مقایسه روش های برآورد چندک‌ها

مهران شمشیربندی[۱] ابوالقاسم بزرگنیا[۲]

چکیده:

در برآورد پارامترهای یک مدل، همیشه انتخاب بهترین براورد پارامتر و به تبع آن بهترین روش برآورد مورد نظر بوده است. در این مقاله با استفاده از سه روش گشتاورهای کلاسیک، گشتاورهای خطی و گشتاورهای خطی خطی، پارامترهای مدل‌هایی با توزیع‌های نمایی و گامبل که با استفاده از چندک‌ها معرفی شده‌اند، مورد مقایسه قرار گرفته‌اند با استناد به اصل کمینگی واریانس، اریبی و مجموع مربع خطای برآورد پارامترهای برآورد شده نشان داده خواهد شد که روش گشتاورهای کلاسیک نسبت به سایر روش‌ها بهتر است.

واژه‌های کلیدی: گشتاورهای کلاسیک، گشتاورهای خطی[۳]، گشتاورهای خطی خطی[۴].

۱ مقدمه

از جمله کارهای اساسی در آمار تطابق یک مدل با مجموعه‌ای از داده‌هاست. بدین منظور پارامترهای مدل باید به گونه‌ای برآورد شوند که برازشی مناسب بین داده‌ها و مدل برازش داده شده ارائه شود. برای تشخیص رفتار یک سری از داده‌ها و در واقع مدل‌بندی آن‌ها می‌توان از توابع چندک توزیع‌ها استفاده کرد.[۱]

روش‌های مختلفی برای برآورد پارامترهای مدل استفاده می‌شود. در این مقاله با استفاده از سه روش برآورد گشتاورهای کلاسیک، گشتاور خطی و گشتاورهای خطی خطی پارامترهای مدل‌هایی که دارای توزیع گامبل و نمایی هستند، در سه اندازه نمونه که به روش مونت کارلو تولید شده‌اند برآورد خواهند شد و سپس با استفاده از سه خاصیت اریبی و کمینگی واریانس و مجموع مربعات خطای برآورد پارامترهای برآورد شده این سه روش برآورد نشان داده‌ایم که روش برآورد گشتاورهای کلاسیک در سه اندازه نمونه یاد شده در سه خاصیت بهترین رفتار را نشان می‌دهد.

۲ مدل‌سازی با استفاده از تابع چندک توزیع

فرض کنید مقادیر مرتب شده $x_1, x_2, \ldots, x_n$ مشاهده شده باشند. توزیع احتمال $F(x)$ را می‌توان با استفاده از پارامترهای مکان A و پارامتر مقیاس B به صورت زیر خطی کرد. تابعی مانند g می‌توان یافت به طوری که:

$$g\left[F\left(x\right)\right] = \frac{x_p - A}{B}$$

---
[۱] دانشگاه آزاد اسلامی واحد اراک
[۲] موسسه عالی غیر انتفایی خیام
[۳] L-Moments
[۴] LL-Moments

بنابراین داریم:

$$A + B.g\left(F\left(x\right)\right) = x_p \qquad (۱)$$

که در آن $g(F(x))$ معادل تابع چندک توزیع است، که با Sp نشان داده شده است. تابع Sp، وارون تابع توزیع، مورد نظر است. p احتمال است و تابع درحالت استاندارد حالتی است که درآن پارامتر مکانی صفر و پارامتر مقیاس یک فرض می‌شود. بنابراین رابطه (۱) را می‌توان به صورت زیر بازنویسی کرد:

$$Q\left(p\right) = A + BSp \qquad (۲)$$

که در آن، $Q(p)$ معادل تابع چندک توزیع است که از خطی کردن تابع چندک توزیع در حالت استاندارد حاصل شده است. دو پارامتر A و B موجود در رابطه (۲) پارامترهای مجهول $Q(p)$ هستند. در این کار هدف مقایسه روش‌های برآوردی است که این دو پارامتر را می‌توان با آن روش‌ها برآورد کرد. [۷]

