تحلیل داده‌های بقا تحت سانسور آگاهنده

اعظم راستین
گروه آمار، دانشگاه شهید بهشتی تهران

چکیده

آنچه تحلیل بقا را از سایر تحلیل‌های آماری متمایز می‌کند پدیده‌ی سانسور است و هنگامی رخ می‌دهد که در جریان مطالعه، برخی از افراد پیشامد مورد نظر را تا پایان زمان پیگیری تجربه نکنند. سانسور به دلایل گوناگون و در مکانیزم‌های مختلفی می‌تواند رخ دهد و در همه‌ی آنها فرض می‌شود که سانسور درباره پیشامد ناآگاهنده است، بدین معنا که زمان رخداد تا پیشامد به زمان سانسور مرتبط نیست. اما اغلب در کاربردهای عملی، سانسور درباره‌ی پیشامد مورد نظر آگاهی‌بخش است که متأسفانه این می‌تواند منجر به یک اریبی در نتایج تحلیل بقا شود. در این مقاله به بررسی کلی پدیده‌ی سانسور و برآورد تابع بقا تحت سانسور آگاهنده خواهیم پرداخت.

واژه‌های کلیدی: تحلیل بقا، تابع بقا، سانسور آگاهنده، مدل KG، تابع مفصل.

۱ مقدمه

تحلیل بقا از نظر علم آمار مجموعه‌ای از روش‌های مختلف آماری در تحلیل متغیرهای تصادفی نامنفی است که مقدار آن فاصله‌ی زمانی ورود به مطالعه تا وقوع حادثه‌ی مورد نظر را شامل می‌شود. از جمله مواردی که می‌تواند مصداق شکست یا پیشامد مورد نظر باشد، طول عمر يك ماشين صنعتی، زمان عود یک بیماری، مرگ و مثال‌هایی از این قبیل است. در تحلیل بقا، اغلب با پدیده‌ای به نام سانسور مواجه هستیم. سانسور ویژگی خاص داده‌های بقا به ویژه در علوم زیستی است و هنگامی رخ می‌دهد که در جریان مطالعه، برخی از افراد پیشامد مورد نظر را تا پایان زمان پیگیری تجربه نکنند. برای مثال اگر پیشامد مورد نظر مرگ باشد، بعضی از بیماران ممکن است در پایان دوره‌ی پیگیری هنوز زنده باشند. گاهی نیز در یک مطالعه افرادی قبل از اتمام دوره‌ی پیگیری، مفقود می‌شوند. در این صورت زمان‌های بقای این افراد نامعلوم بوده و یا به اصطلاح زمان‌های بقای آنها سانسور شده است.

از طرفی دیگر، یک فرض اساسی و متداول در بسیاری از تحلیل‌های بقا، ناآگاهندگی سانسور است، بدین معنا که سانسور درباره‌ی پیشامد مورد نظر آگاهی‌بخش نیست و احتمال سانسور شدن به احتمال رخداد پیشامد مورد علاقه ارتباطی ندارد. اما این فرض اغلب در کاربردهای عملی برقرار نیست و با فرض کلی وابستگی بین تابع توزیع زمان بقا و زمان سانسور مواجهیم. تا کنون روش‌های زیادی برای مدل‌بندی الگوی وابستگی بین زمان پیشامد

و زمان سانسور مطرح شده است که در این مقاله به معرفی برخی از این روش‌ها تحت کلاس مفصل‌ها خواهیم پرداخت. مفصل‌ها برای اولین بار توسط اسکلار [۱۳] به معنی به هم پیوستن و پیوند زدن به کار رفت زیرا این تابع عملاً بین توزیع‌های حاشیه‌ای و توأم نوعی پیوند یا ارتباط برقرار می‌کند. این ارتباط در قضیه‌ی معروف او موسوم به قضیه‌ی اسکلار ارائه شده است. در نتیجه تحت یک تابع مفصل معلوم، مدل‌بندی الگوی وابستگی بین زمان پیشامد و زمان سانسور امکان پذیر خواهد شد که این موضوع در این مقاله مورد مطالعه قرار می‌گیرد.

