

توزیع کاپا سه پارامتری و برازش آن به داده‌های مجموع بارش ماهانه ایستگاه آبدلی استان تهران

مریم شرفی^{۱*}، فریده توانگر^۲، شهره انعامی^۳، حسین نادب^۴

۱. دانشگاه شیراز، دانشکده علوم، گروه آمار

۲. دانشگاه یورک، کانادا، گروه آمار و ریاضی

۳. دانشگاه پیام نور، مرکز بوشهر

۴. دانشگاه یزد، گروه آمار

دریافت ۹۷/۰۲/۰۴

پذیرش ۹۷/۱۰/۲۴

چکیده

توزیع کاپا یکی از توزیع‌های چوله مثبت است که به منظور تجزیه و تحلیل داده‌های بارندگی، سرعت باد و جریان سیلاب استفاده می‌شود. در این مقاله ابتدا به بررسی توزیع کاپا سه پارامتری معرفی شده به وسیله پارک و همکاران [۱] پرداخته و سپس چهار روش برآوردیابی شامل روش گشتاوری، گشتاورهای خطی^۱، بیشینه درست‌نمایی و بیشینه حاصل ضرب فاصله‌ها را برای برآوردیابی پارامترهای این توزیع ارائه داده و با استفاده از یک بررسی شبیه‌سازی، به مقایسه عملکرد آنها پرداخته و در پایان، این روش‌ها، برای داده‌های مربوط به مجموع بارش ماهانه ایستگاه آبدلی استان تهران به کار گرفته می‌شود.

واژه‌های کلیدی: توزیع کاپا سه پارامتری، برآوردگر بیشینه درست‌نمایی، برآوردگر گشتاورهای خطی، برآوردگر بیشینه حاصل ضرب فاصله‌ها.

مقدمه

یک خانواده از توزیع‌های نامتقارن چوله مثبت، توزیع کاپا است که میلک^۲ [۲] و میلک و جانسون^۳ [۳]، برای توصیف داده‌های بارش ارائه کردند. معمولاً از توزیع‌های گاما و لگ نرمال برای برازش داده‌های بارش استفاده می‌شود، اما این توزیع‌ها از نظر محاسباتی آسان نیستند، زیرا تابع توزیع تجمعی و تابع چندکی آنها شکل بسته‌ای ندارد، در حالی که تابع توزیع تجمعی و تابع چندکی خانواده توزیع‌های کاپا دارای شکل بسته است. اشکار^۴ و همکاران [۴] با در نظر گرفتن توزیع کاپا دو پارامتری معرفی شده به وسیله میلک [۲] به عنوان یکی از چندین توزیع دو پارامتری استفاده شده در هواشناسی به این نتیجه رسیدند که این توزیع مناسب‌ترین توزیع برای برازش داده‌های جریان کم در رودخانه است. اشکار و همکاران [۵] به برآورد چندک‌ها و پارامترهای این توزیع با استفاده از روش بیشینه درست‌نمایی پرداختند.

*نویسنده مسئول msharafi@shirazu.ac.ir

1. L-Moments
2. Mielk
3. Johnson
4. Ashkar

پارک^۱ و همکاران [۱] توزیع کاپا دو پارامتری میلک [۲] را با اضافه کردن پارامتر مکان μ تعمیم دادند و توزیع کاپا سه پارامتری را معرفی کردند. آنها این توزیع را برای مدل‌بندی متغیرهایی که دارای کران پایین هستند، مناسب دانستند. سپس به برآورد پارامترهای این توزیع از روش‌های مختلف پرداختند و این توزیع را برای برازش به داده‌های سری زمانی مربوط به بارش روزانه در نظر گرفتند.

آشور^۲ و همکاران [۶] پارامترهای توزیع کاپا سه پارامتری را برای داده‌های سانسور شده نوع دوم برآورد کردند. هاسکینگ [۷] توزیع کاپا چهار پارامتری را به‌عنوان تعمیمی از توزیع کاپا سه پارامتری میلک و جانسون [۳] معرفی کرد. پژوهش‌های بسیاری روی توزیع کاپا چهار پارامتری انجام شده است که از جمله می‌توان به حسن‌زاده و عبدی کردانی [۸] اشاره کرد که این توزیع را در تخمین شاخص بارندگی استاندارد شده به‌کار بردند. در این مقاله، ابتدا در بخش دوم به معرفی توزیع کاپا سه پارامتری معرفی شده به‌وسیله پارک و همکاران [۱] پرداخته و برخی از ویژگی‌های آن را ارائه می‌کنیم. سپس روش‌های مختلف برآوردیابی پارامترهای این توزیع را در بخش سوم شرح داده و در بخش‌های چهارم و پنجم با استفاده از یک بررسی شبیه‌سازی و یک مجموعه داده واقعی به مقایسه روش‌های برآوردیابی ذکر شده در بخش سوم پرداخته و نتیجه‌گیری می‌کنیم.

معرفی توزیع کاپا سه پارامتری و ویژگی‌های آن

متغیر تصادفی X دارای توزیع کاپا سه پارامتری است، هرگاه تابع چگالی آن بدین صورت باشد:

$$f(x) = \frac{\alpha}{\beta} \left[\alpha + \left(\frac{x-\mu}{\beta} \right)^\alpha \right]^{-\left(1+\frac{1}{\alpha}\right)} \quad \alpha > 0, \beta > 0, x > \mu, x > 0, \quad (1)$$

که در آن μ ، β و α به‌ترتیب پارامترهای مکان، مقیاس و شکل هستند. تابع توزیع و تابع چندکی متغیر تصادفی X به‌ترتیب بدین صورت است:

$$F(x) = \left(\frac{x-\mu}{\beta} \right) \left[\alpha + \left(\frac{x-\mu}{\beta} \right)^\alpha \right]^{-\left(\frac{1}{\alpha}\right)},$$

$$x(p) = \mu + \beta \left[\frac{\alpha p^\alpha}{1-p^\alpha} \right]^{\frac{1}{\alpha}}, \quad 0 < p < 1.$$

در رابطه (۱)، اگر $\mu = 0$ آن‌گاه این توزیع به توزیع کاپا دو پارامتری و اگر $\mu = 0$ و $\beta = 1$ ، آن‌گاه به توزیع کاپا یک پارامتری با پارامتر شکل α تبدیل می‌شود. شکل تابع چگالی کاپا سه پارامتری به‌ازای مقادیر متفاوتی از α با $\mu = 1$ و $\beta = 2$ در شکل ۱ نمایش داده شده‌اند.

