مجله علوم آماری، پاییز و زمستان ۱۳۹۷
جلد ۱۲، شماره ۲، ص ۳۴۱ – ۳۵۰
DOI: 10.29252/jss.12.2.341

مشخصه‌سازی توزیع پارتو براساس مشاهدات نزدیک رکورد

معصومه اکبری لاکه و زهره صفرزاده

گروه آمار دانشگاه مازندران

تاریخ دریافت: ۱۳۹۶/۰۶/۲۱   تاریخ آخرین بازنگری: ۱۳۹۷/۰۴/۲۹

چکیده: توزیع پارتو دارای کاربردهای فراوانی در علوم اقتصادی و بیمه‌ای است. تاکنون خواص زیادی از این توزیع براساس داده‌های ترتیبی نظیر آماره‌های مرتب و رکوردها مورد بحث و بررسی قرار گرفته است. در این مقاله ضمن تعریف جدیدی از مفهوم مشاهدات نزدیک رکورد، برخی از نتایج مشخصه‌سازی توزیع پارتو براساس تعداد این مشاهدات به دست آمده است.

واژه‌های کلیدی: توزیع پارتو، رکورد، مشاهدات نزدیک رکورد، مشخصه‌سازی.

۱ مقدمه

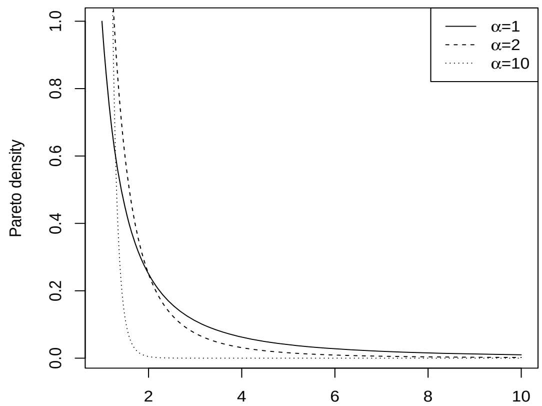
متغیر تصادفی X دارای توزیع پارتو با پارامترهای α و β است، هرگاه تابع چگالی احتمال آن به‌صورت

$$f(x) = \frac{\alpha\beta^{\alpha}}{x^{\alpha+1}} \quad x \geq \beta, \tag{۱}$$

باشد. در حقیقت توزیع پارتو، توزیعی با دم‌های سنگین و چوله است و معمولا برای توصیف توزیع پدیده‌هایی چون درآمد، جمعیت و ... به‌کار می‌رود. نمودار زیر، تابع چگالی پارتو برای $\beta = ۱$ و α های مختلف را نشان می‌دهد. جزئیات بیشتر درباره ویژگی‌ها و کاربردهای توزیع پارتو را می‌توان در آرنولد (۲۰۱۵) و فاس و همکاران (۲۰۱۳) دید. به دلیل برازش خوب این توزیع به داده‌های با دم‌سنگین، تاکنون تعمیم‌هایی از این توزیع معرفی شده است. به عنوان مثال آکینست و همکاران (۲۰۰۸) توزیع بتا – پارتو را معرفی

آدرس الکترونیکی نویسنده مسئول مقاله: معصومه اکبری لاکه، m.akbari@umz.ac.ir

کد موضوع‌بندی ریاضی (۲۰۱۰): 62E10، 62G30



شکل ۱: تابع چگالی پارتو برای $\beta = ۱$ و $\alpha = ۱, ۲, ۱۰$

کرده و نشان دادند که این توزیع دارای برازش بهتری نسبت به توزیع پارتو برای مدل‌بندی داده‌های کرانگین[۱] است. همچنین تاکنون ویژگی‌ها و خواص زیادی از توزیع پارتو براساس داده‌های ترتیبی نظیر آماره‌های مرتب و رکوردها مورد بحث و مطالعه قرار گرفته است.

