



بازه‌های اطمینان آزاد توزیع برای چندک‌های جامعه بر اساس آماره‌های کرانگین در یک طرح چند نمونه‌ای

مصطفی رزمخواه* و جعفر احمدی و بهاره خطیب آستانه

دانشگاه فردوسی مشهد

چکیده. فرض کنید $(X_1, \dots, X_{i n_i})$ نمونه‌ای تصادفی از توزیع F^{α_i} ، $i = 1, \dots, k$ باشد که در آن F تابع توزیع مطلقاً پیوسته است و $\alpha_i > 0$. همچنین فرض کنید نمونه‌ها مستقل از هم باشند و M_{i, n_i} و M'_{i, n_i} به ترتیب ماکسیمم و مینیمم متناظر با نمونه‌ی i ام، $i = 1, \dots, k$ باشند. تعیین بازه‌های اطمینان آزاد توزیع برای چندک‌های جامعه‌ی F با استفاده از توابعی از آماره‌های بالا هدف اصلی این مقاله است. مسئله در حالت‌های مختلف مورد بررسی قرار گرفته و در هر حالت، بازه‌های اطمینان مورد نظر محاسبه شده است.

واژگان کلیدی. آماره‌های ترتیبی کرانگین؛ بازه‌ی اطمینان؛ تابع نرخ خطر معکوس؛ چندک؛ احتمال پوشش؛ مدل F^{α} .

۱ مقدمه

فرض کنید X یک متغیر تصادفی با تابع توزیع $F(x)$ باشد. در این صورت چندک مرتبه‌ی p ام توزیع متغیر تصادفی X که با ξ_p نشان داده می‌شود، عبارت است از کوچک‌ترین عددی مانند x که در نامعادله‌ی

* نویسنده‌ی عهده‌دار مکاتبات.

$F(x) \geq p$ صدق کند؛ یعنی

$$\xi_p = \inf \{x : F(x) \geq p\}.$$

اگر یک متغیر تصادفی پیوسته باشد، آن‌گاه چندک مرتبه‌ی p ام توزیع آن عددی است مانند x که در تساوی $F(x) = p$ صدق کند. در حالت خاص $p = \frac{1}{2}$ ، ξ میانه‌ی جامعه است. چندک‌ها و توابعی از آن‌ها را می‌توان برای ساختن معیارهای مکانی و پراکندگی یک توزیع به کار برد.

هرگاه شکل تابعی توزیع مورد نظر معلوم و فقط مقادیر یک یا چند پارامتر نامعلوم باشد می‌توانیم برآوردگرهای نقطه‌ای، برآوردگرهای بازه‌ای و آزمون فرض را بر اساس شکل تابعی توزیع جامعه انجام دهیم. اما در بیش‌تر مسائل شکل تابعی توزیع جامعه نامعلوم است و باید از فنونی استفاده کنیم که به روش‌های ناپارامتری یا آزاد توزیع معروفند. هرگاه همه‌ی مشاهدات در اختیار باشند، آماره‌های مرتب‌نقش اساسی در ارتباط با استنباط درباره‌ی چندک‌ها ایفا می‌کنند. سرفلینگ (۱۹۸۰)، آرنولد و همکاران (۱۹۹۲) و دیوید و ناگارا (۲۰۰۳) از جمله کتاب‌های معروفی هستند که به اهمیت آماره‌های ترتیبی در مسائل استنباط مربوط و چندک‌ها پرداخته‌اند.

در سال‌های اخیر تحقیقات زیادی در خصوص استنباط آزاد توزیع بر اساس رکوردها مورد توجه پژوهشگران قرار گرفته است. گولاتی و پاجت (۲۰۰۳) مجموعه‌ی کارهای انجام شده در این راستا را جمع‌آوری نموده‌اند. احمدی و ارقامی (۲۰۰۳)، احمدی و بالاکریشنان (۲۰۰۵، ۲۰۰۴) بازه‌های اطمینان آزاد توزیع بر اساس آماره‌های رکوردی برای چندک‌های جامعه را به دست آورده‌اند. اما مقاله‌های زیادی در زمینه‌ی استفاده از آماره‌های مرتب در بازه‌های اطمینان و آزمون‌های فرض آزاد توزیع مربوط به چندک‌ها به چاپ رسیده است که از آن جمله می‌توان به کروسکی (۱۹۷۶)، ساتی و لینگراس (۱۹۸۱)، زو و سان (۲۰۰۰)، جن (۲۰۰۰)، زیلینسکی (۲۰۰۵) و اوزتارک و دشیپاند (۲۰۰۶) اشاره نمود.

اما موارد زیادی وجود دارند که همه‌ی داده‌ها ثبت نمی‌شوند و فقط کوچک‌ترین یا بزرگ‌ترین آن‌ها حائز اهمیت هستند. حالتی را در نظر بگیرید که آزمایشی در دوره‌های متفاوت انجام می‌شود؛ مثل مسابقات علمی که در آن‌ها فقط بالاترین رتبه (ماکسیمم داده‌ها) حائز اهمیت است یا در مسابقات سرعتی که در آن‌ها کم‌ترین زمان (مینیمم مشاهدات) مورد توجه می‌باشد. گاهی نیز ممکن است هم مینیمم و هم ماکسیمم مشاهدات در دوره‌های متوالی ثبت شده باشند؛ مانند بالاترین و پایین‌ترین دمای هوا در طول یک هفته یا یک ماه. بنا بر این، این موضوع که بتوانیم با استفاده از آماره‌های کرانگین در نمونه‌های متوالی، تحلیل آماری و استنباط درباره‌ی جامعه‌ی پایه را انجام دهیم، حائز اهمیت است.

