توزیع رایلی-هندسی دومتغیره و ویژگی‌های آن

وحید نکوخو[1]، اشکان خلیفه[2] و عیسی محمودی[2]

[1]گروه آمار، دانشکده ریاضی و کامپیوتر خوانسار

[2]گروه آمار، دانشکده علوم ریاضی، دانشگاه یزد

چکیده: در این مقاله توزیع سه‌پارامتری 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 ارایه 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: توزیع رایلی، توزیع هندسی، برآوردیابی.

۱ مقدمه

مارشال و الکین (۱۹۹۷) روشی برای اضافه نمودن پارامتری به برخی خانواده توزیع‌ها، به منظور افزایش انعطاف‌پذیری آن‌ها، ارایه نمودند. این پژوهشگران با بررسی خانواده‌های نمایی و وایبل جزییات روش خود

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

°کد موضوع‌بندی ریاضی (۲۰۱۰): 62F10، 62F03، 62H12.

را تشریح کردند. پژوهشگران دیگری نیز با استفاده از خانواده مارشال و الکین تعمیم‌های جدیدی از برخی مدل‌های موجود ارایه 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 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 $\beta > ۰$، به ترتیب، عبارتند از

$$F(x;\beta) = ۱ - e^{-\beta x^۲}, \quad x > ۰,$$
$$f(x;\beta) = ۲\beta x e^{-\beta x^۲}, \quad x > ۰.$$

در ادامه از نماد $\mathrm{R}(\beta)$ 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 ارایه 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[1] بخش مهمی از متون نظریه توزیع‌های آماری را به خود اختصاص داده است. پژوهشگران زیادی با استفاده از شیوه ترکیبی مدل‌های آماری جدیدی ارایه نمودند و یا انعطاف‌پذیری برخی مدل‌های 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 مینیممدنباله‌ای 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 مینیممدنباله‌ای 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 نمایی‌شده[2] و ترکیب آن با توزیع پوآسن، 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 را به حالت دومتغیره 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 دومتغیره معرفی می‌شود. از آن‌جایی که مینیممتغیرهای تصادفی رایلی مجدداً از توزیع رایلی پیروی می‌کند، در ادامه از روش ترکیبی بر اساس تابع مینیمماستفاده خواهد شد.

۱.۲ معرفی مدل

دو دنباله از متغیرهای تصادفی $X_1, X_2, \ldots$ و $Y_1, Y_2, \ldots$ را در نظر بگیرید. فرض کنید X_iها متغیرهای تصادفی مستقل و دارای توزیع مشترک $\mathrm{R}(\beta_1)$ و Y_iها نیز متغیرهای تصادفی مستقل و دارای توزیع مشترک

[1] Compound distributions

[2] Exponentiated Weibull

$\mathrm{R}(\beta_2)$ باشند. X_iها و Y_jها نیز مستقل در نظر گرفته می‌شوند. همچنین فرض کنید N، که از X_iها و Y_jها مستقل است، یک متغیر تصادفی هندسی با تابع جرم احتمال

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

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

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

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

$$\begin{aligned} S_{X,Y}(x,y)=P(X\geq x,Y\geq y) &= \sum_{n=1}^{\infty}P(X\geq x,Y\geq y|N=n)p_n \\ &= p\sum_{n=1}^{\infty}e^{-n\beta_1x^2}e^{-n\beta_2y^2}(1-p)^{n-1} \\ &= \frac{pe^{-\beta_1x^2}e^{-\beta_2y^2}}{1-(1-p)e^{-\beta_1x^2}e^{-\beta_2y^2}}, \end{aligned} \tag{2}$$

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 دومتغیره 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 $\mathrm{BRG}(\beta_1,\beta_2,p)$ نشان داده می‌شود.