فرض کنید $\{X_n : n \geq ۱\}$ دنباله‌ای از متغیرهای تصادفی پیوسته، مستقل و هم‌توزیع با تابع توزیع F و تابع چگالی f باشد. در این صورت، اگر دنباله زمان رکورد بالا به‌صورت $\{T_n^U,\ n \geq ۱\}$ نمایش داده شود، آن‌گاه

$$T_۱^U = ۱, \quad T_n^U = \min\{j :\ X_j > X_{T_{n-۱}^U}\},$$

و دنباله مقادیر رکورد بالا، $\{X^U(n),\ n \geq ۱\}$ به‌صورت $X^U(n) = X_{T_n^U}$ تعریف می‌شود. تابع چگالی nامین رکورد بالا به‌صورت

$$f_n(x) = \frac{\{-\log \bar{F}(x)\}^{n-۱}}{(n-۱)!} f(x),$$

[1]Extreme-value

است، که در آن $\bar{F}(x) = ۱ - F(x)$ تابع بقا متغیر تصادفی X است. به‌طور مشابه، می‌توان دنباله مقادیر رکورد پایین $\{X_{T_n^L}, n \geq ۱\}$ و دنباله زمان رکورد پایین $\{T_n^L, n \geq ۱\}$ را تعریف کرد (آرنولد و همکاران، ۱۹۹۸ و نزرف، ۲۰۰۱). علی‌رغم کاربرد فراوان رکوردها در بسیاری از پدیده‌های طبیعی نظیر مطالعات هیدرولوژی و هواشناسی، در مسائل مربوط به علوم اقتصادی و بیمه‌ای و همچنین آزمون‌های تنش–مقاومت نیز کاربرد فراوان دارد. با این وجود ممکن است در مواردی تعداد رکوردهای ثبت شده، اندک یا جمع‌آوری تعداد بیشتری از آن‌ها سخت و امکان‌پذیر نباشد. به‌عنوان مثال، در آزمون تنش–مقاومت وقتی که نمونه‌های مورد آزمون پرهزینه باشند، این اتفاق رخ می‌دهد. در این صورت، قابلیت اعتماد نتایج آماری به‌دست آمده مبتنی بر چنین دنباله کوتاه از رکوردها کم می‌باشد. گوت و همکاران (۲۰۱۲) نشان دادند که در این گونه موارد استفاده از مشاهدات نزدیک رکورد[۲] منجر به نتایج بهتری خواهد شد. مشاهدات نزدیک رکورد، اولین بار توسط بالاکریشنان و همکاران (۲۰۰۵) معرفی شد به این ترتیب که X_j مشاهده نزدیک nامین رکورد نامیده می‌شود، هرگاه

$$X_j \in (X^U(n) - a,\, X^U(n)), \quad T_n^U < j < T_{n+۱}^U$$

سپس براساس این چنین مشاهدات، متغیر تصادفی گسسته

$$\xi_n(a) = \#\left\{j : T_n^U < j < T_{n+۱}^U,\, X_j \in \left(X^U(n) - a,\, X^U(n)\right)\right\},$$

تعریف شد، که در آن a مقدار ثابت و مثبت است. محققینی چون پاکز (۲۰۰۷)، بوس و گانگوپادهیای (۲۰۱۱) و گوت و همکاران (۲۰۱۲) خواص حدی و مجانبی متغیر تصادفی $\xi_n(a)$ را مورد مطالعه قرار دادند. همچنین اکبری و فشندی (۲۰۱۳) کاربردی از این متغیر تصادفی را در مشخصه‌سازی و مسائل آماری نشان دادند. بالاکریشنان و همکاران (۲۰۰۵) ثابت کردند که تابع جرم احتمال متغیر تصادفی $\xi_n(a)$ به‌صورت

$$P\left(\xi_n(a) = j\right) = \int \beta(x,a)(۱ - \beta(x,a))^j dF_{X^U(n)}(x), \qquad (۲)$$

است، که در آن $\beta(x,a) = \dfrac{\bar{F}(x)}{\bar{F}(x-a)}$. با توجه به رابطه (۲)، جز برای تعداد اندکی از توزیع‌های

[2] near record observations

پیوسته نظیر توزیع‌های نمایی و نمایی دوپارامتری نمی‌توان برای سایر توزیع‌ها مانند توزیع پارتو، تابع احتمال $\xi_n(a)$ را به راحتی محاسبه کرد. برای این منظور تعریف جدید دیگری از مشاهدات نزدیک رکورد و تعداد این چنین مشاهدات به‌صورت

$$\zeta_n(a) = \#\{j : T_n^U < j < T_{n+1}^U,\ X_j \in (aX^U(n),\ X^U(n))\}, \quad 0 < a < 1 \qquad (۳)$$

ارائه می‌شود. در این مقاله برخی از نتایج مشخصه‌سازی توزیع پارتو بر اساس متغیر تصادفی (۳) به‌دست آورده می‌شود. نتایج مربوط به تابع جرم احتمال، تابع بقا و گشتاور مرتبه اول متغیر تصادفی $\zeta_n(a)$ در بخش ۲ آمده است. در بخش ۳ نتایج مشخصه‌سازی توزیع پارتو بر اساس خواص توزیعی متغیر تصادفی $\zeta_n(a)$ ارائه می‌شود. در انتها به بحث و نتیجه‌گیری پرداخته خواهد شد.

