تحلیل بیزی متغیرهای پنهان در مدل‌های آمیخته خطی تعمیم‌یافته فضایی با میدان تصادفی مانای چوله گاوسی

فاطمه حسینی، امید کریمی

گروه آمار، دانشگاه سمنان

چکیده: برای مدل‌بندی داده‌های رسته‌ای فضایی از مدل‌های آمیخته خطی تعمیم‌یافته فضایی استفاده می‌شود که در این مدل‌ها اغلب متغیرهای پنهان که بیان‌گر همبستگی فضایی هستند، با یک میدان تصادفی گاوسی مدل‌بندی می‌شوند. عدم برقراری فرض گاوسی باعث تاثیر روی دقت پیش‌گویی‌ها و برآورد پارامترهای مدل می‌شود. در این مقاله با استفاده یک میدان تصادفی چوله گاوسی مانا و به‌کارگیری یک رهیافت بیزی تقریبی، مدل‌های آمیخته خطی تعمیم‌یافته فضایی مدل‌بندی و برآورد می‌شوند. در یک مثال شبیه‌سازی کارایی مدل و رهیافت بیزی تقریبی بررسی و بر روی یک مثال واقعی پیاده‌سازی می‌شود.

واژه‌های کلیدی: مدل‌های آمیخته خطی تعمیم‌یافته فضایی، تحلیل بیزی تقریبی، میدان تصادفی، میدان تصادفی چوله گاوسی مانا.

آدرس الکترونیکی نویسنده مسئول مقاله: فاطمه حسینی، fatemeh.hoseini@semnan.ac.ir
کد موضوع‌بندی ریاضی (۲۰۱۰): 62J12, 62F15.

۱ مقدمه

برای بررسی داده‌های رسته‌ای فضایی اغلب از مدل های آمیخته خطی تعمیم‌یافته فضایی استفاده می‌شود و همبستگی فضایی با یک متغیر پنهان وارد مدل می‌شود که در اکثر مطالعات با یک میدان تصادفی گاوسی مدل‌بندی می‌شود. یکی از مسائل مهم در این مدل‌ها برآورد پارامترهای مدل و پیش‌گویی متغیرهای پنهان فضایی است. چون اغلب در این مدل‌ها متغیرهای پاسخ متعلق به یک خانواده از توزیع‌ها در نظر گرفته می‌شوند و به دلیل وجود متغیرهای پنهان در این مدل‌ها، برآورد پارامترهای مدل به راحتی و از روش‌های معمول میسر نیست. معمولا در این مدل‌ها متغیرهای پنهان فضایی با میدان تصادفی گاوسی مدل‌بندی می‌شوند (محمدزاده ، ۱۳۹۴). دوری از پذیره نرمال‌بودن روی دقت برآورد مدل و پیش‌گویی‌ها تاثیر می‌گذارد.

در ادامه، برای مدل‌بندی متغیرهای پنهان فضایی از یک میدان تصادفی چوله مانا که در برگیرنده میدان تصادفی گاوسی است، استفاده می‌شود و یک مدل آمیخته خطی تعمیم‌یافته فضایی با متغیرهای پنهان چوله جدید معرفی و با به‌کار بردن یک رهیافت بیزی تقریبی به برآورد و تحلیل این مدل پرداخته می‌شود. اولین بار کیم و مالیک (۲۰۰۴) میدان تصادفی چوله گاوسی را برای تحلیل داده‌های فضایی چوله معرفی نمود. براساس توزیع چوله گاوسی بسته آلارد و ناویو (۲۰۰۷)، کریمی و همکاران (۲۰۱۰) و کریمی و محمدزاده (۲۰۱۱، ۲۰۱۲) به معرفی میدان‌های تصادفی چوله گاوسی پرداختند. در مطالعات اشاره شده میدان‌های چوله گاوسی معرفی شده خوش‌تعریف نیستند و دارای مشکلاتی هستند که در بخش‌های بعد بیان می‌شوند. ریمستاد و امره (۲۰۱۴) با استفاده از میدان تصادفی آلارد و ناویو (۲۰۰۷)، یک میدان تصادفی چوله گاوسی تقریبا مانا تعریف نمودند و کریمی و حسینی (۱۴۰۰) با اصلاح میدان تصادفی چوله گاوسی تقریبا مانا ریمستاد و امره (۲۰۱۴) یک میدان تصادفی چوله گاوسی مانا معرفی نمودند. یکی از اهداف این مقاله استفاده از میدان تصادفی چوله گاوسی مانا معرفی شده توسط کریمی و حسینی (۱۴۰۰) برای مدل‌بندی متغیرهای پنهان در مدل‌های آمیخته خطی تعمیم‌یافته فضایی است. حسینی و محمدزاده (۲۰۱۲) و حسینی و کریمی (۲۰۲۰، ۲۰۲۱) در مدل‌های آمیخته خطی تعمیم‌یافته فضایی، استفاده از توزیع چوله گاوسی بسته برای متغیرهای پنهان فضایی را پیشنهاد کردند و یک تحلیل درست‌نمایی مرکب

برای داده‌های فضایی با متغیرهای پنهان چوله بیان کردند که میدان تصادفی چوله گاوسی تقریبا مانای ریمستاد و امره (۲۰۱۴) را به‌کار گرفتند. حسینی و کریمی (۱۴۰۱) و کریمی (۲۰۲۴) با به کار بردن روش‌های مونت‌کارلویی همیلتونی و الگویتم‌های بیشینه‌سازی امید ریاضی یک الگوریتم جدید برای به‌دست آوردن برآورد ماکسیمم درست‌نمایی پارامترهای این مدل‌ها معرفی نمودند و روی داده‌های واقعی پیاده‌سازی کردند.

در این مقاله از میدان تصادفی چوله گاوسی مانا برای مدل‌بندی متغیرهای پنهان فضایی در مدل‌های آمیخته خطی تعمیم‌یافته فضایی استفاده و سپس با به‌کار گرفتن یک رهیافت بیزی تقریبی، این مدل‌ها برآورد می‌شوند. در بخش ۲ یک میدان تصادفی چوله گاوسی تقریبا مانا و مانا ارائه می‌شوند. در بخش ۳ مدل و در بخش ۴ تحلیل بیزی مدل مورد بررسی قرار می‌گیرد و در بخش ۴ با یک مطالعه شبیه‌سازی کارایی و دقت مدل بررسی می‌شود. در بخش ۶ مدل و رهیافت ارائه شده روی یک مجموعه داده واقعی پیاده‌سازی و در نهایت بحث و نتیجه‌گیری بیان می‌شود.

