پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

انجمن آمار ایران

فرایند چوب‌شکنی رگرسیون-بتا با کوواریانس نامانا وابسته به متغیر کمکی

اسماعیل یارعلی [۱] ، فیروزه ریواز [۲] ، مجید جعفری خالدی [۳]

[۱] دانش‌آموخته دکتری آمار، دانشگاه شهید بهشتی، تهران

[۲] گروه آمار، دانشگاه شهید بهشتی، تهران

[۳] گروه آمار، دانشگاه تربیت مدرس، تهران

چکیده: در این مقاله رویکردی جدیدی برای ساخت فرایند چوب‌شکنی مورد مطالعه قرار می‌گیرد به‌طوری‌که توزیع تصادفی وابسته به متغیر کمکی است. این شیوه مدل‌بندی علاوه برآن‌که نامانایی متاثر از متغیر کمکی را در ساختار وابستگی داده‌ها را پوشش می‌هد، امکان افرازبندی ناحیه مورد مطالعه را نیز فراهم می کند. با در نظر گرفتن این فرایند به عنوان توزیع اثرهای تصادفی در یک چارچوب مدل خطی آمیخته، داده ها مدل شده و استنباط بیزی آن ارائه می شود. سپس با استفاده از یک مجموعه داده‌ی واقعی مربوط به میزان بارش عملکرد مدل معرفی شده در مقایسه با مدل‌های رقیب مورد ارزیابی قرار می‌گیرد.

واژه‌های کلیدی: فرایند چوب‌شکنی، رگرسیون-بتا، نامانایی، متغیر کمکی، ساختار همبستگی.

کد موضوع‌بندی ریاضی (۲۰۲۰): 62H11، 62M30، 62F15.

---

۱ مقدمه

یکی از مؤلفه‌های اصلی تحلیل داده‌های فضایی، تعیین وابستگی فضایی است. این کار عموماً از طریق تابع کوواریانس که متعلق به رده‌ای پارامتری از مدل‌های مانا است، انجام می‌شود. این درحالی است که در بسیاری از کاربردها، فرض مانایی واقع‌بینانه نیست. در واقع، در بسیاری از کاربردها عوامل موضعی مانند توپوگرافیکی منطقه مورد مطالعه بر همبستگی فضایی تاثیرگذار هستند و باعث ایجاد نامانایی در ساختار وابستگی فضایی می‌شوند. لذا ساخت مدل‌های کوواریانس نامانای متأثر از متغیرهای کمکی موضعی بسیار حائز اهمیت است. تاکنون رهیافت‌های مختلفی برای الحاق متغیر کمکی در ساختار مرتبه دوم فرایندهای فضایی ارائه شده است. از جمله می‌توان به روش هموارسازی هسته (ریچ و همکاران ، ۲۰۱۱)، روش پیچش فرایند (رایزر و کالدر ، ۲۰۱۵)، روش دگردیسی (اشمیت

---

[۱] سخنران، e_yarali@sbu.ac.ir

و همکاران ، ۲۰۱۱)، روش بسط توابع پایه (یارعلی و ریواز ، ۲۰۲۰) و استفاده از معادلات مشتقات جزئی تصادفی (اینگریبستن و همکاران ، ۲۰۱۴) اشاره کرد. گرچه الحاق متغیر تصادفی در ساختار وابستگی باعث بهبود در مدل‌های فضایی هم به لحاظ کاهش بعد تعداد پارامتر و به دنبال آن محاسبات ساده و هم تفسیر راحت‌تر نامانایی شده است، اما در تمامی این روش‌ها برای الحاق متغیر کمکی در ساختار وابستگی فضایی فرض شده است میدان تصادفی مورد مطالعه گاوسی است، در حالی‌که چنین فرضی در عمل به سختی برقرار بوده و داده‌ها شواهدی از تخطی مفروضات توزیع گاوسی را از خود نشان می‌دهند. برای رفع این محدودیت‌ها، در سال‌های اخیر مدل‌های فضایی ناگاوسی متعددی از جمله مدل‌های چوله گاوسی، رهیافت پیچش فرایند گاوسی-لگ گاوسی و فرایندهای گاوسی تبدیل یافته g و h توکی معرفی شده است. علی‌رغم ویژگی‌های مناسب این مدل‌ها، ساختار همبستگی القاء شده توسط آنها مبتنی بر فرض محدودکننده مانایی و همسانگردی میدان تصادفی مورد مطالعه است. درحالی‌که در برخی از کاربردها ارتباط ساختار فضایی با متغیر / متغیرهای کمکی موضعی باعث ایجاد نامانیی در ساختار وابستگی داده ها می شود.

