И.К. Пирштук
Учреждение БГУ «НИИ прикладных проблем математики и информатики»
e-mail: pirshtuk@bsu.by

ОБ ИССЛЕДОВАНИИ АСИМПТОТИЧЕСКОГО РЕШАЮЩЕГО ПРАВИЛА РАСПОЗНАВАНИЯ АВТОРЕГРЕССИОННЫХ ВРЕМЕННЫХ РЯДОВ ПРИ НАЛИЧИИ "ВЫБРОСОВ"

Рассматривается задача построения и исследования асимптотических решающих правил распознавания авторегрессионных временных рядов при наличии "выбросов" в классифицируемых наблюдениях. Предложено новое асимптотическое решающее правило. Приводятся результаты численных экспериментов.

Ключевые слова: временной ряд, авторегрессия, искажение наблюдений, распознавание, решающее правило.

1. ВВЕДЕНИЕ. МОДЕЛЬ АВТОРЕГРЕССИОННЫХ ВРЕМЕННЫХ РЯДОВ. ПОСТАНОВКА ЗАДАЧИ

Пусть в пространстве R^1 наблюдается временной ряд $X = (x_1,\dots, x_n) \in R^n$ длительностью n, относящийся к одному из L ($L\geq 2$) классов $\Omega_1, \dots, \Omega_L$ с априорными вероятностями $\pi_1, \dots, \pi_L$ соответственно ($\pi_1 + \dots + \pi_L = 1$). Временной ряд из класса Ω_i описывается моделью авторегрессии (АР(p)):

$$y_t = \theta_{i1}y_{t-1} + \dots + \theta_{ip}y_{t-p} + \upsilon_t, \qquad i=\overline{1,L}, \tag{1}$$

где $\theta_i = (\theta_{i1}, \dots , \theta_{ip})' \in R^p$ – вектор коэффициентов авторегрессии для Ω_i; υ_t – последовательность независимых одинаково распределенных (п.н.о.р.) гауссовских величин с нулевым средним и дисперсией B^2; p – порядок авторегрессии.

Будем предполагать, что корни характеристических уравнений ($i\in D$, $D=\{1, \dots, L\}$):

$$z^p + \sum_{j=1}^{p}\theta_{ij}^0 z^{p-j}=0,$$

лежат внутри единичного круга ($|z|<1$), и случайные процессы АР(p) из (1) являются стационарными [1, 2].

Значения временного ряда наблюдаются на фоне "выбросов":

$$x_t = y_t + \varsigma_t^{(i)}, \tag{2}$$

где y_t определяется соотношением (1), а $\varsigma_t^{(i)}$ – последовательность независимых случайных величин (ошибок), допускающих представление:

$$\varsigma_t^{(i)} = (1+(K_i-1)\eta_t^{(i)})\, \sigma\, \xi_t \qquad (i= \overline{1,L};\quad t= \overline{1,n}), \tag{3}$$

где ξ_t – п.н.о.р. случайных величин с плотностью распределения *q(z)*, имеющих нулевое среднее и единичную дисперсию; $\{\eta_t^{(i)}\}$ – не

И.К. Пирштук
Учреждение БГУ «НИИ прикладных проблем математики и информатики»
e-mail: pirshtuk@bsu.by

зависящая от $\{\xi_t\}$ последовательность бернуллиевских случайных величин:

$$P\{\eta_t^{(i)}=1\}=\varepsilon_i, \qquad P\{\eta_t^{(i)}=0\}=1-\varepsilon_i, \quad 0 \le \varepsilon_i \le \varepsilon_+, \tag{4}$$

$\sigma>0$, $K_i > 1$, ε_i, ε_+ – параметры. Из (3) и (4) видно, что дисперсия случайных ошибок равна σ^2 в те моменты времени, в которые $\eta_t^{(i)}=0$, и равна увеличенной дисперсии $K_i^2\sigma^2$в те моменты времени, в которые $\eta_t^{(i)}=1$ (в эти моменты и могут возникать "выбросы").