در این مقاله طرحی از نمونه‌گیری را مطرح می‌کنیم که در آن k نمونه هریک به اندازه‌ی n_i ، $i = 1, \dots, k$ و به‌طور مستقل از هم از یک جامعه‌ی پایه مانند F استخراج می‌شوند و سپس با استفاده از آماره‌های

کرانگین در این نمونه‌ها سعی می‌کنیم تا برای چندک‌های جامعه‌ی F برآوردگر بازه‌ای پیدا کنیم. توجه شود که نمونه‌ها در طول زمان از جامعه‌ی پایه استخراج می‌شوند و از طرفی ممکن است جامعه‌ی پایه در این خلال اندکی تغییر کند. فرض کنید این تغییرات را به صورت F^{α_i} نشان دهیم؛ یعنی فرض بر این است که نمونه‌ی i ام از توزیع $F_i = F^{\alpha_i}$ استخراج می‌شود که در آن F همان توزیع پایه‌ی جامعه است. در واقع تابع نرخ خطر معکوس نمونه‌ی i ام، α_i برابر تابع نرخ خطر جامعه‌ی پایه است. این مدل در مقاله‌ها به مدل F^α یا مدل نرخ خطر معکوس متناسب معروف است. مدل F^α شامل یک دنباله‌ی نامتناهی $\{Y_n, n \geq 1\}$ از متغیرهای تصادفی مستقل است که تابع توزیع تجمعی Y_n به شکل $F_n(y) = \{F(y)\}^{\alpha(n)}$ ، $\alpha(n) > 0$ می‌باشد. نزروف (۱۹۹۰) به این مدل به عنوان یک طرح F^α اشاره نموده است. در سال‌های اخیر مقاله‌های زیادی در خصوص ترتیب‌های تصادفی این مدل به چاپ رسیده است، از جمله می‌توان به خالدی و کوچار (۲۰۰۴) اشاره نمود.

هدف اصلی ما در این مقاله ابتدا محاسبه‌ی برآوردگر بازه‌ای برای چندک‌های جامعه‌ی پایه در حالت کلی مدل F^α بر اساس آماره‌های ترتیبی و سپس تعمیم نتایج به دست آمده به حالات خاص است. در این مقاله ابتدا در بخش دوم به بیان چند لم ابزاری می‌پردازیم. در بخش سوم به تعیین بازه‌های اطمینان آزاد توزیع برای چندک مرتبه‌ی p ام جامعه بر اساس ماکسیمم نمونه‌ها می‌پردازیم و سپس در بخش چهارم همین کار را بر اساس مینیمم نمونه‌ها انجام می‌دهیم. در بخش پنجم، هم ماکسیمم و هم مینیمم نمونه‌ها را در تعیین بازه‌های اطمینان به کار می‌بریم. سرانجام در بخش ششم، به تفسیر نتایج به دست آمده در بخش‌های پیشین می‌پردازیم.

۲ چند لم اساسی

فرض کنید $X_1, \dots, X_{i n_i}$ نمونه‌ای تصادفی از توزیع F^{α_i} ، $i = 1, \dots, k$ باشد که در آن F تابع توزیع مطلقاً پیوسته‌ای است و $\alpha_i > 0$. همچنین فرض کنید نمونه‌ها مستقل از هم باشند و M_{i, n_i} و M'_{i, n_i} به ترتیب ماکسیمم و مینیمم متناظر با نمونه‌ی i ام، $i = 1, \dots, k$ باشند. در این مقاله سعی بر آن داریم تا با استفاده از آماره‌های معرفی شده، برای چندک‌های جامعه‌ی اصلی؛ یعنی F ، برآوردگر بازه‌ای محاسبه کنیم. در راستای این هدف برخی از ابزار لازم برای محاسبه‌ی فرمول‌های مربوط را در قالب چند لم، بدون اثبات آن‌ها بیان می‌کنیم.

لم ۱ فرض کنید X_1, \dots, X_n متغیرهای تصادفی مستقل و هم‌توزیع با تابع توزیع پیوسته‌ی F و تابع چگالی احتمال f باشند. همچنین فرض کنید $X_{i:n}$ ، نشان دهنده‌ی i امین آماره‌ی مرتب در نمونه‌ای به

اندازه‌ی n از این جامعه باشد. در این صورت تابع چگالی توأم $X_{1:n}$ و $X_{n:n}$ عبارت است از

$$f_{X_{1:n}, X_{n:n}}(x, y) = n(n-1)f(x)f(y)\{F(y) - F(x)\}^{n-2}.$$

لم ۲ فرض کنید $X_i, i = 1, \dots, n$ ، متغیرهای تصادفی مستقل از هم اما ناهم‌توزیع باشند به‌طوری که X_i دارای تابع توزیع F_i باشد. در این صورت

$$P(X_{i:n} \leq x) = \sum_{r=i}^n \sum_{A_r} \prod_{s=1}^r F_{X_{t_s}}(x) \prod_{s=r+1}^n \bar{F}_{X_{t_s}}(x),$$

که در آن A_r مجموع روی تمام جایگشت‌های (t_1, \dots, t_n) از اعداد $1, \dots, n$ است که $t_1 < \dots < t_r < t_{r+1} < \dots < t_n$.

