

مدل سازی پیشامدهای بازگردنده تحت مخاطره رقابتی، همراه با پیشامد بدون بازگشت

علی شریفی، سید رضا هاشمی
گروه آمار، دانشگاه رازی کرمانشاه

تاریخ دریافت: ۱۳۹۲/۷/۳ تاریخ آخرین بازنگری: ۱۳۹۳/۲/۱۹

چکیده: در این مقاله برای تحلیل بقای داده‌های حاصل از پیشامد بازگردنده و مخاطره رقابتی، یک مدل نیمه پارامتری ضربی-جمععی برای تابع نرخ مخاطره در نظر گرفته شده است که دارای یک تابع مخاطره پایه نامعلوم است. مدل شامل ضرایب رگرسیونی برای متغیرهای کمکی موثر در شکست و متغیر شکنندگی جهت بیان همبستگی بین فرایند پیشامد انتهایی و پیشامد بازگردنده است و قابلیت تحلیل داده‌هایی با سانسور راست و سانسور آگاهنده را دارد. برای بیشینه کردن تابع درست‌نمایی به روش درست‌نمایی محلی، از روش عددی و از بسط سری تیلور برای برازش مخاطره پایه استفاده شده است. در نهایت با شبیه‌سازی، روش تایید شده و مدل معرفی شده روی داده‌های واقعی، به کار رفته و پارامترها برآورد شده‌اند.

واژه‌های کلیدی: مدل نیمه پارامتری، تابع نرخ مخاطره، مخاطره پایه، پیشامد بازگردنده، مخاطره رقابتی، سانسور آگاهنده، پیشامد انتها.

در مطالعات پزشکی، گاهی با بیماری‌هایی مواجه هستیم که پس از درمان و بهبود موقت، امکان ابتلا و عود مجدد بیماری وجود دارد. این گونه بیماری‌ها عود پذیر یا بازگردنده هستند و مشاهدات آنها داده‌های پیشامدهای بازگردنده^۱ نامیده می‌شوند. این گونه داده‌ها معمولاً در مطالعات طولی مثل مطالعات زیستی، موارد مربوط به قابلیت اعتماد، اقتصاد، جرم شناسی و جمعیت شناسی، مشاهده می‌شوند. عفونت‌های ناشی از دیالیز، تعداد حملات صرع، بستری شدن‌های متوالی در بیمارستان، تکرار یک سری عادت خاص، انجام جرم‌های مشابه و دستگیر شدن‌های متوالی و خراب شدن‌های پی‌در پی یک سیستم، مثال‌هایی از پیشامدهای بازگردنده هستند. در مطالعات مربوط به تحلیل بقا ممکن است شکست یا وقوع پیشامد مورد نظر، ناشی از چندین رقیب باشد، یعنی چندین بیماری به صورت رقیب‌وار تلاش در شکست فرد داشته باشند، در این صورت مخاطره ناشی از این رقیبها را مخاطره رقابتی^۲ می‌نامند. معمولاً برای سادگی محاسبات و برقراری بسیاری از پذیره‌های تحلیل بقا، تاثیر رقبا مستقل از هم در نظر گرفته می‌شود.

یک رهیافت‌ها متداول برای تحلیل داده‌های بازگردنده استفاده از مدل‌های نرخ جمعی حاشیه‌ای است، اما این مدل‌ها، در اغلب موارد، به دلیل نادیده گرفتن متغیرهای توصیفی که می‌توانند تاثیرگذار باشند، نتیجه مورد انتظار را تولید نمی‌کنند و اربیبی و متوسط توان‌های دوم خطای بالایی در بر دارند. علاوه بر آن وقتی برای داده‌ها پیشامدی پایانی^۳ (مانند مرگ) در نظر گرفته می‌شود و برای برخی از افراد پیشامد بدون تکرار^۴ رخ می‌دهد، به دلیل وجود وابستگی بین وقوع این گونه پیشامدها و پیشامدهای بازگردنده، با مشکلاتی در محاسبات و عدم برآورد پذیری پارامترهای مدل روبه‌رو می‌شویم. هم‌چنین تا به حال مدلی برای داده‌های پیشامدهای بازگردنده که در آن اثر مخاطره‌های رقابتی لحاظ شده باشد و این گونه

^۱ Recurrent events

^۲ Competing risk

^۳ Terminal event

^۴ Zero-recurrence

پیشامدها تحت چند رقیب مطالعه شوند و پیشامد انتها نیز در آن حضور داشته باشد، بررسی نشده است. برای حل این مشکل، در این مقاله مدلی نیمه پارامتری^۵ با یک تابع نرخ جمعی^۶ و یک مخاطره پایه کلی در نظر گرفته می شود. پرنیس و همکاران (۱۹۸۱) و اندرسون و گیل (۱۹۸۲) با استفاده از رویکرد تابع شدت (نرخ مخاطره) مدل کاکسی را در نظر گرفته بودند که در آن سانسور راست اتفاق افتاده بود و توزیع تابع شدت^۷ مربوط به پیشامد بازگردنده، پواسون در نظر گرفته شده بود.

ویی و همکاران (۱۹۸۹)، پیپ و جای (۱۹۹۳) و لیو و همکاران (۲۰۰۴) با استفاده از تابع نرخ مخاطره حاشیه‌ای^۸ برای پیشامدهای بازگردنده و در حضور پیشامد انتها و نیز متغیر کمکی زمان وابسته، تحقیقاتی انجام دادند. جان-ایمرماخر (۲۰۰۸) مدل اندرسون-گیل را با مدل‌های دو جمله‌ای منفی و پواسون برای پیشامد های بازگردنده مقایسه نمود با ذکر این نکته که تنها می توان روشهایی که رویکرد مخاطره حاشیه‌ای را دارند، با هم مقایسه نمود.

در این مقاله یک مدل نیمه پارامتری با نرخ مخاطره جمعی در نظر گرفته شده است که در آن امکان بررسی و برآزش پیشامدهای بدون تکرار وجود دارد، در مدل الگوی تکرار پیشامدهای بازگردنده، یک فرایند نامانای پواسون در نظر گرفته شده است و پیشامدهای انتهایی با استفاده از مخاطره متناسب مدل بندی شده‌اند. در مدل جدید، مخاطره های رقابتی در نظر گرفته می شود و در آن اجازه داده شده است که پیشامدهای بازگردنده و پیشامد انتها به وسیله یک متغیر شکنندگی^۹ وابسته باشند.

