تحلیل بیز مقداری مدل رگرسیون فضایی چوله بر اساس زیر رده منعطفی از توزیع چوله نرمال بسته

امید کریمی، فاطمه حسینی

گروه آمار دانشگاه سمنان

چکیده: مدل‌های رگرسیون فضایی برای تحلیل پاسخ‌های کمی فضایی براساس روابط خطی و غیرخطی با متغیرهای توضیحی به‌کارگرفته می‌شوند. معمولاً همبستگی فضایی پاسخ‌ها با یک میدان تصادفی گاوسی بر اساس توزیع نرمال چند متغیره مدل می‌شوند. اما در عمل با پاسخ‌های چوله‌ای مواجه می‌شویم که برای تحلیل آن‌ها از خانواده توزیع‌های چوله نرمال استفاده می‌شود. توزیع چوله نرمال بسته یکی از خانواده‌های گرده توزیع چوله نرمال است که خواص مشابه توزیع نرمال دارد. در این مقاله تحلیل بیز سلسله مراتبی این مدل‌ها براساس یک زیر رده منعطف از توزیع چوله نرمال بسته ارائه می‌گردد. به دلیل زمان‌بر بودن محاسبات روش‌های مونت کارلویی در تحلیل بیز سلسله مراتبی از رهیافت بیز مقداری برای تقریب توزیع پسین استفاده می‌شود. سپس مدل پیشنهادی روی داده‌های واقعی زمین لرزه‌ای کشور ایران پیاده‌سازی و مورد تحلیل قرار می‌گیرد.

واژه‌های کلیدی: رهیافت بیزی سلسله مراتبی، توزیع مقداری بهینه، هم‌تغییرنگار، داده‌های فضایی.

آدرس الکترونیکی نویسنده مسئول مقاله: امید کریمی، omid.karimi@semnan.ac.ir
کد موضوع‌بندی ریاضی (۲۰۱۰): 60G15, 62M30 .

۱ مقدمه

معمولاً در زمینه‌های مختلفی همچون زمین‌شناسی، اپیدمیولوژی، جغرافیا و پزشکی با پاسخ‌های کمی مواجه می‌شویم که دارای همبستگی فضایی هستند و این همبستگی به فضای مورد مطالعه آن‌ها بستگی دارد. برای مدل‌بندی این نوع پاسخ‌ها از مدل‌های رگرسیون فضایی استفاده می‌شود. انسلین (۱۹۹۰) برای لحاظ کردن همبستگی فضایی در مدل‌های رگرسیون روش‌های مختلفی پیشنهاد کرد. باسو و رینسل (۱۹۹۴) تحلیل درست‌نمایی را برای مدل رگرسیونی با خطاهای خودهمبسته فضایی ارائه کردند. اوه و همکاران (۲۰۰۲) رهیافت بیزی سلسله مراتبی را برای این نوع مدل‌ها با داده‌های گمشده در متغیر پاسخ مورد مطالعه قرار دادند. لی و هوانگ (۲۰۲۲) تاثیر کوید ۱۹ بر بازارهای مسکن ایالات متحده را به‌وسیله مدل‌های رگرسیون فضایی بررسی کردند. در بیشتر مباحث نظری مدل‌های رگرسیون فضایی فرض بر این است که داده‌ها تحققی از یک میدان تصادفی گاوسی هستند، که در آن خطاها دارای توزیع نرمال است. این فرض ممکن است برای بعضی از داده‌ها برقرار نباشد. به‌عنوان مثال: داده‌های بارندگی در مقاله کیم و مالیک (۲۰۰۴)، داده‌های لرزه‌ای در مقاله کریمی و همکاران (۲۰۱۰) و داده‌های آلودگی هوا که توسط کریمی و محمدزاده (۲۰۱۲) مورد استفاده قرار گرفته است. معمولاً در عمل با پاسخ‌هایی مواجه می‌شویم که دارای چولگی هستند و برای مدل‌کردن آن‌ها از توزیع‌های چوله نرمال استفاده می‌شود. کریمی و محمدزاده (۲۰۱۲) تحلیل بیزی برای مدل رگرسیون فضایی چوله بر اساس تحققی از یک میدان تصادفی چوله گاوسی بسته ارائه کردند. اما این میدان‌های تصادفی دارای شرط سازگاری حاشیه‌ای نبودند (کریمی، ۲۰۲۳؛ کریمی و حسینی ، ۱۴۰۰). تان و همکاران (۲۰۱۳) روش‌های بیز مقداری[1] را برای مدل‌های خطی تعمیم‌یافته در مقابل روش‌های بیزی سلسله مراتبی به‌کار گرفتند که برای مدل‌های پیچیده عملکرد مناسبی دارند. ژانگ و همکاران (۲۰۲۳) نیز روش بیز مقداری را برای برازش مدل‌های فضا حالت[2] ارائه کردند. کابرال و همکاران (۲۰۲۴) رهیافت بیز مقداری را برای برازش مدل‌های پنهان فضایی با تقریب لاپلاس به‌کار گرفتند و نشان دادند که زمان اجرا این رهیافت سریع‌تر از رهیافت‌های دیگر است.

در این مقاله تحلیل بیز مقداری روی مدل‌های رگرسیون فضایی بر اساس یک زیر رده منعطف از توزیع

[1] Variational Bayes
[2] State Space Models

چوله نرمال بسته[۳] (CSN) بیان می‌شود که دارای شرط سازگاری حاشیه‌ای و خوش تعریف است. ساختار مقاله به این صورت است که در بخش ۲ یک زیر رده از توزیع CSN بیان می‌گردد. سپس مدل رگرسیون فضایی چوله به همراه تحلیل بیز مقداری و سلسله مراتبی آن در بخش ۳ ارائه می‌گردد. در بخش ۴ یک مطالعه شبیه‌سازی برای ارزیابی مدل پیشنهادی با رهیافت بیز مقداری بیان می‌شود و روش‌های پیشنهادی برای تحلیل یک مثال واقعی مربوط به داده‌های زمین لرزه‌ای کشور طی ۱۵ سال اخیر به کار می‌رود. در نهایت بحث و نتیجه‌گیری ارائه می‌شود.

