

مجله علوم آماری، پاییز و زمستان ۱۳۹۴

جلد ۹، شماره ۲، ص ۲۰۷-۲۲۵

برآورد پارامترهای توزیع تعمیم یافته مقدار کرانگین ماکسیمال و مدل بندی کرانگین سرعت باد شهر زاهدان

فروش عاشوری، ملیحه ابراهیم پور، ابوالقاسم بزرگ نیا
گروه آمار، دانشگاه آزاد اسلامی واحد مشهد

تاریخ دریافت: ۱۳۹۳/۴/۱۱ تاریخ آخرین بازنگری: ۱۳۹۳/۱۲/۲۶

چکیده: رفتار مقادیر کرانگین یک مجموعه داده، به ویژه در پدیده‌های طبیعی مثل دبی رودخانه، سرعت باد، میزان بارندگی و در بسیاری از علوم کاربردی دیگر مانند مطالعات قابلیت اعتماد و تحلیل پیشامدهای کرانگین محیطی کاربرد دارد. اگر بتوان رفتار این گونه داده‌ها را مدل بندی کرد چگونگی رفتار آن‌ها در آینده قابل پیش بینی خواهد بود. معمولاً تحلیل مقادیر کرانگین براساس توزیع تعمیم یافته مقدار کرانگین ماکسیمال انجام می شود. در این مقاله، چهار روش برای برآورد پارامترهای این توزیع و یک روش برای برآورد چندک‌های آن ارائه شده، سپس یک روش برآورد فاصله‌ای و آزمون نیکویی برازش ذکر شده است. برای داده‌های بیشترین سرعت باد شهر زاهدان انواع برآوردها و برآورد فاصله‌ای محاسبه و با هم مقایسه شده‌اند. در آخر، دوره بازگشت بیشترین سرعت باد شهر زاهدان محاسبه شده است.

واژه‌های کلیدی: توزیع تعمیم یافته مقدار کرانگین ماکسیمال، دامنه جذب، گشتاورهای احتمال موزون، روش صدک، آزمون نیکویی برازش، آزمون نسبت درستنمایی.

آدرس الکترونیک مسئول مقاله: فروش عاشوری، fashoori.2000@gmail.com

کد موضوع بندی ریاضی (۲۰۱۰): ۶۰G۷۰

اولین بار فیشر و تیپت (۱۹۲۸) خانواده توابع توزیع مقادیر کرانگین^۱ را معرفی کردند و اولین بار جنکینسن (۱۹۵۵) آن را بررسی کرد. زمانی که وی به دنبال یک مدل برای ماکسیمم سالانه یا روزانه پدیده‌های طبیعی از قبیل ماکسیمم سطح آب یا ارتفاع رودخانه، میزان بارش و مقدار آلودگی هوا بود، به این توزیع توجه کرد.

در بسیاری از مسائل کاربردی، گاهی مطالعه مقادیر کرانگین یعنی مقادیر واقع در دم توزیع برخی از پدیده‌ها، مانند امواج دریا، درجه حرارت، زمین لرزه، سیل، غرامت‌های بیمه، غلظت آلاینده‌های جوی، استحکام مواد و غیره مورد نظر می‌باشد.

با توجه به تغییرات اقلیمی در سال‌های اخیر به علت پیامدهای اقتصادی، اجتماعی و خسارت‌های مربوط به رویدادهای حدی جوی، در بعضی از شاخه‌های علوم نظیر هواشناسی مقادیر ماکسیمم از اهمیت زیادی برخوردارند؛ برای مثال، وقتی یک مجموعه از داده‌های مربوط به وزش باد را بررسی می‌کنیم آنچه مورد نظر ما است تندبادها و بادهای سهمگین هستند که باعث ایجاد خسارت به منازل مسکونی، مزارع و غیره می‌شوند. مواجه شدن با مسائلی مشابه موارد ذکر شده موجب تسریع مدل‌بندی مقادیر کرانگین در سال‌های اخیر شده است. از این رو، مدل‌بندی و پیش‌بینی در مورد این گونه پیشامدهای نادر از اهمیت ویژه‌ای برخوردار است. همچنین، نظریه مقادیر کرانگین^۲ و توزیع‌های مرتبط با آن نقش بسیار مهمی در نظریه آمار دارد. این نظریه بیش از همه با رفتارهای تصادفی مقادیر ماکسیمم و مینیمم نمونه‌های مستقل و هم‌توزیع (i.i.d.) سروکار دارد. ویژگی‌های توزیعی کرانگین‌ها و آماره‌های مرتب که از یک مقدار آستانه‌ای^۳ بالاتر یا پایین‌تر هستند با بررسی دم‌های بالا یا پایین توزیع تعیین می‌شوند. قبل از هر چیز چند تعریف اساسی بیان می‌شود:

^۱ Extreme Values

^۲ Extreme values theory

^۳ Threshold value

تعریف ۱: فرض کنید X_1, \dots, X_n یک نمونه تصادفی با توزیع $F(x)$ باشد. توزیع حدی ناتباهیده^۴ برای آماره مرتب $X_{(n)}$ به صورت

$$\lim_{n \rightarrow \infty} H_n(a_n + b_n x) = \lim_{n \rightarrow \infty} [F(a_n + b_n x)]^n = H(x), \quad \forall x \quad (1)$$

تعریف می شود، که در آن a_n و b_n دنباله ای از اعداد حقیقی هستند و تابع $H_n(x)$ به صورت $H_n(x) = P(X_n \leq x) = [F(x)]^n$ است. اگر رابطه (۱) حداقل برای یک جفت از دنباله های $\{a_n\}$ و $\{b_n > 0\}$ برقرار باشد، آنگاه توزیع $F(x)$ به دامنه جذب ماکسیمال^۵ $H(x)$ متعلق است. توجه کنید که فقط یک خانواده پارامتری به عنوان توزیع حدی برای توزیع ماکسیمم ها وجود دارد (گالامبوس، ۱۹۸۷؛ تیاگو دی الیوریا، ۱۹۵۸).

