مجله علوم آماری، پاییز و زمستان ۱۳۹۱
جلد ۶، شماره ۲، ص ۲۰۱–۲۱۸

برآورد بیزی پارامتر توزیع پارتو تحت توابع زیان توان دوم خطا و لاینکس بر اساس نمونه‌گیری مجموعه رتبه‌دار

رضا علیزاده نوقابی، جعفر احمدی
گروه آمار، دانشگاه فردوسی مشهد

تاریخ دریافت: ۱۳۹۱/۸/۳۰ تاریخ آخرین بازنگری: ۱۳۹۱/۱۱/۴

چکیده : در برخی از مسئله‌های کاربردی، به‌دست آوردن مشاهدات اغلب زمان‌بر و با صرف هزینه همراه می‌باشد. در چنین شرایطی استفاده از یک روش نمونه‌گیری که موجب کاهش هزینه‌ها و افزایش کارآیی برآوردگرها شود، حائز اهمیت است. در این‌گونه موارد روش نمونه‌گیری مجموعه رتبه‌دار جایگزین مناسبی به جای روش نمونه‌گیری تصادفی ساده به نظر می‌رسد. در این مقاله برآورد بیزی پارامتر توزیع پارتو تحت توابع زیان توان دوم خطا و لاینکس، زمانی که داده‌ها به شیوه نمونه‌گیری مجموعه رتبه‌دار جمع‌آوری شده باشد، مورد بررسی قرار گرفته است. به کمک شبیه‌سازی مونت کارلو مخاطره بیزی برآوردگرها محاسبه و در دو روش نمونه‌گیری تصادفی ساده و نمونه‌گیری مجموعه رتبه‌دار هنگامی که از توزیع مزدوج و جفریز به عنوان توزیع پیشین استفاده شود، مقایسه شده‌اند. در انتها با ارائه مثال واقعی برآوردگرهای به‌دست آمده ارزیابی شده‌اند. نتایج به‌دست آمده حاکی از برتری روش نمونه‌گیری مجموعه رتبه‌دار است، بنابراین توصیه می‌شود اگر شرایط برای انجام نمونه‌گیری به روش مجموعه رتبه‌دار فراهم باشد، از این طرح نمونه‌گیری استفاده شود.

° آدرس الکترونیک مسئول مقاله: reza.alizadehn@yahoo.com
° کد موضوع‌بندی ریاضی (۲۰۰۰): ۶۲F۱۵

واژه‌های کلیدی : نمونه‌گیری مجموعه رتبه‌دار، برآورد بیزی، تابع زیان لاینکس، تابع زیان توان دوم خطا، معیار نزدیکی پیتمن.

۱ مقدمه

نمونه‌گیری مجموعه رتبه‌دار اولین بار توسط مک‌اینتایر (۱۹۵۲) معرفی شد. در این شیوه ابتدا به تصادف n^2 واحد نمونه از جامعه مورد بررسی را انتخاب نموده سپس آن‌ها را به طور تصادفی به n زیر نمونه به اندازه n تقسیم کرده و واحدهای نمونه را بر اساس روشی که نیاز به اندازه‌گیری دقیق نباشد به ترتیب صعودی مرتب می‌کنند. واحد دارای کوچکترین رتبه در اولین زیر نمونه و واحد دارای دومین رتبه در زیر نمونه دوم و به همین ترتیب واحد دارنده بزرگترین رتبه در زیر نمونه آخر انتخاب می‌شوند. برای اطلاعات بیشتر در این زمینه به چن و همکاران (۲۰۰۳) مراجعه شود. این روش نمونه‌گیری در بسیاری از مسائل به کار برده می‌شود و علاوه بر افزایش دقت برآوردگرها در مواردی منجر به کاهش هزینه نیز می‌گردد. به عنوان مثال در مسائل زیست محیطی برای برآورد مقدار آلاینده در اطراف یک نیروگاه هسته‌ای بررسی می‌شود که آیا آلایندگی خاک اطراف نیروگاه در سطح نرمال است یا خیر. برای این کار ابتدا نمونه‌هایی انتخاب می‌شوند، سپس با انجام آزمایش‌های بسیار پرهزینه مقدار متغیرهای مورد نظر اندازه‌گیری می‌گردند. روشی مناسب برای این منظور نمونه‌گیری مجموعه رتبه‌دار است. در واقع ممکن است به‌صورت بصری و بر اساس ویژگی‌های فیزیکی مانند رنگ، بافت و بو قبل از تحلیل دقیق بتوان مقدار تقریبی متغیر مورد بررسی در خاک را ارائه داد. نمونه‌هایی که به روش نمونه‌گیری مجموعه رتبه‌دار گزینش شده‌اند، برای اندازه‌گیری دقیق به آزمایشگاه ارسال می‌شوند. برای مثال‌های بیشتر می‌توان به ناسبوآم و سینها (۱۹۹۷)، یو و لم (۱۹۹۷)، مد و همکاران (۱۹۹۹) و پاتیل و همکاران (۱۹۹۹) مراجعه کرد. اگر رتبه‌بندی داده‌ها آسان‌تر از اندازه‌گیری دقیق آن‌ها باشد، استفاده از روش نمونه‌گیری مجموعه رتبه‌دار مفید است. برای مثال بیشتر در منابع فارسی می‌توان به زمانزاده و احمدی (۱۳۹۰) و امینی و ارقامی (۱۳۸۷) مراجعه نمود.

بسیاری از پژوهش‌گران مسئله برآورد پارامترهای برخی از توزیع‌های آماری را بر اساس نمونه‌گیری مجموعه رتبه‌دار مطالعه نموده‌اند. در این میان سینها و همکاران (۱۹۹۶) و شیبو و موتلک (۲۰۰۴) بهترین برآوردگر نااریب خطی برای توزیع‌های نرمال، نمایی و گاما را برآورد کرده‌اند. بیشتر مقاله‌های موجود در مورد برآورد پارامترها به‌صورت کلاسیک می‌باشد و مقالات بسیار کمی به استنباط بیزی زمانی که داده‌ها به شیوه نمونه‌گیری مجموعه رتبه‌دار استخراج شده باشند، پرداخته‌اند که از جمله می‌توان به تحقیقات انجام شده توسط الصالح و همکاران (۲۰۰۰) و سادک و همکاران (۲۰۱۳) اشاره کرد که توزیع آماری تحت مطالعه را نیز نمایی در نظر گرفته‌اند.