برای مثال Sp توزیع گامبل به صورت زیر محاسبه می‌شود:

$$F_X(x) = \exp(-\exp\{-\tfrac{x_p-\xi}{\theta}\}) = p$$

$$\ln p = -\exp\{-\tfrac{x_p-\xi}{\theta}\} - \ln(-\ln p) = \tfrac{x_p-\xi}{\theta}$$

و در نهایت:

$$x_p = \xi - \theta\ln(-\ln p)$$

بنابراین تابع چندک توزیع گامبل به صورت زیر خواهد بود:

$$Q(p) = F^{-۱}(p) = \xi - \theta\ln(-\ln p)$$

با فرض $A = \xi$ و $B = \theta$، تابع چندک توزیع گامبل را به فرم تعریف شده در رابطه (۲) تعریف می‌کنیم:

$$Q(p) = A - B\ln(-\ln p)$$

۳ روش‌های بر آورد

روش‌های برآوردی که در این مقاله مورد مقایسه قرار می گیرند، عبارتند از گشتاورهای کلاسیک، روش گشتاورهای‌$-L$ (گشتاورهای خطی) و روش گشتاورهای‌$-LL$ (گشتاورهای خطی خطی). در ابتدا به اختصار این سه روش را معرفی می‌کنیم:

الف) روش گشتاورهای کلاسیک [۴] : یکی از قدیمی‌ترین روش‌های برآوردیابی، روش برآورد گشتاوری است که در سال ۱۸۹۴ توسط کارل پیرسن [۵] معرفی شد. اساس برآورد، در روش گشتاوری، مساوی قرار دادن خواص نمونه‌ای با خواص جامعه است. خواص جامعه به صورت توابعی از پارامترهای در حال برآورد است. بدین ترتیب که به تعداد پارامترهای مجهول، معادلاتی حاصل می‌شود که حل این معادلات مقادیری را برای این پارامتر ها ارائه می‌دهد. در برآورد مدلی با k پارامتر، اگر خواص جامعه در مدل فرضی، برای برازش را با $G_i(\theta)$ و معادل نمونه‌ای آن را با g_i نمایش دهیم، معادلات برآورد به شکل زیر است:

$$G_i\left(\theta\right) = g_i \qquad i = ۱, ۲, \cdots k$$

ب) روش گشتاورهای خطی: هاسکینگ [۶] گشتاورهای خطی را با استفاده از امید ریاضی یک ترکیب خطی از

---
[۵] Karl Pearson
[۶] Hosking

آماره‌های ترتیبی محاسبه کرد.

استفاده از گشتاورهای خطی بسیار آسان است، زیرا می‌توان آنها را به طور مستقیم به عنوان اندازه‌هایی از موقعیت مقیاس و شکل یک تابع توزیع تفسیر کرد. گشتاورهای خطی مرتبه r ام از متغیر تصادفی X به صورت زیر تعریف شده است:

$$\lambda_r = r^{-1}\sum_{k=0}^{r-1}(-1)^k E\left[X_{r-k|r}\right] \qquad r=1,2,\cdots \tag{3}$$

که در آن $X_{r-k|r}$ عنصر $r-k$ ام از نمونه مرتب شده‌ای به اندازه r است. دو گشتاور خطی اول عبارتند از:

$$\lambda_1 = E\left[X_{1|1}\right]$$

$$\lambda_2 = \frac{1}{2}E\left[X_{2|2} - X_{1|2}\right]$$

λ_1 معادل پارامتر مکانی A و λ_2 به عنوان مقیاسی برای پراکندگی در نظر گرفته شده و معادل پارامتر مقیاسی B است. [6]

رابطه بین گشتاورهای خطی و تابع توزیع و $F(x)$ به صورت زیر نشان داده شده است [2]:

$$E\left[X_{J|r}\right] = \frac{r!}{(j-1)!\,(r-j)!}\int_0^1 Q(p)\, p^{j-1}(1-p)^{r-j}\,dp$$

