



توزیع گسسته متوازن وایبل و مدل گسسته مقدار خودبازگشتی متناظر آن: ویژگی‌ها، برآورد و آنالیز داده‌های شمارشی فوت ناشی از COVID-19

سیده محبوبه حسینی بالادزایی، عین‌الله دیری^{*}، عزت‌الله بالوئی جامخانه

گروه آمار، واحد قائمشهر، دانشگاه آزاد اسلامی، قائمشهر، ایران

دبیر مسئول: حمزه ترابی

تاریخ پذیرش: ۱۴۰۱/۵/۲۲

تاریخ دریافت: ۱۴۰۱/۱/۱۹

چکیده: در این مقاله به معرفی توزیع گسسته جدید وایبل بر اساس روش توزیع گسسته متوازن که حافظ گشتاورهای جزئی بین دو نسخه گسسته و پیوسته توزیع‌ها است، می‌پردازیم. برخی ویژگی‌های آماری توزیع جدید و انواع پراکندگی توزیع مدنظر بر اساس انتخاب‌های گوناگونی از پارامترها ارائه می‌گردد. علاوه بر معرفی نسخه جدید گسسته متوازن وایبل، مدل خودبازگشتی گسسته مقدار با نطفه‌هایی از توزیع گسسته مدنظر را ارائه می‌نماییم و به بررسی روش‌های مختلف برآورد پارامترهای مدل می‌پردازیم. با استفاده از داده‌های فوت ناشی از COVID-19 در کشورهای کوبا، مالاوی و ازبکستان، به بررسی کارایی فرایند جدید در برازش داده‌های واقعی در مقایسه با برخی مدل‌های خودبازگشتی گسسته مقدار کلاسیک می‌پردازیم. در پایان پیش‌بینی فرایند با استفاده از دو رویکرد کلاسیک و بوت‌استرپ غربال، بر اساس داده‌های واقعی نیز بررسی می‌شود.

واژه‌های کلیدی: گسسته متوازن وایبل، حافظ گشتاور جزئی، مدل خودبازگشتی گسسته مقدار، پیش‌بینی فرایند، COVID-19.

رده‌بندی ریاضی: 60E05, 62M05, 62P10

مقدمه ۱

در سال‌های اخیر، معرفی توزیع گسسته و مدل‌سازی رویدادهای شمارشی به دلیل کاربرد وسیع این مدل‌ها در شاخه‌های مختلف علوم مورد توجه بسیاری از محققین واقع شده است. متغیرهای شمارشی در پدیده‌های واقعی مانند تعداد حوادث منجر به خسارت در بیمه، تعداد شرکت‌های زیان‌ده در اقتصاد، تعداد زلزله‌ها در زمین‌شناسی، تعداد سلول‌های سرطانی در پزشکی و غیره مشاهده می‌شود. مدل‌های احتمال پیوسته نمی‌توانند برازش مناسبی را برای برخی از مجموعه‌های داده گسسته ارائه دهند. بنابراین، توزیع‌های گسسته به منظور مدل‌سازی داده‌های

^{*}نویسنده مسئول مقاله

رایانامه: (S. M. Hoseini Baladezaei) Mhoseiny1199@gmail.com (E. Deiri) e.deiri53@Qaemiau.ac.ir,
(E. Baloui Jamkhaneh) e_baloui2008@yahoo.com

شمارشی، مانند دوجمله‌ای، پواسون، هندسی و غیره جایگزین می‌شوند. متأسفانه، توزیع‌های گسسته کلاسیک نیز قادر به مدل‌سازی انواع داده‌ها نمی‌باشند. به عنوان مثال، توزیع دو جمله‌ای دارای دامنه محدود می‌باشد، توزیع پواسون را نمی‌توان بر روی داده‌های بیش‌پراکنده یا کم‌پراکنده اعمال کرد و توزیع هندسی نیز نمی‌تواند مجموعه داده‌ها با تابع نرخ خطر غیر ثابت را پوشش دهد. از این رو، برخی از توزیع‌های گسسته جدید، بر اساس نسخه کلاسیک توزیع‌ها، از طریق روش‌های گسسته‌سازی مختلف ارائه می‌شوند.

یکی از روش‌های پرکاربرد در معرفی توزیع‌های احتمال گسسته، به کارگیری روش گسسته‌سازی تابع بقا است. در این روش، تابع بقا در بین نسخه‌های پیوسته و گسسته ثابت باقی می‌ماند. به عنوان نمونه از توزیع‌های گسسته ارائه شده بر اساس روش گسسته‌سازی بقا، می‌توان به توزیع گسسته نرمال [۲۲]، گسسته رایلی [۲۳]، گسسته گاما [۱۰] و گسسته کومارساتومی معکوس [۱۵] اشاره نمود.

هاگمارک [۱۴] روش گسسته‌سازی حافظ میانگین را معرفی کرد که تابع میانگین بین نسخه پیوسته و گسسته ثابت باقی می‌ماند. اخیراً توپسده و همکاران [۲۶] یک رویکرد گسسته‌سازی جدید مبتنی بر حفظ گشتاور جزئی به نام روش گسسته متوازن (BD) با نمایش تصادفی ساده و انواع مختلف پراکندگی معرفی کرد. آن‌ها توزیع گسسته متوازن گاما را به عنوان یک مدل فرعی معرفی کردند. از ویژگی‌های کاربردی توزیع گسسته بر اساس روش BD، شمول توزیع برای انواع مختلف پراکندگی می‌باشد.

نظر به شیوع بیماری همه‌گیر COVID-19، تحقیقات گسترده‌ای توسط دانشمندان مختلف پزشکی، آماری و اجتماعی پیرامون بررسی تعداد مبتلایان، مرگ و میرها یا بستری در بخش مراقبت‌های ویژه در بیمارستان ناشی از COVID-19 انجام شده است. در حالی که محققین علوم پزشکی به طور خاص بر توسعه واکسن‌ها تمرکز کرده‌اند، از جبهه‌های دیگر آماردانان و محققین علوم داده بر ارائه تحلیل‌ها و پیش‌بینی‌های کارآمد کوشیدند. چندین مدل از داده‌های COVID-19 در مقالات اخیر ارائه شده است. به عنوان مثال، می‌توان به مدل‌های رشد غیرخطی برای COVID-19 در عراق [۱]، مدل اپیدمی قطعی COVID-19 در جامعه همگن آمیخته [۱۳]، مدل COVID-19 با بسط توزیع مماس لگ لجستیک در سومالی [۱۹]، مدل‌سازی سری‌های زمانی خودبازگشتی و پیش‌بینی موارد COVID-19 بر اساس توزیع‌های نرمال مقیاس آمیخته دوتکه‌ای [۱۷] اشاره نمود. اخیراً پوررضا و همکاران [۲۱] خانواده‌ای از توزیع تولید شده گاما را با توجه به توزیع گاما-وایبل ارائه کردند و کاربرد توزیع جدید در برازش مجموعه داده COVID-19 در سارلند را نشان دادند.

از آنجایی که متغیرهای COVID-19 بر اساس زمان‌های متوالی جمع‌آوری می‌شوند، لذا کاوش رویکردهای سری زمانی از اهمیت بسزایی برخوردار است. یکی از مدل‌های کاربردی برای سری‌های زمانی شمارشی، مدل خودبازگشتی گسسته‌مقدار (INAR) است که در شاخه‌های مختلف علوم به کار می‌رود. فرایند خودبازگشتی گسسته‌مقدار مرتبه اول (INAR(1)) بر اساس عملگر رقیق‌ساز معرفی شده است که در سال‌های اخیر در جنبه‌های مختلف گسترش یافته است. مطالعات جامعی پیرامون ارائه عملگرهای رقیق‌ساز مختلف و به موازات آن مدل‌های INAR جدید و همچنین توزیع‌های مختلف حاشیه‌ای فرایند و نوفه‌های آن، انجام شده است. به عنوان مدل‌های INAR(1) پیشگام، می‌توان به مدل‌های INAR(1) با توزیع حاشیه‌ای پواسون [۲]، با توزیع‌های حاشیه‌ای دوجمله‌ای منفی و هندسی [۵]، توزیع حاشیه‌ای پواسون تعمیم‌یافته [۶]، مدل‌های INAR(1) با نوفه‌هایی از توزیع بل [۹]، نوفه‌های از توزیع پواسن-نمایی تبدیل‌یافته [۴] و نوفه‌های با توزیع سارمانوف [۷] اشاره نمود. برخی مقالات جدید مدل‌سازی INAR(1) برای مجموعه داده‌های COVID-19 توسط شما و همکاران [۲۴]، تریاکا و تریاکا [۲۷] و چاتوپادیای و همکاران [۱۱] بررسی شده است.

هدف اصلی این مقاله معطوف به معرفی یک توزیع گسسته جدید با ویژگی‌های برجسته و انعطاف‌پذیر است. از توزیع معروف و پرکاربرد وایبل به عنوان توزیع پایه در روش گسسته‌سازی متوازن استفاده و نسخه گسسته‌ای از توزیع احتمال وایبل به نام توزیع گسسته متوازن وایبل را معرفی می‌نماییم.

توزیع گسسته متوازن وایبل، ویژگی‌های منعطفی را نشان می‌دهد که در مقایسه با بسیاری از توزیع‌های گسسته برتری دارد. توزیع گسسته متوازن وایبل، بر اساس انتخاب مختلف پارامترهای توزیع، انواع مختلفی از پراکندگی شامل بیش‌پراکندگی، کم‌پراکندگی و همسان را شامل می‌شود. در حالی که بسیاری از توزیع‌های کلاسیک، تنها یک حالت از پراکندگی را در برمی‌گیرند و در مدل‌سازی انواع مختلف داده‌های واقعی محدود می‌باشند. از سوی دیگر، توزیع گسسته متوازن وایبل تابع جرم احتمال و نرخ خطر منعطفی را نشان می‌دهد و این توزیع در مدل‌سازی داده‌هایی با نرخ خطر یکنوا (صعودی یا نزولی) و تک‌مدی، کارایی دارد. از این رو، معرفی و بررسی فرایند‌های INAR(1) با استفاده از نوفه‌هایی از توزیع گسسته متوازن وایبل، از جایگاه ویژه‌ای برخوردار می‌باشد. لذا، مدل INAR(1) منعطفی را بر اساس نوآوری‌های گسسته متوازن وایبل ارائه می‌کنیم. همچنین، کارایی مدل پیشنهادی توسط برخی از مجموعه داده‌های COVID-19 نشان داده می‌شود. فرایند INAR(1) مدنظر به عنوان یک رقیب اصلی برای مدل‌های شمارشی کلاسیک محسوب می‌شود، زیرا معیارهای برازش بسیار بهتری را در مدل‌سازی مجموعه داده‌های COVID-19 ارائه می‌کند.

