مجله علوم آماری، بهار و تابستان ۱۳۹۰

جلد ۵، شماره ۱، ص ۲۳-۳۹

بازه اطمینان ناپارامتری با ضریب اطمینان دقیق برای چندک‌های جامعه براساس نمونه‌گیری مجموعه رتبه‌دار

الهام زمان زاده، جعفر احمدی

گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

تاریخ دریافت: ۱۳۹۰/۱/۳۰ تاریخ آخرین بازنگری: ۱۳۸۹/۵/۱۶

چکیده : در این مقاله ضمن معرفی روش نمونه‌گیری مجموعه رتبه‌دار، شیوه ساختن بازه‌های اطمینان برای چندک‌های جامعه براساس آماره‌های ترتیبی حاصل از نمونه‌گیری مجموعه رتبه‌دار، ارائه می‌شود. چون ضریب اطمینان حاصل تابعی پله‌ای است، امکان رسیدن به ضریب اطمینان دقیق از پیش تعیین شده را مشکل می‌سازد. برای این منظور روش جدیدی ارائه می‌شود که با استفاده از آن می‌توان بازه اطمینان بهینه را بدست آورد. در انتها نتایج به دست آمده با نتایج حاصل از سایر روش‌ها مقایسه می‌شود.

واژه‌های کلیدی : آ ماره‌های ترتیبی، ضریب اطمینان، چندک، آزاد توزیع، نمونه‌گیری مجموعه رتبه‌دار.

آدرس الکترونیک مسئول مقاله: ahmadi-j@um.ac.ir

کد موضوع‌بندی ریاضی (۲۰۰۰): ۶۲G۳۰، ۶۲G۱۵، ۶۲D۰۵، ۶۲F۰۷

۱ مقدمه

فرض کنید X یک متغیر تصادفی پیوسته با تابع توزیع $F(x)$ باشد، آنگاه چندک مرتبه pام ($0 < p < 1$) جامعه به صورت

$$\xi_p = F^{-1}(p) = \inf\{x : F(x) \geq p\}$$

تعریف می‌شود. با توجه به نقش مهم چندک و توابعی از آن در استنباط آماری ناپارامتری و پارامتری از جمله برآورد شاخص‌های مرکزی و پراکندگی، شبیه سازی مونت کارلو، آزمون‌های نیکویی برازش و توصیف مدل‌های آماری، مطالعه پارامتر چندک همواره به عنوان یکی از مهمترین موضوعات مطرح مورد علاقه در میان محققان آماری بوده است.

فرض کنید $X_{1:n}$، ...، $X_{n:n}$ آماره‌های ترتیبی حاصل از یک نمونه تصادفی ساده با اندازه n از توزیع $F(x)$ باشند، آنگاه $[X_{r:n}, X_{s:n}]$ یک بازهٔ اطمینان ناپارامتری $100(1-\alpha)\%$ برای چندک مرتبه pام، یعنی ξ_p، است، اگر s و rایی وجود داشته باشند، به طوری که

$$P(X_{r:n} \leq \xi_p \leq X_{s:n}) = 1 - \alpha.$$

تاکنون تحقیقات زیادی دربارهٔ نحوه ساختن بازه اطمینان برای چندک‌ها براساس انواع داده‌های ترتیبی انجام شده است که در این خصوص می‌توان به استین برگر و داویس(۱۹۸۵)، هاتسون(۱۹۹۹)، گول باد (۲۰۰۱)، احمدی و ارقامی (۲۰۰۳)، بالاکریشنان و هان (۲۰۰۷) و بونتر و کرامر (۲۰۱۱) اشاره کرد.

در سال‌های اخیر ساختن بازه اطمینان برای چندک‌ها براساس روش نمونه‌گیری مجموعه رتبه‌دار[۱] (RSS) مورد توجه قرار گرفته است. این روش نمونه‌گیری اولین بار توسط مک اینتایر (۱۹۵۲) برای افزایش دقت برآورد میانگین محصول مراتع، بدون افزایش اریبی آن معرفی شد. تاکاهاشی و واکیموتو (۱۹۶۸) از جمله افرادی بودند که توانستند نظریه روش نمونه‌گیری مجموعه رتبه‌دار را به طور کامل و مبسوط بیان کنند. پس از آن استفاده از نمونه‌گیری مجموعه رتبه‌دار در استنباط‌های

---
۱ Ranked Set Sampling

پارامتری و ناپارامتری رونق گرفت و به دنبال آن ساختن بازه اطمینان برای چندک‌ها براساس مجموعه رتبه‌دار مورد مطالعه قرار گرفت. در این راستا آماره‌های ترتیبی مجموعه رتبه‌دار معرفی شد. مشکل به دست آوردن بازه اطمینان برای ξ_p، در داده‌های مبتنی بر RSS این است که در اغلب موارد، رسیدن به سطح اطمینان $(۱-\alpha)۱۰۰\%$ به طور دقیق امکان پذیر نیست. برای حل این مشکل محققین روش‌های متفاوتی را پیشنهاد دادند. چن (۲۰۰۰) با استفاده از قضیه حد مرکزی، کران‌های تقریبی برای بازه اطمینان ξ_p به دست آورد به طوری که دارای ضریب اطمینان تقریبی $(۱-\alpha)۱۰۰\%$ بود. هرچند نتایج او نشان می‌داد این تقریب در برخی موارد از دقت کافی برخوردار نیست. بالاکریشنان و لی (۲۰۰۶) با استفاده از تابع توزیع آماره‌های ترتیبی مجموعه رتبه‌دار، بازه اطمینان برای ξ_p با ضریب اطمینان حداقل $(۱-\alpha)۱۰۰\%$ پیشنهاد دادند. ازترک و دشپاند (۲۰۰۶) ترکیبی محدب از دو بازه اطمینان در اندازه‌های کوچکتر و بزرگتر از $(۱-\alpha)$ برای ξ_p پیشنهاد کردند که به طور تقریبی دارای ضریب اطمینان $(۱-\alpha)۱۰۰\%$ بود.

