مجله علوم آماری، بهار و تابستان ۱۳۹۷
جلد ۱۲، شماره ۱، ص ۲۲۳ – ۲۳۷
DOI: 10.29252/jss.12.1.223

آزمون-برآوردیابی انقباضی در توزیع رایلی و کاربرد آن در داده‌های سانسور شده نوع دوم

مهران نقی‌زاده قمی[1]، زهره مهدی‌زاده[1] و حمیدرضا زارعی‌فرد[2]

[1] گروه آمار، دانشگاه مازندران
[2] گروه آمار، دانشگاه جهرم

تاریخ دریافت: ۱۳۹۶/۲/۱۸ تاریخ آخرین بازنگری: ۱۳۹۷/۱/۳۰

چکیده: فرض کنید یک نمونه تصادفی از توزیع رایلی تک‌پارامتری در اختیار باشد. در روش‌های کلاسیک آمار، براساس اطلاعات موجود در نمونه و با روش‌های معمول به برآوردیابی پارامتر نامعلوم پرداخته می‌شود. گاهی در عمل، محقق دارای اطلاعاتی درباره پارامتر نامعلوم به‌صورت یک حدس یا گمان می‌باشد. این حدس، اطلاعات غیرنمونه‌ای نامیده می‌شود. در این حالت، برآوردگرهای انقباضی خطی با ترکیب اطلاعات غیرنمونه‌ای و اطلاعات موجود در نمونه معرفی شدند که در نزدیکی مقدار حدسی و واقعی دارای مخاطره کمتری نسبت به برآوردگرهای معمول هستند. در این مقاله، براساس رد یا پذیرش فرضیه صفر نزدیکی مقدار حدسی و مقدار واقعی پارامتر، چند آزمون-برآوردگر انقباضی برای پارامتر مورد بررسی با روش‌های مختلف، معرفی و مخاطره آن‌ها تحت تابع زیان آنتروپی محاسبه می‌شود. سپس رفتار آزمون-برآوردگرهای انقباضی و بهترین برآوردگر خطی براساس کارایی نسبی بین آن‌ها مقایسه می‌شوند. آن‌گاه نتایج به‌دست آمده برای نمونه‌های سانسور شده نوع دوم به‌کار گرفته می‌شود.

واژه‌های کلیدی: آزمون-برآوردگر انقباضی، توزیع رایلی، داده‌های سانسورشده.

آدرس الکترونیکی نویسنده مسئول مقاله: مهران نقی‌زاده قمی، m.naghizadeh@umz.ac.ir
کد موضوع‌بندی ریاضی (۲۰۱۰): 62F10

۱ مقدمه

یک نمونه تصادفی از توزیع رایلی با پارامتر σ با تابع چگالی احتمال

$$f(x;\sigma)=\frac{x}{\sigma}\exp(-\frac{x^{۲}}{۲\sigma}),\quad x>\circ,\ \ \sigma>\circ, \tag{۱}$$

و تابع توزیع

$$F(x;\sigma)=۱-\exp(-\frac{x^{۲}}{۲\sigma}),\quad x>\circ,\ \ \sigma>\circ, \tag{۲}$$

را در نظر بگیرید. تابع نرخ خطر توزیع رایلی به‌صورت

$$r(t)=\frac{f(t;\sigma)}{۱-F(t;\sigma)}=\frac{t}{\sigma},$$

تابعی صعودی نسبت به t است. استفاده از توزیع رایلی از این جهت دارای اهمیت است که دارای نرخ خطر صعودی می‌باشد در حالی‌که توزیع نمایی دارای نرخ خطر ثابت است. گاهی ممکن است محقق دارای اطلاعاتی درباره پارامتر σ به‌صورت یک حدس $\sigma_\circ$ باشد. چنین اطلاعی، به عنوان اطلاعات غیرنمونه‌ای[۱] یا اطلاعات پیشین نامعلوم[۲] نامیده می‌شوند. اولین بار تامسون (۱۹۶۸) مساله منقبض‌کردن یک برآوردگر نااریب به سمت پارامتر واقعی و نزدیک به مقدار حدسی را معرفی و برآوردگر انقباضی خطی[۳] را مطرح کرد. در صورتی‌که مقدار حدسی به مقدار واقعی آن نزدیک باشد، برآوردگرهای انقباضی خطی دارای مخاطره کمتری نسبت به برآوردگرهای معمول مانند برآورگر ماکسیمم درستنمایی هستند. برای تشخیص اینکه مقدار حدسی $\sigma_\circ$ به σ نزدیک است یا خیر، از یک پیش‌آزمون[۴] برای آزمون فرضیه‌های

$$\begin{cases} H_\circ:\ \sigma=\sigma_\circ \\ H_۱:\ \sigma\neq\sigma_\circ, \end{cases} \tag{۳}$$

[1] Nonsample information
[2] Uncertain prior information
[3] Linear shrinkage estimator
[4] Pretest

استفاده می‌شود. براساس رد یا پذیرش فرضیه صفر H_0 در سطح معنی‌داری از پیش تعیین شده α، می‌توان آزمون-برآوردگرهای انقباضی[5] را به‌دست آورد و رفتار آن‌ها را مورد بررسی قرار داد.