۲ برخی از نتایج توزیعی

در این بخش تابع جرم احتمال، تابع بقا و گشتاور مرتبه اول متغیر تصادفی $\zeta_n(a)$ ارائه می‌شود.

قضیه ۱: فرض کنید $\{X_n : n \geq 1\}$ دنباله‌ای از متغیرهای تصادفی پیوسته، مستقل و هم‌توزیع با تابع توزیع F باشند. آن‌گاه تابع جرم احتمال متغیر تصادفی $\zeta_n(a)$ برای هر $j = 0, 1, 2, \ldots$ برابر است با

$$P(\zeta_n(a) = j) = \int \{\frac{\bar{F}(x)}{\bar{F}(ax)}\}\{1 - \frac{\bar{F}(x)}{\bar{F}(ax)}\}^j dF_{X^U(n)}(x). \qquad (۴)$$

برهان: با استفاده از قانون احتمال کل،

$$P(\zeta_n(a) = j) = \int P(\zeta_n(a) = j | X^U(n) = x) dF_{X^U(n)}(x). \qquad (۵)$$

بنابراین کافی است $P(\zeta_n(a) = j | X^U(n) = x)$ به‌دست آورده شود. فرض کنید $X^U(n) = x$، در این صورت $T_{n+1}^U = n + i$ که $i = 1, 2, \ldots$، اگر و فقط اگر $X_{n+i} > x$، $X_{n+i-1} \leq x$، ...، $X_{n+1} \leq x$. لذا طبق (۳)، پیشامد $\{\zeta_n(a) = j | X^U(n) = x\}$ رخ می‌دهد اگر و فقط اگر $n + i$ امین مشاهده بزرگتر از x باشد و j تا از مشاهدات X_{n+1}، ...، X_{n+i-1} در بازه (ax, x) قرار گیرند

و مابقی مشاهدات یعنی $i-۱-j$ تا از مشاهدات کوچکتر از ax باشند. لذا

$$
\begin{aligned}
P(\zeta_n(a)=j|X^U(n)=x) &= \sum_{i=j+۱}^{\infty}\binom{i-۱}{j}\{P(ax<X<x)\}^j \\
&\times \{P(X<ax)\}^{i-۱-j}P(X>x) \\
&= \bar{F}(x)\{F(x)-F(ax)\}^j \\
&\times \sum_{i=j+۱}^{\infty}\binom{i-۱}{j}\{F(ax)\}^{i-۱-j} \\
&= \frac{\bar{F}(x)}{\bar{F}(ax)}\{۱-\frac{\bar{F}(x)}{\bar{F}(ax)}\}^j. \qquad (۶)
\end{aligned}
$$

با جای‌گذاری (۶) در (۵) برهان کامل می‌شود.

با توجه به تابع جرم احتمال به‌دست آمده در قضیه ۱، به‌راحتی می‌توان تابع بقای متغیر تصادفی $\zeta_n(a)$ را محاسبه کرد.

$$
\begin{aligned}
P(\zeta_n(a)\geq j) &= \sum_{r=j}^{\infty}P(\zeta_n(a)=r) \\
&= \int\{۱-\frac{\bar{F}(x)}{\bar{F}(ax)}\}^j dF_{X^U(n)}(x).
\end{aligned}
$$

از آنجا که متغیر تصادفی $\zeta_n(a)$ مثبت است، امید ریاضی آن را می‌توان به‌صورت

$$E[\zeta_n(a)]=\sum_{j=۱}^{\infty}P(\zeta_n(a)\geq j),$$

نوشت. بنابراین با استفاده از (۷)، گشتاور مرتبه اول $\zeta_n(a)$ به‌صورت

$$E[\zeta_n(a)]=\int\frac{\bar{F}(ax)-\bar{F}(x)}{\bar{F}(x)}dF_{X^U(n)}(x), \qquad (۷)$$

به‌دست می‌آید.