لم ۳ فرض کنید متغیر تصادفی X دارای توزیع پیوسته‌ی F باشد. در این صورت $F(X)$ دارای توزیع یکنواخت پیوسته روی بازه‌ی $(0, 1)$ است.

لم ۴ تحت مفروضات لم ۱، $U_{r:n} = F(X_{r:n})$ هم‌توزیع با r امین آماری مرتب در نمونه‌ای به اندازه‌ی n از توزیع یکنواخت پیوسته روی بازه‌ی $(0, 1)$ است.

۳ بازه‌های اطمینان بر اساس ماکسیمم نمونه‌ها

در این بخش برای چندک مرتبه‌ی p ام جامعه‌ی اصلی، $\xi_p = F^{-1}(p)$ ، بر اساس ماکسیمم نمونه‌ها، M_{i,n_i} ها، برآوردگر بازه‌ای پیدا می‌کنیم که به توزیع جامعه بستگی ندارد. ابتدا توجه شود که اگر برای $i = 1, \dots, k$ و $j = 1, \dots, n_i$ ، $X_{ij} \sim F^{\alpha_i}$ آنگاه $M_{i,n_i} \sim F^{\alpha_i n_i}$. در نتیجه بنا به لم ۳

$$\begin{aligned} P(M_{i,n_i} \leq \xi_p) &= P\{M_{i,n_i} \leq F^{-1}(p)\} \\ &= P\{F(M_{i,n_i}) \leq p\} \\ &= P\{F^{\alpha_i n_i}(M_{i,n_i}) \leq p^{\alpha_i n_i}\} \\ &= p^{\alpha_i n_i}. \end{aligned}$$

بنا بر این، بازه‌ی اطمینان مبتنی بر دو ماکسیمم متفاوت $(i, j, i \neq j)$ ، دارای ضریب اطمینان زیر است

$$\begin{aligned} P(M_{i,n_i} \leq \xi_p \leq M_{j,n_j}) &= P(M_{i,n_i} \leq \xi_p) P(M_{j,n_j} \geq \xi_p) \\ (۱) \qquad \qquad \qquad &= p^{\alpha_i n_i} (1 - p^{\alpha_j n_j}), \end{aligned}$$

که تساوی اول از خاصیت مستقل بودن نمونه‌ها به دست آمده است. اما ممکن است $M_{i,n_i} > M_{j,n_j}$ که در این صورت بازه‌ی اطمینان (M_{i,n_i}, M_{j,n_j}) برای ξ_p بی‌معنی خواهد بود. در این حالت می‌توانیم آماره‌های مرتب‌شده‌ی M_{i,n_i} و M_{j,n_j} را به صورت $M_{ij1} \leq M_{ij2}$ در نظر بگیریم که در این صورت ضریب اطمینان عبارت است از

$$\begin{aligned} P(M_{ij1} \leq \xi_p \leq M_{ij2}) &= P(M_{i,n_i} \leq \xi_p \leq M_{j,n_j}) + P(M_{j,n_j} \leq \xi_p \leq M_{i,n_i}) \\ &= P(M_{i,n_i} \leq \xi_p) P(M_{j,n_j} \geq \xi_p) \\ &\quad + P(M_{j,n_j} \leq \xi_p) P(M_{i,n_i} \geq \xi_p) \\ (۲) \qquad \qquad \qquad &= p^{\alpha_i n_i} (1 - p^{\alpha_j n_j}) + p^{\alpha_j n_j} (1 - p^{\alpha_i n_i}). \end{aligned}$$

اما برای رفع مشکل بالا می‌توانستیم از همان ابتدا M_{i,n_i} ها، $i = 1, 2, \dots, k$ را طوری مرتب کنیم که $M_{1:k} \leq M_{2:k} \leq \dots \leq M_{k:k}$. در این صورت با در نظر گرفتن $(M_{i:k}, M_{j:k})$ که در آن $i < j$ ، به عنوان بازه‌ی اطمینانی برای چندک مرتبه‌ی p ام جامعه‌ی اصلی، ضریب اطمینان مربوط بنا به لم ۲ عبارت است از

$$\begin{aligned} P(M_{i:k} \leq \xi_p \leq M_{j:k}) &= P(M_{i:k} \leq \xi_p) - P(M_{j:k} \leq \xi_p) \\ (۳) \qquad \qquad \qquad &= \sum_{r=i}^{j-1} \sum_{A_r} p^{\sum_{s=1}^r n_{ts} \alpha_{ts}} \prod_{s=r+1}^k (1 - p^{n_{ts} \alpha_{ts}}), \end{aligned}$$

که در آن A_r در لم ۲ تعریف شده است.

نتیجه‌ی ۱ روابط (۲) و (۳) در حالات خاص به صورت زیر ساده می‌شوند.

(آ) حالتی را در نظر بگیرید که برای $i, i = 1, \dots, k$ ، داشته باشیم $\alpha_i = 1$ و $n_i = n$ ، در این صورت

$$(۴) \qquad \qquad \qquad P(M_{ij1} \leq \xi_p \leq M_{ij2}) = 2p^n (1 - p^n)$$

و

$$(۵) \quad P(M_{i:k} \leq \xi_p \leq M_{j:k}) = \sum_{r=i}^{j-1} \binom{k}{r} p^{nr} (\lambda - p^n)^{k-r},$$

که در آن $\binom{k}{r} = \frac{k!}{r!(k-r)!}$.

