

مدل خودبازگشتی صحیح مقدار مرتبه اول بر اساس عملگر نازک دو جمله‌ای منفی با نوفه‌های وابسته

معصومه شیروازن، مهرناز محمدپور

گروه آمار، دانشگاه مازندران

تاریخ دریافت: ۱۳۹۷/۰۸/۰۶ تاریخ پذیرش: ۱۳۹۷/۱۲/۲۵

چکیده: در این مقاله، یک مدل جدید خودبازگشتی صحیح مقدار مرتبه اول، بر اساس عملگر نازک دو جمله‌ای منفی معرفی می‌شود که در آن عبارت خطا به طور متوالی به مقدار فرایند در زمان جاری وابسته است. برخی از ویژگی‌های آماری مدل پیشنهادی مورد بحث قرار می‌گیرد و پارامترهای مدل نیز توسط دو روش ماکسیمم درست‌نمایی و یول-واکر برآورد می‌شوند. به کمک شبیه‌سازی، رفتار و کارایی دو روش برآورد مورد مطالعه قرار می‌گیرند. در انتها، برتری مدل معرفی شده در برازش داده‌های واقعی نسبت به سایر مدل‌های صحیح مقدار توسط معیارهای مختلفی بررسی می‌شود.

واژه‌های کلیدی: بیش‌پراکندگی، نازک دو جمله‌ای منفی، مدل خودبازگشتی صحیح مقدار، ماکسیمم درست‌نمایی و یول-واکر.

۱ مقدمه

سری‌های زمانی صحیح مقدار معمولاً دارای دامنه تغییرات متناهی و محدود هستند و نمی‌توان از مدل‌های پیوسته استاندارد برای مدل‌بندی آنها استفاده نمود. از این‌رو در عمل، به مدل‌بندی مناسبی برای برازش و پیش‌بینی این نوع از سری‌های زمانی نیاز است. در مدل‌بندی سری‌های زمانی صحیح مقدار، حفظ طبیعت گسسته داده‌ها از اهمیت بسیاری برخوردار است.

مدل‌بندی سری‌های زمانی صحیح مقدار در سال‌های اخیر رشد و توسعه بسیاری پیدا کرده است و

مدل‌های مختلفی برای مدل‌بندی این‌گونه سری‌های زمانی ارائه شده است. مدل خودبازگشتی صحیح مقدار^۱ (INAR) یکی از مدل‌های پیشنهاد شده است که توسط مکنزی (۱۹۸۵) معرفی شده است. این مدل بر اساس عملگر نازک^۲ دوجمله‌ای ارائه می‌شود که توسط استیوتل و ون‌هارن (۱۹۷۹) مطالعه شده است. مدل خودبازگشتی صحیح مقدار توسط محققان زیادی مورد بررسی قرار گرفته است که در این میان می‌توان به مدل INAR(1) با توزیع حاشیه‌ای پواسن (ال‌اوش و ال‌زید، ۱۹۸۷؛ مکنزی، ۱۹۸۸)، مدل INAR(1) با توزیع حاشیه‌ای هندسی (ال‌زید و ال‌اوش، ۱۹۸۸) و مدل INAR(1) با توزیع حاشیه‌ای دوجمله‌ای منفی (مکنزی، ۱۹۸۶) اشاره کرد. در بین مدل‌هایی که بر اساس تعمیم‌های عملگر نازک دوجمله‌ای معرفی شده‌اند می‌توان به مدل‌های INAR(1) با توزیع حاشیه‌ای دوجمله‌ای منفی (ال‌اوش و علی، ۱۹۹۲؛ ریستیک و همکاران، ۲۰۰۹) و مدل INAR(1) با توزیع حاشیه‌ای پواسن-هندسی (علی و بوذر، ۱۹۹۴) اشاره کرد. هم‌چنین، ژنگ و همکاران (۲۰۰۶ و ۲۰۰۷) و ویب (۲۰۰۸) مدل خودبازگشتی صحیح مقدار با ضریب تصادفی را ارائه نمودند.

تمامی مدل‌های فوق با توزیع‌های حاشیه‌ای گسسته و عملگر نازک‌ساز برای برازش مدل و تحلیل داده‌های شمارشی به کار رفته‌اند. برخی اوقات این مدل‌ها به دلیل ماهیت خاص و برخی ویژگی‌های داده‌های شمارشی هم‌چون بیش‌پراکندگی، چولگی یا کشیدگی، برازش مناسبی را به این داده‌ها ارائه نمی‌دهند. به این دلیل ضرورت دارد که مدل‌های سری زمانی صحیح مقدار جدیدی با توزیع حاشیه‌ای مناسب یا عملگر نازک‌ساز مناسب ارائه شوند. به این منظور، ویب (۲۰۱۵) مدل INAR را بر اساس عملگر نازک دوجمله‌ای اما با خطاهای وابسته با توزیع پواسن در نظر گرفت. بدین صورت که توزیع عبارت خطا در زمان t را تحت یک رابطه تابعی با مقدار سری در زمان $t - 1$ وابسته فرض کرد.

در این مقاله، مدل INAR(1) با عملگر نازک دوجمله‌ای منفی بررسی می‌شود که در آن عبارت خطا به طور متوالی به مقدار گذشته فرایند وابسته است. در بخش ۲، مدل INAR(1) بر اساس عملگر نازک دوجمله‌ای منفی با خطای به‌طور متوالی وابسته معرفی می‌شود. در ادامه، برخی ویژگی‌های آماری این مدل مورد بحث قرار می‌گیرند. در بخش ۳، با استفاده از دو روش ماکسیمم درست‌نمایی و یول-واکر، پارامترهای مدل برآورد می‌شوند. در ادامه با یک مطالعه شبیه‌سازی، عملکرد دو روش فوق برای برآورد پارامترها مورد مطالعه قرار می‌گیرند. سرانجام در بخش ۴، کارایی مدل معرفی شده بر روی یک مجموعه داده واقعی بررسی می‌شود.