۲ میدان تصادفی چوله گاوسی

بردار تصادفی n بعدی $\boldsymbol{x}$ با تابع چگالی

$$f_{n,q}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Gamma},\boldsymbol{\nu},\boldsymbol{\Delta}) = k\phi_n(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\ \Phi_q\big(\boldsymbol{\Gamma}(\boldsymbol{x}-\boldsymbol{\mu});\boldsymbol{\nu},\boldsymbol{\Delta}\big), \qquad (۱)$$

دارای توزیع چوله گاوسی بسته چند متغیره با پارامترهای $\boldsymbol{\mu}$، $\boldsymbol{\Sigma}$، $\boldsymbol{\Gamma}$، $\boldsymbol{\nu}$ و $\boldsymbol{\Delta}$ است، که در آن $\Phi_q\big(\boldsymbol{\Gamma}(\boldsymbol{x}-\boldsymbol{\mu});\boldsymbol{\nu},\Delta\big)$ تابع توزیع تجمعی q متغیره گاوسی با بردار میانگین $\boldsymbol{\nu}$ و ماتریس کوواریانس $\boldsymbol{\Delta}$، $\boldsymbol{\mu}$ بردار پارامتر مکان، $\boldsymbol{\Sigma}$ ماتریس مقیاس $n \times n$ معین مثبت، عناصر ماتریس $\boldsymbol{\Gamma}_{q\times n}$ پارامترهای چولگی هستند و همچنین $k = [\Phi_q(0;\boldsymbol{\nu},\boldsymbol{\Delta}+\boldsymbol{\Gamma\Sigma\Gamma}^\top)]^{-1}$. به‌طور خلاصه این توزیع به‌صورت $CSN_{n,q}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Gamma},\boldsymbol{\nu},\boldsymbol{\Delta})$ نمایش داده می‌شود. وقتی عناصر ماتریس $\boldsymbol{\Gamma}$ صفر در نظر گرفته شوند، تابع چگالی گاوسی حاصل می‌شود. گشتاور مرتبه اول (امید ریاضی) توزیع چوله گاوسی بسته به‌صورت $\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu} + \boldsymbol{\Sigma\Gamma}^\top\boldsymbol{\Psi}$ به‌دست می‌آید که در آن $\boldsymbol{\Psi} = \frac{\Phi_q^*(0;\boldsymbol{\nu},\boldsymbol{\Delta}+\boldsymbol{\Gamma\Sigma\Gamma}^\top)}{\Phi_q(0;\boldsymbol{\nu},\boldsymbol{\Delta}+\boldsymbol{\Gamma\Sigma\Gamma}^\top)}$ است و برای ماتریس

معین مثبت $\boldsymbol{\Omega}$، $\Phi_q^*(\boldsymbol{s};\boldsymbol{\nu},\boldsymbol{\Omega}) = [\nabla_s \Phi_q(\boldsymbol{s};\boldsymbol{\nu},\boldsymbol{\Omega})]^\top$ که در آن $\nabla_s = (\frac{\partial}{\partial s_1}, \cdots, \frac{\partial}{\partial s_q})^\top$ است (گنزالس و همکاران، ۲۰۰۴).

۱.۲ میدان تصادفی چوله گاوسی بسته تقریبا مانا

ریمستاد و امره (۲۰۱۴) میدان تصادفی چوله گاوسی بسته را با تعمیم میدان تصادفی چوله گاوسی تعریف شده توسط آلارد و ناویو (۲۰۰۷) به صورت زیر تعریف کردند، فرض کنید

$$\boldsymbol{U}(\boldsymbol{s}) = \{(U_1(\boldsymbol{s}), U_2(\boldsymbol{s}))^\top, \boldsymbol{s} \in \mathcal{D} \subseteq R^d\},$$

میدان تصادفی گاوسی دو متغیره باشد و $\boldsymbol{U}_2 = [U_2(s'_1), \cdots, U_2(s'_q)]$ که در آن q ثابت و متناهی است. آن‌گاه میدان تصادفی چوله گاوسی بسته به‌صورت $\{\boldsymbol{X}(\boldsymbol{s}) = [\boldsymbol{U}_1(s) | \boldsymbol{U}_2 \leq 0\}$ تعریف می‌شود، اگر برای هر مجموعه متناهی $(\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n)$، $\boldsymbol{X} = (X(\boldsymbol{s}_1), \cdots, X(\boldsymbol{s}_n))^\top$ دارای توزیع چوله گاوسی بسته باشد. آن‌ها نشان دادند که وقتی موقعیت‌های $(\boldsymbol{s}'_1, \cdots, \boldsymbol{s}'_q)$ به اندازه کافی از مرزها دور، $n = q$ و $(\boldsymbol{s}'_1, \cdots, \boldsymbol{s}'_q) = (\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n)$ باشند و با در نظر گرفتن شکل توزیع چوله گاوسی بسته برای $\boldsymbol{X}$ به‌صورت

$$CSN_{n,n}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{C}_\varphi, \frac{\gamma}{\sigma}\mathbf{I_n}, \nu \mathbf{1_n}, (1-\gamma^2)\mathbf{I_n}), \tag{2}$$

که در آن $\boldsymbol{\mu}$ پارامتر مکان، σ^2 پارامتر مقیاس، $1 > |\gamma|$ پارامتر چولگی، $\boldsymbol{C}_\varphi$ ماتریس همبستگی مانا، $\mathbf{I_n}$ ماتریس واحد و $\mathbf{1_n} \in \mathbb{R}^n$ برداری با عناصر یک است، آن‌گاه $\boldsymbol{X}(\boldsymbol{s})$ تقریبا مانا است.

۲.۲ میدان تصادفی چوله گاوسی بسته مانا

کریمی و حسینی (۱۴۰۰) نشان دادند در میدان تصادفی تعریف‌شده توسط ریمستاد و امره (۲۰۱۴) توزیع حاشیه‌ای X_j به ماتریس همبستگی $\boldsymbol{C}_\varphi$ وابسته است که مربوط به کل شبکه فضایی $(\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n)$ است، اما ساختار کلی چگالی X_j یک توزیع چوله گاوسی بسته است. برای رفع این مشکل یعنی

برقراری شرط سازگاری حاشیه‌ای و تعریف یک میدان تصادفی چوله گاوسی مانا، آن‌ها یک تحقق nتایی از میدان تصادفی چوله گاوسی تقریبا مانا به‌صورت $\boldsymbol{X} = (X(\boldsymbol{s}_1), \dots, X(\boldsymbol{s}_n))^\top$ مطابق رابطه (۲) با تفکیک‌های $\boldsymbol{X}_{n\times 1} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}$، $\boldsymbol{C}_\varphi = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$ در نظر گرفتند، که در آن $\boldsymbol{X}_1 \in R^{n_1}$، $\boldsymbol{X}_2 \in R^{n_2}$ و $n_1 + n_2 = n$ و ماتریس همبستگی فضایی $\boldsymbol{C}_\varphi$ تفکیک‌های متناظر با بردارهای $\boldsymbol{X}_1$ و $\boldsymbol{X}_2$ از ماتریس $\boldsymbol{X}$ هستند. سپس نشان دادند که تابع مولد گشتاور حاشیه‌ای $\boldsymbol{X}_1$ به ماتریس همبستگی $\boldsymbol{C}_\varphi$ برای کل ناحیه فضایی وابسته است و شرط سازگاری حاشیه‌ای برقرار نیست. برای این‌که شرط سازگاری حاشیه‌ای برقرار شود، کریمی و حسینی (۱۴۰۰) پیشنهاد نمودند که پارامترهای یک تحقق از میدان تصادفی چوله گاوسی تقریبا مانا به‌صورت

$$\boldsymbol{X} \sim CSN_{n,n}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{C}_\varphi, \frac{\gamma}{\sigma} \boldsymbol{C}_\varphi^{-\frac{1}{2}}, \nu \boldsymbol{1}_n, (1-\gamma^2)\mathbf{I}_n), \tag{3}$$

اصلاح شوند، که در آن $\boldsymbol{C}_\varphi^{-\frac{1}{2}}$ عکس ریشه دوم ماتریس $\boldsymbol{C}_\varphi$ است. با محاسبه تابع مولد گشتاور توزیع حاشیه‌ای $\boldsymbol{X}_1$ نشان دادند که توزیع حاشیه‌ای $\boldsymbol{X}_1$ برای میدان تصادفی چوله گاوسی تعریف‌شده در رابطه (۳) دارای بعد n_1 و از خانواده توزیع‌های چوله گاوسی بسته است، پس دارای شرط سازگاری حاشیه‌ای و یک میدان تصادفی خوش‌تعریف است.

