نگاهی به الگوهای آماری تعمیم یافته

سید محمد تقی کامل میرمصطفایی[۱] و جعفر احمدی[۲]

چکیده:

اخیراً موضوع توزیع‌های کلاسیک آماری تعمیم یافته نظیر بتا نرمال، بتا نمایی، وایبل نمایی شده، نمایی تعمیم یافته و غیره مورد توجه پژوهشگران زیادی قرار گرفته است. نشان داده شده است که در برخی مواقع الگوهای تعمیم یافته از توزیع‌های پایه، انعطاف پذیرتر بوده و در مسائل آزمون نیکویی برازش بهتر عمل می‌کنند. دراین مقاله، ابتدا مروری بر تحقیقات به عمل آمده در این راستا داریم، آن‌گاه توجه خود را به داده‌های ترتیبی معطوف می‌کنیم. الگوهای برخاسته از توزیع‌های آماره‌های مرتب را مورد مطالعه قرار می‌دهیم. در ادامه به معرفی الگوی تعمیم یافته جدیدی از توزیع‌ها بر اساس مقادیر رکوردها می‌پردازیم. خواص الگوی معرفی شده را بررسی نموده، نشان می‌دهیم بعضی از خواص جامعه اولیه توسط الگوی معرفی شده حفظ می‌گردد.

واژه‌های کلیدی: آماره‌های مرتب، رکوردها، توزیع نمایی، نامساوی هولدر، تقارن، چولگی.

۱ مقدمه

رفتار پدیده‌های تصادفی انگیزه‌ای بود که الگوهای آماری معرفی شوند تا بتوان این گونه پدیده‌ها را در چارچوبی خاص بیان و کنترل کرد. این‌جا بود که توزیع‌های کلاسیک مهمی مانند توزیع نمایی، توزیع گاما، توزیع وایبل، توزیع بتا و غیره پا به عرصه‌ی وجود گذاشتند و توانستند الگوهای قابل قبولی برای داده‌های اقتصادی، طول عمر، آزمایشگاهی و غیره باشند. اما با پیشرفت علم و گسترش علوم در رشته‌های مختلف، الگوهای کلاسیک جواب گوی مسائل جدید نبودند، این مسأله آماردانان را بر آن داشت تا مدل‌های جدیدی معرفی نمایند به طوری که شامل توزیع‌های کلاسیک بوده و دارای انعطاف پذیری بیشتری باشند. در این راستا، پژوهشگران علم آمار به تعمیم الگوهای کلاسیک پرداختند. از جمله نخستین آماردانانی که به این مهم پرداخت، آموروسو[۳] [۲] بود که صورت تعمیم یافته‌ی توزیع گاما را معرفی کرد. او این توزیع را به داده‌های نرخ درآمد برازش داد. از آن زمان به بعد، سایر پژوهشگران کلاس‌های متنوعی از توزیع‌های تعمیم

[۱] دانشجوی دکتری آمار—دانشکده علوم ریاضی — دانشگاه فردوسی مشهد
[۲] استاد آمار—دانشکده علوم ریاضی — دانشگاه فردوسی مشهد
[۳] Amoroso

یافته را مانند توزیع گوسین معکوس تعمیم یافته، توزیع پارتوی تعمیم یافته، توزیع بتای تعمیم یافته و غیره معرفی نمودند که هر کدام در نوع خود دارای خواص انعطاف پذیری بیشتری در مقایسه با الگوی کلاسیک متناظر بود. اخیراً گوپتا[۴] و کوندا[۵] [۷] توزیع نمایی نمایی‌شده یا توزیع نمایی تعمیم‌یافته را معرفی نموده، خواص آن را در طی مقالات متعددی (گوپتا و کوندا([۸] و [۹])) بررسی و مطالعه کردند. ایجن و همکاران[۶] [۶] خانواده جدیدی از توزیع های آماری را معرفی کردند که الگویی تعمیم یافته بر اساس متغیر بتا بود. جونز[۷] [۴] از دیدگاه آماره‌های مرتب، الگوی معرفی شده توسط ایجن و همکاران [۶] را مورد بحث قرار داده و نتایج را در یک مقاله‌ی مفصل در مجله‌ی $Test$ به چاپ رساند که توسط آماردانان معروف در زمینه‌ی آماره‌های مرتب نظیر آرنولد[۸]، ناگاراجا[۹]، دیوید[۱۰] و کنت[۱۱] مورد نقد و بررسی قرار گرفت. ادامه‌ی این مقاله به شرح زیر است: در بخش ۲ مروری کوتاه بر آماره‌های مرتب و کاربردهای آن داریم. خواص الگوی معرفی شده توسط جونز در بخش ۳ ارائه می‌شود. در بخش ۴ آماره‌های رکوردی را در نظر می‌گیریم. در نهایت، به معرفی الگوی جدید تعمیم یافته بر اساس آماره‌های رکوردی و برخی از خواص آن می‌پردازیم.