یک رویکرد مناسب استفاده از رهیافت بیزی ناپارامتری است که در آن توزیع داده‌ها نامعلوم و تصادفی در نظر گرفته می شود و برای آن یک پیشینی مناسب انتخاب می شود.در این رویکرد، معمولاً از پیشینی فرایند چوب شکنی استفاده می شود از جمله می‌توان به فرایندهای چوب‌شکنی هسته (ریچ و فوئنتیس ، ۲۰۰۷؛ برزگر و ریواز ، ۲۰۲۰؛ دحدوح و خالدی ، ۲۰۲۰)، فرایندهای چوب‌شکنی پروبیت (رودریگوئز و دانسون ، ۲۰۱۱)، فرایندهای چوب‌شکنی لوجیت (رن و همکارن ، ۲۰۱۱)، فرایندهای چوب‌شکنی تعمیم‌یافته (بارسلا و همکاران ، ۲۰۱۸) و فرایندهای چوب‌شکنی پنهان (رودریگوئز و همکاران ، ۲۰۱۰) اشاره کرد. علی‌رغم ایجاد انعطاف بالا در مدل‌های معرفی شده، ساختار همبستگی القا شده از این مدل‌ها بسیار پیچیده و وابستگی فضایی وابسته به متغیر کمکی نیست. یارعلی و همکاران (۲۰۲۲) برای رفع این مشکل با استفاده از رهیافت بیزی ناپارامتری، رده‌ی جدیدی از فرایند چوب‌شکنی را معرفی کردند به‌طوری‌که نامانایی از طریق متغیر کمکی در همبستگی فضایی القا می‌شود. این شیوه مدل‌بندی علاوه برآن‌که رفتارهای توزیعی شامل عدم تقارن و چند مدی بودن را مدل‌بندی می‌کند، نامانایی متاثر از متغیر کمکی را در ساختار کوواریانس پوشش می‌دهد. همچنین افرازبندی القا شده توسط این پیشینی پیشنهادی وابسته به متغیر کمکی بوده که امکان افرازبندی ناحیه ی مورد مطالعه را بر اساس اطلاعات متغیر کمکی فراهم می کند براین اساس در بخش ۲ فرایند چوب‌شکنی رگرسیون-بتا که توسط یارعلی و همکاران (۲۰۲۲) معرفی شد، بیان می‌شود. در ادامه، با بکارگیری یک مدل بیزی مبتنی بر فرایند چوب‌شکنی رگرسیون-بتا، استنباط‌های پسینی و پیشگویی فضایی ارائه می‌شوند. همچنین با استفاده از یک مجموعه داده‌ی واقعی عملکرد مدل معرفی شده با مدل‌های رقیب از نظر پیشگویی فضایی و افرازبندی ناحیه مورد مطالعه مورد ارزیابی قرار می‌گیرد.

۲ فرایند چوب‌شکنی رگرسیون-بتا

در این بخش با استفاده از رهیافت بیزی ناپارامتری، فرایند فضایی چوب‌شکنی رگرسیون-بتا که توسط یارعلی و همکاران (۲۰۲۲) بیان شد، معرفی می‌شود. بر این اساس فرض کنید $Y(\cdot) = \{Y(\boldsymbol{s}),\ \boldsymbol{s} \in \mathcal{D} \subset \mathcal{R}^d, d \neq ۱\}$ یک میدان تصادفی باشد که روی ناحیه $\mathcal{D}$ تعریف شده است. به‌علاوه فرض کنید $Y(\boldsymbol{s})$ مقدار مشاهد شده در موقعیت $\boldsymbol{s}$ است که از مدل

$$Y(\boldsymbol{s}) = \boldsymbol{w}^{'}(\boldsymbol{s})\boldsymbol{\beta} + \eta_{\boldsymbol{x}}(\boldsymbol{s}) + \epsilon(\boldsymbol{s}), \qquad (۱.۲)$$

پیروی می‌کند که در آن $\boldsymbol{w}(\boldsymbol{s})$ بردار متغیرهای کمکی در مکان $\boldsymbol{s}$ و $\boldsymbol{\beta} \in \mathcal{R}^p$ بردار ضرایب رگرسیونی نامعلوم است. همچنین $\epsilon(\boldsymbol{s})$ نشان‌دهنده‌ی خطای اندازه‌گیری است که فرض می‌شود دارای توزیع نرمال با میانگین صفر و واریانس $\tau^۲$ است. اکنون برای الحاق متغیر

کمکی در ساختار همبستگی فضایی، فرض می‌شود فرایند فضایی $\eta_{\boldsymbol{x}}(\boldsymbol{s})$ به صورت

$$\eta_{\boldsymbol{x}}(\boldsymbol{s}) = G^{-1}(V_{\boldsymbol{x}}(\boldsymbol{s})), \tag{۲.۲}$$

است که در آن $G(\cdot)$ اندازه احتمال تصادفی گسسته است که به صورت $G(\cdot) = \sum_{l=1}^{L} p_l \delta_{\theta_l}(\cdot)$ تعریف می‌شود که در آن $\delta_{\theta_l}(\cdot)$ تابع دیراک، $G^{-1}(\cdot)$ وارون تعمیم‌یافته تابع $G(\cdot)$، و اتم‌های θ_1، $\cdots$، θ_L تحقق‌هایی مستقل از توزیع نرمال $N(0, \sigma^2)$ هستند. همچنین $V_{\boldsymbol{x}}(\cdot) = \{V_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{R}^d\}$ یک فرایند فضایی رگرسیون-بتا است که به صورت زیر تعریف می‌شود.