¹Integer Valued Autoregressive

²Thinning

۲ معرفی مدل

فرایند خودبازگشتی صحیح مقدار مرتبه اول مانای $\{X_t\}$ را به صورت

$$X_t = \alpha * X_{t-1} + \varepsilon_t, \quad (۱)$$

در نظر بگیرید، که در آن عملگر نازک دوجمله‌ای منفی $\alpha * X = \sum_{j=1}^X Y_j$ به صورت تعریف می‌شود که سری شمارشی $\{Y_j\}$ ، دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع هندسی با پارامتر $\frac{1}{1+\alpha}$ است. همچنین $\varepsilon_t | X_{t-1} \sim Po(aX_{t-1} + b)$ ، $X_{t-1} \sim NB(X_{t-1}, \frac{1}{\alpha+1})$ و مستقل $\alpha * X_{t-1} | X_{t-1} \sim NB(X_{t-1}, \frac{1}{\alpha+1})$ و مستقل از هم، $a + \alpha < 1$ و $a, \alpha \in [0, 1)$ ، $b > 0$. مدل (۱) را مدل خودبازگشتی صحیح مقدار بر اساس نازک دوجمله‌ای منفی با خطای متوالی وابسته^۳ نامیده و به اختصار آن را با نماد NBSDINAR(1) نشان می‌دهند. تابع مولد احتمال فرایند NBSDINAR(1) به صورت

$$\begin{aligned} \phi_{X_t}(s) &= E(s^{X_t}) = E[E(s^{\alpha * X_{t-1} + \varepsilon_t} | X_{t-1})] \\ &= E[E(s^{\alpha * X_{t-1}} | X_{t-1})] E[E(s^{\varepsilon_t} | X_{t-1})] \\ &= \phi_X\left(\frac{1}{1+\alpha-\alpha s}\right) e^{b(s-1)} \phi_X(e^{a(s-1)}), \end{aligned}$$

است. تابع احتمال انتقال یک مرحله‌ای مدل NBSDINAR(1) به صورت

$$\begin{aligned} P_{i0} &= P(X_t = i | X_{t-1} = 0) = \frac{e^{-b} b^i}{i!}, \\ P_{ij} &= P(X_t = i | X_{t-1} = j) = P(\alpha * X_{t-1} + \varepsilon_t = i | X_{t-1} = j) \\ &= \sum_{k=0}^i \binom{k+j-1}{j-1} \frac{\alpha^k}{(\alpha+1)^{k+j}} \left(\frac{e^{-(aj+b)} (aj+b)^{i-k}}{(i-k)!} \right), \end{aligned} \quad (۲)$$

محاسبه می‌شود.

ملاحظه ۱: فرض کنید X یک متغیر تصادفی صحیح مقدار با مقادیر صحیح نامنفی و $\alpha > 0$ باشد، برخی خواص عملگر نازک دوجمله‌ای منفی عبارتند از:

³ Negative Binomial Serially Dependent Integer Valued Autoregressive

$$E(\alpha * X|X) = \alpha X \quad -۱$$

$$Var(\alpha * X|X) = \alpha(1 + \alpha)X \quad -۲$$

$$E(\alpha * X) = \alpha E(X) \quad -۳$$

$$Var(\alpha * X) = \alpha(1 + \alpha)E(X) + \alpha^2 Var(X) \quad -۴$$

با توجه به خواص شرطی $\alpha * X$ ، میانگین و واریانس شرطی مرتبه اول فرایند $\{X_t\}$ بترتیب به صورت

$$E(X_t|X_{t-1}) = E(\alpha * X_{t-1} + \varepsilon_t|X_{t-1}) = (a + \alpha)X_{t-1} + b,$$

$$Var(X_t|X_{t-1}) = Var(\alpha * X_{t-1} + \varepsilon_t|X_{t-1}) = (a + \alpha(1 + \alpha))X_{t-1} + b,$$

به دست می‌آیند. بنابراین میانگین و واریانس فرایند بترتیب عبارتند از

$$\mu_X = E(X_t) = \frac{b}{1 - (a + \alpha)},$$

$$\sigma_X^2 = Var(X_t) = \frac{(a + \alpha(1 + \alpha))\mu_X + b}{1 - (a + \alpha)^2} = \frac{\mu_X(1 + \alpha^2)}{1 - (a + \alpha)^2}.$$

ملاحظه ۲: با توجه به میانگین و واریانس فرایند، شاخص پراکندگی فرایند برابر

$$I_X = \frac{\sigma_X^2}{\mu_X} = \frac{1 + \alpha^2}{1 - (a + \alpha)^2},$$

و بزرگتر از یک است. بنابراین NBSDINAR(1) یک فرایند بیش‌پراکنده است و برای برازش داده‌های شمارشی صحیح مقدار بیش‌پراکنده مناسب است.

ملاحظه ۳: فرایند NBSDINAR(1) یک فرایند مارکوفی مرتبه اول با ساختار خودبازگشتی خطی شرطی مرتبه اول (CLAR(1)) به صورت $E(X_{t+1}|X_t) = \gamma X_t + \delta$ است، که در آن $\gamma = a + \alpha$ و $\delta = b$. بر اساس گرانوالد و همکاران (۲۰۰۰) و با توجه به تابع احتمال انتقال یک مرحله‌ای (۲)، برای هر i, j ، $P_{ij} > 0$ ، از این‌رو $\{X_t\}$ یک فرایند تحویل‌ناپذیر^۴ مانا است.

^۴فرایند $\{X_t\}$ با فضای حالت \mathbb{N} را تحویل‌ناپذیر گویند، هرگاه $P_{ij} > 0$ ، $\forall i, j \in \mathbb{N}$.

گزاره ۱: میانگین و واریانس شرطی مرتبه k -ام فرایند $\{X_t\}$ به ترتیب

$$E(X_{t+k}|X_t) = (a + \alpha)^k X_t + b \frac{1 - (a + \alpha)^k}{1 - (a + \alpha)}, \quad (۳)$$

$$\begin{aligned} Var(X_{t+k}|X_t) &= \frac{N^{k-1}(1 - N^k)}{1 - N} (a + \alpha(1 + \alpha)) X_t + b(b + 1) \frac{1 - N^{2k}}{1 - N^2} \\ &+ \frac{Mb}{(1 - N)^2} \left[1 - N^{k-1} - N \frac{1 - N^{2(k-1)}}{1 + N} \right] \\ &- b^2 \frac{(1 - N^k)^2}{(1 - N)^2}, \end{aligned} \quad (۴)$$

