



مدل‌بندی صریح ساختار وابستگی برای تحلیل بیزی داده‌های بقای فضایی: برآورد خطر نسبی مبتلایان سرطان پروستات

زهرا نهاردانی، محسن محمدزاده*

گروه آمار، دانشکده علوم ریاضی، دانشگاه تربیت مدرس، تهران، ایران

دبیر مسئول: سیامک نوربلوچی

تاریخ پذیرش: ۱۴۰۱/۷/۱۸

تاریخ دریافت: ۱۴۰۰/۱۰/۲۰

چکیده: معمولا برای مدل‌بندی رابطه بین متغیر پاسخ با مشاهدات وابسته فضایی و متغیرهای تبیینی، از مدل‌هایی با اثرات تصادفی فضایی استفاده می‌شود. در بسیاری از زمینه‌های کاربردی برخی از متغیرهای تبیینی مدل نیز وابسته‌اند و بسته به نوع وابستگی آنها، استنباط آماری مدل‌های با اثرات تصادفی و کاربرد آنها پیچیده می‌شود. زیرا متغیرهای تبیینی، اثرات تصادفی و عبارت خطای مدل در توضیح تغییرپذیری متغیر پاسخ در رقابت با یکدیگر هستند. در این مقاله به منظور حل این مسئله، روشی برای مدل‌بندی و تحلیل داده‌های بقای فضایی پیشنهاد می‌شود که در آن به جای به‌کارگیری اثرات تصادفی فضایی در مدل، وابستگی فضایی مشاهدات به‌طور صریح در توابع چگالی، بقا و خطر منظور می‌شود و بر اساس آنها توابع درست‌نمایی وابسته فضایی محاسبه و برای استنباط آماری مدل مورد استفاده قرار می‌گیرد. سپس در یک مطالعه شبیه‌سازی برآورد بیزی اثرات متغیرهای تبیینی در مدل با استفاده از الگوریتم متروپولیس-هستینگز تطبیقی محاسبه و مورد ارزیابی قرار می‌گیرند. آنگاه نحوه کاربست روش ارائه شده را برای تحلیل داده‌های بیماران مبتلا به سرطان پروستات نشان داده و خطر نسبی مرگ بیماران با رهیافت بیزی برآورد می‌شود. در انتها بحث و نتیجه‌گیری ارائه خواهد شد.

واژه‌های کلیدی: داده بقای فضایی، مدل‌سازی صریح، اثرات تصادفی آمیخته.

رده‌بندی ریاضی: 62M30; 62H11

* نویسنده مسئول مقاله

رایانامه: mohsen_m@modares.ac.ir (M. Mohammadzadeh), z_nahardani@modares.ac.ir (Z. Nahardani).

۱ مقدمه

یکی از مفروضات اساسی در روش‌های معمول تحلیل بقا، استقلال داده‌ها است. اما گاهی با مواردی مواجه می‌شویم که مشاهدات به موقعیت و مکان نیز بستگی دارند، لذا موقعیت جغرافیایی واحد مورد مطالعه یا محل زندگی یک فرد بیمار می‌تواند در مدت زمان بقای وی اثر داشته باشد [۳]. در حالاتی که موقعیت داده‌های فضایی آدرس محل سکونت افراد است، ممکن است مکان دقیق موقعیت جغرافیایی در دسترس نباشد. اغلب به دلیل محرمانه بودن، این داده‌ها در نواحی مربوطه مانند بلوک، محله یا مناطق شهری جمع‌آوری یا گزارش می‌شوند و لازم است این اطلاعات در تحلیل مشاهدات لحاظ شوند.

در مواردی که داده‌های بقا به موقعیت مکانی مشاهدات بستگی ندارند، برای بررسی و تحلیل آنها از توابع بقا و خطر استفاده می‌شود. احتمال اینکه پیشامدی پس از t واحد زمانی رخ دهد، یعنی $S(t) = P(T > t) = 1 - F(t)$ ، تابع بقا است، که در آن T متغیر زمان بقا، $f(t)$ تابع چگالی T و $F(t)$ تابع توزیع تجمعی آن است. تابع خطر نیز امکان لحظه‌ای وقوع اتفاق مورد نظر به شرط عدم رخداد آن تا زمان t است و به صورت $h(t) = f(t)/S(t)$ تعیین می‌شود. مرسوم‌ترین تابع خطر توسط کاکس [۵] به صورت $h(t|\mathbf{X}) = h_0(t) \exp(\mathbf{X}\beta)$ تعریف شده است، که در آن ماتریس طرح متشکل از متغیرهای تبیینی، $\mathbf{X}_{n \times p} = (X_1, \dots, X_p)$ بردار اثرهای ثابت و $h_0(\cdot)$ تابع خطر پایه است. در برخی مطالعات عوامل تأثیرگذار بر داده‌های بقا قابل مشاهده یا قابل اندازه‌گیری نیستند. در این صورت از مدل‌های با اثرات تصادفی تحت عنوان مدل شکنندگی برای برازش به داده‌های بقا استفاده می‌شود. اما در مواردی که داده‌های بقا وابستگی مکانی یا فضایی داشته باشند، مدل‌های معمول تحلیل بقا، مانند کاکس و شکنندگی نمی‌توانند تحلیل دقیقی از زمان بقا و تابع خطر ارائه دهند زیرا همبستگی فضایی سبب ایجاد عوامل خطر ناشناخته می‌شود. علاوه بر آن این اثر فضایی ممکن است در عمل قابل اندازه‌گیری یا مشاهده نباشد. برای رفع این مشکل، [۸] رده‌ای از مدل‌های شکنندگی را تحت عنوان تابع خطر فضایی معرفی کردند، که با استفاده از ایده مدل‌های آمیخته با اثرات تصادفی، یک میدان تصادفی، مانند $Z(s)$ ، برای در نظر گرفتن همبستگی فضایی داده‌ها به تابع شکنندگی اضافه شد. این تابع خطر فضایی به صورت

$$h(t(s)|X, Z) = h_0(t(s)) \exp(X\beta + Z(s)), \quad s \in D \subset R^d, \quad (1.1)$$

تعریف شد، که در آن $t(s)$ زمان بقای فرد در موقعیت مکانی s ، $h_0(\cdot)$ تابع خطر پایه، X بردار متغیرهای تبیینی، β بردار پارامترهای رگرسیونی، $Z(s)$ یک میدان تصادفی برای منظور کردن همبستگی فضایی در مدل است. [۸] نشان دادند مدل شکنندگی فضایی شناساپذیر است و در صورت گاوسی بودن میدان تصادفی $Z(s)$ و معلوم بودن تابع هم‌تغییرنگار آن به صورت $C(h) = Cov(Z(s+h), Z(s))$ ، نتایج قابل قبولی ارائه می‌کند. اما در برخی مسائل کاربردی ممکن است فرض گاوسی برای اثرات تصادفی واقع بینانه نباشد یا تابع کوواریانس آن نامعلوم باشد. در این صورت استنباط آماری مدل (۱.۱) و به تبع آن تحلیل داده‌ها پیچیده‌تر و دقت نتایج نامطمئن‌تر می‌شوند، زیرا متغیرهای تبیینی و عبارت خطای فضایی در توضیح تغییرپذیری متغیر پاسخ در رقابت با یکدیگر اند. بنابراین ارائه روش‌هایی که از اثرات تصادفی فضایی استفاده نکنند، به‌ویژه در زمینه همبستگی بین متغیرهای تبیینی و ساختار همبستگی فضایی، می‌تواند بسیار راه‌گشا باشد. لذا هدف این مقاله توسعه و به‌کارگیری یک رویکرد بیزی برای تحلیل داده‌های بقای فضایی است که در آن به‌جای استفاده از اثرات تصادفی فضایی در مدل، از مدل‌بندی صریح وابستگی فضایی در سه تابع چگالی احتمال، تابع بقا و تابع خطر برای لحاظ کردن ساختار فضایی داده‌ها بهره گرفته شود. آنگاه بر اساس آنها توابع درستنمایی وابسته فضایی محاسبه و برای استنباط آماری مدل مورد استفاده قرار گیرد. سپس در یک مطالعه شبیه‌سازی برآورد بیزی اثرات متغیرهای تبیینی در مدل با استفاده از الگوریتم متروپولیس-هستینگز تطبیقی محاسبه و مورد ارزیابی قرار می‌گیرند. در ادامه نحوه کاربست روش ارائه شده برای تحلیل داده‌های بیماران مبتلا به سرطان پروستات نشان داده می‌شود و خطر نسبی مرگ بیماران با رهیافت بیزی برآورد خواهد شد. در انتها بحث و نتیجه‌گیری و پیشنهاد برای مطالعات آتی ارائه می‌شود.