بخش دوم این مقاله به معرفی مدل اختصاص داده شده است، در بخش سوم بیشینه سازی تابع درستنمایی و روش برآورد پارامترهای مدل بررسی شده است، در بخش چهارم مطالعه‌ای شبیه سازی ارائه شده و در بخش پنجم داده‌های رد پیوند مغز استخوان بیماران مبتلا به سرطان خون با مدل معرفی شده مدل بندی می شوند.

^۵ Semiparametric model

^۶ Additive rate function

^۷ Intensity function

^۸ Marginal hazard rate

^۹ Frailty variable

۲ معرفی مدل

فرض کنید $N_i(t)$ نشان دهنده تعداد تکرارهای یک پیشامد بازگردنده برای فرد i ام در نمونه مورد مطالعه تا زمان t باشد. همچنین $Y_i = \text{Min}(Y_i^*, C_i, \tau)$ نشان دهنده زمان پایان فرایند برای مشاهده $N_i(t)$ تکرار باشد، که در آن Y_i^* زمان وقوع پیشامد انتها (مثل مرگ یا خرابی دائم) برای فرد i ام و C_i متغیر تصادفی مربوط به سانسور زمان پیشامدها (در آن $C_i = 1$ سانسور شدن را نشان می دهد) و τ حد بالای زمان مشاهده پیشامدها باشد. فرض کنید X بردار متغیر کمکی باشد که مستقل از زمان وقوع پیشامدها است که می تواند سن، جنسیت، وضعیت مالی و سایر متغیرهای پیش بین رگرسیونی باشد. چون متغیر کمکی برای هر فرد مورد مطالعه، منحصر به همان فرد است با نماد x_i نشان داده می شود.

فرض کنید β بردار ضرایب رگرسیونی و $\lambda_0(t)$ تابع پیوسته مخاطره پایه (تابع شدت پایه) است. برای بیان همبستگی بین پیشامدهای بازگردنده، فرایند انتها و بیان تفاوت های فردی بین افراد مورد مطالعه از متغیر شکنندگی Z_i استفاده می شود.

در این مقاله، پیشامدهای بازگردنده که در آنها وقوع هر شکست برای هر تکرار می تواند ناشی از مخاطره های رقیب متفاوتی باشد، مورد بررسی قرار گرفته اند. در حالتی که دو رقیب بر روی تابع مخاطره به صورت هم زمان اثر بگذارند، اثرشان به صورتی فرض شده است که شکست مورد نظر برای فرد i ام تنها به دلیل یکی از مخاطره ها در نظر گرفته می شود. هر کدام از رقبای مخاطره ای، متناسب با نوع و شدت و ذاتشان بر شکست هر کدام از افراد با احتمال خاص خود اثر می گذارند. همچنین وقتی رقیب شماره یک، شروع به تحمیل شکست بر فرد i ام می کند، رقیب دوم از رقابت خارج می شود و وقتی شکست برای فرد مورد نظر رخ داد، هر کدام از رقبا، احتمال متناسب با نوعشان برای شکست های بعدی فرد مورد نظر را همچنان دارا هستند.

احتمال شکست برای هر کدام از افراد، تحت مخاطره k ام عبارت است از

$$\begin{aligned} q_k &= P(K = k) = \lim_{t \rightarrow \infty} \bar{F}_k(t|X) \\ &= \lim_{t \rightarrow \infty} P(T \leq t, K = k|x) \end{aligned}$$

$$= \int_0^t f_k(u|x)du, \quad k = 1, \dots, K, \quad 0 < t < \tau$$

که در آن q_k ها (یعنی احتمال اینکه دلیل وقوع پیشامد، عامل k ام باشد) پارامتر توزیع دو جمله‌ای تعداد پیشامدها در اثر عامل k ام است. برای ساده شدن محاسباتها و نمایش‌های فرمولی، فرض می‌شود تنها دو رقیب مخاطره‌ای وجود دارد و هر کدام از این رقبا برای خود تابع شدتی دارند و از لحاظ آماری از هم مستقل هستند. بنابراین برای $k = 2$ تابع شدت به صورت

$$\lambda_i(t) = \lambda_i(t|x_i, z_i) = (\lambda_{1,0}(t) + x_i^T \beta + z_i)^{\epsilon_i} (\lambda_{2,0}(t) + x_i^T \beta + z_i)^{1-\epsilon_i} \quad (1)$$

خواهد بود، که در آن ϵ_i برابر یک است اگر پیشامد به واسطه مخاطره اول رخ بدهد و برابر صفر است اگر به دلیل مخاطره دوم رخ دهد. فرض کنید m_i تعداد رخداد های بازگردنده تا زمان y_i و t_{i1}, \dots, t_{im_i} زمان مشاهده وقوع پیشامدها برای فرد i ام باشد. در این صورت $N_i(\cdot)$ دارای فرایند پواسون نامانا با نرخ $\lambda_i(t)$ است. بنا به نامانایی فرایند، به جای نرخ مخاطره، می‌توان از عنوان تابع شدت استفاده کرد. در این صورت $N_i(t)$ دارای توزیع پواسون با میانگین $\Lambda_i(t) = \int_0^t \lambda_i(s)ds$ خواهد بود. همچنین اگر $0 \leq a_1 < \dots < a_m \leq \tau$ و $a_j + h_j < a_{j+1}, j = 1, \dots, m$ آنگاه نمودارهای $N_i(a_j + h_j) - N_i(a_j)$ از هم مستقل بوده و دارای فرایند پواسون با میانگین $\Lambda_i(a_j + h_j) - \Lambda_i(a_j) = \int_{a_j}^{a_j+h_j} \lambda_i(s)ds = \lambda_i(a_j)h_j + o(h_j), j = 1, \dots, m$ هستند، که در آن خطای برشی تقریب است به طوری که $\lim_{h \rightarrow 0} o(h)/h = 0$ بنابراین $0 \leq a_{i0} < a_{i1} < \dots < a_{im_i} < a_{im_i+1} = t \leq \tau$ و $N_i(t) = m_i$ آنگاه $a_{ij} + h_j < a_{i,j+1}, j = 1, \dots, m_i$ بنا بر قضیه ۱.۳.۲ داریم:

$$P\{a_{ij} < t_{ij} \leq a_{ij} + h_j, j = 1 \dots m_i | N_i(t) = m_i\} \\ = \frac{(m_i)!}{[\Lambda_{ki}(t)]^{m_i} e^{-\Lambda_{ki}(t)}} \prod_{j=1}^{m_i} [\lambda_{ki}(a_{ij})h_j + o(h_j)]$$