قضیه ۱ (کاستیلو و همکاران، ۱۹۸۹): شرط لازم و کافی برای توزیع پیوسته $F(x)$ که به دامنه جذب ماکسیمال $H_\kappa(x)$ متعلق باشد این است که

$$\lim_{\varepsilon \rightarrow \infty} \frac{F^{-1}(1 - \varepsilon) - F^{-1}(1 - 2\varepsilon)}{F^{-1}(1 - 2\varepsilon) - F^{-1}(1 - 4\varepsilon)} = 2^{-\kappa}$$

توجه شود اگر $\kappa > 0$ به دامنه جذب و ایبول ماکسیمال متعلق است، اگر $\kappa = 0$ به دامنه جذب گامبل ماکسیمال متعلق است و اگر $\kappa < 0$ به دامنه جذب فرشه ماکسیمال متعلق است.

در بخش ۲ توزیع تعمیم یافته مقدار کرانگین ماکسیمال^۶ (GEV_M) و حالت های خاص آن معرفی می شود. در بخش ۳، پارامترهای این توزیع به چهار روش برآورد می شوند. در بخش ۴، آزمون های نیکویی برازش بیان می شوند. در بخش ۵، مقادیر کرانگین بیشترین سرعت باد شهر زاهدان مورد تحلیل قرار می گیرد. نهایتاً بحث و نتیجه گیری در بخش ۶ ارائه می شود.

^۴ Nondegenerate limit distribution

^۵ Maximum Domain of Attraction

^۶ Maximal Generalized Extreme Value

۲ توزیع تعمیم یافته مقدار کرانگین ماکسیمال

معمولاً برای نمایش توزیع‌های مقدار کرانگین، از یک شکل کلی به نام توزیع تعمیم یافته مقدار کرانگین که توسط جنکینسن (۱۹۵۵) ارائه شده است، استفاده می‌شود. در این حالت یک پارامتر اضافی که به پارامتر شکل معروف است، وارد توزیع می‌شود. بنابراین، تابع چگالی توزیع تعمیم یافته مقدار کرانگین ماکسیمال $GEV_M(\lambda, \delta, \kappa)$ که تنها توزیع حدی برای مقادیر ماکسیمم در نمونه‌های i.i.d. است، به صورت

$$f(x) = \begin{cases} \exp[-[1 - \kappa(\frac{x-\lambda}{\delta})]^{\frac{1}{\kappa}}][[1 - \kappa(\frac{x-\lambda}{\delta})]^{\frac{1}{\kappa}} - 1]^{\frac{1}{\delta}}, & 1 - \kappa(\frac{x-\lambda}{\delta}) \geq 0, \kappa \neq 0 \\ \exp[-\exp(\frac{\lambda-x}{\delta})] \exp(\frac{\lambda-x}{\delta})^{\frac{1}{\delta}}, & x \in R, \kappa = 0 \end{cases}$$

تعریف می‌شود، که در آن پارامتر مکان $\lambda \in R$ پارامتر مکان، $\delta > 0$ پارامتر مقیاس و κ پارامتر شکل است به طوری که برای $x \leq \lambda + \frac{\delta}{\kappa}$ ، $\kappa > 0$ و برای $x \geq \lambda + \frac{\delta}{\kappa}$ ، $\kappa \leq 0$ می‌باشد. پارامتر شکل κ رفتار دم توزیع را توصیف می‌کند و برحسب این که این پارامتر مثبت، منفی یا صفر باشد سه توزیع مقدار کرانگین وایبول، گامبل و فرشه ماکسیمال نتیجه می‌شود. حالت‌های خاص توزیع GEV_M شامل توزیع‌های گامبل، وایبول معکوس و فرشه است. چندانک p ام نیز به صورت

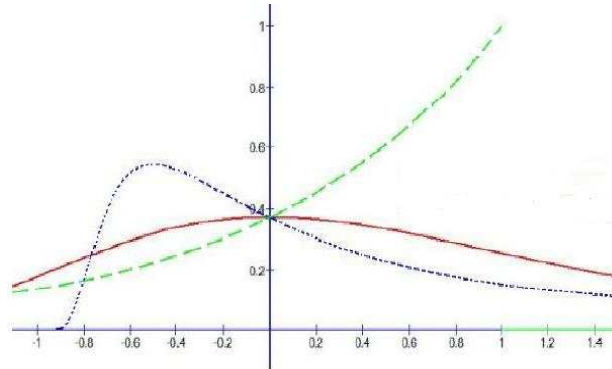
$$x_p = \begin{cases} \lambda + \frac{\delta}{\kappa} [1 - (-\log p)^{\kappa}], & \kappa \neq 0 \\ \lambda - \delta \log(-\log p), & \kappa = 0 \end{cases}$$

تعریف می‌شود. شکل ۱ نمودار تابع چگالی GEV_M را برای مقادیر مختلف پارامتر شکل κ نشان می‌دهد.

۳ برآورد پارامترهای توزیع تعمیم یافته مقدار کرانگین ماکسیمال

در این بخش، چهار روش متفاوت برای برآورد پارامترهای توزیع GEV_M شامل روش‌های برآورد ماکسیمم درست‌نمایی^۷ (MLE)، گشتاورهای احتمال

^۷ Maximum Likelihood Estimation



شکل ۱: چگالی‌های گامبل $\kappa = 0$ (خط ممتد)، وایبول معکوس $\kappa = -1$ (خط چین)، فرشه $\kappa = 1$ (نقطه چین)

موزون^۸ (PWM)، صدک^۹ (EP) و کمترین توان‌های دوم چندک^{۱۰} (QLS) بیان می‌شوند.

