

توزیع رایلی-هندسی دومتغیره و ویژگی‌های آن

وحید نکوخو^۱، اشکان خلیفه^۲ و عیسی محمودی^۳

گروه آمار، دانشکده ریاضی و کامپیوتر خوانسار

گروه آمار، دانشکده علوم ریاضی، دانشگاه یزد

تاریخ دریافت: ۱۳۹۶/۸/۷ تاریخ آخرین بازنگری: ۱۳۹۷/۵/۲۷

چکیده: در این مقاله توزیع سه پارامتری رایلی-هندسی دومتغیره با در نظر گرفتن دنباله‌هایی از متغیرهای تصادفی رایلی، که تعداد آن‌ها خود یک متغیر تصادفی هندسی است، به دست آورده می‌شود. برخی از ویژگی‌های مهم توزیع دومتغیره مورد بحث قرار می‌گیرند. گر چه برآوردهای پارامترهای مدل تحت بررسی با حل معادله‌های ساده و معمولی به دست نمی‌آیند، اما یک الگوریتم مناسب برای دستیابی به برآوردها ارائه می‌گردد. در مطالعه‌ای شبیه‌سازی می‌توان صحت و دقت الگوریتم در زمینه تحلیل داده‌های واقعی را مورد تأیید قرار داد. همچنین توانایی مدل جدید در زمینه تحلیل داده‌های واقعی مورد ارزیابی قرار خواهد گرفت.

واژه‌های کلیدی: برآوردیابی، توزیع رایلی، توزیع هندسی.

۱ مقدمه

مارشال و الکین (۱۹۹۷) روشی برای اضافه نمودن پارامتری به برخی خانواده توزیع‌ها، به منظور افزایش انعطاف‌پذیری آن‌ها، ارائه نمودند. این پژوهشگران با بررسی خانواده‌های نمایی و وایبل جزییات روش خود را تشریح کردند. پژوهشگران دیگری نیز با استفاده از خانواده مارشال و الکین تعمیم‌های جدیدی از برخی مدل‌های موجود ارائه نمودند. برای مثال می‌توان به پژوهش‌های لوزادا و همکاران (۲۰۱۴) و ریستیک و کوندو (۲۰۱۵) و منابع موجود در آن‌ها اشاره نمود. مارشال و الکین (۱۹۹۷) به تعمیم خانواده توزیع‌های

آدرس الکترونیک مسئول مقاله: وحید نکوخو، v.nekoukhou@gmail.com

کد موضوع‌بندی ریاضی (۲۰۱۰): 62H12, 62F03, 62F10

جدید به حالت دومتغیره نیز اهتمام ورزیدند. لیکن آن‌ها به بررسی جزییات حالت دومتغیره، مثال‌هایی در مورد آن و استنباط‌های متناظر نپرداختند. بنابراین می‌توان با مطرح و بررسی نمودن اعضای دیگری از آن خانواده جزییات بیشتری در حالت دومتغیره را تحقیق نمود.

در این مقاله به معرفی توزیع رایلی-هندسی دومتغیره که حاوی سه پارامتر است پرداخته می‌شود. این توزیع دومتغیره با در نظر گرفتن دنباله‌هایی از متغیرهای تصادفی رایلی، که تعداد آن‌ها خود یک متغیر تصادفی هندسی است، به دست می‌آید. تابع توزیع تجمعی و تابع چگالی احتمال متغیر تصادفی رایلی با پارامتر $\beta > 0$ ، به ترتیب، عبارتند از

$$F(x; \beta) = 1 - e^{-\beta x^\gamma}, \quad x > 0,$$

$$f(x; \beta) = \gamma \beta x^{\gamma-1} e^{-\beta x^\gamma}, \quad x > 0.$$

در ادامه از نماد $R(\beta)$ برای نمایش توزیع رایلی استفاده می‌شود.

با وجود سه پارامتر در مدل دومتغیره می‌توان انعطاف‌پذیری قابل قبولی برای تحلیل داده‌های دومتغیره از آن انتظار داشت. علاوه بر آن، به دلیل ساده بودن روند تولید داده از مدل جدید، مطالعه مثال‌های شبیه‌سازی شده به سهولت انجام می‌پذیرد.

در بخش ۲ توزیع رایلی-هندسی دومتغیره معرفی می‌شود، سپس توزیع‌های حاشیه‌ای آن بررسی خواهند شد. در بخش ۳ به برآوردیابی پارامترهای مدل پرداخته می‌شود. گر چه برآوردهای پارامترهای مدل تحت بررسی با حل معادله‌های ساده و معمولی به دست نمی‌آیند، اما با استفاده از الگوریتم EM به راحتی می‌توان به برآورد پارامترها دست یافت. همچنین با مطالعه مثال‌های شبیه‌سازی شده در این بخش می‌توان بر عملکرد الگوریتم رایبه شده مهر تأیید زد و به منظور تحلیل داده‌های واقعی به آن اعتماد نمود. در بخش ۴ توانایی مدل جدید در تحلیل داده‌های واقعی مورد ارزیابی قرار خواهد گرفت. در نهایت، در بخش ۵ بحث و نتیجه‌گیری ارائه خواهد شد.

۲ توزیع رایلی-هندسی دومتغیره

توزیع‌های مرکب^۱ بخش مهمی از متون نظریه توزیع‌های آماری را به خود اختصاص داده است. پژوهشگران زیادی با استفاده از شیوه ترکیبی مدل‌های آماری جدیدی ارائه نمودند و یا انعطاف‌پذیری برخی مدل‌های

^۱Compound distributions

موجود را بهبود بخشیدند. به عنوان مثال، آدامیدیس و لوکاس (۱۹۹۸) با ترکیب کردن توزیع‌های نمایی و هندسی، مدل نمایی-هندسی را معرفی کردند. آن‌ها با در نظر گرفتن مینیمم دنباله‌ای از متغیرهای تصادفی نمایی، که تعداد متغیرهای نمایی موجود در دنباله خود یک متغیر تصادفی هندسی است، مدل نمایی-هندسی را معرفی کردند که به دلیل توانایی تحلیل داده‌هایی با نرخ مخاطره نزولی انعطاف‌پذیری بیشتری نسبت به مدل نمایی دارد.

