

یک مدل فضایی نامخلوط بیزی مبتنی بر مدل منعطف گامای شمارشی

مهسا نادى فر^۱، حسین باغیشنى^۱، افشین فلاح^۲

^۱ گروه آمار دانشگاه صنعتی شاهرود

^۲ گروه آمار دانشگاه بین‌المللی امام خمینی

چکیده: در تحلیل داده‌های شمارشی فضایی عموماً با دو چالش عمده روبه‌رو می‌شویم. اول، نادیده گرفتن هم‌خطی بین اثرات ثابت و فضایی که منجر به مشکل مخلوط فضایی و استنباط‌های غیر قابل تکیه می‌شود. دوم، عدم امکان استفاده از توزیع پواسون به دلیل پراکندگی نامتعادل اکثر داده‌های شمارشی است. برای مواجه شدن با این چالش‌ها، در این مقاله، یک مدل فضایی بیزی جدید با انعطاف بالا را با ترکیب روش‌های نامخلوط فضایی و مدل‌بندی پراکنش داده‌های مبتنی بر توزیع گامای شمارشی، معرفی می‌کنیم. برای برازش و استنباط پیرامون مدل، روش بیزی تقریبی لاپلاس آشیانی جمع‌بسته را به کار می‌گیریم. به‌ویژه، کاربرست مدل را در تحلیل داده‌های سرطان معده در کشور اسلوونی و پهنه‌بندی آن نشان می‌دهیم.

واژه‌های کلیدی: مدل فضایی نامخلوط، داده‌های شمارشی، بیش‌پراکنش، کم‌پراکنش، INLA.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62H11, 62F15, 62J12.

۱ مقدمه

مدل آمیخته خطی تعمیم‌یافته فضایی (SGLM) که اولین بار توسط بی‌سگ و همکاران (۱۹۹۱) برای داده‌های فضایی شبکه‌ای معرفی شدند، یک مدل سلسله‌مراتبی شامل داده‌هایی دارای وابستگی فضایی بر روی واحدهای نمایه شده مکانی، مانند مناطق یک کشور، است. مدل‌های SGLM این وابستگی مکانی را از طریق یک میدان تصادفی مارکوفی گاوسی پنهان، وارد مدل می‌کند (رو و هلد، ۲۰۰۵). در سال‌های اخیر، مساله مهمی که در تحلیل داده‌های فضایی مطرح است، برآوردیابی اثرات ثابت در حضور وابستگی فضایی در یک یا چند متغیر تبیینی است. علیرغم کاربردهای گسترده و انعطاف پذیر مدل‌های SGLM، هنگامی که بین اثرات ثابت و تصادفی همبستگی وجود داشته باشد، این مدل نمی‌تواند در بازیابی اثرات ثابت و تصادفی به خوبی عمل کند. در متون مدل‌بندی فضایی، هم‌خطی بین اثرات تصادفی و اثرات ثابت را به

* نام و آدرس پست الکترونیکی ارائه‌دهنده مقاله: مهسا نادى فر، m.nadifar@shahroodut.ac.ir

اصطلاح مخلوط فضایی^۱ می‌نامند. در این زمینه مطالعات گسترده‌ای به‌ویژه در سال‌های اخیر انجام شده است. ابتدا کلاپتون و همکاران (۱۹۹۳) مساله مخلوط فضایی را که حاصل از هم‌خطی اثرات ثابت و تصادفی است، مطرح کردند. سپس رایس و همکاران (۲۰۰۶) نشان دادند مواقعی که متغیرهای تبیینی از الگوی فضایی خاصی پیروی می‌کنند، ممکن است با اثرات تصادفی آمیخته شوند و مشکل مخلوطی فضایی را ایجاد کنند. آن‌ها برای رفع این مشکل، مدل مخلوط فضایی تحت عنوان RHZ را معرفی کردند. مدل RHZ افزای از اثرات فضایی را که به فضای متعامد ایجاد شده توسط متغیرهای تبیینی تصویر شده را از مدل حذف می‌کنند. پریس و همکاران (۲۰۱۹) نیز مدل جدیدی را تحت عنوان SPOCK معرفی کردند که بدون توجه به اثرات تصادفی و تنها با تغییر ماتریس همسایگی، مشکل مخلوطی فضایی را کاهش می‌دهد. هم‌چنین، دوپونت و همکاران (۲۰۲۰) یک روش ناپارامتری جدید برای مقابله با مشکل مخلوطی فضایی ارائه کردند. آن‌ها با استفاده از توابع هموارساز اسپلاین دو بعدی (TPS) تحت عنوان S+، این مساله را با رویکرد فراوانی‌گرایانه مورد مطالعه قرار دادند.