۲ مدل غیرساختاری بقای فضایی-زمانی

در حالت کلی برای داده‌های بقای فضایی، تابع چگالی احتمال به دو مؤلفه فضایی و زمانی بستگی دارد و در تحلیل داده‌ها لازم است وابستگی بین دو این مولفه نیز در نظر گرفته شود. معمولاً مؤلفه فضایی روی فضای اقلیدسی دو بعدی R^2 و مؤلفه زمان روی فضای اعداد حقیقی R^+ تعریف می‌شود. فعلاً برای سادگی، صرفاً مواردی که مؤلفه‌های فضا و زمان مستقل از هم هستند مطرح می‌شود و حالتی که دو مولفه وابسته باشند به مطالعه‌ای دیگر موقوف خواهد شد. بنابراین فرض می‌شود تابع چگالی احتمال توأم دو مولفه به صورت $f(s, t) = f_1(s)f_2(t)$ است، که در آن $f_1(s)$ و $f_2(t)$ به ترتیب توابع چگالی صرفاً فضایی و صرفاً زمانی‌اند، s نشان‌دهنده موقعیت مکانی است که به صورت طول و عرض جغرافیایی بیان می‌شود و t بیانگر موقعیت زمانی است. حال توابع توزیع تجمعی برای مؤلفه زمانی، مؤلفه فضایی و همچنین توزیع آمیخته مؤلفه‌های فضایی و زمانی تعریف می‌شود. در نماد گذاری، حدود انتگرال‌گیری زمان، فضا و فضا-زمان به ترتیب با t ، s و (s, t) نمایش داده می‌شود. فرض کنید A_s یک زیرناحیه فضایی از R^2 و $t^* \in (0, \infty)$ یک مدت زمان مشخص باشد. تابع توزیع احتمال زمانی به صورت

$$F_t(t^*) = P(T \leq t^*) = \int_0^{t^*} f_2(t) dt$$

و توزیع احتمال روی فضای A_s ، یعنی احتمال آنکه یک موقعیت به طور تصادفی انتخاب شده در A_s باشد، به صورت

$$F_s(A_s) = \int_{A_s} f_1(s) ds$$

تعریف می‌شوند. توزیع توأم فضایی و زمانی، یعنی احتمال اینکه در موقعیت s متغیر تصادفی زمان، مقداری کمتر یا مساوی t^* اختیار کند، به صورت

$$F_t(s, t^*) = \int_0^{t^*} f(s, t) dt = \int_0^{t^*} f_1(s) f_2(t) dt = f_1(s) F_t(t^*),$$

است. احتمال اینکه متغیر تصادفی زمان، مقدار t^* را اختیار کند و موقعیت تصادفی در ناحیه A_s قرار گیرد، به صورت

$$F_s(A_s, t^*) = \int_{A_s} f(s, t) ds = \int_{A_s} f_1(s) f_2(t^*) ds = f_2(t^*) F_s(A_s),$$

و احتمال توأم اینکه متغیر تصادفی زمان مقداری کمتر یا مساوی t^* را اختیار کند و موقعیت تصادفی در ناحیه A_s قرار گیرد. به صورت

$$F_{s,t}(A_s, t^*) = F_s(A_s) F_t(t^*)$$

است. حال می‌توان توابع بقا و خطر کناری و توأم متناظر با آنها را به ترتیب به صورت

$$S_t(s, t^*) = 1 - F_t(s, t^*), \quad S_{s,t}(A_s, t^*) = P(s \in A_s, t > t^*) = F_s(A_s)(1 - F_t(t^*)),$$

$$h_t(s, t^*) = \frac{f(s, t^*)}{S_t(s, t^*)}, \quad h_{s,t}(A_s, t^*) = \frac{\int_{A_s} f(s, t^*) ds}{S_{s,t}(A_s, t^*)} = \frac{F_s(A_s, t^*)}{F_s(A_s)(1 - F_t(t^*))},$$

تعریف کرد، که در آن $S_t(s, t^*)$ ، توزیع تجمعی بقا بیش از زمان t^* برای فردی در موقعیت s ، $S_{s,t}(A_s, t^*)$ توزیع تجمعی توأم بقا بیش از زمان t^* در ناحیه A_s ، $h_t(s, t^*)$ احتمال خطر لحظه‌ای در زمان t^* برای فردی در موقعیت s و $h_{s,t}(A_s, t^*)$ توزیع توأم خطر لحظه‌ای در زمان t^* در ناحیه A_s است.

فرض کنید t_1, \dots, t_n زمان‌های بقای n فرد (یا واحد آماری) باشند که مستقل و هم‌توزیع با تابع چگالی $f(s, t)$ هستند،

مدل‌بندی صریح ساختار وابستگی برای تحلیل بیزی داده‌های بقای فضایی، $i = 1, \dots, n$ ، بردار p بعدی از متغیرهای تبیینی، c_i زمان سانسور برای فرد i ام، $\mathbf{X}_i = (X_i^1, \dots, X_i^p)'$ ، $y_i = \min(c_i, t_i)$ ، δ_i تابع نشانگر داده بقای سانسور شده در زمان سانسور c_i است که به ازای $t_i \leq c_i$ ، $\delta_i = 1$ و به ازای $t_i > c_i$ ، $\delta_i = 0$ است و قرار داده می‌شود $\mathbf{y} = (y_1, \dots, y_n)$ و $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)$. اگر θ بردار پارامترهای مورد نظر، $A = \{A_1, \dots, A_L\}$ کل ناحیه مورد مطالعه و n_ℓ تعداد افراد در ناحیه A_ℓ باشند، آنگاه تابع درستنمایی به صورت

$$\begin{aligned} L_A(\theta|n, \mathbf{y}, \boldsymbol{\delta}, \mathbf{X}_i) &= \prod_{\ell=1}^L \prod_{i=1}^{n_\ell} [F_s(A_\ell, y_{i\ell}, \mathbf{X}_i)]^{\delta_i} [S_{s,t}(A_\ell, y_{i\ell}, \mathbf{X}_i)]^{1-\delta_i} \\ &= \prod_{\ell=1}^L \prod_{i=1}^{n_\ell} [f_\nu(y_{i\ell}, \mathbf{X}_i) \int_{A_\ell} f_\lambda(s) ds]^{\delta_i} \left[\int_{A_\ell} f_\lambda(s) ds \left(1 - \int_0^{y_{i\ell}} f_\nu(t, \mathbf{X}_i) dt \right) \right]^{1-\delta_i}, \end{aligned}$$

است، که در آن $y_{i\ell}$ داده بقای سانسور شده فرد i ام در ناحیه A_ℓ است و $F_s(A_\ell, y_{i\ell}, \mathbf{X}_i)$ به گونه‌ای بیان شده است که وابستگی متغیرهای تبیینی را نشان دهد. با فرض آنکه هر مؤلفه زمانی t_i دارای توزیع وایبول با پارامترهای λ_i و μ باشد و با فرض استقلال فضا و زمان، تابع درستنمایی ناحیه‌ای به صورت

$$\begin{aligned} L_A(\theta|n, \mathbf{y}, \boldsymbol{\delta}) &= \prod_{\ell=1}^L \prod_{i=1}^{n_\ell} [\mu \lambda_i (y_{i\ell})^{\mu-1} e^{-(y_{i\ell})^\mu \lambda_i} \int_{A_\ell} f_\lambda(s) ds]^{\delta_i} \\ &\quad \times \left[\int_{A_\ell} f_\lambda(s) ds \left(1 - \int_0^{y_{i\ell}} \mu \lambda_i t^{\mu-1} e^{-t^\mu \lambda_i} dt \right) \right]^{1-\delta_i}. \end{aligned} \quad (1.2)$$

حاصل می‌شود. فرض کنید متغیرهای تبیینی به پارامترهای λ_i از طریق تابع پیوند $\log(\cdot)$ به صورت $\log(\lambda_i) = \beta_0 + \mathbf{X}_i \boldsymbol{\beta}$ ، مرتبط باشند، که در آن β_0 عرض از مبدأ و $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ بردار پارامترهای رگرسیونی هستند. در این صورت با فرض استقلال مؤلفه‌های فضا و زمان این مدل صریح فضایی، مدل غیرساختاری بقای فضایی-زمانی (STSU) نامیده می‌شود.