با تقسیم خطاهای برشی بر h_{i1}, \dots, h_{im_i} و قرار دادن $h_i = \max\{h_{i1}, \dots, h_{im_i}\} \rightarrow 0$ و میل دادن عبارت حد به صفر، عبارت خطا را

به صفر رسانده و تابع چگالی توام زمان‌های شکست t_{i1}, \dots, t_{im_i} به شرط $N_i(t) = m_i$ به صورت

$$f(t_{i1}, \dots, t_{im_i} | N_i(t) = m_i) = \frac{(m_i)!}{[\Lambda(t)]^{m_i}} \prod_{j=1}^{m_i} \lambda_i(t_{ij})$$

محاسبه می‌شود. این تابع مستقل از نوع ریسک است. زیرا در شکل فرمولی آن می‌توان به جای تابع مخاطره، هر نوع تابع مخاطره دیگری استفاده نمود. در مدل نیمه پارامتری ارائه شده، سعی شده است که اثر مخاطره رقابتی بر روی پیشامدهای بازگردنده مورد بررسی قرار گرفته و مدلی که در آن پیشامدهای بازگردنده دارای نرخ مخاطره رقابتی باشد، برآزش داده شود. حال تابع درستیابی شرطی برای زمان پیشامدهای تکراری به شرط (x_i, z_i, y_i, m_i) که $m_i \geq 1$ (یعنی بعد از اولین شکست و ورود به مطالعه، هر فرد باید دست کم یک شکست داشته باشد) و دارای مخاطره‌های رقابتی مستقل باشد به صورت

$$L(t_{ij} | z_i, x_i, m_i, y_i) = \frac{(m_i)!}{[\Lambda(t)]^{m_i}} \prod_{j=1}^{m_i} \lambda_i(t_{ij}) \\ = \left(\prod_{j=1}^{m_i} \left[\frac{\lambda_{1,0}(t_{ij}) + x_i^T \beta + z_i}{\Lambda_{1,0}(y_i) + (x_i^T \beta + z_i)y_i} \right] \right)^{\epsilon_i} \left(\prod_{j=1}^{m_i} \left[\frac{\lambda_{2,0}(t_{ij}) + x_i^T \beta + z_i}{\Lambda_{2,0}(y_i) + (x_i^T \beta + z_i)y_i} \right] \right)^{1-\epsilon_i} \quad (2)$$

محاسبه می‌شود. به عبارت دیگر اگر مخاطره اول باعث شکست شده باشد تابع درستیابی برابر با بخش اول عبارت (۲) و اگر مخاطره دوم، عامل شکست در زمان خاص باشد تابع درستیابی برابر با بخش دوم خواهد بود. هر شکست خاص، دارای عامل شکست مخصوص به خود است. بنابراین در تابع درستیابی باید از تابع مخاطره مخصوص به آن شکست، در زمان مورد نظر و در مورد هر فرد استفاده شود.

اگر برای هر کدام از افراد، احتمال تکرار پیشامد به تعداد m_i تکرار محاسبه شود، بنابر پواسون بودن توزیع تعداد تکرارها، می‌توان به شرط ثابت بودن متغیر کمکی و زمان پیشامد انتها و متغیر شکنندگی، احتمال را به صورت

$$P\{N_i(Y_i) = m_i | z_i, x_i, y_i\} = [\Lambda_i(y_i)]^{m_i} \frac{\exp\{-\Lambda_i(y_i)\}}{m_i!} \Lambda_i(y_i)$$

$$\begin{aligned}
 &= \int_0^{y_i} \lambda_i(s) ds \\
 &= [\Lambda_{1,0}(y_i) + (x_i^T \beta + z_i) y_i]^{\epsilon_i} [\Lambda_{2,0}(y_i) + (x_i^T \beta + z_i) y_i]^{1-\epsilon_i}
 \end{aligned}$$

محاسبه نمود. در یک مطالعه طولی برای بررسی بیماری صرع، مشاهدات ثبت شده ممکن است چندین حالت داشته باشند. مثلاً ممکن است زمان ثبت شده مربوط به زمان عود بیماری در یک بیمار یا زمان مرگ یک فرد یا زمان بهبود دائمی و ترخیص فرد و عدم عود در همان شخص باشد. در بین این تکرارها و مشاهدات ثبت شده مربوط به داده‌های پیشامدهای بازگردنده، ممکن است مشاهداتی ثبت شده باشند که بر طبق آن فرد یا افرادی بعد از ورود به مطالعه به دلایلی به جز مرگ مشاهده دیگری از آن‌ها ثبت نشده باشد. به این گونه مشاهدات، مشاهدات ناشی از پیشامد بدون بازگشت گفته می‌شود. این نوع پیشامدها ممکن است ناشی از دو اتفاق باشند: ۱- فرد، بعد از ورود به مطالعه کاملاً بهبود یافته و مشاهده‌ای مبنی بر تکرار بیماری در این فرد وجود ندارد.

