برآورد مدل‌های آمیختۀ خطی تعمیم‌یافتۀ فضایی با متغیرهای پنهان چوله نرمال بسته

فاطمه حسینی: دانشگاه سمنان، گروه آمار
محسن محمدزاده*: دانشگاه تربیت مدرس، گروه آمار

چکیده

مدل‌های آمیختۀ خطی تعمیم‌یافتۀ فضایی معمولا برای مدل‌بندی پاسخ‌های فضایی گسسته به‌کار می‌روند، که در آن‌ها ساختار همبستگی فضایی داده‌ها از طریق متغیرهای پنهان در نظر گرفته می‌شود. مسئلۀ مهم در این مدل‌ها، برآورد متغیرهای پنهان فضایی در موقعیت‌های دارای مشاهده پاسخ و پارامترهای مدل و در نهایت پیش‌گویی متغیرهای پنهان در موقعیت‌های فاقد مشاهده است. در این راستا اغلب کاربران برای سهولت، توزیع نرمال را برای متغیرهای پنهان در نظر می‌گیرند. اگرچه این فرض باعث سهولت محاسبات می‌شود، اما گاهی در عمل واقع‌گرایانه نیست، یا به‌دلیل پنهان بودن بررسی آن میسر نیست. لذا در این مقاله استفاده از توزیع چوله‌نرمال بسته که در حالت خاص شامل توزیع نرمال است و تحت حاشیه‌سازی، شرطی کردن و تبدیلات خطی بسته است، برای متغیرهای پنهان پیشنهاد می‌شود. در این مدل‌ها تابع درست‌نمایی فرم بسته‌ای ندارد و به‌دست آوردن برآوردهای ماکسیمم درست‌نمایی پارامترها به راحتی امکان‌پذیر نیست، لذا الگوریتمی تقریبی برای برآورد ماکسیمم درست‌نمایی پارامترهای مدل و پیش‌گوی تقریبی متغیرهای پنهان ارائه می‌شود، که در مقایسه با روش‌های موجود بسیار سریع‌تر است. اعتبار مدل و الگوریتم پیشنهادی در یك مطالعه شبیه‌سازی بررسی می‌شود.

مقدمه

مدل آمیخته خطی تعمیم‌یافتۀ فضایی[1] (SGLM) معمولاً برای مدل‌بندی داده‌های فضایی گسسته روی یك ناحیه پیوسته فضایی به‌کار گرفته می‌شود. اولین بار برسلو[2] و کلایتون[3] از این مدل در بررسی‌های پزشکی استفاده کردند. مسئله‌ای مهم در مدل‌های پیش‌گویی متغیرهای پنهان در موقعیت‌های فاقد مشاهده است. این امر مستلزم برآورد پارامترهای مدل و متغیرهای پنهان است. حضور متغیرهای پنهان در این مدل‌ها و عدم وجود شکل بسته برای توزیع شرطی متغیرهای پنهان به‌شرط متغیرهای پاسخ، برآورد حداکثر درست‌نمایی پارامترهای مدل را

واژه‌های کلیدی: مدل آمیختۀ خطی تعمیم‌یافته فضایی، متغیر پنهان، توزیع چولۀ نرمال بسته
دریافت ۹۰/۵/۲۳ پذیرش ۹۱/٤/٦