با توجه به ماهیت شمارشی اکثر داده‌های مورد بررسی در بیماری‌ها، یکی از مدل‌های رایج رگرسیون پواسون است. یکی از پذیره‌های اصلی این مدل رگرسیونی برابری میانگین و واریانس متغیر پاسخ است. اما این پذیره در بسیاری از کاربردها برقرار نیست و واریانس و میانگین مشاهدات پاسخ تفاوت معنی‌داری دارند. در حالتی که واریانس از میانگین بزرگتر است، بیش‌پراکنندگی و در حالتی که واریانس از میانگین کوچکتر است، کم‌پراکنندگی در داده‌ها وجود دارد. با استفاده از نظریه تجدید، وینکلن (۱۹۹۵) نشان داد با در نظر گرفتن توزیع‌های احتمالی مختلف برای زمان‌های بین رخداد پیشامدها، می‌توان به توزیع‌های شمارشی جدیدی برای تعداد رخدادها دست یافت که انعطاف‌پذیری بیشتری برای تحلیل داده‌ها دارند. اگر توزیع زمان‌های بین رخداد پیشامدها گاما در نظر گرفته شود، توزیعی به نام گامای شمارشی (GC) به‌دست می‌آید. در این توزیع، پارامتری برای مدل‌بندی پراکنش داده‌ها وجود دارد که با در نظر گرفتن مقادیر مختلف برای آن می‌توان بیش‌پراکنش، کم‌پراکنش و پراکنش متعادل را در مدل لحاظ کرد. هم‌چنین، نادی‌فر و همکاران (۲۰۲۱)، مدل S+ را با رویکرد بیزی برای داده‌های شمارشی پراکنده توسعه دادند.

چالش اصلی در رویکرد بیزی، جنبه محاسباتی آن است. به این معنی که در بسیاری از مدل‌های بیزی از جمله مدل‌های فضایی، توزیع‌های پسین و کمیت‌های پسینی مورد علاقه دارای نمایش بسته نیستند. رهیافت معمول برای تقریب این نوع از توزیع پسین، الگوریتم‌های مونت کارلوی زنجیر مارکوفی (MCMC) است. اما این روش‌ها نیز در تقریب مدل‌های پیچیده از جمله مدل‌های فضایی با مشکلات جدی شامل سرعت کند و دشواری تشخیص همگرایی زنجیر مواجه می‌شوند. برای دوری از مشکلات محاسباتی و همگرایی الگوریتم‌های MCMC، رو و همکاران (۲۰۰۹) یک روش بیزی تقریبی با نام تقریب لاپلاس آشیانه‌ای جمع‌بسته (INLA) معرفی کردند که از نظر محاسباتی بسیار سریع‌تر از الگوریتم‌های MCMC است و دقت نتایج حاصل از آن با نتایج روش‌های MCMC رقابت می‌کند.

در این مقاله، یک مدل فضایی نامخلوط شمارشی مبتنی بر توزیع GC را برای تحلیل داده‌های فضایی شبکه‌ای سرطان معده در اسلوانی توسعه می‌دهیم. تحلیل مدل پیشنهادی را با استفاده از رویکرد بیزی و روش INLA انجام دادیم. در بخش ۲، مدل رگرسیونی فضایی GC به‌طور خلاصه شرح داده می‌شود. سپس مدل رگرسیون فضایی مخلوط GC و چگونگی برازش آن با استفاده از روش INLA را در بخش ۳ مورد بررسی قرار می‌دهیم. در بخش ۴، داده‌های مربوط به سرطان معده در اسلوانی را با کمک مدل فضایی مخلوط GC تحلیل می‌کنیم. در پایان، بحث و نتیجه‌گیری را ارائه می‌کنیم.

