

# 一阶混合整数值负二项自回归模型

李 晗<sup>1</sup>, 连 成<sup>1</sup>, 方引芳<sup>1</sup>, 杨 凯<sup>2</sup>

(1. 长春大学 数学与统计学院, 长春 130022; 2. 长春工业大学 数学与统计学院, 长春 130012)

**摘要:** 考虑复杂整数值时间序列数据的建模问题。首先, 提出一类一阶混合整数值负二项自回归模型, 并证明该模型的严平稳遍历性, 讨论该模型的转移概率、期望、方差等概率统计性质; 其次, 研究该模型的最大似然估计问题, 得到了估计量的渐近正态性, 并在数值模拟的基础上进行实证分析。实证分析结果表明, 该模型在拟合毒品犯罪次数数据时性能良好。

**关键词:** 整数值时间序列; 混合负二项自回归模型; 平稳性; 极大似然估计

**中图分类号:** O212 **文献标志码:** A **文章编号:** 1671-5489(2024)03-0547-09

## First-Order Mixed Integer-Valued Negative Binomial Autoregressive Models

LI Han<sup>1</sup>, LIAN Cheng<sup>1</sup>, FANG Yinfang<sup>1</sup>, YANG Kai<sup>2</sup>

(1. School of Mathematics and Statistics, Changchun University, Changchun 130022, China;

2. School of Mathematics and Statistics, Changchun University of Technology, Changchun 130012, China)

**Abstract:** We considered the modeling problem of complex integer-valued time series data. Firstly, we proposed a class of first-order mixed integer-valued negative binomial autoregressive models, proved the strict stationary and ergodicity of the model, and discussed the probabilistic and statistical properties of the model such as transition probability, expectation, variance, etc. Secondly, we studied the maximum likelihood estimation problem of the model, obtained the asymptotic normality of the estimator, and conducted empirical analysis on the basis of numerical simulations. The empirical analysis results show that the model performs well in fitting the drug offense count data.

**Keywords:** integer-valued time series; mixed negative binomial autoregressive model; stationarity; maximum likelihood estimation

## 0 引言

整数值时间序列是由某种现象的某一统计指标在不同时间上的状态形成的相依计数数据, 这类数据广泛应用于通讯保障、医疗卫生、保险精算等领域。自 Al-Osh 等<sup>[1]</sup>提出整数值自回归(INAR)模型以来, 对这类数据的建模和推断研究备受关注<sup>[2-8]</sup>。Du 等<sup>[2]</sup>将 INAR(1)模型进行了推广, 提出了  $p$  阶整数值自回归模型, 即 INAR( $p$ )模型。INAR( $p$ )模型是基于二项稀疏算子<sup>[3]</sup>建立的, 在许多实际问题中具有较强的解释性。Neal 等<sup>[4]</sup>和张哲等<sup>[5]</sup>先后研究了这类模型的估计问题。二项稀疏算子“ $\circ$ ”定义

收稿日期: 2023-08-20.

**第一作者简介:** 李 晗(1988—), 女, 蒙古族, 博士, 讲师, 从事时间序列分析的研究, E-mail: lihancu@163.com. **通信作者简介:** 杨 凯(1984—), 男, 汉族, 博士, 副教授, 从事时间序列分析的研究, E-mail: yangkai@ccut.edu.cn.

**基金项目:** 国家自然科学基金(批准号: 11901053)、吉林省自然科学基金(批准号: YDZJ202301ZYTS393)和吉林省教育厅科学研究项目(批准号: JJKH20230665KJ).

如下:

$$\alpha \circ X = \begin{cases} \sum_{j=1}^X B_j, & X > 0, \\ 0, & X = 0, \end{cases} \quad (1)$$

其中  $\alpha \in [0, 1)$ ,  $X$  是非负整数值随机变量,  $\{B_i\}$  是以  $\alpha$  为参数的独立同分布(i. i. d.)的 Bernoulli 随机变量序列, 并且独立于  $X$ , 满足  $P(B_i=1)=1-P(B_i=0)=\alpha$ . 在式(1)中, 求和序列  $\{B_i\}$  只能取值 0 或 1, 使得  $\alpha \circ X$  的取值总是小于等于  $X$ . 在有些实际数据中, 二项稀疏算子可能不合适, 一个最典型的例子是传染病数据. 若  $X_t$  表示  $t$  时刻某种传染疾病的人数, 则二项稀疏算子“ $\circ$ ”不再适用. 因为在  $t-1$  时刻至  $t$  时刻, 病人可能将病毒传染给不止一个人. 为解决上述问题, Ristić 等<sup>[6]</sup>定义了一种新的算子, 即负二项稀疏算子, 并基于该算子建立了一类以几何分布为边际分布的一阶整数值自回归(NGINAR(1))模型. 相对于 Poisson INAR(1)模型, NGINAR(1)模型更适合拟合具有过渡离散特征的数据. 文献[9-11]进一步考虑了对二项稀疏算子的改进方法和建模问题.

实际应用中有许多整数值时间序列数据, 呈现出多峰、非对称、过渡离散等特征<sup>[12]</sup>. 上述整数值时间序列模型的模型结构过于单一, 无法满足具有多峰、非对称等复杂特征的整数值时间序列的建模需求. 而解决该问题的有效方法是考虑用混合模型的思想对数据进行建模. 例如: Ristić 等<sup>[13]</sup>基于两个稀疏算子建立了具有不同边际分布的一阶和高阶混合整数值自回归模型; Li 等<sup>[14]</sup>基于广义幂级数稀疏算子提出了一类零堆积混合整数自回归模型; Zhang 等<sup>[15]</sup>基于一类相依稀疏算子提出了混合广义二项自回归模型. 本文将基于负二项稀疏算子提出一类一阶混合整数值负二项自回归模型, 研究模型的性质和参数估计等问题, 并讨论其在实际数据中的应用.