در ادامه ابتدا به طور مختصر توزیع آماره‌های ترتیبی مجموعه رتبه‌دار معرفی می‌شود. سپس در بخش ۳ بازه اطمینان برای چندک‌ها براساس آماره‌های ترتیبی مجموعه رتبه‌دار ارائه می‌شود. در بخش ۴ روش جدیدی مطرح می‌شود که با استفاده از آن می‌توان برخلاف روش‌های موجود به مقدار دقیق ضریب اطمینان از پیش تعیین شده دست یافت. در بخش ۵ الگوریتمی معرفی می‌شود که براساس آن می‌توان بازه اطمینان بهینه را تعیین نمود. در بخش ۶ به مقایسه روش پیشنهادی با سایر روش‌های موجود پرداخته خواهد شد. در انتها بخش ۷ به بحث و نتیجه‌گیری اختصاص خواهد یافت.

۲ طرح نمونه‌گیری و ویژگی‌های توزیعی

در این بخش روش نمونه‌گیری مجموعه رتبه‌دار متعادل به اختصار توضیح داده می‌شود. استخراج نمونه‌ای با اندازه n به روش نمونه‌گیری مجموعه رتبه‌دار طی مراحل زیر انجام می‌شود:

۱. نمونه‌ای تصادفی با اندازه $n^۲$ از جامعه استخراج شود.
۲. $n^۲$ واحد انتخاب شده به طور تصادفی به n مجموعه n تایی تقسیم شود.

نمونه	قبل از مرتب سازی	بعد از مرتب سازی
اول	$X_{۱,۱}, X_{۱,۲}, \ldots, X_{۱,n}$	$\underline{X_{[۱,۱]}}, X_{[۱,۲]}, \ldots, X_{[۱,n]}$
دوم	$X_{۲,۱}, X_{۲,۲}, \ldots, X_{۲,n}$	$X_{[۲,۱]}, \underline{X_{[۲,۲]}}, \ldots, X_{[۲,n]}$
$\vdots$	$\vdots$	$\vdots$
nام	$X_{n,۱}, X_{n,۲}, \ldots, X_{n,n}$	$X_{[n,۱]}, X_{[n,۲]}, \ldots, \underline{X_{[n,n]}}$

توجه شود که در طرح نمونه‌گیری مجموعه رتبه‌دار نامتعادل، لزومی ندارد اندازه نمونه‌های هر سطر برابر باشند. در این مقاله طرح متعادل در نظر گرفته شده است.

۳. واحدهای هر مجموعه با روشی به غیر از اندازه‌گیری دقیق (مانند اندازه‌گیری به طور شهودی یا با استفاده از متغیر کمکی) مرتب شود.

۴. اولین کوچکترین واحد از مجموعه اول، دومین کوچکترین واحد از مجموعه دوم، به همین ترتیب تا nامین کوچکترین واحد از مجموعه nام را انتخاب کرده و به طور دقیق اندازه‌گیری شود.

در واقع، نمونه حاصل از طرح نمونه‌گیری مجموعه رتبه‌دار، نمونه‌ای به صورت $\{X_{[i,i]}, i = ۱, \ldots, n\}$ است، که در آن n تعداد واحدهای هر نمونه است. نماد $[\cdot]$ به جای $(\cdot)$ نشان می‌دهد که در رتبه‌بندی مشاهدات ممکن است خطا رخ داده باشد، در غیر این صورت رتبه‌بندی دقیق انجام شده است و مشاهدات به صورت $\{X_{(i,i)}, i = ۱, \ldots, n\}$ نشان داده می‌شود. شایان ذکر است که مشاهدات حاصل از طرح RSS، مستقل از هم هستند، در حقیقت متغیر $X_{(i,i)}$ در سطر iام، هم توزیع با iامین آماره ترتیبی در نمونه‌ای تصادفی با اندازه n است، بنابراین تابع توزیع احتمال آن به صورت

$$F_{X_{(i,i)}}(x) = \sum_{\ell=i}^{n} \binom{n}{\ell} \{F(x)\}^{\ell} \{۱ - F(x)\}^{n-\ell} \qquad (۱)$$

است، که در آن $F(\cdot)$ تابع توزیع احتمال جامعه نمونه‌گیری است.

باید توجه داشت به دلیل این که $\{X_{(i,i)}, i = ۱,\dots,n\}$ در نمونه‌گیری مجموعه رتبه‌دار مستقل هستند، اما هیچ ترتیبی بین آنها وجود ندارد. لذا برای به دست آوردن بازه اطمینان برای چندک‌ها براساس این نمونه‌گیری لازم است ابتدا $X_{(i,i)}$ها از کوچک به بزرگ به صورت $X_{۱:n}^{ORSS} < X_{۲:n}^{ORSS} < \dots < X_{n:n}^{ORSS}$، مرتب شوند. با استفاده از تابع توزیع آماره‌های ترتیبی حاصل از نمونه‌های مستقل و غیر هم توزیع و رابطه (۱)، تابع توزیع $X_{r:n}^{ORSS}$ به صورت

$$\begin{aligned} F_{X_{r:n}^{ORSS}}(x) &= P(X_{r:n}^{ORSS} \le x), \quad x \in R \\ &= \sum_{i=r}^{n} P(\text{دقیقا } i \text{ تا از } X_{(ii)} \text{ ها کوچکتر یا مساوی } x \text{ باشند}) \\ &= \sum_{i=r}^{n} \sum_{S_i^{[n]}} \{\prod_{l=۱}^{i} F_{j_l:n}(x) \prod_{l=i+۱}^{n} [۱ - F_{j_l:n}(x)]\}, \end{aligned} \quad (۲)$$

حاصل می‌شود، که در آن نماد $\sum_{S_i^{[n]}}$ نشان‌دهنده جمع روی تمام جایگشت‌های $(j_۱, j_۲, \dots, j_n)$ از $(۱, ۲, \dots, n)$ زمانی که $j_۱ < \dots < j_i$ و $j_{i+۱} < \dots < j_n$ است. برای جزئیات اثبات رابطهٔ (۲) به دیوید و ناگاراجا (۲۰۰۳) مراجعه شود.