۲ آماره‌های مرتب

فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی از جامعه‌ای پیوسته با تابع توزیع تجمعی $F_X(x)$ و تابع چگالی احتمال $f_X(x)$ باشند. اگر X_i ها را به ترتیب صعودی مرتب کنیم و مقادیر مرتب شده را با $X_{i:n}$ نشان دهیم، آن گاه $X_{i:n}$ را i امین آماره مرتب از نمونه فوق گویند. تابع چگالی احتمال $X_{i:n}$ به صورت زیر به دست می آید:

$$\begin{aligned} f_{X_{i:n}}(x) &= \frac{n!}{(i-1)!(n-i)!}[F_X(x)]^{i-1} \\ &\times[\bar{F}_X(x)]^{n-i} f_X(x). \qquad (1) \end{aligned}$$

که در آن $\bar{F}_X(x) = 1 - F_X(x)$.

امروزه نقش آماره‌های مرتب آن چنان پررنگ شده که بسیاری از مباحث گوناگون آمار را به خود اختصاص داده است. آماره‌های مرتب در قسمت‌های مختلف توصیفی و استنباطی علم آمار دارای کاربرد هستند. به عنوان مثالی از کاربرد این آماره‌ها در آمار توصیفی می‌توان به کشف مشاهدات پرت (داده‌های خیلی کوچک یا خیلی بزرگ) در یک مجموعه از داده‌ها اشاره نمود. همان گونه که می‌دانید یکی از ضعف‌های میانگین به عنوان معیاری جهت تمیز تمرکز داده‌ها میزان

[۴] Gupta
[۵] Kundu
[۶] Eugene et al.
[۷] Jones
[۸] Arnold
[۹] Nagaraja
[۱۰] David
[۱۱] Kent

حساسیت بالای آن نسبت به داده‌های دورافتاده و تغییرات الگو است. در عوض میانه نسبت به مفروضات الگو از حساسیت کمتری برخوردار است. آماره‌های مرتب در بسیاری از موارد به عنوان آماره‌های بسنده و گاه بسنده‌ی کامل معرفی می‌شوند و برآوردگرهای نااریب با کمترین واریانس، فاصله‌های اطمینان مناسب و تواناترین آزمون را برای پارامتر مجهول فراهم می آورند.
در موضوعات کنترل کیفیت برای بررسی در کنترل بودن تولید، اغلب از نمودارهای میانگین و دامنه تغییرات یا از نمودارهای میانه و دامنه تغییرات استفاده می‌شود که برای محاسبه میانه و دامنه تغییرات ناچار به استفاده از آماره‌های مرتب هستیم. برای جزئیات بیشتر در رابطه با کاربردهای وسیع آماره‌های مرتب می‌توان به دو کتاب معروف آرنولد و همکاران [۱۲] [۳] و دیوید و ناگاراجا [۵] مراجعه نمود. در بخش‌های بعدی تعمیمی از رابطه‌ی (۱) ارائه می‌شود.

لم ۱ فرض کنید متغیر تصادفی X دارای توزیع پیوسته‌ی $F_X(x)$ باشد، قرار دهید $Y = F_X(X)$، آن‌گاه Y دارای توزیع یکنواخت بر بازه‌ی $(۰, ۱)$ است.

برهان: بدیهی است. □

لم ۲ فرض کنید $X_۱, X_۲, ..., X_n$ یک نمونه‌ی تصادفی از جامعه‌ای پیوسته با تابع توزیع $F_X(x)$ باشد، آن‌گاه برای $۱ \leq i \leq n$، $U_{i:n}$ با $F_X(X_{i:n})$ هم‌توزیع می‌باشد که در آن $U_{i:n}$، i امین آماره‌ی مرتب در نمونه‌ای به حجم n از جامعه‌ی یکنواخت بر بازه‌ی $(۰, ۱)$ می‌باشد.

برهان: با توجه به صعودی بودن تابع $F_X(x)$، بدیهی است. □

نتیجه ۱ اگر $X_۱, X_۲, ..., X_n$ یک نمونه‌ی تصادفی از جامعه‌ای پیوسته با تابع توزیع $F_X(x)$ باشد، آن‌گاه $F_X(X_{i:n})$ دارای توزیع بتا با پارامترهای $\alpha = i$ و $\beta = n - i + ۱$ می‌باشد.

