

یک برآوردگر ماکسیمم درست‌نمایی برای پارامتر چولگی توزیع چوله نرمال

افشین فلاح، عباس گرامی*

دانشکده ریاضی، آمار و علوم کامپیوتر، پردیس علوم، دانشگاه تهران، تهران، ایران

* مسئول مکاتبات - آدرس الکترونیکی: a.gerami@lycos.com

(دریافت: ۸۵/۹/۲۰؛ پذیرش: ۸۶/۶/۲۴)

چکیده

در این مقاله مساله استنباط در مورد پارامتر چولگی در خانواده توزیع چوله نرمال و مشکلات و دشواریهای آن مورد توجه قرار گرفته است. سپس یک برآوردگر ماکسیمم درست‌نمایی تقریبی که متکی بر برخی اطلاعات پیشین است، برای پارامتر چولگی پیشنهاد شده است. برآوردگر پیشنهادی بصورت تحلیلی و نیز با استفاده از تکنیکهای شبیه‌سازی مورد ارزیابی قرار گرفته است.

واژه‌های کلیدی: چولگی، چوله نرمال، برآورد ماکسیمم درست‌نمایی، اطلاع فیشر.

۱- مقدمه

نتایجی که از روشهای دیگری مانند روش گشتاورها بدست می‌آیند نیز مطلوب نیستند (Cheng & Traylor 1995). در بخش دوم این مقاله ابتدا مروری خواهیم داشت بر تعریف اولیه توزیع چوله نرمال استاندارد، برخی ویژگیهای مهم آن و مطالعاتی که در این زمینه صورت پذیرفته است. در بخش سوم، برآورد درست‌نمایی پارامتر چولگی و مشکلات آن مورد بررسی قرار گرفته و در بخش چهارم یک برآوردگر ماکسیمم درست‌نمایی تقریبی برای آن پیشنهاد می‌شود. در بخش پنجم برآوردگر پیشنهادی با استفاده از تکنیکهای شبیه‌سازی مورد ارزیابی قرار گرفته است. در بخش آخر تاثیر قابل توجه چولگی توزیع بر استنباط آماری در قالب یک مساله کاربردی در چارچوب رگرسیون خطی نشان داده شده است.

۲- توزیع چوله نرمال

اگر X یک متغیر تصادفی پیوسته با تابع چگالی احتمال:

$$f_X(x; \lambda) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \int_{-\infty}^{\lambda x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$= 2\phi(x)\phi(\lambda x)$$

باشد، X را یک متغیر تصادفی چوله نرمال استاندارد با پارامتر چولگی λ می‌نامیم و با نماد $X \sim SN(\lambda)$ نشان می‌دهیم (Azzalini 1985). پارامتر λ که پارامتر شکل توزیع است و بیان‌کننده جهت و میزان چولگی است، در بازه $(-\infty, \infty)$ تغییر می‌کند. چگالی بودن این تابع با استفاده از این واقعیت ثابت می‌شود که اگر f یک تابع چگالی متقارن حول صفر و G یک تابع توزیع پیوسته باشد، آنگاه $f(y) \gamma G(\lambda, y)$ یک تابع چگالی است (Azzalini 1985). این

اهمیت توزیع چوله نرمال و مبنا قرار گرفتن آن در بسیاری از مباحث استنباط آماری، سبب شده است تا تلاشهایی برای دستیابی به خانواده‌ای گسترده‌تر از توزیعها صورت پذیرد، که توزیع نرمال را به عنوان حالتی خاص دربر داشته باشد. به عنوان مثال می‌توان به توزیعهای g و h توکی (Hoaglin 1982) و توزیع نوانی باکس (Box 1953) اشاره کرد. ایده اصلی دستیابی به توزیعی با سه ویژگی، دربر داشتن توزیع نرمال به عنوان حالت خاص، دارا بودن پارامترهایی برای کنترل کشیدگی و چولگی و سادگی انجام اعمال جبری برای آن بوده است. آزالینی (Azzalini 1985) توزیع چوله نرمال (Skew Normal) را معرفی کرد. این توزیع دارای پارامتری است که چولگی را کنترل می‌کند و از اینرو مدل مناسبی برای تحلیل جوامعی است که توزیع آنها تک مدی ولی چوله است. چنین ساختاری در بسیاری از داده‌های واقعی مشاهده می‌شود. برای نمونه توزیع متغیرهایی مانند درآمد در اقتصاد و هوش در روانشناسی، ساختاری چوله دارد. برای انجام استنباط در مورد این متغیرها، این توزیع می‌تواند جایگزین مناسبی برای توزیعهای نرمال، گاما، لگ‌گاما، لگ نرمال و توزیعهایی از این دست باشد.

