

مقایسه دو روش استخراج رکوردها از دیدگاه اطلاع فیشر

مصطفی رزمخواه، جعفر احمدی، بهاره خطیب آستانه

گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

تاریخ دریافت: ۱۳۸۵/۳/۲۹ تاریخ آخرین بازنگری: ۱۳۸۵/۱۲/۲۶

چکیده: رکوردها در یک دنباله از مشاهدات مقادیری هستند که از مشاهدات ماقبل خود بزرگتر (کوچکتر) باشند. اجرای نمونه‌گیری دنباله‌ای تا کسب تعدادی مشخص از داده‌های رکوردی، در مقالات به روش نمونه‌گیری معکوس معروف است. در این حالت امید ریاضی زمان انتظار بین دو رکورد نامتناهی است و تعداد داده‌های رکوردی مستخرج نیز محدود است. یکی از راهکارها برای مقابله با این مشکل، استفاده از طرح نمونه‌گیری تکراری به روش معکوس است. در این طرح با به کار گرفتن روش نمونه‌گیری معکوس ابتدا در مرحله‌ی اول k_1 رکورد را استخراج نموده و نمونه‌گیری را در این مرحله متوقف می‌سازیم. سپس در مرحله‌ی دوم با در نظر گرفتن نمونه‌ای مستقل از نمونه‌ی اول، k_2 رکورد بعدی را استخراج می‌نمائیم و همین روند را تا رسیدن به تعداد رکوردهای مورد نظر ادامه می‌دهیم. در این مقاله حالتی را در نظر می‌گیریم که تعداد رکوردهای مستخرج در هر مرحله ثابت و همگی برابر k باشند. نتایج کلی برای چند رده مختلف از توزیع‌های مهم آماری به دست آمده و برخی از توزیع‌های مهم طول عمر از جمله توزیع نمایی، بر نوع دوازده و وایبول برای تشریح نتایج به دست آمده مورد استفاده قرار گرفته‌اند.

آدرس الکترونیک مسئول مقاله: مصطفی رزمخواه، razmkhah_m@yahoo.com، جعفر احمدی، عضو هسته‌ی قطب علمی آنالیز داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد.

واژه‌های کلیدی : خانواده مقیاسی، خانواده شکل، رکورد بالا (پایین)، توزیع بر.

۱ مقدمه

فرض کنید X_1, X_2, \dots دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع از جامعه‌ای با تابع توزیع $F(x; \theta)$ باشد، در اینصورت گوییم X_j یک رکورد بالا (پایین) است هرگاه از تمام مقادیر ماقبل خود بزرگتر (کوچکتر) باشد. بنابراین X_1 یک رکورد بالا (پایین) است. شماره‌ی سریالی که m امین رکورد بالا (پایین) در آن رخ می‌دهد، خود یک متغیر تصادفی است و به زمان رکورد معروف است. رکوردها در مسائلی از قبیل آزمون طول عمر قطعات الکترونیکی گران‌قیمت و یا آزمون استقامت الوارهای چوب، داده‌های مربوط به مسابقات ورزشی، هواشناسی، ژئوفیزیک، زلزله‌نگاری، تعمیر مینی‌مال در مسائل قابلیت اعتماد، فرآیند پواسون ناهمگن و مانند آن نقش شایان توجهی ایفا می‌کنند. مطالعات آماری رکوردها در سال ۱۹۵۲ توسط چندلر شروع شد. پس از آن مطالعات قابل ملاحظه‌ای روی رکوردها و آماره‌های مربوطه انجام شده است، بطوری‌که تا کنون بیش از ۶۵۰ مقاله در زمینه رکوردها چاپ شده است. علاقه‌مندان می‌توانند به مقالات مروری گللیک (۱۹۷۸)، نوزوروف (۱۹۸۷) و ناگاراچا (۱۹۸۸) یا به کتاب‌های آرنولد و همکاران (۱۹۹۸) و نوزوروف (۲۰۰۱) مراجعه نمایند. برای استخراج داده‌های رکوردی در آزمایش‌های دنباله‌ای، اگر آزمایش‌ها را تا آنجایی انجام دهیم که به تعدادی مشخص از رکوردها دست یابیم، به آن روش نمونه‌گیری معکوس می‌گویند (سامانیگو و ویتاکر، ۱۹۸۸) که در این مقاله از آن به عنوان طرح نمونه‌گیری A یاد می‌کنیم. حالت دیگر که در آن حجم نمونه ثابت اما تعداد رکوردها تصادفی است به طرح نمونه‌گیری مستقیم معروف است؛ از جمله کارهای انجام شده در این مورد می‌توان به هافمن و ناگاراچا (۲۰۰۳) اشاره نمود. فرض کنید در مهندسی عمران، هدف بررسی استقامت بلوک‌هایی باشد که در سازه‌های ساختمانی مورد استفاده قرار می‌گیرند. برای این منظور ابتدا اولین بلوک را به تصادف انتخاب نموده و با اعمال فشار، نقطه‌ی تخریب آن را به دست می‌آوریم. این مقدار به عنوان رکورد اول ثبت

می‌شود. سپس بلوک دوم را نیز به طور تصادفی انتخاب کرده و آن را حداکثر تا مقدار رکورد اول تحت آزمایش قرار می‌دهیم؛ چنانچه تخریب نشود به عنوان نمونه‌ی سالم از آزمایش خارج می‌شود و در غیر این صورت به عنوان رکورد پایین دوم ثبت می‌شود. اگر همین روند را ادامه دهیم تا سرانجام n رکورد به دست آید، به آن روش نمونه‌گیری معکوس (طرح A) اطلاق می‌گردد. بنابراین در این آزمایش دنباله‌ای، فقط نقطه‌ی تخریب بلوک‌هایی ثبت می‌شوند که نامقاوم‌تر از نمونه‌های ماقبل خود باشند. بدیهی است که این روش نسبت به نمونه‌گیری تصادفی ساده مقرون به صرفه‌تر است. اما در این طرح تعداد رکوردها ثابت و حجم نمونه‌ی مورد نیاز تصادفی است. در این مقاله روشی دیگر معرفی می‌شود و آن استفاده از طرح نمونه‌های تکراری به روش معکوس است که آن را با نام طرح نمونه‌گیری B می‌شناسیم. در این طرح با به کار گرفتن روش نمونه‌گیری معکوس ابتدا در مرحله‌ی اول k_1 رکورد را استخراج نموده و نمونه‌گیری را در این مرحله متوقف می‌سازیم. سپس در مرحله‌ی دوم با در نظر گرفتن نمونه‌ای مستقل از نمونه‌ی اول، k_2 رکورد بعدی را استخراج می‌نمائیم و همین روند را تا رسیدن به تعداد رکوردهای مورد نظر ادامه می‌دهیم. طرح B در بسیاری از فرآیندهای صنعتی که کالاهای ساخته شده در بازه‌های زمانی منظم با شرایط ثابت تولید می‌شوند، نتایج سودمندی به دست خواهد داد. به عنوان مثال فرض کنید یک کارخانه‌ی بلوک‌سازی دارای دو خط تولید باشد که در شرایط کاملاً یکسانی فعالیت می‌کنند. در این صورت می‌توانیم به جای انتخاب یک خط تولید برای دستیابی به نقطه‌ی تخریب k رکورد، k_1 مورد را از خط تولید اول و $k_2 = k - k_1$ نمونه را از خط تولید دوم مشاهده نمائیم. در این مقاله حالتی را در نظر می‌گیریم که تعداد رکوردهای مستخرج در هر مرحله ثابت و همگی برابر k باشند. مثلاً برای استخراج mk رکورد در طرح نمونه‌گیری B ، نمونه‌گیری را m مرتبه تکرار می‌کنیم و در هر مرتبه k رکورد استخراج می‌نمائیم. اما برای انجام این کار در طرح نمونه‌گیری A ، تعداد mk رکورد را به طور متوالی استخراج می‌کنیم. طرح نمونه‌گیری B ، به ازای $k = 2$ توسط احمدی و ارقامی (۲۰۰۱) پیشنهاد شد. گولاتی و پاچت (۲۰۰۳) طرح نمونه‌گیری B را برای برآورد تابع توزیع و چندک‌های جامعه بر اساس