مساله برآوردیابی انقباضی خطی در توزیع‌های مختلف و تحت توابع زیان متفاوت، در سال‌های اخیر مورد توجه زیادی قرار گرفته است، که از آن جمله می‌توان به پراکاش و سینگ (۲۰۰۶)، سینگ و همکاران (۲۰۰۷)، پراکاش و سینگ (۲۰۰۸) و پراکاش (۲۰۱۰) اشاره کرد که به ترتیب به برآوردیابی انقباضی خطی در توزیع‌های گاوسی وارونه، پارتو، نمایی، و رایلی معکوس تحت تابع زیان خطی-نمایی[6] پرداختند. از کارهای جدید در این زمینه می‌توان به نقی‌زاده قمی و برموده (۲۰۱۵) و کیاپور و نقی‌زاده قمی (۲۰۱۶) اشاره کرد.

در این مقاله، آزمون-برآوردگرهای انقباضی برای پارامتر توزیع رایلی تحت تابع زیان آنتروپی به صورت

$$L(\sigma,\tilde{\sigma})=\Delta-\ln(\Delta)-1,\quad \Delta=\frac{\tilde{\sigma}}{\sigma} \tag{۴}$$

مورد بررسی قرار می‌گیرد، که در آن $\tilde{\sigma}$ یک برآوردگر دلخواه برای σ است. تابع زیان آنتروپی یک تابع اکیداً محدب و نامتقارن نسبت به Δ و دارای مینیمم یکتا در نقطه $\Delta=1$ است. این تابع زیان در موقعیت‌هایی که کم‌برآوردی دارای اهمیت بیشتری نسبت به بیش‌برآوردی باشد مورد استفاده قرار می گیرد و چون یک تابع زیان ناوردای مقیاس[7] است، به خوبی می‌تواند برای برآورد پارامتر مقیاس توزیع رایلی مورد استفاده قرار گیرد. پارسیان و نعمت‌الهی (۱۹۹۶)، جعفری و پارسیان (۲۰۰۸)، نعمت‌الهی و معتمد-شریعتی (۲۰۰۹ و ۲۰۱۲) و نعمت‌الهی و مغناطیسی (۲۰۱۱) از این تابع زیان در زمینه‌های مختلف برآوردیابی و پیشگویی استفاده کردند.

در بخش ۲ بهترین برآوردگر خطی براساس برآوردگر ماکسیمم درستنمایی به صورت $c\hat{\sigma}$ تحت تابع زیان آنتروپی به‌دست آمده و مخاطره آن محاسبه می‌شود. در بخش ۳ فرم کلی آزمون-برآوردگر انقباضی، معرفی و مخاطره آن تحت تابع زیان آنتروپی محاسبه می‌شود. سپس رفتار آن نسبت به بهترین برآوردگر خطی برای مقادیر مختلف ضریب انقباضی و حجم نمونه با محاسبه کارایی نسبی بین آن‌ها بررسی می‌شود. در بخش ۴ چند آزمون-برآوردگر انقباضی با استفاده از روش‌های مختلف تعیین ضریب انقباضی، معرفی، مخاطره آن‌ها محاسبه و رفتار آن‌ها نسبت به بهترین برآوردگر خطی مورد مقایسه قرار می‌گیرد. در پایان،

[5]Shrinkage test-estimators
[6]Linear-Exponential loss (LINEX)
[7]Scale invariant

نتایج برای تحلیل داده‌های سانسورشده نوع دوم به‌کار می‌روند.

۲ بهترین برآوردگر خطی به فرم $c\hat{\sigma}$

فرض کنید $X_1, ..., X_n$ نمونه‌ای تصادفی از توزیع رایلی با پارامتر σ با تابع چگالی احتمال (۱) باشد. در این‌صورت برآوردگر ماکسیمم درستنمایی σ برابر $\hat{\sigma} = (2n)^{-1} \sum_{i=1}^{n} X_i^2$ است. به راحتی می‌توان نشان داد $U = 2n\hat{\sigma}/\sigma \sim \chi^2_{2n}$. مخاطره برآوردگر $c\hat{\sigma}$ تحت تابع زیان آنتروپی عبارت است از

$$\begin{aligned} R(\sigma, c\hat{\sigma}) &= E[\frac{c\hat{\sigma}}{\sigma}] - E[\ln(\frac{c\hat{\sigma}}{\sigma})] - 1 \\ &= \frac{c}{2n}E[U] - E[\ln(U)] - \ln c + \ln(2n) - 1. \end{aligned}$$

می‌دانیم $E[U] = 2n$. برای محاسبه $E[\ln(U)]$، تابع گاما را به‌صورت

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt. \tag{۵}$$

در نظر بگیرید. با مشتق‌گیری از تابع (۵) نسبت به α، تقسیم طرفین بر $\Gamma(\alpha)$ و تغییر متغیر $y = 2t$ داریم

$$\begin{aligned} \Psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} &= \int_0^\infty [\ln(y) - \ln 2] \frac{y^{\alpha-1} e^{-\frac{y}{2}}}{\Gamma(\alpha) 2^\alpha} dy \\ &= E[\ln Y] - \ln 2, \end{aligned}$$

که در آن $Y \sim Gamma(\alpha, 2)$. تابع $\Psi(\cdot)$، تابع دای گاما[8] نامیده می‌شود که در نرم‌افزار R نسخه ۳.۱.۲ با استفاده از تابع digamma(۰) قابل محاسبه است.

