تحلیل بیزی مدل‌های رگرسیونی عرض از مبدأ تصادفی با توزیع لاپلاس- چوله

راضیه محمدی و ایرج کاظمی[۱]

گروه آمار، دانشگاه اصفهان،

تاریخ دریافت: ۱۳۹۰/۸/۵ تاریخ پذیرش: ۱۳۹۰/۶/۱۲

چکیده: فرض متداول در برازش مدل‌های رگرسیونی عرض از مبدأ تصادفی، نرمال بودن توزیع مؤلفه‌های خطا و اثرهای تصادفی است. با توجه به این‌که غیرنرمال بودن این توزیع‌ها در کاربردهای تجربی امکان‌پذیر است مطالعه بر روی توزیع‌های منعطف‌تر از نرمال در سال‌های اخیر مورد توجه محققین قرارگرفته است. در این مقاله ما با در نظرگرفتن توزیع لاپلاس- چوله برای مؤلفه‌های خطا و اثرهای تصادفی، مدل رگرسیونی منعطفی را در برازش داده‌های وابسته توسط الگوریتم نمونه‌گیری گیبز که مبتنی بر رهیافت بیز است برای استنباط پارامترهای مدل به‌کار می‌بریم. بدین منظور با بهره‌گیری از نمایش سلسله مراتبی توزیع لاپلاس- چوله، توزیع‌های پسین شرطی کامل مورد نیاز الگوریتم را محاسبه می‌کنیم. در نهایت، با تحلیل مجموعه داده‌های تجربی در زمینه‌ی اقتصاد اهمیت مدل پیشنهادی را نشان می‌دهیم.

واژه‌های کلیدی: نمایش سلسله‌مراتبی، اثرهای تصادفی، چگالی پسین شرطی کامل، نمونه‌گیری گیبز، متغیر پنهان.

رده‌بندی ریاضی: ۶۲J۱۲ و ۶۲C۱۰

۱- مقدمه و مشخصات مدل

یکی از انواع مهم داده‌های وابسته در مطالعه‌های طولی جمع‌آوری می‌شود. این داده‌ها شامل مشاهداتی از واحدهای آزمایشی هستند که در دوره‌های زمانی مختلف مورد بررسی قرارگرفته‌اند. ویژگی عمده در تحلیل این داده‌ها، وجود وابستگی بین مشاهدات متوالی از هر

۱- آدرس الکترونیکی نویسنده مسئول مقاله: ایرج کاظمی i.kazemi@stat.ui.ac.ir

واحد است. علاوه بر این، یکسان نبودن رفتار واحدها نسبت به‌هم باعث ایجاد نوعی ناهمگنی خاص در هر دوره زمانی می‌شود. در مدل رگرسیونی عرض از مبدأ تصادفی که از مدل‌های معروف در تحلیل داده‌های وابسته است اثر خاص هر واحد به‌عنوان یک متغیر تصادفی فرض می‌شود. این عمل باعث لحاظ شدن تغییرپذیری بین و درون واحدهای آزمایشی در مدل‌سازی می‌شود که برآمد آن به افزایش دقت بیشتر برآوردها منجر خواهد شد [۱]. مزیت دیگر این مدل، امکان بررسی اثر ویژگی‌های خاص واحدها بر روی متغیر پاسخ است. انواع کاربردهای این مدل در تحقیقات مختلف، از جمله اسکرندال و رب-هسکچ [۲] و مک-کولا و سرل [۳] آمده است.

فرض کنید Y_{ij} متغیر پاسخ مرتبط با iامین مشاهده از فرد jام و $\boldsymbol{x}_{ij}$ بردار p بعدی شامل متغیرهای توضیحی باشد. مدل رگرسیونی عرض از مبدأ تصادفی به‌صورت

$$Y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + b_{\circ j} + \varepsilon_{ij}, \; i = ۱,\ldots,n; \; j = ۱,\ldots,k, \qquad (۱)$$

است که در آن β بردار p بعدی شامل ضرایب رگرسیونی، $b_{\circ j}$ اثر تصادفی واحد jام و ε_{ij} مؤلفه‌ی خطای تصادفی مدل است. به متغیر تصادفی $b_{\circ j}$ در مدل فوق، متغیر پنهان نیز گفته می‌شود. در تحلیل مدل (۱)، متغیرهای تصادفی ε_{ij} و $b_{\circ j}$ از یکدیگر مستقل فرض می‌شوند و نیز

$$\varepsilon_{ij} \overset{iid}{\sim} N\left(\circ, \sigma_e^۲\right), \quad b_{\circ j} \overset{iid}{\sim} N\left(\circ, \sigma_b^۲\right). \qquad (۲)$$

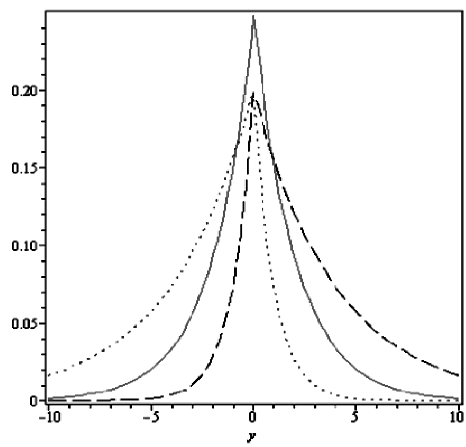
اما در تحلیل بعضی داده‌های واقعی با ساختارهای پیچیده، امکان برقراری توزیع‌های غیرنرمال برای اثرهای تصادفی و مؤلفه‌های خطا، دور از انتظار نمی‌باشد. در این مواقع، استفاده از مدل نرمال می‌تواند منجر به برآوردهایی غیرمنطقی از پارامترهای مدل گردد. یک راه‌کار پیشنهادی ازسوی محققان در مواجه شدن با این موضوع، جایگزین‌کردن توزیع‌های مقاوم است. این توزیع‌ها به‌دلیل داشتن چولگی، کشیدگی بیشتر و دم‌هایی کلفت‌تر از نرمال، می‌توانند داده‌هایی با ساختار پخش غیر نرمال را تحت پوشش قرار دهند. در این زمینه می‌توان به مطالعات مراجع [۴ و ۵] اشاره کرد که از توزیع نرمال- چوله و نرمال/ مستقل- چوله در برازش مدل‌های فوق استفاده کرده‌اند.

ما در این مقاله به‌جای توزیع نرمال در مدل‌های عرض از مبدأ تصادفی، لاپلاس- چوله را قرار داده و نشان می‌دهیم که این توزیع به‌لحاظ وجود چولگی، دم‌های کلفت‌تر و کشیدگی بیشتر، می‌تواند جایگزین مناسبی باشد. برای این‌منظور، در بخش دوم توزیع لاپلاس- چوله را معرفی و ویژگی‌های آن‌را بررسی می‌کنیم. در بخش سوم ابتدا مدل عرض از مبدأ تصادفی با توزیع

لاپلاس-چوله معرفی می‌شود و سپس استنباط بیزی پارامترها توسط روش نمونه‌گیری گیبز که مبتنی بر رهیافت مونت‌کارلوی زنجیر مارکفی است، انجام می‌شود. در بخش چهارم، مدل پیشنهادی را بر مجموعه‌ای از داده‌های واقعی برازش می‌دهیم. بخش پنجم شامل بحث و نتیجه‌گیری است.