## 1 模型的定义和性质

### 1.1 模型的建立

**定义 1** 取值于非负整数集上的随机变量序列  $\{X_t\}$  称为一阶混合整数值负二项自回归模型(记作 MNBINAR(1)模型), 若其满足:

$$X_t = \begin{cases} \alpha_1 * X_{t-1} + \epsilon_{1,t}, & \text{w. p. } p, \\ \alpha_2 * X_{t-1} + \epsilon_{2,t}, & \text{w. p. } 1-p, \end{cases} \quad (2)$$

其中: w. p. 表示以某概率取值;  $\alpha_i \in [0, 1)$ ,  $i=1, 2$ ;  $p$  是混合概率, 为确保模型的可识别性,  $p \in [0.5, 1]$ <sup>[16-17]</sup>; “ $*$ ”是负二项稀疏算子<sup>[6]</sup>, 定义为

$$\alpha * X = \begin{cases} \sum_{j=1}^X W_j, & X > 0, \\ 0, & X = 0, \end{cases}$$

这里  $W_j$  是独立同分布的、参数为  $\frac{\alpha}{1+\alpha}$  的几何分布随机变量序列, 且与  $X$  独立;  $\{\epsilon_{i,t}\}$  为 i. i. d. 的离散随机扰动序列, 其取值范围为自然数集合  $\mathbb{N}$ , 且  $E(\epsilon_{i,t}) = \mu_i$ ,  $\text{Var}(\epsilon_{i,t}) = \sigma_i^2$ , 概率质量函数为  $f(\epsilon_{i,t})$ , 同时对于  $\forall k \in \mathbb{N}$ , 都有  $f(\epsilon_{i,t}=k) > 0$  恒成立; 对固定的  $t$  和  $i$ ,  $\epsilon_{i,t}$  和  $\alpha_i * X_{t-l}, X_{t-l} (l \geq 1)$  独立,  $\{\epsilon_{1,t}\}$  与  $\{\epsilon_{2,t}\}$  相互独立.

引入一个 i. i. d. 的随机变量序列  $\{D_t\}$ , 它与  $\{\alpha_i * X_{t-l}\}, \{X_{t-l}\}, \{\epsilon_{i,t}\} (i=1, 2, l \geq 1)$  独立, 且满足  $P(D_t=1)=1-P(D_t=0)=p$ , 即  $D_t \sim B(1, p)$ , 则由式(2)定义的 MNBINAR(1)模型可等价表示为如下形式:

$$X_t = D_t \cdot (\alpha_1 * X_{t-1} + \epsilon_{1,t}) + (1 - D_t) \cdot (\alpha_2 * X_{t-1} + \epsilon_{2,t}). \quad (3)$$

记  $p_{x_t | x_{t-1}} := P(X_t = x_t | X_{t-1} = x_{t-1})$  为 MNBINAR(1)过程的转移概率, 则由式(3)可知

$$p_{x_t | x_{t-1}} = p \sum_{h=0}^{x_t} [P(\alpha_1 * X_{t-1} = h | X_{t-1} = x_{t-1}) P(\epsilon_{1,t} = x_t - h)] +$$

$$(1 - p) \sum_{h=0}^{x_t} [P(\alpha_2 * X_{t-1} = h | X_{t-1} = x_{t-1}) P(\epsilon_{2,t} = x_t - h)]. \tag{4}$$

该转移概率可用于构建模型的似然函数.

### 1.2 MNBINAR(1)的性质

下面给出 MNBINAR(1)模型一些重要的概率统计性质. 首先给出模型的平稳遍历性, 然后给出模型的一些矩和条件矩.

**命题 1** 由模型(2)定义的 MNBINAR(1)过程是遍历的, 且具有唯一的平稳分布.

证明: 由定义 1 可知, MNBINAR(1)过程  $\{X_t\}$  是状态空间  $\mathbb{N}$  上的 Markov 链. 由式(4)可知, MNBINAR(1)过程的转移概率  $p_{x_t | x_{t-1}}$  恒大于 0, 所以  $\{X_t\}$  是一个不可约且非周期的 Markov 链. 下面证明  $\{X_t\}$  是正常返的. 为此, 首先将式(2)等价表示为如下形式:

$$X_t = \alpha_t * X_{t-1} + \epsilon_t, \tag{5}$$

其中  $\alpha_t = D_t \alpha_1 + (1 - D_t) \alpha_2$ ,  $\epsilon_t = D_t \epsilon_{1,t} + (1 - D_t) \epsilon_{2,t}$ . 将式(5)进行  $(t-1)$ 次迭代, 可得

$$X_t = \alpha_t * \alpha_{t-1} * \dots * \alpha_1 * X_0 + \epsilon_t + \sum_{i=1}^{t-1} \alpha_t * \alpha_{t-1} * \dots * \alpha_{i+1} * \epsilon_i,$$

从而

$$\begin{aligned} P(X_t = 0 | X_0 = 0) &= P(\epsilon_t + \sum_{i=1}^{t-1} \alpha_t * \alpha_{t-1} * \dots * \alpha_{i+1} * \epsilon_i = 0 | X_0 = 0) = \\ &P(\epsilon_t = 0, \alpha_t * \epsilon_{t-1} = 0, \alpha_t * \alpha_{t-1} * \epsilon_{t-2} = 0, \dots, \alpha_t * \alpha_{t-1} * \dots * \alpha_2 * \epsilon_1 = 0 | X_0 = 0) = \\ &P(\epsilon_t = 0) P(\alpha_t * \epsilon_{t-1} = 0) P(\alpha_t * \alpha_{t-1} * \epsilon_{t-2} = 0) \dots P(\alpha_t * \alpha_{t-1} * \dots * \alpha_2 * \epsilon_1 = 0) = \\ &P(\epsilon_t = 0) \left[ \sum_{k=0}^{+\infty} P(\alpha_t * \epsilon_{t-1} = 0 | \epsilon_{t-1} = k) P(\epsilon_{t-1} = k) \right] \times \\ &\left[ \sum_{k=0}^{+\infty} P(\alpha_t * \alpha_{t-1} * \epsilon_{t-2} = 0 | \epsilon_{t-2} = k) P(\epsilon_{t-2} = k) \right] \times \dots \times \\ &\left[ \sum_{k=0}^{+\infty} P(\alpha_t * \alpha_{t-1} * \dots * \alpha_2 * \epsilon_1 = 0 | \epsilon_1 = k) P(\epsilon_1 = k) \right] = \\ &P(\epsilon_t = 0) \left[ P(\epsilon_{t-1} = 0) + \sum_{k=1}^{+\infty} P(\alpha_t * \epsilon_{t-1} = 0 | \epsilon_{t-1} = k) P(\epsilon_{t-1} = k) \right] \times \\ &\left[ P(\epsilon_{t-2} = 0) + \sum_{k=1}^{+\infty} P(\alpha_t * \alpha_{t-1} * \epsilon_{t-2} = 0 | \epsilon_{t-2} = k) P(\epsilon_{t-2} = k) \right] \times \dots \times \\ &\left[ P(\epsilon_1 = 0) + \sum_{k=1}^{+\infty} P(\alpha_t * \alpha_{t-1} * \dots * \alpha_2 * \epsilon_1 = 0 | \epsilon_1 = k) P(\epsilon_1 = k) \right] \geq \\ &P(\epsilon_t = 0) P(\epsilon_{t-1} = 0) \dots P(\epsilon_1 = 0) > 0. \end{aligned}$$