۳ معرفی مدل

در مدل‌های خطی تعمیم‌یافته چون متغیر پاسخ لزوما کمی نیست و می‌تواند کیفی هم باشد، برای متغیر پاسخ یک خانواده از توزیع‌های نمایی که شامل اکثر توزیع‌های آماری گسسته و پیوسته از جمله توزیع گاوسی است، در نظر گرفته می‌شود (مک‌کلاح و نلدر، ۱۹۸۹). در این مقاله نیز متغیرهای پاسخ فضایی متعلق به خانواده از توزیع‌های نمایی فرض می‌شوند. فرض کنید $\boldsymbol{Y}^\top = (y_1, \cdots, y_k)$ بردار متغیرهای پاسخ فضایی در موقعیت‌های $\{s_1, \cdots, s_k\}$، $k \leq n$ باشد که متعلق به یک خانواده از توزیع‌های نمایی

هستند. پس چگالی $\pi(\boldsymbol{y}|\boldsymbol{x})$، به‌طور شرطی مستقل و از یک خانواده از توزیع‌های نمایی به‌صورت

$$\pi(y_i|x_i) = \exp\{y_i x_i - b(x_i) + c(y_i)\}, \quad i = 1, \ldots, k,$$

فرض می‌شود، که مطابق مدل‌های خطی تعمیم یافته $\mathrm{E}(Y_i|x_i) = g^{-1}(x_i)$ است، که در آن $g(\cdot)$ یک تابع پیوند معلوم است. همچنین فرض کنید $\boldsymbol{X} = (x_1, \cdots, x_n)^\top$ یک تحقق از میدان تصادفی چوله گاوسی بسته در n موقعیت $\{s_1, \cdots, s_n\}$ با دامنه $\mathcal{D} \subseteq \mathbb{R}^n$ و دارای توزیع چوله گاوسی به‌صورت (۳) با پارامتر مکان $\boldsymbol{\mu} = \boldsymbol{H}\boldsymbol{\beta}$ باشد، که در آن $\boldsymbol{H}$ ماتریسی با بعد $n \times (p+1)$ شامل متغیرهای کمکی و $\boldsymbol{\beta}$ بردار پارامترهای رگرسیونی است. برای ماتریس $\boldsymbol{C}_\varphi$ ساختار همبستگی فضایی همسانگرد فرض می‌شود. از رابطه (۱) تابع چگالی $\boldsymbol{X}$ به‌صورت

$$\pi(\boldsymbol{x}|\boldsymbol{\eta}) = \phi_n(\boldsymbol{x}; \boldsymbol{H}\boldsymbol{\beta}, \sigma^2 \boldsymbol{C}_\varphi) \frac{\Phi_n(\frac{\gamma}{\sigma} C_\varphi{}^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\beta}), \nu \mathbf{1}_n, (1-\gamma^2)\mathbf{I}_n)}{\Phi_n(\mathbf{0}; \nu \mathbf{1}_n, \mathbf{I}_n)}, \qquad (4)$$

به‌دست می‌آید. بنابراین بردار پارامترهای مدل به‌صورت $\boldsymbol{\eta} = (\boldsymbol{\beta}^\top, \sigma, \varphi, \nu, \gamma)^\top$ در نظر گرفته می‌شود که معمولا $\boldsymbol{\nu}$ را معلوم و صفر در نظر می‌گیرند. اکنون تابع درست‌نمایی را می‌توان به‌صورت

$$\begin{aligned} L(\boldsymbol{\eta} \mid \boldsymbol{y}) = & \int \exp\{\sum_{i=1}^{k} (y_i x_i - b(x_i) + c(y_i))\} \\ & \times \exp\{-\frac{1}{2\sigma^2} (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\beta})^\top \boldsymbol{C}_\varphi{}^{-1} (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\beta})\} \\ & \times \frac{\Phi_n(\frac{\gamma}{\sigma} \boldsymbol{C}_\varphi{}^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\beta}), \nu \mathbf{1}_n, (1-\gamma^2)\mathbf{I}_n)}{\Phi_n(\mathbf{0}; \nu \mathbf{1}_n, \mathbf{I}_n)} dx, \end{aligned}$$

نوشت، که تابعی پیچیده است و دارای شکل بسته‌ای نیست. لذا به‌دست آوردن برآوردها به شکل درست‌نمایی به راحتی امکان‌پذیر نیست. در بخش بعد یک رهیافت بیزی تقریبی برای به‌دست آوردن پارامترهای مدل ارائه می‌شود.

۴ برآورد بیزی مدل

قضیه ۱. فرض کنید متغیرهای پنهان فضایی در مدل آمیخته خطی تعمیم‌یافته فضایی دارای توزیع چوله گاوسی بسته $(\boldsymbol{X}|\boldsymbol{\eta}) \sim CSN_{n,n}(\boldsymbol{H\beta}, \sigma^2\boldsymbol{C}_\varphi, \frac{\gamma}{\sigma}\boldsymbol{C}_\varphi{}^{-\frac{1}{2}}, \nu\mathbf{1_n}, (1-\gamma^2)\mathbf{I_n})$ باشد و $\boldsymbol{X}$ به‌صورت
$\boldsymbol{X} = (\boldsymbol{X}^{obs\top}, \boldsymbol{X}^{pred\top})^\top$ افراز شده باشد. همچنین فرض کنید متغیرهای پاسخ گسسته فضایی متعلق به خانواده از توزیع‌های نمایی

$$\pi(y_i|x_i) = \exp\{y_i x_i - b(x_i)\},\ i = 1, \cdots k,$$

باشند. آن‌گاه با در نظر گرفتن $A = [I_{k\times k}|\mathbf{0}_{k\times n-k}]$ ، $\boldsymbol{X}^{obs} = A\boldsymbol{X}$ و با خطی‌سازی قسمت درست‌نمایی $\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})$ حول یک مقدار ثابت $\boldsymbol{x}$، توزیع $\boldsymbol{X}|\boldsymbol{y},\boldsymbol{\eta}$ به طور تقریبی چوله گاوسی بسته با پارامترهای

$$(\boldsymbol{X}|\boldsymbol{y},\boldsymbol{\eta}) \approx CSN_{n,n}(\boldsymbol{\mu}_{x|y,\eta}, \boldsymbol{\Sigma}_{x|y,\eta}, \frac{\gamma}{\sigma}\boldsymbol{C}_\varphi{}^{-\frac{1}{2}}, \nu_{x|y,\eta}, (1-\gamma^2)\mathbf{I_n}) \qquad (5)$$