در این مقاله با استفاده از تابع زیان لاینکس[۱] پارامتر توزیع پارتو بر مبنای داده‌های نمونه‌گیری مجموعه رتبه‌دار برآورد می‌شود. در بخش‌های ۳ و ۴ به ترتیب برآوردگر بیزی توزیع پارتو بر اساس توزیع‌های پیشین گاما و جفریز به‌دست آورده می‌شوند. همچنین برآوردگر درستنمایی ماکسیمم محاسبه شده و با حالت توزیع پیشین جفریز مقایسه می‌شود. در بخش ۵ نتایج مقایسه مخاطره بیزی برآوردگرها به کمک شبیه‌سازی آورده شده است. برای مقایسه برآوردگرهای بیزی تحت دو تابع زیان متفاوت از معیار نزدیکی پیتمن (پیتمن، ۱۹۳۷) استفاده شده است. اگر $\delta_۱$ و $\delta_۲$ دو برآوردگر پارامتر θ باشند آنگاه نزدیکی پیتمن $\delta_۱$ نسبت به $\delta_۲$ در خصوص θ به‌صورت

$$PMC(\delta_۱,\delta_۲|\theta)=P\left\{|\delta_۱-\theta|<|\delta_۲-\theta|\right\},\ \ \theta\in\Theta$$

تعریف می‌شود، که در آن Θ فضای پارامتر است. بنا به تعریف برآوردگر $\delta_۱$ نسبت به $\delta_۲$ پیتمن نزدیک‌تر به θ است اگر $PMC(\delta_۱,\delta_۲|\theta)\geq ۰/۵$ به‌ازای هر $\theta\in\Theta$ و نامساوی اکید حداقل برای یک $\theta\in\Theta$ برقرار باشد. برای جزئیات بیشتر در خصوص معیار پیتمن، مطالعه تک‌نگاشت کیتینگ و همکاران (۱۹۹۳) به علاقه‌مندان توصیه می‌شود. در بخش ۶ بر اساس داده‌های میزان خسارات باد مربوط به سال ۱۹۷۷ برگرفته از هوگ و کلاگمن (۱۹۸۴) کارآیی برآوردگرهای به‌دست آمده مورد

[۱] Linex

بررسی قرار می‌گیرد.

۲ مفاهیم اولیه

متغیر تصادفی X دارای توزیع پارتو با پارامتر شکل α و پارامتر مقیاس β است، هرگاه تابع چگالی آن به‌صورت

$$f_\alpha(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}};\ \ x \geq \beta,\ \alpha > 0, \beta > 0.$$

باشد. بسیاری از پدیده‌های اجتماعی، علمی و ژئوفیزیکی توسط توزیع پارتو توصیف می‌شوند. این توزیع در مباحث طول عمر و قابلیت اعتماد دارای اهمیت فراوان است. یکی از کاربردهای این توزیع در مباحث اقتصاد است که غالبا برای توصیف توزیع درآمد بین افراد جامعه استفاده می‌شود. برآورد پارامتر توزیع پارتو تحت نمونه‌گیری مجموعه رتبه‌دار به روش‌های کلاسیک بررسی شده است که از جمله می‌توان به ابودیاح و همکاران (۲۰۱۱) اشاره کرد. در این مقاله برآوردگرهای بیزی برای پارامتر α بر اساس نمونه‌گیری مجموعه رتبه‌دار ارائه می‌شود. در ادامه بدون از دست دادن کلیت مسئله مقدار β معلوم و برابر یک فرض شده است. توزیع پیشین مزدوج برای α به‌صورت

$$\pi_1(\alpha) = \frac{\theta^k}{\Gamma(k)}\alpha^{k-1}e^{-\alpha\theta};\quad \alpha > 0. \tag{1}$$

است که توزیع گاما با پارامترهای k و θ است. همچنین توزیع پیشین جفریز به منظور امکان مقایسه با برآوردگر کلاسیک درستنمایی ماکسیمم استفاده شده است که

$$\pi_2(\alpha) \propto \frac{1}{\alpha};\quad \alpha > 0. \tag{2}$$

از توابع زیان متقارن توان دوم خطا و نامتقارن لاینکس برای برآورد α استفاده خواهد شد که به ترتیب به‌صورت

$$\begin{aligned} L_1(\theta,\delta) &= (\theta-\delta)^2, \\ L_2(\theta,\delta) &= b\left\{\exp(c(\theta-\delta)) - c(\theta-\delta) - 1\right\},\ c \neq 0\ b > 0. \end{aligned}$$

هستند، که در آن b و c ثابت هستند و c شکل تابع زیان را مشخص می‌سازد و b برای مقیاس به‌کار برده می‌شود. به ازای $c = 1$، تابع زیان حول صفر نامتقارن و باعث بیش‌برآورد می‌شود. به‌طور کلی اگر $c > 0$، تابع زیان لاینکس به‌صورت خطی برای مقادیر منفی خطا و نمایی برای مقادیر مثبت خطا افزایش می‌یابد. برای مقادیر مثبت و کوچک c یعنی c^j و $j \geq 3$ تابع زیان متقارن و نزدیک به تابع زیان توان دوم خطا است. برای $c = -1$ نیز تابع زیان لاینکس نامتقارن است و باعث کم‌برآوردی می‌شود. اگر $c < 0$ باشد، تابع زیان لاینکس برای مقادیر منفی خطا به‌صورت نمایی و برای مقادیر مثبت خطا به‌صورت خطی زیاد می‌شود. برای اطلاعات بیشتر در مورد تابع زیان لاینکس به چانگ و هانگ (۲۰۰۷) و برای نحوه انتخاب توابع زیان به برگر (۱۹۸۵) مراجعه شود.