于是有  $\sum_{t=1}^{+\infty} P(X_t = 0 | X_0 = 0) = +\infty$ . 根据文献[18]中命题 4.2.3 和推论 4.2.4 可知,  $\{X_t\}$  是一个常返 Markov 链. 因为  $P(X_t = 0 | X_0 = 0) > 0$ , 故由文献[18]中定理 4.3.3 可知, 状态 0 是正常返的. 又因为  $\{X_t\}$  是一个不可约的 Markov 链, 所以  $\{X_t\}$  是一个正常返的 Markov 链. 从而 MNBINAR(1)过程是遍历的. 又根据文献[19]中定理 1 可知,  $\{X_t\}$  存在唯一的平稳分布. 证毕.

**命题 2** 设  $\{X_t\}$  满足式(2), 则对于  $t \geq 1$ , 令  $E(\epsilon_{i,t}) = \mu_i$ ,  $\text{Var}(\epsilon_{i,t}) = \sigma_i^2 (i=1, 2)$ , 有:

- 1)  $E(X_t | X_{t-1}) = [p\alpha_1 + (1-p)\alpha_2] X_{t-1} + p\mu_1 + (1-p)\mu_2$ ;
- 2)  $\text{Var}(X_t | X_{t-1}) = p(1-p)(\alpha_1 - \alpha_2)^2 X_{t-1}^2 + [p\alpha_1(1+\alpha_1) + (1-p)\alpha_2(1+\alpha_2) + 2p(1-p)(\alpha_1 - \alpha_2)(\mu_1 - \mu_2)] X_{t-1} + p\sigma_1^2 + (1-p)\sigma_2^2 + p(1-p)(\mu_1 - \mu_2)^2$ ;
- 3)  $E(X_t) = \frac{p\mu_1 + (1-p)\mu_2}{1 - p\alpha_1 - (1-p)\alpha_2}$ ;
- 4)  $\text{Var}(X_t) = \{p(1-p)(\alpha_1 - \alpha_2)^2 [E(X_t)]^2 + [p\alpha_1(1+\alpha_1) + (1-p)\alpha_2(1+\alpha_2) +$

$$2p(1-p)(\alpha_1 - \alpha_2)(\mu_1 - \mu_2)]E(X_t) + p\mu_1 + (1-p)\mu_2 + p(1-p)(\mu_1 - \mu_2)^2\} / \\ [1 - p\alpha_1^2 - (1-p)\alpha_2^2];$$

$$5) \text{Cov}(X_t, X_{t+h}) = [p\alpha_1 + (1-p)\alpha_2]^h \text{Var}(X_t).$$

证明: 首先证明 1) 成立. 根据式(3)可知,

$$E(X_t | X_{t-1}) = E[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}) + (1 - D_t) \cdot (\alpha_2 * X_{t-1} + \varepsilon_{2,t}) | X_{t-1}] = \\ E[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}) | X_{t-1}] + E[(1 - D_t) \cdot (\alpha_2 * X_{t-1} + \varepsilon_{2,t}) | X_{t-1}] = \\ pE(\alpha_1 * X_{t-1} + \varepsilon_{1,t} | X_{t-1}) + (1 - p)E(\alpha_2 * X_{t-1} + \varepsilon_{2,t} | X_{t-1}) = \\ p(\alpha_1 X_{t-1} + \mu_1) + (1 - p)(\alpha_2 X_{t-1} + \mu_2) = \\ [p\alpha_1 + (1 - p)\alpha_2]X_{t-1} + p\mu_1 + (1 - p)\mu_2.$$

其次, 证明 2) 成立. 由式(3)以及方差性质可知,

$$\text{Var}(X_t | X_{t-1}) = \text{Var}[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}) + (1 - D_t) \cdot (\alpha_2 * X_{t-1} + \varepsilon_{2,t}) | X_{t-1}] = \\ \text{Var}[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}) | X_{t-1}] + \text{Var}[(1 - D_t) \cdot (\alpha_2 * X_{t-1} + \varepsilon_{2,t}) | X_{t-1}] + \\ 2\text{Cov}[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}), (1 - D_t) \cdot (\alpha_2 * X_{t-1} + \varepsilon_{2,t}) | X_{t-1}],$$

其中, 由条件方差公式可得

$$\text{Var}[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}) | X_{t-1}] = \text{Var}\{E[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}) | X_{t-1}, D_t] | X_{t-1}\} + \\ E\{\text{Var}[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}) | X_{t-1}, D_t] | X_{t-1}\} = \\ \text{Var}[D_t \cdot (\alpha_1 X_{t-1} + \mu_1) | X_{t-1}] + E\{D_t^2 \cdot [\alpha_1(1 + \alpha_1)X_{t-1} + \sigma_1^2] | X_{t-1}\} = \\ (\alpha_1 X_{t-1} + \mu_1)^2 \text{Var}(D_t) + [\alpha_1(1 + \alpha_1)X_{t-1} + \sigma_1^2]E(D_t^2) = \\ (\alpha_1 X_{t-1} + \mu_1)^2 p(1 - p) + [\alpha_1(1 + \alpha_1)X_{t-1} + \sigma_1^2]p,$$

