doi: 10.13413/j.cnki.jdxblxb.2024012

k 阶 Poisson 相依驱动随机系数混合算子整数值自回归模型

刘秀芳[1], 张晓磊[2], 王德辉[3]

(1. 太原理工大学 数学学院, 太原 030024; 2. 吉林大学 数学学院, 长春 130012;
3. 辽宁大学 数学与统计学院, 沈阳 110036)

摘要: 采用 k 阶 Poisson 相依驱动随机系数混合算子整数值自回归模型分析具有变量字符元素计数特征的数据, 给出该模型的统计性质和参数的条件极大似然估计, 并证明估计量的渐近正态性. 数值模拟结果表明, 随着样本容量的增加, 参数估计逐渐收敛于真实值. 实际数据分析结果表明了该模型的有效性.

关键词: k 阶自回归模型; Po-DDRCMTINAR(k)模型; 混合稀疏算子; 条件极大似然估计

中图分类号: O212.1 **文献标志码:** A **文章编号:** 1671-5489(2024)04-0866-12

k-Order Poisson Dependent-Driven Random Coefficient Mixed Thinning Integer-Valued Autoregressive Model

LIU Xiufang[1], ZHANG Xiaolei[2], WANG Dehui[3]

(1. *College of Mathematics, Taiyuan University of Technology, Taiyuan* 030024, *China*;
2. *College of Mathematics, Jilin University, Changchun* 130012, *China*;
3. *School of Mathematics and Statistics, Liaoning University, Shenyang* 110036, *China*)

Abstract: By using k-order Poisson dependent-driven random coefficient mixed thinning integer-valued autoregressive model, we analyzed data with the counting of elements of variable character, gave statistical properties of the model and conditional maximum likelihood estimation of parameters, and proved the asymptotic normality of the estimators. The numerical simulation results show that as the sample size increases, the parameter estimation gradually converges to the true value. The actual data analysis results show that the effectiveness of the proposed model.

Keywords: k-order autoregressive model; Po-DDRCMTINAR(k) model; mixed thinning operator; conditional maximum likelihood estimation

0 引 言

由于时间序列数据来源的广泛性和数据类型的多样化, 因此需要更广的随机系数时间序列模型分析数据. 目前, 具有整数特征的时间序列数据在现实生活中应用广泛. 文献[1]基于二项算子提出了一阶整数值自回归(INAR(1))模型, 用于分析保险公司每天收到的索赔数量; 文献[2]讨论了纵向

收稿日期: 2024-01-08.

第一作者简介: 刘秀芳(1990—), 女, 汉族, 博士, 讲师, 从事时间序列分析的研究, E-mail: liuxiufang@tyut.edu.cn. **通信作者简介:** 张晓磊(1985—), 男, 汉族, 博士, 从事时间序列分析的研究, E-mail: zhangxl_jlu@163.com.

基金项目: 国家自然科学基金(批准号: 12271231)和山西省基础研究项目(批准号: 202103021223084).

计数数据的 Poisson INAR(1)模型(PINAR(1))，其中新息序列$\{\varepsilon_t\}$服从 Poisson 分布；文献[3]提出了一种基于 Pegram 和二项算子的平稳非负一阶整数值随机变量模型(PegramINAR(1))；文献[4]进一步将 INAR(1)模型推广到高阶，研究了 INAR(p)模型的存在性和平稳遍历性；文献[5]得到了满足INAR(p)模型的时间序列$\{X_t\}$的高阶矩和累积量的差分方程，并以状态空间形式获得了该模型的谱和双谱密度函数，从而在频域中对其进行表征. 特别地，文献[1]中参数φ可能是随时间变化的随机变量，这是因为φ可能受各种环境因素的影响，例如，患者的生存概率受卫生保健质量和患者自身健康状况等因素的影响. 因此，研究随机系数整数值自回归模型有一定的应用价值. 文献[6]提出的一阶随机系数整数值自回归模型 RCINAR(1)分析了到医院就医的门诊病人数；文献[7]研究了观测驱动随机系数整数值自回归模型 OD-RCINAR(1)；文献[8]提出了一种新的具有随机系数和零膨胀几何边际分布的平稳一阶整数值自回归模型 ZIGINARRC(1)，用于分析过度离散、过多零点和自相关特征的计数数据；文献[9]将 RCINAR(1)模型扩展为高阶随机系数整数值自回归模型(RCINAR(p))，并用极大似然法、条件最小二乘法、修正拟似然法和广义矩法估计模型参数；文献[10]提出了 p 阶 Poisson 相依驱动随机系数整数值自回归模型(Po-DDRCINAR(p))，采用最小二乘法和拟似然法估计参数，并通过数值模拟及分析癫痫病数据验证了模型的有效性；文献[11]考虑了一个具有负二项边际的一阶平稳整数值自回归模型(NBINAR(1))，用于分析盗窃计数数据；文献[12]基于负二项算子提出了一阶几何整数值自回归模型(NGINAR(1))，用于分析细菌分解数量的数据；文献[13]提出了一种新的季节性几何整数值自回归模型 SGINAR(1)，并讨论了该模型的一些基本概率统计特性；文献[14]将上述模型扩展为高阶情形，提出了一个新的组合整数值自回归模型 CGINAR(p)，该模型具有几何边际分布，并分析了美国每月的脊髓灰质症发病率数据及年太阳黑子相对数数据；文献[15]提出了 p 阶混合整数值自回归模型(SDMINAR(p))，其随机误差项是序列相依的，该模型可用于分析过度离散的数据；文献[16]提出了一种 p 阶二态自激励门限整数值自回归(SETINAR(2,p))模型，给出了参数估计和假设检验方法，并将该模型用于分析德国脑膜炎球菌病患者人数；文献[17]研究了一种由独立的负二项分布随机变量驱动的一阶门限整数值自回归模型 NBTINAR(1)；文献[18]在文献[17]的基础上研究了随机项分布未知的情形，考虑了参数的点估计和区间估计问题，并提供了一种检验数据非线性的方法；文献[19]针对具有变量字符元素计数特征的数据，提出了一阶混合算子整数值自回归模型，其中变量字符元素计数特征即为元素可能在一段时间内处于被动状态，并以这种方式对总数只添加 0 或 1，而在下一个计数间隔内，由于某些原因，元素可能处于主动状态，对计数值做出的贡献更大且大于 1；文献[20]将上述模型推广到一阶随机系数混合算子整数值自回归模型(RCMTINAR(1))，用于分析生存概率受各种环境因素影响的数据；为分析具有可变字符元素计数特征的多变量数据，文献[21]提出了高阶相依驱动随机系数混合算子整数值自回归模型. 本文基于文献[21]，考虑随机误差项服从 Poisson 分布情形的 k 阶相依驱动随机系数混合算子整数值自回归模型：

$$X_t=\sum_{i=1}^{k}\alpha_{ti}\bullet_{p_i}X_{t-i}+\varepsilon_t,\qquad t\geqslant 1, \tag{1}$$

其中：$\alpha_{t1},\alpha_{t2},\cdots,\alpha_{tk}$的概率分布为

$$\begin{cases} P(\alpha_{t1}=\alpha_1,\ \alpha_{t2}=0,\ \cdots,\ \alpha_{tk}=0)=\varphi_1,\\ P(\alpha_{t1}=0,\ \alpha_{t2}=\alpha_2,\ \cdots,\ \alpha_{tk}=0)=\varphi_2,\\ \qquad\vdots\\ P(\alpha_{t1}=0,\ \alpha_{t2}=0,\ \cdots,\ \alpha_{tk}=\alpha_k)=\varphi_k,\\ P(\alpha_{t1}=0,\ \alpha_{t2}=0,\ \cdots,\ \alpha_{tk}=0)=\varphi_0, \end{cases} \tag{2}$$

