بازه های اطمینان درستنمایی تجربی

سید جواد نقیبی　　ناصر رضا ارقامی [۱]

چکیده

در این مقاله ضمن معرفی روش درستنمایی تجربی، طریقهٔ محاسبهٔ بازه های اطمینان ناپارامتری برای میانگین جامعه به وسیلهٔ این روش و روش بوت استرپ بررسی شده و نتایج به کمک شبیه سازی با سایر روشهای ناپارامتری دیگر مقایسه شده اند.

واژه های کلیدی: روشهای ناپارامتری، درستنمایی ماکسیمم، تابع توزیع تجربی، بوت استرپ (خود گردان)، درستنمایی تجربی

۱ مقدمه

شاید یکی از مهمترین مسائلی که در آمار ناپارامتری مطرح است مسألهٔ پیدا کردن یک بازهٔ اطمینان برای یک پارامتر، بخصوص برای میانگین جامعه است. همان طور که می دانید تاکنون روشهای مختلفی برای محاسبهٔ یک بازهٔ اطمینان در حالت ناپارامتری ابداع شده اند. برای مثال، اگر پارامتر مورد نظر میانگین باشد، با پذیرش فرض تقارن توزیع، می توان از روش ویلکاکسون استفاده کرد یا در صورت قبول فرض نرمال بودن داده ها، می توان بازه های t ی استیودنت را به کار برد. یا در حالتی دیگر ممکن است آمارگر در به دست آوردن بازهٔ اطمینان برای چندکهای یک توزیع، از یک روش کلاسیک که مبتنی بر خواص آماره های ترتیبی توزیع یکنواخت است استفاده کند. اما دو روش کلیتر و جامعتر که اخیراً به این روشهای ناپارامتری اضافه شده اند عبارت اند از روش بوت استرپ (۱۹۷۹) و روش درستنمایی تجربی (۱۹۸۸). پایه و اساس هر دوی این روشها مبتنی بر تابع توزیع تجربی و خواص آن است و هر دو روش، خواص نمونه ای تقریباً یکسانی دارند. اما ما در اینجا نشان می دهیم روش درستنمایی تجربی، هم از نظر زمان و هم از نظر کارایی، در اکثر موارد از روش بوت استرپ (و سایر روشهای ناپارامتری دیگر) بهتر عمل می کند.

۲ بوت استرپ (خود گردان)

در سال ۱۹۷۹ افران [۲] روشی را بر مبنای اصل جایگذاری تابع توزیع تجربی و کاربرد روش جک نایف ابداع کرد و روش خود را بوت استرپ نام نهاد. به طور کلی اصل اساسی در استفاده از روش بوت استرپ، برآورد هر تابعی از تابع توزیع برابر با مقدار همان تابع از تابع توزیع تجربی است. فرض کنید به منظور برآورد $\theta = T(F)$، می خواهیم از برآوردگر $\hat{\theta} = S(X)$

[۱] دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

[۲] Efron

استفاده کنیم. با استفاده از روش بوت استرپ علاوه بر برآورد θ به وسیلهٔ $\hat{\theta}^* = S(x^*)$ (که آن را برآورد بوت استرپ θ می نامیم)، همهٔ ویژگیهای $\hat{\theta}$ (توزیع، واریانس، انحراف معیار و اریبی و ...) را می توان به وسیلهٔ ویژگیهای متغیر تصادفی $\hat{\theta}^* = S(X^*)$ برآورد کرد که در آن X^*، یک نمونهٔ تصادفی با جایگذاری، استخراج شده از توزیع تجربی F_n، است که به آن نمونهٔ بوت استرپ می گوییم:

$$X^* = (X_1^*,\ldots,X_n^*) \quad , \quad X_1^*,\ldots,X_n^* \overset{iid}{\sim} F_n$$

برای به دست آوردن X^* کافی است که از نمونهٔ اولیهٔ X، نمونه های تصادفی n تایی با جایگذاری استخراج کنیم. تعداد نمونه های ممکنی که از توزیع F_n استخراج می شوند، متناهی است. بنابراین می توان از دید نظری، کلیهٔ نمونه های ممکن را به دست آورده و برای هر یک از آنها $\hat{\theta}^* = S(X^*)$ را محاسبه کرد، ولی با افزایش مقدار n، تعداد نمونه های ممکن افزایش می یابد و محاسبات، طولانی و وقت گیر خواهد شد. لذا در عمل، نمونه های خود را از توزیع F_n به B نمونهٔ بوت استرپ محدود می کنیم. طبیعی است که هر چه B بزرگتر باشد دقت برآورد ما بیشتر است و وقتی B به بزرگترین مقدار خود میل می کند، مانند این است که همهٔ نمونه های ممکن را استخراج کرده ایم.

۲ـ۱ بازه های اطمینان صدکی بوت استرپ

فرض کنید $G_{\hat{\theta}}^*$ تابع توزیع بوت استرپ $\hat{\theta} = T(F_n)$ برای برآورد $\theta = T(F)$ باشد. روش صدکی، بازهٔ اطمینان در سطح $1-2\alpha$ را برای θ بر اساس صدکهای توزیع بالا به صورت زیر پیشنهاد می کند:

$$(\underline{\hat{\theta}}\ ,\ \hat{\bar{\theta}}) = (G^{*-1}(\alpha) \text{ و } G^{*-1}(1-\alpha))$$

روشن است که اگر $\hat{\theta}_\alpha^*$، چندک α ام توزیع G^* باشد طبق تعریف داریم: $G^*(\hat{\theta}_\alpha^*) = \alpha$. بنابراین، بازهٔ بالا را می توان به صورت زیر نشان داد:

$$(\underline{\hat{\theta}}\ ,\ \hat{\bar{\theta}}) = (\hat{\theta}_\alpha^* \text{ و } \hat{\theta}_{1-\alpha}^*)$$

افران در مقاله های خود توجیهی برای استفاده از روش بالا آورده است ولی متأسفانه به اثبات آن نپرداخته است. ما در اینجا به کمک قضیهٔ زیر، اثباتی برای توجیهی که افران ارائه کرده عرضه می کنیم:

قضیهٔ ۱ ـ فرض کنید $\hat{\theta}$ برآورد درستنمایی ماکسیمم θ بر اساس نمونهٔ تصادفی $X_1,\ldots,X_n$ از چگالی $f(x,\theta)$ باشد که دارای تابع توزیع G_θ است. همچنین فرض کنید $G_\theta(t)$ نسبت به θ اکیداً نزولی باشد، و تابع اکیداً صعودی ψ و ثابت τ وجود دارند به طوری که توزیع $Z = \tau[\psi(\hat{\theta}) - \psi(\theta)]$ یک توزیع متقارن (نسبت به مبدأ) مانند H است.[۳] در این صورت یک حد اطمینان بالای $1-\alpha$ برای θ عبارت است از $G_{\hat{\theta}}^{-1}(1-\alpha)$، یعنی چندک متناظر با احتمال $1-\alpha$ در توزیع $\hat{\theta}$ به ازای $\theta = \hat{\theta}$.