*نویسنده مسئول mohsen_m@modares.ac.ir

۱. Spatial Generalized Linear Mixed Model ۲. Breslow ۳. Clayton

دشوار و گاهی غیرممکن می‌سازد. دیگل[1] و همکاران [۴] با رهیافت بیزی و مینیمم کردن میانگین توان دوم خطاها، پیش‌گوی بهینه را برای متغیرهای پنهان به‌دست آوردند. زانگ [۸] با ترکیب روش مونت‌کارلو و الگوریتم گرادیانت بیشینه‌سازی امید ریاضی (EMG)، الگوریتم جدید گرادیانت بیشینه‌سازی امید ریاضی مونت کارلویی[2] (MCEMG) را برای برآورد ماکسیمم درست‌نمایی پارامترهای هم‌بستگی و پیش‌گویی متغیرهای پنهان نرمال در مدل SGLM ارائه کرد. کریستین‌سن [۲] با روش ماکسیمم درست‌نمایی و الگوریتم مونت کارلو، پارامترها و پیش‌گوی بهینه را در این مدل‌های همراه با فرض نرمال بودن متغیرهای پنهان به‌دست آورد. اما در عمل، به‌دلیل غیرقابل مشاهده بودن متغیرهای پنهان در مدل‌های SGLM، بررسی نرمال بودن آن‌ها مقدور نیست و پذیرش ناصحیح این فرض می‌تواند بر دقت برآورد پارمترها و پیش‌گوها تأثیر سو داشته باشد. دامین‌گوئز توزیع چولهٔ نرمال بسته[3] (CSN) معرفی کرد[۳]. این خانواده از توزیع‌ها از خانوادهٔ توزیع نرمال انعطاف‌پذیرتر و تحت حاشیه‌سازی، شرطی کردن و تبدیلات خطی بسته هستند. محمدزاده وحسینی الگوریتم MCEMG را برای مدل‌های با متغیرهای پنهان فضایی چوله‌نرمال بسته تعمیم دادند [۷]. حسینی و همکاران توزیع متغیرهای پنهان به شرط متغیرهای گسسته پاسخ را به‌طور تقریبی به دست آوردند [۵]. در این مقاله با استفاده از این توزیع تقریبی و الگوریتم EMG، الگوریتم جدید گرادیانت بیشینه‌سازی امید ریاضی تقریبی[4] (AEMG) برای برآورد ماکسیمم درست‌نمایی پارامترهای مدل با متغیرهای پنهان چولهٔ نرمال بسته معرفی می‌شود. سپس پیش‌گوی تقریبی مینیمم میانگین توان دوم خطا[5] (MMSE) برای متغیرهای پنهان ارائه می‌گردد. در انتها اعتبار روش‌ها و الگوریتم ارائه شده در یك بررسی شبیه‌سازی ارزیابی می‌شود.

مدل SGLM با متغیرهای پنهان چوله‌نرمال بسته

برای تعریف مدل SGLM ابتدا باید یك مدل درست‌نمایی برای مشاهدات فضایی گسسته ($\mathbf{y}$) و سپس یك توزیع برای متغیرهای پنهان فضایی ($\mathbf{x}$) در نظر گرفته شوند. توزیع $\mathbf{y}$ متعلق به خانوادهٔ نمایی [٦] و توزیع $\mathbf{x}$ را چولهٔ نرمال بسته در نظر می‌گیریم. یعنی بردار تصادفی n بعدی $\mathbf{x}$ دارای تابع چگالی

$$f_{n,q}(x,\mu,\Sigma,D,\nu,\Delta)=\Phi_q^{-1}(0;\nu,\Delta+D\Sigma D')\phi_n(x;\mu,\Sigma)\Phi_q(D(x-\mu);\nu,\Delta) \qquad (۱)$$

است، که در آن $\boldsymbol{\mu}$ بردار پارامتر مکان، Σماتریس $n\times n$ معین مثبت مقیاس، D ماتریس $q\times n$ چولگی، $\phi_n(\cdot;\boldsymbol{\mu},\Sigma)$ تابع چگالی نرمال nمتغیره با میانگین $\boldsymbol{\mu}$ و ماتریس کواریانس Σو $\Phi_q(\cdot,\mathbf{v},\Delta)$ تابع توزیع تجمعی نرمال qمتغیره با میانگین $\boldsymbol{\mu}$ و ماتریس کواریانس Δ است. متغیر $\mathbf{x}$ که دارای تابع چگالی (1) است، به‌صورت $\mathbf{x}\sim CSN_{n,q}(\boldsymbol{\mu},\Sigma,D,\mathbf{v},\Delta)$ نمایش داده می‌شود. بدیهی است اگر D ماتریس صفر باشد، تابع چگالی فوق تبدیل به‌توزیع نرمال n متغیره خواهد شد. همچنین اگر $q=1$، $\mathbf{v}=\mathbf{0}$، $\Delta=1$ و $D=\boldsymbol{\lambda}'\Sigma^{-\frac{1}{2}}$ آن‌گاه چگالی فوق،

[1]. Diggle [2]. Monte Carlo Expectation Maximization Gradiant [3].Closed skew Normal
[4]. Approximate Expectation Maximization Gradiant [5]. Minimum Meam Square Error

فوق، چگالی توزیع چوله نرمال می‌شود. میانگین توزیع چوله‌نرمال بسته به‌صورت

$$E(X) = \boldsymbol{\mu} + \Sigma D'\boldsymbol{\psi} \tag{۲}$$

است [۳]، که در آن $\boldsymbol{\psi} = \dfrac{\Phi_q^*(\mathbf{0};\mathbf{v},\Delta + D\Sigma D')}{\Phi_q(\mathbf{0};\mathbf{v},\Delta + D\Sigma D')}$ و برای هر ماتریس معین مثبت Ω، $\Phi_q^*(\mathbf{r};\mathbf{v},\Omega) = \left[\nabla_r \Phi_q(\mathbf{r};\mathbf{v},\Omega)\right]'$ و $\nabla_r = (\dfrac{\partial}{\partial r_1},\ldots,\dfrac{\partial}{\partial r_q})'$ است.