۳ فرایندی فضایی برای محاسبه تابع درستنمایی

در تابع درستنمایی (۱.۲)، محاسبه عبارت $F_s(A_\ell) = \int_{A_\ell} f_\lambda(s) ds$ ، که وابسته به نواحی فضایی A_ℓ است نیازمند ایجاد یک فرایند فضایی برای مدل‌بندی ساختار وابستگی نواحی است. انتخاب‌های متنوعی برای مشخص کردن مدل فضایی وجود دارد، یک رویکرد در نظر گرفتن یک میدان تصادفی مانای مرتبه دوم است که وابستگی فضایی آن توسط یک تابع کوواریانس فضایی تعیین می‌شود. مدل‌بندی فضایی با استفاده از تابع کوواریانس از نظر محاسباتی دشوار و گاهی محدود کننده است. جایگزینی مناسب برای ساده‌سازی محاسباتی و مشخص کردن تابع کوواریانس، می‌تواند یک مدل فرایند پیچش باشد [۶]، زیرا هر فرایند گاوسی ثابت می‌تواند به عنوان پیچشی از یک فرایند نوفه سفید $V(s)$ با یک تابع هسته مشخص $k(s)$ بیان شود. در این مقاله برای ساختن مدل فضایی، از یک فرایند پیچش با یک تابع هسته گاوسی استفاده می‌شود که این امکان را فراهم می‌سازد که در ناحیه مورد مطالعه مجاز به تغییر باشد. [۶] نشان داد که می‌توان مدل فرایند فضایی را با مشخص کردن فرایند نوفه سفید $V(s)$ و هسته هموارسازی $k(s)$ مشخص کرد. با انتخاب هسته گاوسی $k(s) \propto \exp(-\frac{1}{\nu} \|s\|^2)$ ، ماتریس کوواریانس ایجاد می‌شود که تابعی از توان دوم فاصله بین دو مکان فضایی است و به تدریج با افزایش فاصله کوچک می‌شود. برای هسته گاوسی، تابع کوواریانس فضایی به صورت $C(h) = \text{Cov}(Z(s), Z(s')) \propto \exp(-\frac{1}{\nu} \|h/\sqrt{\nu}\|^2)$ در نظر گرفته می‌شود، که در آن $h = s - s'$. همچنین [۶] نشان داد فرایند گاوسی $Z(s)$ روی ناحیه فضایی \mathbb{S} می‌تواند به صورت $Z(s) = \int_{\mathbb{S}} k(u-s)V(u)du$ ساخته شود و به دلیل آنکه این انتگرال به صورت تحلیلی قابل حل نیست با شبکه کردن ناحیه \mathbb{S} ، می‌توان $Z(s)$ را به صورت

$$Z(s) = \sum_{j=1}^{n_g} \exp(V_j) k(w_j - s), \quad V_j \sim N(0, \sigma^2), \quad j = 1, \dots, n_g$$

تقریب زد، که در آن w_j ، نقاط مشبکه و n_g تعداد کل نقاط مشبکه روی ناحیه مورد مطالعه است. بنابراین

$$z_{A_\ell} = \int_{A_\ell} Z(s) ds \approx \sum_{s \in A_\ell} \sum_{j=1}^{n_g} \exp(V_j) k(w_j - s).$$

برای هر ناحیه A_ℓ ، $\ell = 1, \dots, L$ ، هسته به ازای فاصله بین مرکز ناحیه و هر نقطه مشبکه در تمام ناحیه محاسبه می‌شود. این مقادیر در ماتریس هسته $K = (K_{ij})$ ، $i = 1, \dots, L$ و $j = 1, \dots, n_g$ با ابعاد $L \times n_g$ ذخیره می‌شود.

برای هر $\ell = 1, \dots, L$ انتگرال تابع چگالی احتمال فضایی روی هر ناحیه A_ℓ برابر نسبت بین z_{A_ℓ} و z_{A_T} می‌شود. بنابراین مقدار احتمال خطر هر محل و اندازه‌های مربوط به آن به تعداد نواحی مورد مطالعه بستگی دارد. در این صورت تابع درستنمایی ناحیه‌ای (۱.۲) به صورت

$$L_A(\mu, \lambda | n, \mathbf{y}, \delta) = \prod_{\ell=1}^L \prod_{i=1}^{n_\ell} \frac{z_{A_\ell}}{z_{A_T}} [\mu \lambda_i (y_{i\ell})^{\mu-1} e^{-(y_{i\ell})^\mu \lambda_i}]^{\delta_i} \left[\left(1 - \int_0^{y_{i\ell}} \mu \lambda_i t^{\mu-1} e^{-t^\mu \lambda_i} dt \right) \right]^{1-\delta_i}.$$

بازنویسی می‌شود.

۴ تحلیل بیزی مدل STSU

برای هر یک از ضرایب رگرسیونی β_0, \dots, β_p در مدل STSU توزیع پیشین نرمال $N(\circ, 1^\circ)$ در نظر گرفته شده است. همچنین برای اثرات نوفه سفید V_k ، $k = 1, \dots, n_g$ ، که n_g تعداد کل نقاط مشبکه در ناحیه مورد مطالعه است، توزیع پیشین نرمال $N(\circ, \Sigma_x)$ منظور می‌شود، که در آن Σ_x یک ماتریس قطری است. برای لگاریتم پارامتر شکل توزیع وایبول، یعنی $\log(\mu)$ ، توزیع نرمال با میانگین صفر و واریانس کوچک در نظر گرفته می‌شود. تحلیل‌های حساسیت نشان داده است که برآورد پارامترها تحت تأثیر انتخاب توزیع‌های پیشین نیستند [۷]. برای محاسبات برآورد بیزی پارامترها با استفاده از الگوریتم متروپلیس هستینگز از توزیع پسین پارامترها تصادفی تولید نموده و براساس تابع زیان درجه دوم میانگین نمونه‌ها به عنوان برآورد بیزی پارامترها محاسبه می‌شوند. همگرایی زنجیرها با استفاده از نمودار اثر و روش تشخیص همگرایی گلن-روبین [۴] مورد بررسی قرار می‌گیرد. بردار پارامترهای مورد نظر θ ، شامل ضرایب و انحرافات استاندارد آنها، لگاریتم پارامتر شکل توزیع وایبول $\log(\mu)$ و فرایند نوفه سفید V_k ، $k = 1, \dots, n_g$ است. برای برآورد پارامترهای مدل STSU ابتدا مقادیر اولیه θ برای θ تعیین می‌شود، سپس برای $t = 1, \dots, T$ سه گام زیر تکرار می‌شود:

گام ۱: مقادیر جدید θ' از توزیع‌های نامزد متقارن $h(\theta'|\theta)$ تولید شوند.

گام ۲: $\log(\alpha) = \min(\circ, \log(L_A(y|\theta')\pi(\theta')/L_A(y|\theta)\pi(\theta)))$ محاسبه شود، که در آن L_A تابع درستنمایی ناحیه‌ای و $\pi(\circ)$ توزیع پیشین θ است.

گام ۳: با احتمال α ، مقدار پارامتر به صورت $\theta^t = \theta'$ به‌روزرسانی شود.