هستند، که در آن‌ها $N = a + \alpha$ و $M = \alpha^2 + N + 2Nb$

برهان: میانگین شرطی به راحتی قابل محاسبه است، اما برای واریانس شرطی، ابتدا میانگین شرطی توان دوم فرایند محاسبه می‌شود. می‌توان نوشت:

$$\begin{aligned} E(X_{t+1}^2|X_t) &= E[(\alpha * X_t + \varepsilon_{t+1})^2 | X_t] = N^2 X_t^2 + M X_t + b(b + 1), \\ E(X_{t+2}^2|X_t) &= E[E(X_{t+2}^2|X_{t+1}) | X_t] \\ &= E(N^2 X_{t+1}^2 + M X_{t+1} + b(b + 1) | X_t) \\ &= N^2 X_t^2 + M(N^2 + N) X_t + (N^2 + 1)b(b + 1) + Mb, \\ E(X_{t+3}^2|X_t) &= E[E(X_{t+3}^2|X_{t+2}) | X_t] \\ &= E[N^2 X_{t+2}^2 + M(N^2 + N) X_{t+2} + (N^2 + 1)b(b + 1) + Mb | X_t] \\ &= N^2 X_t^2 + M(N^2 + N^2 + N^2 + N^2) X_t + (N^2 + N^2 + 1)b(b + 1) \\ &+ M(N^2 + N + 1)b. \end{aligned}$$

بنابراین با استقرا و با توجه به این‌که $N < 1$ ، نتیجه می‌شود

$$\begin{aligned} E(X_{t+k}^2|X_t) &= N^{2k} X_t^2 + M N^{k-1} \frac{1 - N^k}{1 - N} X_t + b(b + 1) \frac{1 - N^{2k}}{1 - N^2} \\ &+ \frac{Mb}{1 - N} \left[\frac{1 - N^{k-1}}{1 - N} - N \frac{1 - N^{2(k-1)}}{1 - N^2} \right]. \end{aligned}$$

۲۲۰ مدل خودبازگشتی صحیح مقدار مرتبه اول

فرع ۱: با توجه به روابط (۳) و (۴) داریم:

$$\lim_{k \rightarrow \infty} E(X_{t+k} | X_t) = \mu_X, \quad \lim_{k \rightarrow \infty} Var(X_{t+k} | X_t) = \sigma_X^2.$$

گزاره ۲: تابع خودکواریانس و خطای فرایند $\{X_t\}$ از مدل NBSDINAR(1) به ترتیب عبارتند از

$$\begin{aligned} \gamma_X(k) &:= Cov(X_t, X_{t-k}) = (a + \alpha)^k \gamma_X(0), \\ \gamma_\varepsilon(j) &:= Cov(\varepsilon_t, \varepsilon_{t-j}) = a(a + \alpha)^{j-1} (a\alpha\sigma_X^2 + \sigma_\varepsilon^2). \end{aligned}$$

برهان: تابع خودکواریانس فرایند NBSDINAR(1) به صورت

$$\begin{aligned} Cov(X_t, X_{t-k}) &= E[Cov(X_t, X_{t-k} | X_{t-k})] + Cov[E(X_t | X_{t-k}), E(X_{t-k} | X_{t-k})] \\ &= Cov[(\alpha + a)^k X_{t-k} + b \frac{1 - (a + \alpha)^k}{1 - (a + \alpha)}, X_{t-k}] \\ &= (\alpha + a)^k Cov(X_{t-k}, X_{t-k}) \\ &= (\alpha + a)^k Var(X_t), \end{aligned}$$

به دست می‌آید. برای محاسبه تابع خودهمبستگی عبارت خطا، محاسبات زیر نیاز است

$$\begin{aligned} Cov(\varepsilon_t, X_{t-1}) &= E[Cov(\varepsilon_t, X_{t-1} | X_{t-1})] + Cov[E(\varepsilon_t | X_{t-1}), E(X_{t-1} | X_{t-1})] \\ &= Cov(aX_{t-1} + b, X_{t-1}) = aVar(X_t), \\ Cov(X_t, \varepsilon_t) &= Cov(\alpha * X_{t-1} + \varepsilon_t, \varepsilon_t) = a\alpha\sigma_X^2 + \sigma_\varepsilon^2, \\ Cov(X_t, \varepsilon_{t-j}) &= Cov[E(X_t | X_{t-1}, \varepsilon_{t-j}), E(\varepsilon_{t-j} | X_{t-1}, \varepsilon_{t-j})] \\ &\quad + E[Cov(X_t, \varepsilon_{t-j} | X_{t-1}, \varepsilon_{t-j})] \\ &= Cov[(a + \alpha)X_{t-1} + b, \varepsilon_{t-j}] = (\alpha + a)Cov(X_{t-1}, \varepsilon_{t-j}) \\ &= (\alpha + a)^j Cov(X_{t-j}, \varepsilon_{t-j}) = (\alpha + a)^j (a\alpha\sigma_X^2 + \sigma_\varepsilon^2). \end{aligned}$$

بنابراین تابع خودکواریانس خطا به صورت

$$\begin{aligned} Cov(\varepsilon_t, \varepsilon_{t-j}) &= Cov[E(\varepsilon_t | X_{t-1}, \varepsilon_{t-1}, \dots, \varepsilon_{t-j}), E(\varepsilon_{t-j} | X_{t-1}, \varepsilon_{t-1}, \dots, \varepsilon_{t-j})] \\ &+ E[Cov(\varepsilon_t, \varepsilon_{t-j} | X_{t-1}, \varepsilon_{t-1}, \dots, \varepsilon_{t-j})] \\ &= Cov(aX_{t-1} + b, \varepsilon_{t-j}) = a(\alpha + a)^{j-1} (a\alpha\sigma_X^2 + \sigma_\varepsilon^2), \end{aligned}$$

محاسبه می‌شود. میانگین و واریانس عبارت خطا در رابطه (۱)، به ترتیب به صورت

$$\begin{aligned} \mu_\varepsilon &= E(\varepsilon_t) = a\mu_X + b = (1 - \alpha)\mu_X, \\ \sigma_\varepsilon^2 &= Var(\varepsilon_t) = \mu_\varepsilon + a^2\sigma_X^2, \end{aligned}$$

به دست می‌آیند. به دلیل بزرگتر بودن واریانس از میانگین، فرایند ε_t بیش‌پراکنده است.

۳ برآوردیابی و شبیه‌سازی

در این بخش پارامترهای مدل NBSDINAR(1) با دو روش ماکسیم درستنمایی و یول-واکر برآورد می‌شوند.

