The 45th Annual Iranian Mathematics Conference
August 26-29 2014

تحلیل درستنمایی تاوانیده برای توزیع چوله نرمال و چوله تی

طاهره الیاسی ، دانشجوی کارشناسی ارشد آمار ریاضی، دانشگاه سمنان، taherehelyasi@gmail.com

امید کریمی ، عضو هیأت علمی گروه آمار، دانشگاه سمنان، omid.karimi@profs.semnan.ac.ir

چکیده:
توزیع‌های چوله‌نرمال و چوله‌تی خانواده‌های پارامتری هستند که اخیرا به‌طور گسترده‌ای مورد بررسی قرار گرفته‌اند به‌طوری که خانواده کلی‌تری در مقایسه با توزیع‌های نرمال وt کلاسیک هستند که پارامتر چولگی در آنها مشخص شده است. مشکل اصلی این خانواده‌ها آن است که برآورد درستنمایی ماکسیمم ممکن است دچار واگرایی‌هایی به‌خاطر چولگی شود، که برای نمونه‌های محدود با احتمال قابل توجهی اتفاق می‌افتد، این مسأله اثرات ناخوشایندی روی تحلیل‌ها دارد. راه‌هایی برای غلبه یافتن بر این مشکل پیش روی هر دو رهیافت کلاسیک و بیزی گذاشته شده است. اما قابلیت اجرایی شدن آنها به وضعیت نمونه‌ها وابسته است. دراین مقاله بر اساس درستنمایی تاوانیده روشی ارائه می‌شود که با بعضی از روش‌های موجود معادل است و با این حال کاربردهای کلی‌تری در حالت چند متغیره دارد.

کلمات کلیدی: توزیع چوله نرمال، توزیع چوله‌تی، درستنمایی تاوانیده.

مقدمه

توزیع چوله‌نرمال متعلق به خانواده بزرگ نرمال است با یک پارامتراضافهα که چولگی را نشان می‌دهد. در حالت یک متغیره دارای تابع چگالی به‌صورت

$$f(x)=\frac{۲}{\omega}\phi(z)\Phi(\alpha z),\ \ z=(x-\xi)/w \qquad (۱)$$

است که درآن ξ پارامتر مکان، $(w\geq ۰)$ پارامتر مقیاس و α پارامتر چولگی است. همچنین $\phi(.)$ تابع چگالی و $\Phi(.)$ تابع توزیع نرمال استاندارد را نشان می‌دهند. برای نمایش توزیع چوله‌نرمال از نماد $SN(\xi,\omega^2,\alpha)$ استفاده می‌کنیم. اگر $\alpha=۰$، آنگاه $N(\xi,\omega^2)$ را خواهیم داشت و برای هر نمونه نمودار تابع لگاریتم درستنمایی برای α در یک نقطه دارای خمیدگی است و ماتریس‌اطلاع مورد انتظار ویژه است حتی اگر توزیع معلوم باشد. در این مقاله از پارامترهای مرکزی شده آزالینی (۱۹۸۵) استفاده می‌شود. حالت چند متغیره توزیع چوله‌نرمال توسط آزالینی و دالاواله (۲۰۱۳) مورد بررسی قرار گرفته است.

در این مقاله ابتدا تحلیل درستنمایی پارامتر چولگی و مشکلات موجود بیان می‌شود. سپس در بخش ۲ تحلیل درستنمایی تاوانیده برای توزیع چوله‌نرمال معرفی می‌گردد. توزیع چوله‌تی چند متغیره در بخش ۳ ارائه و برای آن تحلیل درستنمایی تاوانیده در بخش ۴ بیان می‌شود. در بخش ۵ یک مطالعه شبیه‌سازی انجام و نهایتا در بخش ۶ بحث و نتیجه‌گیری ارائه می‌شود.