۲- توزیع لاپلاس-چوله

یکی از توزیع‌های منعطف در تحلیل داده‌هایی که پخش داده‌ها در مرکز بیشتر است، لاپلاس-چوله است. در دهه‌ی اخیر مطالعات زیادی بر روی ویژگی‌های این توزیع صورت گرفته و شکل‌های متفاوتی برای تابع چگالی آن ارائه شده است (به‌عنوان مثال مراجع [۶، ۷، ۸ و ۹] را ببینید). ما در این مقاله توجه خود را به شکلی از این توزیع که توسط ارسلان [۱۰] ارائه شده‌است، معطوف می‌کنیم. اهمیت این توزیع از آن‌روست که تابع چگالی آن، برخلاف توزیع‌های لاپلاس- چوله‌ی دیگر، شکل صریحی دارد. این توزیع به‌صورت زیر تعریف می‌شود.



شکل ۱: تابع چگالی لاپلاس- چوله با $\mu=0$، $\sigma=2$ و $\gamma=0$ (خط ثابت)، $\gamma=-1/5$ (خط نقطه‌ای) و $\gamma=1/5$ (خط خط‌چین)

تعریف ۱. متغیر تصادفی Y دارای توزیع لاپلاس- چوله است هرگاه تابع چگالی آن برابر

$$f(y)=\frac{1}{2\sigma\alpha}\exp\left[\gamma\left(\frac{y-\mu}{\sigma^{2}}\right)-\alpha\left|\frac{y-\mu}{\sigma}\right|\right],\ -\infty<y<\infty, \qquad (3)$$

باشد که در آن $\mu\in\mathbb{R}$ پارامتر مکان، $\sigma^{۲}$ پارامتر مقیاس $(\sigma>۰)$ ، $\gamma\in\mathbb{R}$ پارامتر چولگی و $\alpha=\sqrt{۱+\frac{\gamma^{۲}}{\sigma^{۲}}}$ است. برای نمایش این توزیع از نماد $Y\sim SL\left(\mu,\sigma^{۲},\gamma\right)$ استفاده می‌کنیم. نمایش‌های مختلفی از این توزیع در شکل ۱ مشاهده می‌شود.

برخی از ویژگی‌های مهم توزیع لاپلاس- چوله با تابع چگالی (۳) به‌صورت زیر است:

۱- اگر $Y\sim SL\left(\mu,\sigma^{۲},\gamma\right)$ آن‌گاه امید ریاضی و واریانس آن عبارت است از

$$E\left(Y\right)=\mu+۲\gamma \tag{۴}$$

$$Var\left(Y\right)=۲\left(\sigma^{۲}+۲\gamma^{۲}\right) \tag{۵}$$

۲- اگر $Y\sim SL\left(\mu,\sigma^{۲},\gamma\right)$ آن‌گاه Y دارای نمایشی تصادفی به‌صورت

$$Y\overset{d}{=}\mu+\gamma W^{-۱}+\sigma\sqrt{W^{-۱}}Z \tag{۶}$$

است که در آن متغیر تصادفی Z دارای توزیع نرمال استاندارد و متغیر تصادفی W مستقل از Z و دارای توزیع وارون-گاما با تابع چگالی $f\left(w\right)=\frac{۱}{۲}w^{-۲}\exp\left(-\frac{۱}{۲}w^{-۱}\right)$ برای $w>۰$ می‌باشد و نماد $W\sim IGamma\left(۱,\ \frac{۱}{۲}\right)$ را به‌کار می‌بریم.

اثبات: با توجه به نمایش تصادفی (۶)، توزیع Y به شرط $W=w$، نرمال با میانگین $\mu+w^{-۱}\gamma$ و واریانس $\frac{\sigma^{۲}}{w}$ است. لذا تابع چگالی توام (Y, W) برابر با

$$f\left(y,w\right)=\frac{۱}{۲\sqrt{۲\pi\sigma^{۲}}}w^{-\frac{۳}{۲}}\exp\left[-\frac{۱}{۲}\left(sw+\alpha w^{-۱}\right)+\frac{\gamma}{\sigma^{۲}}\left(y-\mu\right)\right], \tag{۷}$$

است که در آن $w>۰$ ، $y\in\mathbb{R}$، $s=\left(\frac{y-\mu}{\sigma}\right)^{۲}$ و $\alpha=\sqrt{۱+\frac{\gamma^{۲}}{\sigma^{۲}}}$. با انتگرال‌گیری از این تابع نسبت به w و با استفاده از تعریف تابع بسل تعدیل‌شده‌ی نوع سوم به‌صورت

$$K_{-۱/۲}\left(\sqrt{\alpha^{۲}s}\right)=K_{۱/۲}\left(\sqrt{\alpha^{۲}s}\right)=\sqrt{\frac{\pi}{۲}}\left(\alpha^{۲}s\right)^{-۱/۴}e^{-\sqrt{\alpha^{۲}s}} \tag{۸}$$

تابع چگالی (۳) حاصل می‌شود.

۳- اگر $Y \sim SL(\mu,\sigma^{2},\gamma)$ و $X = cY + b$ که c و b اعداد حقیقی هستند، آن‌گاه

$$X \sim SL\left(c\mu + b, c^{2}\sigma^{2}, c\gamma\right). \quad (9)$$

حال با استفاده از این توزیع، یک مدل عرض از مبدأ تصادفی را که در آن مؤلفه‌های خطا و اثرهای تصادفی دارای توزیع لاپلاس- چوله هستند، معرفی کرده و نشان می‌دهیم که این مدل در برازش داده‌های واقعی می‌تواند عملکرد بهتری نسبت به فرض متداول نرمال داشته باشد.