گشتاورهای غیرمرکزی اول تا سوم X به‌ترتیب بدین صورت محاسبه می‌شوند:

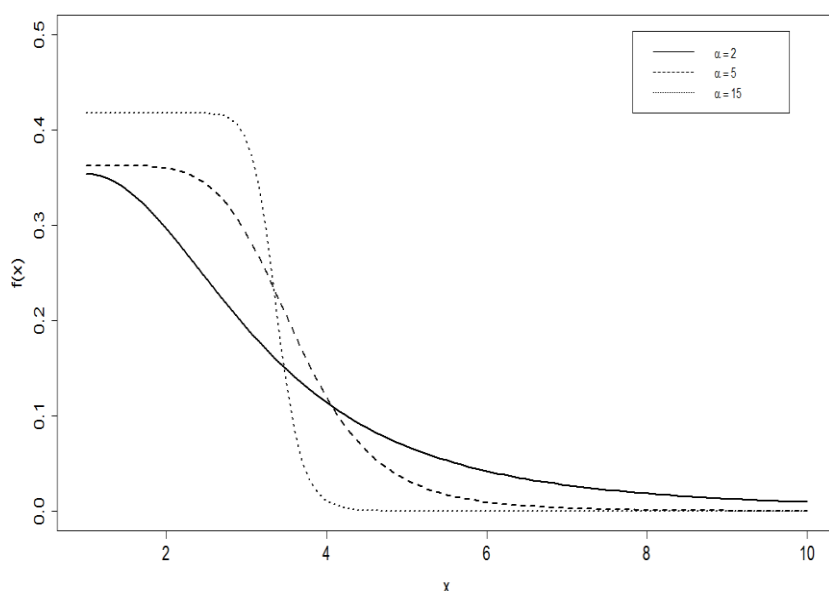
$$\mu_1 = E(X) = \int_0^1 [x(p)] dp = \int_0^1 \left(\mu + \beta \left[\frac{\alpha F^\alpha}{1-F^\alpha} \right]^{\frac{1}{\alpha}} \right) dp = \mu + \beta \alpha^{\frac{1}{\alpha}-1} B\left(\frac{2}{\alpha}, 1 - \frac{1}{\alpha}\right), \alpha > 1$$

$$\begin{aligned}\mu_2 = E(X^2) &= \int_0^1 [x(p)]^2 dp \\ &= \mu^2 + 2\mu\beta\alpha^{\frac{1}{\alpha}-1}B\left(\frac{2}{\alpha}, 1 - \frac{1}{\alpha}\right) + \beta^2\alpha^{\frac{2}{\alpha}-1}B\left(\frac{3}{\alpha}, 1 - \frac{2}{\alpha}\right) \quad \alpha > 2\end{aligned}$$

$$\begin{aligned}\mu_3 = E(X^3) &= \int_0^1 [x(p)]^3 dp \\ &= \mu^3 + 3\mu^2\beta\alpha^{\frac{1}{\alpha}-1}B\left(\frac{2}{\alpha}, 1 - \frac{1}{\alpha}\right) + 3\mu\beta^2\alpha^{\frac{2}{\alpha}-1}B\left(\frac{3}{\alpha}, 1 - \frac{2}{\alpha}\right) \\ &\quad + \beta^3\alpha^{\frac{3}{\alpha}-1}B\left(\frac{4}{\alpha}, 1 - \frac{3}{\alpha}\right), \quad \alpha > 3\end{aligned}$$

که در آن

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$



شکل ۱. نمودار تابع چگالی کاپا سه پارامتری به‌ازای مقادیر مختلفی از α ، $\mu = 1$ و $\beta = 2$

به‌طور کلی گشتاور مرتبه k ام متغیر تصادفی X را می‌توان بدین‌صورت به‌دست آورد:

$$\mu_k = E(X^k) = \int_0^1 [x(p)]^k dp = \sum_{i=0}^k \binom{k}{i} \mu^i \beta^{k-i} g_{(k-i)}(\alpha) \quad (۲)$$

به‌طوری‌که

$$g_{(k-i)}(\alpha) = \alpha^{\left(\frac{k-i}{\alpha}\right)-1} B\left(\frac{1+(k-i)}{\alpha}, 1 - \frac{k-i}{\alpha}\right), \quad \alpha > k-i.$$

با استفاده از روابط مذکور می‌توان واریانس، ضرایب چولگی و کشیدگی متغیر تصادفی X را نیز بدین‌صورت محاسبه کرد:

$$V(X) = \mu_2 - \mu_1^2 = \beta^2 \left(g_{(2)}(\alpha) - g_{(1)}^2(\alpha) \right), \quad \alpha > 2,$$

$$\gamma_1 = \frac{\mu_3 - 3\mu_2\mu_1 + 2\mu_1^3}{[\mu_2 - \mu_1^2]^{\frac{3}{2}}} = \frac{g_{(3)}(\alpha) - 3g_{(1)}(\alpha)g_{(2)}(\alpha) + 2g_{(1)}^3(\alpha)}{[g_{(2)}(\alpha) - g_{(1)}^2(\alpha)]^{\frac{3}{2}}}, \quad \alpha > 3,$$

$$\begin{aligned} \gamma_2 &= \frac{\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4}{[\mu_2 - (\mu_1)^2]^2} - 3 \\ &= \frac{g_{(4)}(\alpha) - 4g_{(1)}(\alpha)g_{(3)}(\alpha) + 6g_{(1)}^2(\alpha)g_{(2)}(\alpha) - 3g_{(1)}^4(\alpha)}{[g_{(2)}(\alpha) - g_{(1)}^2(\alpha)]^2} - 3, \quad \alpha > 4. \end{aligned}$$

روش‌های برآوردیابی پارامترهای توزیع کاپا سه پارامتری

در این بخش چهار روش برآوردیابی برای پارامترهای توزیع کاپا سه پارامتری ارائه می‌گردد. سه روش اول در پارک و همکاران [۱] آمده است، ولی روش بیشینه حاصل ضرب فاصله‌ها به عنوان روش جدید در این مقاله در نظر گرفته می‌شود.