برتو-سوزا و همکاران (۲۰۱۱) با در نظر گرفتن مینیمم دنباله‌ای از متغیرهای تصادفی وایبل و ترکیب آن با توزیع هندسی مدل وایبل-هندسی را معرفی نمودند. همچنین وانگ و الباتال (۲۰۱۴) توزیع وایبل-هندسی بهبودیافته را مطالعه کردند و نکوخو و بیدرام (۲۰۱۷) تعمیمی از مدل وایبل-هندسی به دست آوردند. محمودی و سپهدار (۲۰۱۳) با در نظر گرفتن ماکسیمم دنباله‌ای از متغیرهای تصادفی وایبل نمایی شده^۲ و ترکیب آن با توزیع پواسن، مدل ترکیبی جدیدی معرفی کردند. بیدرام و همکاران (۲۰۱۳) با در نظر گرفتن ماکسیمم دنباله‌ای از متغیرهای تصادفی نمایی تعمیم‌یافته و ترکیب آن با توزیع هندسی، توسعه جدیدی از توزیع نمایی تعمیم‌یافته معرفی نمودند. کوندو (۲۰۱۵) مدل بیدرام و همکاران (۲۰۱۳) را به حالت دو متغیره تعمیم داد و ویژگی‌های آن را بررسی کرد.

در این بخش با در نظر گرفتن دنباله‌ای از متغیرهای تصادفی رایلی، که تعداد متغیرهای موجود در آن تصادفی و از توزیع هندسی پیروی می‌کند، مدل رایلی-هندسی دو متغیره معرفی می‌شود. از آنجایی که مینیمم متغیرهای تصادفی رایلی مجدداً از توزیع رایلی پیروی می‌کند، در ادامه از روش ترکیبی بر اساس تابع مینیمم استفاده خواهد شد.

۱.۲ معرفی مدل

دو دنباله از متغیرهای تصادفی X_1, X_2, \dots و Y_1, Y_2, \dots را در نظر بگیرید. فرض کنید X_i ها متغیرهای تصادفی مستقل و دارای توزیع مشترک $R(\beta_1)$ و Y_i ها نیز متغیرهای تصادفی مستقل و دارای توزیع مشترک $R(\beta_2)$ باشند. X_i ها و Y_j ها نیز مستقل در نظر گرفته می‌شوند. همچنین فرض کنید N ، که از X_i ها و Y_j ها مستقل است، یک متغیر تصادفی هندسی با تابع جرم احتمال

$$P(N = n) = p_n = p(1 - p)^{n-1}, \quad n \in \mathbb{N} = \{1, 2, \dots\}, \quad 0 \leq p \leq 1, \quad (1)$$

²Exponentiated Weibull

باشد، که با نماد $Ge(p)$ نمایش داده می‌شود. متغیر تصادفی دوبعدی (X, Y) به صورت

$$Y = \min\{Y_1, \dots, Y_N\} \quad \text{و} \quad X = \min\{X_1, \dots, X_N\}$$

را در نظر بگیرید، که منجر به تعریف توزیع رایلی-هندسی دو متغیره می‌شود. تابع بقای توأم بردار تصادفی (X, Y) به صورت

$$\begin{aligned} S_{X,Y}(x, y) = P(X \geq x, Y \geq y) &= \sum_{n=1}^{\infty} P(X \geq x, Y \geq y | N = n) p_n \\ &= p \sum_{n=1}^{\infty} e^{-n\beta_1 x} e^{-n\beta_2 y} (1-p)^{n-1} \\ &= \frac{pe^{-\beta_1 x} e^{-\beta_2 y}}{1 - (1-p)e^{-\beta_1 x} e^{-\beta_2 y}}, \end{aligned} \quad (2)$$

است، که در آن $\beta_1 > 0, \beta_2 > 0$ و $0 \leq p \leq 1$ پارامترهای مدل هستند.

تعریف ۱: بردار تصادفی (X, Y) دارای توزیع رایلی-هندسی دو متغیره با پارامترهای $\beta_1 > 0, \beta_2 > 0$ و $0 \leq p \leq 1$ است هرگاه تابع بقای آن به صورت (۲) باشد و با نماد $BRG(\beta_1, \beta_2, p)$ نشان داده می‌شود.

شایان ذکر است اگر $p = 1$ ، آنگاه

$$S_{X,Y}(x, y) = S_R(x; \beta_1) S_R(y; \beta_2),$$

که در آن $S_R(\cdot; \beta_i), i = 1, 2$ ، تابع بقای توزیع $R(\beta_i)$ است. به عبارت دیگر به ازای $p = 1$ متغیره‌های تصادفی X و Y مستقل هستند. بنابراین پارامتر p می‌تواند به عنوان پارامتر همبستگی ایفای نقش نماید. توزیع رایلی-هندسی دو متغیره می‌تواند در تحلیل سیستم‌ها مورد استفاده قرار گیرد. به عنوان مثال، دو سیستم سری که هر یک دارای N مؤلفه مستقل و هم‌توزیع هستند را در نظر بگیرید، به طوری که تعداد مؤلفه‌ها، N ، یک متغیر تصادفی باشد.