فرض کنید $\mathbf{x} = (x_1,\ldots,x_n)'$ بردار متغیرهای پنهان فضایی در n موقعیت $\{s_1,\ldots,s_n\}$ با چگالی $\mathbf{x} \mid \boldsymbol{\eta} \sim CSN_{n,1}(H\boldsymbol{\beta},\Sigma_\theta,\lambda'\Sigma_\theta^{-\frac{1}{2}},0,1)$ باشد، که در آن $\boldsymbol{\eta} = (\boldsymbol{\beta}',\boldsymbol{\theta},\boldsymbol{\lambda})'$ پارامترهای مدل، H ماتریس $n \times (p+1)$ متغیرهای تبیینی، $\boldsymbol{\beta} = (\beta_0,\ldots,\beta_p)'$ بردار پارامترهای رگرسیونی، $\boldsymbol{\lambda}$ بردار پارامترهای چولگی و $\boldsymbol{\theta}$ بردار پارامترهای همبستگی فضایی مدل هستند. بنا بر این با توجه به (۱) چگالی متغیرهای پنهان به‌صورت

$$f(\mathbf{x} \mid \boldsymbol{\eta}) = \frac{2}{(2\pi)^{n/2} \mid \Sigma_\theta \mid^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - H\boldsymbol{\beta})'\Sigma_\theta^{-1}(\mathbf{x} - H\boldsymbol{\beta})\right\} \cdot \Phi_q\left(\boldsymbol{\lambda}'\Sigma^{-\frac{1}{2}}(\mathbf{x} - H\boldsymbol{\beta})\right), \tag{۳}$$

می‌شود. اکنون فرض کنید مشاهدات در موقعیت‌های فضایی $\{s_1,\ldots,s_k\}$ در اختیار باشند و هدف پیش‌گویی متغیرهای پنهان در موقعیت‌های فاقد مشاهده $\{s_{k+1},\ldots,s_n\}$ باشد. متغیرهای پنهان در k موقعیت مشاهده شده به صورت $\mathbf{x}^{obs} = A\mathbf{x}$ نمایش داده می‌شود که در آن $A = [I_{k\times k} \mid 0_{k\times n-k}]$ است. در این صورت بردار $\mathbf{x}$ را می‌توان به‌صورت $\mathbf{x} = (\mathbf{x}^{obs'},\mathbf{x}^{pred'})'$ تجزیه کرد، که در آن $\mathbf{x}^{pred}$ بردار متغیرهای پنهان در $n-k$ موقعیت انتخاب شده برای پیش‌گویی است. همچنین $\mathbf{y} = (y_1,\ldots,y_k)'$ بردار متغیرهای پاسخ فضایی گسسته در موقعیت‌های دارای مشاهده $\{s_1,\ldots,s_k\}$ است، با فرض استقلال شرطی این متغیرها روی متغیرهای پنهان، $\pi(\mathbf{y} \mid \mathbf{x}^{obs})$ عضو خانوادهٔ نمایی با تابع چگالی $\pi(y_i \mid x_i) = \exp\{y_i x_i - b(x_i) + c(y_i)\}$، $i = 1,\ldots,k$ است ([6]). بنا بر این مدل به‌صورت $E(y_i \mid x_i) = g^{-1}(x_i)$ است و به‌طور خلاصه مؤلفه‌های مدل SGLM بدین‌صورت خواهد شد:

$$\begin{aligned}\pi(\mathbf{y},\mathbf{x} \mid \boldsymbol{\eta}) = \pi(\mathbf{y} \mid \mathbf{x})\pi(\mathbf{x} \mid \boldsymbol{\eta}) \propto \mid \Sigma_\theta \mid^{-1/2} \exp\left\{\sum_{i=1}^{n}[y_i x_i - b(x_i) + c(y_i)] - \frac{1}{2}(\mathbf{x} - H\boldsymbol{\beta})'\Sigma_\theta^{-1}(\mathbf{x} - H\boldsymbol{\beta})\right\} \\ \times \Phi\left(\boldsymbol{\lambda}'\Sigma^{-\frac{1}{2}}(\mathbf{x} - H\boldsymbol{\beta})\right)\end{aligned} \tag{۴}$$