برآورد ماکسیم درستنمایی: بر اساس نمونه تصادفی X_1, \dots, X_n از مدل NBSDINAR(1)، تابع لگ درستنمایی را می‌توان به صورت

$$\ell(\alpha, a, b) = \ln(L(X_1, \dots, X_n; \alpha, a, b)) = \sum_{t=1}^n \ln(P(X_t | X_{t-1})), \quad (5)$$

نوشت، که در آن تابع احتمال انتقال یک مرحله‌ای معرفی شده در (۲) است. برآوردگرهای ماکسیم درستنمایی (ML) پارامترهای α, a, b یعنی $\hat{\alpha}_{ML}, \hat{a}_{ML}, \hat{b}_{ML}$ از ماکسیم ساختن تابع لگ درستنمایی (۵) به دست می‌آید. در این مقاله با استفاده از دستور "nlm" در نرم‌افزار R، برآوردهای فوق محاسبه شده‌اند.

قضیه ۱: برآوردگرهای ماکسیم درستنمایی $\hat{\theta} = (\hat{\alpha}_{ML}, \hat{a}_{ML}, \hat{b}_{ML})^T$ دارای توزیع مجانبی نرمال به

صورت

$$\sqrt{n}(\hat{\theta} - \theta)^T \sim N(\underline{\circ}, \Sigma^{-1}), \quad n \rightarrow \infty,$$

است، که در آن $\underline{\circ} = (\circ, \circ, \circ)^T$ بردار میانگین و

$$\Sigma = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

ماتریس اطلاع فیشر مشاهده شده را نشان می‌دهند و

$$a_{11} = \sum_{t=1}^T \ddot{\ell}_{\alpha\alpha} = \sum_{t=1}^T \frac{\partial^2 \ln P(X_t|X_{t-1})}{\partial \alpha \partial \alpha},$$

$$a_{12} = a_{21} = \sum_{t=1}^T \ddot{\ell}_{a\alpha} = \sum_{t=1}^T \frac{\partial^2 \ln P(X_t|X_{t-1})}{\partial a \partial \alpha},$$

$$a_{13} = a_{31} = \sum_{t=1}^T \ddot{\ell}_{b\alpha} = \sum_{t=1}^T \frac{\partial^2 \ln P(X_t|X_{t-1})}{\partial b \partial \alpha},$$

$$a_{22} = \sum_{t=1}^T \ddot{\ell}_{aa} = \sum_{t=1}^T \frac{\partial^2 \ln P(X_t|X_{t-1})}{\partial a \partial a},$$

$$a_{23} = a_{32} = \sum_{t=1}^T \ddot{\ell}_{ab} = \sum_{t=1}^T \frac{\partial^2 \ln P(X_t|X_{t-1})}{\partial a \partial b},$$

$$a_{33} = \sum_{t=1}^T \ddot{\ell}_{bb} = \sum_{t=1}^T \frac{\partial^2 \ln P(X_t|X_{t-1})}{\partial b \partial b},$$

$$\begin{aligned}\ddot{\ell}_{\alpha\alpha} &= \frac{1}{(1+\alpha)^2} \left[X_{t-1} + \frac{X_{t-1}(X_{t-1}+1)P(X_t-2|X_{t-1}+2)}{P(X_t|X_{t-1})} \right. \\ &\quad \left. - \frac{2X_{t-1}P(X_t-1|X_{t-1}+1)}{P(X_t|X_{t-1})} - \left(\frac{X_{t-1}P(X_t-1|X_{t-1}+1)}{P(X_t|X_{t-1})} \right)^2 \right],\end{aligned}$$

$$\ddot{\ell}_{aa} = X_{t-1}^2 \left[\frac{P(X_t-2|X_{t-1})}{P(X_t|X_{t-1})} - \left(\frac{P(X_t-1|X_{t-1})}{P(X_t|X_{t-1})} \right)^2 \right],$$

$$\ddot{\ell}_{bb} = \frac{P(X_t-2|X_{t-1})}{P(X_t|X_{t-1})} - \left(\frac{P(X_t-1|X_{t-1})}{P(X_t|X_{t-1})} \right)^2,$$

$$\begin{aligned}\ddot{\ell}_{a\alpha} &= \ddot{\ell}_{\alpha a} = \frac{X_{t-1}}{1+\alpha} \left[\frac{(X_{t-1}+1)P(X_t-2|X_{t-1}+1)}{P(X_t|X_{t-1})} - \frac{P(X_t-1|X_{t-1}+1)}{P(X_t|X_{t-1})} \right. \\ &\quad \left. - \frac{X_{t-1}P(X_t-1|X_{t-1}+1)P(X_t-1|X_{t-1})}{P^2(X_t|X_{t-1})} \right],\end{aligned}$$

$$\ddot{\ell}_{b\alpha} = \ddot{\ell}_{\alpha b} = \frac{X_{t-1}}{1+\alpha} \left[\frac{P(X_t-2|X_{t-1}+1)}{P(X_t|X_{t-1})} - \frac{P(X_t-1|X_{t-1}+1)P(X_t-1|X_{t-1})}{P^2(X_t|X_{t-1})} \right],$$

$$\ddot{\ell}_{ab} = \ddot{\ell}_{ba} = X_{t-1} \left[\frac{P(X_t-2|X_{t-1})}{P(X_t|X_{t-1})} - \left(\frac{P(X_t-1|X_{t-1})}{P(X_t|X_{t-1})} \right)^2 \right].$$

برهان: با استفاده از بو و همکاران (۲۰۰۸)، می‌توان نتیجه گرفت توزیع مجانبی $(\hat{\alpha}_{ML}, \hat{a}_{ML}, \hat{b}_{ML})^T$ نرمال است. مشتق‌های مرتبه اول تابع احتمال انتقال یک مرحله‌ای نسبت به پارامترهای a, b و α به

صورت زیر محاسبه می‌شوند:

$$\begin{aligned}
 \frac{\partial P(X_t|X_{t-1})}{\partial \alpha} &= \sum_{k=0}^{X_t} \binom{k+X_{t-1}-1}{X_{t-1}-1} \frac{\alpha^k}{(\alpha+1)^{k+X_{t-1}}} \left[\frac{k-\alpha X_{t-1}}{\alpha(\alpha+1)} \right] P(\varepsilon_t = X_t - k) \\
 &= \frac{1}{\alpha+1} \sum_{k=1}^{X_t} X_{t-1} \binom{k+X_{t-1}-1}{X_{t-1}} \frac{\alpha^{k-1}}{(\alpha+1)^{k+X_{t-1}}} P(\varepsilon_t = X_t - k) \\
 &\quad - \frac{X_{t-1}}{\alpha+1} P(X_t|X_{t-1}) \\
 &= \frac{X_{t-1}}{\alpha+1} [P(X_t - 1|X_{t-1} + 1) - P(X_t|X_{t-1})],
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial P(X_t|X_{t-1})}{\partial a} &= \sum_{k=0}^{X_t} \binom{k+X_{t-1}-1}{X_{t-1}-1} \frac{\alpha^k}{(\alpha+1)^{k+X_{t-1}}} X_{t-1} \left[\frac{X_t - k}{aX_{t-1} + b} - 1 \right] \\
 &\quad \times \frac{e^{-(aX_{t-1}+b)} (aX_{t-1} + b)^{X_t-k}}{(X_t - k)!} \\
 &= \sum_{k=0}^{X_t} \binom{k+X_{t-1}-1}{X_{t-1}-1} \frac{\alpha^k}{(\alpha+1)^{k+X_{t-1}}} X_{t-1} \left[\frac{X_t - k}{aX_{t-1} + b} \right] \\
 &\quad \times \frac{e^{-(aX_{t-1}+b)} (aX_{t-1} + b)^{X_t-k}}{(X_t - k)!} - X_{t-1} P(X_t|X_{t-1}) \\
 &= X_{t-1} [P(X_t - 1|X_{t-1}) - P(X_t|X_{t-1})],
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial P(X_t|X_{t-1})}{\partial b} &= \sum_{k=0}^{X_t} \binom{k+X_{t-1}-1}{X_{t-1}-1} \frac{\alpha^k}{(\alpha+1)^{k+X_{t-1}}} \left[\frac{X_t - k}{aX_{t-1} + b} - 1 \right] \\
 &\quad \times \frac{e^{-(aX_{t-1}+b)} (aX_{t-1} + b)^{X_t-k}}{(X_t - k)!} \\
 &= \sum_{k=0}^{X_t} \binom{k+X_{t-1}-1}{X_{t-1}-1} \frac{\alpha^k}{(\alpha+1)^{k+X_{t-1}}} \left[\frac{X_t - k}{aX_{t-1} + b} \right] \\
 &\quad \times \frac{e^{-(aX_{t-1}+b)} (aX_{t-1} + b)^{X_t-k}}{(X_t - k)!} - P(X_t|X_{t-1}) \\
 &= P(X_t - 1|X_{t-1}) - P(X_t|X_{t-1}),
 \end{aligned}$$

بنابراین

$$\dot{\ell}_{\alpha} = \frac{\partial \ln P(X_t|X_{t-1})}{\partial \alpha} = \left(\frac{X_{t-1}}{1+\alpha}\right) \frac{P(X_t-1|X_{t-1}+1) - P(X_t|X_{t-1})}{P(X_t|X_{t-1})}, \quad (۶)$$

$$\dot{\ell}_a = \frac{\partial \ln P(X_t|X_{t-1})}{\partial a} = (X_{t-1}) \frac{P(X_t-1|X_{t-1}) - P(X_t|X_{t-1})}{P(X_t|X_{t-1})},$$

$$\dot{\ell}_b = \frac{\partial \ln P(X_t|X_{t-1})}{\partial b} = \frac{P(X_t-1|X_{t-1}) - P(X_t|X_{t-1})}{P(X_t|X_{t-1})}.$$

درایه‌های ماتریس اطلاع فیشر مشاهده شده با محاسبه مشتقات مرتبه دوم $\ln P(X_t|X_{t-1})$ به کمک (۶)، به راحتی محاسبه می‌شوند.

برآورد یول-واکر: از آن‌جاکه $E(X_t) = \frac{b}{1-(a+\alpha)}$ و $Corr(X_t, X_{t-1}) = a + \alpha$ و $Var(X_t) = \frac{\mu_X(\alpha^2 + 1)}{1 - (a + \alpha)^2}$ ، با استفاده از میانگین نمونه‌ای (\bar{X}) ، تابع خودهمبستگی نمونه‌ای (C) و واریانس نمونه‌ای (S_X^2) ، برآوردهای یول-واکر (YW) پارامترهای (a, b, α) به صورت

$$\hat{b}_{YW} = \bar{X}(1 - C), \quad \hat{\alpha}_{YW} = \sqrt{\frac{S_X^2(1 - C^2)}{\bar{X}}} - 1, \quad \hat{a}_{YW} = C - \hat{\alpha}_{YW}$$

به دست می‌آیند.

۱.۳ مطالعه شبیه‌سازی

در این بخش، یک مطالعه شبیه‌سازی برای مقایسه عملکرد دو روش ماکسیم درستنمایی و یول-واکر برای برآورد پارامترهای α, a, b انجام می‌شود. برای این منظور، بر اساس الگوریتم ۱ نمونه‌های مختلف با اندازه‌های $n = 50, 100, 500, 1000, 5000$ از فرایند NBSDINAR(1) شبیه‌سازی شده است و بر اساس نمونه‌های شبیه‌سازی شده، برآورد پارامترها و جذر میانگین توان‌های دوم خطای (RMSE) آن‌ها با دو روش محاسبه شده است. برای شبیه‌سازی و محاسبه برآوردها از الگوریتم ۱ استفاده شده است.

الگوریتم ۱:

گام ۱: یک نمونه n تایی از فرایند NBSDINAR(1) شبیه‌سازی کنید.

گام ۲: برآوردهای مختلف ماکسیمم درستنمایی و یول-واکر را محاسبه کنید.
 گام ۳: گام‌های ۱ تا ۲ را ۲۰۰۰ بار تکرار کنید.
 گام ۴: میانگین برآوردها و جذر میانگین توان‌های دوم خطای آن‌ها را در ۲۰۰۰ بار تکرار محاسبه کنید.
 جدول ۱: میانگین و جذر میانگین توان‌های دوم خطا برای برآوردگر پارامترهای مدل (1)NBSDINAR.

