

مقایسه روش های برآورد چندک ها

مهران شمشیربندی^۱ ابوالقاسم بزرگنیا^۲

چکیده:

در برآورد پارامترهای یک مدل، همیشه انتخاب بهترین برآورد 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 توزیع های نمایی و گامبل که با استفاده از چندک ها معرفی شده اند، مورد مقایسه قرار گرفته اند 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 روش ها بهتر است. واژه های کلیدی: گشتاورهای کلاسیک، گشتاورهای خطی^۳، گشتاورهای خطی خطی^۴.

۱ مقدمه

مربعات خطای برآورد پارامترهای برآورد شده این سه روش برآورد نشان داده ایم که روش برآورد گشتاورهای کلاسیک در سه اندازه نمونه یاد شده در سه خاصیت بهترین رفتار را نشان می دهد.

۲ مدل سازی با استفاده از تابع چندک توزیع

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

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

از جمله کارهای اساسی در آمار تطابق یک مدل با مجموعه ای از داده ها است. بدین منظور پارامترهای مدل باید به گونه ای برآورد شوند که برازشی مناسب بین داده ها و مدل برازش داده شده ارائه شود. برای تشخیص رفتار یک سری از داده ها و در واقع مدل بندی آن ها می توان از توابع چندک توزیع ها استفاده کرد. [۱]

روش های مختلفی برای برآورد پارامترهای مدل استفاده می شود. 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 شده اند برآورد خواهند شد و سپس با استفاده از سه خاصیت اریبی و کمینگی واریانس و مجموع

^۱دانشگاه آزاد اسلامی واحد اراک

^۲موسسه عالی غیر انتفاعی خیام

^۳L-Moments

^۴LL-Moments

بنابراین داریم:

$$A + B.g(F(x)) = x_p \quad (۱)$$

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 در حالت استاندارد حالتی است که در آن پارامتر مکانی صفر و پارامتر مقیاس یک فرض می‌شود. بنابراین رابطه (۱) را می‌توان به صورت زیر بازنویسی کرد:

$$Q(p) = A + BSp \quad (۲)$$

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:

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

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

و در نهایت:

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

بنابراین تابع چندک توزیع گامبل به صورت زیر خواهد بود:

$$Q(p) = F^{-1}(p) = \xi - \theta \ln(-\ln p)$$

Karl Pearson^۵
Hosking^۶

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 برآورد

روش‌های برآوردی که در این مقاله مورد مقایسه قرار می‌گیرند، 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 پارامترها 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:

$$G_i(\theta) = g_i \quad i = 1, 2, \dots, k$$

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

از آن جا که در بسیاری از توزیع ها، فرم بسته ای برای انتگرال روابط (۴) و (۵) نمی توان یافت و یا در برخی موارد محاسبه انتگرال بسیار مشکل است، لذا وانگ^۷ گشتاورهای خطی را با استفاده از نمونه های مرتب شده به اندازه n توسط روابط زیر برآورد کرد. [۲]:

$$\lambda_1 = \sum_{i=1}^n x_{p_i} / \binom{n}{1} \quad (6)$$

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

ج) روش گشتاورهای خطی خطی:

اونوز^۸ و بایازیت^۹ 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 داده های کوچکتر داده می شود. در نمونه های ترتیبی به اندازه n گشتاورهای خطی خطی برای نمونه ای به اندازه ریاضی گرفتن از r تا، از کوچکترین عناصر نمونه به صورت زیر محاسبه می شود:

$$\lambda_{Lr}^m = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k}$$

$$E[X_{r-k|r+m}] \quad r = 1, 2, 3, \dots$$

و دو گشتاور اول به صورت زیر معرفی می شوند:

آماره های ترتیبی محاسبه کرد.

