

تحلیل بیزی داده‌های بقا با همبستگی فضایی

فاطمه حسینی^۱، امید کریمی^۲، عهدیه عزیز^۳

تاریخ دریافت: ۱۳۹۶/۳/۱۵

تاریخ پذیرش: ۱۳۹۷/۶/۲۶

چکیده:

اغلب در عمل داده‌های مربوط به زمان مرگ و میر یک واحد زنده دارای همبستگی ناشی از موقعیت قرار گرفتن مشاهدات در فضای مورد مطالعه است. یکی از موضوعات مهم در تحلیل این نوع از داده‌های بقا با وابستگی فضایی، برآورد پارامترها و پیشگویی مقادیر نامعلوم در موقعیت‌های مشخص بر اساس بردار مشاهدات است. در این مقاله، برای تحلیل این نوع از داده‌های بقا، مدل رگرسیونی کاکس با تابع خطر به صورت تکه‌ای نمایی استفاده و وابستگی فضایی به صورت یک میدان تصادفی گاوسی و یک متغیر پنهان به مدل اضافه می‌شود. به دلیل عدم وجود صورت صریح برای توزیع پسین و توزیع‌های شرطی کامل و طولانی بودن محاسبات با الگوریتم‌های مونته کارلوی زنجیر مارکوفی برای تحلیل این مدل از رهیافت بیزی تقریبی استفاده می‌شود. در یک مثال کاربردی نحوه پیاده‌سازی رهیافت بیزی تقریبی ارائه می‌شود.

واژه‌های کلیدی: تحلیل بیز تقریبی، مدل پارامتری، رگرسیون کاکس، میدان تصادفی گاوسی پنهان.

۱ مقدمه

برای رگرسیون نیمه پارامتری کاکس باشند. [۲] با استفاده از اسپلاین تاوانیده^۶ در مدل کاکس به مقایسه این مدل با مدل کاکس معمولی پرداختند. مطالعات فراوان دیگری پیرامون مدل نیمه پارامتری کاکس و مدل‌های پارامتری وجود دارد که برای تحلیل داده‌های بقا پیشنهاد شده‌اند، اما به دلیل پیچیدگی توابع درست‌نمایی استنباط بیزی داده‌های بقا بیش‌تر مورد توجه قرار گرفته است. [۱۳] به‌طور مفصل به مطالعه مدل‌های بقای بیزی پرداخته‌اند. گاهی در عمل با داده‌های بقایی مواجه می‌شویم که وابستگی از نوع مکانی دارند، به این گونه از داده‌ها، داده‌های بقای فضایی^۷ می‌گویند. [۴، ۵] برای تحلیل داده‌های بقای فضایی از نوع زمین‌آمار دو رهیافت نیمه پارامتری رگرسیون کاکس

گاهی در عمل با داده‌هایی مواجه می‌شویم که بیان‌کننده زمان رخداد مثل زمان شکست یک مؤلفه فیزیکی یا زمان مرگ یک واحد زنده است. به این گونه داده‌ها، داده‌های بقا^۴ و به مجموعه روش‌های مختلف آماری که به تحلیل این نوع داده‌ها می‌پردازد، تحلیل بقا می‌گویند. در اکثر تحقیقات پزشکی که هدف بررسی توزیع بقا است، از روش‌های کلاسیک مانند رگرسیون کاکس^۵ و مدل‌های پارامتری استفاده می‌شود، [۳] به مقایسه دو رهیافت کلاسیک رگرسیون کاکس و مدل‌های پارامتری برای داده‌های بقای مربوط به سرطان معده پرداختند و نشان دادند در برخی شرایط مدل‌های پارامتری می‌توانند جایگزین مناسبی

^۱ عضو هیئت علمی گروه آمار دانشگاه سمنان، سمنان، ایران.

^۲ عضو هیئت علمی گروه آمار دانشگاه سمنان، سمنان، ایران.

^۳ دانش‌آموخته کارشناسی ارشد آمار دانشگاه سمنان، سمنان، ایران.

^۴ Survival Data

^۵ Cox Regression

^۶ Penalized Spline

^۷ Spatial Survival Data

آمیخته کاکس برای داده‌های بقای فضایی پرداخته می‌شود و مدل و رهیافت بیزی ارائه شده روی یک مجموعه داده واقعی پیاده‌سازی می‌شود.

ساختار مقاله به این صورت است که در بخش دوم به تعریف داده‌های بقا، مشاهدات سانسور شده و تابع درست‌نمایی مدل پرداخته شده است. در بخش سه مدل رگرسیونی کاکس و مدل‌های پارامتری معرفی شده‌اند. مدل کاکس با تابع خطر تکه‌ای نمایی در بخش چهار بیان شده است و مدل و روش توضیح داده شده در بخش پنج روی یک مثال پیاده‌سازی می‌شود. در نهایت بحث و نتیجه‌گیری آورده شده است.

۲ تحلیل بقا

مجموعه‌ای از روش‌های مختلف آماری برای تحلیل متغیرهای تصادفی نامنفی که مقدار آن می‌تواند زمان شکست یک مؤلفه فیزیکی یا زمان مرگ یک واحد زنده باشد، را تحلیل بقا می‌گویند و فاصله زمانی ورود به مطالعه تا وقوع حادثه را زمان بقا گویند و با T نشان می‌دهند. در تحلیل داده‌های بقا گاهی با داده‌های سانسور شده^{۱۰} مواجه هستیم. سانسور شده به مواردی اطلاق می‌شود که در طول مدت بررسی از دست رفته و ثبت وضعیت نهایی آنها امکان‌پذیر نیستند. به عبارت دیگر در تحلیل بقا، مجموعه‌ای از عناصر تحت بررسی در زمان‌های مختلف وارد مطالعه شده و تا زمان رخداد مورد نظر تا پایان مطالعه مورد بررسی قرار می‌گیرند که در این میان ممکن است پیامد مورد انتظار برای بعضی عناصر رخ ندهد که زمان بقای آنها به عنوان سانسور شده نامیده می‌شود. داده‌های سانسور شده در اثر گمشدگی در طول مطالعه، خارج شدن از مطالعه یا پایان مطالعه رخ می‌دهد. اگر زمان بقای واقعی آزمودنی کمتر از زمان ثبت شده باشد، یعنی $T < T_0$ ، که در آن T_0 زمان ورود آزمودنی به مطالعه و T زمان بقای واقعی است، آن‌گاه مشاهده سانسور شده چپ نامیده می‌شود. اگر زمان بقای واقعی آزمودنی بیشتر از زمان ثبت شده باشد، یعنی $T > C > T_0$ (C زمان ثبت شده) باشد، سانسور شده

و مدل پارامتری وایبل استفاده نمودند. [۱۱] به تعمیم مطالعات [۱۰، ۱۶] پرداختند و یک مدل جمعی کاکس فضایی معرفی نمودند. [۱۴، ۱۵] مدل‌های ترکیبی جدید بر اساس مدل کاکس برای زمان‌های بقا سانسور شده ارائه دادند. [۱۷] به مطالعه بیزی مدل‌های بقا با رهیافت بیزی تقریبی با به کار بردن تقریب لاپلاس آشیانه‌ای جمع‌بسته (INLA)^۸ پرداختند و نشان دادند رهیافت بیزی تقریبی زمان محاسبات را نسبت به رهیافت بیزی کامل و الگوریتم‌های مونته کارلوی زنجیر مارکوفی^۹ کاهش می‌دهد. [۹] به تحلیل داده‌های بقای فضایی با رهیافت بیزی و مدل‌های چندسطحی پرداختند. [۱۸] یک مدل نیمه پارامتری بیزی را برای تحلیل داده‌های بقای فضایی پیشنهاد دادند.

در این مقاله، داده‌های بقا با وابستگی فضایی با مدل کاکس با تابع خطر به صورت تکه‌ای نمایی مدل‌بندی و وابستگی فضایی به صورت یک متغیر پنهان با اثر غیرخطی وارد مدل می‌شود و این متغیر پنهان به صورت یک میدان تصادفی مارکوفی گاوسی در نظر گرفته می‌شود. برای تحلیل این مدل از رهیافت بیزی تقریبی استفاده می‌شود. برای به کار بردن رهیافت بیزی ابتدا توزیع‌های پیشین برای پارامترهای و ابرپارامترهای مدل در نظر گرفته می‌شوند و سپس با استفاده از تابع درست‌نمایی و توزیع‌های پیشین، توزیع پسین تشکیل می‌شود. به دلیل عدم وجود صورت صریح برای توزیع پسین می‌توان با تشکیل توزیع‌های شرطی کامل از الگوریتم گیبز برای تولید نمونه استفاده کرد و در صورتی که توزیع‌های شرطی کامل نمایش مشخصی از توزیع‌ها نداشتند، برای تولید نمونه الگوریتم متروپلیس هستینگز را به کار گرفت. به دلیل پیچیدگی این مدل‌ها، نداشتن صورت صریح برای توزیع پسین و برخی توزیع‌های شرطی کامل، محاسبات و اجرای برنامه‌های کامپیوتری بسیار زمان‌بر است. برای حل این مشکل و به دست آوردن برآورد پارامترهای مدل و پیشگویی از یک رهیافت بیزی تقریبی بر اساس تقریب لاپلاس استفاده می‌شود که بسیار سریع‌تر از الگوریتم‌های مونته کارلوی زنجیر مارکوفی است. در این مقاله با رهیافت بیزی تقریبی به برآورد یک مدل