شایان ذکر است اگر $p=1$، آنگاه

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

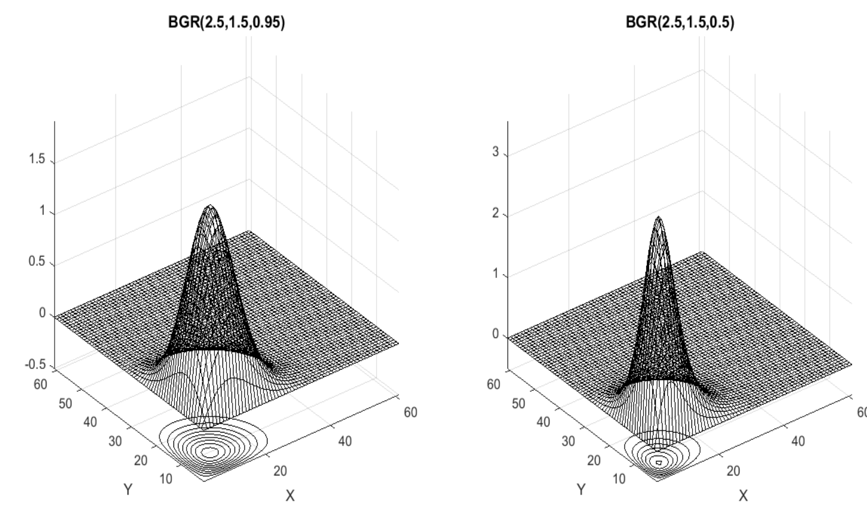
که در آن $S_R(.;\beta_i)$، $i=1,2$، تابع بقای توزیع $\mathrm{R}(\beta_i)$ است. به عبارت دیگر به ازای $p=1$ متغیرهای تصادفی X و Y مستقل هستند. بنابراین پارامتر p می‌تواند به عنوان پارامتر همبستگی ایفای نقش نماید.

توزیع رایلی-هندسی دومتغیره 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، به‌طوری‌که تعداد مؤلفه‌ها، N، یک متغیر تصادفی باشد. اگر $X_1,X_2,...$ و $Y_1,Y_2,...$ به ترتیب نمایانگر طول عمر مؤلفه‌های دو سیستم

مفروض باشند، آنگاه طول عمر سیستم‌ها عبارت است از (X, Y) که در آن $X = \min\{X_1, ..., X_N\}$ و $Y = \min\{Y_1, ..., Y_N\}$. تابع چگالی احتمال توأم (X, Y) به صورت

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\partial^2}{\partial x \partial y} S_{X,Y}(x,y) \\ &= p\frac{f_R(x;\beta_1) f_R(y;\beta_2)[1+(1-p)e^{-\beta_1 x^2} e^{-\beta_2 y^2}]}{[1-(1-p)e^{-\beta_1 x^2} e^{-\beta_2 y^2}]^3}, \end{aligned}$$

است، که در آن $f_R(.;\beta_i)$ تابع چگالی احتمال $\mathrm{R}(\beta_i)$ است. شکل ۱ تابع چگالی احتمال توأم (۳) 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

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

است. بنابراین تابع بقای توأم آن به ازای $x \geq ۰$ و $n \in \mathbb{N}$، عبارت است از

$$
\begin{aligned}
S_{X,N}(x,n) = P(X \geq x, N \geq n) &= \sum_{j=n}^{\infty} P(X \geq x | N = j) p_j \\
&= \frac{p}{۱-p} \sum_{j=n}^{\infty} e^{-j\beta_۱ x^۲} (۱-p)^j \\
&= \frac{p(۱-p)^{n-۱} e^{-n\beta_۱ x^۲}}{۱-(۱-p)e^{-\beta_۱ x^۲}}. \qquad (۴)
\end{aligned}
$$

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

$$S_X(x) = S_{X,N}(x,۱) = \frac{pe^{-\beta_۱ x^۲}}{۱-(۱-p)e^{-\beta_۱ x^۲}}, \quad x \geq ۰, \qquad (۵)$$

به دست می‌آید. به ازای یک مقدار مشخص x و $p = ۱$ تابع بقای $S_X(x)$ در (۵) همان تابع بقای توزیع رایلی با پارامتر $\beta_۱$ خواهد بود. به عبارت دیگر هنگامی‌که $p \to ۱$، شکل تابع چگالی احتمال متغیر تصادفی کناری X مانند تابع چگالی رایلی است.

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

$$S_X(x) = S_{X,Y}(x,۰),$$

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

$$f_X(x) = -\frac{d}{dx} S_X(x) = w(x) f_R(x;\beta_۱), \quad x \geq ۰, \qquad (۶)$$

است، که در آن

$$w(x) = \frac{p}{[۱-(۱-p)e^{-\beta_۱ x^۲}]^۲}. \qquad (۷)$$

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 $\frac{۱}{p}$ 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 هر گاه تابع بقای X به صورت (۵) یا به طور معادل، تابع چگالی آن به صورت (۶) باشد و با نماد $\mathrm{URG}(\beta_۱, p)$ نمایش داده می‌شود.