برخلاف ویژگیهای مطلوب این مدل احتمالاتی، در زمینه برآورد پارامتر شکل این خانواده مشکلاتی وجود دارد. هنگامی که همه مشاهدات مثبت هستند، تابع درست‌نمایی بطور یکنوا افزایشی است و برآورد ماکسیمم درست‌نمایی برای پارامتر چولگی با احتمال مثبت، بینهایت خواهد بود (Liseo 1990). بعلاوه روش ماکسیمم درست‌نمایی تنها روشی نیست که در این خانواده با مشکلاتی مواجه می‌شود،

که در آنها:

$$w(x_i) = \frac{\varphi(\lambda(\frac{x_i - \mu}{\sigma}))}{\phi(\lambda(\frac{x_i - \mu}{\sigma}))}, \quad i = 1, 2, \dots, n$$

جدا از این واقعیت که معادلات نرمال باید با استفاده از روشهای عددی حل شوند، با نگاه دقیقتر به تابع درستنمایی رفتار غیرعادی و مشکل-ساز آن خصوصا به ازای مقادیر نزدیک به صفر پارامتر چولگی که متناظر با توزیع نرمال است، بیشتر مشخص می‌شود. این رفتار غیر معمول تابع درستنمایی در همسایگی صفر، توسط آزالینی (۱۹۸۵) و چیوگنا (۱۹۹۷)، آزالینی و کاپیتانیو (۱۹۹۹) و پیوسی (۲۰۰۰)، مورد بررسی قرار گرفته است. بوضوح مشاهده می‌شود که برای مقادیر کوچک یا متوسط n ، تابع درستنمایی ممکن است بطور یکنوا کاهشی یا افزایشی باشد و این به معنی بینهایت بودن برآوردگر ماکسیمم درستنمایی λ است، که برآوردی کاملا غیرمعقول است. راه‌حلهای متفاوتی توسط افراد مختلف ارائه شده‌اند. از جمله لیزنو و لوپرفیدو (۲۰۰۶) در چهارچوب بیزی و بر این اساس که توزیع پیشین جفریز برای λ و پسین متناظر با آنها هر دو سره هستند، استفاده از میانه یا مد توزیع پسین را به عنوان برآورد پیشنهاد کرده‌اند. این برآوردگرهای بیزی براساس معیارهای بسامدی خواص مطلوبی دارند. سارتوری (۲۰۰۶) روشی بر اساس اصلاح معادلات درستنمایی ارائه کرده است. از طرفی ماتریس اطلاع فیشر در نقطه صفر برحسب λ تکین است (Azzalini 1985). شکل ۱ نمودار چند تابع لگ درستنمایی شبیه‌سازی شده را به ازای $\lambda = 4$ و اندازه نمونه ۲۰ نشان می‌دهد. ملاحظه می‌شود که در تمام موارد در نقطه صفر یک ناهمواری وجود دارد. در واقع هرچند توزیع مجانبی برآوردگر ماکسیمم درستنمایی و آماره آزمون نسبت درستنمایی آزمون فرض ساده هنگامی که ماتریس اطلاع ناتکین است، موجود هستند، اما وقتی ماتریس اطلاع تکین است، وضعیت چنین نیست. رفتار تابع درستنمایی پیشنهاد می‌کند که ممکن است به اطلاعات کمکی دیگری در مورد مدل نیاز باشد. (Liseo & Loperfido 2006).

۴- برآوردگر پیشنهادی

در راستای ایده استفاده از اطلاعات کمکی که توسط لیزنو و لوپرفیدو (۲۰۰۶) مطرح شده است، در این بخش با فرض اینکه تحلیلیگر بر اساس تجربه یا به نحو دیگری از جهت چولگی جامعه تحت مطالعه آگاه می‌باشد، با استفاده از خواص تابع توزیع تجمعی توزیع نرمال، برای پارامتر λ یک برآوردگر ماکسیمم درستنمایی تقریبی که از دقت مطلوبی برخوردار است، معرفی می‌شود. چنین فرضی در بسیاری از حالات چندان محدود کننده نیست و همانگونه که پیشتر ذکر شد، در

تابع چگالی به ازای $\lambda = 0$ ، به چگالی نرمال استاندارد تبدیل می‌شود، به ازای $\lambda = 1$ ، توزیع ماکسیمم دو متغیر تصادفی نرمال استاندارد است و میزان چولگی با λ رابطه مستقیم دارد. بعلاوه $\log f_X(x; \lambda)$ قویا تک مدی است، یعنی تابعی کاو از x است و اگر $X \sim SN(\lambda)$ آنگاه $-X \sim SN(-\lambda)$.