¹ Spatial confounding

۲ مدل رگرسیون فضایی GC

فرض کنید $\{\tau_k\}$ ، $k \in \mathbb{N}$ ، دنباله زمانی مورد انتظار بین پیشامدهای $(k-1)$ ام و k ام باشد، به طوری که τ_k ها متغیرهایی مستقل و هم توزیع با توزیع گاما، میانگین $E(\tau) = \alpha/\gamma$ و واریانس $\text{Var}(\tau) = \alpha/\gamma^2$ هستند. می توان نشان داد (وینکلن، ۱۹۹۵) اگر Y_T نشان دهنده تعداد پیشامدها در بازه زمانی (\bullet, T) باشد، آنگاه متغیر تصادفی Y_T دارای توزیع گامای شمارشی با پارامترهای α و γ است، که با نماد $Y_T \sim \text{GC}(\alpha, \gamma)$ نشان داده می شود. تابع جرم احتمال این توزیع به صورت

$$p_{Y_T}(y) = G(y\alpha, \gamma T) - G((y+1)\alpha, \gamma T), \quad y = \bullet, 1, 2, \dots$$

است، که در آن

$$G(n\alpha, \gamma T) = \frac{1}{\Gamma(n\alpha)} \int_0^{\gamma T} x^{n\alpha-1} e^{-x} dx.$$

میانگین این توزیع نیز به صورت $E(Y_T) = \sum_{i=1}^{\infty} G(i\alpha, \gamma T)$ محاسبه می شود. پارامتر α در توزیع GC وضعیت پراکنش داده ها را کنترل می کند. زمانی که $\alpha = 1$ توزیع زمان های بین رخدادها به توزیع نمایی تبدیل می شود و بنابراین توزیع Y_T پواسون خواهد بود. برای مقادیر $\alpha < 1$ توزیع GC بیش پراکنش و در حالی که $\alpha > 1$ ، توزیع GC کم پراکنش ممکن در داده ها را مدل بندی خواهد کرد.

برای تعریف یک مدل رگرسیونی مبتنی بر توزیع GC، بر اساس نظریه مدل های خطی تعمیم یافته، باید میانگین متغیر پاسخ و متغیرهای تبیینی به کمک یک تابع پیوند مناسب، پیوند داده شوند. بدون کاستن از کلیت مسئله، فرض کنید طول بازه زمانی مورد انتظار برای رخداد پیشامدها، برای همه مشاهدات یکسان و برابر واحد است. در مدل رگرسیون GC، به دلیل نامتناهی بودن میانگین متغیر پاسخ، نمی توان مدل رگرسیونی را به طور مستقیم بر حسب میانگین پاسخ Y_i ، $i = 1, \dots, n$ ، نوشت. از این رو وینکلن (۱۹۹۵) مدل رگرسیونی را بر حسب میانگین زمان های مورد انتظار، τ_i ، تعریف کردند.

مدل رگرسیون GC برای داده های شبکه ای را با وارد کردن یک متغیر تصادفی فضایی، ϕ_i ، که فرض می کنیم از یک فرآیند اتورگرسیو شرطی ذاتی^۲ (ICAR) پیروی می کند، توسعه می دهیم. برای این منظور، فرض کنید مشاهدات در مکان های اندازه گیری شده، به شرط معلوم بودن میدان تصادفی، مستقل شرطی هستند و از توزیع GC پیروی می کنند. پیشگوی خطی در یک مدل رگرسیون GC فضایی به صورت

$$\log(E(\tau_i | \mathbf{X}_i)) = \log\left(\frac{\alpha}{\gamma(\mathbf{X}_i)}\right) = -\eta_i = -(\mathbf{X}_i' \boldsymbol{\beta} + \phi_i), \quad i = 1, \dots, n, \quad (1.2)$$

تعریف می شود. از این رو، مدل رگرسیون GC فضایی به صورت $Y_i | \eta_i, \alpha \sim \text{GC}(\alpha, \alpha \exp(\eta_i))$ ، $i = 1, \dots, n$ ، به دست می آید، که در آن $Y_i = Y(s_i)$ متغیر پاسخ شمارشی در نواحی $\mathbf{S} = (s_1, \dots, s_n)'$ است و $\boldsymbol{\Phi} = (\phi_1, \dots, \phi_n)'$ اثرات فضایی را نشان می دهند که

$$\boldsymbol{\Phi} | \tau_{\phi} \sim N\left(\bullet, (\tau_{\phi} \mathbf{A})^{-1}\right), \quad (2.2)$$