است، که در آن $\boldsymbol{\mu}_{x|y,\eta} = \boldsymbol{H\beta} + \boldsymbol{C}_\varphi\boldsymbol{R}(z(\boldsymbol{y},\boldsymbol{x}^{obs}) - \boldsymbol{AH\beta})$، برای $i = 1,\ldots,k$، $z_i(y_i,x_i) = [y_i - b'(x_i) + x_i b''(x_i)]/b''(x_i)$ خطی‌سازی درست‌نمایی در مقدار ثابت $\boldsymbol{x}$، $\boldsymbol{R} = \boldsymbol{A}^\top(\boldsymbol{AC}_\varphi\boldsymbol{A}^\top + \frac{1}{\sigma^2}\boldsymbol{P})^{-1}$، $\boldsymbol{P}$ یک ماتریس قطری با عناصر $P(i,i) = 1/b''(x_i)$، $\boldsymbol{\Sigma}_{x|y,\eta} = \sigma^2\boldsymbol{C}_\varphi(\mathbf{I_n} - \mathbf{RAC}_\varphi)$ و $\boldsymbol{\nu}_{x|y,\eta} = \nu\mathbf{1_n} - \sigma\gamma\mathbf{C}_\varphi{}^{\frac{1}{2}}\mathbf{R}(\mathbf{z}(\mathbf{y},\mathbf{x^{obs}}) - \mathbf{AH}\boldsymbol{\beta})$ است (حسینی و کریمی، ۱۴۰۱).

برای به‌دست آوردن توزیع تقریبی (۵) می‌توان ابتدا یک مقدار اولیه برای $\boldsymbol{x}^{(0)}$ در نظر گرفت و $m = 0$ قرار داده شود و برآورد $\hat{\pi}(\boldsymbol{X}|\boldsymbol{y},\boldsymbol{\eta})$ به‌صورت

$$CSN_{n,n}(\boldsymbol{\mu}_{x|y,\eta}(\boldsymbol{x}^{(m)}), \boldsymbol{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m)}), \frac{\gamma}{\sigma}\boldsymbol{C}_\varphi{}^{-\frac{1}{2}}, \nu_{x|y,\eta}(\boldsymbol{x}^{m}), (1-\gamma^2)\mathbf{I_n})$$

محاسبه و میانگین این توزیع تقریبی به عنوان مقدار جدید $\boldsymbol{x}^{(m)}$ منظور و $m = m + 1$ قرار داده شود و الگوریتم تا رسیدن به همگرایی تکرار شود. برای پارامتر $\boldsymbol{\beta}$ توزیع پیشین گاوسی به‌صورت $\boldsymbol{\beta} \sim N(\boldsymbol{a}, B)$ به‌کار گرفته می‌شود که باعث می‌شود توزیع شرطی کامل متغیرهای پنهان از خانواده توزیع چوله گاوسی بسته پیروی کند که در قضیه ۲ بیان می‌شود. با در نظر گرفتن توزیع‌های پیشین معمول برای سایر پارامترهای مدل به‌صورت $\sigma \sim IG(\alpha, \tau)$، $\varphi \sim \Gamma(\lambda, \omega)$، $\gamma \sim U(-1, 1)$ و فرض استقلال پیشین‌ها، $\pi(\boldsymbol{\eta}) = \pi(\boldsymbol{\beta})\pi(\sigma)\pi(\varphi)\pi(\gamma)$، توزیع پسین به‌صورت

$$\pi(\boldsymbol{x}, \boldsymbol{\eta}|\boldsymbol{y}) = \pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})\pi(\boldsymbol{\eta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{y}|x)\pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{\eta})}{\pi(\boldsymbol{y})} \tag{6}$$

است، بنابراین از (۶) و برآورد توزیع شرطی کامل تقریبی $\hat{\pi}(\boldsymbol{X}|\boldsymbol{y}, \boldsymbol{\eta})$ توزیع تقریبی حاشیه‌ای پسین به‌صورت $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{\eta})}{\hat{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})}$ به‌دست می‌آید. چون توزیع پسین در این خانواده پیچیده است با استفاده از الگوریتم‌های نمونه‌گیر گیبز و متروپولیس–هستینگس برآورد بیزی پارامترها به‌دست آورده می‌شوند. با در نظر گرفتن توزیع پیشین گاوسی برای پارامترهای رگرسیونی $\boldsymbol{\beta}$، ثابت می‌شود توزیع شرطی کامل برای متغیرهای پنهان شکل بسته‌ای دارد اما توزیع شرطی کامل پارامترها شکل مشخصی ندارند. در ادامه یک قضیه برای اثبات این مورد بیان می‌شود. برای سادگی نوشتاری قضیه بردار پارامترهای $\boldsymbol{\eta}$ را به‌صورت $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\theta})$ تفکیک می‌کنیم.

قضیه ۲. فرض کنید $\boldsymbol{X}|\boldsymbol{\beta}, \boldsymbol{\theta} \sim CSN_{n,n}(\boldsymbol{H}\boldsymbol{\beta}, \sigma^2\boldsymbol{C}_\varphi, \frac{\gamma}{\sigma}\boldsymbol{C}_\varphi{}^{-\frac{1}{2}}, \nu\mathbf{1_n}, (1-\gamma^2)\mathbf{I_n})$ و $\boldsymbol{\beta} \sim N(\boldsymbol{a}, \boldsymbol{B})$ باشد، آن‌گاه $\boldsymbol{X}|\boldsymbol{\theta} \sim CSN_{n,n}(\boldsymbol{H}\boldsymbol{a}, \Sigma_x, \Gamma_x, \nu\mathbf{1_n}, \Delta_x)$ است. که در آن $\Sigma_x = \sigma^2\boldsymbol{C}_\varphi + HBH^\top$، $\Gamma_x = \sigma\gamma\boldsymbol{C}_\varphi^{\frac{1}{2}}\Sigma_x^{-1}$ و $\Delta_x = \mathbf{I_n} - \sigma^2\gamma^2\boldsymbol{C}_\varphi\Sigma_x^{-1}$ است.

برهان: بدون از دست دادن کلیت مساله و برای سادگی در نوشتن اثبات، شرط روی پارامتر θ را کنار می‌گذاریم. طبق ساختار توزیع چوله گاوسی بسته می‌توان نوشت:

$$\begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix}|\boldsymbol{\beta} \sim N_{n+n}(\begin{pmatrix} H\boldsymbol{\beta} \\ \boldsymbol{\nu} \end{pmatrix}, \begin{pmatrix} \sigma^2\boldsymbol{C}_\varphi & \sigma\gamma\boldsymbol{C}_\varphi^{\frac{1}{2}} \\ \sigma\gamma\boldsymbol{C}_\varphi^{\frac{1}{2}} & \mathbf{I_n} \end{pmatrix})$$