توابع بقا و چگالی احتمال توزیع $\mathrm{URG}(\beta_1, p)$ 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

$$S_X(x) = p \sum_{k=0}^{\infty} (1 - p)^k e^{-(k+1)\beta_1 x^2}, \tag{8}$$

ارایه می‌شود. بنابراین تابع چگالی X به صورت

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

نیز قابل بیان خواهد بود. گشتاورها و تابع مولد گشتاور توزیع $\mathrm{URG}(\beta_1, p)$ 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

$$\begin{aligned} E(X^m) &= p \sum_{k=0}^{\infty} (1 - p)^k E(Y^m), \\ &= p \sum_{k=0}^{\infty} (1 - p)^k \left\{(k + 1)\beta_1\right\}^{-\frac{m}{2}} \Gamma(1 + \frac{m}{2}), \end{aligned} \tag{10}$$

ارایه می‌شود، که در آن $Y \sim \mathrm{R}((k + 1)p)$. با استفاده از (۵) می‌توان چندک γ-ام توزیع $\mathrm{URG}(\beta_1, p)$ را، پس از محاسبه، به صورت

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

ارایه نمود. بنابراین میانه توزیع $\mathrm{URG}(\beta_1, p)$ با جایگزین نمودن $\gamma = \frac{1}{2}$ در این رابطه قابل دسترسی است.

توزیع N به شرط $X = x$ با استفاده از (۳) به صورت

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

به دست می‌آید، که امید ریاضی شرطی آن عبارتست از

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

قضیه ۱. فرض کنید $\{U_i : i \geq ۱\}$ دنباله‌ای از متغیرهای تصادفی مستقل با توزیع مشترک $\text{URG}(\beta_۱, p)$ و متغیر تصادفی N دارای توزیع $\text{Ge}(q)$، $۰ < q < ۱$، و مستقل از U_iها باشد. اگر $U = \min\{U_۱, ..., U_N\}$، آنگاه $U \sim \text{URG}(\beta_۱, pq)$.

برهان.

$$\begin{aligned} P(U \geq u) &= P(U_۱ \geq u, ..., U_N \geq u) \\ &= \sum_{n=۱}^{\infty} P(U_۱ \geq u, ..., U_N \geq u | N = n) p_n \\ &= \sum_{n=۱}^{\infty} \frac{p^n e^{-n\beta_۱ u^۲}}{[۱ - (۱ - p)e^{-\beta_۱ u^۲}]^n} q(۱ - q)^{n-۱} \\ &= \frac{pqe^{-\beta_۱ u^۲}}{۱ - (۱ - pq)e^{-\beta_۱ u^۲}}. \end{aligned}$$

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 $(X, Y) \sim \text{BRG}(\beta_۱, \beta_۲, p)$، آنگاه ویژگی‌های زیر برقرارند:

الف) $X \sim \text{URG}(\beta_۱, p)$ و $Y \sim \text{URG}(\beta_۲, p)$.

ب) $X \geq x | Y \geq y \sim \text{URG}(\beta_۱, p^*)$، به‌طوری‌که $p^* = ۱ - (۱ - p)e^{-\beta_۲ y^۲}$.

ج) $\min\{X, Y\} \sim \text{URG}(\beta_۱ + \beta_۲, p)$.

د) $P(X \geq x | Y = y) = \frac{e^{-\beta_۱ x^۲}[۱-(۱-p)e^{-\beta_۲ y^۲}]^۲}{[۱-(۱-p)e^{-\beta_۱ x^۲} e^{-\beta_۲ y^۲}]^۲}$.

