نشریه پژوهش‌های اقلیم شناسی
سال چهاردهم | شماره پنجاه و سوم | بهار ۱۴۰۲
وصول: ۱۴۰۱/۰۹/۲۲ پذیرش: ۱۴۰۲/۰۳/۲۲
صص ۵۶-۵۱

برآورد پارامتر مدل در تغییرات اقلیمی با بکارگیری داده های رکوردی حاصل از طرح نمونه گیری مجموعه رتبه دار رکوردی

احسان گل‌زاده گروی[*1]
استادیار، گروه آمار، دانشگاه پیام نور، تهران، ایران

چکیده

امروزه از آنجایی که عوامل اقلیمی در میان عوامل گوناگون، اثری جدی بر مطالعات مربوط به زندگی بشری می‌گذارد. ضرورت دارد، در برنامه ریزی‌های مختلف، نقش پارامترهای اقلیمی به عنوان عاملی تاثیرگذار در روند اجرایی برنامه ها مورد بررسی قرارگیرد. یکی از مباحث مهم در اقلیم شناسی، برآورد پارامترهای مجهول مدل می‌باشد. در این مقاله مدل مورد بررسی توزیع نمایی تعمیم یافته است. برآورد پارامترهای مدل بر اساس اطلاعات نمونه در دسترس و استفاده از یک طرح نمونه گیری که منجر به کاهش هزینه و افزایش دقت برآوردگرها گردد بسیار مفید و لازم است. در این پژوهش با استفاده از روش های ماکسیمم درستنمایی و روش بیزی و استفاده از داده های رکوردی حاصل از طرح نمونه گیری مجموعه رتبه دار رکوردی (RRSS)، پارامترهای مدل برآورد می شوند. در ادامه به کمک شبیه سازی مونت کارلو، معیار مخاطره برآوردگرها مورد ارزیابی قرار می گیرد. در انتها، نتایج به کمک تحلیل داده های واقعی مربوط به رکوردهای حاصل از داده‌های درجه حرارت در مقیاس زمانی روزانه از ماه دی طی سال های ۱۳۹۷ تا ۱۴۰۰ که از ایستگاه هواشناسی شهر ساری در استان مازندران به دست آمده‌اند بررسی شده است. نتایج ارزیابی نشان می دهد که در استفاده از طرح RRSS برای برآورد پارامتر مدل، با افزایش رکوردها برآوردگر بیزی دقت بالاتری در مقایسه با برآوردگر ماکسیمم درستنمایی از خود نشان می دهد.

کلید واژه‌ها: برآورد ماکسیمم درست نمایی، برآورد بیزی، مقادیر رکورد پایین، طرح *RRSS*، استان مازندران.

* نویسنده مسئول: ۰۹۱۱۱۱۱۴۰۶۵۰
Email: e_golzadeh_g@pnu.ac.ir

مقدمه

آزمایش‌های کاربردی متعددی وجود دارد که به دلایلی نامعلوم، تعدادی از مشاهدات مورد مطالعه در آن محدود و ناپدید شده‌اند. برای نمونه، بعضی از این محدودیت‌ها عبارتند از: فرصت کم برای اعلام نتایج، طولانی شدن زمان آزمایش، عدم دسترسی به همه واحدها و یا مایوس شدن از نتیجه دادن همه واحدها. این عوامل سبب می‌شود محقق دسترسی به تمامی داده‌های مورد مطالعه نداشته باشد. در بررسی دنباله‌ای از پدیده‌های تصادفی، مانند ثبت داده‌های درجه حرارت، میزان بارش، پیشامدهای طبیعی مانند زلزله، سیل، طغیان رودخانه‌ها، فعال شدن آتشفشان و غیره، فقط تعداد محدودی مشاهدات ثبت و مورد توجه قرار می‌گیرند که مقادیر آنها از مشاهدات ماقبل خود بزرگتر یا کوچکتر هستند. همین سبب می‌شود که محقق به تمامی مشاهدات دسترسی نداشته باشد. بنابراین ناچار است در چنین وضعیتی برای تحلیل داده‌ها از رفتار همان تعداد محدود مشاهدات که به مقادیر رکوردی معروفند به عنوان یک مجموعه داده مشاهده شده استفاده کند. مقادیری مثل رکورد میزان بارش، سرعت باد، ارتفاع رودخانه، مقدار آلودگی هوا، دمای متوسط هوا، سونامی و غیره. برای مثال، وقتی یک مجموعه از داده‌های مربوط به وزش باد را بررسی می کنیم آنچه مورد نظر است تندبادها و بادهای سهمگینی است که باعث خسارت به منازل مسکونی، منازل، مزارع و غیره می شوند. با توجه به تغییرات اقلیمی، در سال‌های اخیر به علت پیامدهای اقتصادی، اجتماعی و خسارت‌های مربوط به رویدادهای جوی در بعضی از شاخه‌های علوم نظیر هواشناسی، استفاده از مقادیر رکوردی مورد توجه قرار گرفته و روش‌های آماری مختلفی برای توصیف آن به کار رفته است. بعضی پیامدهای ناشی از تغییرات اقلیمی عبارتند از: افزایش مهاجرت، انقراض و نابودی جانوران، بالا آمدن آب سطح دریاها وآب شدن یخچال ها (شاهی نژاد و همکاران، ۲۰۱۶)، (عاقل پور و همکاران، ۲۰۱۸) و (سیدنژاد و همکاران، ۲۰۲۲). مواجهه با چنین مسائلی مشابه موارد ذکر شده، موجب تسریع مدل بندی پیشامدهای نادر شده است. اولین بار چاندلر در سال ۱۹۵۲ نظریه رکوردها را پیشنهاد داد که در چند دهه اخیر نگاه آمارانان و هیدرولوژیست ها را به خود معطوف نموده است. پژوهش بسیاری در زمینه رکوردها انجام شده است. برای اطلاع بیشترمی‌توان به مقالات (صالحی و همکاران، ۲۰۱۵)، (رزمخواه و همکاران، ۱۳۹۸) و (گل زاده و همکاران، ۲۰۲۱) مراجعه کرد. نظریه رکوردها علاوه بر ویژگی‌های مهم تئوری دارای کاربردهای خاص عملی است. در اینجا مقادیر رکوردی و نحوه به دست آوردن رکوردهای پایین و بالا به اختصار توضیح داده می‌شود:

یک دنباله نامتناهی از متغیرهای تصادفی مستقل و هم توزیع $\{X_i,\ i \geq 1\}$ با تابع توزیع تجمعی مشترک $F(x;\theta)$ و تابع چگالی احتمال $f(x;\theta)$ فرض کنید. باشد، آنگاه مشاهده‌ی X_j یک رکورد پایین (بالا) است، اگر برای همه مقادیر i که $j > i$ ، $X_j(>) < X_i$ برقرار باشد. به بیان دیگر، برای اینکه یک رکورد جدید رخ دهد باید بتوان آن را با مشاهدات قبلی مقایسه نمود.گوئیم مشاهده X_j یک رکورد پایین (بالا) است، هرگاه از تمام مشاهدات قبل از خود کوچکتر (بزرگتر) باشد. بنابراین اولین مشاهده یعنی X_1 اولین رکورد (پایین و بالا) است که به آن رکورد بدیهی نیز گفته می‌شود. فرض کنید یک نمونه دنباله تصادفی $x_1 = 24$ ، $x_2 = 12$ ، $x_3 = 29$ ، $x_4 = 21$، $x_5 = 32$ ، $x_6 = 58$، $x_7 = 18$ ، $x_8 = 8$ به حجم n=8 داشته باشیم. اولین مشاهده یعنی مقدار۲٤، اولین رکورد بالا است. برای ثبت دومین رکورد بالا، باید مشاهده ای در ادامه دنباله اعداد انتخاب شود که مقدار آن ازمقدار۲٤ بزرگتر باشد یعنی مقدار ۲۹. برای ثبت سومین رکورد بالا، باید مشاهده ای در ادامه دنباله اعداد انتخاب شود که مقدار آن از ۲٤ و۲۹ بزرگتر باشد یعنی مقدار۳۲. به همین ترتیب برای ثبت رکورد چهارم بالا، باید مشاهده ای در دنباله اعداد انتخاب شود که مقدار آن از تمام رکوردهای قبلی ۲٤، ۲۹ و۳۲ بزرگتر باشد یعنی مقدار ۵۸ و الی آخر. در اینصورت برای دنباله تصادفی فوق، مقادیر رکوردهای بالا عبارتند از :

$$U_1 = 24, U_2 = 29, U_3 = 32, U_4 = 58$$

به طور مشابه می‌توان گفت، اولین مشاهده یعنی مقدار۲٤، اولین رکورد پایین است. برای ثبت دومین رکورد پایین، باید مشاهده ای در ادامه دنباله اعداد انتخاب شود که مقدار آن از ۲٤ کوچکتر باشد یعنی مقدار ۱۲. برای ثبت سومین رکورد پایین، باید مشاهده ای در ادامه دنباله اعداد انتخاب شود که

مقدار آن از ۲٤ و ۱۲ کوچکتر باشد یعنی مقدار ۸. به همین ترتیب الی آخر.

