توزیع چوله لاپلاس-نرمال و انتخاب مدل در این خانواده از توزیع‌ها

قباد برمال‌زن - عابدین حیدری - منیره کریمی

دانشگاه زابل

چکیده: اکثر تحلیل‌ها و استنباط‌های آماری براساس فرض نرمال بودن مشاهدات انجام می‌گیرد. در حالی‌که در بسیاری از موارد، توزیع مشاهدات نامتقارن بوده و در نتیجه تحلیل این مجموعه از مشاهدات با استفاده از توزیع نرمال، ممکن است خیلی مناسب نباشد. در چنین مواقعی معمولاً سعی می‌شود ابتدا با استفاده از تبدیلی، توزیع مشاهدات را تا حدامکان به توزیع نرمال نزدیک کرده و سپس به تحلیل مشاهدات پرداخته‌می‌شود. اما این روش خود با دشواری‌هایی از جمله چگونگی انتخاب تبدیل مناسب و اریبی برآورد کننده‌ها مواجه است. لذا در چنین مواردی، توزیع چوله‌نرمال که توزیعی نامتقارن است می‌تواند نقش اساسی در تحلیل این مشاهدات ایفا نماید. دلایل خوبی و استفاده از توزیع چوله نرمال، در برداشتن توزیع نرمال به عنوان یک حالت خاص و همچنین دارا بودن پارامترهایی برای کنترل چولگی و کشیدگی ذکر شده است. آزالینی (۱۹۸۵) نشان دادند که مقدار ضریب چولگی و ضریب کشیدگی برای این توزیع محدود است و این مسأله می‌تواند به عنوان یک ضعف از این توزیع تلقی شود. زیرا برای مشاهداتی که دارای چولگی یا کشیدگی شدیدی باشند غیر قابل استفاده است. در این مقاله، به توزیع چوله لاپلاس-نرمال و خواص آن پرداخته شده است که دارای ضریب کشیدگی با دامنه وسیع‌تری نسبت به توزیع چوله نرمال آزالینی است و چندین روش برای شبیه سازی از این توزیع ارائه شده است. همچنین به مقایسه بین تابع چگالی احتمال‌های چوله لاپلاس-نرمال و چوله‌نرمال، براساس معیار کولبک-لیبلر، به منظور تقریب بهتری، برای تابع چگالی مشاهدات نامتقارن پرداخته شده است.

واژه‌های کلیدی: توزیع چوله نرمال، توزیع چوله لاپلاس-نرمال، انتخاب مدل، معیار کولبک-لیبلر

۱ مقدمه

انتخاب مدل یکی از مفاهیم اساسی در استنباط‌آماری است که در روش کلاسیک محدود به بررسی پارامترهای جامعه می‌شود. پیشنهاد یک مدل قطعی، براساس تعداد محدودی

Archive of SID

www.SID.ir

از مشاهدات، ممکن است موجب بوجود آمدن ریسک بزرگی در انتخاب مدل شود. زمانی‌که مشاهدات نامتقارن باشند آزالینی مدل توزیع چوله نرمال را با تابع چگالی احتمال $f(z;\lambda) = ۲\phi(z)\,\Phi(\lambda z),\ z \in \Re$ برای تحلیل مشاهدات معرفی نمودند که در آن $\lambda \in \Re$ بیانگر پارامتر کنترل چولگی است. آزالینی (۱۹۸۵) نشان داد چولگی این توزیع بین (۰/۹۹۵, ۰/۹۹۵−) و کشیدگی آن بین (۰/۸۶۹, ۰/۸۶۹−) است و این محدویت می‌تواند به عنوان یک ضعف از این توزیع تلقی شود. زیرا برای یک مجموعه از مشاهدات که دارای چولگی یا کشیدگی شدیدی باشند غیر قابل استفاده است. تابع چگالی احتمال توزیع چوله لاپلاس-نرمال به صورت زیر است:

$$f(z;\lambda) = ۲\psi(z)\,\Phi(\lambda z), \qquad z \in \Re,\ \lambda \in \Re,$$

که در آن $\psi(z) = \frac{۱}{۲}e^{-|z|},\ z \in \Re$ بیانگر تابع چگالی احتمال توزیع لاپلاس و Φ تابع توزیع نرمال استاندارد است.

توزیع چوله-نرمال یک متغیره، نخستین بار توسط آزالینی (۱۹۸۵) معرفی گردید. خواص و ویژگی‌های این توزیع، بعد از آزالینی، توسط هنز(۱۹۸۶)، آزالینی و دالاواله (۱۹۹۶) آزالینی و کاپیتانو(۱۹۹۹)، برانکو و دی (۲۰۰۱)، آرنولد و بیور(۲۰۰۲)، بالاکریشنان(۲۰۰۲) و آزالینی و چیوگنا(۲۰۰۴) مورد بررسی قرار گرفته است.