۲ توزیع چوله نرمال بسته

متغیر تصادفی $\boldsymbol{Y}$ دارای توزیع CSN است (گنزالس و همکاران، ۲۰۰۴)، اگر دارای تابع چگالی

$$f_Y(\boldsymbol{y}) = \phi_p(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma) \frac{\Phi_q(\Gamma(\boldsymbol{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta)}{\Phi_q(\mathbf{0}; \boldsymbol{\nu}, \Delta + \Gamma\Sigma\Gamma')}, \tag{۱}$$

باشد و با نماد $\mathrm{CSN}_{p,q}(\boldsymbol{\mu}, \Sigma, \Gamma, \boldsymbol{\nu}, \Delta)$ نمایش داده می‌شود، که در آن $\boldsymbol{\mu} \in R^p$، Σ ماتریس معین مثبت $p \times p$ است که شامل پارامتر مقیاس است، Γ ماتریس $q \times p$ است که شامل پارامتر چولگی است، $\boldsymbol{\nu} \in R^q$ و Δ یک ماتریس $q \times q$ معین مثبت است. ϕ_p و Φ_q به‌ترتیب تابع چگالی و تابع توزیع تجمعی توزیع نرمال p و q متغیره هستند. توزیع CSN تحت تبدیلات خطی و مجموع بردارهای تصادفی مستقل بسته است. هر چند این توزیع دارای خواص بسته بودن و انعطاف‌پذیر است اما تابع چگالی آن دارای پیچیدگی و پارامترهای زیادی است. به عنوان مثال برای q های بزرگ محاسبه Φ_q دشوار و چالش برانگیز است. به همین خاطر یک زیر رده از این توزیع بر اساس ایده مارکز و گنزالس (۲۰۲۲) برای تحلیل مدل رگرسیون فضایی چوله ارائه می‌شود. برای اطلاعات بیشتر در مورد این زیر رده منعطف و نحوه تعریف یک میدان تصادفی فضایی چوله خوش تعریف (دارا بودن خاصیت سازگاری حاشیه‌ای) به مقاله کریمی و حسینی (۱۴۰۰) مراجعه شود. فرض کنید متغیر تصادفی X_i دارای توزیع CSN به‌صورت $X_i \sim \mathrm{CSN}_{۱,۱}(۰, ۱, \lambda_i, ۰, ۱)$ و X_i به ازای $i = ۱, \ldots, n$ مستقل باشند، آن‌گاه تابع چگالی

[3] Closed Skew Normal

X به‌صورت $f_X(\boldsymbol{x}) = ۲^n \phi_n(\boldsymbol{x}; \mathbf{0}, \mathbf{I_n}) \Phi_n(\Gamma \boldsymbol{x})$ به دست می‌آید که در آن $\Gamma = \mathrm{Diag}(\boldsymbol{\lambda})$، $\boldsymbol{\lambda} = (\lambda_۱, \ldots, \lambda_n)'$، $\Phi_n(\Gamma \boldsymbol{x}) = \prod_{i=۱}^n \Phi(\lambda_i x_i)$ و $\lambda_i \in R$ است. در آن صورت $\boldsymbol{X} \sim \mathrm{CSN}_{n,n}(۰, I_n, \Gamma, \mathbf{0}, \mathbf{I_n})$، بنابراین $E(\boldsymbol{X}) = b\boldsymbol{\delta}$ و $V(\boldsymbol{X}) = \Omega$ که در آن $\boldsymbol{\delta} = (\delta_۱, \ldots, \delta_n)'$، $\boldsymbol{\tau} = (\tau_۱, \ldots, \tau_n)'$، $b = (۲/\pi)^{\frac{۱}{۲}}$، $\delta_i = \lambda_i(۱ + \lambda_i^۲)^{-\frac{۱}{۲}}$، $\tau_i = (۱ - b^۲ \delta_i^۲)^{-\frac{۱}{۲}}$ و $\Omega = \mathrm{Diag}(\boldsymbol{\tau})$ است. بردار تصادفی $\boldsymbol{X}$ را می‌توان به‌صورت $\boldsymbol{Z} = \Omega^{-۱}(\boldsymbol{X} - b\boldsymbol{\delta})$ استاندارد کرد، به این ترتیب $\mathrm{E}(\boldsymbol{Z}) = \mathbf{0}$ و $\mathrm{V}(\boldsymbol{Z}) = \mathbf{I_n}$ است. در ادامه با استفاده از خاصیت بسته بودن توزیع‌های CSN نسبت به تبدیلات خطی، مدل رگرسیون فضایی چوله تعریف و برآورد بیزی آن ارائه می‌شود.

۳ مدل رگرسیون فضایی چوله

با در نظر گرفتن بردار تصادفی استاندارد شده $\boldsymbol{Z}$، مدل رگرسیون فضایی چوله برای بردار پاسخ $\boldsymbol{y}_n$ در n موقعیت فضایی $(\boldsymbol{s}_۱, \ldots, \boldsymbol{s}_n)$ به‌صورت

$$\boldsymbol{y}_n = H\boldsymbol{\beta} + \sigma C \boldsymbol{Z}. \tag{۲}$$

تعریف می‌شود، که در آن H یک ماتریس $n \times p$ از متغیرهای توضیحی، $\boldsymbol{\beta}$ یک بردار p بعدی از ضرائب رگرسیونی، $\sigma > ۰$ پارامتر مقیاس و C یک ماتریس $n \times n$ که از تجزیه چولسکی ماتریس همبستگی فضایی به‌صورت $\Sigma_n = CC'$ به‌دست می‌آید. در واقع Σ_n ماتریس کوواریانس فضایی بردار تصادفی $\boldsymbol{y}_n$ با مولفه‌های $\sigma_{ij} = \sigma^۲ \rho\big(y(\boldsymbol{s}_i), y(\boldsymbol{s}_j); \varphi\big)$ است که در آن $\rho(\cdot; \varphi)$ یک تابع همبستگی فضایی با پارامتر دامنه فضایی φ است. بنابراین بردار پاسخ‌های فضایی $\boldsymbol{y}_n$ با توجه به خاصیت تبدیلات خطی توزیع CSN دارای توزیع CSN به‌صورت $\boldsymbol{y}_n \sim \mathrm{CSN}_{n,n}(\boldsymbol{\mu}_y, \sigma_y, \frac{۱}{\sigma}\Gamma\Omega C^{-۱}, \mathbf{0}, \mathbf{I_n})$ است، که در آن $\boldsymbol{\mu}_y = H\boldsymbol{\beta} - b\sigma C \Omega^{-۱} \boldsymbol{\delta}$ و $\Sigma_y = \sigma^۲ C \Omega^{-۲} C'$. از این رو تابع چگالی به‌صورت

$$f_\eta(\boldsymbol{y}) = ۲^n \phi_n(\boldsymbol{y}; \boldsymbol{\mu}_y, \sigma_y) \Phi_n\big(\frac{۱}{\sigma}\Gamma\Omega C^{-۱}(\boldsymbol{y} - \boldsymbol{\mu}_y)\big)$$

است، که در آن $\boldsymbol{\eta} = (\boldsymbol{\beta}, \sigma, \varphi, \boldsymbol{\lambda})'$ پارامترهای مدل است. لگاریتم تابع درست‌نمایی به‌صورت

$$\begin{aligned}\ell(\boldsymbol{\eta}|\boldsymbol{y}) &= \log\big(f_\eta(\boldsymbol{y})\big)\\ &= n\log(۲) - \frac{n}{۲}\log(۲\pi) - \frac{V'V}{۲} - n\log(\sigma) - \log(|C|)\\ &\quad + \sum_{i=۱}^{n}\big\{\log\big(\Phi(\lambda_i v_i)\big) + \log(\tau_i)\big\}, \end{aligned} \tag{۳}$$

ساده می‌شود، که در آن $V = \frac{۱}{\sigma}\Omega C^{-۱}(\boldsymbol{y} - \boldsymbol{\mu}_y)$ و v_i مولفه i ام بردار V است. با توجه به پیچیدگی تابع درست‌نمایی برای برازش پارامترهای مدل رگرسیون فضایی چوله از رهیافت‌های بیز سلسله مراتبی استفاده می‌شود و رهیافت بیز مقداری نیز برای تقریب توزیع پسین به کار گرفته می‌شود که در ادامه به آن می‌پردازیم.