در اینصورت برای دنباله تصادفی فوق، مقادیر رکوردهای پایین عبارت است از

$$R_1 = 24, \quad R_2 = 12, \quad R_3 = 8$$

برای آشنایی بیشتر درباره رکوردها می توان به منابع کلیدی (Gulati and Padgett, 2003)، (Arnold et al., 1998) ، (Nevzorov, 2001) و (Ahsanullah, 1995) مراجعه کرد. از آنجا که روش‌های ریاضی و توزیع‌های آماری، نتایجی دقیق در محاسبات اقلیمی و فرایندهای هیدرولوژیکی ارائه می‌دهند، آگاهی از توزیع احتمال، زمینه مناسبی برای برنامه‌ریزی فراهم می‌آورد. رفتار پدیده‌های تصادفی انگیزه‌ای بود که الگوهای آماری معرفی شوند تا بتوان این گونه پدیده‌ها را در چارچوبی خاص بیان وکنترل کرد. در این پژوهش $\{X_i; i \geq 1\}$ دنباله‌ای نامتناهی از متغیرهای تصادفی پیوسته، مستقل و هم‌توزیع دارای توزیع نمایی تعمیم یافته با تابع چگالی احتمال

$$f(x;\theta) = \theta e^{-x}(1-e^{-x})^{\theta-1}, \qquad x > 0, \qquad \theta > 0 \qquad (1)$$

و تابع توزیع تجمعی

$$F(x;\theta) = (1-e^{-x})^{\theta}, \qquad x > 0, \qquad \theta > 0 \qquad (2)$$

می‌باشد. این یکی از توزیع های مهم در خانواده نمایی است. الگوی آماری نمایی بیشتر در تخمین زدن مدت زمان لازم برای رخداد یک پیشامد خاص استفاده می شود. برای مثال، مدت زمان لازم (ازهمین حالا) تا رخداد یک زمین لرزه، آغاز یک جنگ، دریافت یک تماس تلفنی اشتباه و ... متغیرهای تصادفی با توزیع نمایی می باشند. تغییر پارامترهای اقلیمی به اشکال مختلفی قابل ردیابی است. در بسیاری مواقع این تغییرات نهان و از طریق مطالعه در شکل توزیع فراوانی نمایان می‌شوند. بنابراین بررسی شکل توزیع فراوانی عناصر اقلیمی، نکات بسیار مهمی در خصوص رفتار بلند مدت این عناصر نشان می‌دهد و به لحاظ علمی و عملی از اهمیت شایان توجهی برخوردار است. به دلیل خاصیت بی حافظگی توزیع نمایی و سادگی در محاسبات ریاضی مربوط به آن در تغییرات اقلیمی، این توزیع مورد توجه محققان بسیاری قرار گرفته است. از آن جمله می‌توان به (حسین عساکره، ۲۰۱۸)، (حسام سیدکابلی و همکاران، ۲۰۱۲) و (ذبیح الله خانی تملیه و همکاران، ۲۰۲۰) اشاره کرد. استفاده از این توزیع در برازش داده ها و پس از برآورد پارامترهای توزیع، در حوزه‌های نظیر سیلاب، خشکسالی، مدیریت حوضه آبریز، کشاورزی و ...ابزار مناسبی در اختیار مدیران حوزه های مختلف قرار می دهد، تا با در نظر گرفتن این برآوردها، سیاستهای آینده را در جهت بهینه نمودن هزینه ها و بهره وری حداکثر، طرح ریزی کنند. منظور از برآورد، پیدا کردن برآوردگر یا تابعی از نمونه تصادفی است که مقدار مشاهده شده آن به مقدار واقعی پارامتر مجهول ولی ثابت جامعه نزدیک‌تر باشد. در یک طرح نمونه گیری، مساله برآورد و تحلیل داده ها از دو رهیافت کلاسیک و بیزی بهره گرفته می‌شود. در رهیافت کلاسیک، پارامتر مورد بررسی ثابت اما نامعلوم بوده و بر اساس نمونه ای تصادفی در مورد آن تصمیم گیری می شود. این در حالی است که در رهیافت بیزی، برخلاف نگرش کلاسیک، پارامتر مورد بررسی را می توان به چشم یک متغیر تصادفی نگاه کرد. با توجه به این که پارامتر جامعه ممکن است تحت تاثیر عوامل آزمایش تغییر کند لذا همیشه نمی‌توان آنرا مقداری ثابت تلقی کرد. در چنین شرایطی تغییرات آن توسط یک توزیع احتمال، تحت عنوان توزیع پیشین توصیف می شود. این توزیع پیشین به عنوان برآیند اطلاعات پژوهش‌گر درباره پارامتر، بر اساس تجربیات قبلی آزمایشگر تعیین شده است. بنابراین استفاده از اطلاعات پیشین راجع به پارامتر، قبل از به دست آوردن اطلاعات نمونه ای که در دسترس است، اهمیت پیدا می کند. آنچه حائز اهمیت است، توانایی انتخاب یک روش مناسب نمونه‌گیری بر اساس طرح‌های مختلف است که موجب به حداکثر رساندن دقت خصیصه‌های جامعه ، کاهش هزینه‌ها و اعتمادپذیری بیشتر می‌شود. در این پژوهش، طرحی پرکاربرد و کم هزینه تحت عنوان طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی (RRSS) را در نظرگرفته ایم که نخستین بار توسط صالحی و احمدی (۲۰۱٤) معرفی شد. تحقیق بر اساس طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی تاکنون مورد توجه محققین بسیاری واقع شده است. در این میان می توان به (صالحی و احمدی، ۲۰۱۵)، (صالحی و همکاران،

۲۰۱٦)، (اسکندرزاده و همکاران، ۲۰۱٦)، (پاوول و توماس، ۲۰۱۷)، (صفریان و همکاران، ۲۰۱۹) و (گل زاده و همکاران، ١٤٠٠) اشاره نمود.