۳ برآوردگر بیزی

در این بخش برآوردگر بیزی پارامتر α تحت توابع زیان توان دوم خطا و لاینکس با استفاده از شیوه نمونه‌گیری تصادفی ساده و نمونه‌گیری مجموعه رتبه‌دار به‌دست آورده می‌شود. در ادامه از $\pi_{SRS}(\alpha|x)$ و $\pi_{RSS}(\alpha|y)$ به ترتیب برای نمایش توابع چگالی پسین نمونه‌گیری تصادفی ساده و نمونه‌گیری مجموعه رتبه‌دار استفاده می‌شود. همچنین نمادهای $\hat{\alpha}_S(SRS)$، $\hat{\alpha}_L(SRS)$، $\hat{\alpha}_S(RSS)$ و $\hat{\alpha}_L(RSS)$ به ترتیب برای برآوردگرهای بیزی α تحت توابع زیان توان دوم خطا و لاینکس بر اساس نمونه‌گیری تصادفی ساده و رتبه‌دار استفاده خواهند شد.

۱.۳ برآوردگر بیزی بر اساس نمونه گیری به شیوه تصادفی ساده

فرض کنید $X_1, \ldots, X_n$ نمونه‌ای تصادفی از توزیع پارتو با پارامتر α و توزیع پیشین مزدوج $\pi_1(\alpha)$ در رابطه (۱) باشد. توزیع پسین را می‌توان به‌صورت

$$\pi_{SRS}(\alpha|\boldsymbol{x}) = \frac{1}{\Gamma(n+k)}\Big(\sum_{i=1}^{n}\log(x_i)+\theta\Big)^{n+k}\alpha^{n+k-1}e^{-\alpha(\sum_{i=1}^{n}\log(x_i)+\theta)}. \qquad (۳)$$

نوشت، که در آن $\boldsymbol{x} = (x_۱, \ldots, x_n)$ بردار مشاهدات است.

برآوردگر بیزی تحت تابع زیان توان دوم خطا، میانگین توزیع پسین، به‌صورت

$$\hat{\alpha}_S(SRS) = \frac{n+k}{\sum_{i=۱}^{n} \log(X_i) + \theta}. \qquad (۴)$$

است. واریان (۱۹۷۵) نشان داد که برآوردگر بیزی تحت تابع زیان لاینکس به‌صورت

$$\hat{\alpha}_L(SRS) = \frac{-۱}{c} \log(E(e^{-c\alpha}|X)). \qquad (۵)$$

است. با استفاده از رابطه (۳) داریم

$$\begin{aligned} E(e^{-c\alpha}|\boldsymbol{x}) &= A \int_{\circ}^{\infty} \frac{e^{-c\alpha}}{\Gamma(n+k)} (\sum_{i=۱}^{n} \log(x_i) + \theta)^{n+k} \alpha^{n+k-۱} e^{-\alpha(\sum_{i=۱}^{n} \log(x_i) + \theta)} d\alpha \\ &= \frac{(\sum_{i=۱}^{n} \log(x_i) + \theta)^{n+k}}{\Gamma(n+k)} \int_{\circ}^{\infty} \alpha^{n+k-۱} e^{-\alpha(\sum_{i=۱}^{n} \log(x_i) + \theta + c)} d\alpha \qquad (۶) \\ &= (۱ + \frac{c}{\sum_{i=۱}^{n} \log(x_i) + \theta})^{-(n+k)}. \end{aligned}$$

با توجه به (۵) نتیجه می‌شود که

$$\hat{\alpha}_L(SRS) = \frac{n+k}{c} \log\ (۱ + \frac{c}{\sum_{i=۱}^{n} \log(X_i) + \theta}). \qquad (۷)$$

از (۴) و (۷) به‌سادگی می‌توان نشان داد که اگر $c \to \circ$ آنگاه $\hat{\alpha}_L(SRS) \to \hat{\alpha}_S(SRS)$

۲.۳ برآوردگر بیزی بر اساس نمونه‌گیری مجموعه رتبه‌دار

فرض کنید $Y_1,\dots,Y_n$ نمونه‌های یک چرخه بر اساس نمونه‌گیری مجموعه رتبه‌دار از توزیع پارتو باشند. تابع چگالی jامین آماره مرتب به‌صورت

$$\begin{aligned} h_\alpha(y_j) &= j\tbinom{n}{j} f_\alpha(y_j)[F_\alpha(y_j)]^{j-1}[1-F_\alpha(y_j)]^{n-j} \\ &= j\tbinom{n}{j}\tfrac{\alpha}{(1+y_j)}(1-(\tfrac{1}{1+y_j})^{\alpha})^{j-1}(\tfrac{1}{1+y_j})^{\alpha(n-j-1)} \\ &= \sum_{t=0}^{j-1} C_t(j)g_t(y_j),\ \ y_j>0, \end{aligned}$$

است، که در آن

$$\begin{aligned} C_t(j) &= j\binom{n}{j}\binom{j-1}{t}(-1)^t, \\ g_t(y_j) &= \alpha\exp(-\alpha(1+n-j+t)\log(1+y_j))\exp(-\log(1+y_j)). \end{aligned}$$

بنابراین تابع درستنمایی نمونه‌هایی که به روش نمونه‌گیری مجموعه رتبه‌دار استخراج شده است، به‌صورت

$$\begin{aligned} L_{RSS}(\alpha) &= \prod_{j=1}^{n} h_\alpha(y_j) \\ &= \prod_{j=1}^{n}\sum_{t=0}^{j-1} C_t(j)g_t(y_j) \qquad (8) \\ &= \sum_{i_1=0}^{0}\sum_{i_2=0}^{1}\dots\sum_{i_n=0}^{n-1}[\prod_{j=1}^{n} C_{i_j}(j)]\alpha^n\exp(-\alpha A_{\boldsymbol{i}}(\boldsymbol{y}))e^{(-\sum_{j=1}^{n}\log(1+y_j))}, \end{aligned}$$