برای اثبات این قضیه، ابتدا دو لم زیر را ارائه می دهیم:

لم ۱ـ فرض کنید $\hat{\theta} = T = T(F_n)$ برآوردگر θ و $G_\theta(t)$ تابع توزیع $\hat{\theta}$ باشد که نسبت به θ اکیداً نزولی است. آنگاه جواب معادلهٔ $G_\theta(T) = \alpha$ (که در آن θ مجهول معادله است) یک کران اطمینان بالای $1-\alpha$ برای θ است.

برهان: فرض کنید $\bar{\theta}$ جواب معادلهٔ $G_\theta(T) = \alpha$ باشد و تابع $K_t(\theta)$ همان $G_\theta(t)$ باشد که در آن θ متغیر محسوب می شود. طبق فرض، $K_t(\theta)$ اکیداً نزولی است و با تعریف بالا $K_T^{-1}(\alpha) = \bar{\theta}$. لذا داریم:

$$P_\theta(\theta < \bar{\theta}) = P_\theta[\theta < K_T^{-1}(\alpha)] = P_\theta(K_T(\theta) > \alpha)$$

و چون $K_t(\theta)$ نسبت به θ نزولی است،

---

[۳] چون برآوردگرهای درستنمایی ماکسیمم به طور مجانبی نرمال اند، فرض بالا زیاد بیجا نیست. علاوه بر این خیلی از توزیعهای نامتقارن با یک تبدیل مناسب، به یک توزیع متقارن تبدیل می شوند. توزیع ضریب همبستگی یک نمونه از نرمال دو متغیره شاهدی بر این ادعاست.

$$= P_\theta(G_\theta(T) > \alpha) = P_\theta(U > \alpha) = 1 - \alpha$$

که در آن U، متغیری تصادفی است که توزیع یکنواخت در بازهٔ (۰,۱) دارد.

لم ۲ ـ تحت شرایط قضیهٔ (۱) همواره داریم:

$$G_\theta(t) = 1 - G_t(\theta) \qquad \forall t,\theta$$

برهان: با توجه به شرایط قضیهٔ (۱) داریم:

$$Z = \tau[\psi(\hat\theta) - \psi(\theta)] \sim H$$

که H یک توزیع متقارن نسبت به مبدأ است.

از طرفی می توان نوشت:

$$\begin{aligned} G_\theta(t) &= P_\theta(\hat\theta \le t) = P_\theta[\psi(\hat\theta) \le \psi(t)] \\ &= P_\theta\{\tau[\psi(\hat\theta) - \psi(\theta)] \le \tau[\psi(t) - \psi(\theta)]\} \\ &= P_\theta\{Z \le \tau[\psi(t) - \psi(\theta)]\} \\ &= H\{\tau[\psi(t) - \psi(\theta)] \end{aligned} \qquad (1)$$

چون H یک توزیع متقارن نسبت به مبدأ است نتیجه می شود:

$$H\{\tau[\psi(t) - \psi(\theta)]\} = 1 - H\{\tau[\psi(\theta) - \psi(t)]\}$$

که با مقایسه با رابطهٔ (۱) نتیجه می شود:

$$= 1 - G_t(\theta) \qquad \forall t,\theta .$$

برهان قضیهٔ ۱ :

بنابر تعریف $\bar\theta$ داریم: $K_T(\bar\theta) = G_{\bar\theta}(T) = \alpha$

که در آن $T = \hat\theta$ یک متغیر تصادفی و $\bar\theta$ تابعی از T است. برای اثبات قضیه باید نشان دهیم:

$$G_{\hat\theta}(\bar\theta) = 1 - \alpha$$

ولی طبق لم ۲ ثابت کردیم: $G_{\hat\theta}(\bar\theta) = 1 - G_{\bar\theta}(\hat\theta)$

و بنا به لم ۱ داریم: $= 1 - G_{\bar\theta}(T) = 1 - \alpha$

لذا حکم ثابت است.

با روشی مشابه می توان نشان داد $\underline{\theta}$ ، یعنی چندک متناظر با احتمال α در توزیع $\hat\theta$ ، به ازای $\theta = \hat\theta$ ، یک حد اطمینان پایین در سطح α برای θ است. علاوه بر این $G^*_{\hat\theta}$ (یعنی توزیع بوت استرپ $\hat\theta$) یک تقریب مناسب برای $G_{\hat\theta}$ است (زیرا به ازای $G_{\hat\theta}$ که از جامعه به دست می آید، $G^*_{\hat\theta}$ را داریم که از نمونه به دست می آید) و با به کار بردن $G^*_{\hat\theta}$ به جای $G_{\hat\theta}$ مقادیر $\bar\theta$ و $\underline{\theta}$ برآورد می شوند که ما مقادیر این برآوردها را به ترتیب با $\hat{\bar\theta}$ و $\hat{\underline{\theta}}$ نشان داده ایم.

کاملاً روشن است که در عمل نیازی به دانستن توابع H و ψ و ثابت τ نداریم. این مطلب کارایی این روش را بهتر روشن می کند. بنابراین به طور خلاصه الگوریتم محاسبهٔ بازهٔ اطمینان صدکی بوت استرپ را می توان به صورت زیر خلاصه کرد:

۱) استخراج B نمونهٔ مستقل بوت استرپ $\underset{\sim}{X}^{*1}, \underset{\sim}{X}^{*2}, \ldots, \underset{\sim}{X}^{*B}$ از نمونهٔ اولیهٔ $\underset{\sim}{X} = (X_1, \ldots, X_n)$.