توزیع بتا دارای خواص جالب توجه متعددی است. در حالتی که $\alpha = \beta = ۱$ باشد، توزیع بتا همان توزیع یکنواخت بر بازه‌ی $(۰, ۱)$ خواهد بود و در صورتی که $\alpha = \beta \neq ۱$ توزیع بتا به شکل توزیعی متقارن بر بازه‌ی $(۰, ۱)$ در خواهد آمد که دم‌های آن همراه با افزایش $\alpha > ۱$ نازک‌تر و با کاهش $\alpha < ۱$ پهن‌تر خواهد شد. اگر $\alpha \neq \beta$، آن‌گاه توزیع بتا، یک توزیع چوله خواهد بود. توزیع بتا با بسیاری از توزیع‌ها در ارتباط است اما زمانی که α و β هر دو عددی مثبت و صحیح باشند، شاید رابطه این توزیع با توزیع دوجمله‌ای بیشتر مورد توجه باشد. برای جزئیات بیشتر به کتاب توزیع‌های پیوسته جانسون [۱۳] و کوتز [۱۴] [۱۰] مراجعه کنید.

[۱۲] Arnold et al.
[۱۳] Johnson
[۱۴] Kotz

۳ الگوی کلی تعمیم یافته بر اساس آماره‌های مرتب

فرض کنید Y دارای توزیع بتا با پارامترهای α و β باشد. ایجن و همکاران [۶] با فرض یک تابع توزیع F، متغیر تصادفی $X = F^{-1}(Y)$ را در نظر گرفتند که در این صورت تابع چگالی احتمال X به صورت زیر می‌باشد:

$$g_F(x;\alpha,\beta)= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times[F(x)]^{\alpha-1}[\bar{F}(x)]^{\beta-1}f(x). \quad (۲)$$

که در آن $\Gamma(.)$ تابع گامای کامل است و به صورت زیر تعریف می‌شود:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy.$$

سپس آن‌ها با فرض این‌که F تابع توزیع متغیر نرمال باشد، به مطالعه‌ی خواص متغیرتصادفی $X = F^{-1}(Y)$ پرداختند و الگوی ایجاد شده را بتانرمال نام نهادند. جونز [۴] با نگرش به توزیع آماره‌های مرتب، الگوی جدیدی مشابه (۲) معرفی نمود. لازم به ذکر است که با فرض $\alpha = i$ و $\beta = n - i + 1$ در (۱)، الگوی جونز به دست می‌آید. در ادامه خواص الگوی جونز را مطالعه می‌کنیم.

۱.۳ تقارن

فرض کنید X دارای تابع چگالی احتمال $f_X(x)$ و تابع توزیع تجمعی $F_X(x)$ باشد، گوییم توزیع X حول a متقارن است هرگاه $f_X(a-x) = f_X(a+x)$ یا $F_X(a+t) = 1 - F_X(a-t)$. در لم زیر نشان داده می‌شود که خاصیت تقارن از جامعه‌ی اصلی به الگوی جونز منتقل می‌گردد.

لم ۳ اگر F توزیعی متقارن حول صفرباشد و G_F کلاس توزیع‌های تولید شده به فرم (۲) توسط F باشد، آن‌گاه توزیع G_F نیز حول صفر متقارن است اگر و فقط اگر $\alpha = \beta$.

برهان:

$$\begin{aligned} g_F(-x;\alpha,\beta) &= \frac{1}{B(\alpha,\beta)}[F(-x)]^{\alpha-1} \\ &\times[\bar{F}(-x)]^{\beta-1}f(-x) \\ &= \frac{1}{B(\alpha,\beta)}[\bar{F}(x)]^{\alpha-1} \\ &\times[F(x)]^{\beta-1}f(x) \\ &= g_F(x;\beta,\alpha). \end{aligned}$$

پس اگر $\alpha = \beta$، آن‌گاه $g_F(-x) = g_F(x)$ و در نتیجه توزیع g_F متقارن خواهد بود. □

در رابطه با خواص چولگی و رفتار دم ها مشابه آن‌چه که در رابطه با توزیع بتا بیان شد، می‌توان گفت در صورتی که g_F متقارن باشد، دم‌ها همراه با افزایش $\alpha > 1$ نازک‌تر و با کاهش $\alpha < 1$ پهن‌تر خواهند شد. اگر $\alpha \neq \beta$ ، آن‌گاه توزیع g_F ، یک توزیع چوله خواهد بود که میزان چولگی به میزان تفاوت α و β وابسته است و علامت آن به علامت $\alpha - \beta$ بستگی دارد.