برهان. قسمت‌های (الف)-(ج) به راحتی و با اندکی محاسبه اثبات می‌گردند. برای اثبات قسمت (د) ابتدا فرض کنید $c(y) = \frac{۲\beta_۲ y[۱-(۱-p)e^{-\beta_۲ y^۲}]^۲}{f_R(y;\beta_۲)}$. توزیع شرطی به‌صورت

$$\begin{aligned} P(X \geq x | Y = y) &= \sum_{n=۱}^{\infty} P(X \geq x | Y = y, N = n) P(N = n | Y = y) \\ &= \sum_{n=۱}^{\infty} c(y) n e^{-n\beta_۱ x^۲} (۱ - p)^{n-۱} e^{-n\beta_۲ y^۲} \\ &= \frac{e^{-\beta_۱ x^۲}[۱ - (۱ - p)e^{-\beta_۲ y^۲}]^۲}{[۱ - (۱ - p)e^{-\beta_۱ x^۲} e^{-\beta_۲ y^۲}]^۲}, \end{aligned}$$

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 $\{(U_i, V_i) : i \geq ۱\}$ دنباله‌ای از متغیرهای تصادفی مستقل با توزیع مشترک $\mathrm{BRG}(\beta_۱, \beta_۲, p)$ و $N \sim \mathrm{Ge}(q)$، $۰ < q < ۱$ باشد. بعلاوه زوج‌های (U_i, V_i) و N مستقل باشند. اگر $U = \min\{U_۱, ..., U_N\}$ و $V = \min\{V_۱, ..., V_N\}$ آنگاه $(U, V) \sim \mathrm{BRG}(\beta_۱, \beta_۲, pq)$.

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

$$f_{X,Y,N}(x, y, n) = ۴n^۲\beta_۱\beta_۲ p(۱ - p)^{n-۱} xye^{-n\beta_۱ x^۲} e^{-n\beta_۲ y^۲}, \qquad (۱۳)$$

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

$$P(N = n|X = x, Y = y) = n^۲ g^{n-۱}(x, y) \frac{\{۱ - g(x, y)\}^۳}{۱ + g(x, y)}, \qquad (۱۴)$$

است، که در آن $g(x, y) = (۱ - p)e^{-\beta_۱ x^۲} e^{-\beta_۲ y^۲}$. چون $M \sim \mathrm{Ge}(q)$، بنابراین $E(M^۳) = \frac{(۱-q)^۲+۴(۱-q)+۱}{q^۳}$. در نتیجه امید ریاضی توزیع شرطی (۱۴) به صورت

$$\begin{aligned} E(N|X = x, Y = y) &= \frac{\{۱ - g(x, y)\}^۳}{۱ + g(x, y)} \sum_{n=۱}^{\infty} n^۳ g^{n-۱}(x, y) \\ &= \frac{g^۲(x, y) + ۴g(x, y) + ۱}{۱ - g^۲(x, y)}, \end{aligned}$$

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 $(X,Y) \sim \text{BRG}(\beta_۱,\beta_۲,p)$، آنگاه پارامتر تنش-مقاومت توزیع رایلی-هندسی دومتغیره عبارت است از

$$\begin{aligned} R = P(X<Y) &= \sum_{i=۱}^{n} P(X<Y, N=n) \\ &= \sum_{i=۱}^{n} \int_{۰}^{\infty}\int_{۰}^{y} f_{X,Y,N}(x,y,n)dxdy \\ &= \sum_{i=۱}^{n} p(۱-p)^{n-۱}\frac{\beta_۱}{\beta_۱+\beta_۲} \\ &= \frac{\beta_۱}{\beta_۱+\beta_۲}. \end{aligned} \tag{۱۵}$$

۳ برآورد پارامترهای مدل با الگوریتم EM

فرض کنید نمونه تصادفی $(x_۱,y_۱),...,(x_n,y_n)$ در دسترس باشد. لگاریتم تابع درستنمایی بر اساس این مشاهدات عبارت است از

$$\begin{aligned} \ell(\beta_۱,\beta_۲,p) &= \sum_{i=۱}^{m} \ln f_{X,Y}(x_i,y_i) \\ &= m\ln p + m\ln\beta_۱ + \sum_{i=۱}^{m}\ln(۲x_i) - \beta_۱\sum_{i=۱}^{m} x_i^۲ \\ &+ m\ln\beta_۲ + \sum_{i=۱}^{m}\ln(۲y_i) - \beta_۲\sum_{i=۱}^{m} y_i^۲ \\ &+ \sum_{i=۱}^{m}\ln\{۱+(۱-p)e^{-\beta_۱ x_i^۲}e^{-\beta_۲ y_i^۲}\} \\ &- ۳\sum_{i=۱}^{m}\ln\{۱-(۱-p)e^{-\beta_۱ x_i^۲}e^{-\beta_۲ y_i^۲}\}. \end{aligned} \tag{۱۶}$$

برای دست‌یابی به برآوردگرهای ماکسیمم درستنمایی پارامترها لازم است ماکسیمم رابطه (۱۶) 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 در استفاده از این الگوریتم کار ساده‌ای نیست. برای اجتناب از این چالش از الگوریتم EM برای محاسبه برآوردهای ماکسیمم درستنمایی استفاده می‌شود.