۲۲ مقایسه دو روش استخراج رکوردها از دیدگاه اطلاع فشر

رکوردها به کار برده‌اند. فرض کنید X یک متغیر تصادفی پیوسته با تابع چگالی احتمال $f(x; \theta)$ باشد. در این صورت میزان اطلاع فشر نهفته در X در خصوص پارامتر θ بصورت $I_X(\theta) = E\left(\frac{\partial}{\partial \theta} \log f(X; \theta)\right)^2$ تعریف می‌شود. تحت شرایطی (که به شرایط نظم معروفند) برابری زیر را داریم:

$$I_X(\theta) = -E\left(\frac{\partial^2}{\partial \theta^2} \log f(X; \theta)\right).$$

اطلاع فشر که موضوع اصلی این مقاله می‌باشد، در نامساوی کرامر-رائو برای به‌دست آوردن کران پایین واریانس برآوردگرها و همچنین در واریانس توزیع حدی برآوردگرهای ماکسیمم درست‌نمایی ظاهر می‌شود. اخیراً بررسی اطلاع فشر نهفته در آماره‌های رکوردی مورد توجه محققین زیادی قرار گرفته است. در ایران می‌توان به احمدی (۲۰۰۰)، احمدی و ارقامی (۲۰۰۱ و ۲۰۰۳) و رزمخواه (۱۳۸۳) اشاره نمود. همچنین احمدی و رزمخواه (۱۳۸۱) اطلاع فشر نهفته در مشاهدات مستقل و هم‌توزیع (*i.i.d.*) از توزیع بر نوع دوازده را برای پارامتر شکل و رزمخواه و همکاران (۱۳۸۲) اطلاع فشر نهفته در آماره‌های رکوردی همان توزیع را مورد بررسی قرار داده‌اند. در خارج از ایران مقالاتی توسط هافمن و ناگاراچا (۲۰۰۳)، هافمن و همکاران (۲۰۰۵) و بالاکریشنان و استپانوف (۲۰۰۶) در این راستا چاپ شده است.

در این مقاله گوئیم توزیع $F(x; \theta)$ دارای خاصیت **UBGA** (**LBGA**)، **UBEA** (**LBEA**) یا **UBLA** (**LBLA**) است، هرگاه اطلاع فشر نهفته در رکوردهای بالا (پایین) در خصوص پارامتر θ در طرح B به ترتیب بیشتر، مساوی یا کمتر از طرح A باشد. در ادامه، ابتدا در بخش دوم ابزار و نتایج کمکی بیان می‌شوند و سپس در بخش سوم، نتایج کلی برای محاسبه‌ی میزان اطلاع فشر نهفته در رکوردهای بالا (پایین) در رده‌های مختلف از توزیع‌ها ارائه می‌گردند. در بخش چهارم، میزان اطلاع فشر نهفته در توزیع‌های نمایی، بر نوع دوازده و وایبول، در هر دو طرح نمونه‌گیری مقایسه و چند خانواده‌ی یک‌پارامتری بر این اساس دسته‌بندی می‌شوند.

۲ نتایج کمی

با توجه به تعریف دو طرح نمونه‌گیری A و B که در بخش مقدمه ذکر شد، بدیهی است که اطلاع فیشر نهفته در رکوردها و زمان بین آنها در طرح A بیشتر از یا مساوی با طرح B می‌باشد، زیرا در طرح B عملاً اطلاعات مربوط به زمان آخرین رکورد در هر مرتبه‌ی تکرار از دست می‌رود ولی در طرح A چنین اتفاقی نمی‌افتد. بنابراین در این مقاله فقط به بررسی رکوردها بدون در نظر گرفتن زمان آنها می‌پردازیم. در این بخش سعی شده است تا ابزار لازم برای حصول نتایج بخش‌های بعدی ارائه شوند.

لم ۱: فرض کنید X_1, X_2, \dots دنباله‌ای از متغیرهای تصادفی $i.i.d.$ با تابع توزیع $F_X(x; \theta)$ و تابع چگالی $f_X(x; \theta)$ باشند. در این صورت در طرح نمونه‌گیری A ,

الف: تابع چگالی احتمال نخستین mk رکورد بالا عبارت است از

$$f_{\underline{R}}(\underline{r}; \theta) = f_X(r_{mk}; \theta) \prod_{i=1}^{mk-1} \frac{f_X(r_i; \theta)}{\bar{F}_X(r_i; \theta)},$$

که در آن $\underline{R} = (R_1, \dots, R_{mk})$ معرف نخستین mk رکورد بالا و $\underline{r} = (r_1, \dots, r_{mk})$ نشان‌دهنده‌ی مقادیر مشاهده شده‌ی آنها می‌باشد.

ب: تابع چگالی احتمال نخستین mk رکورد پایین به صورت زیر است

$$f_{\underline{R}'}(\underline{r}'; \theta) = f_X(r'_{mk}; \theta) \prod_{i=1}^{mk-1} \frac{f_X(r'_i; \theta)}{F_X(r'_i; \theta)},$$

که در آن $\underline{R}' = (R'_1, \dots, R'_{mk})$ معرف نخستین mk رکورد پایین و $\underline{r}' = (r'_1, \dots, r'_{mk})$ نشان‌دهنده‌ی مقادیر مشاهده شده‌ی آنهاست.

ج: تابع چگالی احتمال زامین رکورد بالا R_j عبارت است از

$$f_{R_j}(r; \theta) = \frac{(-\log \bar{F}_X(r; \theta))^{j-1}}{(j-1)!} f_X(r; \theta).$$

د: تابع چگالی احتمال زامین رکورد پایین R'_j به صورت زیر است

$$f_{R'_j}(r'; \theta) = \frac{(-\log F_X(r'; \theta))^{j-1}}{(j-1)!} f_X(r'; \theta).$$

برهان : برای جزئیات برهان به کتاب آرنولد و همکاران (۱۹۹۸) مراجعه شود.

لم ۲ : تحت مفروضات لم ۱، در طرح نمونه‌گیری B ، روابط زیر برقرار هستند:
الف: تابع چگالی احتمال نخستین mk رکورد بالا عبارت است از

$$f_{\underline{R}_1, \dots, \underline{R}_m}(\underline{r}_1, \dots, \underline{r}_m; \theta) = \prod_{i=1}^m f_X(r_{ik}; \theta) \prod_{j=1}^{k-1} \frac{f_X(r_{ij}; \theta)}{F_X(r_{ij}; \theta)}, \quad (1)$$

که در آن به ازای $i = 1, \dots, m$ معرف k رکورد بالای استخراج شده در مرحله i ام طرح نمونه‌گیری B و $\underline{r}_i = (r_{i1}, \dots, r_{ik})$ نشان دهنده‌ی مقادیر مشاهده شده‌ی آنهاست.