داده‌ها و روش‌شناسی

در این بخش، پس از معرفی طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی، پارامتر توزیع نمایی تعمیم یافته به روش‌های ماکسیمم درستنمایی و بیزی تحت تابع زیان توان دوم خطا برآورد می شود. سپس با استفاده از مقادیر رکوردی حاصل از طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی، برآوردگرها با استفاده از معیاری آماری مورد مقایسه قرار می گیرد.

معرفی طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی

روش‌های مختلفی برای به‌دست آوردن مشاهدات وجود دارد. در مسائل نمونه گیری، محقق همواره با مساله انتخاب روش مناسب جمع آوری اطلاعات روبرو است، به طوری که دقت را برای برآورد مشخصه های یک جامعه با حداقل هزینه تامین کند و هنگامی که بودجه معینی برای نمونه گیری دارد، حداکثر دقت را برای مشخصه‌های مجهول فراهم نماید. با بسط علم آمار و ارتباط تمام شاخه های آمار با نمونه گیری، این شاخه آمار برای برآورد مشخصه های جامعه و بررسی ویژگی های این برآورد به سرعت گسترش یافت. لذا در شرایطی که اندازه گیری واحدهای جامعه مشکل یا پرهزینه باشد از طرح نمونه گیری مجموعه رتبه‌دار رکوردی استفاده می شود که در این بخش با چگونگی استفاده از این طرح نمونه‌گیری به اختصار آشنا می‌شویم:

فرض کنید، n دنباله‌ی مستقل از متغیرهای تصادفی در اختیار است و نمونه‌گیری در هر دنباله به نحوی است که تنها رکوردهای پایینی دنباله ها ثبت می‌شوند. اولین رکورد پایین در دنباله اول و دومین رکورد پایین در دنباله دوم و به همین ترتیب nامین رکورد پایین در دنباله nام انتخاب می شوند. به عبارت دیگر، برای تحلیل آماری در دنباله iام، تنها مقدار i امین رکورد پایین حفظ می‌شود و iامین نمونه‌گیری دنباله‌ای زمانی متوقف می‌شود که iامین رکورد پایین ثبت شود. به این طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی پایین گفته می‌شود. مراحل استخراج نمونه‌ای با حجم n به روش نمونه‌گیری مجموعه رتبه‌دار رکوردی پایین را می توان به کمک دیاگرام زیر دقیق‌تر نشان داد

$$\begin{array}{llllll}
1: & R_{(1)1} & & & \rightarrow & \boldsymbol{R_{1,1}} = R_{(1)1} \\
2: & R_{(1)2} & R_{(2)2} & & \rightarrow & \boldsymbol{R_{2,2}} = R_{(2)2} \\
\dots & & \dots & & \rightarrow & \dots \quad \dots \\
n: & R_{(1)n} & R_{(2)n} & R_{(n)n} & \rightarrow & \boldsymbol{R_{n,n}} = R_{(n)n}
\end{array} \qquad (3)$$

متغیر $R_{(i)j}$ نشان دهنده iامین رکورد پایین در دنباله jام است. در واقع به بردار $R = (\boldsymbol{R_{1,1}}, \boldsymbol{R_{2,2}}, \dots, \boldsymbol{R_{n,n}})$ یک نمونه مجموعه رتبه دار رکوردی پایین با حجم n گفته می‌شود. نکته قابل توجه این است که مشاهدات حاصل از این طرح مستقل و غیر همتوزیع هستند. فرض کنید $\boldsymbol{R_{i,i}}$ ها رکورد پایین باشند، آن‌گاه تابع چگالی احتمال iامین رکورد پایین و تابع چگالی احتمال توام n رکورد پایین به‌ترتیب به فرم زیر هستند:

$$f_{i,i}(x) = \frac{\{-\log F(x)\}^{i-1}}{(i-1)!} f(x) \qquad (4)$$

$$f_R(r;\theta) = \prod_{i=1}^{n} \frac{\{-\log F(r_{i,i};\theta)\}^{i-1}}{(i-1)!} f(r_{i,i};\theta) \quad , \quad \theta \in \Theta \qquad (5)$$

که در آن‌ها، $r = (\boldsymbol{r_{1,1}}, \boldsymbol{r_{2,2}}, \dots, \boldsymbol{r_{n,n}})$ مقدار مشاهده شده بردار $R = (\boldsymbol{R_{1,1}}, \boldsymbol{R_{2,2}}, \dots, \boldsymbol{R_{n,n}})$ و Θ فضای پارامتر است. برای جزئیات بیشتر به (Arnold et al., 1998) مراجعه شود. در سراسر این مقاله منظور از رکورد، رکورد پایین است.