۳ بازه اطمینان براساس آماره‌های ترتیبی مجموعه رتبه‌دار

برای ساختن یک بازه اطمینان برای ξ_p براساس آماره‌های ترتیبی مجموعه رتبه‌دار، ابتدا احتمال پوشش بازه $[X_{r:n}^{ORSS}, X_{s:n}^{ORSS}]$ با استفاده از این واقعیت که $F(X_{r:n}^{ORSS}) \overset{d}{=} U_{r:n}^{ORSS}$، به صورت زیر به دست آورده می‌شود. در اینجا $X \overset{d}{=} Y$، نماد هم توزیع بودن دو متغیر تصادفی X و Y است، بعلاوه، $U_{r:n}^{ORSS}$، rامین آماره ترتیبی حاصل از مجموعه رتبه‌دار از جامعه یکنواخت $(۰, ۱)$ است. چون برای $r < s$ آماره $X_{r:n}^{ORSS}$ با احتمال یک از $X_{s:n}^{ORSS}$ کوچکتر و هر دو متغیر تصادفی پیوسته هستند، می‌توان نوشت

$$\begin{aligned} P(X_{r:n}^{ORSS} \le \xi_p) &= P(X_{r:n}^{ORSS} \le \xi_p, X_{s:n}^{ORSS} < \xi_p) \\ &+ P(X_{r:n}^{ORSS} \le \xi_p, X_{s:n}^{ORSS} \ge \xi_p) \\ &= P(X_{s:n}^{ORSS} < \xi_p) + P(X_{r:n}^{ORSS} \le \xi_p \le X_{s:n}^{ORSS}). \end{aligned}$$

در نتیجه داریم

$$\begin{aligned} P(X_{r:n}^{ORSS} \leq \xi_p \leq X_{s:n}^{ORSS}) &= P(X_{r:n}^{ORSS} \leq F^{-۱}(p) \leq X_{s:n}^{ORSS}) \\ &= P(F(X_{r:n}^{ORSS}) \leq p \leq F(X_{s:n}^{ORSS})) \\ &= P(U_{r:n}^{ORSS} \leq p \leq U_{s:n}^{ORSS}) \\ &= P(U_{r:n}^{ORSS} \leq p) - P(U_{s:n}^{ORSS} \leq p) \\ &= \pi(r, s, n, p). \end{aligned}$$

بالاکریشنان و لی (۲۰۰۶) با استفاده از رابطه‌های (۱) و (۲)، نشان دادند که مقدار دقیق $\pi(r, s, n, p)$ برابر است با

$$\sum_{i=r}^{s-۱} \sum_{S_i^{[n]}} \{ \prod_{\ell=۱}^{i} [\sum_{t=j_\ell}^{n} \binom{n}{t} p^t (۱-p)^{n-t}] \prod_{\ell=i+۱}^{n} [۱ - \sum_{t=j_\ell}^{n} \binom{n}{t} p^t (۱-p)^{n-t}] \}.$$

از رابطه فوق واضح است که ضریب اطمینان بازه $[X_{r:n}^{ORSS}, X_{s:n}^{ORSS}]$ به توزیع جامعه یعنی F بستگی ندارد. بنابراین یک بازه اطمینان توزیع آزاد است. از سوی دیگر باید توجه شود که برای یک ضریب اطمینان از پیش تعیین شده η، یک بازه اطمینان دو طرفه برا ی ξ_p براساس آماره‌های ترتیبی مجموعه نمونه‌گیری رتبه‌دار وجود دارد اگر و تنها اگر نامساوی زیر برقرار باشد

$$P(U_{۱:n}^{ORSS} \leq p \leq U_{n:n}^{ORSS}) \geq \eta. \qquad (۳)$$

بنابراین شرط (۳) معادل است با

$$\prod_{i=۱}^{n} [۱ - \sum_{s=i}^{n} \binom{n}{s} p^s (۱-p)^{n-s}] + \prod_{i=۱}^{n} [\sum_{s=i}^{n} \binom{n}{s} p^s (۱-p)^{n-s}] \leq ۱ - \eta.$$

با فرض برقراری نامساوی اخیر، به دلیل پله‌ای بودن تابع $\pi(r, s, n, p)$ نسبت به r و s، معمولاً بازه اطمینانی به صورت $[X_{r:n}^{ORSS}, X_{s:n}^{ORSS}]$ نمی‌توان یافت که احتمال پوشش آن دقیقا برابر با ضریب اطمینان از پیش تعیین شده η باشد.

۴ بازه اطمینان تصادفی برحسب آماره‌های ترتیبی مجموعه رتبه‌دار

زلینسکی و زلینسکی (۲۰۰۵) برای رفع مشکل مشابه مطرح شده در بخش قبل برای آماره‌های ترتیبی معمولی روشی ارائه دادند، که در اینجا برای آماره‌های ترتیبی مجموعه رتبه‌دار تحقق داده می‌شود. فرض کنید اندیس‌های r و s آماره‌های ترتیبی مجموعه نمونه‌گیری رتبه‌دار، متغیرهای تصادفی R، S و مستقل از X_iها باشند، به طوری که

$$P(R=r,S=s)=\lambda_{rs}.$$

حال بازه اطمینان با اندیس‌های تصادفی R و S را به صورت $[X_{R:n}^{ORSS},X_{S:n}^{ORSS}]$ در نظر بگیرید، مجدداً با استفاده از خاصیت تبدیل آماره‌های ترتیبی، می‌توان نوشت

$$\begin{aligned}P(X_{R:n}^{ORSS}\le\xi_p\le X_{S:n}^{ORSS}) &= P(U_{R:n}^{ORSS}\le p\le U_{S:n}^{ORSS})\\ &= \sum_{r=1}^{n-1}\sum_{s=r+1}^{n}\lambda_{rs}\pi(r,s,n,p).\end{aligned}$$