ب: تابع چگالی احتمال نخستین mk رکورد پایین به صورت زیر است

$$f_{\underline{R}'_1, \dots, \underline{R}'_k}(\underline{r}'_1, \dots, \underline{r}'_k; \theta) = \prod_{i=1}^m f_X(r'_{ik}; \theta) \prod_{j=1}^{k-1} \frac{f_X(r'_{ij}; \theta)}{F_X(r'_{ij}; \theta)}. \quad (2)$$

که در آن به ازای $i = 1, \dots, m$ معرف k رکورد پایین استخراج شده در مرحله i ام طرح نمونه‌گیری B و $\underline{r}'_i = (r'_{i1}, \dots, r'_{ik})$ نشان دهنده‌ی مقادیر مشاهده شده‌ی آنهاست.

ج: برای $j = 1, \dots, k$ ، زامین رکورد بالا (پایین) در تکرارهای مختلف طرح B هم‌توزیع است با زامین رکورد بالا (پایین) در طرح A . به عبارتی، به ازای $R'_{ij} \stackrel{d}{=} R'_j$ و $R_{ij} \stackrel{d}{=} R_j$ ، $j = 1, \dots, k$ و $i = 1, \dots, m$

برهان : با توجه به اینکه تکرارها در m مرحله به طور مستقل از هم انجام می‌شوند، توابع چگالی احتمال توأم k رکورد مستخرج از هر مرحله بایستی در هم ضرب شوند. یعنی، برای رکوردهای بالا داریم

$$f(\underline{r}_1, \dots, \underline{r}_k; \theta) = \prod_{i=1}^m f(\underline{r}_i; \theta).$$

برای رکوردهای پایین نیز به طور مشابه است. بنابراین با استفاده از لم ۱ قسمت‌های الف و ب، روابط (۱) و (۲) نتیجه می‌شوند. قسمت ج نیز با توجه به نحوه‌ی

استخراج رکوردها در طرح‌های A و B بدیهی است.

لم ۳: تحت مفروضات لم ۱، اطلاع فیشر نهفته در نخستین mk رکورد بالا در طرح نمونه‌گیری B عبارت است از

$$I_{\underline{R}}^B(\theta) = -m \sum_{j=1}^k E \left(\frac{\partial^2}{\partial \theta^2} \log f_X(R_j; \theta) \right) + m \sum_{j=1}^{k-1} E \left(\frac{\partial^2}{\partial \theta^2} \log \bar{F}_X(R_j; \theta) \right). \quad (۳)$$

برهان: از رابطه‌ی (۱)، تساوی زیر را داریم

$$\frac{\partial^2}{\partial \theta^2} \log f_{\underline{R}}(\underline{x}; \theta) = \sum_{i=1}^m \sum_{j=1}^k \frac{\partial^2}{\partial \theta^2} \log f_X(R_{ij}; \theta) - \sum_{i=1}^m \sum_{j=1}^{k-1} \frac{\partial^2}{\partial \theta^2} \log \bar{F}_X(R_{ij}; \theta).$$

با توجه به تعریف اطلاع فیشر و لم ۲ قسمت ج، برهان کامل است.

تذکر ۱: مشابه فرمول (۳)، اطلاع فیشر نهفته در نخستین mk رکورد بالا در طرح A عبارت است از

$$I_{\underline{R}}^A(\theta) = - \sum_{j=1}^{mk} E \left(\frac{\partial^2}{\partial \theta^2} \log f_X(R_j; \theta) \right) + \sum_{j=1}^{mk-1} E \left(\frac{\partial^2}{\partial \theta^2} \log \bar{F}_X(R_j; \theta) \right). \quad (۴)$$

لم ۴: تحت مفروضات لم ۱، اطلاع فیشر نهفته در نخستین mk رکورد پایین در طرح نمونه‌گیری B عبارت است از

$$I_{\underline{R}'}^B(\theta) = -m \sum_{j=1}^k E \left(\frac{\partial^2}{\partial \theta^2} \log f_X(R'_j; \theta) \right) + m \sum_{j=1}^{k-1} E \left(\frac{\partial^2}{\partial \theta^2} \log F_X(R'_j; \theta) \right). \quad (۵)$$

برهان با استفاده از رابطه‌ی (۲)، مشابه لم ۳ ثابت می‌شود.

۲۶ مقایسه دو روش استخراج رکوردها از دیدگاه اطلاع فیشر

تذکر ۲ : مشابه فرمول (۵)، اطلاع فیشر نهفته در نخستین mk رکورد پایین در طرح A عبارت است از

$$I_{\underline{R}'}^A(\theta) = - \sum_{i=1}^{mk} E \left(\frac{\partial^2}{\partial \theta^2} \log f_X(R'_i; \theta) \right) + \sum_{i=1}^{mk-1} E \left(\frac{\partial^2}{\partial \theta^2} \log F_X(R'_i; \theta) \right). \quad (۶)$$

لم ۵ : (لهمن و کسلا، ۱۹۹۸) فرض کنید متغیر تصادفی X دارای تابع توزیع احتمال $F_X(x; \theta)$ باشد و $\theta = h(\eta)$ که h تابعی مشتق‌پذیر است. در این صورت

$$I_X^*(\eta) = I_X(h(\eta)) \left[\frac{\partial}{\partial \eta} h(\eta) \right]^2,$$

که در آن $I_X^*(\eta)$ میزان اطلاع فیشر نهفته در X در خصوص پارامتر η می‌باشد.

لم ۶ : اگر X دارای تابع چگالی احتمال $f(x; \theta)$ باشد و $Y = g(X)$ که g تابعی یک به یک، مشتق‌پذیر و مستقل از θ باشد، آنگاه،

$$I_X(\theta) = I_Y(\theta).$$

لم ۷ : فرض کنید X و Y دو متغیر تصادفی باشند به طوری که با احتمال یک $X \leq Y$ و ϕ یک تابع افزایشی باشد. در این صورت با احتمال یک داریم

$$E[\phi(X)] \leq E[\phi(Y)].$$

لم‌های ۶ و ۷ به سادگی ثابت می‌شوند.

۳ نتایج اصلی

در این بخش نتایج اصلی برای مقایسه‌ی دو طرح نمونه‌گیری A و B از دیدگاه اطلاع فیشر ارائه می‌شوند. برای این منظور ابتدا چند رده از توزیع‌های مهم آماری

را به شرح ذیل معرفی می‌کنیم.

الف: C_1 را رده توزیع‌های F در نظر می‌گیریم طوری که

$$\bar{F}(x; \theta) = e^{-\alpha(\theta)a(x)},$$

که در آن $\alpha(\cdot)$ فقط تابعی از θ و $a(\cdot)$ فقط تابعی از x و هر دو مثبت هستند.

ب: C_2 را رده توزیع‌های F در نظر می‌گیریم طوری که

$$F(x; \theta) = e^{-\beta(\theta)b(x)},$$

که در آن $\beta(\cdot)$ فقط تابعی از θ و $b(\cdot)$ فقط تابعی از x و هر دو مثبت هستند.

لازم به ذکر است که C_1 و C_2 شامل بسیاری از توزیع‌های معروف طول عمر از جمله نمایی، وایبول، پارتو، بر نوع دوازده، توزیع توانی، توزیع مقادیر فرین و ... می‌باشند.

ج: C_3 را خانواده‌ی نمایی با تابع چگالی (جرم) احتمال زیر در نظر می‌گیریم

$$f(x; \theta) = A(x)B(\theta)e^{C(\theta)D(x)}.$$

د: C_4 را خانواده‌ی توزیع‌های مکانی F در نظر می‌گیریم طوری که

$$F(x; \theta) = F_0(x - \theta).$$

ه: C_5 را خانواده‌ی توزیع‌های مقیاسی F در نظر می‌گیریم. به عبارتی

$$C_5 = \{F : F(x; \theta) = F_0(\theta x), \theta > 0\}.$$

و: C_6 را خانواده‌ی توزیع‌های شکل F در نظر می‌گیریم طوری که

$$F(x; \theta) = F_0(x^\theta), \theta > 0.$$

در بندهای د تا و، F_0 به θ بستگی ندارد.

