

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ НА ОСНОВЕ ПРОСТРАНСТВЕННО-ВРЕМЕННЫХ ДАННЫХ

М. К. Журак

1. ПУАССОНОВСКАЯ УСЛОВНО АВТОРЕГРЕССИОННАЯ МОДЕЛЬ

Модели пространственно-временных статистических данных находят широкое применение при решении актуальных прикладных задач. В данной статье предлагается Пуассоновская условно авторегрессионная модель пространственно-временных наблюдений и решается задача статистического оценивания параметров модели.

Введем следующие обозначения: (Ω, F, P) – вероятностное пространство; $s \in S = \{1, 2, \dots, n\}$ – индексная переменная, кодирующая пространственные координаты географического объекта, n – число географических объектов; $t \in \mathfrak{T} = \{1, 2, \dots, T\}$ – дискретное время, T – длительность временного промежутка наблюдений; $x_{s,t} \in N_0 = N \cup \{0\}$ – дискретная случайная величина в момент t в точке s ; $F_{\bar{s}, < t} = \sigma\{x_{u,\tau} : u \neq s, \tau < t\} \subset F$ – σ -алгебра, порожденная указанными в скобках случайными величинами; $z_{s,t} \geq 0$ – наблюдаемый (известный) уровень фактора в точке s в момент t , влияющий на $x_{s,t}$; $\{\varphi_k(t) : 1 \leq k \leq K\}$ – заданный набор базисных функций, определяющих тренд ($\varphi_1(t) \equiv 1$); $\{\xi_{s,t}\}$ – независимые гауссовские случайные величины, $L\{\xi_{s,t}\} = N_1(0, \sigma_s^2)$.

Аналогично [1] построим Пуассоновскую условно авторегрессионную модель пространственно-временных наблюдений $\{x_{s,t}\}$:

$$L\{x_{s,t} | F_{\bar{s}, < t}\} = \Pi(\lambda_{s,t}), \quad \lambda_{s,t} = \Lambda_{s,t} e^{\xi_{s,t}}, \quad s \in S, t \in \mathfrak{T},$$

$$\Lambda_{s,t} = \exp\left\{\sum_{\tau=1}^n \sum_{i=1}^p \alpha_{s\tau i} x_{\tau, t-i} + \sum_{j=1}^q \beta_{sj} z_{s, t-j-1} + \sum_{k=1}^K \gamma_{sk} \varphi_k(t)\right\}, \quad (1)$$

где $\{\alpha_{s\tau i}, \beta_{sj}, \gamma_{sk}, \sigma_s^2\}$ – параметры модели, $p, q, K \in N_0$ – заданные числа.

Задача состоит в статистическом оценивании указанных выше параметров модели (1) по наблюдениям $\{x_{s,t} : s \in S, t \in \mathfrak{T}\}$.

2. ВЫЧИСЛЕНИЕ ОЦЕНОК МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Учитывая, что оптимальные статистические выводы [2] основаны на функции правдоподобия, рассмотрим задачу построения функции правдоподобия.

Примем матричные обозначения:

$$D_s = (\alpha_{s11}, \dots, \alpha_{s1p}, \dots, \alpha_{sn1}, \dots, \alpha_{snp}, \beta_{s1}, \dots, \beta_{sq}, \gamma_{s1}, \dots, \gamma_{sK})' \in R^{pn+q+K},$$

$$\theta_s = (D_s', \sigma_s^2)' \in R^{pn+q+K+1}, \quad \theta = (\theta_1', \dots, \theta_n')' \in R^{n(pn+q+K+1)},$$

$$\xi_s = (\xi_{s,1}, \dots, \xi_{s,T})' \in R^T, \quad \xi = (\xi_1', \dots, \xi_n')' \in R^{nT},$$

$$X_{(t)} = (x_{1,t}, x_{2,t}, \dots, x_{n,t})' \in R^n, \quad s \in S, t \in \mathfrak{T},$$

где штрих обозначает транспонирование матрицы, θ – составной вектор параметров модели, подлежащих оцениванию.