۳- مدل رگرسیونی عرض از مبدأ تصادفی با خطاها و اثرات تصادفی لاپلاس- چوله

۳-۱ مشخصات مدل

مدل رگرسیونی عرض از مبدأ تصادفی (۱) را با این فرض که متغیرهای تصادفی ε_{ij} و b_{0j} به‌ازای همه‌ی مقادیر i و j از هم مستقل و دارای توزیع لاپلاس- چوله به‌صورت

$$\varepsilon_{ij} \overset{iid}{\sim} SL\left(0,\sigma_e^{2},\gamma_e\right), \quad b_{0j} \overset{iid}{\sim} SL\left(0,\sigma_b^{2},\gamma_b\right) \quad (10)$$

باشند، در نظر بگیرید. با توجه به ویژگی سوم در بخش قبل، از فرض (۱۰) نتیجه می‌گیریم که Y_{ij} به‌شرط b_{0j} نیز دارای توزیع لاپلاس- چوله با پارامترهای $\mathbf{x}'_{ij}\boldsymbol{\beta} + b_{0j}$، σ_e^{2} و γ_e است. امید ریاضی و واریانس حاشیه‌ای Y_{ij} برابر با

$$E\left(Y_{ij}\right) = \mathbf{x}'_{ij}\boldsymbol{\beta} + 2\left(\gamma_e + \gamma_b\right) \quad (11)$$

$$Var\left(Y_{ij}\right) = 2\left(\sigma_e^{2} + \sigma_b^{2}\right) + 4\left(\gamma_e^{2} + \gamma_b^{2}\right) \quad (12)$$

محاسبه می‌شوند. هم‌چنین

$$Cov\left(Y_{ij}, Y_{i'j}\right) = \operatorname{var}\left(b_{0j}\right) = 2\left(\sigma_b^{2} + 2\gamma_b^{2}\right). \quad (13)$$

بنابراین ضریب همبستگی درون- خوشه‌ای که بیانگر همبستگی بین دو واحد متفاوت i و i' $(i \neq i')$ در واحد jام است، را به‌صورت

$$\rho=\frac{\sigma_b^{2}+2\gamma_b^{2}}{\sigma_e^{2}+\sigma_b^{2}+2\left(\gamma_e^{2}+\gamma_b^{2}\right)} \qquad (14)$$

به‌دست می‌آوریم. این ضریب نسبتی از واریانس کل را نشان می‌دهد که مربوط به تغییرپذیری بین واحدها می‌شود. لذا مقدار آن همواره نامنفی است. عدم پذیرش فرض $H_0:\rho=0$، گواه بر وجود ساختار همبسته در داده‌ها می‌باشد.

۳-۲ تحلیل بیزی مدل با عرض از مبدأ تصادفی توسط توزیع لاپلاس- چوله

در آمار بیزی استنباط بر مبنای توزیع‌های پسین حاشیه‌ای انجام می‌گیرد. از آن‌جا که در مدل‌های پیچیده، محاسبه‌ی این توزیع‌ها به‌سادگی امکان‌پذیر نیست، رهیافت مونت‌کارلوی زنجیر مارکفی به‌کار گرفته می‌شود [۱۱]. در روش‌های مبتنی بر این رهیافت به‌جای شبیه‌سازی مستقیم از توزیع پسین حاشیه‌ای، از توزیع‌های دیگر به‌قسمی شبیه‌سازی می‌شود که توزیع زنجیر تولید شده به این توزیع پسین مانا باشد. یکی از این روش‌های معروف نمونه‌گیری گیبز است که با در نظرگرفتن مقادیر اولیه برای پارامترها، از توزیع‌های پسین شرطی کامل استفاده کرده و برآورد بیز پارامترها را طبق الگوریتم تکراری مشخصی محاسبه می‌کند [۱۲]. مزیت این روش نسبت به روش‌های دیگر مونت‌کارلوی زنجیر مارکفی این است که در هر مرحله از الگوریتم، نمونه‌ها از توزیع‌های شناخته‌شده تولید می‌شوند که این موجب سهولت فرآیند شبیه‌سازی است. اگر بعضی از این توزیع‌ها صریح و شناخته‌شده نباشند آن‌گاه الگوریتم‌های دیگری مانند متروپولیس-هستینگز درون الگوریتم گیبز جهت شبیه‌سازی به‌کار می‌رود. علاوه برآن، این توزیع‌های پسین شرطی کامل متناسب با توزیع پسین توام هستند و از این‌رو محاسبه‌ی آن‌ها با استفاده از توزیع پسین توام امکان‌پذیر است.

در این قسمت با استفاده از روش نمونه‌گیری گیبز، به برآورد پارامترهای مدل عرض از مبدأ تصادفی با توزیع لاپلاس- چوله می‌پردازیم. نظر به این‌که استفاده از نمایش سلسله‌مراتبی مدل، در آسان‌تر شدن محاسبات مربوط به نمونه‌گیری گیبز مفید است، لذا به‌جای استفاده‌ی مستقیم از تابع چگالی توزیع لاپلاس- چوله، شکل سلسله‌مراتبی آن را در نظر می‌گیریم. با-توجه به نمایش تصادفی توزیع لاپلاس- چوله در رابطه‌ی (۶)، متغیرهای تصادفی W_{eij} و W_{bj} برای $i=1,\dots,n$ و $j=1,\dots,k$ وجود دارند به‌قسمی که

$$Y_{ij}\left|W_{eij}=w_{eij},b_{0j}\right.\overset{ind}{\sim}N\left(\mathbf{x}_{ij}'\boldsymbol{\beta}+b_{0j}+w_{eij}^{-1}\gamma_e,\frac{\sigma_e^{2}}{w_{eij}}\right) \qquad (15)$$

$$W_{eij} \overset{iid}{\sim} IGamma\left(۱,\ \frac{۱}{۲}\right) \qquad (۱۶)$$

$$b_{\circ j} \left| W_{bj} = w_{bj} \overset{ind}{\sim} N\left(w_{bj}^{-۱}\gamma_b, w_{bj}^{-۱}\sigma_b^۲\right)\right. \qquad (۱۷)$$

$$W_{bj} \overset{iid}{\sim} IGamma\left(۱,\ \frac{۱}{۲}\right). \qquad (۱۸)$$

در این رابطه‌ها $IGamma(\theta,\lambda)$ معرف توزیع وارون-گاما با هسته‌ی چگالی $w^{-(\theta+۱)}\exp\left(-\frac{\lambda}{w}\right)$ است که در آن $w>\circ$، $\lambda>\circ$ و $\theta>\circ$.

با توجه به شکل سلسله‌مراتبی فوق، تابع درست‌نمایی توام مدل برابر با

$$L(\boldsymbol{\eta}|\mathbf{y}) = \prod_{j=۱}^{k}\prod_{i=۱}^{n} f\left(y_{ij}\left|w_{eij}, b_{\circ j}\right.\right) f\left(w_{eij}\right) f\left(b_{\circ j}\left|w_{bj}\right.\right) f\left(w_{bj}\right) \qquad (۱۹)$$

است که در آن $\boldsymbol{\eta} = \left(\boldsymbol{\beta}', \sigma_e^۲, \sigma_b^۲, \gamma_e, \gamma_b\right)'$ و بردار $\boldsymbol{y}$ معرف کل مشاهدات متغیر پاسخ است. با جایگذاری تابع‌های چگالی از رابطه‌های (۱۸)- (۱۵) در رابطه (۱۹) داریم

$$\begin{aligned} L(\boldsymbol{\eta}|\mathbf{y}) \propto & \prod_{j=۱}^{k}\prod_{i=۱}^{n} w_{eij}^{-۳/۲} w_{bj}^{-۳/۲} \left(\sigma_e^۲\right)^{-۱/۲} \left(\sigma_b^۲\right)^{-۱/۲} \\ & \times\exp\left[-\frac{w_{eij}}{۲\sigma_e^۲}\left(y_{ij} - \mathbf{x}'_{ij}\boldsymbol{\beta} - b_{\circ j} - w_{eij}^{-۱}\gamma_e\right)^۲ - \frac{۱}{۲w_{eij}}\right] \\ & \times\exp\left[-\frac{w_{bj}}{۲\sigma_b^۲}\left(b_{\circ j} - w_{bj}^{-۱}\gamma_b\right)^۲ - \frac{۱}{۲w_{bj}}\right]. \end{aligned} \qquad (۲۰)$$