استفاده از گشتاورهای خطی بسیار آسان است، زیرا می توان آنها را به طور مستقیم به عنوان اندازه هایی 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:

$$\lambda_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k E[X_{r-k|r}] \quad r = 1, 2, \dots \quad (3)$$

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

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

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

λ_1 معادل پارامتر مکانی A و λ_2 به عنوان مقیاسی برای پراکندگی در نظر گرفته شده و معادل پارامتر مقیاسی B است. [۶]

رابطه بین گشتاورهای خطی و تابع توزیع و $F(x)$ به صورت زیر نشان داده شده است [۲]:

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

پس λ_1 و λ_2 عبارتند از:

$$\lambda_1 = \int_0^1 Q(p) dp \quad (4)$$

$$\lambda_2 = \int_0^1 Q(p) (2p-1) dp \quad (5)$$

Wang^۷
Önöz^۸
Bayazit^۹

۴ بهترین برآوردها

روش‌های بسیاری برای تمایز قائل شدن و ترجیح دادن روش‌های برآورد، بریکدیگر 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:

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

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

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

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

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

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

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:

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

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

است که دازای 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

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

$$\lambda_{L2}^m = E[X_{1|1+m}]$$

توزیع هستند:

$$\lambda_{L1}^m = (1+m) \int_0^1 Q(p)(1-p)^m dp \quad (8)$$

و

$$\begin{aligned} \lambda_{L2}^m &= \frac{1}{r} (1+m)(2+m) \int_0^1 Q(p)p(1-p)^m dp \\ &- \frac{1}{r} (2+m) \int_0^1 Q(p)(1-p)^{1+m} dp \end{aligned}$$

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:

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

و

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

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

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

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

$$Q(p) = A - B \ln(1 - p) \quad (12)$$

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

$$s(p) = \frac{\partial Q(p)}{\partial p} = \frac{B}{1 - p}$$

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

$$Q(p) = A - B \ln(-\ln p) \quad (13)$$

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

۵ مقایسه روش‌های برآورد چنک‌ها برای دو توزیع نمایی و گامبل

۱.۵ روش گشتاورهای کلاسیک:

الف) توزیع نمایی: توزیع نمایی دارای تابع چنک توزیعی به صورت (۱۲) است. اگر چنک‌های نمونه‌ای $q = 1 - p, p$ به ترتیب برابر با x_q, x_p باشند، آنگاه با مساوی قرار دادن این مقادیر با مقادیر متناظر برای جامعه‌ای از توزیع نمایی داریم:

$$\begin{cases} Q(p) = A - B \ln(1 - p) = x_p \\ Q(1 - p) = A - B \ln(1 - 1 + p) = x_q \end{cases}$$

از حل این دستگاه داریم:

$$\hat{A} = x_q + \left[\frac{x_p - x_q}{\ln p - \ln(1 - p)} \right] \ln p$$

$$\hat{B} = \frac{x_p - x_q}{\ln p - \ln(1 - p)} \quad (15)$$

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

ب) توزیع گامبل: توزیع گامبل دارای تابع چنک توزیعی به صورت رابطه (۱۳) است. اگر چنک‌های نمونه‌ای $q = 1 - p, p$ را به ترتیب با x_q, x_p نشان دهیم، از حل دستگاه زیر:

$$\begin{cases} Q(p) = A - B \ln(-\ln p) = x_p \\ Q(1 - p) = A - B \ln(-\ln(1 - p)) = x_q \end{cases}$$

خواهیم داشت:

$$\hat{A} = x_p + \frac{x_p - x_q}{\ln \left[\frac{\ln(1 - p)}{\ln p} \right]} \cdot \ln[-\ln p] \quad (29)$$

$$\hat{B} = \frac{x_p - x_q}{\ln \left[\frac{\ln(1 - p)}{\ln p} \right]}$$

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

۲.۵ روش گشتاورهای خطی:

برای برآورد پارامترهای تابع چنک هر دو توزیع چون با استفاده از روابط (۶) و (۷) به فرم بسته‌ای از انتگرال نمی‌توان دست یافت. بنابراین با استفاده از روابط (۶) و (۷) مقادیر این دو پارامتر را برآورد و با استفاده از آن دو، مقادیر تابع چنک برآورد می‌شود [۲].