۱. روش گشتاوری (MME)

فرض کنید $X_1, X_2, \dots, X_n$ یک نمونه تصادفی از توزیع کاپا سه پارامتری با مقادیر $x_1, x_2, \dots, x_n$ باشد. در این بخش برآوردهای گشتاوری پارامترهای α ، β و μ محاسبه و با $\tilde{\alpha}$ ، $\tilde{\beta}$ و $\tilde{\mu}$ نمایش داده می‌شود. می‌دانیم گشتاورهای غیرمرکزی نمونه‌ای بدین صورت تعریف می‌شوند:

$$m_1 = n^{-1} \sum_{i=1}^n x_i, \quad m_2 = n^{-1} \sum_{i=1}^n x_i^2, \quad m_3 = n^{-1} \sum_{i=1}^n x_i^3.$$

اینک به حل دستگاه $m_i = \mu_i$ ، $i = 1, 2, 3$ می‌پردازیم.

بنابر رابطه (۲) داریم:

$$\mu_1 = \mu + \beta g_{(1)}(\alpha). \quad (3)$$

با توجه به روابط (۲) و (۳)

$$\begin{aligned} \mu_2 &= \mu^2 + 2\mu\beta g_{(1)}(\alpha) + \beta^2 g_{(2)}(\alpha) \\ &= (\mu_1 - \beta g_{(1)}(\alpha))^2 + 2\mu\beta g_{(1)}(\alpha) + \beta^2 g_{(2)}(\alpha) \\ &= \mu_1^2 + \beta^2 (g_{(2)}(\alpha) - g_{(1)}^2(\alpha)). \end{aligned} \quad (4)$$

حال با استفاده از روابط (۲)، (۳)، و (۴) و انجام محاسبات ساده داریم:

$$\mu_3 = \mu^3 + 3\mu^2\beta g_{(1)}(\alpha) + 3\mu\beta^2 g_{(2)}(\alpha) + \beta^3 g_{(3)}(\alpha) \quad (5)$$

$$= \mu_1^3 + \left(\frac{\mu_2 - \mu_1^2}{g_{(2)}(\alpha) - g_{(1)}^2(\alpha)} \right)^{\frac{3}{2}} (2g_{(1)}^3(\alpha) - 3g_{(1)}(\alpha)g_{(2)}(\alpha) + g_{(3)}(\alpha)).$$

اکنون با به‌کارگیری روابط (۳)، (۴) و (۵)، جواب دستگاه $m_i = \mu_i$ ، $i = 1, 2, 3$ ، که همان برآوردهای گشتاوری هستند در این روابط صدق می‌کند:

$$\begin{aligned}\tilde{\mu} &= m_1 - \tilde{\beta} g_{(1)}(\tilde{\alpha}), \\ \tilde{\beta} &= \left(\frac{m_2 - m_1^2}{g_{(2)}(\tilde{\alpha}) - g_{(1)}^2(\tilde{\alpha})} \right)^{\frac{1}{2}}, \\ \frac{g_{(3)}(\tilde{\alpha}) - 3g_{(1)}(\tilde{\alpha})g_{(2)}(\tilde{\alpha}) + 2g_{(1)}^3(\tilde{\alpha})}{\left(g_{(2)}(\tilde{\alpha}) - g_{(1)}^2(\tilde{\alpha}) \right)^{\frac{3}{2}}} &= \frac{m_3 - m_1^3 - 3m_1(m_2 - m_1^2)}{(m_2 - m_1^2)^{\frac{3}{2}}}.\end{aligned}$$

واضح است که شرط وجود گشتاورهای اول، دوم و سوم این است که $\alpha > 3$. همچنین اگر $\tilde{\mu}$ بزرگ‌تر از اولین آماره ترتیبی باشد یعنی $\tilde{\mu} > x_{1:n}$ ، آن‌گاه برآوردهای گشتاوری وجود ندارد.

۲. روش گشتاورهای خطی (LME)

گشتاورهای خطی امید ریاضی ترکیب‌های خطی معینی از آماره‌های ترتیبی است که می‌توان آنها را برای هر متغیر تصادفی با میانگین متناهی تعریف کرد. مزیت عمده این روش نسبت به روش گشتاوری این است که برآوردهای گشتاورهای خطی تابعی خطی از داده‌ها بوده است و کم‌تر تحت تأثیر داده‌های پرت قرار می‌گیرد. این روش را هاسکینگ^۱ [۹] معرفی کرده است.

اگر X یک متغیر تصادفی با تابع توزیع تجمعی $F(\cdot)$ و تابع چنک $x(\cdot)$ باشد و آماره‌های ترتیبی $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ حاصل از یک نمونه تصادفی انتخابی به حجم n از X باشند، گشتاورهای خطی متغیر تصادفی بدین‌صورت تعریف می‌شوند:

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:r}) \quad r = 1, 2, 3, \dots \quad (۶)$$

همچنین امید ریاضی یک آماره ترتیبی را می‌توان به‌صورت (۷) محاسبه کرد:

$$E(X_{j:r}) = \frac{r!}{(j-1)!(r-j)!} \int_0^1 x(p) p^j (1-p)^{r-j} dp. \quad (۷)$$

با جای‌گذاری رابطه (۷) در رابطه (۶) و انجام محاسبات داریم:

$$\begin{aligned}\lambda_1 &= E(X) = \int_0^1 x(p) dp, \\ \lambda_2 &= \frac{E(X_{2:2} - X_{1:2})}{2} = \int_0^1 x(p) (2p - 1) dp, \\ \lambda_3 &= \frac{E(X_{3:3} - 2X_{2:3} + X_{1:3})}{3} = \int_0^1 x(p) (6p^2 - 6p + 1) dp.\end{aligned}$$

اغلب بهتر است گشتاورهای مراتب بالاتر $r \geq 3$ ، λ_r را استاندارد کرد، در این صورت مستقل از واحد اندازه‌گیری X خواهند شد. بنابراین نسبت‌های گشتاور خطی را می‌توان به صورت $\tau_r = \frac{\lambda_r}{\lambda_2}$ ، $r \geq 3$ تعریف کرد. از آن‌جاکه λ_r تابعی از امید ریاضی آماره‌های ترتیبی یک نمونه به اندازه r است، طبیعی است که با استفاده از آماره‌های ناریبی برآورد شوند. این آماره‌های ناریب را هافدینگ^۱ [۱۰] معرفی کردند که در زیر به آنها اشاره می‌کنیم.