ساختار کلی مقاله به شرح زیر می‌باشد. در بخش ۲، توزیع گسسته متوازن وایبل و فرایند خودبازگشتی گسسته‌مقدار متناظر آن معرفی شده است که از ویژگی‌های منعطف و کاربردی برخوردار است. برآورد پارامترهای فرایند مدنظر با استفاده از سه رویکرد حداکثر درست‌نمایی شرطی، کمترین مربعات شرطی و بول‌واکر در بخش ۳ ارائه شده است. در بخش ۴، با استفاده از نمونه‌های شبیه‌سازی شده، روش‌های مختلف برآورد محاسبه و کارایی آن‌ها بررسی شده است. در پایان در بخش ۵، به بررسی قابلیت فرایند جدید در برازش داده‌های واقعی فوت ناشی از COVID-19، در مقایسه با سایر مدل‌های INAR(1) کلاسیک پرداخته شده است و همچنین کفایت مدل با استفاده از آنالیز باقیمانده‌های پیرسن نشان داده و با دو رویکرد کلاسیک و بوت‌استرپ، پیش‌بینی‌های فرایند ارائه شده است.

۲ توزیع گسسته متوازن وایبل و فرایند خودبازگشتی گسسته مقدار جدید

روش گسسته سازی متوازن توسط توپسده و همکاران [۲۶] معرفی شده است. در این بخش با تمرکز به روی روش گسسته سازی متوازن و بر اساس توزیع پایه وایبل، توزیع گسسته جدیدی و متعاقبا فرایند خودبازگشتی گسسته مقدار جدیدی را با نوفه های از توزیع گسسته متوازن وایبل (BDW) معرفی می نماییم.

تابع چگالی احتمال (PDF)، تابع توزیع تجمعی (CDF) و تابع نرخ خطر (HRF) وایبل به صورت زیر ارائه می گردد

$$g_Y(y, \beta, \theta) = \frac{\beta}{\theta} \left(\frac{y}{\theta}\right)^{\beta-1} e^{-\left(\frac{y}{\theta}\right)^\beta}, \quad y \geq 0, \beta, \theta > 0,$$

$$G_Y(y, \beta) = 1 - e^{-\left(\frac{y}{\theta}\right)^\beta},$$

$$H(y, \alpha, \theta) = \frac{g(y, \beta)}{1 - G(y, \beta)} = \frac{\beta}{\theta} \left(\frac{y}{\theta}\right)^{\beta-1}.$$

بنابراین، تابع HRF توزیع وایبل نسبت به y صعودی می باشد.

امید ریاضی و واریانس توزیع وایبل به ترتیب به صورت زیر نشان داده می شود

$$E(Y) = \theta \Gamma\left(1 + \frac{1}{\beta}\right),$$

$$Var(Y) = \theta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right].$$

که در آن $\Gamma(t)$ تابع گاما و به صورت $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx = (t-1)!$ تعریف می شود. در این بخش، برخی مفاهیم اولیه روش گسسته سازی متوازن را ارائه می نماییم [۲۶].

۱.۲ مفاهیم اساسی روش گسسته سازی متوازن

روش گسسته سازی، روشی جامع است که حتی در اندازه گیری متغیر پیوسته نیز کاربرد دارد. در واقع، هیچ مشاهده ای یک پیوستار کامل را پوشش نمی دهد، زیرا در هر صورت شامل تعداد متناهی رقم اعشار می باشد. از این رو، می توان ادعا نمود که تقریباً همه مقادیر مشاهده شده در واقع گسسته اند زیرا تنها با تعداد محدود اعشار اندازه گیری می شوند و نمی توانند همه نقاط یک پیوستار را شامل شوند.

نحوه عملکرد روش گسسته سازی متوازن را با استفاده از مثالی شرح می دهیم. فرض کنید متغیر تصادفی Y به صورت قطر یک لوله تعریف می شود که با استفاده از یک دستگاه با دقت میلی متر (mm) اندازه گیری شده است. با توجه به ماهیت پیوسته متغیر تصادفی Y ، احتمال مشاهده $Y = y$ (mm) برابر با صفر است. هنگامی که مقدار واقعی قطر یک لوله y در واقع بین دو درجه بندی متوالی v و $v+1$ قرار می گیرد، مقدار $z = v$ mm یا $z = v+1$ mm گزارش می شود، بنابراین فقط یک نسخه گسسته Z از Y مشاهده می شود. واضح است که هرچه y به v نزدیکتر باشد، احتمال گزارش $z = v$ بیشتر است و برعکس هر چه y به $v+1$ نزدیکتر باشد، احتمال گزارش $z = v+1$ بیشتر است. گسسته سازی متوازن از این فرض حاصل می شود که با توجه به $v \leq y < v+1$ ، احتمال گزارش $z = v+1$ دقیقاً $z = v - v$ است.

تعریف ۱.۲ (روش گسسته سازی متوازن). متغیر تصادفی پیوسته Y را از توزیع پیوسته دلخواه $CD(\delta)$ و متغیر تصادفی گسسته U به شرط Y ، را به صورت $U|Y = y \sim Ber(r)$ در نظر بگیرید، آن گاه متغیر تصادفی گسسته Z ، به صورت نسخه گسسته متوازن از توزیع پیوسته $CD(\delta)$ تعریف می شود، اگر دارای نمایش تصادفی زیر باشد

$$Z|U = u, Y = y \stackrel{d}{=} v + u,$$

که در آن $Ber(\cdot)$ ، نماد توزیع برنولی، $[y]$ و $v = [y]$ و $r = y - v$ است.

متغیر تصادفی پیوسته Y با توابع PDF و CDF به ترتیب $g(\cdot, \delta)$ و $G(\cdot, \delta)$ را با بردار پارامترهای مدل δ در نظر بگیرید. تابع جرم احتمال (PMF) و CDF توزیع BD به صورت زیر تعریف می شود

$$f(z, \delta) = (z-1)G(z-1, \delta) - 2zG(z, \delta) + (z+1)G(z+1, \delta) + E_Y(1, z-1|\delta) - E_Y(1, z|\delta), \quad z \in \mathbb{Z}, \quad (1.2)$$

$$F(z, \delta) = G(z, \delta) + (z+1)H(z, \delta) - E_Y(1, z|\delta), \quad (2.2)$$

که در آن $H(z, \delta) = G(z+1, \delta) - G(z, \delta)$ و $E_Y(n, z|\delta)$ بر n -امین گشتاور جزئی به صورت زیر دلالت دارد

$$E_Y(n, z|\delta) = \int_z^{z+1} y^n g(y, \delta) dy.$$

در ادامه، توزیع پیوسته وایبل را به عنوان مولفه پیوسته CD در روش گسسته‌سازی متوازن در نظر می‌گیریم و توزیع گسسته جدیدی را بر اساس توزیع پایه وایبل معرفی می‌کنیم.

۲.۲ توزیع گسسته متوازن وایبل

با توجه به روابط (۱.۲) و (۲.۲)، توزیع گسسته متوازن وایبل (BDW) را با در نظر گرفتن توزیع پایه وایبل معرفی می‌کنیم. گشتاور جزئی مرتبه اول توزیع وایبل به صورت زیر محاسبه می‌شود

$$\begin{aligned} E_Y(1, z|\beta, \theta) &= \int_z^{z+1} y g(y, \beta, \theta) dy \\ &= \theta \left[\Gamma\left(1 + \frac{1}{\beta}, \left(\frac{z}{\theta}\right)^\beta\right) - \Gamma\left(1 + \frac{1}{\beta}, \left(\frac{z+1}{\theta}\right)^\beta\right) \right]. \end{aligned} \quad (3.2)$$

که در آن $\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt$ تابع گامای ناکامل به صورت $\Gamma(s, x)$ است. بنابراین، با استفاده از (۱.۲)، (۲.۲) و گشتاور جزئی مرتبه اول (۳.۲)، توابع PMF و CDF توزیع BDW به صورت زیر ارائه می‌گردد. برای $z = 0$

$$f_z(0, \beta, \theta) = 1 - e^{-\left(\frac{1}{\theta}\right)^\beta} - \theta \left[\Gamma\left(1 + \frac{1}{\beta}\right) - \Gamma\left(1 + \frac{1}{\beta}, \left(\frac{1}{\theta}\right)^\beta\right) \right]$$

و برای $z = 1, 2, \dots$ داریم

$$\begin{aligned} f_z(z, \beta, \theta) &= 2ze^{-\left(\frac{z}{\theta}\right)^\beta} - (z-1)e^{-\left(\frac{z-1}{\theta}\right)^\beta} - (z+1)e^{-\left(\frac{z+1}{\theta}\right)^\beta}, \quad z = 1, 2, \dots \\ &+ \theta \left[\Gamma\left(1 + \frac{1}{\beta}, \left(\frac{z-1}{\theta}\right)^\beta\right) + \Gamma\left(1 + \frac{1}{\beta}, \left(\frac{z+1}{\theta}\right)^\beta\right) - 2\Gamma\left(1 + \frac{1}{\beta}, \left(\frac{z}{\theta}\right)^\beta\right) \right], \end{aligned} \quad (4.2)$$

$$\begin{aligned} F_z(z, \beta, \theta) &= 1 - (z+1)e^{-\left(\frac{z+1}{\theta}\right)^\beta} + ze^{-\left(\frac{z}{\theta}\right)^\beta} \\ &- \theta \left[\Gamma\left(1 + \frac{1}{\beta}, \left(\frac{z}{\theta}\right)^\beta\right) - \Gamma\left(1 + \frac{1}{\beta}, \left(\frac{z+1}{\theta}\right)^\beta\right) \right]. \end{aligned}$$

نمودارهای PMF و HRF توزیع BDW برای مقادیر مختلف پارامترهای β, θ ، به ترتیب در شکل ۱ و شکل ۲ رسم شده است.

