



INAR(1)-PK: Modeling Discrete-Time Count Data with Overdispersion

Rasekhi, M. 

Department of Statistics, Faculty of Mathematical Sciences and Statistics
Malayer University, Malayer, Iran.

Corresponding author: M. Rasekhi, m.rasekhi@malayeru.ac.ir

Received: 29/6/2025 **Revised:** 22/11/2025 **Accepted and Published Online:** 23/11/2025.

Introduction

Count time series arise naturally in many real-world applications, such as weekly airline passenger counts, annual deaths from respiratory diseases, and short-interval stock transaction volumes. They are especially common in veterinary and biomedical research, where data such as the number of animals with malnutrition or skin lesions exhibit correlation and overdispersion. To better model such data, a new discrete distribution, the Poisson-Komal, is introduced. This distribution flexibly handles overdispersion and offers favorable statistical properties for count modeling. Leveraging the Poisson-Komal distribution, a novel INAR(1) (Integer-valued Autoregressive of order 1) model is developed using the binomial thinning operator, a foundational technique established by McKenzie (1986) and Al-Osh and Alzaid (1987). The resulting model maintains the interpretability of classical INAR processes while accommodating complex dispersion patterns typical in empirical count data. Its performance is evaluated on real veterinary datasets concerning animal health indicators, with comparisons to existing models showing improved fit and predictive accuracy. By integrating the Poisson-Komal distribution into the INAR framework, this approach enhances the ability to analyze overdispersed, correlated count time series across diverse fields, including epidemiology, finance, and veterinary science, offering practitioners a more robust and adaptable modeling tool.

Material and Methods

This paper introduces a novel INAR(1) model based on the discrete Poisson–

Komal distribution. Various parametric and non-parametric estimation methods, including conditional maximum likelihood (parametric) and Yule-Walker (non-parametric), are examined. Moreover, several INAR(1) processes are considered for modeling real-world count data on the number of animals affected by malnutrition or skin lesions. Model fit quality is assessed using the Akaike Information Criterion and the Bayesian Information Criterion.

Results and Discussion

The simulation comparison is conducted based on two criteria: bias and mean squared error. Both measures perform well as the sample size increases, with results confirming the superiority of the conditional maximum likelihood estimation method in terms of mean squared error. To evaluate estimator performance, 1,000 replications are carried out across various sample sizes. Furthermore, based on model fit criteria, it is concluded that the INAR(1)-PK process is preferred over other INAR(1) models.

Conclusion

In this paper, a new first-order non-negative integer-valued autoregressive model with innovations following the Poisson-Komal distribution, denoted as INAR(1)-PK, is proposed for modeling overdispersed count time series. To estimate the model parameters, two distinct estimation methods are investigated, and their performance is compared through simulation studies. The proposed model is applied to two real-world datasets, and its results are compared against six two- and three-parameter INAR(1) models. Empirical findings indicate that the INAR(1)-PK model performs better in the presence of overdispersion and is therefore a more suitable choice for such data.

Keywords: Poisson-Komal distribution, INAR(1) process, Conditional Maximum Likelihood, binomial thinning operator, Yule-Walker estimate.1

Mathematics Subject Classification (2020): 62M10, 62F10.



©The Author(s). The Publisher is Iranian Statistical Society.

This is an open access article distributed under the terms and conditions of [CC BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

فرآیند INAR(1)-PK: مدل‌بندی داده‌های سری زمانی گسسته با بیش پراکندگی

مهدی راسخی

گروه آمار، دانشکده علوم ریاضی و آمار، دانشگاه ملایر

نویسنده مسئول: مهدی راسخی، m.rasekhi@malayeru.ac.ir

تاریخ دریافت: ۱۴۰۴/۴/۸ تاریخ بازنگری: ۱۴۰۴/۹/۱ تاریخ پذیرش و انتشار: ۱۴۰۴/۹/۲

چکیده: در این مقاله، یک فرایند خودبازگشتی از مرتبه اول با مقادیر صحیح نامنفی بر اساس عملگر دقیق‌ساز دوجمله‌ای و با نوفه‌ای از نوع پواسن-کومال معرفی شده است. به‌منظور برآورد پارامترهای این مدل، دو روش برآوردیابی شامل برآورد درستنمایی شرطی و روش یول-والکر مورد بررسی قرار گرفته‌اند. همچنین، با انجام یک مطالعه شبیه‌سازی، روش‌های برآوردیابی ارزیابی شده‌اند. علاوه بر این با استفاده از دو مجموعه داده واقعی در علوم دامپزشکی عملکرد فرآیند پیشنهادی نشان داده شده است. واژه‌های کلیدی: توزیع پواسن-کومال، فرآیند INAR(1)، برآورد ماکسیمم درستنمایی شرطی، برآورد یول-والکر.

کد موضوع‌بندی ریاضی (۲۰۲۰): 62F10، 62M10.

۱ مقدمه

سری‌های زمانی متشکل از داده‌های شمارشی در حوزه‌های مختلفی مانند بررسی بازارهای مالی، پژوهش‌های پزشکی، بیمه و سایر زمینه‌ها به طور رایج دیده می‌شوند. نمونه‌هایی از این داده‌ها عبارتند از: تعداد مسافران هواپیما در هر هفته، تعداد مرگ‌ومیر سالانه ناشی از بیماری‌های ریوی یا تعداد معاملات سهام در بازه‌های زمانی کوتاه. برای مدل‌سازی بهینه دنباله‌ی داده‌های شمارشی، مدل خودبازگشتی خطی از مرتبه اول با مقادیر صحیح نامنفی ((INAR(1) ارائه شده است که به طور مستقل توسط مکنزی (۱۹۸۵، ۱۹۸۶) و الزید و الاش (۱۹۸۸) توسعه یافت. این مدل‌ها به ویژه در مواقعی کاربرد دارند که داده‌ها دارای ویژگی بیش پراکندگی باشند، یعنی واریانس داده‌ها از میانگین آنها بیشتر



©نویسندگان). ناشر انجمن آمار ایران است.