同理可得

$$\text{Var}[(1 - D_t) \cdot (\alpha_2 * X_{t-1} + \varepsilon_{2,t}) | X_{t-1}] = (\alpha_2 X_{t-1} + \mu_2)^2 p(1 - p) + [\alpha_2(1 + \alpha_2)X_{t-1} + \sigma_2^2](1 - p).$$

又因为  $D_t(1 - D_t)$  恒为 0, 从而

$$\text{Cov}[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}), (1 - D_t) \cdot (\alpha_2 * X_{t-1} + \varepsilon_{2,t}) | X_{t-1}] = \\ -E[D_t \cdot (\alpha_1 * X_{t-1} + \varepsilon_{1,t}) | X_{t-1}]E[(1 - D_t) \cdot (\alpha_2 * X_{t-1} + \varepsilon_{2,t}) | X_{t-1}] = \\ -pE(\alpha_1 * X_{t-1} + \varepsilon_{1,t} | X_{t-1}) \cdot (1 - p)E(\alpha_2 * X_{t-1} + \varepsilon_{2,t} | X_{t-1}) = \\ -p(1 - p)(\alpha_1 X_{t-1} + \mu_1)(\alpha_2 X_{t-1} + \mu_2),$$

综上, 整理可得

$$\text{Var}(X_t | X_{t-1}) = p(1 - p)(\alpha_1 - \alpha_2)^2 X_{t-1}^2 + [p\alpha_1(1 + \alpha_1) + (1 - p)\alpha_2(1 + \alpha_2) + \\ 2p(1 - p)(\alpha_1 - \alpha_2)(\mu_1 - \mu_2)]X_{t-1} + \\ p\sigma_1^2 + (1 - p)\sigma_2^2 + p(1 - p)(\mu_1 - \mu_2)^2.$$

第三, 证明 3) 成立. 由双期望公式计算可得

$$E(X_t) = E[E(X_t | X_{t-1})] = E\{[p\alpha_1 + (1 - p)\alpha_2]X_{t-1} + p\mu_1 + (1 - p)\mu_2\} = \\ [p\alpha_1 + (1 - p)\alpha_2]E(X_{t-1}) + p\mu_1 + (1 - p)\mu_2.$$

在平稳意义下有  $E(X_t) = E(X_{t-1})$ , 所以有

$$E(X_t) = \frac{p\mu_1 + (1 - p)\mu_2}{1 - p\alpha_1 - (1 - p)\alpha_2}.$$

第四, 证明 4) 成立. 由条件方差公式可得

$$\text{Var}(X_t) = \text{Var}[E(X_t | X_{t-1})] + E[\text{Var}(X_t | X_{t-1})],$$

其中,

$$\text{Var}[E(X_t | X_{t-1})] = \text{Var}\{[p\alpha_1 + (1 - p)\alpha_2]X_{t-1} + p\mu_1 + (1 - p)\mu_2\} = \\ [p\alpha_1 + (1 - p)\alpha_2]^2 \text{Var}(X_{t-1}),$$

$$E[\text{Var}(X_t | X_{t-1})] = p(1 - p)(\alpha_1 - \alpha_2)^2 E(X_{t-1}^2) + [p\alpha_1(1 + \alpha_1) + (1 - p)\alpha_2(1 + \alpha_2) + \\ 2p(1 - p)(\alpha_1 - \alpha_2)(\mu_1 - \mu_2)]E(X_{t-1}) +$$

$$\begin{aligned}
& p\mu_1 + (1 - p)\mu_2 + p(1 - p)(\mu_1 - \mu_2)^2 = \\
& p(1 - p)(\alpha_1 - \alpha_2)^2 \text{Var}(X_{t-1}) + p(1 - p)(\alpha_1 - \alpha_2)^2 [E(X_{t-1})]^2 + \\
& [p\alpha_1(1 + \alpha_1) + (1 - p)\alpha_2(1 + \alpha_2) + 2p(1 - p)(\alpha_1 - \alpha_2)(\mu_1 - \mu_2)]E(X_{t-1}) + \\
& p\mu_1 + (1 - p)\mu_2 + p(1 - p)(\mu_1 - \mu_2)^2,
\end{aligned}$$

所以

$$\begin{aligned}
\text{Var}(X_t) = & [p\alpha_1^2 + (1 - p)\alpha_2^2]\text{Var}(X_{t-1}) + p(1 - p)(\alpha_1 - \alpha_2)^2 [E(X_{t-1})]^2 + \\
& [p\alpha_1(1 + \alpha_1) + (1 - p)\alpha_2(1 + \alpha_2) + 2p(1 - p)(\alpha_1 - \alpha_2)(\mu_1 - \mu_2)]E(X_{t-1}) + \\
& p\mu_1 + (1 - p)\mu_2 + p(1 - p)(\mu_1 - \mu_2)^2.
\end{aligned}$$

在平稳意义下有  $E(X_t) = E(X_{t-1})$ ,  $\text{Var}(X_t) = \text{Var}(X_{t-1})$ , 所以有

$$\begin{aligned}
\text{Var}(X_t) = & \{p(1 - p)(\alpha_1 - \alpha_2)^2 [E(X_t)]^2 + [p\alpha_1(1 + \alpha_1) + (1 - p)\alpha_2(1 + \alpha_2) + \\
& 2p(1 - p)(\alpha_1 - \alpha_2)(\mu_1 - \mu_2)]E(X_t) + \\
& p\mu_1 + (1 - p)\mu_2 + p(1 - p)(\mu_1 - \mu_2)^2\} / [1 - p\alpha_1^2 - (1 - p)\alpha_2^2].
\end{aligned}$$

最后证明 5) 成立. 由全协方差公式可得

$$\begin{aligned}
\text{Cov}(X_t, X_{t+h}) = & E(X_t X_{t+h}) - E(X_t)E(X_{t+h}) = \\
& E[E(X_t X_{t+h} | X_t, X_{t+h-1})] - E(X_t)E[E(X_{t+h} | X_{t+h-1})] = \\
& E[X_t E(X_{t+h} | X_{t+h-1})] - E(X_t)E[E(X_{t+h} | X_{t+h-1})] = \\
& E\{[p\alpha_1 + (1 - p)\alpha_2]X_t X_{t+h-1} + [p\mu_1 + (1 - p)\mu_2]X_t\} - \\
& E(X_t)E\{[p\alpha_1 + (1 - p)\alpha_2]X_{t+h-1} + p\mu_1 + (1 - p)\mu_2\} = \\
& [p\alpha_1 + (1 - p)\alpha_2]E(X_t X_{t+h-1}) - [p\alpha_1 + (1 - p)\alpha_2]E(X_t)E(X_{t+h-1}) = \\
& [p\alpha_1 + (1 - p)\alpha_2]\text{Cov}(X_t, X_{t+h-1}) = \dots = [p\alpha_1 + (1 - p)\alpha_2]^h \text{Var}(X_t).
\end{aligned}$$

证毕.