به‌ازای $\alpha = n$ داریم $E[\ln(U)] = \ln 2 + \Psi(n)$. در نتیجه تابع مخاطره $c\hat{\sigma}$ به‌صورت

$$R(\sigma, c\hat{\sigma}) = c - \ln c + \ln n - \Psi(n) - 1. \tag{۶}$$

---
[8] Digamma function

است که تابعی محدب نسبت به c و دارای نقطه مینیمم در $c = ۱$ می‌باشد. بنابراین برآوردگر $\hat{\sigma}$، برآوردگر دارای کمترین مخاطره در کلاس برآوردگرهای به فرم $c\hat{\sigma}$ تحت تابع زیان آنتروپی می‌باشد. همچنین مخاطره برآوردگر $\hat{\sigma}$ تحت تابع زیان آنتروپی به‌صورت

$$R(\sigma, \hat{\sigma}) = \ln n - \Psi(n). \qquad (۷)$$

محاسبه می‌شود. طبق تعریف لهمن (۱۹۵۱)، برآوردگر $\tilde{\sigma}$ مخاطره–نااریب[۹] نامیده می‌شود اگر

$$E[L(\sigma, \tilde{\sigma})] \leq E[L(\sigma', \tilde{\sigma})], \quad \sigma \neq \sigma'. \qquad (۸)$$

یافتن برآوردگرهای مخاطره–نااریب به تابع زیان بستگی دارد. با توجه به تابع زیان آنتروپی داریم

$$\begin{aligned} E[L(\sigma, \tilde{\sigma})] - E[L(\sigma', \tilde{\sigma})] &= E[\frac{\tilde{\sigma}}{\sigma} - \ln(\frac{\tilde{\sigma}}{\sigma}) - ۱] - E[\frac{\tilde{\sigma}}{\sigma'} - \ln(\frac{\tilde{\sigma}}{\sigma'}) - ۱] \\ &= (\frac{۱}{\sigma} - \frac{۱}{\sigma'})E[\tilde{\sigma}] + \ln(\frac{\sigma}{\sigma'}). \end{aligned}$$

با قرار دادن $E(\tilde{\sigma}) = \sigma$ داریم

$$E[L(\sigma, \tilde{\sigma})] - E[L(\sigma', \tilde{\sigma})] = \ln(\frac{\sigma}{\sigma'}) - \frac{\sigma}{\sigma'} + ۱ < ۰.$$

بنابراین برآوردگر $\tilde{\sigma}$ مخاطره–نااریب است اگر $E(\tilde{\sigma}) = \sigma$. چون $E(\hat{\sigma}) = \sigma$، پس $\hat{\sigma}$ مخاطره نااریب است، لذا $\hat{\sigma}$ بهترین برآوردگر خطی (برآوردگر مخاطره نااریب با کمترین مخاطره) در کلاس برآوردگرهای به فرم $c\hat{\sigma}$ است.

۳ آزمون–برآوردگرهای انقباضی

با فرض در اختیار داشتن مقدار حدسی $\sigma_۰$ برای پارامتر σ و با پیروی از تامسون (۱۹۶۸) برآوردگر انقباضی خطی به صورت

$$\hat{\sigma}_S = k\hat{\sigma} + (۱ - k)\sigma_۰, \quad k \in [۰, ۱],$$

[9] Risk-unbiased

را برای σ درنظر بگیرید، که در آن مقدار k ضریب انقباضی نامیده می‌شود که توسط محقق با توجه به عقیده‌اش نسبت به مقدار حدسی σ_0 تعیین می‌شود. ایده استفاده از برآوردگرهای انقباضی خطی، بهبود بخشیدن برآوردگرهای نااریب با استفاده از برآوردگرهای اریب است که دارای مخاطره کمتری باشند. برآوردگر $\hat{\sigma}_S$ دارای اریبی به‌صورت

$$B(\sigma, \hat{\sigma}_S) = E(\hat{\sigma}_S) - \sigma = k\sigma + (1-k)\sigma_0 - \sigma = (k-1)(\sigma - \sigma_0),$$

است که به ازای $k = 1$ یا $\sigma = \sigma_0$ برابر صفر می‌شود. برای بررسی میزان نزدیکی مقدار حدسی σ_0 به σ می‌توان فرضیه‌های داده شده در (۳) را در سطح معنی‌داری α مورد آزمون قرار داد. برای انجام آزمون می‌توان از آماره آزمون $U = 2n\hat{\sigma}/\sigma$ استفاده کرد که دارای توزیع کی‌دو با $2n$ درجه آزادی است. هدف از این مقاله، ساختن برآوردگرهایی براساس رد یا پذیرش فرضیه صفر $H_0 : \sigma = \sigma_0$ در سطح معنی‌داری از پیش تعیین شده α است. این برآوردگرها آزمون-برآوردگرهای انقباضی نامیده شده و با $\hat{\sigma}_{ST}$ نشان داده می‌شوند. نحوه ساختن $\hat{\sigma}_{ST}$ به این صورت است که اگر فرضیه صفر $H_0 : \sigma = \sigma_0$ پذیرفته شود آن‌گاه $\hat{\sigma}_{ST} = k\hat{\sigma} + (1-k)\sigma_0$ و اگر $H_0 : \sigma = \sigma_0$ پذیرفته نشود آن‌گاه $\hat{\sigma}_{ST} = \hat{\sigma}$. یعنی اگر مقدار حدسی σ_0 به σ نزدیک باشد از برآوردگر انقباضی خطی $\hat{\sigma}_S$ و در غیر این‌صورت از بهترین برآوردگر مخاطره-نااریب خطی یعنی $\hat{\sigma}$ استفاده می‌شود. برای نمایش بهتری از $\hat{\sigma}_{ST}$، فرض کنید H_0 در سطح معنی‌داری α پذیرفته شود. در این‌صورت

$$Pr\left(q_1 \leq \frac{2n\hat{\sigma}}{\sigma_0} \leq q_2\right) = 1 - \alpha, \tag{۹}$$