این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

باشد. در سال‌های اخیر، محققان تلاش کرده‌اند مدل INAR(1) را با استفاده از نوفه‌هایی با توزیع‌های انعطاف‌پذیر بهبود دهند که سبب مدل‌بندی مناسب‌تر داده‌ها با مقادیر متفاوت بیش پراکندگی می‌شود.

الزید و الاش (۱۹۸۸) به کمک نوفه با توزیع پواسن یک فرایند خودبازگشتی خطی از مرتبه اول با مقادیر صحیح نامنفی برای مدل‌سازی داده‌های سری زمانی گسسته معرفی نمودند (INAR(1)-P). **لیویو و همکاران (۲۰۱۸)** یک فرایند INAR(1) مبتنی بر نوفه با توزیع پواسن-لیندلی را برای مدل‌بندی تعداد کالای فروخته شده در یک مغازه پیشنهاد دادند (INAR(1)-PL). **آلتون (۲۰۲۰)** نیز به معرفی توزیع پواسن-بلال دارای خاصیت بیش پراکندگی، پرداخت و از آن به عنوان توزیع نوفه در فرایند INAR(1) بهره برد (INAR(1)-PB). **آلتون و همکاران (۲۰۲۱)** نیز فرایند INAR(1) جدیدی را بر اساس نوفه با توزیع پواسن-کواسی گاما ارائه کردند و کارایی آن را بر داده‌های زلزله نشان دادند (INAR(1)-PQX). همچنین، **مایا و همکاران (۲۰۲۲)** مدل INAR(1) مبتنی بر نوفه با توزیع پواسن-نمایی توسعه‌یافته را برای مدل‌بندی تعداد بیماران دارای بیماری سیفلیس معرفی نمودند (INAR(1)-PEE). به‌علاوه فرایند دیگری مبتنی بر نوفه با توزیع پواسن-میرا برای مدل‌سازی سری زمانی داده‌های زلزله در ژاپن ارائه کردند (INAR(1)-PMI). **نوراللهی و همکاران (۱۴۰۱)** فرایند مشابهی را بر اساس نوفه با توزیع نمایی وایل گسسته برای مدل‌بندی تعداد فوت‌های ناشی از بیماری COVID-19 ایجاد نمودند.

در این مقاله، فرایند INAR(1) با عملکرد رقیق‌ساز دوجمله‌ای مرور شده و یک فرایند جدید بر اساس نوفه با توزیع پواسن-کومال (PK) که توانایی مدل‌کردن داده‌های سری‌زمانی با بیش پراکندگی را داراست، معرفی می‌شود. توزیع PK که توسط **الامیر و احسان الحق (۲۰۲۵)** پیشنهاد شده است، انعطاف‌پذیری بالایی در مدل‌بندی داده‌های دارای بیش پراکندگی را دارد. ویژگی‌های آماری و برآوردیابی پارامترهای مدل پیشنهادی مورد بررسی قرار گرفته، عملکرد روش‌های برآوردیابی با یک مطالعه شبیه‌سازی ارزیابی شده و مدل روی دو مجموعه داده واقعی بکارگرفته می‌شود. نتایج حاصل از مدل‌بندی فرایند پیشنهادی با دیگر فرایندهای موجود مقایسه می‌شوند.

بقیه مقاله به شرح زیر سازماندهی شده است: در بخش ۲، ویژگی‌های آماری اصلی توزیع پواسن-کومال مورد بررسی قرار می‌گیرد. بخش ۳ به معرفی فرایند INAR(1)-PK و ویژگی‌های آن اختصاص دارد. در بخش ۴، دو روش برآوردیابی برای استنباط پارامترها ارائه می‌شوند و در بخش ۵، یک مطالعه شبیه‌سازی به منظور ارزیابی و مقایسه این روش‌های برآورد انجام می‌پذیرد. کاربرد عملی و اهمیت مدل پیشنهادی نیز در بخش ۶ از طریق دو مثال مبتنی بر داده‌های واقعی نشان داده می‌شود. در نهایت، بخش ۷ مقاله را با جمع‌بندی اصلی یافته‌ها به پایان می‌رساند.

۲ توزیع پواسن-کومال

الامیر و احسان الحق (۲۰۲۵) یک توزیع گسسته ترکیبی با توزیع پواسن و توزیع کومال (**شانکر، ۲۰۲۳**) ارائه کردند

و آن را با PK نشان دادند. توابع جرم احتمال و توزیع تجمعی این توزیع به ترتیب به صورت

$$f(x; \delta) = \frac{\delta^2}{\delta^2 + \delta + 1} \left[\frac{1}{(\delta + 1)^x} + \frac{x + 1}{(\delta + 1)^{x+2}} \right], \quad x = 0, 1, \dots$$

$$F(x; \delta) = 1 - \frac{\delta^2 + 2\delta^2 + \delta x + 3\delta + 1}{(\delta^2 + \delta + 1)(\delta + 1)^{x+2}}$$

و r -امین گشتاور فاکتوریل این توزیع را به صورت

$$\mu'_{(r)} = \frac{\delta^2}{\delta^2 + \delta + 1} \left[\frac{(1 + \delta)\Gamma(r + 1)}{\delta^{r+1}} + \frac{\Gamma(r + 2)}{\delta^{r+2}} \right].$$

بدست آوردند. حال بر اساس رابطه بدیهی بین گشتاور فاکتوریل و گشتاور ساده، دو گشتاور اول ساده این توزیع به ترتیب به صورت

$$E(X) = \frac{2 + \delta + \delta^2}{\delta(\delta^2 + \delta + 1)}, \quad E(X^2) = \frac{6 + 4\delta + 3\delta^2 + \delta^3}{\delta^2(\delta^2 + \delta + 1)^2}.$$

و واریانس آن به صورت

$$V(X) = \frac{2 + 6\delta + 8\delta^2 + 6\delta^3 + 4\delta^4 + \delta^5}{\delta^2(\delta^2 + \delta + 1)^2}.$$