۱.۳ برازش مدل

برای برآورد پارامترهای مدل رگرسیون فضایی (۲) از رهیافت بیزی سلسله مراتبی با توجه به پیچیدگی تابع درست‌نمایی (۳) استفاده می‌شود. با در نظر گرفتن پیشین‌های معمول برای پارامترها به‌صورت

$$\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_\beta, \sigma_\beta^۲\mathbf{I_p}), \quad \sigma \sim IG(a,b), \quad \varphi \sim G(\alpha,\gamma), \quad \lambda_i \sim N(\mu_{\lambda_i}, \ \sigma_{\lambda_i}^۲),$$

توزیع پسین به‌صورت

$$\begin{aligned}\pi(\boldsymbol{\eta}|\boldsymbol{y}) &\propto \phi_n(\boldsymbol{y};\boldsymbol{\mu}_y,\sigma_y)\Phi_n\big(\frac{۱}{\sigma}\Gamma\Omega C^{-۱}(\boldsymbol{y}-\boldsymbol{\mu}_y)\big)\phi_p(\boldsymbol{\beta};\boldsymbol{\mu}_\beta,\sigma_\beta^۲\mathbf{I_p})\pi(\sigma;a,b)\\ &\times \pi(\varphi;\alpha,\gamma)\Pi_{i=۱}^n\phi(\lambda_i;\mu_{\lambda_i},\sigma_{\lambda_i}^۲)\pi(\boldsymbol{\mu}_\beta)\pi(\sigma_\beta^۲)\\ &\times \pi(a)\pi(b)\pi(\alpha)\pi(\gamma)\Pi_{i=۱}^n\pi(\mu_{\lambda_i})\pi(\sigma_{\lambda_i}^۲),\end{aligned}$$

مقاله پذیرفته شده

خلاصه می‌شود، که در آن $\pi(\cdot)$ توزیع پیشین ابر پارامترهای $\boldsymbol{\mu}_\beta$، σ^2_β، a، b، α، γ، μ_{λ_i} و $\sigma^2_{\lambda_i}$ است. توزیع پیشین ابرپارامترها به‌صورت

$$\boldsymbol{\mu}_\beta \sim N_p(\mathbf{0}, ۵\mathbf{I_p}), \quad \sigma^۲_\beta \sim \mathrm{IG}(۳, ۲), \quad a, b \sim U(۱, ۵),$$
$$\alpha, \gamma \sim U(۱, ۱۰), \quad \mu_{\lambda_i} \sim N(۰, ۵), \quad \sigma^۲_{\lambda_i} \sim \Gamma(۰/۵, ۰/۵),$$

لحاظ شده‌اند. لازم به ذکر است که مقادیر معلوم برای هر یک از پارامترهای ابر پیشین‌ها طوری در نظر گرفته شده‌اند که حسایت پیشین‌ها با بالا بردن پراکندگی آن‌ها کم شود. به عنوان مثال در توزیع پیشین ابر پارامتر μ_{λ_i} واریانس ۵ لحاظ شده است که پارامتر میانگین توزیع پیشین پارامتر چولگی از اطلاع کمی برخوردار باشد و حساسیت آن در استنباط کمتر شود. با توجه به پیچیده بودن توزیع پسین لازم است از روش‌های مونت کارلویی استفاده شود. در این مقاله از رهیافت بیز مقداری (تان و همکاران ، ۲۰۱۳) برای تقریب توزیع پسین با سرعت محاسبات نسبتا بالا به‌عنوان یک جایگزین مناسب برای روش‌های MCMC استفاده می‌شود.

۲.۳ رهیافت بیز مقداری

هدف بیز مقداری یافتن توزیع مقداری $q(\boldsymbol{\eta})$ است، که توزیع پسین $\pi(\boldsymbol{\eta}|\boldsymbol{y})$ را از طریق مینیمم کردن فاصله احتمال بین توزیع مقداری و توزیع پسین تقریب می‌زند (تان و همکاران ، ۲۰۱۳). فاصله احتمال توسط رابطه کولبک - لایبلر[۴] (KL) به‌صورت

$$\begin{aligned}\mathrm{KL}(q(\boldsymbol{\eta})\|\pi(\boldsymbol{\eta} \mid \boldsymbol{y})) &= \int \ln\left(\frac{q(\boldsymbol{\eta})}{\pi(\boldsymbol{\eta} \mid \boldsymbol{y})}\right) q(\boldsymbol{\eta}) d\boldsymbol{\eta} \\ &= \mathbb{E}_q[\ln q(\boldsymbol{\eta})] - \mathbb{E}_q[\ln \pi(\boldsymbol{\eta} \mid \boldsymbol{y})] \\ &= \mathbb{E}_q[\ln q(\boldsymbol{\eta})] - \mathbb{E}_q[\ln \pi(\boldsymbol{\eta}, \boldsymbol{y})] + \ln f(\boldsymbol{y}).\end{aligned} \tag{۴}$$

[4] Kullback-Leibler

اندازه‌گیری می‌شود. در این روش با مینیمم کردن رابطه KL به توزیع هدف

$$q^*(\eta) = \arg\min_q \mathrm{KL}\left(q(\boldsymbol{\eta}) \| \pi(\boldsymbol{\eta} \mid \boldsymbol{y})\right).$$

به‌عنوان تقریبی از توزیع پسین می‌رسد. چون $\ln f(\boldsymbol{y})$ شکل بسته‌ای ندارد، رابطه KL از نظر تحلیلی قابل حل نیست. معادله (۴) را به‌صورت

$$\mathrm{RKL}(q(\boldsymbol{\eta})) = \ln f(\boldsymbol{y}) - \mathrm{KL}\big(q(\eta) \| \pi(\eta \mid \boldsymbol{y})\big). \tag{۵}$$