پس λ_1 و λ_2 عبارتند از:

$$\lambda_1 = \int_0^1 Q(p)dp \tag{4}$$

$$\lambda_2 = \int_0^1 Q(p)(2p-1)\,dp \tag{5}$$

از آن‌جا که در بسیاری از توزیع‌ها، فرم بسته‌ای برای انتگرال روابط (۴) و (۵) نمی‌توان یافت و یا در برخی موارد محاسبه انتگرال بسیار مشکل است، لذا وانگ[7] گشتاورهای خطی را با استفاده از نمونه‌های مرتب شده به اندازه n توسط روابط زیر برآورد کرد. [2]:

$$\lambda_1 = \sum_{i=1}^{n} x_{p_i} \Big/ \binom{n}{1} \tag{6}$$

$$\lambda_2 = \frac{\sum_{i=1}^{n}\left[\binom{i-1}{1} - \binom{n-i}{1}\right]x_{p_i}}{2\binom{n}{2}} \tag{7}$$

ج) روش گشتاورهای خطی خطی:

اونوز[8] و بایازیت[9] در سال ۲۰۰۱ روش برآورد پارامتری گشتاورهای خطی خطی را معرفی کردند. گشتاورهای خطی خطی ترکیب خطی از گشتاورهای خطی هستند. در این روش وزن بیشتری به داده‌های کوچکتر داده می‌شود. در نمونه‌های ترتیبی به اندازه n گشتاورهای خطی خطی برای نمونه‌ای به اندازه $r+m(m=1,2,\cdots)$ تعریف می‌شوند و با امید ریاضی‌گرفتن از r تا، از کوچکترین عناصر نمونه به صورت زیر محاسبه می شود:

$$\lambda_{Lr}^{m} = r^{-1}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k} E\left[X_{r-k|r+m}\right] \qquad r=1,2,3,\cdots$$

و دو گشتاور اول به صورت زیر معرفی می‌شوند:

---
[7] Wang
[8] Önöz
[9] Bayazit

$$\lambda_{L1}^{m} = \frac{1}{2} r^{-1} E\left[X_{2|2+m} - X_{1|2+m}\right]$$
$$\lambda_{L2}^{m} = E\left[X_{1|1+m}\right]$$

توزیع هستند:

$$\lambda_{l1}^{m} = (1+m)\int_{0}^{1} Q(p)\,(1-p)^{m}\,dp \qquad (8)$$

و

$$\lambda_{l2}^{m} = \frac{1}{2}(1+m)(2+m)\int_{0}^{1} Q(p)p(1-p)^{m}dp \qquad (9)$$
$$- \frac{1}{2}(2+m)\int_{0}^{1} Q(p)(1-p)^{1+m}dp$$

در این روش برآورد نیز چون در بسیاری از موارد فرم بسته و تعریف شده‌ای برای انتگرال روابط (۸) و (۹) نمی‌توان یافت و یا انتگرال‌گیری از آن کار بسیار مشکلی است، بنابراین با استفاده از روابط معرفی شده توسط وانگ روابط زیر را برآورد پارامترها با استفاده از نمونه، بیان شده است:

$$\lambda_{L1}^{m} = \frac{\sum_{i=1}^{n} \binom{n-i}{m} x_{p_i}}{\binom{n}{1+m}} \qquad (10)$$

و

$$\lambda_{L2}^{m} = \frac{\sum_{i=1}^{n} \left[\binom{n-i}{m} - \binom{i-1}{1} - \binom{n-i}{1+m}\right] x_{p_i}}{2\binom{n}{2+m}} \qquad (11)$$

با فرض $m = 0$ گشتاورهای خطی خطی به گشتاورهای خطی تبدیل می‌شوند. [۲]

۴ بهترین برآوردها

روش های بسیاری برای تمایز قائل شدن و ترجیح دادن روش های برآورد، بر یکدیگر وجود دارد. در این کار با مقایسه مجموع مربعات خطاها، اریبی و واریانس، در هر سه روش، بهترین روش برگزیده می‌شود.