از دیگر ویژگیهای جالب این توزیع آن است که همانند توزیع نرمال استاندارد معمولی، مربع آن دارای توزیع خی دو با یک درجه آزادی است. سایر گشتاورهای این توزیع را می‌توان به کمک تابع مولد گشتاورهای آن:

$$M_X(t) = \sqrt{2} e^{\frac{t^2}{2}} \phi(\delta t)$$

بدست آورد، که δ در آن تابعی از پارامتر چولگی است. بعلاوه گشتاورهای اول و دوم و واریانس این توزیع عبارتند از:

$$E(X) = \sqrt{\frac{2}{\pi}} \delta, \quad E(X^2) = 1, \quad Var(X) = 1 - \left(\sqrt{\frac{2}{\pi}} \delta\right)^2$$

خانواده مکان-مقیاس چوله نرمال بصورت:

$$f_X(x; \lambda, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \int_{-\infty}^{\lambda x + \mu(1-\lambda)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

$$= \sqrt{2} \phi(x; \mu, \sigma^2) \phi\left(\lambda \left(\frac{x-\mu}{\sigma}\right)\right)$$

تعریف و با نماد $X \sim SN(\lambda, \mu, \sigma^2)$ نشان داده می‌شود. تابع مولد گشتاورهای این توزیع بصورت زیر بدست می‌آید.

$$M_X(t) = \sqrt{2} e^{\lambda t + \frac{t^2 \sigma^2}{2}} \phi(\delta t \sigma)$$

شکل تابع مولد گشتاورها نشان می‌دهد خانواده توزیعهای چوله نرمال تحت تبدیلات خطی بسته است.

۳- برآورد پارامتر چولگی

برآورد ماکسیمم درستنمایی پارامترها با در نظر گرفتن لگاریتم تابع درستنمایی:

$$\ell(\lambda, \mu, \sigma^2) = \ln(L(\lambda, \mu, \sigma^2)) = c - \frac{n}{2} \ln(\sigma^2) -$$

$$\frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2 + \sum_{i=1}^n \ln(\phi(\lambda \left(\frac{x_i - \mu}{\sigma}\right)))$$

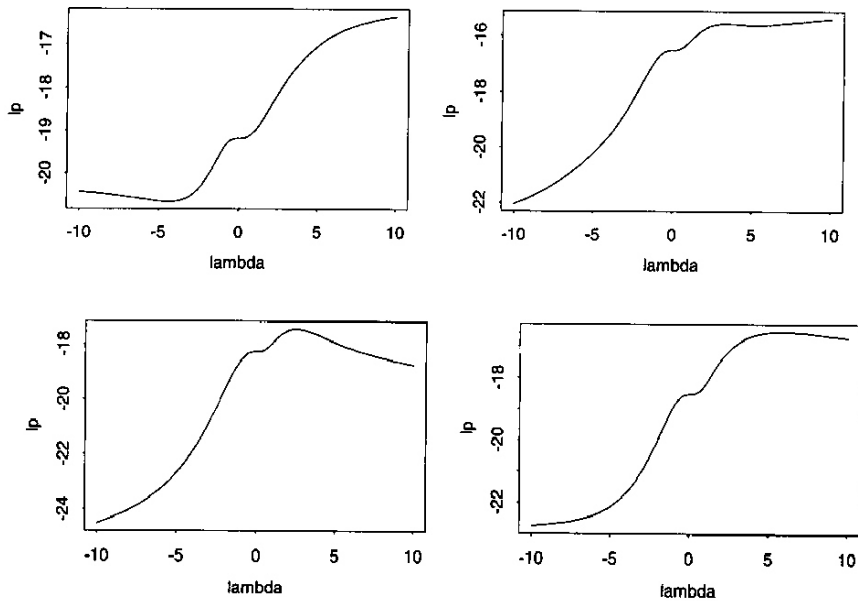
و از حل معادلات نرمال زیر بدست می‌آید.

$$\frac{\partial \ell(\lambda, \mu, \sigma^2)}{\partial \lambda} = \sum_{i=1}^n w(x_i) \left(\frac{x_i - \mu}{\sigma}\right) = 0$$

$$\frac{\partial \ell(\lambda, \mu, \sigma^2)}{\partial \mu} = \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right) + \frac{\lambda}{\sigma} \sum_{i=1}^n w(x_i) = 0$$

$$\frac{\partial \ell(\lambda, \mu, \sigma^2)}{\partial \sigma} = -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^3} - \frac{\lambda}{2\sigma} \sum_{i=1}^n w(x_i) (x_i - \mu) = 0$$

شکل ۱- توابع لگ درست‌نمایی شبیه‌سازی شده برای به ازای $\lambda = 4$ و اندازه نمونه ۲۰.



هنگامی ماکسیمم می‌شود که روابط:

$$\begin{cases} \lambda \geq \frac{\Delta\sigma}{x_i} & \lambda > 0 \\ \lambda \leq \frac{\Delta\sigma}{x_i} & \lambda < 0 \end{cases}, \quad i = 1, 2, \dots, n$$

برقرار باشند، اگر $X_{(1)}$ و $X_{(n)}$ به ترتیب کوچکترین و بزرگترین آماره ترتیبی را نشان دهند، برای برقراری روابط فوق کافی است روابط:

$$\begin{cases} \lambda \geq \frac{\Delta\sigma}{|x_{(1)}|} & \lambda > 0 \\ \lambda \leq \frac{\Delta\sigma}{|x_{(n)}|} & \lambda < 0 \end{cases}$$