### 2 参数估计

采用极大似然法进行参数估计. 设  $\{X_1, \dots, X_T\}$  为来自 MNBINAR(1) 过程的一组样本, 令  $\theta = (\alpha_1, \alpha_2, \mu_1, \mu_2, p)^T$  为待估参数, 则样本的对数似然函数为

$$\begin{aligned}
l(\theta) = & \log\left(\prod_{t=1}^T p_{x_t | x_{t-1}}\right) = \sum_{t=1}^T \log\left\{p \sum_{h=0}^{x_t} [P(\alpha_1 * X_{t-1} = h | X_{t-1} = x_{t-1})P(\epsilon_{1,t} = x_t - h)] + \right. \\
& \left. (1 - p) \sum_{h=0}^{x_t} [P(\alpha_2 * X_{t-1} = h | X_{t-1} = x_{t-1})P(\epsilon_{2,t} = x_t - h)]\right\}. \tag{6}
\end{aligned}$$

则参数  $\theta$  的条件极大似然估计量  $\hat{\theta}_{\text{CML}} := (\hat{\alpha}_1^{\text{CML}}, \hat{\alpha}_2^{\text{CML}}, \hat{\mu}_1^{\text{CML}}, \hat{\mu}_2^{\text{CML}}, \hat{p}^{\text{CML}})^T$  可通过极大化式(6)得到, 即 
$$\hat{\theta}_{\text{CML}} = \underset{\theta}{\text{argmax}} l(\theta). \tag{7}$$

由于似然函数较复杂, 难以找到式(7)的解析解, 因此需采用数值算法进行求解, 如 Newton-Raphson 迭代法等. 本文采用全局优化算法, 基于 R 软件进行求解.

**定理 1** 由式(7)给出的条件极大似然估计  $\hat{\theta}_{\text{CML}}$  是相合的, 并且有如下渐近分布:

$$\sqrt{T}(\hat{\theta}_{\text{CML}} - \theta_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}^{-1}(\theta_0)), \tag{8}$$

其中  $T$  是样本量,  $\theta_0$  为  $\theta$  的真值,  $\mathbf{I}(\theta)$  为 Fisher 信息矩阵.

证明: 由于 MNBINAR(1) 过程是一个 Markov 过程, 由命题 1 可知, 定理 1 是文献[20]中定理 2.1 和定理 2.2 的特殊情形, 因此, 在具体的模型设置中, 只需验证模型(2)的误差项  $\epsilon_{1,t}$  和  $\epsilon_{2,t}$  是否满足文献[20]中定理 2.1 和定理 2.2 的正则条件, 即可得到估计量的渐近正态性. 证明细节略.

### 3 数值模拟

下面通过数值模拟验证极大似然估计的效果. 考虑如下两种模型:

模型 I: 假设  $\epsilon_{i,t} \sim P(\mu_i)$ ,  $i = 1, 2$ ;

模型 II: 假设  $\epsilon_{i,t} \sim NB\left(v, \frac{\alpha_i}{1+\alpha_i}\right)$ ,  $i=1,2$ , 其概率质量函数为

$$P(\epsilon_{i,t} = z) = \frac{\Gamma(v+z)}{\Gamma(v)\Gamma(z+1)} \frac{\alpha_i^z}{(1+\alpha_i)^{v+z}}, \quad i=1,2, \quad z=0,1,2,\dots$$

容易验证,  $\mu_i = E(\epsilon_{i,t}) = \alpha_i v$ . 由于模型 II 的特殊结构, 因此只需要估计参数  $\alpha_1, \alpha_2, p$ . 针对两个模型分别选取如下一组参数进行模拟, 取样本量  $T=100, 300, 500$ ,  $v=3$ (已知).

模型 I:  $\alpha_1=0.9, \alpha_2=0.3, \mu_1=3.0, \mu_2=1.0, p=0.5$ ;

模型 II:  $\alpha_1=0.8, \alpha_2=0.2, p=0.7$ .

在样本量  $T=100$  时分别绘制出模型 I 和模型 II 的样本路径图和自相关函数(ACF)图, 结果如图 1 所示. 由图 1 可见, 两个序列均为平稳序列. 对于每组参数, 在 R 软件环境下进行 1 000 次重复实验, 分别计算其偏差(Bias)、标准偏差(standard deviation, SD)、标准误差(standard error, SE), 并在样本量  $T=500$  时分别绘制出模型 I 和模型 II 极大似然估计结果的 QQ(quantile-quantile)图, 以验证这两组估计结果的正态性. 模拟结果列于表 1, 两个模型的 QQ 图分别如图 2 和图 3 所示.