که در آن $q_1 = \chi^2_{2n,\frac{\alpha}{2}}$ و $q_2 = \chi^2_{2n,1-\frac{\alpha}{2}}$ چندک‌های چپ توزیع کی‌دو با $2n$ درجه آزادی هستند، یعنی $Pr(\chi^2_{2n} < \chi^2_{2n,\nu}) = \nu$. در نتیجه آزمون-برآوردگر انقباضی $\hat{\sigma}_{ST}$ را می‌توان به‌صورت

$$\hat{\sigma}_{ST} = \begin{cases} k\hat{\sigma} + (1-k)\sigma_0 & a_1 \leq \hat{\sigma} \leq a_2 \\ \hat{\sigma} & \hat{\sigma} < a_1 \; or \; \hat{\sigma} > a_2, \end{cases} \tag{۱۰}$$

نوشت، که در آن $a_i = \frac{q_i\sigma_0}{2n}$، $i = 1, 2$. اگر قرار داده شود $A = \{a_1 \leq \hat{\sigma} \leq a_2\}$ و $B = \{u_1 \leq U \leq u_2\}$، که در آن $u_i = q_i\delta$، $i = 1, 2$ و $\delta = \frac{\sigma_0}{\sigma}$، آن‌گاه مخاطره آزمون-برآوردگر انقباضی $\hat{\sigma}_{ST}$

تحت تابع زیان آنتروپی برابر است با

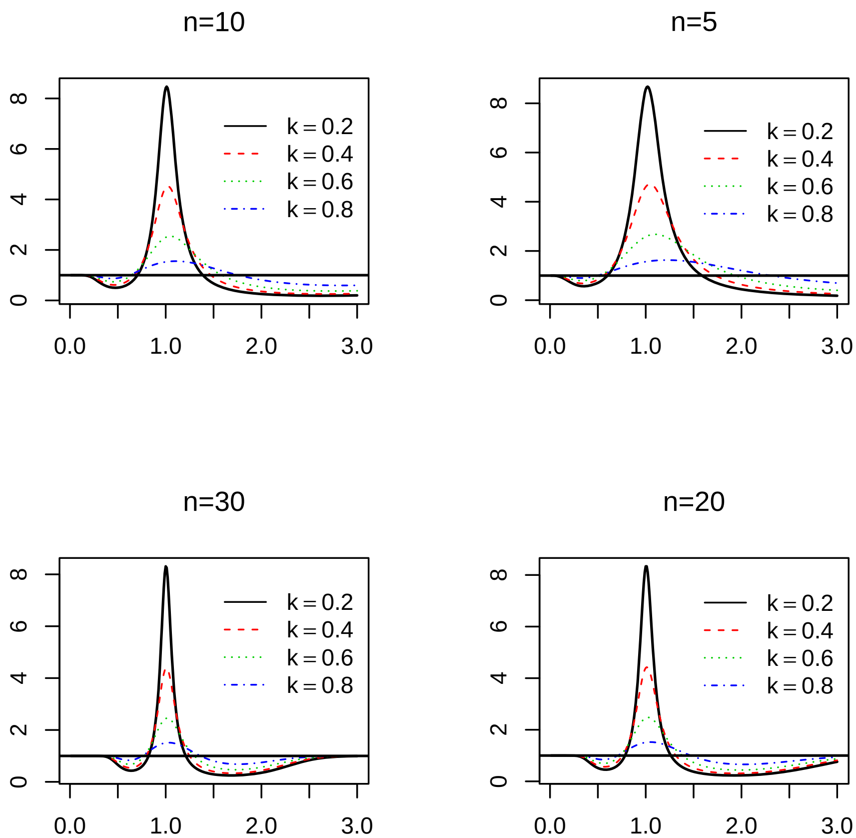
$$\begin{aligned}
R(\sigma,\hat{\sigma}_{ST}) &= E[\frac{\hat{\sigma}_{ST}}{\sigma}]-E[\ln(\frac{\hat{\sigma}_{ST}}{\sigma})]-1\\
&= E[\{\frac{k\hat{\sigma}+(1-k)\sigma_0}{\sigma}-\ln(\frac{k\hat{\sigma}+(1-k)\sigma_0}{\sigma})-1\}I(A)]\\
&\quad +E[\{\frac{\hat{\sigma}}{\sigma}-\ln(\frac{\hat{\sigma}}{\sigma})-1\}I(A^c)]\\
&= E[\{\frac{kU}{2n}+(1-k)\delta-\ln(\frac{kU}{2n}+(1-k)\delta)-1\}I(B)]\\
&\quad -E[\{\frac{U}{2n}-\ln(\frac{U}{2n})-1\}I(B)]\\
&\quad +E[\frac{U}{2n}-\ln(\frac{U}{2n})-1]\\
&= \int_{u_1}^{u_2}\{(1-k)\delta+\frac{ku}{2n}-\ln((1-k)\delta+\frac{ku}{2n})\}h(u)du\\
&- \int_{u_1}^{u_2}[\frac{u}{2n}-\ln(\frac{u}{2n})]h(u)du+\ln n-\Psi(n),
\end{aligned} \qquad (۱۱)$$

که در آن $h(u)$ تابع چگالی متغیر تصادفی $U=2n\hat{\sigma}/\sigma\sim\chi^2_{2n}$ است. مخاطره (۱۱) با روش‌های عددی در نرم‌افزار R برای مقادیر انتخابی n، k، α و δ محاسبه می‌شود.