۳ مشخصه‌سازی توزیع پارتو

تاکنون نتایج متعددی از مشخصه‌سازی توزیع پارتو به‌دست آمده است. به عنوان مثال افیفی (۲۰۰۶) مشخصه‌سازی از آمیخته دو توزیع پارتو بر اساس گشتاور بریده‌شده از چپ ارائه کرده است. احسن‌الله و شکیل (۲۰۱۲) توانستند توزیع پارتو را براساس مقادیر رکورد بالا مشخصه‌سازی کنند. همچنین احسن‌الله و همکاران (۲۰۱۶) یک قاعده کلی برای مشخصه‌سازی توزیع‌های پیوسته استاندارد نظیر توزیع پارتو ارائه نمودند. آنها نشان دادند که تساوی گشتاور مرتبه اول بریده شده از چپ یا راست با ضریبی از تابع نرخ خطر آن برای مشخصه‌سازی توزیع پارتو شرط لازم و کافی است. نتایجی از مشخصه‌سازی توزیع پارتو بر اساس گشتاورهای شرطی آماره‌های مرتب توسط نوفال و ال جبلی (۲۰۱۷) ارائه شده است. در این بخش برخی از نتایج مشخصه‌سازی توزیع پارتو براساس مشاهدات نزدیک رکورد با استفاده از خواص دنباله توابع کامل بیان و اثبات می‌شود. لذا ابتدا برخی مفاهیم اولیه مورد نیاز دنباله توابع کامل بیان می‌شوند.

تعریف ۱: دنباله $\{\phi_n,\ n \geq ۱\}$ در فضای هیلبرت $\mathcal{H}$ کامل نامیده می‌شود اگر و فقط اگر تنها عضو $\mathcal{H}$ که با ϕ_n متعامد است، عضو خنثی آن باشد. به‌عبارت دیگر،

$$n \in \mathbb{N};\ \langle f, \phi_n \rangle = ۰ \Longleftrightarrow f = ۰ \in \mathcal{H},$$

که در آن نماد $\langle \cdot, \cdot \rangle$، ضرب داخلی فضای هیلبرت $\mathcal{H}$ را نشان می‌دهد.

در این مقاله فضای هیلبرت $L^۲[۰, ۱]$ در نظر گرفته شده و ضرب داخلی آن به‌صورت

$$\langle f, g \rangle = \int_۰^۱ f(x)g(x)dx,$$

تعریف می‌شود، که در آن f و g توابع حقیقی مقدار و انتگرال‌پذیر از مرتبه دوم روی بازه $[۰, ۱]$ است. یکی از دنباله‌های توابع کامل معروف در فضای $L^۲[۰, ۱]$، دنباله $\{x^n,\ n \geq ۱\}$ است که در این مقاله مورد استفاده قرار گرفته است. قضیه زیر که به قضیه مونتز [3] معروف است شرط لازم و کافی برای کامل بودن زیر دنباله فوق را بیان می‌کند.

قضیه ۲: (هیگنز، ۲۰۰۴) دنباله $\{x^{n_۱}, x^{n_۲}, \ldots, ۱ \leq n_۱ \leq n_۲ < \ldots\}$ در فضای $L^۲[۰, ۱]$ کامل است اگر و فقط اگر $\sum_{j=۱}^{\infty} n_j^{-۱} = \infty$.

[3] Múntz

لم ۱: اگر X یک متغیر تصادفی پیوسته با تکیه‌گاه (β,∞) دارای تابع توزیع F باشد، آن‌گاه X دارای توزیع پارتو با پارامترهای (α,β) است اگر و فقط اگر برای هر $۰ < \alpha < ۱$، داشته باشیم

$$\bar{F}(aF^{-۱}(۱-e^{-y})) = e^{-y}a^{-\alpha} \tag{۸}$$

برهان: اگر X دارای توزیع پارتو باشد، آن‌گاه رابطه (۸) به راحتی به‌دست می‌آید. حال فرض کنید رابطه (۸) برقرار باشد، آن‌گاه با در نظر گرفتن تغییر متغیر $t = F^{-۱}(۱-e^{-y})$، نتیجه می‌شود

$$\bar{F}(at) = \bar{F}(t)a^{-\alpha} \tag{۹}$$

تابع $\bar{F}(x) = cx^{-\alpha}$، جواب کلی معادله تابعی (۹) است که در آن c ثابت مثبت است. بنابراین اثبات کامل می‌شود.