فرض کنید $x_n, \dots, x_2, x_1$ مشاهدات یک نمونه تصادفی و $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ نمونه مرتب شده باشند. r -امین گشتاور خطی نمونه برابر است با

$$l_r = \binom{n}{r}^{-1} \sum_{1 \leq i_1} \sum_{i_2 \leq \dots} \dots \sum_{i_r \leq n} r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{i_r-k:n} \quad r = 1, 2, \dots, n.$$

بنابراین در حالت خاص داریم:

$$l_1 = n^{-1} \sum_{i=1}^n x_i,$$

$$l_2 = \frac{1}{2} \binom{n}{2}^{-1} \sum_{i>j} (x_{i:n} - x_{j:n}),$$

$$l_3 = \frac{1}{3} \binom{n}{3}^{-1} \sum_{i>j>k} (x_{i:n} - 2x_{j:n} + x_{k:n}).$$

در نتیجه نسبت‌های گشتاور خطی $\tau_r = \frac{\lambda_r}{\lambda_2}$ را می‌توان با $t_r = \frac{l_r}{l_2}$ ، $r \geq 3$ برآورد کرد.

گشتاورهای خطی توزیع کاپا سه پارامتری برابرند با

$$\lambda_1 = \mu + \beta \alpha^{\frac{1}{\alpha}-1} B\left(\frac{2}{\alpha}, 1 - \frac{1}{\alpha}\right),$$

$$\lambda_2 = 2\beta \alpha^{\frac{1}{\alpha}-1} B\left(\frac{3}{\alpha}, 1 - \frac{1}{\alpha}\right) - \beta \alpha^{\frac{1}{\alpha}-1} B\left(\frac{2}{\alpha}, 1 - \frac{1}{\alpha}\right),$$

$$\lambda_3 = 6\beta \alpha^{\frac{1}{\alpha}-1} B\left(\frac{4}{\alpha}, 1 - \frac{1}{\alpha}\right) - 6\beta \alpha^{\frac{1}{\alpha}-1} B\left(\frac{3}{\alpha}, 1 - \frac{1}{\alpha}\right) + \beta \alpha^{\frac{1}{\alpha}-1} B\left(\frac{2}{\alpha}, 1 - \frac{1}{\alpha}\right),$$

و در این صورت نسبت گشتاور خطی τ_3 تابعی از α می‌شود که برابر است با:

$$\tau_3 = \frac{\lambda_3}{\lambda_2} = \frac{6B\left(\frac{4}{\alpha}, 1 - \frac{1}{\alpha}\right) - 6B\left(\frac{3}{\alpha}, 1 - \frac{1}{\alpha}\right) + B\left(\frac{2}{\alpha}, 1 - \frac{1}{\alpha}\right)}{2B\left(\frac{3}{\alpha}, 1 - \frac{1}{\alpha}\right) - B\left(\frac{2}{\alpha}, 1 - \frac{1}{\alpha}\right)}$$

برای برآورد پارامترهای α ، β و μ از این روش، گشتاورهای خطی در توزیع کاپا را با کمیت‌های متناظر آن در نمونه مساوی قرار داده و برآوردهای گشتاور خطی پارامترهای α ، β و μ که با $\bar{\alpha}$ ، $\bar{\beta}$ و $\bar{\mu}$ نمایش می‌دهیم را بدین صورت به دست می‌آوریم:

$$l_1 = \mu + \beta g_{(1)}(\alpha)$$

در نتیجه

$$\bar{\mu} = l_1 - \hat{\beta} g_{(1)}(\bar{\alpha})$$

و

$$l_2 = \beta [2g_{(2)}(\alpha) - g_{(1)}(\alpha)]$$

بنابراین

$$\bar{\beta} = \frac{l_2}{2g_{(2)}(\bar{\alpha}) - g_{(1)}(\bar{\alpha})}$$

و $\bar{\alpha}$ را با برابر قرار دادن نسبت گشتاور خطی توزیع و نمونه محاسبه می‌کنیم. بنابراین

$$\tau_3 = t_3,$$

در نتیجه $\bar{\alpha}$ ریشه معادله $g_L(\bar{\alpha}) = \frac{l_3}{l_2}$ است. واضح است که اگر $\bar{\mu}$ بزرگ‌تر از اولین آماره ترتیبی باشدیعنی $\bar{\mu} > x_{1:n}$ ، آن‌گاه برآورد گشتاور خطی وجود ندارد.

۳. روش بیشینه درست‌نمایی (MLE)

تابع درست‌نمایی برای n مشاهده مستقل از توزیع کاپا سه پارامتری عبارت است از:

$$L(\mu, \alpha, \beta) = \begin{cases} \left(\frac{\alpha}{\beta}\right)^n \prod_{i=1}^n \left[\alpha + \left(\frac{x_i - \mu}{\beta}\right)^\alpha\right]^{-\left(\frac{\alpha+1}{\alpha}\right)} & \alpha, \beta > 0, x_i > \mu, \\ 0 & o.w \end{cases}$$

برآوردگرهای بیشینه درست‌نمایی پارامترهای α ، β و μ که با $\hat{\alpha}$ ، $\hat{\beta}$ و $\hat{\mu}$ نمایش داده می‌شوند تابع $L(\mu, \alpha, \beta)$ را به‌ازای پارامترهای α ، β و μ بیشینه می‌کند. از آن‌جاکه این تابع برای α و β ثابت، نسبت به پارامتر μ صعودی است، پس $\hat{\mu} = X_{1:n}$ حال با جای‌گذاری $\hat{\mu}$ در تابع $L(\mu, \alpha, \beta)$ و مشتق‌گیری از تابع $\ln L(\hat{\mu}, \alpha, \beta)$ نسبت به α و β و صفر قرار دادن آنها داریم:

$$\frac{n}{\hat{\alpha}} + \frac{1}{\hat{\alpha}^2} \sum_{i=2}^n \ln \left[\hat{\alpha} + \left(\frac{x_i - x_{1:n}}{\beta} \right)^{\hat{\alpha}} \right] - \left(\frac{\hat{\alpha} + 1}{\hat{\alpha}} \right) \sum_{i=2}^n \frac{1}{\hat{\alpha}_i} = 0,$$

$$\frac{-n}{\hat{\beta}} + \left(\frac{\hat{\alpha} + 1}{\hat{\alpha}} \right) \sum_{i=2}^n \frac{\hat{\alpha} \hat{\beta}^{-1} \hat{b}_i^{\hat{\alpha}}}{\hat{\alpha}_i} = 0,$$

که در آن $\hat{\alpha}_i = \hat{\alpha} + \hat{b}_i^{\hat{\alpha}}$ و $\hat{b}_i = \frac{x_i - x_{(1)}}{\beta}$ دو معادله مذکور به‌روش‌های عددی قابل حل هستند.