برای مقایسه آزمون-برآوردگر انقباضی $\hat{\sigma}_{ST}$ و بهترین برآوردگر مخاطره-نااریب خطی $\hat{\sigma}$، کارایی نسبی آزمون-برآوردگر انقباضی $\hat{\sigma}_{TS}$ نسبت به $\hat{\sigma}$ به صورت

$$RE(\hat{\sigma}_{ST},\hat{\sigma})=\frac{R(\sigma,\hat{\sigma})}{R(\sigma,\hat{\sigma}_{ST})}. \qquad (۱۲)$$

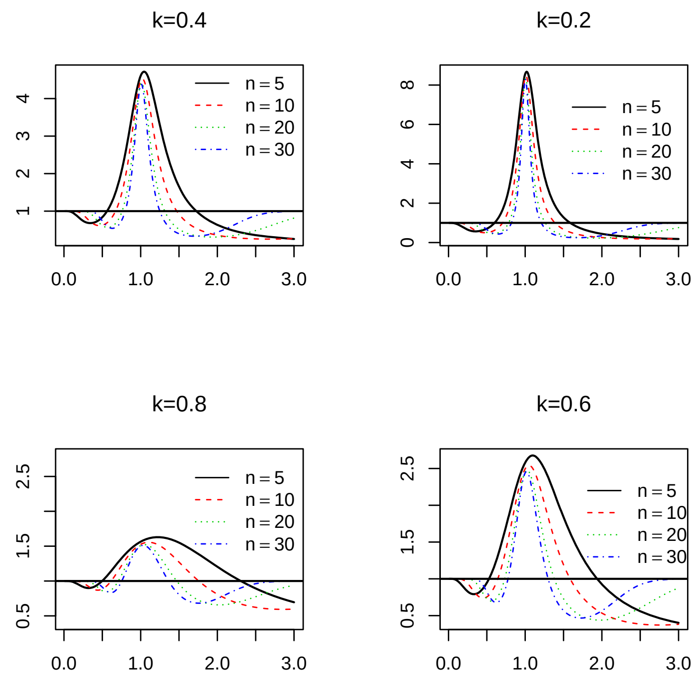
محاسبه می‌شود. شکل‌های ۱ و ۲ کارایی نسبی (۱۲) را نسبت به δ به ازای مقادیر مختلف $n=5,10,20,30$، $k=0/2(0/2)0/8$ و $\alpha=0/01$ نشان می‌دهد. همان‌طور که در شکل ۱ ملاحظه می‌شود آزمون-برآوردگرهای انقباضی برای مقادیر δ نزدیک به یک دارای مخاطره کمتر نسبت به برآوردگر $\hat{\sigma}$ هستند و برای n و α ثابت، با افزایش ضریب انقباضی k، از کارایی نسبی آزمون-برآوردگرهای انقباضی کاسته می‌شود. در شکل ۲ ملاحظه می‌شود که به ازای k و α ثابت و برای مقادیر δ نزدیک به یک، با افزایش حجم نمونه n، کارایی نسبی آزمون-برآوردگرهای انقباضی کاهش می‌یابد. بنابراین برای مقادیر δ نزدیک به یک (مقادیر σ_0 نزدیک به σ) و برای ترکیبی از حجم نمونه و ضریب انقباضی کوچک‌تر، آزمون-برآوردگرهای انقباضی با کارایی بیشتر خواهیم داشت.



شکل ۱. نمودارهای کارایی نسبی به ازای مقادیر مختلف n، k و $\alpha = ۰/۰۱$

۴ ضرایب انقباضی و آزمون-برآوردگرهای انقباضی

در این بخش، ۳ ضریب انقباضی مختلف معرفی می‌شوند که با روش‌های مختلف به دست می‌آیند و آزمون-برآوردگرهای انقباضی متناظر را می‌سازند. سپس مخاطره آن‌ها تحت تابع زیان آنتروپی محاسبه و رفتار آن‌ها نسبت به برآوردگر $\hat{\sigma}$ با محاسبه کارایی نسبی بین آن‌ها ارزیابی می‌شود.



شکل ۲. نمودارهای کارایی نسبی به ازای مقادیر مختلف n، k و $\alpha = ۰/۰۱$

مخاطره برآوردگر انقباضی $\hat{\sigma}_S$ تحت تابع آنتروپی به‌صورت

$$\begin{aligned} R(\sigma, \hat{\sigma}_S) &= E[\frac{\hat{\sigma}_S}{\sigma}] - E[\ln(\frac{\hat{\sigma}_S}{\sigma})] - ۱ \\ &= E[\frac{k\hat{\sigma} + (۱-k)\sigma_{۰}}{\sigma}] - E[\ln(\frac{k\hat{\sigma} + (۱-k)\sigma_{۰}}{\sigma})] - ۱ \\ &= k + (۱-k)\delta - \int_{۰}^{\infty} \ln[(۱-k)\delta + \frac{ku}{۲n}]h(u)du - ۱ \end{aligned}$$

است که با روش‌های عددی نسبت به k مینیمم می‌شود. مقدار $k = k_{۱}$ که از مینیمم‌کردن تابع مخاطره به‌دست می‌آید، به عنوان اولین ضریب انقباضی در نظر گرفته می‌شود. آزمون–برآوردگر انقباضی متناظر با $k_{۱}$ را با $\hat{\sigma}_{ST۱}$ نشان می‌دهیم. مخاطره آزمون–برآوردگر انقباضی $\hat{\sigma}_{ST۱}$ با قرار دادن $k = k_{۱}$ در رابطه (۱۱) به دست می‌آید.