نتیجه ۲.۲. امید ریاضی و واریانس توزیع BD به صورت زیر نشان داده می‌شود

$$E(Z) = E(Y), \quad Var(Z) = Var(Y) + \zeta_0(\delta),$$

که در آن $\zeta_0(\delta) = \sum_{i=0}^{\infty} \zeta_0(i, \delta)$ با

$$\zeta_0(i, \delta) = (2i+1)E_Y(1, i|\delta) - E_Y(2, i|\delta) - i(i+1)H(i, \delta).$$

طبق نتیجه ۲.۲، امید ریاضی و واریانس توزیع BDW به صورت زیر محاسبه می‌گردد

$$\begin{aligned} \mu_Z &= E(Z) = \theta \Gamma\left(1 + \frac{1}{\beta}\right), \\ \sigma_Z^2 &= Var(Z) = \theta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right] + \zeta_0(\beta, \theta), \end{aligned} \quad (5.2)$$

که در آن

$$E_Y(\nu, z|\beta, \theta) = \theta^\nu \left[\Gamma\left(\nu + \frac{\nu}{\beta}, \left(\frac{z}{\theta}\right)^\beta\right) - \Gamma\left(\nu + \frac{\nu}{\beta}, \left(\frac{z+1}{\theta}\right)^\beta\right) \right].$$

و با کمی محاسبات مقدماتی نتیجه می‌شود که

$$\begin{aligned} \zeta_\circ(i, \beta, \theta) &= (\nu i + 1)\theta \left[\Gamma\left(\nu + \frac{1}{\beta}, \left(\frac{i}{\theta}\right)^\beta\right) - \Gamma\left(\nu + \frac{1}{\beta}, \left(\frac{i+1}{\theta}\right)^\beta\right) \right] \\ &\quad - \theta^\nu \left[\Gamma\left(\nu + \frac{\nu}{\beta}, \left(\frac{i}{\theta}\right)^\beta\right) - \Gamma\left(\nu + \frac{\nu}{\beta}, \left(\frac{i+1}{\theta}\right)^\beta\right) \right] \\ &\quad - i(i+1) \left(e^{-\left(\frac{i}{\theta}\right)^\beta} - e^{-\left(\frac{i+1}{\theta}\right)^\beta} \right) \end{aligned}$$

شاخص پراکندگی فیشر (FDI) به صورت نسبت واریانس به میانگین تعریف می‌شود که بیانگر این است که داده‌ها بیش‌پراکنده (واریانس بیشتر از میانگین)، کم‌پراکندگی (واریانس کمتر از میانگین) یا هم‌پراکنده (واریانس و میانگین مساوی) هستند. بدین صورت که اگر $FDI < 1$ ، توزیع داده‌ها کم‌پراکنده، $FDI = 1$ توزیع داده‌ها هم‌پراکنده و اگر $FDI > 1$ ، توزیع داده‌ها بیش‌پراکنده است. میانگین، واریانس و شاخص پراکندگی فیشر برای توزیع BDW و انتخاب‌های گوناگونی از پارامترهای توزیع در جدول ۱ ارائه شده است.

۳.۲ مدل INAR(1) بر اساس نوفه‌هایی از توزیع BDW

در این بخش، به معرفی مدل INAR(1) بر اساس نوفه‌هایی از توزیع BDW می‌پردازیم. عملگر رقیق‌ساز دوجمله‌ای به صورت زیر تعریف شده است.

اکنون فرایند INAR(1) ایستای $\{X_t\}$ را توسط معادله بازگشتی زیر معرفی می‌نماییم

$$X_t = \alpha \circ X_{t-1} + Z_t, \quad 0 < \alpha < 1, \quad t \geq 1, \quad (۶.۲)$$

که در آن $\{Z_t\}$ دنباله‌ای از متغیرهای تصادفی از توزیع BDW با پارامترهای β و θ است. هم‌چنین "۰" نماد عملگر رقیق‌ساز دوجمله‌ای به صورت $\alpha \circ X = \sum_{i=1}^X W_i$ که W_i دنباله‌ای از متغیرهای تصادفی برنولی با پارامتر α ، $0 < \alpha < 1$ است. هم‌چنین متغیرهای تصادفی $\{Z_t\}$ و $\alpha \circ X_{t-1}$ به شرط مقدار معین X_{t-1} ، از یکدیگر مستقل می‌باشند. در ادامه، فرایند ایستای INAR(1) ارائه شده در (۶.۲) را به اختصار با نماد INAR-BDW(1) نشان می‌دهیم.

بر اساس (۶.۲)، تابع انتقال مرتبه یک به صورت زیر ارائه می‌گردد

$$P_{ij} = P(X_t = j | X_{t-1} = i) = \sum_{k=0}^{\min(i,j)} \binom{i}{k} \alpha^k (1-\alpha)^{i-k} f_Z(j-k, \beta, \theta). \quad (۷.۲)$$

که $f_Z(\cdot, \beta, \theta)$ در (۴.۲) تعریف شده است.

لم ۳.۲. فرایند $\{X_t\}$ ارائه شده در (۶.۲)، برگشت‌ناپذیر، نادره‌ای و زنجیر مارکف بازگشتی مثبت است، از این‌رو، $\{X_t\}$ ارگودیک و ایستا است.

□

اثبات. اثبات بدیهی و حذف شده است.

گزاره ۴.۲. فرایند ایستای INAR-BDW(1) به صورت $\{X_t\}$ را در نظر بگیرید، امید ریاضی و واریانس فرایند به ترتیب به صورت زیر ارائه می‌گردد

$$\begin{aligned} E(X_t) &= E(\alpha \circ X_{t-1} + Z_t) = \alpha E(X_{t-1}) + E(Z_t) = \frac{\mu_Z}{1-\alpha} \\ Var(X_t) &= Var(\alpha \circ X_{t-1} + Z_t) = Var(\alpha \circ X_{t-1}) + Var(Z_t) \\ &= \alpha(1-\alpha)E(X) + \alpha^\nu Var(X) + \sigma_Z^\nu \\ &= \frac{\alpha(1-\alpha)E(X) + \sigma_Z^\nu}{1-\alpha^\nu} = \frac{\alpha\mu_Z + \sigma_Z^\nu}{1-\alpha^\nu} \end{aligned}$$

که در آن μ_Z و σ_Z^ν در (۵.۲) تعریف شده است.

گزاره ۵.۲. فرایند ایستای $INAR-BDW(1)$ را در نظر بگیرید، برای $0 < \alpha < 1$ و $t \geq 1$ (الف) امید ریاضی شرطی به صورت زیر ارائه می‌گردد

$$E(X_t | X_{t-k}) = \alpha^k X_{t-k} + \frac{1 - \alpha^k}{1 - \alpha} \mu_Z. \quad (8.2)$$

به راحتی نتیجه می‌شود که برای $k \rightarrow \infty$ ، مقدار امید ریاضی شرطی به امید ریاضی غیر شرطی $\frac{\mu_Z}{1 - \alpha}$ همگرا است. (ب) واریانس شرطی به صورت زیر نشان داده می‌شود

$$Var(X_t | X_{t-1}) = \alpha(1 - \alpha)X_{t-1} + \sigma_Z^2, \quad (9.2)$$

و برای $k = 2, 3, \dots$

$$Var(X_t | X_{t-k}) = \alpha^k(1 - \alpha^k)X_{t-k} + \mu_Z \left[\frac{\alpha(1 - \alpha^{k-1})}{1 - \alpha} - \frac{\alpha^2(1 - \alpha^{2(k-1)})}{1 - \alpha^2} \right] + \sigma_Z^2 \frac{1 - \alpha^{2k}}{1 - \alpha^2},$$

که در آن σ_Z^2 واریانس نوفه‌های $\{Z_t\}$ است که در نتیجه ۲.۲ محاسبه شده است. همچنین

$$\lim_{k \rightarrow \infty} Var(X_t | X_{t-k}) = \mu_Z \frac{\alpha}{1 - \alpha^2} + \frac{\sigma_Z^2}{1 - \alpha^2},$$

که در واقع واریانس غیر شرطی فرایند است.

(ج) تابع خودبازگشتی فرایند $\{X_t\}$ به صورت زیر نشان داده می‌شود

$$\rho(k) = Corr(X_t, X_{t-k}) = \alpha^k.$$

۳ رویکردهای برآورد پارامترهای فرایند $INAR-BDW(1)$

این بخش، به بررسی روش‌های مختلف برآورد پارامترهای نامعلوم فرایند $INAR-BDW(1)$ ، شامل روش‌های برآورد حداکثر درستنمایی شرطی، کمترین مربعات شرطی تعمیم‌یافته و یول-واکر می‌پردازیم.

۱.۳ برآورد حداکثر درستنمایی شرطی

برآورد حداکثر درستنمایی شرطی (CML) پارامترهای فرایند $INAR-BDW(1)$ ، از طریق بیشینه نمودن تابع لگاریتم درستنمایی شرطی بر حسب پارامترهای $\delta = (\alpha, \beta, \theta)$ ، به دست می‌آید. تابع لگاریتم درستنمایی شرطی به صورت زیر ارائه می‌گردد

$$\ell(\delta; X_2, \dots, X_n) = \log L(X_2, \dots, X_n | \delta) = \sum_{t=2}^n \log P(X_t = j | X_{t-1} = i),$$

که در آن، $P(X_t = j | X_{t-1} = i)$ تابع انتقال (۷.۲) است. برآوردهای CML پارامترهای نامعلوم، با استفاده از روش‌های بهینه‌سازی عددی در نرم‌افزار آماری R مانند دستوره‌های `nlm` و `nlminb` محاسبه می‌گردد.

۲.۳ برآورد کمترین مربعات شرطی تعمیم‌یافته

برآوردهای کمترین مربعات شرطی تعمیم‌یافته (MCLS) بر اساس ترکیبی از برآوردهای کمترین مربعات شرطی و برآوردهای CML توسط میلیتیک ایلیک و همکاران [۱۸] معرفی شده است. به طوری که از برآوردهای CML به عنوان مقدار اولیه در برآورد کمترین مربعات شرطی استفاده می‌شود.