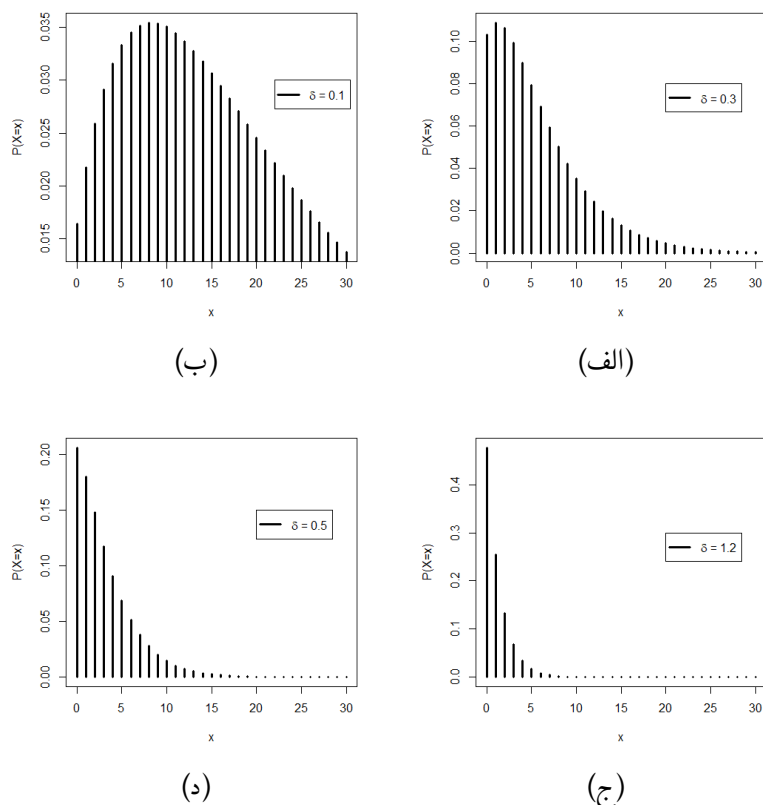
است. همچنین شاخص پراکندگی براساس فرمول

$$DI(X) = \frac{2 + 6\delta + 8\delta^2 + 6\delta^3 + 4\delta^4 + \delta^5}{\delta(\delta^2 + \delta + 1)(\delta^2 + \delta + 2)}.$$

حاصل می شود. شکل ۱ نمودارهای تابع جرم احتمال توزیع PK را تحت مقادیر مختلف پارامتر نشان می دهد. شکل ۲ بیان کننده آن است که برای تمامی مقادیر پارامتر δ ، شاخص پراکندگی بزرگتر از ۱ است. این رفتار، کارایی این توزیع در مدل سازی داده های شمارشی با بیش پراکندگی را تأیید می کند.

شانکر (۲۰۲۳) توزیع کومال را بر اساس توزیع آمیخته نمایی و گاما ایجاد نمود. با بهره بردن از روش تولید داده از توزیع آمیخته، الگوریتم تولید داده از این توزیع ارائه می شود.

- گام ۱: تولید $K_i, i = 1, \dots, n$ بر اساس نمونه با جایگذاری از اعداد ۱ و ۲ با بردار احتمال $(\frac{\delta(\delta+1)}{\delta^2+\delta+1}, \frac{1}{\delta^2+\delta+1})$.
- گام ۲: اگر $K_i = 1$ مقدار λ_i عبارتست از یک نمونه تولیدی از $\exp(\delta)$ و در غیر این صورت این مقدار برابر است با یک نمونه تولیدی از $\Gamma(2, \delta)$.
- گام ۳: تولید X_i از توزیع $Poisson(\lambda_i)$.



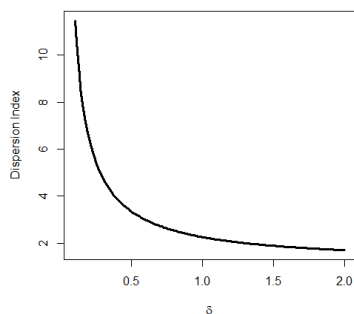
شکل ۰۱. نمودار تابع جرم احتمال توزیع PK تحت مقادیر مختلف پارامتر

۳ مدل INAR(1) براساس نوفه‌هایی از توزیع PK

فرض کنید $\{X_t\}_{t \in \mathbb{Z}}$ نشان دهنده فرایند INAR(1) بر اساس تعریف

$$X_t = p \circ X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad (۱)$$

باشد، که در آن $0 \leq p < 1$ و $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ دنباله ای از متغیرهای تصادفی هم توزیع و مستقل با مقادیر صحیح نامنفی و میانگین $E(\varepsilon_t) = \mu_\varepsilon$ و واریانس $V(\varepsilon_t) = \sigma_\varepsilon^2$ است. فرض کنید مولد $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ مستقل از مشاهدات گذشته $X_{t-k}, k \geq 1$ است. عملگر رقیق‌ساز $\circ$ نشان دهنده عملگر رقیق‌ساز دو جمله ای است که به



شکل ۲. نمودار شاخص پراکندگی توزیع PK تحت مقادیر مختلف پارامتر

فرم $X_{t-1} := \sum_{j=1}^{X_{t-1}} W_j$ تعریف می‌شود که در آن $\{W_j\}_{j \geq 1}$ نشان دهنده دنباله‌ای از متغیرهای تصادفی هم‌توزیع و مستقل با توزیع برنولی با احتمال موفقیت p است. اگر $0 \leq p < 1$ این فرآیند ایستا و در غیر این صورت فرآیند نایستا است. برای تشخیص ساختار حرکت در فرآیند، احتمال انتقال یک مرحله‌ای

$$P(X_t = k | X_{t-1} = \ell) = \sum_{i=1}^{\min(k, \ell)} \binom{\ell}{i} p^i (1-p)^{\ell-i} P(\varepsilon_t = k-i), \quad k, \ell \geq 0$$