فرض کنید $(x_۱, y_۱, n_۱), ..., (x_m, y_m, n_m)$ مشاهدات کامل باشند. در اینجا اطلاعات مربوط به n_iها ممکن است در دسترس نباشند. ابتدا فرض کنید p معلوم باشد. لگاریتم تابع درستنمایی بر اساس مشاهدات کامل و بدون در نظر گرفتن مقادیری که به پارامترها بستگی ندارند عبارت است از

$$\ell_c(\beta_۱, \beta_۲) = m \ln \beta_۱ + m \ln \beta_۲ - \beta_۱ \sum_{i=۱}^{m} n_i x_i^۲ - \beta_۲ \sum_{i=۱}^{m} n_i y_i^۲. \quad (۱۷)$$

بر اساس مشاهدات کامل برآوردگرهای ماکسیمم درستنمایی پارامترهای $\beta_۱$ و $\beta_۲$ به ترتیب به صورت

$$\hat{\beta}_۱ = \frac{m}{\sum_{i=۱}^{m} n_i x_i^۲} \quad \text{و} \quad \hat{\beta}_۲ = \frac{m}{\sum_{i=۱}^{m} n_i y_i^۲}$$

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 ${\beta_۱}^{(k)}$ و ${\beta_۲}^{(k)}$ مقادیر پارامترهای $\beta_۱$ و $\beta_۲$ در مرحله k-ام الگوریتم باشند. بنابراین مرحله $(k+۱)$-ام، مرحله E الگوریتم، متناظر است با تشکیل شبهه لگاریتم تابع درستنمایی[۵] که بدون در نظر گرفتن مقادیر ثابت به صورت

$$\ell_s(\beta_۱, \beta_۲) = m \ln \beta_۱ + m \ln \beta_۲ - \beta_۱ \sum_{i=۱}^{m} \tilde{n_i}^{(k)} x_i^۲ - \beta_۲ \sum_{i=۱}^{m} \tilde{n_i}^{(k)} y_i^۲, \quad (۱۸)$$

ارایه می‌شود، که در آن

$$\tilde{n_i}^{(k)} = E(N | X = x_i, Y = y_i, \beta_۱^{(k)}, \beta_۲^{(k)}, p).$$

به منظور دست‌یابی به $\beta_۱^{(k+۱)}$ و $\beta_۲^{(k+۱)}$، مرحله M الگوریتم، کافیست ماکسیمم رابطه (۱۸) بر حسب پارامترها محاسبه شود. بنابراین خواهیم داشت

$$\hat{\beta_۱}^{(k+۱)} = \frac{m}{\sum_{i=۱}^{m} \tilde{n_i}^{(k)} x_i^۲} \quad \text{و} \quad \hat{\beta_۲}^{(k+۱)} = \frac{m}{\sum_{i=۱}^{m} \tilde{n_i}^{(k)} y_i^۲}.$$

[3] Newton-Raphson

[4] profile log-likelihood

[5] pseudo log-likelihood function

این فرآیند تا همگرایی الگوریتم ادامه می‌یابد. به ازای یک مقدار ثابت p برآوردهای به‌دست آمده با $\hat{\beta}_1(p)$ 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 مطالعه‌ای شبیه‌سازی با اندازه‌های مختلف حجم نمونه و مقادیر پارامترها انجام می‌شود. توجه کنید که تولید داده از مدل $\mathrm{BRG}(\beta_1, \beta_2, p)$ ساده است. برای این منظور، ابتدا N از مدل هندسی تولید می‌شود. سپس با مشاهده $N = n$، X و Y می‌توانند به ترتیب از توزیع‌های $\mathrm{R}(\beta_1)$ و $\mathrm{R}(\beta_2)$ تولید شوند. مجموعه پارامترهای در نظر گرفته شده در این شبیه‌سازی عبارتند از

الف) $\beta_1 = \beta_2 = ۲$ و $p = ۰/۲۵$،

ب) $\beta_1 = ۱$، $\beta_2 = ۲$ و $p = ۰/۵$،

ج) $\beta_1 = \beta_2 = ۲$ و $p = ۰/۷۵$.