۵ مطالعه شبیه‌سازی

در این بخش نحوه کاربست مدل پیشنهادی STSU و مدل بقای فضایی مبتنی بر تابع خطر فضایی (۱.۱) بر اساس داده‌های شبیه‌سازی شده نمایش داده می‌شود و کارایی دو مدل برای تحلیل داده‌های بقای فضایی مورد ارزیابی و مقایسه قرار می‌گیرد. برای این منظور ابتدا $n = 50$ موقعیت s_1, \dots, s_n به‌طور تصادفی در ناحیه $A = [0, 1] \times [0, 1]$ به عنوان ناحیه مورد مطالعه، ایجاد شده است. سپس برای شبیه‌سازی داده‌ها از مدل بقای فضایی، برای تابع خطر پایه $h_\circ(t(s))$ در موقعیت فضایی s ، توزیع Weibull (μ', λ') و برای ساختار همبستگی فضایی میدان تصادفی $Z(s)$ تابع

کوواریانس فضایی گاوسی به صورت $C(h; \sigma, \phi) = \sigma^2 \exp(-\|h\|^2/\phi^2)$ ، در نظر گرفته می‌شود. برای پارامترها مقادیر $(\mu' = 1, \lambda' = 2)$ و $(\sigma = 1, \phi = 0.1)$ منظور شده است. سپس تعداد $p = 5$ متغیر تبیینی در مدل (۱.۱) به صورت (X_1, \dots, X_5) در نظر گرفته شده است. دو متغیر X_1 و X_2 دو حالتی و از توزیع برنولی با احتمال‌های موفقیت 0.1 و 0.5 تولید شده‌اند. متغیر X_3 از توزیع نرمال استاندارد و متغیرهای همبسته فضایی X_4 و X_5 هر یک با استفاده از تابع کوواریانس فضایی نمایی $C(h, \sigma, \phi)$ با میانگین صفر و پارامترهای $(\sigma = 4, \phi = 0.1)$ و $(\sigma = 10, \phi = 0.1)$ شبیه‌سازی شده‌اند. زمان‌های سانسور (c_1, \dots, c_n) نیز بر اساس توزیع $U(0, 2)$ شبیه‌سازی شده‌اند. توزیع‌های پیشین ضرایب رگرسیونی نیز به صورت $\beta_j \sim N(0, 10)$ ، $j = 0, \dots, p$ ، در نظر گرفته شده است. همچنین با در نظر گرفتن یک میدان تصادفی گاوسی $Z(s)$ با میانگین صفر و تابع کوواریانس گاوسی $C(h) = \sigma^2 \exp(-\|h\|^2/\phi^2)$ ، برای تولید داده بقای فضایی در موقعیت $s \in A$ از رابطه

$$T(s) = - \left[\frac{\log(U)}{\gamma \exp(X\beta + Z(s))} \right]^{\frac{1}{\nu}}$$

استفاده شده است [۱]، که در آن U از توزیع $U(0, 1)$ تولید می‌شود.

برای شبیه‌سازی، داده‌ها از مدل STSU همان موقعیت‌های تولید شده در ناحیه A را در نظر گرفته و n مقدار از توزیع $Weibull(\mu, \lambda_i)$ برای زمان وقوع $t(s_i)$ داده تولید شده است. متغیرهای تبیینی به پارامتر $\log(\lambda_i)$ ، به صورت $\log(\lambda_i) = X\beta$ مرتبط شده‌اند، که در آن بردار ضرایب رگرسیونی و $X = (1, X_1, \dots, X_p)$ ماتریس $(p+1) \times n$ بعدی متناظر با p متغیر تبیینی $1 \times n$ بعدی و عبارت عرض از مبدأ مدل است. تعداد $p = 5$ متغیر تبیینی همانند مدل قبل منظور شده‌اند. توزیع‌های پیشین برای ضرایب رگرسیون همانند مدل قبل، برای فرایند نوفه سفید $(V_k \sim N(0, 25))$ ، $k = 1, \dots, n_g$ ، و برای لگاریتم پارامتر شکل توزیع وایبول $\log(\mu) \sim N(0, 0.1)$ در نظر گرفته شده است.

برای ارزیابی و مقایسه دقت حاصل از دو رویکرد استفاده از مدل STSU و تابع خطر فضایی (۱.۱)، پارامترهای این دو مدل برآورد و دقت آنها مقایسه می‌شود. برای برآورد بیزی پارامترهای دو مدل، الگوریتم ۱ با تکرار ۳۰۰۰۰ اجرا شد و ۱۰۰۰۰ تکرار اول به عنوان دوره داغیدن کنار گذاشته شده‌اند. تعداد گام‌ها نیز ۲۰ در نظر گرفته شده است. بنابراین برای هر پارامتر نمونه‌ای تصادفی به حجم ۱۰۰۰ تولید شده و میانگین آنها به عنوان برآورد پارامترها محاسبه شده است. مقادیر واقعی پارامترهای رگرسیونی، بازه‌های باورمندی ۹۵٪ به همراه میانگین توان دوم خطای MSE برآوردها در جدول ۱ ارائه شده است. همان طور که ملاحظه می‌شود ضرایب رگرسیونی هر دو مدل که برای محاسبه ریسک هر عامل مورد استفاده قرار می‌گیرند با دقت مناسبی برآورد شده‌اند، اما MSE های برآوردهای پارامترهای رگرسیونی برای مدل STSU همواره کوچکتر از MSE های مدل با تابع خطر فضایی (۱.۱) است. همچنین نتایج حاصل از محاسبه معیار اطلاع انحراف (DIC) و تعداد پارامترهای موثر (p_D) هر دو مدل که پیچیدگی آنها از طریق واریانس پسین انحرافات مشخص می‌کند، نشان دهنده برازش بهتر مدل STSU نسبت به مدل بقای فضایی است. اگرچه می‌توان یکی از دلایل این اختلاف در مقدار DIC را در تعداد بیشتر پارامترهای مدل بقای فضایی دید. برآورد پارامترهای تابع کوواریانس و توزیع خطر پایه برای مدل بقای فضایی به ترتیب به صورت $\hat{\mu}' = 1.431$ ، $\hat{\lambda}' = 1632$ ، $\hat{\sigma} = 1505$ و $\hat{\phi} = 0.104$ محاسبه شده است. برای بررسی همگرایی زنجیرهای تولید شده به توزیع مانای خود از نمودار اثر و روش تشخیص همگرایی گلن-روبین استفاده می‌شود و بر این اساس، یک زنجیر به توزیع مانا همگرا است اگر عامل کاهش مقیاس چندمتغیره (MPSRF) آن نزدیک یک (کمتر از ۱/۲) باشد. ضریب یک به این معنی است که واریانس بین زنجیرها و واریانس درون زنجیرها برابر است و مقادیر بزرگتر به این معنی است که هنوز تفاوت قابل توجهی بین زنجیرها وجود دارد. همان طور که در جدول ۲ ملاحظه می‌شود، برای مدل STSU مقدار معیار همگرایی برابر ۱/۰۱ به دست آمده است، بنابراین زنجیر تولید شده همگرا است و نمونه‌های تولید شده از توزیع پسین توأم پی‌روی می‌کنند. همچنین مقادیر نزدیک به یک معیار همگرایی گلن-روبین برای هر یک از پارامترها در جدول ۲ نشان‌دهنده همگرایی زنجیر تولید شده برای هر پارامتر است، به علاوه نمودار اثر در شکل ۱ بیانگر آن است که زنجیر برای هر پارامتر از آمیختگی خوبی برخوردار است، یعنی نواحی چگال پسین به خوبی مورد کاوش قرار گرفته‌اند.