است، که در آن $0 < p < 1$. برای ایجاد یک فرآیند تصادفی INAR(1) جدید، دنباله متغیرهای تصادفی $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ که از توزیع PK پیروی می‌کند را به عنوان متغیرهای نوفه فرآیند در برابری (۱) لحاظ می‌شود. در نتیجه احتمال انتقال یک مرحله فرآیند تصادفی جدید INAR(1)-PK به صورت

$$P(X_t = k | X_{t-1} = \ell) = \sum_{i=1}^{\min(k, \ell)} \binom{\ell}{i} p^i (1-p)^{\ell-i} \frac{\delta^2}{\delta^2 + \delta + 1} \left(\frac{1}{(\delta + 1)^{k-i}} + \frac{k-i+1}{(\delta + 1)^{k-i+2}} \right)$$

حاصل می‌شود. بر اساس روش ویب (۲۰۱۸) میانگین، واریانس و شاخص پراکندگی فرآیند INAR(1)-PK به ترتیب به صورت

$$\mu_X = \frac{2 + \delta + \delta^2}{\delta(\delta^2 + \delta + 1)(1-p)}$$

$$\sigma_X^2 = \frac{(1+p)\delta^5 + (3+2p)\delta^4 + (6+4p)\delta^3 + (8+3p)\delta^2 + (6+2p)\delta + 2}{(1-p^2)\delta^2(\delta^2 + \delta + 1)^2}$$

$$DI_X = \frac{(1+p)\delta^5 + (3+2p)\delta^4 + (6+4p)\delta^3 + (8+3p)\delta^2 + (6+2p)\delta + 2}{(1+p)\delta(\delta^2 + \delta + 1)(\delta^2 + \delta + 2)}$$

بیان می‌شوند. تحت فرضیات مدل، بر اساس روش ارائه شده در [الزید و الاش \(۱۹۸۸\)](#)، میانگین و واریانس شرطی فرآیند به ترتیب به صورت

$$\begin{aligned} E(X_t | X_{t-1}) &= E(p \circ X_{t-1} + \varepsilon_t | X_{t-1}) \\ &= E(p \circ X_{t-1} | X_{t-1}) + E(\varepsilon_t | X_{t-1}) \\ &= pX_{t-1} + \frac{2 + \delta + \delta^2}{\delta(\delta^2 + \delta + 1)} \end{aligned} \quad (2)$$

$$\begin{aligned} V(X_t | X_{t-1}) &= V(p \circ X_{t-1} + \varepsilon_t | X_{t-1}) \\ &= V(p \circ X_{t-1} | X_{t-1}) + V(\varepsilon_t | X_{t-1}) \\ &= p(1-p)X_{t-1} + \frac{2 + 6\delta + 8\delta^2 + 6\delta^3 + 4\delta^4 + \delta^5}{\delta^2(\delta^2 + \delta + 1)^2} \end{aligned} \quad (3)$$

حاصل شده است.

۴ برآوردیابی پارامترهای فرآیند

دو روش متداول در سری‌های زمانی گسسته مقدار، برآوردیابی روش یول-والکر (YW) و ماکسیمم درستنمایی شرطی (CML) برای برآورد پارامترهای مدل INAR-PK(1) به کار گرفته شده‌اند. بر اساس یافته‌های گزارش شده [لیویو و همکاران \(۲۰۱۸\)](#) و [برگنیون و همکاران \(۲۰۱۹\)](#)، روش (CML) از دقت برآوردی بالاتری نسبت به روش (YW) برخوردار است.

۴.۱ روش ماکسیمم درستنمایی شرطی

برآورد پارامترهای مدل INAR(1) اغلب با روش درستنمایی شرطی انجام می‌شود، زیرا توزیع دقیق مقدار اولیه X_1 ممکن است ناشناخته یا پیچیده باشد. این روش با قراردادن شرط بر روی مشاهدات اولیه، محاسبات را ساده‌تر کرده و از نیاز به مدل‌سازی توزیع X_1 اجتناب می‌کند. این رویکرد محاسبات را کاهش داده و برای نمونه‌های بزرگ بدون از دست دادن کارایی آماری است. برای انجام روش ماکسیمم درستنمایی شرطی تحت فرآیند INAR-PK(1) و برابری (۱) تابع درستنمایی شرطی

$$\begin{aligned}\ell(p, \delta) &= \sum_{t=2}^T \ln (P(X_t = k | X_{t-1} = \ell)) \\ &= \sum_{t=2}^T \ln \left[\left(\frac{\delta^2}{\delta^2 + \delta + 1} \right)^{\min(X_t, X_{t-1})} \binom{X_{t-1}}{i} \right. \\ &\quad \times \left. p^i (1-p)^{X_{t-1}-i} \left(\frac{1}{(\delta+1)^{X_t-i}} + \frac{(k-i+1)}{(\delta+1)^{X_t-i+2}} \right) \right] \quad (4)\end{aligned}$$

تعریف شده است. برآوردگر ماکسیمم درستنمایی شرطی (CML) فرآیند INAR(1)-PK، که با $\hat{p}_{CLS}$ و $\hat{\delta}_{CLS}$ نشان داده می‌شود، با ماکسیمم کردن تابع درستنمایی شرطی (۴) حاصل می‌شود. بدلیل ساختار پیچیده تابع درستنمایی، یافتن جواب تحلیلی برای حل آن غیر ممکن است، بنابراین باید روش های بهینه سازی عددی مورد استفاده قرار گیرد.

برآوردیابی پارامترها به کمک تابع optim در نرم افزار R انجام شده است. برای مدیریت دامنه پارامترهای p و δ از الگوریتم L-BFGS-B استفاده می‌شود. ماتریس اطلاع مشاهده شده در پارامترهای برآورد شده با استفاده از خروجی ماتریس هسیان حاصل از بهینه سازی محاسبه می‌شود. خروجی حاصل امکان محاسبه خطاهای استاندارد تقریبی برای برآوردگرها را فراهم می‌نماید. [جو \(۱۹۹۷\)](#) ثابت کرد که تحت شرایط نظم، برآوردگرهای CML دارای سازگاری، توزیع نرمال مجانبی و کارایی هستند.