$\varphi_0,\varphi_1,\cdots,\varphi_k$ 非负，且 $\sum\limits_{i=0}^{k}\varphi_i=1$；在给定 α_{ti} 条件下，$\alpha_{ti}\bullet_{p_i}=\sum\limits_{j=1}^{X_{t-i}}W_j^{(ti)}$，这里非负整数值随机变量 X_{t-i} 与计数序列 $W_j^{(ti)}$ 相互独立，且$\{W_j^{(ti)}\}$是一个独立同分布的随机变量序列，其分布为

$$W_j^{(ti)}=\begin{cases} B_j^{(ti)},\ \text{w. p. } p_i,\\ G_j^{(ti)},\ \text{w. p. } 1-p_i, \end{cases}\qquad p_i\in[0,1],$$

$B_j^{(ti)}$ 服从 Bernoulli(α_{ti})分布，$G_j^{(ti)}$ 服从 Geometric($\alpha_{ti}/(1+\alpha_{ti})$)分布，其概率质量函数为 $P(G_j^{(ti)}=x)=\frac{\alpha_{ti}^x}{(1+\alpha_{ti})^{x+1}}$，$x=0,1,\cdots$，$B_j^{(ti)}$ 和 $G_j^{(ti)}$ 对任意时间 t 相互独立；时变系数$\{\alpha_{ti},1\leqslant i\leqslant k\}$对任意时间 t 是相互独立同分布的，$\{\varepsilon_t\}$是独立同分布的非负整数值序列，并服从参数为 λ 的 Poisson 分布，且$\{\alpha_{ti},1\leqslant i\leqslant k\}$与$\{\varepsilon_t\}$相互独立. 该模型称为 Po-DDRCMTINAR(k)模型，其参数是 $\boldsymbol{\theta}=(\alpha_1,\cdots,\alpha_k,\varphi_1,\cdots,\varphi_k,p_1,\cdots,p_k,\lambda)'$. 根据方程(2)，可得 α_{it}的均值、方差以及 α_{it}与$\alpha_{jt}(i\neq j)$之间的协方差分别为

$$E(\alpha_{ti})=\alpha_i\varphi_i,\quad \mathrm{Var}(\alpha_{ti})=\alpha_i^2\varphi_i(1-\varphi_i),\quad \mathrm{Cov}(\alpha_{ti},\alpha_{tj})=-\alpha_i\alpha_j\varphi_i\varphi_j. \tag{3}$$

特别地，当参数 $p_1=p_2=\cdots p_k=1$ 时，本文模型退化为文献[10]提出的 Po-DDRCINAR(k)模型. 本文对参数 $p_1,p_2,\cdots,p_k$ 是否等于 1，即该模型是否退化为 Po-DDRCINAR(k)模型进行检验. 最后，通过数值模拟和实例分析验证模型的有效性.

1 模型的统计性质

假设$\{X_t\}$是由模型(1)-(2)定义的 Po-DDRCMTINAR(k)模型，下面给出该模型的一些统计性质.

命题 1 1) $E(X_t\mid X_{t-1},\cdots,X_{t-k})=\sum_{i=1}^{k}\alpha_i\varphi_iX_{t-i}+\lambda$；

2) 如果该模型是 k 阶平稳的，则 $E(X_t)=\lambda\left(1-\sum_{i=1}^{k}\alpha_i\varphi_i\right)^{-1}$；

3) $\mathrm{Var}(X_t\mid X_{t-1},\cdots,X_{t-k})=\sum_{i=1}^{k}[\alpha_i\varphi_i(1+(1-2p_i)\alpha_i)X_{t-i}+\alpha_i^2\varphi_iX_{t-i}^2]+2\lambda\sum_{i=1}^{k}\alpha_i\varphi_iX_{t-i}+\lambda^2+\lambda-\left(\sum_{i=1}^{k}\alpha_i\varphi_iX_{t-i}+\lambda\right)^2$；

4) 令 $\gamma_j=\mathrm{Cov}(X_t,X_{t-j})$，则 $\gamma_j=\sum_{i=1}^{k}\alpha_i\varphi_i\gamma_{j-i}$.

证明：令 $S_i=\alpha_{ti}\cdot{}_{p_i}X_{t-i}$，则 S_i 的条件矩母函数为

$$M_{S_i\mid\alpha_{ti},X_{t-i}}=\left[p_i(1-\alpha_{ti}+\alpha_{ti}\mathrm{e}^t)+(1-p_i)\frac{1}{1+\alpha_{ti}-\alpha_{ti}\mathrm{e}^t}\right]^{X_{t-i}},$$

因此，可得 S_i 的 l 阶条件矩为$E(S_i^l\mid\alpha_{ti},X_{t-i})=M_{S_i\mid\alpha_{ti},X_{t-i}}^{(l)}(0)$. 特别地，有

$$E(S_i\mid\alpha_{ti},X_{t-i})=\alpha_{ti}X_{t-i},\qquad E(S_i^2\mid\alpha_{ti},X_{t-i})=\alpha_{ti}^2X_{t-i}(1-X_{t-i})+(\alpha_{ti}+2(1-p_i)\alpha_{ti}^2)X_{t-i},$$

进一步，可得 $E(S_i\mid X_{t-i})=\alpha_i\varphi_iX_{t-i}$，$E(S_i^2\mid X_{t-i})=\alpha_i\varphi_i(1+(1-2p_i)\alpha_i)X_{t-i}+\alpha_i^2\varphi_iX_{t-i}^2$，从而可推出统计性质 1)～4)成立. 证毕.

根据文献[21]中定理 1，可知在 $\sum_{i=1}^{k}\alpha_i\varphi_i<1$，$0<\alpha_i<1$，$0<\varphi_i<1$，$0<p_i<1$，$0<\varphi_1+\varphi_2+\cdots+\varphi_k<1$及 $E(A_t\otimes A_t)$的最大特征值小于 1 成立的条件下，模型$\{X_t\}$是平稳遍历的. 进一步，根据文献[22]，通过以下结果给出模型$\{X_t\}$的严平稳遍历性.

定理 1 满足方程(1)-(2)的模型$\{X_t\}$是严平稳遍历的.

证明：首先证明模型$\{X_t\}$是严平稳的. 只需证明对 $t,k'>0$ 和任意的 $x_1,x_2,\cdots,x_{k'}$，

$$P(X_i=x_i,\ 1\leqslant i\leqslant k')=P(X_{t+i}=x_i,\ 1\leqslant i\leqslant k').$$

基于模型的定义，X_t 是一个 k 阶 Markov 过程. 不失一般性，假设$k\leqslant k'$，则有

$$P(X_i=x_i,\ 1\leqslant i\leqslant k')=P(X_1=x_1)\prod_{l=2}^{k'}P(X_l=x_l\mid X_{l-j}=x_{l-j},\ 1\leqslant j\leqslant\min\{k,l-1\}), \tag{4}$$

$$P(X_{t+i}=x_i,\ 1\leqslant i\leqslant k')=P(X_{t+1}=x_1)\prod_{l=2}^{k'}P(X_{t+l}=x_l\mid X_{t+l-j}=x_{l-j},\ 1\leqslant j\leqslant\min\{k,l-1\}). \tag{5}$$

再根据模型的定义及 ε_t 是一个独立同分布的 Poisson 随机变量序列，可得方程(4)，(5)的等号右边相等，从而证得该过程是严平稳的.

下面证明该过程的遍历性. 令 A 为序列 X_t 的任意不可变事件，根据文献[23]中定义 2 知，只需证明 $P(A)=0$ 或 $P(A)=1$ 即可. 由于 A 是任意不可变事件，因此存在 $B\in B(\mathbb{R}^\infty)$，使得对任意的

$t\in\mathbb{Z}$，满足 $A=\{w\,|\,(X_t,X_{t-1},\cdots)\in B\}$. 因为 $X_t\in N$，所以 $A=\{w\,|\,(X_t,X_{t-1},\cdots)\in B\bigcap N\}$. 如果 $F(X_t,X_{t-1},\cdots)$是一个由$\{X_t,X_{t-1},\cdots\}$生成的 σ-域，则对任意的 $t\in\mathbb{Z}$，有

$$A\in F(X_t,X_{t-1},\cdots),\tag{6}$$

根据模型(1)-(2)，有

$$F(X_t,X_{t-1},\cdots)\subseteq F(\varepsilon_t,V^{(t)},\varepsilon_{t-1},V^{(t-1)},\cdots),\tag{7}$$

其中 $V^{(t)}$是 X_t 构造中使用计数序列的符号，即混合算子运算符序列. 根据 X_t 的定义知，$\{\varepsilon_t,V^{(t)}\}$是一个独立的随机变量序列，结合方程(6),(7)，对任意的 $t\in\mathbb{Z}$，有 $A\in F(\varepsilon_t,V^{(t)},\varepsilon_{t-1},V^{(t-1)},\cdots)$或者等价地有 $A\in\bigcap\limits_{t=0}^{-\infty}F(\varepsilon_t,V^{(t)},\varepsilon_{t-1},V^{(t-1)},\cdots)$. 因此，得到集合的无穷交集是一个尾 σ-代数满足 Kolmogorov 0-1 律定理，所以 A 是一个尾事件，且有 $P(A)=0$ 或 $P(A)=1$. 从而遍历性得证. 证毕.