در بخش ۲ این مقاله، توزیع چوله لاپلاس-نرمال و برخی از خواص آن آورده شده است. بخش ۳ شامل تبدیل‌هایی روی متغیرهای تصادفی مختلف، برای ساخت و تولید داده تصادفی از توزیع چوله لاپلاس-نرمال است و سرانجام در بخش ۴ به انتخاب مدل براساس معیار کولبک-لیبلر برای مشاهداتی که توزیع آن‌ها نامتقارن است پرداخته شده است.

۲ خواص توزیع چوله لاپلاس-نرمال

در این بخش، باایده گرفتن از ویژگی‌های توزیع چوله نرمال، به خواصی از توزیع چوله لاپلاس- نرمال پرداخته شده است که در قالب چند لم بیان شده است.

لم ۱ فرض کنید Z دارای توزیع چوله لاپلاس-نرمال با پارامتر λ باشد. در این صورت می‌توان خواص زیر را از این توزیع برشمرد:

الف- متغیر تصادفی $-Z$ دارای توزیع چوله لاپلاس-نرمال با پارامتر $-\lambda$ است.

ب- اگر $\lambda = ۰$ باشد آنگاه Z دارای توزیع لاپلاس خواهد شد.

پ- اگر $\lambda \to \pm\infty$ آنگاه توزیع Z به توزیع نیم لاپلاس میل می‌کند.

Archive of SID

www.SID.ir

برهان: الف- فرض کنید Z دارای توزیع چوله لاپلاس-نرمال با پارامتر λ باشد. در این صورت تابع چگالی احتمال متغیر تصادفی $-Z$ عبارت است از:

$$\begin{aligned} f(-z;\lambda) &= 2\psi(-z)\,\Phi(-\lambda z) \\ &= 2\psi(z)\,\Phi((-\lambda)z) \\ &= f(z;\,-\lambda). \end{aligned}$$

ب- اثبات این قسمت با جایگذاری $\lambda = 0$ در تابع چگالی احتمال چوله لاپلاس-نرمال براحتی حاصل می‌شود.
پ- اگر $\lambda \to +\infty$ میل کند داریم:

$$\begin{aligned} \lim_{\lambda\to+\infty} f(z;\lambda) &= \lim_{\lambda\to+\infty} 2\psi(z)\,\Phi(\lambda z) \\ &= 2\psi(z) \lim_{\lambda\to+\infty}\int_{-\infty}^{\lambda z}\Phi(t)\,dt \\ &= e^{-|z|}\,I_{(0,+\infty)}(z). \end{aligned}$$

بطور مشابه می‌توان نشان داد که اگر $\lambda \to -\infty$ میل کند آنگاه توزیع Z به توزیع نیم لاپلاس $e^{-|z|}\,I_{(-\infty,0)}(z)$ میل می‌کند.
کارلین (۱۹۶۸) نشان داد تابع توزیع ناتباهیده F تک مدی است اگر و فقط اگر لگاریتم تابع چگالی احتمال متناظر آن، به ازای تکیه گاه متغیر، تابعی مقعر باشد. به عبارت دیگر مشتق دوم لگاریتم تابع چگالی احتمال آن به ازای تمام مقادیر تکیه گاه منفی باشد.

لم ۲ تابع توزیع چوله لاپلاس-نرمال به ازای یک λ ی ثابت، تک مدی است.

برهان: برای اینکه نشان دهیم که $f(z;\lambda)$ تک مدی است کافیست نشان دهیم که $\frac{\partial^2}{\partial z^2}\ln f(z;\lambda)$ برای هر مقدار z همواره منفی است. بنابراین

$$\begin{aligned} \frac{\partial^2}{\partial z^2}\ln f(z;\lambda) &= \frac{\partial^2}{\partial z^2}\{\ln 2 + \ln\psi(z) + \ln\Phi(\lambda z)\} \\ &= \frac{\partial^2}{\partial z^2}\ln\psi(z) + \frac{\partial^2}{\partial z^2}\ln\Phi(\lambda z) \\ &= \frac{\partial^2}{\partial z^2}\{-|z|\} + \frac{\partial}{\partial z}(\frac{\lambda\,\phi(\lambda z)}{\Phi(\lambda z)} \\ &= 0 + \frac{\lambda^2\,\phi(\lambda z)}{\Phi(\lambda z)}\left\{\frac{\phi(\lambda z)}{\Phi(\lambda z)} + \lambda z\right\}. \end{aligned}$$

Archive of SID

برای اینکه نشان دهیم عبارت فوق همواره منفی است کافیست نشان دهیم که عبارت $\lambda z + \frac{\phi(\lambda z)}{\Phi(\lambda z)}$ مثبت است. بدین منظور دو حالت زیر ممکن است اتفاق افتد:

۱– اگر $\lambda z \geq 0$ باشد آنگاه $\lambda z + \frac{\phi(\lambda z)}{\Phi(\lambda z)} \geq 0$ است و نتیجه برقرار است.