تعریف ۱.۲. فرض کنید $\{Z(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{R}^d\}$ یک فرایند گاوسی مانا با میانگین صفر، واریانس یک و تابع همبستگی $\rho(\boldsymbol{s}_i, \boldsymbol{s}_j; \lambda)$ باشد. اگر $V_{\boldsymbol{x}}(\boldsymbol{s})$ به شرط $Z(\boldsymbol{s})$، یک فرایند مستقل و دارای توزیع $Beta(\mu_x(\boldsymbol{s})\psi, \psi(1-\mu_x(\boldsymbol{s})))$ با تابع چگالی

$$f(v_x) = \frac{\Gamma[\psi]}{\Gamma[\mu_x(\boldsymbol{s})\psi]\Gamma[\psi(1-\mu_x(\boldsymbol{s}))]} v_x^{\mu_x(\boldsymbol{s})\psi-1}(1-v_x)^{\psi(1-\mu_x(\boldsymbol{s}))-1}, \quad 0 < \mu_x(\boldsymbol{s}) < 1, \qquad \psi > 0, \tag{۳.۲}$$

باشد که در آن $\mu_{\boldsymbol{x}}(\boldsymbol{s}) = \mathbb{E}[V_{\boldsymbol{x}}(\boldsymbol{s})|Z(\boldsymbol{s})]$ به صورت

$$\mu_x(\boldsymbol{s}) = \frac{\exp\left(\boldsymbol{x}'(\boldsymbol{s})\boldsymbol{\gamma} + Z(\boldsymbol{s})\right)}{1 + \exp\left(\boldsymbol{x}'(\boldsymbol{s})\boldsymbol{\gamma} + Z(\boldsymbol{s})\right)}, \tag{۴.۲}$$

تعریف می‌شود به طوری‌که $\boldsymbol{x}(\boldsymbol{s}) = (x_1(\boldsymbol{s}), \cdots, x_q(\boldsymbol{s}))'$ برداری از متغیرهای کمکی در مکان $\boldsymbol{s}$، $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_q)'$ برداری از ضرایب نامعلوم که اثر متغیرهای کمکی را در همبستگی فضایی کنترل می‌کند، آن‌گاه $V_{\boldsymbol{x}}(\cdot) = \{V_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{R}^d\}$ را یک فرایند فضایی رگرسیون- بتا گویند.

حال، برای هر مجموعه متناهی از موقعیت‌های $\boldsymbol{s}_1$، $\cdots$، $\boldsymbol{s}_n$ روی ناحیه $\mathcal{D}$، احتمال شرطی

$$P_{\boldsymbol{x}}\Big(\eta_{\boldsymbol{x}}(\boldsymbol{s}_1) = \theta_{l_1}, \cdots, \eta_{\boldsymbol{x}}(\boldsymbol{s}_n) = \theta_{l_n} | \{v_l\}_{l=1}^{L}, \{\theta_l\}_{l=1}^{L}\Big)$$

به صورت

$$\begin{aligned} &P_{\boldsymbol{x}}\Big(\eta_{\boldsymbol{x}}(\boldsymbol{s}_1) = \theta_{l_1}, \cdots, \eta_{\boldsymbol{x}}(\boldsymbol{s}_n) = \theta_{l_n} | \{v_l\}_{l=1}^{L}, \{\theta_l\}_{l=1}^{L}\Big) \\ &\quad = P\Big(\sum_{j=1}^{l_1-1} p_j \le V_{\boldsymbol{x}}(\boldsymbol{s}_1) < \sum_{j=1}^{l_1} p_j; \cdots; \sum_{j=1}^{l_n-1} p_j \le V_{\boldsymbol{x}}(\boldsymbol{s}_n) < \sum_{j=1}^{l_n} p_j\Big), \end{aligned} \tag{۵.۲}$$

تعریف می‌شود به طوری‌که l_1،$\cdots$،l_n یکی از مقادیر $\{1, \cdots, L\}$ را می‌گیرد. در این صورت توزیع پیشینی القاشده توسط مدل سلسله‌مراتبی بالا تحت اندازه احتمال $P_{\boldsymbol{x}}$ را یک فرایند چوب‌شکنی رگرسیون-بتا نامیده و با نماد $P_{\boldsymbol{x}} \sim BRSBP(\alpha, \psi, \boldsymbol{\gamma}, \lambda, \sigma^2)$ نمایش می‌دهیم.