حال باید توزیع‌های پیشین مناسبی به پارامترهای مدل تخصیص دهیم. با این فرض که پارامترها از یکدیگر مستقل هستند، توزیع‌های پیشین نیم-مزدوج شرطی زیر را در نظر می‌گیریم:

$$\boldsymbol{\beta} \sim N_p\left(\boldsymbol{\beta}_\circ, \boldsymbol{\Sigma}_\circ\right), \qquad (۲۱)$$

$$\sigma_e^۲ \sim IGamma\left(\frac{\tau_e}{۲}, \frac{T_e}{۲}\right), \qquad (۲۲)$$

$$\sigma_b^{۲} \sim IGamma\left(\frac{\tau_b}{۲},\frac{T_b}{۲}\right), \tag{۲۳}$$

$$\gamma_e \sim N\left(\mu_e,\delta_e^{۲}\right), \tag{۲۴}$$

$$\gamma_b \sim N\left(\mu_b,\delta_b^{۲}\right). \tag{۲۵}$$

توزیع پسین توام: با توجه به این‌که توزیع پسین توام متناسب با حاصل‌ضرب تابع درست-نمایی در توزیع پیشین توام پارامترهای مدل است، داریم:

$$\begin{aligned} f\left(\boldsymbol{\beta},\sigma_e^{۲},\sigma_b^{۲},\gamma_e,\gamma_b,\mathbf{b},\mathbf{w}_e,\mathbf{w}_b\middle|\mathbf{y}\right) \propto & \prod_{j=۱}^{k}\prod_{i=۱}^{n} w_{eij}^{-۳/۲} w_{bj}^{-۳/۲}\left(\sigma_e^{۲}\right)^{-۱/۲}\left(\sigma_b^{۲}\right)^{-۱/۲} \\ & \times\exp\left[-\frac{w_{eij}}{۲\sigma_e^{۲}}\left(y_{ij}-\mathbf{x}'_{ij}\boldsymbol{\beta}-b_{\circ j}-w_{eij}^{-۱}\gamma_e\right)^{۲}-\frac{۱}{۲w_{eij}}\right] \\ & \times\exp\left[-\frac{w_{bj}}{۲\sigma_b^{۲}}\left(b_{\circ j}-w_{bj}^{-۱}\gamma_b\right)^{۲}-\frac{۱}{۲w_{bj}}\right] \\ & \times\exp\left[-\frac{۱}{۲}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_\circ\right)'\boldsymbol{\Sigma}_\circ^{-۱}\left(\boldsymbol{\beta}-\boldsymbol{\beta}_\circ\right)\right] \\ & \times\left(\sigma_e^{۲}\right)^{-\left(\frac{\tau_e}{۲}+۱\right)}\left(\sigma_b^{۲}\right)^{-\left(\frac{\tau_b}{۲}+۱\right)}\exp\left(-\frac{T_e}{۲\sigma_e^{۲}}+-\frac{T_b}{۲\sigma_b^{۲}}\right) \\ & \times\exp\left[-\frac{۱}{۲\delta_e^{۲}}\left(\gamma_e-\mu_e\right)^{۲}-\frac{۱}{۲\delta_b^{۲}}\left(\gamma_b-\mu_b\right)^{۲}\right]. \end{aligned} \tag{۲۶}$$

که در آن $\mathbf{w}_e=\left[W_{e۱۱},...,W_{enk}\right]'$، $\mathbf{w}_b=\left[W_{b۱},..,W_{bk}\right]'$، $\mathbf{b}=\left[b_{\circ۱},..b_{\circ k}\right]'$ و $\mathbf{Y}=\left[\mathbf{Y}'_۱,..,\mathbf{Y}'_k\right]'$ با $\mathbf{Y}_j=\left[Y_{۱j},..,Y_{nj}\right]'$ است. از آن‌جا که الگوریتم نمونه‌گیر گیبز از چگالی‌های پسین شرطی کامل پارامترها جهت یافتن برآوردها استفاده می‌کند، لذا توجه خود را به محاسبه‌ی این توزیع‌ها معطوف می‌کنیم. نتایج حاصل از محاسبه‌ی این چگالی‌ها را در زیر آورده‌ایم.

توزیع‌های پسین شرطی کامل: مدل رگرسیونی (۱) را با فرض آن‌که اثرهای تصادفی و مؤلفه‌های خطا دارای توزیع لاپلاس-چوله باشند، در نظر می‌گیریم. با استفاده از چگالی پسین

توام (۲۶)، چگالی‌های پسین پارامترها و کمیت‌های تصادفی مرتبط به شرط بقیه پارامترها و کمیت‌ها به‌صورت زیر حاصل می‌شوند. ابتدا، با توجه به مستقل بودن دو به دو W_{eij} ها برای همه‌ی i و jها، تابع چگالی پسین شرطی کامل توام آن‌ها برابر با حاصل‌ضرب تابع چگالی پسین شرطی کامل هر یک از آن‌ها خواهد بود. با انجام چند محاسبه جبری چگالی پسین شرطی کامل W_{eij} برای هر i و j متناسب با

$$f\left(w_{eij}\middle|\boldsymbol{\beta},\sigma_e^{۲},\sigma_b^{۲},\gamma_e,\gamma_b,\mathbf{b},\mathbf{w}_{\backslash eij},\mathbf{w}_b,\mathbf{y}\right)\propto w_{eij}^{-۳/۲}\exp\left[-\frac{۱}{۲}\left(w_{eij}s_{۱ij}+\frac{\alpha_e^{۲}}{w_{eij}}\right)\right]$$

است که در آن $s_{۱ij}=\frac{۱}{\sigma_e^{۲}}\left(y_{ij}-\mathbf{x}'_{ij}\boldsymbol{\beta}-b_{۰j}\right)^{۲}$ و $\alpha_e=\sqrt{۱+\frac{\gamma_e^{۲}}{\sigma_e^{۲}}}$ و $\mathbf{w}_{\backslash eij}$ بردار $\boldsymbol{W}_e$ است که W_{eij} از آن حذف شده است. رابطه فوق نشانگر توزیع وارون- گاوسین با هسته‌ی چگالی $w^{-۳/۲}\exp\left[-\frac{۱}{۲}\left(aw+\frac{b}{w}\right)\right]$ با $a,b>۰$ است. به‌عبارت دیگر

$$W_{eij}\middle|\boldsymbol{\beta},\sigma_e^{۲},\sigma_b^{۲},\gamma_e,\gamma_b,\mathbf{b},\mathbf{w}_{\backslash eij},\mathbf{w}_b,y_{ij}\sim IGaussian\left(s_{۱ij},\alpha_e^{۲}\right). \qquad (۲۷)$$