می‌نویسیم. از آنجا که رابطه KL همیشه مثبت است، معادله (۴) نشان می‌دهد که توزیع مقداری بهینه را می‌توان با ماکسیمم کردن رابطه (۵) به‌دست آورد. توزیع مقداری باید توسط تحلیلگر انتخاب شود. که معمولاً از تقریب‌های گاوسی استفاده می‌شود. یک راه حل برای تعیین توزیع مقداری استفاده از بیز مقداری میدان میانگین [۵] است. در این روش چگالی توزیع مقداری به‌صورت حاصلضرب مولفه‌های مجزا یه‌صورت

$$q(\boldsymbol{\eta}) = \prod_{j=1}^{J} q(\eta_j) = q(\boldsymbol{\beta})q(\sigma)q(\phi)q(\boldsymbol{\lambda}),$$

مشخص می‌شود، که در آن $j \in \{1, \ldots, J\}$ مجموعه اندیس‌گذار پارامترهای مدل است. این مشخصه‌سازی باعث استقلال پسینی بین پارامترها می‌شود. اورمرود و وند (۲۰۱۰) چگالی مقداری بهینه را به‌صورت $q^*(\eta_j) \propto \exp[\mathbb{E}_{-\eta_j}\left(\log\pi(\boldsymbol{y}, \boldsymbol{\eta})\right)]$ به‌دست آوردند، که در آن نماد $-\eta_j$ یعنی همه مولفه‌های بردار $\boldsymbol{\eta}$ بجز jامین مولفه آن است و

$$\mathbb{E}_{-\eta_j}\left(\log\pi(\boldsymbol{y}, \boldsymbol{\eta})\right) = \int \prod_{\substack{k=1 \\ k\neq j}}^{J} q(\eta_k) \log\pi(\boldsymbol{y}, \boldsymbol{\eta}) d_{\eta_1} \ldots d_{\eta_{j-1}} d_{\eta_{j+1}} \ldots d_{\eta_J}.$$

[5] Mean Field Variational Bayes

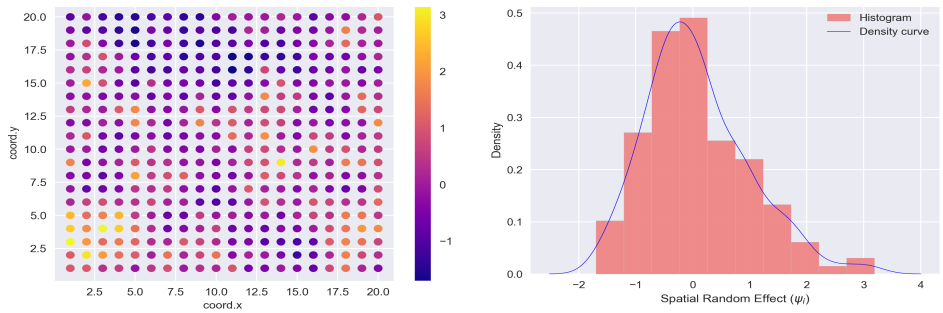
در صورتی که $q^*(\eta_j)$ دارای شکل توزیعی مشخصی باشد، توزیع مقداری آن به‌راحتی محاسبه می‌شود و به آن مزدوج شرطی گفته می‌شود.

۴ مطالعه شبیه‌سازی

در این بخش یک مطالعه شبیه‌سازی بر اساس مدل رگرسیون فضایی چوله (۲) در مشبکه‌های منظم ۵×۵، ۱۰×۱۰ و ۲۰×۲۰ با رهیافت بیز مقداری برای ارزیابی مدل پیشنهادی صورت می‌گیرد. برای این منظور متغیر پاسخ را از مدل رگرسیون فضایی چوله به‌صورت زیر تولید می‌کنیم:

$$y_i = \beta_۰ + \beta_۱ x_i + \psi_i,\ i = ۱, ۲, \ldots, n \in \{۲۵, ۱۰۰, ۴۰۰\},$$

که در آن متغیر توضیحی x_i از توزیع نرمال با میانگین صفر و انحراف معیار یک تولید می‌شود و ψ_i مولفه iام عبارت $\sigma C \boldsymbol{Z}$ در مدل (۲) است. ساختار همبستگی فضایی همسانگرد نمایی برای محاسبه ماتریس $\boldsymbol{C}$ و پارامترهای مدل را با مقادیر معلوم $\beta_۰ = ۵$، $\beta_۱ = ۲$، $\sigma^۲ = ۱$، $\varphi = ۶$، $\lambda_i = \lambda = ۳$ در نظر می‌گیریم. همچنین برای ارزیابی عملکرد مدل پیشنهادی، مدل رگرسیون فضایی گاوسی به عنوان یک مدل رقیب پایه جهت مقایسه به کار گرفته می‌شود. اگر در مدل چوله پارامتر چولگی λ را صفر قرار دهیم، مدل گاوسی به‌دست می‌آید که نشان می‌دهد مدل رگرسیون فضایی گاوسی حالت خاص مدل رگرسیون فضایی چوله است. از این‌رو فرایند شبیه‌سازی را ۱۰۰ بار تکرار کردیم و یک نمونه از اثرات فضایی شبیه‌سازی شده مشبکه‌ی منظم ۲۰×۲۰ در شکل ۱ رسم شده است. برای ارزیابی برآورد بیز مقداری پارامترهای دو مدل از معیار ریشه دوم میانگین توان دوم خطاها $RMSE(\hat{\eta}) = \sqrt{\frac{۱}{m}\sum_{i=۱}^{m}(\hat{\eta}_i - \eta_i)^۲}$ با ۱۰۰ مجموعه داده شبیه‌سازی استفاده شده است، که در آن $m = ۱۰۰$ تعداد مجموعه داده‌های شبیه‌سازی برای هر حالت و $\hat{\eta}_i$ برآورد بیز مقداری پارامترهای مدل برای مجموعه داده شبیه‌سازی‌شده iام است. نتایج برآورد پارامترها برای ۱۰۰ مجموعه داده شبیه‌سازی‌شده با مشبکه منظم ۲۰×۲۰ در جدول ۱ ارایه شدند. نتایج نشان می‌دهد که مدل رگرسیون فضایی چوله بر اساس رهیافت بیز مقداری نسبت به مدل گاوسی به خوبی عمل نموده است. البته نتایج RMSE بیان می‌کنند که پارامترهای همبستگی فضایی در دو مدل از



(الف) (ب)

شکل ۱: نمونه‌ای از داده‌های شبیه‌سازی شده مدل رگرسیون فضایی چوله: (الف) هیستوگرام اثرات فضایی چوله، (ب) نمودار پراکنش اثرات فضایی در مشبکه منظم ۲۰ × ۲۰.