۴.۲ روش یول-والکر

روش برآوردیابی یول-والکر براساس برابری توابع خودکواریانس و میانگین فرآیند با حالت نمونه‌ای آنها و برآورد پارامترها با حل دستگاه معادلات ایجاد شده حاصل می‌شود. برای فرآیند INAR(1) تابع خود همبستگی نمونه‌ای $\rho_X(h) = p^h$ (ACF) که $h \geq 1$ و $p \in (0, 1)$ عملگر رقیق ساز فرآیند است. بنابراین برآورد یول-والکر p به صورت

$$\hat{p}_{YW} = \frac{\sum_{t=2}^T (X_t - \bar{X})(X_{t-1} - \bar{X})}{\sum_{t=2}^T (X_t - \bar{X})^2},$$

است. برآوردگر یول-والکر $\hat{\delta}_{YW}$ از برابر قرار دادن میانگین نظری فرآیند INAR-PK(1) با مقدار نمونه‌ای آن به دست می‌آید. معادله برآوردیابی با قرار دادن تفاضل این دو کمیت برابر با صفر تشکیل شده و سپس $\hat{\delta}_{YW}$ از حل

معادله

$$((1 - \hat{p}_{YW})\bar{x})\delta^3 + ((1 - \hat{p}_{YW})\bar{x} - 1)\delta^2 + ((1 - \hat{p}_{YW})\bar{x} - 1)\delta - 2 = 0.$$

نسبت به δ حاصل می‌شود، که ریشه آن به صورت عددی به کمک تابع uniroot در نرم افزار R بدست می‌آید.

۵ مطالعه شبیه‌سازی

یک مطالعه شبیه‌سازی به منظور ارزیابی رفتار برآوردهای CML و YW بر مبنای حجم متفاوت نمونه ای برای پارامترهای مدل INAR-PK(1) انجام شده و این کار بر اساس ۱۰۰۰ تکرار برای حجم های نمونه $n = 100, 300, 500, 1000, 2000$ اجرا شده است. نمونه‌های تصادفی با توزیع PK با استفاده از الگوریتم معرفی شده در بخش ۲.۱ تولید شده‌اند. دو مجموعه انتخابی از مقادیر پارامتر برای انجام شبیه‌سازی عبارت است از $p = 0.6, \delta = 0.1$ و $p = 0.5, \delta = 0.6$. عملکرد برآوردهای CML و YW با استفاده از دو معیار اریبی (Bias) و میانگین مربعات خطا (MSE) ارزیابی می‌شود. این معیارها به فرم $Bias = \sum_{j=1}^N \frac{\hat{\gamma}_{i,j} - \gamma_i}{N}$ و $MSE = \sum_{j=1}^N \frac{(\hat{\gamma}_{i,j} - \gamma_i)^2}{N}$ هستند، که $i = 1, 2$ و $\gamma = (p, \delta)$. انتظار می‌رود که اریبی و میانگین مربعات خطا نزدیک به صفر باشند. نتایج شبیه‌سازی برای دو مجموعه پارامتر انتخابی در جدول (۱) خلاصه شده است. مقادیر محاسبه شده برای Bias و MSE که در جدول گزارش شده‌اند، به طور کلی نزدیک به صفر هستند و با افزایش اندازه نمونه کاهش می‌دهند. این موضوع میزان اطمینان هر دو روش برآوردیابی را در حجم‌های محدود تأیید می‌کند. این نتایج نشان می‌دهند که روش CML از لحاظ معیار MSE در تمامی حجم‌های نمونه‌ای، عملکرد بهتری نسبت به روش YW دارد.

۶ کاربرد فرایند پیشنهادی

قابلیت مدل‌بندی فرایند INAR-PK را با استفاده از دو داده واقعی در علوم دامپزشکی با برخی از فرایندهای موجود مقایسه می‌نماییم. از جمله این فرایندها INAR(1)-P, INAR(1)-PL, INAR(1)-PB, INAR(1)-PQX, INAR(1)-PEE و INAR(1)-PMI هستند. برای انتخاب بهترین مدل از معیارهای رایج انتخاب مدل مانند معیار اطلاع آکایکی AIC و معیار اطلاع بیزی BIC استفاده می‌نماییم. کمترین مقادیر این معیارها بهترین مدل انتخابی را تعیین می‌نماید. در ادامه دو مجموعه داده برای مدل بندی استفاده شده است که به ترتیب داده اول عبارت است از تعداد موجودات دارای آسیب پوستی که از تاریخ ژانویه ۲۰۰۳ تا دسامبر ۲۰۰۹ در منطقه نیوزلند به صورت ماهانه گردآوری شده است و داده دوم عبارتست از تعداد موجودات دارای سو تغذیه در همان بازه زمانی در منطقه

جدول ۰۱. نتایج شبیه سازی برای فرآیند INAR-PK(1): بر اساس مجموعه‌های انتخابی پارامترها

n		$p = ۰.۵$		$\delta = ۰.۶$		$p = ۰.۶$		$\delta = ۰.۸$	
		CML	YW	CML	YW	CML	YW	CML	YW
۱۰۰	Bias	-۰/۰۰۸۹	-۰/۰۳۰۷	۰/۰۱۸۰	۰/۰۰۳۹	-۰/۰۰۳۳	-۰/۰۲۴۵	۰/۰۰۲۱	-۰/۰۰۰۷
	MSE	۰/۰۰۱۷	۰/۰۰۷۵	۰/۰۰۴۲	۰/۰۰۷۷	۰/۰۰۰۳	۰/۰۰۵۴	۰/۰۰۰۱	۰/۰۰۰۲
۳۰۰	Bias	-۰/۰۰۴۶	-۰/۰۱۲۳	۰/۰۰۸۳	۰/۰۰۳۱	-۰/۰۰۱۸	-۰/۰۰۸۹	۰/۰۰۱۱	۰/۰۰۰۱
	MSE	۰/۰۰۰۵	۰/۰۰۲۰	۰/۰۰۱۱	۰/۰۰۲۳	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۰	۰/۰۰۰۱
۵۰۰	Bias	-۰/۰۰۳۵	-۰/۰۰۶۷	۰/۰۰۶۳	۰/۰۰۴۱	-۰/۰۰۱۴	-۰/۰۰۵۴	۰/۰۰۰۹	۰/۰۰۰۱
	MSE	۰/۰۰۰۳	۰/۰۰۱۰	۰/۰۰۰۷	۰/۰۰۱۳	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۰	۰/۰۰۰۱
۱۰۰۰	Bias	-۰/۰۰۳۰	-۰/۰۰۵۱	۰/۰۰۳۴	۰/۰۰۲۰	-۰/۰۰۱۰	-۰/۰۰۳۴	۰/۰۰۰۶	۰/۰۰۰۱
	MSE	۰/۰۰۰۱	۰/۰۰۰۶	۰/۰۰۰۳	۰/۰۰۰۷	۰/۰۰۰۰	۰/۰۰۰۱	۰/۰۰۰۰	۰/۰۰۰۰
۲۰۰۰	Bias	-۰/۰۰۱۶	-۰/۰۰۳۹	۰/۰۰۲۳	۰/۰۰۱۱	-۰/۰۰۰۶	-۰/۰۰۲۱	۰/۰۰۰۱	۰/۰۰۰۰
	MSE	۰/۰۰۰۰	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۱	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰	۰/۰۰۰۰