به طوریکه $\boldsymbol{X}|\boldsymbol{\beta} \overset{d}{=} \boldsymbol{U}_1|(\boldsymbol{U}_2 < \mathbf{0}|\boldsymbol{\beta})$ است ($\overset{d}{=}$ یعنی هم‌توزیعی). همچنین طبق فرض توزیع پارامتر $\boldsymbol{\beta}$ به‌صورت $N(\boldsymbol{a}, \boldsymbol{B})$ است، بنابراین

$$\begin{aligned}
\mathrm{E}\begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix} &= \mathrm{E}\Big[\mathrm{E}\Big(\begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix}|\boldsymbol{\beta}\Big)\Big] = \mathrm{E}\Big[\begin{pmatrix} H\boldsymbol{\beta} \\ \boldsymbol{\nu} \end{pmatrix}\Big] = \begin{pmatrix} H\boldsymbol{a} \\ \boldsymbol{\nu} \end{pmatrix} \\
\mathrm{Var}\begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix} &= \mathrm{E}\Big[\mathrm{Var}\Big(\begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix}|\boldsymbol{\beta}\Big)\Big] + \mathrm{Var}\Big[\mathrm{E}\Big(\begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix}|\boldsymbol{\beta}\Big)\Big] \\
&= \begin{pmatrix} \sigma^2\boldsymbol{C}_\varphi & \sigma\gamma\boldsymbol{C}_\varphi^{\frac{1}{2}} \\ \sigma\gamma\boldsymbol{C}_\varphi^{\frac{1}{2}} & \mathbf{I_n} \end{pmatrix} + \mathrm{Var}\Big[\begin{pmatrix} H\boldsymbol{\beta} \\ \boldsymbol{\nu} \end{pmatrix}\Big] \\
&= \begin{pmatrix} \sigma^2\boldsymbol{C}_\varphi + HBH^\top & \sigma\gamma\boldsymbol{C}_\varphi^{\frac{1}{2}} \\ \sigma\gamma\boldsymbol{C}_\varphi^{\frac{1}{2}} & \mathbf{I_n} \end{pmatrix}.
\end{aligned}$$

با قرار دادن $\Sigma_x = \sigma^2\boldsymbol{C}_\varphi + HBH^\top$، داریم

$$\begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix} \sim N_{n+n}\Big(\begin{pmatrix} H\boldsymbol{a} \\ \boldsymbol{\nu} \end{pmatrix}, \begin{pmatrix} \Sigma_x & \Sigma_x\Gamma_x^\top \\ \Gamma_x\Sigma_x & \Delta_x + \Gamma_x\Sigma_x\Gamma_x^\top \end{pmatrix}\Big),$$

که در آن $\Gamma_x = \sigma\gamma\boldsymbol{C}_\varphi^{\frac{1}{2}}\Sigma_x^{-1}$ و $\Delta_x = \mathbf{I_n} - \Gamma_x\Sigma_x\Gamma_x^\top$ است. بنابراین طبق ساختار توزیع چوله گاوسی بسته $\boldsymbol{X} \overset{d}{=} \boldsymbol{U}_1|\boldsymbol{U}_2 < \mathbf{0}$ است.

۱.۴ پیشگویی فضایی

پیشگویی متغیرهای پنهان فضایی یکی از هدف‌های مهم در مدل‌های آمیخته خطی تعمیم‌یافته فضایی است. برای این منظور متغیرهای پنهان در $n-k$ موقعیت فاقد مشاهده $\{\boldsymbol{s}_{k+1}, \ldots, \boldsymbol{s}_n\}$ را با $\boldsymbol{X}^{pred} = A\boldsymbol{X}$ مشخص می‌کنیم. با توجه به قضیه ۱ و استفاده از خاصیت حاشیه‌ای خانواده توزیع‌های

چوله گاوسی بسته، توزیع شرطی کامل $\boldsymbol{X}^{pred}$ به شرط $(\boldsymbol{y}, \boldsymbol{\eta})$ به صورت

$$\begin{aligned}
&\boldsymbol{X}^{pred}|\boldsymbol{y}, \boldsymbol{\eta} \sim CSN_{n-k,n}(\boldsymbol{\mu}_p, \Sigma_p, \Gamma_p, \boldsymbol{\nu}_p, \Delta_p),\\
&\boldsymbol{\mu}_p = A\boldsymbol{\mu}_{x|y,\eta}, \quad \Sigma_p = A\Sigma_{x|y,\eta}A^\top,\\
&\Gamma_p = \Gamma_{x|y,\eta}\Sigma_{x|y,\eta}A^\top(A\Sigma_{x|y,\eta}A^\top)^{-1},\\
&\boldsymbol{\nu}_p = \boldsymbol{\nu}_{x|y,\eta}, \qquad \Gamma_{x|y,\eta} = \frac{\gamma}{\sigma}\boldsymbol{C}_\varphi{}^{-\frac{1}{2}},\\
&\Delta_p = (1-\gamma^2)\mathbf{I_n} + \Gamma_{x|y,\eta}\Sigma_{x|y,\eta}\Gamma_{x|y,\eta}^\top - \Gamma_p A\Sigma_{x|y,\eta}\Gamma_{x|y,\eta}^\top,
\end{aligned}$$

به‌دست می‌آید. که از آن برای محاسبه توزیع‌های حاشیه‌ای پیشگوی بر اساس نمونه‌های پسین پارامترها استفاده می‌شود. بنابراین توزیع حاشیه‌ای پیشگوی متغیرهای پنهان را می‌توان به‌صورت

$$\hat{\pi}(x_j|\boldsymbol{y}) = \sum_\ell \pi(x_j|\boldsymbol{y}, \boldsymbol{\eta}_\ell)\pi(\boldsymbol{\eta}_\ell|\boldsymbol{y}), \quad j = k+1, \ldots, n,$$

تقریب زد که در آن جمع روی نمونه‌های مونت کارلویی توزیع پسین $\pi(\boldsymbol{\eta}|\boldsymbol{y})$ است. سپس توزیع پیشگوی متفیر پاسخ در موقعیت‌های فاقد مشاهده به‌صورت زیر محاسبه می‌شود.

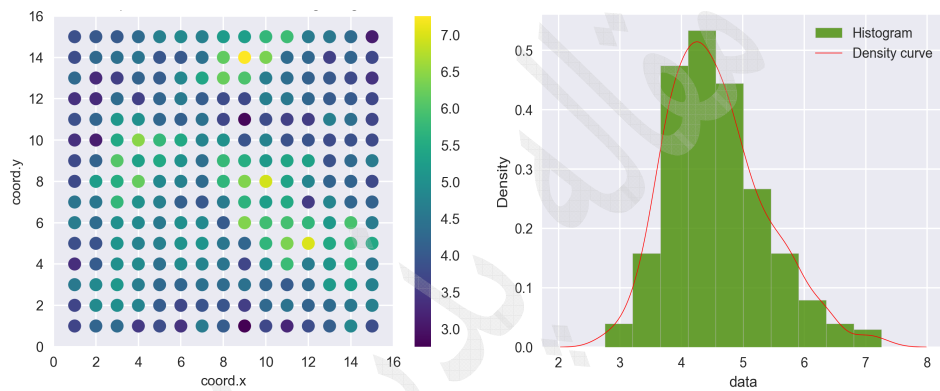
$$\pi(y_j|\boldsymbol{y}) = \int_{x_j} \pi(y_j|x_j)\pi(x_j|\boldsymbol{y})dx_j, \quad j = k+1, \ldots, n.$$

۵ مطالعه شبیه‌سازی

ابتدا یک مشبکه منظم ۱۵ × ۱۵ در نظر گرفته و تعداد ۲۲۵ موقعیت روی آن تولید شده است. با در نظر گرفتن ساختار همسانگرد نمایی برای $\boldsymbol{C}_\varphi$ و $\beta_0 = 2$، $\beta_1 = 1$، $\sigma^2 = 1$، $\varphi = 5$ و $\gamma = 0.85$ متغیرهای پنهان از توزیع چوله گاوسی به‌صورت

$$(\boldsymbol{x}|\boldsymbol{\eta}) \sim CSN_{n,n}(\beta_0 + \beta_1\boldsymbol{h}, \sigma^2\boldsymbol{C}_\varphi, \frac{\gamma}{\sigma}\boldsymbol{C}_\varphi{}^{-\frac{1}{2}}, \nu\mathbf{1_n}, (1-\gamma^2)\mathbf{I_n})$$

استخراج شد، که در آن $\boldsymbol{h}$ بردار مقادیر استاندارد شده عرض جغرافیایی است که به عنوان متغیر کمکی وارد مدل می‌شود. متغیر پاسخ به شرط متغیرهای پنهان، $y_j, j = ۱, \cdots, ۲۲۵$ از توزیع پواسون به‌صورت $y_j \sim Poiss(p_j)$ و $p_j = \exp(x_j)$ تولید و فرایند شبیه‌سازی ۱۰۰ بار تکرار شد. یک نمونه از موقعیت‌ها و داده‌های تولید شده در شکل ۱ رسم شده است.