جدول ۱: نتایج برآورد بیز مقداری برای دو مدل رگرسیون فضایی چوله و مدل گاوسی براساس ۱۰۰ مجموعه داده شبیه‌سازی شده در مشبکه منظم ۲۰ × ۲۰

		مدل چوله			مدل گاوسی		
پارامتر	واقعی	برآورد	انحراف معیار	RMSE	برآورد	انحراف معیار	RMSE
β_0	۵	۵/۳۷۲	۰/۶۳۸	۰/۳۷۶	۵/۸۸۲	۰/۷۲۴	۰/۴۱۳
β_1	۲	۱/۸۹۳	۰/۲۵۲	۰/۱۲۲	۲/۶۱۱	۰/۵۱۳	۰/۴۱۹
σ^2	۱	۱/۳۰۶	۰/۰۵۱	۰/۸۲۸	۱/۸۲۲	۰/۱۴۷	۰/۹۱۱
φ	۶	۶/۲۷۷	۰/۹۵۷	۰/۱۳۳	۴/۷۵۸	۱/۵۴۴	۱/۸۷۱
λ	۳	۲/۵۲۹	۰/۳۱۳	۰/۰۲۷	—	—	—

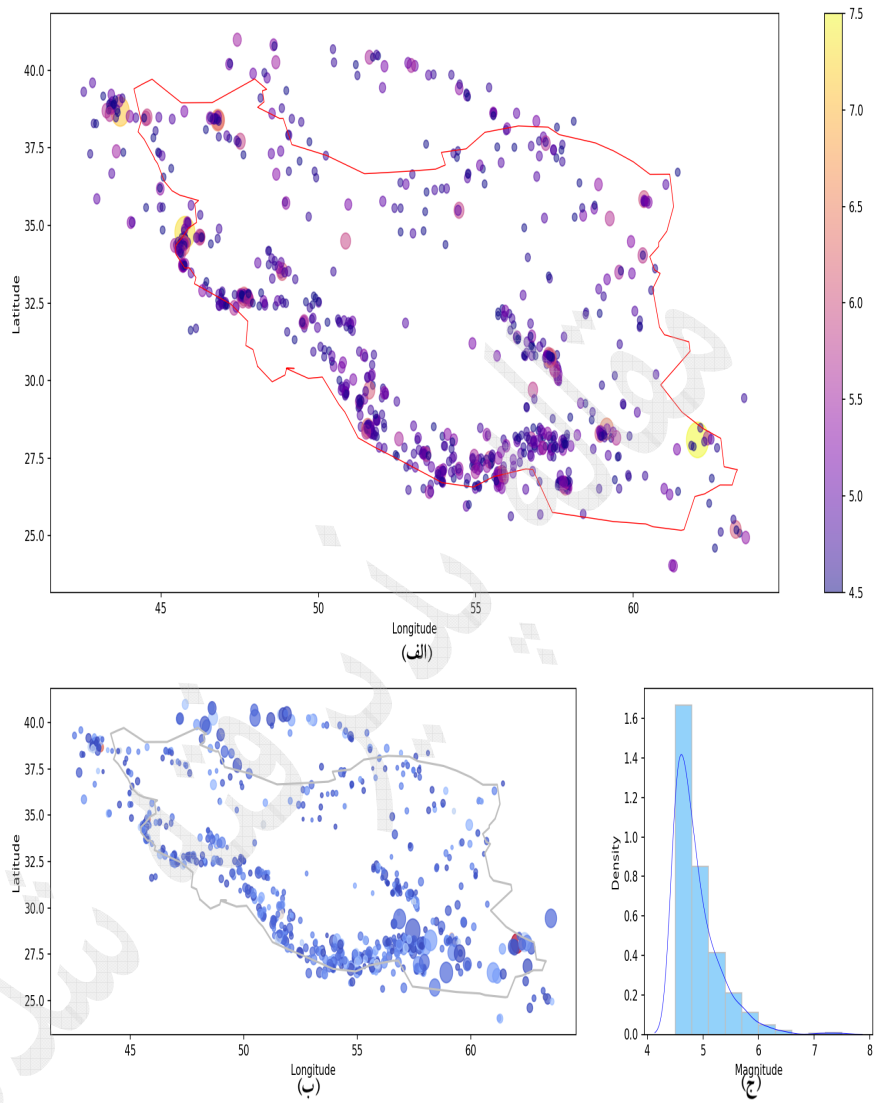
دقت کمتری برخوردار هستند. با توجه به نتایج می‌توان گفت که دقت برآورد پارامترهای مدل با رهیافت بیز مقداری با توجه به سرعت محاسبات بالاتر نسبت به روش‌های MCMC نسبتا خوب است. لازم به ذکر است که برای یک مجموعه داده شبیه‌سازی شده در این مطالعه شبیه‌سازی به‌طور متوسط رهیافت بیز مقداری ۳ الی ۴ برابر سریعتر است. همچنین نتایج روش بیز مقداری دو مدل رگرسیون فضایی چوله و گاوسی برای مشبکه‌های منظم ۵ × ۵ و ۱۰ × ۱۰ در جدول ۲ خلاصه شده‌اند. نتایج RMSE نشان می‌دهد که برآورد بیز مقداری برای مشبکه‌های منظم با حجم کم نیز عملکرد مناسبی دارد و می‌توان گفت که با افزایش حجم نمونه برآوردهای ضرایب رگرسیونی و دامنه فضایی نسبتا بهتر می‌شود.

جدول ۲: نتایج برآورد بیز مقداری برای دو مدل رگرسیون فضایی چوله و مدل گاوسی براساس ۱۰۰ مجموعه داده‌ی شبیه‌سازی شده در مشبکه‌های منظم ۵ × ۵ و ۱۰ × ۱۰

		مدل چوله				مدل گاوسی			
		۵ × ۵		۱۰ × ۱۰		۵ × ۵		۱۰ × ۱۰	
پارامتر	واقعی	برآورد	RMSE	برآورد	RMSE	برآورد	RMSE	برآورد	RMSE
β_0	۵	۵٫۹۶۱	۰٫۵۱۱	۵٫۶۱۵	۰٫۳۶۵	۶٫۵۲۲	۰٫۹۰۳	۵٫۹۸۲	۰٫۵۳۳
β_1	۲	۱٫۶۶۲	۰٫۲۸۸	۱٫۷۱۲	۰٫۱۲۱	۲٫۹۵۵	۰٫۷۱۲	۲٫۸۶۸	۰٫۶۰۹
σ^2	۱	۱٫۴۹۷	۱٫۰۸۷	۱٫۴۳۷	۰٫۹۲۲	۱٫۹۸۵	۱٫۰۴۹	۱٫۸۹۹	۱٫۰۵۵
φ	۶	۷٫۵۲۲	۱٫۹۵۶	۶٫۸۶۵	۰٫۴۹۱	۴٫۵۸۷	۱٫۸۸۵	۴٫۶۶۹	۱٫۸۹۷
λ	۳	۲٫۴۷۸	۰٫۰۴۵	۲٫۴۴۳	۰٫۰۳۹	–	–	–	–