قضیه ۱: فرض کنید F متعلق به کلاس C_1 باشد. در این صورت این توزیع دارای خواص $UBEA$ و $LBGA$ می‌باشد.

برهان : با استفاده از لم‌های ۵ و ۶، می‌توان بدون از دست دادن کلیت مسأله فرض نمود $\alpha(\theta) = \theta$ و $a(x) = x$. بنابراین می‌توان توزیع نمایی با تابع چگالی احتمال $f(x) = \theta e^{-\theta x}$ از کلاس C_1 را در نظر گرفت. در این حالت تساوی‌های زیر را داریم:

$$\frac{\partial^2}{\partial \theta^2} \log f(x; \theta) = -\frac{1}{\theta^2}, \quad (۷)$$

$$\frac{\partial^2}{\partial \theta^2} \log \bar{F}(x; \theta) = 0, \quad (۸)$$

$$\frac{\partial^2}{\partial \theta^2} \log F(x; \theta) = -\frac{x^2 e^{-\theta x}}{(1 - e^{-\theta x})^2}. \quad (۹)$$

بنابراین با توجه به روابط (۳)، (۴)، (۷) و (۸) داریم

$$I_{\underline{R}}^B(\theta) = I_{\underline{R}}^A(\theta) = \frac{mk}{\theta^2}.$$

در نتیجه خاصیت *UBEA* برای کلاس C_1 بر اساس رکوردهای بالا برقرار است. برای اثبات خاصیت *LBGA* برای این رده، با توجه به روابط (۵)، (۶)، (۷) و (۹) داریم

$$I_{\underline{R}'}^B(\theta) = \frac{mk}{\theta^2} - m \sum_{j=1}^{k-1} E[g(R'_j)]$$

$$I_{\underline{R}'}^A(\theta) = \frac{mk}{\theta^2} - \sum_{j=1}^{mk-1} E[g(R'_j)], \quad \text{و}$$

$$g(R'_j) = \frac{R_j'^2 e^{-\theta R'_j}}{(1 - e^{-\theta R'_j})^2}. \quad \text{که در آن}$$

از طرفی به ازای $j \geq 1$ ، با احتمال یک $R'_j > R'_{j+1}$ و تابع $g(x)$ برای $x > 0$ یک تابع نزولی بر حسب x است. لذا با احتمال یک داریم $g(R'_j) < g(R'_{j+1})$. در این صورت با توجه به لم ۷،

$$m \sum_{j=1}^{k-1} E[g(R'_j)] < \sum_{j=1}^{mk-1} E[g(R'_j)].$$

در نتیجه

$$I_{R'}^B(\theta) > I_{R'}^A(\theta).$$

با توجه به نامساوی فوق اثبات کامل است.

قضیه ۲: فرض کنید F متعلق به رده C_2 باشد. در این صورت این توزیع دارای خواص $UBGA$ و $LBEA$ می باشد. **برهان** مشابه قضیه ۱ ثابت می شود، با این تفاوت که در این حالت $b(x)$ باید تابعی کاهشی از x باشد.

قضیه ۳: فرض کنید F متعلق به رده C_3 باشد. قرار دهید

$$L(x; \theta) = \frac{\partial^2}{\partial \theta^2} \log \bar{F}_X(x; \theta),$$

در این صورت موارد زیر را داریم:

الف: اگر به ازای مقدار ثابت c داشته باشیم $L(x; \theta) = c$ ، آنگاه F دارای خاصیت $UBEA$ ، $UBLA$ یا $UBGA$ است هر گاه c به ترتیب مثبت، صفر یا منفی باشد.

ب: اگر $L(x; \theta)$ یک تابع افزایشی از x باشد، آنگاه F دارای خاصیت $UBLA$ می باشد.

برهان با استفاده از لم (۵)، می توان بدون از دست دادن کلیت مسأله فرض نمود $C(\theta) = \theta$ در این صورت

$$\frac{\partial^2}{\partial \theta^2} \log f(x; \theta) = \frac{\partial^2}{\partial \theta^2} \log B(\theta).$$

در نتیجه با توجه به روابط (۳) و (۴) تساوی های زیر را داریم

$$I_{\underline{R}}^B(\theta) = -mk \frac{\partial^2}{\partial \theta^2} \log B(\theta) + m \sum_{j=1}^{k-1} E \left(\frac{\partial^2}{\partial \theta^2} \log \bar{F}_X(R_j; \theta) \right),$$

$$I_{\underline{R}}^A(\theta) = -mk \frac{\partial^2}{\partial \theta^2} \log B(\theta) + \sum_{j=1}^{mk-1} E \left(\frac{\partial^2}{\partial \theta^2} \log \bar{F}_X(R_j; \theta) \right).$$

۳۰ مقایسه دو روش استخراج رکوردها از دیدگاه اطلاع فیشر

با استفاده از روابط فوق قسمت الف نتیجه می‌شود. قسمت ب نیز به طور مشابه و با توجه به اینکه با احتمال یک $R_j < R_{j+1}$ ، ثابت می‌شود.

قضیه ۴: فرض کنید F متعلق به رده C_3 باشد. قرار دهید

$$K(x; \theta) = \frac{\partial^2}{\partial \theta^2} \log F(x; \theta).$$

در این صورت موارد زیر را داریم:

الف: اگر به ازای مقدار ثابت c داشته باشیم $K(x; \theta) = c$ ، آنگاه F دارای خواص $LBEA$ یا $LBGA$ است هرگاه c به ترتیب مثبت، صفر یا منفی باشد.
ب: اگر $K(x; \theta)$ یک تابع کاهشی از x باشد، آنگاه F دارای خاصیت $LBLA$ می‌باشد.

برهان قسمت الف مشابه برهان قضیه ۳ قسمت الف می‌باشد. برای برهان قسمت ب باید از این خاصیت استفاده شود که برای رکوردهای پایین با احتمال یک رابطه‌ی $R'_j > R'_{j+1}$ برقرار است.

قضیه ۵: فرض کنید F متعلق به رده C_4 باشد. در این صورت روابط زیر برقرار هستند:

الف: اطلاع فیشر نهفته در نخستین mk رکورد بالا از این رده در طرح نمونه‌گیری B عبارت است از

$$I_{\bar{R}}^B(\theta) = m \sum_{j=1}^k \Psi_1(j, \circ | \bar{F}_\circ) + m \sum_{j=1}^{k-1} \Psi_2(j, \circ | \bar{F}_\circ),$$

که در آن

$$\Psi_1(j, t | F_\circ) = \int_{-\infty}^{\infty} y^t [f_\circ'(y) - f_\circ''(y)f_\circ(y)] \frac{[-\log F_\circ(y)]^{j-1}}{f_\circ(y)(j-1)!} dy \quad (10)$$

و

$$\Psi_2(j, t | F_\circ) = \int_{-\infty}^{\infty} y^t f_\circ(y) [F_\circ''(y)F_\circ(y) - F_\circ'(y)] \frac{[-\log F_\circ(y)]^{j-1}}{[F_\circ(y)]^2(j-1)!} dy. \quad (11)$$

ب: اطلاع فیشر نهفته در نخستین mk رکورد پایین از این رده در طرح نمونه‌گیری B به صورت زیر است

$$I_{\underline{R}}^B(\theta) = m \sum_{j=1}^k \Psi_1(j, \circ | F_{\circ}) + m \sum_{j=1}^{k-1} \Psi_2(j, \circ | F_{\circ}),$$

که در آن Ψ_1 و Ψ_2 به ترتیب در روابط (۱۰) و (۱۱) تعریف شده‌اند.