上述统计性质有利于获得模型中参数估计量的渐近性质. 为得到参数的条件极大似然估计，下面给出模型的转移概率：

$$P_\theta(X_t\,|\,X_{t-1},\cdots,X_{t-k})=\sum_{j=1}^{k}\sum_{m_j=0}^{M_t}\varphi_j f_\varepsilon(X_t-m_j)P(\alpha_j\bullet_{p_j}X_{t-j}=m_j\,|\,X_{t-1},\cdots,X_{t-k})+\varphi_0 f_\varepsilon(X_t),$$

其中 $f_\varepsilon(X_t-m_j)=\dfrac{\lambda^{X_t-m_j}\mathrm{e}^{-\lambda}}{(X_t-m_j)!}$，$m_0=0$，$j=0,1,\cdots,k$. 并且

$$\begin{aligned}P(\alpha_j\bullet_{p_j}X_{t-j}=m_j\,|\,X_{t-1},\cdots,X_{t-k})=&(1-p_j)^{X_{t-j}}P\left(\mathrm{NB}\left(X_{t-j},\frac{\alpha_j}{1+\alpha_j}\right)=m_j\right)+\\&p_j^{X_{t-j}}P(\mathrm{Bin}(X_{t-j},\alpha_j)=m_j)+\sum_{i=1}^{X_{t-j}-1}\binom{X_{t-j}}{i}p_j^i(1-p_j)^{X_{t-j}-i}\times\\&P\left(\mathrm{Bin}(i,\alpha_j)+\mathrm{NB}\left(X_{t-j}-i,\frac{\alpha_j}{1+\alpha_j}\right)=m_j\right),\end{aligned}$$

其中 $(1-p_j)^{X_{t-j}}P\left(\mathrm{NB}\left(X_{t-j},\dfrac{\alpha_j}{1+\alpha_j}\right)=m_j\right)$对应的 M_t 是 X_t，$p_j^{X_{t-j}}P(\mathrm{Bin}(X_{t-j},\alpha_j)=m_j)$和 $P\left(\mathrm{Bin}(i,\alpha_j)+\mathrm{NB}\left(X_{t-j}-i,\dfrac{\alpha_j}{1+\alpha_j}\right)=m_j\right)$对应的 M_t 是 $\min\{X_t,X_{t-j}\}$. 用$\binom{a}{b}$表示二项组合，则

$$P\left(\mathrm{Bin}(i,\alpha_j)+\mathrm{NB}\left(X_{t-j}-i,\frac{\alpha_j}{1+\alpha_j}\right)=m_j\right)=\sum_{l=0}^{\min\{m_j,i\}}\binom{i}{l}\binom{X_{t-j}-i+m_j-l-1}{X_{t-j}-i-1}\frac{\alpha_j^{m_j}(1-\alpha_j)^{i-l}}{(1+\alpha_j)^{m_j-l+X_{t-j}-i}}.$$

为方便计算，下面给出一些表达式：

$$Y(d_j,e_j)=\binom{d_j+e_j-1}{d_j-1}\left(\frac{\alpha_j}{1+\alpha_j}\right)^{e_j}\left(\frac{1}{1+\alpha_j}\right)^{d_j},$$

$$Z(d_j,e_j)=\binom{d_j}{e_j}\alpha_j^{e_j}(1-\alpha_j)^{d_j-e_j},\qquad j=1,2,\cdots,k,$$

$$r(d_j,e_j)=\binom{d_j}{e_j}p_j^{e_j}(1-p_j)^{d_j-e_j},\quad q(d_i,e_i)=\frac{\lambda^{d_i-e_i}\mathrm{e}^{-\lambda}}{(d_i-e_i)!},\quad i=0,1,\cdots,k,\quad e_0=0,$$

从而模型的转移概率也可表示为

$$\begin{aligned}P_\theta(X_t\,|\,X_{t-1},\cdots,X_{t-k})=&\sum_{j=1}^{k}\varphi_j\Big\{\sum_{m_j=0}^{X_t}q(X_t,m_j)(1-p_j)^{X_{t-j}}Y(X_{t-j},m_j)+\sum_{m_j=0}^{\min\{X_t,X_{t-j}\}}q(X_t,m_j)\times\\&\Big[p_j^{X_{t-j}}Z(X_{t-j},m_j)+\sum_{i=1}^{X_{t-j}-1}r(X_{t-j},i)\sum_{l=0}^{\min\{m_j,i\}}Z(i,l)Y(X_{t-j}-i,m_j-l)\Big]\Big\}+\\&\Big(1-\sum_{j=1}^{k}\varphi_j\Big)q(X_t,m_0).\end{aligned}\tag{8}$$

2　参数的条件极大似然估计

下面基于模型的统计性质和转移概率讨论模型中未知参数的条件极大似然估计.

假设 $X_0,X_1,\cdots,X_n$ 是由 Po-DDRCMTINAR(k)过程生成的随机变量，其真实参数值为 $\boldsymbol{\theta}_0$. 对数似然函数表示为 $l(\boldsymbol{\theta})=\sum\limits_{t=k+1}^{n}\log\ P_{\boldsymbol{\theta}}(X_t\mid X_{t-1},\cdots,X_{t-k})$，其中 $P_{\boldsymbol{\theta}}(X_t\mid X_{t-1},\cdots,X_{t-k})$ 由方程(6)给出，简记为 $P_{\boldsymbol{\theta}}$. 本文模型的得分函数为

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial\alpha_j}=&\sum_{t=k+1}^{n}\frac{1}{P_{\boldsymbol{\theta}}}\Big\{\frac{\varphi_jX_{t-j}}{1+\alpha_j}\Big[\sum_{m_j=0}^{X_t-1}(1-p_j)^{X_{t-j}}Y(X_{t-j}+1,m_j)q(X_t-1,m_j)-\\&\sum_{m_j=0}^{X_t}(1-p_j)^{X_{t-j}}Y(X_{t-j},m_j)q(X_t,m_j)\Big]+\frac{\varphi_jX_{t-j}}{1-\alpha_j}\times\\&\Big[\sum_{m_j=0}^{\min\{X_t-1,X_{t-j}-1\}}p_j^{X_{t-j}}Z(X_{t-j}-1,m_j)q(X_t-1,m_j)-\sum_{m_j=0}^{\min\{X_t,X_{t-j}\}}p_j^{X_{t-j}}Z(X_{t-j},m_j)q(X_t,m_j)\Big]+\\&\sum_{m_j=0}^{\min\{X_t,X_{t-j}\}}\varphi_jq(X_t,m_j)\sum_{i=1}^{X_{t-j}-1}r(X_{t-j},i)\times\\&\Big[\frac{i}{1-\alpha_j}\Big[\sum_{l=0}^{\min\{m_j-1,i-1\}}Z(i-1,l)-\sum_{l=0}^{\min\{m_j,i\}}Z(i,l)\Big]Y(X_{t-j}-i,m_j-l)+\\&\frac{X_{t-j}-i}{1+\alpha_j}\Big[\sum_{l=0}^{\min\{m_j-1,i\}}Y(X_{t-j}-i+1,m_j-l-1)-\sum_{l=0}^{\min\{m_j,i\}}Y(X_{t-j}-i,m_j-l)\Big]Z(i,l)\Big]\Big\},\end{aligned}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial\varphi_j}=&\sum_{t=k+1}^{n}\frac{1}{P_{\boldsymbol{\theta}}}\Big\{\sum_{m_j=0}^{X_t}(1-p_j)^{X_{t-j}}Y(X_{t-j},m_j)q(X_t,m_j)+\sum_{m_j=0}^{\min\{X_t,X_{t-j}\}}q(X_t,m_j)\times\\&\Big[p_j^{X_{t-j}}Z(X_{t-j},m_j)+\sum_{i=1}^{X_{t-j}-1}r(X_{t-j},i)\sum_{l=0}^{\min\{m_j,i\}}Z(i,l)Y(X_{t-j}-i,m_j-l)\Big]-q(X_t,m_0)\Big\},\end{aligned}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial p_j}=&\sum_{t=k+1}^{n}\frac{1}{P_{\boldsymbol{\theta}}}\Big\{-\varphi_jX_{t-j}\sum_{m_j=0}^{X_t}(1-p_j)^{X_{t-j}-1}Y(X_{t-j},m_j)q(X_t,m_j)+\varphi_jX_{t-j}\sum_{m_j=0}^{\min\{X_t,X_{t-j}\}}q(X_t,m_j)\times\\&\Big[p_j^{X_{t-j}-1}Z(X_{t-j},m_j)+\frac{1}{1-p_j}\Big(\sum_{i=0}^{X_{t-j}-2}r(X_{t-j}-1,i)-\sum_{i=1}^{X_{t-j}-1}r(X_{t-j},i)\Big)\sum_{l=0}^{\min\{m_j,i\}}Z(i,l)Y(X_{t-j}-i,m_j-l)\Big]\Big\},\end{aligned}$$