۲) برآورد کردن پارامتر θ در هر یک از این B نمونه، یعنی محاسبه:

$$\hat\theta^{*b} = S(\underset{\sim}{x}^{*b}) \qquad b = 1,2,\ldots,B.$$

۳) محاسبهٔ صدکی α ۱۰۰ ام و $(1-\alpha)$۱۰۰ ام توزیع $\hat\theta^{*b}$ که آنها را به ترتیب با $\hat\theta^*_\alpha$ و $\hat\theta^*_{1-\alpha}$ نشان می دهیم.

۴) ساختن بازهٔ اطمینان صدک بوت استرپ در سطح $(1-2\alpha)$ برای θ به صورت: $(\hat\theta^*_\alpha, \hat\theta^*_{1-\alpha})$.

گاهی اوقات در نبود پیش فرضهای قضیهٔ (۱)، از روشهای دیگر بوت استرپ که در حقیقت تعمیمی از روش صدکی اند استفاده می شود. برای آشنایی با این روشها می توانید به [۴] و [۵] مراجعه کنید.

۳ درستنمایی تجربی

همانطور که می دانید اصل درستنمایی، یکی از اصول کاهش داده هاست که روشهای مبتنی بر آن، خواص ارزنده ای در استنباط پارامتری دارند. برآوردگرهای درستنمایی ماکسیمم نسبت به تبدیلها، پایا بوده و تحت بعضی شرایط ساده دارای خواص مجانبی ارزشمندی چون کارایی، سازگاری و نرمال بودن هستند. آزمونهای (LRT) در بیشتر موارد ناااریب بوده و در ردهٔ آزمونهای UMP یا

UMPU قرار می گیرند. مهمترین مشخصهٔ آماره های نسبت درستنمایی، توزیع مجانبی χ^2 برای تبدیل ساده ای از آنهاست.

یکی از روشهای محاسبهٔ بازهٔ اطمینان در حالت پارامتری، استفاده از ناحیهٔ پذیرش آزمون LR است، اما استفاده از روش درستنمایی در حالت ناپارامتری، رهیافتی کاملاً جدید است. در نگاه اول چنین به نظر می رسد که چون شکل و تعریف تابع درستنمایی شدیداً متأثر از شکل توزیع است، به دست آوردن بازهٔ اطمینان با این فن، فقط مخصوص حالت پارامتری است. استفاده از تابع درستنمایی در استنباط ناپارامتری، سالها مورد توجه دانشمندان آمار بوده است که از آن جمله می توان به تامس وگرانکمیر[4]، کاکس و افران[5] اشاره کرد. اما سرانجام در سال ۱۹۸۸ فردی به نام اون[6] توانست با تعریف نسبت درستنمایی تجربی، خواص آزمونهای نسبت درستنمایی را به حالت ناپارامتری تعمیم دهد و به کمک آن قضیهٔ ویلکس را در حالت ناپارامتری اثبات کند.

۱ـ۳ نسبت درستنمایی تجربی F

تعریف ـ فرض کنید $x_1, x_2, \ldots, x_n$ یافته های یک نمونهٔ تصادفی مستقل و همتوزیع از توزیع F باشند. تابع درستنمایی تجربی F را به صورت زیر تعریف می کنیم:

$$L(F) = \prod_{i=1}^{n} dF(x_i) = \prod_{i=1}^{n} w_i \qquad (2)$$

که در آن $w_i = dF(x_i) = P(X = x_i)\ ;\ i = 1,2,\ldots,n$.

لم ـ ماکسیمم مقدار $L(F)$ به ازای $F = F_n$ حاصل می شود.

برهان: فرض کنید $X_1, X_2, \ldots, X_n$ یک نمونهٔ تصادفی n تایی از F باشد، در این صورت اگر $\hat{F}$ برآوردگر درستنمایی ماکسیمم تجربی F باشد، طبق تعریف داریم:

$$L(\hat{F}) = \sup_{F \in \mathcal{F}} L(F) < \infty$$

که در آن $\mathcal{F}$ مجموعهٔ همهٔ توابع توزیع گسسته است.

اولاً اگر $S = \{x_1, x_2, \ldots, x_n\}$ مجموعهٔ مشاهدات نمونهٔ تصادفی باشد، به ازای هر $x_i \in S$ داریم: $f(x_i) > 0$ که f تابع چگالی احتمال F است. بنابراین اگر $\hat{f}$ برآورد f باشد باید $\hat{f}(x_i) > 0\ ;\ \forall x_i \in S$ یعنی تکیه گاه $\hat{F}$ باید شامل S باشد، زیرا در غیر این صورت اگر

$$\exists\ x_i \in S \qquad x_i \notin S_{\hat{F}}$$

فوراً نتیجه می شود $\hat{f}(x_i) = 0$ و از آنجا داریم $L(\hat{F}) = 0$.

ثانیاً تکیه گاه $\hat{F}$ باید دقیقاً برابر S باشد، زیرا اگر فرض کنید

$$S_{\hat{F}} = \{x_1, x_2, \ldots, x_n\} \cup \{x_{n'}\}$$

می توان $\hat{F}$ جدیدی ساخت که تکیه گاه آن S باشد و احتمال $x_{n'}$ بین مقادیر داخل S توزیع شود و در نتیجه به مقدار $L(\hat{F})$ بزرگتری دست پیدا کرد. پس برای یافتن $\sup L(F)$ روی مجموعهٔ همهٔ توابع توزیع گسسته، فقط باید به توابع توزیعی که تکیه گاه آنها S است توجه کرد. ما مجموعهٔ توابع توزیعی را که روی فضای نمونه ای تعریف می شوند به صورت زیر نشان می دهیم:

$$\mathcal{F}_n = \{F \mid F \ll F_n\}$$

حال نشان می دهیم که بیشترین مقدار $L(F)$ روی مجموعهٔ بالا به ازای $\hat{F} = F_n$ حاصل می شود. روشن است که اعضای $\mathcal{F}_n$ را می توان به صورت زیر نشان داد: $f(x_i) = w_i$ وقتی که

$$\sum_{i=1}^{n} w_i = 1 \quad ، \quad 0 \le w_i \le 1 \quad ، \quad i = 1,2,\ldots,n\ .$$