^۸ Integrated Nested Laplace approximations

^۹ Markov Chain Monte Carlo

^{۱۰} Censored Data

بر این اساس، لگاریتم درست‌نمایی ترکیبی برای داده i -ام
 $(T_{lo}, T_{up}, \delta_i)$ به صورت کلی

$$l_i = \delta_i \log h(T_{up}) - \int_0^{T_{up}} h(u) du + \log \{1 - \exp(-\int_{T_{lo}}^{T_{up}} h(u) du)\}.$$

قابل بیان است. در بسیاری مطالعات پزشکی هدف بررسی توزیع بقای بیماران مبتلا به بیماری‌های خاص و عوامل مرتبط با بقای آنها است. معمولاً از دو مدل رگرسیونی برای بررسی داده‌های بقا استفاده می‌شود مدل خطرات متناسب کاکس که به عنوان یک مدل نیمه پارامتری شناخته می‌شود و مدل‌های پارامتری که از جمله مدل‌های استاندارد پارامتری می‌توان به مدل‌های وایبل، نمایی و لگ‌نرمال اشاره کرد. چون مدل رگرسیونی کاکس پیش فرض کمتری نسبت به مدل‌های پارامتری نیاز دارد، اغلب در عمل این مدل را انتخاب می‌کنند. [۳] نشان دادند که مدل‌های پارامتری در برخی شرایط کارتر از مدل رگرسیونی کاکس هستند. در مدل‌های پارامتری معمولاً از روش ماکسیمم درست‌نمایی برای برآورد پارامترهای مجهول استفاده می‌شود و خطر نسبی این مدل تفسیر مشابه نسبت بخت در رگرسیون کاکس را دارد.

۳ رگرسیون کاکس و مدل‌های پارامتری

مدلی که معمولاً برای تحلیل داده‌های بقا به کار می‌رود، رگرسیون کاکس نامیده می‌شود که به عنوان مدل خطرات متناسب نیز شناخته شده است [۸]. مدل کاکس یک مدل نیمه پارامتری برای تحلیل داده‌های سانسور شده بقا است. در این مدل همواره فرض می‌شود که اثر متغیرهای توضیحی روی لگاریتم تابع خطر، تابعی خطی از متغیرهای توضیحی است و برای تابع خطر توزیع خاصی فرض نمی‌شود و تابع خطر پایه به عنوان تابعی نامنفی و نامعلوم در نظر گرفته و از روی داده‌ها برآورد می‌شود. چون در مدل کاکس نمایش پارامتری برای تابع خطر پایه فرض نمی‌شود، مدل کاکس را یک مدل نیمه پارامتری برای تابع خطر می‌گویند. تابع خطر کاکس به صورت $h(t) = h_0(t) \exp(\mathbf{z}'\beta)$

از راست می‌گویند. اگر ترکیبی از هر دو نوع سانسور شده چپ و راست باشد، به عبارت دیگر زمان بقای واقعی دقیقاً مشاهده نشده است، اما می‌دانیم این زمان در یک فاصله $T_{lo} < T < T_{up}$ قرار دارد، سانسور شده فاصله‌ای^{۱۱} نامیده می‌شود. بنا بر این هر مشاهده به صورت (T_{lo}, T_{up}, δ) قابل بیان است، که در آن δ بیان‌کننده وضعیت واحد مورد آزمایش از نظر سانسور یا رخداد است، یعنی اگر واحد مورد آزمایش سانسور شده است، آن گاه $\delta = 0$ و اگر رخداد مورد نظر برای واحد مورد آزمایش رخ داده است، آن گاه $\delta = 1$. پس اگر مشاهده سانسور نشده باشد، یعنی $\delta = 1$ و $T_{lo} = T_{up} = T$ و اگر مشاهده از راست سانسور شده باشد، آن گاه $\delta = 0$ و $T_{lo} = T_{up} = C$ و اگر مشاهده سانسور شده فاصله‌ای باشد، آن گاه $\delta = 0$ و $T_{lo} < T_{up}$. برای مطالعه بیشتر به [۷] مراجعه شود.

فرض کنید T یک زمان بقای تصادفی با تابع چگالی $f(t)$ باشد، آن گاه تابع بقا به صورت

$$P(T > t) = S(t) = 1 - F(t) = \int_t^\infty f(u) du,$$

و تابع خطر به صورت

$$h(t) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} P(t < T < t + \delta | T > t),$$

تعریف می‌شود، بنا بر این با تابع بقا رابطه‌ای به صورت

$$h(t) = \lim_{\delta \rightarrow 0} \frac{S(t) - S(t + \delta)}{S(t)} = \frac{f(t)}{S(t)},$$

دارد. تابع درست‌نمایی به صورت $L = \prod_{i=1}^m L_i$ ، که در آن m تعداد مشاهدات است و اگر آزمودنی i -ام سانسور نشده باشد و T زمان بقای آن باشد، آن گاه

$$L_i = h(T)S(T) = h(T) \exp\{-\int_0^T h(u) du\},$$

و اگر آزمودنی i -ام سانسور شده باشد، آن گاه $h(t) = 1$ و

$$L_i = S(c) = \exp\{-\int_0^c h(u) du\}$$

و اگر آزمودنی i -ام سانسور شده فاصله‌ای باشد، آن گاه

$$L_i = S(T_{lo}) - S(T_{up}) = \exp\{-\int_0^{T_{lo}} h(u) du\} \times \{1 - \exp(-\int_{T_{lo}}^{T_{up}} h(u) du)\}. \quad (1)$$

^{۱۱} Interval Censored

است، که در آن h_0 تابع خطر پایه^{۱۲}، $\mathbf{z}' = (z_1, \dots, z_{\eta_\beta})$ بردار متغیرهای توضیحی و $\beta' = (\beta_1, \dots, \beta_{\eta_\beta})$ بردار پارامترهای رگرسیونی هستند. در این مدل فرض می‌شود متغیرهای توضیحی اثرات ثابتی را روی الگوهای شکست دارند اما تأثیرات این متغیرهای توضیحی با زمان قابل تغییر است. بر خلاف مدل نیمه پارامتری کاکس، اگر برای تابع خطر پایه یک نمایش پارامتری مثل وایبل یا نمایی در نظر گرفته شود، مدل بقا، یک مدل پارامتری است. فرض کنید داده‌ها $(T_{lo}, T_{up}, \delta_i)$ ، $(i = 1, \dots, m)$ مشاهده شده‌اند. در مدل وایبل تابع خطر به صورت $h(t, \alpha) = \alpha t^{\alpha-1} \exp(\mathbf{z}'\beta)$ ، یعنی تابع خطر پایه $h_0(t) = \alpha t^{\alpha-1}$ فرض می‌شود و تابع لگ درست‌نمایی به صورت

$$\begin{aligned} h_i(t) &= h_0(t) \exp(\mathbf{z}'_i \beta) \\ &= \exp\{\mathbf{z}'_i \beta + b_k\} \\ &= \exp(\eta_{ik}), \quad t \in (a_{k-1}, a_k). \end{aligned} \quad (2)$$

است، که در آن $b_k = \log(\lambda_k)$ در مدل (۱) فرض می‌شود که اثر متغیرهای توضیحی روی لگاریتم تابع خطر، تابعی خطی از متغیرهای توضیحی است. گاهی در عمل برخی از متغیرهای توضیحی پیوسته می‌توانند به شکل‌های غیرخطی روی میزان خطر اثر بگذارند. در مدل کاکس با تابع خطر تکه‌ای نمایی به صورت (۱) می‌توان مستقیماً نمایش غیرخطی متغیرها را وارد کرد. در این حالت در پیشگو، علاوه بر b_k و $\mathbf{z}'_i \beta$ جمله $f^r(\cdot)$ اضافه و به صورت

$$\eta_{ik} = \beta_0 + \sum_{r=1}^{\eta_r} w_{ir} f_r(u_{ir}) + \sum_{j=1}^{\eta_\beta} \beta_j z_{ji} + b_k. \quad (3)$$

بیان می‌شود، که در آن η_r تعداد متغیرهای توضیحی با اثرات غیرخطی، η_β تعداد متغیرهای توضیحی با اثرات خطی، β_j ضرایب رگرسیونی و نشان‌دهنده اثر خطی متغیرهای توضیحی و $f_r(\cdot)$ توابع ناشناخته‌ای هستند که برای در نظر گرفتن همبستگی‌های فضایی، همبستگی‌های خوشه‌ای یا همبستگی‌های زمانی و اثرات غیرخطی متغیرهای توضیحی هستند. بنا بر این اثرات تصادفی فضایی، اثرات تصادفی خوشه‌ای و غیره از طریق تابع $f_r(\cdot)$ در مدل قابل بیان هستند و w_{ir} وزن‌های معلوم و ε_i اثرات تصادفی یا خطاها هستند. برای تحلیل این مدل از رهیافت بیزی تقریبی استفاده می‌شود که در بخش بعد این رهیافت معرفی شده است.