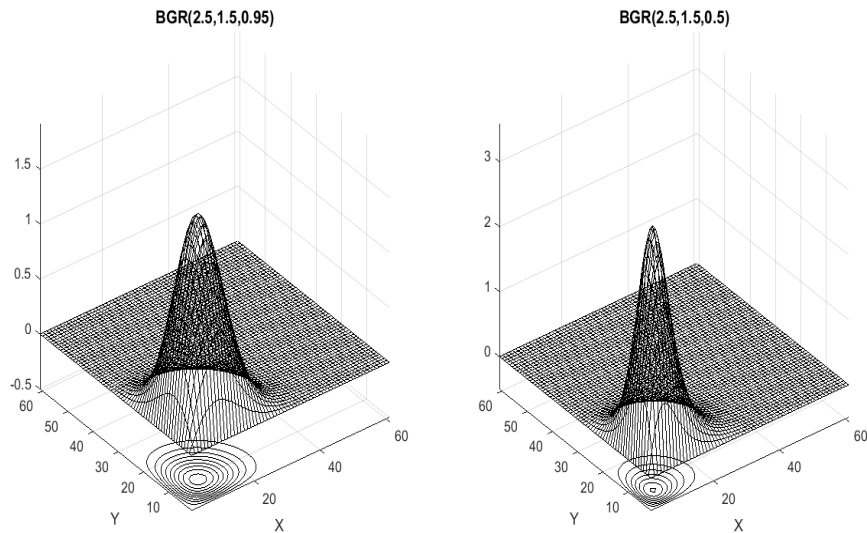
فرض کنید $X_1, X_2, \dots, Y_1, Y_2, \dots$ به ترتیب نمایانگر طول عمر مؤلفه‌های دو سیستم مفروض باشند، آنگاه طول عمر سیستم‌ها عبارت است از (X, Y) که در آن $X = \min\{X_1, \dots, X_N\}$ و

$Y = \min\{Y_1, \dots, Y_N\}$. تابع چگالی احتمال توأم (X, Y) به صورت

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} S_{X,Y}(x,y)$$

$$= p \frac{f_R(x; \beta_1) f_R(y; \beta_2) [\lambda + (\lambda - p) e^{-\beta_1 x} e^{-\beta_2 y}]}{[\lambda - (\lambda - p) e^{-\beta_1 x} e^{-\beta_2 y}]^2},$$

است، که در آن تابع چگالی احتمال $R(\beta_i)$ است. شکل ۱ تابع چگالی احتمال توأم (؟؟) را به ازای برخی مقادیر پارامترهای مدل نمایش می‌دهد.



شکل ۱: تابع چگالی احتمال توأم رایلی-هندسی دومتغیره به ازای برخی مقادیر پارامترهای آن.

۲.۲ توزیع‌های حاشیه‌ای

تابع چگالی احتمال توأم متغیر تصادفی دو بعدی (X, N) به صورت

$$f_{X,N}(x,n) = f_{X|N}(x|n) p_n = 2n\beta_1 x e^{-n\beta_1 x} p(\lambda - p)^{n-1}, \quad x > 0, n \in \mathbb{N}, \quad (3)$$

است. بنابراین تابع بقای توأم آن به ازای $x \geq 0$ و $n \in \mathbb{N}$ ، عبارت است از

$$\begin{aligned} S_{X,N}(x, n) = P(X \geq x, N \geq n) &= \sum_{j=n}^{\infty} P(X \geq x | N = j) p_j \\ &= \frac{p}{1-p} \sum_{j=n}^{\infty} e^{-j\beta_1 x^\gamma} (1-p)^j \\ &= \frac{p(1-p)^{n-1} e^{-n\beta_1 x^\gamma}}{1 - (1-p)e^{-\beta_1 x^\gamma}}. \end{aligned} \quad (۴)$$

با استفاده از (۴) تابع بقای کناری متغیر تصادفی X به صورت

$$S_X(x) = S_{X,N}(x, 1) = \frac{pe^{-\beta_1 x^\gamma}}{1 - (1-p)e^{-\beta_1 x^\gamma}}, \quad x \geq 0, \quad (۵)$$

به دست می‌آید. به ازای یک مقدار مشخص x و $p = 1$ تابع بقای $S_X(x)$ در (۵) همان تابع بقای توزیع رایلی با پارامتر β_1 خواهد بود. به عبارت دیگر هنگامی که $p \rightarrow 1$ ، شکل تابع چگالی احتمال متغیر تصادفی کناری X مانند تابع چگالی رایلی است.

تابع بقای حاشیه‌ای متغیر X را می‌توان از تابع بقای توأم $S_{X,Y}(x, y)$ در رابطه (۲) به صورت

$$S_X(x) = S_{X,Y}(x, 0),$$

به دست آورد. تابع چگالی احتمال کناری متغیر تصادفی X به صورت

$$f_X(x) = -\frac{d}{dx} S_X(x) = w(x) f_R(x; \beta_1), \quad x \geq 0, \quad (۶)$$

است که در آن

$$w(x) = \frac{p}{[1 - (1-p)e^{-\beta_1 x^\gamma}]^2}. \quad (۷)$$

بنابراین تابع چگالی کناری X را می‌توان نوعی توزیع رایلی وزنی با وزن $w(x)$ در (۷) در نظر گرفت.

تابع وزن $w(x)$ بر حسب x نزولی است و با تغییرات x از صفر به ∞ ، از $\frac{1}{p}$ به p کاهش می‌یابد. تابع چگالی X می‌تواند شکل‌های مختلفی به صورت نزولی و تک‌مدی داشته باشد.

تعریف ۲: متغیر تصادفی X دارای توزیع رایلی-هندسی تک‌متغیره است هرگاه تابع بقای X به صورت (۵) یا به طور معادل، تابع چگالی آن به صورت (۶) باشد و با نماد $URG(\beta_1, p)$ نمایش داده می‌شود.