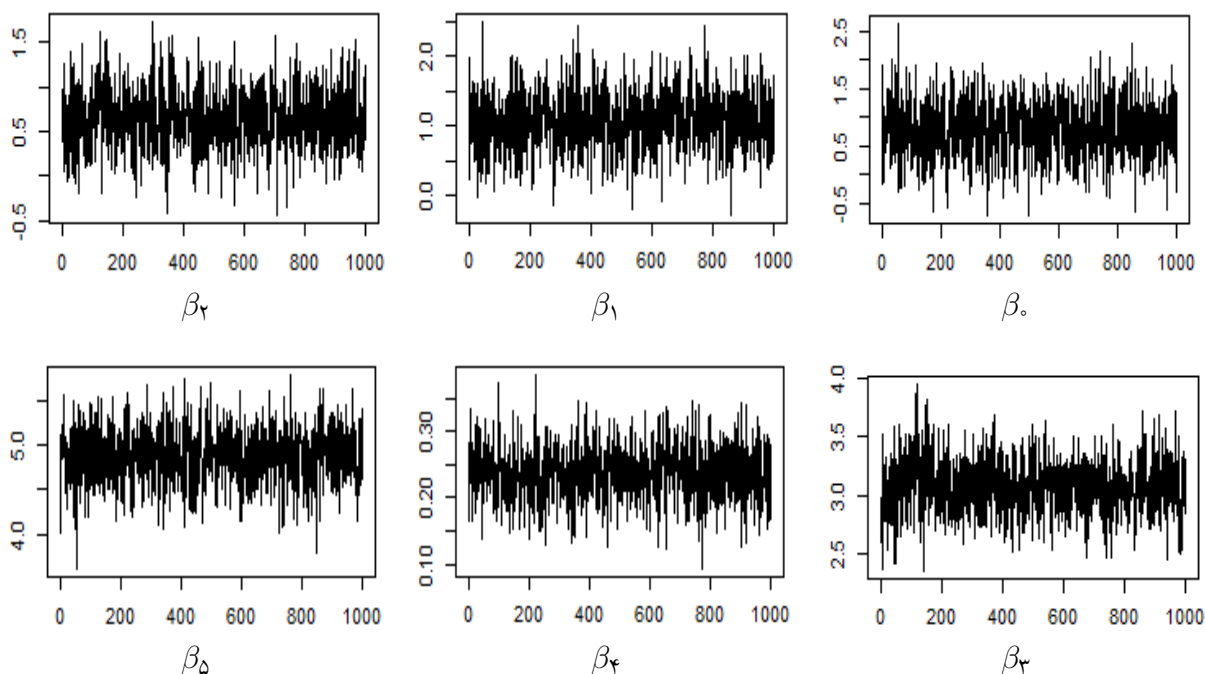
نمودارهای خودهمبستگی ابزاری متداول برای بررسی تصادفی بودن مجموعه داده‌ها هستند. به عبارتی روشی دیگر برای ارزیابی همگرایی، ارزیابی خودهمبستگی بین زنجیرهای مارکوف است. از آنجا که یک مقدار نمونه در زمان t وابسته به مقدار قبلی آن در زمان $t-1$ است، نمونه‌ها همواره همبسته خواهند بود. ارزیابی همبستگی در نمونه‌های موجود در

جدول ۱: برآورد، بازه باورمندی و MSE ضرایب رگرسیونی مدل‌های خطر فضایی و STSU

مدل STSU			مدل خطر فضایی			مقدار واقعی	پارامتر
بازه باورمندی ۹۵٪	MSE	برآورد	بازه باورمندی ۹۵٪	MSE	برآورد		
(-۰/۱۶, ۱/۷۵۲)	۰/۰۴۲	۰/۷۹۶	(-۱۲/۱۹۱, ۱۵/۳۲۷)	۰/۳۵۲	۱/۵۹۳	۱	β_0
(۰/۳۰۱, ۱/۸۹۰)	۰/۰۱۱	۱/۰۹۳	(-۰/۴۴۰, ۲/۳۳۴)	۰/۰۷۷	۰/۹۲۲	۱/۲	β_1
(۰/۰۲۸, ۱/۲۸۸)	۰/۰۱۶	۰/۶۲۸	(-۰/۲۹۱, ۲/۱۸۳)	۰/۲۰۲	۰/۹۵۰	۰/۵	β_2
(۲/۵۹۱, ۳/۵۴۳)	۰/۰۰۴	۳/۰۶۶	(۳/۸۴۵, ۶/۷۱۸)	۵/۴۵۸	۵/۳۳۶	۳	β_3
(۰/۱۵۴, ۰/۳۱۹)	۰/۰۰۱	۰/۲۳۵	(-۰/۲۹۵, ۰/۸۴۰)	۰/۱۳۴	۰/۵۶۶	۰/۲	β_4
(۴/۲۵۶, ۵/۴۸۲)	۰/۰۱۷	۴/۸۷۱	(۷/۰۹۷, ۱۱/۲۵۳)	۹/۳۸۳	۹/۴۰۳	۵	β_5
۲۳۵/۶۵۶			۳۲۸/۷۷۲			DIC	
۲۷/۹۶۳			۳۶/۶۷۲			\hat{p}_D	

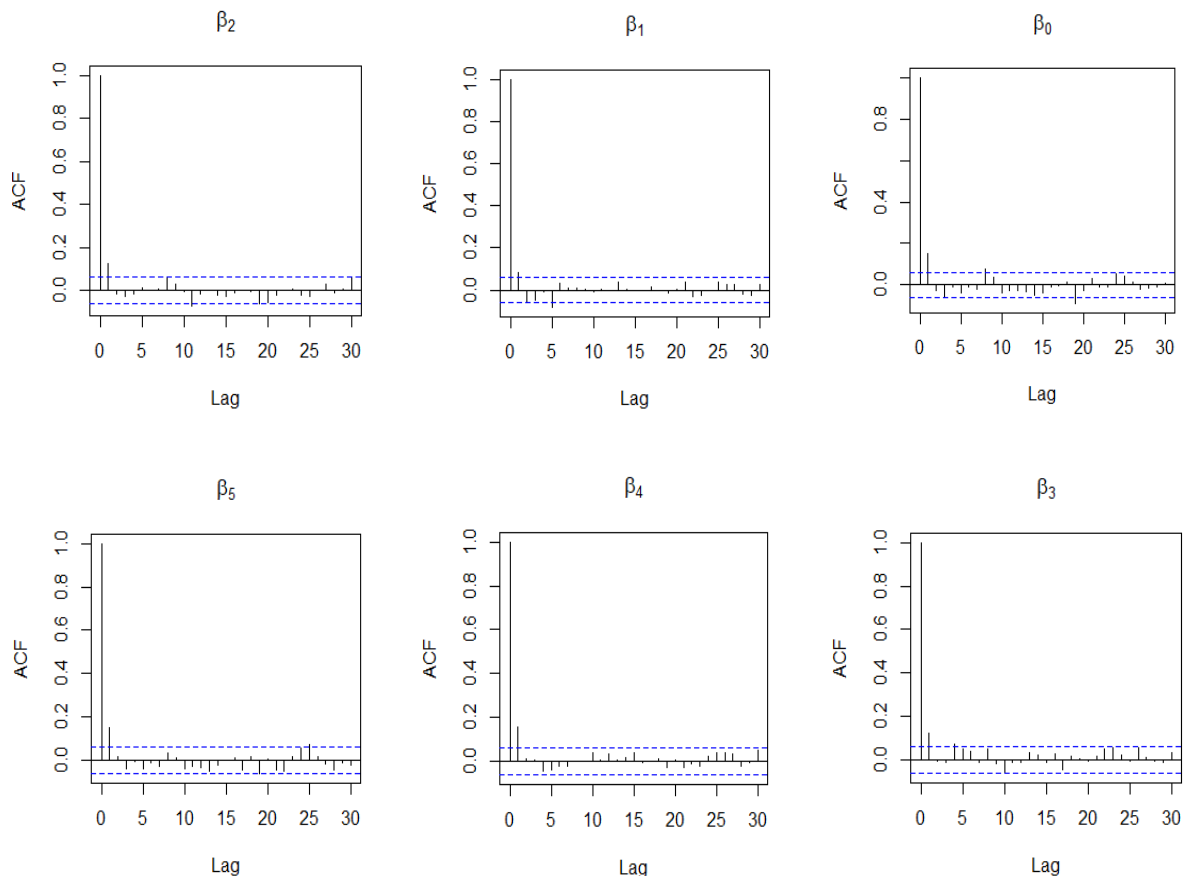
جدول ۲: نتایج آزمون همگرایی گلن-روبین برای مدل STSU

آماره	پارامتر
۱/۰۰۰	β_0
۱/۰۰۱	β_1
۱/۰۰۱	β_2
۱/۰۰۰	β_3
۰/۹۹۹	β_4
۱/۰۰۰	β_5
۱/۰۰۱	MPSRF



شکل ۱: نمودار اثر پارامترهای مدل STSU

زنجیرها با نمودار تابع خودهمبستگی در مقابل تاخیر صورت می‌پذیرد، که در آن انتظار داریم با افزایش تاخیر، همبستگی به سرعت کاهش یابد. همان‌طور که در نمودارهای خودهمبستگی پارامترها در شکل ۲ ملاحظه می‌شود، وابستگی نمونه‌های تولید شده به سرعت به سمت صفر میل می‌کند. به عبارت دیگر، مقادیر زنجیر برای پارامترهای β_1, \dots, β_5 با فاصله ۱۰ ناهمبسته می‌شوند و برای پارامتر β_0 با فاصله ۲۰ ناهمبسته شده‌اند.