برآورد پارامتر مدل با استفاده از طرح نمونه گیری RRSS

در این بخش بر روی مساله برآورد پارامتر مدل با بکارگیری داده‌های رکوردی حاصل از طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی متمرکز می‌شویم. ابتدا به سراغ برآورد پارامتر مدل به روش ماکسیمم درستنمایی می رویم.

برآورد ماکسیمم درستنمایی

در میان برآوردهای کلاسیک، برآورد به روش ماکسیمم درستنمایی حائز اهمیت می‌باشد که اولین بار توسط گوس در سال ۱۸۲۱ به کار گرفته شد. برآورد ماکسیمم درستنمایی عبارت است از دستورالعملی برای به دست آوردن برآوردی به نام برآورد ماکسیمم درستنمایی که آن را به اختصار با (MLE) نشان می‌دهیم و مبتنی بر یک تابع آماری مهم به نام تابع درستنمایی است. تابع درستنمایی نمونه را به صورت تابع چگالی توام مشاهدات تعریف می‌کنیم که تابعی از پارامتر توزیع در نظر گرفته می‌شود. اصل ماکسیمم درستنمایی بیان می‌کند که برآورد ماکسیمم درستنمایی پارامتر، مقداری است که تابع درستنمایی را روی فضای پارامتر ماکسیمم مطلق می کند. یکی از روش‌های به دست آوردن برآورد ماکسیمم درستنمایی، استفاده از تکنیک مشتق گیری از لگاریتم تابع درستنمایی می‌باشد. اگر نماد $L(\theta/r)$ تابع درستنمایی مشاهدات استخراج شده بر اساس طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی باشد، با جایگذاری روابط (۱) و (۲) در رابطه (۵) داریم.

$$L(\theta/r) = K\theta^N e^{\theta \sum_{i=1}^{n}(1-e^{-r_{i,i}})} \qquad (٦)$$

که در آن $N = \frac{n(n+1)}{2}$ و K ضریب تناسب است و به پارامتر θ بستگی ندارد، یعنی می‌توان نوشت:

$$L(\theta/r) \propto \theta^N e^{\theta \sum_{i=1}^{n}(1-e^{-r_{i,i}})} \qquad (٧)$$

با توجه به رابطه (۷)، لگاریتم تابع درستنمایی برابر است با

(۸)

$$logL(\theta/r) \propto N\, log\theta + \theta \sum_{i=1}^{n}(1-e^{-r_{i,i}})$$

برای برآورد پارامتر مدل از لگاریتم تابع درستنمایی مشاهدات نسبت به θ مشتق می گیریم و مساوی صفر قرار می دهیم.

(۹)

$$\frac{\partial logL(\theta/r)}{\partial\theta} = \frac{N}{\theta} + \sum_{i=1}^{n}(1-e^{-r_{i,i}}) = 0$$

از حل معادله (۹)، برآورد ماکسیمم درستنمایی θ برابر است با

$$\hat{\theta}_{MLE} = \frac{N}{T} \qquad (١٠)$$

که در آن $T = -\sum_{i=1}^{n}(1-e^{-r_{i,i}})$ است. می‌توان نشان داد متغیر تصادفی T دارای توزیع گاما با پارامترهای N و θ می‌باشد. اگر $\hat{\theta}_{MLE}$ برآورد ماکسیمم درستنمایی پارامتر θ جامعه باشد آنگاه معیار میانگین توان دوم خطا (MSE) برای ارزیابی آن به صورت زیر تعریف می‌شود

(۱۱)

$$MSE(\hat{\theta}_{MLE}, \theta) = \mathrm{Var}(\hat{\theta}_{MLE}) + (\mathrm{E}(\hat{\theta}_{MLE}) - \theta)^2$$

که در آن نمادهای Var ، E و $(\mathrm{E}(\hat{\theta}_{MLE}) - \theta)^2$ به ترتیب نشان دهنده واریانس، امید ریاضی (میانگین) و توان دوم اریبی برآورد می‌باشند. این رابطه نشان می‌دهد که معیار MSE در صورتی کوچک می باشد که واریانس و توان دوم اریبی برآورد، هر دو کوچک باشند. از روابط (۱۰) و (۱۱) داریم

$$MSE(\hat{\theta}_{MLE}, \theta) = \mathrm{Var}(\frac{N}{T}) + (\mathrm{E}(\frac{N}{T}) - \theta)^2) = \frac{(N\theta)^2}{(N-1)^2(N-2)} + (\frac{\theta}{N-1})^2.$$