توابع بقا و چگالی احتمال توزیع $URG(\beta_1, p)$ را می‌توان به صورت آمیخته‌های نامتناهی از توزیع‌های رایلی نمایش داد. در واقع با استفاده از بسط

$$(1 - z)^{-1} = \sum_{k=0}^{\infty} z^k, \quad |z| < 1,$$

تابع بقای X به صورت

$$S_X(x) = p \sum_{k=0}^{\infty} (1 - p)^k e^{-(k+1)\beta_1 x^r}, \quad (8)$$

ارایه می‌شود. بنابراین تابع چگالی X به صورت

$$f_X(x) = p \sum_{k=0}^{\infty} (1 - p)^k f_R(x; (k + 1)\beta_1), \quad (9)$$

نیز قابل بیان خواهد بود. گشتاورها و تابع مولد گشتاور توزیع $URG(\beta_1, p)$ را می‌توان با استفاده از گشتاورها و توابع مولد گشتاور توزیع‌های رایلی موجود در (۹) محاسبه نمود. به عنوان مثال، گشتاور m -ام متغیر تصادفی X به صورت

$$\begin{aligned} E(X^m) &= p \sum_{k=0}^{\infty} (1 - p)^k E(Y^m), \\ &= p \sum_{k=0}^{\infty} (1 - p)^k \{(k + 1)\beta_1\}^{-\frac{m}{r}} \Gamma\left(1 + \frac{m}{r}\right), \end{aligned} \quad (10)$$

ارایه می‌شود، که در آن $Y \sim R((k + 1)p)$.

با استفاده از (۵) می‌توان چندک γ -ام توزیع $URG(\beta_1, p)$ را، پس از محاسبه، به صورت

$$\xi_\gamma = \left\{ -\frac{1}{\beta_1} \ln \frac{1-\gamma}{p + (1-\gamma)(1-p)} \right\}^{\frac{1}{2}},$$

ارایه نمود. بنابراین میانه $URG(\beta_1, p)$ با جایگزین نمودن $\frac{1}{2}$ در این رابطه قابل دسترسی است. توزیع N به شرط $X = x$ با استفاده از (۳) به صورت

$$f_{N|X}(n|x) = \frac{2\beta_1 x [1 - (1-p)e^{-\beta_1 x^2}]^2}{f_R(x; \beta_1)} n(1-p)^{n-1} e^{-n\beta_1 x^2}, \quad (11)$$

به دست می‌آید، که امید ریاضی شرطی آن عبارتست از

$$E(N|X = x) = \frac{1 + (1-p)e^{-\beta_1 x^2}}{1 - (1-p)e^{-\beta_1 x^2}}. \quad (12)$$

قضیه ۱: فرض کنید دنباله $\{U_i : i \geq 1\}$ ، دنباله‌ای از متغیرهای تصادفی مستقل با توزیع مشترک $URG(\beta_1, p)$ و متغیر تصادفی N از توزیع $Ge(q)$ ، $0 < q < 1$ ، و مستقل از U_i ها باشد. اگر $U \sim URG(\beta_1, pq)$ ، آنگاه $U = \min\{U_1, \dots, U_N\}$

برهان:

$$\begin{aligned} P(U \geq u) &= P(U_1 \geq u, \dots, U_N \geq u) \\ &= \sum_{n=1}^{\infty} P(U_1 \geq u, \dots, U_N \geq u | N = n) p_n \\ &= \sum_{n=1}^{\infty} \frac{p^n e^{-n\beta_1 u^2}}{[1 - (1-p)e^{-\beta_1 u^2}]^n} q(1-q)^{n-1} \\ &= \frac{pqe^{-\beta_1 u^2}}{1 - (1-pq)e^{-\beta_1 u^2}}. \end{aligned}$$

۳.۲ ویژگی‌های مدل دومتغیره

با بررسی ویژگی‌های توزیع‌های حاشیه‌ای اکنون به راحتی می‌توان ویژگی‌های توزیع رایلی-هندسی دومتغیره را بررسی نمود.

قضیه ۲: اگر $(X, Y) \sim \text{BRG}(\beta_1, \beta_2, p)$ ، آنگاه ویژگی‌های زیر برقرارند:

(الف) $Y \sim \text{URG}(\beta_2, p)$ و $X \sim \text{URG}(\beta_1, p)$

(ب) $X \geq x | Y \geq y \sim \text{URG}(\beta_1, p^*)$ که به طوری که $p^* = 1 - (1 - p)e^{-\beta_2 y}$

(ج) $\min\{X, Y\} \sim \text{URG}(\beta_1 + \beta_2, p)$

(د) $P(X \geq x | Y = y) = \frac{e^{-\beta_1 x} [1 - (1 - p)e^{-\beta_2 y}]^2}{[1 - (1 - p)e^{-\beta_1 x} e^{-\beta_2 y}]^2}$

برهان: قسمت‌های (الف)-(ج) به راحتی و با اندکی محاسبه اثبات می‌گردند. برای اثبات قسمت (د) ابتدا فرض کنید

$$c(y) = \frac{2\beta_2 y [1 - (1 - p)e^{-\beta_2 y}]^2}{f_R(y; \beta_2)}$$

توزیع شرطی به صورت

$$\begin{aligned} P(X \geq x | Y = y) &= \sum_{n=1}^{\infty} P(X \geq x | Y = y, N = n) P(N = n | Y = y) \\ &= \sum_{n=1}^{\infty} c(y) n e^{-n\beta_1 x} (1 - p)^{n-1} e^{-n\beta_2 y} \\ &= \frac{e^{-\beta_1 x} [1 - (1 - p)e^{-\beta_2 y}]^2}{[1 - (1 - p)e^{-\beta_1 x} e^{-\beta_2 y}]^2}, \end{aligned}$$

محاسبه می‌شود.

در قضیه بعد، که مشابه قضیه ۱ قابل اثبات است، نشان داده می‌شود توزیع رایلی-هندسی دومتغیره نیز مانند حالت تک‌متغیره، تحت ترکیب با توزیع هندسی بسته است.

قضیه ۳: فرض کنید $\{(U_i, V_i) : i \geq 1\}$ دنباله‌ای از متغیرهای تصادفی مستقل با توزیع مشترک $\text{BRG}(\beta_1, \beta_2, p)$ و $N \sim \text{Ge}(q)$ ، $0 < q < 1$ باشد. بعلاوه زوج‌های (U_i, V_i) و N مستقل باشند. اگر $U = \min\{U_1, \dots, U_N\}$ و $V = \min\{V_1, \dots, V_N\}$ آنگاه $(U, V) \sim \text{BRG}(\beta_1, \beta_2, pq)$.