برآوردگرهای کمترین مربعات شرطی تعمیم یافته پارامترهای فرایند INAR-BDW(1) با کمینه نمودن تابع زیر به دست می آید

$$Q(\delta_1) = \sum_{t=2}^n (X_t - E(X_t|X_{t-1}))^2 = \sum_{t=2}^n \left(X_t - \alpha X_{t-1} - \theta \Gamma\left(1 + \frac{1}{\beta}\right) \right)^2.$$

برآوردگرهای MCLS به صورت زیر به دست می آید

$$\hat{\alpha}_{MCLS} = \frac{(n-1) \sum_{t=2}^n X_t X_{t-1} - \sum_{t=2}^n X_t \sum_{t=2}^n X_{t-1}}{(n-1) \sum_{t=2}^n X_{t-1}^2 - \left(\sum_{t=2}^n X_{t-1}\right)^2},$$

$$\hat{\theta}_{MCLS} = \frac{\sum_{t=2}^n X_t - \hat{\alpha}_{MCLS} \sum_{t=2}^n X_{t-1}}{(n-1) \Gamma\left(1 + \frac{1}{\hat{\beta}_{CML}}\right)},$$

به طوری که در آن $\hat{\beta}_{CML}$ برآورد CML پارامتر β است و به عنوان مقدار اولیه در برآورد پارامتر MCLS به کار برده می شود [۲۵]. بنابراین

$$\hat{\beta}_{MCLS} = \frac{1}{\Gamma^{-1}\left(\frac{\sum_{t=2}^n X_t - \hat{\alpha}_{MCLS} \sum_{t=2}^n X_{t-1}}{(n-1)\hat{\theta}_{MCLS}}\right) - 1},$$

که در آن $\Gamma^{-1}(\cdot)$ معکوس تابع گاما می باشد.

۳.۳ برآورد یول واکر

امید ریاضی $E(X_t) = \frac{\mu_Z}{1-\alpha} = \frac{\theta \Gamma\left(1 + \frac{1}{\beta}\right)}{1-\alpha}$ و تابع خودهمبستگی $Corr(X_t, X_{t-1}) = \alpha$ فرایند INAR-BDW(1) را در نظر بگیرید، با استفاده از میانگین نمونه ای و تابع خودهمبستگی نمونه ای برآوردهای یول واکر YW پارامترهای (α, μ_Z) به صورت زیر نتیجه می شود

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

$$\hat{\theta}_{YW} = \frac{\bar{X}(1 - \hat{\alpha}_{YW})}{\Gamma\left(1 + \frac{1}{\hat{\beta}_{YW}}\right)}.$$

برای محاسبه برآورد یول واکر پارامتر β از روش ارائه شده در آلتون و مامودخان [۴] استفاده می نماییم. شاخص پراکندگی فیشر فرایند INAR-BDW(1) به صورت

$$FDI_X = \frac{1}{1+\alpha} (FDI_Z + \alpha) = \frac{1}{1+\alpha} \left(\frac{\mu_Z}{\sigma_Z^2} + \alpha \right),$$

ارائه می شود و تابعی از دو پارامتر β و θ است. لذا با جایگذاری $\hat{\theta}_{YW}$ در FDI_Z و حل معادله زیر، برآورد یول واکر پارامتر β به دست می آید

$$\widehat{FDI}_X - \frac{1}{1 + \hat{\alpha}_{YW}} (FDI_Z + \hat{\alpha}_{YW}) = 0,$$

که در آن \widehat{FDI}_X مقدار تجربی شاخص پراکندگی فیشر به صورت $\widehat{FDI}_X = \frac{\bar{X}}{S_X^2}$ و S_X^2 واریانس نمونه ای فرایند است.

۴ مطالعات شبیه سازی روش های برآورد پارامترهای فرایند INAR-BDW(1)

بر اساس مطالعات شبیه سازی مونت کارلو، کارایی روش های برآورد پارامترهای فرایند INAR-BDW(1) را تحت مقادیر مختلف نمونه $n = 100, 200, 500, 1000$ و با $h = 1000$ تکرار بررسی و مقایسه می نماییم. در ادامه دو ترکیب مختلف از پارامترهای مدل به صورت $(\alpha, \phi, \beta) = (0.8, 0.6, 1.5), (0.4, 0.3, 0.5)$ را در نظر می گیریم. برای ارزیابی عملکرد روش های مختلف برآورد پارامترهای فرایند، از معیار ریشه میانگین مربعات خطا (RMSE) استفاده می کنیم. نتایج در **جدول ۲** خلاصه شده است که گویای همگرایی مقادیر برآورد به مقدار واقعی پارامترها است. علاوه بر این، با افزایش حجم نمونه، مقادیر RMSE رو به کاهش است. در میان روش های مختلف برآورد، روش پارامتری حداکثر درستنمایی، کمترین مقدار RMSE را نشان می دهد که بیانگر برتری این روش، نسبت به سایر روش های مورد بررسی است.

۵ آنالیز داده های واقعی گسسته مقدار

در این بخش به بررسی کارایی مدل INAR-BDW(1) در برازش داده های واقعی گسسته مقدار می پردازیم. بدین منظور، داده های فوت روزانه ناشی از COVID-19 در کوبا در بازه زمانی ۹ نوامبر ۲۰۲۱ الی ۱۳ ژانویه ۲۰۲۲ شامل $n = 66$ مشاهده، مالاوی در بازه زمانی ۱۹ سپتامبر الی ۱۲ دسامبر ۲۰۲۱ شامل $n = 85$ مشاهده، ازبکستان در بازه زمانی ۲۸ سپتامبر ۲۰۲۱ الی ۲۱ ژانویه ۲۰۲۲ شامل $n = 116$ مشاهده را در نظر می گیریم. داده های COVID-19 برگرفته از اطلاعات منتشر شده توسط سازمان جهانی بهداشت (WHO) در درگاه اینترنتی <https://covid19.who.int> می باشد.

نمودارهای مسیر نمونه ای، تابع خودهمبستگی (ACF) و تابع خودهمبستگی جزئی (PACF) هر سه سری داده های COVID-19 در **شکل ۳** نشان داده شده که موید ساختار فرایند خودبازگشتی مرتبه اول است.

برخی از ویژگی های آماری داده های واقعی در **جدول ۳** ارائه شده است. همچنین، ایستایی داده های واقعی با استفاده از آزمون دیکی فولر افزوده (ADF) بررسی و نتایج آن در **جدول ۳** ارائه شده که موید ایستایی هر سه سری داده واقعی است. طبق نتایج **جدول ۳**، داده های فوت ناشی از COVID-19 در کوبا و مالاوی بیش پراکنده و در ازبکستان کم پراکنده است. همان طور که پیشتر نیز اشاره شد، فرایند INAR-BDW(1) انواع مختلف پراکندگی مشاهدات شامل بیش پراکندگی و کم پراکندگی را پوشش می دهد و برای برازش داده های مذکور مناسب است.

در ادامه به مقایسه فرایند INAR-BDW(1) با سایر مدل ها INAR(1) که انواع مختلف پراکندگی (بیش پراکندگی و کم پراکندگی) را شامل می شوند، ارجحیت INAR-DBL(1) [۳]، INAR-DP(1) [۱۶]، INAR-DLi-3P(1) [۱۲]، INAR-DP(1) [۸] و INAR-GP(1) [۸] می پردازیم. در ادامه مقادیر برآورد CML و آماره های نیکویی برازش (GOF) شامل AIC، BIC، HQIC و CAIC را برای هر یک از مدل های INAR(1) مربوطه را در **جدول ۴** تا **۶** گزارش می نماییم.

نتایج به دست آمده، حاکی از برتری فرایند INAR-BDW(1) نسبت به سایر مدل های رقیب است، زیرا کمترین مقادیر آماره های GOF متعلق به فرایند INAR-BDW(1) است. بنابراین، فرایند INAR-BDW(1) نسبت به سایر مدل ها بهترین برازش را برای داده های فوت ناشی از COVID-19 ارائه می نماید.

۱.۵ آنالیز باقیمانده های داده های فوت ناشی از COVID-19

به منظور بررسی کفایت مدل مدنظر، به ارائه آنالیز باقیمانده های پیرسن داده های واقعی می پردازیم. با در نظر گرفتن امید ریاضی شرطی و واریانس شرطی معرفی شده در (۸.۲) و (۹.۲) و جایگذاری مقادیر برآورد پارامترها، باقیمانده های پیرسن به صورت زیر تعریف می شود

$$e_t = \frac{X_t - E(X_t | X_{t-1})}{\sqrt{Var(X_t | X_{t-1})}}$$

نمودار ACF نمونه ای باقیمانده های پیرسن در **شکل ۴** نشان داده شده که بیانگر عدم همبستگی باقیمانده ها است. همچنین با استفاده از آزمون جانگ-باگس و p -مقدارهای به دست آمده به صورت $(0.559, 0.441, 0.783)$ ، عدم همبستگی باقیمانده ها نیز تأیید می گردد. نمودار دوره نگار تجمعی باقیمانده ها نیز در **شکل ۵** ارائه شده که نشان دهنده توزیع تصادفی باقیمانده ها و عدم مشاهده هرگونه روندی در آنها است.

در **شکل ۶**، نتایج روش نمونه گیری پارامتری را نشان می دهد. ابتدا نمونه های با حجم $n = (66, 85, 116)$ و 5000 تکرار را با استفاده از برازش فرایند INAR-BDW(1) بر اساس مقادیر برآورد CML پارامترها تولید می نماییم. با استفاده از مجموعه داده های بوت استرپ، 5000 تابع خودهمبستگی و در هر لگ، چندک های $100(0.975)$ و $100(0.25)$ تابع ACF را به عنوان نواحی

پذیرش به دست می‌آوریم. کران ناحیه پذیرش با نماد "+" و مقادیر ACF نمونه‌ای نیز با نماد "●" نشان داده شده است. بر اساس شکل ۶، تمامی توابع خودهمبستگی نمونه‌ای میان ناحیه پذیرش قرار گرفته است که دلالت بر کفایت مدل مدنظر دارد.

۲.۵ روش‌های پیش‌بینی فرایند

در این بخش، به منظور بررسی قابلیت و توانایی پیش‌بینی فرایند INAR-BDW(1)، به ارائه پیش‌بینی با استفاده از دو رویکرد کلاسیک و بوت‌استرپ غربال می‌پردازیم. پیش‌بینی k -گام کلاسیک برای فرایند INAR-BDW(1) به صورت زیر نشان داده می‌شود

$$\hat{X}_t = E(X_t | X_{t-k}) = \alpha X_{t-k} + \frac{1-\alpha^k}{1-\alpha} E(Z),$$

که در آن α و $E(Z)$ با استفاده از مقادیر برآورد CML متناظرشان جایگزین می‌شوند.