الف) مجموع مربعات خطا: اگر $\hat{Q}(p)$ یک برآوردگر نامعلوم x_p باشد، آنگاه:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left|\hat{Q}(p) - x_{p_i}\right|^2$$

مجموع مربعات خطاهاست و هدف مینیمم کردن این مقدار است. [۱]

ب) اریبی: اگر $\hat{Q}(p)$ برآوردگر نامعلوم x_p باشد، آنگاه:

$$B\left(\hat{Q}(P)\right) = E\left(\hat{Q}(p)\right) - x_p$$

اریبی است. بهترین روش برآورد، دارای کمترین اریبی است و برای داده‌های مستقل و هم توزیع، داریم:

$$E\left(\hat{Q}(p)\right) = x_p - \frac{p(1-p)\,s(p)}{2(n+2)\,s^3(p)}$$

که در آن $s(p)$ معادل تابع چگالی چندک توزیع است. $s(p)$ از مشتق $Q(p)$ نسبت به p حاصل می‌شود.[۵]

ج) واریانس: اگر $s(p)$ معادل تابع چگالی باشد، آنگاه:

$$Var\left(\hat{Q}(p)\right) = \frac{p(1-p)}{(n+2)\,s^2(p)}$$

واریانس $Q(p)$ است [۵]. بهترین روش برآورد، روشی است که دارای کمترین واریانس باشد.

توزیع‌های مورد مطالعه:

در این کار با استفاده از شبیه‌سازی مونت کارلو سه نمونه تصادفی برای هر یک از توزیع‌های نمایی و گامبل با اندازه‌های ۱۰، ۳۰ و ۱۰۰ تولید شده است. هدف یافتن

پاسخ هر یک از اندازه‌های این دو توزیع در سه روش برآورد یاد شده است.

الف) توزیع نمایی: توزیع نمایی ($E(\lambda)$) دارای تابع چندک توزیعی به صورت زیر است:

$$Q(p) = A - B\ln(۱-p) \qquad (۱۲)$$

که از وارون تابع توزیع حاصل شده است. تابع چگالی چندک توزیع نمایی نیز به صورت زیر است:

$$s(p) = \frac{\partial Q(p)}{\partial p} = \frac{B}{۱-p}$$

ب) توزیع گامبل: توزیع گامبل دارای تابع چندک توزیعی به صورت زیر است:

$$\begin{aligned} Q(p) &= A - B\ln(-\ln p) \qquad (۱۳) \\ s(p) &= \frac{\partial Q(p)}{\partial p} = \frac{B}{p\ln p} \end{aligned}$$

۵ مقایسه روش‌های برآورد چندک‌ها برای دو توزیع نمایی و گامبل

۱.۵ روش گشتاورهای کلاسیک:

الف) توزیع نمایی: توزیع نمایی دارای تابع چندک توزیعی به صورت (۱۲) است. اگر چندک‌های نمونه‌ای $p, q = ۱-p$ به ترتیب برابر با x_p, x_q باشند، آنگاه با مساوی قرار دادن این مقادیر با مقادیر متناظر برای جامعه‌ای از توزیع نمایی داریم:

$$\begin{cases} Q(p) = A - B\ln(۱-p) = x_p \\ Q(۱-p) = A - B\ln(۱-۱+p) = x_q \end{cases}$$

از حل این دستگاه داریم:

$$\hat{A} = x_q + \left[\frac{x_p - x_q}{\ln p - \ln(۱-p)}\right]\ln p$$

$$\hat{B} = \frac{x_p - x_q}{\ln p - \ln(۱-p)} \qquad (۱۵)$$

با جایگزینی روابط (۱۴) و (۱۵) در رابطه (۱۲) مقدار چندک جامعه برای هر سه اندازه تولید شده، برآورد خواهد شد.

ب) توزیع گامبل: توزیع گامبل دارای تابع چندک توزیعی به صورت رابطه (۱۳) است. اگر چندک‌های نمونه‌ای $p, q = ۱-p$ را به ترتیب با x_q, x_p نشان دهیم، از حل دستگاه زیر:

$$\begin{cases} Q(p) = A - B\ln(-\ln(p)) = x_p \\ Q(۱-p) = A - B\ln(-\ln(۱-p)) = x_q \end{cases}$$

خواهیم داشت:

$$\hat{A} = x_p + \frac{x_p - x_q}{\ln\left[\frac{\ln(۱-p)}{\ln p}\right]} . \ln[-\ln p] \qquad (۲۹)$$

$$\hat{B} = \frac{x_p - x_q}{\ln\left[\frac{\ln(۱-p)}{\ln p}\right]}$$

با جایگزینی دو رابطه اخیر در رابطه (۱۳)، تابع چندک توزیع در جامعه را برآورد می‌کنیم.