۱.۳ روش ماکسیمم درستنمایی

لگاریتم تابع درستنمایی به صورت

$$\ell(\lambda, \delta, \kappa) = \begin{cases} -n \log \delta - (1 - \kappa) \sum_{i=1}^n z_i - \sum_{i=1}^n \exp[z_i], & \kappa \neq 0 \\ -n \log \delta - \sum_{i=1}^n \exp[-(\frac{x_i - \lambda}{\delta})] - \sum_{i=1}^n (\frac{x_i - \lambda}{\delta}), & \kappa = 0 \end{cases} \quad (2)$$

است، که در آن $z_i = -\frac{1}{\kappa} \log[1 - \kappa(\frac{x_i - \lambda}{\delta})]$ در روش MLE معادلات درستنمایی فرم بسته‌ای ندارند و حل این معادلات به روش‌های عددی مانند نیوتن-رافسون امکان پذیر است. لذا می‌توان از برنامه بهینه‌سازی استاندارد در نرم افزار R برای یافتن برآورد پارامترهایی که تابع لگاریتم درستنمایی را ماکسیمم می‌کند، استفاده نمود. توجه شود که اگر κ به صفر نزدیک است، از ضابطه دوم تابع لگاریتم درستنمایی (۲) استفاده کنید و اگر κ به اندازه کافی از صفر فاصله دارد و در شرط $1 - \kappa(\frac{x_i - \lambda}{\delta}) \geq 0, i = 1, \dots, n$ صدق می‌کند، از ضابطه اول تابع لگاریتم

^۸ Probability Weighted Moments

^۹ Elemental Percentile

^{۱۰} Quantile Least Squares

درست‌نمایی (۲) استفاده کنید. در سایر نقاط لگاریتم درست‌نمایی نامتناهی است. برای استفاده از برنامه‌های بهینه‌سازی استاندارد به صورت زیر عمل می‌شود: برآوردهای ماکسیمم درست‌نمایی $\theta = (\lambda, \delta, \kappa)$ با استفاده از روش تکرار

$$\theta_{j+1} = \theta_j + I_{\theta}^{-1} \nabla_{\theta} \ell, \quad j = 0, 1, \dots$$

به دست می‌آید، که در آن θ_0 یک برآورد اولیه، I_{θ}^{-1} معکوس ماتریس اطلاع فیشر و $\nabla_{\theta} \ell$ بردار شیب ℓ نسبت به θ در θ_j است. در این صورت: الف (وقتی $\kappa \neq 0$): مولفه‌های ماتریس اطلاع فیشر I_{θ} عبارتند از (پریسکات و والدن، ۱۹۸۰):

$$\begin{aligned} m_{11} &= E\left(-\frac{\partial^2 \ell}{\partial \lambda^2}\right) = \frac{n}{\delta^2 \gamma^2} p \\ m_{22} &= E\left(-\frac{\partial^2 \ell}{\partial \delta^2}\right) = \frac{n}{\delta^2 \kappa^2} [\gamma - 2\Gamma(\gamma - \kappa) + p] \\ m_{33} &= E\left(-\frac{\partial^2 \ell}{\partial \kappa^2}\right) = \frac{n}{\kappa^2} \left[\frac{\pi^2}{6} + \left(\gamma - \frac{1}{\kappa}\right)^2 + \frac{2q}{\kappa} + \frac{p}{\kappa^2}\right] \\ m_{12} &= m_{21} = E\left(-\frac{\partial^2 \ell}{\partial \lambda \partial \delta}\right) = \frac{n}{\delta^2 \kappa} [p - \Gamma(\gamma - \kappa)] \\ m_{13} &= m_{31} = E\left(-\frac{\partial^2 \ell}{\partial \lambda \partial \kappa}\right) = -\frac{n}{\delta \kappa} \left(q + \frac{p}{\kappa}\right) \\ m_{23} &= m_{32} = E\left(-\frac{\partial^2 \ell}{\partial \delta \partial \kappa}\right) = -\frac{n}{\delta \kappa^2} \left[\gamma - \frac{1 - \Gamma(\gamma - \kappa)}{\kappa} - q - \frac{p}{\kappa}\right] \end{aligned}$$

که در آن

$$\begin{aligned} \Gamma(u) &= \int_0^{\infty} y^{u-1} e^{-y} dy \\ \psi(u) &= \frac{d \log \Gamma(u)}{du} \\ p &= (\gamma - \kappa)^2 \Gamma(\gamma - \kappa) \\ q &= \Gamma(\gamma - \kappa) \left[\psi(\gamma - \kappa) - \frac{\gamma - \kappa}{\kappa}\right] \\ \gamma &= 0.5772157 \end{aligned}$$

در این حالت شرایط نظم فقط برای $\frac{1}{\kappa} < \kappa$ برقرار است و واریانس‌های مجانبی و کوواریانس‌ها به صورت $\Sigma_{\hat{\theta}} = I_{\theta}^{-1}$ هستند.

ب (وقتی $\kappa = 0$): مولفه‌های ماتریس اطلاع فیشر I_θ عبارتند از:

$$I_\theta = \begin{bmatrix} \frac{n(1-\kappa)^2 \Gamma(1-2\kappa)}{\delta^2} & \frac{n(\gamma-1)}{\delta^2} \\ \frac{n(\gamma-1)}{\delta^2} & \frac{n(1-12\gamma+17\gamma^2+n^2)}{17\gamma^2} \end{bmatrix}$$

در این حالت شرایط نظم برقرار است و واریانس‌های مجانبی و کوواریانس‌ها به صورت $\Sigma_{\hat{\theta}} = I_\theta^{-1}$ است. بازه اطمینان $100(1-\alpha)\%$ برای پارامترهای λ ، δ و κ و برای چندک p ام، x_p ، به ترتیب عبارتند از:

$$\begin{aligned} \lambda &\in (\hat{\lambda} \pm Z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\lambda}}) & \kappa &\in (\hat{\kappa} \pm Z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\kappa}}) \\ \delta &\in (\hat{\delta} \pm Z_{\frac{\alpha}{2}} \hat{\sigma}_{\hat{\delta}}) & x_p &\in (\hat{x}_p \pm Z_{\frac{\alpha}{2}} \hat{\sigma}_{x_p}) \end{aligned}$$