۱.۲.۵ رویکرد پیش‌بینی بوت‌استرپ غربال

با توجه ساختار گسسته مشاهدات شمارشی، لذا پیش‌بینی‌کننده کلاسیک ویژگی گسسته داده‌های شمارشی را حفظ نمی‌کند. در این راستا، به ارائه روش پیش‌بینی بوت‌استرپ غربال می‌پردازیم که یک روش ناپارامتری و با ساختار گسسته مطابق با ماهیت داده‌های شمارشی می‌باشد (پاسکال و همکاران [۲۰] را ببینید). با استفاده از الگوریتم زیر، مراحل محاسبه پیش‌بینی‌های بوت‌استرپ غربال را ارائه می‌نماییم.

۱. پارامتر عملگر رقیق‌ساز α را با استفاده از رویکرد یول‌واکر برآورد نمایید.
۲. باقیمانده‌های $\hat{Z}_t = X_t - \hat{\alpha} X_{t-1}$ را برای $t = 2, \dots, n$ محاسبه کنید.
۳. توزیع تجربی باقیمانده‌های تعمیم‌یافته \tilde{Z}_t را به دست آورید، به طوری که $\tilde{Z}_t = [\hat{Z}_t]$ و نماد $[\cdot]$ ، دلالت بر گرد کردن اعداد به نزدیک‌ترین عدد صحیح دارد.
۴. سری نمونه‌های بوت‌استرپی X_t^b را به صورت زیر به دست آورید

$$X_t^b = \hat{\alpha} \circ X_{t-1}^b + Z_t^b, \quad b = 1, \dots, B,$$

که در آن B ، حجم نمونه بوت‌استرپی است و Z_t^b برای $t = 1, 2, \dots, n$ با استفاده از توزیع تجربی ارائه شده در گام سوم، تولید شده است.

۵. برآورد یول‌واکر پارامتر $\hat{\alpha}^b$ با استفاده از رابطه $\alpha = 1 - \frac{E(Z_t)}{E(X_t)}$ و جایگذاری میانگین نمونه‌ای محاسبه نمایید.

۶. بر اساس میانگین نمونه‌ای $\hat{\alpha}_i^b$ $\hat{\alpha}^b = \frac{1}{B} \sum_{i=1}^B \hat{\alpha}_i^b$ ، برآورد پارامتر α را به دست آورید.

۷. مشاهدات آینده بوت‌استرپی را با استفاده از معادله بازگشتی زیر به دست آورید

$$\hat{X}_{t+1}^b = \hat{\alpha}^b \circ X_t^b + Z_{t+1}^b.$$

به منظور ارزیابی دو روش پیش‌بینی، مقادیر پیش‌بینی کلاسیک و بوت‌استرپ غربال داده‌های فوت ناشی از COVID-19، در جدول ۷ گزارش شده‌اند که مقادیر مشاهده‌شده برای آن‌ها را در اختیار داریم. با استفاده از معیار میانگین درصد مطلق خطای متقارن (SMAPE)، به مقایسه عملکرد دو روش پیش‌بینی می‌پردازیم. بر اساس نتایج ارائه شده در جدول ۷، مقادیر SMAPE پیش‌بینی بوت‌استرپ غربال کوچک‌تر از کلاسیک است و علاوه بر این، پیش‌بینی‌کننده‌های بوت‌استرپ غربال مقادیر صحیح سازگار با ماهیت داده‌های واقعی ارائه می‌نماید. لذا به منظور پیش‌بینی فرایند، روش ناپارامتری بوت‌استرپ غربال ارجحیت دارد.

نتیجه‌گیری

با در نظر گرفتن روش گسسته‌سازی متوازن، یک نسخه گسسته از توزیع وایبل ارائه شده که از ویژگی‌های منطقی برخوردار است و برای برآزش انواع داده‌های واقعی کاربرد دارد. همچنین، یک فرایند خودبازگشتی گسسته‌مقدار مرتبه اول بر اساس توزیع گسسته متوازن وایبل ارائه شده است. ویژگی‌های مدل و همچنین رویکردهای برآورد متفاوت برای پارامترهای مدل مورد مطالعه قرار می‌گیرد. ارزیابی خواص و رویکردهای برآورد از طریق برخی مطالعات شبیه‌سازی انجام می‌شود که حاکی از برتری روش برآورد حداکثر درستنمایی شرطی نسبت به روش برآورد کمترین مربعات شرطی و بول واکر است. کفایت مدل $INAR(1)$ مدنظر توسط سه مجموعه داده‌های فوت ناشی از COVID-19 بررسی و با سایر مدل‌های رقیب مقایسه می‌شود. برای هر مجموعه داده COVID-19، تجزیه و تحلیل باقیمانده‌ها (باقیمانده‌های پیرسن)، و همچنین روش‌های پیش‌بینی کلاسیک و بوت استرپ غربال ارائه شده است که نتایج بیانگر برتری کفایت مدل مدنظر و برتری روش پیش‌بینی بوت‌استرپ غربال می‌باشد.

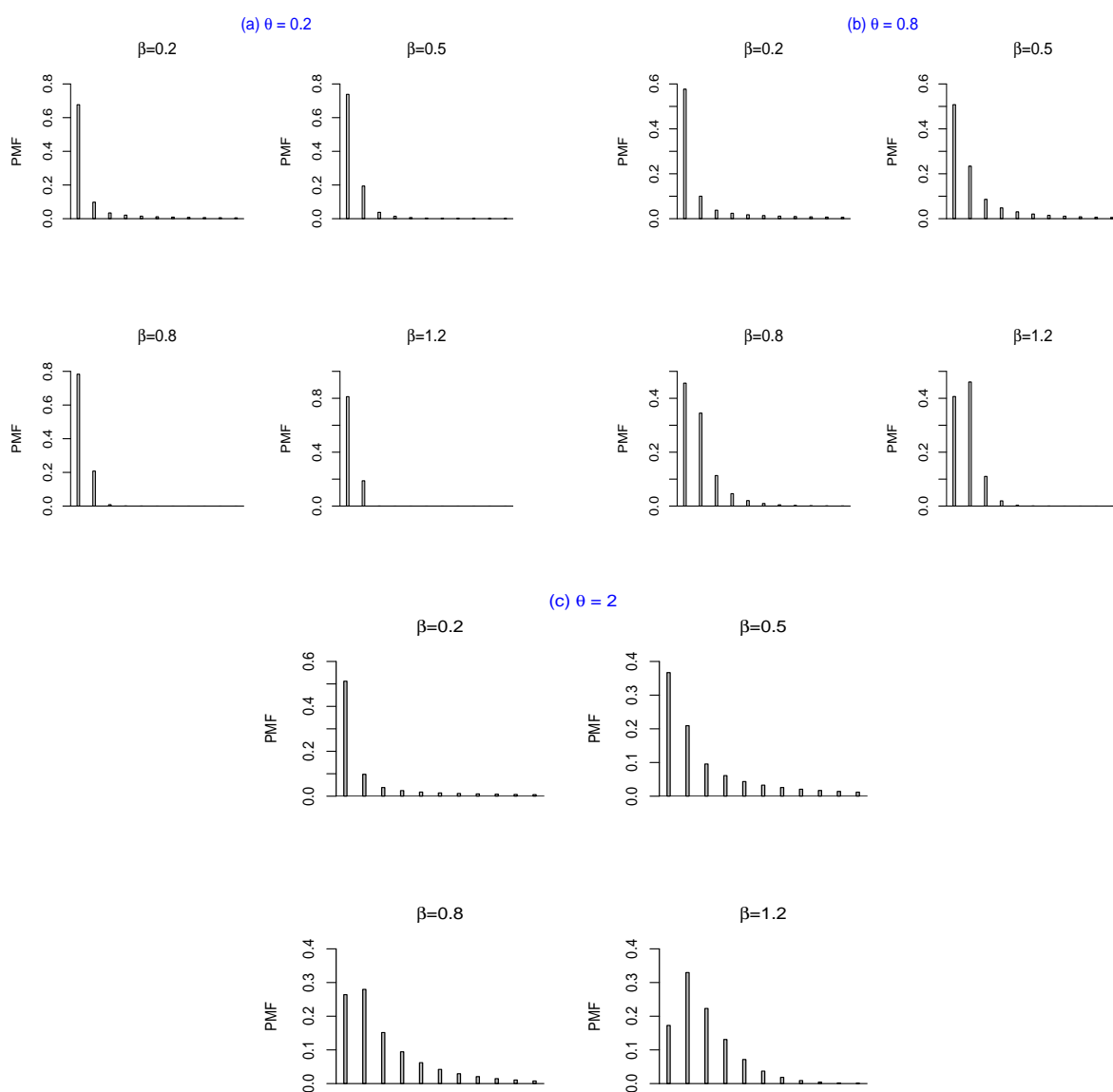
فهرست منابع

- [1] Al-Ani G.B., *Statistical modeling of the novel COVID-19 epidemic in Iraq*, Epidemiol. Methods., 10(s1) (2021), 20200025.
- [2] Al-Osh M.A. and Alzaid A.A., *First-order integer-valued autoregressive (INAR(1)) process*, J. Time Ser. Anal., 8 (1987), 261–275.
- [3] Altun E., El-Morshedy M. and Eliwa M.S., *A study on discrete Bilal distribution with properties and applications on integer-valued autoregressive process*, Revstat Stat. J., (2020).
- [4] Altun E. and Mamode Khan N.A., *Modelling with the novel INAR(1)-PTE process*, Methodol. Comput. Appl. Probab., (2021) <https://doi.org/10.1007/s11009-021-09878-2>.
- [5] Alzaid A.A. and Al-Osh M.A., *First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties*, Stat. Neerl., 42 (1988), 53–61.
- [6] Alzaid A.A. and Al-Osh M.A., *Some autoregressive moving average processes with generalized Poisson marginal distributions*, Ann. Inst. Stat. Math., 45 (1993), 223–232.
- [7] Bermúdez L. and Dimitris A.K., *Multivariate INAR(1) regression models based on the Sarmanov distribution*, Mathematics., 9(5) (2021), 505.
- [8] Bourguignon M., Rodrigues J. and Santos-Neto M., *Extended Poisson INAR(1) processes with equidispersion, underdispersion and overdispersion*, J. Appl. Stat., 46(1) (2019), 101–118.
- [9] Castellares F., Ferrari S.L.P. and Lemonte A.J., *On the Bell distribution and its associated regression model for count data*, Appl. Math. Model., 56 (2018), 172–185.
- [10] Chakaraborty S. and Chakravarty D., *Discrete gamma distribution: properties and parameter estimation*, Commun. Stat. - Theory Methods., 41 (2012), 3301–3324.
- [11] Chattopadhyay S., Maiti R., Das S. and Biswas A., *Change-point analysis through INAR process with application to some COVID-19 data*, Stat. Neerl., 76(1) (2022), 4–34.
- [12] Eliwa M.S., Altun E., El-Dawoody M. and El-Morshedy M., *A new three-parameter discrete distribution with associated INAR(1) process and applications*, IEEE Access, 8 (2020), 91150–91162.
- [13] Gumel A.B., Iboi E.A., Ngonghala C.N. and Elbasha E.H., *A primer on using mathematics to understand COVID-19 dynamics: Modeling, analysis and simulations*, Infect. Dis. Model., 6 (2021), 148–168.