۱ تحلیل درستنمایی پارامتر چولگی

در این بخش نشان داده می‌شود که با احتمال مثبت برآورد درستنمایی ماکسیمم (MLE) برای α واگرا است. این مسأله به آسانی در حالت یک پارامتری توزیع $SN(۰,۱,\alpha)$ مورد بررسی قرار می‌گیرد و لگاریتم درستنمایی روی نمونه $z = (z_۱,\ldots,z_n)$ به‌صورت $l(\alpha) = \sum_{i=۱}^{n} \zeta_۰(\alpha z_i)$ تعریف می‌شود. که در آن $\zeta_۰(x) = \log ۲\Phi(x)$ است. از آنجایی که $\zeta_۰$ یک تابع یکنواخت صعودی است، ماکسیمم l در $\alpha = \infty$ وقتی همه $z_i \geq ۰$ و وقتی $\alpha = -\infty$ همه $z_i \leq ۰$، اتفاق می‌افتد. (آزالینی و دالاواله، ۲۰۱۳). لذا وقتی تمام مقادیر نمونه هم‌علامت‌اند به یک برآوردMLE واگرا می‌رسیم که این موضوع نشان می‌دهد وقتی مشاهدات هم‌علامت نباشند MLE متناهی است . احتمال $P(z \leq ۰) = \frac{۱}{۲} + \pi^{-۱}\arctan(\alpha)$ وقتی $Z \sim SN(۰,۱,\alpha)$ را در نظر بگیرید، احتمال MLE واگرا به‌صورت زیر محاسبه می‌شود:

$$P_{n,\alpha} = (\frac{۱}{۲} - \frac{\arctan(\alpha)}{\pi})^n + (\frac{۱}{۲} + \frac{\arctan(\alpha)}{\pi})^n \quad (۲)$$

این احتمال وقتی $n \to \infty$ و $| \alpha | \leq \infty$ باشد، به صفر می‌رسد اما برای نمونه‌های کوچک و به‌خصوص حالتی که $\alpha \neq 0$ باشد، قابل چشم‌پوشی نیست. بنابراین در خانواده ۳ پارامتری به‌صورت $SN(\zeta,\omega^۲,\alpha)$ مقادیر نامتناهی برای MLE می‌تواند حاصل شود.

۲ تابع لگاریتم درستنمایی تاوانیده

تابع لگاریتم درستنمایی تاوانیده راهی است که در بعضی از مسائل برای اصلاح رفتارنامطلوب MLE به آن اشاره شده است، هدف رسیدن به برآوردگرهای واگرای α است. برای این منظور تابعی به شکل $l_p(\theta) = l(\theta) - Q$ در نظر بگیرید. که در آن $l(\theta)$ تابع لگاریتم درستنمایی و Q کمیت نامنفی است. کمیت $\tilde{\theta}$ که $l_p(\theta)$ را ماکسیمم می‌کند یک برآورد درستنمایی تاوانیده[۱] (MPLE) نامیده می‌شود. داریم: $Q \geq ۰$ $Q \mid_{\alpha=0} = 0$, , $lim_{\alpha_j \to \pm\infty} Q = \infty$ که در آن α_j،jامین مولفه α برای $j = ۱,\ldots,d$ است. برای توزیع چوله‌نرمال و چوله‌تی $logL$ به ∞ واگرا نیست حتی اگر برآورد ماکسیمم درستنمایی α واگرا باشد. در واقع با ترکیب مواردی که درباره Q در بالا اشاره شد می‌توانیم مطمئن باشیم که $l_p(\theta)$ یک ماکسیمم متناهی که در فضای پارامتر قرار بگیرد را ارائه می‌دهد.

۳ توزیع چوله‌تی چندمتغیره

تابع چگالی توزیع تی‌ـ استیودنت d بعدی $t_d(x;\Omega,\nu)$، با ν درجه آزادی و ماتریس مقیاس Ω با بعد $d \times d$ به‌صورت زیر تعریف می‌شود:

$$\begin{aligned} t_d(x;\Omega,\nu) &= \frac{\Gamma(\nu+d)/۲}{det(\Omega)^{۱/۲}(\pi\nu)^{d/۲}\Gamma(\nu/۲)} \\ &\times (۱ + \frac{x^T\Omega^{-۱}x}{\nu})^{-(\nu+d)/۲} \quad (۳) \end{aligned}$$