$$\begin{aligned}\frac{\partial l(\boldsymbol{\theta})}{\partial\lambda}=&\sum_{t=k+1}^{n}\frac{1}{P_{\boldsymbol{\theta}}}\Big\{\sum_{j=1}^{k}\Big[\sum_{m_j=0}^{X_t-1}\varphi_jq(X_t-1,m_j)(1-p_j)^{X_{t-j}}Y(X_{t-j},m_j)+\sum_{m_j=0}^{\min\{X_t-1,X_{t-j}\}}\varphi_jq(X_t-1,m_j)\times\\&\Big[p_j^{X_{t-j}}Z(X_{t-j},m_j)+\sum_{i=1}^{X_{t-j}-1}r(X_{t-j},i)\sum_{l=0}^{\min\{m_j,i\}}Z(i,l)Y(X_{t-j}-i,m_j-l)\Big]\Big]+\\&\Big(1-\sum_{j=1}^{k}\varphi_j\Big)q(X_t-1,m_0)-P_{\boldsymbol{\theta}}\Big\}.\end{aligned}$$

因此，通过最大化条件对数似然函数或通过求解与得分函数相关的方程，可推出未知参数 $\boldsymbol{\theta}$ 的条件极大似然估计量 $\hat{\boldsymbol{\theta}}$.

下面给出 Y,Z,r,q,P 及其偏导数的相关递推关系，用于证明参数估计量的渐近正态性. 根据转移概率和得分函数，可得 Y,Z,r,q,P 的表达式如下：

$$Y(d_j,e_j)=\frac{\alpha_jd_j}{e_j}Y(d_j+1,e_j-1),\tag{9}$$

$$Z(d_j,e_j)=\alpha_jZ(d_j-1,e_j-1)+(1-\alpha_j)Z(d_j-1,e_j),\tag{10}$$

$$r(d_j,e_j)=p_jr(d_j-1,e_j-1)+(1-p_j)r(d_j-1,e_j),\tag{11}$$

$$\frac{\partial Y(d_j,e_j)}{\partial\alpha_j}=\frac{d_j}{1+\alpha_j}[Y(d_j+1,e_j-1)-Y(d_j,e_j)],\tag{12}$$

$$\frac{\partial Z(d_j,e_j)}{\partial\alpha_j}=\frac{d_j}{1-\alpha_j}[Z(d_j-1,e_j-1)-Z(d_j,e_j)],\tag{13}$$

$$\frac{\partial r(d_j,e_j)}{\partial p_j}=\frac{d_j}{1-p_j}[r(d_j-1,e_j-1)-r(d_j,e_j)], \tag{14}$$

$$\frac{\partial q(d_j,e_j)}{\partial \lambda}=q(d_j-1,e_j)-q(d_j,e_j), \tag{15}$$

$$\frac{\partial^2 q(d_j,e_j)}{\partial \lambda^2}=q(d_j-2,e_j)-2q(d_j-1,e_j)+q(d_j,e_j), \tag{16}$$

$$\begin{aligned}\frac{\partial P_{\boldsymbol{\theta}}}{\partial \alpha_j}=&\sum_{m_j=0}^{X_t}\varphi_j q(X_t,m_j)(1-p_j)^{X_{t-j}}\frac{\partial Y(X_{t-j},m_j)}{\partial \alpha_j}+\\&\sum_{m_j=0}^{\min\{X_t,X_{t-j}\}}\varphi_j q(X_t,m_j)\Bigg[p_j^{X_{t-j}}\frac{\partial Z(X_{t-j},m_j)}{\partial \alpha_j}+\sum_{i=1}^{X_{t-j}-1}r(X_{t-j},i)\times\\&\sum_{l=0}^{\min\{m_j,i\}}\left(\frac{\partial Z(i,l)}{\partial \alpha_j}Y(X_{t-j}-i,m_j-l)+Z(i,l)\frac{\partial Y(X_{t-j}-i,m_j-l)}{\partial \alpha_j}\right)\Bigg].\end{aligned} \tag{17}$$

根据式(9)～(17)可得如下命题.

命题 2　$-\max\left\{\frac{X_{t-j}}{1+\alpha_j},\frac{X_{t-j}}{1-\alpha_j},\frac{2X_{t-j}}{1-\alpha_j^2}\right\}\leqslant\frac{\partial \log P_{\boldsymbol{\theta}}}{\partial \alpha_j}\leqslant\max\left\{\frac{X_{t-j}}{\alpha_j(1+\alpha_j)},\frac{X_{t-j}}{\alpha_j(1-\alpha_j)},\frac{2X_{t-j}}{\alpha_j(1-\alpha_j^2)}\right\}$.

证明：由于

$$\begin{aligned}\frac{\partial \log P_{\boldsymbol{\theta}}}{\partial \alpha_j}\geqslant&\frac{1}{P_{\boldsymbol{\theta}}}\Bigg[-\frac{\varphi_j X_{t-j}}{1+\alpha_j}\sum_{m_j=0}^{X_t}q(X_t,m_j)(1-p_j)^{X_{t-j}}Y(X_{t-j},m_j)-\\&\frac{X_{t-j}}{1-\alpha_j}\sum_{m_j=0}^{\min\{X_t,X_{t-j}\}}q(X_t,m_j)p_j^{X_{t-j}}Z(X_{t-j},m_j)-\sum_{m_j=0}^{\min\{X_t,X_{t-j}\}}\varphi_j q(X_t,m_j)\times\\&\sum_{i=1}^{X_{t-j}-1}\left(\frac{i}{1-\alpha_j}+\frac{X_{t-j}-i}{1+\alpha_j}\right)r(X_{t-j},i)\sum_{l=0}^{\min\{m_j,i\}}Z(i,l)Y(X_{t-j}-i,m_j-l)\Bigg],\end{aligned}$$

所以 $\frac{\partial \log P_{\boldsymbol{\theta}}}{\partial \alpha_j}\geqslant-\max\left\{\frac{X_{t-j}}{1+\alpha_j},\frac{X_{t-j}}{1-\alpha_j},\frac{2X_{t-j}}{1-\alpha_j^2}\right\}$.