به‌طور مشابه توزیع پسین شرطی کامل کمیت W_{bj} برای هر j وارون-گاوسی به‌صورت

$$W_{bj}\middle|\boldsymbol{\beta},\sigma_e^{۲},\sigma_b^{۲},\gamma_e,\gamma_b,\mathbf{b},\mathbf{w}_{\backslash bj},\mathbf{w}_e,y_{ij}\sim IGaussian\left(\frac{b_{۰j}^{۲}}{\sigma_b^{۲}},\alpha_b^{۲}\right) \qquad (۲۸)$$

است که در آن $\alpha_b=\sqrt{۱+\frac{\gamma_b^{۲}}{\sigma_b^{۲}}}$ و $\mathbf{w}_{\backslash bj}$ بردار $\boldsymbol{W}_b$ با حذف مؤلفه W_{bj} از آن است. هم‌چنین برای اثرهای تصادفی به‌دست می‌آوریم که

$$b_{۰j}\middle|\boldsymbol{\beta},\sigma_e^{۲},\sigma_b^{۲},\gamma_e,\gamma_b,\mathbf{b}_{\backslash j},\mathbf{w}_b,\mathbf{w}_e,y_{ij}\sim N\left(b_j^{*},\sigma_j^{۲*}\right) \qquad (۲۹)$$

که در آن $\mathbf{b}_{\backslash j}$ بردار $\mathbf{b}$ با حذف مؤلفه $b_{۰j}$ از آن است،

$$b_j^{*}=\sigma_j^{۲*}\left(\frac{۱}{\sigma_e^{۲}}\sum_{i=۱}^{n}s_{۲ij}+\frac{\gamma_b}{\sigma_b^{۲}}\right),\sigma_j^{۲*}=\left(\sigma_e^{-۲}\sum_{i=۱}^{n}w_{eij}+\sigma_b^{-۲}w_{bj}\right)^{-۱}$$

و $s_{2ij}=w_{eij}\left(y_{ij}-\mathbf{x}'_{ij}\boldsymbol{\beta}-w_{eij}^{-1}\gamma_e\right)$. برای یافتن توزیع پسین شرطی کامل بردار ضرایب رگرسیون $\boldsymbol{\beta}$ ابتدا ماتریس $\mathbf{X}=\left[\mathbf{X}'_1,\dots,\mathbf{X}'_k\right]'$ را که در آن $\mathbf{X}_j$ ماتریسی $(n\times p)$ است، تعریف می‌کنیم. با انجام چند محاسبه جبری داریم

$$\boldsymbol{\beta}\,\middle|\,\sigma_e^2,\sigma_b^2,\gamma_e,\gamma_b,\mathbf{b},\mathbf{w}_b,\mathbf{w}_e,\mathbf{y}\sim N_p\left(\boldsymbol{\beta}_1,\boldsymbol{\Sigma}_1\right) \qquad (30)$$

که درآن بردار میانگین آن به‌صورت

$$\boldsymbol{\beta}_1=\boldsymbol{\Sigma}_1\left[\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}-\mathbf{X}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{b}^*+\mathbf{w}_e^*\gamma_e\right)+\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\beta}_0\right]$$

و ماتریس واریانس-کوواریانس آن به‌صورت $\boldsymbol{\Sigma}_1=\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}+\boldsymbol{\Sigma}_0^{-1}\right)^{-1}$ است که $\mathbf{b}^*=\left[\mathbf{b}_{01}\mathbf{1}'_n,\dots,\mathbf{b}_{0k}\mathbf{1}'_n\right]'$، $\mathbf{w}_e^*=\left[\mathbf{w}_{e1}^{*\prime},\dots,\mathbf{w}_{ek}^{*\prime}\right]'$، $\mathbf{w}_{ej}^*=\left[w_{e1j}^{-1},\dots w_{enj}^{-1}\right]'$ و $\boldsymbol{\Sigma}$ یک ماتریس بلوکی قطری با مقادیر $\frac{\sigma_e^2}{w_{eij}}\mathbf{I}_n$ است. علاوه بر آن، توزیع شرطی کامل مؤلفه‌ی واریانس σ_e^2 وارون-گاما به‌صورت

$$\sigma_e^2\,\middle|\,\boldsymbol{\beta},\sigma_b^2,\gamma_e,\gamma_b,\mathbf{b},\mathbf{w}_b,\mathbf{w}_e,\mathbf{y}\sim IGamma\left(\frac{N+\tau_e}{2},\frac{T_e+s_3}{2}\right) \qquad (31)$$

است که در آن $N=kn$ و $s_3=\sum_{j=1}^{k}\sum_{i=1}^{n}w_{eij}\left(y_{ij}-\mathbf{x}'_{ij}\boldsymbol{\beta}-b_{0j}-w_{eij}^{-1}\gamma_e\right)^2$. هم‌چنین

$$\sigma_b^2\,\middle|\,\boldsymbol{\beta},\sigma_e^2,\gamma_e,\gamma_b,\mathbf{b},\mathbf{w}_b,\mathbf{w}_e,\mathbf{y}\sim IGamma\left(\frac{k+\tau_b}{2},\frac{T_b+s_4}{2}\right) \qquad (32)$$

که در آن $s_4=\sum_{j=1}^{k}w_{bj}\left(b_{0j}-w_{bj}^{-1}\gamma_b\right)^2$. برای پارامتر چولگی γ_e به‌دست می‌آوریم که

$$\gamma_e\,\middle|\,\boldsymbol{\beta},\sigma_e^2,\sigma_b^2,\gamma_b,\mathbf{b},\mathbf{w}_b,\mathbf{w}_e,\mathbf{y}\sim N\left(\mu_{e1},\delta_{e1}^2\right) \qquad (33)$$

که در آن $\mu_{e1}=\delta_{e1}^2\left(\frac{s_5}{\sigma_e^2}+\frac{\mu_e}{\delta_e^2}\right)$، $\delta_{e1}^2=\left(\sigma_e^{-2}\sum_{j=1}^{k}\sum_{i=1}^{n}w_{eij}^{-1}+\delta_e^{-2}\right)^{-1}$ و $s_5=\sum_{j=1}^{k}\sum_{i=1}^{n}\left(y_{ij}-\mathbf{x}'_{ij}\boldsymbol{\beta}-b_{0j}\right)$. هم‌چنین، به‌طور مشابه داریم:

$$\gamma_b \left| \boldsymbol{\beta}, \sigma_e^2, \sigma_b^2, \gamma_e, \mathbf{b}, \mathbf{w}_b, \mathbf{w}_e, \mathbf{y} \sim N\left(\mu_{b1}, \delta_{b1}^2\right) \right. \tag{۳۴}$$

که در آن $\mu_{b1} = \delta_{b1}^2 \left(\sigma_b^{-2} \sum_{j=1}^{k} b_{\circ j} + \delta_b^{-2} \right)$ و $\delta_{b1}^2 = \left(\sigma_b^{-2} \sum_{j=1}^{k} w_{bj}^{-1} + \delta_b^{-2} \right)^{-1}$.