بنابراین تابع چگالی چوله‌تی، d بعدی به شرح زیر است:

$$\begin{aligned} &۲t_d(x-\zeta;\Omega,\nu)T\{\alpha^T\omega^{-۱}(x-\zeta) \\ &\times(\frac{\nu+d}{(x-\zeta)^T\Omega^{-۱}(x-\zeta)+d})^{۱/۲};\nu+d\} \quad (۴) \end{aligned}$$

که در آن α وζ بردارهای dبعدی و ω ماتریس قطری تشکیل شده توسط مربع ریشه‌ها که عناصر قطری Ω هستند، در این حالت پارامترهای مستقیم (DP)[۲] به‌صورت $\rho = (\zeta^T,\nu(\Omega)^T,\alpha^T,\nu)^T$ اگر Y یک متغیر تصادفی با تابع چگالی (۴) باشد می نویسیم:$Y \sim ST_d(\zeta,\Omega,\alpha,\nu)$ ماتریس واریانس به‌صورت زیر هستند:

$$\Sigma_Y = \nu_۲\Omega - \mu_۰\mu_۰^T, \quad \mu_Y = \zeta + \mu_۰, \quad (\nu \geq ۱)$$

[۱] Maximum Penalized Likelihood Estimation

[۲] Direct parameterization

که در آن $\nu_2 = \frac{\nu}{\nu-2}$، $\mu_0 = b_\nu \omega \delta$ و

$$\delta = \frac{1}{(1+\alpha^T \bar{\Omega}\alpha)^{1/2}}\bar{\Omega}\alpha, \quad \bar{\Omega} = \omega^{-1}\Omega\omega^{-1} \qquad (5)$$

است. برای تولید مجموعه پارامترهای مرکزی شده (CP)[3] در حالت چند متغیره μ, Σ عناصر طبیعی هستند،مولفه سوم CP شاخص چولگی را مانند حالت چند متغیره چوله‌نرمال مشخص می‌کند. بردار γ_1را توسط رابطه زیر:

$$\begin{aligned}
\gamma_1 &= \gamma_1(\delta,\nu) = \frac{\mu_3(\delta,\nu)}{\mu_2(\delta,\nu)^{3/2}}, \\
\mu_2(\delta,\nu) &= \mathrm{Var}\{Z_T\} = \frac{\nu}{\nu-2} - b_\nu^2\delta^2, \\
\mu_3(\delta,\nu) &= E\{[Z_T - \mu_1(\delta,\nu)]^3\} \\
&= b_\nu\delta[\frac{3\nu}{(\nu-3)(\nu-2)} - \delta^2(\frac{\nu}{\nu-3} - 2b_\nu^2])
\end{aligned} \qquad (6)$$

تعریف می‌کنیم. که در آن باید $\nu \geq 3$ باشد. طی مراحل زیر ρ را به $\bar{\rho}$ تبدیل می‌کنیم:

$$\theta = (\zeta^T, \nu(\Omega)^T, \eta^T, \nu)^T \rightarrow \phi = (\mu^T, \nu(\Sigma)^T, \mu_0^T, \nu)^T$$
$$\rho = (\zeta^T, \nu(\Omega)^T, \alpha^T, \nu)^T \rightarrow \bar{\rho} = (\mu^T, \nu(\Sigma)^T, \gamma_1^T, \gamma_2^M)$$

که در آن $\eta = \omega^{-1}\alpha$ پارامترهای میانه θ, ϕ اصلی نیستند و فقط تکنیک‌هایی برای دستیابی به نتیجه هستند. این محاسبات توسط آرنالو و آزالینی (۲۰۱۳) همانند توزیع چوله‌نرمال چند متغیره با یک مؤلفه اضافی در مجموعه پارامترها که رفتار دنباله رامشخص می‌کنند، انجام شده که این پارامتر اضافه در سه مجموعه اول بالا ν و در مجموعه آخری γ_2^M به‌صورت زیر است:

$$\gamma_2^M = E\{[(X-\mu)^T\Sigma^{-1}(X-\mu)]^2\} - d(d+2),$$

از $I(\bar{\rho}) = D_{\bar{\rho}}\rho^T I(\rho) D_{\bar{\rho}}\rho$ ماتریس ژاکوبین است که از تبدیل DP به CP به دست آمده و $D_{\bar{\rho}}\rho = D_\theta \rho D_\phi \theta D_{\bar{\rho}}\phi$ است.