۲.۳ روش گشتاورهای احتمال موزون

در روش PWM مناسب است که از گشتاورهای β_s توزیع GEV_M ، یعنی، $\beta_s = M(1, s, 0) = \frac{1}{s+1} [\lambda + \frac{\delta}{\kappa} (1 - \frac{\Gamma(1+\kappa)}{(s+1)^\kappa})]$ برای $\kappa \leq -1$ استفاده شود. اما سایر گشتاورهای β_s وجود ندارند. برای $\kappa > -1$ ، گشتاورهای β_s توزیع با گشتاورهای b_s نمونه $b_s = m(1, s, 0) = \frac{1}{n} \sum_{i=1}^n x_{(i)} p_{(i)}^s$ می‌شوند (گرینوود و همکاران، ۱۹۷۹):

$$\beta_0 = b_0 = \lambda + \frac{\delta}{\kappa} [1 - \Gamma(1 + \kappa)] \quad (3)$$

$$2\beta_1 - \beta_0 = 2b_1 - b_0 = \frac{\delta}{\kappa} \Gamma(1 + \kappa) (1 - 2^{-\kappa}) \quad (4)$$

$$\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} = \frac{3b_2 - b_0}{2b_1 - b_0} = \frac{1 - 3^{-\kappa}}{1 - 2^{-\kappa}} \quad (5)$$

برآوردگرهای PWM از حل معادلات (۳) تا (۵) برای λ ، δ و κ به دست می‌آیند. چون رابطه (۵) تابعی فقط از پارامتر κ است از این رو، با یک روش عددی مناسب برآورد پارامتر κ به دست می‌آید. با جایگذاری $\hat{\kappa}$ در (۴) برآورد پارامتر δ به صورت

$$\hat{\delta} = \frac{\hat{\kappa}(2b_1 - b_0)}{\Gamma(1 + \hat{\kappa})(1 - 2^{-\hat{\kappa}})}$$

حاصل می‌شود. در نهایت، $\hat{\kappa}$ و $\hat{\delta}$ در (۳) جایگزین می‌شوند و برآورد پارامتر λ به صورت

$$\hat{\lambda} = b_0 - \frac{\hat{\delta}}{\hat{\kappa}} [1 - \Gamma(1 + \hat{\kappa})]$$

به دست می‌آید. برآورد پارامترهای توزیع GEV_M به روش‌های مختلف دارای مشکلاتی است. برای مثال، در روش MLE دامنه توزیع به پارامتر بستگی دارد و در نتیجه MLE ویژگی‌های مجانبی معمولی را ندارد، MLE نیاز به محاسبات عددی دارد و تابع درست‌نمایی ممکن است برای بعضی از نمونه‌ها ماکسیمم موضعی نداشته باشد. همچنین، روش PWM می‌تواند برآوردهایی از θ را تولید کند که در فضای پارامتر قرار ندارند. در این گونه موارد از روش برآورد EP استفاده می‌شود؛ زیرا در بررسی شبیه‌سازی‌هایی که توسط کاستیلو و هادی (۱۹۹۷، ۱۹۹۵a، ۱۹۹۵b) صورت گرفته است نشان می‌دهد که هیچ روشی به طور یکنواخت برای همه $\theta \in \Theta$ بهترین نیست اما روش برآورد EP مقایسه خوبی را با دیگر روش‌ها انجام می‌دهد.

۳.۳ روش صدک

روش EP در دو مرحله انجام می‌شود: در مرحله اول برآوردهای اولیه برای پارامترها که براساس زیر مجموعه‌های انتخاب شده از داده‌ها هستند، محاسبه می‌شوند و با ترکیب آن‌ها در مرحله دوم برآوردهایی برای همه پارامترها برای تولید برآوردهای بسنده و استوار نهایی به دست می‌آید.

مرحله اول: برآوردهای اولیه صدک

برای حالت $\kappa \neq 0$ توزیع GEV_M دارای سه پارامتر λ ، δ و κ است. بنابراین، هر مجموعه مولفه شامل سه آماره مرتب مجزا است. فرض کنید چندک‌های نمونه ونظری را مساوی هم قرار دهید:

$$x_{(i)} = \lambda + \frac{\delta}{\kappa} [1 - (-\log p_{(i)})^\kappa]$$

$$\begin{aligned} x_{(j)} &= \lambda + \frac{\delta}{\kappa} [1 - (-\log p_{(j)})^\kappa] \\ x_{(r)} &= \lambda + \frac{\delta}{\kappa} [1 - (-\log p_{(r)})^\kappa] \end{aligned} \quad (۶)$$

که در آن $p_{(i)} = \frac{(i-\circ/۳۵)}{n}$ اکنون تساوی سوم را از دوم کم کنید و سپس تساوی سوم را از اول کم کنید. در این صورت داریم:

$$\frac{x_{(j)} - x_{(r)}}{x_{(i)} - x_{(r)}} = \frac{C_r^\kappa - C_j^\kappa}{C_r^\kappa - C_i^\kappa} = \frac{1 - A_{jr}^\kappa}{1 - A_{ir}^\kappa} \quad (۷)$$

که در آن $A_{ir} = \frac{C_i}{C_r}$ و $C_i = -\log(p_{(i)})$ با حل رابطه (۷) برای κ با استفاده از روش دو بخشی برآورد $\hat{\kappa}_{ijr}$ ، که تابعی از سه مشاهده $x_{(i)}$ ، $x_{(j)}$ و $x_{(r)}$ است، به دست می آید (کاستیلو و همکاران، ۲۰۰۵). با جایگذاری $\hat{\kappa}_{ijr}$ در روابط (۶) برآورد پارامترهای λ و δ به صورت

$$\begin{aligned} \hat{\delta}_{ijr} &= \frac{\hat{\kappa}_{ijr}(x_{(i)} - x_{(r)})}{C_r^{\hat{\kappa}} - C_i^{\hat{\kappa}}} \\ \hat{\lambda}_{ijr} &= x_{(i)} - \frac{\hat{\delta}_{ijr}(1 - C_i^{\hat{\kappa}_{ijr}})}{\hat{\kappa}_{ijr}} \end{aligned}$$

به دست می آیند. توجه شود که برآوردهای $\hat{\kappa}_{ijr}$ ، $\hat{\delta}_{ijr}$ و $\hat{\lambda}_{ijr}$ برای $\hat{\kappa} > 0$ باید در $x_{(n)} \leq \hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}}$ و برای $\hat{\kappa} < 0$ باید در $x_{(1)} \geq \hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}}$ صدق کنند.