(ب) فرض کنید که برای $i = 1, \dots, k$ ، داشته باشیم، $\alpha_i = \frac{1}{n_i}$. در این صورت:

$$(۶) \quad P(M_{ij\lambda} \leq \xi_p \leq M_{ij\lambda}) = 2p(\lambda - p)$$

و

$$(۷) \quad P(M_{i:k} \leq \xi_p \leq M_{j:k}) = \sum_{r=i}^{j-1} \binom{k}{r} p^r (\lambda - p)^{k-r}.$$

(ج) اگر برای $i = 1, \dots, k$ ، $\alpha_i = \frac{n^*}{n_i}$ که در آن $n^* = \sum_{i=1}^k n_i$ ، آنگاه

$$(۸) \quad P(M_{ij\lambda} \leq \xi_p \leq M_{ij\lambda}) = 2p^{n^*} (\lambda - p^{n^*})$$

و

$$(۹) \quad P(M_{i:k} \leq \xi_p \leq M_{j:k}) = \sum_{r=i}^{j-1} \binom{k}{r} p^{rn^*} (\lambda - p^{n^*})^{k-r}.$$

سؤال: i و j مطلوب را چگونه انتخاب کنیم؟

واضح است که i و j را باید به گونه ای انتخاب کنیم که برای ضریب اطمینان α و مرتبه ی چندک داده شده، بازه ی اطمینان مربوط دارای کوتاه ترین طول باشد به طوری که احتمال پوشش آن حد اقل به اندازه ی α باشد. ملاحظه می شود که در هر یک از حالت های (آ)، (ب) و (ج) مقدار $P(M_{i:k} \leq \xi_p \leq M_{j:k})$ بر اساس یک توزیع دوجمله ای محاسبه می شود و در هر حالت می توانیم i و j را طوری تعیین کنیم که بنا به مطلوبیت موجود $P(M_{i:k} \leq \xi_p \leq M_{j:k})$ بزرگ تر از $P(M_{ij\lambda} \leq \xi_p \leq M_{ij\lambda})$ و یا بزرگ تر از احتمال از پیش تعیین شده ی α باشد. اما انتخاب i و j یکتا نیست و انتخابی که $j - i$ را حد اقل می کند، مناسب به نظر می رسد. از طرفی می دانیم تابع جرم احتمال دوجمله ای $C(r, k)p^r(\lambda - p)^{k-r}$ یک تابع تک مدی با مد $[kp]$ ، بزرگ ترین مقدار صحیح کوچک تر یا مساوی با kp است که این تابع با معلوم بودن k و p تا حدود $[kp]$ افزایش و سپس کاهش می یابد. بنا بر این، اگر بخواهیم به حد اقل $j - i$ برسیم، باید با i و j نزدیک به $[kp]$ شروع کنیم و به تدریج $j - i$ را افزایش دهیم تا به وضعیت مطلوب برسیم. توجه

شود که در حالت (ب)، وقتی که $p = 0.5$ ، کوچک‌ترین مقدار $i - j$ با انتخاب $j = k - i + 1$ به دست می‌آید. گاهی ممکن است با انتخاب بازه‌ی اطمینان $(M_{i:k}, M_{j:k})$ دقیقاً به سطح اطمینان مورد نظر α_0 برسیم. در این صورت می‌توانیم بازه‌ای به شکل (M_L, M_U) را طوری تعیین کنیم که برای $0 \leq \lambda \leq 1$ داشته باشیم $M_L = (1 - \lambda)M_{i:k} + \lambda M_{i+1:k}$ و $M_U = (1 - \lambda)M_{k-i+1:k} + \lambda M_{k-i:k}$. فرض کنید بازه‌های $(M_{i:k}, M_{k-i+1:k})$ ، $(M_{i+1:k}, M_{k-i:k})$ و (M_L, M_U) به ترتیب دارای ضرایب اطمینان γ_i ، γ_{i+1} و α_0 باشند. در این صورت $\alpha_0 \in [\gamma_{i+1}, \gamma_i]$ و λ از رابطه‌ی زیر محاسبه می‌شود.

$$\lambda = \frac{(k-i)I}{i + (k-i)I}, \quad I = \frac{\gamma_i - \alpha_0}{\gamma_i - \gamma_{i+1}}.$$

۴ بازه‌های اطمینان بر اساس مینیمم نمونه‌ها

در این بخش برای چندک مرتبه‌ی p ام جامعه‌ی اصلی، $\xi_p = F^{-1}(p)$ ، بر اساس مینیمم نمونه‌ها، M'_{i,n_i} ، برآوردگر بازه‌ای آزاد توزیع پیدا می‌کنیم. ابتدا توجه شود که اگر برای $i = 1, \dots, k$ و $j = 1, \dots, n_i$ ، $X_{ij} \sim F^{\alpha_i}$ آن‌گاه $1 - (1 - F^{\alpha_i})^{n_i} \sim M'_{i,n_i}$. فرض کنید بخواهیم بازه‌ی اطمینان مبتنی بر مینیمم دو نمونه‌ی متفاوت (i و j ، $i \neq j$) را مورد بررسی قرار دهیم. در این حالت مرتب شده‌ی این دو آماره را به صورت $M'_{ij1} < M'_{ij2}$ نشان می‌دهیم که احتمال پوشش بازه‌ی اطمینان مربوط به صورت زیر محاسبه می‌شود.

$$\begin{aligned} P(M'_{ij1} \leq \xi_p \leq M'_{ij2}) &= P(M'_{i,n_i} \leq \xi_p \leq M'_{j,n_j}) + P(M'_{j,n_j} \leq \xi_p \leq M'_{i,n_i}) \\ &= \{1 - (1 - p^{\alpha_i})^{n_i}\}(1 - p^{\alpha_j})^{n_j} \\ &\quad + \{1 - (1 - p^{\alpha_j})^{n_j}\}(1 - p^{\alpha_i})^{n_i}. \end{aligned} \quad (10)$$