۲ اگر $\lambda z < 0$ باشد با استفاده از تغیر متغیر $t = -\lambda z$ و نامساوی میل[1] $1 - \Phi(x) \leq (\frac{1}{x}\,\phi(x))$ بسادگی می‌توان نشان داد که عبارت $\lambda z + \frac{\phi(\lambda z)}{\Phi(\lambda z)}$ مثبت است. بنابراین بدین ترتیب حکم ثابت می‌شود.

۳ روش ساخت‌وتولید داده تصادفی از توزیع چوله لاپلاس–نرمال

در این بخش، با استفاده از تبدیلاتی روی متغیرهای تصادفی مختلف، به روش ساخت متغیر چوله لاپلاس–نرمال پرداخته شده است که با استفاده از نرم افزارهای آماری، براحتی می‌توان از این توزیع، داده تصادفی تولید نمود.

لم **۳** فرض کنید X و Y متغیرهای تصادفی مسقل به ترتیب با توزیع‌های نرمال استاندارد و لاپلاس باشند. آنگاه متغیر تصادفی Z باتعریف Y به شرط $X \leq \lambda Y$ دارای توزیع توزیع چوله لاپلاس–نرمال با پارامتر λ است.

برهان: تابع توزیع Z به صورت زیر است:

$$\begin{aligned} F(z;\lambda) &= P(Z \leq z) \\ &= P\left(Y \leq z | X \leq \lambda Y\right) \\ &= \frac{P\left(Y \leq z, X \leq \lambda Y\right)}{P(X \leq \lambda Y)} \\ &= \frac{\int_0^z P\left(X \leq \lambda Y | Y = y\right)\ \psi(y)dy}{P(X \leq \lambda Y)} \\ &= \frac{\int_0^z P\left(X \leq \lambda Y\right)\ \psi(y)dy}{\frac{1}{2}}) \\ &= 2\int_0^z \Phi(\lambda y)\,\psi(y)\,dy. \end{aligned}$$

[1] Mill's Inequality

Archive of SID

www.SID.ir

همچنین تابع چگالی احتمال Z به صورت زیر است:

$$f(z;\lambda) = \frac{\partial}{\partial z} F(z;\lambda) = ۲\Phi(\lambda z)\,\psi(z).$$

لم **۴** فرض کنید X و Y متغیرهای تصادفی مسقل به ترتیب با توزیع‌های نرمال استاندارد و لاپلاس باشند. آنگاه متغیر تصادفی

$$Z = \begin{cases} Y & X \leq \lambda Y, \\ -Y & X > \lambda Y, \end{cases}$$

دارای توزیع چوله لاپلاس-نرمال است.

برهان: تابع توزیع Z عبارت است از:

$$\begin{aligned} F(z;\lambda) &= P(Z \leq z) \\ &= P(Z \leq z, X \leq \lambda Y) + P(Z \leq z, X > \lambda Y) \\ &= P(Y \leq z, X \leq \lambda Y) + P(-Y \leq z, X > \lambda Y) \\ &= \int_{-\infty}^{z} P(X \leq \lambda Y | Y = y)\,\psi(y)\,dy + \int_{-z}^{+\infty} P(X > \lambda Y | Y = y)\,\psi(y)\,dy \\ &= \int_{-\infty}^{z} \Phi(\lambda y)\psi(y)\,dy + \int_{-z}^{+\infty} (۱ - \Phi(\lambda y))\psi(y)\,dy. \end{aligned}$$

بنابراین تابع چگالی احتمال Z به صورت زیر است:

$$\begin{aligned} f(z;\lambda) &= \frac{\partial}{\partial z} F(z;\lambda) \\ &= \Phi(\lambda z)\psi(z) + (۱ - \Phi(\lambda z))\psi(z) \\ &= \Phi(\lambda z)\psi(z) + \Phi(\lambda z)\psi(z) \qquad (۱ - \Phi(-t) = \Phi(t)) \\ &= ۲\Phi(\lambda z)\,\psi(z). \end{aligned}$$

لم **۵** فرض کنید X و Y متغیرهای تصادفی مسقل به ترتیب با توزیع‌های یکنواخت $(۰,۱)$ و لاپلاس باشند. در اینصورت متغیر تصادفی Z به صورت Y با شرط $X \leq \Phi(\lambda Y)$ دارای توزیع چوله لاپلاس-نرمال با پارامتر λ است.