حاصل می‌شود، که در آن $\boldsymbol{y}=(y_1,\dots,y_n)$، $\boldsymbol{i}=(i_1,\dots,i_n)$ و

$$A_{\boldsymbol{i}}(\boldsymbol{y})=\sum_{j=1}^{n}(1+n-j+i_j)\log(1+y_j),$$

با استفاده از (۸) تابع چگالی پسین را می‌توان به‌صورت

$$\pi_{RSS}(\alpha|\boldsymbol{y})=\frac{\sum_{i_1=0}^{0}\sum_{i_2=0}^{1}\dots\sum_{i_n=0}^{n-1}[\prod_{j=1}^{n} C_{i_j}(j)]\alpha^{n+k-1}\exp(-\alpha(A_{\boldsymbol{i}}(\boldsymbol{y})+\theta))}{\sum_{i_1=0}^{0}\sum_{i_2=0}^{1}\dots\sum_{i_n=0}^{n-1}[\prod_{j=1}^{n} C_{i_j}(j)]\Gamma(n+k)(A_{\boldsymbol{i}}(\boldsymbol{y})+\theta)^{-(n+k)}}. \qquad (9)$$

نوشت. از رابطه (۹) برآورد بیزی پارامتر α تحت تابع زیان توان دوم خطا به‌صورت

$$\hat{\alpha}_S(RSS) = \frac{\sum\limits_{i_1=0}^{0} \sum\limits_{i_2=0}^{1} \dots \sum\limits_{i_n=0}^{n-1} [\prod\limits_{j=1}^{n} C_{i_j}(j)](n+k)(A_{\boldsymbol{i}}(\boldsymbol{Y})+\theta)^{-(n+k+1)}}{\sum\limits_{i_1=0}^{0} \sum\limits_{i_2=0}^{1} \dots \sum\limits_{i_n=0}^{n-1} [\prod\limits_{j=1}^{n} C_{i_j}(j)]\Gamma(n+k)(A_{\boldsymbol{i}}(\boldsymbol{Y})+\theta)^{-(n+k)}}$$

به‌دست می‌آید. به کمک رابطه (۵) برآورد بیزی پارامتر α تحت تابع زیان لاینکس به‌دست آورده می‌شود، در این صورت داریم

$$E(e^{-c\alpha}|\boldsymbol{y}) = \frac{\sum\limits_{i_1=0}^{0} \sum\limits_{i_2=0}^{1} \dots \sum\limits_{i_n=0}^{n-1} [\prod\limits_{j=1}^{n} C_{i_j}(j)](A_{\boldsymbol{i}}(\boldsymbol{y})+\theta+c)^{-(n+k)}}{\sum\limits_{i_1=0}^{0} \sum\limits_{i_2=0}^{1} \dots \sum\limits_{i_n=0}^{n-1} [\prod\limits_{j=1}^{n} C_{i_j}(j)](A_{\boldsymbol{i}}(\boldsymbol{y})+\theta)^{-(n+k)}}.$$

بنابراین

$$\hat{\alpha}_L(RSS) = \frac{-1}{c}\log[E(e^{-c\alpha}|\boldsymbol{Y})].$$

برای ریسک برآوردگرها صورت تابعی بسته‌ای به‌دست نیامده اما به‌وسیله شبیه‌سازی انجام شده و نتایج حاصل در بخش ۵، جدول ۲ آمده است.

۳.۳ برآوردگر بیزی بر اساس توزیع پیشین ناآگاهی بخش

فرض کنید α دارای توزیع پیشین ناآگاهی بخش (۲) باشد. برآوردگرهای بیزی در حالت نمونه‌گیری تصادفی ساده مشابه بخش ۱.۳ به‌صورت

$$\hat{\alpha}_S^J(SRS) = \frac{n}{\sum\limits_{j=1}^{n}\log(X_j)},$$
$$\hat{\alpha}_L^J(SRS) = \frac{n}{c}\log\left(1+\frac{c}{\sum\limits_{j=1}^{n}\log(X_j)}\right),$$

به‌دست می‌آیند، که در آن‌ها اندیس بالای J برای توزیع پیشین جفریز استفاده شده است. در اینجا نیز اگر $c \to 0$ ملاحظه می‌شود که $\hat{\alpha}_L^j(SRS) \to \hat{\alpha}_S^j(SRS)$. در

طرح نمونه‌گیری مجموعه رتبه‌دار بعد از محاسبه توزیع پسین نتایج زیر تحت دو تابع زیان توان دوم خطا و لاینکس حاصل می‌شود.

$$\hat{\alpha}_S^J(RSS) = n\frac{\sum\limits_{i_1=0}^{0}\sum\limits_{i_2=0}^{1}\cdots\sum\limits_{i_n=0}^{n-1}[\prod\limits_{j=1}^{n} C_{i_j}(j)](A_{\boldsymbol{i}}(\mathbf{Y}))^{-(n+1)}}{\sum\limits_{i_1=0}^{0}\sum\limits_{i_2=0}^{1}\cdots\sum\limits_{i_n=0}^{n-1}[\prod\limits_{j=1}^{n} C_{i_j}(j)](A_{\boldsymbol{i}}(\mathbf{Y}))^{-n}},$$

$$\hat{\alpha}_L^J(RSS) = \frac{-1}{c}\log[\frac{\sum\limits_{i_1=0}^{0}\sum\limits_{i_2=0}^{1}\cdots\sum\limits_{i_n=0}^{n-1}[\prod\limits_{j=1}^{n} C_{i_j}(j)]}{\sum\limits_{i_1=0}^{0}\sum\limits_{i_2=0}^{1}\cdots\sum\limits_{i_n=0}^{n-1}[\prod\limits_{j=1}^{n} C_{i_j}(j)]}(1+\frac{c}{A_{\boldsymbol{i}}(\mathbf{Y})})^{-n}].$$

لازم به ذکر است اگر نمونه‌گیری بر مبنای m چرخه انجام شود، نتایج مشابهی به‌دست می‌آید.