اما اگر بخواهیم از بین آماره‌های $M'_{1:k} < M'_{2:k} < \dots < M'_{k:k}$ دو آماره را به عنوان یک برآوردگر بازه‌ای برای ξ_p در نظر بگیریم، در این صورت ضریب اطمینان بازه‌ی مورد نظر، بنا به لم ۲ عبارت است از

$$(11) \quad P(M'_{i:k} \leq \xi_p \leq M'_{j:k}) = \sum_{r=i}^{j-1} \sum_{A_r} \{1 - (1 - p^{\alpha_{t_s}})^{n_{t_s}}\} \prod_{s=r+1}^k (1 - p^{\alpha_{t_s}})^{n_{t_s}},$$

که در آن A_r در لم ۲ تعریف شده است.

نتیجه‌ی ۲ در حالت خاص که برای $i = 1, \dots, k$ داشته باشیم، $\alpha_i = 1$ و $n_i = n$ ، روابط بالا

به صورت زیر ساده می شوند.

$$(۱۲) \quad P(M'_{ij1} \leq \xi_p \leq M'_{ij2}) = 2\{1 - (1-p)^n\}(1-p)^n$$

و

$$(۱۳) \quad P(M'_{i:k} \leq \xi_p \leq M'_{j:k}) = \sum_{r=i}^{j-1} \binom{k}{r} \{1 - (1-p)^n\}^r (1-p)^{n(k-r)}.$$

ملاحظه می شود که مقدار $P(M'_{i:k} \leq \xi_p \leq M'_{j:k})$ بر اساس یک توزیع دوجمله ای با پارامترهای k و $1 - (1-p)^n$ محاسبه می شود و بنا بر این، تمام مطالبی را که در بخش ۳ بیان کردیم برای این بخش نیز قابل تعمیم هستند.

۵ بازه های اطمینان بر اساس مینیمم و ماکسیمم نمونه ها

منطقی به نظر می رسد که در طرح نمونه گیری مورد بررسی در این مقاله، بازه های اطمینان به دست آمده بر اساس ماکسیمم نمونه ها برای چندک های بالا ($p \geq 0.5$) و انواع مبتنی بر مینیمم نمونه ها برای چندک های پایین ($p \leq 0.5$)، بهتر عمل کنند که البته این موضوع در بخش ششم به طور مفصل تحلیل خواهد شد. اما گاهی هم مینیمم و هم ماکسیمم نمونه ها را در اختیار داریم که در این صورت ترجیح می دهیم از اطلاعات موجود در هر دو آماره استفاده کنیم. در این بخش برای چندک مرتبه p ام جامعه ی اصلی، $\xi_p = F^{-1}(p)$ ، در دو حالت زیر طوری برآوردگر بازه ای آزاد توزیع پیدا می کنیم که هم مینیمم و هم ماکسیمم نمونه ها در آن مؤثر باشند.

۵/۱ برآورد بر اساس مینیمم و ماکسیمم های مستقل از هم

در این حالت اگر آماره ی مینیمم را از یک نمونه و آماره ی ماکسیمم را از نمونه ای دیگر انتخاب کنیم، از آنجایی که نمونه ها مستقل از هم می باشند، لذا آماره های مورد نظر نیز مستقل از هم هستند. بنا بر این، بازه ی اطمینانی به صورت (M'_{i,n_i}, M_{j,n_j}) ، دارای ضریب اطمینان زیر است

$$(۱۴) \quad \begin{aligned} P(M'_{i,n_i} \leq \xi_p \leq M_{j,n_j}) &= P(M'_{i,n_i} \leq \xi_p)P(M_{j,n_j} \geq \xi_p) \\ &= \{1 - (1-p^{\alpha_i})^{n_i}\}(1-p^{\alpha_j n_j}). \end{aligned}$$

اما ممکن است در اینجا نیز مشکلی مشابه بخش سوم رخ دهد، یعنی $M'_{i,n_i} > M_{j,n_j}$. در این حالت بازه‌ی اطمینان (M'_{i,n_i}, M_{j,n_j}) برای ξ_p بی‌معنی خواهد بود. حال اگر آماره‌های مرتب شده‌ی M_{j,n_j} و M'_{i,n_i} را به صورت $V_{ij1} \geq V_{ij2}$ در نظر بگیریم، آنگاه

$$\begin{aligned} P(V_{ij1} \leq \xi_p \leq V_{ij2}) &= P(M'_{i,n_i} \leq \xi_p \leq M_{j,n_j}) + P(M_{j,n_j} \leq \xi_p \leq M'_{i,n_i}) \\ &= P(M'_{i,n_i} \leq \xi_p)P(M_{j,n_j} \geq \xi_p) \\ &\quad + P(M_{j,n_j} \leq \xi_p)P(M'_{i,n_i} \geq \xi_p) \\ (15) \quad &= \{1 - (1 - p^{\alpha_i})^{n_i}\}(1 - p^{\alpha_j n_j}) + p^{\alpha_j n_j}(1 - p^{\alpha_i})^{n_i}. \end{aligned}$$

نتیجه‌ی ۳ در حالتی خاص که برای $i = 1, \dots, k$ داشته باشیم، $\alpha_i = 1$ و $n_i = n$ ، ضریب اطمینان موجود در رابطه‌ی (۱۵) به صورت زیر ساده می‌شود.