که در آن \mathbf{D} ، $\mathbf{A} = (\mathbf{D} - \mathbf{W})$ ماتریس قطری با درایه های w_{i+} ، $i = 1, \dots, n$ ، ماتریس متشکل از w_{ij} ها و τ_{ϕ} پارامتر دقت مدل است. لازم به ذکر است که اگر نواحی i و j همسایه باشند $w_{ij} = 1$ و در غیر این صورت صفر تعریف می شود. محاسبه می شود.

۳ مدل رگرسیون فضایی نامخلوط GC

در این بخش برخی از مدل‌های فضایی نامخلوط را که برای برطرف کردن مشکل هم‌خطی اثرات ثابت و تصادفی ارائه شده‌اند را با رویکرد بیزی و روش INLA، برای مدل GC می‌نویسیم.

مدل RHZ. رایش و همکاران (۲۰۰۶) مدل RHZ را برای مقابله با مساله مخلوط فضایی معرفی کردند. ایده RHZ مبتنی بر افراز اثرات تصادفی فضایی به دو زیرفضا و حذف زیرفضای متعامد با اثرات ثابت است. بدین منظور، اثر تصادفی فضایی در پیشگویی خطی (۱.۲)، را می‌توان به صورت $\Phi = \Phi^X + \Phi^\perp = \mathbf{H}\Phi_1 + \mathbf{B}\Phi_2$ نوشت، که در آن \mathbf{H} یک ماتریس $(n \times q)$ بعدی است که اندازه‌ای مشابه با ماتریس طرح \mathbf{X} دارد، \mathbf{B} یک ماتریس $n \times (n - q)$ بعدی است که ستون‌های آن در فضای متعامد \mathbf{X} قرار دارد و افزایهای Φ_1 و Φ_2 به ترتیب بردارهایی با بعد q و $n - q$ هستند. آن‌ها نشان دادند که ماتریس H باعث مخلوطی فضایی در مدل می‌شود. بنابراین با حذف مولفه مربوط به H ، پیشگویی خطی برای مدل RHZ مبتنی بر توزیع GC به صورت زیر نوشته می‌شود:

$$\eta_i = \mathbf{x}'_i \beta + \mathbf{B}_i \Phi_2, \quad \Phi_2 | \tau_\phi \sim N\left(0, (\tau_\phi \mathbf{B}' \mathbf{A} \mathbf{B})^{-1}\right). \quad (1.3)$$

مدل SPOCK. پریتس و همکاران (۲۰۱۹) مدل جایگزین دیگری به نام SPOCK معرفی کردند که مشکلات حاصل از همگرایی مدل RHZ را مرتفع می‌کند. مدل SPOCK بدون افراز فضای اثرات تصادفی و با تغییر ماتریس همسایگی، مشکل مخلوط فضایی را کاهش می‌دهد. مدل SPOCK، ماتریس همسایگی را در فضای متعامد ماتریس طرح، تصویر می‌کند.