یک رهیافت دیگر برای ساخت فرایند چوب‌شکنی رگرسیون-بتا استفاده از فرایندهای برچسب‌گذاری است. برای این منظور، فرض کنید $\xi_{\boldsymbol{x}}(\cdot) = \{\xi_{\boldsymbol{x}}(\boldsymbol{s}), \boldsymbol{s} \in \mathcal{D}\}$ یک فرایند برچسب‌گذاری وابسته به متغیرهای کمکی است که مقدار l را می‌گیرد اگر

$$\sum_{j=1}^{l-1} p_j \le V_{\boldsymbol{x}}(\boldsymbol{s}) < \sum_{j=1}^{l} p_j. \tag{۶.۲}$$

حال برای هر مجموعه متناهی $\boldsymbol{s}_1$، $\cdots$، $\boldsymbol{s}_n$ وزن‌های توأم وابسته فضایی برای فرایند برچسب‌گذاری به صورت

$$Q_{\boldsymbol{x}(\boldsymbol{s}_1),\cdots,\boldsymbol{x}(\boldsymbol{s}_n)}(l_1,\cdots,l_n) = P\Big(\xi_{\boldsymbol{x}}(\boldsymbol{s}_1)=l_1,\cdots,\xi_{\boldsymbol{x}}(\boldsymbol{s}_n)=l_n|\{v_l\}_{l=1}^{L}\Big)$$
$$= P\Big(\sum_{j=1}^{l_1-1} p_j \le V_{\boldsymbol{x}}(\boldsymbol{s}_1) < \sum_{j=1}^{l_1} p_j;\cdots;\sum_{j=1}^{l_n-1} p_j \le V_{\boldsymbol{x}}(\boldsymbol{s}_n) < \sum_{j=1}^{l_n} p_j\Big), \quad (۷.۲)$$

تعریف می‌شود. در این صورت به شرط فرایند برچسب‌گذاری، اثر تصادفی فضایی (۲.۲) را می‌توان به صورت $\eta_{\boldsymbol{x}}(\boldsymbol{s}) = \theta_{\xi_{\boldsymbol{x}}(\boldsymbol{s})}$ نوشت. بنابراین، با حاشیه‌سازی روی $\xi_{\boldsymbol{x}}$ داریم $\eta_{\boldsymbol{x}}|P_{\boldsymbol{x}} \sim P_{\boldsymbol{x}}$ و $P_{\boldsymbol{x}} \sim BRSBP(\alpha,\psi,\gamma,\lambda,\sigma^2)$. از این‌رو با استفاده از فرایند برچسب‌گذاری (۷.۲) می‌توان یک تعریف معادل و یکسان برای $BRSBP$ ارائه کرد.

۱.۲ ویژگی‌های فرایند چوب‌شکنی رگرسیون-بتا

- فرایند چوب‌شکنی رگرسیون-بتا با توزیع متناهی بعد (۵.۲) در اصول سازگاری کولموگوروف صدق می‌کند.
- فرایند چوب‌شکنی رگرسیون-بتا رده‌ی بزرگتری از فرایندهای چوب‌شکنی است که در حالت خاص فرایند چوب‌شکنی پنهان (رودریگوئز و همکاران ، ۲۰۱۰) را در خود دارد. به بیان واضح‌تر، اگر در مدل $BRSBP$ به جای فرایند فضایی رگرسیون-بتا، فرایند فضایی یکنواخت $U(\boldsymbol{s})$ در نظر گرفته شود آنگاه فرایند چوب‌شکنی پنهان $(LaSBP)$ حاصل می‌شود. علاوه براین اگر در مدل پیشنهادی، ψ به بینهایت میل کند آنگاه مدل $BRSBP$ به یک مدل رگرسیون لوچستیک تبدیل می‌شود. همچنین اگر $\mu_{\boldsymbol{x}}(\boldsymbol{s})$ به صورت مدل رگرسیون پروبیت مدل‌بندی شود و ψ به بینهایت میل کند آنگاه مدل $BRSBP$ به مدل رگرسیون پروبیت تبدیل می‌شود. بنابراین مدل پیشنهادی رده‌ی بزرگتری است که در حالت خاص مدل‌های رگرسیون لوچستیک و پروبیت را در خود دارد. همچنین توزیع بتا نسبت به توزیع یکنواخت انعطاف بیشتری دارد و پارامترهای این توزیع می‌تواند رفتارهای مختلفی از داده‌ها را مدل‌بندی کند.
- افرازبندی القا شده توسط مدل $BRSBP$ روی اثر فضایی مبتنی بر متغیرکمکی است که با گسسته‌سازی فرایند فضایی رگرسیون-بتا حاصل می‌شود. همچنین در تعریف مدل $BRSBP$ از یک فرایند گاوسی برای القای وابستگی فضایی استفاده شده است که این کار موجب سادگی در محاسبات و برازش مدل نسبت به فرایندهای چوب‌شکنی دیگر می‌شود.