۲.۳ پارامترهای مکان و مقیاس

سه خانواده‌ی پارامتری معروف در آمار داریم که عبارتند از:

الف) خانواده‌ی مکانی: F را متعلق به خانواده‌ی مکان و θ را پارامتر مکان گویند هرگاه $F(x;\theta) = F_{\circ}(x-\theta)$.

ب) خانواده‌ی مقیاس: F را متعلق به خانواده‌ی مقیاس و $\theta > \circ$ را پارامتر مقیاس گویند هرگاه $F(x;\theta) = F_{\circ}(\theta x)$.

ج)خانواده‌ی شکل: F را متعلق به خانواده‌ی شکل و $\theta > \circ$ را پارامتر شکل گویند هرگاه $F(x;\theta) = F_{\circ}(x^{\theta})$.

در لم زیر نشان داده می‌شود که اگر F متعلق به خانواده‌ی مکان—مقیاس باشد، آن‌گاه این خواص برای الگوی (۲) حفظ می‌گردد.

لم ۴ (جونز، [۴]) اگر F متعلق به خانواده‌ی مکان—مقیاس باشد، خانواده‌ی توزیع‌های معرفی شده توسط الگوی (۲) متعلق به خانواده مکان—مقیاس است.

برهان:

$$\begin{aligned} g_F(x;\alpha,\beta,\mu,\sigma) &= \frac{۱}{B(\alpha,\beta)}\left[F_{\circ}(\frac{x-\mu}{\sigma})\right]^{\alpha-۱} \\ &\times \left[\bar{F}_{\circ}(\frac{x-\mu}{\sigma})\right]^{\beta-۱}\frac{۱}{\sigma}f_{\circ}(\frac{x-\mu}{\sigma}) \\ &= \frac{۱}{\sigma}g_{F_{\circ}}(\frac{x-\mu}{\sigma};\alpha,\beta,\circ,\circ) \\ &= \frac{۱}{\sigma}g_{F_{\circ}}(\frac{x-\mu}{\sigma};\alpha,\beta). \end{aligned} \tag{۳}$$

اگر Z دارای تابع چگالی احتمال استاندارد شده الگوی (۲) و U دارای تابع چگالی احتمال به فرم (۳) باشد، بدیهی است که $E(U) = \mu + \sigma\mu_G$ و $Var(U) = \sigma^{۲}\sigma_G^{۲}$ که $\mu_G = E(Z)$ و $\sigma_G^{۲} = Var(Z)$.

۳.۳ تابع توزیع

فرض کنید $G_F(x)$ تابع توزیع احتمال الگوی (۲) باشد، آن‌گاه بلافاصله از تبدیل $U = F(X)$ نتیجه می‌شود که

$$\begin{aligned} G_F(x) &= P(Y \le F(x)) \\ &= \int_{\circ}^{F(x)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}u^{\alpha-۱}(۱-u)^{\beta-۱}du. \end{aligned}$$

با توجه به رابطه توزیع بتا و دوجمله‌ای چنانچه α و β اعداد صحیح و مثبت باشند، تابع توزیع G_F به صورت زیر در خواهد آمد:

$$\begin{aligned} G_F(x) &= \sum_{i=\alpha}^{\alpha+\beta+۱}\binom{\alpha+\beta+۱}{i} \\ &\times [F(x)]^i[\bar{F}(x)]^{\alpha+\beta-i-۱}. \end{aligned}$$

۴.۳ گشتاورها

فرض کنید X دارای تابع چگالی احتمال $f_X(x)$ و تابع توزیع تجمعی $F_X(x)$ باشد، آن‌گاه داریم

$$E(X) = \int x f_X(x)dx = \int_{\circ}^{۱} F_X^{-۱}(x)dx.$$

در لم زیر شرایط وجود گشتاورهای الگوی جونز ارائه می‌شود.

لم ۵ وجود یک گشتاور مرتبه‌ی $k+\delta$ با فرض $\delta > 0$ در جامعه اصلی، وجود گشتاورهای مرتبه‌ی k را در جامعه تولید شده توسط الگوی (۲)، نتیجه می‌دهد.