به دست می‌آید. در عمل گاهی با مواردی مواجه می‌شویم که علاوه بر متغیرهای توضیحی با اثرات خطی، برخی متغیرهای توضیحی اثرات غیرخطی بر روی مدل دارند. در بخش بعد یک مدل مفید برای مطالعه داده‌های بقا در حضور انواع متغیر توضیحی با اثرات خطی و غیرخطی و وابستگی‌ها بیان و سپس شیوه تحلیل بیزی مدل ارائه می‌شود.

$$\begin{aligned} l &= \delta \log h(T_{up}) - \int_0^{T_{up}} h(t) dt, \\ &+ \log \{1 - \exp(-\int_{T_{lo}}^{T_{up}} h(t) dt)\}, \\ &= \delta [\log \alpha + (\alpha - 1) \log T_{up} + \mathbf{z}'\beta], \\ &- e^{\mathbf{z}'\beta} T_{up}^\alpha + \log \{1 - e^{(-e^{\mathbf{z}'\beta} (T_{up}^\alpha - T_{lo}^\alpha))}\}. \end{aligned}$$

۴ مدل کاکس با تابع خطر پایه تکه‌ای نمایی

فرض کنید محور زمان به K بازه زمانی

$$0 < a_1 < a_2 < \dots < a_k < \infty$$

افراز شده باشد که در آن a_k از هر زمان بقا مشاهده شده در مطالعه بزرگ‌تر است. بنا بر این k بازه زمانی به صورت $I_k = (a_{k-1}, a_k]$ ، $k = 1, \dots, K$ خواهیم داشت به طوری که

^{۱۲} Baseline hazard

گرافی نشان‌دار باشد، که در آن ε شامل همه زوج‌های $\{i, j\}$ است به طوری که رأس‌های i و j هیچ یال مشترکی نداشته باشند، اگر و تنها اگر $x_i \perp x_j \mid \mathbf{x}_{-ij}$ ، آن‌گاه \mathbf{x} یک میدان تصادفی مارکوفی گاوسی نسبت به گراف \mathfrak{G} نامیده می‌شود. [۱۹] نشان دادند اگر \mathbf{x} دارای توزیع نرمال چندمتغیره با میانگین μ و ماتریس دقت (عکس ماتریس کوواریانس) $\mathbf{Q} > \mathbf{0}$ باشد، آن‌گاه درایه i - j ام ماتریس \mathbf{Q} یعنی Q_{ij} برابر صفر است، اگر و تنها اگر $x_i \perp x_j \mid \mathbf{x}_{-ij}$. پس درایه i - j ام ماتریس دقت، یعنی Q_{ij} ، مخالف صفر است اگر و تنها اگر $\{i, j\} \in \varepsilon$. [۲۰] نشان دادند در بیشتر موارد \mathbf{Q} ماتریسی $n \times n$ است و فقط $O(n)$ درایه از n^2 درایه ماتریس \mathbf{Q} غیرصفر است که این به علت ویژگی مارکوفی میدان تصادفی است. $n \times n$ بودن این ماتریس باعث افزایش سرعت در محاسبات و الگوریتم‌های تکراری مانند نمونه‌گیری از میدان تصادفی مارکوفی می‌شود. در نهایت بردار تصادفی $\mathbf{x} = (x_1, \dots, x_n)' \in R^n$ یک میدان تصادفی مارکوفی گاوسی تحت گراف نشان‌دار $G = (\nu, \varepsilon)$ با میانگین μ و ماتریس دقت $\mathbf{Q}_{n \times n} > \mathbf{0}$ است اگر و تنها اگر تابع چگالی آن به صورت $\pi(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\mathbf{Q}|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{x} - \mu)^T \mathbf{Q}(\mathbf{x} - \mu)\}$ باشد.

۲.۵ تقریب لاپلاس

با استفاده از تقریب لاپلاس می‌توان هر تابع $g(X)$ را با توزیع نرمال تقریب زد. ابتدا از سری تیلور $\log g(X)$ به صورت زیر است:

$$\log g(X) \approx \log g(\hat{X}) + \frac{\partial \log g(\hat{X})}{\partial X} (X - \hat{X}) + \frac{1}{2} \frac{\partial^2 \log g(\hat{X})}{\partial X^2} (X - \hat{X})^2.$$

از طرفی چون $\frac{\partial \log g(\hat{X})}{\partial X} = 0$ پس

$$\log g(X) \approx \log g(\hat{X}) + \frac{1}{2} \frac{\partial^2 \log g(\hat{X})}{\partial X^2} (x - \hat{X})^2,$$

۵ تحلیل بیزی تقریبی مدل کاکس با تابع خطر پایه تکه‌ای نمایی

رهیافت بیزی تقریبی در [۲۰] معرفی شده است. این رهیافت اغلب در مدل‌های گاوسی پنهان کاربرد دارد. مدل کاکس فضایی با تابع خطر پایه نمایی چون شامل توابع ناشناخته‌ای است که برای در نظر گرفتن اثرات تصادفی مثل اثرات فضایی، خوشه‌ای و... به مدل اضافه می‌شوند، بنا بر این به صورت یک مدل گاوسی پنهان قابل بیان است و می‌توان با استفاده از رهیافت بیزی تقریبی به تحلیل بیزی این مدل پرداخت.

[۱۹] نشان دادند میدان تصادفی گاوسی پنهان به صورت یک میدان تصادفی مارکوفی گاوسی^{۱۳} (GMRF) با یک ماتریس دقت $\mathbf{Q}(\theta)$ قابل بیان است و به همین دلیل می‌توان از روش‌های عددی مربوط به ماتریس‌های $n \times n$ که بسیار سریع‌تر از روش‌های محاسباتی ماتریس‌های چگال هستند استفاده کرد. [۲۰] با به کار بردن میدان تصادفی مارکوفی گاوسی و تقریب لاپلاس یک رهیافت بیزی تقریبی برای مدل‌های گاوسی پنهان ارائه نمودند. برای معرفی این رهیافت ابتدا به معرفی مفاهیم میدان تصادفی مارکوفی گاوسی، تقریب لاپلاس و تقریب لاپلاس آشیانه‌ای جمع‌بسته^{۱۴} (INLA) پرداخته می‌شود.

۱.۵ میدان تصادفی مارکوفی گاوسی

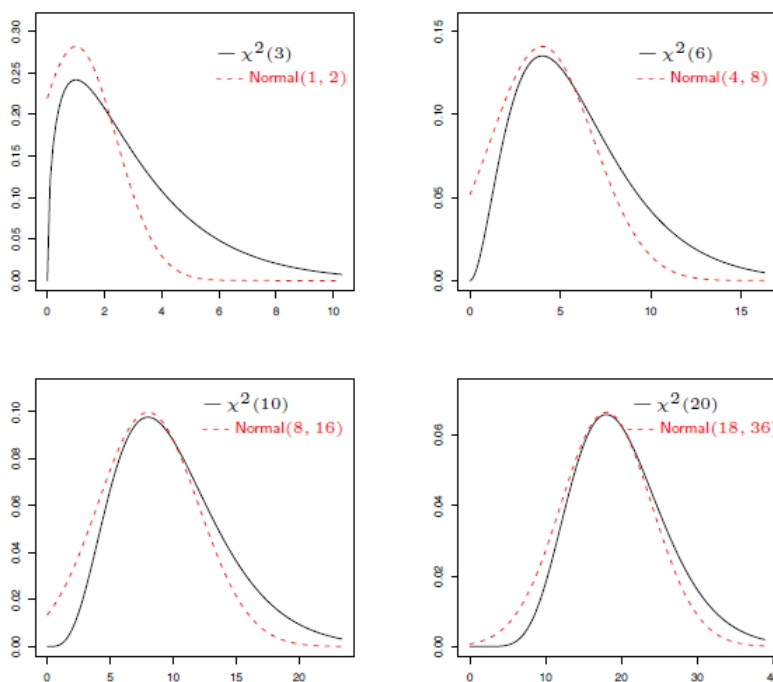
میدان تصادفی مارکوفی گاوسی بر اساس گراف قابل تعریف است. گراف \mathfrak{G} مجموعه‌ای از رأس‌ها است که توسط خانواده‌ای از یال‌ها به هم وصل شده‌اند و به صورت زوج مرتب (ν, ε) نشان داده می‌شود، که در آن ν مجموعه‌ای متناهی و غیر تهی از رئوس و ε یال‌های آن است. اگر رأس‌ها به صورت $\nu = \{1, \dots, n\}$ باشند، آن‌گاه گراف را نشان‌دار^{۱۵} گویند. فرض کنید بردار تصادفی $\mathbf{x} = (x_1, \dots, x_n)'$ دارای توزیع نرمال چندمتغیره با میانگین μ و ماتریس کوواریانس Σ باشد. به علاوه $\Sigma = (\nu, \varepsilon)$

^{۱۳} Gaussian Markov Random Field

^{۱۴} Integrated Nested Laplace Approximation

^{۱۵} Labeled Graph

^{۱۶} Sparse Matrix



شکل ۱. تقریب لاپلاس توزیع χ^2 با نرمال

به طوری که، می‌توان $P(X)$ را به صورت

$$\tilde{P}(X) = N(k - 2, 2(k - 2)),$$

تقریب زد. در شکل ۱ به ازای مقادیر مختلف k توزیع χ^2 دو با تقریب لاپلاس گاوسی، تقریب زده شده است. از تقریب لاپلاس می‌توان هر توزیع پسین شرطی را به صورت

$$\tilde{P}(z | \omega) = \frac{P(x, z | \omega)}{\tilde{P}(x | z, \omega)},$$

تقریب زد، که در آن $\tilde{P}(x | z, \omega)$ تقریب لاپلاس توزیع شرطی x به شرط z, ω است.