۳ استنباط بیزی و پیشگویی فضایی

از آن‌جا که هر یک از پارامترها ویژگی خاصی از میدان تصادفی را بیان می‌کنند، فرض می‌شود کلیه‌ی پارامترها مستقل از هم هستند. برای ضرایب رگرسیونی $\boldsymbol{\beta}$ و γ توزیع‌های پیشینی ناآگاهی‌بخش نرمال چندمتغیره در نظر گرفته می‌شود. برای پارامترهای واریانس σ^2 و τ^2 به‌ترتیب پیشینی مزدوج گامای وارون ناآگاهی‌بخش $IG(s_{\tau^2},r_{\tau^2})$ و $IG(s_{\sigma^2},r_{\sigma^2})$ استفاده می‌شود. همچنین پیشینی گاما $G(s_\alpha,r_\alpha)$ و $G(s_\psi,r_\psi)$ به‌ترتیب برای پارامترهای α و ψ در نظر گرفته می‌شود. همچنین برای برازش مدل پیشنهادی، تابع همبستگی نمایی با پارامتر λ به صورت $\rho(\boldsymbol{s},\boldsymbol{s}';\lambda) = \exp\{-\frac{||\boldsymbol{s}-\boldsymbol{s}'||}{\lambda}\}$ فرض می‌شود. برای پارامتر همبستگی λ نیز توزیع گاما $G(a_\lambda,b_\lambda)$ در نظر گرفته می‌شود.

فرض کنید $\boldsymbol{\Theta} = \{\boldsymbol{\beta},\gamma,\sigma^2,\tau^2,\lambda,\alpha,\psi\}$ بردار پارامترهای نامعلوم باشد. با توجه به اینکه توزیع پسینی دارای فرم بسته نیست، با اتخاذ روش نمونه‌گیری گیبز و داده‌افزایی توزیع‌های تمام‌شرطی تعیین و از آن‌ها نمونه‌گیری می‌شود. با توجه به اینکه برخی از

توزیع‌های تمام‌شرطی دارای فرم استاندارد نیستند از الگوریتم متروپولیس-هستینگس برای نمونه‌گیری از آن‌ها استفاده می‌شود. به منظور خلاصه‌سازی جزئیات محاسبات بیان نشده است.

یکی از اهداف اصلی تحلیل داده‌های فضایی پیشگویی فرایند مورد مطالعه در مکان‌های فاقد نمونه است. برای این منظور فرض کنید $y^* = y(\boldsymbol{s}_0)$ و $\boldsymbol{y} = (y(\boldsymbol{s})_1, \cdots, y(\boldsymbol{s}_n))'$ به‌ترتیب نشان‌دهنده‌ی متغیر پاسخ در هر موقعیت دلخواه جدید $\boldsymbol{s}_0$ و برداری از مشاهدات از فرایند Y در مکان‌های $\boldsymbol{s}_1$، $\cdots$، $\boldsymbol{s}_n$ باشند. در این صورت تحت مدل معرفی شده، توزیع پیشگوی فضایی برای مکان $\boldsymbol{s}_0$ به صورت

$$p(y^*|\boldsymbol{y}) = \int p(y^*|\xi_{\boldsymbol{x}}(\boldsymbol{s}_0), \theta_{\xi_{\boldsymbol{x}}(\boldsymbol{s}_0)}, \boldsymbol{\beta}, \tau^2) p(\xi_{\boldsymbol{x}}(\boldsymbol{s}_0), \theta_{\xi_{\boldsymbol{x}}(\boldsymbol{s}_0)}, \boldsymbol{\beta}, \tau^2|\boldsymbol{y}) d\boldsymbol{\beta} d\tau^2 d\xi_{\boldsymbol{x}}(\boldsymbol{s}_0) d\theta_{\xi_{\boldsymbol{x}}(\boldsymbol{s}_0)} \quad (۱.۳)$$

است که در آن

$$p(y^*|\xi_{\boldsymbol{x}}(\boldsymbol{s}_0), \theta_{\xi_{\boldsymbol{x}}(\boldsymbol{s}_0)}, \boldsymbol{\beta}, \tau^2) \sim N(\boldsymbol{w}'\boldsymbol{\beta} + \theta_{\xi_{\boldsymbol{x}}(\boldsymbol{s}_0)}, \tau^2). \quad (۲.۳)$$