۲.۵ روش گشتاورهای خطی:

برای برآورد پارامترهای تابع چندک هر دو توزیع چون با استفاده از روابط (٦) و (۷) به فرم بسته‌ای از انتگرال نمی‌توان دست یافت. بنابراین با استفاده از روابط (٦) و (۷) مقادیر این دو پارامتر را برآورد و با استفاده از آن دو، مقادیر تابع چندک برآورد می‌شود [۲].

۳.۵ روش گشتاورهای خطی خطی:

در این روش نیز مانند روش گشتاورهای خطی رسیدن به فرم بسته‌ای از انتگرال‌های روابط (۸) و (۹) دشوار است. بنابراین برای برآورد پارامترهای مذکور از روابط (۱۰) و (۱۱) استفاده می‌شود [۲].

۶ نتیجه‌گیری

با توجه به معیارهایی که در بخش (۳) بیان شد، چندک‌های توزیع هر دو توزیع مورد نظر در سه اندازه نمونه و با استفاده از سه روش یاد شده با یکدیگر مقایسه شده‌اند. ملاحظه می‌شود که روش گشتاورهای کلاسیک در هر سه اندازه نمونه دارای کمترین مقدار مجموع مربعات خطاست. (جدول شماره ۱). همچنین، در هر سه اندازه نمونه یاد شده و در هر دو توزیع، گشتاورهای کلاسیک دارای کمترین واریانس است. (جدول شماره ۲).

در بین این سه روش در هر سه اندازه نمونه روش گشتاورهای کلاسیک دارای کمترین اریبی است. (جدول شماره ۳).

جدول ۱: مقادیر مجموع مربعات خطای سه روش برآورد در دو توزیع نمایی و گامبل

نوع توزیع		توزیع گامبل			توزیع نمایی		
		حجم نمونه					
روش براورد ↓		$n=۱۰$	$n=۳۰$	$n=۱۰۰$	$n=۱۰$	$n=۳۰$	$n=۱۰۰$
گشتاورهای کلاسیک		$۳/۸۶۷\times ۱۰^{-۱۱}$	$۶/۱۴\times ۱۰^{-۷}$	$۸/۴۱\times ۱۰^{-۵}$	$۲/۲۸\times ۱۰^{-۶}$	$۷/۹۴\times ۱۰^{-۵}$	$۱/۶۵\times!۱۰^{-۷}$
گشتاورهای L		۲۹/۸۹۵	۱۹/۵۳۶	۲/۴	۵۱/۹۷	۹/۲۹۳	۴/۵۶۴۴
گشتاورهای LL	$m=۱$	۵۶/۶۵	۱۶/۷۱	۱/۹۳	۳۴/۳۹۱	۶/۳۸۷	۱/۵۹۴
گشتاورهای LL	$m=۲$	۸۸/۹۵	۱۵/۴۵	۲/۸۴۶	۵۰/۹۷	۵/۸۱	۲/۱۲۵