برآورد بیزی

در مساله برآورد، یک برآوردگرخوب، برآوردگری است که میزان خطا در آن با احتمال یک نزدیک صفر باشد. داشتن یک اندازه دقت برآوردیابی، برای برآوردکننده بسیار مهم است. تابع زیان $L(\theta, \delta)$، نمایانگر مقدار زیانی است که آماردان متحمل می‌شود، وقتی که θ در وضعیت درست بوده و او عمل δ را اتخاذ می‌کند. ایده ال این است که برای هر تابع زیان داده شده، علاقه‌مند به پیدا کردن برآوردگری باشیم که متوسط تابع زیان را مینیمم کند. $R(\theta, \delta)$ تابع مخاطره است که نمایانگر میانگین مقدار تابع زیان است و به صورت

$R(\theta,\delta) = E(L(\theta,\delta))$ تعریف می شود. رایج ترین تابع زیان در برآورد پارامتر، تابع زیان توان دوم خطا، $L(\theta,\hat\theta) = (\hat\theta-\theta)^2$، است که در آن $\hat\theta$ برآورد θ است. برآورد معیار مخاطره تحت تابع زیان توان دوم خطا در مثلا ۱۰۰۰۰ بار تکرار شبیه سازی به صورت زیر

$$R(\theta,\hat\theta) = \frac{1}{10000}\sum_{i=1}^{10000}(\hat\theta_i-\theta)^2$$

تعریف می‌شود. اکنون فرض کنید θ دارای تابع چگالی احتمال پیشین گاما به فرم زیر

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \quad (۱۳)$$

باشد که در آن α و β پارامترهای توزیع پیشین با مقادیری مثبت و معلوم اند و $\Gamma(.)$ تابع گامای کامل است. استفاده از روش بیزی، این امکان را به محقق می‌دهد تا اطلاعات پیشین را در تحلیل وارد کند. در این بخش برآورد بیزی پارامتر θ تحت تابع زیان توان دوم خطا و با در نظر گرفتن تابع چگالی احتمال پیشین گاما به دست می‌آید. از مولفه‌های مهم رهیافت بیزی تابع چگالی پسین θ است که متناسب با حاصلضرب تابع چگالی احتمال پیشین و تابع درستنمایی مشاهدات به فرم زیر

$$\Pi(\theta/r) \propto L(\theta/r)\pi(\theta) \quad (۱۴)$$

است. با بکارگیری روابط (۷) و (۱۳) در رابطه (۱۴)، تابع چگالی پسین θ، تابعی گاما با پارامترهای $N+\alpha$ و $\beta - \sum_{i=1}^{n} log(1-e^{-r_{i,i}})$ به فرم

(۱۵)

$$\Pi(\theta/r) \propto \theta^{N+\alpha-1}e^{-\theta(\beta-\sum_{i=1}^{n}\log(1-e^{-r_{i,i}}))}$$

می‌شود. تحت تابع زیان توان دوم خطا، برآورد بیزی پارامتر θ می‌شود، میانگین تابع چگالی پسین θ یعنی $E(\theta/r)$. به کمک رابطه (15) برآورد بیزی θ می‌شود.

$$\hat\theta_{BS} = E(\theta/r) = \frac{N+\alpha}{N+\alpha+\beta-\sum_{i=1}^{n} log(1-e^{-r_{i,i}})}$$

نتایج و بحث

در این بخش ابتدا با استفاده از شبیه‌سازی مونت کارلو و سپس به کمک داده‌ای واقعی، عملکرد طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی مورد ارزیابی قرار می‌گیرد.

شبیه‌سازی مونت کارلو

در این بخش به کمک شبیه سازی مونت کارلو به مقایسه عملکرد روش ماکسیمم درستنمایی و روش بیزی پرداختیم. برای این منظور، مقادیر $\alpha = 2$ و $\beta = 1$ را برای پارامترهای توزیع پیشین در رابطه (13)، در نظرگرفتیم. مقدار واقعی پارامتر θ را ۱/۳۸۳ و اندازه‌های ۵،۳،۲ و ۸ را برای حجم نمونه n اختیار کرده‌ایم. به کمک نمونه های شبیه سازی شده با تکرار ۱۰۰۰۰، پارامتر مدل برآورد شده است. برآورد ماکسیمم درستنمایی، برآورد بیزی، معیار میانگین توان دوم خطا برای برآوردگر ماکسیمم درستنمایی و معیار مخاطره برای برآوردگر بیزی در جدول ۱ زیرآمده است. با توجه به نتایج جدول ۱ می‌توان گفت:

۱- با افزایش حجم نمونه، مقادیر مخاطره بیزی و میانگین توان دوم خطا برآوردگرها کاهش می‌یابد.

۲- با افزایش حجم نمونه، مقادیر برآورد شده پارامتر θ به مقدار واقعی ۱/۳۸۳ نزدیک می‌شود.

۳- برآوردگر بیزی دقت بالاتری در مقایسه با برآوردگر ماکسیمم درستنمایی از خود نشان می‌دهد.