فرضیه صفر $H_0 : \sigma = \sigma_0$ در سطح α پذیرفته می‌شود اگر

$$q_1 \leq \frac{2n\hat{\sigma}}{\sigma_0} \leq q_2 \Leftrightarrow 0 \leq \frac{\frac{U}{\delta} - q_1}{q_2 - q_1} \leq 1.$$

بنابراین انتخاب دیگری برای ضریب انقباضی به صورت

$$k_2 = \frac{\frac{U}{\delta} - q_1}{q_2 - q_1},$$

خواهد بود. آزمون-برآوردگر انقباضی متناظر با ضریب انقباضی k_2 را با $\hat{\sigma}_{ST2}$ نشان می‌دهیم که مخاطره آن با قرار دادن $k = k_2$ در رابطه (۱۱) به دست می‌آید.

اگر فرضیه صفر $H_0 : \sigma = \sigma_0$ پذیرفته شود آن‌گاه با پیروی از پراکاش و سینگ (۲۰۰۸) داریم

$$q_1 \leq \frac{2n\hat{\sigma}}{\sigma_0} \leq q_2 \Rightarrow q_1 \leq E[\frac{2n\hat{\sigma}}{\sigma_0}] \leq q_2 \Rightarrow q_1 \leq 2n \leq q_2.$$

بنابراین $\frac{q_1}{2n} \leq 1$. اگر بخواهیم ضریب انقباضی کوچکتری در نظر بگیریم، فرض می‌کنیم $\frac{q_1}{2n} \simeq 1$. بنابراین داریم

$$2n\frac{\frac{U}{2n\delta} - \frac{q_1}{2n}}{q_2 - q_1} \simeq \frac{2n}{q_2 - q_1}(\frac{U}{2n\delta} - 1)$$

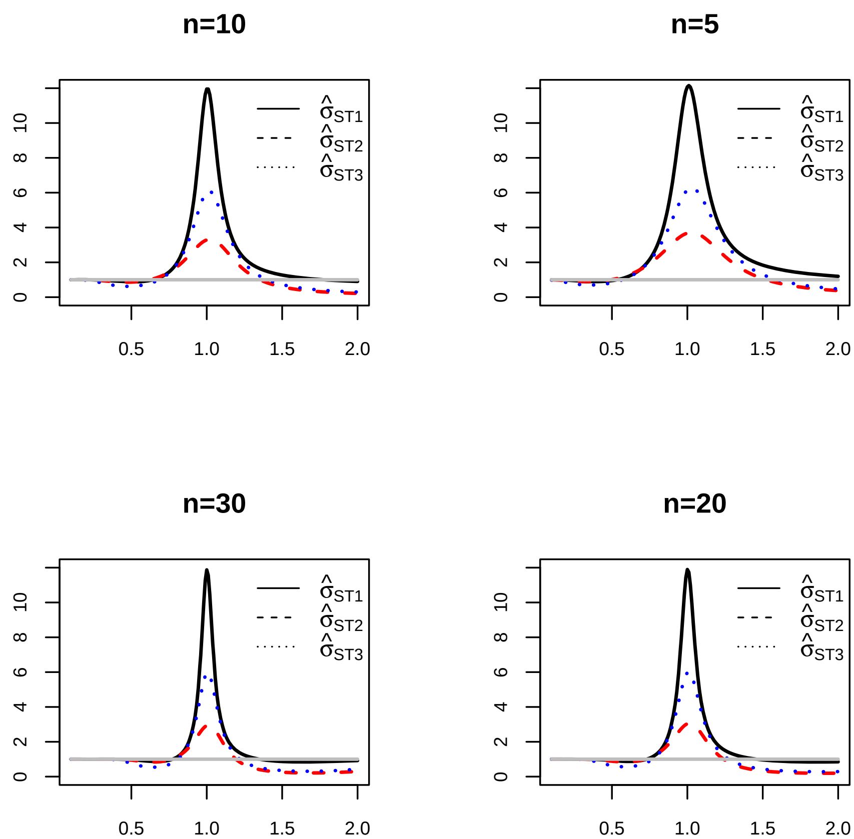
که ممکن است منفی باشد. بنابراین ضریب انقباضی k_3 به‌صورت

$$k_3 = \frac{2n}{q_2 - q_1}|\frac{U}{2n\delta} - 1|.$$

در نظر گرفته می‌شود. آزمون-برآوردگر انقباضی متناظر با ضریب انقباضی k_3 یعنی $\hat{\sigma}_{ST3}$ دارای مخاطره (۱۱) با قراردادن $k = k_3$ خواهد بود.

برای مقایسه عملکرد آزمون-برآوردگرهای انقباضی $\hat{\sigma}_{STi}$، $i = 1, 2, 3$ با بهترین برآوردگر مخاطره-نااریب خطی، $\hat{\sigma}$، کارایی نسبی $\hat{\sigma}_{STi}$ نسبت به $\hat{\sigma}$ همانند (۱۲) محاسبه شده و در شکل ۳ به ازای $n = 5, 10, 20, 30$ و $\alpha = 0/01$ نمایش داده شده است. همان‌طور که ملاحظه می‌شود کارایی نسبی

برای هر سه آزمون-برآوردگر در $\delta = ۱$ به مقدار بیشینه خود می‌رسد. برای مقادیر δ نزدیک به یک ($\sigma_۰$ نزدیک به σ) هر سه آزمون-برآوردگر دارای کارایی نسبی بیشتری نسبت به $\hat{\sigma}$ هستند، هر چند با افزایش n، از دامنه مقادیر δ نزدیک به یک که منجر به کارایی بیشتر می‌شود کاسته می‌شود. همچنین برآوردگر $\hat{\sigma}_{ST۱}$ که با استفاده از ضریب انقباضی k_{min} حاصل از مینیمم‌کردن مخاطره برآوردگر انقباضی $\hat{\sigma}_S$ به‌دست می‌آید برای مقادیر δ نزدیک به یک دارای کارایی نسبی بیشتری در مقایسه با آزمون-برآوردگرهای دیگر است.