مدل S+. دوپونت و همکاران (۲۰۲۰) یک روش جدید مبتنی بر روش‌های ناپارامتری و توابع TPS در رویکرد فراوانی‌گرایانه تحت عنوان S+ معرفی کردند. برای معرفی مدل S+ ابتدا یک مدل فضایی متناظر با مدل (۱.۲) با استفاده از تابع TPS را شرح می‌دهیم. بدین منظور، پیشگویی خطی ارائه شده در رابطه (۱.۲) را به صورت $\eta_i = \mathbf{x}'_i \beta + f_{tps}^+(\mathbf{s}_i)$ بازنویسی می‌کنیم، که در آن $f_{tps}^+(\mathbf{s}_i)$ تابع TPS برای نواحی $i = 1, \dots, n$ است و معادل ناپارامتری اثر تصادفی ICAR در رابطه (۲.۲) است. سپس مدل S+ به این صورت تعریف می‌شود که ابتدا متغیرهای تبیینی مخلوط شده با اثر فضایی را با توابع هموارساز TPS به صورت $\mathbf{x}_i = f_{tps_x}(\mathbf{s}_i) + \varepsilon$ ، مدل می‌کنیم، که در آن $f_{tps_x}(\cdot)$ به این معنی است که \mathbf{x} به $f_{tps}^+(\cdot)$ از طریق مولفه $f_{tps_x}(\cdot)$ وابسته است. از این رو، ابتدا باقیمانده‌های معادله رگرسیون $r(\mathbf{X}) = (r(\mathbf{x}_1), \dots, r(\mathbf{x}_n))'$ را به صورت $r(\mathbf{X}) = \mathbf{X} - \hat{f}_{tps_x}(\mathbf{s})$ به دست آورده و به جای متغیر تبیینی \mathbf{x} در پیشگویی خطی جایگزین می‌کنیم. بنابراین، پیشگویی خطی برای یک مدل S+ به صورت زیر نوشته می‌شود:

$$\eta_i = r(\mathbf{x}_i) \beta + f_{tps}^+(\mathbf{s}_i), \quad (2.3)$$

به دلیل وجود انتگرال چندبعدی در توابع درستنمایی حاصل از پیشگوهای خطی در رابطه‌های (۱.۳)، (۲.۳) و بعد معمولاً بالای میدان تصادفی، n ، استنباط مدل با رویکرد مبتنی بر درستنمایی همواره یک چالش جدی است. در این راستا، **نادی‌فر و همکاران (۲۰۲۱)**، مدل S+ را با رویکرد بیزی و روش INLA توسعه دادند که در بخش بعدی توضیح خواهیم داد.

فرض کنید $\mathbf{y}(s)$ یک میدان تصادفی از توزیع گامای شمارشی باشد، برای توسعه مدل S+ در رویکرد بیزی، برای $f_{tps}(\cdot)$ مدل قدم زدن تصادفی مرتبه دوم دو بعدی (RW2D) را در نظر می‌گیریم (رو و هلد، ۲۰۰۵). برای انجام استنباط بیزی پیرامون مدل پیشنهادی، لازم است که توزیع‌های پیشین مناسب برای پارامترها و ابرپارامترهای مدل تعیین شوند. با در نظر گرفتن بردار پارامترهای $\theta = (\alpha, \beta, \tau_\phi)$ و با فرض استقلال پیشینی پارامترها می‌توان نوشت $\pi(\theta) = \pi(\alpha)\pi(\beta)\pi(\tau_\phi)$. با توجه به محدودیت روش INLA (رو و همکاران، ۲۰۰۹)، مولفه‌های بردار ضرایب رگرسیونی β ، به صورت مستقل، دارای توزیع پیشین گاوسی با میانگین صفر و واریانس 10^2 در نظر گرفته شده است. همچنین برای

α ، نادى‌فر و همکاران (۲۰۲۱) پیشین جدیدی را مبتنی بر پیشین‌های با پیچیدگی جریمه‌شده (PC) تعریف کردند که با INLA قابل اجرا نیز است. برای پارامتر دقت τ_ϕ نیز توزیع پیشین PC توصیه می‌شود (سیمپسون و همکاران، ۲۰۱۷). در نهایت مدل S+ بیزی مبتنی بر توزیع GC به صورت

$$\pi(\alpha, \beta, f_{tps_x}(\mathbf{s}), \tau_x, f_{tps}^+(\mathbf{s}) | \mathbf{y}) \propto \prod_{i=1}^n \{GC(\alpha, \alpha \exp(\eta_i)) \pi(\beta) \pi(\alpha) RW2D(f_{tps_x}(\mathbf{s})) \pi(\tau_x) RW2D(f_{tps}^+(\mathbf{s}))\}.$$

به دست می‌آید که فرم بسته‌ای ندارد و انجام استنباط بیزی مبتنی بر MCMC دشوار و از کارایی ضعیفی برخوردار است. در نتیجه از رهیافت بیزی تقریبی مبتنی بر روش INLA، برای برآورد کمیت‌های مورد علاقه توزیع پسین استفاده می‌کنیم. این روش برای انجام استنباط بیزی تقریبی توسط رو و همکاران (۲۰۰۹) معرفی و پیشنهاد شد، که در آن تکنیک‌های تقریب لاپلاس و انتگرال‌گیری عددی به طریقی بسیار کارا با هم ترکیب می‌شوند، به طوری که شبیه‌سازی‌های سنگین MCMC با تقریب‌های دقیق جایگزین می‌شوند. روش INLA، در اغلب موارد، با دقتی عمل می‌کند که حتی اختلاف‌های بین نتایج حاصل از آن و نمونه‌گیری‌های MCMC، جزئی و غیرقابل تشخیص هستند.