[3] Centred parameterization
[4] Euler's Costant

۴ تعمیم درستنمایی تاوانیده به توزیع چوله‌تی

برای توزیع چوله‌تی در مورد خانواده سه پارامتری (ζ, ω, α) با ν معلوم معادلات زیر

$$M(\alpha) = -\frac{\alpha a_4(\alpha)}{2a_2(\alpha)}, \quad a_p(\alpha) = E\{Z^p \zeta_1(\alpha Z)^2\}$$

برای محاسبه Q ارائه شده است. آلوارز و گامرو(۲۰۱۲) نشان داده‌اند معادله $M(\alpha) = 0$ وقتی $\nu \geq 2$، یک جواب متناهی ارائه می‌دهد. در توزیع چوله‌تی مشابه توزیع SN به محاسبه یک تقریب مناسب ازتابع $M(\alpha)$ می‌پردازیم. آزالینی و آرلانو (۲۰۱۳) برای هر مقدار ثابت ν تقریب $\frac{\alpha}{-2M(\alpha)} \approx e_{1\nu} + e_{2\nu}\alpha^2$ را ارائه نمودند، که در آن ضرایب $e_{1\nu}, e_{2\nu}$ با استفاده از رفتار $\alpha^2 = 0, \alpha^2 \rightarrow \infty$ به‌دست می‌آیند که در نهایت به رابطه زیر می‌رسیم:

$$e_{2\nu} = g_\nu^2 \frac{E\{X_1^2 \zeta_1(X_1; \nu+1)\}}{E\{X_3^4 \zeta_1(X_3\sqrt{\frac{(\nu+1)}{\nu+3}}; \nu+1)\}}, e_{1\nu} = \frac{g_\nu}{3}$$

که در آن $g_\nu = \frac{(\nu+2)(\nu+3)}{(\nu+1)^2}$, $\zeta_1(x;\nu) = \frac{t(x;\nu)}{T(x;\nu)}$ $X_k \sim t_{\nu+k}$, است.

ثابت نباشد باید برای بهینه‌سازی عددی محاسبه کنیم

بعضی از نتایج عددی نشان می‌دهد که $log(\frac{e_{2\nu}}{e_2 - 1})$ خیلی به تابع خطی $log(\nu+\gamma)$ نزدیک است، که در آن $\gamma = 0/57721...$ مقدار ثابت اولر[4] است و e_2 به‌صورت زیر محاسبه می‌شود:

$$\begin{aligned}
e_2 &= lim_{\alpha^2 \rightarrow \infty}\{\frac{1+\alpha^2}{\alpha^2}\frac{E\{X^2\zeta_1(\delta X)\}}{E\{X^4\zeta_1(\delta X)\}} - \frac{e_1}{\alpha^2}\} \\
&= \frac{E\{X^2\zeta_1(\delta X)\}}{E\{X^4\zeta_1(\delta X)\}} \approx 0/854
\end{aligned}$$

که در آن $e_1 = \frac{a_2(0)}{a_4(0)} = \frac{E(X^2)}{E(X^4)} = \frac{1}{3}$ و تقریب $e_{2\nu}$ به‌صورت $e_{2\nu} \approx e_2(1 + \frac{4}{\nu+\gamma})$، به‌دست می‌آید. در

حالت چند متغیره مشابه توزیع چوله‌نرمال یک تابع لگاریتم درستنمایی تاوانیده را در نظر می‌گیریم که در آن $Q(\alpha) = c_1 log(1 + c_2\alpha^2)$ و ضرایب c_1, c_2 مقادیر مثبت و به ν وابسته‌اند. که آزالینی و دالاواله (۲۰۱۳) آنها را به‌صورت $c_1 = \frac{1}{(4e_{2\nu})}$، $c_2 = \frac{e_{2\nu}}{e_{1\nu}}$ محاسبه کرده‌اند. بنابراین با توجه به این ضرائب می‌توان برآورد درستنمایی تاوانیده را برای توزیع چوله‌تی محاسبه نمود.