۴ برآوردگر درستنمایی ماکسیمم بر اساس نمونه‌گیری مجموعه رتبه‌دار

فرض کنید $X_1,\ldots,X_n$ نمونه‌ای تصادفی باشد که بر اساس نمونه‌گیری مجموعه رتبه‌دار از توزیع پارتو استخراج شده باشد، داریم

$$L_{RSS}(\alpha) = \prod_{j=1}^{n}[j\binom{n}{j}]\alpha^n\prod_{j=1}^{n}[1-\frac{1}{x_j^{\alpha}}]^{j-1}\prod_{j=1}^{n}[\frac{1}{x_j}]^{\alpha(n-j)}\prod_{j=1}^{n}[\frac{1}{x_j}]^{\alpha+1}.$$

بنابراین

$$\frac{\partial\log(L_{RSS}(\alpha))}{\partial\alpha} = \frac{n}{\alpha}+\sum_{j=1}^{n}[(j-1)\log(x_j)(\frac{1}{1-\frac{1}{{x_j}^{\alpha}}})-n\log(x_j)] = 0. \quad (10)$$

از طرفی $\frac{\partial^2\log(L_{RSS}(\alpha))}{\partial\alpha^2} < 0$، در نتیجه ریشه معادله (۱۰) برآوردگر درستنمایی ماکسیمم α است. در اینجا با استفاده از روش نیوتن–رافسون جواب‌ها را به‌دست آورده و با برآوردگر بیزی زمانی که از توزیع پیشین ناآگاهی بخش استفاده شود، مقایسه شده است. نتایج شبیه‌سازی در جدول ۳ آورده شده است.

۵ مطالعه شبیه‌سازی

برای مقایسه برآوردگرها دو معیار مخاطره آن‌ها و معیار نزدیکی پیتمن در نظر گرفته شده است. چون محاسبه آن‌ها به‌صورت تحلیلی امکان‌پذیر نبود، این کار به‌وسیله شبیه‌سازی مونت کارلو انجام شده است. برای این منظور مراحل زیر دنبال می‌شود.

۱. از توزیع پیشین (۱)، $\alpha_{\circ}$ تولید می‌شود.

۲. از توزیع پارتو با پارامتر $\alpha_{\circ}$، داده به اندازه n تولید می‌شود.

۳. عبارت‌های $d_{۱}(i) = (\hat{\alpha}_i - \alpha_{\circ})^{۲}$ و $d_{۲}(i) = e^{c(\hat{\alpha}_i - \alpha_{\circ})} - c(\hat{\alpha}_i - \alpha_{\circ}) - ۱$ محاسبه می‌شود. در اینجا منظور از $\hat{\alpha}_i$ یکی از برآوردگرهای معرفی شده در این مقاله است.

۴. مراحل ۱ تا ۳ را $N = ۱۰۰۰۰$ بار تکرار نموده و عبارت‌های

$$ER(SEL)(\hat{\alpha}_i) = \frac{۱}{N}\sum_{i=۱}^{N} d_{۱}(i),$$

$$ER(LINEX)(\hat{\alpha}_i) = \frac{۱}{N}\sum_{i=۱}^{N} d_{۲}(i).$$

محاسبه می‌شوند، که در آن نماد ER برای برآورد مخاطره استفاده شده است.

۵. معیار پیتمن نیز به‌صورت

$$EPMC(\hat{\alpha}_{۱}, \hat{\alpha}_{۲}|\alpha) = \frac{۱}{N}\#[|\hat{\alpha}_{۱_۱} - \alpha_{\circ}| < |\hat{\alpha}_{۲_i} - \alpha_{\circ}|]$$

محاسبه می‌شود. حال برای پارامترهای توزیع پیشین نیز سه مورد در نظر گرفته می‌شود. در اینجا با ثابت نگه داشتن میانگین توزیع پیشین، سه حالت برای واریانس آن در نظر گرفته‌شده است. به جدول ۱ دقت شود.

نتایج شبیه‌سازی در جداول ۲ تا ۵ آمده است. از جداول ۲ تا ۵ موارد زیر ملاحظه می‌شود:

۱. جدول ۲ بیانگر آن است که با افزایش حجم نمونه مخاطره بیزی کاهش می‌یابد.

۲. برآوردگرهای بیزی بر مبنای نمونه‌گیری مجموعه رتبه‌دار، مخاطره بیزی بسیار کمتری نسبت به حالت نمونه‌گیری تصادفی ساده دارند.

۳. در هر دو طرح نمونه‌گیری مخاطره بیزی در حالت توزیع پیشین ناآگاهی بخش بیشتر از حالتی است که از توزیع پیشین مزدوج استفاده گردد.

جدول ۱: میزان اطلاع در حالت میانگین ثابت

	k	θ	$E(\alpha)$	$Var(\alpha)$
اطلاع کم	۱	۲	۰/۵	۰/۲۵
اطلاع متوسط	۵	۱۰	۰/۵	۰/۰۵
اطلاع زیاد	۱۰	۲۰	۰/۵	۰/۰۰۵

۴. با افزایش میزان اطلاع (با تغییر k و θ) مخاطره بیزی کاهش می‌یابد.

۵. جدول ۳ بیانگر این است که برآوردگر درستنمایی ماکسیمم و برآوردگر بیزی تحت تابع زیان توان دوم خطا و توزیع پیشین جفریز تقریبا با هم برابرند، که مطابق انتظار است.

۶. بنابر جداول ۴ و ۵ از نظر معیار نزدیکی پیتمن برآوردگرهای بیزی تحت نمونه‌گیری مجموعه رتبه‌دار بهتر از برآوردگرهای بیزی تحت نمونه‌گیری تصادفی ساده می‌باشند.

۷. بنابر جداول ۴ و ۵ برآوردگر بیزی تحت تابع زیان توان دوم خطا بهتر از برآوردگر بیزی تحت تابع زیان لاینکس برای $c = 1$ است و برای $c = -1$ برآوردگر بیزی تحت تابع زیان لاینکس بهتر از برآوردگر بیزی تحت تابع زیان توان دوم خطا است.