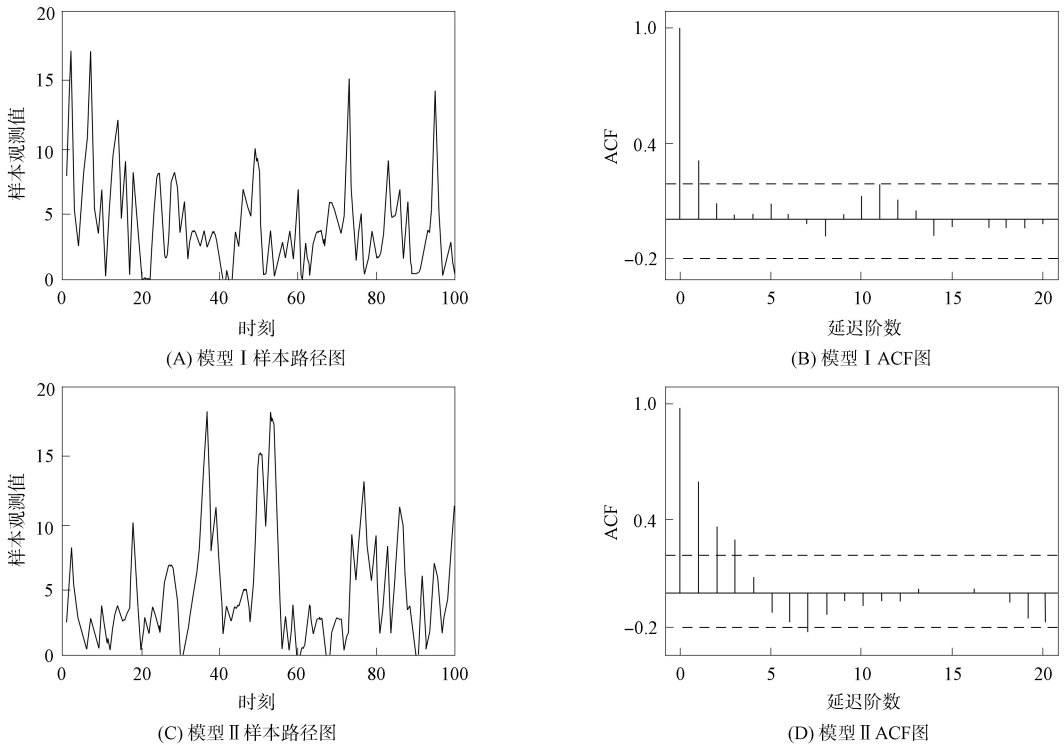


图 1 模型 I 和模型 II 的样本路径图和 ACF 图

Fig. 1 Sample path and ACF plots of model I and model II

表 1 模型 I 和模型 II 的估计结果

Table 1 Estimation results of model I and model II

模型	参数	T=100			T=300			T=500		
		Bias	SD	SE	Bias	SD	SE	Bias	SD	SE
I	$\alpha_1=0.9$	-0.029 6	0.168 1	0.178 4	-0.009 5	0.087 5	0.085 9	-0.004 1	0.063 9	0.064 6
	$\alpha_2=0.3$	-0.022 6	0.090 3	0.097 7	-0.005 1	0.053 5	0.051 1	-0.002 4	0.038 8	0.039 2
	$\mu_1=3.0$	0.187 5	0.882 4	0.940 5	0.087 6	0.484 9	0.476 4	0.025 6	0.350 8	0.356 8
	$\mu_2=1.0$	0.110 8	0.476 8	0.446 4	0.036 8	0.254 5	0.239 7	0.026 7	0.186 2	0.183 6
	$p=0.5$	-0.005 3	0.115 1	0.118 7	-0.004 0	0.070 5	0.067 5	-0.002 7	0.052 1	0.052 8
II	$\alpha_1=0.8$	0.007 7	0.086 7	0.085 4	0.003 9	0.041 3	0.041 9	0.001 7	0.032 2	0.032 0
	$\alpha_2=0.2$	0.003 7	0.070 0	0.068 5	0.001 4	0.037 3	0.036 0	0.001 1	0.027 7	0.027 6
	$p=0.7$	-0.019 5	0.110 8	0.109 6	-0.003 9	0.057 4	0.056 5	-0.003 8	0.043 3	0.043 6

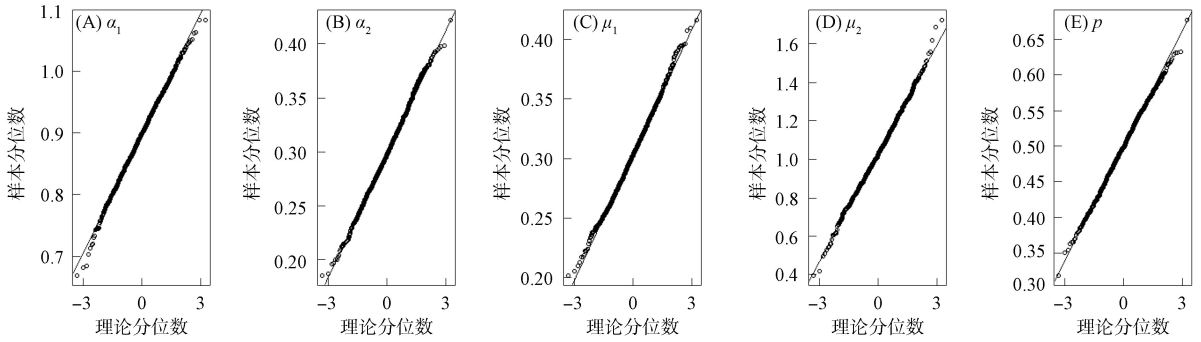


图 2 模型 I 极大似然估计结果的 QQ 图

Fig. 2 QQ plots of maximum likelihood estimation results for model I

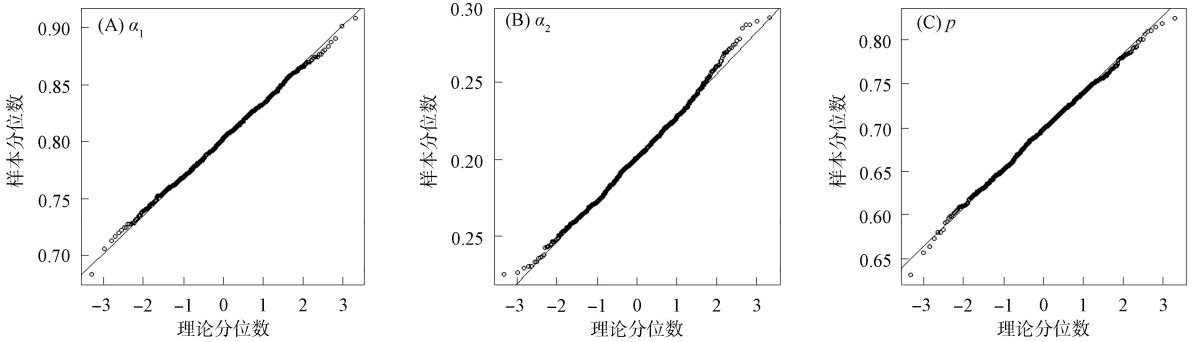


图 3 模型 II 极大似然估计结果的 QQ 图

Fig. 3 QQ plots of maximum likelihood estimation results for model II

由表 1 可见, 在不同参数组合下, 随着样本量增大, 所有的 Bias, SD 和 SE 都减小, 每个参数的 SD 和 SE 越来越接近. 由图 2 和图 3 可见, 几乎所有的点都与 45° 线重合. 从而验证了极大似然估计的相合性和渐近正态性.