در رابطه (۱.۳) محاسبه‌ی انتگرال به صورت تحلیلی ناممکن است و با استفاده از روش‌های مونت‌کارلو می‌توان آن را تقریب زد. در این راستا، به ازای هر مشاهده پسینی از $\xi_{\boldsymbol{x}}(\boldsymbol{s}_0)$، $\theta_{\xi_{\boldsymbol{x}}(\boldsymbol{s}_0)}$، $\boldsymbol{\beta}$ و τ^2 یک مشاهده از توزیع (۲.۳) تولید می‌شود. سپس با تکرار مراحل فوق به اندازه مورد نیاز، امکان دستیابی به مشاهدات از توزیع پیشگو به صورت $\{y(\boldsymbol{s}_0)^{(m)}, j = 1, \cdots, M\}$ مهیا می‌شود. در نتیجه پیشگویی فضایی بیزی در مکان $\boldsymbol{s}_0$ به صورت زیر بدست می‌آید:

$$\hat{y}(\boldsymbol{s}_0) \equiv \hat{\mathbb{E}}(y(\boldsymbol{s}_0)|\boldsymbol{y}) = \frac{1}{M} \sum_{m=1}^{M} \boldsymbol{y}^{(m)}(\boldsymbol{s}_0).$$

۴ مثال کاربردی

در این بخش از مقاله به عملکرد و ارزیابی مدل معرفی شده از دیدگاه پیشگویی فضایی و افرازبندی ناحیه مورد مطالعه در مقایسه با مدل‌های معرفی شده توسط ریچ و همکاران (۲۰۱۱)، رایزر و کالدر (۲۰۱۵)، گرمسی و لی (۲۰۰۸)، پیچ و کوئینتانا (۲۰۱۶) و رودریگوئز و همکاران (۲۰۱۰) پرداخته می‌شود. برای این منظور، این مدل‌ها به اختصار به‌ترتیب به صورت $M1$، $M2$، TGP، PPM، $LaSBP$ و $BRSBP$ نمایش داده می‌شوند. داده‌های مورد بررسی متوسط میزان بارش برحسب میلی‌متر مربوط به ۴۱۳ ایستگاه هواشناسی در سال ۲۰۱۶ برای کشور ایران است. این داده‌ها به وسیله‌ی سازمان هواشناسی و وزارت نیرواندازه‌گیری شده است. در ادامه برای بررسی نامانایی داده‌ها در ساختار دوم فرایند از آزمون نامانایی که توسط سویتریاندیوپدی و سوبارائو (۲۰۱۷) بیان گردید، استفاده شده است که در سطح ۵ درصد مانایی داده‌ها رد می‌شود. همچنین با توجه به این‌که میزان بارش به متغیرهای هواشناسی از جمله دما وابسته است، دما به عنوان متغیر کمکی برای برازش مدل‌ها انتخاب می‌کنیم. از آن‌جا که هدف این مقاله بررسی تاثیر متغیر کمکی در همبستگی فضایی است، ابتدا روند از داده‌ها حذف و با داده‌های بدون روند برازش مدل‌ها را انجام می‌دهیم.

در ادامه از معیار میانگین توان دوم خطا ($MSPE$) برای ارزیابی قدرت پیشگویی مدل‌های بیان شده، استفاده می‌شود. برای این منظور ۱۰ مجموعه داده با حذف تصادفی ۱۰ درصد از داده‌ها تولید شده و سپس با استفاده از رهیافت بیزی مدل‌های $M1$، $M2$، TGP، PPM، $LaSBP$ و $BRSBP$ به داده‌ها برازش داده شده است. به منظور خلاصه‌سازی، جزئیات انتخاب پیشینی و نحوه‌ی استنباط بیزی این مدل‌ها بیان نشده است. در جدول ۱ نتایج هر یک از مدل‌ها براساس معیار $MSPE$ آمده است. همان‌طور که

جدول ۱: مقایسه‌ی عملکرد پیشگویی مدل‌های: $M۱$، $M۲$، $LaSBP$، TGP، PPM و $BRSBP$.

مدل	M۱	M۲	LaSBP	BRSBP	PPM	TGP
MSPE	۱٫۹۲	۱٫۳۶	۱٫۰۸	۰٫۶۴	۰٫۸۸	۰٫۷۶

مشاهده می‌شود مدل $BRSBP$ عملکرد بهتری نسبت به دیگر مدل‌ها دارد. به‌خصوص، مدل‌های $M۱$ و $M۲$ عملکرد ضعیف‌تری نسبت به مدل پیشنهادی دارند. در واقع مدل $BRSBP$ مقدار $MSPE$ کمتری نسبت به مدل‌های $M۱$، $M۲$، PPM و TGP داراست. مقدار $MSPE$ برای مدل پیشنهادی ۰٫۶۴ و برای مدل $LaSBP$ برابر با ۱٫۰۸ است که تقریباً ۴۱٪ کوچکتر است.