پیش‌گوی تقریبی در مدل‌های SGLM

با فرض معلوم بودن پارامترهای مدل پیش‌گوی مینیمم میانگین مربع خطای (MMSE) متغیرهای پنهان در موقعیت jام به‌صورت $E(x_j \mid \mathbf{y})$ به‌دست می‌آید که برای محاسبهٔ آن نیاز به توزیع کناری $\pi(x_j \mid \mathbf{y})$ است. چون توزیع $\pi(\mathbf{x} \mid \mathbf{y}) \propto \pi(\mathbf{y} \mid \mathbf{x})\pi(\mathbf{x})$، فرم بسته‌ای ندارد، نمی‌توان $E(x_j \mid \mathbf{y})$ را مستقیماً محاسبه کرد و به

طور تقریبی به‌دست آورده می‌شود. حسینی و همکاران [۵] نشان دادند اگر $\mathbf{x}$ دارای توزیع چولۀ نرمال بسته و توزیع $\mathbf{y}$ عضو خانوادۀ نمایی باشد، با خطی کردن $\pi(\mathbf{y}\,|\,\mathbf{x})$ حول یك مقدار ثابت $\mathbf{x}^0$، توزیع $\pi(\mathbf{x}\,|\,\mathbf{y})$ را می‌توان با توزیع چولۀ نرمال بسته به‌صورت

$$\hat{\pi}(\mathbf{x}\,|\,\mathbf{y}) \approx CSN_{n,1}(\hat{\boldsymbol{\mu}}_{x|y,\eta}(\mathbf{y},\mathbf{x}^0), \hat{\Sigma}_{x|y,\eta}(\mathbf{x}^0), \hat{D}_{x|y,\eta}, \hat{\mathbf{v}}_{x|y,\eta}, 1) \qquad (۵)$$

تقریب زد، که در آن $\hat{\boldsymbol{\mu}}_{x|y,\eta}(\mathbf{y},\mathbf{x}^0) = H\boldsymbol{\beta} + \Sigma_\theta A' R^{-1}(z(\mathbf{y},\mathbf{x}^{obs}) - AH\boldsymbol{\beta})$، $R = A\Sigma_\theta A' + P$، P ماتریسی قطری با درایه‌های $P(i,i) = 1/b''(x_i),\ i=1,\ldots,k$،

$$\ln\pi(\mathbf{x}\,|\,\boldsymbol{\eta}) = -\frac{1}{2}\ln|\Sigma_\theta| - \frac{1}{2}(\mathbf{x}-H\boldsymbol{\beta})'\Sigma_\theta^{-1}(\mathbf{x}-H\boldsymbol{\beta}) + \ln\{2\Phi(\boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(\mathbf{x}-H\boldsymbol{\beta}))\},$$

$$z_i(y_i, x_i^0) = [y_i - b'(x_i^0) + x_i b''(x_i^0)]/b''(x_i^0)$$

$\hat{\Sigma}_{x|y,\eta}(\mathbf{x}^0) = \Sigma_\theta - \Sigma_\theta A' R^{-1} A \Sigma_\theta$، $\hat{D}_{x|y,\eta} = \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}$ و $\hat{\mathbf{v}}_{x|y,\eta}(\mathbf{y},\mathbf{x}^0) = \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(H\boldsymbol{\beta} - \boldsymbol{\mu}_{x|y,\eta})$ هستند.

با در نظر گرفتن $\mathbf{x}^* = \begin{pmatrix} x_j & \mathbf{x}_{-j}\end{pmatrix}' = A_j\mathbf{x},\ j=1,\ldots,n$، که در آن $\mathbf{x}_{-j} = (x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n)$ و A_j ماتریس واحدی است که jامین سطر آن به سطر اول منتقل شده است، بنا بر خاصیت بسته بودن توزیع چولۀ نرمال بسته نسبت به تبدیلات خطی، $[\mathbf{x}^*\,|\,\mathbf{y},\boldsymbol{\eta}]$ دارای توزیع تقریبی

$$CSN(\hat{\boldsymbol{\mu}}_{x^*|y,\eta}, \hat{\Sigma}_{x^*|y,\eta}, \hat{D}_{x^*|y,\eta}, \hat{\mathbf{v}}_{x|y,\eta}, \hat{\Delta}_{x^*|y,\eta})$$