### 4 实证分析

下面基于 MNBINAR(1) 模型拟合一组实际数据 (<http://www.forecastingprinciples.com>), 该数据集记录了美国匹兹堡第 63 号警车从 1990 年 1 月至 2001 年 12 月期间巡逻过程中每个月所报告的毒品犯罪次数, 包括 144 个观测值. 观测数据的样本均值和样本方差分别为 2.972 和 8.447, 呈现出典型的过渡离散特征. 数据的样本路径图和 ACF 图如图 4 所示.

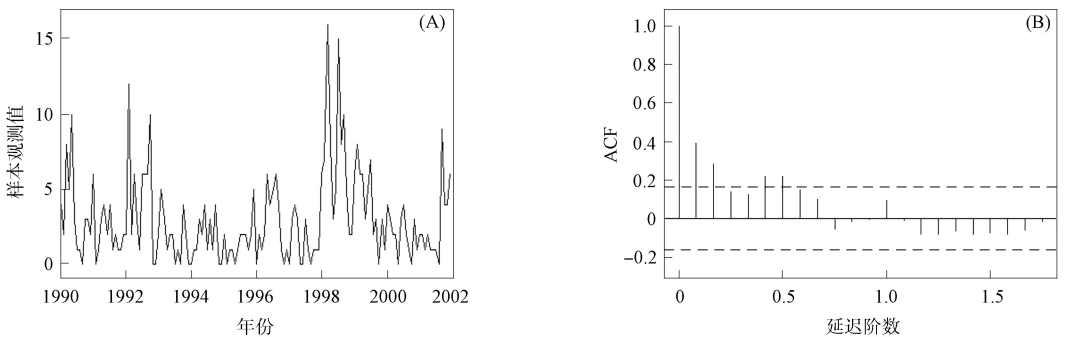


图 4 毒品犯罪次数数据的样本路径图(A)和 ACF 图(B)

Fig. 4 Sample path (A) and ACF plots (B) of drug offense count data

由图 4 可见, 所分析的数据是一个平稳的时间序列. 结合数据实际应用背景, 分别用 NGINAR(1) 模型<sup>[6]</sup> 和 MNBINAR(1) 模型对数据进行拟合. 在拟合 MNBINAR(1) 模型时,  $\epsilon_t$  分别取 Poisson 分布和  $v=1, 2, 3, 4, 5$  的负二项分布, 并将基于这两种分布的  $\epsilon_t$  所对应的模型分别记为 P-MNBINAR(1) 模

型和 NB-MNBINAR(1)模型. 利用极大似然估计法得到 NGINAR(1)模型和 MNBINAR(1)模型的参数估计值, 计算其相应的标准差, 并基于 AIC 准则<sup>[21]</sup>对拟合模型进行比较, 其中  $AIC=2k-2\ln(L)$ , 这里  $k$  为参数个数,  $L$  为模型的极大似然函数值. 拟合结果列于表 2.

表 2 不同模型的拟合结果比较

Table 2 Comparison of fitting results of different models

模型	估计结果	标准差	AIC 值
NGINAR(1)	$\hat{\mu}=2.76$	0.480	629.704
	$\hat{\alpha}=0.43$	0.071	
P-MNBINAR(1)	$\hat{\alpha}_1=0.33$	0.073	617.572
	$\hat{\alpha}_2=0.87$	0.253	
	$\hat{\mu}_1=1.13$	0.223	
	$\hat{\mu}_2=3.80$	0.800	
	$\hat{p}=0.80$	0.086	
NB-MNBINAR(1), $v=1$	$\hat{\alpha}_1=0.51$	0.111	634.136
	$\hat{\alpha}_2=1.29$	0.350	
	$\hat{p}=0.63$	0.227	
NB-MNBINAR(1), $v=2$	$\hat{\alpha}_1=0.44$	0.081	617.846
	$\hat{\alpha}_2=1.12$	0.343	
	$\hat{p}=0.74$	0.202	
NB-MNBINAR(1), $v=3$	$\hat{\alpha}_1=0.38$	0.059	613.067
	$\hat{\alpha}_2=0.99$	0.299	
	$\hat{p}=0.80$	0.162	
NB-MNBINAR(1), $v=4$	$\hat{\alpha}_1=0.33$	0.045	612.122
	$\hat{\alpha}_2=0.88$	0.244	
	$\hat{p}=0.82$	0.133	
NB-MNBINAR(1), $v=5$	$\hat{\alpha}_1=0.28$	0.036	612.708
	$\hat{\alpha}_2=0.79$	0.196	
	$\hat{p}=0.82$	0.116	

由表 2 可见, 在  $\{\epsilon_{i,t}\}$  依次服从  $v=1\sim 5$  的负二项分布所对应的模型中,  $v=1$  和  $v=2$  所对应的模型的  $\hat{\alpha}_2$  均大于 1, 说明  $v=1, 2$  是不合理的. 此外, 这两个模型的 AIC 值也较大. 因此, 首先排除这两个拟合结果. 在其余拟合模型中, NGINAR(1)模型的 AIC 值最大, 说明采用混合模型对此数据集进行拟合十分必要. 对比所有拟合模型,  $\{\epsilon_{i,t}\}$  服从  $v=4$  的负二项分布的 NB-MNBINAR(1)模型 AIC 值最小, 说明 NB-MNBINAR(1)模型最适合拟合该组数据. 可见, 本文提出的模型在实际应用中性能最佳, 是一个有竞争力的模型. 未来的研究可考虑类似文献[22-23]中的方法, 将数据分解后再建立模型, 以进一步增强模型的灵活性.