برهان: فرض کنید $p > 1$ و $\frac{1}{p}+\frac{1}{q}=1$. اگر $T \sim G_F(\alpha,\beta)$ داریم:

$$\begin{aligned} E(T^k) &= \frac{1}{B(\alpha,\beta)} \\ &\quad \times \int_{-\infty}^{\infty} t^k [F(t)]^{\alpha-1}[\bar{F}(t)]^{\beta-1} f(t)dt \\ &= \frac{1}{B(\alpha,\beta)} \\ &\quad \times \int_0^1 [F^{-1}(y)]^k y^{\alpha-1}(1-y)^{\beta-1} dy \end{aligned}$$

که تساوی اخیر از تبدیل $y=F(t)$ نتیجه می‌شود. با استفاده از نامساوی هولدر داریم:

$$\begin{aligned} E(T^k) &= E\left\{[F^{-1}(y)]^k\right\} \\ &\leq \frac{1}{B(\alpha,\beta)} \\ &\quad \times \left\{\int_0^1 \left[y^{\alpha-1}(1-y)^{\beta-1}\right]^q dy\right\}^{\frac{1}{q}} \\ &\quad \times \left\{\int_0^1 [F^{-1}(y)]^{kp} dy\right\}^{\frac{1}{p}} \\ &= \frac{B(\alpha q-q+1,\beta q-q+1)}{B(\alpha,\beta)} \\ &\quad \times \left\{\int_{-\infty}^{\infty} x^{k+\delta} f(x)dx\right\}^{\frac{1}{p}} \\ &= \frac{B(\alpha q-q+1,\beta q-q+1)}{B(\alpha,\beta)} \\ &\quad \times \left[E(X^{k+\delta})\right]^{\frac{1}{p}} \end{aligned}$$

که در آن $X \sim F$ و $kp=k+\delta$. □

۵.۳ مد

فرض کنید X دارای تابع چگالی احتمال $f_X(x)$ و تابع توزیع تجمعی $F_X(x)$ باشد، گوییم نقطه‌ی M مد F است اگر $\delta > 0$ موجود باشد به گونه‌ای که برای هر $x \in (M-\delta, M+\delta)$ داشته باشیم $f_X(M) \geq f_X(x)$. اکنون خاصیت تک مدی بودن الگوی (۲) را بررسی می‌کنیم.

لم ۶ اگر F به طور پیوسته مشتق پذیر و تک مد باشد و در الگوی (۲)، شرط $\alpha=\beta\geq 1$ برقرار باشد، آن‌گاه g_F نیز تک مد است.

برهان: فرض کنید که M مد توزیع F باشد، آن‌گاه بنا به لم ۳ و فرض، $F(M)=\frac{1}{2}$. برای $\alpha=\beta$ از (۲) داریم

$$g_F(x;\alpha,\alpha)=\frac{1}{B(\alpha,\alpha)}[F(x)\bar{F}(x)]^{\alpha-1}f(x).$$

بنابراین کافی است که نشان دهیم عبارت داخل براکت تنها ماکزیمم خود را در $x=M$ اختیار می‌کند. فرض کنید $Q(x)=F(x)\bar{F}(x)$، آن‌گاه

$$Q'(x)=f(x)[\bar{F}(x)-F(x)].$$

$Q'(x)=0$ اگر و فقط اگر $1-2F(x)=0$ یا $f(x)=0$. اما اگر $f(t)=0$، بدیهی است که $x=t$ نمی‌تواند مد F باشد. بنابراین حالتی را در نظر می‌گیریم که $1-2F(x)=0$ که در این صورت $F(x)=\frac{1}{2}$. از آن‌جا که F تابعی اکیداً صعودی است، $x=M$ و اگر $y < M$، آن‌گاه $1-2F(y) > 0$ و اگر $y > M$

آن‌گاه $۰ < ۲F(y) - ۱$. واضح است که $Q(x)$ با $۲F(x) - ۱$ هم‌علامت است و در نتیجه برهان کامل است. □

با این وجود، در حالت کلی تک مد بودن f به تک مد بودن g_F منجر نمی‌شود. در حالتی که $\alpha < ۱$ یا $\beta < ۱$ باشد، نمی‌توان یک قانون کلی بیان کرد. ایجن و همکاران [٦] در حالتی خاص که F توزیع نرمال باشد، نشان داده‌اند چنانچه α و β هر دو کمتر از ۰/۲۱۴ باشند، توزیع g_F دو مد خواهد داشت. با این وجود، لم زیر که تعمیمی از قضیه‌ای است که آلم[۱۵] [۱] در رابطه با تک مد بودن آماره‌های ترتیبی بیان و اثبات نمود، یک قانون کلی برای تک مد بودن g_F بیان می‌دارد.

لم ۷ اگر X دارای توزیع F و تابع چگالی احتمال f باشد و g_F الگوی معرفی شده در (۲) باشد و $\frac{۱}{f}$ تابعی محدب باشد، آن‌گاه g_F به ازای هر مقدار $\alpha \geq ۱$ و $\beta \geq ۱$، تک مد خواهد بود.