Построим вначале условную функцию правдоподобия $L(\theta | \xi)$ при фиксированном случайном векторе ξ .

Теорема 1. Для определенной выше модели (1) пространственно-временных данных $\{x_{s,t} : s \in S, t = 1, 2, \dots, T\}$ условная функция правдоподобия при условии, что фиксирован случайный вектор ξ , имеет вид:

$$L(\theta | \xi) = P\{X_{(1)}, \dots, X_{(p)} | \xi\} \prod_{s=1}^n \prod_{t=p+1}^T \frac{\Lambda_{s,t}^{x_{s,t}}}{x_{s,t}!} \cdot e^{x_{s,t} \xi_{s,t} - \Lambda_{s,t}} e^{\xi_{s,t}}. \quad (2)$$

Введем в рассмотрение специальную функцию $f(k, a, b, \sigma^2) = E\{\eta^{2k} e^{a\eta - b e^\eta}\}$, где $a \geq 0, b \geq 0, k \geq 0, \sigma^2 > 0, L\{\eta\} = N_1(0, \sigma^2)$:

$$f(k, a, b, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} y^{2k} e^{a y - b e^y} e^{-\frac{y^2}{2\sigma^2}} dy. \quad (3)$$

Теорема 2. В условиях теоремы 1 логарифмическая функция правдоподобия для совокупности всех наблюдений $\{x_{s,t} : s \in S, t = 1, 2, \dots, T\}$ имеет следующий вид:

$$l(\theta) = \sum_{s=1}^n \sum_{t=p+1}^T \left(x_{s,t} \ln \Lambda_{s,t} - \ln x_{s,t}! + \ln f(0, x_{s,t}, \Lambda_{s,t}, \sigma_s^2) \right).$$

Из модели (1) и теоремы 2 следует, что задача нахождения ОМП параметров модели распадается на n задач:

$$l_s(\theta_s) \rightarrow \max_{\theta_s}, s \in S. \quad (4)$$

Необходимое условие для локального максимума в (4) имеет вид:

$$\nabla_{\theta_s} l_s(\theta_s) = 0. \quad (5)$$

Задачу (5) будем решать численно, применяя итерационный метод Ньютона [3], который обладает квадратичной сходимостью. Для этого метода $(k + 1)$ -ая итерация имеет вид $(k = 0, 1, 2, \dots)$:

$$\theta_s^{(k+1)} = \theta_s^{(k)} - (H^{(k)})^{-1} \cdot (\nabla l_s)^{(k)}, \quad (6)$$

где $(\nabla l_s)^{(k)}$ – вектор первых производных в точке $\theta_s^{(k)}$, $H^{(k)}$ – матрица вторых производных в точке $\theta_s^{(k)}$, $\theta_s^{(0)}$ – начальное приближение. Итерационные вычисления заканчиваем, если $\|\nabla l_s(\theta_s^{(k+1)})\| < \varepsilon$, где $\varepsilon \geq 0$ – наперед заданная достаточно малая величина, определяющая точность вычисления ОМП; при этом в качестве решения принимаем $\theta_s^{(k+1)}$.

Для задания вектора начального приближения $\theta_s^{(0)}$ для итерационного алгоритма (6) вместо модели (1) построим аппроксимационную модель, учитывая, что $L\{x_{s,t} | F_{\bar{s}, < t}\} = \Pi(\lambda_{s,t})$, то $E\{x_{s,t} | F_{\bar{s}, < t}\} = \lambda_{s,t}$:

$$\ln x_{s,t} = \sum_{\tau=1}^n \sum_{i=1}^p \tilde{\alpha}_{s\tau i} x_{\tau, t-i} + \sum_{j=1}^q \tilde{\beta}_{sj} z_{s, t-j-1} + \sum_{k=1}^K \tilde{\gamma}_{sk} \Phi_k(t) + \tilde{\xi}_{s,t}, \quad s \in S, t \in \mathfrak{T}, \quad (7)$$

где $\{\tilde{\xi}_{s,t}\}$ – независимые гауссовские случайные величины, $L\{\tilde{\xi}_{s,t}\} = N_1(0, \tilde{\sigma}_s^2)$. Оценки параметров модели (7) можно вычислить по методу наименьших квадратов.