قضیه ۳: فرض کنید $\{X_n : n \geq ۱\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با تابع توزیع پیوسته F با تکیه‌گاه، $x \geq \beta$ باشد. در این صورت F، تابع توزیع متغیر تصادفی پارتو با پارامترهای (α,β) است، اگر و فقط اگر برای هر $n \geq ۱$ و $j = ۰,۱,۲,\ldots$ یکی از شرایط زیر برقرار باشد.

(الف) $P(\zeta_n(a) = j) = a^{\alpha}(۱-a^{\alpha})^j$

(ب) $P(\zeta_n(a) \geq j) = a^{\alpha j}$

(ج) $E[\zeta_n(a)] = (a)^{-\alpha} - ۱$

برهان: (الف) فرض کنید برای هر $j = ۰,۱,۲,\ldots$ داشته باشیم

$$P(\zeta_n(a) = j) = a^{\alpha}(۱-a^{\alpha})^j \tag{۱۰}$$

با توجه به تابع جرم احتمال متغیر تصادفی $\zeta_n(a)$، رابطه (۱۰) به‌صورت

$$\int_{\beta}^{\infty} \frac{\bar{F}(x)}{\bar{F}(ax)}(۱-\frac{\bar{F}(x)}{\bar{F}(ax)})^j \frac{\{-\log \bar{F}(x)\}^{n-۱}}{(n-۱)!} f(x)dx = a^{\alpha}(۱-a^{\alpha})^j,$$

قابل بیان است، که با تغییر متغیر $y=-\log \bar{F}(x)$، به‌صورت

$$\int_{\circ}^{\infty} \frac{e^{-y}}{\bar{F}(aF^{-1}(1-e^{-y}))}(1-\frac{e^{-y}}{\bar{F}(aF^{-1}(1-e^{-y}))})^{j}\frac{y^{n-1}}{(n-1)!}dy=a^{\alpha}(1-a^{\alpha})^{j}, \tag{11}$$

قابل بازنویسی است. از طرفی سمت راست رابطه (۱۱) را می‌توان به‌صورت

$$a^{\alpha}(1-a^{\alpha})^{j}\int_{\circ}^{\infty}\frac{y^{n-1}}{(n-1)!}e^{-y}dy \tag{12}$$

در نظر گرفت. با استفاده از این حقیقت، رابطه (۱۱) پس از ساده شدن برابر است با

$$\int_{\circ}^{\infty}\{\frac{e^{-y}}{\bar{F}(aF^{-1}(1-e^{-y}))}(1-\frac{e^{-y}}{\bar{F}(aF^{-1}(1-e^{-y}))})^{j}-a^{\alpha}(1-a^{\alpha})^{j}\}y^{n-1}e^{-y}dy=\circ$$

با توجه به کامل بودن دنباله تابعی $\{y^{n-1},\, n\geq 1\}$، نتیجه

$$\frac{e^{-y}}{\bar{F}(aF^{-1}(1-e^{-y}))}(1-\frac{e^{-y}}{\bar{F}(aF^{-1}(1-e^{-y}))})^{j}-a^{\alpha}(1-a^{\alpha})^{j}=\circ, \tag{13}$$

برای هر $j=\circ,1,2,\ldots$ به‌دست می‌آید. با قرار دادن $j=\circ$ در رابطه (۱۳)، داریم

$$\frac{e^{-y}}{\bar{F}(aF^{-1}(1-e^{-y}))}=a^{\alpha}$$

بر اساس لم ۱، اثبات قسمت (الف) کامل می‌شود.
(ب) فرض کنید $P(\zeta_n(a)\geq j)=a^{\alpha j}$، در این صورت با استفاده از رابطه (۷) نتیجه می‌شود.

$$\int_{\beta}^{\infty}\{1-\frac{\bar{F}(x)}{\bar{F}(ax)}\}^{j}\frac{\{-\log\bar{F}(x)\}^{n-1}}{(n-1)!}f(x)dx=a^{\alpha j}$$

ادامه برهان (ب) مانند قسمت قبل است. همچنین قسمت (ج) نیز به طور مشابه ثابت می‌شود.
برای اثبات شرط کافی، اگر X دارای توزیع پارتو با پارامترهای (α,β) باشد، آن‌گاه با استفاده از روابط (۴)، (۷) و (۷) به ترتیب نتایج قسمت‌های (الف)، (ب) و (ج) به راحتی به‌دست می‌آیند.