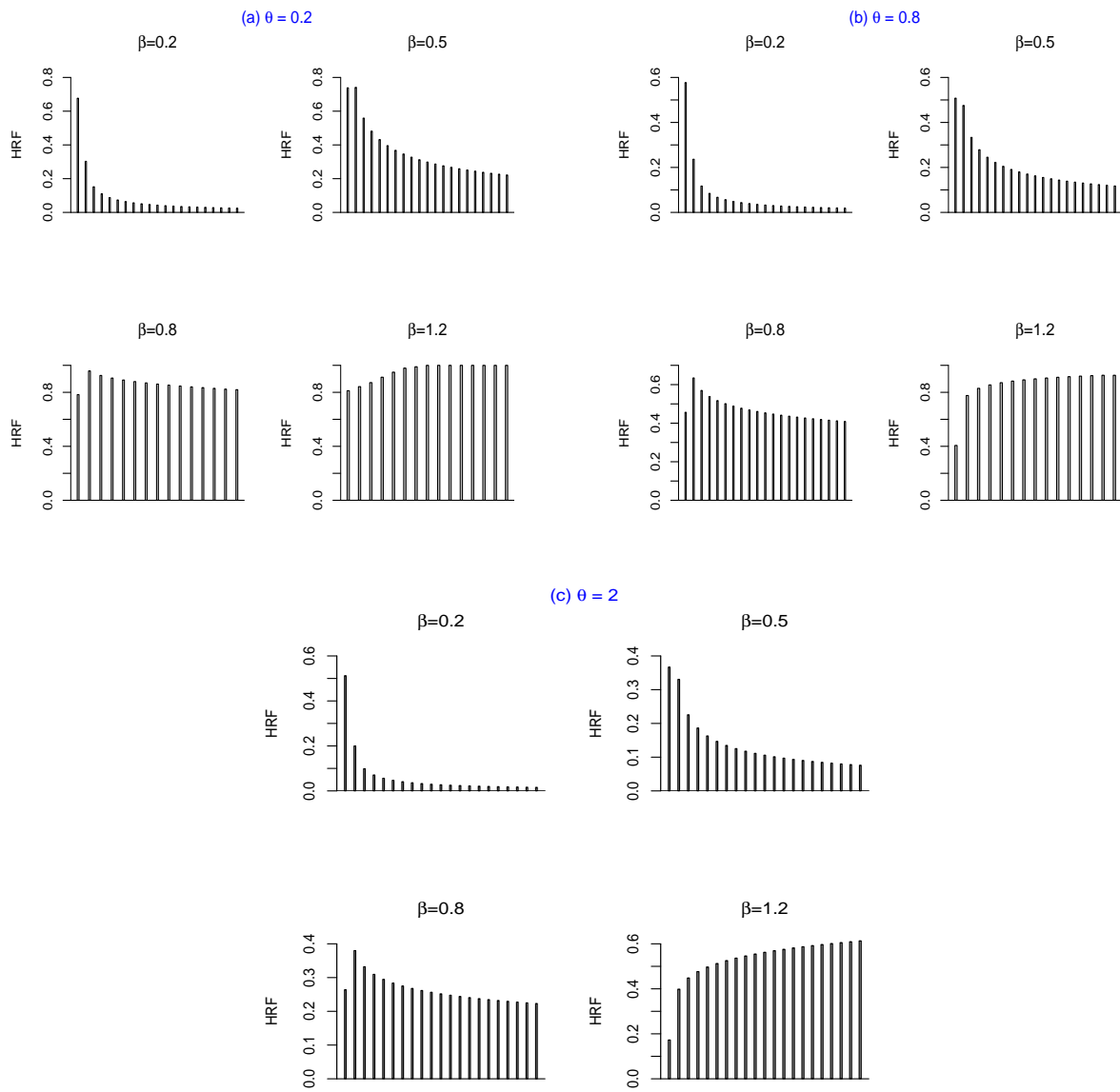
- [14] Hagmark P.E., *On construction and simulation of count data models*, Math. Comput. Simul., 77(1) (2008), 72–80.
- [15] Hegazy M., Abd EL-Kader R., AL-Dayian G. and EL-Helbawy A.A.-A., *Discrete inverted Kumaraswamy distribution: Properties and estimation*, Pak. J. Stat. Oper. Res., 18(1) (2022), 297–328.
- [16] Irshad M.R. و Chesneau C., D'cruz V. and Maya R., *Discrete pseudo Lindley distribution: properties, estimation and application on INAR(1) process*, Math. Comput. Appl., 26(4) 2021, 76.
- [17] Maleki M., Mahmoudi M.R., Wraith D. and Pho K.H., *Time series modelling to forecast the confirmed and recovered cases of COVID-19*, Travel. Med. Infect. Dis., 37 (2020), 101742.
- [18] Miletić Ilić A.V., Ristić M.M., Nastić A.S. and Bakouch H.S., *An INAR(1) model based on a mixed dependent and independent counting series*, J. Statist. Comput. Simul., 88(2) (2018), 290–304.
- [19] Muse A.H., Tolba A.H., Fayad E., Abu Ali O.A., Nagy M. and Yusuf M., *Modelling the COVID-19 mortality rate with a new versatile modification of the log-Logistic distribution*, Comput. Intell. Neurosci., 2021 (2021), 8640794.
- [20] Pascual L., Romo J. and Ruiz E., *Bootstrap predictive inference for ARIMA processes*, J. Time Ser. Anal., 25(4) (2004), 449–65.
- [21] Pourreza H., Baloui Jamkhaneh E. and Deiri E., *A family of Gamma-generated distributions: Statistical properties and applications*, Stat. Methods. Med. Res., 30(8) (2021), 1850–1873.
- [22] Roy D., *The discrete normal distribution*, Commun. Stat. - Theory Methods., 32(10) (2003), 1871–1883.
- [23] Roy D., *Discrete Rayleigh distribution*, IEEE Trans. Reliab., 53 (2004), 255–260.
- [24] Shamma N., Mohammadpour M. and Shirozhan M., *A time series model based on dependent zero inflated counting series*, Comput. Stat., 35 (2020), 1737–1757.
- [25] Shirozhan M. and Mohammadpour M., *A new class of INAR(1) model for count time series*, J. Stat. Comput. Simul., 88(7) (2018), 1348–1368.
- [26] Tovissodé C.F., Honfo S.H., Doumaté J.T. and Glélé Kakai R., *On the discretization of continuous probability distributions using a probabilistic rounding mechanism*, Mathematics., 9 (2021) 555.
- [27] Triacca M. and Triacca U., *Forecasting the number of confirmed new cases of COVID-19 in Italy for the period from 19 May to 2 June 2020*, Infect. Dis. Model., 6 (2021), 362–369.

جدول ۱: برخی شاخص‌های آماری توزیع BDW برای ترکیب‌های گوناگونی از پارامترهای β و θ .

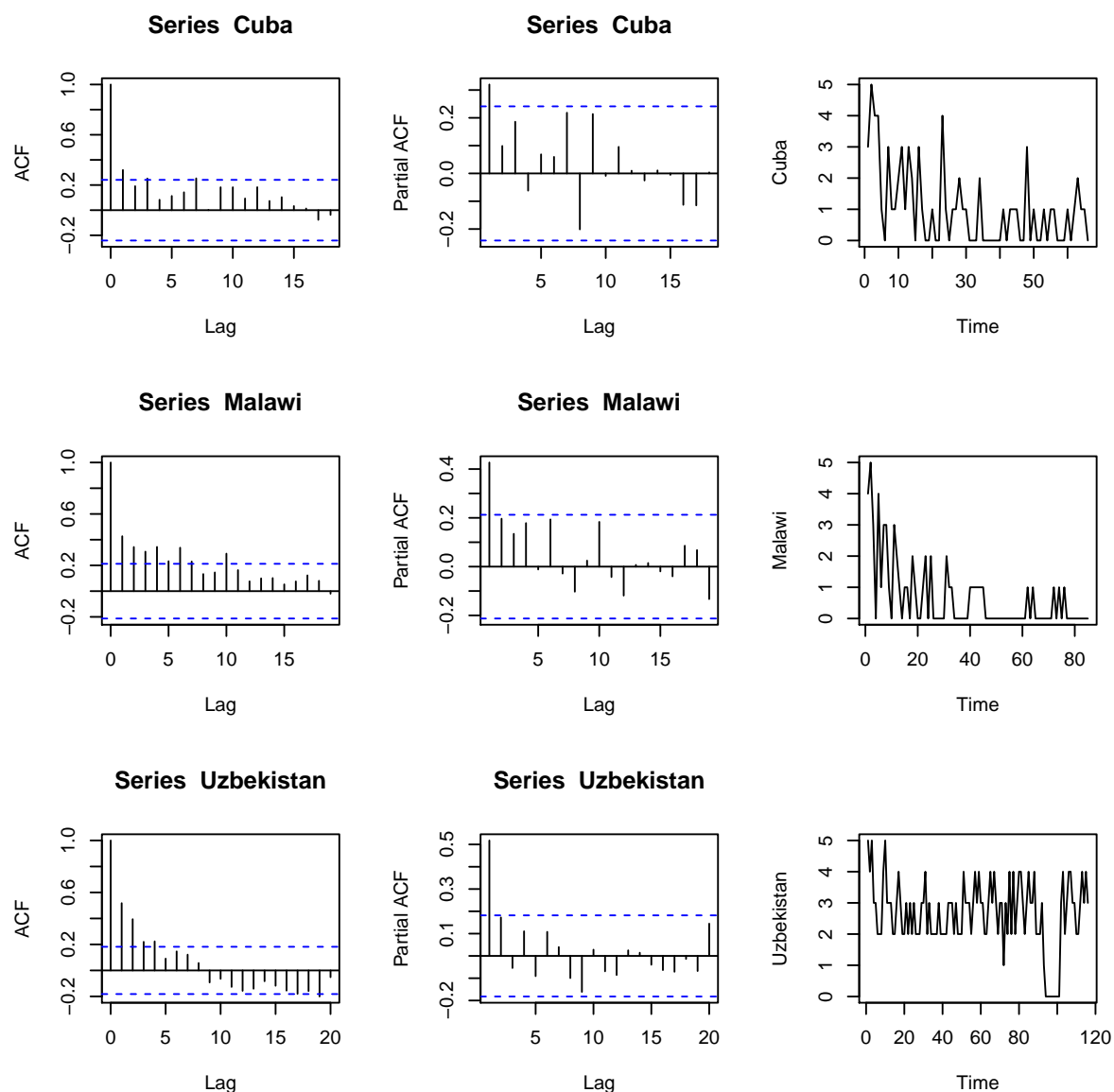
شاخص‌ها	$\theta = 0.5$			$\theta = 1.5$			$\theta = 2$		
	$\beta = 0.5$	$\beta = 1.44$	$\beta = 2$	$\beta = 0.5$	$\beta = 1.44$	$\beta = 2$	$\beta = 0.5$	$\beta = 1.44$	$\beta = 2$
میانگین	۱/۰۰۰	۰/۴۵۳	۰/۴۴۳	۳/۰۰۰	۱/۳۶۱	۱/۳۲۹	۴/۰۰۰	۱/۸۱۵	۱/۷۷۲
واریانس	۵/۱۲۱	۰/۲۷۹	۰/۲۵۱	۴۵/۱۳۹	۱/۰۹۱	۰/۶۵۲	۸۰/۱۴۳	۱/۸۰۷	۱/۰۲۶
FDI	۵/۱۲۱	۰/۶۱۵	۰/۵۶۶	۱۵/۰۴۶	۰/۸۰۲	۰/۴۹۱	۲۰/۰۳۵	۰/۹۹۵	۰/۵۷۹