۴. روش بیشینه حاصل‌ضرب فاصله‌ها (MPSE)

روش MPSE یک روش کلی برای برآورد پارامترهای توزیع‌های تک‌متغیره است که به‌وسیله چنگ و امین [۱۲]،

[۱۱] معرفی شد. این روش به‌ویژه برای برآوردیابی مواقعی که تکیه‌گاه متغیر تصادفی بزرگ‌تر از یک پارامتر مجهول

است، مناسب است. این روش را به‌طور خلاصه شرح می‌دهیم:

فرض کنید $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ مشاهدات آماره‌های ترتیبی یک نمونه از توزیع $F(x; \theta)$ باشند. فاصله‌ها را به صورت

$$D_i(\theta) = F(x_{i:n}; \theta) - F(x_{i-1:n}; \theta), \quad i = 1, 2, \dots, n+1,$$

تعریف می‌کنیم، وقتی که $x_{0:n} = 0$ و $x_{n+1:n} = \infty$. در واقع D_i ها نشان‌دهنده مقدار اختلاف تابع توزیع در دو نقطه مجاور از نمونه تصادفی مرتب شده است. به سادگی می‌توان نشان داد که $\sum_{i=1}^{n+1} D_i(\theta) = 1$ روش MPSE روشی است که در آن با بیشینه کردن تابع $G(\theta) = (\prod_{i=1}^{n+1} D_i(\theta))^{n+1}$ یا به طور معادل لگاریتم آن نسبت به θ ، به برآوردگر کردن پارامتر θ پرداخته می‌شود. واضح است که این برآوردگر شکل بسته‌ای ندارد و با استفاده از روش‌های عددی محاسبه می‌شود. به طور کلی مزیت عمده روش MPSE آن است که برخلاف تابع درست‌نمایی که در برخی مواقع غیر کراندار است و نمی‌تواند برآوردگرهای سازگار تولید کند، $G(\theta)$ همواره کراندار است $G(\theta) \leq (n+1)^{-1}$ و می‌تواند تحت شرایط کلی تری نسبت به روش بیشینه درست‌نمایی برآوردگرهای سازگار ارائه دهد. همچنین برآوردگرهای حاصل از این روش، مانند برآوردگرهای بیشینه درست‌نمایی در صورت برقراری شرایط نظم، به طور مجانبی دارای توزیع نرمال و کاراست. در اینجا با جای‌گذاری تابع توزیع کاپا سه پارامتری و $\theta = (\mu, \alpha, \beta)$ به برآوردگرهای MPSE می‌رسیم. توجه کنید که مساوی نبودن داده‌ها یک شرط اساسی در این روش است.

پژوهش‌های شبیه‌سازی

در این بخش با ۱۰۰۰ بار شبیه‌سازی روش‌های مختلف برآورد از نظر اریبی (*bias*) و جذر میانگین توان دوم خطا (*rmse*) به ازای مقادیر مختلف n مقایسه می‌شوند. چون این توزیع برای داده‌های آب و هوایی به کار می‌رود و معمولاً حجم نمونه زیاد است، مقادیر n برابر با ۱۰۰، ۱۲۰، ۱۵۰، ۲۰۰، ۳۰۰ و ۵۰۰ در نظر گرفته شده است. *bias* و *rmse* روش‌های مختلف برآوردیابی در حالت $\mu = 0$ ، $\alpha = 4$ و $\beta = 0.5$ در جدول ۱ و در حالت $\mu = 1$ ، $\alpha = 5$ و $\beta = 2$ در جدول ۲ آمده است. توجه شود که روش‌های گشتاوری و گشتاوری خطی در حالتی که موجود بوده‌اند در نظر گرفته شده‌اند و تمام نتایج با استفاده از نرم افزار R به دست آمده‌اند. نتایج جدول‌های ۱ و ۲ نشان می‌دهند که برای تمامی روش‌های برآوردیابی، با افزایش حجم نمونه، *rmse* کاهش می‌یابد. همچنین برای حجم نمونه ثابت، روش MPSE علاوه بر این که در برآورد کردن تمامی پارامترها دارای کم‌ترین قدر مطلق اریبی است (به جز چند مورد که MLE دارای قدر مطلق اریبی کم‌تری است) دارای کم‌ترین *rmse* (به جز چند موارد اندکی که MLE دارای *rmse* کم‌تری است) است. بنابراین می‌توان نتیجه گرفت که روش برآوردیابی MPSE دارای بهترین عملکرد نسبت به سه روش برآوردیابی دیگر است.