تذکر ۱: براساس قضیه ۲، نتایج قضیه ۳ برای هر دنباله صعودی و مثبت از n_j که $\sum_{j=1}^{\infty} n_j^{-1} = \infty$ ، برقرار است.

۴ بحث و نتیجه‌گیری

توزیع پارتو یکی از مهم‌ترین توزیع‌ها در مسائل اقتصادی و بیمه‌ای است. لذا ارائه راهکارهایی جهت مدل‌بندی آن، می‌تواند یکی از اصلی‌ترین موضوعات باشد. آزمون‌هایی چون کای-دو یا کلموگروف اسمیرنوف می‌تواند در این گونه موارد مورد استفاده قرار گیرد. اما اخیرا محققینی چون نیک تین (۲۰۱۷) آزمون‌های نکویی برازش مبتنی بر مشخصه‌سازی را یک روش موثر برای مدل‌بندی توزیع‌ها دانسته‌اند. لذا امیدواریم که نتایج مشخصه‌سازی ارائه شده در این مقاله، بتواند به عنوان یک ابزار مناسب برای مدل‌سازی توزیع پارتو مورد استفاده قرار گیرد.

۵ تقدیر و تشکر

نویسندگان مقاله از سردبیر، داوران و ویراستار محترم مجله که با رهنمودهای ارزنده خود باعث بهتر شدن مقاله گردیده‌اند، کمال تشکر و قدردانی دارند.

مراجع

Aczél, J., (1966), *Lectures on Functional Equations and Their Applications*, Academic Press, New York and London.

Afify, E., E., (2006), Characterization of Mixtures from Exponential and Pareto Distributions Using Left Truncated Moments, *International Journal of Statistical Sciences*, **5**, 1-18.

Ahsanullah, M. and Shakil, M., (2012), A Note on the Characterization of Pareto Distribution by Upper Record Values, *Communication of the Korean Mathematical Society*, **27**, 835-842.

Ahsanullah, M., Shakil, M. and Golam Kibria, B. M., (2016), Characterization of Continuous Distributions by Truncated Moment, *Journal of Modern Applied Statistical Methods*, **15**, 316-331.

Akbari, M. and Fashandi, M., (2013), On Characterization Results Based on the Number of Observations Near the k-Records, *Statistics*, **48**, 633-640.

Akinset, A., Famoye, F. and Lee, C., (2008), The Beta-Pareto Distribution, *Statistics*, **42**, 547-563.

Arnold, B., C., (2015), *Pareto Distributions. CRC press, Boca Raton*, London and New York.

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N., (1998), *Records*, John Wiley, New York.

Balakrishnan, N., Pakes, A. and Stepanov, A., (2005), On the Number and Sum of Near Record Observations, *Advances in Applied Probability*, **37**, 765-780.

Bose, A. and Gangopadhyay, (2011), Asymptotic Properties of Near Pfeifer Records, *Extremes*, **14**, 235-265.

Foss, S., Korshunov, D. and Zachary, S., (2013), *An Introduction to Heavy-Tailed Subexponential Distributions*, Springer, United States.

Gouet, R., López, F. J. and Sanz, G., (2012), On δ -Record Observations: Asymptotic Rates for the Counting Process and Elements of Maximum Likelihood Estimation, *Test*, **21**, 188-214.

Higgins, J. R., (2004), *Completeness and Basis Properties of Sets of Special Functions*, Cambridge University Press, New York.

Nevzorov, V. B., (2001), *Records: Mathematical Theory*, American Mathematical Society, Providence, RI, Transl. Math. Monogr ,**194** .

Nikitin, Ya. Yu., (2017), Test based on Characterizations, and Their Efficiencies: A survey, *Cta et Commrntations Universitatis Tartuensis De Mathematica*, **21**, 3-24.

Nofal, Z. M. and El Gebaly, Y. M., (2017), New Characterization of the Pareto Distribution, *Pakistan Journal of Statistical and Operation Research*, **13**, 63-74.

Pakes, A. G., (2007), Limit Theorems for Numbers of Near Records, *Extremes*, **10**, 207-224.