۶ داده‌های مربوط به خسارات باد

در این بخش به منظور تشریح کارآیی برآوردگرهای پیشنهاد شده، داده‌های میزان خسارات باد در سال ۱۹۷۷ از کتاب هوگ و کلاگمن (۱۹۸۴) مورد بررسی قرار می‌گیرند. این داده‌ها شامل ۳۶ مقدار گرد شده خسارات باد بر حسب میلیون دلار است و در جدول ۶ آمده است. ریزو (۲۰۰۹) نشان داد که داده‌ها از توزیع پارتو با پارامتر مقیاس ۱/۵ و پارامتر شکل ۰/۷۶۴ پیروی می‌کنند. بعد از تغییر مقیاس داده‌ها نمونه‌ای به حجم ۶ به دو روش نمونه‌گیری مجموعه رتبه‌دار و نمونه‌گیری تصادفی ساده از بین همین داده‌ها استخراج کرده و برآورد پارامتر α و مخاطره بیزی

جدول ۲: مخاطره بیزی برآوردگرها به کمک شبیه‌سازی مونت کارلو

			SEL					LINEX			
			جفریز		گاما			جفریز		گاما	
n	k	θ	SRS	RSS	SRS	RSS	c	SRS	RSS	SRS	RSS
۴	۱	۲	۰/۲۲۹۰	۰/۰۴۰۸	۰/۰۶۳۵	۰/۰۳۲۷	۱	۰/۱۰۷۰	۰/۰۲۳۱	۰/۰۲۶۰	۰/۰۱۵۳
							۱-	۰/۰۳۸۸	۰/۰۲۵۲	۰/۰۳۸۸	۰/۰۱۷۴
	۵	۱۰	۰/۲۵۸۸	۰/۰۵۰۱	۰/۰۱۳۳	۰/۰۱۲۹	۱	۰/۱۱۷۵	۰/۰۲۴۳	۰/۰۰۶۸	۰/۰۰۶۰
							۱-	۰/۱۵۷۶	۰/۰۲۶۱	۰/۰۰۷۸	۰/۰۰۶۹
	۱۰	۲۰	۰/۲۴۳۶	۰/۰۵۰۱	۰/۰۰۶۶	۰/۰۰۶۳	۱	۰/۱۰۹۵	۰/۰۲۴۳	۰/۰۰۳۱	۰/۰۰۳۰
							۱-	۰/۱۴۷۱	۰/۰۲۶۱	۰/۰۰۳۶	۰/۰۰۳۳
۵	۱	۲	۰/۱۳۳۷	۰/۰۲۷۱	۰/۰۵۶۱	۰/۰۲۱۷	۱	۰/۰۶۱۱	۰/۰۱۳۱	۰/۰۲۴۱	۰/۰۱۰۴
							۱-	۰/۰۷۶۶	۰/۰۱۴۰	۰/۰۳۲۴	۰/۰۱۱۲
	۵	۱۰	۰/۱۵۶۳	۰/۰۲۷۳	۰/۰۱۳۰	۰/۰۱۱۲	۱	۰/۰۷۹۶	۰/۰۱۳۶	۰/۰۰۵۷	۰/۰۰۵۳
							۱-	۰/۰۶۹۵	۰/۰۱۴۰	۰/۰۰۷۴	۰/۰۰۵۸
	۱۰	۲۰	۰/۱۵۵۲	۰/۰۲۷۳	۰/۰۰۶۷	۰/۰۰۶۴	۱	۰/۰۷۷۹	۰/۰۱۳۳	۰/۰۰۳۱	۰/۰۰۳۰
							۱-	۰/۰۶۰۱	۰/۰۱۴۰	۰/۰۰۳۵	۰/۰۰۳۳
۶	۱	۲	۰/۰۹۸۶	۰/۰۱۸۰	۰/۰۴۸۷	۰/۰۱۵۴	۱	۰/۰۴۸۶	۰/۰۰۸۸	۰/۰۲۱۸	۰/۰۰۷۵
							۱-	۰/۰۵۳۳	۰/۰۰۹۱	۰/۰۲۷۱	۰/۰۰۷۹
	۵	۱۰	۰/۱۰۱۱	۰/۰۱۷۶	۰/۰۱۳۰	۰/۰۰۹۳	۱	۰/۰۴۷۸	۰/۰۰۸۶	۰/۰۰۶۰	۰/۰۰۴۵
							۱-	۰/۰۵۴۴	۰/۰۰۹۰	۰/۰۰۷۴	۰/۰۰۴۸
	۱۰	۲۰	۰/۱۰۶۳	۰/۰۱۷۶	۰/۰۰۶۵	۰/۰۰۵۹	۱	۰/۰۴۶۷	۰/۰۰۸۶	۰/۰۰۲۵۰	۰/۰۰۲۰
							۱-	۰/۰۵۱۳	۰/۰۰۸۹	۰/۰۰۳۹	۰/۰۰۲۷

جدول ۳: مقایسه برآوردگر $\hat{\alpha}_{ML}$ با $\hat{\alpha}^J_S(RSS)$ و میانگین توان دوم خطای آن‌ها برای $\alpha = ۲$

n	$\hat{\alpha}_{ML}$	$\hat{\alpha}^J_S(RSS)$	$ER(\hat{\alpha}^j_S(RSS))$	$MSE(\hat{\alpha}_{ML}(RSS))$
۱۰	۲/۲۱۹۶	۲/۲۲۹۸	۰/۶۸۷۱	۰/۷۱۳۳
۲۰	۲/۱۰۳۲	۲/۱۰۴۹	۰/۲۵۸۶	۰/۲۹۹۱
۳۰	۲/۰۶۷۸	۲/۰۶۸۷	۰/۱۵۶۱	۰/۱۷۸۲