另一方面，根据式(9)～(17)有

$$\begin{aligned}\frac{\partial \log P_{\boldsymbol{\theta}}}{\partial \alpha_j}\leqslant&\frac{1}{P_{\boldsymbol{\theta}}}\Bigg[\sum_{m_j=0}^{X_t}\varphi_j q(X_t,m_j)\frac{m_j}{\alpha_j(1+\alpha_j)}(1-p_j)^{X_{t-j}}Y(X_{t-j},m_j)+\\&\sum_{m_j=0}^{\min\{X_t,X_{t-j}\}}\varphi_j q(X_t,m_j)\Bigg[\frac{X_{t-j}}{\alpha_j(1-\alpha_j)}p_j^{X_{t-j}}Z(X_{t-j},m_j)+\sum_{i=1}^{X_{t-j}-1}r(X_{t-j},i)\times\\&\sum_{l=0}^{\min\{m_j,i\}}\left[\frac{i}{\alpha_j(1-\alpha_j)}+\frac{m_j-l}{\alpha_j(1+\alpha_j)}\right]Z(i,l)Y(X_{t-j}-i,m_j-l)\Bigg]\Bigg]\leqslant\\&\max\left\{\frac{X_{t-j}}{\alpha_j(1+\alpha_j)},\frac{X_{t-j}}{\alpha_j(1-\alpha_j)},\frac{2X_{t-j}}{\alpha_j(1-\alpha_j^2)}\right\}.\end{aligned}$$

证毕.

3　参数估计量的渐近正态性

下面根据定理 1、命题 2 和式(9)～(17)，讨论参数估计量 $\hat{\boldsymbol{\theta}}$ 的渐近正态性. 采用文献[24-25]的方法证明估计量 $\hat{\boldsymbol{\theta}}$ 的渐近正态性. 首先给出如下正则条件：

(H_1) 集合 $\left\{m:P(\varepsilon_t=m)=f(m,\lambda)=\frac{\lambda^m}{m!}e^{-\lambda}>0\right\}$ 不依赖于 λ；

(H_2) $E(\varepsilon_t^3)=\lambda^3+3\lambda^2+\lambda<\infty$；

(H_3) $P(\varepsilon_t=m)$是三次连续可微的；

(H_4) 对任意的 $\lambda'\in B$(这里 B 是集合 $\mathbb{R}$ 的一个开子集)，存在 λ'的一个领域 U，使得

$$\sum_{k=0}^{\infty}\sup_{\lambda\in U}f(k,\lambda)<\infty,\quad\sum_{k=0}^{\infty}\sup_{\lambda\in U}\left|\frac{\partial f(k,\lambda)}{\partial \lambda}\right|<\infty,\quad\sum_{k=0}^{\infty}\sup_{\lambda\in U}\left|\frac{\partial^2 f(k,\lambda)}{\partial \lambda^2}\right|<\infty;$$

(H_5) 对于 $\lambda'\in B$，存在 λ' 的一个邻域 U 和常数 C_1,C_2,C_3，使得序列 $\varphi_1(n)=C_1n$，$\varphi_{11}(n)=C_2n^2$，$\varphi_{111}(n)=C_3n^3$，当 $n\geqslant 0$ 时，对 $\forall\lambda\in U$ 和 $\forall m\leqslant n$，在过程 $\{X_t\}$ 具有平稳遍历分布的条件下，$P(\varepsilon_t)$ 的非零概率密度函数 $f(m,\lambda)$ 满足：

$$\left|\frac{\partial f(k,\lambda)}{\partial\lambda}\right|\leqslant\varphi_1(n)f(m,\lambda),\quad \left|\frac{\partial^2 f(k,\lambda)}{\partial\lambda^2}\right|\leqslant\varphi_{11}(n)f(m,\lambda),\quad \left|\frac{\partial^3 f(k,\lambda)}{\partial\lambda^3}\right|\leqslant\varphi_{111}(n)f(m,\lambda);$$

(H_6) 记 $\boldsymbol{I}(\boldsymbol{\theta})=(\sigma_{ij})_{(3k+1)(3k+1)}$ 为 Fisher 信息矩阵，这里 $i,j=1,2,\cdots,3k+1$，且 $\boldsymbol{I}(\boldsymbol{\theta})$ 是非奇异的，并有 $\boldsymbol{I}(\boldsymbol{\theta})=E\left(\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\theta_i}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\theta_j}\right)$.

下面给出文献[24]中定理 1.3 的一种特殊形式.

定理 2 假设正则条件(H_1)～(H_6)成立，则 $\boldsymbol{\theta}$ 的条件极大似然估计量 $\hat{\boldsymbol{\theta}}$ 具有渐近正态性，且有

$$\sqrt{N}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\xrightarrow{d}N(0,\boldsymbol{I}^{-1}(\boldsymbol{\theta})),\qquad N\to\infty.$$

证明：根据文献[25]，显然满足条件(H_1)～(H_4). 由 k 阶相依驱动随机系数混合算子整数值自回归模型的定义知，随机项 ε_t 服从参数为 λ 的 Poisson 分布，因此也满足条件(H_5). 下面只需证明满足条件(H_6)，从而只需证明下列结论成立：

1) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\alpha_j}\right|^2<\infty$;

2) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\varphi_j}\right|^2<\infty$;

3) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial p_j}\right|^2<\infty$, $j=1,2,\cdots,k$;

4) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\lambda}\right|^2<\infty$;

5) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\alpha_j}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\alpha_q}\right|<\infty$, $j\neq q$, $j,q=1,2,\cdots,k$;

6) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\alpha_j}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\varphi_m}\right|<\infty$, $j,m=1,2,\cdots,k$;

7) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\alpha_j}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial p_m}\right|<\infty$;

8) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\alpha_j}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\lambda}\right|<\infty$;

9) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\varphi_j}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\varphi_q}\right|<\infty$;

10) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\varphi_j}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial p_m}\right|<\infty$;

11) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\varphi_j}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\lambda}\right|<\infty$;

12) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial p_j}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial p_q}\right|<\infty$;

13) $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial p_j}\cdot\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\lambda}\right|<\infty$.

首先证明结论 1). 根据定理 1 和命题 1，可得 $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\alpha_j}\right|^2<C\cdot E(X_{k+1-j}^2)<\infty$，这里 C 是一个合适的常数. 其次证明结论 2). 根据得分函数 $\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\varphi_j}$ 的表达式，可得 $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\varphi_j}\right|<\frac{1}{\varphi_j}<\infty$，从而 $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\varphi_j}\right|^2<\infty$. 再次，对于结论 3)，根据得分函数 $\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial p_j}$ 的表达式，有 $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial p_j}\right|<C\cdot E(X_{k+1-j})<\infty$，因此 $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial p_j}\right|^2<C\cdot E(X_{k+1-j}^2)<\infty$；对于结论 4)，根据条件($H_5$)，有 $\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\lambda}\right|<\varphi_1(X_{k+1})$，因此可得 $E\left|\frac{\partial\log P_{\boldsymbol{\theta}}}{\partial\lambda}\right|^2<E\varphi_1^2(X_{k+1})<\infty$. 最后，基于结论 1)～4)和

条件(H_5)，可得结论 5)～13)成立. 从而可定义 Fisher 信息矩阵 $\boldsymbol{I}(\boldsymbol{\theta})$，同时根据条件($H_6$)，可得 $\boldsymbol{I}(\boldsymbol{\theta})$ 是非奇异的. 证毕.