است، که در آن

$$\hat{\boldsymbol{\mu}}_{x^*|y,\eta} = A_j\hat{\boldsymbol{\mu}}_{x|y,\eta}(\mathbf{y},\mathbf{x}^0) = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\ \hat{\Sigma}_{x^*|y,\eta} = A_j\hat{\Sigma}_{x|y,\eta}(\mathbf{x}^0) = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix},$$
$$\hat{D}_{x^*|y,\eta} = \hat{D}_{x|y,\eta}A_j' = [D_1 \quad D_2],\ \hat{\Delta}_{x^*|y,\eta} = 1.$$

طبق خاصیت بسته بودن توزیع چولۀ نرمال بسته نسبت به حاشیه‌سازی، $[x_j\,|\,\mathbf{y}]$ دارای توزیع تقریبی

$$CSN(\mu_1, \Sigma_{11}, D^*, \hat{\mathbf{v}}_{x|y,\eta}, \Delta^*)$$

است، که در آن $D^* = D_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}$، $\Delta^* = 1 + D_2\Sigma_{22.1}D_2'$ و $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

بنا بر این از (۲)

$$E(x_j\,|\,\mathbf{y}) = \mu_1 + \Sigma_{11}D^{*\prime}\psi \qquad (٦)$$

که پیش‌گوی MMSE تقریبی متغیرهای پنهان فضایی در موقعیت jام است، که در آن

$$\boldsymbol{\psi} = \frac{\Phi_q^*(r; \hat{\mathbf{v}}_{x|y}, \Delta^* + D^*\Sigma_{11}D^{*\prime})}{\Phi_q(\mathbf{0}; \hat{\mathbf{v}}_{x|y}, \Delta + D_2\Sigma_{11}D_2')}\Big|_{r=0}.$$

و با فرض معلوم بودن بردار پارامترها قابل محاسبه است. اما در عمل اغلب پارامترهای مدل نامعلوم هستند و به‌دلیل وجود متغیرهای پنهان فضایی، برآورد پارامترهای مدل به‌راحتی قابل محاسبه نیستند. محمدزاده و حسینی [۷] الگوریتم MCEMG زانگ [۸] را برای مدل‌های SGLM با متغیرهای پنهان فضایی چوله تعمیم

دادند. اما اجرای این الگوریتم نیازمند تولید نمونه‌های مونت‌کارلویی و صرف وقت زیاد برای همگرایی است. در این‌جا با استفاده از الگوریتم EMG و تقریب ارائه شده به‌وسیلهٔ محمدزاده و حسینی [۵] برای $\pi(\mathbf{x}\,|\,\mathbf{y})$، الگوریتم جدید AEMG پیشنهاد می‌شود که به نمونه‌های مونت‌کارلویی نیاز ندارد و از سرعت محاسبات بیش‌تری نسبت به الگوریتم MCEMG برخوردار است.

فرض کنید $(\mathbf{y},\mathbf{x})$ شامل بردار متغیرهای پاسخ گسسته فضایی و متغیرهای پنهان باشد، آن‌گاه از رابطهٔ (٤) تابع درست‌نمایی کامل به‌صورت

$$L_c(\boldsymbol{\eta}\,|\,\mathbf{y},\mathbf{x}) = \prod_{i=1}^{k} \pi(y_i\,|\,\mathbf{x})\pi(\mathbf{x}\,|\,\boldsymbol{\eta})$$

است. برای اجرای الگوریتم AEMG، مقادیر اولیه پارامتر به‌صورت $\boldsymbol{\eta}^{(0)}$را در نظر گرفته و با قرار دادن $m=0$، با الگوریتم EMG بردار

$$\begin{aligned}\boldsymbol{\eta}^{(m+1)} &= \boldsymbol{\eta}^{(m)} - \left[E\left\{\frac{\partial^2 \ln L_c}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\,|\,\mathbf{y}\right\}\right]^{-1}_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}} \left[E(\frac{\partial \ln L_c}{\partial\boldsymbol{\eta}}\,|\,\mathbf{y})\right]_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}} \\ &= \boldsymbol{\eta}^{(m)} - \left[E\left\{\frac{\partial^2 \ln \pi(\mathbf{x}\,|\,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\,|\,\mathbf{y}\right\}\right]^{-1}_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}} \left[E\left\{\frac{\partial \ln \pi(\mathbf{x}\,|\,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\,|\,\mathbf{y}\right\}\right]_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(m)}} \qquad (7)\end{aligned}$$