برقرار باشند. چون در توزیع نرمال استاندارد عمده احتمال در بازه (۳) و (۳-) قرار دارد، هر مقدار نزدیک به ۳ می‌تواند جایگزین مناسبی برای Δ باشد. مطالعات شبیه‌سازی نشان می‌دهد که مقدار ۲/۹ منطقی است و به برآوردهایی با دقت مناسب منتهی می‌شود. لذا به عنوان نتیجه بحث فوق و با توجه به اینکه برای تشخیص علامت λ می‌توان میانگین و میانه نمونه را با هم مقایسه کرد، برآوردگر ML (تقریبی) پارامتر λ بصورت:

$$\lambda_{(X_1, \dots, X_n)}^* = \begin{cases} \frac{\Delta\sigma}{|x_{(1)}|} & \bar{X}_n > m \\ 0 & \bar{X}_n = m \\ \frac{\Delta\sigma}{|x_{(n)}|} & \bar{X}_n < m \end{cases}$$

پیشنهاد می‌شود، که در آن \bar{X}_n و m به ترتیب میانگین و میانه نمونه

بسیاری از موارد مانند مطالعه درآمد در اقتصاد، همین اطلاع پیشین است که ایده استفاده از توزیع چوله نرمال را مطرح می‌سازد. برای بدست آوردن یک برآوردگر تحلیلی، ابتدا حالت خاص را در نظر می‌گیریم. فرض کنید X_1, \dots, X_n نمونه‌ای تصادفی از توزیع $SN(\lambda, \sigma^2)$ باشد. در اینصورت تابع درست‌نمایی بصورت زیر می‌باشد.

$$\begin{aligned} L(\lambda) &= \prod_{i=1}^n f_X(x_i; \lambda, \sigma^2) \\ &= \prod_{i=1}^n \{2\phi(x_i; \lambda, \sigma^2)\phi(\lambda x_i)\} \\ &\propto \prod_{i=1}^n \phi\left(\frac{\lambda x_i}{\sigma}\right) \end{aligned} \quad (1)$$

چون $\phi(\cdot)$ تابعی نانزولی است و $0 \leq \phi(\cdot) \leq 1$ ، بنابراین هر جمله حاصلضرب (۱) به ازای بزرگترین مقدار ممکن برای λx_i ، ماکسیمم خود را کسب می‌کند. به ازای x_i های داده شده، λ بی انتخاب می‌کنیم که $L(\lambda) \propto \prod_{i=1}^n \phi\left(\frac{\lambda x_i}{\sigma}\right)$ ماکسیمم شود. اگر عدد حقیقی مثبتی باشد که برای آن $\phi(\Delta) = 1$ ، برای این منظور لازم است برای هر i نابرابری:

$$\frac{\lambda x_i}{\sigma} \geq \Delta > 0$$

برقرار باشد. چون λx_i مثبت است، در این رابطه λ و x_i هم علامت هستند، بنابراین با توجه به نانزولی بودن تابع $\phi(\cdot)$ و بسته به مثبت یا منفی بودن λ که براساس اطلاعات پیشین معلوم است، تابع $L(\lambda)$

همانطور که ملاحظه می‌شود، هر سه برآوردگر برای مقادیر کوچک λ دارای دقت بیشتری هستند و با بزرگتر شدن λ از دقت آنها کاسته می‌شود. بویژه برآوردگر گشتاوری برای نمونه‌های کوچک و چولگی بزرگ، عملکرد کاملاً نامطلوبی دارد. لازم به ذکر است که این برآوردگر تنها تحت شرایط محدود کننده $\bar{X} < \frac{2}{\pi}$ قابل محاسبه است. از طرفی با افزایش حجم نمونه بر دقت هر سه برآوردگر افزوده می‌شود، بطوریکه برای نمونه‌های بزرگتر از ۱۰۰ دقت برآوردگر پیشنهادی از برآوردگر ماکسیمم درست‌نمایی بیشتر است. توزیع شبیه-سازی شده این برآوردگر نیز براساس ۵۰۰۰ بار محاسبه λ^* برای نمونه‌های ۲۰۰ تایی از توزیع $SN(1, 0, \frac{1}{3})$ در شکل ۲ رسم شده است. هر سه برآوردگر دارای چولگی هستند و هیچکدام دقیقاً بر مقدار واقعی λ متمرکز نیستند.

۶- بررسی بیشتر

همانطور که اشاره شد به دلیل شکل پیچیده توزیع برآوردگر معرفی شده، امکان مطالعه دقیق این برآوردگر بصورت تحلیلی وجود ندارد. با این وجود در این بخش بطور تقریبی و با استفاده از توزیع مجانبی آماره‌های ترتیبی نشان داده می‌شود که شکل کلی برآوردگر پیشنهادی صحیح است.

