



تحلیل مدت زمان اجرای رویدادها مبتنی بر زنجیره مارکوف و مدل نیمه مارکوف

سیما حسنوند

کارشناسی ارشد کامپیوتر گرایش نرم افزار، دانشگاه آزاد اسلامی واحد علوم و تحقیقات خراسان رضوی
Sima.hasanvand@yahoo.com

لقمان رشیدی

کارشناسی ارشد کامپیوتر گرایش نرم افزار، دانشگاه بین المللی امام رضا
Loghman.rashidi@yahoo.com

سونیا شیخی بیات

کارشناسی ارشد کامپیوتر گرایش نرم افزار، دانشگاه آزاد اسلامی واحد علوم و تحقیقات خراسان رضوی
Sonia_sheykhi@yahoo.com

چکیده

انواع روش های آماری در دسترس محققان برای تجزیه و تحلیل مدت زمان رویداد داده ها، اغلب به عنوان بررسی حفظ بقاء اشاره شده است. برآورد Kaplan-Meier و Cox hazard regression معمولاً ابزاری مناسب برای تمام مطالعات نیست. مدل زنجیره مارکوف می تواند اطلاعات سانسور شده، خطرات رقابتی (سانسور آموزنده)، نتایج متعدد، نتایج مکرر، سستی و احتمالات غیرثابت بقاء را در خود جای دهد. مدل زنجیره مارکوف، هر چند اغلب توسط محققان در آنالیز مدت زمان رویدادها نادیده گرفته شده، اما مدت طولانی است که در مطالعات بالینی استفاده می شود و دارای کاربرد گسترده ای در زمینه های دیگر است.

واژگان کلیدی: زنجیره مارکوف، فرایند مارکوف، مدل نیمه مارکوف، Cox hazard regression، Kaplan-Meier



مقدمه

تجزیه و تحلیل داده های زمان به رویداد در مطالعات انسانی و حیوانی چالش های چند آماری را نشان می دهد. علاوه بر مشکلات و مشاهدات سانسور شده، ممکن است انواع مختلفی از شکست ها هم تحت بررسی قرار بگیرد ("مشکل خطر رقابتی") (Abner EL, Kryscio RJ, Cooper GE, Fardo DW, Jicha GA, et al. 2012). نتایج بالینی دیگری از شکست ممکن است در طول پیگیری مشاهده شود، از جمله آنهایی که خطر ابتلا به نارسایی را تغییر می دهند و یا می توانند بیش از یکبار رخ دهند، و اینکه حساسیت فرد به شکست (به عنوان مثال سستی) امکان دارد در طول زمان ثابت باشد (Kryscio RJ, Abner EL, Lin Y, Cooper GE, Fardo DW, et al. 2013). در حالی که مدت زمان رویداد روش تجزیه و تحلیل معمول مانند Kaplan-Meier برآورد محصول - حد و Cox hazard regression می تواند براحتی اجرا و از داده های سانسور شده مؤثر که فرضیه سانسور را بی ارزش نگه می دارد استفاده شود. تجزیه و تحلیل مربوط به سانسور اطلاعات، نتایج متعدد و یا احتمال غیر ثابت بقاء به خوبی از فرآیندهای مارکوف مناسب استفاده کند (Crowder M, 2012). رویکرد معاصر در مشکل سانسور شده اطلاعات در Cox hazard regression شامل تجزیه و تحلیل بقاء چند متغیره است.

فرآیند مارکوف

فرآیند مارکوف فرآیندی تصادفی است که حرکت یک فرد را از طریق تعداد متناهی از حالات تعریف کند، یکی (و تنها یکی) است که می بایست هر فرد در هر زمان خاص آن را توصیف کند. حرکات ممکن است در میان حالتها با یک ماتریس انتقال و یا نمودار حالت به تصویر کشیده شود (Crowder M, 2012).

به منظور فسخ فرآیند، حداقل یکی از حالات باید جذب شود، به عنوان مثال، افراد صفر احتمال ترک حالتی که یکبار به آن وارد شده است (Salazar JC, Schmitt FA, Yu L, Mendiondo MS, Kryscio RJ, 2007). مرگ، به عنوان مثال، حالت جذب معمولاً در مطالعات بالینی استفاده می شود، اما آن نیز در معرض خطر رقابتی شناخته شده و برای نتایج بالینی در مطالعات افراد مسن تر می باشد. فرآیندهای مارکوف به طور مداوم و یا گسسته و همچنین دارای زمان همگن و یا زمان ناهمگن باشد. تمرکز این مقاله به طور گسسته می باشد، زیرا فرآیندهای زمان همگن مارکوف به نام زنجیره مارکوف می باشد

زنجیره مارکوف

مدل زنجیره مارکوف به تحلیلگران برای محاسبه و میزان (یا شدت) حرکت در ارتباط با هر یک از انتقالات بین حالات در یک سیکل تک اجازه می دهد و همچنین تعداد تقریبی چرخه در حالت خاصی به سر می برد. هنگامی که مشاهدات در فواصل منظم ساخته می شود، تعداد چرخه می تواند به عنوان زمانی در حالت، تفسیر شده باشد. زمان صرف شده در تمام کشورهای قبل از جذب را می توان به منظور برآورد زمان زنده ماندن خلاصه نمود.