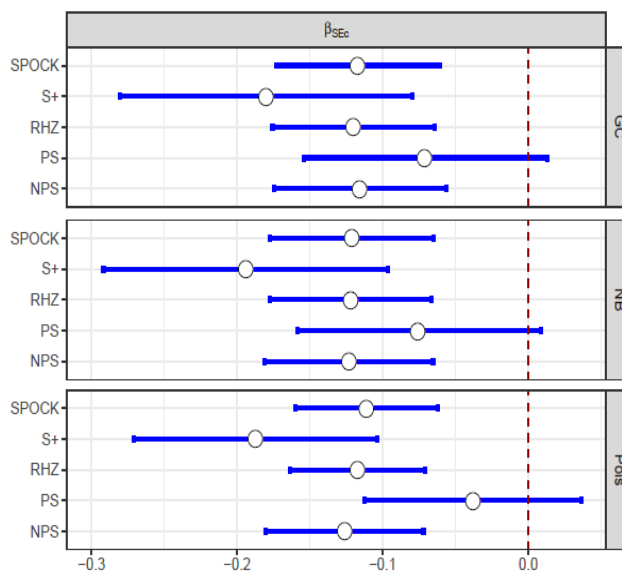
۴ تحلیل داده‌های بیش‌پراکنده سرطان معده اسلوانی

برای کاربست مدل پیشنهادی، تحلیل داده‌های مربوط به تعداد مرگ و میر بیماران مبتلا به سرطان معده در ۱۹۲ ناحیه در اسلوانی طی سال‌های ۱۹۹۵ تا ۲۰۰۵ را مد نظر قرار دادیم. این مجموعه داده دارای یک متغیر تبیینی مربوط به عامل اقتصادی-اجتماعی است. پیشگوی خطی را برای تحلیل این داده‌ها به صورت $\eta_i = \log(E_i) + \beta_0 + \beta_{SEC} SEC_i + \phi_i$ ، $i = 1, \dots, 192$ ، تعریف می‌کنیم که در آن، E_i ، β_{SEC} ، β_0 و ϕ_i به ترتیب نشان‌دهنده عرض از مبدأ، ضریب اثر ثابت برای متغیر تبیینی SEC، متغیر تبیینی فرعی و اثر فضایی هستند. سه مدل رگرسیونی GC، پواسون و دوجمله‌ای منفی را تحت مدل‌های بیزی S+، SPOCK، RHZ و مدل فضایی پارامتری PS و ناپارامتری NPS بدون در نظر گرفتن مساله مخلوط فضایی را به مدل برازش دادیم. برای نشان دادن تاثیر مشکل مخلوط فضایی در برآورد پارامترها، نمودار جنگل را برای برآورد ضریب اثر SEC در شکل ۱ نمایش دادیم. این شکل فاصله HPD در سطح ۹۵٪ و برآورد میانگین پسینی با دایره سفید رنگ را به‌ازای مدل‌های مختلف نشان می‌دهد.

همان‌طور که مشاهده می‌شود فاصله HPD برای مدل PS به‌ازای همه توزیع‌ها، وقتی که اثر مخلوط فضایی را در مدل لحاظ نمی‌کنیم، دربردارنده ۰ است و به این معنی است که رابطه متغیر اقتصادی-اجتماعی، SEC، با تعداد افراد مبتلا به سرطان معده، معنی‌دار نیست.