در این حالت نیز ضریب اطمینان به توزیع F وابسته نیست و علاوه بر r، s، n و p به مقدار λ_{rs} نیز بستگی دارد. تحت شرط (۳) تعداد نامتناهی λ_{rs} $(1\le r<s\le n)$ وجود دارد به طوری که $P(U_{R:n}^{ORSS}\le p\le U_{S:n}^{ORSS})=\eta$، بنابراین بی‌نهایت بازه اطمینان ناپارامتری تصادفی برای ξ_p در سطح اطمینان داده شده η در اختیار است. حال مسئله انتخاب بهترین آنهاست که در بخش بعد به آن پرداخته می‌شود.

۵ انتخاب بهترین بازه اطمینان براساس آماره‌های ترتیبی مجموعه رتبه‌دار

فرض کنید شرط (۳) برقرار باشد. برای بازه اطمینان $[U_{R:n}^{ORSS},U_{S:n}^{ORSS}]$ با اندیس‌های تصادفی R و S داریم

$$P(R=r,S=s)=\lambda_{rs},\quad 1\le r<s\le n,\ \ \lambda_{rs}\ge 0$$

به طوری که

$$\sum_{r=1}^{n-1}\sum_{s=r+1}^{n}\lambda_{rs}=1.$$

در این صورت به سادگی می‌توان نشان داد که امید ریاضی طول بازه برابر است با

$$E(S-R)=\sum_{r=1}^{n-1}\sum_{s=r+1}^{n}\lambda_{rs}(s-r).$$

برای پیداکردن بهترین اندیس‌هاس تصادفی R و S باید عبارت $\sum_{r=1}^{n-1}\sum_{s=r+1}^{n}\lambda_{rs}(s-r)$, تحت شرط‌های زیر مینیمم شود

$$\sum_{r=1}^{n-1}\sum_{s=r+1}^{n}\lambda_{rs}\pi(r,s,n,p)=\eta,$$

$$\sum_{r=1}^{n-1}\sum_{s=r+1}^{n}\lambda_{rs}=1,\quad 1\leq r<s\leq n,\ \ \lambda_{rs}\geq 0.$$

روش‌های عددی مختلفی برای حل این مسئله وجود دارد، که در اینجا گام‌های یکی از آن‌ها معرفی می‌شود:

گام ۱: با در نظر گرفتن $B(k)$ به صورت

$$B(k)=\sum_{S_k^{[n]}}\prod_{\ell=1}^{k}[\sum_{t=j_\ell}^{n}\binom{n}{t}p^t(1-p)^{n-t}]\prod_{\ell=k+1}^{n}[1-\sum_{t=j_\ell}^{n}\binom{n}{t}p^t(1-p)^{n-t}]$$

مقدار $k\ (k=1,\ldots,n-1)$ طوری انتخاب شود که $B(k)$ ماکسیمم گردد.

گام ۲: به عنوان اولین تقریب بازه اطمینان، $[U_{k:n}^{ORSS},U_{k+1:n}^{ORSS}]$ در نظر گرفته شود. اگر $B(k)=B(k+1)$ و $B(k)>B(\ell)$، برای هر $\ell\notin\{k,k+1\}$ آنگاه بازه $[U_{k:n}^{ORSS},U_{k+2:n}^{ORSS}]$ انتخاب شود.

گام ۳: فرض شود $P(U_{k:n}^{ORSS}\leq p\leq U_{k+1:n}^{ORSS})<\eta$. زیرا اگر

$$P(U_{k:n}^{ORSS}\leq p\leq U_{k+1:n}^{ORSS})=\eta$$

آنگاه بازه $[U_k,U_{k+1}]$ بهترین بازه اطمینان خواهد بود.

گام ۴: اگر $B(k-1)>B(k+1)$، آنگاه بازه $[U_{k-1:n}^{ORSS},U_{k+1:n}^{ORSS}]$ به عنوان تقریب بعدی در نظر گرفته شود، در غیر این صورت تقریب بعدی بازه $[U_{k:n}^{ORSS},U_{k+2:n}^{ORSS}]$ خواهد بود.

گام ۵: این فرایند تا زمانی که برای یک $(r_\circ = r - ۱,\ s_\circ = s)$ یا $(r_\circ = r,\ s_\circ = s + ۱)$ نامساوی‌های

$$P(U_{r:n}^{ORSS} \leq p \leq U_{s:n}^{ORSS}) < \eta,\ P(U_{r_\circ:n}^{ORSS} \leq p \leq U_{s_\circ:n}^{ORSS}) > \eta$$

برقرار شوند، ادامه داده شود.

گام ۶: λ طوری محاسبه شود که رابطه زیر برقرار باشد

$$\lambda P(U_{r:n}^{ORSS} \leq p \leq U_{s:n}^{ORSS}) + (۱ - \lambda)P(U_{r_\circ:n}^{ORSS} \leq p \leq U_{s_\circ:n}^{ORSS}) = \eta.$$

بنابراین بازه اطمینانی حاصل شد که با احتمال λ برابر با $[U_{r:n}^{ORSS}, U_{s:n}^{ORSS}]$ و با احتمال $۱ - \lambda$ برابر با $[U_{r_\circ:n}^{ORSS}, U_{s_\circ:n}^{ORSS}]$ است. با روش پیشنهاد شده، بازه اطمینان برای ξ_p با ضریب اطمینان دقیق ۹۵٪ به ازای مقادیر ممکن p و برخی از مقادیر n همراه با متوسط طول بازه (L) و مقدار λ محاسبه شده و در جدول ۱ ارائه شده‌اند.