استفاده از زنجیره مارکوف نیاز به دو فرضیه اساسی دارد:

- احتمال انتقال ثابت در طول زمان (زمان همگن)
- احتمال انتقال بعدی تنها به حالت فعلی (به مرتبه اول خاصیت مارکوف) بستگی دارد.

این مدل برای تجزیه و تحلیل زمان رویدادها می باشد. تجزیه و تحلیل همزمان چندین وقایع موردنظر و رقابت های خطری را از طریق حالات تعریف شده در مدل می گنجانند، و همچنین ضعف های فردی را از طریق اثرات تصادفی موضوعات خاص در نظر می گیرد. اطلاعات سانسور شده، در دو سمت چپ و راست برای زنجیره مارکوف مناسب است. از آنجایی که افراد نیاز ندارند برای ورود به ماتریس انتقال در هر حالت خاص، داده های سمت چپ سانسور شده را جایگزین کنند. سانسور فاصله ای رسمی در زنجیره مارکوف می باشد، که فرض می کنیم عمل انتقال تنها یکبار در هر سیکل، یا در ابتدا یا در انتهای آن جایگزین نیست (Salazar JC, Schmitt FA, Yu L, Mendiondo MS, Kryscio RJ, 2007). در واقع انتقالات مطمئن در هر زمانی در نظر گرفته می شوند، و انتقالات مشاهده نشده ممکن است که در بین چرخه های ارزیابی باشد. روشی مانند اصلاح نیمه چرخه، که در آن فرضیه انتقالات در وسط سیکل رخ می دهد، و برای کاهش تعصب ناشی از فرض اینکه انتقال را تنها در آغاز چرخه یا پایان آن پیشنهاد شده است.

اگر مدل بالینی و ساختار داده ها پشتیبانی شود، این فرضیه را به وجود می آورد که تمام انتقال ها یک طرفه است (یعنی، هیچ انتقال معکوسی ممکن نیست)، یک مدل نیمه مارکوف است، که یک مورد از زنجیره مارکوف که در آن زمان سپری شده در حال حاضر به حالت آن در هر دو حالت مجاور قبل و آینده بستگی دارد، و می توان برای داده های سانسور شده بازه ای را در نظر گرفت (Song C, Kuo L, Derby CA, Lipton RB, Hall CB, 2011).

در نهایت، بر خلاف تجزیه و تحلیل زمان های سنتی رویداد که در آن تنها یک نتیجه برای هر فرد ممکن است، زنجیره مارکوف به تحلیلگران برای محاسبه زمان زنده ماندن در حالات متعدد این اجازه را می دهد. این امر به ویژه برای مطالعات بیماری های مزمن که با مراحل آنرا به خوبی تعریف شده، مانند سرطان و بیماری های خود ایمن، که در آن بهبودی و عود کمردرد مورد توجه بقای کلی هستند، و زوال عقل به علت بیماری های عصبی، که در آن حالات بیماری های پیش بالینی و خفیف و علامت آن به طور فزاینده مورد توجه پژوهشگران برای شناسایی درمان و استراتژی های پیشگیرانه آن می باشد. مانند تجزیه و تحلیل سنتی زمان که منحنی بقا آن ممکن است از نتایج مدل برآورد شود. میانگین زمان بقا ممکن با استفاده از راه حل ماتریس، شبیه سازی Covariate-adjusted regression، یا شبیه سازی زنجیره ای مونت کارلو نتیجه گرفته شود. این محاسبات دست و پاگیرتر است، اما هنوز هم ممکن که طبق برآوردهای احتمال انتقال از مدل رگرسیون همگام تنظیم و مشتق گرفته شود. در مقابل، مدل نیمه مارکوف برآورد معنی زمان بقا به طور مستقیم و بدون نیاز به محاسبات اضافی است (Pan F, Goh JW, Cutter G, Su W, Pleimes D, et al., 2012).



مدل متفاوتی از فرضیات

تأیید فرضیات مدل زمان همگن را می توان با استفاده از آزمون نسبت احتمال ارزیابی، و مرتبه اول مارکوف فرضیه اموال را می توان با آزمون χ^2 مورد بررسی قرار داد.

فرضیه زمان همگن اغلب برای پاسخگویی دشوار است، به ویژه در مطالعات بیماری های مزمن که در آن مطالعات سال های طولانی، چرخه مشاهده تنها می تواند یک سال یا بیشتر از آن را پوشش دهد، و افزایش سن به طور کلی بیشتر مربوط به خطر ابتلا به بیماری یا مرگ می باشد (Yu L, Griffith WS, Tyas SL, Snowdon DA, Kryscio RJ, 2010).