شکل ۲: نمودارهای خودهمبستگی پارامترهای مدل STSU تحت زنجیر اول

محاسبات شبیه‌سازی با لپ تاپ با مشخصات 16GB Ram, Intel Core i9 11900H انجام شده است. زمان اجرا برای حجم نمونه‌های ۵۰، ۷۰، ۱۰۰ و ۱۵۰ و تعداد تکرار ۱۰۰۰، به ترتیب ۲، ۳، ۶ و ۱۳ دقیقه به طول انجامیده است. لازم به ذکر است که دقت نتایج با افزایش حجم نمونه به میزان جزیی افزایش یافته است که برای اجتناب از طولانی شدن متن نتایج گزارش نشده‌اند.

۶ تحلیل داده‌های سرطان پروستات

سرطان پروستات دومین سرطان شایع بعد از سرطان ریه در مردان است. که میانگین سن تشخیص بیماری ۶۵ سال است. میزان شیوع و نفوذ سرطان پروستات در مناطق مختلف دنیا با بیشترین میزان در آمریکای شمالی و کمترین میزان در جنوب آسیا، متفاوت است. به طور مثال تقریباً از هر ۶ مرد آمریکایی یک نفر به این بیماری مبتلا می‌شود. اگرچه دلایل ابتلا به این بیماری ناشناخته است، اما مهمترین عواملی که خطر ابتلا را بالا می‌برد، شامل سن، نژاد، سابقه خانوادگی یا ژنتیک و موقعیت جغرافیایی افراد است. در این مقاله از داده‌های سرطان نظارت، همه‌گیرشناسی و نتایج نهایی (SEER) (<https://seer.cancer.gov>) در سال ۲۰۱۵ استفاده شده است. داده‌های مورد مطالعه، شامل موارد سرطان پروستات است. مرحله تشخیص این بیماری، معمولاً به سه دسته عود موضعی، عود منطقه‌ای و عود دوردست تقسیم می‌شود. موضعی به سرطانی گفته می‌شود که در غده پروستات محدود شده است. عود منطقه‌ای به سرطانی گفته می‌شود که شامل غدد لنفاوی منطقه‌ای یا نفوذ به غده پروستات با یا بدون گسترش مستقیم از محدوده غده پروستات به اندام‌ها یا بافت‌های اطراف می‌شود. عود دوردست به این معنی است که سرطان به قسمت‌هایی از بدن که از تومور اولیه دور است، گسترش یافته است. برای سرطان پروستات، سیستم مرحله‌بندی SEER موارد موضعی و منطقه‌ای را در یک گروه قرار می‌دهد. در این مطالعه ابتلا به سرطان پروستات ۵۰۰ نفر در سال ۲۰۱۵ با تاریخ کامل انتخاب شده است، یعنی بیماران با زمان بقا صفر از مطالعه حذف شده‌اند. برای زمان بقا، از مرگ و میر ناشی از هر علتی استفاده شده است زیرا مرگ و

میر ناشی از سرطان پروستات کم است. به این دلیل که تومورهای پروستات اغلب کند رشد می‌کنند و مردان مبتلا به این بیماری با احتمال زیاد به دلایلی غیر از سرطان پروستات فوت می‌کنند، اما به اشتباه علت اصلی مرگ را به این بیماری نسبت می‌دهند. مدل مورد نظر برای متغیرهایی که در دسترس بودند و به صورت استقرایی تصور می‌شد که با نتیجه حیاتی مرتبط اند، تنظیم شده است. متغیرهای تبیینی برای هر فرد بیمار، شامل: وضعیت تأهل در هنگام تشخیص بیماری، که به صورت متأهل در مقابل مجرد، جدا شده، مطلقه و بیوه است، سن بیمار بر حسب سال در هنگام تشخیص بیماری، نژاد، این متغیر شامل افراد سیاه پوست، سفید پوست و سایر نژادها است که افراد سیاه پوست در یک گروه و سفیدپوستان و سایر نژادها در گروه دیگر قرار گرفته‌اند، مرحله تشخیص که شامل موضعی یا منطقه‌ای در مقابل دوردست است، درجه تمایز یافتگی تومور که شامل درجه ۱ (به خوبی متمایز شده) و ۲ (نسبتاً متمایز شده) در مقابل درجه ۳ (تمایز ضعیف) و ۴ (بدون تمایز) است، متغیر ناحیه‌ای شامل میانه درآمد خانواده در سال ۲۰۱۵ است (که اطلاعات مربوط به درآمد از طریق <https://www.census.gov> اخذ شده است).

جدول ۳: مشخصات جمعیتی داده‌های سرطان پروستات

متغیر تبیینی	وضعیت	میان	دامنه میان چارکی	دامنه
زمان بقا	ماه	۲۰	۵	۱ - ۲۳
سن	سال	۶۶	۱۱	۴۰ - ۹۴
درآمد	میان	۴۵/۴۳	۱۱/۱۸	۲۶/۳۲ - ۷۴/۸۵
		تعداد	درصد	
تأهل	مجرد	۷۱	۱۴/۲	
	متأهل	۳۵۱	۷۰/۲	
	مطلقه یا بیوه	۷۸	۱۵/۶	
نژاد	سفید	۳۵۶	۷۱/۲	
	سیاه	۱۴۲	۲۸/۴	
	سایر	۲	۰/۴	
درجه تمایز یافتگی	۱	۱۲۸	۲۵/۶	
	۲	۲۴۶	۴۹/۲	
	۳	۱۲۶	۲۵/۲	
	۴	۰	۰	
مرحله تشخیص	دوردست	۳۱	۶/۲	
	موضعی	۴۶۹	۹۳/۸	
وضعیت بقا	زنده	۴۷۹	۹۵/۸	
	مرده	۲۱	۴/۲	

برای برازش مدل مورد نظر تعداد ثابت ۵۰۰ موقعیت در ناحیه $A = [0, 1] \times [0, 1]$ در نظر گرفته شده است. برآورد پارامترها یعنی میانگین پسینی و بازه باورمندی ۹۵٪ ضرایب در جدول ۴ گزارش شده است. با توجه به اینکه متغیرهای تبیینی از طریق لگاریتم پارامتر مقیاس توزیع وایبول مرتبط شده‌اند، بنابراین ضرایب بالاتر برای متغیرهای تبیینی نشان‌دهنده میزان بقای کمتر در ارتباط با افزایش مقادیر متغیرهای تبیینی است. پس از کنترل سایر متغیرهای موجود در مدل، و انتخاب حالت‌های غیرمتأهل، نژاد سفید، عود موضعی/منطقه‌ای و درجه تمایز یافتگی ۱ و ۲ به عنوان مرجع، خطر نسبی برای هر متغیر تبیینی محاسبه شده است. خطر نسبی مرگ در جمعیت سیاه پوستان نسبت به سفید پوستان و سایر نژادها ۲/۴۳۸ برابر و نرخ خطر در مرحله عود دوردست ۷/۵۸۴ برابر نرخ خطر در مرحله عود موضعی یا منطقه‌ای است و همچنین با توجه به نرخ خطر سن می‌توان گفت، به ازای افزایش یکسال، رخداد مرگ ۱/۵۶۴ برابر می‌شود. همانطور که ملاحظه می‌شود بیشترین اثر را متغیر مرحله تشخیص با ۷/۵۸۴ خطر نسبی برای افراد داشته است. وضعیت تأهل، درجه تمایز یافتگی و میانه درآمد خانواده ارتباطی با نتیجه نشان نمی‌دهند.