۳.۵ تقریب لاپلاس آشیانه‌ای جمع بسته برای مدل

کاکس با تابع خطر تکه‌ای نمایی

در این قسمت ابتدا تحلیل بیزی تقریبی مدل‌های گاوسی پنهان که در [۲۰] ارائه شده است و به رهیافت INLA معروف است، بیان و سپس برای تحلیل بیزی مدل کاکس با تابع خطر تکه‌ای نمایی استفاده می‌شود. فرض کنید x میدان گاوسی پنهان و θ بردار پارامترها است و θ_{-z} بردار حاصل از حذف درایه z -ام است. ابتدا توزیع‌های پسین کناری متغیرهای پنهان و پارامترها

با در نظر گرفتن $\hat{\sigma}^2 = \frac{-1}{\partial^2 \log g(\hat{X})}$ می‌توان نوشت

$$\log g(X) \approx \log g(\hat{X}) - \frac{1}{2\hat{\sigma}^2}(X - \hat{X})^2,$$

و به طور معادل

$$\int g(X) dx = \int e^{\log g(X)} dx \approx \text{const} \int \exp\left[-\frac{(X - \hat{X})^2}{2\hat{\sigma}^2}\right] dx$$

و بنا بر این

$$g(X) \approx N(\hat{X}, \hat{\sigma}^2).$$

مثال: توزیع χ^2 را در نظر بگیرید،

$$P(X) = \frac{g(X)}{c} = \frac{X^{\frac{k}{2}-1} e^{-\frac{X}{2}}}{c}$$

بنا بر این

$$L(X) = \log g(X) = \left(\frac{k}{2} - 1\right) \log x - \frac{X}{2}$$

$$L'(X) = \frac{\partial \log g(X)}{\partial X} = \left(\frac{k}{2} - 1\right) X^{-1} - \frac{1}{2}$$

$$L''(X) = \frac{\partial^2 \log g(X)}{\partial X^2} = -\left(\frac{k}{2} - 1\right) X^{-2}.$$

که با حل $L'(X) = 0$ نما به صورت $\hat{X} = k - 2$ است و با جایگذاری نما در $L''(X)$ ، $\hat{\sigma}^2 = 2(K - 2)$ به دست می‌آید.

به صورت

□ مبدأ مختصات را به نما θ^* انتقال داده، فرمول مختصات در

مبدأ θ^* به صورت $\theta(z) = \theta^* + \mathbf{V}\Lambda^{-1}\mathbf{z}$ تعریف می شود که در آن \mathbf{z} مقادیر استاندارد شده هستند.

$$\pi(\mathbf{x}_i | \mathbf{y}) = \int \pi(\mathbf{x}_i | \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\mathbf{y} \boldsymbol{\theta}, \quad (۴)$$

$$\pi(\boldsymbol{\theta}_j | \mathbf{y}) = \int \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-j}. \quad (۵)$$

□ با شروع از مبدأ مختصات جدید، روی هر یک از محورها

نقاطی به فاصله مقادیر صحیح مناسب δ_z به گونه ای اختیار می شوند، که شرط

$$\log \tilde{\pi}(\boldsymbol{\theta}(\circ) | \mathbf{y}) - \log \tilde{\pi}(\boldsymbol{\theta}(\mathbf{z}) | \mathbf{y}) < \delta_\pi, \quad (۶)$$

برقرار باشد. معمولاً $\delta_\pi = ۲/۵$ مناسب است، [۲۰]. به طور مشابه نقاط درون صفحات نیز تعیین می شوند.

با به کار بردن این الگوریتم، θ_e هایی از توزیع $\pi(\boldsymbol{\theta} | \mathbf{y})$ تولید می شوند که می توان از آنها برای تقریب توزیع $\pi(\boldsymbol{\theta} | \mathbf{y})$ استفاده کرد. به عنوان مثال فرض کنید $\mathbf{x} = (x_1, x_2)'$ دارای توزیع نرمال با پارامتر مکان $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ و ماتریس کوواریانس همسانی باشد. همچنین توزیع پیشین $\pi(\boldsymbol{\theta}) \propto ۱$ و مشاهدات $\mathbf{y} = (۲, ۵)'$ را برای متغیر پاسخ دو جمله ای با تعداد آزمایشه $u = (۱۰, ۱۰)'$ در نظر بگیرید. با در نظر گرفتن $\delta_z = ۰/۵$ ، نقاط روی محورهای t_1 و t_2 به گونه ای تعیین شوند تا شرط (۶) برقرار باشد. شکل ۲ (چپ) خطوط تراز $\hat{\pi}(\boldsymbol{\theta} | \mathbf{y})$ را نشان می دهد، که در آن مبدأ مختصات به نما $\theta^* = (\theta_1^*, \theta_2^*)'$ انتقال یافته است. بردارهای ویژه بر اساس ماتریس هسین در نما محاسبه شده اند. شکل ۲ (راست) نمودار خطوط تراز و ۶۸ نقطه تولید شده با الگوریتم بالا را نشان می دهد.

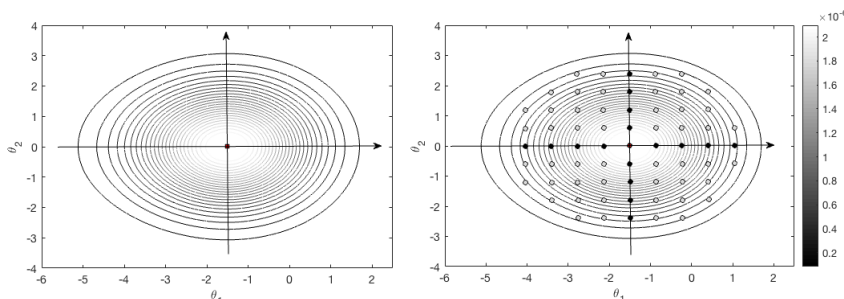
برای $i, j = 1, \dots, m$ تشکیل می شوند. در روش INLA برای چگالی های پسینی حاشیه ای (۴)، (۵) تقریب هایی به صورت $\tilde{\pi}(x | \mathbf{y})$ و $\tilde{\pi}(\boldsymbol{\theta}_j | \mathbf{y})$ ارائه می شود. ابتدا با استفاده از بسط تیلور و خطی سازی قسمت درست نمایی حول یک مقدار ثابت، توزیع $\pi(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta})$ با استفاده از یک توزیع نرمال چندمتغیره $\tilde{\pi}_G(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta})$ تقریب زده و مقدار چگالی در نما توزیع محاسبه می شود. سپس برای به دست آوردن توزیع پسین تقریبی پارامترها با به کار بردن تقریب لاپلاس، توزیع تقریبی پسین به صورت

$$\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x} | \mathbf{y}, \boldsymbol{\theta})} \Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}$$

می شود، که در آن $\mathbf{x}^*(\boldsymbol{\theta})$ نما برای توزیع شرطی کامل \mathbf{x} به شرط $\boldsymbol{\theta}$ است. برای تولید نقاط از توزیع پسین تقریبی پارامترها از الگوریتم زیر استفاده می شود:

□ ابتدا نما برای تابع $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$ با روش های بهینه سازی به صورت θ^* تعیین می شود.

□ ماتریس هسین^{۱۷} در نما محاسبه می شود و عکس آن به صورت $\mathbf{H}^{-1} = \mathbf{V}\Lambda\mathbf{V}'$ تجزیه می شود، که در آن \mathbf{V} ماتریس بردارهای ویژه و Λ ماتریس قطری مقادیر ویژه آن است.



شکل ۲. اکتشاف پسین پارامتر $\boldsymbol{\theta}$ ، (چپ) نما معلوم شده و مبدأ مختصات به نما انتقال داده شده است. (راست) روی محورهای مختصات و درون صفحات نقاط تولید شده اند.