الگوریتم نمونه‌گیر گیبز جهت محاسبه‌ی برآوردهای پسین،

$$\boldsymbol{\theta}^{(0)} = \left(\boldsymbol{\beta}^{(0)}, \sigma_e^{2(0)}, \sigma_b^{2(0)}, \gamma_e^{(0)}, \gamma_b^{(0)}, \mathbf{b}^{(0)}, \mathbf{w}_e^{(0)}, \mathbf{w}_b^{(0)} \right)$$

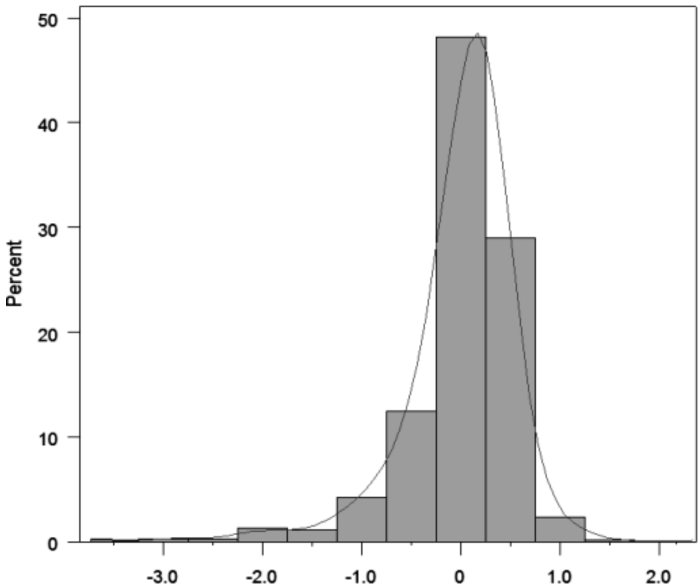
را به‌عنوان بردار مقدار اولیه‌ی پارامترها در نظر می‌گیرد. سپس برای $t=1,\dots,T$ که T بزرگ است، مقادیر $\boldsymbol{\theta}^{(1)},\dots,\boldsymbol{\theta}^{(T)}$ را از توزیع‌های پسین شرطی کامل (۲۷) تا (۳۴) شبیه‌سازی می‌کند تا مرحله همگرایی الگوریتم حاصل شود. با محاسبه‌ی میانگین مقادیر تولیدشده برای هر پارامتر، می‌توان برآورد بیز آن را محاسبه کرد.

۴- تحلیل داده‌های بدهی‌های مالیاتی

در این بخش با ارائه‌ی یک مثال تجربی، اهمیت مدل معرفی‌شده در بخش قبل را نشان می‌دهیم. داده‌ها شامل اطلاعات ۲۳۵ پرداخت‌کننده‌ی مالیات در ایالت متحده است که در طی سال‌های ۱۹۸۴- ۱۹۸۲ و ۱۹۸۷- ۱۹۸۶ جمع‌آوری شده‌اند. این داده‌ها به‌منظور بررسی اثر ویژگی‌های فردی و اقتصادی افراد بر روی میزان بدهی مالیاتی آن‌ها، در تحقیقات مختلف مورد مطالعه قرار گرفته‌اند. در مطالعات مختلف، فرض نرمال بودن برای توزیع مؤلفه‌های خطا و اثرهای تصادفی، در نظر گرفته شده‌است [۱ و ۱۳]. ما با استفاده از توزیع لاپلاس- چوله مدل‌های رگرسیونی متفاوتی را برازش می‌دهیم و نشان می‌دهیم که مدل پیشنهادی ما در تحلیل داده‌ها، عملکرد بهتری نسبت به مدل‌های قبلی دارد. مجموعه‌ی داده‌ها شامل هفت متغیر است. متغیر پاسخ، LNTAX، لگاریتم طبیعی بدهی‌های مالیاتی فرد است که بر پایه‌ی دلار سال ۱۹۸۳ ثبت شده‌است. متغیرهای توضیحی مرتبط با ویژگی‌های فردی عبارتند از: ۱- متغیر نشانگر MS برای متأهلین ۲- متغیر نشانگر HH برای سرپرست خانوارها ۳- متغیر نشانگر AGE برای افراد ۶۵ سال یا بیشتر از آن ۴- متغیر DEPEND که تعداد افراد تحت تکفل بیمه‌ای مالیات‌دهنده را نشان می‌دهد. متغیرهای توضیحی مربوط به ویژگی‌های اقتصادی عبارتند از: ۱- LNTPI که لگاریتم طبیعی مجموع درآمدهای خالص فرد را بر پایه‌ی دلار سال ۱۹۸۳ نشان می‌دهد ۲- MR که نرخ نهایی مالیات را نشان می‌دهد ۳- متغیر نشانگر EMP برای استفاده از تسهیلات تخفیفی توسط مالیات دهندگان ۴- PREP که میزان پرداخت‌های اولیه‌ی مالیات‌دهنده است. مدل را به‌صورت

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 \mathrm{LNTPI}_{ij} + \beta_2 \mathrm{MS}_{ij} + \beta_3 \mathrm{HH}_{ij} + \beta_4 \mathrm{DEPEND}_{ij} + \beta_5 \mathrm{AGE}_{ij} \\ &+ \beta_6 \mathrm{MR}_{ij} + \beta_7 \mathrm{EMP}_{ij} + \beta_8 \mathrm{PREP}_{ij} + b_{.j} + \varepsilon_{ij} \end{aligned} \tag{۳۵}$$

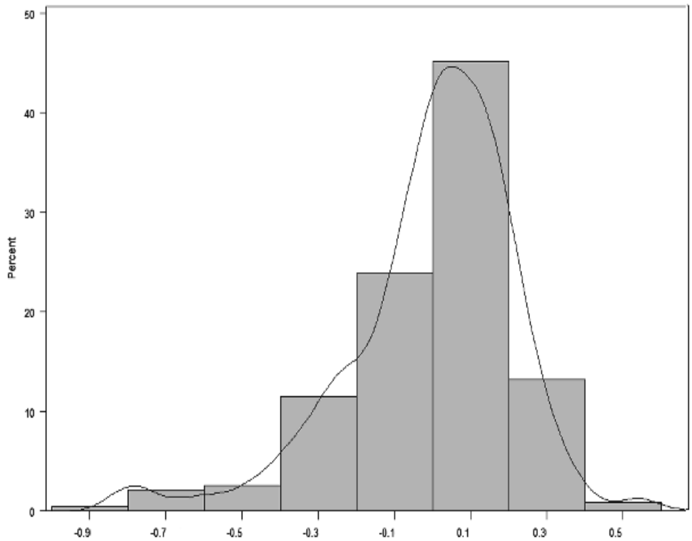
در نظر گیرید که در آن ۲۳۵,...,۱=j و i=۱ ,...,n_j و n_j تعداد مشاهدات هر فرد است. ابتدا مدل فوق را با فرض نرمال بودن خطاها و اثرهای تصادفی، بر روی داده‌ها برازش داده‌ایم. برآورد ضریب همبستگی درون خوشه‌ای برابر با ۰/۲۲۸ به‌دست آمده است و نشان می‌دهد که حدود ۲۳ درصد از کل تغییرپذیری مشاهدات از تغییرپذیری بین افراد ناشی می‌شود.



