудно видеть, что $\hat{\Lambda} = f(\hat{\varphi}_l)$ при $\nu = l$ и соответствующие $1 \times N^2$-матрицы производных имеют вид:

$$d^{(l)} = \left. \left(\frac{\partial f}{\partial \hat{\varphi}_l} \right) \right|_{\hat{\varphi}_l = \varphi_l} : \left. \frac{\partial f}{\partial \hat{\xi}_{l,ij}} \right|_{\hat{\varphi}_l = \varphi_l} = \ln \frac{p_{2,ij}}{p_{1,ij}}, \quad \left. \frac{\partial f}{\partial \eta_{ij}} \right|_{\hat{\varphi}_l = \varphi_l} = P_{l,ij}.$$

Тогда, согласно лемме 2, (14) и ввиду независимости наблюдаемых ОЦМ, аналогично получаем:

$$L\left\{\sqrt{T_0}\left(\hat{\Lambda} - a_l\right) \middle| \nu = l\right\} \to N(0, \tilde{\sigma}_l^2),$$

где $\tilde{\sigma}_l^2 = d^{(l)}\Sigma(\varphi_l)d^{(l)\prime}$ и имеет вид (11). Тогда из (13) нетрудно видеть, что

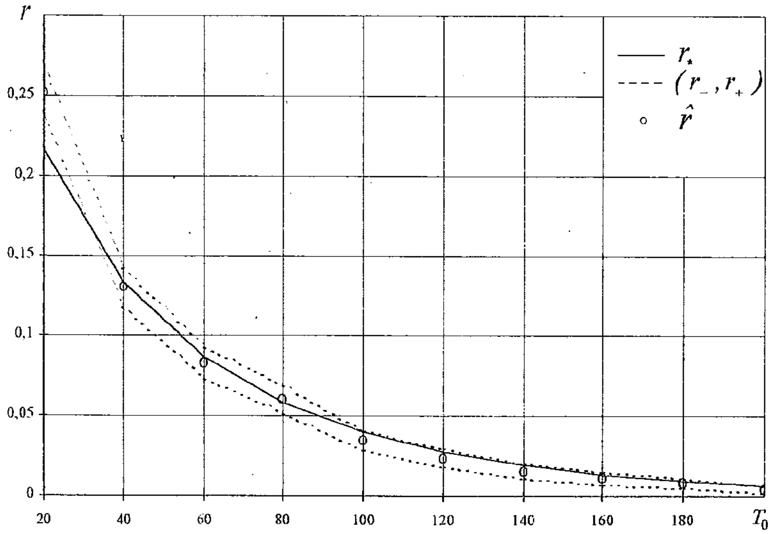
$$r_l \to \Phi\left((-1)^{l-1}\tilde{\delta}_l\right), \quad T_0 \to \infty,$$

откуда следует утверждение теоремы.■

Сравнивая (8) и (12), следует отметить, что если $\lambda_l \to \infty$ (т.е. при $T_0 \to \infty$ длины обучающих выборок увеличиваются быстрее T_0), то разность между вероятностями ошибочной классификации ПРП (5), (9) и БРП (5), (6) стремится к нулю.

4. Результаты численных экспериментов. Была проведена серия вычислительных экспериментов по оцениванию вероятности ошибочной классификации r (3), результаты которой хорошо согласуются с результатами (8), (12). К примеру, на рисунке приведены значения оценок $\hat{r}$ и 95%-доверительный интервал (r_-,r_+) для r, а также предельные значения r_*, вычисленные согласно (12) в случае $\lambda_1=\lambda_2=3$ и для следующих значений параметров классов:

$$q_1 = q_2 = 0{,}5, \quad P^{(1)} = \begin{pmatrix} 0{,}6 & 0{,}4 \\ 0{,}4 & 0{,}6 \end{pmatrix}, \quad P^{(2)} = \begin{pmatrix} 0{,}4 & 0{,}6 \\ 0{,}6 & 0{,}4 \end{pmatrix}.$$



Результаты численного эксперимента по оцениванию риска ПРП

Таким образом, в работе получены асимптотические выражения для вероятности ошибочной классификации ОЦМ при различных уровнях априорной неопределенности. Результаты численных экспериментов подтверждают работоспособность приведенных выражений.

1. Kharin Yu. Robustness in Statistical Pattern Recognition // Dordrecht, 1996.
2. Basawa I.V., Rao B.L.S. Statistical Inference for Stochastic Processes // New York, 1980.
3. Kostevich A.L. // Pattern recognition and Information Processing // Szczecin, 1997. Vol.1. P.82.
4. Serfling R.J. Approximation Theorems of Mathematical Statistics // New York, 1980.

Поступила в редакцию 19.06.97.

УДК 519.95

Н.М.ДМИТРУК

ОПТИМИЗАЦИЯ НЕОКЛАССИЧЕСКОЙ МОДЕЛИ ЭКОНОМИЧЕСКОГО РОСТА

The optimal behavior of the neo-classical economic growth model with constraints on consumption and investment is described.

1. Экономический рост – одна из центральных проблем макроэкономики. Ее первые математические модели исследованы в работах [1–3]. С теорией оптимального роста можно ознакомиться по [4]. Цель данной работы – с помощью достаточных условий оптимальности построить оптимальные управления типа обратной связи для неоклассической модели экономического роста.

2. Неоклассическая модель экономического роста Солоу описывает агрегированную замкнутую экономику. Рассмотрим развитие экономики