ابتدا داده‌هایی از مدل رایلی-هندسی دومتغیره به ازای حجم نمونه‌های مختلف تولید می‌شوند. سپس با الگوریتم EM برآوردهای ماکسیمم درستنمایی پارامترها محاسبه می‌شوند. در همه حالت‌ها از مقادیر اولیه $\beta_1^{(0)} = \beta_2^{(0)} = ۰/۵۰$ و $p^{(0)} = ۰/۵۰$ استفاده می‌شود. زمانی الگوریتم به کار خود خاتمه می‌دهد که تفاضل تابع درستنمایی در دو تکرار متوالی کمتر از $۱۰^{-۵}$ 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.

همان طور که ملاحظه می‌شود با افزایش حجم نمونه از میزان اریبی و میانگین‌های توان دوم خطاها کاسته می‌شود، که به نوعی تصدیق‌کننده سازگاری برآوردگرهای ماکسیمم درستنمایی هستند. همچنین با افزایش مقدار p میانه تعداد تکرارها تا همگرایی الگوریتم کاهش می‌یابد. در واقع هنگامی‌که $p \to ۱$، 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.

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

n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{p}$	$\mathrm{MSE}(\beta_1)$	$\mathrm{MSE}(\beta_2)$	$\mathrm{MSE}(p)$	میانه تعداد تکرارها
$(\beta_1, \beta_2, p) = (۲, ۲, ۰/۲۵)$							
۲۵	۲/۵۷۲	۲/۵۴۴	۰/۳۳۷	۱/۱۸۳	۱/۲۶۲	۰/۰۲۶	۸۷
۵۰	۲/۲۵۸	۲/۲۸۳	۰/۲۹۵	۰/۵۲۴	۰/۵۳۹	۰/۰۱۱	۹۲
۷۵	۲/۱۶۹	۲/۱۷۶	۰/۲۷۸	۰/۳۳۹	۰/۳۴۳	۰/۰۰۸	۹۷
۱۰۰	۲/۱۲۵	۲/۱۱۱	۰/۲۶۷	۰/۲۴۹	۰/۲۴۳	۰/۰۰۵	۱۰۰
$(\beta_1, \beta_2, p) = (۱, ۲, ۰/۵)$							
۲۵	۱/۱۴۳	۲/۲۵۵	۰/۵۸۶	۰/۱۴۹	۰/۵۸۹	۰/۰۴۷	۴۰
۵۰	۱/۰۶۶	۲/۱۲۰	۰/۵۴۲	۰/۰۶۸	۰/۲۹۰	۰/۰۲۹	۴۲
۷۵	۱/۰۵۵	۲/۱۱۱	۰/۵۳۷۶	۰/۰۴۸	۰/۲۰۵	۰/۰۲۰	۴۳
۱۰۰	۱/۰۴۰	۲/۰۸۴	۰/۵۳۱	۰/۰۳۳	۰/۱۴۳	۰/۰۱۴	۴۴
$(\beta_1, \beta_2, p) = (۲, ۲, ۰/۷۵)$							
۲۵	۲/۰۹۰	۲/۰۸۲	۰/۷۷۰	۰/۳۳۳	۰/۳۱۴	۰/۰۴۱	۳۵
۵۰	۲/۰۷۵	۲/۰۸۳	۰/۷۸۰	۰/۱۷۵	۰/۱۷۰	۰/۰۲۸	۳۷
۷۵	۲/۰۴۷	۲/۰۴۰	۰/۷۷۰	۰/۱۲۰	۰/۱۱۰	۰/۰۲۲	۳۸
۱۰۰	۲/۰۴۱	۲/۰۴۰	۰/۷۶۸	۰/۰۸۹	۰/۰۸۴	۰/۰۱۸	۳۹

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.

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 برآوردگرهای ماکسیمم درستنمایی پارامترها استفاده می‌شود. حدس زدن

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

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

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 ادعاست.

پرسش مهمی که مطرح می‌شود آن است که آیا مدل رایلی-هندسی دومتغیره قادر است برازش مناسبی به داده‌ها ارایه دهد یا خیر. 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 می‌نیمم آن‌ها برازش داده می‌شود.