مرحله دوم: برآوردهای نهایی

برآوردهای اولیه فوق براساس سه آماره مرتب مجزا می باشند. از این رو، برآوردهای بسنده و استوار با استفاده از آماره های مرتب دیگر به صورت زیر به دست می آیند. عدد معلوم N را از زیر مجموعه های اولیه هر کدام با حجم ۳ به طور تصادفی یا با استفاده از همه زیر مجموعه های ممکن انتخاب کنید. برای هر یک از این سه زیر مجموعه اولیه، یک برآورد اولیه از پارامترهای $\theta = \{\lambda, \delta, \kappa\}$ محاسبه کنید و با $\hat{\theta}_1, \dots, \hat{\theta}_N$ نشان دهید. برآوردهای اولیه ای که در تناقض با داده ها هستند کنار گذاشته می شوند. این برآوردهای اولیه را می توان با استفاده از توابع مناسبی

(ترجیحاً استوار^{۱۱} مانند میانه^{۱۲} (MED) و α -میانگین پیراسته^{۱۳} (TM α) که در آن α درصد پیرایش را نشان می‌دهد) برای به دست آوردن یک برآورد نهایی کلی با هم ترکیب کرد. بنابراین، یک برآورد نهایی از $\theta = \{\lambda, \delta, \kappa\}$ به صورت

$$\begin{aligned}\hat{\kappa}_{MED} &= \text{Median}(\hat{\kappa}_1, \dots, \hat{\kappa}_N) \\ \hat{\delta}_{MED} &= \text{Median}(\hat{\delta}_1, \dots, \hat{\delta}_N) \\ \hat{\lambda}_{MED} &= \text{Median}(\hat{\lambda}_1, \dots, \hat{\lambda}_N)\end{aligned}\quad (۸)$$

یا

$$\begin{aligned}\hat{\kappa}_{TM_\alpha} &= TM_\alpha(\hat{\kappa}_1, \dots, \hat{\kappa}_N) \\ \hat{\delta}_{TM_\alpha} &= TM_\alpha(\hat{\delta}_1, \dots, \hat{\delta}_N) \\ \hat{\lambda}_{TM_\alpha} &= TM_\alpha(\hat{\lambda}_1, \dots, \hat{\lambda}_N)\end{aligned}\quad (۹)$$

تعریف می‌شود، که در آن $MED(y_1, \dots, y_N)$ میانه مجموعه اعداد $\{y_1, \dots, y_N\}$ و $TM_\alpha(y_1, \dots, y_N)$ میانگین به دست آمده بعد از پیرایش $(100 - \alpha)\%$ درصد بزرگترین و $(100 - \alpha)\%$ درصد کوچکترین آماره‌های مرتب $\{y_{(1)}, \dots, y_{(N)}\}$ است. برآوردگرهای MED استوارند اما کارا نیستند. برآوردگرهای TM_α کمتر استوارند اما کارا تر از برآوردگرهای MED هستند. هر چه میزان پیراستگی بیشتر باشد، برآوردگرهای TM_α استوارتر و کارایی کمتری دارند. تجربه نشان می‌دهد که برآوردهای MED و TM_α دارای ویژگی‌های مشابه هستند، اما جذر میانگین توان دوم خطاهای TM_α اندکی کوچکتر از MED است (کاستیلو و همکاران، ۲۰۰۵). توجه شود که برای دوری از برآوردهایی که در تناقض با داده‌ها هستند، بهتر است نقطه انتهایی $\lambda + \frac{\delta}{\kappa}$ به جای λ برآورد شود و براساس آن λ دوباره به دست آید، یعنی، برآورد زیر را جایگزین برآورد (۸) کنید:

$$\left(\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}}\right)_{MED} = \text{Median}\left(\left(\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}}\right)_{1,2,n}, \left(\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}}\right)_{1,3,n}, \dots, \left(\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}}\right)_{1,n-1,n}\right)$$

^{۱۱} Robust

^{۱۲} Median

^{۱۳} α -trimmed mean

$$\hat{\lambda}_{MED} = Median(\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}})_{MED} - \frac{\hat{\delta}_{MED}}{\hat{\kappa}_{MED}}$$

به طور مشابه، برآورد زیر را جایگزین برآورد (۹) کنید:

$$(\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}})_{TM_\alpha} = TM_\alpha((\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}})_{1,2,n}, (\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}})_{1,2,n}, \dots, (\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}})_{1,n-1,n})$$

$$\hat{\lambda}_{TM_\alpha} = TM_\alpha(\hat{\lambda} + \frac{\hat{\delta}}{\hat{\kappa}})_{TM_\alpha} - \frac{\hat{\delta}_{TM_\alpha}}{\hat{\kappa}_{TM_\alpha}}$$

۴.۳ روش کمترین توان‌های دوم چندک

در روش QLS پارامترهای توزیع GEV_M ، با مینیمم کردن توان‌های دوم اختلاف بین چندک‌های نظری و مشاهده شده برآورد می‌شوند. در این صورت، پارامترهای این توزیع با حل معادلات زیر که با استفاده از برنامه بهینه سازی استاندارد در نرم افزار R قابل محاسبه است، به دست می‌آیند:

$$Minimize_{\lambda, \delta, \kappa} \sum_{i=1}^n [x_{(i)} - \lambda - \frac{\delta}{\kappa} [1 - [-\log(p_{(i)})]^\kappa]^2, \quad \kappa \neq 0$$

$$Minimize_{\lambda, \delta} \sum_{i=1}^n [x_{(i)} - \lambda + \delta \log[-\log(p_{(i)})]^2, \quad \kappa = 0$$