جدول ۱- مخاطره بیزی و میانگین توان دوم خطای برآوردگر θ به کمک شبیه سازی با تکرار ۱۰۰۰۰

حجم نمونه ↓	برآورد بیزی	مخاطره بیزی	حجم نمونه ↓	برآورد ماکسیمم درستنمایی	میانگین توان دوم خطا
۲	۱/۷۹۶	۰/۶۳۴	۲	۲/۰۳۳	٤/۳٤٤
۳	۱/٦٦٤	۰/٤۲٦	۳	۱/٦٥۷	۰/۸۹۷
۵	۱/٥۱۷	۰/۱٦٦	٥	۱/٤۸۱	۰/۱۸۸
۸	۱/٤٤۰	۰/۰٦۱	۸	۱/٤۱۹	۰/۰٦۲

داده‌های واقعی درجه حرارت

در این بخش، برای تشریح کارایی برآوردگرهای به دست آمده در بخش‌های قبل، داده‌های میزان درجه حرارت در مقیاس زمانی روزانه از ماه دی طی سال‌های ۱۳۹۷، ۱۳۹۸، ۱۳۹۹ و ۱۴۰۰ که از ایستگاه هواشناسی شهر ساری در استان مازندران به دست آمده‌اند، استفاده شد. در برازش مدل، p – مقدار آزمون کولموگروف– اسمیرنوف برای بررسی نیکویی برازش توزیع نمایی تعمیم یافته بر پایه این داده‌ها برابر ۰/۷۱۸ است که گویای مناسب بودن برازش است. با استفاده از نرم افزار آماری R می‌توان دید این داده‌ها از توزیع نمایی تعمیم یافته با پارامتر ۳/۱ پیروی می کند. بردار زیر

$$r = (\boldsymbol{r_{1,1}} = 3/6\ \boldsymbol{r_{2,2}} = 0/4, \boldsymbol{r_{3,3}} = 2/8, \boldsymbol{r_{4,4}} = 0/1)$$

بردار رکوردهای پایین به‌دست آمده از داده‌های درجه حرارت براساس طرح RRSS می‌باشد. پر واضح است که برای n=4 به راحتی می‌توان دید، $T=-\sum_{i=1}^{4}\log(1-e^{-r_{i,i}}) = 1/543$. براساس توضیحات فوق، نتایج برآورد پارامتر به روش ماکسیمم درستنمایی و روش بیزی، در جدول ۲ آمده است.

جدول ۲- برآورد پارامتر و مخاطره برآوردگر برای داده‌های درجه حرارت (مقدار واقعی پارامتر ۳/۱ است)

روش	مقدار برآورد پارامتر	مخاطره
ماکسیمم درستنمایی	٦/٤٨	۱/٦۲
بیزی	٤/۱٥	۰/۹۷

این مثال عددی، بار دیگر نتایج حاصل از مطالعه شبیه‌سازی در بخش قبل را تایید می کند. نکته ای که در جدول ۲ دوباره روی آن تاکید می شود، کوچکتر بودن مخاطره برآوردگر به روش بیزی نسبت به مخاطره برآوردگر به روش ماکسیمم درستنمایی است.

نتیجه‌گیری

با توجه به اینکه تغییرات اقلیمی به صورت تصادفی می‌باشند، بنابراین اساس تجزیه و تحلیل پدیده‌های اقلیمی، استفاده از تکنیک‌ها و فنون آماری است. وجود داده‌های آماری مناسب در منطقه مورد مطالعه، در پردازش مسائل و دریافت خروجی‌های قابل اعتماد امری بسیار مهم و تاثیر گذارمی باشد. از آن جا که میزان درجه حرارت در ایستگاه‌های هواشناسی استاندارد اندازه گیری می شود، یکی از توصیف کننده های اصلی وضعیت محیط زمین است. بنابراین برآورد پارامتر مدل بر اساس داده‌های درجه حرارت در هر منطقه، یکی از پیش نیازهای مهم برای برنامه ریزی کشاورزی و مدیریت منابع آب می باشد. درجه حرارت یکی از اصلی ترین عناصر اقلیمی است که تأثیر مهمی بر جنگل‌ها، مراتع، اراضی کشاورزی و در نهایت زندگی افراد منطقه دارد. در مسائل آماری به دست آوردن مشاهدات از طریق فرایندی که بتوان با صحت بیشتر در مورد خصیصه های جامعه اظهار نظر کرد یکی از اصول اساسی است. در شرایطی که به دلایلی چون هزینه و زمان، جمع آوری داده‌ها دشوار باشد روش نمونه‌گیری مجموعه رتبه‌دار رکوردی روش مناسبی است. در این پژوهش برای تخمین پارامتر مدل نمایی تعمیم یافته به کمک نرم افزار آماری R ، از رکوردهای درجه حرارت در دوره زمانی روزانه از ماه دی طی سال‌های ۱۳۹۷ تا ۱۴۰۰ مربوط به ایستگاه هواشناسی شهر ساری در استان مازندران که از فرایند طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی حاصل شد، استفاده شده است. نتایج عددی نشان داد، برآوردگر بیز حاصل از طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی کاراتر و دارای مخاطره کوچکتری نسبت به برآوردگر ماکسیمم درستنمائی است. به طوری که با افزایش رکوردها هم، به عنوان نمونه های به دست آمده، برآورد پارامترها افزایش یافته و به مقدار واقعیشان نزدیک تر می شود. پیشنهاد می گردد، در تغییرات اقلیمی اگر شرایط برای انجام طرح نمونه‌گیری مجموعه رتبه‌دار رکوردی مهیا باشد، از این طرح نمونه‌گیری برای برآورد پارامترهای مدل استفاده شود.