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

جدول ۳: نتایج برازش توزیع‌های حاشیه‌ای

متغیرها	مقدار آماره آزمون K–S	p–مقدار
$X_۱$	۰/۱۳۳۵	۰/۴۸۳۵
$X_۲$	۰/۱۲۳۳	۰/۵۸۴۷
$\min\{X_۱, X_۲\}$	۰/۱۵۰۲	۰/۳۳۹۷

حال توزیع رایلی–هندسی دومتغیره به داده‌ها برازش داده می‌شود. با استفاده از داده‌های فوتبال، برآوردهای ماکسیمم درستنمایی پارامترها در مدل دومتغیره عبارتند از $\hat{\beta}_۱ = ۰/۰۰۰۴۳۸$، $\hat{\beta}_۲ = ۰/۰۰۰۲۹۳$ و $\hat{p} = ۰/۴۳۰۲$. بر اساس برآوردهای به دست آمده مقدار لگاریتم تابع درستنمایی برابر ۳۲۶/۹۳۸– با ضریب اطلاع بیزی $\mathrm{BIC} = ۶۶۴/۷۱$ و ضریب اطلاع آکائیک $\mathrm{AIC} = ۶۵۹/۸۷$ است.

برآورد پارامترهای مدل نرمال دومتغیره به داده‌های فوتبال نیز عبارتند از $\boldsymbol{\mu} = (۳۲/۸۶, ۴۰/۸۹)'$ و $\Sigma = \begin{bmatrix} ۲۱۰/۲۰ & ۵۰۷/۲۵ \\ ۳۹۴/۵۰ & ۲۱۰/۲۰ \end{bmatrix}$. 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. گر چه برآوردگرهای 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 برخود 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

[1] Adamidis, K. and Loukas, S. (1998), A Lifetime Distribution with Decreasing Failure Rate, *Statististics & Probabability Letters*, **39**, 35–42.

[2] 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.

[3] Bidram, H., Behboodian, J. and Towhidi, M. (2013), A New Generalized Exponential Geometric Distribution, *Communications in Statistics—Theory and Methods*, **42**, 528–542.

[4] Kundu, D. (2015), Bivariate Geometric (Maximum) Generalized Exponential Distribution, *Journal of Data Sciences*, **13**, 693-712.

[5] Kundu, D. and Gupta, A. K. (2014), On Bivariate Weibull-Geometric Distribution, *Journal of Multivariate Analysis*, **123**, 19-29.

[6] 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.

[7] 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.

[8] Mahmoudi, E. and Sepahdar, A. (2013), Exponentiated Weibull–Poisson Distribution: Model, Properties and Applications, *Mathematics and Computers in Simulation*, **92**, 76–97.

[9] Meintanis, S. G. (2007), Test of Fit for Marshall-Olkin Distributions with Applications, *Journal of Statistical Planning and inference* **137**, 3954-3963.

[10] 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.

[11] Ristic, M. M. and Kundu, D. (2015), Marshall-Olkin Generalized Exponential Distribution, *METRON*, **73**, 317-333.

[12] Wang, M. and Elbatal, I. (2014), The Modified Weibull Geometric Distribution, *METRON*, DOI: 10.1007/s40300-014-0052-1.

Bivariate Rayleigh-geometric distribution

V. Nekoukhou[1], **A. Khalifeh**[2] **and E. Mahmoudi**[2]

[1]Department of Statistics, Khansar Faculty of Mathematics and Computer Science, Khansar, Iran.

[2]Department of Statistics, Yazd University, Yazd, Iran.

Abstract: In this paper, we study a three-parameter bivariate distribution obtained by taking Geometric minimum of Rayleigh distributions. Some important properties of this bivariate distribution have been investigated. It is observed that the maximum likelihood estimators of the parameters cannot be obtained in closed forms. We propose to use the EM algorithm to compute the maximum likelihood estimates of the parameters, and it is computationally quite tractable. Based on an extensive simulated study, the effectiveness of the proposed algorithm is confirmed. We also analyze one real data set for illustrative purposes. Finally, we conclude the paper.

Keywords: Rayleigh distribution, Geometric distribution, Estimation.

Mathematics Subject Classification (2010): 62F10, 62F03, 62H12.