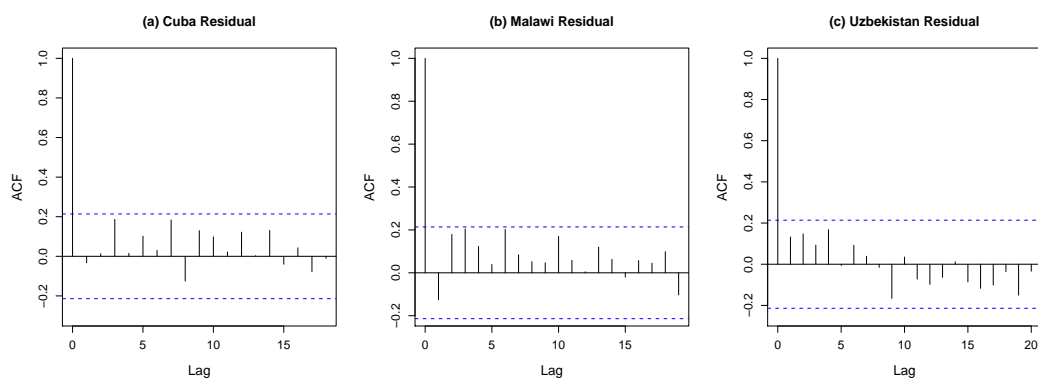
شکل ۱: نمودار PMF توزیع BDW برای مقادیر مختلف پارامترهای θ و β .



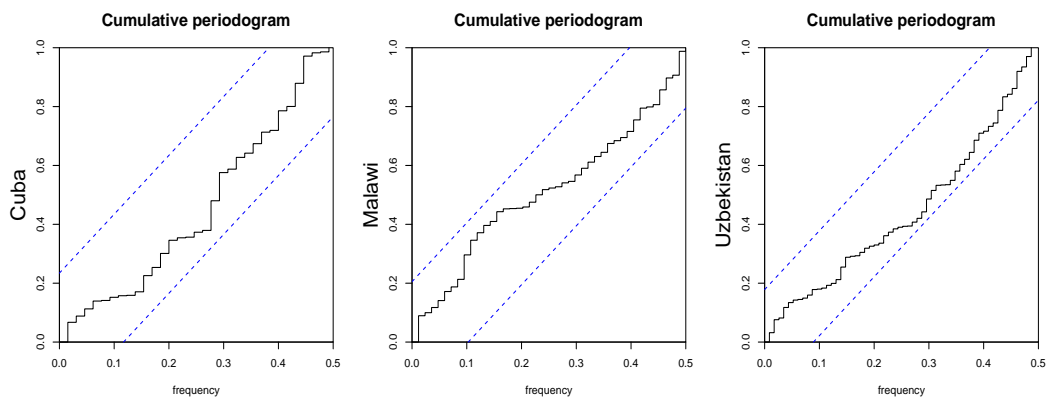
شکل ۲: نمودار HRF توزیع BDW برای مقادیر مختلف پارامترهای β و θ .



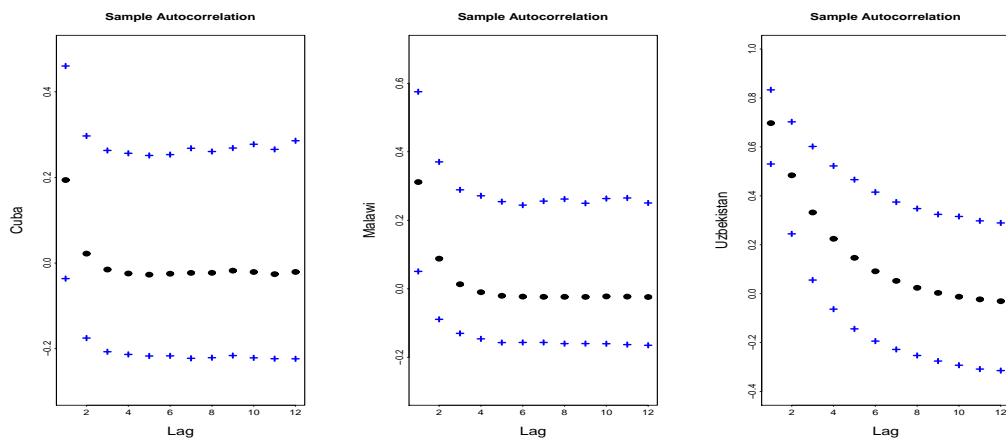
شکل ۳: نمودارهای مسیر نمونه‌ای، ACF و PACF داده‌های فوت ناشی از COVID-19.



شکل ۴: نمودار ACF نمونه‌ای باقیمانده‌های پیرسن داده‌های فوت ناشی از COVID-19.



شکل ۵: نمودار دوره‌نگار باقیمانده‌های پیرسن داده‌های فوت ناشی از COVID-19.



شکل ۶: ACF بوت‌استرپ و ناحیه پذیرش.

جدول ۲: نتایج شبیه سازی برآورد پارامترهای مدل INAR-BDW(1) و مقادیر RMSE در پراتنز.

n	CML			MCLS			YW		
	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\phi}$	$\hat{\beta}$
	$(\alpha, \beta, \theta) = (0.8, 0.6, 1.5)$								
۱۰۰	۰.۷۸۲۳	۰.۵۵۶۳	۱.۳۹۳۵	۰.۷۴۸۷	۰.۵۲۰۴	۱.۳۷۱۷	۰.۷۳۰۸	۰.۶۹۵۵	۱.۳۳۴۲
RMSE	(۰.۰۳۶۱)	(۰.۱۲۲۲)	(۰.۳۵۸۶)	(۰.۰۵۹۸)	(۰.۲۱۱۹)	(۰.۵۹۸۵)	(۰.۰۷۶۸)	(۰.۱۹۵۱)	(۰.۶۴۶۳)
۲۰۰	۰.۷۹۱۴	۰.۵۸۰۱	۱.۴۴۵۷	۰.۷۷۵۷	۰.۵۴۲۵	۱.۳۹۲۱	۰.۷۶۱۸	۰.۶۵۰۵	۱.۴۰۷۵
RMSE	(۰.۰۲۵۹)	(۰.۰۹۳۴)	(۰.۳۰۸۲)	(۰.۰۴۷۶)	(۰.۱۷۰۸)	(۰.۵۵۲۹)	(۰.۰۵۲۲)	(۰.۱۷۵۵)	(۰.۴۷۹۸)
۵۰۰	۰.۷۹۴۳	۰.۵۹۳۱	۱.۴۶۱۶	۰.۷۹۷۸	۰.۶۱۲۴	۱.۴۳۱۳	۰.۷۹۰۵	۰.۶۱۵۹	۱.۵۶۷۷
RMSE	(۰.۰۲۳۱)	(۰.۰۸۲۶)	(۰.۱۹۶۶)	(۰.۰۳۷۴)	(۰.۱۰۶۱)	(۰.۳۱۲۹)	(۰.۰۳۶۶)	(۰.۱۱۱۸)	(۰.۳۱۵۵)
۱۰۰۰	۰.۷۹۹۴	۰.۵۹۸۸	۱.۴۸۵۲	۰.۷۹۶۱	۰.۵۹۰۶	۱.۴۸۵۲	۰.۷۹۲۶	۰.۵۹۱۳	۱.۴۹۱۷
RMSE	(۰.۰۱۲۶)	(۰.۰۳۸۵)	(۰.۰۹۰۴)	(۰.۰۲۶۹)	(۰.۰۸۰۳)	(۰.۱۹۰۲)	(۰.۰۲۹۵)	(۰.۰۸۰۱)	(۰.۱۹۳۳)
	$(\alpha, \beta, \theta) = (0.4, 0.3, 0.5)$								
۱۰۰	۰.۳۹۱۲	۰.۲۸۹۲	۰.۴۸۸۱	۰.۳۷۴۶	۰.۲۹۰۸	۰.۵۳۱۸	۰.۳۸۸۲	۰.۲۸۳۵	۰.۴۷۷۴
RMSE	(۰.۰۲۸۱)	(۰.۰۲۱۷)	(۰.۳۱۹۸)	(۰.۰۸۳۹)	(۰.۰۵۹۹)	(۰.۴۳۶۹)	(۰.۰۹۳۹)	(۰.۰۵۸۲)	(۰.۵۱۸۳)
۲۰۰	۰.۳۹۴۵	۰.۲۹۰۶	۰.۵۰۸۳	۰.۴۱۰۲	۰.۲۹۳۴	۰.۵۱۹۹	۰.۳۸۴۷	۰.۲۸۳۸	۰.۵۱۰۸
RMSE	(۰.۰۱۹۵)	(۰.۰۱۹۴)	(۰.۲۵۹۴)	(۰.۰۶۳۷)	(۰.۰۴۴۱)	(۰.۳۷۱۹)	(۰.۰۶۰۳)	(۰.۰۳۱۲)	(۰.۳۲۶۹)
۵۰۰	۰.۳۹۶۱	۰.۲۹۸۱	۰.۴۹۰۹	۰.۳۹۰۲	۰.۲۹۴۹	۰.۵۱۰۹	۰.۳۹۳۷	۰.۲۹۸۳	۰.۴۹۸۴
RMSE	(۰.۰۱۱۲)	(۰.۰۰۶۹)	(۰.۱۷۹۳)	(۰.۰۴۲۶)	(۰.۰۲۹۷)	(۰.۲۷۵۱)	(۰.۰۴۵۴)	(۰.۰۲۰۵)	(۰.۲۴۶۵)
۱۰۰۰	۰.۳۹۸۷	۰.۲۹۹۲	۰.۴۹۵۹	۰.۳۹۸۴	۰.۲۹۶۷	۰.۵۰۹۷	۰.۳۹۲۵	۰.۲۹۷۷	۰.۴۹۶۸
RMSE	(۰.۰۰۸۵)	(۰.۰۰۲۲)	(۰.۰۹۲۸)	(۰.۰۲۷۳)	(۰.۰۱۲۸)	(۰.۱۳۸۶)	(۰.۰۳۰۶)	(۰.۰۱۹۸)	(۰.۱۶۰۸)