برهان: می دانیم

$$g_F(x;\alpha,\beta) = \frac{۱}{B(\alpha,\beta)}[F(x)]^{\alpha-۱}[\bar{F}(x)]^{\beta-۱}f(x).$$

شرط محدب بودن $\frac{۱}{f}$ بیان می‌کند که f از راست مشتق پذیر و $\frac{f'(x)}{f^۲(x)}$ غیرصعودی است که $f'(x)$ مشتق راست f است. بنابراین

$$g'_F(x;\alpha,\beta) = \frac{۱}{B(\alpha,\beta)}[F(x)]^{\alpha-۱}[\bar{F}(x)]^{\beta-۱}f^۲(x) \times\left\{\frac{f'(x)}{f^۲(x)}+\frac{\alpha-۱}{F(x)}-\frac{\beta-۱}{\bar{F}(x)}\right\} \quad (۴)$$

[۱۵] Alam

که g'_F مشتق راست g است. چون $\frac{f'(x)}{f^۲(x)}$ تابعی غیرصعودی از x است، کمیت داخل آکولاد در سمت راست (۴) غیر صعودی است. پس زمانی که x از $-\infty$ به ∞ تغییر یابد، g'_F حداکثر یک تغییر علامت دارد و اگر g'_F تغییر علامت دهد، این تغییر علامت از مثبت به منفی خواهد بود. لذا g_F تک مد است. □

۶.۳ آماره‌های بسنده

اگر $X_۱, X_۲, ..., X_n$ یک نمونه‌ی تصادفی از الگوی (۲) با شاخص‌های پارامتری α و β و توزیع F باشند، آن‌گاه چنانچه F شامل پارامتر نباشد، داریم

$$\begin{aligned} g_F(x_۱,...,x_n,\alpha,\beta) &= \prod_{i=۱}^{n} g_F(x_i,\alpha,\beta) \\ &= [B(\alpha,\beta)]^{-n}\prod_{i=۱}^{n}[F(x_i)]^{\alpha-۱} \\ &\times[\bar{F}(x_i)]^{\beta-۱}f(x_i) \\ &= [B(\alpha,\beta)]^{-n}e^{u(x_۱,...,x_n)} \\ &\times e^{(\alpha-۱)\sum_{i=۱}^{n}\log F(x_i)} \\ &\times e^{(\beta-۱)\sum_{i=۱}^{n}\log \bar{F}(x_i)} \qquad (۵) \end{aligned}$$

که $u(x_۱,...,x_n) = \log\prod_{i=۱}^{n} f(x_i)$. با توجه به فرم (۵) مشاهده می‌کنیم که g_F متعلق به خانواده‌ی نمایی ۲ پارامتری است. فرض کنید $T_۱(X_۱,...X_n) = \log\prod_{i=۱}^{n} F(X_i)$ و

$T_۲(X_۱,...X_n)=\log\prod_{i=۱}^{n}\bar{F}(X_i)$ دراین صورت

$$g_F(x_۱,...,x_n,\alpha,\beta) = [B(\alpha,\beta)]^{-n}e^{u(x_۱,...,x_n)} \times e^{(\alpha-۱)T_۱(x_۱,...x_n)} \times e^{(\beta-۱)T_۲(x_۱,...x_n)}$$

و در نتیجه

$$T=(T_۱(X_۱,...,X_n),T_۲(X_۱,...,X_n))$$

آماره‌ی بسنده‌ی کامل برای $\theta=(\alpha,\beta)$ خواهد بود.

۴ الگوی کلی تعمیم یافته بر اساس آماره‌های رکوردی

فرض کنید $\{X_i, i\geq ۱\}$ دنباله‌ای از متغیرهای تصادفی باشد، گوییم X_j رکورد بالاست هرگاه از تمام مقادیر ماقبل خود بزرگ‌تر باشد، یعنی

$$X_j > Max\{X_۱,...,X_{j-۱}\}.$$

به طور مشابه گوییم X_k رکورد پایین است هرگاه از تمام ماقبل خود کوچک‌تر باشد، یعنی

$$X_k < Min\{X_۱,...,X_{k-۱}\}.$$

در این مقاله n امین رکورد بالا و پایین را به ترتیب با U_n و L_n نشان می‌دهیم.
این نوع داده‌ها در آزمایش‌های دنباله‌ای نظیر آزمایش‌های تخریبی، مهندسی عمران، در مسائل زلزله‌شناسی، هواشناسی و رشته‌های ورزشی رخ می‌دهند.برای مطالعه‌ی جزئیات بیشتر در خصوص خواص و کاربرد آماره‌های رکوردی می‌توان به کتاب آرنولد و همکاران [۴] مراجعه نمود.
اگر $\{X_i, i\geq ۱\}$ دنباله‌ای از متغیرهای تصادفی پیوسته‌ی هم‌توزیع و مستقل با تابع توزیع تجمعی $F(x)$ و تابع چگالی احتمال $f(x)$ باشد. آن‌گاه تابع چگالی احتمال n امین مقدار رکورد بالایی، U_n، عبارت است از

$$f_{U_n}(x)=\frac{۱}{\Gamma(n)}[-\log\bar{F}(x)]^{n-۱}f(x), \quad -\infty<x<\infty. \quad (۶)$$