قضیه ۶: فرض کنید F متعلق به رده C_5 باشد. در این صورت روابط زیر نتیجه می‌شوند:

الف: اطلاع فیشر نهفته در نخستین mk رکورد بالا از این کلاس در طرح نمونه‌گیری B به صورت زیر است

$$\theta^{\vee} I_{\underline{R}}^B(\theta) = mk + m \sum_{j=1}^k \Psi_1(j, \vee | \bar{F}_{\circ}) + m \sum_{j=1}^{k-1} \Psi_2(j, \vee | \bar{F}_{\circ}),$$

که در آن Ψ_1 و Ψ_2 در روابط (۱۰) و (۱۱) تعریف شده‌اند.

ب: اطلاع فیشر نهفته در نخستین mk رکورد پایین از این رده در طرح نمونه‌گیری B به صورت زیر است

$$\theta^{\vee} I_{\underline{R}}^B(\theta) = mk + m \sum_{j=1}^k \Psi_1(j, \vee | F_{\circ}) + m \sum_{j=1}^{k-1} \Psi_2(j, \vee | F_{\circ}).$$

قضیه ۷: فرض کنید F متعلق به رده C_6 باشد. در این صورت روابط زیر را داریم:
الف: اطلاع فیشر نهفته در نخستین mk رکورد بالا از این رده در طرح نمونه‌گیری B به صورت زیر است

$$\theta^{\vee} I_{\underline{R}}^B(\theta) = mk - m \sum_{j=1}^k \phi_1(j | \bar{F}_{\circ}) + m \sum_{j=1}^{k-1} \phi_2(j | \bar{F}_{\circ}),$$

که در آن

$$\phi_1(j|F_0) = \int_{-\infty}^{\infty} \frac{y \log^{\vee}(y)}{f_0(y)} \left\{ f_0(y)[y f_0''(y) + f_0'(y)] - y f_0'^{\vee}(y) \right\} \times \frac{[-\log F_0(y)]^{j-1}}{(j-1)!} dy \quad (12)$$

و

$$\phi_2(j|F_0) = \int_{-\infty}^{\infty} \frac{y \log^{\vee}(y)}{[F_0(y)]^{\vee}} \left\{ F_0(y)[y F_0''(y) + F_0'(y)] - y F_0'^{\vee}(y) \right\} \times \frac{[-\log F_0(y)]^{j-1}}{(j-1)!} f_0(y) dy. \quad (13)$$

ب: اطلاع فیشر نهفته در نخستین mk رکورد پایین از این رده در طرح نمونه‌گیری B به صورت زیر است

$$\theta^{\vee} I_{\underline{R}'}^B(\theta) = mk - m \sum_{j=1}^k \phi_1(j|F_0) + m \sum_{j=1}^{k-1} \phi_2(j|F_0),$$

که در آن ϕ_1 و ϕ_2 به ترتیب در روابط (۱۲) و (۱۳) تعریف شده‌اند.

تذکر ۳: برهان قضیه‌های ۵، ۶ و ۷ به علت محاسبات جبری طولانی و به منظور کاستن از حجم مقاله حذف شده‌اند.

۴ مثال‌ها

در این بخش، سعی می‌کنیم تا با استفاده از لم‌ها و قضایای ارائه‌شده در بخش ۳، توزیع‌های خاص را مورد بررسی قرار داده و در هر مورد تعیین کنیم که کدامیک از طرح‌های نمونه‌گیری A یا B حاوی اطلاع فیشر بیشتری درباره‌ی پارامتر جامعه می‌باشند.

۱.۴ توزیع نمایی

فرض کنید X_1, X_2, \dots دنباله‌ای از متغیرهای تصادفی $i.i.d.$ با تابع توزیع زیر باشد

$$F_X(x; \theta) = 1 - e^{-\theta x}, \quad x > 0, \theta > 0.$$

در این صورت با توجه به قضیه ۱ این توزیع دارای خواص $UBEA$ و $LBGA$ می باشد. در نتیجه اگر طول عمر یک قطعه الکترونیکی از توزیع نمایی پیروی کند، با توجه به اینکه در انجام آزمایش های مربوط به طول عمر رکوردهای پایین حائز اهمیت هستند، بهتر است که طرح نمونه گیری B مورد استفاده قرار گیرد. در توزیع نمایی برای رکوردهای پایین با استفاده از قضیه ۶ می توان نشان داد

$$\theta^{\nu} I_{R'}^B(\theta) = mk - m \sum_{j=1}^{k-1} \varphi(j), \quad (14)$$

که در آن

$$\varphi(j) = \frac{1}{(j-1)!} \int_0^1 \frac{y}{(1-y)^2} \log^{\nu}(y) (-\log(1-y))^{j-1} dy.$$

همچنین

$$\theta^{\nu} I_{R'}^A(\theta) = mk - \sum_{j=1}^{mk-1} \varphi(j). \quad (15)$$

در جدول ۱، نتایج عددی مربوط به اطلاع فیشر نهفته در رکوردهای پایین مستخرج از توزیع نمایی با میانگین $\frac{1}{\theta}$ در خصوص پارامتر θ آمده است. با توجه به روابط (۱۴) و (۱۵) و جدول ۱، نتایج زیر در مورد رکوردهای پایین مستخرج از توزیع نمایی با میانگین $\frac{1}{\theta}$ در خصوص پارامتر θ عاید می گردند:

۱. در طرح A ، مقدار $\theta^{\nu} I_{R'}^A(\theta)$ از بالا کراندار و تقریباً برابر با $1/144$ می باشد.
۲. به ازای m (k) ثابت، $\theta^{\nu} I_{R'}^B(\theta)$ بر حسب k (m) افزایشی است.
۳. به ازای $mk = c_0$ ثابت، با افزایش m $\theta^{\nu} I_{R'}^B(\theta)$ افزایش می یابد. مثلاً اطلاع فیشر نهفته در نخستین ۱۰ رکورد پایین در طرح B با مرتبه ی تکرار ۲ برابر $2/289$ ولی با مرتبه ی تکرار ۵ برابر $5/571$ می باشد.

۲.۴ توزیع توانی

فرض کنید X_1, X_2, \dots دنباله ای از متغیرهای تصادفی $i.i.d.$ با تابع توزیع زیر باشد

$$F_X(x; \theta) = x^{\theta}, \quad 0 < x < 1, \theta > 0.$$

۳۴ مقایسه دو روش استخراج رکوردها از دیدگاه اطلاع فشر

در این صورت می توان نوشت $F_X(x; \theta) = e^{-\theta \log(\frac{1}{x})}$. در نتیجه بنا به قضیه ی ۲ این توزیع دارای خواص $UBGA$ و $LBEA$ می باشد.

۳.۴ توزیع بر نوع دوازده

خانواده توزیع های بر در سال ۱۹۴۲ توسط بر معرفی شد، توزیع بر نوع دوازده معروف ترین عضو این خانواده می باشد. صورت تابعی توزیع بر نوع دوازده برای پارامترهای مثبت α و β عبارت است از:

$$F(x; \theta) = 1 - (1 + x^\alpha)^{-\beta}, \quad x \geq 0 \quad (16)$$

است (جزئیات بیشتر را می توانید در تیم (۱۹۹۳) ببینید). این توزیع در عمل بسیار پر کاربرد است که از جمله می توان به نقش آن در بخش هایی از کنترل کیفیت، طول عمر یا مدل بندی زمان نقص و درآمد اشاره نمود. در این مقاله نماد $B_{12}(\alpha, \beta)$ را برای توزیع (۱۶) بکار برده و دو حالت زیر را در نظر می گیریم:

الف: α معلوم باشد

در این صورت با در نظر گرفتن $\beta = \theta$ ، این توزیع متعلق به رده C_1 می باشد که طبق قضیه ی ۱، دارای خواص $UBEA$ و $LBGA$ است.