جدول ۱: بازه اطمینان بهینه ۹۵٪ برای چندک مرتبه pام براساس مجموعه رتبه‌دار

n		p: ۰/۳	/۰۴	۰/۵	۰/۶	۰/۷
۴	بازه			[۱, ۳][۱, ۴]		
				[۲, ۴][۱, ۴]*		
	λ			۰/۱۱		
	L			۲/۸۹		
۵	بازه		[۱, ۴][۱, ۳]	[۱, ۴][۱, ۵]	[۳, ۵][۲, ۵]	
	λ		۰/۰۶	۰/۵	۰/۰۶	
	L		۲/۹۴	۳/۵	۲/۹۴	
۶	بازه	[۱, ۳][۱, ۴]	[۱, ۴][۱, ۵]	[۲, ۵][۱, ۵]	[۳, ٦][۲, ٦]	[۴, ٦][۳, ٦]
				[۲, ۵][۲, ٦]*		
	λ	۰/۰۳	۰/۵۸	۰/۹۸	۰/۵۸	۰/۰۳
	L	۲/۹۷	۳/۴۲	۳/۰۲	۳/۴۲	۲/۹۷
۷	بازه	[۱, ۴][۱, ۵]	[۲, ۵][۱, ۵]	[۲, ۵][۲, ٦]	[۳, ٦][۳, ۷]	[۴, ۷][۳, ۷]
	λ	۰/۹۸	۰/٦۵	۰/۳۷	۰/٦۵	۰/۹۸
	L	۳/۰۲	۳/۳۵	۳/٦۳	۳/۳۵	۳/۰۲
۸	بازه	[۱, ۴][۱, ۵]	[۲, ۵][۲, ٦]	[۳, ٦][۲, ٦]	[۴, ۷][۳, ۷]	[۵, ۸][۴, ۸]
				[۳, ٦][۳, ۷]*		
	λ	۰/۴۸	۰/۵۱	۰/۳۵	۰/۵۱	۰/۴۸
	L	۳/۵۲	۳/۴۹	۳/٦۵	۳/۴۹	۳/۵۲

* احتمال پوشش بازه [۱, ۳] با [۲, ۴]، [۱, ۵] با [۲, ٦]، [۲, ٦] با [۳, ۷] برابر است.

لازم به ذکر است برای مقایسه با سایر روش‌ها برای p مقادیر ۰/۲ تا ۰/۸ و برای n نیز مقادیر ۳ تا ۸ در نظر گرفته شده است. برای خانه‌های خالی جدول ۱،

شرط (۳) برقرار نیست در نتیجه بازه اطمینان با ضریب اطمینان ۹۵٪ وجود ندارد. اعداد زیر بازه‌ها در جدول ۱، ردیف اول مقدار λ و ردیف دوم متوسط طول بازه یعنی L است.

۶ مقایسه روش‌ها

در این بخش روش پیشنهادی با سایر روش‌های موجود که در ادامه به اختصار معرفی می‌شوند، مورد مقایسه قرار می‌گیرد.

الف- روش چن: فرض کنید $X^{ORSS}_{1:n},\ldots,X^{ORSS}_{n:n}$ مشاهدات مرتب شده حاصل از طرح نمونه‌گیری از مجموعه رتبه‌دار باشد، دراین صورت چن (۲۰۰۰) بازه اطمینان تقریبی $(1-\alpha)100\%$ با دم‌های برابر را برای چندک مرتبه pام به صورت $[X^{ORSS}_{\ell_1:N},X^{ORSS}_{\ell_2:N}]$ پیشنهاد داد، ℓ_1 و ℓ_2 به ترتیب نزدیک‌ترین عدد صحیح مثبت عبارت‌های زیر در نظر گرفت

$$\begin{aligned}\ell_1 &\approx np - Z_{1-\frac{\alpha}{2}}\left(\sum_{r=1}^{n} I_p(r,n-r+1)[1-I_p(r,n-r+1)]\right)^{1/2}\\ \ell_2 &\approx np + Z_{1-\frac{\alpha}{2}}\left(\sum_{r=1}^{n} I_p(r,n-r+1)[1-I_p(r,n-r+1)]\right)^{1/2},\end{aligned}$$

که در آن Z_a چندک مرتبه aام از توزیع نرمال استاندارد و $I_p(a,b)$ تابع توزیع بتا با پارامترهای a و b است، یعنی $I_p(a,b)=\frac{1}{B(a,b)}\int_0^p t^{a-1}(1-t)^{b-1}dt$.

جدول ۲، بازه اطمینان ۹۵٪ برای چندک مرتبه pام به همراه احتمال پوشش و متوسط طول بازه (L) به ازای مقادیر مختلف n و مقادیر ممکن p که از روش پیشنهادی چن بدست آمده، را نشان می‌دهد. اعداد زیر بازه‌ها در جدول ۲، ردیف اول مقدار احتمال پوشش و ردیف دوم متوسط طول بازه است. به دلیل اینکه چن از تقریب نرمال برای پیدا کردن کران‌های تقریبی برای چندک مرتبه pام استفاده می‌کند، چنین به نظر می‌رسد روش وی در برخی از موارد عملکرد خوبی از خود نشان نداده و احتمال پوشش برخی بازه‌ها تفاوت زیادی با ضریب اطمینان مورد نظر دارد. به عنوان مثال به ازای $p=0/3$ و $n=6$، احتمال پوشش ۰/۸۳ حاصل شد در