اکنون توزیع N به شرط $\{X = x, Y = y\}$ به دست آورده می‌شود. ابتدا توجه کنید تابع چگالی احتمال توأم (X, Y, N) به صورت

$$f_{X,Y,N}(x, y, n) = 2n^2 \beta_1 \beta_2 p (1 - p)^{n-1} x y e^{-n\beta_1 x} e^{-n\beta_2 y}, \quad (13)$$

است. بنابراین توزیع N به شرط $\{X = x, Y = y\}$ به صورت

$$P(N = n | X = x, Y = y) = n^{\nu} g^{n-1}(x, y) \frac{\{1 - g(x, y)\}^{\nu}}{1 + g(x, y)}, \quad (14)$$

است، که در آن $g(x, y) = (1 - p)e^{-\beta_1 x^{\nu}} e^{-\beta_2 y^{\nu}}$.

چون $M \sim \text{Ge}(q)$ ، بنابراین $E(M^{\nu}) = \frac{(1-q)^{\nu} + \nu(1-q) + 1}{q^{\nu}}$. در نتیجه امید ریاضی توزیع شرطی (۱۴) به صورت

$$\begin{aligned} E(N | X = x, Y = y) &= \frac{\{1 - g(x, y)\}^{\nu}}{1 + g(x, y)} \sum_{n=1}^{\infty} n^{\nu} g^{n-1}(x, y) \\ &= \frac{g^{\nu}(x, y) + \nu g(x, y) + 1}{1 - g^{\nu}(x, y)}, \end{aligned}$$

به دست می‌آید.

۴.۲ پارامتر تنش-مقاومت

به طور ساده پارامتر تنش-مقاومت را می‌توان ارزیابی قابلیت اعتماد یک جزء از سیستم با در نظر گرفتن متغیر تصادفی X به عنوان تنش تجربه شده توسط سیستم و متغیر Y به عنوان مقاومت آن جزء برای غلبه بر تنش توصیف کرد. براساس این تعریف ساده اگر میزان تنش از مقاومت تجاوز کند آن جزء از سیستم دچار خرابی یا شکست خواهد شد، برعکس. بنابراین در چنین حالتی قابلیت اعتماد، احتمال شکست نخوردن آن جزء یعنی $P(X < Y)$ تعریف خواهد شد. در دو دهه اخیر، پارامتر تنش-مقاومت، که معمولاً آن را با R نمایش می‌دهند، در تحقیقات مربوط به رشته‌های مهندسی، روان‌شناسی، پزشکی و به ویژه داروسازی و همچنین در علوم آموزش و پرورش نقش پررنگی داشته است. اگر $(X, Y) \sim \text{BRG}(\beta_1, \beta_2, p)$ ، آنگاه پارامتر تنش-مقاومت توزیع رایلی-هندسی دومتغیره عبارت است از

$$\begin{aligned} R = P(X < Y) &= \sum_{i=1}^n P(X < Y, N = n) \\ &= \sum_{i=1}^n \int_0^{\infty} \int_0^y f_{X,Y,N}(x, y, n) dx dy \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n p(1-p)^{n-1} \frac{\beta_1}{\beta_1 + \beta_2} \\
 &= \frac{\beta_1}{\beta_1 + \beta_2}.
 \end{aligned}$$

۳ برآورد پارامترهای مدل با الگوریتم EM

فرض کنید نمونه تصادفی $(x_1, y_1), \dots, (x_n, y_n)$ در دسترس باشد. لگاریتم تابع درستنمایی بر اساس این مشاهدات عبارت است از

$$\begin{aligned}
 \ell(\beta_1, \beta_2, p) &= \sum_{i=1}^m \ln f_{X,Y}(x_i, y_i) \\
 &= m \ln p + m \ln \beta_1 + \sum_{i=1}^m \ln(\beta_1 x_i) - \beta_1 \sum_{i=1}^m x_i \\
 &+ m \ln \beta_2 + \sum_{i=1}^m \ln(\beta_2 y_i) - \beta_2 \sum_{i=1}^m y_i \\
 &+ \sum_{i=1}^m \ln\{1 + (1-p)e^{-\beta_1 x_i} e^{-\beta_2 y_i}\} \\
 &- \sum_{i=1}^m \ln\{1 - (1-p)e^{-\beta_1 x_i} e^{-\beta_2 y_i}\}.
 \end{aligned}$$

برای دستیابی به برآوردهای ماکسیمم درستنمایی پارامترها لازم است ماکسیمم رابطه (۱۵) بر حسب پارامترها محاسبه شود. برای انجام این کار با یک مسأله بهینه‌سازی سه‌بعدی روبرو هستیم که برای حل آن می‌توان، به عنوان مثال، از الگوریتم نیوتن-رافسون^۳ استفاده نمود. لیکن حدس زدن مقادیر اولیه استفاده از این الگوریتم کار ساده‌ای نیست. برای اجتناب از این چالش از الگوریتم EM برای محاسبه برآوردهای ماکسیمم درستنمایی استفاده می‌شود.

فرض کنید $(x_1, y_1, n_1), \dots, (x_m, y_m, n_m)$ مشاهدات کامل باشند. در اینجا اطلاعات مربوط به n_i ها ممکن است در دسترس نباشند. ابتدا فرض کنید p معلوم باشد. لگاریتم تابع درستنمایی بر اساس

³Newton-Raphson

مشاهدات کامل و بدون در نظر گرفتن مقادیری که به پارامترها بستگی ندارند عبارت است از

$$\ell_c(\beta_1, \beta_2) = m \ln \beta_1 + m \ln \beta_2 - \beta_1 \sum_{i=1}^m n_i x_i^{\downarrow} - \beta_2 \sum_{i=1}^m n_i y_i^{\downarrow}. \quad (15)$$

بر اساس مشاهدات کامل برآوردگرهای ماکسیمم درستنمایی پارامترهای β_1 و β_2 به ترتیب به صورت

$$\hat{\beta}_2 = \frac{m}{\sum_{i=1}^m n_i y_i^{\downarrow}} \quad \text{و} \quad \hat{\beta}_1 = \frac{m}{\sum_{i=1}^m n_i x_i^{\downarrow}}$$

هستند. برآورد ماکسیمم درستنمایی پارامتر p نیز با ماکسیمم کردن لگاریتم تابع درستنمایی نیمرخ^۴ بر حسب p به دست می‌آید.