۲- بازه زمانی تکرار پیشامد بازگردنده برای آن دسته از افراد بزرگتر از محدوده زمانی مطالعه مورد نظر بوده و اولین تکرار آن بعد از اتمام مطالعه اتفاق افتاده است. فرض کنید p_i احتمال این است که فرد i ام دارای پیشامد تکرار باشد. اگر تمام افراد دارای احتمال پیشامد بازگردنده یکسان باشند، این احتمال به صورت $p = p_i$ نشان داده می‌شود. در این صورت p می‌تواند نسبت تکرار پیشامد بازگردنده در یک جامعه باشد (ژائو و همکاران، ۲۰۱۲). احتمال وقوع پیشامد تکراری برای هر فرد که دارای مخاطره رقابتی مستقل هستند عبارت است از

$$p_i = \left[\frac{\exp\{x_i^T \alpha_1\}}{1 + \exp\{x_i^T \alpha_1\}} \right]^{\epsilon_i} \left[\frac{\exp\{x_i^T \alpha_2\}}{1 + \exp\{x_i^T \alpha_2\}} \right]^{1-\epsilon_i}$$

که برای هر رخدادهای تابعی از متغیرهای کمکی و ضریب رگرسیونی α بوده و تنها عامل تفاوت در احتمال وقوع پیشامد بازگشتی برای هر کدام از مخاطره‌های رقابتی همان ضریب رگرسیونی است. تابع درستیابی برای تمام افرادی که دارای پیشامد بازگردنده هستند برابر است با:

$$L_i = p_i \left(\prod_{j=1}^{m_i} \left[\frac{\lambda_{1,0}(t_{ij}) + x_i^T \beta + z_i}{\Lambda_{1,0}(y_i) + (x_i^T \beta + z_i) y_i} \right] \right)^{\epsilon_i}$$

$$\times \left(\prod_{j=1}^{m_i} \frac{\lambda_{\tau, \circ}(t_{ij}) + x_i^T \beta + z_i}{\Lambda_{\tau, \circ}(y_i) + (x_i^T \beta + z_i) y_i} \right)^{1 - \epsilon_i} \Lambda_i(y_i)$$

اگر پیشامد بدون تکرار رخ داده باشد، $m_i = \circ$ و این اتفاق دارای احتمال وقوع $1 - p_i$ خواهد بود. اما عدم تکرار پیشامد بازگشتی ممکن است به دلیلی چون عدم تکرار پیشامد به طور کلی نیز باشد، در مثال پزشکی فرد مورد بررسی به دلیل بهبود کامل بعد از اولین مراجعه به پزشک، تا زمان τ به مطالعه بازنگشته و بعد از اتمام نیز برنخواهد گشت. احتمال وقوع چنین پیشامدی متفاوت از عبارت احتمالی قبلی بوده و تابعی از خود فرایند پیشامد بازگشتی است. احتمال این که برای فرد خاصی اتفاق مشابهی بیفتد برابر است با:

$$P(N_i(y_i) = \circ) = \exp\{-\Lambda_i(y_i)\}$$

که فرم آن برای انواع مخاطره‌های رقابتی یکسان بوده و تنها تابع مخاطره پایه‌ی جمععی تغییر خواهد کرد. بنابراین تابع درست‌نمایی شرطی برای حالت $m_i = \circ$ عبارت است از

$$\begin{aligned} L &= (1 - p_i) + p_i P(N_i(y_i) = \circ) \\ &= (1 - p_i) + p_i \exp\{-\Lambda_i(y_i)\} \end{aligned}$$

چون مخاطره رقابتی بر عبارت احتمال پیشامد بدون بازگشت تاثیر گذاشته است، بنابراین در فرمول کلی به صورت ظاهری، مخاطره رقابتی دیده نمی‌شود. تابع درست‌نمایی شرطی با در نظر گرفتن پیشامد انتها و پیشامدهای بدون تکرار عبارت است از

$$\begin{aligned} L(t_{ij}|z_i, x_i, Y_i) &= \prod_{i=1}^n L_i(\cdot|z_i, x_i, y_i, m_i = 1)^{\delta_i} L_i(\cdot|z_i, x_i, y_i, m_i = \circ)^{1 - \delta_i} \\ &= \prod_{i=1}^n (p_i \prod_{j=1}^{m_i} \lambda_i(t_{ij}) \exp\{-\Lambda_i(y_i)\})^{\delta_i} (1 - p_i + p_i \exp\{-\Lambda_i(y_i)\})^{1 - \delta_i} \end{aligned}$$

۳ پیشینه‌سازی تابع درست‌نمایی

برای ماکسیمم کردن تابع درست‌نمایی به برآورد تابع مخاطره پایه، متغیر شکنندگی و پارامترهای رگرسیونی نیاز است. معمولاً برای متغیر شکنندگی از توزیع گاما استفاده می‌شود و برای برآورد پارامترهای آن الگوریتم EM به کار گرفته می‌شود. تابع مخاطره پایه به صورت ناپارامتری با استفاده از تابع هسته و تابع درست‌نمایی محلی برازش داده می‌شود. اما برآورد کردن با استفاده از تابع هسته کار پیچیده‌ای است و به حوصله و دقت زیادی احتیاج دارد (ژائو و همکاران، ۲۰۱۲). در این مقاله از تابع درست‌نمایی چندمتغیره محلی برای برازش پارامترها استفاده می‌شود.

۱.۳ برازش تابع مخاطره پایه

برای برآورد تابع نرخ مخاطره پایه ابتدا تابع نرخ مخاطره به صورت

$$\lambda_i(t) = \left[\frac{\lambda_{1,0}(t)}{\lambda_{r,0}(t)} \right]^{\epsilon_i} [\lambda_{r,0}(t)] [x_i^T \beta + z_i]$$

بازنویسی می‌شود و با روش برآورد تابع با سری تیلور برآورد می‌شود. تابع نرخ مخاطره پایه در تابع نرخ مخاطره، برابر با $[\lambda_{r,0}(t)]^{\epsilon_i}$ است و بسط تیلور آن عبارت است از

$$\lambda_0(t|\theta) = \theta_0 + \theta_1 t + \frac{\theta_2 t^2}{2!} + \dots + \frac{\theta_v t^v}{v!} \quad (3)$$

در این صورت تابع نرخ مخاطره تجمعی به صورت

$$\Lambda_0(t|\omega) = \int_0^{y_i} \lambda_0(s) ds = \int_0^{y_i} \left\{ \theta_0 + \theta_1 s + \frac{\theta_2 s^2}{2!} + \dots + \frac{\theta_v s^v}{v!} \right\} ds$$