هم‌چنین، برآورد پارامترها به همراه فاصله معتبر بیزی آن‌ها و برخی از معیارهای انتخاب مدل مانند DIC و WAIC، برای مدل‌های نامخلوط فضایی در جدول ۱ گزارش شده‌اند. با توجه به برآورد تقریبی ۰/۵ برای α و فاصله اعتبار بیزی آن و با توجه به ویژگی توزیع GC، مشاهدات بیش‌پراکنده هستند. به‌طور کلی نتایج جدول بیانگر این هستند که مدل RHZ بر اساس معیار WAIC و مدل SPOCK بر اساس معیار DIC بهترین عملکرد را دارند.

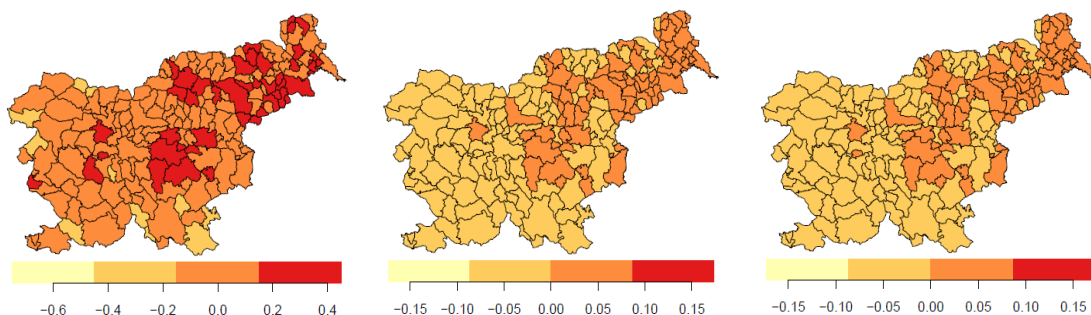
شکل ۲ پهنه‌بندی اثرات فضایی برآورد شده تحت مدل‌های RHZ، SPOCK و S+ و برای همه توزیع‌ها نشان می‌دهد. همان‌طور که مشاهده می‌شود عملکرد مدل‌های GC و دوجمله‌ای منفی، که برای مشاهدات بیش‌پراکنده مناسب هستند، شبیه اما مدل پواسون کمی متفاوت است و در برخی نواحی وابستگی بیشتری را نشان می‌دهد.



شکل ۱: برآورد میانگین پسینی به همراه فاصله ۹۵٪ HPD برای ضریب اثر متغیر SEC تحت مدل‌های مختلف شمارشی و روش‌های نامخلوط فضایی.

جدول ۱: مقادیر میانگین پسینی پارامترها به همراه فاصله اعتبار ۹۵٪ HPD و معیارهای انتخاب مدل برای توزیع‌ها و مدل‌های مخلوط فضایی مختلف.

WAIC	DIC	τ (فاصله HPD)	پارامتر پراکندگی (فاصله HPD)	مدل	توزیع
۱۱۲۶/۲۷۷	۹۸۵/۵۴۳	۸/۸۲۱(۳/۰۹۹, ۱۶/۵۸۸)	—	S+	پواسون
۱۱۱۹/۰۵۷	۹۷۹/۲۴۹	۹/۹۹۹(۲/۸۸۳, ۲۱/۷۲۰)	—	SPOCK	
۱۰۸۹/۹۶۰	۹۵۹/۴۲۱	۷/۳۵۲(۳/۲۰۱, ۱۲/۶۹۵)	—	RHZ	
۱۱۰۶/۷۷۵	۸۷۶/۹۳۷	۵۱/۴۶۴(۲/۹۱۶, ۱۴۲/۴۷۶)	۰/۵۵۳(۰/۴۲۳, ۰/۶۸۸)	S+	GC
۱۱۰۳/۴۴۴	۸۶۹/۳۰۳	۴۴۴/۳۰۰(۰/۲۹۵, ۲۹۴۳/۸۵۵)	۰/۵۴۳(۰/۴۱۹, ۰/۶۷۰)	SPOCK	
۱۱۰۲/۷۳۹	۸۷۳/۳۷۱	۳۲۲۹۹/۱۶۰(۰/۱۴۲, ۴۳۴۸۹۱/۷۳۴)	۰/۵۵۴(۰/۴۲۲, ۰/۶۹۱)	RHZ	
۱۲۰.۱۱۱۴	۷۱۰.۹۹۰	(۳۹۹.۲۲۶, ۴۱۹.۰) ۰.۹۲.۷۷	(۹۷۴.۳۹, ۶۴۶.۱۰) ۸۸۵.۲۳	S+	دوجمله‌ای منفی
۱۱۰۹/۷۲۴	۹۸۶/۶۱۱	۶۶۹/۲۳۸(۰/۱۱۷, ۴۷۹۳/۶۹۲)	۲۴/۰.۱۷(۱۱/۱۷۶, ۳۹/۶۰۹)	SPOCK	
۱۱۰۹/۶۶۰	۹۸۶/۵۳۳	۵۰۱۴/۹۷۰(۰/۱۳۸, ۲۵۷۸۵۸/۸۸۱)	۲۲/۵۵۲(۱۰/۸۳۸, ۳۶/۳۴۴)	RHZ	