برای بررسی همگرایی زنجیر، از آزمون همگرایی گلن-روبین استفاده کرده و نتایج حاصل در جدول ۵ گزارش شده

جدول ۴: برآورد پارامترها و بازه باورمندی ۹۵٪ در مدل STSU

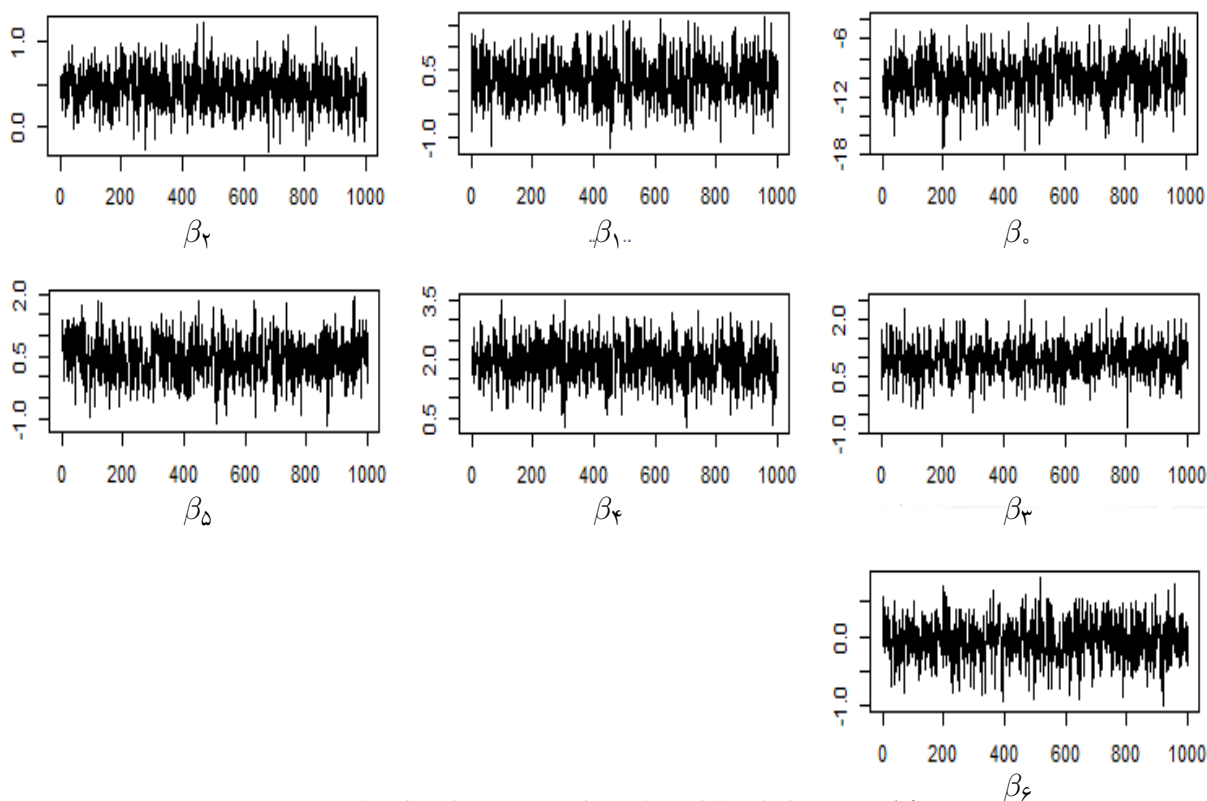
متغیر تبیینی	وضعیت	برآورد ضریب	خطر نسبی	بازه باورمندی ۹۵٪
عرض از مبدأ	-	-۹/۶۵۴	-	(-۱۳/۴۶۱, -۵/۶۶۰)
تأهل	غیر متأهل متأهل	Ref. ۰/۴۷۹	۱/۶۱۴	(-۰/۴۰۳, ۱/۳۷۵)
نژاد	سفید سیاه	Ref. ۰/۸۹۱	۲/۴۳۸	(۰/۱۳۸, ۱/۶۶۷)
مرحله تشخیص	موضعی دوردست	Ref. ۲/۰۲۶	۷/۵۸۴	(۰/۹۸۸, ۲/۹۰۵)
درجه تمایز یافتگی	۱ و ۲ ۳ و ۴	Ref. ۰/۳۳۴	۱/۳۹۷	(-۰/۴۹۸, ۱/۱۹۷)
سن درآمد	- میانه	۰/۴۴۷ -۰/۱۶۸	۱/۵۶۴ ۰/۸۴۵	(۰/۰۳۵, ۰/۸۸۵) (-۰/۷۶۰, ۰/۳۳۹)

است. با توجه به مقادیر PSRF برای هر پارامتر، که مقداری نزدیک به یک دارد، می‌توان گفت که زنجیر تولید شده برای هر پارامتر، همگرا است. همچنین همگرایی زنجیر به توزیع مانا برای هر یک از پارامترها، با رسم نمودار اثر بررسی شده است. با توجه به شکل ۳ می‌توان همگرایی هر پارامتر به مقدار واقعی آن را مشاهده کرد. همچنین نمودارهای خودهمبستگی، شکل ۴، نشان می‌دهند که زنجیر برای هر پارامتر ناهمبسته است.

جدول ۵: نتایج آزمون همگرایی گلمن-روبین برای داده‌های سرطان پروستات

پارامتر	آماره
β_0	۱/۰۰۰
β_1	۱/۰۰۰
β_2	۱/۰۰۱
β_3	۱/۰۰۰
β_4	۱/۰۰۳
β_5	۱/۰۰۰
β_6	۰/۹۹۹
$MPSRF$	۱/۰۰۰

جدول ۳ مشخصات جمعیتی متغیرهای در نظر گرفته شده را نشان می‌دهد. محدوده سن تشخیص ۹۴ - ۴۰ سال با میانه تشخیص ۶۶ سال است. میانه زمان مرگ ۲۰ ماه با دامنه میان چارکی ۵ است که محدوده آن ۱ تا ۲۳ ماه است. در بین این تعداد از بیماران در نظر گرفته شده، اکثر افراد سفید پوست (۷۱/۲٪) در مقابل ۲۸/۴٪ سیاه پوست و تنها ۰/۴٪ از نژادهای دیگر اند. همچنین ۹۳/۸٪ از مرحله‌های تشخیص موضعی یا منطقه‌ای، در حالی که تنها ۶/۲٪ دارای عود دوردست اند. از بین تمام بیماران ۹۵/۸٪ از آنها تا پایان دوره مطالعه زنده بودند و تنها ۴/۲٪ به هر دلیلی فوت کردند. اکثر سرطان‌ها یعنی ۴۹/۲٪ در درجه تمایز یافتگی ۲ است در حالی که ۲۵/۶٪ در درجه ۱ و ۲۵/۲٪ در درجه ۳ قرار دارند و در بین این ۵۰۰ نفر هیچکدام در درجه ۴ که مرحله پیشرفته‌ی بیماری است قرار نداشتند. در مورد وضعیت تأهل، ۷۰/۲٪ از افراد متأهل، ۱۴/۲٪ مجرد و ۱۵/۶٪ جداشده، مطلقه و یا بیوه بودند.