Archive of SID

www.SID.ir

برهان: تابع توزیع Z عبارت است از:

$$\begin{aligned} F(z;\lambda) &= P(Z \le z) \\ &= P(Y \le z | X \le \Phi(\lambda Y)) \\ &= \frac{P(Y \le z, X \le \Phi(\lambda Y))}{P(X \le \Phi(\lambda Y))} \\ &= \frac{\int_{-\infty}^{z} P(X \le \Phi(\lambda Y)|Y=y)\,\psi(y)\,dy}{P(X \le \Phi(\lambda Y))} \\ &= \frac{\int_{-\infty}^{z} P(X \le \Phi(\lambda Y)|Y=y)\,\psi(y)\,dy}{\frac{۱}{۲}} \\ &= ۲\int_{-\infty}^{z} P(X \le \Phi(\lambda Y)|Y=y)\,\psi(y)\,dy. \end{aligned}$$

همچنین تابع چگالی احتمال Z به صورت زیر است:

$$\begin{aligned} f(z;\lambda) &= \frac{\partial}{\partial z} F(z;\lambda) \\ &= ۲\Phi(\lambda z)\,\psi(z). \end{aligned}$$

مخرج کسر بالا یعنی $P(X \le \Phi(\lambda Y))$ برابر $\frac{۱}{۲}$ است. زیرا

$$\begin{aligned} P(X \le \Phi(\lambda Y)) &= \int_{-\infty}^{+\infty} P(X \le \Phi(\lambda Y)|Y=y)\,\psi(y)\,dy \\ &= \int_{-\infty}^{+\infty} P(X \le \Phi(\lambda y))\,\psi(y)\,dy \\ &= \frac{۱}{۲}\int_{-\infty}^{+\infty} ۲\Phi(\lambda y)\,\psi(y)\,dy \\ &= \frac{۱}{۲} \times ۱ \\ &= \frac{۱}{۲}. \end{aligned}$$

Archive of SID

۴ انتخاب‌مدل براساس معیارکولبک-لیبلر برای مشاهدات نامتقارن

فرض کنید یک نمونه تصادفی $Z_n, ..., Z_۲, Z_۱$ از توزیعی با تابع چگالی درست $h(.)$ در اختیار داریم. در حالت کلی h معمولاً نامعلوم است. گام اول در تشخیص و تقریب چگالی این مشاهدات، استفاده از هیستوگرام مشاهدات است. در صورتی که هیستوگرام مشاهدات، تک مدی و نامتقارن باشد ممکن است توزیع نرمال، تقریب مناسبی برای توصیف مشاهدات نباشد. در چنین مواقعی، توزیع چوله نرمال و چوله لاپلاس-نرمال می‌توانند به عنوان تقریب‌هایی از این چگالی درست، پیشنهاد شوند. بطور بدیهی می بایست تقریب‌های مورد نظر، به چگالی درست h نزدیک باشند تا به استنباط معتبر در مورد جامعه برسیم. در این بخش، ما به بررسی این سؤال اساسی پرداخته‌ایم که چگونه می‌توان از بین دو مدل پیشنهاد شده فوق، نزدیکترین وبه عبارتی بهترین مدل را از لحاظ نزدیکی به چگالی درست h پیدا کرد. برای پاسخ به این سؤال از معیار کولبک لیبلر استفاده شده است.

فرض کنید Z دارای چگالی درست h باشد که از آن بی‌خبریم. همچنین فرض کنید مدل $\mathcal{F}_\Lambda = \{f(z;\lambda) = ۲\psi(z)\,\Phi(\lambda z), \quad \lambda \in \Lambda \subset \Re\}$ یک مدل پیشنهاد شده برای تقریب چگالی درست h باشد. معیار کولبک-لیبلر بین h و f_λ به صورت زیر تعریف شده است:

$$KL(h; f_\lambda) = E_h\left\{\log \frac{h(Z)}{f(Z;\lambda)}\right\},$$

که در آن E_h بیانگر امید ریاضی تحت چگالی درست h می‌باشد. این معیار به اختصار با نماد KL نشان داده می‌شود و دارای خواص زیر است:

الف) $KL(h; f_\lambda) \geq ۰$.