参 考 文 献

[1] AL-OSH M A, ALZAID A A. First-Order Integer-Valued Autoregressive (INAR(1)) Process [J]. Journal of Time Series Analysis, 1987, 8(3): 261-275.

[2] DU J G, LI Y. The Integer-Valued Autoregressive (INAR( $p$ )) Model [J]. Journal of Time Series Analysis, 1991, 12(2): 129-142.

[3] STEUTEL F W, VAN HARN K. Discrete Analogues of Self-decomposability and Stability [J]. The Annals of Probability, 1979, 7(5): 893-899.

[4] NEAL P, RAO T S. MCMC for Integer-Valued ARMA Processes [J]. Journal of Time Series Analysis, 2007, 28(1): 92-110.

[5] 张哲, 张海祥, 张卓飞, 等. INAR(1)模型参数的 Bayes 估计[J]. 吉林大学学报(理学版), 2010, 48(6): 931-935. (ZHANG Z, ZHANG H X, ZHANG Z F, et al. Bayesian Estimation of Parameters in the INAR(1) Model [J]. Journal of Jilin University (Science Edition), 2010, 48(6): 931-935.)

- [6] RISTIĆ M M, BAKOUCH H S, NASTIĆ A S. A New Geometric First-Order Integer-Valued Autoregressive (NGINAR(1)) Process [J]. *Journal of Statistical Planning and Inference*, 2009, 139(7): 2218-2226.
- [7] SCOTTO M G, WEIß C H, GOUVEIA S. Thinning-Based Models in the Analysis of Integer-Valued Time Series: A Review [J]. *Statistical Modelling*, 2015, 15(6): 590-618.
- [8] DAVIS R A, FOKIANOS K, HOLAN S H, et al. Count Time Series: A Methodological Review [J]. *Journal of the American Statistical Association*, 2021, 116: 1533-1547.
- [9] YANG K, KANG Y, WANG D H, et al. Modeling Overdispersed or Underdispersed Count Data with Generalized Poisson Integer-Valued Autoregressive Processes [J]. *Metrika*, 2019, 82(7): 863-889.
- [10] 张洁, 张玉, 董小刚. 自激励广义二项门限自回归模型的统计推断 [J]. *吉林大学学报(理学版)*, 2023, 61(2): 275-284. (ZHANG J, ZHANG Y, DONG X G. Statistical Inference for Self-exciting Generalized Binomial Threshold Autoregressive Model [J]. *Journal of Jilin University (Science Edition)*, 2023, 61(2): 275-284.)
- [11] 邵志艳, 王佳聪, 杨凯, 等. 带有相依稀疏算子的一阶随机系数二项自回归模型 [J]. *吉林大学学报(理学版)*, 2023, 61(5): 1083-1089. (TAI Z Y, WANG J C, YANG K, et al. First-Order Random Coefficient Binomial Autoregressive Model with Dependent Thinning Operator [J]. *Journal of Jilin University (Science Edition)*, 2023, 61(5): 1083-1089.)
- [12] 刘子健, 桂尚珂, 陈硕, 等. 一阶混合整数负二项自回归模型 [J]. *吉林大学学报(理学版)*, 2021, 59(6): 1395-1399. (LIU Z J, GUI S K, CHEN S, et al. First-Order Mixed Integer-Valued Binomial Autoregressive Model [J]. *Journal of Jilin University (Science Edition)*, 2021, 59(6): 1395-1399.)
- [13] RISTIĆ M M, NASTIĆ A S. A Mixed INAR( $p$ ) Model [J]. *Journal of Time Series Analysis*, 2012, 33(6): 903-915.
- [14] LI C, WANG D H, ZHANG H X. First-Order Mixed Integer-Valued Autoregressive Processes with Zero-Inflated Generalized Power Series Innovations [J]. *Journal of the Korean Statistical Society*, 2015, 44(2): 232-246.
- [15] ZHANG J, SHAO S Y, YANG K, et al. A Statistical Study for Some Classes of First-Order Mixed Generalized Binomial Autoregressive Models [J/OL]. *Communications in Statistics: Theory and Methods*, (2023-04-27) [2023-06-26]. <https://doi.org/10.1080/03610926.2023.2205046>.
- [16] MAO H Y, ZHU F K, CUI Y. A Generalized Mixture Integer-Valued GARCH Model [J]. *Statistical Methods and Applications*, 2020, 29: 527-552.
- [17] ZHU F K, LI Q, WANG D H. A Mixture Integer-Valued ARCH Model [J]. *Journal of Statistical Planning and Inference*, 2010, 140(7): 2025-2036.
- [18] ROSS S M. *Stochastic Processes* [M]. 2nd ed. New York: Wiley, 1996: 170-177.
- [19] ROSENBLATT M. *Markov Processes, Structure and Asymptotic Behaviour* [M]. Berlin: Springer, 1971: 7-10.
- [20] BILLINGSLEY P. *Statistical Inference for Markov Processes* [M]. Chicago: University of Chicago Press, 1961: 10-14.
- [21] AKAIKE H. A New Look at the Statistical Model Identification [J]. *IEEE Transactions on Automatic Control*, 1974, 19(6): 716-723.
- [22] 秦喜文, 邢婷婷, 董小刚, 等. 基于 EEMD 的公路客流量波动性与周期性研究 [J]. *东北师大学报(自然科学版)*, 2016, 48(1): 44-48. (QIN X W, XING T T, DONG X G, et al. The Volatility and Periodicity of Highway Passenger Volume Based on Ensemble Empirical Mode Decomposition [J]. *Journal of Northeast Normal University (Natural Science Edition)*, 2016, 48(1): 44-48.)
- [23] 秦喜文, 张瑜, 董小刚, 等. 基于 EEMD 和 SVR 的人民币汇率预测 [J]. *东北师大学报(自然科学版)*, 2016, 49(2): 47-51. (QIN X W, ZHANG Y, DONG X G, et al. Forecasting RMB Exchange Rate Based on EEMD and SVR [J]. *Journal of Northeast Normal University (Natural Science Edition)*, 2016, 49(2): 47-51.)

(责任编辑: 赵立芹)