جدول ۲: بازه اطمینان ۹۵٪ برای چندک مرتبه pام براساس مجموعه رتبه‌دار با روش چن

n		p						
		۰/۲	۰/۳	۰/۴	۰/۵	۰/۶	۰/۷	۰/۸
۳	بازه احتمال پوشش L		[۰, ۲] ۰/۸۴ ۲	[۰, ۳] ۰/۹۸ ۳	[۰, ۳] ۰/۹۵ ۳	[۰, ۳] ۰/۸۷ ۳	[۱, ۳] ۰/۷۳ ۲	
۴	بازه احتمال پوشش L	[۰, ۲] ۰/۸۸ ۲	[۰, ۳] ۰/۹۷ ۳	[۰, ۳] ۰/۹۱ ۳	[۱, ۳] ۰/۷۵ ۲	[۱, ۴] ۰/۹۵ ۳	[۱, ۴] ۰/۸۶ ۳	[۲, ۴] ۰/۶۷ ۲
۵	بازه احتمال پوشش L	[۰, ۲] ۰/۷۹ ۲	[۰, ۳] ۰/۹۲ ۳	[۰, ۴] ۰/۹۸ ۴	[۱, ۴] ۰/۹۱ ۳	[۱, ۵] ۰/۹۸ ۴	[۲, ۵] ۰/۹۳ ۳	[۳, ۵] ۰/۷۶ ۲
۶	بازه احتمال پوشش L	[۰, ۳] ۰/۹۷ ۳	[۰, ۳] ۰/۸۳ ۳	[۱, ۴] ۰/۹۲ ۳	[۱, ۵] ۰/۹۷ ۴	[۲, ۵] ۰/۸۷ ۳	[۳, ۶] ۰/۹۶ ۳	[۳, ۶] ۰/۸۵ ۳
۷	بازه احتمال پوشش L	[۰, ۳] ۰/۹۴ ۳	[۰, ۴] ۰/۹۶ ۴	[۱, ۴] ۰/۸۱ ۳	[۲, ۵] ۰/۸۸ ۳	[۳, ۶] ۰/۹۳ ۳	[۳, ۷] ۰/۹۸ ۴	[۴, ۷] ۰/۹ ۳
۸	بازه احتمال پوشش L	[۰, ۳] ۰/۸۹ ۳	[۱, ۴] ۰/۹ ۳	[۱, ۵] ۰/۹۴ ۴	[۲, ۶] ۰/۹۳ ۴	[۳, ۷] ۰/۹۸ ۴	[۴, ۷] ۰/۸۶ ۳	[۵, ۸] ۰/۹۳ ۳

حالیکه ضریب اطمینان ۹۵٪ است.

ب- روش بالاکریشنان و لی: بالاکریشنان و لی (۲۰۰۶) برای ساختن یک بازه اطمینان برای چندک مرتبه pام براساس مشاهدات مرتب شده حاصل از طرح نمونه‌گیری از مجموعه رتبه‌دار، ابتدا احتمال پوشش بازه $[X_{r:n}^{ORSS}, X_{s:n}^{ORSS}]$ را به دست آوردند. با توجه به گسسته بودن احتمال بدست آمده، بازه اطمینان برای چندک مرتبه pام را به صورت زیر مشخص کردند.

گام ۱: r و s را طوری انتخاب کرده که $s-r$ تا حد امکان کوچک و ضریب اطمینان بازه اطمینان فوق بزرگتر یا مساوی $1-\alpha$ باشد.

گام ۲: اگر برای دو مقدار متفاوت (r, s)، مقدار $s-r$ برابر شد، r و s طوری انتخاب شوند که امید ریاضی طول بازه $[U_{r:n}^{ORSS}, U_{s:n}^{ORSS}]$ تا حد امکان کوچک شود.

برای انجام مقایسه، همان مقادیر n و p در جداول ۱ و ۲ در اینجا نیز در نظر گرفته شده‌اند. بازه اطمینان با ضریب اطمینان ۹۵٪ برای ξ_p با استفاده از روش بالاکریشنان و لی به همراه متوسط طول بازه (L) به دست آمده و نتایج در جدول ۳ خلاصه شده است. اعداد زیر بازه‌ها در جدول ۳ متوسط طول بازه‌ها را نشان می‌دهند.

جدول ۳: بازه اطمینان ۹۵٪ برای چندک مرتبه pام براساس مجموعه رتبه‌دار با روش بالاکریشنان و لی

		p				
n		۰/۳	۰/۴	۰/۵	۰/۶	۰/۷
۴	بازه		[۱, ۴]	[۱, ۴]	[۱, ۴]	
	L		۳	۳	۳	
۵	بازه		[۱, ۴]	[۱, ۵]	[۲, ۵]	
	L		۳	۴	۳	
٦	بازه	[۱, ۴]	[۱, ۵]	[۲, ۵]	[۲, ٦]	[۳, ٦]
	L	۳	۴	۳	۴	۳
۷	بازه	[۱, ۴]	[۱, ۵]* [۲, ٦]	[۲, ٦]	[۳, ۷]* [۲, ٦]	[۴, ۷]
	L	۳	۴	۴	۴	۳
۸	بازه	[۱, ۵]	[۲, ٦]	[۳, ۷]	[۳, ۷]	[۴, ۸]
	L	۴	۴	۴	۴	۴

* بازه اطمینان براساس مینیمم متوسط طول انتخاب شده است.

عیب اصلی این روش این است که در اینجا نیز نمی‌توان به ضریب اطمینان دقیق دست یافت و متوسط طول بازه تقریباً در همه موارد بیشتر از روش پیشنهادی است.