(ب) (الف)

شکل ۱: تحقق از میدان تصادفی چوله گاوسی بسته مانا: الف– بافتنگار مقادیر متغیرهای پنهان شبیه‌سازی شده، ب– نمودار پراکنش متغیرهای پنهان روی مشبکه منظم ۱۵×۱۵.

برای بررسی عملکرد مدل چوله گاوسی پیشنهادی، دو رویکرد بر اساس مشبکه‌های ۵×۵ و ۱۵×۱۵ اجرا شد. در رویکرد اول ۱۰۰ مجموعه داده از مدل چوله گاوسی پیشنهادی تولید و رهیافت بیزی تقریبی معرفی شده بر اساس دو مدل چوله گاوسی و گاوسی به‌کار گرفته شد و در رویکرد دوم ۱۰۰ مجموعه داده از مدل گاوسی (یعنی حالت خاص مدل چوله گاوسی وقتی پارامتر چولگی صفر است) تولید شدند و رهیافت بیزی تقریبی نیز بر اساس دو مدل رقیب به‌کار گرفته شد. برای ارزیابی برآورد بیزی پارامترها از معیارهای اریبی و ریشه دوم میانگین توان دوم خطاها

$$RMSE(\hat{\eta}) = \sqrt{\frac{۱}{m}\sum_{i=۱}^{m}(\hat{\eta}_i - \eta_i)^۲}, \quad m = ۱۰۰,$$

با ۱۰۰ مجموعه داده شبیه‌سازی استفاده شده است، که در آن m تعداد مجموعه داده‌های شبیه‌سازی برای

جدول ۱: نتایج شبیه‌سازی براساس ۱۰۰ مجموعه داده تولید شده از مدل چوله گاوسی

پارامتر	واقعی	چوله گاوسی				گاوسی			
		۵ × ۵		۱۵ × ۱۵		۵ × ۵		۱۵ × ۱۵	
		اریبی	RMSE	اریبی	RMSE	اریبی	RMSE	اریبی	RMSE
β_0	۲	۰٫۴۸	۰٫۷۴۱	۰٫۱۹	۰٫۶۳۲	−۰٫۵۶	۰٫۹۰۳	−۰٫۲۳	۰٫۷۷۱
β_1	۱	−۰٫۰۷	۰٫۰۵۲	−۰٫۰۲	۰٫۰۳۹	۰٫۰۸	۰٫۰۶۹	۰٫۰۶	۰٫۰۶۱
σ^2	۱	۰٫۱۲	۰٫۰۶۱	۰٫۰۱	۰٫۰۳۳	۰٫۳۷	۰٫۴۱۶	۰٫۱۹	۰٫۱۳۱
φ	۵	۱٫۳۱	۱٫۳۵۲	۰٫۱۳	۰٫۹۰۱	۱٫۸۴	۱٫۶۱	۰٫۹۷	۱٫۲۱۶
γ	۰٫۸۵	−۰٫۰۹	۰٫۰۱۳	−۰٫۰۸	۰٫۰۰۲	−	−	−	−

جدول ۲: نتایج شبیه‌سازی براساس ۱۰۰ مجموعه داده تولید شده از مدل گاوسی

پارامتر	واقعی	چوله گاوسی				گاوسی			
		۵ × ۵		۱۵ × ۱۵		۵ × ۵		۱۵ × ۱۵	
		اریبی	RMSE	اریبی	RMSE	اریبی	RMSE	اریبی	RMSE
β_0	۲	۰٫۲۱	۰٫۶۹۸	۰٫۱۳	۰٫۵۴۱	−۰٫۱۸	۰٫۶۸۳	۰٫۱۱	۰٫۵۱۹
β_1	۱	−۰٫۰۷	۰٫۰۶۱	−۰٫۰۴	۰٫۰۳۱	۰٫۰۶	۰٫۰۵۵	۰٫۰۲	۰٫۰۲۷
σ^2	۱	۰٫۳۲	۰٫۱۰۲	۰٫۰۹	۰٫۰۲۹	۰٫۲۳	۰٫۰۹۹	۰٫۰۸	۰٫۰۲۵
φ	۵	−۱٫۵۳	۱٫۲۴۱	۰٫۵۶	۰٫۷۸۷	−۱٫۴۱۱	۱٫۱۷	۰٫۳۱	۰٫۷۴۳
γ	۰	۰٫۰۵	۰٫۰۰۲	۰٫۰۲	۰٫۰۰۱	−	−	−	−

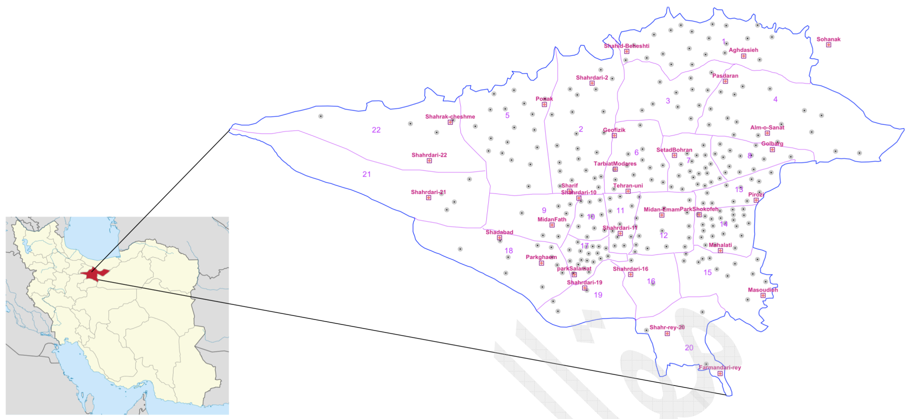
هر حالت و $\hat{\eta}_i$ برآورد بیزی پارامترهای میدان تصادفی برای مجموعه داده iام است. پس از همگرایی الگوریتم‌های مونت کارلویی و کنار گذاشتن مقادیر سوخته، برآورد بیزی پارامترها محاسبه گردید. در این جا از معیار همگرایی گلمن–روبین (گلمن و روبین، ۱۹۹۲) برای چک کردن همگرایی نمونه‌های مونت کارلویی استفاده شد که مقدار آن هرچه به یک نزدیکتر باشد نشان دهنده آن است که الگوریتم‌ها همگرا شده‌اند. این مقدار برای پارامترهای رگرسیونی یک و برای سایر پارامترها ۱٫۰۱ به دست آمد. خلاصه نتایج داده‌های شبیه‌سازی شده از مدل چوله گاوسی در جدول ۱ و داده‌های شبیه‌سازی شده از مدل گاوسی در جدول ۲ ارایه شدند. نتایج نشان می‌دهد که مدل چوله گاوسی بر اساس رویکرد اول به خوبی عمل نموده است. برای مشبکه‌های کوچک نیز مدل چوله گاوسی همچنان عملکرد مناسبی دارد، البته پارامترهای همبستگی فضایی در مشبکه‌های کوچکتر از دقت کمتری برخوردار هستند. در رویکرد دوم (جدول ۲) نیز مدل چوله گاوسی با توجه به اینکه داده‌ها از مدل گاوسی تولید شده‌اند و چولگی ندارند نیز عملکرد نسبتا