ب: اگر β معلوم باشد

در این حالت با فرض $\beta = 1$ و $\alpha = \theta$ ، به مقایسه ی دو طرح A و B در توزیع $B_{12}(\theta, 1)$ می پردازیم.

لم ۸: فرض کنید X_1, X_2, \dots دنباله ای از متغیرهای تصادفی *i.i.d.* با توزیع $B_{12}(\theta, 1)$ باشند، در این صورت اطلاع فشر نهفته در نخستین mk رکورد بالا در توزیع $B_{12}(\theta, 1)$ به صورت زیر است:

الف: در طرح نمونه گیری B ،

$$\theta^\gamma I_{\underline{R}}^B(\theta) = mk + m \sum_{j=1}^{k-1} \xi(j) + 2m\xi(k). \quad (17)$$

که در آن به ازای $j = 1, \dots, k$

$$\xi(j) = \frac{1}{(j-1)!} \int_0^{\infty} \frac{y \log^2(y)}{(1+y)^4} (\log(1+y))^{j-1} dy. \quad (18)$$

ب: در طرح نمونه گیری A ,

$$\theta^2 I_{\underline{R}}^A(\theta) = mk + \sum_{j=1}^{mk-1} \xi(j) + 2\xi(mk), \quad (19)$$

که در آن $\xi(j)$ در رابطه‌ی (۱۸) تعریف شده است.

برهان با استفاده از قضیه‌ی ۷ قسمت الف و با توجه به اینکه در این توزیع $F_0(y) = 1 - (1+y)^{-1}$ می‌توان با انجام محاسبات جبری به نتیجه‌ی مورد نظر دست یافت.

لم ۹: تحت مفروضات لم ۸، اطلاع فیشر نهفته در نخستین mk رکورد پایین در توزیع $B(1, \theta)$ به صورت زیر است:

الف: در طرح نمونه گیری B ، داریم

$$\theta^2 I_{\underline{R}}^B(\theta) = mk + m \sum_{j=1}^{k-1} \psi(j) + 2m\psi(k).$$

که در آن به ازای $j = 1, \dots, k$

$$\psi(j) = \frac{1}{(j-1)!} \int_0^{\infty} \frac{y \log^2(y)}{(1+y)^4} \left(\log\left(1 + \frac{1}{y}\right)\right)^{j-1} dy. \quad (20)$$

ب: در طرح نمونه گیری A ,

$$\theta^2 I_{\underline{R}}^A(\theta) = mk + \sum_{j=1}^{mk-1} \psi(j) + 2\psi(mk),$$

که در آن $\psi(j)$ در رابطه‌ی (۲۰) تعریف شده است.

برهان با استفاده از قضیه‌ی ۷ قسمت ب و انجام محاسبات جبری، نتیجه‌ی مورد نظر عاید می‌گردد.

هرگاه توزیع جامعه مقارن باشد، با توجه به نتایج ارائه شده در احمدی و ارقامی (۲۰۰۱)، میزان اطلاع فیشر نهفته در رکوردهای بالا و پایین با هم برابر است.

۳۶ مقایسه دو روش استخراج رکوردها از دیدگاه اطلاع فشر

علیرغم اینکه توزیع $B_{12}(\theta, 1)$ نامتقارن است، در لم ۱۰ نشان خواهیم داد که عکس مطلب فوق برقرار نیست.

لم ۱۰: تحت مفروضات لم ۸، اطلاع فشر نهفته در نخستین mk رکورد بالا در هر دو طرح نمونه‌گیری A و B برابر با mk رکورد پایین است. به عبارتی

$$I_{\underline{R}}^A(\theta) = I_{\underline{R}'}^A(\theta),$$

$$I_{\underline{R}}^B(\theta) = I_{\underline{R}'}^B(\theta).$$

برهان مشابه برهان ارائه شده در رزمخواه و همکاران (۱۳۸۲).

قضیه ۸: تحت مفروضات لم ۸، اطلاع فشر نهفته در نخستین mk رکورد بالا (پایین) توزیع $B_{12}(\theta, 1)$ در طرح نمونه‌گیری B بیشتر از طرح A می‌باشد. یعنی این توزیع دارای خواص $UBGA$ و $LBGA$ می‌باشد.

برهان برهان را برای رکوردهای بالا ارائه می‌دهیم؛ با توجه به لم ۱۰، نتیجه برای رکوردهای پایین نیز برقرار خواهد بود. با توجه به (۱۷) و (۱۹) باید ثابت کنیم

$$m \sum_{j=1}^{k-1} \xi(j) + 2m\xi(k) > \sum_{j=1}^{mk-1} \xi(j) + 2\xi(mk). \quad (21)$$

از طرفی با استفاده از رابطه‌ی (۱۸) و محاسبات عددی، نتیجه می‌شود که $\xi(j)$ به ازای $3 \leq j$ ، یک تابع نزولی از j می‌باشد. مقادیر $\xi(j)$ به ازای $1, \dots, 7$ در جدول ۲ آمده است. با توجه به جدول ۲، ملاحظه می‌شود که $\xi(1) > \xi(5) > \dots > \xi(4) > \xi(3) > \xi(2) < \xi(1)$. همچنین $\xi(1) > \xi(5) > \dots$

رابطه‌ی (۲۱) به ازای مقادیر مختلف k ، به صورت زیر ثابت می‌شود.

الف: به ازای $k = 2$ ، نامساوی (۲۱) به صورت زیر تغییر می‌یابد.

$$m\xi(1) + 2m\xi(2) > \sum_{j=1}^{2m-1} \xi(j) + 2\xi(2m). \quad (22)$$

اکنون به روش استقرا نسبت به m نشان می‌دهیم که (۲۲) برقرار است.
گام اول: فرض کنید $m = 2$ در این صورت کفایت رابطه‌ی زیر برقرار باشد.

$$\xi(1) + 3\xi(2) > \xi(3) + 2\xi(4).$$

درستی رابطه‌ی فوق از جدول ۲ نتیجه می‌شود.

گام دوم: فرض کنید (۲۲) به ازای $m = m_0$ برقرار باشد. با فرض $m = m_0 + 1$ و بعد از ساده نمودن عبارات جبری کفایت رابطه‌ی زیر ثابت شود.

$$\xi(1) + 2\xi(2) > \xi(2m_0 + 1) + 2\xi(2m_0 + 2) - \xi(2m_0).$$

نابرابری اخیر با توجه به اینکه $\xi(1) > \xi(2m_0 + 1)$ ، $\xi(2) > \xi(2m_0 + 2)$ و $\xi(2m_0) > 0$ بدیهی است.

ب: در حالت $k = 3$ برقراری رابطه‌ی (۲۱) مشابه الف به روش استقرا ثابت می‌شود.

ج: با فرض $k \geq 4$ و با توجه به رفتار تابع $\xi(\cdot)$ ، نابرابری‌های $m \sum_{j=1}^{k-1} \xi(j) > \sum_{j=mk-m+1}^{mk} \xi(j)$ و $m \sum_{j=1}^{k-1} \xi(j) > \sum_{j=1}^{mk-m} \xi(j)$ نتیجه می‌شوند.
 بنابراین

$$m \sum_{j=1}^{k-1} \xi(j) + 2m\xi(k) > \sum_{j=1}^{mk-m} \xi(j) + 2 \sum_{j=mk-m+1}^{mk-1} \xi(j) + 2\xi(mk).$$

نابرابری اخیر، رابطه‌ی (۲۱) را به طور قوی تأیید می‌کند.