ج- روش ازترک و دشپاند: ازترک و دشپاند (۲۰۰۶) بازه اطمینان برای چندک مرتبه pام براساس مشاهدات مرتب شده حاصل از نمونه‌گیری مجموعه رتبه‌دار را به صورت ترکیب خطی محدب از دو بازه اطمینان ارائه دادند. فرض کنید دو بازه اطمینان $I_{r,s} = [X_{r:n}^{ORSS}, X_{s:n}^{ORSS}]$ و $II_{r+۱,s-۱} = [X_{r+۱:n}^{ORSS}, X_{s-۱:n}^{ORSS}]$ با ضریب اطمینان $(۱-\alpha_I)۱۰۰$٪ و $(۱-\alpha_{II})۱۰۰$٪ برای چندک مرتبه pام جامعه باشند، به طوری که داشته باشیم $۱-\alpha_{II} \leq ۱-\alpha \leq ۱-\alpha_I$. در این صورت بازه اطمینان

پیشنهادی ازترک و دشپاند به صورت

$$I_{\epsilon_1,\epsilon_2} = [(1-\epsilon_1)X_{r:n}^{ORSS} + \epsilon_1 X_{r+1:n}^{ORSS}, (1-\epsilon_2)X_{s:n}^{ORSS} + \epsilon_2 X_{s-1:n}^{ORSS}]$$

است، که در آن

$$\epsilon_1 \approx \left[1 + \frac{r(1-p)(\alpha_{II}/2 - \alpha/2)}{(n-r)p(\alpha/2 - \alpha_1/2)}\right]^{-1} \qquad (4)$$

$$\epsilon_2 \approx \left[1 + \frac{(n-(s-1))p(\alpha/2 - \alpha_{II}/2)}{(s-1)(1-p)(\alpha_I/2 - \alpha/2)}\right]^{-1}. \qquad (5)$$

جدول۴ شامل نتایج مربوط به ضرایب بازه اطمینان ۹۵٪ با روش ازترک و دشپاند را برای چندک مرتبه pام همراه با متوسط طول بازه (L)، به ازای همان مقادیر n و p در نظر گرفته شده در سایر جداول، می‌باشد. مقادیر هر خانه در جدول ۴، ردیف اول $r(\epsilon_1)$، ردیف دوم $s(\epsilon_2)$، ردیف سوم $\alpha_I - 1$، ردیف چهارم $\alpha_{II} - 1$ و ردیف پنجم متوسط طول بازه است.

به عنوان مثال وقتی $n = 5$ و $p = 0/5$ باشد، بازه $I_{1,5} = [X_{1:5}^{ORSS}, X_{5:5}^{ORSS}]$ دارای احتمال پوشش دقیق $0/995$ و بازه کوچکتر $II_{2,4} = [X_{2:5}^{ORSS}, X_{4:5}^{ORSS}]$ دارای احتمال پوشش $0/827$ است. با استفاده از روابط (۴) و (۵)، مقادیر ϵ_1 و ϵ_2 به صورت $\epsilon_1 = 0/595$ و $\epsilon_2 = 0/595$ به دست آمده است. بنابراین بازه اطمینان ۹۵٪ برای $\xi_{0/5}$ برای اندازه نمونه ۵، به صورت حاصل می‌شود

$$I_{\epsilon_1,\epsilon_2} = [0/405X_{1:5}^{ORSS} + 0/595X_{2:5}^{ORSS}, 0/405X_{5:5}^{ORSS} + 0/595X_{4:5}^{ORSS}].$$

مشکلی که در روش پیشنهادی ازترک و دشپاند مشاهده می‌شود این است که به دلیل سختی محاسبات آنها احتمال پوشش بازه پیشنهادی را به طور دقیق محاسبه نکردند و به طور شهودی استنباط کردند احتمال پوشش بازه اطمینان نزدیک به ۹۵٪ است.

اکنون با مقایسه جداول ۱ و ۲، ملاحظه می‌شود که روش پیشنهادی چن در برخی موارد تقریب ضعیفی از ضریب اطمینان ارائه می‌کند اما در مواردی که احتمال پوشش فاصله اطمینان پیشنهادی چن نزدیک به ۰/۹۵ است، روش پیشنهادی

جدول ۴: ضرایب بازه اطمینان ۹۵٪ برای چندک مرتبه pام براساس مجموعه رتبه‌دار با روش ازترک و دشپاند

n		p				
		۰/۳	۰/۴	۰/۵	۰/۶	۰/۷
۴	$r(\epsilon_1)$			۱(۰/۱۵۰)		
	$s(\epsilon_2)$			۴(۰/۱۵۰)		
	$1-\alpha_I$			۰/۹۷۵		
	$1-\alpha_{II}$			۰/۵۲۹		
	L			۲/۷		
۵	$r(\epsilon_1)$		۱(۰/۰۸۱)	۱(۰/۵۹۵)	۲(۰/۰۶۹)	
	$s(\epsilon_2)$		۴(۰/۰۶۹)	۵(۰/۵۹۵)	۵(۰/۰۸۱)	
	$1-\alpha_I$		۰/۹۶۵	۰/۹۹۵	۰/۹۶۵	
	$1-\alpha_{II}$		۰/۵۰۵	۰/۸۲۷	۰/۵۰۵	
	L		۲/۸۵	۲/۸۱	۲/۸۵	
۶	$r(\epsilon_1)$	۱(۰/۰۲۳)	۱(۰/۴۹۳)	۱(۰/۳۴۱)	۲(۰/۴۶۷)	۳(۰/۰۲۵)
	$s(\epsilon_2)$	۴(۰/۰۲۵)	۵(۰/۴۶۷)	۵(۰/۱۷۱)	۶(۰/۴۹۳)	۶(۰/۰۲۳)
	$1-\alpha_I$	۰/۹۵۵	۰/۹۹۳	۰/۹۷۵	۰/۹۹۳	۰/۹۵۵
	$1-\alpha_{II}$	۰/۴۸۰	۰/۸۰۴	۰/۷۱۴	۰/۸۰۴	۰/۴۸۰
	L	۲/۹۵	۳/۰۴	۳/۴۹	۳/۰۴	۲/۹۵
				۲(۰/۱۷۱)*		
				۶(۰/۳۴۱)		
				۰/۹۷۵		
				۰/۷۱۴		
۷	$r(\epsilon_1)$	۱(۰/۲۹۸)	۱(۰/۴۰۱)	۲(۰/۳۶۳)	۳(۰/۲۵۱)	۳(۰/۳۴۰)
	$s(\epsilon_2)$	۵(۰/۳۴۰)	۵(۰/۲۵۱)	۶(۰/۳۶۳)	۷(۰/۴۰۱)	۷(۰/۲۹۸)
	$1-\alpha_I$	۰/۹۸۴	۰/۹۸۲	۰/۹۸۸	۰/۹۸۲	۰/۹۸۴
	$1-\alpha_{II}$	۰/۷۴۳	۰/۷۵۸	۰/۷۸۲	۰/۷۵۸	۰/۷۴۳
	L	۳/۳۶	۳/۳۵	۳/۲۷	۳/۳۵	۳/۳۶
۸	$r(\epsilon_1)$	۱(۰/۴۱۵)	۲(۰/۲۱۵)	۲(۰/۱۲۳)	۳(۰/۲۵۵)	۴(۰/۳۵۶)
	$s(\epsilon_2)$	۵(۰/۳۵۶)	۶(۰/۲۵۵)	۶(۰/۰۷۲)	۷(۰/۲۱۵)	۸(۰/۴۱۵)
	$1-\alpha_I$	۰/۹۸۹	۰/۹۷۹	۰/۹۶۲	۰/۹۷۹	۰/۹۸۹
	$1-\alpha_{II}$	۰/۷۸۴	۰/۷۳۹	۰/۶۸۶	۰/۷۳۹	۰/۷۸۴
	L	۳/۲۳	۳/۵۳	۳/۸۰	۳/۵۳	۳/۲۳
				۳(۰/۰۷۲)*		
				۷(۰/۱۲۳)		
				۰/۹۶۲		
				۰/۶۸۶		