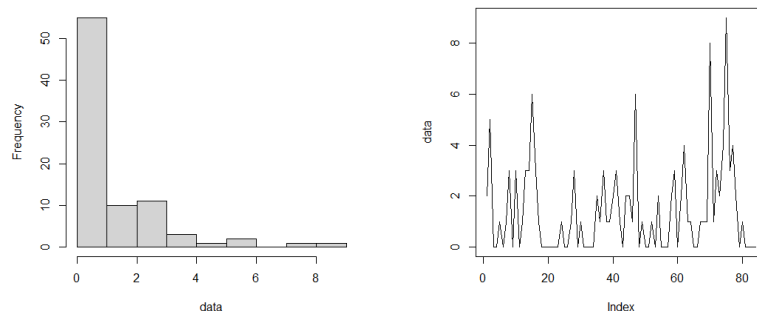
نیوزلند بصورت ماهانه می باشد. این داده ها در دو جدول در مقاله محمدپور و همکاران (۲۰۱۸) ارائه شده است. هر مجموعه داده شامل ۸۴ مشاهده است.

۶.۱ آسیب‌های پوستی

در کاربرد اول، ابتدا بر اساس آماره آزمون ارائه شده توسط شویر و ویب (۲۰۱۴)، بیش پراکندگی داده بررسی می‌شود. میانگین و واریانس مجموعه داده آسیب پوستی به ترتیب عبارتست از $2/4$ و $8/48$ و در نتیجه معیار پراکندگی برابر با $3/53$ است. همچنین آماره آزمون و P - مقدار آزمون مذکور به ترتیب برابر است با $11/37$ و $0/01 <$ است که تاکید بر بیش پراکندگی این مجموعه داده را دارد. شکل ۳ نمودارهای بافت‌نگار، تابع خودهمبستگی ACF و تابع خودهمبستگی جزئی PACF این مجموعه داده را نشان می‌دهد. جهش گام اول نمودار تابع خودهمبستگی جزئی بیان‌کننده مناسب بودن استفاده از مدل $AR(1)$ است. نتایج برازش فرآیندهای $INAR(1)$ به مجموعه داده آفات پوستی در جدول ۲ ارائه شده است. در این جدول به وضوح مشخص است که کمترین مقادیر AIC و BIC مربوط به مدل پیشنهادی $INAR(1)$ -PK است. مقدار میانگین، واریانس و شاخص پراکندگی این مجموعه داده به ترتیب برابر $3/35$ ، $1/42$ و $2/34$ است و همین مقادیر برای مدل برازش شده $INAR(1)$ -PK به ترتیب عبارتست از $1/412$ ، $2/926$ و $2/071$. بر اساس نتایج جدول و نزدیک بودن مقادیر میانگین، واریانس و شاخص پراکندگی برآورد شده مدل پیشنهادی با مقدار تجربی آن، برتری برازش مدل پیشنهادی نسبت به سایر مدل‌ها را نشان می‌دهد. معمولاً برای اطمینان از برازش مناسب مدل به داده‌ها تحلیل باقیمانده انجام می‌شود. در اینجا با استفاده از (۲) و (۳) از باقیمانده پیرسون که به صورت

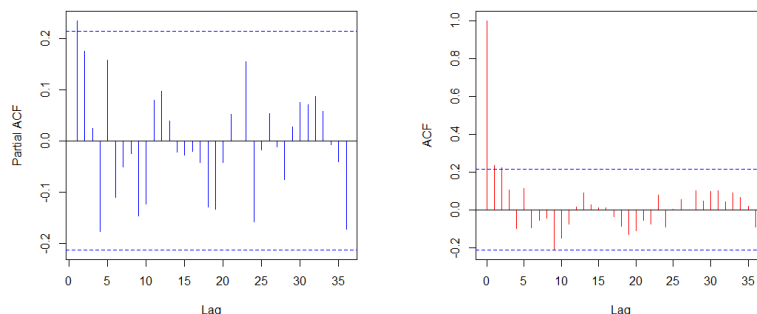
$$r_t = \frac{X_t - E(X_t | X_{t-1})}{(V(X_t | X_{t-1}))^{(1/2)}}$$

استفاده نموده‌ایم. وقتی مدل بطور مناسب برازش شده باشد، انتظار می‌رود که مقادیر میانگین و واریانس این باقیمانده ها به ترتیب نزدیک ۰ و ۱ باشد. برای مدل پیشنهادی برازش شده بر مجموعه داده آسیب پستی این مقادیر برابر ۰ و ۱/۲ است که مقادیر نزدیکی به مقادیر مورد انتظار است.