جدول ۱. مقایسه روش‌های مختلف برآوردیابی در حالت $\mu = 0$ و $\alpha = 4$ و $\beta = 0.5$

n	روش	μ		α		β	
		bias	rmse	bias	rmse	bias	rmse
۱۰۰	MME	-۰/۰۴۲۴	۰/۰۶۲۴	۱/۲۸۹۸	۱/۶۳۴۷	۰/۱۰۶۸	۰/۱۲۵۲
	LME	-۰/۰۱۲۲	۰/۰۱۹۰	۰/۳۷۰۹	۱/۰۲۳۰	۰/۰۲۱۷	۰/۰۶۳۰
	MLE	-۰/۰۰۷۰	۰/۰۱۰۰	۰/۲۳۵۷	۰/۹۹۷۸	-۰/۰۰۲۳	۰/۰۵۷۲
	MPSE	-۰/۰۰۰۱	۰/۰۰۶۸	-۰/۰۹۵۶	۰/۸۶۲۲	-۰/۰۰۶۸	۰/۰۵۵۵
۱۲۰	MME	-۰/۰۳۹۹	۰/۰۵۴۷	۱/۲۴۴۷	۱/۵۶۸۱	۰/۰۹۷۶	۰/۱۱۲۹
	LME	-۰/۰۱۱۱	۰/۰۱۶۵	۰/۳۴۲۵	۰/۹۸۳۰	۰/۰۱۷۸	۰/۰۵۵۳
	MLE	-۰/۰۰۵۴	۰/۰۰۷۸	۰/۱۹۳۷	۰/۹۵۵۳	۰/۰۰۰۱	۰/۰۵۲۱
	MPSE	۰/۰۰۰۲	۰/۰۰۶۰	-۰/۱۱۴۹	۰/۷۷۱۲	-۰/۰۰۶۵	۰/۰۴۹۸
۱۵۰	MME	-۰/۰۳۷۴	۰/۰۵۶۱	۱/۱۳۲۶	۱/۴۱۴۷	۰/۰۹۵۸	۰/۱۱۲۸
	LME	-۰/۰۱۰۵	۰/۰۱۵۵	۰/۲۸۲۸	۰/۷۷۳۱	۰/۰۱۴۸	۰/۰۴۸۷
	MLE	-۰/۰۰۴۴	۰/۰۰۶۳	۰/۱۰۹۲	۰/۷۸۴۳	-۰/۰۰۴۱	۰/۰۴۵۳
	MPSE	۰/۰۰۰۲	۰/۰۰۴۸	-۰/۰۶۰۱	۰/۶۴۱۲	-۰/۰۰۳۶	۰/۰۴۴۰
۲۰۰	MME	-۰/۰۳۵۵	۰/۰۴۴۴	۱/۰۳۶۴	۱/۲۸۱۳	۰/۰۸۵۶	۰/۰۹۶۲
	LME	-۰/۰۰۹۶	۰/۰۱۳۵	۰/۲۰۳۳	۰/۶۳۹۵	۰/۰۱۳۴	۰/۰۴۱۰
	MLE	-۰/۰۰۳۶	۰/۰۰۵۱	۰/۰۷۸۰	۰/۶۰۲۲	-۰/۰۰۲۶	۰/۰۳۸۰
	MPSE	-۰/۰۰۰۱	۰/۰۰۳۳	۰/۰۱۷۰	۰/۶۲۸۱	-۰/۰۰۱۹	۰/۰۳۷۷
۳۰۰	MME	-۰/۰۳۲۵	۰/۰۴۶۲	۰/۸۴۱۷	۰/۰۸۴۷	۰/۰۷۷۹	۰/۰۹۰۰
	LME	-۰/۰۰۷۵	۰/۰۱۰۷	۰/۱۵۷۰	۰/۵۱۹۵	۰/۰۱۰۱	۰/۰۳۷۷
	MLE	-۰/۰۰۲۳	۰/۰۰۳۳	۰/۰۲۷۹	۰/۴۹۸۹	-۰/۰۰۱۶	۰/۰۳۰۸
	MPSE	-۰/۰۰۰۰	۰/۰۰۲۲	-۰/۰۷۱۸	۰/۴۸۰۳	-۰/۰۰۴۵	۰/۰۳۰۳
۵۰۰	MME	-۰/۰۲۸۹	۰/۰۳۴۵	۰/۶۹۴۰	۰/۹۱۷۲	۰/۰۶۶۵	۰/۰۷۵۹
	LME	-۰/۰۰۳۵	۰/۰۰۸۳	۰/۱۱۶۳	۰/۴۱۰۳	۰/۰۰۰۱	۰/۰۱۰۲
	MLE	-۰/۰۰۱۳	۰/۰۰۱۸	۰/۰۴۰۸	۰/۳۸۱۴	-۰/۰۰۰۴	۰/۰۲۴۹
	MPSE	-۰/۰۰۰۰	۰/۰۰۱۳	-۰/۰۳۸۱	۰/۳۶۳۱	-۰/۰۰۲۹	۰/۰۲۳۲

مثال کاربردی

در این بخش، سری زمانی مربوط به مجموع بارش ماهانه ایستگاه آبدلی استان تهران بر حسب میلی‌متر در ماه ژانویه طی سال‌های ۱۹۸۳ تا ۲۰۱۴ یعنی ۳۲ سال در نظر گرفته شده است که در سایت سازمان هواشناسی کشور قابل دسترسی است و این داده‌ها تا یک رقم اعشار گرد شده‌اند. از آن‌جاکه داده‌های در نظر گرفته شده مربوط به مجموع بارش ماه ژانویه طی سال‌های مختلف است، بنابراین می‌توان آنها را به‌صورت مستقل در نظر گرفت.

هر داده را با عددی تصادفی تولید شده از توزیع یکنواخت در بازه $(0, 0.5)$ با حداکثر خطای مطلق 0.05 جمع کرده و تا ۴ رقم اعشار گرد کرده‌ایم. داده‌های جدید را داده‌های تبدیل یافته می‌نامیم. لازم به یادآوری است که داده‌های اصلی تا یک رقم اعشار گرد شده‌اند و چون برای گرد کردن اعداد تا یک رقم اعشار، خطای مطلق با بیش‌ترین مقدار 0.05 در نظر گرفته می‌شود، بنابراین از چنین تبدیلی استفاده شده است. به‌علاوه، استفاده از این تبدیل باعث از بین رفتن داده‌های تکراری موجود در نمونه می‌شود و می‌توان برآوردیابی به‌روش MPSE را نیز انجام داد. برای این داده‌ها برآوردهای MME، LME، MLE و MPSE پارامترهای توزیع کاپا سه پارامتری، آماره آزمون کولموگروف اسمیرنوف (KS) و هم‌چنین محاسبه p -value شده و نتایج در جدول ۳ ارائه شده است. محاسبات و نتایج مربوط به این مثال نیز با استفاده از نرم افزار R به‌دست آمده است.