$$(16) \quad P(V_{ij1} \leq \xi_p \leq V_{ij2}) = \{1 - (1 - p)^n\}(1 - p^n) + p^n(1 - p)^n.$$

۵/۲ برآورد بر اساس مینیمم و ماکسیمم‌های وابسته

حالتی را در نظر بگیرید که هر دو آماره‌ی مینیمم و ماکسیمم را از یک نمونه انتخاب کرده باشیم. در این صورت بدیهی است که این آماره‌ها به هم وابسته هستند. فرض کنید M_{i,n_i} ماکسیمم نمونه‌ای تصادفی به اندازه‌ی n_i از جامعه‌ی F^{α_i} و M'_{i,n_i} مینیمم همان نمونه باشد. در این صورت با توجه به توزیع آماره‌های یاد شده، که به ترتیب در بخش‌های ۳ و ۴ آمده است و لم ۴، $M_{n_i}^* = F^{\alpha_i}(M_{i,n_i})$ و $M'_{n_i} = F^{\alpha_i}(M'_{i,n_i})$ به ترتیب هم‌توزیع با ماکسیمم و مینیمم نمونه‌ای تصادفی به اندازه‌ی n_i از توزیع $U(0, 1)$ هستند. بنا بر این، احتمال پوشش بازه‌ی اطمینانی به صورت (M'_{i,n_i}, M_{i,n_i}) برای ξ_p بنا به لم ۱ به صورت زیر محاسبه می‌شود.

$$\begin{aligned} P(M'_{i,n_i} \leq \xi_p \leq M_{i,n_i}) &= P(M'_{i,n_i} \leq F^{-1}(p) \leq M_{i,n_i}) \\ &= P\{F^{\alpha_i}(M'_{i,n_i}) \leq p^{\alpha_i} \leq F^{\alpha_i}(M_{i,n_i})\} \\ &= P(M'_{n_i} \leq p^{\alpha_i} \leq M_{n_i}^*) \end{aligned}$$

$$\begin{aligned}
 &= \int_0^{p^{\alpha_i}} \int_{p^{\alpha_i}}^1 n_i(n_i - 1)(y - x)^{n_i-2} dy dx \\
 &= \int_0^{p^{\alpha_i}} \int_{p^{\alpha_i}}^1 f_{M'_{n_i}, M_{n_i}}^*(x, y) dy dx \\
 (17) \quad &= 1 - p^{\alpha_i n_i} - (1 - p^{\alpha_i})^{n_i}.
 \end{aligned}$$

نتیجه‌ی ۴ در حالت خاصی که برای $i = 1, \dots, k$ ، داشته باشیم، $\alpha_i = 1$ و $n_i = n$ ، ضریب اطمینان بالا به صورت زیر محاسبه می‌شود.

$$(18) \quad P(M'_{i,n} \leq \xi_p \leq M_{i,n}) = 1 - p^n - (1 - p)^n.$$

۶ بحث و نتیجه‌گیری

در این بخش با توجه به مطالب بیان شده در بخش‌های پیشین، نقش آماره‌های مینیمم و ماکسیمم را در پیدا کردن بازه‌ی اطمینان‌های آزاد توزیع برای چندک‌های جامعه‌ی اصلی؛ یعنی F ، مشخص نموده و تا حد امکان تعیین می‌کنیم که در چه مواقعی، کدام نوع از آن‌ها بهتر عمل می‌کنند. مطالب خود را در قالب چند نتیجه در حالتی خاص که برای $i = 1, \dots, k$ ، داشته باشیم $\alpha_i = 1$ و $n_i = n$ ، به تفصیل بیان می‌کنیم.

نتیجه‌ی ۵ با توجه به روابط (۴) و (۱۲) و با تعریف تابع

$$h(p) = P(M_{ij1} \leq \xi_p \leq M_{ij2}) - P(M'_{ij1} \leq \xi_p \leq M'_{ij2}),$$

ملاحظه می‌شود که این تابع نسبت به نقطه‌ی $(0.5, 0)$ متقارن است؛ یعنی $h(p) > 0$ ، $p > 0.5$ و از طرفی برای $p > 0.5$ ، $h(0.5 + p) = -h(0.5 - p)$. بنا بر این، اگر $p > 0.5$ ، آن‌گاه $P(M_{ij1} \leq \xi_p \leq M_{ij2}) > P(M'_{ij1} \leq \xi_p \leq M'_{ij2})$. به طور مشابه می‌توان نشان داد که اگر $p > 0.5$ ، آن‌گاه $P(M_{i:k} \leq \xi_p \leq M_{i:k}) > P(M'_{i:k} \leq \xi_p \leq M'_{i:k})$ و اگر $p < 0.5$ ، آن‌گاه جهت نابرابری بالا عوض خواهد شد؛ یعنی، در این حالت اگر $p > 0.5$ ، آن‌گاه برآوردگر بازه‌ای بر اساس ماکسیمم دو نمونه‌ی متفاوت دارای ضریب اطمینان بیشتری است و اگر $p < 0.5$ ، آن‌گاه برآوردگر بازه‌ای بر اساس مینیمم دو نمونه‌ی متفاوت بهتر است. همچنین توجه شود که $h(0.5) = 0$ ، لذا در برآورد میانه‌ی جامعه؛ یعنی، $\xi_{0.5}$ ، مینیمم و ماکسیمم دو نمونه‌ی متفاوت رفتاری یکسان دارند.