خواهد شد. برای برآورد تابع نرخ مخاطره پایه که در آن مخاطره‌های رقابتی حضور ندارند، ژائو و همکاران (۲۰۱۲) از معادله (۳) برای برازش تابع مخاطره پایه استفاده کرده و از یک تابع هموارساز برای کاهش میزان خطای برازش بسط تیلور استفاده نموده‌اند. در اینجا برای برازش تابع مخاطره رقابتی پایه از سری تیلور استفاده کرده

و بسط تیلور آن برآزش داده می شود. بنابراین خواهیم داشت:

$$\lambda_0(t_{ij}) \approx \theta_0^\circ(s_0) + \theta_1^\circ(s_0)(t_{ij} - s_0) + \dots + \theta_v^\circ(s_0)(t_{ij} - s_0)^v = \bar{T}_{ij}^T \theta^\circ$$

که در آن $\theta^\circ = (\theta_0^\circ(s_0), \dots, \theta_v^\circ(s_0))^T$ و $\bar{T}_{ij} = (1, (t_{ij} - s_0), \dots, (t_{ij} - s_0)^v)^T$ و $\theta_k^\circ = \lambda_0^{(k)}(s_0)/k!$, $k = 0, 1, \dots, v$ و انتگرال گیری از تابع نرخ مخاطره رقابتی پایه به صورت

$$\Lambda_0(y_i) = \int_0^{y_i} \left\{ \lambda_0(s_0) + \lambda_0'(s_0)(t - s_0) + \dots + \frac{\lambda_0^{(v)}(s_0)(t - s_0)^v}{v!} \right\} dt = \bar{U}_i^T \theta^\circ$$

به دست می آید، که در آن $\bar{U}_i = (u_0^{(i)}, u_1^{(i)}, \dots, u_v^{(i)})^T$ و $u_j^{(i)} = \int_0^{y_i} (t - s_0)^j dt$ و s_0 نقطه پایه انتگرال ریمان است. اگر مجموعه پارامترهای قابل برآورد به صورت $\gamma = (\theta^T, \alpha^T, \beta)$ خلاصه نویسی شود، تابع لگاریتم درست‌نمایی شرطی محلی به صورت

$$\begin{aligned} \ell_i^* &= \delta_i \sum_{j=1}^{m_i} \{ \log[\bar{T}_{ij}^T \theta^\circ + x_i^T \beta + z_i] - [\bar{U}_{ij}^T \theta + (x_i^T \beta + z_i) y_i] \} + \delta_i \log(p_i) \\ &+ (1 - \delta_i) \sum_{j=1}^{m_i} \{ \log[(1 - p_i) + p_i \exp\{-\bar{U}_{ij}^T \theta + (x_i^T \beta + z_i) y_i\}] \} \quad (4) \end{aligned}$$

خواهد شد. که تابع نرخ مخاطره رقابتی پایه به صورت ماتریسی در تابع درست‌نمایی حضور دارد. همچنین تابع درست‌نمایی شرطی شده محلی کلی برابر خواهد بود با:

$$L(s_0, \gamma) = \sum_{i=1}^n \ell_i^*(s_0, \gamma)$$

بنابراین برای برآورد پارامترهای مدل، عبارت

$$\frac{\partial L(s_0, \gamma)}{\partial \gamma} = \sum_{i=1}^n \left\{ \frac{\partial}{\partial \gamma} \log f_i(\gamma) \right\} = 0$$

که تابعی از ℓ_i^* است باید به صورت عددی حل شود.

۲.۳ برآورد پارامترهای مدل

انتخاب درجه بسط تیلور در برآورد کردن تابع مخاطره پایه چالش قابل توجهی در این روش است، از طرفی انتخاب درجه کوچک بسط، امکان نزدیک شدن بسط به

خود تابع اصلی را کم می‌کند و از طرف دیگر، با بزرگ شدن این درجه تعداد پارامترهای مورد نیاز برای برآورد تابع مخاطره پایه افزایش می‌یابد که محاسبات را با دشواری روبه‌رو می‌کند، برای کنار آمدن با این مشکل (چن و همکاران، ۲۰۱۳) پیشنهاد استفاده از اصل نسبت درست‌نمایی را مطرح نمودند به این صورت که با استفاده از توزیع مجانبی منهای دوبرابر لگاریتم نسبت درست‌نمایی که کای دو است، معنی‌داری نسبت درست‌نمایی برای دو درجه متوالی از بسط تیلورآزمون شده و در صورت معنی دار بودن، درجه کوچکتر، برای بسط انتخاب می‌شود، در غیر این صورت دو درجه متوالی بعدی آزمون شده و این کار تا زمان معنی دار شدن اولین نسبت درست‌نمایی ادامه می‌یابد. در این بخش با استفاده از این روش و با برآورده شدن شرط $\chi^2_{df=1,0/95} < -2(l_2 - l_3)$ از بسط تیلور مرتبه دوم استفاده خواهد شد، در این صورت پارامترهای قابل برآورد برای تابع درست‌نمایی شامل $\alpha_1, \alpha_2, \beta, \theta_0^1, \theta_0^2, \theta_1^1, \theta_1^2, \theta_2^1, \theta_2^2$ خواهند بود.

معمولاً با مشتق گرفتن از معادله (۴) و برابر صفر قرار دادن عبارت حاصل از مشتق، پارامترهای مورد نظر برآورد می‌شوند. به دلیل مشخص نبودن توابع نرخ مخاطره پایه نمی‌توان از روش‌های عادی برای برآورد پارامترها استفاده کرد. همچنین اگر توابع مخاطره پایه به صورت مشخص شده در نظر گرفته شوند باز هم به دلیل حل نشدن عبارات حاصل از مشتق، باید پارامترها با روش‌های عددی نیوتون-گوس، نیوتون رافسون یا لونیبرگ-مارگوارت حل شوند. این کار با استفاده از بسته‌های نرم‌افزاری در برنامه R شبیه maxLik و pracma، که از سایت رسمی نرم افزار R قابل دسترس است انجام می‌شود.