۵ مطالعه شبیه‌سازی

در این بخش روش‌های پیشنهاد شده در یک مثال شبیه‌سازی‌های به کار می‌روند. داده‌ها از یک توزیع چوله‌نرمال $SN(\zeta, \omega^2, \alpha)$ شبیه‌سازی شده‌اند برای سهولت در مقایسه‌ها، مقادیر پارامتر به‌صورت: $\zeta = 0, \omega = 1, \alpha = 5, 10$ با اندازه‌های نمونه $n = 50, 200$ با تعداد تکرارهای 10^5 در نظر گرفته شده است. در همه حالت‌ها پارامترهای مقیاس و مکان $\zeta = 0, \omega = 1$ هستند. برآوردگرهای MLE کلاسیک و $MPLE$ با تابع تاوان:

$$Q(\alpha) = c_1 log(1 + c_2\alpha^2), Q'(\alpha) = 2c_1c_2\frac{\alpha}{1 + c_2\alpha^2}$$

و ضرایب $c_1 = \frac{1}{(4e_2)} \approx 0.875913, c_2 = \frac{e_2}{e_1} \approx 0.856250$، محاسبه شده اند. بررسی آنها اهمیت $\tilde{\alpha}$ را بر $\hat{\alpha}$ مشخص می‌کند دقت کنید که در اینجا $\tilde{\alpha}$ و $\hat{\alpha}$ اساسا منطبق نیستند. در عمل $|\hat{\alpha}| \geq 100$ در نظر گرفته‌ایم که به برآوردهی واگرا دلالت می‌کند، و احتمال این رویداد را بهتر از حالت یک پارامتری ارزیابی می‌کند؛ به عنوان مثال برای $n = 50, \alpha = 5$ این احتمال ۰/۱۳۹ است در حالی که در حالت یک پارامتری این احتمال فقط ۰/۰۳۹ است، اریبی و تغییرپذیری $\tilde{\xi}$ و$\tilde{\omega}$ آنها قدری بیشتر از MLE آنهاست.

نتایج

برای توزیع‌های چوله‌تی و چوله‌نرمال در حالت‌های یک متغیره و چند متغیره اصولی مطرح شد که از مسأله برآوردهای مرزی فضای پارامتر که می‌تواند برای برآورد ماکسیمم درستنمایی اتفاق بیفتد، اجتناب می‌کند. مسأله برآوردهای مرزی وقتی اندازه نمونه خیلی بزرگ باشد از بین می‌رود، اما در تکرار برای اندازه نمونه‌های کوچک و محدود با احتمال ناچیزی وجود دارد. این موضوع اثرات مخربی روی تحلیل پروژه‌ها دارد. در این مقاله روشی ارائه شده که به نتایج موجود نزدیک است اما تفاوت‌هایی هم وجود دارد، یکی از این تفاوت‌ها تغییر پیدا کردن از نااریبی بااستفاده از تابع تاوان Q است که می‌تواند آزادانه انتخاب شود و تا حدی قانع‌کننده باشد. نکته مفید دیگری که از فرمول‌بندی درستنمایی تاوانیده حاصل شد، سازگاری آن با اختیار پارامترهای مرکزی شده(CP) اشاره شد که به عنوان ابزاری برای غلبه یافتن بر ویژه بودن اطلاع توزیع چوله‌نرمال است.

مراجع

[1] Arellano-Valle,R.B., and Azzalini,A.(2013). The centred parameterization and related quantities of the skew-t distribution. Jour. of Multiva. Analy., 113: 73-90.

[2] Azzalini,A.,(1985). A class of distributions which includes the normal ones. Scandin. Jour. of Statis. 12,171-178.

[3] Azzalini,A. and Arellano-Valle,R.B.(2013). Maximum penalized likelihood estimation for skew-normal and skew-t distributions, Jour. of Statis. Plan. and Infer., 143: 419-433.

[4] lvarez,B., Jimenez Gamero,M.D (2012).A note on bias reduction of maximum likelihood estimates for the scalar Skew t distribution. Jour. of statis. plan. and Infer. 142(2),608-612.