۵ تحلیل داده‌های زمین لرزه‌ای ایران

در این بخش به تحلیل داده‌های زمین لرزه‌ای کشور طی ۱۵ سال اخیر بر اساس مدل رگرسیون فضایی چوله و روش‌های معرفی شده می‌پردازیم. داده‌های زمین لرزه‌ای که توسط مرکز لرزه‌نگاری کشور (مؤسسه‌ی ژئوفیزیک تهران) طی سال‌های ۲۰۰۶ الی ۲۰۲۰ ثبت شده است برای تحلیل در این مقاله در نظر گرفته شده است. داده‌های زمین لرزه‌ای شامل زمان وقوع زلزله (Date)، عرض جغرافیایی (Lat)، طول جغرافیایی (Long)، عمق (Depth)، شهر (City)، استان (Province) و بزرگی زمین لرزه (Mg) است. نقشه‌ی نقاط بزرگی زمین لرزه‌های بالای ۴٫۵ در شکل ۲(الف) به همراه هیستوگرام آن‌ها (شکل ۲(ج)) و عمق زمین لرزه‌ها (شکل ۲(ب)) رسم شده‌اند. با توجه به شکل ۲ می‌توان همبستگی مکانی (فضایی) داده‌ها را مشاهده کرد. با توجه به نقشه عمق زمین لرزه‌ها بیشتر زمین لرزه‌ها عمیق در جنوب کشور رخ داده است. در مطالعات زمین لرزه‌ای، بررسی عمق زمین لرزه‌ها از اهمیت زیادی برخوردار است چرا که هرچه عمق زمین لرزه‌ها کم‌تر باشد شدت خسارت آن بیش‌تر می‌شود. همچنین چولگی در نمودار هیستوگرام بزرگی زمین لرزه‌ها دیده می‌شود. در ادامه مدل رگرسیون فضایی چوله (۲) به‌منظور ارایه نقشه پیشگویی بزرگی زمین لرزه‌ها در کل نقاط کشور برای مشاهده نقاط پرخطر کشور به‌کار گرفته می‌شود. با در نظر گرفتن بزرگی زمین لرزه به عنوان متغیر پاسخ (Mg_i) برای هر موقعیت s_i و عمق (D_i)، طول جغرافیایی $(Long_i)$

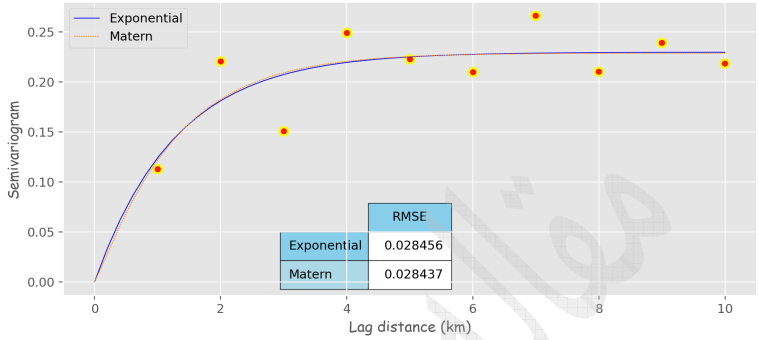


شکل ۲: (الف) نقشه موقعیت زمین لرزه‌های کشور بیشتر از ۴٫۵ طی پانزده سال اخیر، (ب) نقشه عمق زمین لرزه‌ها و (ج) نمودار هیستوگرام آن‌ها

و عرض جغرافیایی (Lat_i) زمین لرزه‌ها به عنوان متغیرهای توضیحی مدل رگرسیون فضایی به‌صورت

$$Mg_i = \beta_۰ + \beta_۱ D_i + \beta_۱ Long_i + \beta_۲ Lat_i + \psi_i,\ i = ۱, ۲, \ldots, \qquad (۶)$$

تعریف می‌شود، که در آن ψ_i مولفه iام عبارت σCZ در مدل (۲) است. در این مدل به دلیل شناسایی پذیر بودن پارامتر چولگی به‌صورت $\lambda_i = \lambda$ در نظر گرفتیم. برای تعیین مدل همبستگی فضایی داده‌ها از



شکل ۳: نمودار تغییرنگار تجربی داده‌های زمین لرزه‌های کشور

تغییرنگار تجربی داده‌ها استفاده کرده‌ایم. شکل ۳ نمودار تغییرنگار تجربی به همراه برازش دو مدل تغییرنگار نمایی و ماترن را نشان می‌دهد. همانطور که مشاهده می‌شود مدل نمایی به خوبی برازش شده است بنابراین این مدل برای تعیین ساختار همبستگی فضایی ماتریس Σ_n در نظر گرفته می‌شود. سپس رهیافت بیز مقداری به همراه روش‌های برآورد پیشنهادی در بخش قبل روی داده‌ها پیاده‌سازی شد. بدلیل پیچیدگی توزیع پسین در روش بیز سلسله مراتبی لازم است از روش‌های MCMC استفاده کرد که این روش‌ها نسبت به بیز مقداری زمان‌بر بود به طوری که در این مطالعه زمان اجرا با یک سیستم لب‌تاپ شخصی معمولی (Intel Core i5 CPU, 4 GB RAM) به‌طور متوسط حدود ۵ برابر بیشتر طول می‌کشد و نتایج برازش مدل هم مشابه است.

مدل رگرسیون فضایی با اثرات فضایی گاوسی را به عنوان مدل رقیب پایه برای مقایسه با مدل رگرسیون فضایی چوله (۶) در نظر گرفته‌ایم. مدل رگرسیون فضایی گاوسی حالت خاصی از مدل چوله است زمانی که پارامتر چولگی λ صفر در نظر گرفته شود. نتایج برآورد پارامترهای دو مدل به روش بیز مقداری در جدول ۳ ارائه شده است که با توجه به فاصله‌های قابل باور بیزی در سطح ۹۵ درصد، پارامترهای β_2 و β_3 معنی‌دار نیستند یعنی طول و عرض جغرافیایی تاثیر معنی‌داری روی بزرگی زمین لرزه‌ها ندارد. در واقع از شمال به جنوب و از شرق به غرب کشور روند معنی‌داری از نظر بزرگی زمین لرزه‌ها مشاهده نمی‌شود.