ب) $KL(h; f_\lambda) = ۰$ اگر و فقط اگر $\lambda_\star \in \Lambda$ چنان وجود داشته باشد که $h = f(x, \lambda_\star)$ باشد.

فرض کنید $\mathcal{F}_\Lambda$ و $\mathcal{G}_\Gamma = \{g(z;\gamma) = ۲\psi(z)\,\Phi(\gamma z), \quad \gamma \in \Gamma \subset \Re\}$ که به ترتیب خانواده توزیع‌های چوله نرمال و چوله لاپلاس-نرمال هستند، دو مدل پیشنهاد شده برای تقریب چگالی درست h باشند. اختلاف معیار کولبک-لیبلر بین این دو مدل را با نماد $\Delta(f_\lambda; g_\gamma)$ نشان داده و عبارت است از:

$$\begin{aligned}\Delta(f_\lambda; g_\gamma) &= KL(h; f_\lambda) - KL(h; g_\gamma)\\ &= E_h[\log g(Z;\gamma)] - E_h[\log f(Z;\lambda)]\\ &= E_h[\log \psi(z)\,\Phi(\gamma z)] - E_h[\log \phi(z)\,\Phi(\lambda z)].\end{aligned}$$

Archive of SID

www.SID.ir

کمیت $\Delta(f_\lambda; g_\gamma)$ یک کمیت آگاهی بخش است، بدین معنا که اگر $\Delta(f_\lambda; g_\gamma) = 0$ باشد می‌توان نتیجه گرفت که دو مدل پیشنهاد شده فوق، به یک اندازه چگالی درست h را توصیف می‌کنند. اگر $\Delta(f_\lambda; g_\gamma) > 0$ باشد آنگاه مدل چوله لاپلاس-نرمال بهتر از مدل چوله نرمال است. با استدلالی مشابه، همچنین می‌توان نتیجه گرفت که اگر $\Delta(f_\lambda; g_\gamma) < 0$ باشد آنگاه مدل چوله نرمال بهتر از مدل چوله لاپلاس-نرمال است.

بحث و نتیجه‌گیری

توزیع چوله نرمال دارای مقادیر ضریب چولگی و کشیدگی کراندار است. لذا در هنگام انتخاب مدل، برای مجموعه‌ایی از مشاهدات که شدیداً چوله یا شدیداً کشیده هستند نمی‌توان این توزیع را بکار برد. به همین دلیل چون توزیع لاپلاس-نرمال چوله دارای ضریب کشیدگی وسیعتری نسبت به توزیع نرمال-چوله است لذا در چنین مواردی بکار بردن این توزیع، منطقی‌تر بنظر می‌رسد. همچنین تمایز دیگری که توزیع چوله لاپلاس- نرمال با توزیع چوله t- نرمال دارد این است که این توزیع تک مدی است در حالی‌که توزیع چوله t-نرمال دو مدی است. لذا در مقایسه بین مشاهدات نامتقارن تک مدی، دو مدل چوله نرمال و چوله لاپلاس-نرمال پیشنهاد می شوند که با استفاده از معیار کولبک لیبلر می‌توان تقریب بهتر راàز بین این دو مدل پیدا نمود.

مراجع

Azzalini, A., (1985). A Class of Distribution which Includes the Normal Ones. *Scandinavian Journal of Statistics*, **12**, 171-178.

Azzalini, A., (1986). Further Results on a Class of Distribution which Includes the Normal Ones. *Statistica*, **46**, 199-208.

Azzalini, A., Capitanio, A., (2003). Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew Distribution. *Journal of the Royal Statistical Society:* Series B, **65**,367-389.

Azzalini, A., Chiogna, M., (2004). Some New Results on the Steress-Strength Model for Skew Normal Variate. *Metron LXII*, 315-326.

Commenges. D., Sayyareh. A, Letenneure. L., Guedji. J. and Bar-Hen. A.(2008). *In The Annal of Applied Statistics.* Vol.**2**. No. 3, 1123-1142.

Archive of SID

www.SID.ir

Henze, N. A., (1986). A Probabilistic Representation of the Skew- Normal Distribution. *Scandinavian Journal of Statistics*, **13**, 271-275.

Kullback. S. (1986), *Information Theory and Statistics*. Dover, New York. MR 0866144.

Lin, J, G., Xie, F. C., Wei, B. C., (2009) Statistical Diagnostic for Skew-t-Normal

Wang, J., Boyer., Genton, M. G., (2004). A Skew Symmetric Representation of Multivariate Distributions.*Statistica Sinica*, **14**, 1259-1270

Archive of SID

www.SID.ir