بنابراین

$$L(F) = \prod_{i=1}^{n} w_i$$

با استفاده از روش لاگرانژ داریم:

$$G = \sum_{i=1}^{n} \log w_i - \lambda\left(\sum_{i=1}^{n} w_i - 1\right)$$

---

4. Thomas and Grunkemeyer

5. Cox and Efron

6. Owen

$$\frac{\partial G}{\partial w_i}=\frac{1}{w_i}-\lambda=0 \Rightarrow w_i=\frac{1}{\lambda}$$

$$\sum_{i=1}^{n} w_i=1 \Rightarrow \sum_{i=1}^{n}\frac{1}{\lambda}=1 \Rightarrow \lambda=n \Rightarrow w_i=\frac{1}{n}$$

یعنی $\hat{F}$ تابع توزیع تجربی است، و

$$\operatorname*{Sup}_{F\in \mathcal{F}} L(F)=L(F_n)$$

تعریف ـ فرض کنید F_n تابع توزیع تجربی F باشد. نسبت درستنمایی تجربی F را به صورت زیر تعریف می کنیم:

$$R(F)=\frac{L(F)}{L(F_n)} \qquad 0\le R(F)\le 1$$

با به کار بردن رابطهٔ (۲)، داریم:

$$R(F)=\prod_{i=1}^{n} n w_i \qquad (۳)$$

حال فرض کنید به برآورد پارامتر $\theta=T(F)$ علاقمندیم. اصل جایگذاری، تابع توزیع تجربی $\hat{\theta}=T(F_n)$ را به عنوان برآورد (نقطه ای) درستنمایی ماکسیمم ناپارامتری θ پیشنهاد می کند ولی اگر به دنبال برآورد فاصله ای θ باشیم، ذیلاً نشان می دهیم تحت شرایطی معقول، مجموعهٔ $\{T(F)\mid R(F)\ge c, F\ll F_n\}$ را می توان به عنوان یک بازهٔ اطمینان برای θ به کار برد. برای سهولت، فرض کنید θ میانگین F باشد.

۳ـ۲ محاسبهٔ بازهٔ اطمینان برای میانگین به روش مستقیم

قضیهٔ ۲ ـ فرض کنید $X_1, X_2, \ldots$ و X متغیرهای تصادفی مستقل با تابع توزیع ناتباهیدهٔ F_0 باشند. همچنین برای $c<1$ فرض کنید:

$$\mathcal{F}_{c,n}=\{F\mid R(F)\ge c, F\ll F_n\} \qquad (۴)$$

تعریف می کنیم:

$$X_{u,n}=\operatorname*{Sup}_{F\in \mathcal{F}_{c,n}}\int x\,dF \quad \text{و} \quad X_{l,n}=\inf_{F\in \mathcal{F}_{c,n}}\int x\,dF$$

آنگاه فاصلهٔ $(X_{l,n}, X_{u,n})$ یک بازهٔ اطمینان برای $\mu=T(F_0)=E_{F_0}(X)$ است.

برهان: مسألهٔ آزمون فرض $H_0: F=F_0$ در مقابل $H_1: F\ne F_0$ را در نظر بگیرید، که در آن $F_0\in \mathcal{F}_n$. طبق اصل آزمون نسبت درستنمایی، H_0 را رد می کنیم اگر و فقط اگر

$$R(F_0)=\frac{L(F_0)}{L(F_n)}<c$$

با معکوس کردن ناحیهٔ رد این آزمون، یک نوار اطمینان برای F_0 به صورت زیر به دست می آید:

$$\mathcal{F}_{c,n}=\{F\mid \frac{L(F)}{L(F_n)}\ge c, F\in \mathcal{F}_{-}\}$$

حال داریم:

$$P(F_0\in \mathcal{F}_{c,n})\ge 1-\alpha \qquad (۵)$$

با تعریف $X_{u,n}$ و $X_{l,n}$ به صورت

$$X_{u,n}=\operatorname*{Sup}_{F\in \mathcal{F}_{c,n}}\int x\,dF, \quad \text{و} \quad X_{l,n}=\inf_{F\in \mathcal{F}_{c,n}}\int x\,dF$$

داریم:

$$F_0\in \mathcal{F}_{c,n} \Rightarrow \mu=E_{F_0}(X)\in (X_{l,n}, X_{u,n})$$

و از آنجا

$$P(F_0\in \mathcal{F}_{c,n})\le P(X_{l,n}\le \mu\le X_{u,n}) \qquad (۶)$$

با مقایسهٔ روابط (۵) و (۶) نتیجه می شود:

$$P(X_{l,n}\le \theta\le X_{u,n})\ge 1-\alpha$$

یعنی $(X_{l,n}, X_{u,n})$ یک بازهٔ اطمینان $(1-\alpha)100\%$ برای μ است. با به کار بردن فرمولهای (۲)، (۳) و (۴) در قضیهٔ بالا، مقادیر $X_{u,n}$ و $X_{l,n}$ به صورت زیر تغییر می یابند:

$$X_{u,n}=Sup\sum_{i=1}^{n} w_i x_i \quad , \quad X_{l,n}=\inf\sum_{i=1}^{n} w_i x_i$$

که مقادیر Sup و inf روی مقادیری از w_i هستند که در شرایط زیر صدق می کنند:

$$w_i \ge 0 \quad , \quad \sum_{i=1}^{n} w_i = 1 \quad , \quad \prod_{i=1}^{n} n w_i \ge c \qquad (۷)$$

بنابراین، مسألهٔ پیدا کردن بازهٔ اطمینان، به مسألهٔ برنامه ریزی ریاضی زیر منتهی می شود:

مقدار $\sum_{i=1}^{n} w_i x_i$ را با توجه به شرایط زیر بیشینه کنید:

(i) $w_i \ge 0$

(ii) $\sum_{i=1}^{n} w_i = 1$

(iii) $\prod_{i=1}^{n} n w_i \ge c$

با استفاده از روش لاگرانژ داریم:

$$G = \sum_{i=1}^{n} w_i x_i + \lambda_1 \left(1 - \sum_{i=1}^{n} w_i\right) + \lambda_2 \left(\log c - \sum_{i=1}^{n} \log n w_i\right)$$

$$\frac{\partial G}{\partial w_i} = x_i - \lambda_1 - \frac{\lambda_2}{w_i} \qquad i = 1,2,\dots,n$$

$$\frac{\partial G}{\partial w_i} = 0 \Rightarrow w_i = \frac{\lambda_2}{x_i - \lambda_1} \qquad (۸)$$