^{۱۷} Hessian Matrix

است. این توزیع پسین معمولاً صورت صریحی ندارد، بنا بر این می‌توان از روش‌های تکرارشونده مثل الگوریتم مونته کارلوی زنجیر مارکوفی برای به دست آوردن برآورد پارامترها استفاده کرد که بسیار زمان‌بر و گاهی اوقات همگرایی به راحتی به دست نمی‌آید و از رهیافت بیزی تقریبی و الگوریتمی که در بالا توضیح داده شد برای به دست آوردن توزیع‌های پسین پارامترها و متغیرهای پنهان به صورت تقریبی استفاده و (۹) به صورت تقریبی مشخص و در نهایت از رابطه (۸) توزیع‌های پسین کناری محاسبه و برآوردهای بیزی تقریبی به دست آورده می‌شود. در ادامه شیوه تحلیل بیزی تقریبی مدل کاکس با تابع خطر تکه‌ای نمایی در یک مثال کاربردی مورد بررسی قرار می‌گیرد.

۶ مثال واقعی

داده‌های مورد بررسی مشاهدات مربوط به $m = 1043$ بیمار مبتلا به سرطان خون بین سال‌های ۱۹۸۲ الی ۱۹۹۸ در شمال غرب انگلستان در $n_s = 24$ ناحیه و $t = 1, \dots, 4977$ است، که حدود ۱۶ درصد موارد سانسور شده از سمت راست هستند. برای تحلیل این داده‌ها از نرم‌افزار R استفاده شد که کدهای آن در پیوست موجود است، [۱۵، ۱۲] به مطالعه این داده‌ها پرداخته‌اند.

متغیرهای توضیحی برای هر بیمار i شامل سن بیمار (age_i)، تعداد سلول‌های سفید خون (wbc_i)، شاخص محرومیت تاونزند^{۱۹}، (tpi_i) جنسیت بیمار (sex_i) و منطقه سکونت (s_i) هستند. نمودار داده‌ها در شکل مشخص شده است. در این‌جا زمان به $K = 20$ بازه زمانی مساوی افزاز می‌شود و چهار مدل کاکس با تابع خطر پایه تکه‌ای نمایی و با همبستگی فضایی، بدون همبستگی فضایی، بدون اثرات غیرخطی و با افزایش اثرات غیرخطی در نظر گرفته می‌شود و سپس با معیار مدل‌گزینی DIC مورد مقایسه قرار می‌گیرند. برای افزازهای زمانی $K = 10$ و $K = 50$ محاسبات تکرار و تفاوت محسوسی در نتایج ایجاد نشد.

پس از تولید مقادیر θ از روش فوق، اکنون تقریب لاپلاس

توزیع شرطی کامل $\pi(x_i | \mathbf{y}, \theta)$ به صورت

$$\tilde{\pi}_{LA}(x_i | \theta, \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \theta, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}_{-i} | x_i, \theta, \mathbf{y})} \Big|_{x_i = x_{-i}^*(x_i, \theta)}, \quad (7)$$

است، که در آن بردار حاصل از حذف درایه i -ام است. $\tilde{\pi}_G(\mathbf{x}_{-i} | x_i, \theta, \mathbf{y})$ تقریب گاوسی از $\pi(\mathbf{x}_{-i} | x_i, \theta, \mathbf{y})$ که x_i یک مقدار ثابت و $\mathbf{x}_{-i}^*(x_i, \theta)$ نما برای $\pi(\mathbf{x}_{-i} | x_i, \theta, \mathbf{y})$ است. در نهایت پس از محاسبه توزیع تقریبی پسین ابرپارامترها و توزیع شرطی کامل متغیرهای پنهان اکنون می‌توان توزیع پسین کناری متغیرهای پنهان را از رابطه

$$\begin{aligned} \pi(x_i | \mathbf{y}) &= \int \pi(x_i | \mathbf{y}, \theta) \pi(\theta | \mathbf{y}) d\theta \\ &\approx \sum_{\ell=1}^L \tilde{\pi}(x_i | \theta_\ell, \mathbf{y}) \tilde{\pi}(\theta_\ell | \mathbf{y}) \Delta_\ell. \end{aligned} \quad (8)$$

محاسبه کرد، که در آن جمع روی نقاط تولید شده برای θ بسته شده است، با وزن‌های ناحیه‌ای Δ_ℓ برای $\ell = 1, \dots, L$. چون تولید نقاط از روش توضیح داده شده، با فواصل یکسان در یک شبکه منظم تولید می‌شوند، Δ_ℓ ها با هم برابر هستند. به دلیل مشابه بودن رابطه (۷) به تقریب لاپلاس و تقریب توزیع پسین (۸) با جمع روی مقادیر θ تولید شده این روش، رهیافت تقریبی لاپلاس آشیانه‌ای جمع بسته نامگذاری شده است.

اکنون برای تحلیل بیزی تقریبی مدل (۲) و به دست آوردن برآوردهای بیزی پارامترهای این مدل، توزیع پیشین گاوسی مبهم^{۱۸} برای بردار پارامترهای رگرسیونی β و یک پیشین گاوسی برای بردار (b_1, \dots, b_k) و $\mathbf{f}_r(\cdot)$ پیشگوی η_{ik} به صورت گاوسی توزیع می‌شود و با تعریف بردار \mathbf{x} به صورت

$$\mathbf{x} = \{\{\mathbf{f}_r(\cdot)\}, \{\beta_j\}, \{\eta_{ik}\}, \{b_k\}\},$$

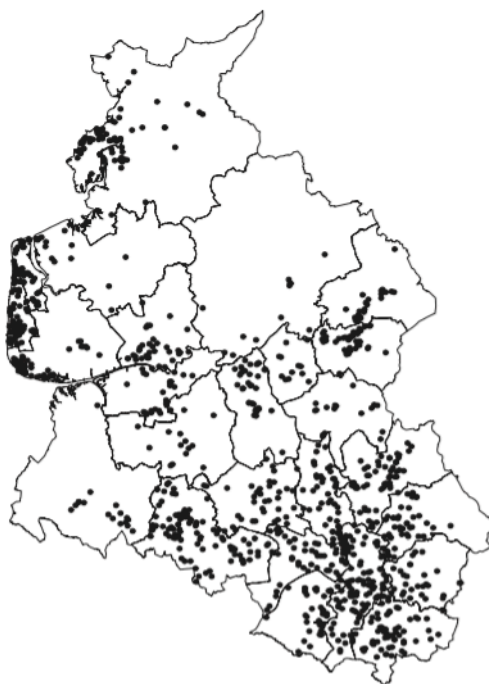
آن‌گاه می‌توان بردار پنهان \mathbf{x} را با یک میدان پنهان گاوسی با ماتریس دقت $Q^{-1}(\theta)$ ، به صورت $\mathbf{x} | \theta \sim N(\mathbf{0}, Q^{-1}(\theta))$ مدل‌بندی نمود که در آن بردار ابر پارامترها است. فرض کنید $y_i = ((t_i, \delta_i), i = 1, \dots, m)$ داده‌های بقا هستند، که به شرط

میدان پنهان \mathbf{x} مستقلند. بنا بر این توزیع پسین به صورت

$$\pi(\mathbf{x}, \theta | \mathbf{y}) \propto \pi(\theta) \pi(\mathbf{x} | \theta) \prod_{i=1}^m \pi(y_i | \mathbf{x}, \theta), \quad (9)$$

^{۱۸} Vague Prior

^{۱۹} Townsend



شکل ۳. موقعیت داده‌های سرطان خون در شمال غرب انگلستان

در مدل اول اثرات سن و جنس به صورت خطی و اثرات تعداد سلول‌های سفید و شاخص تاونزند به صورت غیرخطی به صورت قدم زدن تصادفی مرتبه دو با پارامتر دقت τ_{wbc} و τ_{tpi} بدون اثر فضایی در نظر گرفته شد، بنا بر این مدل (۱) به صورت

$$M_2 : \eta_{ik} = \beta_0 + \beta_1 sex_i + \beta_2 age_i + \beta_3 wbc_i + f_1(tpi) + f_{spat}(s_i) + b_k.$$

در مدل دوم اثرات سن، جنس و تعداد سلول‌های سفید به صورت خطی و اثر شاخص تاونزند به صورت غیرخطی قدم زدن تصادفی مرتبه دو با پارامتر دقت τ_{tpi} و اثر فضایی به صورت

$$M_3 : \eta_{ik} = \beta_0 + \beta_1 sex_i + \beta_2 age_i + \beta_3 wbc_i + b_k$$

است. در مدل ۳ فقط اثرات سن، جنس و تعداد سلول‌های سفید به صورت خطی در مدل در نظر گرفته شد، لذا مدل به صورت

است. در مدل ۴، اثرات سن، جنس و تعداد سلول‌های سفید به صورت خطی و اثر شاخص تاونزند به صورت غیرخطی قدم زدن تصادفی مرتبه دو با پارامتر دقت τ_{tpi} و اثر فضایی به صورت مدل BYM^{۲۰} معرفی شده است [۶]، در نظر گرفته شد. در مدل BYM دو نوع اثر تصادفی وارد مدل می‌شود، اثر تصادفی فضایی و اثر ناهمگنی به طوری که اثر فضایی، همبستگی فضایی مشاهدات را در مدل بیان می‌کند و اثر ناهمگنی برای این فرض وارد مدل می‌شود که برآوردهای به دست آمده بین نواحی مختلف مستقل از یکدیگر هستند. یعنی دو اثر همبستگی فضایی درونی و استقلال برونی را با مدل BYM می‌توان در نظر گرفت. مدل