در مباحث استنباط آماری در مورد توزیع مجانبی آماره‌های ترتیبی دو قضیه معروف وجود دارد (David 2003). این قضایا تحت شرایط کلی که مهمترین آنها پیوسته بودن توزیع تحت مطالعه است، برقرارند. توزیع بزرگترین آماره ترتیبی، $X_{(n)}$ ، به توزیع مقدار اکسترمم یا گامبل معروف است.

$$P(X_{(n)} \leq y) = \exp(-\exp(-(\frac{y - a_n}{b_n})))$$

که در آن a_n و b_n اعداد حقیقی هستند. ثابت می‌شود که a_n بطور تقریبی چندک $\frac{n}{n+1}$ ام توزیع است و رابطه امید $X_{(n)}$ با a_n خطی است.

از طرفی اگر تابع توزیع تجمعی پیوسته و روی $(0, 1)$ اکیدا یکنوا و np عددی صحیح باشد، در اینصورت آماره ترتیبی $X_{(np)}$ بطور مجانبی دارای توزیع نرمال است.

$$X_{(np)} \sim AN\left(\zeta_p, \frac{n}{[f(\zeta_p)]^2}\right)$$

که در آن ζ_p چندک p ام توزیع را نشان می‌دهد (David 2003). با انتخاب $p = \frac{1}{n}$ ، توزیع کوچکترین آماره ترتیبی بصورت:

را نشان می‌دهند. در حالت کلی، اگر X_1, \dots, X_n نمونه‌ای تصادفی از توزیع $SN(\lambda, \mu, \sigma^2)$ باشد، در اینصورت تابع درست‌نمایی بصورت:

$$L(\lambda) = \prod_{i=1}^n f_X(x_i; \lambda, \mu, \sigma^2) \\ = \prod_{i=1}^n \left\{ \frac{2}{\sigma} \phi(x_i; \lambda, \mu, \sigma^2) \phi\left(\lambda \left(\frac{x_i - \mu}{\sigma}\right)\right) \right\}$$

است. با استدلالی مشابه، برآوردگر ML تقریبی برای پارامتر λ در این حالت عبارت است از:

$$\lambda_{(X_1, \dots, X_n)}^* = \begin{cases} \frac{2/9\sigma}{|\min(x_i - \mu)|} & \bar{X}_n > m \\ \cdot & \bar{X}_n = m \\ -\frac{2/9\sigma}{|\max(x_i - \mu)|} & \bar{X}_n < m \end{cases} \quad (2)$$

بدیهی است که در صورت نامعلوم بودن μ و σ^2 این پارامترها باید به نحوی برآورد و در رابطه (۲) جایگزین شوند. برای توزیع نرمال چوله استاندارد این برآوردگر به شکل زیر در می‌آید:

$$\lambda_{(X_1, \dots, X_n)}^* = \begin{cases} \frac{2/9}{|x_{(1)}|} & \bar{X}_n > m \\ \cdot & \bar{X}_n = m \\ -\frac{2/9}{|x_{(n)}|} & \bar{X}_n < m \end{cases} \quad (2)$$

۵- شبیه‌سازی

بررسی خواص این برآوردگر با توجه به شکل پیچیده آن دشوار است. اما می‌توان خواص آن را بوسیله شبیه‌سازی مورد بررسی قرار داد. برآوردگر مونت کارلویی امید این برآوردگر عبارت است از:

$$E(\lambda^*) \approx \frac{1}{M} \sum_{i=1}^M \lambda_i^*$$

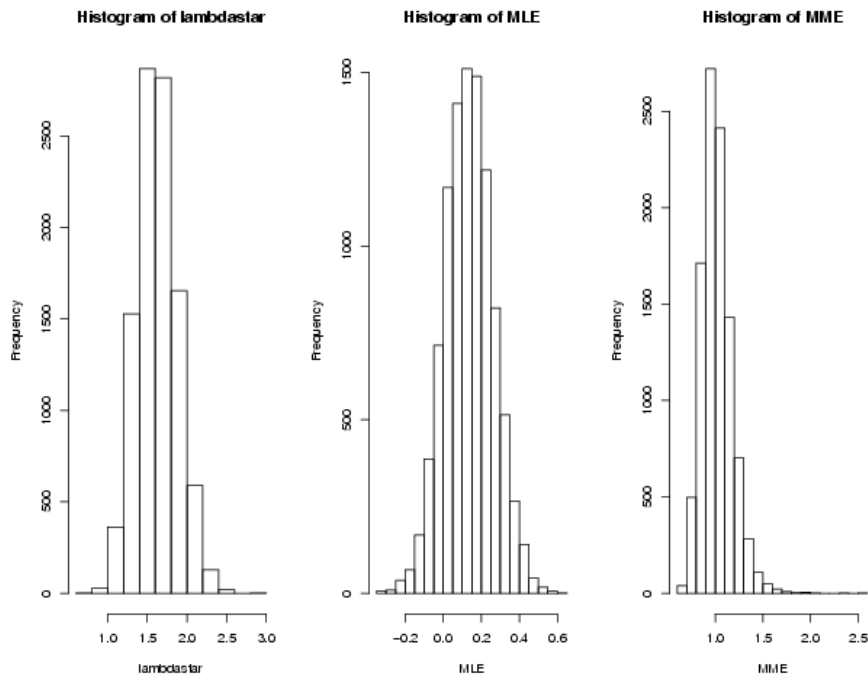
که در آن λ_i^* براساس یک نمونه n تایی از رابطه (۳) محاسبه می‌شود. میانگین توان دوم خطای آن را نیز می‌توان بصورت:

$$MSE(\lambda_i^*) = \frac{1}{M} \sum_{i=1}^M (\lambda - \lambda_i^*)^2$$

محاسبه نمود. این مقادیر به همراه مقادیر متناظرشان برای برآوردگرهای ماکسیمم درست‌نمایی و گشتاوری در جدول ۱ به ازای اندازه‌های مختلف نمونه و چند مقدار از پارامتر چولگی نمایش داده شده‌اند. برآوردگر ماکسیمم درست‌نمایی با استفاده از بسته SN در نرم افزار R و برآوردگر گشتاوری از رابطه $\frac{\bar{X}}{\sqrt{(\frac{1}{\pi}) - \bar{X}^2}}$ محاسبه شده است. در کلیه محاسبات تعداد تکرارها $M = 5000$ در نظر گرفته شده است.

جدول ۱: میانگین توان دوم خطا و اریبی برآوردگر پیشنهادی (λ^*)، ماکسیمم درست‌نمایی (MLE) و گشتاوری (MME).

n	λ	MSE			BIAS		
		λ^*	MLE	MME	λ^*	MLE	MME
100	2	2/7005	1/8468	18/9710	-0/9897	1/1642	-0/7015
	1	0/8067	-0/6091	-0/5145	-0/5822	0/6447	-0/0774
	-1	0/7967	-0/6107	-0/3735	0/5781	-0/6470	-0/0879
	-2	2/7624	1/8447	47/1700	1/0092	-1/1628	-0/7078
200	1	1/1954	1/8293	32/4850	-0/6743	1/1645	-0/5710
	2	0/4462	-0/5866	-0/554	0/4396	0/6488	-0/290
	-1	0/4429	-0/5874	-0/439	0/4384	-0/6494	-0/234
	-2	1/1928	1/8246	18/5673	0/6799	-1/1634	-0/5536
300	1	0/7874	1/8131	5/8107	0/5476	1/1619	-0/3418
	2	0/4465	-0/5866	-0/298	0/3801	0/6474	-0/287
	-1	0/3319	-0/5860	-0/271	0/3803	-0/6480	-0/182
	-2	0/8976	1/8141	3/0832	0/5464	-1/1619	-0/3210
400	1	0/5832	1/2088	10/1305	0/7078	1/1629	-0/2978
	2	0/2679	-0/5708	-0/200	-0/3419	0/6472	-0/0116
	-1	0/2676	-0/5718	-0/183	0/3417	-0/6479	-0/0125
	-2	0/5701	-0/8152	1/4878	0/4636	-1/8637	-0/2091



شکل ۲- توزیع شبیه‌سازی شده برآوردگر پیشنهادی، ماکسیمم درست‌نمایی و گشتاوری با فرض $\lambda = \frac{1}{3}$ و $n = 20$.

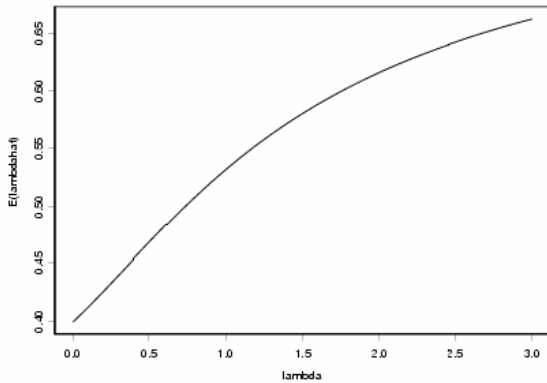
برقرار هستند و با در نظر گرفتن توضیحات فوق، می‌توان برای n های به اندازه کافی بزرگ، نوشت:

$$E(X_{(1)}) = \zeta_{\frac{1}{n}}^* , \quad E(X_{(n)}) = \zeta_{\frac{n}{n+1}}^*$$

که در آنها ζ_p^* چندک p ام توزیع چوله نرمال است. متأسفانه هیچ

$$X_{(n)} \sim AN \left(\zeta_{\frac{1}{n}}^* , \frac{\frac{1}{n} \left(1 - \frac{1}{n}\right)}{\left[f \left(\zeta_{\frac{1}{n}}^* \right) \right]^2} \right)$$

خواهد بود. با توجه به اینکه شرایط این قضایا برای توزیع چوله نرمال



شکل ۳- رابطه بین امید ریاضی برآوردگر پیشنهادی و λ .