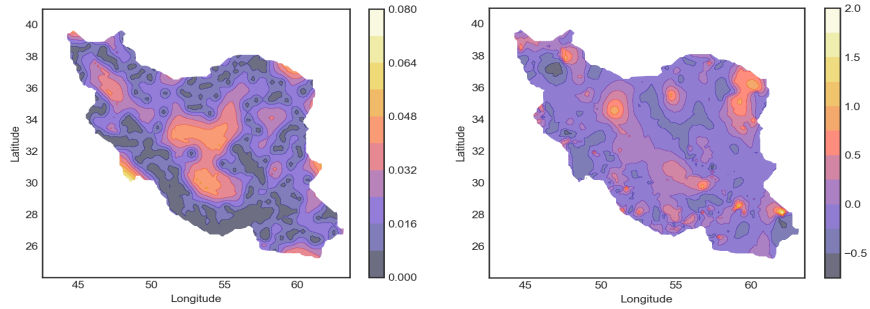
جدول ۳: نتایج برآورد بیز مقداری پارامترهای مدل رگرسیون فضایی برای داده‌های زمین لرزه‌ای کشور

	مدل چوله				مدل گاوسی			
پارامتر	برآورد	انحراف استاندارد	چندک ۰٫۰۲۵	چندک ۰٫۹۷۵	برآورد	انحراف استاندارد	چندک ۰٫۰۲۵	چندک ۰٫۹۷۵
β_0	۳٫۶۱۹	۰٫۳۷۲	۲٫۸۹۱	۴٫۳۴۸	۴٫۵۲۳	۰٫۴۷۲	۳٫۰۶۳	۵٫۷۵۲
β_1	۰٫۰۷۵	۰٫۰۱۰	۰٫۰۵۵	۰٫۰۹۵	۰٫۰۸۹	۰٫۰۳۱	۰٫۰۶۲	۰٫۱۱۴
β_2	−۰٫۰۲۳	۰٫۰۴۹	−۰٫۱۱۹	۰٫۰۷۲	−۰٫۰۱۶	۰٫۰۷۲	−۰٫۲۱۱	۰٫۱۵۲
β_3	۰٫۰۴۸	۰٫۰۶۴	−۰٫۰۷۶	۰٫۱۷۴	۰٫۰۳۹	۰٫۰۹۳	−۰٫۱۴۵	۰٫۱۹۸
σ	۰٫۲۳۱	۰٫۵۶۱	۰٫۰۸۱	۱٫۱۴۱	۰٫۵۶۲	۰٫۹۸۴	۰٫۰۱۱	۲٫۰۹۳
φ	۳٫۸۶۱	۱٫۰۱۷	۰٫۲۴۳	۶٫۸۶۴	۶٫۱۵۱	۲٫۲۱۵	۰٫۰۷۸	۹٫۴۵۳
λ	۲٫۱۳۱	۱٫۵۲۳	۰٫۶۸۱	۴٫۴۷۲	–	–	–	–
$RMSPE$	۰٫۰۳۲۴۸				۰٫۰۸۷۹۴			

اما عمق لرزه‌ها تاثیر معنی‌داری روی بزرگی لرزه‌ها طی پانزده سال اخیر داشته است چون برآورد فاصله‌ای ۹۵ درصد شامل صفر نیست. با استفاده از روش اعتبار سنجی متقابل دو مدل رگرسیون فضایی چوله و گاوسی مورد مقایسه قرار گرفتند. برای این منظور از معیار ریشه دوم میانگین توان دوم خطای پیشگویی (RMSPE) به‌صورت

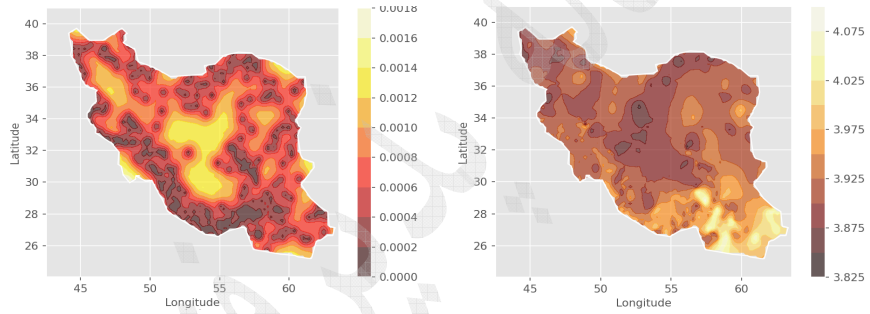
$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{(i)})},$$

استفاده شد، که در آن $\hat{y}_{(i)}$ مقدار پیشگویی iامین پاسخ بر اساس تمام مشاهدات بدون حضور iامین مشاهده است. نتایج RMSPE برای دو مدل نیز در جدول ۳ ارائه شده است که نشان می‌دهد مدل رگرسیون فضایی چوله نسبت به مدل گاوسی برای این داده‌ها بهتر است. نقشه پیشگویی فضایی و دقت پیشگویی برای اثرات فضایی مدل رگرسیون فضایی چوله در شکل ۴ رسم شده است که نقشه پیشگویی نواحی همبسته فضایی مشابه و نواحی با اثرات فضایی بیشتر را نشان می‌دهد. با توجه به نقشه پیشگویی شکل ۴ (ب)، نواحی مرکزی کشور دقت پیشگویی فضایی کمتری دارند. همچنین نقشه پیشگویی فضایی برای متغیر پاسخ (بزرگی زمین لرزه‌ها) به همراه دقت آن در شکل ۵ ارائه شده است که نقاط پر خطر از نظر بزرگی زمین لرزه‌ها را روی نقشه ایران نشان می‌دهد. در واقع نقاط هم‌رنگ از نظر میزان بزرگی لرزه



(ب) (الف)

شکل ۴: (الف) نقشه پیشگویی اثرات فضایی و (ب) دقت پیشگویی آن‌ها برای مدل رگرسیون چوله



(ب) (الف)

شکل ۵: (الف) نقشه پیشگویی فضایی و (ب) دقت پیشگویی داده‌های زمین لرزه‌های کشور طی پانزده سال اخیر

مشابه هستند و به نوعی همبستگی مکانی داده‌ها را نشان می‌دهد. همچنین نقاط جنوب و جنوب شرق کشور بیشتر در خطر وقوع زمین لرزه‌های بزرگ هستند. نقشه واریانس پیشگویی شکل ۵(ب) هم نشان می‌دهد که دقت پیشگویی در نقاط مرکزی و برخی نقاط شمال غربی نسبت به بقیه نقاط کمتر است.