محاسبه می‌شود، که در آن

$$\ln \pi(\mathbf{x}\,|\,\boldsymbol{\eta}) = -\frac{1}{2}\ln\left|\Sigma_\theta\right| - \frac{1}{2}(\mathbf{x}-H\boldsymbol{\beta})'\Sigma_\theta^{-1}(\mathbf{x}-H\boldsymbol{\beta}) + \ln\{2\Phi(\lambda'\Sigma_\theta^{-\frac{1}{2}}(\mathbf{x}-H\boldsymbol{\beta}))\}$$

و امید ریاضی‌ها در رابطهٔ (۷) به‌طور تقریبی با به‌کار بردن توزیع تقریبی (۵) قابل محاسبه هستند. در صورت همگرایی، $\boldsymbol{\eta}^{(m+1)}$ برآورد به‌صورت پارامتر خواهد بود، در غیر این صورت الگوریتم تکرار می‌شود.

بررسی شبیه‌سازی

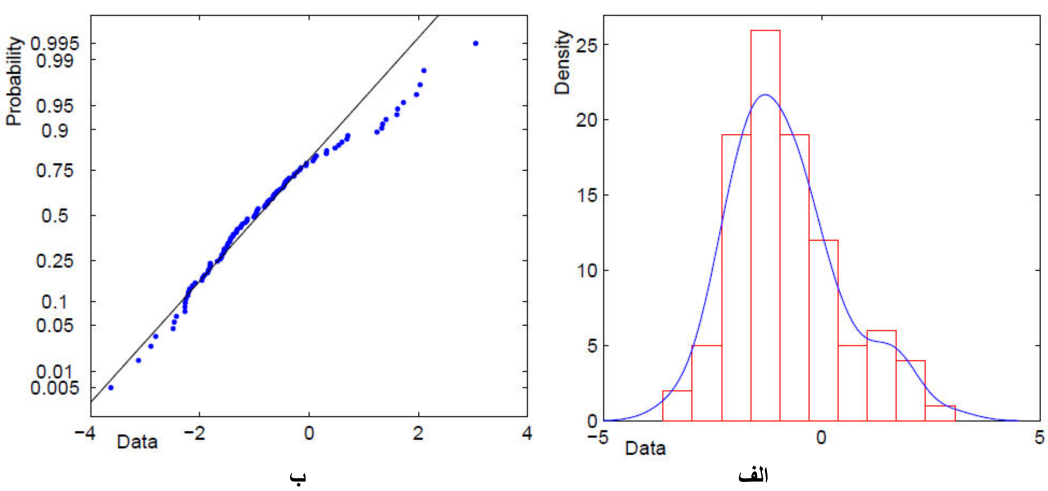
در این بخش عمل‌کرد مدل و روش‌های پیشنهاد شده در یك بررسی شبیه‌سازی براساس داده‌های تولید شده در مشبکه‌ای منظم ۱۰×۱۰، به‌صورت $\{(\ell,k),\ell,k=1,\ldots,10\}$ ارزیابی می‌شود. برای مقادیر معین پارامترهای توزیع CSN، متغیرهای پنهان فضایی $\mathbf{x}$ از توزیع $CSN_{100,1}(\beta_1 z,\Sigma_\theta,\lambda_0\mathbf{1}',\Sigma_\theta^{-\frac{1}{2}},0,1)$ تولید شده‌اند مقادیر $\beta_1=۰/۵$ ، $\sigma^2=۲$ ، $\varphi=٤$ و $\lambda_0=۲$ در نظر گرفته شده‌اند. تابع کواریانس همسان‌گرد به‌صورت $C(h)=\sigma^2\exp(-h/\varphi),\ \text{h}>0$ و متغیر تبیینی در موقعیت (ℓ,k) ام به‌صورت $z_{\ell k}=\log(1+\ell)$ در نظر گرفته شده است. متغیرهای پاسخ $y_{\ell k}$ نیز با شرطی کردن روی متغیرهای پنهان فضایی از $y_{\ell k}\sim Bin(100,\exp(x_{\ell k})/[1+\exp(x_{\ell k})])$ تولید شده‌اند. بافت‌نگار و نمودار Q-Q مقادیر تولید شده برای متغیرهای پنهان در شکل ۱ نشان دهنده چوله بودن توزیع داده‌ها است.