شکل ۲: برآورد میانگین پسینی اثرات فضایی، مدل GC (راست)، دوجمله‌ای منفی (میانی) و پواسون (چپ) تحت مدل SPOCK.

بحث و نتیجه‌گیری

زمانی که بین اثرات ثابت و تصادفی، هم‌خطی وجود داشته، یا به عبارتی مشکل مخلوط فضایی برقرار باشد، مدل‌های فضایی رایج عملکرد مناسبی نخواهند داشت. مشکل مخلوط فضایی می‌تواند بر استنباط‌ها، به‌ویژه معنی‌داری اثرات ثابت، تاثیر زیادی داشته باشند. از طرفی، اکثر داده‌های شمارشی، دچار بیش‌پراکنش یا کم‌پراکنش هستند و مدل پواسون برای مدل‌بندی آن‌ها مناسب نیست. در این مقاله، یک مدل فضایی نامخلوط بیزی جدید بر اساس سک مدل پواسون منعطف، با نام گامای شمارشی، ارائه کردیم که هر دو مشکل مخلوطی فضایی و پراکندگی نامتعادل مشاهدات را مرتفع می‌کند. مدل پیشنهادی با یک مثال کاربردی برای تحلیل و پهنه‌بندی نرخ افراد مبتلا سرطان معده در کشور اسلوانی مورد ارزیابی قرار داده شد. نتایج به‌دست آمده نشان دادند که مدل پیشنهادی برای برازش مشاهدات بیش‌پراکنده، نسبت به مدل‌های متداول پواسون و دو جمله‌ای منفی، از برتری خوبی برخوردار است.

مراجع

- Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics, *Annals of the Institute of Statistical Mathematics*, **43(1)**, 1-20.
- Clayton, D., Bernardinelli, L., and Montomoli, C. (1993). Correlation in ecological analysis, *International Journal of Epidemiology*, **6**, 1193-1202.
- Dupont, E., Wood, S. M., and Augustin, N. (2020). Spatial+: a novel approach to spatial confounding, Available from: arxiv.org/abs/2009.09420v1.
- Nadifar, M., Baghishani, H., Kneib, T., and Fallah, A. (2021). Flexible Bayesian modeling of counts: constructing penalized complexity priors, Available from: arxiv.org/abs/2105.08686.
- Nadifar, M., Baghishani, H. and Fallah, A. (2021). A flexible Bayesian non-confounding spatial model for analysis of dispersed count data in clinical studies, Available from: arxiv.org/abs/2105.09893.
- Prates, M. O., Assuncao, R. M., and Rodrigues, E. C. (2019). Alleviating spatial confounding for areal data problems by displacing the geographical centroids, *Bayesian Analysis*, **14(2)**, 623-647.
- Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models, *Biometrics*, **62**, 1197-1206.
- Rue, H., Martino, S., Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using Integrated Nested Laplace Approximations, *Journal of the Royal Statistical Society, B*, **71**, 319-392.
- Rue, H. and Held, L. (2005). *Gaussian markov random fields: Theory and Applications*. Chapman & Hall/CRC Press, London.

Simpson, D., Rue, H., Riebler, A., Martins, T. G. and Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors, *Statistical Science*, **32**, 1-28.

Winkelmann R. (1995), Duration dependence and dispersion in count-data models, *Journal of Business and Economic Statistics*, **13**, 467 - 474.