خوبی داشته است. لازم به ذکر است که مدل چوله گاوسی حالت کلی‌تری از مدل گاوسی است. با توجه به نتایج دو رویکرد اجرا شده می‌توان گفت که دقت برآورد پارامترهای مدل تحت تاثیر فرض توزیعی در نظر گرفته شده روی متغیرهای پنهان فضایی است و در صورت عدم برقراری آن، از دقت برآوردها کاسته می‌شود.

۶ داده‌های آلودگی هوای شهر تهران

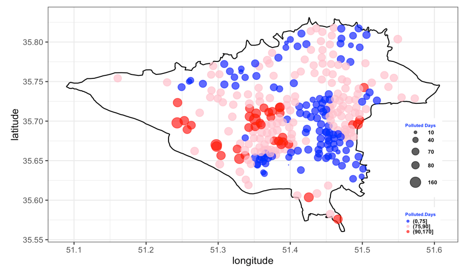
آلودگی هوا یکی از مشکلات مهم شهرهای بزرگ مثل تهران است. وضعیت نامناسب سلامت و تعداد روزهای ناسالم یکی از عوامل مهم در انتخاب محل زندگی برای بسیاری از افراد است، به ویژه کسانی که به بیماری‌های تنفسی و ریه‌ای دچار هستند. در این بخش، تعداد روزهای ناسالم در مناطق مختلف تهران بر اساس اطلاعات به دست آمده از ایستگاه‌های نظارت بر آلودگی هوا مورد بررسی قرار گرفته است. شهر تهران با مساحت کلی ۸۱۴،۱۸ کیلومتر مربع و یکی از بزرگ‌ترین و پرجمعیت‌ترین شهرهای جهان با جمعیتی حدود ۹ میلیون نفر است. تهران در شمال کوهستان مرکزی ایران واقع شده است. در ایستگاه‌ها معیارهای مختلف آلودگی هوا ثبت می‌شوند و وضعیت آلودکی هوا با شاخص کیفیت هوا[1] (AQI) اندازه‌گیری می‌شود. سطح AQI بر اساس اندازه‌گیری ذرات معلق (مانند SO2، NO2، PM10، PM2.5، CO، O3) در ایستگاه‌های نظارتی تعیین می‌شود. در این مطالعه مقادیر این شاخص به ۳ دسته تقسیم می‌شود (https://www.airnow.gov/aqi/aqi-basics). زمانی که $AQI > ۱۰۰$، کیفیت هوا ناسالم (روز آلوده) است، اگر $۵۰ < AQI \leq ۱۰۰$ کیفیت هوا قابل قبول و روز خوب است و اگر $AQI \leq ۵۰$ خطر آلودگی هوا وجود ندارد به عبارت دیگر روز خوبی است. در این بخش، داده‌های AQI در ۲۹۹ مکان از ۱۳ اسفند ۱۳۹۷ تا ۲۹ اسفند ۱۳۹۸ برای ۳۱ ایستگاه سینوپتیک در ۲۲ منطقه شهرداری تهران از شرکت کنترل کیفیت هوای تهران (http://air.tehran.ir) دریافت و مورد تحلیل قرار گرفته است. مکان‌های داده‌ها بر روی نقشه تهران در شکل ۲ با دایره‌های پرشده نشان داده شده است. تعداد روزهای ناسالم بر اساس AQI در تهران در شکل ۳ (الف) نشان داده شده است. رنگ و اندازه

[1] Aie Quality Index

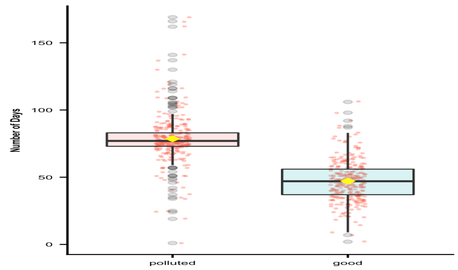


شکل ۲: نمایش موقعیت ۲۹۹ مکان ثبت شده در ۲۲ منطقه شهرداری تهران توسط دایره‌های پر شده

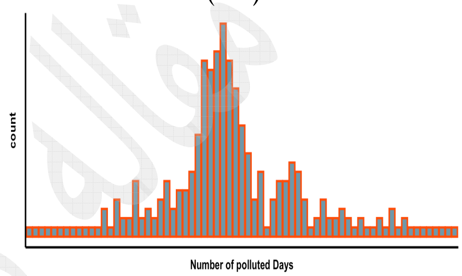
دایره‌ها نسبت روزهایی با AQI بیش از ۱۰۰ را نشان می‌دهد. تعداد روزهای ناسالم را به عنوان متغیر پاسخ مدنظر قرار داده‌ایم. در شکل ۳ (ب) نمودارهای جعبه‌ای تعداد روزها با کیفیت هوای ناسالم و تعداد روزها با کیفیت هوا خوب و همچنین بافتنگار تعداد روزها با کیفیت هوای ناسالم و نمودار پراکندگی این دو متغیر در مقابل هم را مشاهده می‌نمایید که بین متغیر پاسخ و روزهای خوب روندی مشاهده می‌شود و روزهای خوب به‌عنوان متغیر کمکی وارد مدل شده‌است. برای ساختار همبستگی فضایی داده‌ها تابع همتغییرنگار نمایی همسانگرد به‌صورت $C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \exp(- \parallel \boldsymbol{s}_i - \boldsymbol{s}_j \parallel /\varphi)$ در نظر گرفته شد. فرض می‌شود متغیر پاسخ دارای توزیع پواسون شرطی مستقل $Y_i|X_i \sim Poisson(P_i),\ i = ۱, \ldots, ۲۹۹,$ است، که در آن $P_i = \mathrm{E}(Y_i|X_i)$، $\log(P_i) = X_i$ و Y_i تعداد روزهای ناسالم به عنوان متغیر پاسخ و X_i یک متغیر پنهان با توزیع چوله گاوسی بسته به‌صورت (۳) با پارامتر مکان $\beta_۰ + \beta_۱ z_i$ است، که در آن z_i تعداد روزهای خوب و متغیر کمکی مدل است. با به‌کار بردن رهیافت بیزی معرفی‌شده و پس از همگرایی الگوریتم‌های مونت کارلویی برآورد بیزی پارامترهای مدل $\beta_۰ = ۱/۶۰۰۱$، $\beta_۱ = ۰/۰۰۹۷$، $\sigma^۲ = ۷۶/۹۸۰۶$، $\varphi = ۵/۱۰۸۶$ و $\gamma = ۰/۸۸۸۳$ به‌دست آمد. با استفاده از مدل برآوردشده نقشه پیش‌گویی متغیر پاسخ فضایی در شکل ۴ آورده شده است. این نمودار شدت آلودگی هوا در مناطق مختلف را در بازه زمانی ۱۳ اسفند ۱۳۹۷ تا ۲۹ اسفند ۱۳۹۸ در ۲۲ منطقه شهرداری تهران نشان می‌دهد. شدت آلودگی هوا در مناطق مرکزی، جنوب‌غربی، جنوبی، شرقی و شمال‌شرقی از سایر مناطق