در نظر گرفته می‌شود، به طوری که

$$M_1 : \eta_{ik} = \beta_0 + \beta_1 sex_i + \beta_2 age_i + f_1(wbc_i) + f_2(tpi_i) + b_k$$

در نظر گرفته می‌شود، به طوری که

$$\pi(f_2 | \tau_{tpi}) \propto \tau_{tpi}^{\frac{(n_{tpi} - 2)}{2}} \exp \left\{ -\frac{1}{\tau_{tpi}} \sum_{i=3}^{n_{tpi}} (f_{2i} - \frac{1}{2} f_{2(i-1)} + f_{2(i-2)})^2 \right\},$$

$$\pi(f_1 | \tau_{wbc}) \propto \tau_{wbc}^{\frac{(n_{wbc} - 2)}{2}} \exp \left\{ -\frac{1}{\tau_{wbc}} \sum_{i=3}^{n_{wbc}} (f_{1i} - \frac{1}{2} f_{1(i-1)} + f_{1(i-2)})^2 \right\}.$$

هستند. در مدل دوم اثرات سن، جنس و تعداد سلول‌های سفید به صورت اثرات خطی و اثر شاخص تاونزند به صورت غیرخطی قدم زدن تصادفی مرتبه دو با پارامتر دقت τ_{tpi} و اثر فضایی به صورت یک میدان تصادفی مارکوفی گاوسی با پارامتر دقت

^{۲۰} Besag, York and Mollie Model

(۴) به صورت

$$M_{\tau} : \eta_{ik} = \beta_0 + \beta_1 sex_i + \beta_2 age_i + \beta_3 wbc_i + f_1(tpi) + f_{i.i.d.}(H_i) + f_{spat}(s_i) + b_k \quad (10)$$

است. در هر چهار مدل b_k فرایند قدم زدن تصادفی مرتبه اول با توزیع نرمال به صورت $b_{k+1} - b_k \sim N(0, \tau_b)$, $k = 1, \dots, 19$ فرض شد. همچنین در مدل دوم و چهارم، اثر فضایی به صورت فرایند اتورگرسیو شرطی ذاتی^{۲۱} توسط یک میدان تصادفی مارکوفی گاوسی مدل‌بندی شده است یعنی

$$[(f_{spat})_i | (f_{spat})_{-i}, \tau_s] \sim N\left(\frac{1}{n_i^s} \sum_{j \in \partial_i} (f_{spat})_j, \frac{1}{n_i^s \tau_s}\right)$$

که در آن ∂_i مجموعه همسایگی مربوط به موقعیت i -ام و n_i^s تعداد آنها، برای تمام پارامترهای دقت توزیع پیشین مبهم گاما با پارامترهای $(1, 0.001)$ و برای پارامترهای رگرسیونی توزیع پیشین مبهم نرمال با واریانس بزرگ در نظر گرفته شده است. برای مقایسه چهار مدل از معیار DIC بیزی استفاده شد و به صورت زیر $DIC = \bar{D}(\theta) + P_D = D(\bar{\theta}) + 2P_D$ تعریف

می‌شود، که در آن $\bar{D}(\theta)$ میانگین پسین انحراف است، که میزان برازش مدل را بیان می‌کند. $P_D = \bar{D}(\theta) - D(\bar{\theta})$ میزان پیچیدگی مدل را نشان می‌دهد و تعداد پارامترهای مؤثر مدل نامیده می‌شود. مدلی که DIC کمتری دارد، مدل بهتری است. نتایج در جدول ۱ ارائه شده است.

با توجه به تعداد پارامترهای مؤثر، DIC و لگاریتم درست‌نمایی، مدل دوم و چهارم از دو مدل اول و سوم مناسب‌تر به نظر می‌رسند. مدل دوم و مدل چهارم به عنوان مدل‌های مناسب انتخاب می‌شوند، که بیان‌کننده لزوم وجود اثر فضایی در مدل پرداخته می‌شود. ابتدا با انتخاب مدل چهارم به بررسی بیشتر این مدل می‌پردازیم. نتایج مدل ۴ در جدول ۲ ارائه شده است. τ_s و τ_H به ترتیب پارامترهای دقت اثر فضایی و اثر ناهمگنی هستند. شکل ۴ نمودار توزیع چگالی‌های پسین برای چهار اثر خطی، اثر ثابت، اثر تعداد سلول‌های سفید خون، جنس و سن ارائه گردیده است. در شکل ۵، نمودار چگالی پسین برای τ_H ، τ_{tpi} و τ_b رسم شده است.

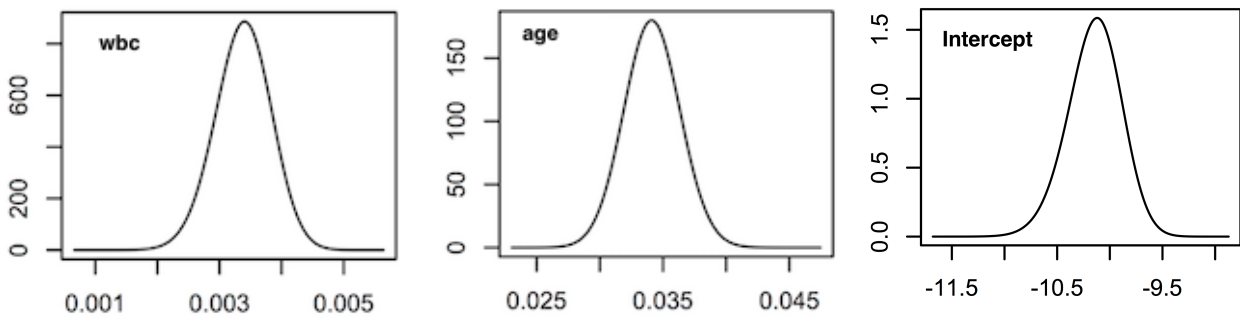
جدول ۱. مقایسه مدل‌ها

log - likelihood	DIC	P_D	Model
-۲۶۵۶/۲۹	۵۳۸۱/۱۱	۱۶/۴۹	M_1
-۲۷۶۷/۳۹	۵۳۵۴/۷۷	۲۴/۵۱	M_2
-۲۶۷۴/۸۶	۵۳۹۴/۸۶	۱۱/۷۳	M_3
-۲۷۴۴/۱۰	۵۳۵۶/۶۷	۲۶/۶۷	M_4

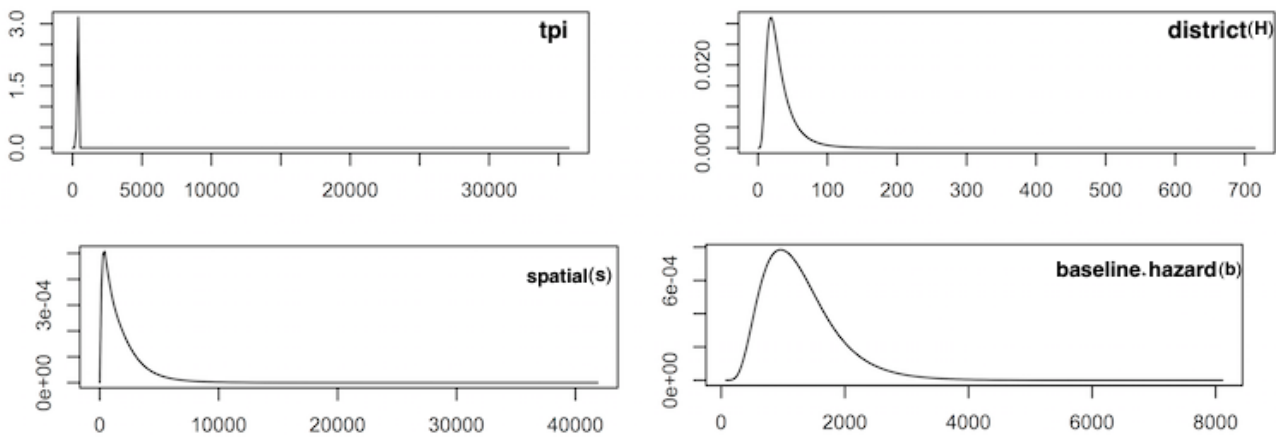
جدول ۲. برآورد بیزی تقریبی پارامترهای مدل ۴، (میانگین پسین، انحراف معیار و چندک‌ها)