با به کار بردن قید (ii) در رابطهٔ (۸) خواهیم داشت:

$$\sum_{i=1}^{n} \frac{\lambda_2}{x_i - \lambda_1} = 1 \quad \Rightarrow \quad \lambda_2 = \frac{1}{\sum_{i=1}^{n} \frac{1}{x_i - \lambda_1}}$$

و از آنجا

$$w_i = \frac{1}{(x_i - \lambda_1) \sum_{i=1}^{n} \frac{1}{x_i - \lambda_1}} \qquad (۹)$$

که در آن λ_1 ریشهٔ معادلهٔ زیر است:

$$0 = \sum_{i=1}^{n} \log \frac{n}{(x_i - \lambda) \sum_{i=1}^{n} \frac{1}{x_i - \lambda}} - \log c \equiv f(\lambda). \qquad (۱۰)$$

دو لم زیر برای به دست آوردن λ_1 بسیار مفیدند.

لم ۱ـ مقادیر قابل قبول برای λ در معادلهٔ $f(\lambda) = 0$ عبارت است از:

$$\mathcal{A}_\lambda = \mathcal{A}_1 \cup \mathcal{A}_2$$

که در آن $\mathcal{A}_1 = (-\infty, x_{(1)})$ و $\mathcal{A}_2 = (x_{(n)}, +\infty)$.

برهان: اولاً اگر فرض کنیم $\lambda = x_{(1)}$ یا $\lambda = x_{(n)}$، آنگاه از رابطهٔ (۹) نتیجه می شود $w_i > 1$، که خلاف فرض است. پس فرض کنید $x_{(1)} < \lambda < x_{(n)}$. بدون از دست دادن کلیت، می توان فرض کرد $x_{(1)} < \lambda < x_{(2)} < \dots < x_{(n)}$، در این صورت با قرار دادن $S = \sum_{i=1}^{n} \frac{1}{x_i - \lambda}$، مقادیر S از دو حالت زیر خارج نیست:

۱) اگر $S > 0$ باشد، چون $X_{(1)} - \lambda < 0$، داریم:

$$w_1 = \frac{1}{(x_{(1)} - \lambda) S} < 0$$

۲) اگر $S < 0$ باشد، چون به ازای هر $i \neq 1$، $x_{(i)} - \lambda > 0$، در نتیجه $w_i < 0$. بنابراین در هر دو حالت مقادیر w_i غیر مجازند. لذا حکم ثابت است.

لم ۲ ـ تابع $f(\lambda)$ در هر یک از مجموعه های $\mathcal{A}_1$ و $\mathcal{A}_2$ فقط یک ریشه دارد.

برهان: فرض کنید $\lambda \in \mathcal{A}$. بنابراین $x_i - \lambda > 0$، $\forall i = 1,2,\dots,n$. پس می توان نوشت:

$$f = n \ln n - \sum_{i=1}^{n} \ln(x_i - \lambda) - n \ln \sum_{i=1}^{n} \frac{1}{x_i - \lambda} - \ln c$$

$$\frac{\partial f}{\partial \lambda} = \sum_{i=1}^{n} \frac{1}{x_i - \lambda} - n \frac{\sum_{i=1}^{n} \frac{1}{(x_i - \lambda)^2}}{\sum_{i=1}^{n} \frac{1}{x_i - \lambda}}$$

که با فرض $y_i = \frac{1}{x_i - \lambda}$، داریم: $\sum_{i=1}^{n} y_i > 0$ و

$$\frac{\partial f}{\partial \lambda}=\frac{\left(\sum_{i=1}^{n} y_i\right)^2-n\sum_{i=1}^{n} y_i^2}{\sum_{i=1}^{n} y_i}$$

$$=\frac{-n\left(\sum_{i=1}^{n} y_i^2-n\bar{y}^2\right)}{\sum_{i=1}^{n} y_i}<0$$

بنابراین f روی $\mathcal{A}_1$ نزولی است و علاوه بر این

$$\lim_{\lambda\to+\infty} f(\lambda)=\infty \quad , \quad \lim_{\lambda\to x_{(1)}} f(\lambda)=0$$

یعنی f در $\mathcal{A}_1$ فقط یک ریشه دارد. به روشی مشابه می توان نشان داد f در $\mathcal{A}_2$ نیز فقط یک ریشه دارد.

الگوریتم زیر دستورالعمل ساختن بازهٔ اطمینان درستنمایی تجربی برای میانگین را نشان می دهد:

۱) با انتخاب سطح آزمون α ، مقدار c را از رابطهٔ زیر پیدا کنید:

$$1-\alpha=P_r(\chi^2_{(1)}<-2\log c)$$

۲) با استفاده از یافته های نمونهٔ تصادفی $X_1,\dots,X_n$ و به کمک روشی تکراری با یک نقطه شروع (نیوتن ـ رافسون) ریشه های معادلهٔ

$$f(\lambda)=\sum_{i=1}^{n}\log\frac{n}{(x_i-\lambda)\sum_{i=1}^{n}\frac{1}{x_i-\lambda}}-\log c=0$$

را در هر یک از بازه های $\mathcal{A}_1=(-\infty,x_{(1)})$ و $\mathcal{A}_2=(x_{(n)},+\infty)$ پیدا کنید. (معادلهٔ بالا در هر یک از بازه های بالا فقط یک ریشه دارد).