شکل ۲: بافت‌نگار مانده‌های سطح اول

جهت بررسی درستی فرض‌های این مدل، بافت‌نگار مانده‌ها را برای هر دوسطح (سطح اول شامل مؤلفه‌های خطا و سطح دوم شامل اثرهای تصادفی است) رسم کرده‌ایم که در شکل‌های ۲ و ۳ مشاهده می‌شوند. با توجه به این نمودارها، به‌نظر می‌رسد که مدل نرمال برازش مناسبی برای این داده‌ها نباشد. نتایج حاصل از آزمون نیکویی برازش مانده‌ها نیز فرض نرمال بودن آن‌ها را در سطح پنج درصد رد می‌کند. علاوه برآن، بافت‌نگارهای فوق که حاکی از وجود چولگی و کشیدگی بیشتر توزیع داده‌ها نسبت به نرمال هستند، شبیه به توابع چگالی توزیع لاپلاس- چوله می‌باشند. با این انگیزه به برازش مدل‌های مختلف براساس توزیع لاپلاس-چوله می‌پردازیم. مدل‌های برازش‌یافته عبارتند از:

مدل اول: خطاها و اثرهای تصادفی دارای توزیع نرمال.
مدل دوم: خطاها دارای توزیع لاپلاس-چوله و اثرهای تصادفی دارای توزیع نرمال.
مدل سوم: خطاها دارای توزیع نرمال و اثرهای تصادفی دارای توزیع لاپلاس- چوله.
مدل چهارم: خطاها و اثرهای تصادفی دارای توزیع لاپلاس- چوله.



شکل ۳: بافت‌نگار مانده‌های سطح دوم

برای برآورد مدل‌های فوق، از نرم‌افزار اُپن‌باگز[1] نسخه ۳.۲.۱ استفاده کرده و توزیع‌های پیشین $N\,(۰, ۱۰۰۰)$ را برای بردار ضرایب رگرسیونی، $IGamma$ (۰,۱/۱) را برای مؤلفه‌های واریانس و $N\,(۰, ۱۰۰۰)$ را برای پارامترهای چولگی منظور کرده‌ایم. در برازش مدل‌ها با توزیع‌های چوله، برای حل مسئله‌ی تشخیص پذیری عرض از مبدأ، یعنی پارامتر، در مدل‌های دوم، سوم و چهارم را به‌ترتیب برابر با ۲γ_e، ۲γ_b و ۲$(\gamma_e+\gamma_b)$ در نظر گرفتیم. این موضوع در اکثر تحقیقات انجام شده بر روی توزیع‌های چوله، منظور نشده است. با در نظر گرفتن تعداد ۱۰۰۰ دورریز و تولید نمونه‌ی ۳۰۰۰تایی با مقدار رقیق‌سازی ۳۰ برای هر مدل و هم‌چنین پس از اطمینان از همگرایی الگوریتم گیبز که توسط معیارها و نمودارهای مرتبط انجام شده است،

[1] Open Bugs

برآورد پارامترهای هر مدل را به‌دست آوردیم. جهت مقایسه‌ی مدل‌های فوق از معیار اطلاع آکائیک (AIC) استفاده می‌کنیم. مقادیر این معیار برای چهار مدل فوق، به‌ترتیب برابر با ۲۰۷۱/۶، ۱۴۹۴/۷، ۲۳۴/۴ و ۲۲۶/۲ به‌دست آمده‌است.

جدول ۱: برآورد بیز پارامترهای مدل عرض از مبدأ تصادفی چهار مدل (انحراف معیارهای بیزی در پرانتز گزارش شده‌اند)

	مدل اول	مدل دوم	مدل سوم	مدل چهارم
CONSTANT	-۰/۰۱۴ (۰/۴۴۲)	-۰/۱۹۴ (۰/۰۱۹)	-۰/۱۸۱ (۰/۰۴۸)	-۰/۲۳۴ (۰/۰۳۲)
LNTPI	۰/۶۵۲ (۰/۰۵۲)	۰/۶۹۹ (۰/۰۰۶)	۰/۶۶۷ (۰/۰۱۰)	۰/۷۱۱ (۰/۰۰۵)
MS	۰/۴۷. (۰/۰۷۵)	۰/۱۵۶ (۰/۰۴۰)	۰/۲۱۹ (۰/۰۶۸)	۰/۱۲۰ (۰/۰۳۷)
HH	۰/۰۸۱ (۰/۰۹۹)	۰/۰۲۷ (۰/۰۵۵)	۰/۱۱۶ (۰/۰۹۱)	۰/۰۴۸ (۰/۰۵۱)
DEPEND	-۰/۰۹۹ (۰/۰۲۴)	-۰/۰۹۱ (۰/۰۱۳)	-۰/۰۸۳ (۰/۰۲۳)	-۰/۰۷۸ (۰/۰۱۲)
AGE	-۰/۱۶۶ (۰/۰۷۷)	-۰/۱۲۰ (۰/۰۵۱)	-۰/۱۰۵ (۰/۰۷۹)	-۰/۰۵۷ (۰/۰۴۵)
MR	۰/۰۵۱ (۰/۰۰۴)	۰/۰۳۹ (۰/۰۰۲)	۰/۰۴۹ (۰/۰۰۳)	۰/۰۴۰ (۰/۰۰۱)
EMP	-۰/۲۱۴ (۰/۰۶۸)	-۰/۰۸۶ (۰/۰۴۳)	-۰/۱۵۶ (۰/۰۶۶)	-۰/۰۵۲ (۰/۰۴۱)
PREP	-۰/۰۳۱ (۰/۰۴۶)	-۰/۰۰۵ (۰/۰۲۵)	۱/۲e-۴ (۰/۰۴۴)	۰/۰۱۰ (۰/۰۲۳)
σ_e^2	۰/۳۰۶ (۰/۰۱۵)	۰/۰۷۱ (۰/۰۰۵)	۰/۲۹۳ (۰/۰۱۴)	۰/۰۶۸ (۰/۰۰۵)
σ_b^2	۰/۰۹۵ (۰/۰۱۶)	۰/۰۴۵ (۰/۰۰۸)	۰/۰۳۷ (۰/۰۰۹)	۰/۰۱۹ (۰/۰۰۵)
γ_e	-	-۰/۰۹۷ (۰/۰۱۰)	-	-۰/۰۸۴ (۰/۰۱۰)
γ_b	-	-	-۰/۰۹۱ (۰/۰۲۴)	-۰/۰۹۹ (۰/۰۱۷)