دو مدل SGLM، یکی با توزیع چوله‌نرمال بسته و دیگری با توزیع نرمال برای متغیرهای پنهان فضایی

شبیه‌سازی و براساس معیار میانگین توان دوم خطا، دقت برآوردها و براساس معیار میانگین توان دوم خطای پیش‌گویی دقت پیش‌گویی‌های حاصل مقایسه شده‌اند. برآورد پارامترها از دو الگوریتم AEMG و MCEMG محاسبه شده‌اند.

برای محاسبهٔ مقادیر MSE تحت شرایط بالا، ۱۰۰ مجموعه داده تولید شده است. متوسط مقادیر برآورد شده پارامترها و MSE در جدول ۱ ارائه شده‌اند. چنان‌که ملاحظه می‌شود، برآورد پارامترهای مدل SGLM با توزیع چوله‌نرمال بسته برای متغیرهای پنهان فضایی از اریبی و MSE کوچک‌تری نسبت به مدل SGLM با متغیرهای پنهان فضایی نرمال برخوردار است. برای بررسی معنی‌داری تفاوت بین MSEهای به دست آمده دو مدل نرمال و چوله وقتی پارامترها با الگوریتم AEMG برآورد شده‌اند، از آزمون تی زوجی استفاده شده است و مقادیر آمارهٔ آزمون، مقادیر احتمال و بازه‌های اطمینان ۹۵% تفاوت‌ها در جدول ۱ آورده شده است. چنان‌که ملاحظه می‌شود این آزمون‌ها تفاوت معنی‌دار بین مقادیر MSE برای دو مدل نرمال و چوله را تایید می‌کند.

هم‌چنین چنان‌که در جدول۱ ملاحظه می‌شود، نتایج به‌دست آمده از الگوریتم‌های AEMG و MCEMG مشابه هم است با این تفاوت که اجرای الگوریتم پیشنهاد شده بسیار سریع است، در صورتی که اجرای الگوریتم MCEMG ساعت‌ها به‌طول می‌انجامد. برای مقایسهٔ دقت پیش‌گویی دو مدل، در موقعیت (۵/۵ و ۵/۵) مقدار پیش‌گو برای هر ۱۰۰ مجموعه داده محاسبه شده است. سپس مقدار میانگین توان دوم خطای پیش‌گویی برای دو مدل SGLM با توزیع‌های چوله‌نرمال بسته و نرمال محاسبه و به‌ترتیب مقادیر ۱/۶۴۳۲ و ۱/۸۸۴۲ حاصل شده است. این نتایج بیانگر آن است که توزیع چوله‌نرمال بسته برای متغیرهای پنهان فضایی موجب کاهش مقدار میانگین توان دوم خطای پیش‌گویی می‌شود.



نمودار ۱. الف. نمودار Q-Q، ب. بافت‌نگار متغیرهای پنهان تولید شده

جدول ۱.متوسط برآورد و میانگین مربع خطای پارامترها براساس ۱۰۰ مجموعه داده

الگوریتم	پارامتر	مدل: CSN SGLM		مدل: N SGLM		آزمون معنی داری تفاوت مقادیرMSE دو مدل		
		اریبی	MSE	اریبی	MSE	مقدار آماره	مقدار احتمال	بازه اطمینان تفاوت دو مقدار
AEMG	β_1	-۰/۰۵۷۳	۰/۲۶۳۱	-۰/۰۹۱۸	۰/۳۴۵۵	۶/۰۴	۰/۰۰۰۰	(۰/۱۱۷۳ و ۰/۰۵۹۳)
	σ	-۰/۰۳۷۳	۰/۰۵۲۳	-۰/۱۳۲۶	۰/۰۶۱۳	۵/۹۴	۰/۰۰۰۰	(۰/۰۱۱۸ و ۰/۰۰۵۹)
	φ	۰/۱۶۸۶	۱/۴۹۳۱	۰/۱۸۱۹	۱/۸۲۸۹	۲۷/۱	۰/۰۰۰۰	(۰/۴۷۵۹ و ۰/۳۲۴۶)
	λ	-۰/۴۷۲۲	۱/۳۱۲۴	-	-	-	-	
MCEMG	β_1	-۰/۰۷۶۳	۰/۲۱۴۶	-۰/۰۸۸۷	۰/۳۳۲۱	-	-	
	σ	-۰/۰۳۰۵	۰/۰۵۱۷	-۰/۱۲۳۶	۰/۰۶۲۱	-	-	
	φ	۰/۱۵۲۳	۱/۴۸۳۴	۰/۳۹۱۹	۱/۹۱۲۵	-	-	
	λ	-۰/۴۵۷۴	۱/۲۸۴۱	-	-	-	-	