جدول ۴: مقایسه برآوردگرها از نظر معیار پیتمن برای $k = ۵$ و $\theta = ۱۰$

معیار پیتمن	c	n: ۴	۵	۶
$\hat{\alpha}_L(RSS)\ v.s.\ \hat{\alpha}_L(SRS)$	۱	۰/۴۲۴۰	۰/۴۰۷۳	۰/۳۷۸۲
	−۱	۰/۴۲۷۵	۰/۴۰۹۳	۰/۳۸۰۵
$\hat{\alpha}_S(RSS)\ v.s.\ \hat{\alpha}_S(SRS)$		۰/۴۲۳۲	۰/۴۰۷۱	۰/۳۷۷۸
$\hat{\alpha}_S(SRS)\ v.s.\ \hat{\alpha}_L(SRS)$	۱	۰/۴۷۲۵	۰/۴۷۰۹	۰/۴۸۲۵
	−۱	۰/۵۶۱۷	۰/۵۵۶۶	۰/۵۴۵۷
$\hat{\alpha}_S(RSS)\ v.s.\ \hat{\alpha}_L(RSS)$	۱	۰/۴۷۸۸	۰/۴۸۱۵	۰/۴۷۶۴
	−۱	۰/۵۴۶۳	۰/۵۴۱۳	۰/۵۴۴۱

جدول ۵: مقایسه برآوردگرها از نظر معیار پیتمن برای $k = ۱۰$ و $\theta = ۲۰$

معیار پیتمن	c	n: ۴	۵	۶
$\hat{\alpha}_L(RSS)\ v.s.\ \hat{\alpha}_L(SRS)$	۱	۰/۴۳۳۸	۰/۴۱۵۱	۰/۴۰۸۹
	−۱	۰/۴۳۶۰	۰/۴۱۳۸	۰/۴۰۸۴
$\hat{\alpha}_S(RSS)\ v.s.\ \hat{\alpha}_S(SRS)$		۰/۴۳۴۷	۰/۴۱۷۸	۰/۴۱۰۴
$\hat{\alpha}_S(SRS)\ v.s.\ \hat{\alpha}_L(SRS)$	۱	۰/۴۷۸۶	۰/۴۷۷۷	۰/۴۸۵۲
	−۱	۰/۵۵۰۰	۰/۵۴۵۴	۰/۵۴۱۲
$\hat{\alpha}_S(RSS)\ v.s.\ \hat{\alpha}_L(RSS)$	۱	۰/۴۸۰۲	۰/۴۸۱۲	۰/۴۸۷۹
	−۱	۰/۵۴۴۶	۰/۵۳۹۴	۰/۵۲۹۴

برآوردگرها در این دو روش به کمک شبیه‌سازی مونت کارلو با ۱۰۰۰۰ بار تکرار به‌دست آورده و در جداول ۷ و ۸ خلاصه شده‌اند.

جدول ۶: خسارات باد بر حسب میلیون دلار

۳۲	۶	۶	۲	۲	۶	۴	۲۴	۲۲
۲	۱۵	۲	۲	۶	۹	۵	۵	۲
۲۴	۲	۳	۳	۸	۲۵	۱۷	۲	۲
۲۴	۲	۴۳	۵	۵	۲۷	۲	۴	۳

همان‌طور که در جدول ۷ ملاحظه می‌شود مقدار برآورد شده به مقدار واقعی نزدیک است. جدول ۸ نیز نتایج بخش شبیه‌سازی شامل جدول ۲ را تایید می‌کند و مهمترین نکته‌ای که دوباره روی آن تاکید می‌شود، کوچک بودن ریسک بیزی برآوردگرها در روش نمونه‌گیری مجموعه رتبه‌دار نسبت به روش نمونه‌گیری تصادفی ساده است.

جدول ۷: برآورد پارامتر برای داده‌های خسارات باد (مقدار واقعی α برابر ۰/۷۶۴ است)

		SEL					LINEX			
		جفریز		گاما			جفریز		گاما	
k	θ	SRS	RSS	SRS	RSS	c	SRS	RSS	SRS	RSS
۱	۲	۰/۸۳۸۶	۰/۷۳۹۳	۰/۷۵۰۷	۰/۷۱۹۶	۱	۰/۷۷۹۸	۰/۷۲۴۴	۰/۷۱۱۱	۰/۷۰۶۴
						۱−	۰/۹۱۴۰	۰/۷۵۵۱	۰/۷۹۷۲	۰/۷۳۳۶
۵	۱۰	۰/۸۳۸۶	۰/۷۳۷۸	۰/۶۲۵۸	۰/۶۶۵۸	۱	۰/۷۷۹۷	۰/۷۲۳۹	۰/۶۰۸۴	۰/۶۵۶۶
						۱−	۰/۹۱۳۷	۰/۷۵۴۵	۰/۶۴۴۶	۰/۶۷۵۴
۱۰	۲۰	۰/۸۳۳۱	۰/۷۳۹۶	۰/۵۷۷۶	۰/۶۲۸۲	۱	۰/۷۷۵۳	۰/۷۴۴۷	۰/۵۶۷۴	۰/۶۲۱۵
						۱−	۰/۹۰۶۴	۰/۷۵۵۴	۰/۵۸۸۴	۰/۶۳۵۱

۷ بحث و نتیجه‌گیری

در مسائل آماری به‌دست آوردن داده‌ها از طریق فرآیندی دقیق که بتوان با صحت و دقت بیشتر در مورد خصیصه‌های جامعه اظهار نظر کرد، یکی از اصول اساسی است. در شرایطی که در آن اندازه‌گیری دقیق مشاهدات نمونه به دلایلی چون هزینه،

جدول ۸: مخاطره بیزی برآوردگرها برای داده‌های خسارات باد

LINEX					SEL					
Gamma		Jeffrys			Gamma		Jeffrys			
RSS	SRS	RSS	SRS	c	RSS	SRS	RSS	SRS	θ	k
۰/۰۰۸۷	۰/۰۱۵۳	۰/۰۰۹۹	۰/۰۳۵۶	۱	۰/۰۱۷۸	۰/۰۳۴۷	۰/۰۲۰۵	۰/۰۸۸۲	۲	۱
۰/۰۰۹۰	۰/۰۲۰۷	۰/۰۱۰۶	۰/۰۵۵۲	−۱						
۰/۰۰۸۸	۰/۰۱۳۹	۰/۰۰۹۹	۰/۰۳۵۱	۱	۰/۰۱۷۲	۰/۰۲۵۳	۰/۰۲۰۴	۰/۰۸۷۳	۱۰	۵
۰/۰۰۸۳	۰/۰۱۱۳	۰/۰۱۰۶	۰/۰۵۴۸	−۱						
۰/۰۱۱۴	۰/۰۱۸۹	۰/۰۱۰۰	۰/۰۳۳۳	۱	۰/۰۲۲۶	۰/۰۳۶۸	۰/۰۲۰۶	۰/۰۸۲۷	۲۰	۱۰
۰/۰۱۱۱	۰/۰۱۷۶	۰/۰۱۰۶	۰/۰۵۲۱	−۱						