با توجه به این‌که مدل چهارم، یعنی مدلی که در آن خطاها و اثرهای تصادفی دارای توزیع لاپلاس- چوله هستند، کمترین مقدار AIC را دارد، لذا می‌توان نتیجه گرفت که این مدل نسبت به سایر مدل‌های فوق، برازش بهتری برای داده‌ها می‌باشد. برآورد پارامترهای چهار مدل در جدول ۱ گزارش شده است. مقایسه‌ی نتایج جدول ۱ با نتایج حاصل از مدل نرمال، نشان می‌دهد که بعضی از برآوردها و انحراف استاندارد آن‌ها تغییرات زیادی از لحاظ اندازه و جهت مثبت یا منفی بودن اثر متغیرهای توضیحی موردنظر دارند. به‌خصوص، ضریب همبستگی درون- خوشه‌ای با مدل چهارم (معادله ۱۴) برابر با ۰/۹۹۸ برآورد می‌شود که تفاوت محسوسی

با برآورد آن در مدل نرمال دارد. با توجه به آن‌که معیارهای انتخاب مدل، توزیع نرمال را جهت تحلیل داده‌های فوق نامناسب تشخیص می‌دهد، لذا نتایج مدل نرمال که به‌طورمتداول توسط محققان قبلی گزارش شده است، بایستی با احتیاط در نظر گرفته شود.

۵- بحث و نتیجه‌گیری

در این مقاله مدل رگرسیونی عرض از مبدأ تصادفی با توزیع لاپلاس- چوله به‌عنوان جایگزینی برای مدل نرمال که به‌صورت متداول در تحلیل داده‌های وابسته به‌کار می‌رود، معرفی گردید. این مدل به‌دلیل وجود چولگی، کشیدگی زیاد و دم‌های کلفت‌تر، از انعطاف‌پذیری بیشتری برخوردار بوده و می‌تواند در برازش داده‌هایی با توزیع غیر نرمال عملکرد مناسبی داشته باشد. از سوی دیگر شکل ساده‌ی تابع چگالی به‌کار برده شده در این مقاله، انجام محاسبات بیزی را آسان می‌کند که این امر با در نظر گرفتن مدل‌های جایگزین معرفی شده توسط محققان قبلی، در خور توجه است. در مثال ارائه‌شده در این مقاله، با توجه به عدم تقارن مانده‌ها و نیز تجمع بیشتر داده‌ها در مرکز و هم‌چنین آزمون نیکویی برازش، از توزیع لاپلاس- چوله به‌عنوان یک جایگزین استفاده گردید که در نهایت بنابر معیار انتخاب مدل، مدلی که در آن خطاها دارای توزیع لاپلاس و اثرهای تصادفی دارای توزیع لاپلاس-چوله هستند، به‌عنوان مدل مناسب پیشنهاد شد. قابل توجه است که با در نظر گرفتن سایر توزیع‌های مقاوم، ممکن است مدل‌های جدید دیگر در مقایسه با مدل پیشنهادی در این مقاله، عملکرد بهتری داشته باشند که از اهداف این مقاله خارج است.

تشکر و قدردانی

نویسندگان مقاله از داوران محترم، به‌خاطر ارائه نظرها وپیشنهادهای به‌جا و مناسب کمال تشکر را دارند. هم‌چنین از مسئول محترم تحصیلات تکمیلی و پژوهشی دانشگاه اصفهان تشکر می‌شود.

مراجع

[1] Frees, E.W. (2004), *Longitudinal and Panel Data, Analysis and Applications in the Social Sciences*, Cambridge University Press, New York.

[2] Skrondal, A. and Rabe-Heskheth, S. (2004), *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC.

[3] McCulloch, C.E., Searle, S.R. and Neuhaus, J.M. (2008), *Generalized, Linear, and Mixed Models*, John Wiley & Sons, Inc.

[4] Arellano-Valle, R.B., Bolfarine, H. and Lachos, V.H. (2007), Bayesian inference for skew-normal linear mixed models, *Journal of Applied Statistics*, **34**(6), 663-682.

[5] Lachos, V.H., Ghosh, P. and Arellano-Valle, R.B. (2010), Likelihood based inference for skew-normal independent linear mixed models, *Statistica Sinica*, **20**, 303-322.

[6] Kotz, S., Kozubowski, T.J. and Podgorski, K. (2001), *The Laplace Distribution and Generalizations,* Birkhauser, Boston.

[7] Kozubowski, T.J. and Podgorski, K. (2000a), A multivariate and asymmetric generalization of Laplace distribution, *Comput Stat*, **15**, 531-540.

[8] Kozubowski, T.J. and Podgorski, K. (2000b), Asymmetric Laplace distribution, *Math Sci*, **25**, 37-46.

[9] Kozubowski, T.J. and Nadarajah, S. (2010), Multitude Laplace distributions, *Stat Papers*, **51**, 127-148.

[10] Arslan, O. (2009), An alternative multivariate skew Laplace distribution: Properties and estimation, *Statistical Papers*, **49**(1), 1-23.

[11] Gelfand, A. and Smith, A. (1990), Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398-409.

[12] Gelfand, A. (2000), Gibbs sampling. *Journal of the American Statistical Association*, **95**(452), 1300-1304.

[13] Rabe-Heskheth, S. and Skrondal, A. (2008), *Multilevel and Longitudinal Modeling Using Stata*, 2nd, StataCorp LP.

Bayesian Analysis of Random-Intercept Models with the Skew-Laplace Distribution

Razieh Mohammadi, Iraj Kazemi

Department of Statistics, University of Isfahan, Isfahan, Iran

Abstract

In fitting random-intercept models, it is commonly assumed that the random effects and the error terms follow the normal distribution. In many empirical applications, the true distribution of data obeys non-normality and thus the main concern of most recent studies is the use of alternative distributions. In this paper, we propose a new class of random-intercept models using the Skew-Laplace distribution. The new regression model is flexible in the analysis of correlated data and simple in the implementation of Markov Chain Monte Carlo methods, such as the Gibbs sampling approach. Using the stochastic representation of the Skew-Laplace distribution we derive the full conditional posteriors distributions in order to present the Bayesian inference of model parameters. A real data analysis is illustrated from the economic contexts to show the usefulness of the proposed model.

Keywords: Full conditional posterior density, Gibbs sampling, Hierarchical representation, Latent variable, Random effect.

Mathematics Subject Classification (2000): 62J12, 62C10