При реализации выше представленного алгоритма вычисления ОМП возникает задача вычисления функции $f(\cdot)$, определенной в (3).

Лемма. Функция $f(k, a, b, \sigma^2)$, $a \geq 0$, $b \geq 0$, $k \geq 0$, $\sigma^2 > 0$, входящая в выражение функции правдоподобия и ее производных и определяемая формулой (3), ограничена и может быть вычислена приближенно при

помощи квадратурной формулы $f(k, a, b, \sigma^2) = \sum_{j=0}^n A_j g(y_j) + R_n(g)$, где

$$g(y) = \frac{(\sqrt{2\sigma^2})^{2k}}{\sqrt{\pi}} y^{2k} e^{a\sqrt{2\sigma^2}y - be^{\sqrt{2\sigma^2}y}}, \quad \{y_j\} - \text{корни многочлена Эрмита,}$$

$\{A_j\}$ – коэффициенты квадратурной формулы, которые могут быть

вычислены по формуле $A_j = \frac{2^n(n+1)!\sqrt{\pi}}{(n+1)^2(H_n(y_j))^2}$, $j=0,1,\dots,n$. Остаток

имеет вид $R_n(g) = \frac{(n+1)!\sqrt{\pi}}{2^{n+1}(2n+2)!} \max_{\eta \in (-\infty, \infty)} g^{(2n+2)}(\eta)$.

3. РЕЗУЛЬТАТЫ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

Компьютерные эксперименты проводились на модельных данных. Для модели (1) при $n=3$, $p=1$, $q=0$, $K=2$, $S=\{1, 2, 3\}$, $\varphi_1(t)=1$, $\varphi_2(t)=t$, $\theta_1=(-0.1, -0.01, -0.01, 2.2, -0.01, 0.4)'$, $\theta_2=(-0.1, -0.014, -0.01, 2.5, -0.01, 0.5)'$, $\theta_3=(-0.1, -0.01, -0.015, 2.1, -0.01, 0.3)'$.

Оценим среднеквадратическую погрешность оценивания параметров методом статистического моделирования:

$$r_T(\theta) = \frac{1}{F} \sum_{k=1}^F \|\hat{\theta}^{(k)} - \theta\|^2,$$

где $\hat{\theta}^{(k)} \in R^{18}$ – оценка параметров по k -ой реализации пространственно-временных данных, $\theta \in R^{18}$ – истинное значение параметров, F – количество реализаций. На рисунке 1 изображены графики зависимостей $r_T(\theta)$ от T для ОМП модели (1) и для МНК-оценок аппроксимационной модели (7) (T изменяется от 20 до 240, итерационный алгоритм (6) применялся для 10 начальных значений, $F=500$).