شکل ۳. نمودارهای کارایی نسبی به ازای مقادیر مختلف n و $\alpha = ۰/۰۱$

۵ آزمون-برآوردیابی انقباضی داده‌های سانسور شده نوع دوم از راست

سانسور نوع دوم اغلب در انجام آزمون‌های طول عمر بکار می‌رود. در مجموعه نمونه سانسورشده‌ی نوع دوم، فقط r کوچکترین مشاهدات از یک نمونه تصادفی nتایی قابل مشاهده هستند. به عبارت دیگر تمام n عضو نمونه در آزمایش قرار می‌گیرند اما به جای انجام آزمایش تا پایان ازکارافتادگی تمامی اعضای نمونه، این کار تا زمانی انجام می‌شود که r عضو از کار بیفتد. این کار موجب صرفه‌جویی در زمان و هزینه‌های آزمایش خواهد شد. باید توجه کرد که در این نوع سانسور، تعداد مشاهدات لازم یعنی r مقداری ثابت است و پیش از شروع آزمایش تعیین می‌شود، اما طول دوره‌ی آزمایش یک متغیر تصادفی است (لاولس، ۲۰۰۳).

با فرض مشاهده نمونه سانسور شده $x_{1:n}, ..., x_{r:n}$ از توزیع رایلی، تابع درستنمایی برابر است با

$$L(\theta|x_{1:n}, ..., x_{r:n}) = \frac{n!\prod_{i=1}^{r} x_i}{(n-r)!\sigma^r}\exp\{-\frac{\sum_{i=1}^{r} x_{i:n}^2 + (n-r)x_{r:n}^2}{2\sigma}\}, \qquad (13)$$

در نتیجه، برآوردگر ماکسیمم درستنمایی σ برابر است با

$$T_r = \frac{\sum_{i=1}^{r} X_{i:n}^2 + (n-r)X_{r:n}^2}{2r}. \qquad (14)$$

همچنین $2rT_r/\sigma \sim \chi^2_{2r}$. با نگاهی به نتایج به دست آمده در بخش‌های ۲ تا ۴ می‌توان نتایج را با داده‌های سانسور شده نوع دوم با جایگذاری r و T_r به جای n و $\hat{\sigma}$ به کار برد.

۱.۵ مثال کاربردی

داده‌های زیر (لاولس، ۲۰۰۳، صفحه ۲۶۳)، زمان‌های شکست (بر حسب دقیقه) برای یک نمونه ۱۵تایی از مولفه‌های الکتریکی در یک آزمون طول عمر شتابنده[10] می‌باشد.

۱/۴ ۵/۱ ۶/۳ ۱۰/۸ ۱۲/۱ ۱۸/۵ ۱۹/۷ ۲۲/۲
۲۳ ۳۰/۶ ۳۷/۳ ۴۶/۳ ۵۳/۹ ۵۹/۸ ۶۶/۲.

[10] Accelerated life test

با استفاده از آزمون کلموگروف-اسمیروف با مقدار آماره آزمون $D = ۰/۲۳۴۱$ و مقدار معنی‌داری $p - value = ۰/۳۳۰۴$، می‌توان توزیع رایلی با پارامتر $\sigma = ۵۸۰/۵۹$ را برای این داده‌ها مناسب دانست. فرض کنید آزمایش تا شکست ۴ مولفه ادامه یابد، یعنی $r = ۴$ باشد. در نتیجه از رابطه (۱۴)، برآورد ماکسیمم درستنمایی σ که در واقع بهترین برآوردگر خطی نیز هست برابر است با $T_۴ = ۱۷۵۷/۰۹$. با توجه به رابطه (۷)، مخاطره این برآوردگر برابر است با ۰/۱۳۰۲. اگر محقق دارای اطلاع پیشین به‌صورت حدس $\sigma = ۱۷۰۰$ باشد، آنگاه مقدار آماره آزمون برای آزمودن فرض $H_۰ : \sigma = ۱۷۰۰$ در مقابل $H_۱ : \sigma \neq ۱۷۰۰$ برابر $\chi^۲ = ۲rT_r/\sigma_۰ = ۸/۲۷$ است. با فرض $\alpha = ۰/۰۵$ داریم $q_۱ = \chi^۲_{۸,۰/۰۲۵} = ۲/۱۸$ و $q_۲ = \chi^۲_{۸,۰/۹۷۵} = ۱۷/۵۳$. در نتیجه فرض صفر در سطح معنی‌داری ۰/۰۵ پذیرفته می‌شود. ضریب انقباضی $k_۱$ که از مینیمم‌کردن مخاطره برآوردگر انقباضی $\hat{\sigma}_S$ به‌دست می‌آید برابر ۰/۰۰۴ است. چون مقدار $k_۱$ به مقدار $\delta = \sigma_۰/\sigma$ وابسته است، از برآورد δ به‌صورت

$$\hat{\delta} = \frac{\sigma_۰}{T_۴} = \frac{۱۷۰۰}{۱۷۵۷/۰۹} = ۰/۹۶۸$$

برای برآورد $k_۱$ استفاده می‌شود. ضرایب انقباضی $k_۲$ و $k_۳$ نیز به‌صورت

$$\begin{aligned} k_۲ &= \frac{U/\delta - q_۱}{q_۲ - q_۱} = \frac{۲rT_r/\sigma_۰ - q_۱}{q_۲ - q_۱} = ۰/۳۹, \\ k_۳ &= \frac{۲r}{q_۲ - q_۱}|\frac{T_r}{\sigma_۰} - ۱| = ۰/۰۰۲. \end{aligned}$$

جدول ۱ مخاطره برآوردگرهای محاسبه‌شده و همچنین کارایی نسبی بین آزمون-برآوردگرها و برآورد $T_۴$ را نشان می‌دهد.