۴ مطالعه شبیه‌سازی

فرض کنید تابع مخاطره به فرم (۱) باشد، که در آن تابع مخاطره پایه برای رقیب اول به صورت $\lambda_{1,0}(t) = 2e^{-2t}$ و تابع مخاطره پایه برای رقیب دوم برابر با $\lambda_{2,0}(t) = \eta\omega e^{\eta t} e^{\omega} e^{-\omega e^{\eta t}}$ باشد که در آن $\eta = 2/322$ و $\omega = 1$ است. همچنین متغیر کمکی از توزیع $N(1, 1)$ تولید می‌شود و دارای ضریب رگرسیون با مقادیر

تابع مخاطره برابر با $3/35$ و زمان‌های شکست از توزیع $W(1/5, 2)$ تولید می‌شود. متغیر شکنندگی نیز دارای توزیع $G(2, 1)$ و $\alpha_1, \alpha_2 = 1$ بوده ϵ_i و δ_i از توزیع دو جمله‌ای با احتمال موفقیت 0.5 تبعیت کنند. به ازای $n = 50$ ، تعداد تکرار شکست برای این افراد حداکثر ۵ در نظر گرفته شده است. برای برازش پارامترهای تابع مخاطره، ابتدا توابع مخاطره پایه با بسط تیلور از روی مقادیر مخاطره و زمان‌های وقوع شکست برازش داده شده و نتایج در جدول ۱ آورده شده است. به دلیل طولانی بودن محاسبات و لزوم استفاده چند باره از اصل نسبت درست‌نمایی، درجه بسط سری تیلور ۶ در نظر گرفته و متوسط توان‌های دوم خطای ناشی از برازش محاسبه شده است که برای هر دو تابع مخاطره پایه، عددی بسیار کوچک است. لازم به ذکر است که متوسط توان‌های دوم خطا برای بسط تیلور در 2000 تکرار شبیه‌سازی در نقطه زمانی تولید شده از تفاوت بین مقدار تابع مخاطره اصلی و مقدار بسط تیلور برآورد شده محاسبه شده است. مرتبه‌های باقیمانده این بسط‌ها دارای مقادیر ناچیزی هستند که بر عدم لزوم استفاده از آزمون درست‌نمایی در شبیه‌سازی صحنه می‌گذارد. برازش انجام شده توسط بسته نرم افزاری "pracma" انجام شده است.

جدول ۱: پارامترهای برازش شده برای بسط تیلور با مرتبه ۶ برای توابع مخاطره پایه

مخاطره پایه	θ_0°	θ_1°	θ_2°	θ_3°
$\lambda_{1,0}(t)$	۱/۵۸۳	-۲/۵۲۴	۱/۷۳۸	-۰/۶۵۲
$\lambda_{2,0}(t)$	۰/۰۳۳	-۰/۰۷۶	۰/۰۷۰	-۰/۰۳۳
مخاطره پایه	θ_4°	θ_5°	θ_6°	MSE
$\lambda_{1,0}(t)$	۰/۱۳۸	-۰/۰۱۵	۰/۰۰۱	۰/۰۰۰
$\lambda_{2,0}(t)$	۰/۰۰۸	-۰/۰۰۱	۰/۰۰۰	۰/۰۰۰

در این صورت توابع برازش شده برای توابع مخاطره پایه عبارتند از

$$\lambda_{1,0} = -1/583 - 2/524t + 1/738t^2 - 0/652t^3 + 0/138t^4 + \dots$$

$$\lambda_{2,0} = 0/326 - 0/076t + 0/070t^2 - 0/033t^3 + 0/008t^4 + \dots$$

حال بسط تیلور برآزش شده به جای توابع مخاطره پایه در تابع درست‌نمایی قرار داده می‌شود. در نهایت توسط بسته نرم افزاری "maxLik" پارامترهای رگرسیون برآورد می‌شوند، جدول ۲ نتایج حاصل از ۲۰۰۰ تکرار شبیه‌سازی نمونه‌های ۵۰ تایی که هر کدام حداکثر ۵ شکست داشته‌اند را نشان می‌دهد، همان‌گونه که ملاحظه می‌شود، پارامترهای مدل با میزان آریبی و MSE تجربی بسیار کوچک برآورد شده‌اند که دلالت بر درستی روش و دقت قابل قبول روش‌های برآورد دارد. لازم به ذکر است که شبیه‌سازی برای مقادیر پارامترهای متنوع‌تری انجام شده است اما به دلیل محدودیت فضا از آوردن جداول بیشتر خودداری شده است.

جدول ۲: نتیجه ۲۰۰۰ بار تکرار شبیه‌سازی با نمونه ۵۰ تایی

مقدار واقعی	آریبی برآورد شده	MSE تجربی	
β	۰/۰۱۲	۰/۰۰۰	۲
α_1	۰/۰۴۴	۰/۰۰۰	۲
α_2	۰/۰۰۱	۰/۰۰۰	۱/۵
β	۰/۰۰۲	۰/۰۰۰	۳
α_1	۰/۰۰۰	۰/۰۰۰	۰/۵
α_2	۰/۰۰۲	۰/۰۰۰	۴
β	۰/۰۱۳	۰/۰۰۰	۱۰
α_1	۰/۰۰۵	۰/۰۰۰	۵
α_2	۰/۰۰۱	۰/۰۰۰	۵

۵ مدل‌سازی داده‌های پیوند مغز استخوان

طی دهه گذشته، استفاده از پیوند مغز استخوان به طور وسیعی برای درمان بیماران مبتلا به بیماری‌های خونی گسترش یافته است. یکی از اهداف اصلی پیگیری پس از پیوند، پیش‌بینی عوارض منفی این درمان شامل عود بیماری، رد پیوند و مقاومت علیه پیوند به منظور تعیین درمان پیش‌گیرانه است. در این زمینه ارزیابی کایمریسم مختلط، به عنوان یک روش مهم در پیگیری نتایج پس از پیوند گزارش شده است.