با این حال، این نگرانی را می توان با طبقه بندی داده ها (به عنوان مثال، با گروه سنی و یا دوره مطالعه) و یا مدل رگرسیون، که در آن اثر متغیرهای کمکی در برآورد احتمال انتقال کاهش می یابد، در نظر گرفت. در رگرسیون، متغیرهای کمکی ممکن یا ثابت و یا وابسته به زمان باشند. حتی هنگامی که فرضیه های اساسی مدل برآورده شده می باشند، استفاده از مدل زنجیره مارکوف هنوز هم امکان دارد ناموفق باشد (van Raalte AA, Caswell H, 2013).

چگالی داده ها، به عنوان مثال، فرکانس مشاهده هر نوع انتقال، ممکن است بیش از حد در برخی از سلول های پیاده سازی مدل رگرسیون پراکنده باشد. سلول های پراکنده، که در آن چند حادثه مشاهده شده، احتمال دارد به برآورد نادرست و یا منجر به جلوگیری از همگرایی مدل شود (Hochreiter R, Wozabal D, 2010).

بحث و نتیجه گیری

اگرچه مدل مارکوف در کاربردهای بالینی برای بیشتر از 60 سال استفاده شده، اما ترکیب اثرات تصادفی موضوعات خاص در زنجیره مارکوف برای محاسبه تمایلات فرد به انتقالات توسعه نسبتاً تازه است. با این حال، گنجانیدن اثرات تصادفی در برآورد احتمال بسیار پیچیده است، و اتصالات مدل مانند می تواند وقت گیر باشد. از همه مهمتر باید مفهوم خود را به دقت در نظر بگیرد. مدل هایی که از حالات تونل استفاده می کنند (به عنوان مثال، حالاتی که در آن ها معکوس انتقالات ممکن نیست و از اثرات تصادفی پیچیده استفاده می کنند).

در پایان، زنجیره مارکوف ابزاری مفید برای آنالیز بقا که به مدل سازی های متنوع تر این اجازه را می دهد که در روش های استاندارد زمان رویدادها در دسترس باشد. در حالی که تمرکز مقاله بر مطالعات بالینی بوده است، زنجیره مارکوف کاربردهای واضح و روشنی در زمینه های مختلف از جمله پژوهش کار، امور مالی، علوم سیاسی، مهندسی شیمی و جمعیت شناسی دارد. با این حال، در حالی که بسیاری از خوانندگان مجله و داوران به آسانی نتایجی از مدل مارکوف را درک کنند، آنها ممکن است با مفروضات در زمینه آماری نداشته باشند، به ویژه در زمینه هایی که در آن استفاده از مدل مارکوف هنوز گسترده نیست. اگر چنین باشد، آنها ممکن چالش محققان را نادیده بگیرند و نشان بدهند که این فرضیات قابل دفاع هستند. با توجه به اینکه استفاده نادرست از مدل مارکوف ممکن نتایج متعصبانه ای را در بر داشته باشد، و شاید برخی از استانداردها در گزارش نتایج مدل مارکوف و تأیید فرضیات آن مورد نیاز باشد.



منابع

- Abner EL, Kryscio RJ, Cooper GE, Fardo DW, Jicha GA, et al. (2012) Mild Cognitive Impairment: Statistical models of transition using longitudinal clinical data. *Int J Alzheimers Dis*. 2012: 291920.
- Kryscio RJ, Abner EL, Lin Y, Cooper GE, Fardo DW, et al. (2013) Adjusting for mortality when identifying risk factors for transitions to MCI and dementia. *J Alzheimers Dis* 35: 823-832.
- Crowder M (2012) *Multivariate survival analysis and competing risks*. CRC Press, Taylor and Francis Group: Boca Raton, Florida.
- Salazar JC, Schmitt FA, Yu L, Mendiondo MS, Kryscio RJ (2007) Shared random effects analysis of multi-state Markov models: application to a longitudinal study of transitions to dementia. *Statist Med* 26: 568-580.
- Song C, Kuo L, Derby CA, Lipton RB, Hall CB (2011) Multi-stage transitional models with random effects and their application to the Einstein Aging Study. *Biometrical J* 53: 938-955.
- Pan F, Goh JW, Cutter G, Su W, Pleimes D, et al. (2012) Long-term cost-effectiveness model of interferon beta-1b in the early treatment of multiple sclerosis in the United States. *Clin Ther* 34: 1966-1976.
- Yu L, Griffith WS, Tyas SL, Snowdon DA, Kryscio RJ (2010) A nonstationary Markov transition model for computing the relative risk of dementia before death. *Statist Med* 29: 639-648.
- Hochreiter R, Wozabal D (2010) Evolutionary estimation of a coupled Markov chain credit risk model. *Natural Computing in Computational Finance* 293: 31-44.
- van Raalte AA, Caswell H (2013) Perturbation analysis of indices of lifespan variability. *Demography* 50: 1615-1640.