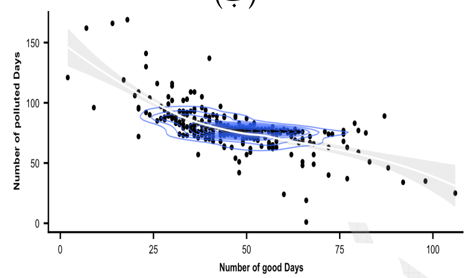
(الف)



(ب)

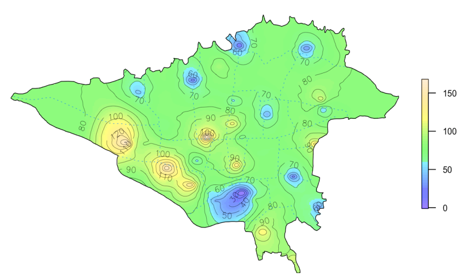


(ج)

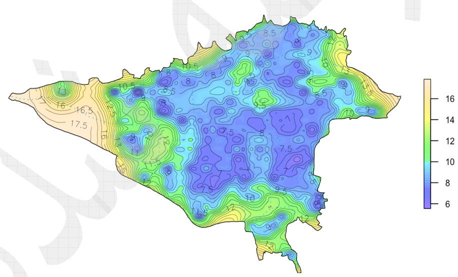


(د)

شکل ۳: الف– تعداد روزهای ناسالم بر اساس AQI $(AQI > ۱۰۰)$، رنگ و اندازه دایره‌ها نسبت روزها را نشان می‌دهد. ب– نمودارهای جعبه‌ای روزهای ناسالم و روزهای خوب. ج– بافتنگار تعداد روزهای با کیفیت هوای ناسالم، د– پراکندگی روزهای خوب در مقابل روزهای ناسالم



(الف)



(ب)

شکل ۴: الف– نقشه پیش‌گویی متغیر پاسخ فضایی و ب– خطای معیار پیش‌گویی‌ها.

بیشتر است. شکل ۴ (الف) نقشه خطای معیار پیش‌گویی‌ها را نشان می‌دهد، که نشان‌دهنده این است که دقت پیش‌گویی در مناطقی که داده‌های بیشتری در دسترس هستند، بیشتر از سایر مناطق است.

بحث و نتیجه‌گیری

در این مقاله از یک میدان تصادفی چوله گاوسی بسته مانا برای مدل‌بندی متغیرهای پنهان فضایی در مدل‌های آمیخته خطی تعمیم‌یافته استفاده و یک رهیافت بیزی برای برآورد این مدل‌ها معرفی شد. برای به‌دست آوردن برآوردهای بیزی پارمترهای مدل از یک رهیافت بیزی تقریبی و الگوریتم‌های مونت‌کارلویی گیبز و متروپولیس–هستینگز استفاده شد. نتایج شبیه‌سازی نشان می‌دهد که مدل معرفی شده براساس میدان تصادفی چوله گاوسی مانا و رهیافت تقریبی بیزی معرفی شده برای به‌دست آوردن برآوردهای مدل به خوبی عمل نموده است. به دلیل پیچیدگی مدل معرفی شده تحلیل بیزی با به‌کار بردن روش‌های مونت‌کارلویی زمان بر است و گاهی همگرایی‌ها به خصوص برای داده‌های حجیم به راحتی حاصل نمی‌شود. به عنوان پیشنهاد می‌توان برای به‌دست آوردن برآورد پارامترها به روش بیزی از رهیافت INLA یا الگوریتم‌های مونت‌کارلویی همیلتونی استفاده کرد که باعث افزایش سرعت محاسبات می‌شوند.

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

محمدزاده، م.، (۱۳۹۸)، آمار فضایی و کاربردهای آن، چاپ سوم، مرکز نشر آثار علمی دانشگاه تربیت مدرس، تهران،

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

حسینی، ف. و کریمی، الف. (۱۴۰۱) برآورد مدل‌های آمیخته خطی تعمیم‌یافته فضایی با میدان تصادفی چوله گاوسی مانا، اندیشه آماری، **۲۷**(۱)، ۷۳-۷۹.

Allard, D. and Naveau, P. (2007), A New Spatial Skew-Normal Random Field Model, *Communications in Statistics - Theory and Methods*, **36**, 1821-1834.

Gelman, A. and Rubin, D. B. (1992), Inference from Iterative Simulation using Multiple Sequences. *Statistical Science*, **7**, 457–472.

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, 25-42.

Hosseini, F., Mohammadzadeh, M., (2012). Bayesian Prediction for Spatial GLMM's with Closed Skew Normal Latent Variables, *Australian & New Zealand Journal of Statistics*, **54**, 43-62.

Hosseini, F. and Karimi, O. (2020), Approximate Likelihood Inference in Spatial Generalized Linear Mixed Models with Closed Skew Normal Latent Variables, *Communication in Statistics- Simulation and Computation*, **49**, 121-134.

Hosseini, F. and Karimi, O. (2021), Approximate Pairwise Likelihood Inference in SGLM Models with Skew Normal Latent Variables, *Journal of Computational and Applied Mathematics*, **398**, 113692.

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.

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

Karimi, O. and Mohammadzadeh, M. (2011), Bayesian Spatial Prediction for Discrete Closed Skew Gaussian Random Field, *Mathematical Geosciences*, **43**, 565–582.

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

Karimi, O. (2024), A Hamiltonian Monte Carlo EM Algorithm for Generalized Linear Mixed Models with Spatial Skew Latent Variables, *Statistical Papers*, **65**(2), 1065–1084.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, Chapman and Hall, London.

Rimstad, K. and Omre, H. (2014), Skew-Gaussian Random Fields, *Spatial Statistics*, **10**, 43-62.

Bayesian Analysis of Latent Variables in Spatial GLM Models with Stationary Skew Gaussian Random Field

Fatemeh Hosseini and Omid Karimi

Department of Statistics, Semnan University, Semnan , Iran.

Abstract: The spatial generalized linear mixed models are often used, where the latent variables representing spatial correlations are modeled through a Gaussian random field to model the categorical spatial data. The violation of the Gaussian assumption affects the accuracy of predictions and parameter estimates in these models. In this paper, the spatial generalized linear mixed models are fitted and analyzed by utilizing a stationary skew Gaussian random field and employing an approximate Bayesian approach. The performance of the model and the approximate Bayesian approach is examined through a simulation example, and implementation on an actual data set is presented.

Keywords: Spatial Generalized Linear Mixed Models, Approximate Bayesian Analysis, Gaussian Random Field, Stationary Skew Gaussian Random Field.

Mathematics Subject Classification (2010): 62F15, 62J12.