(ب)

(الف)



(د)

(ج)

شکل ۳. نمودار سری زمانی، بافت نگار، تابع خودهمبستگی (ACF) و تابع خودهمبستگی جزئی (PACF) برای تعداد آفات پستی

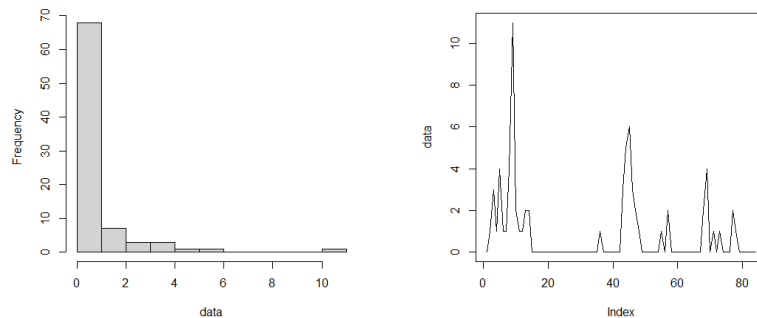
۶.۲ کاهش اشتها

برای بررسی دومین کاربرد مدل پیشنهادی، آزمون بررسی بیش پراکندگی داده را بر اساس آماره آزمون ارائه شده توسط **شویرو ویب (۲۰۱۴)** انجام می‌دهیم. میانگین و واریانس مجموعه داده کاهش اشتها به ترتیب عبارتست از ۰/۸۲ و ۲/۸۹ و در نتیجه معیار پراکندگی برابر با ۳/۵۲ است. همچنین آماره آزمون و P – مقدار آزمون مذکور به ترتیب برابر

جدول ۲. برآورد ماکسیمم درستنمایی شرطی پارامترهای مدل، خطای استاندارد برای تعداد آفات پوستی

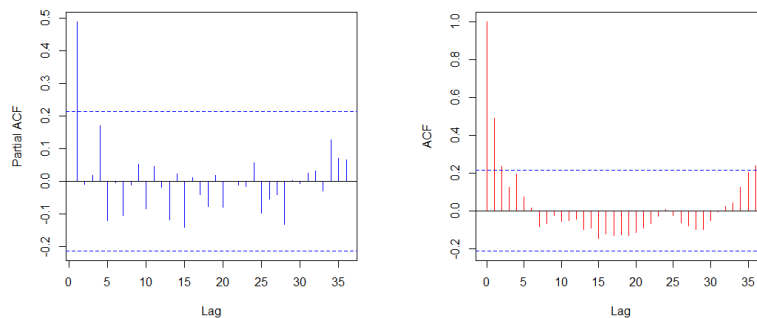
مدل	پارامتر	خطای استاندارد	AIC	BIC
INAR(1)-PK	p	۰/۱۱۵	۲۷۴/۲۷۸	۲۷۹/۱۲۴
	δ	۱/۰۵۲		
INAR(1)-P	p	۰/۱۷۲	۳۰۳/۴۰۷	۳۰۸/۲۴۴
	λ	۱/۱۷۱		
INAR(1)-PB	p	۰/۱۰۰	۲۷۸/۱۸۹	۲۸۳/۰۲۷
	θ	۰/۶۵۳		
INAR(1)-PL	p	۰/۱۱۱	۲۷۴/۷۴۸	۲۷۹/۵۸۶
	θ	۱/۱۶۴		
INAR(1)-PQX	p	۰/۱۲۲	۲۷۵/۷۶۲	۲۸۳/۰۱۹
	α	۶/۲۱۱		
	θ	۱/۰۲۶		
INAR(1)-PEE	p	۰/۱۱۸	۲۷۵/۹۳۳	۲۸۳/۱۸۹
	α	۰/۸۰۹		
	β	۰/۰۱۰		
INAR(1)-PMI	p	۰/۱۲۲	۲۷۵/۷۶۲	۲۸۳/۰۱۹
	α	۰/۱۶۹		
	θ	۱/۰۲۶		

با ۱۲/۹۷ و ۰/۰۰۱ < است که تاکید بر بیش پراکندگی این مجموعه داده را دارد. شکل ۴ نمودارهای یافت‌نگار، تابع خودهمبستگی ACF و تابع خودهمبستگی جزئی PACF این مجموعه داده را نشان می‌دهد. مشابه کاربرد اول، جهش گام اول نمودار تابع خودهمبستگی جزئی بیان‌کننده مناسب بودن استفاده از مدل AR(1) است. نتایج برازش فرایندهای INAR(1) به مجموعه داده کاهش اشتها در جدول ۳ ارائه شده است. در این جدول نیز به وضوح مشخص است که کمترین مقادیر AIC و BIC مربوط به مدل پیشنهادی INAR(1)-PK است. مقدار میانگین، واریانس و شاخص پراکندگی برای مدل برازش شده INAR(1)-PK به ترتیب عبارتست از ۰/۸۲۱، ۱/۲۴۸ و ۰/۵۲۰. نتایج جدول برتری برازش مدل پیشنهادی نسبت به سایر مدل‌ها را نشان می‌دهد. مشابه کاربرد اول، برای اطمینان از برازش مناسب مدل به داده‌ها تحلیل باقیمانده به کمک باقیمانده پیرسون انجام می‌شود. همانطور که در کاربرد اول بیان شد، زمانی که مدل بطور مناسب برازش شده باشد، انتظار می‌رود که مقادیر میانگین و واریانس این باقیمانده‌ها به ترتیب نزدیک ۰ و ۱ باشد. برای مدل پیشنهادی برازش شده بر مجموعه داده کاهش اشتها این مقادیر برابر ۰/۰۳۷ و ۱/۲۵ است که مقادیر نزدیک به مقادیر مورد انتظار است.



(ب)

(الف)



(د)

(ج)

شکل ۴. نمودار سری زمانی، بافت نگار، تابع خودهمبستگی (ACF) و تابع خودهمبستگی جزئی (PACF) برای تعداد کاهش اشتها

۷ بحث و نتیجه‌گیری

در این مقاله، یک مدل جدید خودبازگشتی عدد صحیح نامنفی مرتبه اول با نوفه‌هایی دارای توزیع $PK(1)$ (INAR(1)) برای مدل‌سازی سری‌های زمانی شمارشی با بیش‌پراکندگی پیشنهاد شده است. به منظور برآورد پارامترهای مدل، دو روش مختلف مورد بررسی قرار گرفته و عملکرد آنها از طریق شبیه‌سازی مقایسه شده است. مدل پیشنهادی روی دو مجموعه داده واقعی اعمال شده و نتایج آن با شش مدل $INAR(1)$ دو و سه پارامتری مقایسه گردیده است. یافته‌های تجربی نشان می‌دهد که مدل $INAR(1)$ -PK در مواجهه با بیش‌پراکندگی عملکرد بهتری دارد و بنابراین برای این نوع داده‌ها گزینه مناسب‌تری محسوب می‌شود.