Неучет "выбросов" в модели (1) приводит к тому, что оптимальное (байесовское) решающее правило (БРП) [4], минимизирующее риск классификации при отсутствии "выбросов", теряет свою оптимальность при их появлении. Поэтому актуальной является задача построения и исследования эффективности робастных решающих правил (РРП).

Вектор параметров в модели (1) – (4) может содержать коэффициенты $\{\theta_i, K_i, \varepsilon_i\}$, σ, B. Если априорная информация о параметрах рассматриваемой модели отсутствует, то предполагается наличие обучающей выборки, на основе которой построено подстановочное БРП (ПБРП) [4].

Задача заключается в построении и исследовании асимптотического решающего правила для классификации наблюдаемой реализации $X = (x_1,\ldots, x_n) \in R^n$ длительностью n.

Рассмотренная выше модель допускает обобщение на случай многомерной авторегрессии, однако в данной работе, не ограничивая общности, рассматриваются одномерные временные ряды авторегрессии.

Примеры прикладных задач, для успешного решения которых необходимо разрешить выше сформулированные задачи:

– распознавание случайных сигналов [5];
– выделение и анализ сейсмических сигналов [6].

2. АСИМПТОТИЧЕСКОЕ РЕШАЮЩЕЕ ПРАВИЛО РАСПОЗНАВАНИЯ АВТОРЕГРЕССИОННЫХ ВРЕМЕННЫХ РЯДОВ ПРИ НАЛИЧИИ "ВЫБРОСОВ"

Для модели (1)–(4) при отсутствии искажений ($\varepsilon_+=0$), в настоящей статье рассматривается задача построения асимптотических робастных решающих правил распознавания временных рядов при наличии "выбросов" ($\varepsilon_+>0$) в классифицируемых наблюдениях, задаваемых соотношением (3).

В этой ситуации для построения асимптотических решающих правил воспользуемся вспомогательным результатом, который сформулируем в виде теоремы [3].

И.К. Пирштук
Учреждение БГУ «НИИ прикладных проблем математики и информатики»
e-mail: pirshtuk@bsu.by

Теорема. Если имеют место искажения модели авторегрессии (1) –(4), то совместная плотность распределения вероятностей выборки X = ($x_1,\ldots, x_n$) для i -го класса допускает асимптотическое разложение:

$$p_i(x) = n_n(x \mid 0,\Sigma^{(i)} + \sigma^2 \mathbf{I}_n) + \varepsilon_i \,\mathbf{(}\, n_n(x\mid 0,C^{(i)}) - n_n(x\mid 0, \Sigma^{(i)}+ \sigma^2 \mathbf{I}_n)\mathbf{)}, +O(\varepsilon_i^2), \quad (5)$$

где $C^{(i)}$ – известная матрица:

$$C^{(i)} = K_i^2\, \sigma^2\, \mathbf{I}_n + \Sigma^{(i)}, \quad (6)$$

а ковариационная матрица $\Sigma^{(i)}$ находится из уравнений Юла-Уокера для i-го класса, $\mathbf{I}_n$ – единичная матрица.

Асимптотическое решающее правило примет вид:

$$d = \arg \max_{i\in D}\ \pi_i\, p_i(X). \quad (7)$$

Для исследования риска асимптотического решающего правила (7), (5), (6) проводились численные эксперименты. Результаты исследований приведены ниже.

3. СРАВНИТЕЛЬНЫЙ АНАЛИЗ РАСПОЗНАВАНИЯ АВТОРЕГРЕССИОННЫХ ВРЕМЕННЫХ РЯДОВ

С целью исследования эффективности классификации авторегрессионных временных рядов проводились численные эксперименты на ПЭВМ для двух классов (L=2) с реализациями авторегрессии 2-го порядка. Число моделируемых временных рядов M=100 (по 50 из каждого класса). Ряды авторегрессии различались коэффициентами: θ_1 =(-0.7; 0.49)$'$, θ_2=(- 0.2; 0.04)$'$ соответственно. Здесь «$'$» – символ транспонирования.