注 1　当新息序列$\{\varepsilon_t\}$的分布未知时，可令 $\mu_\varepsilon=E(\varepsilon_t)$，$\sigma_\varepsilon^2=\mathrm{Var}(\varepsilon_t)$，此时参数为 $\boldsymbol{\theta}=(\alpha_1,\cdots,\alpha_k,\varphi_1,\cdots,\varphi_k,p_1,\cdots,p_k,\mu_\varepsilon,\sigma_\varepsilon^2)'$. 从而可采用文献[26]中的修正拟似然估计法估计未知参数，基于修正拟似然估计量 $\check{\boldsymbol{\theta}}$ 重新获得方差的估计量，记为 $T^{-1}(\boldsymbol{\theta})$，这里 $T(\boldsymbol{\theta})$ 的估计量为 $\hat{T}(\boldsymbol{\theta})=\frac{1}{n}\sum_{t=k+1}^{n}m_t(\check{\boldsymbol{\theta}})m_t(\check{\boldsymbol{\theta}})'$，其中 $m_t(\boldsymbol{\theta})=(m_{t1}(\boldsymbol{\theta}),\cdots,m_{t,3k+2}(\boldsymbol{\theta}))'$，

$$m_{ti}(\boldsymbol{\theta})=\frac{X_t-E(X_t\mid X_{t-1},\cdots,X_{t-k})}{\mathrm{Var}_{\boldsymbol{\theta}}(X_t\mid X_{t-1},\cdots,X_{t-k})}\cdot\frac{\partial E(X_t\mid X_{t-1},\cdots,X_{t-k})}{\partial\theta_i},\qquad i=1,2,\cdots,3k+2.$$

4　数值模拟

下面通过数值模拟评估条件极大似然估计量的小样本性能. 考虑 Po-DDRCMTINAR(2)模型. 采用估计量的有限样本平均绝对偏差误差(MADE)和均方误差(MSE)进行评价，其中

$$\mathrm{MADE}=\frac{1}{M}\sum_{i=1}^{M}|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}|,\qquad \mathrm{MSE}=\frac{1}{M}\sum_{i=1}^{M}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^2,$$

这里 $\hat{\boldsymbol{\theta}}$ 表示 $\boldsymbol{\theta}$ 的估计量，M 表示重复模拟次数. 考虑下列 4 组参数组合：

1) $\alpha_1=0.3$，$\alpha_2=0.4$，$\varphi_1=0.4$，$\varphi_2=0.5$，$p_1=0.45$，$p_2=0.5$，$\lambda=0.3$；

2) $\alpha_1=0.3$，$\alpha_2=0.4$，$\varphi_1=0.4$，$\varphi_2=0.5$，$p_1=0.45$，$p_2=0.5$，$\lambda=0.5$；

3) $\alpha_1=\alpha_2=0.4$，$\varphi_1=\varphi_2=0.3$，$p_1=0.45$，$p_2=0.5$，$\lambda=0.3$；

4) $\alpha_1=\alpha_2=0.4$，$\varphi_1=\varphi_2=0.3$，$p_1=p_2=0.5$，$\lambda=0.5$.

样本容量的取值分别为 $n=100,300,500,800,1\,000$，模拟实验重复 500 次. 图 1 为样本量 $n=100$ 时，Po-DDRCMTINAR(2)模型在 4 组参数下的路径图.

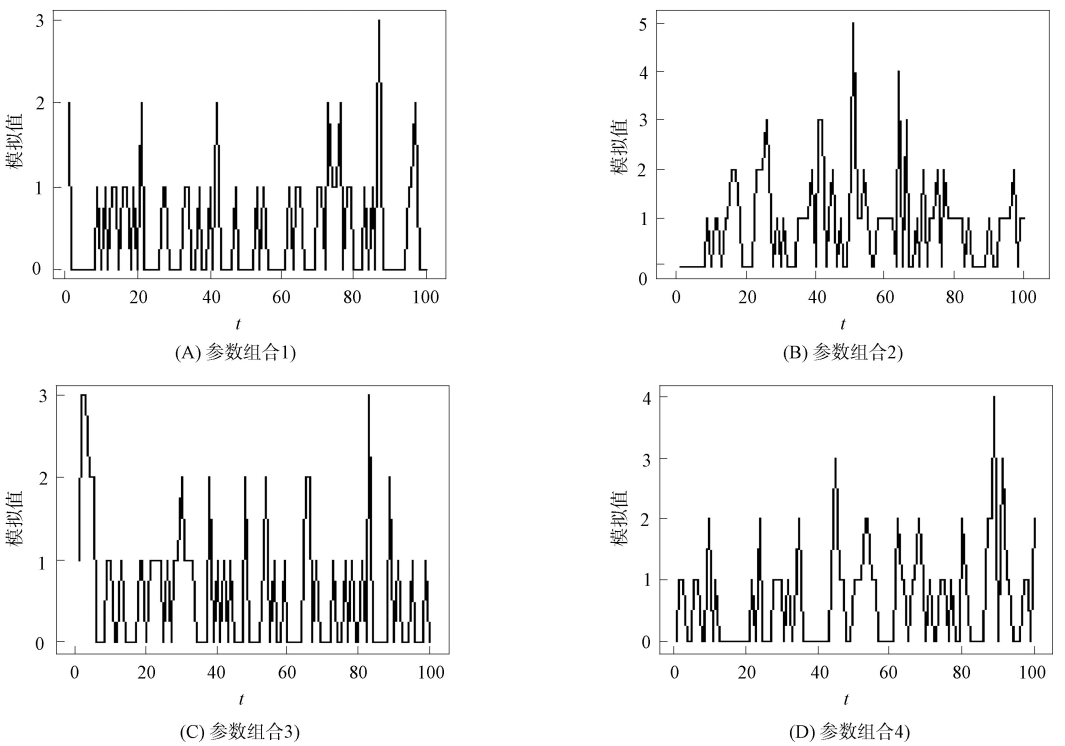


图 1　当 $n=100$ 时，Po-DDRCMTINAR(2)模型在 4 组参数下的路径图

Fig. 1　Path diagrams of Po-DDRCMTINAR(2) model under four sets of parameters when $n=100$

表 1 和表 2 列出了 4 组参数下该模型的条件极大似然估计结果. 由表 1 和表 2 可见，随着样本量

的增大，绝对误差和均方误差都逐渐减小，表明参数估计逐渐收敛到真实值.

表 1 在不同参数组合下不同样本量参数估计的 MADE

Table 1 MADE of parameter estimation of different sample sizes under different parameter combinations

参数组合	n	MADE						
		α_1	α_2	φ_1	φ_2	p_1	p_2	λ
1)	100	0.483 1	0.331 6	0.272 9	0.298 7	0.137 8	0.106 9	0.048 8
	300	0.295 5	0.190 8	0.233 9	0.232 3	0.113 6	0.087 4	0.027 7
	500	0.207 5	0.159 3	0.202 1	0.183 5	0.110 5	0.084 7	0.022 1
	800	0.186 7	0.144 9	0.182 9	0.169 6	0.109 8	0.082 9	0.019 8
	1 000	0.183 8	0.138 9	0.173 5	0.162 3	0.108 0	0.081 7	0.017 5
2)	100	0.398 3	0.262 0	0.261 0	0.280 2	0.100 9	0.064 0	0.058 0
	300	0.273 3	0.167 0	0.225 4	0.214 9	0.098 2	0.059 1	0.038 4
	500	0.224 5	0.143 6	0.198 8	0.185 3	0.098 1	0.058 8	0.034 6
	800	0.182 4	0.132 6	0.174 8	0.162 8	0.098 0	0.058 4	0.031 4
	1 000	0.176 2	0.127 5	0.164 3	0.158 0	0.097 3	0.058 0	0.031 3
3)	100	0.610 7	0.489 3	0.216 7	0.217 1	0.134 9	0.133 5	0.048 6
	300	0.266 4	0.298 6	0.188 4	0.194 7	0.092 9	0.093 2	0.027 0
	500	0.207 2	0.244 9	0.182 2	0.178 6	0.087 4	0.089 7	0.021 5
	800	0.178 5	0.218 2	0.173 8	0.169 9	0.085 7	0.084 5	0.017 0
	1 000	0.173 0	0.199 8	0.159 1	0.166 1	0.083 5	0.083 5	0.016 0
4)	100	0.487 2	0.473 9	0.209 3	0.223 1	0.068 6	0.069 6	0.053 3
	300	0.249 4	0.250 3	0.193 0	0.202 0	0.059 4	0.059 7	0.036 7
	500	0.202 7	0.212 0	0.177 2	0.181 9	0.058 2	0.059 0	0.031 5
	800	0.175 2	0.198 3	0.168 9	0.163 9	0.058 2	0.058 1	0.027 9
	1 000	0.170 7	0.190 4	0.163 9	0.162 7	0.058 1	0.058 0	0.025 6

表 2 在不同参数组合下不同样本量参数估计的 MSE

Table 2 MSE of parameter estimation of different sample sizes under different parameter combinations