۴ آزمون نیکویی برازش

روش‌های متفاوتی برای بررسی خوبی برازش مدل وجود دارد که از آن جمله می‌توان به آزمون کلموگروف-اسمیرنوف اشاره کرد. در مواقعی که دم توزیع برای ما اهمیت دارد، پیشنهاد می‌شود که از آزمون آندرسون-دارلینگ که آماره این آزمون عضوی از خانواده کرامر-فون میزس است، استفاده شود. علت انتخاب این آزمون برای توزیع GEV_M این است که آماره آن بیشتر از سایر آماره‌ها به مشاهدات واقع در دم توزیع اهمیت می‌دهد. همچنین، برای آزمون نیکویی برازش می‌توان از آزمون فرض برای توزیع GEV_M که شامل روش‌هایی براساس درست‌نمایی است، استفاده کرد. برای آزمون نیکویی برازش می‌توان از روش‌های دیگری نیز مانند

نمودارهای گزارش احتمال و یا آزمون فرض برای توزیع که شامل آزمون‌های والد^{۱۴} و خمیدگی^{۱۵} است، استفاده کرد (کاستیلو و همکاران، ۱۹۸۹).

۵ مدل‌بندی مقادیر کرانگین بیشترین سرعت باد شهر زاهدان

با توجه به اینکه سرعت‌های باد بین ۷۰ تا ۱۰۰ کیلومتر در ساعت باعث ایجاد خسارت‌هایی در محورهای مواصلاتی و کمبود دید افقی در حد ۶ کیلومتر می‌شود، به بررسی بیشترین سرعت باد سالانه شهر زاهدان می‌پردازیم. در این تحقیق، مشاهدات مربوط به بیشترین سرعت باد سالانه (برحسب نات) برای دوره زمانی ۱۳۳۰ تا ۱۳۹۰ از ایستگاه هواشناسی زاهدان مورد تحلیل قرار گرفته است. به این ترتیب که ابتدا مدل مناسب برازش داده شده و سپس نیکویی برازش مورد ارزیابی قرار گرفته است. در نهایت برخی پیش‌بینی‌های لازم انجام شده است. در انجام محاسبات از نرم افزار R استفاده شده است.

توزیع GEV_M به این مجموعه داده برازش داده شده و پارامترهای آن به چهار روش MLE, PWM, EP و QLS برآورد می‌شوند. برای بررسی نیکویی برازش از میانگین قدر مطلق خطای مقیاس‌بندی شده^{۱۶} (ASAE) استفاده شده است.

$$ASAE = \frac{1}{n} \sum_{i=1}^n \frac{|x_{(i)} - \hat{x}_{(i)}|}{x_{(n)} - x_{(1)}}$$

همان‌طور که در جدول ۱ ملاحظه می‌شود، در مقایسه روش‌های برآورد براساس معیار ASAE برآوردهای روش MLE و EP نزدیک به هم هستند. در نهایت روش MLE بهتر از سه روش دیگر است. با بررسی نمودارهای احتمال^{۱۷} (P-P) و احتمال تجمعی^{۱۸} (Q-Q) در شکل‌های ۲، ۳، ۴ و ۵ مشاهده می‌شود که، در شکل ۲ در هر دو نمودار P-P و Q-Q بیشترین روند خطی وجود دارد و این یعنی، مدل GEV_M با استفاده از روش MLE به خوبی به داده‌های بیشترین سرعت باد شهر زاهدان برازش

^{۱۴} Wald test

^{۱۵} Curvature test

^{۱۶} Average Scaled Absolute Error

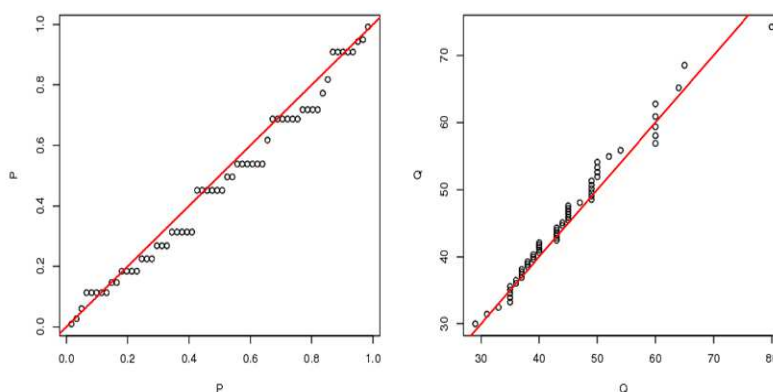
^{۱۷} Percentile-Percentile

^{۱۸} Quantile-Quantile

جدول ۱: برآورد پارامترهای توزیع GEV_M با استفاده از سه روش MLE, PWM و QLS برای بیشترین سرعت باد شهر زاهدان

روش	$\hat{\lambda}$	$\hat{\delta}$	$\hat{\kappa}$	ASAE
MLE	۴۰/۵	۷/۰۴	-۰/۰۴	۰/۰۱۹
PWM	۴۱/۴۵	۸/۳۸	۰/۲	۰/۰۳۲۶
EP	۴۰/۴۲	۷/۵۵	۰/۰۳	۰/۰۱۹۶
QLS	۴۰/۳۵	۸/۱۲	-۰/۰۲	۰/۰۲۱۹

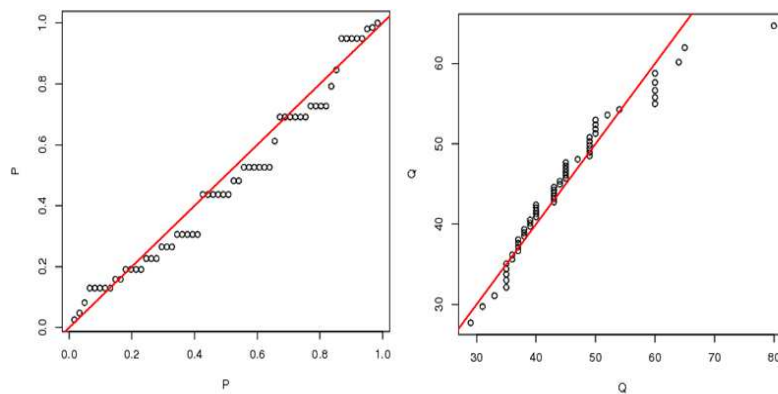
یافته است. اما در شکل ۳ روند خطی به خوبی مشاهده نمی شود از این رو برآزش PWM به خوبی دیگر روش ها نیست.