۳.۵ روش گشتاورهای خطی خطی:

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 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

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

با توجه به معیارهایی که در بخش (۳) بیان شد، چندک‌های توزیع هر دو توزیع مورد نظر در سه اندازه

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

نوع توزیع		توزیع نمایی				توزیع گامبل	
		حجم نمونه					
↓ روش برآورد		$n = 100$	$n = 30$	$n = 10$	$n = 100$	$n = 30$	$n = 10$
گشتاورهای کلاسیک		$1/65 \times 10^{-7}$	$7/94 \times 10^{-5}$	$2/28 \times 10^{-1}$	$8/41 \times 10^{-5}$	$6/14 \times 10^{-7}$	$3/867 \times 10^{-11}$
گشتاورهای L		۴/۵۶۴۴	۹/۲۹۳	۵۱/۹۷	۲/۴	۱۹/۵۳۶	۲۹/۸۹۵
گشتاورهای LL	$m = 1$	۱/۵۹۴	۶/۳۸۷	۳۴/۳۹۱	۱/۹۳	۱۶/۷۱	۵۶/۶۵
گشتاورهای LL	$m = 2$	۲/۱۲۵	۵/۸۱	۵۰/۹۷	۲/۸۴۶	۱۵/۴۵	۸۸/۹۵

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

نوع توزیع	توزیع نمایی				توزیع گامبل			
	حجم نمونه				حجم نمونه			
↓ روش برآورد	$n = 10$	$n = 30$	$n = 100$	$n = 10$	$n = 30$	$n = 100$	$n = 10$	$n = 30$
گشتاورهای کلاسیک	$9/7295 \times 10^{-5}$	$4/812 \times 10^{-5}$	$2/266 \times 10^{-1}$	$2/266 \times 10^{-5}$	$2/266 \times 10^{-5}$	$2/266 \times 10^{-5}$	$9/7295 \times 10^{-1}$	$1/89 \times 10^{-5}$
گشتاورهای L	$9/249 \times 10^{-4}$	$3/251 \times 10^{-4}$	$4/81 \times 10^{-5}$	$4/17 \times 10^{-5}$	$4/89 \times 10^{-5}$	$4/89 \times 10^{-5}$	$3/249 \times 10^{-5}$	$4/167 \times 10^{-5}$
گشتاورهای LL	$m = 1$	$1/484 \times 10^{-4}$	$7/777 \times 10^{-4}$	$4/89 \times 10^{-4}$	$4/89 \times 10^{-4}$	$4/89 \times 10^{-4}$	$1/184 \times 10^{-4}$	$4/167 \times 10^{-5}$
گشتاورهای LL	$m = 2$	$2/681 \times 10^{-4}$	$1/503 \times 10^{-4}$	$7/714 \times 10^{-4}$	$7/714 \times 10^{-4}$	$7/714 \times 10^{-4}$	$1/184 \times 10^{-4}$	$4/167 \times 10^{-5}$

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

نوع توزیع	توزیع نمایی				توزیع گامبل			
	حجم نمونه				حجم نمونه			
↓ روش برآورد	$n = 10$	$n = 30$	$n = 100$	$n = 10$	$n = 30$	$n = 100$	$n = 10$	$n = 30$
گشتاورهای کلاسیک	$9/729 \times 10^{-1}$	$4/481 \times 10^{-1}$	$1/159 \times 10^{-1}$	$4/428 \times 10^{-1}$	$1/537 \times 10^{-1}$	$9/003 \times 10^{-7}$	$9/003 \times 10^{-7}$	$9/003 \times 10^{-7}$
گشتاورهای L	$9/249 \times 10^{-5}$	$3/254 \times 10^{-3}$	$1/547 \times 10^{-1}$	$2/598 \times 10^{-4}$	$2/598 \times 10^{-1}$	$3/428 \times 10^{-1}$	$3/428 \times 10^{-1}$	$3/428 \times 10^{-1}$
گشتاورهای LL	$m = 1$	$1/484 \times 10^{-2}$	$7/777 \times 10^{-5}$	$1/831 \times 10^{-5}$	$2/794 \times 10^{-4}$	$2/43 \times 10^{-5}$	$9/249 \times 10^{-1}$	$9/249 \times 10^{-1}$
گشتاورهای LL	$m = 2$	$2/681 \times 10^{-4}$	$1/503 \times 10^{-4}$	$3/617 \times 10^{-5}$	$6/682 \times 10^{-4}$	$4/197 \times 10^{-5}$	$1/656 \times 10^{-5}$	$1/656 \times 10^{-5}$

مراجع

- [۱] شفيعی، 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 distribution functions in civil engineering applications.