سپاس گزاری

نویسنده مقاله از زحمات سردبیر محترم مجله و دیگر همکاران گرامی کمال تشکر را دارد.

منابع

1. Aghelpour, P., and Nadi, M., 2018, Assessing the accuracy of SARIMA model in modeling and long-term forecast of average monthly temperature in different climates of Iran, Climatological Research, 35: 113-126.
2. Ahsanullah, M., 1995, Record Statistics, Nova Science Publishers Inc, Commack, New York.
3. Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N., 1998, Records, John Wiley & Sons, New York.
4. Asakereh, H., 2012, Frequency Distribution Change of Extreme precipitation in Zanjan City, Geography and Environmental Planning Journal, 1: 13-18.
5. Chandler, K. N., 1952, the distribution and frequency of record values, Journal of Royal Statistical Society, 14: 220-228.
6. Eskandarza deh, M., Tahmasebi, S., and Afshari, M., 2016, Information measures for Record ranked set samples, Ciencia e Natura, 38 :554-563.
7. Golzade -۷ Gervi, E., Nasiri, P., and Salehi, M., 2021, Comparison of Empirical Bayesian Estimations and Predictions Based on Record Ranked Set Sampling Scheme with inverse Sampling Scheme, Journal of Statistical Sciences, 1: 193-218.
8. Golzade -۸ Gervi, E., Nasiri, P., and Salehi, M., 2021, An overview of Bayesian prediction of future record statistics using upper record ranked set sampling scheme, Int. J. Nonlinear Anal. Appl., 1: 493-507.
9. Nevzorov, V., 2001, Records, Mathematical Theory, Translation of Mathematical Monographs No, 194, American Mathematical Society, Providence, RI.
10. Paul, J., and Thomas, PY., 2017, Concomitant Record ranked set sampling,
11. Communications in Statistics-Theory and Methods, 46: 9518-9540.
12. Razmkhah, M., Ahmadi, J., and Khatib Astaneh, B., 2019, Comparison of Two Sampling Schemes for Extracting Record Data with Regard to Fisher Information, Journal of Statistical Sciences, 1: 19-44.
13. Safaryian, A., Arashi, M., and Arabi, R., 2019, Improved Estimators for stress-strength Reliability using Record ranked set sampling scheme, Communications in statistics-simulation and computation, DOI: 10.1080/03610918.2018.1468451.
14. Salehi, M., and Ahmadi, J., 2014, Record ranked set sampling scheme, Metron, 72: 351-365.
15. Salehi, M., and Ahmadi, J., 2015, Estimation of stress-strength Reliability using Record ranked set sampling scheme from the Exponential distribution, Filomat, 29: 1149-1162.
16. Salehi, M., Ahmadi, J., and Balakrishnan, N., 2015, Prediction of order Statistics and Record values based on ordered Ranked set sampling, Statistical Computation and Simulation, 85: 77-88.
17. Salehi, M., Ahmadi, J., and Dey, S., 2016, Comparison of two sampling schemes for generating Record-breaking Data from the Proportional hazard rate models, Communications in Statistics-Theory and Methods, 45: 3721-3733.
18. Seyyed Kaboli, H., AkhodAli, A.M., Masah Bavani, A.R., and Radmanesh, A.R., 2012, A Downscaling Model Based on K-nearest neighbor (K-NN) Non-parametric Method, Journal of Water and Soil, 4: 779-808.
19. SeyyedNezhad Golkhatmi, N., Abbasi N., and Rezaee-Pazhand, H., 2022, Forecasting of 125 years annual precipitation time series in Mashhad , Climatological Research, 50: 83-94.
20. Shahinejad, B., and Dehghani, R., 2016, Application of wavelet neural network for estimation of means daily temperature in Sari area, Climatological Research, 27: 75-86.
21. Zabihollah Khani, T., Rezaei, H., and Mirabbasi Najafabadi, R., 2020, Application of the Nested Copula Functions for Analysis of Four variate of Meteorological Droughts (Case Study: West of Iran), Climatological Research, 1: 93-111.