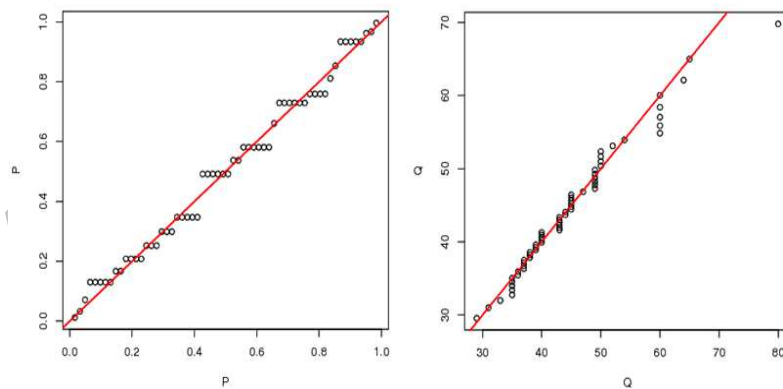
شکل ۲: نمودارهای P-P و Q-Q توزیع GEV_M با روش MLE برای بیشترین سرعت باد

در محاسبه بازه اطمینان ML برای پارامترهای λ , δ و κ به خطای استاندارد برآوردها نیاز است که پس از محاسبه آنها بازه اطمینان ۹۵٪ برای این پارامترها به صورت زیر حاصل می شوند:

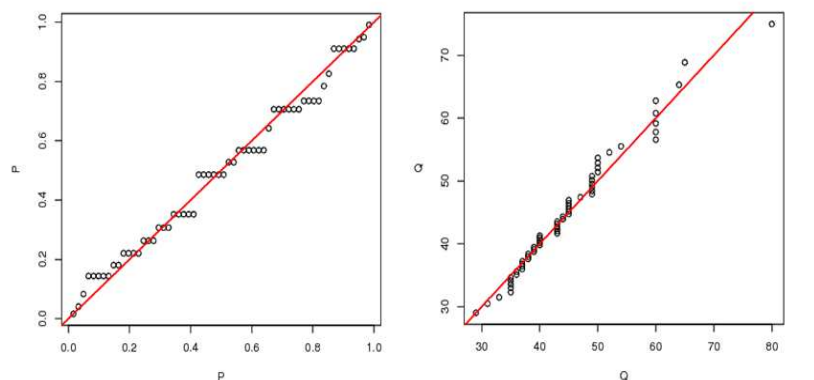
۲۲۰ مدل‌بندی مقادیر کرانگین سرعت باد شهر زاهدان



شکل ۳: نمودارهای P-P و Q-Q توزیع GEV_M با روش PWM برای بیشترین سرعت باد



شکل ۴: نمودارهای P-P و Q-Q توزیع GEV_M با روش EP برای بیشترین سرعت باد



شکل ۵: نمودارهای P-P و Q-Q توزیع GEV_M با روش QLS برای بیشترین سرعت باد

$$\lambda \in (38/5, 42/5)$$

$$\delta \in (5/57, 8/51)$$

$$\kappa \in (-0/23, 0/15)$$

که نشان می‌دهد بازه پارامتر شکل κ شامل صفر است، در نتیجه فرض این که توزیع بیشترین سرعت باد شهر زاهدان در سطح معنی‌داری ۵٪ از دامنه جذب گامبل پیروی می‌کند را تأیید می‌کند. همچنین، بازه اطمینان ML برای چندک‌های ۰/۹۰، ۰/۹۵ و ۰/۹۹ در سطح معنی‌داری ۰/۰۵ و ۰/۰۱ α محاسبه و در جدول ۲ ارائه شده‌اند.

همان‌طور که ملاحظه می‌شود بازه اطمینان برای $x_{0/99}$ بزرگتر از بازه‌های اطمینان برای $x_{0/95}$ و $x_{0/9}$ است. مقادیر آماره آزمون نسبت درست‌نمایی برای فرضیه (گامبل) $H_0: \kappa = 0$ در مقابل (وایبول یا فرشه) $H_1: \kappa \neq 0$ و p-value (به اختصار به صورت PV نشان داده شده است) در جدول ۳ ارائه شده‌اند.

براساس آماره نسبت درست‌نمایی (LR) و هم‌مین‌طور براساس آماره نسبت

جدول ۲: فاصله اطمینان ML برای چندک‌های ۰/۹۰، ۰/۹۵ و ۰/۹۹

α	p	\hat{x}_p	CI
۰/۰۱	۰/۹۰	۵۹/۲۱	(۵۴/۳۹, ۶۲/۰۳)
	۰/۹۵	۶۵/۰۳	(۵۶/۹۸, ۷۳/۰۹)
	۰/۹۹	۷۸/۳۳	(۵۸/۶۸, ۹۷/۹۷)
۰/۰۵	۰/۹۰	۵۹/۲۱	(۵۵/۵۴, ۶۲/۸۷)
	۰/۹۵	۶۵/۰۳	(۵۸/۹, ۷۱/۱۶)
	۰/۹۹	۷۸/۳۳	(۶۳/۳۸, ۹۳/۲۷)