۳) فرض کنید λ_1 و λ_2 به ترتیب ریشه های معادلهٔ بالا در $\mathcal{A}_1$ و $\mathcal{A}_2$ باشند. مقادیر w_{i1} و w_{i2} ، $i=1,2,\dots,n$ ، را از روابط زیر به دست آورید:

$$w_{i1}=\frac{1}{(x_i-\lambda_1)\sum_{i=1}^{n}\frac{1}{x_i-\lambda_1}}$$

$$w_{i2}=\frac{1}{(x_i-\lambda_2)\sum_{i=1}^{n}\frac{1}{x_i-\lambda_2}}$$

٤) متناظر با w_{i1} و w_{i2} قرار دهید:

$$X_{l,n}=\sum_{i=1}^{n} w_{i1}x_i \quad , \quad X_{u,n}=\sum_{i=1}^{n} w_{i2}x_i$$

۵) بازهٔ درستنمایی تجربی در سطح $1-\alpha$ برای میانگین عبارت است از: $(X_{l,n},X_{u,n})$.

از طرف دیگر یک راهبرد جالب برای به دست آوردن بازهٔ اطمینان بالا، استفاده از تابع نسبت درستنمایی تجربی θ (پارامتر F) است. (استفاده از این روش برای مشخصه هایی که در قالب یک M ـ برآوردگر [٦] قابل بیان اند، ساده تر به نظر می رسد). بخش بعدی با استفاده از تعریف و قضایا، رهنمون ما در استفاده از این روش است.

۳ـ۳ استفاده از تابع نسبت درستنمایی تجربی θ برای محاسبهٔ بازهٔ اطمینان

تعریف: مسألهٔ آزمون فرض $H_0: T(F)=\theta_0$ در مقابل $H_1: T(F)\neq\theta_0$ را که در آن $\theta=T(F)$ یک پارامتر F است در نظر بگیرید. تابع نسبت درستنمایی تجربی θ_0 را به صورت زیر تعریف می کنیم:

$$R_E(\theta_0)=\sup_F\{R(F)\mid T(F)=\theta_0, F\ll F_n\} \quad (11)$$

چون قبلاً نشان دادیم $L(F)$ در F_n ماکسیمم می شود، بنابراین فوراً نتیجه می شود که $\sup R_E(\theta)$ با توجه به θ در $\hat{\theta}=T(F_n)$ به ماکسیمم مقدار خود می رسد.

در حالتی که $\theta=\mu$ میانگین F است، با استفاده از روابط (۳) و (۷)، $R_E(\theta_0)$ به شکل زیر در می آید:

$$R_E(\mu_0)= \quad (12)$$

$$\sup\left\{\prod_{i=1}^{n} nw_i \mid w_i>0, \sum_{i=1}^{n} w_i=1, \sum_{i=1}^{n} w_ix_i=\mu_0\right\}$$

با توجه به استدلال بالا داریم:

$$\operatorname*{Sup}_{\mu} R_E(\mu) = R_E(\bar{x}) = ۱$$

و کاملاً روشن است که $\mu_\circ \in (X_{l,n}, X_{u,n})$ اگر و فقط اگر $R_E(\mu_\circ) > c$، (زیرا ناحیهٔ پذیرش آزمون بالا را می توان به عنوان بازهٔ اطمینان برای μ به کار برد).

بنابراین راهی مناسب در به دست آوردن بازهٔ اطمینان برای میانگین، جستجوی نقاطی مانند $\mu_\circ$ است که به ازای آنها داشته باشیم $R_E(\mu_\circ) > c$. فرض کنید $\hat{\mu}_l$ و $\hat{\mu}_u$ دو عدد باشند به طوری که دقیقاً داشته باشیم:

$$R_E(\hat{\mu}_u) = R_E(\hat{\mu}_l) = c$$

در این صورت استدلال بالا بازهٔ $(\hat{\mu}_l, \hat{\mu}_u)$ را به عنوان بازهٔ اطمینان برای μ معرفی می کند، زیرا

$$\forall \mu_\circ \in (\hat{\mu}_l, \hat{\mu}_u) \qquad R_E(\mu_\circ) > c$$

نمودار زیر درستی این رابطه را بهتر نشان می دهد.

قبل از اینکه نحوهٔ مشخص کردن $\hat{\mu}_l$ و $\hat{\mu}_u$ را بیان کنیم، باید چگونگی انتخاب c به عنوان نقطهٔ بحرانی مشخص شود. قضیهٔ زیر ما را در این مورد راهنمایی می کند.

قضیهٔ ۳ـ تحت شرایط قضیهٔ (۲)، اگر $\int |X|^۳ dF_\circ < \infty$ وقتی $n \to \infty$، داریم:

$$P_r(X_{l,n} \le E(X) \le X_{u,n}) \to P_r(\chi^۲_۱ \le -۲\log c)$$

برهان: بدون از دست دادن کلیت می توان فرض کرد $E(X) = \circ$. چون $F_\circ$ ناتباهیده است، وقتی $n \to \infty$ با احتمال یک داریم:

$$X_{(۱)} < \circ < X_{(n)}$$

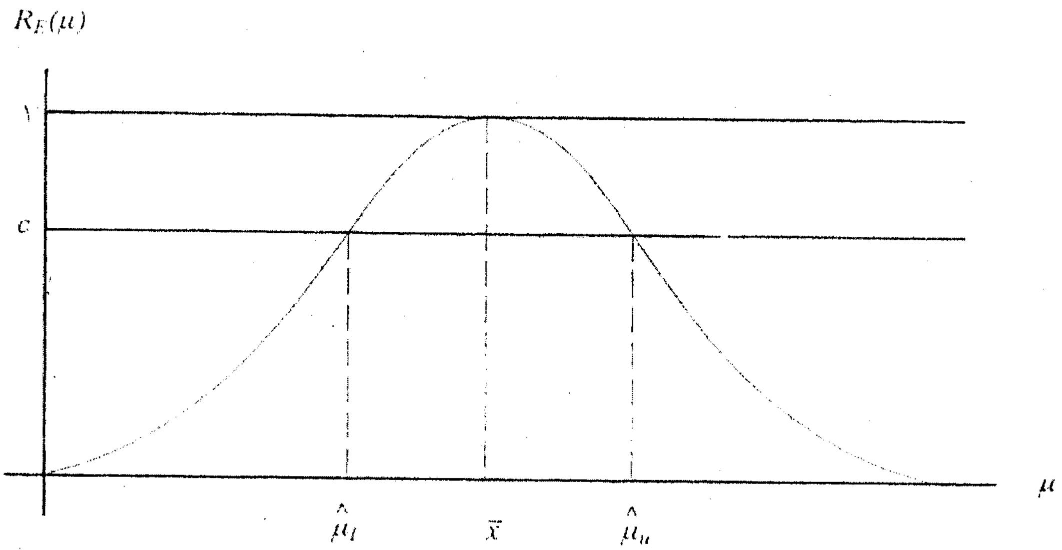
بنابراین از رابطهٔ (۱۲)، $R_E(\circ)$ وجود دارد و

$$R_E(\circ) = \operatorname{Sup}\left\{\prod_{i=۱}^{n} nw_i \mid w_i > \circ, \sum_{i=۱}^{n} w_i = ۱, \sum_{i=۱}^{n} w_i x_i = \circ\right\}$$

و مشابه استدلالی که در قبل داشتیم:

$X_{l,n} \le \circ \le X_{u,n}$ اگر و فقط اگر $R_E(\circ) \ge c$.