جدول ۲. مقایسه روش‌های مختلف برآوردیابی در حالت $\mu = 1$ ، $\alpha = 5$ و $\beta = 2$

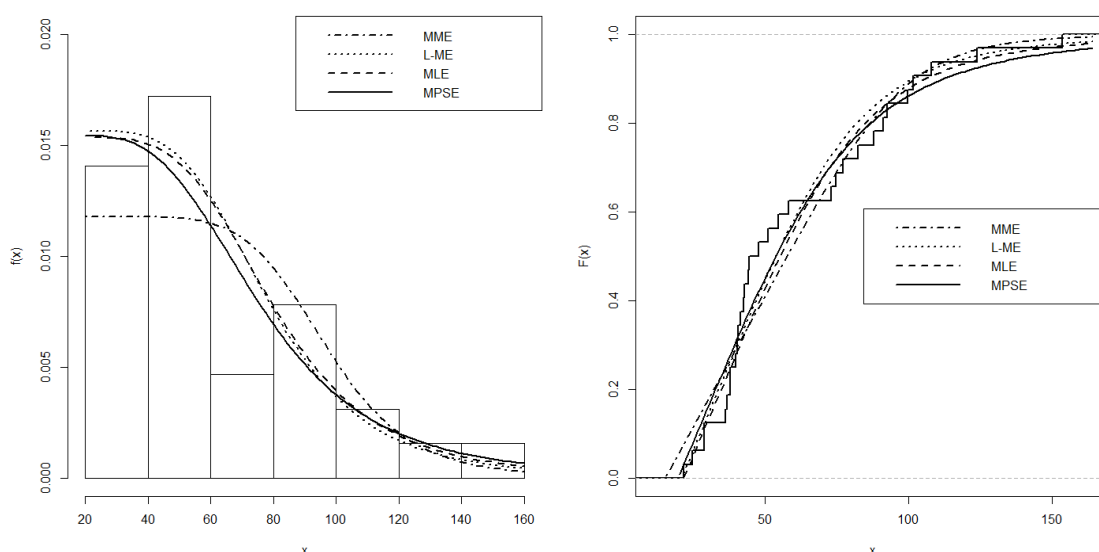
n	روش	μ		α		β	
		bias	rmse	bias	rmse	bias	rmse
۱۰۰	MME	-۰/۰۵۹۰	۰/۱۱۳۷	۱/۳۰۸۰	۱/۹۴۹۱	۰/۲۰۳۸	۰/۲۸۵۴
	LME	-۰/۰۵۴۵	۰/۰۷۰۲	۰/۳۹۸۴	۱/۴۴۵۳	۰/۰۴۹۵	۰/۲۱۹۷
	MLE	-۰/۰۲۸۸	۰/۰۳۹۸	۰/۴۴۹۶	۱/۴۱۳۰	-۰/۰۰۵۳	۰/۲۱۷۷
	MPSE	-۰/۰۰۲۲	۰/۰۳۰۱	-۰/۱۳۴۹	۱/۳۶۵۴	-۰/۰۳۰۷	۰/۲۱۷۴
۱۲۰	MME	-۰/۰۵۸۱	۰/۱۱۱۴	۱/۱۳۶۱	۱/۷۴۹۹	۰/۱۸۹۲	۰/۲۶۷۱
	LME	-۰/۰۴۰۶	۰/۰۶۴۳	۰/۳۱۴۳	۱/۳۷۷۶	۰/۰۳۵۵	۰/۲۰۹۳
	MLE	-۰/۰۲۳۲	۰/۰۳۲۹	۰/۳۷۴۶	۱/۳۸۳۶	-۰/۰۰۳۴	۰/۲۰۶۳
	MPSE	-۰/۰۰۰۸	۰/۰۲۳۶	-۰/۱۵۷۱	۱/۰۴۸۱	-۰/۰۳۴۱	۰/۱۹۰۵
۱۵۰	MME	-۰/۰۵۸۹	۰/۱۰۴۰	۱/۰۱۶۷	۱/۵۴۴۸	۰/۱۸۰۱	۰/۲۵۰۶
	LME	-۰/۰۳۶۷	۰/۰۵۷۴	۰/۳۲۹۱	۱/۱۱۸۴	۰/۰۵۱۴	۰/۱۸۴۰
	MLE	-۰/۰۱۸۵	۰/۰۲۵۷	۰/۲۱۷۷	۱/۰۴۱۶	-۰/۰۰۲۰	۰/۱۶۵۳
	MPSE	-۰/۰۰۰۱	۰/۰۱۸۱	-۰/۱۵۰۱	۰/۹۲۸۴	-۰/۰۲۳۰	۰/۱۷۱۴
۲۰۰	MME	-۰/۰۵۶۷	۰/۰۹۳۵	۰/۸۶۲۲	۱/۳۴۰۶	۰/۱۶۱۴	۰/۲۲۵۶
	LME	-۰/۰۳۵۶	۰/۰۵۱۴	۰/۲۱۰۵	۰/۹۱۱۴	۰/۰۳۷۷	۰/۱۶۰۴
	MLE	-۰/۰۱۳۳	۰/۰۱۸۵	۰/۱۸۲۳	۰/۹۱۸۹	-۰/۰۰۱۷	۰/۱۵۰۷
	MPSE	-۰/۰۰۰۴	۰/۰۱۳۴	-۰/۱۲۴۲	۰/۷۹۱۴	۰/۰۱۸۱	۰/۱۴۲۰
۳۰۰	MME	-۰/۰۴۳۹	۰/۰۸۲۵	۰/۶۳۴۸	۱/۰۵۷۷	۰/۱۲۵۴	۰/۱۹۳۶
	LME	-۰/۰۳۱۱	۰/۰۴۲۶	۰/۱۴۷۲	۰/۷۱۲۷	۰/۰۳۴۵	۰/۱۲۶۱
	MLE	-۰/۰۰۹۱	۰/۰۱۳۲	۰/۱۰۹۸	۰/۷۰۳۷	-۰/۰۰۳۳	۰/۱۱۶۰
	MPSE	-۰/۰۰۰۰	۰/۰۰۹۲	-۰/۱۲۴۴	۰/۶۵۴۰	-۰/۰۲۲۱	۰/۱۲۲۸
۵۰۰	MME	-۰/۰۳۸۳	۰/۰۶۳۳	۰/۴۸۳۵	۰/۸۸۱۵	۰/۱۰۲۵	۰/۱۶۶۲
	LME	-۰/۰۰۷۹	۰/۰۰۷۲	۰/۰۲۴۶	۰/۵۲۴۸	-۰/۰۱۹۸	۰/۰۹۰۳
	MLE	-۰/۰۰۵۸	۰/۰۰۷۴	۰/۰۲۵۲	۰/۵۲۶۰	-۰/۰۱۰۴	۰/۰۹۰۱
	MPSE	-۰/۰۰۰۰	۰/۰۰۵۲	-۰/۰۸۴۰	۰/۴۶۶۰	-۰/۰۱۴۳	۰/۰۸۹۵

چنان‌که مشاهده می‌شود به کارگیری هر چهار روش برآوردیابی منجر به برازش مناسب توزیع کاپا سه پارامتری به داده‌ها می‌شود، اما برآورد MPSE بیش‌ترین برازش را از نظر آماره KS و $p\text{-value}$ به‌دست می‌دهد.