نتیجه‌ی ۶ با توجه به رابطه‌ی (۱۶) ملاحظه می‌شود که تابع $f(p) = P(V_{ij1} \leq \xi_p \leq V_{ij2})$ نسبت به خط $p = 0.5$ متقارن است؛ یعنی، $f(0.5 + p) = f(0.5 - p)$. از طرفی این تابع با افزایش p تا 0.5 افزایش و سپس کاهش می‌یابد. بنا بر این، بازه‌ی اطمینان برای چندک مرتبه‌ی p ام جامعه‌ی F بر اساس مینیم یک نمونه و ماکسیمم نمونه‌ای دیگر برای مقادیر نزدیک به 0.5 ، دارای ضریب اطمینان بیش‌تری است.

در نتایج ۷، ۸ و ۹ به تشریح این مطلب می‌پردازیم که با داشتن هم مینیمم و هم ماکسیمم نمونه‌ها، استفاده از کدام نوع آن‌ها مطلوب‌تر است.

نتیجه‌ی ۷ با توجه به روابط (۴) و (۱۶) و با تعریف تابع

$$g(p) = P(V_{ij1} \leq \xi_p \leq V_{ij2}) - P(M_{ij1} \leq \xi_p \leq M_{ij2}),$$

ملاحظه می‌شود که $g(\sqrt[n]{0.5}) = 0$. از طرفی برای $0 < p < \sqrt[n]{0.5}$ ، $g(p) > 0$ و برای $\sqrt[n]{0.5} < p < 1$ ، $g(p) < 0$. بنا بر این، اگر $0 < p < \sqrt[n]{0.5}$ ، آن‌گاه به منظور تعیین برآوردگر بازه‌ای برای ξ_p ، در صورت امکان از مینیمم یک نمونه و ماکسیمم نمونه‌ای دیگر استفاده می‌کنیم، در حالی که اگر $\sqrt[n]{0.5} < p < 1$ ، آن‌گاه ترجیح می‌دهیم از ماکسیمم‌های دو نمونه‌ی مستقل برای رسیدن به منظور بالا استفاده نماییم. همچنین برای $p = \sqrt[n]{0.5}$ دو نسخه‌ی بالا رفتار یکسانی دارند.

نتیجه‌ی ۸ مشابه نتیجه‌ی ۷، با توجه به روابط (۱۲) و (۱۶) و با تعریف تابع

$$u(p) = P(V_{ij1} \leq \xi_p \leq V_{ij2}) - P(M'_{ij1} \leq \xi_p \leq M'_{ij2}),$$

ملاحظه می‌شود که $u(1 - \sqrt[n]{0.5}) = 0$. از طرفی برای $0 < p < 1 - \sqrt[n]{0.5}$ ، $u(p) < 0$ و برای $1 - \sqrt[n]{0.5} < p < 1$ ، $u(p) > 0$. بنا بر این، اگر $0 < p < 1 - \sqrt[n]{0.5}$ ، آن‌گاه به منظور تعیین برآوردگر بازه‌ای برای ξ_p ، در صورت امکان از مینیمم یک نمونه و ماکسیمم نمونه‌ای دیگر استفاده می‌کنیم، در حالی که اگر $1 - \sqrt[n]{0.5} < p < 1$ ، آن‌گاه ترجیح می‌دهیم از مینیمم‌های دو نمونه‌ی مستقل برای رسیدن به منظور بالا استفاده کنیم. همچنین برای $p = 1 - \sqrt[n]{0.5}$ دو نسخه‌ی بالا رفتاری یکسان دارند. اکنون در نتیجه‌ی ۹ تلفیقی از نتایج ۷ و ۸ را در حالت کلی‌تر بیان می‌کنیم.

نتیجه‌ی ۹ با توجه به نتایج ۷ و ۸، در صورتی که محدودیتی از لحاظ استفاده از مینیمم یا ماکسیمم نمونه‌ها نداشته باشیم، آن‌گاه به منظور تعیین برآوردگر بازه‌ای برای ξ_p به روش زیر عمل می‌کنیم: اگر $0 < p < 1 - \sqrt[n]{0.5}$ ، آن‌گاه از مینیمم‌های دو نمونه‌ی مستقل و در صورتی که $\sqrt[n]{0.5} < p < 1$ ، از ماکسیمم‌های دو نمونه‌ی مستقل و در غیر این صورت از مینیمم یک نمونه و ماکسیمم نمونه‌ای دیگر استفاده می‌کنیم.

نتیجه‌ی ۱۰ با توجه به روابط (۱۶) و (۱۸) ملاحظه می‌شود مقدار $P(V_{ij1} \leq \xi_p \leq V_{ij2})$ به اندازه‌ی $2p^n(1-p)^n$ از $P(M'_{i,n} \leq \xi_p \leq M_{i,n})$ بیش‌تر است؛ یعنی، مینیمم و ماکسیمم‌های مستقل از هم بهتر از نوع وابسته‌ی آن‌ها عمل می‌کنند.

نتیجه‌ی ۱۱ در رابطه‌ی (۵) می‌توانیم i و j را طوری تعیین کنیم که برای مقادیر مختلف k ، $P(M_{i:k} \leq \xi_p \leq M_{j:k})$ از ۰/۹۵ بیش‌تر باشد. این مقادیر را با استفاده از نسخه‌ی ۸ نرم‌افزار Maple برای $p = ۰/۸۵, ۰/۹$ و $n = ۵$ ، $k = ۱۰, ۲۰$ محاسبه نموده و نتایج آن را در جدول ۱ آورده‌ایم.