به منظور استفاده از الگوریتم EM، فرض کنید $\beta_1^{(k)}$ و $\beta_2^{(k)}$ مقادیر پارامترهای β_1 و β_2 در مرحله k -ام الگوریتم باشند. بنابراین مرحله $(k+1)$ -ام، مرحله E الگوریتم، متناظر است با تشکیل شبهه لگاریتم تابع درستنمایی^۵ که بدون در نظر گرفتن مقادیر ثابت به صورت

$$\ell_s(\beta_1, \beta_2) = m \ln \beta_1 + m \ln \beta_2 - \beta_1 \sum_{i=1}^m \tilde{n}_i^{(k)} x_i^{\downarrow} - \beta_2 \sum_{i=1}^m \tilde{n}_i^{(k)} y_i^{\downarrow}, \quad (16)$$

ارایه می‌شود، که در آن

$$\tilde{n}_i^{(k)} = E(N|X = x_i, Y = y_i, \beta_1^{(k)}, \beta_2^{(k)}, p).$$

به منظور دستیابی به $\beta_1^{(k+1)}$ و $\beta_2^{(k+1)}$ ، مرحله M الگوریتم، کفایت ماکسیمم رابطه (۱۶) بر حسب پارامترها محاسبه شود. بنابراین خواهیم داشت

$$\hat{\beta}_2^{(k+1)} = \frac{m}{\sum_{i=1}^m \tilde{n}_i^{(k)} y_i^{\downarrow}} \quad \text{و} \quad \hat{\beta}_1^{(k+1)} = \frac{m}{\sum_{i=1}^m \tilde{n}_i^{(k)} x_i^{\downarrow}}$$

این فرآیند تا همگرایی الگوریتم ادامه می‌یابد. به ازای یک مقدار ثابت p برآوردهای به دست آمده با $\hat{\beta}_1(p)$

^۴profile log-likelihood

^۵pseudo log-likelihood function

و $\hat{\beta}_2(p)$ نمایش داده می‌شوند. در نهایت برای دستیابی به برآورد ماکسیمم درست‌نمایی پارامتر p ، ماکسیمم لگاریتم تابع درست‌نمایی نیم‌رخ $\ell(\hat{\beta}_1(p), \hat{\beta}_2(p), p)$ بر حسب p به دست آورده می‌شود. لازم به ذکر است با ایجاد تعدیل‌های لازم در الگوریتم ارایه شده می‌توان از آن برای برآورد پارامترهای توزیع‌های حاشیه‌ای، یعنی مدل‌های رایلی-هندسی تک‌متغیره، نیز استفاده نمود.

۱.۳ مطالعه شبیه‌سازی

در این بخش به منظور نشان دادن توانایی الگوریتم EM شبیه‌سازی با اندازه‌های مختلف حجم نمونه و مقادیر پارامترها انجام می‌شود. توجه کنید، تولید داده از مدل $BRG(\beta_1, \beta_2, p)$ ساده است. بدین منظور، ابتدا N از مدل هندسی تولید و سپس با مشاهده $N = n$ ، X و Y می‌توانند به ترتیب از توزیع‌های $R(\beta_2)$ و $R(\beta_1)$ تولید شوند. مجموعه پارامترهای در نظر گرفته شده در این شبیه‌سازی عبارتند از:

$$\text{الف) } \beta_1 = \beta_2 = 2 \text{ و } p = 0.25$$

$$\text{ب) } \beta_1 = 1, \beta_2 = 2 \text{ و } p = 0.5$$

$$\text{ج) } \beta_1 = \beta_2 = 2 \text{ و } p = 0.75$$

ابتدا داده‌هایی از مدل رایلی-هندسی دو متغیره به ازای حجم نمونه‌های مختلف تولید می‌شوند. سپس با الگوریتم EM برآوردهای ماکسیمم درست‌نمایی پارامترها محاسبه می‌شوند. در همه حالت‌ها از مقادیر اولیه $\beta_1^{(0)} = \beta_2^{(0)} = 0.5$ و $p^{(0)} = 0.5$ استفاده می‌شود. زمانی الگوریتم به کار خود خاتمه می‌دهد که تفاضل تابع درست‌نمایی در دو تکرار متوالی کمتر از 10^{-5} باشد. این فرآیند ۱۰۰۰ مرتبه تکرار می‌شود. متوسط برآوردها، میانگین توان‌های دوم خطاهای متناظر و میانه تعداد تکرارها تا همگرایی الگوریتم در جدول ۱ ارایه شده‌اند.

همان‌طور که ملاحظه می‌شود با افزایش حجم نمونه از میزان اریبی و میانگین‌های توان دوم خطاها کاسته می‌شود، که به نوعی تصدیق‌کننده سازگاری برآوردهای ماکسیمم درست‌نمایی هستند. همچنین با افزایش مقدار p میانه تعداد تکرارها تا همگرایی الگوریتم کاهش می‌یابد. در واقع هنگامی که $p \rightarrow 1$ ، متغیرهای X و Y به سمت استقلال از یکدیگر متمایل و باعث ساده‌تر شدن مدل می‌شوند. در نتیجه تعداد تکرارها تا همگرایی الگوریتم کاهش می‌یابد.

در صورت تغییر مقادیر اولیه، به جز میانه تعداد تکرارها، نتایج در تمامی حالت‌ها بدون تغییر باقی می‌مانند. بنابراین با توجه به بررسی مثال‌های شبیه‌سازی شده می‌توان گفت الگوریتم EM ارایه شده عملکرد مناسبی دارد و می‌توان برای تحلیل داده‌های واقعی به آن اعتماد کرد.