در شکل ۱ نمودار افرازبندی ایستگاه‌های هواشناسی حاصل از برازش مدل $BRSBP$ آمده است. همانطور که مشاهده می‌شود ایستگاه‌هایی که رفتار دمایی یکسانی دارند در خوشه‌های یکسان قرار گرفته‌اند (شکل ۱ب). به عنوان مثال، مناطق شرقی و جنوب شرقی ایران که دارای متوسط دمای بالا و میزان بارش کم هستند، در خوشه‌های یکسان قرارگرفته‌اند. همچنین، در شکل‌های ۱ج، ۱د و ۱ه به‌ترتیب افرازبندی ایستگاه‌های هواشناسی حاصل از برازش مدل‌های $LaSBP$، PPM و TGP آمده است. همانطور که مشاهده می‌شود، افرازبندی ایجاد شده توسط مدل‌های $LaSBP$ و PPM تقریباً شبیه مدل $BRSBP$ است اما تعداد خوشه‌های ایجاد شده براساس مدل $BRSBP$ کمتر است. اما در مدل TGP ایستگاه‌های هواشناسی نزدیک به هم، در خوشه‌های یکسان قرارگرفته‌اند. با این حال، با توجه به تنوع آب و هوایی که در ایستگاه‌های نزدیک به هم وجود دارد، افرازبندی ایجاد شده توسط مدل TGP منطقی و معقول به‌نظر نمی‌رسد.

بحث و نتیجه‌گیری

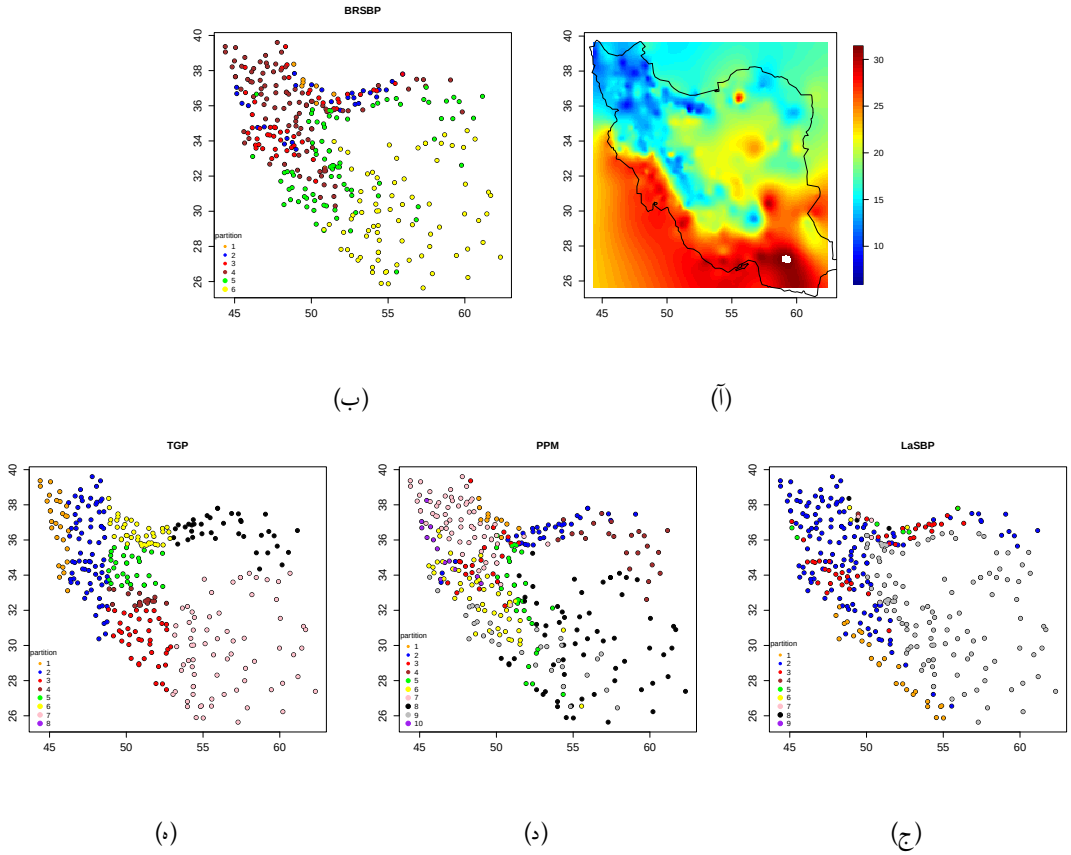
در این مقاله با استفاده از فرایند رگرسیون-بتا، رویکرد جدیدی برای ساخت فرایندهای فضایی مورد مطالعه قرار گرفت به‌طوری‌که توزیع تصادفی وابسته به متغیر کمکی است. همانند فرایندهای چوب‌شکنی تعمیم‌یافته تابع کوواریانس مدل معرفی شده نامانا و وابسته به متغیر کمکی است. از ویژگی‌های مهم مدل معرفی شده می‌توان به خوشه‌بندی بر اساس متغیر کمکی اشاره کرد که باعث بهبود در افرازبندی ناحیه مورد مطالعه می‌شود. همچنین با استفاده از یک مجموعه داده واقعی عملکرد مدل از نظر پیشگویی فضایی و خوشه‌بندی مورد ارزیابی قرار گرفت.