۶- یک مساله کاربردی

همانطور که در مقدمه ذکر شد، موارد زیادی وجود دارد که در آن توزیع جامعه تحت مطالعه متقارن نیست. در این قسمت با ارائه یک مساله کاربردی نشان داده می‌شود که مد نظر قرار دادن این عدم تقارن تا چه حد می‌تواند در استنباط آماری حائز اهمیت باشد. داده‌های جدول ۲ میزان درآمد (Y) و تعداد ساعات اضافه کار (X) را برای ۷۰ کارمند استان قزوین در سال ۱۳۸۲ نشان می‌دهد. بافت نگار داده‌های درآمد در شکل ۴ نشان دهنده چولگی جامعه تحت مطالعه است. نتایج حاصل از برازش مدل رگرسیونی در دو حالت مانده‌های نرمال و چوله نرمال، در جدول ۳ ارائه شده است. در هر دو حالت، پارامترهای مدل با استفاده از روش ماکسیمم درست‌نمایی برآورد شده‌اند. میزان برازش هر مدل براساس آماره پرس (Press Statistics) مورد ارزیابی قرار گرفته است.

رابطه بسته‌ای برای چندکهای توزیع چوله نرمال وجود ندارد، با این وجود براساس نابرابری زیر (Azzalini 1985):

$$-\frac{1}{\pi} \arctan \lambda \leq \phi^*(z; \lambda) + \phi(z) \leq \frac{1}{\pi} \arctan \lambda$$

و با در نظر گرفتن یکی از کرانه‌ها می‌توان نوشت:

$$\zeta_p^* \approx \zeta_p - \frac{1}{\pi} \arctan |\lambda|$$

که در آن $\phi^*(z; \lambda)$ و ζ_p^* به ترتیب نشان دهنده تابع توزیع تجمعی و چندک pام توزیع چوله نرمال هستند. اکنون با جایگذاری مقادیر $\frac{1}{n}$ و $\frac{n}{n+1}$ برای p، خواهیم داشت:

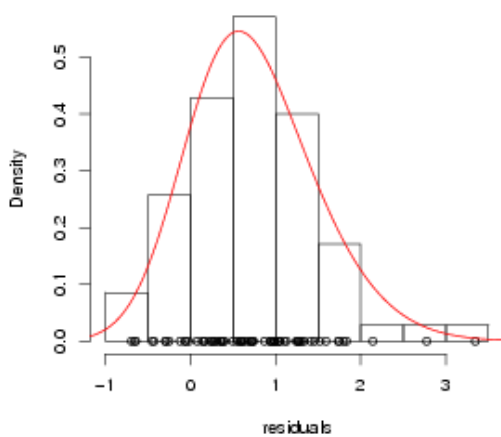
$$E(\lambda^*) \approx \frac{1}{E(X_{(1)})} \approx \frac{1}{\zeta_{\frac{1}{n}} - \frac{1}{\pi} \arctan \lambda}, \quad \lambda > 0$$

$$E(\lambda^*) \approx \frac{1}{E(X_{(n)})} \approx -\frac{1}{\zeta_{\frac{n}{n+1}} + \frac{1}{\pi} \arctan \lambda}, \quad \lambda < 0$$

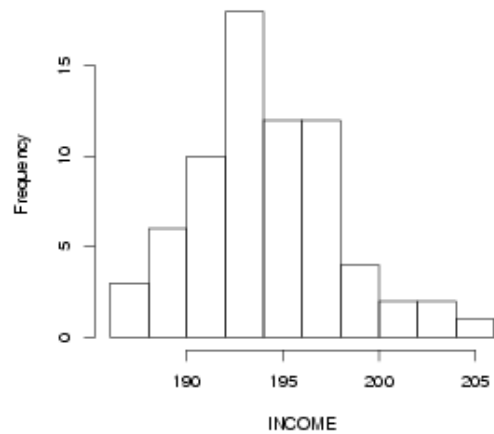
شکل ۳ نشان می‌دهد که به ازای $|\lambda| < 2$ ، خط $E(\lambda^*) = \lambda$ تقریباً بر نیمساز ناحیه اول منطبق است، یعنی اگر چولگی زیاد نباشد، λ^* برآوردگری تقریباً ناریب برای λ است و همانطور که بررسیهای شبیه‌سازی نیز نشان داد، به ازای مقادیر بزرگ پارامتر چولگی از دقت برآوردگر معرفی شده کاسته می‌شود. البته این یک مشکل عمومی در بین برآوردگرهای موجود می‌باشد و مطالعات بیشتری در این زمینه لازم است. از دیگر مشکلات این برآوردگر تاثیر شدید داده‌های پرت بر آن است.

جدول ۲: میزان درآمد (Y) و تعداد ساعات اضافه کار (X) برای ۷۰ کارمند استان قزوین در سال ۱۳۸۲.