در شکل ۲، نمودار سمت چپ، هیستوگرام داده‌های تبدیل یافته میزان بارش به‌همراه تابع چگالی کاپا سه پارامتری برازش داده شده به‌وسیله چهار روش برآوردیابی را نشان می‌دهد و در نمودار سمت راست، تابع توزیع تجربی داده‌ها و تابع توزیع کاپا سه پارامتری برازش یافته بر اساس این چهار روش برآوردیابی ارائه شده است که هر دو نمودار برازش مناسب روش MPSE را تأیید می‌کنند.

جدول ۳. برآوردهای MME، LME، MLE و MPSE پارامترهای توزیع کاپا سه پارامتری، آماره KS و $p\text{-value}$ متناظر برای داده‌های تبدیل یافته

روش	μ	α	β	KS	$p\text{-value}$
MME	۱۵/۱۸۶۴	۵/۸۸۶۳	۶۲/۷۵۷۴	۰/۱۵۶۲	۰/۳۷۶۲
LME	۲۱/۰۳۲۶	۳/۴۹۸۹	۴۴/۶۸۵۹	۰/۱۳۸۲	۰/۵۳۸۸
MLE	۲۱/۶۳۲۶	۳/۳۳۹۰	۴۵/۴۳۴۱	۰/۱۵۴۶	۰/۳۸۹۱
MPSE	۱۹/۶۵۳۹	۲/۸۸۷۲	۴۴/۸۹۹۸	۰/۱۲۹۸	۰/۶۰۷۹



شکل ۲. میزان برازش توزیع کاپا سه پارامتری روی داده‌ها با استفاده از روش‌های مختلف برآوردیابی

بحث و نتیجه‌گیری

در این مقاله به بررسی توزیع کاپا سه پارامتری پرداختیم. سپس چهار روش برآوردیابی شامل روش گشتاوری، گشتاورهای خطی، بیشینه درست‌نمایی و بیشینه حاصل‌ضرب فاصله‌ها را برای برآوردیابی پارامترهای این توزیع در نظر گرفتیم. با انجام یک بررسی شبیه‌سازی، به مقایسه عملکرد این چهار روش برآوردیابی پرداختیم. نتایج این شبیه‌سازی نشان می‌دهد که روش بیشینه حاصل‌ضرب فاصله‌ها (MPSE) دارای قدر مطلق اریبی و جذر میانگین توان دوم خطای کمتری نسبت به بقیه روش‌ها است. در پایان، این روش‌ها، برای یک مجموعه داده واقعی به کار گرفته شد که با به‌دست آوردن آماره و p -value آزمون کولموگروف اسمیرنوف حاصل از برازش روش‌های مختلف برآوردیابی توزیع کاپا سه پارامتری، روش MPSE برازش بهتری نسبت به سایر روش‌ها به این داده‌ها دارد که به این ترتیب نتیجه حاصل از شبیه‌سازی نیز تأیید می‌شود.

سپاس‌گزاری

از داوران محترم به‌دلیل پیشنهادهای ارزنده‌شان که باعث بهبودی مقاله شد، تشکر می‌کنیم.

منابع

1. Park J. S., Seo S. C., Kim T. Y., "A kappa distribution with a hydrological application", Stochastic environmental research and risk assessment, 23 (2009) 579-586.
2. Mielke Jr P. W., "Another family of distributions for describing and analyzing precipitation data", Journal of Applied Meteorology, 12 (1973) 275-280.
3. Mielke P. W, Johnson E. S, "Three-Parameter kappa distribution maximum likelihood estimates and likelihood ratio tests", Monthly Weather Review, 101 (1973) 701-707.

4. Ashkar F., Bayentin J., Savoie N., El-Jabi N., "Frequency analysis of low stream flow characteristics using statistical distributions", Critical Transitions in Water and Environmental Resources Management, 4th International Water Association World Water Congress(2004).
5. Ashkar F., Aucoin F., Bayentin L., "Parameter and quantile estimation of the 2-parameter kappa distribution by maximum likelihood", Stochastic Environmental Research and Risk Assessment, 26 (2012) 1025-1039.
6. Ashour S. K., Elsherpieny E. A., Abdel-all Y. Y., "Parameter estimation for three-parameter kappa distribution under Type-II censored samples", Journal of Applied Sciences Research, 5 (2009) 1762-1766.
7. Hosking J. R. M., "The four-parameter kappa distribution", IBM Journal of Research and Development, 38 (1994) 251-258.
8. حسن‌زاده یوسف، عبدی کردانی امین، "استفاده از توزیع چهار پارامتری کاپا در تخمین شاخص بارندگی استاندارد شده"، علوم و مهندسی آبیاری (مجله علمی کشاورزی)، ۳۵ (۱۳۹۱) ۲۱-۳۱.
9. Hosking J. R. M., "L-moments: analysis and estimation of distributions using linear combinations of order statistics", Journal of the Royal Statistical Society B, (1990) 105-124.
10. Hoeffding, W., "A class of statistics with asymptotically normal distribution", The Annals of Mathematical Statistics, 19 (1948) 293-325.
11. Cheng R. C. H., Amin N. A. K., "Maximum product of spacings estimation with application to the lognormal distribution", Math Report, 791 (1979).
12. Cheng R. C. H., Amin N. A. K., "Estimating parameters in continuous univariate distributions with a shifted origin", Journal of the Royal Statistical Society B, (1983) 394-403.