بحث و نتیجه‌گیری

در این مقاله یک مدل رگرسیون چوله براساس زیر رده منعطفی از توزیع چوله نرمال بسته برای مدل‌بندی داده‌های فضایی چوله ارائه شد. با توجه به پیچیدگی تابع درست‌نمایی مدل، رهیافت بیز سلسله مراتبی به

کار گرفته شد و برای تقریب توزیع پسین از رهیافت بیز مقداری به‌خاطر سرعت محاسبات بالا به عنوان جایگزینی مناسب برای روش‌های مونت کارلویی استفاده شد. در یک مطالعه شبیه‌سازی مدل رگرسیون فضایی پیشنهادی مورد ارزیابی قرار گرفت و نتایج شبیه‌سازی برای ۱۰۰ مجموعه داده در دامنه‌های فضایی مختلف نشان دادند که مدل پیشنهادی با رهیافت بیز مقداری عملکرد مناسبی داشت و در مقایسه با مدل گاوسی به عنوان یک مدل پایه به‌طور متوسط $RMSE$ برآوردهای آن کمتر بودند. همچنین سرعت محاسبات رهیافت بیز مقداری در مقایسه با الگوریتم‌های MCMC به‌طور متوسط ۳ الی ۴ برابر بیشتر بود (از نرم‌افزار پایتون ۳ روی یک سیستم لب‌تاپ شخصی i5 CPU با 4GB RAM استفاده شده است). در نهایت مدل پیشنهادی روی داده‌های واقعی زمین لرزه‌ای کشور طی ۱۵ سال اخیر اجرا گردید و مدل رگرسیون فضایی گاوسی نیز به‌عنوان یک مدل رقیب به داده‌ها برازش شد. نتایج اعتبارسنجی متقابل نشان داد که مدل پیشنهادی عملکرد بهتری دارد. فاصله‌های قابل باور بیزی ۹۵ درصد برای بررسی تاثیر متغیرهای مدل محاسبه شدند که بیان کردند عمق زمین لرزه‌ها تاثیر معنی‌داری در بزرگی زمین لرزه‌های ثبت شده طی ۱۵ سال اخیر داشته است. در نهایت نقشه پیشگویی فضایی روی کل ناحیه فضایی رسم شد که نشان داد بخش‌هایی از جنوب و شرق کشور بیشتر از بقیه نواحی در معرض زمین لرزه‌های به بزرگی ۴/۵ و بیشتر هستند. همچنین نقشه‌های دقت پیشگویی نشان دادند که پیشگویی فضایی بخشی‌هایی از نقاط مرکزی و شمال غربی کشور از دقت کمتری برخوردار هستند.

تقدیر و تشکر

نویسندگان مقاله از پیشنهادات و نظرات ارزنده داوران محترم، ویرایش ادبی سردبیر، هیئت تحریریه و ویراستار مجله که در بهبود کیفیت مقاله بسیار مؤثر واقع شد، تقدیر و تشکر می‌نمایند.

مراجع

کریمی، الف. و حسینی، ف. (۱۴۰۰)، معرفی یک میدان تصادفی مانای چوله گاوسی، مجله علوم آماری، ۱۵(۲)، ۵۴۹-۵۶۶.

Anselin, L. (1990), Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis, *Journal of Regional Science*, **30**, 185–207.

Basu, S. and Reinsel, G.C. (1994), Regression Models with Spatially Correlated Errors, *Journal of the American Statistical Association*, **89**, 88–99.

Cabral, R., Bolin, D. and Rue, H., (2024), Fitting Latent Non-Gaussian Models Using Variational Bayes and Laplace Approximations, *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2023.2296704.

Cressie, N. (1993), *Statistics for Spatial Data*, Revised Edition, John Wiley, New York.

Gonzalez-Farias, G., Dominguez-Molina, A. and Gupta, A. K. (2004), The Closed Skew Normal Distribution. In: *Genton M. G., ed. Skew-elliptical distributions and their applications: A journey beyond normality. Boca Raton, FL: Chapman and Hall CRC,* 2542.

Karimi, O. (2023), A Hamiltonian Monte Carlo EM algorithm for Generalized Linear Mixed Models with Spatial Skew Latent Variables, *Statistical Paper*, https://doi.org/10.1007/s00362-023-01419-y.

Karimi, O. and Mohammadzadeh, M. (2012), Bayesian Spatial Regression Models with Closed Skew Normal Correlated Errors and Missing, *Statistical Papers,* **53(1)**, 205-218.

Karimi, O., Omre, H. and Mohammadzadeh, M. (2010), Bayesian Closed-skew Gaussian Inversion of Seismic AVO Data for Elastic Material Properties, *Geophysics,* **75**, R1-R11.

Márquez-Urbina, O.U. and González-Farías, G. (2022), A Flexible Special Case of the CSN for Spatial Modeling and Prediction, *Spatial Statistics*, **47**, 100556.

Kim, H.M. and Mallick, B.K. (2004), A Bayesian Prediction using the Skew Gaussian Distribution, *Journal of Statistical Planning and Inference*, **120**, 85–101.

Lee, J. and Huang, Y. (2022), Covid-19 Impact on US Housing Markets: Evidence from Spatial Regression Models, *Spatial Economic Analysis*, **17**:3, 395-415.

Oh, M., Shina, D.W. and Kim, H.J. (2002), Bayesian Analysis of Regression Models with Spatially Correlated Errors and Missing Observations, *Computational Statistics and Data Analysis*, **39**, 387–400.

Ormerod, J. T. and Wand, M. P. (2010), Explaining Variational Approximations, *The American Statistician*, **64**(2), 140–153.

Tan, L. S. and Nott, D. J. (2013), Variational Inference For Generalized Linear Mixed Models Using Partially Noncentered Parametrizations, *Statistical Science,* **28**, 168–188.

Zhang, Q., Shan, L., Xie, L., Shaowu G. and Hongye S. (2023), Variational Bayesian State Space Model for Dynamic Process Fault Detection, *Journal of Process Control*, **124**, 129-141.

Variational Bayesian Analysis of Skew Spatial Regression Model Based on a flexible Subclass of Closed Skew-Normal Distribution

Omid Karimi and Fatemeh Hosseini

Department of statistics, Semnan University, Semnan, Iran.

Abstract: Spatial regression models are used to analyze quantitative spatial responses based on linear and non-linear relationships with explanatory variables. Usually, the spatial correlation of responses is modeled with a Gaussian random field based on a multivariate normal distribution. However, in practice, we encounter skewed responses, which are analyzed using skew-normal distributions. Closed skew-normal distribution is one of the extended families of skew-normal distributions, which has similar properties to normal distributions. This article presents a hierarchical Bayesian analysis based on a flexible subclass of closed skew-normal distributions. Given the time-consuming nature of Monte Carlo methods in hierarchical Bayes analysis, we have opted to use the variational Bayes approach to approximate the posterior distribution. This decision was made to expedite the analysis process without compromising the accuracy of our results. Then, the proposed model is implemented and analyzed based on the real earthquake data of Iran.

Keywords: Hierarchical Bayesian approach, Optimal variational distribution, Variogram, Spatial data.

Mathematics Subject Classification (2010): 60G15, 62M30.