پارامتر	میانگین	انحراف معیار	چندک‌ها		
			۰/۵	۰/۵	۰/۹۷۵
β_0	-۱۰/۱۴۸۶	۰/۲۵۵۳	-۱۰/۶۷۷۲	-۱۰/۱۳۹۲	-۹/۶۷۳۲
β_1	۰/۰۳۴۱	۰/۰۰۲۲	۰/۰۲۹۸	۰/۰۳۴۱	۰/۰۳۸۵
β_2	۰/۰۰۳۴	۰/۰۰۰۵	۰/۰۰۲۵	۰/۰۰۳۴	۰/۰۰۴۲
τ_{tpi}	۷۱۵۵/۷۰	۷۳۸۷/۹۸	۳۷۰/۷۵۱	۴۹۱۰/۴۸	۲۷۵۵۴/۶۱
τ_H	۳۲/۵۷	۲۲/۱۱	۹/۰۱۹	۲۶/۶۸	۹۰/۶۹
τ_s	۱۷۶۲/۴۰	۱۷۹۵/۴۷	۱۰۲/۱۰۳	۱۲۲۲/۲۸	۶۵۰۴/۹۸
τ_b	۱۲۸۳/۵۴	۶۰۶/۳۷	۴۴۵/۴۸۶	۱۱۷۳/۷۶	۲۷۶۷/۶۹

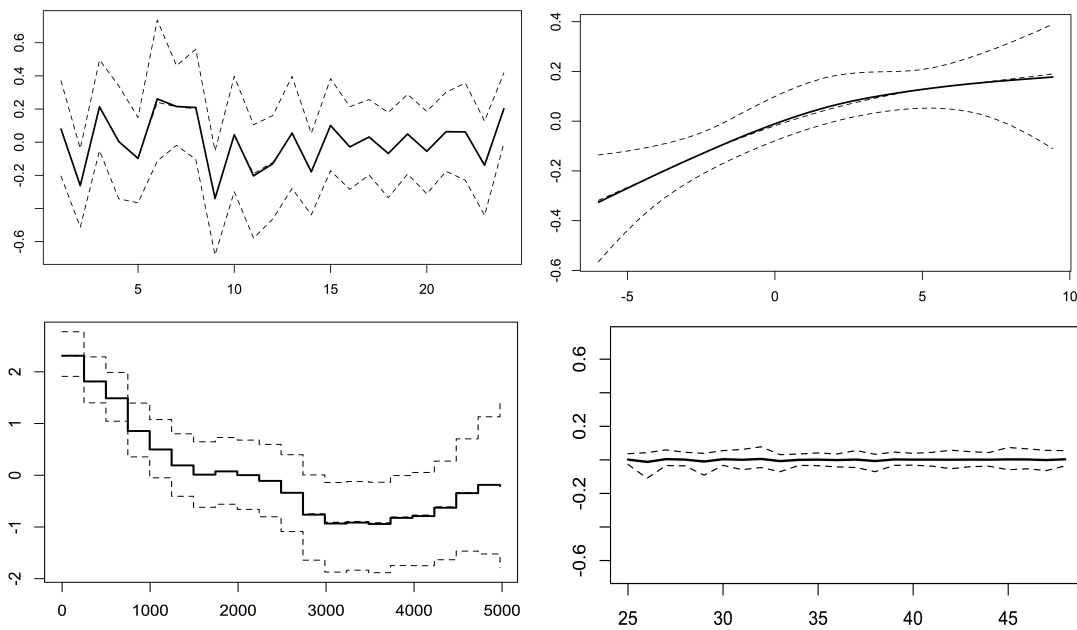
^{۲۱} Intrinsic Conditiond Aut- Regressive



شکل ۴. نمودار توزیع‌های چگالی پسین برای اثرات خطی، اثر ثابت، سن و تعداد سلول‌های سفید.



شکل ۵. نمودار چگالی پسین پارامترهای دقت τ_{tpi} ، τ_H ، τ_S و τ_b .



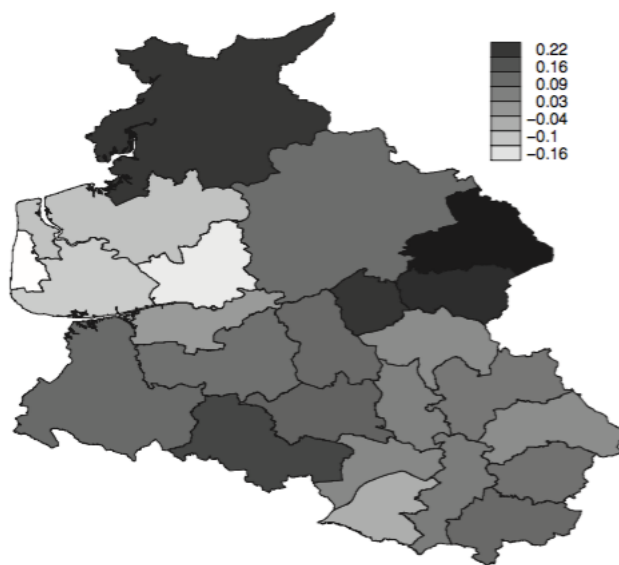
شکل ۶. نمودار پیشگویی و فاصله پیشگویی ۹۵ درصد برای (راست-بالا) شاخص تاونزند (چپ-بالا) اثر فضایی (راست-پایین) اثر ناهمگنی فضایی (چپ-پایین) اثر لگاریتم خطر پایه.

در منطقه غرب متمرکز شده‌اند و نواحی با ریسک بالا بر روی نقشه پراکنده‌تر هستند. برای مدل ۲ و به‌ازای مقادیر مختلف K ، $K = 20$ ، $K = 50$ و $K = 100$ نمودار میانگین پسین و فاصله پیشگویی ۹۵ درصد برای تابع خطر پایه لگاریتمی در شکل ۸ رسم شده است. همان‌طور که ملاحظه می‌شود با افزایش تعداد بازه‌ها یعنی K برآورد با جزئیات بیشتر برای تابع خطر پایه به‌دست آمده است. در این نمودارها ملاحظه می‌شود که تابع خطر پایه در بیشتر دوره مطالعه یعنی تا نقطه زمانی حدود ۳۷۰۰۰ یعنی ۱۰ سال پس از مطالعه افزایشی است، به‌طوری‌که در سال ۱۰-م فقط ۲۶ نفر در مطالعه باقی‌مانده است و به‌همین دلیل در نقاط انتهایی نمودار، به نظر می‌رسد که خطر پایه افزایشی شده است.

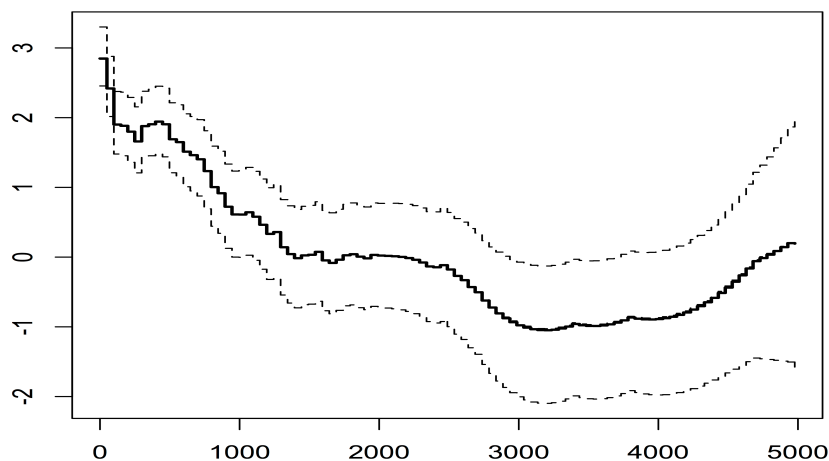
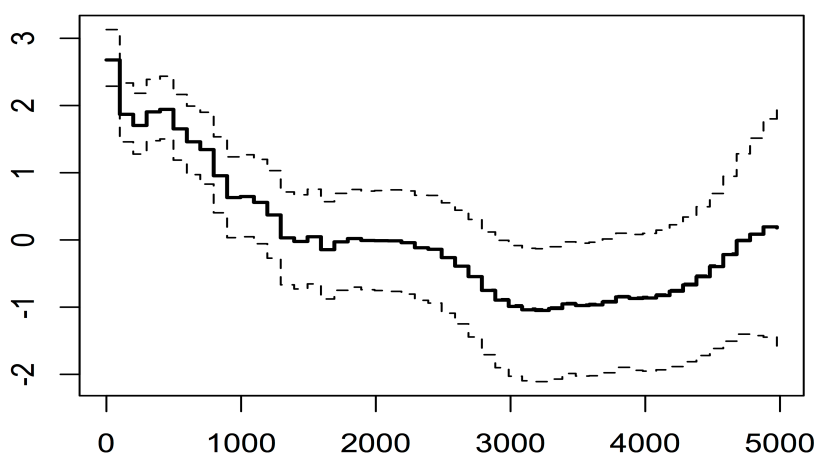
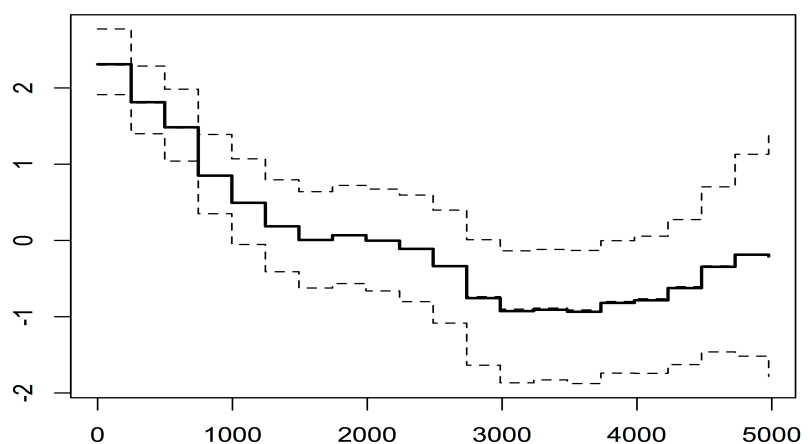
شکل ۵ (بالا)، نمودارهای پیشگویی و فاصله پیشگویی ۹۵ درصد شاخص تاونزند و اثر فضایی هستند. اثر شاخص تاونزند ابتدا افزایشی است و سپس تقریباً بعد از رسیدن به مقدار حدود صفر ثابت است. در شکل ۶ (پایین)، پیشگویی و فاصله پیشگویی ۹۵ درصد برای تابع خطر پایه لگاریتمی و اثر ناهمگنی فضایی رسم شده است. با توجه به نمودار پیشگویی اثر ناهمگنی فضایی به نظر می‌رسد برآورد این اثر صفر است و وجود آن در مدل غیر ضروری است و لذا مدل دوم که مشابه مدل چهارم بدون اثر ناهمگنی فضایی است، مدل مناسب داده‌ها است. نتایج تحلیل بیزی تقریبی مدل دوم در جدول ۳ ارائه شده است. در شکل ۷ برآورد اثرات فضایی بر روی نقشه منطقه مورد بررسی نشان داده شده است. ملاحظه می‌شود که نواحی با ریسک کمتر