بنابراین کافی است نشان دهیم: $-۲\log R_E(\circ) \xrightarrow{D} \chi^۲_۱$



بازهٔ اطمینان برای μ

برای ادامهٔ اثبات می توانید به [۱] مراجعه کنید. با استفاده از تعریف $R_E(\circ)$ با به کار بردن روش لاگرانژ داریم:

$$G=\sum_{i=1}^{n}\log nw_i+\gamma(1-\sum_{i=1}^{n}w_i)+n\lambda(0-\sum_{i=1}^{n}w_ix_i)$$

$$\frac{\partial G}{\partial w_i}=0 \quad\Rightarrow\quad \frac{1}{w_i}-\gamma-n\lambda x_i=0 \quad (13)$$

$$\Rightarrow w_i=\frac{1}{\lambda+n\lambda x_i} \quad (14)$$

اگر دو طرف تساوی (۱۳) را در w_i ضرب کنیم و روی i جمع ببندیم، داریم:

$$\sum w_i\frac{\partial G}{\partial w_i}=0 \Rightarrow n-\gamma=0 \Rightarrow \gamma=n \quad (15)$$

با جایگذاری (۱۵) در (۱٤) نتیجه می شود:

$$w_i=\frac{1}{n(1+\lambda x_i)} \quad (16)$$

$$\log R_E(0)=\sum_{i=1}^{n}\log nw_i=-\sum_{i=1}^{n}\log(1+\lambda_0 x_i) \quad (17)$$

که در آن λ_0 ریشهٔ معادلهٔ زیر است:

$$0=n^{-1}\sum_{i=1}^{n}\frac{x_i}{1+\lambda x_i}\equiv g(\lambda).$$

مشابه روش قبلی برای انتخاب مناسب λ_0 ، دو لم زیر به راحتی قابل اثبات اند.

لم ۱ـ مقادیر قابل قبول λ در معادلهٔ $g(\lambda)=0$ عبارت اند از:

$$J_n=(-X_{(n)}^{-1},X_{(1)}^{-1})$$

لم ۲ـ تابع $g(\lambda)$ روی مقادیر J_n فقط یک ریشه دارد.

پس به طور خلاصه الگوریتم زیر دستورالعمل این روش را بهتر نشان می دهد:

۱) با انتخاب سطح آزمون α، مقدار c ($0<c<1$) را از رابطهٔ زیر به دست آورید:

$$1-\alpha=P(\chi^2_{(1)}\le -2\log c)$$

۲) با فرض $E(X)=t$ مقادیر $y_i^t=x_i-t$ ، $i=1,\dots,n$ ، را از روی x_i به دست آورید. (به عنوان اولین نقطهٔ شروع از $t=\bar{x}$ استفاده کنید).

۳) با استفاده از مقادیر y_i^t و به وسیلهٔ یک روش عددی (مانند روش نیوتن ـ رافسون) تنها ریشهٔ معادلهٔ

$$g(\lambda)=\frac{1}{n}\sum_{i=1}^{n}\frac{y_i^t}{1+\lambda y_i^t}=0$$

را روی بازهٔ $J_n=(-Y_{(n)}^{-1},-Y_{(1)}^{-1})$ پیدا کنید.

٤) فرض کنید λ_0 ریشهٔ معادلهٔ بالا روی J_n باشد. مقادیر w_i ، $i=1,\dots,n$ ، و R را از روابط زیر به دست آورید:

$$w_i(t)=\frac{1}{n(1+\lambda_0 y_i^t)} \qquad i=1,2,\dots,n$$

$$R_E^X(t)=\prod_{i=1}^{n}nw_i(t)$$

۵) با تغییر دادن مقادیر t ، دو نقطهٔ t_1 و t_2 ، ($t_1<t_2$) را پیدا کنید که برای این دو نقطه دقیقاً داشته باشیم:

$$R_E^X(t_1)=R_E^X(t_2)=c$$

(طبیعتاً این عمل مستلزم چندین بار اجرای این الگوریتم از گام (۲) تا گام (۵) است).

٦) بازهٔ اطمینان درستنمایی تجربی در سطح $1-\alpha$ برای میانگین عبارت است از:

$$(X_{l,n},X_{u,n})=(t_1,t_2)$$

توجه کنید جوابهایی که از این الگوریتم و الگوریتم قبلی به دست می آیند، دقیقاً برابرند. (شایان ذکر است که در شبیه سازی به دلیل عناصر تصادفی، جوابهای حاصل ممکن است اندکی تفاوت داشته باشند).در حقیقت مزیت استفاده از این الگوریتم در مواقعی

که به دنبال بازهٔ اطمینان برای پارامترهایی مانند میانه و چندکها هستیم، بهتر مشاهده می شود.[۷]

٤ شبیه سازی نتایج

در اینجا برای بررسی میزان کارایی روش درستنمایی تجربی در مقایسه با سایر روشها، از یک آزمایش شبیه سازی استفاده کرده ایم. نتایج به دست آمده بر اساس ۱۰۰۰ نمونهٔ ۲۰ تایی از توزیع $\chi^2_{(1)}$ ، در جدول شمارهٔ ۱ جمع آوری شده اند. در این آزمایش برای هر نمونهٔ ۲۰ تایی، یک بازهٔ اطمینان ۹۰٪ برای میانگین توسط روشهای مختلف بوت استرپ ، درستنمایی تجربی و t ی استیودنت محاسبه شده است. (در روشهای بوت استرپ برای هر نمونهٔ ۲۰ تایی، ۱۰۰۰ نمونهٔ بوت استرپ استخراج شده و این عمل ۱۰۰۰ بار تکرار شده است). در این آزمایش یک روش برتر تحت عنوان «بوت استرپ تعدیل شده با درستنمایی تجربی» که تلفیقی از دو روش بوت استرپ و درستنمایی تجربی است نیز آورده شده است. در این روش به جای استفاده از توزیع مجانبی $-2\log R_E(t)$ ، با فرض $F_\circ = F_n$ از توزیع بوت استرپ آن استفاده شده است، یعنی پس از استخراج B نمونهٔ بوت استرپ در هر یک از این نمونه ها، مقادیر $-2\log R_E(\circ)$ از رابطهٔ (۱۷) محاسبه شده و به جای استفاده از توزیع χ^2 (برای مشخص کردن c)، مستقیماً از توزیع بوت استرپ آن استفاده کرده ایم.