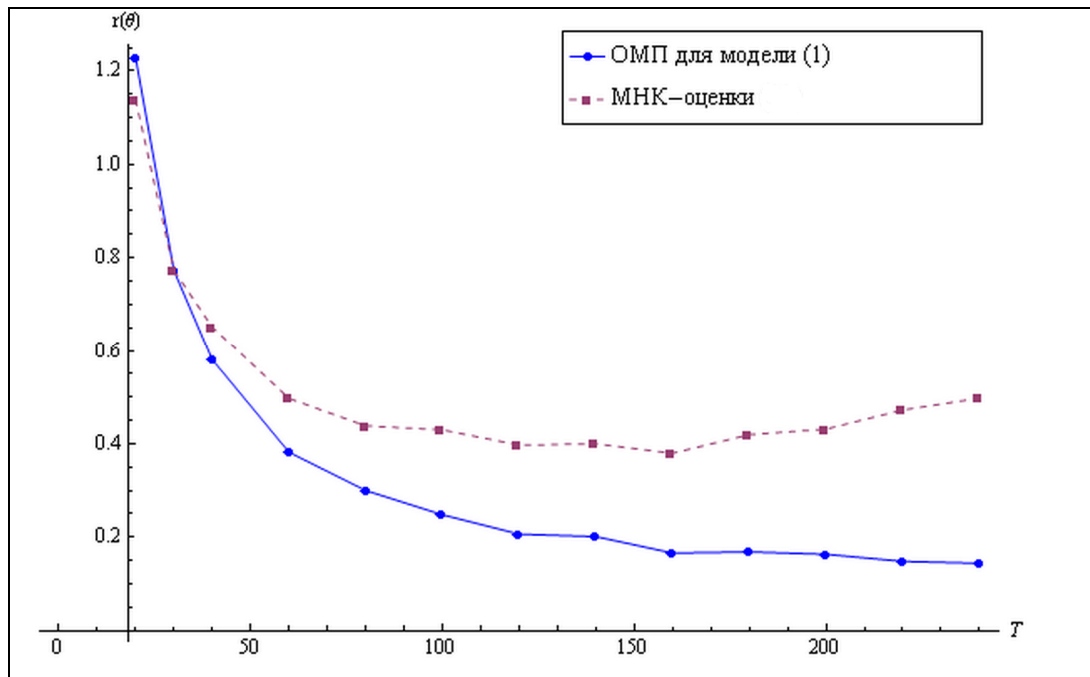


Рис. 1. Среднеквадратическая погрешность ОМП модели (1) и МНК-оценок модели (7)

Рисунок 1 иллюстрирует свойство состоятельности ОМП для модели (1). Видно также, что при $T > 30$ среднеквадратическая погрешность оценивания параметров для модели (1) меньше, чем для МНК-оценок аппроксимационной модели (7).

Литература

1. *Mariella L., Tarantino M.* Spatial temporal conditional Auto-Regressive Model: A New Autoregressive Matrix // Austrian journal of Statistic. 2010. Vol. 3. P. 223–244.
2. *Боровков А. А.* Математическая статистика. М., 1984.
3. *Альсевич В. В., Крахотко В. В.* Методы оптимизации: упражнения и задания: учебное пособие. Минск, 2005.

КОНВЕЙЕРНЫЕ ВЫЧИСЛЕНИЯ С ОБРАТНЫМИ СВЯЗЯМИ

И. В. Жураховский

Конвейер – один из способов обработки данных, заключающийся в том, что программа разбивается на последовательность независимых стадий, каждая из которых обрабатывается на своем узле конвейера. Мы рассмотрим модель конвейера, узлами которого служат процессы, поддерживающую сложную структуру программы.

Использование конвейера на процессах имеет свои преимущества. Если есть в наличии несколько ядер процессора, то можно распределить узлы-процессы между ними, повысив скорость обработки. Отдельные узлы могут содержать законченные подпрограммы, что позволяет проверять работоспособность каждого узла по отдельности, а также позволяет представить главную программу как конструкцию из готовых модулей.

В качестве узлов модели для простоты мы возьмем консольные приложения Windows, а качестве связей между ними – анонимные каналы. Сам конвейер предполагается работающим на единственном компьютере.

Анонимные каналы обеспечивают межпроцессное взаимодействие на локальном компьютере [1]. Каждый канал характеризуется дескрипторами входа и выхода. Данные, попадающие на вход канала из одного процесса, можно получить в порядке FIFO на выходе другого процесса.

Консольное приложение по умолчанию имеет три стандартных дескриптора – ввода, вывода и ошибок [2]. Для обеспечения связи мы будем заменять стандартные дескрипторы дескрипторами каналов.

Прежде всего, мы построим классический линейный конвейер. Запуск процессов-узлов и установление связей между ними мы организуем в главном процессе. Для переопределения стандартных дескрипторов можно поступить, например, следующим образом: сделать требуемые дескрипторы каналов наследуемыми, установить их вместо стандарт-