جدول ۳. برآورد بیزی تقریبی پارامترهای مدل ۲، (میانگین پسین، انحراف معیار و چندک‌ها)

پارامتر	میانگین	انحراف معیار	چندک‌ها		
	۰/۲۵	۰/۵	۰/۹۷۵	۰/۵	۰/۹۷۵
β_0	-۱۰/۱۶۳۸	۰/۲۵۱۶	-۱۰/۶۸۵۵	-۱۰/۱۵۴۳	-۹/۶۹۶۳
β_1	۰/۰۳۴۵	۰/۰۰۲۲	۰/۰۳۰۲	۰/۰۳۴۵	۰/۰۳۸۹
β_2	۰/۰۰۳۴	۰/۰۰۰۵	۰/۰۰۲۵	۰/۰۰۳۴	۰/۰۰۴۳
$T_{t_{pi}}$	۸۷۹۱/۱۵	۹۷۶۷/۰۹	۳۹۱/۷۲۴	۵۷۳۴/۹۴	۳۵۳۸۹/۴۴
τ_s	۱۵/۳۸	۱۱/۴۰	۳/۹۰۲	۱۲/۲۳	۴۵/۴۸
τ_b	۱۲۹۲/۰۶	۶۰۵/۱۴	۴۴۳/۹۵۰	۱۱۸۶/۸۲	۲۷۶۲/۸۲



شکل ۷. برآورد اثرات فضایی بر روی نقشه منطقه مورد بررسی



شکل ۸. میانگین پسین (خط ممتد) و فواصل پیشگویی ۹۵ درصد (خط پسین) برای تابع خطر پایه لگاریتمی با $K = 20$ (بالا)، $K = 50$ (وسط) و $K = 100$ (پایین).

بحث و نتیجه گیری

حجم نمونه کم باشد یا اطلاعاتی پیرامون پارامترها در دسترس باشد، استفاده از رهیافت بیزی نسبت به روشهای کلاسیک برتری دارد. همچنین در عمل گاهی داده‌ها وابسته به مکان هستند، بنا بر این برای در نظر گرفتن همبستگی فضایی یک اثر

اغلب در مطالعات تحلیل بقا از روشهای کلاسیک پارامتری و نیمه پارامتری کاکس استفاده می‌شود، اما در برخی مطالعات اگر

غیر خطی فضایی به مدل اضافه و این اثر فضایی به صورت یک میدان تصادفی گاوسی مارکوفی مدل‌بندی شد. در این مقاله یک نوع مدل کاکس با تابع خطر پایه تکه‌ای نمایشی ارائه شد و در راستای کاهش زمان محاسبات و اجرای برنامه‌های کامپیوتری این مدل با رهیافت بیزی تقریبی مورد مطالعه قرار گرفت. به دلیل پنهان بودن برخی اثرات مثل اثر فضایی، توزیع این اثرات معمولاً نرمال فرض می‌شود. به عنوان پیشنهاد می‌توان در راستای افزایش دقت تحلیل‌ها این اثرات را با یک کلاس بزرگ‌تری از توزیع‌ها مثل چوله نرمال مدل‌بندی کرد.

مراجع

- [۱] باغستانی، ا.، حاجی‌زاده، ا.، و فاطمی، ر.، (۱۳۸۹). مقایسه مدل‌های بیزی پارامتری در تحلیل عوامل مؤثر بر میزان بقای بیماران مبتلا به سرطان معده، مجله تخصصی اپیدمیولوژی ایران، دوره ۶، شماره ۳، صفحات ۲۱-۱۸.
- [۲] نورکجوری، ه.، حاجی‌زاده، ا.، باغستانی، ا.، و پورحسینقلی، م. ا.، (۱۳۹۱). استفاده از مدل اسپلاین جریمه شده در تعیین عوامل مؤثر بر بقای بیماران مبتلا به سرطان معده، مجله علمی پژوهشی دانشگاه علوم پزشکی اراک، سال ۱۵، شماره ۶۹، صفحات ۹۵-۱۰۳.
- [۳] پورحسینقلی، م. ا.، حاجی‌زاده، ا.، ابدی، ع. ر.، صفایی، آ.، مقیمی‌دهکردی، ب. و زالی، م. ر.، (۱۳۸۶). مقایسه رگرسیون کاکس و مدل‌های پارامتری در تحلیل بقای بیماران مبتلا به سرطان معده، مجله تخصصی اپیدمیولوژی ایران، دوره ۳، شماره ۲، صفحات ۲۵-۲۹.
- [4] Banerjee, S. and Carlin, B. P. (2003). Semiparametric spatio-temporal frailty modelling, *Environmetrics* **14**, 523–535.
- [5] Banerjee, S., Wall, M. M. and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data with application to infant mortality in Minnesota, *Biostatistics*, **4**, 123–142.
- [6] Besag, J., York, J. and Mollie, A. (1991). Bayesian image r-storation, with two applications in spatial statistics, *Annals of the institute of statistical Mathematics*, **43 (1)**, 1-59.
- [7] Biswas A, Datta S, Fine JP, Segal MR. (2008). *Statistical Advances in The Biomedical Sciences: Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics*, John Wiley and Sons, Inc.
- [8] Cox, D. R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society Series. B*, **34**, 187–220.
- [9] Dasgupta, P., Cramb, S. M., Aitken, J. F., Turrell, G., Baade, P. D. (2014). Comparing multilevel and Bayesian spatial random effects survival models to assess geographical inequalities in colorectal cancer survival: A case study, *International Journal of Health Geographics*, **13**, 36-50.
- [10] Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors, *Journal of the Royal Statistical Society Series C*, **50(2)**, 201–220.

- [11] Hannerfeind, A. Brezger, A. and Fahrmeir, L. (2006). Geoadditive survival models, *Journal of the American Statistical Association*, **101(475)**, 1065–1075.
- [12] Henderson, R. Shimakura, S. and Gorst, D. (2002). Modeling spatial variation in leukemia survival data, *Journal of the American Statistical Association*, **97**, 965–972.
- [13] Ibrahim, J., Chen, M. and Sinha, D. (2001). *Bayesian Survival Analysis*, Springer, New York.
- [14] Kneib, T. (2006). Geoadditive hazard regression for interval censored survival times, *Computational Statistical and Data Analysis*, **51**, 777–792.
- [15] Kneib, T. and Fahrmeir, L. (2007). A mixed model approach for geoadditive hazard regression, *Scandinavian Journal of Statistics*, **34**, 207–228.
- [16] Lang, S. and Brezger, A. (2004). Bayesian P-Splines, *Journal of Computational and Graphical Statistics*, **13**, 183–212.
- [17] Martino, S., Akerkar, R., and Rue, H. (2011). Approximate Bayesian Inference for Survival Models, *Scandinavian Journal of Statistics*, **38**, 514-528.
- [18] Pana, C., Caia, B., Wangb, L., Lin, X. (2014). Bayesian semiparametric model for spatially correlated interval-censored survival data, *Computational Statistics and Data Analysis*, **74**, 198-208
- [19] Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, Volume 104 of Monographs on Statistics and Applied. Chapman and Hall, London.
- [20] Rue, H. and Martino, S. (2009). Approximate bayesian inference for hierachical gaussian markov random fields models, *Journal of Statistical Planning and Inferenc*, **137**, 3177-3199.
- [20] Rue, H., Martino, S. and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations, *Journal of the Royal Statistical Society*, **71**, 319-392.