جدول ۱: مقادیر برآوردها همراه میانگین توان دوم خطاها و میانه تعداد تکرارها.

$(\beta_1, \beta_2, p) = (2, 2, 0.25)$							
میانۀ تعداد تکرارها	$MSE(p)$	$MSE(\beta_2)$	$MSE(\beta_1)$	\hat{p}	$\hat{\beta}_2$	$\hat{\beta}_1$	n
۸۷	۰/۰۲۶	۱/۲۶۲	۱/۱۸۳	۰/۳۳۷	۲/۵۴۴	۲/۵۷۲	۲۵
۹۲	۰/۰۱۱	۰/۵۳۹	۰/۵۲۴	۰/۲۹۵	۲/۲۸۳	۲/۲۵۸	۵۰
۹۷	۰/۰۰۸	۰/۳۴۳	۰/۳۳۹	۰/۲۷۸	۲/۱۷۶	۲/۱۶۹	۷۵
۱۰۰	۰/۰۰۵	۰/۲۴۳	۰/۲۴۹	۰/۲۶۷	۲/۱۱۱	۲/۱۲۵	۱۰۰
$(\beta_1, \beta_2, p) = (1, 2, 0.5)$							
۴۰	۰/۰۴۷	۰/۵۸۹	۰/۱۴۹	۰/۵۸۶	۲/۲۵۵	۱/۱۴۳	۲۵
۴۲	۰/۰۲۹	۰/۲۹۰	۰/۰۶۸	۰/۵۴۲	۲/۱۲۰	۱/۰۶۶	۵۰
۴۳	۰/۰۲۰	۰/۲۰۵	۰/۰۴۸	۰/۵۳۷۶	۲/۱۱۱	۱/۰۵۵	۷۵
۴۴	۰/۰۱۴	۰/۱۴۳	۰/۰۳۳	۰/۵۳۱	۲/۰۸۴	۱/۰۴۰	۱۰۰
$(\beta_1, \beta_2, p) = (2, 2, 0.75)$							
۳۵	۰/۰۴۱	۰/۳۱۴	۰/۳۳۳	۰/۷۷۰	۲/۰۸۲	۲/۰۹۰	۲۵
۳۷	۰/۰۲۸	۰/۱۷۰	۰/۱۷۵	۰/۷۸۰	۲/۰۸۳	۲/۰۷۵	۵۰
۳۸	۰/۰۲۲	۰/۱۱۰	۰/۱۲۰	۰/۷۷۰	۲/۰۴۰	۲/۰۴۷	۷۵
۳۹	۰/۰۱۸	۰/۰۸۴	۰/۰۸۹	۰/۷۶۸	۲/۰۴۰	۲/۰۴۱	۱۰۰

۴ برازش مدل جدید به داده‌های واقعی

به منظور تشریح توانایی مدل BRG در زمینه تحلیل داده‌های واقعی از داده‌های فوتبال در لیگ قهرمانان اروپا مربوط به سال‌های ۲۰۰۶-۲۰۰۴ که شامل ۳۷ مسابقه است استفاده می‌شود. داده‌ها اولین بار توسط مینتانیس (۲۰۰۷) مورد استفاده قرار گرفت. در این مجموعه داده، که در جدول ۲ بدون نام بردن اسامی تیم‌ها ارایه شده است، X_1 نمایانگر زمان به ثمر رسیدن اولین گل بر حسب دقیقه توسط یک ضربه مستقیم مانند ضربه خطا یا پنالتی توسط هریک از دو تیم و X_2 زمان به ثمر رسیدن اولین گل به هر شکل توسط تیمی است که در زمین خانه خود بازی می‌کند.

از الگوریتم EM برای محاسبه برآوردهای ماکسیمم درست‌نمایی پارامترها استفاده می‌شود. حدس زدن مقادیر اولیه با توجه به همگرایی الگوریتم چالش جدی محسوب نمی‌شود. استفاده از مقادیر اولیه مختلف و به دست آوردن برآوردهای یکسان پارامترها دلیل بر این ادعا است.

پرسش مهمی که مطرح می‌شود آن است که آیا مدل رایلی-هندسی دومتغیره قادر است برازش مناسبی به داده‌ها ارایه دهد؟ اگر چه آزمون‌های نیکویی برازش زیادی برای توابع توزیع تک‌متغیره وجود دارند، اما برای یک تابع توزیع دومتغیره آزمون جامعی موجود نیست (کوندو و گوپتا، ۲۰۱۴). با استناد به قضیه ۲ توزیع‌های رایلی-هندسی تک‌متغیره به متغیرهای کناری و همچنین می‌نیم آن‌ها برازش داده می‌شود.

جدول ۲: داده‌های فوتبال مربوط به لیگ قهرمانان اروپا در سال‌های ۲۰۰۴-۲۰۰۶

X_2	X_1	X_2	X_1
۳۴	۳۴	۲۰	۲۶
۳۹	۵۳	۱۸	۶۳
۷	۵۴	۱۹	۱۹
۲۸	۵۱	۸۵	۶۶
۶۴	۷۶	۴۰	۴۰
۱۵	۶۴	۴۹	۴۹
۴۸	۲۶	۸	۸
۱۶	۱۶	۷۱	۶۹
۱۳	۴۴	۳۹	۳۹
۱۴	۲۵	۴۸	۸۲
۱۱	۵۵	۷۲	۷۲
۴۹	۴۹	۶۲	۶۶
۲۴	۲۴	۹	۲۵
۳۰	۴۴	۳	۴۱
۳	۴۲	۷۵	۱۶
۴۷	۲۷	۱۸	۱۸
۲۸	۲۸	۱۴	۲۲
۲	۲	۴۲	۴۲
		۵۲	۳۶

مقدار آماره آزمون کلموگروف-اسمیرنوف و p -مقدار متناظر آن در جدول ۳ ارائه شده است. با توجه به مقادیر آماره آزمون و p -مقدارهای متناظر به نظر می‌رسد مدل‌های رایلی-هندسی تک‌متغیره برازش‌های قابل قبولی ارائه می‌دهند. بنابراین می‌توان انتظار داشت مدل رایلی-هندسی دو متغیره نیز برازش مناسبی به داده‌های توأم ارائه نماید.