۴.۴ توزیع وایبول

فرض کنید متغیر تصادفی X دارای توزیع وایبول با تابع توزیع

$$F(x; \alpha, \lambda) = 1 - e^{-\lambda x^\alpha}, \quad x > 0 \quad (23)$$

باشد. در این توزیع اگر پارامتر شکل α معلوم باشد، آنگاه F متعلق به رده C_1 می‌باشد که خواص آن قبلاً بررسی شده است. اکنون با فرض معلوم بودن پارامتر λ

۳۸ مقایسه دو روش استخراج رکوردها از دیدگاه اطلاع فیشر

بدون از دست دادن کلیت مسأله قرار می‌دهیم $\lambda = 1$ و $\alpha = \theta$. در این حالت رابطه‌ی (۲۲) به صورت زیر خواهد بود

$$F(x; \theta) = 1 - e^{-x^\theta}, \quad x > 0. \quad (24)$$

لم ۱۱: فرض کنید X_1, X_2, \dots دنباله‌ای از متغیرهای تصادفی *i.i.d.* با تابع توزیع (۲۳) باشد، در این صورت اطلاع فیشر نهفته در نخستین mk رکورد بالا به صورت زیر است:

الف: در طرح نمونه‌گیری B ,

$$\theta^2 I_{\underline{R}}^B(\theta) = mk + m \frac{\Gamma''(k+1)}{(k-1)!}.$$

ب: در طرح نمونه‌گیری A ,

$$\theta^2 I_{\underline{R}}^A(\theta) = mk + \frac{\Gamma''(mk+1)}{(mk-1)!}.$$

در قسمت‌های الف و ب، $\Gamma''(\cdot)$ مشتق مرتبه‌ی دوم تابع گامای کامل می‌باشد.

برهان با توجه به قضیه‌ی ۷ قسمت الف و با توجه به اینکه در این توزیع $F_0(y) = 1 - e^{-y}$ می‌توان با به‌کارگرفتن محاسبات جبری نتیجه‌ی مورد نظر را به دست آورد.

در جدول ۳ نتایج عددی مربوط به $\theta^2 I_{\underline{R}}^B(\theta)$ و $\theta^2 I_{\underline{R}}^A(\theta)$ برای توزیع وایبول در خصوص پارامتر شکل آمده است. با توجه به جدول ۳، در مورد رکوردهای بالای توزیع وایبول در خصوص پارامتر شکل نتایج زیر عاید می‌شوند:

۱. توزیع وایبول نسبت به پارامتر شکل خود، دارای خاصیت *UBLA* می‌باشد.
۲. با افزایش مرتبه‌ی تکرار m اطلاع فیشر نهفته در یک تعداد مشخص از رکوردهای بالا در طرح B ، کاهش می‌یابد. مثلاً اطلاع فیشر نهفته در نخستین ۸ رکورد بالا با مرتبه‌ی تکرار ۲ برابر $27/92$ ولی با مرتبه‌ی تکرار ۴ برابر $17/97$ می‌باشد.

جدول ۱: مقادیر عددی $\theta^2 I_{R'}^A(\theta)$ و $\theta^2 I_{R'}^B(\theta)$ برای توزیع نمایی (پارامتر مقیاس).

m	k=۲		k=۳		k=۵	
	طرح A	طرح B	طرح A	طرح B	طرح A	طرح B
۲	۱/۱۴۲	۲/۲۲۸	۱/۱۴۴	۲/۲۷۴	۱/۱۴۴	۲/۲۸۹
۳	۱/۱۴۴	۳/۳۴۳	۱/۱۴۴	۳/۴۱۱	۱/۱۴۴	۳/۴۳۳
۴	۱/۱۴۴	۴/۴۵۷	۱/۱۴۴	۴/۵۴۸	۱/۱۴۴	۴/۵۷۷
۵	۱/۱۴۴	۵/۵۷۱	۱/۱۴۴	۵/۶۸۵	۱/۱۴۴	۵/۷۲۱
۸	۱/۱۴۴	۸/۹۱۴	۱/۱۴۴	۹/۰۹۶	۱/۱۴۴	۹/۱۵۴
۱۰	۱/۱۴۴	۱۱/۱۴۲	۱/۱۴۴	۱۱/۳۷۱	۱/۱۴۴	۱۱/۴۴۳

جدول ۲: مقادیر $\xi(j)$ به ازای $j = ۱, \dots, ۷$

j	۱	۲	۳	۴	۵	۶	۷
$\xi(j)$	۰/۲۱۵	۰/۲۴۷	۰/۲۸۰	۰/۲۶۰	۰/۲۰۹	۰/۱۵۳	۰/۱۰۵

جدول ۳: مقادیر عددی $\theta^2 I_{R'}^A(\theta)$ و $\theta^2 I_{R'}^B(\theta)$ برای توزیع وایبول (پارامتر شکل).

m	k=۲		k=۳		k=۴	
	طرح A	طرح B	طرح A	طرح B	طرح A	طرح B
۲	۱۳/۹۶	۸/۹۹	۲۷/۹۷	۱۷/۱۷	۴۵/۶۰	۲۷/۹۲
۳	۲۷/۹۷	۱۳/۴۸	۵۵/۵۸	۲۵/۷۶	۸۹/۵۳	۴۱/۸۸
۴	۴۵/۶۰	۱۷/۹۷	۸۹/۵۳	۳۴/۳۴	۱۴۲/۷۲	۵۵/۸۴
۵	۶۶/۲۶۰	۲۲/۴۷	۱۲۸/۶۷	۴۲/۹۳	۲۰۳/۴۵	۶۹/۷۹
۸	۱۴۲/۷۲	۳۵/۹۴	۲۷۰/۵۵	۶۸/۶۸	۴۲۰/۸۰	۱۱۱/۶۷
۱۰	۲۰۳/۴۵	۴۴/۹۳	۳۸۱/۴۲	۸۵/۸۵	۵۸۸/۹۸	۱۳۹/۵۹

۴۰. مقایسه دو روش استخراج رکوردها از دیدگاه اطلاع فشر

لم ۱۲: تحت مفروضات لم ۱۰، اطلاع فشر نهفته در نخستین mk رکورد پایین در توزیع وایبول درباره‌ی پارامتر شکل به صورت زیر است:

الف: در طرح نمونه‌گیری B ,

$$\theta^2 I_{R'}^B(\theta) = mk - m \sum_{j=1}^k \beta_1(j) + m \sum_{j=1}^{k-1} \beta_2(j),$$

که در آن

$$\beta_1(j) = \frac{1}{(j-1)!} \int_0^\infty ye^{-y} \log^2(y) [-\log(1-e^{-y})]^{j-1} dy \quad (25)$$

و

$$\beta_2(j) = \frac{1}{(j-1)!} \int_0^\infty ye^{-2y} \log^2(y) [-\log(1-e^{-y})]^{j-1} \frac{1-y-e^{-y}}{(1-e^{-y})^2} dy. \quad (26)$$

ب: در طرح نمونه‌گیری A ,

$$\theta^2 I_{R'}^A(\theta) = mk - \sum_{j=1}^{mk} \beta_1(j) + \sum_{j=1}^{mk-1} \beta_2(j),$$

که در آن $\beta_1(j)$ و $\beta_2(j)$ به ترتیب در روابط (۲۳) و (۲۴) معرفی شده‌اند. برهان با توجه به قضیه‌ی ۷ قسمت ب و انجام محاسبات جبری، نتیجه‌ی مورد نظر عاید می‌گردد.