مراجع

Bandyopadhyay, S., and Rao, S. S. (2017), *A test for stationarity for irregularly spaced spatial data*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79, 95–123.

Barcella, W., M. De Iorio, S. Favaro, and G. L. Rosner (2018), *Dependent generalized dirichlet process priors for the analysis of acute lymphoblastic leukemia*, Biostatistics, 19(3), 342–358.

Barzegar, Z., and Rivaz, F. (2020), *A scalable Bayesian nonparametric model for large spatio-temporal data*, Computational Statistics, 35, 153–173.



(ب) (آ)

(ه) (د) (ج)

شکل ۱: (آ) نمودار رویه دما و (ب، ج، د ، ه) نمودارهای افرازبندی ایستگاه‌های هواشناسی از برازش مدل‌های $BRSBP$، $LaSBP$، PPM و TGP.

Dahdouh, O., and Khaledi, M. J., (2020), *Generalized spatial stick-breaking processes*, Communications in Statistics Simulation and Computation, 1-20.

Dunson, D. B. and J.-H. Park (2008),*Kernel stick-breaking processes*, Biometrika 95(2), 307–323.

Gramacy, R. B. and H. K. H. Lee (2008), *Bayesian treed gaussian process models with an application to computer modeling*. Journal of the American Statistical Association, 103(483), 1119–1130.

Ingebrigtsen, R., F. Lindgren, and I. Steinsland (2014), *Spatial models with explanatory variables in the dependence structure*, Spatial Statistics 8, 20–38.

Page, G. L. and F. A. Quintana (2016), *Spatial product partition models*, Bayesian Analysis 11(1), 265–298. 149.

Reich, B. J. and M. Fuentes (2007), *A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields*, The Annals of Applied Statistics 1(1), 249–264.

Reich, B.J., Eidsvik, J., Guindani, M., Nail, A.J., Schmidt, A.M., (2011), *A class of covariate-dependent spatiotemporal covariance functions for the analysis of daily ozone concentration*, The annals of applied statistics. 5 (4), 2425—2447.

Ren, L., L. Du, L. Carin, and D. B. Dunson (2011), *Logistic stick-breaking process*, Journal of Machine Learning Research 12(1).

Risser, M.D., Calder, C.A., (2015), *Regression-based covariance functions for nonstationary spatial modeling*, Environmetrics 26 (4), 284—297.

Rodriguez, A. and D. B. Dunson (2011), *Nonparametric Bayesian models through probit stick-breaking processes*, Bayesian Analysis 6(1).

Rodríguez, A., Dunson, D.B. and Gelfand, A.E., (2010), *Latent stick-breaking processes*, Journal of the American Statistical Association, 105(490), pp.647-659.

Schmidt, A. M., P. Guttorp, and A. O'Hagan (2011), *Considering covariates in the covariance structure of spatial processes*, Environmetrics 22(4), 487–500.

Yarali, E., Rivaz, F., (2020), *Incorporating covariate information in the covariance structure of misaligned spatial data*, Environmetrics 31 (6), e2623.

Yarali, E., Rivaz, F., and Khaledi, M. J. (2022), *A Bayesian nonparametric spatial model with covariate-dependent joint weights*, Spatial Statistics, 100662.

Beta regression stick-breaking process with covariate dependent nonstationary covariance

Esmail Yarali, Firoozeh Rivaz[1], and Majid Jafari Khaledi[2]

[1]Department of Statistics, Shahid Beheshti University, Tehran.

[2]Department of Statistics, Tarbiat Modares University, Tehran.

Abstract: This paper introduces a new class of stick-breaking processes that incorporate covariate information into the random joint distributions of spatial processes. The proposed modeling approach allows for a nonstationary covariance function that is driven by covariates, and it also induces a covariate-dependent random partitioning scheme. Using this process as a random effect distribution within a linear mixed model framework, the data can be effectively modeled, and a Bayesian inference approach is developed. The performance of the proposed model is evaluated and compared with that of competing spatial models using a real-world example.

Keywords: Stick-breaking process, Beta-Regression, Covariate information, Non-stationarity, Covariance structure.

Mathematics Subject Classification (2020): 62H11, 62M30, 62F15.