جدول ۳: برخی معیارهای آماری و p -مقدار آزمون ADF برای داده های فوت ناشی از COVID-19.

	FDI	p-value	خودهمبستگی	واریانس	میانگین	مجموعه داده ها
کوبا	۱.۴۹۲	۰.۰۱۶	۰.۳۲	۱.۵۳۷	۱.۰۳۱	
مالاوی	۱.۷۹۲	۰.۰۱۳	۰.۴۲۷	۱.۱۵۹	۰.۶۴۷	
ازبکستان	۰.۴۴۸	۰.۰۴۳	۰.۵۱۸	۱.۱۸۷	۲.۶۴۶	

جدول ۴: برآورد CML و مقادیر GOF برای داده‌های فوت ناشی از COVID-19 در کوبا.

مدل	CML	AIC	BIC	HQIC	CAIC
INAR-DBL(1)	$\hat{\alpha} = 0.2088, \hat{p} = 0.5199$	187.34	191.72	190.07	187.53
INAR-DPsL(1)	$\hat{\alpha} = 0.235, \hat{\beta} = 2.7909, \hat{\theta} = 1.1389$	185.09	191.65	187.68	186.47
INAR-DLi-3P(1)	$\hat{\alpha} = 0.6931, \hat{\beta} = 0.4407, \hat{\lambda} = 0.3201, \hat{p} = 0.2355$	181.09	189.84	184.55	181.74
INAR-DP(1)	$\hat{\alpha} = 0.2381, \hat{\phi} = 0.4837, \hat{\mu} = 0.5761$	188.99	195.56	191.59	189.38
INAR-GP(1)	$\hat{\alpha} = 0.2337, \hat{\phi} = 0.1951, \hat{\mu} = 0.6081$	184.21	190.77	186.81	184.59
INAR-BDW(1)	$\hat{\alpha} = 0.2238, \hat{\beta} = 0.7225, \hat{\theta} = 0.6319$	179.97	186.53	182.56	180.46

جدول ۵: برآورد CML و مقادیر GOF برای داده‌های فوت ناشی از COVID-19 در مالاوی.

مدل	CML	AIC	BIC	HQIC	CAIC
INAR-DBL(1)	$\hat{\alpha} = 0.3224, \hat{p} = 0.3846$	176.02	180.91	177.98	176.17
INAR-DPsL(1)	$\hat{\alpha} = 0.3329, \hat{\beta} = 5670.32, \hat{\theta} = 1.2727$	175.11	182.44	178.06	175.41
INAR-DLi-3P(1)	$\hat{\alpha} = 1.3892, \hat{\beta} = 0.0098, \hat{\lambda} = 0.2801, \hat{p} = 0.332$	175.11	184.88	179.04	175.61
INAR-DP(1)	$\hat{\alpha} = 0.6403, \hat{\phi} =, \hat{\mu} = 0.3594$	175.11	182.73	177.47	175.19
INAR-GP(1)	$\hat{\alpha} = 0.3461, \hat{\phi} = 0.2721, \hat{\mu} = 0.2769$	176.85	184.18	179.79	177.14
INAR-BDW(1)	$\hat{\alpha} = 0.3405, \hat{\beta} = 0.4726, \hat{\theta} = 0.1829$	171.87	179.21	174.82	172.25

جدول ۶: برآورد CML و مقادیر GOF برای داده‌های فوت ناشی از COVID-19 در ازبکستان.

مدل	CML	AIC	BIC	HQIC	CAIC
INAR-DBL(1)	$\hat{\alpha} = 0.7653, \hat{p} = 0.4641$	322.19	327.69	324.42	322.29
INAR-DPsL(1)	$\hat{\alpha} = 0.7391, \hat{\beta} = 0.5064, \hat{\theta} = 2.3031$	322.24	330.51	325.59	322.45
INAR-DLi-3P(1)	$\hat{\alpha} = -0.1659, \hat{\beta} = 0.7745, \hat{\lambda} = 0.9999, \hat{p} = 0.7391$	324.24	335.25	328.71	324.61
INAR-DP(1)	$\hat{\alpha} = 0.7374, \hat{\phi} = 1.5571, \hat{\mu} = 0.6911$	323.02	331.28	326.37	323.23
INAR-GP(1)	$\hat{\alpha} = 0.7388, \hat{\phi} = -0.1428, \hat{\mu} = 0.7691$	321.11	329.36	324.45	321.31
INAR-BDW(1)	$\hat{\alpha} = 0.7347, \hat{\beta} = 1.2267, \hat{\theta} = 0.7322$	317.15	325.41	320.51	317.42

جدول ۷: پیش‌بینی k -گام داده‌های فوت ناشی از COVID-19.

k	کوبا			مالاوی			ازبکستان		
	داده‌های واقعی	بوت‌استرپ	کلاسیک	داده‌های واقعی	بوت‌استرپ	کلاسیک	داده‌های واقعی	بوت‌استرپ	کلاسیک
۱	۰	۰/۷۷۶۸	۱	۱	۰/۷۴۸۶	۰	۴	۳/۶۲۳۷	۵
۲	۰	۰/۷۷۶۸	۰	۰	۰/۴۰۸۱	۰	۳	۲/۸۸۹۱	۳
۳	۰	۰/۷۷۶۸	۲	۰	۰/۴۰۸۱	۰	۳	۲/۸۸۹۱	۲
۴	۱	۱/۰۰۰۰۶	۱	۰	۰/۴۰۸۱	۱	۲	۲/۱۵۴۳	۲
۵	۰	۰/۷۷۶۸	۱	۰	۰/۴۰۸۱	۱	۲	۲/۱۵۴۳	۱
۶	۱	۱/۰۰۰۰۶	۱	۰	۰/۴۰۸۱	۰	۳	۲/۸۸۹۱	۴
۷	۲	۱/۲۲۴۴	۲	۰	۰/۴۰۸۱	۱	۴	۳/۶۲۳۷	۴
۸	۱	۱/۰۰۰۰۶	۱	۰	۰/۴۰۸۱	۱	۳	۲/۸۸۹۱	۵
۹	۱	۱/۰۰۰۰۶	۱	۰	۰/۴۰۸۱	۰	۴	۳/۶۲۳۷	۴
۱۰	۰	۰/۷۷۶۸	۱	۰	۰/۴۰۸۱	۰	۳	۲/۸۸۹۱	۳
	SMAPE	۱/۰۴۸	۰/۷۹۳		۱/۸۲۸	۰/۹۹۱		۰/۳۲۴۶	۰/۲۰۷



The Balanced Discrete Weibull Distribution and Its Corresponding Integer-value Autoregressive Model: Properties, Estimation and Analysis of Counting Death of COVID-19 Data

Seyedeh Mahbubeh Hoseini Baladezaei , Einolah Deiri*, Ezzatallah Baloui Jamkhaneh

Department of Statistics, Qaemshahr Branch, Islamic Azad University, Qaemshahr, Iran

Communicated by: H. Torabi

Received: 2022/4/8

Accepted: 2022/8/13

Abstract: In this paper, we introduce a new discrete Weibull distribution based on the balanced discretization method, which preserves the partial moments between the two discrete and continuous versions of the distributions. Some statistical features of the new distribution and different kinds of dispersion of the proposed distribution are presented based on various selections of parameters. In addition to introducing the new version of balanced discrete Weibull, we provide the integer-valued autoregressive model with the innovation of the proposed discrete distribution and evaluate different methods for estimating the model parameters. Using the counts of death of the COVID-19 data in Cuba, Malawi and Uzbekistan, we appraise the performance of the new process in fitting real data to some classical integer-valued autoregressive models. Finally, the forecasting of the process is checked based on real data using both classical and sieve bootstrap approaches.

Keywords: Balanced discrete Weibull, Partial mean preserving, Integer-valued autoregressive model, Forecasting of process, COVID-19.



©2022 Shahid Chamran University of Ahvaz, Ahvaz, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0 license) (<http://creativecommons.org/licenses/by-nc/4.0/>).

*Corresponding author.

E-mail addresses: Mhoseiny1199@gmail.com (S. M. Hoseini Baladezaei), e.deiri53@qaemiau.ac.ir (E. Deiri), e_baloui2008@yahoo.com (E. Baloui Jamkhaneh)