جدول ۱. مقادیر مطلوب i و j برای n ، p و k مختلف

(p, k, n)	i	j
$(۰/۸۵, ۱۰, ۵)$	$= ۱$	≥ ۸
	$= ۲$	≥ ۹
$(۰/۹۰, ۱۰, ۵)$	≤ ۲	≥ ۹
	$= ۳$	$= ۱۰$
$(۰/۸۵, ۲۰, ۵)$	≤ ۵	≥ ۱۴
$(۰/۹۰, ۲۰, ۵)$	≤ ۵	≥ ۱۶
	$۶ \leq i \leq ۸$	≥ ۱۷

نتیجه‌ی ۱۲ در رابطه‌ی (۷) می‌توانیم i و j را طوری تعیین کنیم که برای مقادیر مختلف k ، $P(M_{i:k} \leq \xi_p \leq M_{j:k})$ از ۰/۹۵ بیش‌تر باشد. این مقادیر را با استفاده از نسخه‌ی ۸ نرم‌افزار Maple برای $p = ۰/۲۵, ۰/۵, ۰/۶, ۰/۷۵$ و $k = ۱۰, ۲۰$ محاسبه نموده و نتایج آن را در جدول ۲ آورده‌ایم. توجه شود که رابطه‌ی (۷) به اندازه‌ی نمونه بستگی ندارد و لذا برای هر اندازه‌ی نمونه‌ای برقرار است.

جدول ۲. مقادیر مطلوب i و j برای p و k مختلف

(p, k)	i	j
$(\circ/5^\circ, 1^\circ)$	≤ 2	≥ 9
$(\circ/6^\circ, 1^\circ)$	≤ 2 $= 3$	≥ 9 $= 10$
$(\circ/5^\circ, 2^\circ)$	≤ 6	≥ 15
$(\circ/6^\circ, 2^\circ)$	≤ 8	≥ 17
$(\circ/25^\circ, 2^\circ)$	$= 1$ $= 2$	≥ 9 ≥ 10
$(\circ/75^\circ, 2^\circ)$	≤ 11	≥ 19

تقدیر و تشکر

نویسندگان مقاله از پیشنهادات و نظرات داوران محترم که در بهبود مقاله مؤثر واقع شد، تقدیر و تشکر می‌کنند.

مرجع‌ها

- Arnold, B.C.; Balakrishnan, N.; Nagaraja, H.N. (1992). *A First Course in Order Statistics*. Wiley, New York.
- Ahmadi, J.; Arghami, N.R. (2003). Nonparametric confidence and tolerance interval from record values data. *Statist. Papers*, **44**, 455-468.
- Ahmadi, J.; Balakrishnan, N. (2004). Confidence intervals for quantiles in terms of record range. *Statist. Probab. Lett.*, **68**, 395-405.
- Ahmadi, J.; Balakrishnan, N. (2005). Distribution-free confidence intervals for quantile intervals based on current records. *Statist. Probab. Lett.*, **75**, 190-202.
- Chen, Z. (2000). On ranked-set sample quatiles and their applications. *J. Statist. Plann. Inference*, **83**, 125-135.
- David, H.A.; Nagaraja, H.N. (2003). *Order Statistics*. third edition, Wiley, New York.
- Gulati, S.; Padgett, W.J. (2003). *Parametric and Nonparametric Inference from Record-Breaking Data*. Lecture Notes in Statistics, **172**, Springer, New York.
- Khaledi, B.; Kochar, S. (2004). Ordering convolutions of gamma random variables. *Sankhya*, **66**, 466-473.

- Krewski, D. (1976). Distribution-free confidence intervals for quantile intervals. *J. Amer. Statist. Assoc.*, **71**, 420-422.
- Nevzorov, V.B. (1990). Records for nonidentical distributed random variables. *Proceeding of the 5th Vilnius conference*, **2**, 227-233, VSP, Mokslas.
- Ozturk, O.; Deshpande, J. V. (2006). Ranked-set sample nonparametric quantile confidence intervals. *J. Statist. Plann. Inference*, **136**, 570-577.
- Sathe, Y.S.; Lingras, S.R. (1981). Bounds for the confidence coefficient of outer and inner confidence intervals for quantile interval. *J. Amer. Statist. Assoc.*, **76**, 473-475.
- Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley series in Probability and Mathematical Statistics, Wiley, New York.
- Zhou, X.; Sun, L.; Ren, H. (2000). Quantile estimation for left truncated and right censored data. *Statist. Sinica*, **10**, 1217-1229.
- Zielinski, R.; Zielinski, W. (2005). Best exact nonparametric confidence intervals for quantiles. *Statistics*, **39**, 67-71.

دریافت: ۱۱ دی ۱۳۸۴
آخرین اصلاح: ۱۳ اردیبهشت ۱۳۸۵

مصطفی رزمخواه	جعفر احمدی
گروه آمار،	گروه آمار،
دانشکده‌ی علوم ریاضی،	دانشکده‌ی علوم ریاضی،
دانشگاه فردوسی،	دانشگاه فردوسی،
مشهد، ایران.	مشهد، ایران.
پایان‌نگار: razmkhah_m@yahoo.com	پایان‌نگار: ahmadi@math.um.ac.ir

بهاره خطیب آستانه
گروه آمار،
دانشکده‌ی علوم ریاضی،
دانشگاه فردوسی،
مشهد، ایران.
پایان‌نگار: khatib_b@yahoo.com