* احتمال پوشش دو بازه [۱, ۵]، [۲, ۶] و دو بازه [۲, ۶]، [۳, ۷] برابر است.

عملکرد عموماً بهتری از بازه اطمینان پیشنهادی چن دارد. با مقایسه جداول ۱ و ۳، بازه اطمینان پیشنهادی تقریبا در تمام حالات متوسط طول کوتاهتری از بازه اطمینان پیشنهادی بالاکریشنان و لی دارد. با مقایسه جداول ۱ و ۴، نمی‌توان به طور کلی در مورد برتر بودن روش پیشنهادی اظهار نظر کرد. اما باید به خاطر داشت ضریب اطمینان روش پیشنهادی برخلاف روش ازترک و دشپاند دقیقاً برابر با ۹۵٪ است.

۷ بحث و نتیجه‌گیری

یکی از روش‌هایی که در سال‌های اخیر در مبحث نمونه‌گیری مورد توجه قرار گرفته، روش نمونه‌گیری مجموعه رتبه‌دار است. استفاده از اطلاعات نهفته در داده‌های جمع‌آوری شده بر مبنای این روش، برای استنباط پارامترهای جامعه از جمله چندک‌ها حائز اهمیت است. بر این اساس، در این مقاله ابتدا روش‌های موجود در خصوص نحوه ساختن بازه اطمینان برای پارامتر چندک با استفاده از نمونه‌گیری مجموعه رتبه‌دار را بررسی و مشاهده شد که بر هریک از آنها عیبی وارد است. بازه اطمینان با ضریب اطمینان دقیق $(1-\alpha)100$٪ برای چندک مرتبه pام، براساس مشاهدات مرتب شده حاصل از طرح نمونه‌گیری مجموعه رتبه‌دار ارائه شد. سپس روش پیشنهادی با سایر روشهای موجود مقایسه گردید. نتایج مقایسه نشان داد که روش پیشنهادی عملکرد نسبتاً خوبی دارد با این امتیاز که دارای ضریب اطمینان دقیق $(1-\alpha)100$٪ است.

تقدیر و تشکر

نویسندگان مقاله از پیشنهادات و نظرات داوران محترم که در بهبود مقاله موثر واقع شد، همچنین از قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد به خاطر حمایت مالی، تقدیر و تشکر می‌کنند.

مراجع

Ahmadi, J. and Arghami, N. R. (2003), Nonparametric Confidence and Tolerance Intervals from Record Values Data, *Statistical Papers*, **44**, 455–468.

Balakrishnan, N. and Han, Donghoon (2007), Optimal Progressive Type-II Censoring Schemes for Nonparametric Confidence Intervals of Quantiles, *Communications in Statistics: Simulation and Computation*, **36**, 1247–1262.

Balakrishnan, N. and Li, T. (2006), Confidence Interval for Quantiles and Tolerance Intervals based on Ordered Ranked Set Samples, *Annals of the Institute of Statistical Mathematics*, **58**, 757–777.

Beutner, E. and Cramer, E. (2011), Confidence Intervals for Quantiles in a Minimal Repair Set-up, *International Journal of Applied Mathematics and Statistics*, **24**, 86–97.

Chen, Z. (2000), On Ranked-set Sample Quantiles and Their Applications, *Journal of Statistical Planning and Inference*, **83**, 125–135.

David, H. A. and Nagaraja, H. N. (2003), *Order Statistics* (3rd ed.), Wiley, New York

Hutson, A. (1999), Calculating Nonparametric Confidence Interval for Quantiles Using Fractional Order Statistics, *Journal of Applied Statistics*, **26**, 343-353.

Guilbaud, O. (2001), Exact Nonparametric Confidence Intervals for Quantiles with Progressive Type-II Censoring, *Scandinavian Journal of*

Statistics, **28**, 699-713.

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

Ozturk, O. and V. Deshpande, J. (2006), Ranked-Set Sample Nonparametric Quantile Confidence Interval, *Journal of Statistical planning and Inference*, **136**, 570-577.

Steinberg, Seth M. and Davis, C. E. (1985), Distribution-Free Confidence Intervals for Quantiles in Small Samples, *Communications in Statistics: Theory Methods*, **14**, 979–990.

Tkahasi, K. and Wakimoto, K. (1968), On Unbiased Estimates of the Population Mean Based on the Stratified by Means of Ordering, *Annals of the Institute of Statistical Mathematics*, **20**, 1-31.

Zielinski, R. and Zielinski, W. (2005), Best Exact Nonparametric Confidence Interval for Quantiles, *Statistics*, **39**, 67-71.