اصطلاح کایمریسم^{۱۰} به عنوان حضور سلول‌های خون‌ساز با منشا غیر میزبان تلقی می‌شود. کایمریسم کامل به معنی جایگزینی کامل خون‌سازی گیرنده، توسط سلول‌های خون‌ساز دهنده و کایمریسم مختلط به معنی حضور هم‌زمان سلول‌های خون‌ساز دهنده و گیرنده می‌باشد. افزایش تعداد سلول‌های گیرنده پس از پیوند، به صورت کاملاً محسوس به نفع خطر عود بیماری است و به همین خاطر تشخیص زودرس، جنبه مهمی از پیش‌آگهی است. بیمارانی که دچار عود می‌شوند، می‌توانند با تزریق انفوسیت دهنده وارد رمیسیون (فاز تلاش مجدد برای پیوند) شوند. تحلیل کایمریسم روش بسیار ارزشمندی برای ارزیابی رژیم‌های آماده‌سازی مختلف، مقاومت علیه پیوند (GVHD)^{۱۱}، رژیم پیش‌گیرانه و سلول درمانی برای افزایش موفقیت پیوند دارد.

در این ارتباط اطلاعات مورد نیاز مربوط به ۸۰ بیمار شناسایی شده، که دارای تشخیص اولیه لوسمی میلوئیدی حاد (AML)^{۱۲} و لوسمی حاد (ALL)^{۱۳} بودند، ثبت و دسته‌بندی شده‌اند. بیماران مربوطه، به دلایلی مانند سردرد و بی‌حوصلگی و خواب‌آلودگی و کسلی و منگی به بیمارستان مراجعه نموده و بعد از تشخیص بیماری، آزمایش‌ها و درمان‌های مورد نظر انجام شده است.

اطلاعات مربوط به زمان شکست، مربوط به بیمارانی که بیشتر از ۳ بازگشت داشتند، بعد از سومین عود بیماری سانسور شده است (به دلایل پزشکی). بعضی از بیماران نیز در طول مطالعه به دلایلی معلوم یا نامعلوم از مطالعه کنار گذاشته شده و اطلاعات مربوط به زمان شکست آنها سانسور شده‌اند، در نهایت جمعاً ۵۶ مشاهده و ۲۴ مورد سانسور شده در داده‌ها باقی مانده است. مطالعه مربوط به دوره زمانی یک ماهه بعد از پیوند بوده و محدوده سنی بیماران ۳ تا ۵۰ سال با میانه سنی ۲۳ سال است. در آن ارتباط کایمریسم مختلط با بیماری پیوند علیه میزبان حاد، عود بیماری، بقا و بقای بدون عود در بیماران دریافت‌کننده از دو روش مغز استخوان و خون محیطی بررسی شده است پیش‌بینی تابع بقا و تابع احتمال بقای افراد، بعد از

^{۱۰} Chimerism

^{۱۱} Graft versus host disease

^{۱۲} Acute Myeloid Leukemia

^{۱۳} Acute Lymphocytic Leukemia

پیوند و احتمال بازگشت بیماری و پیش بینی تابع مخاطره از مهمترین اهدافی است که در اینجا به آن پرداخته می شود.

از ۸۰ بیمار انتخاب شده ۳۳ نفر مرد و ۴۷ نفر زن بودند. ۳۶ نفر از بیماران دارای تشخیص اولیه ALL و ۴۴ نفر دارای تشخیص اولیه AML بودند. ۲۷ نفر با استفاده از رژیم بوسولفان و سیکلوفسفاماید، ۵۰ نفر با رژیم بوسولفان و فلودارابین و ۳ نفر با رژیم بوسولفان و فلودارابین و آنتی تیموسیت گلوبین آماده سازی شده بودند. همچنین ۳۵ نفر دارای GVHD درجه یک، ۱۵ نفر درجه دو، ۲۴ نفر درجه ۳ و ۶ نفر به GVHD درجه ۴ دچار شده بودند.

در ادامه احتمال های مربوط به شکست ناشی از هر یک از مخاطره های AML و ALL بررسی می شوند. با توجه به برنامه نوشته شده توسط اسکروجا و همکاران (۲۰۰۷)، برای محاسبه احتمال وقوع شکست بر اساس هر کدام از مخاطره ها به روش گرای (۱۹۸۸)، خروجی های زیر حاصل شدند. بیماری ALL با حداکثر احتمال ۰/۳۲۵، انحراف معیار ۰/۰۳۶ در بازه اطمینان ۹۵ درصد (۰/۲۵۶ و ۰/۳۹۷) موجب وقوع شکست در افراد شده و بیماری AML با حداکثر احتمال ۰/۴۲۷، انحراف معیار ۰/۰۳۸ در بازه اطمینان ۹۵ درصد (۰/۳۵۲ و ۰/۴۹۹) موجب وقوع شکست در افراد می شود. بیماری ALL در زمان $t = ۴۲/۵$ واحد به همگرایی رسیده و بیماری AML در زمان $t = ۲۹$ واحد به همگرایی می رسد. بنابراین حداکثر مقدار خود را در این واحدهای زمانی دارا هستند. خروجی نشان دهنده تاثیر بیشتر بیماری AML در شکست افراد نسبت به بیماری ALL است. نتایج حاصل از برازش مدل ۱ به داده ها در جدول ۳ خلاصه شده اند. لازم به ذکر است که با استفاده از رابطه یک به یک بین تابع مخاطره و تابع بقا می توان به سادگی با استفاده از تابع مخاطره برآورد شده تابع بقا را محاسبه نمود. با استفاده از رابطه بین تابع مخاطره و تابع بقا می توان به سادگی اطلاعات بقای بیماران را نیز پیش بینی نمود، نتایج این کار در جدول ۴ خلاصه شده است.