参数组合	n	MSE						
		α_1	α_2	φ_1	φ_2	p_1	p_2	λ
1)	100	0.500 0	0.270 0	0.088 2	0.112 9	0.029 1	0.020 6	0.003 9
	300	0.214 0	0.058 5	0.069 1	0.072 2	0.014 1	0.009 3	0.001 3
	500	0.071 2	0.039 7	0.055 2	0.047 3	0.012 6	0.007 8	0.000 8
	800	0.058 1	0.033 3	0.045 0	0.040 5	0.012 2	0.007 2	0.000 6
	1 000	0.054 7	0.031 7	0.040 7	0.037 3	0.011 8	0.006 9	0.000 5
2)	100	0.312 0	0.120 7	0.083 8	0.100 7	0.010 7	0.005 1	0.005 3
	300	0.208 3	0.045 4	0.067 1	0.062 6	0.009 7	0.036 0	0.002 2
	500	0.083 9	0.032 5	0.052 3	0.047 2	0.009 7	0.003 5	0.001 7
	800	0.054 8	0.027 5	0.041 4	0.036 5	0.009 6	0.003 5	0.001 3
	1 000	0.053 1	0.026 6	0.037 4	0.035 9	0.009 5	0.003 4	0.001 2
3)	100	0.983 6	0.823 9	0.060 1	0.057 9	0.036 7	0.034 1	0.003 8
	300	0.163 4	0.155 2	0.048 6	0.049 7	0.012 1	0.011 7	0.001 2
	500	0.069 2	0.089 5	0.045 7	0.043 6	0.008 7	0.009 3	0.000 7
	800	0.047 6	0.069 9	0.044 4	0.043 0	0.007 9	0.007 6	0.000 4
	1 000	0.045 8	0.061 9	0.041 6	0.041 7	0.007 4	0.007 4	0.000 4
4)	100	0.578 8	0.741 2	0.056 3	0.062 2	0.007 4	0.006 8	0.004 6
	300	0.119 5	0.095 8	0.052 9	0.055 8	0.003 7	0.003 8	0.002 1
	500	0.080 3	0.068 9	0.045 9	0.049 1	0.003 5	0.003 6	0.001 4
	800	0.046 4	0.060 5	0.044 6	0.042 3	0.003 5	0.003 4	0.001 1
	1 000	0.040 8	0.056 0	0.042 4	0.040 0	0.003 4	0.003 4	0.001 0

本文对参数 $p_i(i=1,2)$是否为 1 进行假设检验. 原假设 H_0：$p_i-1=0$，备择假设 H_1：$p_i-1<0$. 在参数组合 1)下，基于样本量 $n=100,300,500,800,1\ 000$ 和 500 次重复实验，计算检验统计量

$t=\dfrac{\hat{p}_i-1}{\sqrt{\mathrm{Var}(\hat{p}_i)}}$ 的值. 参数 p_1 相应的检验统计量值分别为 $-1.328\ 7$，$-2.259\ 3$，$-3.010\ 0$，$-3.704\ 2$，$-4.273\ 9$；参数 p_2 相应的检验统计量值分别为 $-1.516\ 8$，$-2.797\ 1$，$-3.703\ 8$，$-4.857\ 6$，$-5.498\ 0$. 当显著性水平 $\alpha=0.05$ 时，在小样本 $n=100$ 情况下，参数 $p_i(i=1,2)$ 未通过显著性检验，即接受 $p_i=1$ 的假设；在其他样本下参数 $p_i(i=1,2)$ 通过了显著性检验，即 $p_i-1<0$. 可见，样本量越大，参数 $p_i(i=1,2)$ 统计量的值越小于临界值，参数 $p_i-1<0$ 可信度越高，因此本文模型不会退化为 Po-DDRCINAR(k)模型的可信度较高.

5　实例分析

考虑一组我国从 2020 年 5 月 4 日到 2020 年 9 月 23 日共 143 d 的全国新冠感染每日疑似病例数量. 数据变化范围是 0～18(http://www.nhc.gov.cn/wjw/xwdt/list_83.shtml). 该组数据的均值和方差分别是 3.944 1 和 10.827 8. 图 2 为该组数据的散点图、自相关(ACF)图和偏自相关(PACF)图. 由图 2，本文考虑下列 4 种模型：Po-DDRCMTINAR(1)，Po-DDRCMTINAR(2)，Po-DDRCINAR(1)[10]，Po-DDRCINAR(2)[10]. 表 3 列出了上述 4 种模型参数的极大似然估计和用于评价模型优劣的 AIC，BIC 和 HQ 准则[10]. 由表 3 可见，本文提出的 Po-DDRCMTINAR(1)模型分析该组数据的 AIC，BIC 和 HQ 值最小. 基于 Po-DDRCMTINAR(1)模型，图 3 为拟合残差的路径图、直方图、ACF 图和 PACF 图. 由图 3 可见，该模型的拟合残差近似服从标准正态分布. 表明本文提出的一阶 Po-DDRCMTINAR(1)模型更适合分析该组数据.

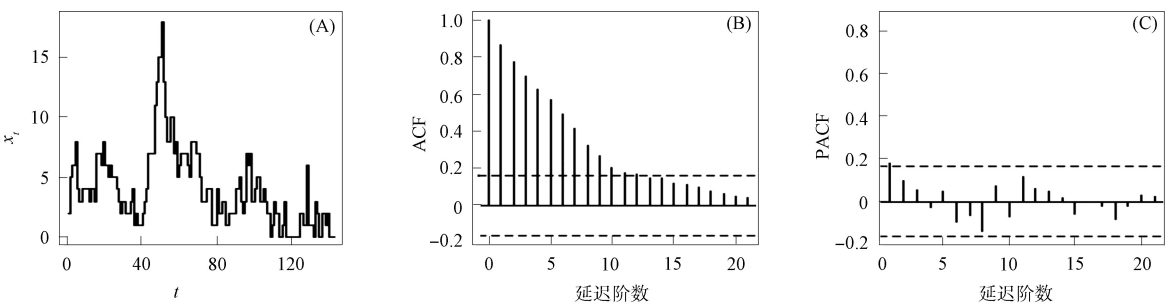


图 2　新冠感染疑似病例数据的计数图(A)、ACF 图(B)和 PACF 图(C)

Fig. 2　Count plot (A), ACF plot (B) and PACF plot (C) of suspected case data of COVID-19

表 3　不同模型对新冠感染疑似病例数据的参数极大似然估计和 AIC，BIC，HQ 值

Table 3　Maximum likelihood estimation of parameters and AIC，BIC，HQ values of different models for suspected case data of COVID-19

Po-DDRCMTINAR(1)	Po-DDRCMTINAR(2)	Po-DDRCINAR(1)	Po-DDRCINAR(2)
$\hat{\alpha}_1=0.848\ 7$	$\hat{\alpha}_1=0.896\ 5$	$\hat{\alpha}_1=0.922\ 8$	$\hat{\alpha}_1=0.674\ 1$
$\hat{\varphi}_1=0.970\ 6$	$\hat{\alpha}_2=0.897\ 9$	$\hat{\varphi}_1=0.790\ 3$	$\hat{\alpha}_2=0.427\ 5$
$\hat{p}_1=0.567\ 8$	$\hat{\varphi}_1=0.885\ 1$	$\hat{\lambda}=0.984\ 4$	$\hat{\varphi}_1=0.851\ 4$
$\hat{\lambda}=0.757\ 0$	$\hat{\varphi}_2=0.106\ 4$	AIC=530.353 5	$\hat{\varphi}_2=0.415\ 7$
AIC=36.789 0	$\hat{p}_1=0.565\ 9$	BIC=539.242 1	$\hat{\lambda}=1.374\ 2$
BIC=48.640 4	$\hat{p}_2=0.555\ 1$	HQ=529.159 5	AIC=546.694 8
HQ=35.196 9	$\hat{\lambda}=0.683\ 8$		BIC=561.509 1
	AIC=60.426 7		HQ=544.704 7
	BIC=81.166 6		
	HQ=57.640 5		

综上，本文提出了 k 阶 Poisson 相依驱动随机系数混合算子整数值自回归模型，讨论了该模型的统计性质和参数的条件极大似然估计，证明了估计量的渐近正态性. 最后通过数值模拟和实例分析验证了模型的有效性.