$(a, b, \alpha) = (0.4, 2, 0.3)$						
\hat{a}_{YW}	\hat{b}_{YW}	\hat{a}_{YW}	$\hat{\alpha}_{ML}$	\hat{b}_{ML}	$\hat{\alpha}_{ML}$	n
0.3638	2.0631	0.3178	0.2837	2.0439	0.4104	50
0.1369	0.3559	0.1565	0.0827	0.1217	0.0919	RMSE
0.3559	2.0557	0.3311	0.2954	2.0217	0.4082	100
0.1357	0.3506	0.1533	0.0764	0.1145	0.0902	RMSE
0.3436	2.0509	0.3481	0.2974	1.9829	0.3917	500
0.1131	0.3159	0.1316	0.0712	0.0915	0.0836	RMSE
0.2853	2.0434	0.3571	0.3075	2.0142	0.3963	1000
0.1061	0.2898	0.1197	0.0585	0.0880	0.0660	RMSE
0.3170	2.0322	0.3745	0.3029	2.0081	0.4002	5000
0.0821	0.2239	0.0930	0.0493	0.0828	0.0586	RMSE
$(a, b, \alpha) = (0.2, 4, 0.6)$						
0.5508	4.0792	0.1275	0.5832	3.9505	0.1901	50
0.1519	0.4710	0.1082	0.0986	0.1407	0.0789	RMSE
0.5573	4.0731	0.1295	0.5872	3.9582	0.1928	100
0.1498	0.4276	0.0921	0.0931	0.1395	0.0754	RMSE
0.5602	4.0651	0.1401	0.5911	3.9702	0.2052	500
0.1472	0.3931	0.0836	0.0892	0.1284	0.0681	RMSE
0.5694	4.0592	0.1472	0.6029	4.0268	0.1982	1000
0.1273	0.3618	0.0813	0.0732	0.1061	0.0547	RMSE
0.5721	4.0497	0.1582	0.5994	4.0199	0.1997	5000
0.1105	0.3037	0.0779	0.0622	0.1003	0.0485	RMSE

جدول ۱، میانگین و جذر میانگین توان‌های دوم خطای آن‌ها را براساس ۲۰۰۰ بار شبیه‌سازی نشان می‌دهد. برای شبیه‌سازی از فرایند، مقادیر a, b, α به صورت $(0.4, 2, 0.3)$ و $(0.2, 4, 0.6)$ در نظر گرفته

شده‌اند. بر اساس این جدول، برآوردها به مقدار واقعی پارامتر همگرا هستند. هم‌چنین با افزایش حجم نمونه، جذر میانگین توان‌های دوم خطا کاهش می‌یابد. افزون بر این، می‌توان نتیجه گرفت برآوردهای ماکسیمم درست‌نمایی کمترین مقدار جذر میانگین توان‌های دوم خطا را نسبت به برآوردهای یول-واکر دارند و از این‌رو برآوردهای بهتری هستند.

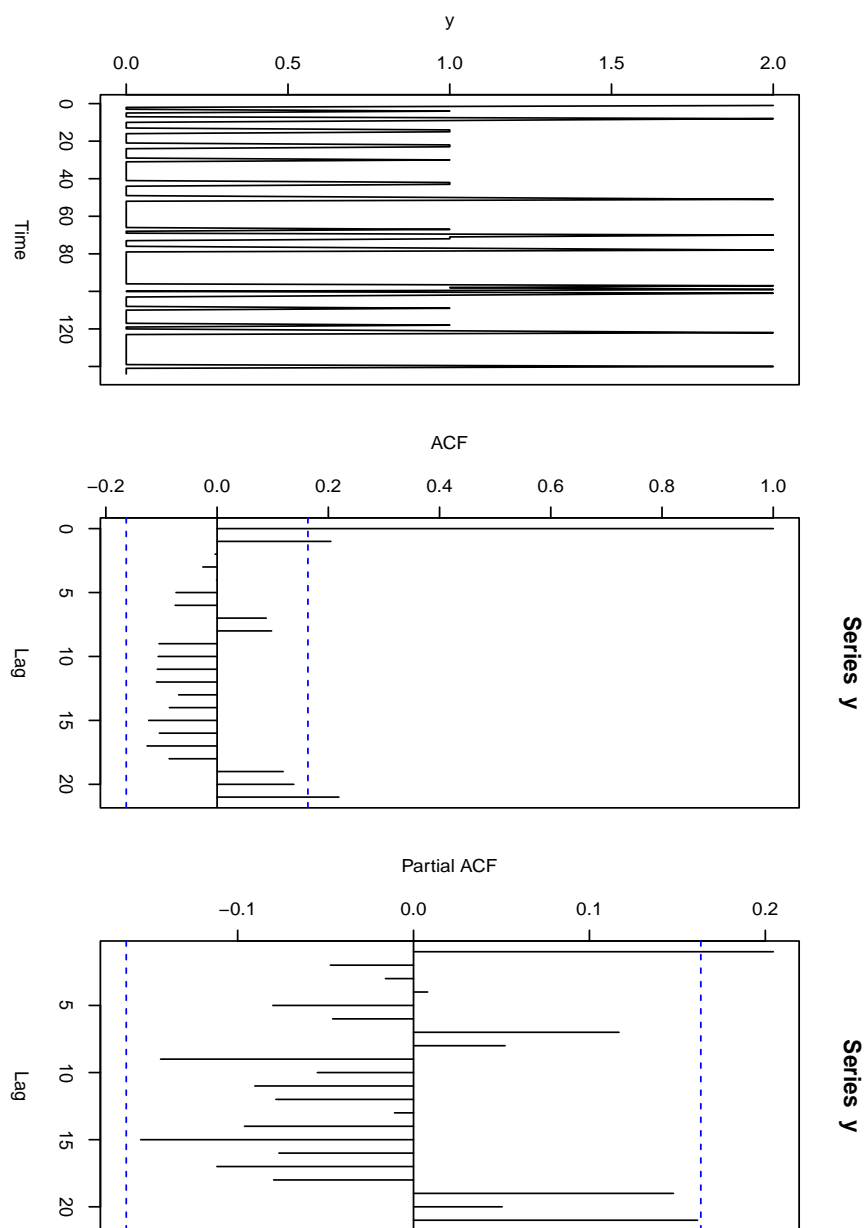
۴ تحلیل داده‌های جرایم جنسی

در این بخش، کارایی فرایند (1) NBSDINAR در مدل‌سازی داده‌های واقعی، توسط یک سری داده‌های شمارشی مورد بررسی قرار می‌گیرد. داده‌های شمارشی در نظر گرفته شده در این جا، تعداد جرایم جنسی ماهانه‌ای است که طی سال‌های ۱۹۹۰ تا ۲۰۰۱ در ۳۱۴-امین ایستگاه پلیس در پترزبورگ در روسیه گزارش شده‌اند. این داده‌ها در سایت (<http://www.forecastingprinciples.com>) ارائه شده‌اند. مسیر نمونه‌ای، تابع خودهمبستگی و خودهمبستگی جزئی داده‌های فوق در شکل ۱ نمایش داده شده است. همان‌طور که ملاحظه می‌شود شرایط مدل خودبازگشتی مرتبه اول برای این داده‌ها برقرار است. هم‌چنین برای بررسی مانایی فرایند، از آزمون دیکی فولر افزایشی^۵ استفاده شده است که p -مقدار این آزمون برابر ۰/۰۱ به دست آمده و نشان‌دهنده مانایی داده‌های فوق است. این داده‌ها دارای میانگین ۰/۲۷۰۸، واریانس ۰/۳۳۸۷ و شاخص پراکندگی $\hat{I}_x = ۱/۲۵$ هستند. بنابراین بیش‌پراکنده‌اند.