همان طور که ملاحظه می شود GVHD و جنسیت در زمان مخاطره رد پیوند و به تبع در تابع بقا به صورت معنی داری موثر هستند و سن و نوع رژیم پیشنهادی تاثیر معنی داری بر تابع بقا ندارد. در خصوص متغیر تصادفی شکنندگی که در مدل لحاظ

جدول ۳: پارامترهای برازش شده برای تابع بقای مدل

متغیر	ضرایب	انحراف معیار	z	$P(0 > z)$
جنسیت	۰/۷۳۹	۰/۲۴۹	۲/۸۰۸	۰/۰۰۵
سن	۰/۰۰۳	۰/۰۱۰	۰/۱۸۱	۰/۸۵۶
رژیم	-۰/۴۴۶	۰/۲۱۱	-۱/۹۱۶	۰/۰۵۵
GVHD	-۰/۴۴۰	۰/۱۳۹	-۲/۸۶۴	۰/۰۰۴

جدول ۴: پارامترهای برازش شده برای تابع مخاطره مدل

متغیر	$\exp(-)$ ضرایب	کران پایین	کران بالا
جنسیت	۰/۴۷۷	۱/۲۵۰	۳/۵۷۱
سن	۰/۹۹۸	۰/۹۷۸	۱/۰۲۷
رژیم	۱/۵۶۳	۰/۴۰۵	۱/۰۱۰
GVHD	۱/۵۵۳	۰/۴۷۷	۰/۸۷۰

شده بود که در واقع معادل با اجازه دادن به وجود تغییراتی متناسب با واریانس متغیر شکنندگی در تابع مخاطره است که بتواند نوسانات تصادفی ناشی از نادیده گرفتن متغیرهای توصیفی ناشناخته را جبران نماید، وجود این متغیر در مدل معنی دار نشد که نشان‌دهنده عدم تاثیر متغیرهای توصیفی ناشناخته و تفاوت‌های فردی خاص بر تابع بقا در افراد مورد بررسی است.

بحث و نتیجه‌گیری

در این مقاله یک مدل نیمه پارامتری جمعی تابع شدت، حاصل از پیشامدهای بازگردنده تحت مخاطره رقابتی همراه با فرایند انتها در نظر گرفته شد. با استفاده از فرایند شمارشی پواسون ناهمگن پیشامدهای تکراری برای هر فرد که دارای متغیرهای توصیفی مخصوص به خود می‌باشد را به تفکیک نوع مخاطره می‌شمارد، این مدل قابلیت تحلیل داده‌هایی با سانسور تصادفی و سانسور آگاهنده را دارا است و افرادی که دارای پیشامد بازگشتی نبودند نیز در مطالعه وارد شده‌اند. در این مطالعه اثر مخاطره رقابتی وارد تحلیل‌های مربوط به پیشامدهای بازگردنده شده

است. اضافه کردن نوع مخاطره به تحلیل پیشامدهای بازگردنده باعث می شود وقوع پیشامد تکراری با در نظر گرفتن نوع بیماری بررسی شود. این مطلب سبب می شود تشخیص عوامل و متعیرهای توصیفی که در وقوع پیشامد در هر یک از دو نوع بیماری تاثیرگذار است امکان پذیر شود و به صورت مفهومی افزایش دقت در برآورد پارامترها را در پی خواهد داشت و در نتیجه امکان پیش بینی دقیق تر را فراهم می کند. نتایج مطالعه شبیه سازی به لحاظ کوچک بودن آریبی ها و متوسط توان های دوم خطا مؤید این امر هستند، در این مدل برآوردها با آریبی و MSE تجربی بسیار کوچک بدست آمدند که نشان دهنده دقت قابل قبول برآوردها است. استفاده این مدل برای تحلیل داده های پیوند مغز استخوان برای بیماران مبتلا به سرطان خون قابلیت کاربرد مدل را در داده های حاصل از مطالعات هم گروهی و آینده نگر با ماهیت پیچیده (به دلیل وجود سانسور، پیشامد بازگردنده و پیشامد انتها) مشخص می نماید اگر چه کم بودن تعداد افراد مورد مطالعه در این داده ها سبب شد تاثیر نوع رژیم آماده سازی معنی دار ارزیابی نشود. در صورتی که بتوان اطلاعات جامع تری به همراه متغیرهای توصیفی بیشتر و احتمالاً رژیم های آماده سازی جدیدتر، از این گونه بیماران در سطح کشور به دست آورد، می توان با استفاده از مدل پیشنهادی در این مقاله به نتایج مفیدتری نیز دست یافت.

تقدیر و تشکر

مؤلفین کمال تشکر و قدردانی را از داوران محترم و هیئت تحریریه مجله وزین علوم آماری که سبب ارتقا و روان تر شدن مقاله شد ابراز می دارند.

مراجع

Anderson, P. K. and Gill, R. D. (1982), Cox Regression Models for Counting Processes: a Large Study, *Annals of Statistics*, **10**, 1100-1120.

Chen, D. G., Yu, L., Peace, K. E., Lio, Y. L. and Wang, Y. (2013), Approximating the Baseline Hazard Function by Taylor Series for Interval

- Censored Time-to-Event Data, *Journal of Biopharmaceutical Statistics*, **23**, 695-708.
- Gray, R. (1988), A Class of K-Sample Test for Comparing the Cumulative Incidence of a Competing Risk, *Annals of Statistics*, **16**, 695-708.
- Liu, L., Wolfe, R. A. and Huang, X. (2004), Shared Frailty Model for Recurrent Events and a Terminal Event, *Biometrics*, **60**, 747-756.
- Pepe, M. S. and Cai, J. (1993), Some Graphical Display and Marginal Regression Analyses for Recurrent Times and Dependent Covariates, *Journal of American Statistical Association*, **88**, 811-820.
- Prentice, R. I., Williams, B. J. and Peterson, A. V. (1981), On the Regression Analysis of Multivariate Failure Time Data, *Biometrika*, **68**, 79-373.
- Ross, S. M. (1996), *Stochastic Processes*, 2nd Edt., Wiley, New York.
- Scrucca, L., Santucci, A. and Aversa, F. (2007), Competing Risk Analysis Using R: An Easy Guide for Clinicians, *Bone Marrow Transplantation*, **40**, 381-387.
- Wei, L. G., Lin, D. Y. and Weissfeld, L. (1989), Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distribution, *Journal of American Statistical Association*, **84**, 1065-1073.
- Zhao, X. B., Zhou, X. and Wang, J. L. (2012), Semiparametric Model for Recurrent Event Data with Excess Zeros and Informative Censoring, *Journal of Statistical Planning and Inference*, **142**, 289-300.