جدول ۳: نتایج برازش توزیع‌های حاشیه‌ای

متغیرها	مقدار آماره آزمون K-S	p -مقدار
X_1	۰/۱۳۳۵	۰/۴۸۳۵
X_2	۰/۱۲۳۳	۰/۵۸۴۷
$\min\{X_1, X_2\}$	۰/۱۵۰۲	۰/۳۳۹۷

حال توزیع رایلی-هندسی دو متغیره به داده‌ها برازش داده می‌شود. با استفاده از داده‌های فوتبال، برآوردهای ماکسیمم درست‌نمایی پارامترها در مدل دو متغیره عبارتند از $\hat{\beta}_1 = ۰/۰۰۰۴۳۸$ ، $\hat{\beta}_2 = ۰/۰۰۰۲۹۳$ و $\hat{p} = ۰/۴۳۰۲$. بر اساس برآوردهای به دست آمده مقدار لگاریتم تابع درست‌نمایی برابر -۳۲۶۹۳۸ و ضریب اطلاع بیزی $BIC = ۶۶۴/۷۱$ و ضریب اطلاع آکائیک $AIC = ۶۵۹/۸۷$ است.

برآورد پارامترهای مدل نرمال دو متغیره به داده‌های فوتبال نیز عبارتند از $\mu = (۳۲/۸۶, ۴۰/۸۹)'$ و $\Sigma = \begin{bmatrix} ۵۰۷/۲۵ & ۲۱۰/۲۰ \\ ۲۱۰/۲۰ & ۳۹۴/۵۰ \end{bmatrix}$. این در حالی است که مقدار لگاریتم تابع درست‌نمایی و ضرایب اطلاع بیزی و

آکائیک مدل نرمال دومتغیره برای داده‌های جدول ۲، به ترتیب ۳۲۶/۲۱۳-، ۶۷۰/۴۸ و ۶۶۲/۴۳ هستند. بنابراین با مقایسه نتایج فوق می‌توان ادعا نمود که مدل رایلی-هندسی دومتغیره برازش قابل قبولی به داده‌های فوتبال ارایه می‌نماید.

بحث و نتیجه‌گیری

در این مقاله به بررسی توزیع رایلی-هندسی دومتغیره و ویژگی‌های آن پرداخته شد. از جمله ویژگی‌های مدل آن است که توزیع‌های حاشیه‌ای رایلی-هندسی تک‌متغیره هستند. گرچه برآوردهای ماکسیمم درست‌نمایی با حل معادله‌های معمولی و به صورت ساده‌ای به دست نمی‌آیند، اما الگوریتم EM به خوبی پاسخ‌گوی دستیابی به برآوردهای ماکسیمم درست‌نمایی است. همچنین تحلیل یک مجموعه داده واقعی و مقایسه نتیجه‌های به دست آمده با مدل نرمال دومتغیره گویای این حقیقت است. تعمیم توزیع رایلی-هندسی دومتغیره به حالت چند متغیره می‌تواند منجر به شکل‌گیری پژوهش‌های آتی گردد.

تقدیر و تشکر

نویسندگان مقاله بر خود لازم می‌دانند از زحمات سردبیر محترم مجله و داوران محترم که نظرات ارزشمندشان کمک شایانی به بهتر شدن مقاله نمود تشکر و قدردانی نمایند.

مراجع

- Adamidis, K. and Loukas, S. (1998), A Lifetime Distribution with Decreasing Failure Rate, *Statistics & Probability Letters*, **39**, 35–42.
- Barreto-Souza, W., Lemos-Morais, A. and Cordeiro, G.M. (2011), The Weibull-Geometric Distribution, *Journal of Statistical Computation and Simulation*, **81(5)**, 645–657.
- Bidram, H., Behboodian, J. and Towhidi, M. (2013), A New Generalized Exponential Geometric Distribution, *Communications in Statistics—Theory and Methods*, **42**, 528–542.
- Kundu, D. (2015), Bivariate Geometric (Maximum) Generalized Exponential Distribution, *Journal of Data Sciences*, **13**, 693-712.

- Kundu, D. and Gupta, A. K. (2014), On Bivariate Weibull-Geometric Distribution, *Journal of Multivariate Analysis*, **123**, 19-29.
- Louzada, F., Marchi, V. A. A., and Roman, M. (2014), The Exponentiated Exponential-Geometric Distribution; a Distribution with Decreasing, Increasing and Unimodal Failure Rate, *Statistics*, **48**, 167-181.
- Marshall, A. W. and Olkin, I. (1997), A New Method of Adding a Parameter to a Family of Distributions with Application to the Exponential and Weibull Families, *Biometrika*, **84**, 641-652.
- Mahmoudi, E. and Sepahdar, A. (2013), Exponentiated Weibull–Poisson Distribution: Model, Properties and Applications, *Mathematics and Computers in Simulation*, **92**, 76–97.
- Meintanis, S. G. (2007), Test of Fit for Marshall-Olkin Distributions with Applications, *Journal of Statistical Planning and inference* **137**, 3954-3963.
- Nekoukhou, V. and Bidram, H. (2017), A New Generalization of the Weibull-Geometric Distribution with Bathtub Failure Rate, *Communications in Statistics - Theory and Methods*, **46**, 4296-4310.
- Ristic, M. M. and Kundu, D. (2015), Marshall-Olkin Generalized Exponential Distribution, *METRON*, **73**, 317-333.
- Wang, M. and Elbatal, I. (2014), The Modified Weibull Geometric Distribution, *METRON*, DOI: 10.1007/s40300-014-0052-1.