در این مثال، مدل (1) NBSDINAR به همراه برخی مدل‌های (1) INAR به داده‌ها برازش داده شده است. مدل‌های (1) INAR در نظر گرفته شده عبارتند از مدل (1) PINAR (ال‌اوش و ال‌زید، ۱۹۸۷)، مدل (1) GINAR (ال‌زید و ال‌اوش، ۱۹۸۸)، مدل (1) NBIINAR (ال‌اوش و علی، ۱۹۹۲)، مدل (1) NBRCINAR (ویب، ۲۰۰۸)، مدل (1) NGINAR (ریستیک و همکاران، ۲۰۰۹)، مدل (1) GPQINAR (ال‌زید و ال‌اوش، ۱۹۹۳)، مدل (1) EPINAR (ویب، ۲۰۱۵).

برای هر مدل، برآوردهای ماکسیمم درست‌نمایی پارامترها، چهار آماره نیکویی برازش AIC, BIC, HQIC و CAIC، جذر میانگین توان‌های دوم خطا محاسبه و نتایج در جدول ۲ گزارش شده است. همان‌طور که ملاحظه می‌شود، کمترین مقدار آماره‌های نیکویی برازش و RMSE برای مدل (1) NBSDINAR نسبت به سایر مدل‌ها (1) INAR حاصل می‌شود. بنابراین، می‌توان نتیجه گرفت که فرایند (1) NBSDINAR برای برازش داده‌های جرایم جنسی بهترین عملکرد را دارد.

⁵ Augmented Dickey-Fuller Test



شکل ۱: مسیر نمونه‌ای، تابع خودهمبستگی و تابع خودهمبستگی جزئی نمونه‌ای داده‌های جرایم جنسی.

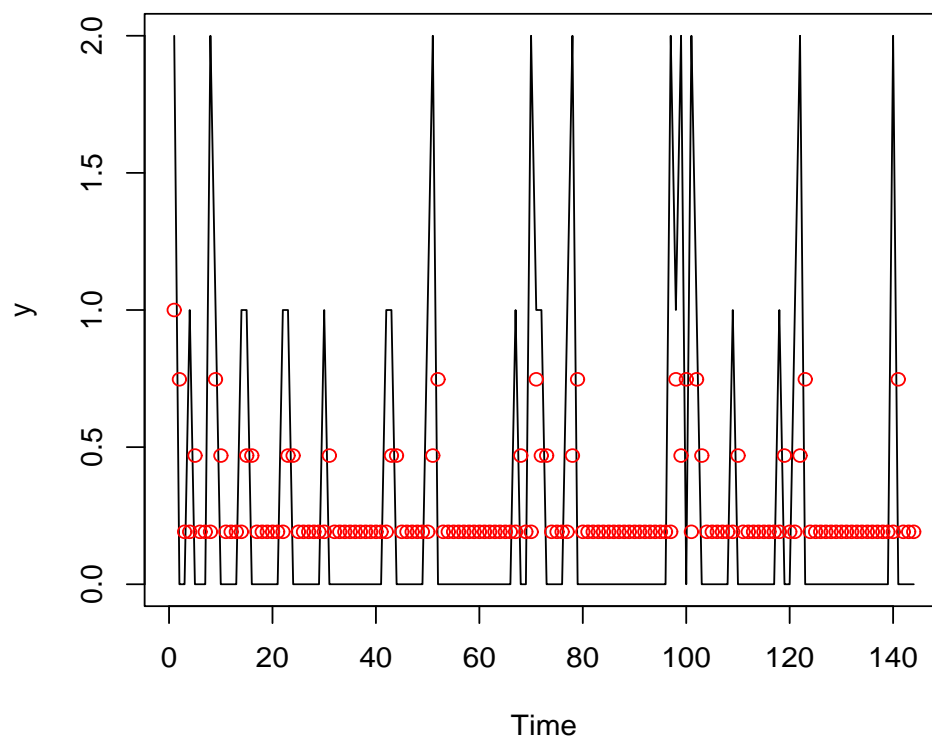
جدول ۲: برآورد پارامترها و برخی آماره‌های نیکویی برازش برای داده‌های مربوط به جرایم جنسی.

مدل	پارامتر	برآورد	AIC	BIC	HQIC	CAIC	RMSE
i.i.d Poisson	λ	۰/۲۷۰۸	۱۹۵/۷۵	۱۹۸/۷۲	۱۹۶/۹۵	۱۹۵/۷۷	-
i.i.d Geom	μ	۰/۲۷۰۸	۱۹۱/۶۱	۱۹۴/۵۷	۱۹۲/۸۱	۱۹۱/۶۳	-
i.i.d NegBin	n	۰/۷۴۶۳	۱۹۳/۴۴	۱۹۹/۳۸	۱۹۵/۸۵	۱۹۳/۵۳	-
	p	۰/۷۳۳۷					
GINAR(1)	p	۰/۲۱۶۲	۱۸۷/۳۹	۱۹۳/۳۳	۱۸۹/۸۱	۱۸۷/۴۸	۰/۵۵۱۱
	α	۰/۲۱۳۲					
PINAR(1)	λ	۰/۲۱۵۱	۱۹۰/۸۷	۱۹۶/۸۱	۱۹۳/۲۸	۱۹۰/۹۶	۰/۵۵۱۱
	α	۰/۲۱۳۶					
NGINAR(1)	p	۰/۲۷۴۳	۱۸۷/۴۲	۱۹۳/۳۶	۱۸۹/۸۳	۱۸۷/۵۱	۰/۵۵۱۹
	α	۰/۲۵۸۱					
NBRCINAR(1)	n	۱/۲۰۳۱	۱۹۰/۶۷	۱۹۹/۵۸	۱۹۴/۲۹	۱۹۰/۸۴	۰/۵۵۱۳
	p	۰/۸۱۰۳					
	ρ	۰/۲۱۹۹					
NBIINAR(1)	n	۰/۶۴۴۸	۱۸۷/۹۱	۱۹۶/۸۲	۱۹۱/۵۳	۱۸۸/۰۸	۰/۵۵۳۴
	p	۳/۳۱۵۶					
	ρ	۰/۲۹۳۰					
GPQINAR(1)	λ	۰/۲۰۲۳	۱۹۱/۶۸	۲۰۰/۵۹	۱۹۵/۳۰	۱۹۱/۸۵	۰/۵۵۱۴
	θ	۰/۰۹۶۹					
	ρ	۰/۱۹۳۸					
EPINAR(1)	ρ	۰/۰۰۰۷	۱۹۲/۹۷	۲۰۱/۸۸	۱۹۶/۵۹	۱۹۳/۱۴	۰/۵۵۱۲
	a	۰/۲۳۴۳					
	b	۰/۲۰۶۴					
NBSDINAR(1)	a	۰/۱۸۵۴	۱۸۳/۷۳	۱۹۲/۶۴	۱۸۷/۳۵	۱۸۳/۹۰	۰/۵۵۱۱
	b	۰/۱۹۱۱					
	α	۰/۰۹۲۶					