جدول ۱. مخاطره برآوردگرها و کارایی نسبی بین آزمون-برآوردگرهای انقباضی و $T_۴$

	$T_۴$	$\hat{\sigma}_{ST۱}$	$\hat{\sigma}_{ST۲}$	$\hat{\sigma}_{ST۳}$
مخاطره	۰/۱۳۰۲	۰/۰۳۶۶۵	۰/۰۴۹۶۶	۰/۰۳۶۶۴
کارایی نسبی	—	۳/۵۵۱۸۳	۲/۶۲۰۹۹	۳/۵۵۱۹۱

همان‌طور که ملاحظه می‌شود هر سه آزمون-برآوردگر دارای مخاطره کمتری نسبت به برآوردگر $T_۴$ هستند. همچنین آزمون-برآوردگر $\hat{\sigma}_{ST۳}$ دارای مخاطره کمتری نسبت به دیگر آزمون-برآوردگرها است.

بحث و نتیجه‌گیری

در این مقاله به آزمون-برآوردیابی انقباضی برای پارامتر مقیاس توزیع رایلی براساس داده‌های سانسور شده و مقایسه آن‌ها با بهترین برآوردگر خطی به فرم $c\hat{\sigma}$ تحت تابع زیان آنتروپی پرداخته شده است. آزمون-برآوردگرهای انقباضی براساس رد یا پذیرش فرضیه صفر $H_{\circ} : \sigma = \sigma_{\circ}$ به‌دست می‌آیند. نتایج نشان می‌دهد که آزمون-برآوردیابی انقباضی وقتی که مقدار حدسی پارامتر به مقدار واقعی نزدیک باشد دارای مخاطره کمتری نسبت به بهترین برآوردگر خطی است. یکی از مسایل مهم در ساختن چنین برآوردگرهایی، انتخاب مقدار k است که به مقدار حدسی و نمونه بستگی دارد. یکی از روش‌های محاسبه k، مینیمم‌سازی مخاطره برآوردگر انقباضی $\hat{\sigma}_S$ نسبت به k است. همچنین دو مقدار دیگر برای k معرفی شده که فقط در زمانی که فرضیه صفر پذیرفته شود مورد استفاده قرار می‌گیرند. به‌طور کلی با توجه به شکل‌های رسم‌شده، آزمون-برآوردگر $\hat{\sigma}_{ST\text{۱}}$ مبتنی بر $k = k_{min}$ دارای مخاطره کمتری نسبت به سایر آزمون-برآوردگرهای معرفی‌شده در نزدیکی مقدار حدسی و مقدار واقعی است.

تقدیر و تشکر

نویسندگان از نظرات و پیشنهادات ارزنده داوران، سردبیر و ویراستار محترم مجله که باعث بهبود کیفیت مقاله گردید تقدیر و تشکر می‌کنند.

مراجع

Jafari, M., and Parsian, A. (2008), Posterior Regret Gamma-Minimax and Prediction with Applications on k-Records Data under Entropy Loss Function, *Communications in Statistics-Theory and Method*, **37**, 2202-2212.

Kiapour, A., and Naghizadeh Qomi, M. (2016), Shrinkage Preliminary Test Estimation under a Precautionary Loss Function with Applications on Records and Censored Data, *Journal of the Iranian Statistical Society*, **15**, 73-85.

Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data.* John Wiley and Sons, New York.

Lehmann, E. L. (1951). A General Concept of Unbiasedness, *Annals of Mathematical Statistics,* **22**, 578-592.

Naghizadeh Qomi, M., and Barmoodeh, L. (2015), Shrinkage Testimation in Exponential Distribution Based on Records under Asymmetric Squared Log Error Loss, *Journal of Statistical Research of Iran,* **12**, 225-238.

Nematollahi, N., and Meghnatisi, Z. (2011), On the Admissibility of Estimators of Two Ordered Gamma Scale Parameters under Entropy Loss Function, *REVSTAT-Statistical Journal*, **9**, 227-245.

Nematollahi, N., and Motamed-Shariati, F. (2009), Estimation of the Scale Parameter of the Selected Gamma Population under the Entropy Loss Function, *Communications in Statistics-Theory and Method,* **38**, 208-221.

Nematollahi, N., and Motamed-Shariati, F. (2012), Estimation of the Parameter of the Selected Uniform Population under the Entropy Loss Function, *Journal of Statistical Planning and inference,* **142**, 2190-2202.

Parsian, A., and Nematollahi, N. (1996), Estimation of Scale Parameter under Entropy Loss Function, *Journal of Statistical Planning and Inference,* **52**, 77-91.

Prakash, G., and Singh, D. C. (2006), Shrinkage Testimators for the Inverse Dispersion of the Inverse Gaussian Distribution under the LINEX Loss Function, *Austrian Journal of Statistics*, **35**, 463-470.

Prakash, G., and Singh, D. C. (2008). Shrinkage Estimation in Exponential Type-II Censored Data under LINEX Loss, *Journal of the Korean Statistical Society*, **37**, 53-61.

Singh, D. C., Prakash, G., and Singh, P. (2007), Shrinkage Testimators for the Shape Parameter of Pareto Distribution Using the LINEX Loss Function, *Communication in Statistics-Theory and Methods*, **36**, 741-753.

Thompson, J. R. (1968). Some Shrunken Techniques for Estimating the Mean. *Journal of the American Statistical Association*, **63**, 113-122.