X	۸	۴	۱۰	۶	۸	۹	۱۴	۵	۴	۹
Y	۱۹۱/۴۱	۱۹۱/۷۲	۱۹۶/۵۰	۱۹۰/۵۷	۱۹۵/۳۱	۱۹۲/۰۴	۱۹۹/۱۶	۱۹۱/۳۶	۱۹۷/۲۲	۲۰۵/۵۲
X	۷	۵	۱۰	۵	۸	۷	۶	۹	۷	۵
Y	۱۸۷/۳۶	۱۹۷/۱۳	۱۹۱/۳۶	۱۹۷/۷۵	۱۹۵/۴۱	۱۹۵/۳۱	۱۹۲/۴۹	۱۹۴/۵۰	۱۹۶/۴۹	۱۹۵/۶۳
X	۷	۷	۱۳	۹	۶	۸	۶	۸	۷	۱۰
Y	۱۹۴/۲۶	۱۸۹/۱۶	۱۹۶/۸۲	۱۹۳/۴۹	۱۸۷/۵۹	۱۹۲/۱۱	۱۹۶/۴۵	۱۹۶/۴۱	۱۹۲/۹۱	۱۹۴/۶۰
X	۴	۹	۶	۷	۶	۷	۵	۷	۹	۷
Y	۱۹۰/۴۱	۱۹۰/۵۳	۱۹۲/۱۶	۱۹۳/۵۴	۱۹۴/۰۷	۱۸۸/۲۷	۱۹۵/۵۹	۱۸۹/۵۰	۱۹۶/۰۸	۱۹۱/۸۸
X	۷	۷	۱۰	۹	۶	۱۰	۹	۱۰	۹	۶
Y	۱۹۲/۷۹	۱۹۳/۱۵	۱۹۲/۱۲	۱۹۵/۴۶	۱۹۰/۴۸	۱۸۸/۵۵	۱۹۲/۲۶	۱۸۶/۴۰	۱۸۹/۶۸	۱۹۳/۶۸
X	۱۰	۹	۸	۶	۸	۷	۹	۸	۷	۴
Y	۱۸۹/۷۴	۲۰۰/۶۵	۱۹۳/۵۴	۲۰۰/۰۲	۱۹۳/۲۸	۱۹۹/۳۳	۱۹۴/۹۵	۱۹۳/۴۲	۱۹۷/۸۵	۱۹۳/۲۶
X	۹	۷	۶	۳	۵	۱۰	۸	۷	۶	۱۲
Y	۱۹۳/۸۱	۱۹۲/۶۱	۱۹۲/۹۱	۱۹۹/۱۲	۱۹۰/۸۵	۲۰۳/۴۸	۱۹۷/۱۴	۲۰۳/۱۶	۱۹۷/۶۹	۱۹۹/۶۷



شکل ۵: بافت‌نگار باقیمانده‌های مدل رگرسیونی.



شکل ۴: بافت‌نگار داده‌های درآمد.

جدول ۳: مقایسه مدل‌های رگرسیونی تحت فرض خطاهای نرمال و چوله -

نرمال.

توزیع مانده‌ها	نرمال	چوله نرمال
β_0	۱۹۲/۷۶	۱۹۳/۴۰
$S(\beta_0)$	۱/۷۱۶۲	۱/۴۹۹۰
β_1	۰/۱۸۷۱	۰/۱۰۳۴
$S(\beta_1)$	۰/۲۲۰۰	۰/۲۳۱۷
$Press(MSE)$	۲۲/۹۰۰۵	۱۵/۸۹۸۹

تفاوت قابل توجه بین آماره پرس در دو مدل، به خوبی اهمیت در نظر گرفتن چولگی داده‌های درآمد را نشان می‌دهد. بافت‌نگار شکل ۵ نشان دهنده چولگی توزیع مانده‌ها می‌باشد.

منابع:

- Azzalini A. 1985: A class of distributions which includes the normal ones. *Scan. J. Statist.* **12**: 171-178.
- Azzalini A., Capitanio A. 1999: Statistical applications of multivariate skew normal distribution. *J. Roy. Statist. Soc. B*, **61**: 579-602.
- Box G.E.P. 1953: A note on regions of kurtosis. *Biometrika*. **40**: 465-468.
- Chiogna M. 1997: Notes on estimation problems with scalar skew-normal distribution. Technical Report 15, Dept. of statistical Sciences, university of Padua.
- Cheng C.H., Traylor L. 1995: non-Regular maximum likelihood problems (with discussion). *J. Roy. Statist. Soc. B*. **57**: 3-44.
- David H.A. 2003: Order Statistics. Third Edition, John Wiley & Sons, Inc., New York.
- Hoaglin P. 1982: g- and h-distributions. In Encyclopedia of Statistical Science (ed. N. Johnson & S. Kotz). Wiley. New York.
- Liseo B. 1990: The skew-normal class of densities: inferential aspects a Bayesian view of point. *Statistica*. **50**: 59-70.
- Liseo B. Loperfido N. 2006: A note on reference priors for scalar skew-normal distribution. *J. Statist. Plann. Inference*. **136**: 378-389.
- Pewsey A. 2000: Problems of inference for Azzalini's skew-normal distribution. *J. Appl. stat.* **27**: 859-870. *J. Statist. Plann. Inference*.
- Sartori n. 2006: Bias prevention of maximum likelihood estimator: skew normal and t distribution. *J. Statist. Plann. Inference*, **136**, 12, 4259-4275.