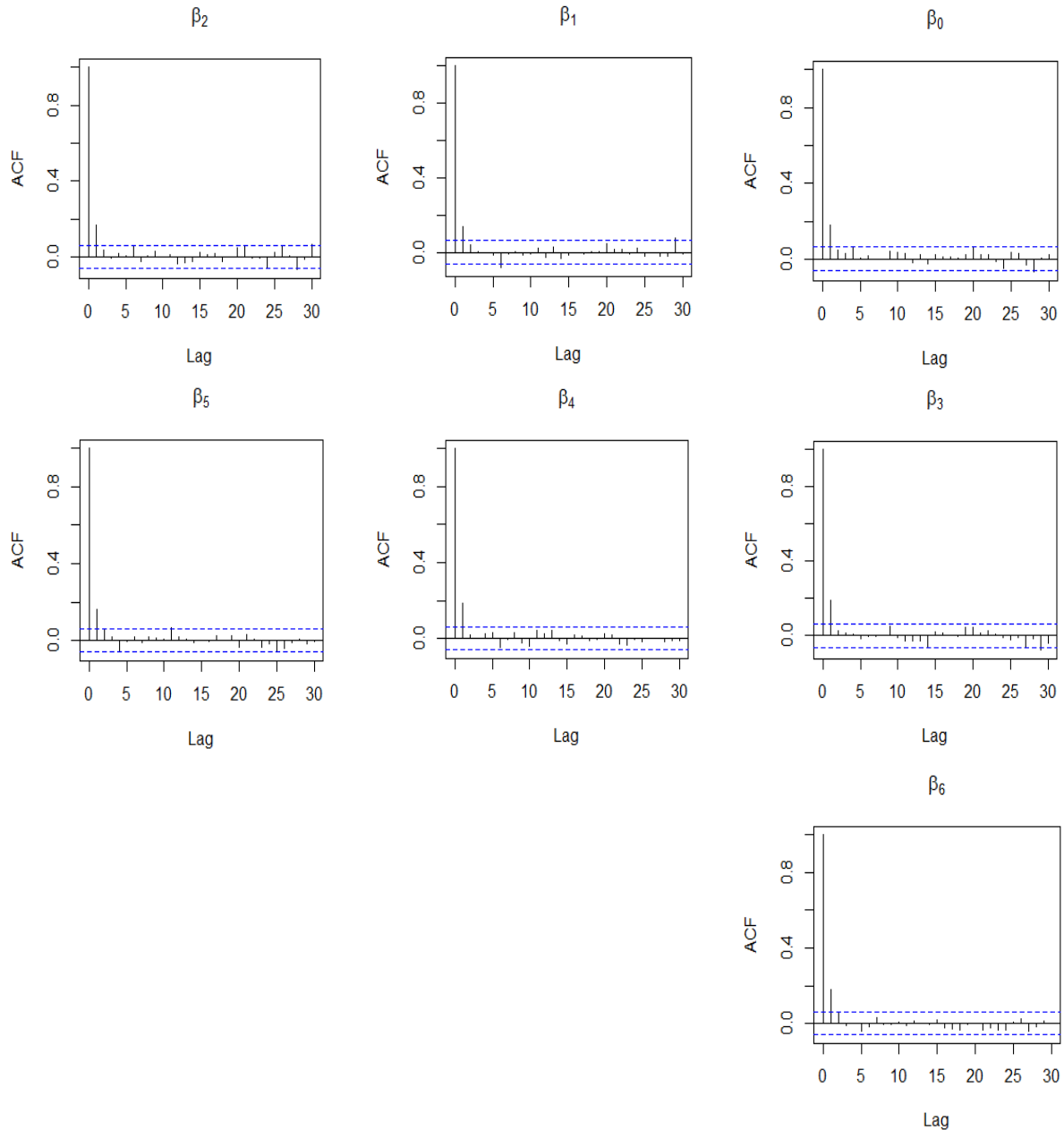
شکل ۳: نمودار اثر پارامترهای مدل STSU برای داده‌های PrCa

۷ بحث و نتیجه‌گیری

در این مقاله یک مدل با متغیر تصادفی فضایی پنهان تحت عنوان ”مدل غیرساختاری بقای فضایی-زمانی” برای تحلیل داده‌های بقای فضایی پیشنهاد شد، که در آن وابستگی فضایی به‌طور صریح در توابع بقا، چگالی و خطر منظور گردید. در مطالعه شبیه‌سازی میزان دقت پیش‌گویی‌های دو مدل STSU و مدل بقای فضایی مورد بررسی قرار گرفت. نتایج بیانگر عملکرد بهتر مدل STSU نسبت به مدل بقای فضایی است. در مدل STSU برآورد بیزی ضرایب و اثرات متغیرهای تبیینی مدل با استفاده از الگوریتم متروپلیس-هستینگز تطبیقی محاسبه شده است که مزیت استفاده از این الگوریتم سادگی محاسباتی آن نسبت به سایر روش‌ها است. زیرا برای اجرای آن نیاز به محاسبه توزیع پسین و تعیین توزیع نامزد توسط کاربر نیست. همچنین برازش این مدل به داده‌ها، برخلاف مدل بقای فضایی، مستلزم برآورد پارامترهای تابع کواریانس فضایی نیست، که هزینه برازش و دقت مدل را تحت تاثیر قرار می‌دهد. از طرفی با توجه به اینکه مدل بقای فضایی مخاطرات متناسب است، بنابراین این مدل توانایی تحلیل داده‌هایی که مخاطرات متناسب نباشد را ندارد. در حالی که مدل STSU مخاطرات متناسب نیست و این موضوع نقطه قوت این مدل است. بنابراین مدل STSU هنگامی که تسلط یا اطلاعات کافی برای تعیین تابع کواریانس داده‌های فضایی نداریم، کاربرد مناسبی دارد. از آنجا که روش ارائه شده با فرض استقلال مؤلفه‌های فضا و زمان و در یک ناحیه منظم مطرح شده است، توسعه مدل STSU برای حالت مؤلفه‌های وابسته و نواحی نامنظم می‌تواند موضوع مطالعات آتی قرار گیرد.

تقدیر و تشکر

نویسندگان از سردبیر، داوران و هیئت تحریریه محترم مجله به دلیل ارائه پیشنهادها و ارزنده که موجب ارائه بهتر مقاله شد تشکر می‌نمایند. از حمایت قطب علمی تحلیل داده‌های وابسته فضایی و فضایی-زمانی دانشگاه تربیت مدرس نیز قدردانی می‌شود.



شکل ۴: نمودارهای خودهمبستگی پارامترهای مدل STSU برای داده‌های PrCa تحت زنجیر اول

فهرست منابع

- [۱] مترجم، ک.، محمدزاده، م. و آبیاری، آ. مدل‌های شکنندگی و خطرهای متناسب برای تحلیل داده‌های بقای فضایی، مجله مدل‌سازی پیشرفته ریاضی، ۴(۲) (۱۳۹۳)، ۱۰۱-۱۲۳.
- [2] Banerjee, S. and Dey, D. K. *Semiparametric Proportional Odds Models for Spatially Correlated Survival Data*, Lifetime Data Anal, **11**(2) (2005), 175-191.
- [3] Banerjee, S., Wall, M. M. and Carlin, B. P. *Frailty Modeling for Spatially Correlated Survival data, with Application to Infant Mortality in Minnesota*, Biostatistics, **4**(1) (2003), 123-142.
- [4] Gelman, A. and Rubin, D. *Inference from Iterative Simulation Using Multiple Sequences*, Statistical Science, **7**(4) (1992), 457-472.

- [5] Cox, D. R. *Regression Models and Life-Tables*, Journal of the Royal Statistical Society, Series B, **34**(2) (1972), 187-220.
- [6] Higdon, D. *Space and Space-Time Modeling Using Process Convolutions*, in *Quantitative Methods for Current Environmental Issues*, Springer, London, (2002), 37-56.
- [7] Onicescu, G., Lawson, A. B., Zhang, J., Gebregziabher, M., Wallace, K. and Eberth, J. M. *Spatially Explicit Survival Modeling for Small Area Cancer Data*, Journal of Applied Statistics, **45**(3) (2018), 568-585.
- [8] Motarjem, K., Mohammadzadeh, M. and Abyar, A. *Geostatistical Survival Model with Gaussian Random Effect*, Statistical Papers, **61**(1) (2020), 85-107.



Explicit Modeling of Correlation Structure For Bayesian Analysis of Spatial Survival Data: Estimating the Relative Risk of Prostate Cancer Patients

Zahra Nahardani, Mohsen Mohammadzadeh[†]

Department of Statistics, Tarbiat Modares University, Tehran, Iran.

Communicated by: Siamak Noorbaloochi

Received: 2022/1/10

Accepted: 2022/10/10

Abstract: Models with spatial stochastic effects are commonly used to model the relationship between response variables and spatially dependent observations and explanatory variables. In many applications, some model's explanatory variables are dependent. Depending on the type of dependence, the statistical inference of the models with random effects and their applications are complicated; because the explanatory variables, random effects, and model error expression compete with each other in explaining the variability of the response variable. In this paper, a method for modeling and analyzing spatial survival data is proposed to solve this problem. Instead of using spatial stochastic effects in the model, the spatial dependence of observations is explicitly included in density, survival, and hazard functions. Then, in a simulation study, the effects of explanatory variables in the model are calculated and evaluated using the comparative Metropolis-Hastings algorithm. The proposed method is then used to analyze the data of patients with prostate cancer, and the Bayesian approach is used to estimate the death relative risk of patients. Finally, a discussion and conclusion will be presented.

Keywords: Spatial Survival Data, Explicit Modeling, Mixed Random Effects.

Mathematics Subject Classification (2010): 62M30, 62H11.



©2022 Shahid Chamran University of Ahvaz, Ahvaz, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0 license) (<http://creativecommons.org/licenses/by-nc/4.0/>).

[†]Corresponding author.

E-mail addresses: z_nahardani@modares.ac.ir (Z. Nahardani), mohsen_m@modares.ac.ir (M. Mohammadzadeh).