از آنجا که امید شرطی یک روش معمول برای پیش‌بینی مقدار میانگین داده‌های سری زمانی است، پیش‌بینی مقدار X در زمان $t + 1$ بر اساس مشاهدات تا زمان t برابر میانگین شرطی یک گام بعد به صورت

$$\hat{X}_{t+1} = E(X_{t+1} | X_t) = (a + \alpha) X_t + b,$$

۲۳۰ مدل خودبازگشتی صحیح مقدار مرتبه اول



شکل ۲: داده‌های جرایم جنسی و مقادیر پیش‌بینی شده توسط مدل NBSDINAR(1).

است، که در عمل، پارامترهای (a, b, α) با مقادیر برآورد ماکسیمم درستنمایی متناظرشان یعنی $(\hat{a}, \hat{b}, \hat{\alpha})$ جایگزین می‌شوند. بنابراین رابطه

$$\hat{X}_{t+1} = (\hat{a} + \hat{\alpha})X_t + \hat{b}, \quad t = 1, \dots, n \quad (7)$$

برقرار است. یادآور می‌شود که $\hat{X}_1 = \hat{E}(X_1) = \frac{\hat{b}}{1 - (\hat{a} + \hat{\alpha})}$ و n تعداد مشاهدات است. برای بررسی کفایت و توانایی پیش‌بینی مدل NBSDINAR(1)، سری زمانی واقعی و مقادیر پیش‌بینی شده آن بر اساس رابطه (۷) در شکل ۲ ارائه شده است.

۵ بحث و نتیجه‌گیری

با ارائه یک فرایند خودبازگشتی صحیح مقدار جدید به برازش آن به داده‌های بیش‌پراکنده پرداخته شده است. پارامترهای فرایند را با دو روش مختلف برآورد و کارایی فرایند در برازش داده‌ها توسط یک سری داده واقعی بیش‌پراکنده بررسی شده است. نتایج برازش مدل به داده‌های واقعی نشان داد که مدل جدید دارای کمترین مقدار RMSE و کمترین مقدار آماره‌های نیکویی برازش است.

تقدیر و تشکر

نویسندگان مقاله ضمن تشکر از سردبیر و ویراستار محترم مجله، از پیشنهادها و نظرات ارزشمند داوران محترم مقاله که باعث بهبود کیفیت مقاله گردید، کمال تشکر و قدردانی را دارند.

مراجع

- Al-Osh, M. A. and Aly, E-E. A. A. (1992), First Order Autoregressive Time Series with Negative Binomial and Geometric Marginals, *Communications in Statistics: Theory and Methods*, **21**, 2483-2492.
- Al-Osh, M. A. and Alzaid, A. A. (1987), First-Order Integer-Valued Autoregressive (INAR(1)) Process, *Journal of Time Series Analysis*, **8**, 261-275.
- Aly, E-E. A. A. and Bouzar, N. (1994), Explicit Stationary Distributions for Some Galton-Watson Processes with Immigration, *Communications in Statistics. Stochastic Models*, **10**, 499-517.
- Alzaid, A. A. and Al-Osh, M. A. (1988), First-Order Integer-Valued Autoregressive (INAR(1)) Process: Distributional and Regression Properties, *Statistica Neerlandica*, **42**, 53-61.
- Alzaid, A. A. and Al-Osh, M. A. (1993), Some Autoregressive Moving Average Processes with Generalized Poisson Marginal Distributions, *Annals of the Institute of Statistical Mathematics*, **45**, 223-232.
- Bu, R., McCabe, B. and Hadri, K. (2008), Maximum Likelihood Estimation of Higher Order Integer Valued Autoregressive Processes, *Journal of Time Series Analysis*, **29**, 973-994.

- Grunwald, G. K., Hyndman, R. J., Tedesco, L. and Tweedie, R. L. (2000), Non-Gaussian Conditional Linear AR(1) Models, *Australian and New Zealand Journal of Statistics*, **42**, 479–495.
- McKenzie, E. (1985), Some Simple Models for Discrete Variate Time Series, *American Water Resources Association*, **21**, 645-650.
- McKenzie, E. (1986), Autoregressive Moving-Average Processes with Negative Binomial and Geometric Distributions, *Advances in Applied Probability*, **18**, 679-705.
- McKenzie, E. (1988), Some ARMA Models for Dependent Sequences of Poisson Counts, *Advances in Applied Probability*, **20**, 822-835.
- Ristić, M. M., Bakouch, H. S. Nastić, A. S. (2009), A New Geometric First-Order Integer-Valued Autoregressive (NGINAR(1)) Process, *Journal of Statistical Planning and Inference*, **139**, 2218-2226.
- Steutel, F. W. and Van Harn, K. (1979), Discrete Analogues of Self-Decomposability and Stability, *Annals of Probability*, **7**, 893-899.
- Weiβ, C. H. (2008), Thining Operations for Modeling Time Series of Count-A Survey, *Advances in Statistical Analysis*, **92**, 319-341.
- Weiβ, C. H. (2015), A Poisson INAR(1) Model with Serially Dependent Innovations, *Metrika*, **78**, 829-851.
- Zheng, H., Basawa, I. V. and Datta, S. (2006), Inference for Pth-Order Random Coefficient Integer-Valued Autoregressive Processes, *Journal of Time Series Analysis*, **27**, 411-440.
- Zheng, H., Basawa, I. V. and Datta, S. (2007), First-Order Random Coefficient Integer-Valued Autoregressive Processes, *Journal of Statistical Planning and Inference*, **137**, 212-229.