به طور مشابه تابع چگالی احتمال n امین رکورد پایین به صورت زیر می‌باشد:

$$f_{L_n}(x)=\frac{۱}{\Gamma(n)}[-\log F(x)]^{n-۱}f(x), \quad -\infty<x<\infty. \quad (۷)$$

رکوردهای بالا شامل اطلاعات بیشتری در خصوص دم سمت راست توزیع نسبت به رکوردهای پایین می‌باشند، همچنین رکوردهای پایین نیز شامل اطلاعاتی بیشتر در خصوص دم سمت چپ توزیع می‌باشند. لذا در نظر داریم یک خانواده‌ی کلی جدیدی از توزیع‌ها بر حسب توزیع رکوردها بسازیم که تلفیقی توأم از رکوردهای بالا و پایین باشد. این خانواده را به صورت زیر معرفی می‌کنیم

$$h_F(t;\alpha,\beta) = \frac{۱}{\gamma(\alpha,\beta)}[-\log\bar{F}(t)]^{\alpha-۱} \times[-\log F(t)]^{\beta-۱}f(t) \quad (۸)$$

که α و β مثبت بوده و $\gamma(\alpha,\beta)$ به شکل زیر تعریف می‌شود:

$$\gamma(\alpha,\beta)=\int_{0}^{\infty} y^{\alpha-1}[-\log(1-e^{-y})]^{\beta-1}e^{-y}dy$$

که آن را تابع گامای توسعه یافته می نامیم. خانواده‌ی معرفی شده در (۸) توسط شاخص‌های پارامتری α و β و تابع توزیع F مشخص می‌شود. بدیهی است که $\gamma(\alpha,1)=\Gamma(\alpha)$ و $\gamma(1,\beta)=\Gamma(\beta)$. از (۶) و (۷) و (۸) واضح است که اگر $\beta=1$ و $\alpha\in N$ ، که N مجموعه اعداد طبیعی است، آن‌گاه تابع چگالی احتمال خانواده‌ی (۸) همان تابع چگالی احتمال α امین مقدار رکورد بالایی است و اگر $\alpha=1$ و $\beta\in N$، آن‌گاه تابع چگالی احتمال خانواده‌ی (۸) همان تابع چگالی احتمال β امین مقدار رکورد پایینی است که توسط توزیع F و تابع چگالی احتمال f تولید شده است. همچنین توجه داریم که $h_F(x;1,1)=f(x)$.

لم ۸ برای هر مقدار مثبت α و β داریم

$$[\gamma(\alpha,\beta)]^2\leq\Gamma(2\alpha-1)\Gamma(2\beta-1).$$

به علاوه تساوی زیر را داریم

$$\gamma(\alpha,\beta)=\gamma(\beta,\alpha).$$

در ادامه‌ی مقاله خواص خانواده‌ی (۸) را به طور مختصر بیان می‌کنیم. ویژگی‌های تقارن و تک مدی بودن الگوی (۸) در لم‌های زیر آمده است.

لم ۹ اگر F توزیعی متقارن حول صفرباشد و H_F کلاس توزیع‌های تولید شده به فرم (۸) توسط F باشد، آن‌گاه توزیع H_F نیز حول صفر متقارن است اگر و فقط اگر $\alpha=\beta$.

لم ۱۰ اگر F متقارن و تک مد باشد، h_F نیز تک مد است با شرط آن‌که $\alpha=\beta\geq 1$.

پارامترهای α و β پارامترهای شکل هستند. اگر F متقارن باشد، h_F چوله به راست است اگر $\beta>\alpha$. میزان چولگی راست همراه با افزایش α افزایش می‌یابد. همچنین h_F چوله به چپ خواهد بود اگر $\alpha>\beta$. میزان چولگی چپ همراه باکاهش β ، افزایش می‌یابد.
هرگاه F متعلق به خانواده‌ی مکان—مقیاس باشد، آن‌گاه این ویژگی توسط الگوی (۸) حفظ می‌شود. در این ارتباط، لم زیر را داریم.