بحث و نتیجه‌گیری

وجود متغیرهای پنهان فضایی در مدل‌های SGLMM و نامعلوم بودن توزیع واقعی آن‌ها، روی برآورد پارامترهای مدل و دقت پیش‌گویی تأثیرگذار است. لذا در عمل فرض نرمال بودن متغیرهای پنهان فضایی گاهی فرض نادرست و گمراه‌کننده‌ای است و استنباط‌های آماری براساس این فرض، از جمله برآورد پارامترها و پیش‌گوها را بی‌اعتبار می‌سازد. برای افزایش دقت برآورد پارامترها و پیش‌گوها، استفاده از توزیع چوله‌نرمال بسته که کلاس بزرگ‌تر و انعطاف‌پذیرتری از کلاس توزیع نرمال است، برای متغیرهای پنهان پیشنهاد گردید و الگوریتمی برای برآورد تقریبی پارامترهای مدل و پیش‌گوی تقریبی بهینه برای متغیرهای پنهان معرفی شدند. در بررسی شبیه‌سازی نشان داده شد که استفاده از توزیع چولهٔ نرمال بسته به‌عنوان توزیع متغیرهای پنهان فضایی در مدل‌های SGLM به‌جای توزیع نرمال، موجب افزایش دقت برآورد پارامترها و پیش‌گوها می‌گردد. همچنین الگوریتم AEMG برای به‌دست آوردن پارامترهای مدل معرفی شد به‌طوری که استفاده از الگوریتم AEMG در مقایسهٔ با الگوریتم MCEMG باعث ارائه نتایج مشابه و کاهش زمان محاسبات می‌گردد. زمان اجرای الگوریتم MCEMG با رایانه شخصی با مشخصات Cpu=2.5 Ghz و نرم‌افزار Matlab برای هر مجموعه داده حدود چهل دقیقه و برای ۱۰۰ مجموعه داده حدود سه روز و اجرای الگوریتم AEMG برای هر مجموعه داده حدود بیست ثانیه و برای ۱۰۰ مجموعه داده حدود نیم ساعت به طول انجامید. لذا تفاوت زیادی در زمان اجرای الگوریتم‌ها وجود دارد، در صورتی‌که در نتایج تفاوت زیادی مشاهده نمی‌شود.

قدردانی و تشکر

نویسندگان از داوران محترم مجله که نظرات ارزنده آن‌ها موجب بهبود مقاله گردید و از حمایت قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد قدردانی می‌کنند.

منابع

1. NE Breslow, D. G. Clayton, "Approximate Inference in Generalized Linear Mixed Models", *Journal of the American Statistical Association*, 88 (1993) 9-25.
2. O. F. Christensen, "Monte Carlo Maximum Likelihood in Model-Based Geostatistics", *Journal of Computaional and Graphical Statistics*, 13 (2004) 702-718.
3. J. Dominguez-Molina, G. Gonzalez-Farias, A. Gupta, "The Multivariate Closed Skew Normal Distribution", *Technical Report 03-12*, Department of Mathematics and Statistics, Bowling Green State University (2003).
4. P. Diggle, J. A. Tawn, R. A. Moyeed, "Model-Based Geostatistics (with Discussio)", *Journal of the Royal Statistical Society. Series C. Applied Statistics* 47 (1998) 299-350.
5. F. Hosseini, J. Eidsvik, M. Mohammadzadeh, "Approximate Bayesian Inference in Spatial GLMM with Skew Normal Latent Variables", *Computational Statistics and Data Analysis*, 55 (2011) 1791-1806.
6. P. McCullagh, J. A. Nelder, Generalized Linear Model, *Chapman and Hall*, London (1989).
7. M. Mohammadzadeh, F. Hosseini, "Maximum-Likelihood Estimation for Spatial GLM Models with Closed-Skew Normal Latent Variables", *Procedia Environmental Science,* 3 (2011) 63-68.
8. H. Zhang, "On Estimation and Prediction for Spatial Generalized Linear Mixed Models", *Biometrics*, 58 (2002) 129-136.