جدول ۲: مقادیر واریانس سه روش برآورد در دو توزیع نمایی و گامبل

نوع توزیع		توزیع نمایی			توزیع گامبل		
		حجم نمونه					
روش برآورد ↓		$n = ۱۰$	$n = ۳۰$	$n = ۱۰۰$	$n = ۱۰$	$n = ۳۰$	$n = ۱۰۰$
گشتاورهای کلاسیک		$۹/۷۲۹۵ \times ۱۰^{-۵}$	$۴/۸۱۲ \times ۱۰^{-۵}$	$۲/۲۶۶ \times ۱۰^{-۶}$	$۲/۲۶۶ \times ۱۰^{-۵}$	$۲/۲۶۶ \times ۱۰^{-۵}$	$۹/۷۱۳ \times ۱۰^{-۶}$
گشتاورهای L		$۹/۳۴۹ \times ۱۰^{-۴}$	$۳/۳۵۱ \times ۱۰^{-۴}$	$۴/۸۱ \times ۱۰^{-۵}$	$۴/۸۱ \times ۱۰^{-۵}$	$۵/۷۷ \times ۱۰^{-۵}$	$۱/۸۹ \times ۱۰^{-۵}$
گشتاورهای LL	$m = ۱$	$۱/۴۸۴ \times ۱۰^{-۳}$	$۷/۷۷۷ \times ۱۰^{-۴}$	$۴/۸۹۸ \times ۱۰^{-۴}$	$۴/۸۹۸ \times ۱۰^{-۴}$	$۹/۰۰۹ \times ۱۰^{-۵}$	$۳/۱۱۱ \times ۱۰^{-۵}$
گشتاورهای LL	$m = ۲$	$۲/۶۸۱ \times ۱۰^{-۳}$	$۱/۵۰۳ \times ۱۰^{-۳}$	$۷/۷۱۴ \times ۱۰^{-۴}$	$۷/۷۱۴ \times ۱۰^{-۴}$	$۱/۱۸۴ \times ۱۰^{-۴}$	$۴/۱۶۷ \times ۱۰^{-۵}$

جدول ۳: مقادیر اریبی سه روش برآورد در دو توزیع نمایی و گامبل

نوع توزیع		توزیع نمایی			توزیع گامبل		
		حجم نمونه					
روش برآورد ↓		$n = ۱۰$	$n = ۳۰$	$n = ۱۰۰$	$n = ۱۰$	$n = ۳۰$	$n = ۱۰۰$
گشتاورهای کلاسیک		$۹/۷۲۹ \times ۱۰^{-۶}$	$۴/۴۸۱ \times ۱۰^{-۶}$	$۱/۱۵۹ \times ۱۰^{-۶}$	$۴/۹۲۸ \times ۱۰^{-۹}$	$۱/۵۳۷ \times ۱۰^{-۶}$	$۹/۰۰۳ \times ۱۰^{-۷}$
گشتاورهای L		$۹/۳۴۹ \times ۱۰^{-۵}$	$۳/۳۵۴ \times ۱۰^{-۳}$	$۶/۵۴۷ \times ۱۰^{-۶}$	$۲/۵۹۸ \times ۱۰^{-۴}$	$۲/۵۹۸ \times ۱۰^{-۶}$	$۳/۴۲۸ \times ۱۰^{-۶}$
گشتاورهای LL	$m = ۱$	$۱/۴۸۴ \times ۱۰^{-۴}$	$۷/۷۷۷ \times ۱۰^{-۵}$	$۱/۸۳۱ \times ۱۰^{-۵}$	$۲/۶۹۴ \times ۱۰^{-۴}$	$۲/۴۳ \times ۱۰^{-۵}$	$۹/۲۳۵ \times ۱۰^{-۶}$
گشتاورهای LL	$m = ۲$	$۲/۶۸۱ \times ۱۰^{-۴}$	$۱/۵۰۳ \times ۱۰^{-۴}$	$۳/۶۱۷ \times ۱۰^{-۵}$	$۶/۶۸۲ \times ۱۰^{-۴}$	$۴/۱۹۷ \times ۱۰^{-۵}$	$۱/۶۵۷ \times ۱۰^{-۵}$

مراجع

[۱] شفیعی، م، ع و جمال زاده، م، ا، ۱۳۸۲ مقدمه‌ای بر مدلسازی توزیعی با استفاده از توابع چندک، دانشگاه صنعتی اصفهان.

[2] Bayazit, M. and ÖnÖz, B. 2002, LL-Moments for estimating low flood quantiles

[3] Berger, J. 1980, *Statistical Decision Theory,* Springer Verlag.

[4] Chen, E.J and Kelton, W.D. Quantile and histogram estimation, .

[5] Hosking, JRM., 1990, L-Moments: Analysis and estimation of distributions using linear combinations of order statistics.

[6] Van Gelder, P.H.A and Vrijling, J.K. A comparative study of different parameter estimation methods for Statistical disribution functions in civil engineering applications.