برای داشتن یک معیار نسبتاً دقیق از روش پارامتری (که در عمل به دلیل مجهول بودن توزیع جامعه مقدور نیست) نیز برای محاسبهٔ این بازه ها استفاده شده است.

جدول (۱) درصد پوشش مشاهده در ۱۰۰۰ بازهٔ اطمینان به دست آمده در هر روش را نشان می دهد. یک معیار برای مقایسهٔ دقت روشها، متوسط طول بازه در هر روش است. مقادیر $N_{L>1}$ و $N_{U<1}$ به ترتیب تعداد بازه های اطمینانی هستند که حد پایین آنها از یک بیشتر (مقدار میانگین واقعی) و حد بالای آنها از یک کمتر است. چون بازه های اطمینان در سطح ۹۰٪ ساخته شده اند، به طور طبیعی انتظار می رود مقادیر $N_{L>1}$ و $N_{U<1}$ برابر ۵٪ تعداد این بازه ها باشند. بنابراین منطقی است که خطای یکطرفهٔ (اریبی یکطرفهٔ) هر روش را به وسیلهٔ رابطهٔ زیر به دست آوریم:

$$|N_{L>1} - 50| + |N_{U<1} - 50|$$

با بررسی نتایج، دیده می شود که دو روش مبتنی بر درستنمایی تجربی از اغلب روشهای ناپارامتری دیگر بهتر عمل می کنند. در بین روشهای ناپارامتری، روش بوت استرپ تعدیل شده با درستنمایی تجربی، بهترین روش ممکن در این جدول است، اگرچه این روش برای بازه های اطمینان یکطرفه مناسب نیست. (این اریبی یکطرفه نوعاً از خواص اجتناب ناپذیر روشهای مبتنی بر درستنمایی است).

در بررسیهای دیگر مشاهده می شود که در اندازهٔ نمونهٔ ۱۰ یا بیشتر از آن، به خصوص در مورد توزیعهای چوله، بازه های درستنمایی تجربی از بازه های t ی استیودنت دقیقتر عمل می کنند، اگرچه برای اندازه های کوچک ($n < 10$) این بازه ها بهتر و یا لااقل به خوبی بازه های درستنمایی تجربی هستند. بنابراین در این حالت به علت سهولت محاسبه، بازه های t ی استیودنت توصیه می شود.

نتایج شبیه سازی بر روی توزیعهای مختلفی مانند یکنواخت، نرمال، لوژستیک و نمایی (که نتایج آنها در این مقاله ذکر نشده است) حاکی از این است که روش درستنمایی تجربی هم از نظر زمان و هم از نظر نتیجه، از روش بوت استرپ بهتر عمل می کند.

[۷] در بیشتر موارد همهٔ پارامترهای یک توزیع از قبیل میانگین، میانه، چارکها و ... را می توان در قالب یک M ـ برآوردگر نشان داد ([۶]،[۷]). در این صورت با فرض $Y_i = \psi(X_i, t)$ می توان الگوریتمی مشابه بالا به دست آورد.

جدول ۱

خطای (اریبی) یکطرفه	$N_{U<1}$	$N_{L>1}$	متوسط طول بازه	پوشش مشاهده شده	
۱۱٤	۱۲۱	۷	۰/۸۹۳	۰/۸۷۲	درستنمایی تجربی
۸٤	۸۹	۵	۰/۹۱	۰/۹۰٦	بوت استرپ تعدیل شده با درستنمایی تجربی
۱۲۷	۱۵۰	۲۳	۰/۹۵۱	۰/۸۲۷	صدکی بوت استرپ
۹۹	۱۳۵	۳٦	۰/۹۷٦	۰/۸۲۹	بوت استرپ با تصحیح اریبی
۵۵	۱۰۵	۵۰	۱/۰۱	۰/۸٤۵	بوت استرپ با تصحیح اریبی شتابدار
٦۹	۱۱۲	۵۷	۱/۳۷	۰/۸۳۱	بوت استرپ t
۱۳۵	۱٤۸	۱۳	۱/۰۸	۰/۸۳۹	t ی استیودنت
۷	۵٦	۵۱	۰/۸۱۰	۰/۸۹۳	پارامتری

منابع

[1] Owen, A. B., *Empirical Likelihood Ratio Confidence Intervals for a Single Functional,* Biometrika, Vol. 75, No. 2, 237-249, 1988.

[2] Owen, A. B., *Empirical Likelihood Ratio Confidence Regions,* Ann.Statist., Vol. 18, No. 1, 90-120, 1990.

[3] Qin, J. and Lawless, J., *Empirical Likelihood and Estimating General Equations,* Ann. Statist., Vol. 22, 300-325, 1994.

[4] Schenker, N., *Qualms about Bootstrap Confidence Interval,* J.Am.Statist.Assoc. 80, 360-361, 1985.

[5] Efron, B. and Tibshirani, R.J., *An Introduction to the Bootstrap,* New York, London, (C) 1993, Chapman & Hall, ISBN, 0-412-04231-2.

[6] Serfling, R.J., *Approximation Theorem of Mathematical Statistic,* (C) 1980, John Wiely & Sons, ISBN, 0-471-02403-1.

[7] Lehman, E.L., *Theory of Point Estimation,* (C) 1983, By John Wiely & Sons, Inc., ISBN, 0-471-05849-1.