جدول ۳. برآورد ماکسیمم درستنمایی شرطی پارامترهای مدل و خطای استاندارد برای تعداد کاهش اشتها

مدل	پارامتر	برآورد	خطای استاندارد	AIC	BIC
INAR(1)-PK	p	۰/۳۱۹	۰/۰۸۳	۱۹۸/۶۴۰	۲۰۳/۴۷۸
	δ	۲/۰۳۶	۰/۳۴۸		
INAR(1)-P	p	۰/۳۸۴	۰/۰۷۳	۲۲۷/۴۱۷	۲۳۲/۲۵۵
	λ	۰/۵۱۱	۰/۰۸۶		
INAR(1)-PB	p	۰/۳۳۷	۰/۰۸۱	۲۰۶/۹۷۲	۲۱۱/۸۱۰
	θ	۱/۵۱۸	۰/۲۸۷		
INAR(1)-PL	p	۰/۳۲۳	۰/۰۸۲	۱۹۹/۸۰۲	۲۰۴/۶۴۰
	θ	۲/۳۴۸	۰/۴۰۵		
INAR(1)-PQX	p	۰/۳۱۲	۰/۰۸۴	۱۹۹/۱۸۵	۲۰۶/۴۴۱
	α	۹/۴۳۵	۱۳/۱۴۲		
	θ	۲/۰۸۶	۰/۵۸۱		
INAR(1)-PEE	p	۰/۳۱۵	۰/۸۴۰	۲۰۰/۲۰۱	۲۰۷/۴۵۸
	α	۱/۷۶۶	۰/۴۹۲		
	β	۰/۰۱۱	۰/۳۴۱		
INAR(1)-PMI	p	۰/۳۱۲	۰/۰۸۴	۱۹۹/۱۸۵	۲۰۶/۴۴۱
	α	۰/۴۶۱	۰/۸۳۷		
	θ	۲/۰۸۶	۰/۵۸۱		

تقدیر و تشکر

نویسنده از سردبیر، داوران و ویراستار گرامی مجله که با دیدگاه‌های ارزشمندشان به بهبود چشمگیر کیفیت این مقاله کمک کردند، صمیمانه تشکر و قدردانی می‌نماید.

مراجع

نوراللهی، م.ج.، دیری، ع. و بالوئی جامخانه، ع. (۱۴۰۱)، مدل خودبازگشتی گسسته مقدار براساس نوفه‌هایی با توزیع نمایی-وایبل گسسته، مجله علوم آماری، ۱۶، ۴۹۲-۴۶۹.

Alomaira, A.M. and Ahsan-ul-Haq, M. (2025), A New Mixed Poisson Komal Distribution with Application on Radiation, Agricultural and Medical Sciences Data, *Journal of Radiation Research and Applied Sciences*, **Published Online:** <https://doi.org/10.1016/j.jrras.2025.101500>.

- Al-Osh, M. and Alzaid, A. (1987), First-Order Integer-Valued Autoregressive (INAR (1)) Process, *Journal of Time Series Analysis*, **8**(3), 261-275.
- Altun, E. (2020), A New One-Parameter Discrete Distribution with Associated Regression and Integer-Valued Autoregressive Models, *Mathematica Slovaca*, **70**(4), 979-994.
- Altun, E., Bhati, D. and Khan, N.M. (2021), A New Approach to Model the Counts of Earthquakes: INARPQX(1) Process, *SN Applied sciences*, **274**(3), 1-17.
- Alzaid, A. and Al-Osh, M. (1988), First-Order Integer-Valued Autoregressive (INAR (1)) Process: Distributional and Regression Properties, *Statistica Neerlandica*, **42**(1), 53-61.
- Bourguignon, M., Rodrigues, J. and Santos-Neto, M. (2019), Extended Poisson INAR (1) Processes with Equidispersion, Underdispersion and Overdispersion, *Journal of Applied Statistics*, **46**(1), 101-118.
- Maya, R., Chesneau, C., Krishna, A. and Irshad, M.R. (2022), Poisson Extended Exponential Distribution with Associated INAR(1) Process and Applications, *Stats*, **5**, 755-772.
- Maya, R., Irshad, M.R., Chesneau, C., Nitin, S.L. and Shibu, D.S. (2022), On Discrete Poisson-Mirra Distribution: Regression, INAR(1) Process and Applications, *Axioms*, **11**, 193-220.
- McKenzie, E. (1985), Some Simple Models for Discrete Variate Time Series *JAWRA Journal of the American Water Resources Association*, **21**, 645-650.
- McKenzie, E. (1986), Autoregressive Moving-Average Processes with Negative Binomial and Geometric Marginal Distributions, *Advances in Applied Probability*, **18**, 679-705.
- Mohammadpour, M., Bakouch, S.H. and Shirozhan, M. (2018), Poisson-Lindley INAR(1) Model with Applications, *Brazilian Journal of Probability and Statistics*, **32**(2), 262-280.

- Livio, T., Khan, N.M., Bourguignon, M. and Bakouch, H.S. (2018), An INAR(1) Model with Poisson-Lindley Innovations, *Economics Bulletin*, **38**(3), 1505-1513.
- Schweer, S. and WeiB, C.H. (2014), Compound Poisson INAR (1) Processes: Stochastic Properties and Testing for Overdispersion, *Computational Statistics and Data Analysis*, **77**, 267-284.
- Shanker, R. (2023), Komal Distribution with Properties and Application in Survival Analysis, *Biometrics and Biostatistics International Journal*, **12**(2), 40-44.
- Joe, H. (1997), *Multivariate Models and Multivariate Dependence Concepts*, Chapman and Hall, London.
- WeiB, C.H. (2018), *An Introduction to Discrete-Valued Time Series*, Wiley, London.