جدول ۳: مقادیر آماره‌های آزمون LR و LR* برای داده‌های سرعت باد

Critical.Value	LR*	PV	LR	$\hat{\kappa}$	n
۳/۸۴	۰/۱۶	۰/۶۹	۰/۱۷	-۰/۰۴	۶۰

درست‌نمایی تعدیل یافته^{۱۹} هاسکینگ (۱۹۸۴) $(LR^* = (1 - \frac{\lambda}{n})LR)$ که تقریب دقیق‌تری به توزیع مجانبی LR می‌دهد، فرض H_0 پذیرفته می‌شود. بنابراین، توزیع گامبل ماکسیمال با شرایط ارائه شده به‌عنوان توزیع تعمیم یافته مقدار کرانگین مورد نظر به داده‌های ماکسیمم برازنده می‌شود. سپس با استفاده از آزمون‌های آندرسون-دارلینگ و کلموگروف-اسمیرنوف آزمون‌های نیکویی برازش اجرا شده است که با آماره آزمون $A-D = ۰/۳۶$ و $p\text{-value} = ۰/۸۹$ و با آماره آزمون $K-S = ۰/۰۸$ و $p\text{-value} = ۰/۸۵$ برازندگی این مدل تأیید می‌شود.

۶ بحث و نتیجه‌گیری

پس از تحلیل و بررسی آزمون نیکویی برازش نسبت درست‌نمایی به این نتیجه رسیدیم که مدل گامبل ماکسیمال به داده‌های بیشترین سرعت باد شهر زاهدان برازش خوبی یافته است و این نتیجه را آزمون‌های آندرسون-دارلینگ و کلموگروف-اسمیرنوف نیز تأیید کردند. همچنین، بازه اطمینان برای پارامتر شکل κ این فرض که توزیع بیشترین سرعت باد شهر زاهدان از دامنه جذب گامبل ماکسیمال پیروی می‌کند، را تأیید کرد. از طرفی، با مقایسه چهار روش برآورد

^{۱۹} Modified Likelihood Ratio

MLE، PWM، EP و QLS براساس معیار ASAE و نمودارهای P-P و Q-Q این نتیجه حاصل شد که روش MLE بهتر از سه روش دیگر است. بررسی مقادیر کرانگین سرعت باد در طی سال‌های ۱۳۹۰-۱۳۳۰ بیانگر آن است که بیشترین سرعت‌های باد در طی این ۶۰ سال به میزان ۸۰ و ۶۰ نات بوده است. اکنون با مشخص شدن مدل سرعت باد این شهر (یعنی مدل گامبل ماکسیمال) که یکی از حالت‌های خاص توزیع GEV_M است، دوره بازگشت این میزان سرعت باد به ترتیب برابر ۱۵۹ و ۱۴ می‌باشد. بنابراین با توجه به تعریف دوره بازگشت که میانگین زمان مورد انتظار برای رخداد یک پیشامد کرانگین است، به طور متوسط هر ۱۵۹ سال یک بار باید انتظار وزش بادی به سرعت ۸۰ نات و هر ۱۴ سال یک بار باید انتظار وزش بادی به سرعت ۶۰ نات را داشته باشیم.

در پایان ذکر چند نکته ضروری است:

۱. راه حل‌های عددی برای برآورد ماکسیمم درست‌نمایی (ML) به یک مقدار اولیه $\theta_0 = \{\lambda_0, \delta_0, \kappa_0\}$ نیاز دارد که می‌توان از برآوردهای گامبل برای θ_0 یا برای سادگی از روش‌های EP و QLS استفاده کرد.
۲. برآوردهای ML برای حالت $\kappa > -\frac{1}{2}$ ویژگی‌های کلاسیک و خوب برآوردهای ML را دارا هستند ولی برای حالت $\kappa \leq -\frac{1}{2}$ این ویژگی را ندارند و پیشنهاد می‌شود از روش دیگری مانند روش PWM استفاده شود.

تقدیر و تشکر

از داوران محترم این مقاله به خاطر نظرات سازنده‌شان در بهبود این اثر، تشکر و قدردانی می‌شود.

مراجع

- Castillo, E. and Hadi, A. S. (1995a), Modeling Lifetime Data with Applications to Fatigue Models, *Journal of the American Statistical Association*, **90**, 1041-1054.

- Castillo, E. and Hadi, A. S. (1995b), Parameter and Quantile Estimation for the Generalized Extreme-Value Distribution, *Environmetrics*, **5**, 417-432.
- Castillo, E. and Hadi, A. S. (1997), Fitting the Generalized Pareto Distribution to Data , *Journal of the American Statistical Association*, **92**, 1609-1620.
- Castillo, E., Galambos, J. and Sarabia, J. M. (1989), The Selection of the Domain of Attraction of an Extreme Value Distribution from a Set of Data, In *Extreme Value Theory*, vol. 51, of *Lecture Notes in Statistics*, pp.181-190, Springer Verlag, New York.
- Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2005), *Extreme Value and Related Models with Applications in Engineering and Science*, John Wiley, New York.
- Fisher, R. A. and Tippett, L. H. C. (1928), Limiting Forms of the Frequency Distributions of the Largest or Smallest Member of a Sample, *Proceedings of the Cambridge Philosophical Society*, **24**, 180-190.
- Galambos, J. (1987), *Asymptotic Theory of Extreme Order Statistics*, Robert E. Krieger, Malabar, Florida, 2nd Ed.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979), Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form, *Water Resources Research*, **15**, 1049-1054.
- Hosking, J. R. M. (1984), Testing Whether the Shape Parameter is Zero in the Generalized Extreme-Value Distribution, *Biometrika*, **71**, 367-374.

Jenkinson, A. F. (1955), The Frequency Distribution of the Annual Maximum (or Minimum) of Meteorological Elements, *Quarterly Journal of the Royal Meteorological Society*, **81**, 158-171.

Prescott, P. and Walden, A. T. (1980), Maximum Likelihood Estimation of the Parameters of the Generalized Extreme-Value Distribution, *Biometrika*, **67**, 723-724.

Tiago de Oliveria, J. (1958), Extremal Distributions, *Revista de la Facultad de Ciencias*, **A7**, 215-227.

Archive of SID