در جدول ۴ نتایج عددی مربوط به $\theta^2 I_{R'}^A(\theta)$ و $\theta^2 I_{R'}^B(\theta)$ برای توزیع وایبول در خصوص پارامتر شکل آمده است. با توجه به جدول ۴ در مورد رکوردهای پایین مستخرج از توزیع وایبول نتایج زیر عاید می‌گردند:

۱. توزیع وایبول نسبت به پارامتر شکل خود، دارای خاصیت $LBGA$ می‌باشد.
۲. با افزایش مرتبه‌ی تکرار m اطلاع فشر نهفته در یک تعداد مشخص از رکوردها در طرح B افزایش می‌یابد. مثلاً اطلاع فشر نهفته در نخستین ۶ رکورد پایین با مرتبه‌ی تکرار ۲ برابر ۸/۴۹ ولی با مرتبه‌ی تکرار ۳ برابر ۹/۱۴ می‌باشد.

فرع ۱: با توجه به جدول ۳ و جدول ۴ در مورد رکوردهای بالا و پایین مستخرج از توزیع وایبول ملاحظه می‌شود که اطلاع فشر نهفته در رکوردهای بالا در هر دو طرح بیشتر از اطلاع فشر نهفته در رکوردهای پایین است.

جدول ۴: مقادیر عددی $\theta^2 I_R^A(\theta)$ و $\theta^2 I_R^B(\theta)$ برای توزیع وایبول (پارامتر شکل).

m	k=۲		k=۳		k=۴	
	طرح A	طرح B	طرح A	طرح B	طرح A	طرح B
۲	۵/۳۸	۶/۰۹	۷/۵۲	۸/۴۹	۹/۵۶	۱۰/۷۶
۳	۷/۵۲	۹/۱۴	۱۰/۵۷	۱۲/۷۴	۱۳/۵۸	۱۶/۱۵
۴	۹/۵۶	۱۲/۱۹	۱۳/۵۸	۱۶/۹۸	۱۷/۵۸	۲۱/۵۳
۵	۱۱/۵۸	۱۵/۲۳	۱۶/۵۸	۲۱/۲۳		۲۶/۹۱
۸	۱۷/۸۵	۲۴/۳۷		۳۳/۹۶		۴۳/۰۶

جدول ۵: دسته‌بندی چند توزیع یک‌پارامتری از منظر طرح A و B.

نام یا تابع توزیع	رکوردهای بالا	رکوردهای پایین
$F \in C_1$	UBEA	LBGA
$F \in C_2$	UBGA	LBEA
$\Gamma(\alpha, \beta)$, معلوم α		
$\alpha = 1$	UBEA	LBGA
$0 < \alpha < 1$	UBLA	LBLA
$\alpha > 1$	UBGA	LBGA
$B_{12}(\theta, 1)$	UBGA	LBGA
$1 - \exp(-\alpha x^\theta), x > 0$		
معلوم α	UBLA	LBGA
معلوم θ	UBEA	LBGA
$\exp(-x^{-\theta}), x > 0$	UBGA	LBLA
$\exp(-e^{-\theta(x-\mu)})$		
معلوم μ	UBGA	LBLA
معلوم θ	UBGA	LBEA
$(1 + e^{-\beta(x-\alpha)})^{-1}$		
معلوم α	UBGA	LBGA
معلوم β	UBGA	LBGA

۵ بحث و نتیجه گیری

در این مقاله دو نوع طرح نمونه‌گیری استخراج داده‌های رکوردی از دیدگاه اطلاع فیشر مورد مقایسه قرار گرفته‌اند. نتایج برای چند مدل معروف از توزیع‌های طول عمر در بخش ۴ ارائه شد. به طور مشابه می‌توان خواص توزیع‌های بیشتری را بررسی نمود. در جدول ۵ چند توزیع مهم آماری یک‌پارامتری از منظر دو طرح A و B دسته‌بندی شده‌اند. همانطوری که از جدول ۵ مشاهده می‌شود، اطلاع فیشر نهفته در آماره‌های رکوردی در دو نوع طرح بستگی به توزیع جامعه دارد. مثلاً برای رکوردهای بالا، در توزیع نمایی (پارامتر مقیاس) میزان اطلاع در هر دو طرح مساوی، در توزیع بر نوع دوازده (پارامتر شکل) در طرح B بیشتر از طرح A و در توزیع وایبول (پارامتر شکل) در طرح A بیشتر از طرح B می‌باشد.

تقدیر و تشکر

نویسندگان از پیشنهادات ارزنده‌ی داوران محترم که باعث اصلاحات سازنده‌ای در این مقاله شد، کمال تشکر و سپاسگزاری را دارند.

مراجع

- احمدی، ج.، رزمخواه، م. (۱۳۸۱)، اطلاع در توزیع $BURR XII$ ، مجموعه مقالات ششمین کنفرانس آمار، جلد ۱، ۲۶-۴۰.
- رزمخواه، م. (۱۳۸۳)، اطلاع فیشر نهفته در آماره‌های رکوردی، پایان‌نامه‌ی کارشناسی ارشد، دانشگاه فردوسی مشهد.
- رزمخواه، م.، خطیب، ب. و احمدی، ج. (۱۳۸۲)، اطلاع فیشر نهفته در آماره‌های رکوردی توزیع $BURR XII$ ، اندیشه آماری، سال هشتم، شماره‌ی دوم، ۱۷-۲۵.

مصطفی رزمخواه، جعفر احمدی، بهاره خطیب آستانه ۴۳

Ahmadi, J. (2000), *Record Values, Theory and Applications*, Ph.D. Thesis Ferdowsi University of Mashhad, Iran.

Ahmadi, J. and Arghami, N.R. (2001), *On the Fisher Information in Record Values*, *Metrika* **53**, 195–206.

Ahmadi, J. and Arghami, N.R. (2003), *Comparing the Fisher Information in Record Values and iid Observations*, *Statistics* **37(5)**, 435-441.

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

Balakrishnan, N. and Stepanov, A. (2006), *On the Fisher Information in Record Data*, *Statistics and Probability Letters*, **76**, 537-545.

Burr, I. W. (1942), *Cumulative Frequency Functions*, *Annals Math. Statist.*, **13**, 215-232.

Chandler, K. N. (1952), *The Distribution and Frequency of Record Values*, *J. Roy. Statist. Soc. Ser. B* **14**, 220-228.

Glick, N. (1978), *Breaking Record and Breaking Boards*, *American Mathematical Monthly*, **85**, 2-26.

Gulati, S. and Padgett, W. J. (2003), *Parametric and Nonparametric Inference from Record-breaking Data*, *Lecture Note in Statistics*, V. **172**, Springer, New York.

Hofmann, G., Balakrishnan, N. and Ahmadi, J. (2005), *Characterization of Hazard Function Factorization by Fisher Information in Minima and Upper Record Values*, *Statistics and Probability Letters*, **72**, 51-57.

۴۴ مقایسه دو روش استخراج رکوردها از دیدگاه اطلاع فیشر

Hofmann, G., Nagaraja, H. N. (2003), *Fisher Information in Record Data*,
Metrika, **57**, 177-193.

Lehmann, E. L. and Casella, G. (1998), *Theory of Point Estimation - 2nd
Edition*, Springer, New York.

Nagaraja, H. N. (1988), *Record Value and Related Statistics a Review*,
Comm. Statist. Theory Methods **17**, 2223-2238.

Nevzorov, V. B. (1987), *Records*, Theory Probab. Appl. **32**, 201-228.

Nevzorov, V. B. (2001), *Records: Mathematical Theory*, Amer. Math.
Soc., Providence, Rhode Island.

Samaniego, F. J. and Whitaker L. R. (1988), *On Estimating Population
Characteristics from Record-breaking Observations*, II: Nonparametric
Results, Naval Research Logistics Quarterly, **35**, 221-236.

Tim, R. L. Fry (1993), *Univariate and Multivariate Burr Distributions*,
Pakistan J. Statist., **A**, 1-24.