تخریب کنندگی نمونه و زمان‌بر بودن دشوار باشد اما رتبه‌گذاری مجموعه‌هایی از نمونه به نسبت آسان و قابل اطمینان باشد، روش نمونه‌گیری مجموعه رتبه‌دار روش مناسبی است. یکی از برتری‌های این روش این است که برآوردگر میانگین جامعه حاصل از آن کاراتر از نمونه‌گیری تصادفی ساده عمل می‌کند. همچنین برآوردگرهای بیزی پارامترهای توزیع پاریتو بر اساس مشاهدات حاصل از این طرح نمونه‌گیری ارائه شد و با برآوردگرهای حاصل از روش نمونه‌گیری تصادفی ساده مقایسه گردید. نتایج مقایسه نشان از برتری روش نمونه‌گیری مجموعه رتبه‌دار بر روش تصادفی ساده داشت.

تقدیر و تشکر

نویسندگان مقاله لازم می‌دانند از زحمات سر دبیر مجله و داوران محترم که با نقطه نظرات و پیشنهادات سازنده‌ی خود، باعث بهبود مقاله شدند تقدیر و تشکر کنند.

مراجع

زمانزاده، ا.، احمدی، ج. (۱۳۹۰)، بازه اطمینان ناپارامتری با ضریب اطمینان دقیق برای چندک‌های جامعه بر اساس نمونه‌گیری مجموعه رتبه‌دار، مجله علوم آماری، جلد **۵**، ۳۳-۲۹.

امینی، م.، ارقامی، ن. (۱۳۸۷)، دقت برآورد نسبتی درنمونه‌گیری از مجموعه‌ی رتبه‌دار ، گزیده مطالب آماری، سال **۱۹**، ۴۷–۲۷.

Abu-Dayyeh, W., Assrhani, A. and Ibrahim, K. (2011), Estimation of the Shape and Scale Parameters of Pareto Distribution Using Ranked Set Sampling, *Statistical Papers*, **54**, 207-225.

Al-Saleh, M. F. and Al-Hadhrami, S. A. (2003), Estimation of the Mean of the Exponential Distribution Using Moving Extremes Ranked Set Sampling, *Statistical Papers*, **44**, 367-382.

Al-Ssaleh, M. F., Al-Shrafat, K. and Muttlak, H. (2000), Bayesian Estimation Using Ranked Set Sampling, *Biometrical Journal*, **42**, 489-500.

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Second Edition, Springer, New York.

Chang, Y. C, Hung, W. L. (2007), LINEX Loss Functions with Applications to Determining the Optimum Process Parameters, *Quality & Quantity* , **41**, 291-301.

Chen, Z., Bai, Z. and Sinha, B. K. (2003), *Ranked Set Sampling: Theory and Applications*, Springer, New York.

Hogg, R. V. and Klugman, S. A. (1984), *Loss Distributions*, Wiley, New York.

Keating, J. P., Mason, R. L. and Sen, P. K. (1993), Pitman Measure of Closeness: A Comprision of Statistical Estimators, *Society for Industrial and Applied Mathematics*, Philadelphia.

McIntyre, G. A. (1952), a Method for Unbiased Selective Sampling Using Ranked Sets, *Australian Journal of Agriculture Research*, **3**, 358-390.

McIntyre, G. A. (1952), A Method for Unbiased Selective Sampling Using Ranked Sets, *Australian Journal of Agriculture Research*, **3**, 358-390.

Mode, N., Conquest, L. and Marker, D., (1999), Ranked Set Sampling for Ecological Research: Accounting for the Total Cost of Sampling, *Environmetrics*, **10**, 179-194.

Nussbaum, B. D. and Sinha, B. K. (1997), Cost Effective Gasoline Sampling Using Ranked Set Sampling, *In Proceedings of the Section on Statistics and the Environment*, American Statistical Association, Alexandria, VA. 83-87.

Patil, G. P., Sinha, A. K. and Taillie, C., (1999), Ranked Set Sampling: A Bibliography, *Environmental and Ecological Statistics*, **6**, 91-98.

Pitman, E. J. G. (1937), The Closest Estimates of Statistical Parameters, *Proceedings of the Cambridge Philosophical Society*, **33**, 212-222.

Rizzo, L. (2009), New Goodness-of-Fit Tests for Pareto Distributions, *ASTIN Bulletin*, **39**, 691-715.

Sadek, A., Sultan, K. S. and Balakrishnan, N. (2013), Bayesian Estimation Based on Ranked Set Sampling Using Asymetric Loss Function, *Bulletin of the Malaysian Mathematical Sciences Society*, to appear.

Shaibu, A. B. and Muttlak, H. A. (2004), Estimating the Parameters of the Normal, Exponential and Gamma Distributions Using Median and Extreme Ranked Set Samples, *Statistica*, **64**, 75-98.

Sinha, B. K., Sinha, B. K. and Purkayastha, S. (1996), On Some Aspects of Ranked Set Sampling for Estimation of Normal and Exponential Parameters, *Statistical Decisions*, **14**, 223-240.

Varian, H. R. (1975), A Bayesian Approach to Real Estate Assessment, In Finberg, S. E., Zellner, A., eds., *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J.Savege*, Amesterdam: North-Holland, 195-208.

Yu, P. L. H. and Lam, K. (1997), Regression Estimator in Ranked Set Sampling, *Biometrics*, **53**, 1070-1080.