Численные результаты сравнительного статистического анализа вероятности ошибочного распознавания авторегрессионных временных рядов при наличии искажений для байесовского, подстановочного байесовского и асимптотического робастного решающих правил (БРП, ПБРП, АРРП) приведены ниже в таблице; здесь же указаны точечные оценки вероятности ошибки и 90%-ные доверительные интервалы. Временные ряды из двух равновероятных классов различались векторами коэффициентов авторегрессии: θ_1=(0,4; 0,16)$'$, θ_2=(-0,4; 0,16)$'$. Искажения определялись параметрами: ε_i =0.2; K_i=10 (i= $\overline{1,L,}$). Из каждого класса было смоделировано по 50 реализаций (всего 100 реализаций) при изменении длительности наблюдения временного ряда от 10 до 100 отсчетов с шагом равным 10. Как видно из таблицы, эффективность асимптотического робастного решающего правила при наличии искажений, особенно при малых длительностях реализации ряда выше, чем эффективность байесовских решающих правил.

И.К. Пирштук
Учреждение БГУ «НИИ прикладных проблем математики и информатики»
e-mail: pirshtuk@bsu.by

Таблица – **Оценка эффективности статистического распознавания авторегрессионных временных рядов.**

Длительность ряда	Частота ошибок классификации		
	БРП	ПБРП	АРРП
10	0.10 [0.050, 0.150]	0.27 [0.197, 0.343]	0.06 [0.021, 0.099]
20	0.04 [0.008, 0.072]	0.03 [0.002, 0.058]	0.04 [0.008, 0.072]
30	0.01 [0.0, 0.026]	0.04 [0.008, 0.072]	0.01 [0.0,0.3]
40	0.02 [0.0, 0.043]	0.03 [0.002, 0.058]	0.01 [0.0,0.3]
50	0.01 [0.0, 0.026]	0.01 [0.0, 0.026]	0.0 [0.0,0.3]
60	0.0 [0.0, 0.03]	0.0 [0.0, 0.04]	0.0 [0.0,0.3]
70	0.0 [0.0, 0.03]	0.0 [0.0, 0.03]	0.0 [0.0,0.3]
80	0.0 [0.0, 0.03]	0.0 [0.0, 0.03]	0.0 [0.0,0.3]
90	0.0 [0.0, 0.03]	0.0 [0.0, 0.03]	0.0 [0.0,0.3]
100	0.0 [0.0, 0.03]	0.0 [0.0, 0.03]	0.0 [0.0,0.3]

Заметим, что особенностью временных рядов авторегрессии является факт зависимости элементов (отсчетов) ряда. В связи с этим распознаванию подлежит вся реализация временного ряда, а не один отсчет как при поточечном распознавании.

ЛИТЕРАТУРА

1. Андерсон Т. Статистический анализ временных рядов. М.: Мир., 1976. – 755 с.

2. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Вып. 1. – М.: Мир, 1974. – 406 с. Вып. 2. – М.: Мир, 1974. – 197 с.

3. Пирштук И.К. Асимптотическое решающее правило классификации авторегрессионных временных рядов при наличии "выбросов" // Статистический и прикладной анализ временных рядов: Материалы междунар. научн. конф. SAATS-97 (11-13 ноября 1997 г., Брест).– Брест, БрГУ, 1997. – С. 56-58.

4. Жук Е.Е., Пирштук И.К. Статистическая классификация временных рядов авторегрессии // Весці АН Беларусі. Сер. фіз.-мат. навук. – 1999. – № 2. – С. 28-34.

5. Сенин А.Г. Распознавание случайных сигналов. – Новосибирск: Наука, 1974. – 76 с.

6. Буа П. Некоторые вопросы применения аппарата распознавания образов при разведке на нефть и газ // Анализ и выделение сейсмических сигналов / Под ред. Ч. Чжаня.– М.: 1986.– С. 176-190.