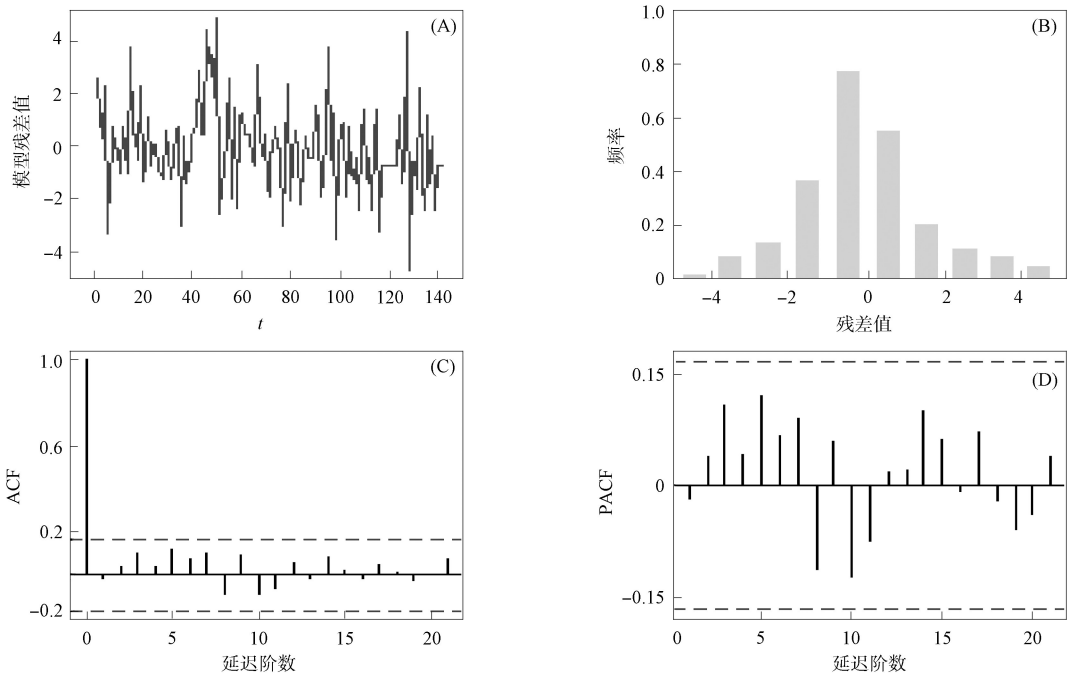


图 3 新冠感染疑似病例数据模型的残差图(A)、残差直方图(B)、ACF 图(C)和 PACF 图(D)

Fig. 3 Residual plot (A), residual histogram (B), ACF plot (C) and PACF plot (D) of suspected case data model of COVID-19

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] BÖCKENHOLT U. An INAR(1) Negative Multinomial Regression Model for Longitudinal Count Data [J]. Psychometrika, 1999, 64(1): 53-67.

[3] KHOO W C, ONG S H, BISWAS A. Modeling Time Series of Counts with a New Class of INAR(1) Model [J]. Statistical Papers, 2017, 58(2): 393-416.

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

[5] SILVA M E, OLIVEIRA V L. Difference Equations for the Higher Order Moments and Cumulants of the INAR(p) Model [J]. Journal of Time Series Analysis, 2005, 26(1): 17-36.

[6] ZHENG H T, BASAWA I, DATTA S. First-Order Random Coefficient Integer-Valued Autoregressive Processes [J]. Journal of Statistical Planning and Inference, 2007, 137(1): 212-229.

[7] ZHENG H T, BASAWA I V. First-Order Observation-Driven Integer-Valued Autoregressive Processes [J]. Statistics and Probability Letters, 2008, 78(1): 1-9.

[8] BAKOUCH H S, MOHAMMADPOUR M, SHIROZHAN M. A Zero-Inflated Geometric INAR(1) Process with Random Coefficient [J]. Applications of Mathematics, 2018, 63(1): 79-105.

[9] ZHENG H T, BASAWA I V, DATTA S. Inference for pth-Order Random Coefficient Integer-Valued Autoregressive Processes [J]. Journal of Time Series Analysis, 2006, 27(3): 411-440.

[10] LIU X F, WANG D H, DENG D L, et al. Maximum Likelihood Estimation of the DDRCINAR(p) Model [J]. Communications in Statistics: Theory and Methods, 2021, 50(24): 6231-6255.

[11] RISTIĆ M M, NASTIĆ A S, BAKOUCH H S. Estimation in an Integer-Valued Autoregressive Process with Negative Binomial Marginals (NBINAR(1)) [J]. Communications in Statistics: Theory and Methods, 2012, 41(4): 606-618.

[12] 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.

[13] TIAN S Q, WANG D H, CUI S. A Seasonal Geometric INAR Process Based on Negative Binomial Thinning Operator [J]. Statistical Papers, 2020, 61(6): 2561-2581.

[14] NASTIĆ A S, RISTIĆ M M, BAKOUCH H S. A Combined Geometric INAR(p) Model Based on Negative Binomial Thinning [J]. Mathematical and Computer Modelling, 2012, 55(5/6): 1665-1672.

[15] LIU X F, ZHANG W K, YIN W Z, et al. Maximum Likelihood Estimation of the SDMINAR(p) Model to Analyze Some COVID-19 Data [J/OL]. Communications in Statistics: Simulation and Computation, (2022-11-04)[2023-11-25]. https://doi.org/10.1080/03610918.2022.2145486.

[16] YANG K, LI A, LI H, et al. High-Order Self-excited Threshold Integer-Valued Autore Gressive Model: Estimation and Testing [J/OL]. Communications in Mathematics and Statistics, (2023-05-13)[2023-11-25]. https://doi.org/10.1007/s40304-022-00325-3.

[17] YANG K, WANG D H, JIA B T, et al. An Integer-Valued Threshold Autoregressive Process Based on Negative Binomial Thinning [J]. Statistical Papers, 2018, 59: 1131-1160.

[18] WANG X H, WANG D H, YANG K, et al. Estimation and Testing for the Integer-Valued Threshold Autoregressive Models Based on Negative Binomial Thinning [J]. Communications in Statistics: Simulation and Computation, 2021, 50(6): 1622-1644.

[19] NASTIĆ A S, RISTIĆ M M, JANJIĆ A D. A Mixed Thinning Based Geometric INAR(1) Model [J]. Filomat, 2017, 31(13): 4009-4022.

[20] CHANG L Y, LIU X F, WANG D H, et al. First-Order Random Coefficient Mixed-Thinning Integer-Valued Autoregressive Model [J]. Journal of Computational and Applied Mathematics, 2022, 410: 114222-1-114222-24.

[21] LIU X F, WANG D H, CHEN H P, et al. A Class of kth-Order Dependence-Driven Random Coefficient Mixed Thinning Integer-Valued Autoregressive Process to Analyse Epileptic Seizure Data and COVID-19 Data [J]. Australian and New Zealand Journal of Statistics, 2024, 66(2): 249-280.

[22] RISTIĆ M M, NASTIĆ A S. A Mixed INAR(p) Model [J]. Journal of Time Series Analysis, 2012, 33(6): 903-915.

[23] SHIRYAEV A N. Veroyatnost' [M]. 2nd ed. Moscow: Nauka, 1989: 441.

[24] BILLINGSLEY P. Statisrical Inference for Markov Processes [M]. Chicago: University of Chicago Press, 1961: 1-72.

[25] FRANKE J, SELIGMANN T. Conditional Maximum Likelihood Estimates for INAR(1) Processes and Their Application to Modelling Epileptic Seizure Counts [M]. Developments in Time Series Analysis. New York: Champman and Hall, 1993: 310-330.

[26] LI H, YANG K, WANG D H. Quasi-likelihood Inference for Self-exciting Threshold Integer-Valued Autoregressive Processes [J]. Computational Statistics, 2017, 32(4): 1597-1620.

(责任编辑：李　琦)