لم ۱۱ خانواده‌ی توزیع‌های معرفی شده توسط الگوی (۸) متعلق به خانواده‌ی مکان—مقیاس است اگر F متعلق به خانواده‌ی مکان—مقیاس باشد.

اگر Z دارای تابع چگالی احتمال استاندارد شده الگوی (۸) و U دارای تابع چگالی احتمال به صورت $\frac{1}{\sigma}h_{F_0}(\frac{x-\mu}{\sigma})$ باشد، بدیهی است که $E(U)=\mu+\sigma\mu_H$ و $Var(U)=\sigma^2\sigma_H^2$ که $\mu_H=E(Z)$ و $\sigma_H^2=Var(Z)$.

اکنون درصدد برمی‌آییم تا آماره‌های بسنده خانواده‌ی (۸) را بیابیم. اگر $X_1, X_2, ..., X_n$ یک نمونه‌ی تصادفی از الگوی (۸) با شاخص‌های پارامتری α و β و توزیع F باشند، آن‌گاه چنان‌چه F شامل پارامتر نباشد، داریم:

$$\begin{aligned} h_F(x_1,...,x_n,\alpha,\beta) &= \prod_{i=1}^{n} h_F(x_i,\alpha,\beta) \\ &= \frac{1}{[\gamma(\alpha,\beta)]^n}\prod_{i=1}^{n} f(x_i) \\ &\quad \times \prod_{i=1}^{n}[-\log \bar{F}(x_i)]^{\alpha-1} \\ &\quad \times \prod_{i=1}^{n}[-\log F(x_i)]^{\beta-1} \\ &= [\gamma(\alpha,\beta)]^{-n} e^{u(x_1,...,x_n)} \\ &\quad \times e^{(\alpha-1)\sum_{i=1}^{n}\log(-\log \bar{F}(x_i))} \\ &\quad \times e^{(\beta-1)\sum_{i=1}^{n}\log(-\log F(x_i))} \end{aligned} \qquad (۹)$$

که $u(x_1,...,x_n) = \log \prod_{i=1}^{n} f(x_i)$. با توجه به فرم (۹) مشاهده می‌شود که h_F متعلق به خانواده‌ی نمایی ۲ پارامتری است. فرض کنید

$$T_1(X_1,...X_n) = \sum_{i=1}^{n}\log(-\log \bar{F}(X_i))$$

و

$$T_2(X_1,...X_n) = \sum_{i=1}^{n}\log(-\log F(X_i))$$

دراین صورت

$$\begin{aligned} h_F(x_1,...,x_n,\alpha,\beta) &= [\gamma(\alpha,\beta)]^{-n} e^{u(x_1,...,x_n)} \\ &\quad \times e^{(\alpha-1)T_1(x_1,...x_n)} \\ &\quad \times e^{(\beta-1)T_2(x_1,...x_n)} \end{aligned}$$

و در نتیجه

$$T = (T_1(X_1,...,X_n), T_2(X_1,...,X_n))$$

آماره‌ی بسنده‌ی کامل برای $\theta = (\alpha,\beta)$ خواهد بود.

مراجع

[1] Alam, K. (1972), Unimodality of the distribution of an order statistic, *The Annals of Mathematical Statistics*, 43(6), 2041-2044.

[2] Amoroso, L. (1925), Ricerche intorno alla curva dei redditi. *Annali de Mathematica*, 4(2), 123-159.

[3] Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1992). *A First Course in Order Statistics*, John Wiley & Sons, New York.

[4] Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1998), *Records*, John Wiley & Sons, New York.

[5] David, H.A. and Nagaraja, H.N. (2003), *Order Statistics*, 3rd Ed., Wiley, Hoboken, NJ.

[6] Eugene, N., Lee, C. and Famoye, F. (2002), Beta-normal distribution and its applications, *Communications in Statistics - Theory and Methods*, 31, 497-512.

[7] Gupta, R.D. and Kundu, D. (1999), Generalized exponential distribution, *Australian and New Zealand Journal of Statistics*, 41, 173-188.

[8] Gupta, R.D. and Kundu, D. (2001), Exponentiated exponential family: An alternative to gamma and Weibull distributions, *Biometrical Journal*, 43, 117-130.

[9] Gupta, R.D. and Kundu, D. (2007), Generalized exponential distribution: Existing results and some recent developments, *Journal of Statistical Planning and Inference*, 137, 3537-3547.

[10] Johnson, N.L. and Kotz, S. (1970), *Continuous Univariate Distributions-2*, John Wiley & Sons, New York.

[11] Jones, M.C. (2004), Families of distributions arising from distributions of order statistics, *Test*, 13, 1-43.