به طور کلی هدف این مقاله بحث پیرامون سانسور آگاهنده است و نیز مسائلی که این پدیده در برآورد تابع بقا ایجاد می‌کند. در ادامه‌ی این مقاله ابتدا در بخش ۲ تعریف سانسور و انواعی از آن را مطرح خواهیم کرد، در بخش ۳ به معرفی سانسور آگاهنده و سانسور ناآگاهنده خواهیم پرداخت و در بخش ۳ برآورد تابع بقا تحت سانسور آگاهنده را مورد مطالعه قرار می‌دهیم.

۲ سانسور در تحلیل بقا

تحلیل بقا بر داده‌هایی متمرکز است که نشان‌دهنده‌ی فاصله‌ی زمانی بین شروع پیگیری تا وقوع یک پیشامد معین هستند. این فاصله‌ی زمانی را با T نمایش می‌دهند و به آن زمان بقا می‌گویند [۸]. اما آنچه تحلیل بقا را از سایر تحلیل‌های آماری متمایز می‌کند پدیده‌ی سانسور است که به طور کلی به سه دلیل می‌تواند رخ دهد،

۱. شخص پیشامد را تا پایان دوره‌ی پیگیری تجربه نکند.

۲. شخص قبل از اتمام دوره‌ی پیگیری، مفقود شود.

۳. شخص از دوره‌ی پیگیری انصراف دهد.

هنگامی که مشاهده‌ی سانسورشده‌ای وجود ندارد، مجموعه‌ی زمان‌های بقا را کامل نامند.

در ادامه‌ی این مقاله T را زمان بقای واقعی (غیرقابل مشاهده)، C را زمان سانسور، $\delta = I(T \leq C)$ را نشانگر سانسور و $Y = min(T, C)$ را زمان بقای مشاهده شده در نظر گرفته‌ایم.

۱.۲ انواع سانسور

سانسور در اشکال مختلف و به دلایل گوناگونی می تواند اتفاق افتد،

۱. سانسور از راست: اگر زمان بقا از زمان ورود فرد به مطالعه بیشتر باشد یا به عبارتی هرگاه زمان سانسور یک مشاهده پس از ورود به مطالعه اتفاق افتد آن مشاهده را سانسور از راست گویند.

به علت سانسور راست تنها Y_i و δ_i قابل مشاهده هستند که در آن $Y_i = \min(T_i, C_i)$ و داشته باشیم،

$$\delta_i = \begin{cases} ۰, & T_i \leq C_i, \\ ۱, & T_i > C_i. \end{cases}$$

سانسور راست متداول‌ترین نوع سانسور در تحلیل بقا است. برای طرح سانسور راست، سه نوع متفاوت در نظر گرفته می‌شود،

(آ) سانسور نوع I: در این نوع سانسور، زمان‌های سانسور (C_i) تمام مشاهدات یکسان بوده و زمان‌های سانسور شده برابر با طول دوره‌ی پیگیری است.

(ب) سانسور نوع II: در این نوع سانسور $C_i = T_r$ زمان r-امین شکست بوده و زمان مشاهدات سانسور شده برابر با زمان بزرگترین مشاهده‌ی سانسور نشده است.

(ج) سانسور نوع III یا تصادفی: در این نوع سانسور، C_iها متغیرهای تصادفی، مستقل و هم‌توزیع با توزیعی مشترک هستند.

سانسور تصادفی یکی از متداول‌ترین طرح‌های سانسور از راست است. در واقع سانسور تصادفی به مواردی گفته می‌شود که زمان توقف مطالعه برای همه‌ی افراد نمونه یکسان و از قبل مشخص شده است ولی زمان ورود آنها به مطالعه متفاوت و تصادفی است. در بیشتر مطالعات همه‌گیر شناسی [1] و پزشکی دوره‌ی پیگیری ثابت بوده و در طول این دوره، بیماران در زمان‌های متفاوت وارد دوره‌ی پیگیری می‌شوند. از آنجایی که زمان‌های ورودی یکسان نیستند، زمان‌های سانسور شده نیز متفاوت‌اند [۱].

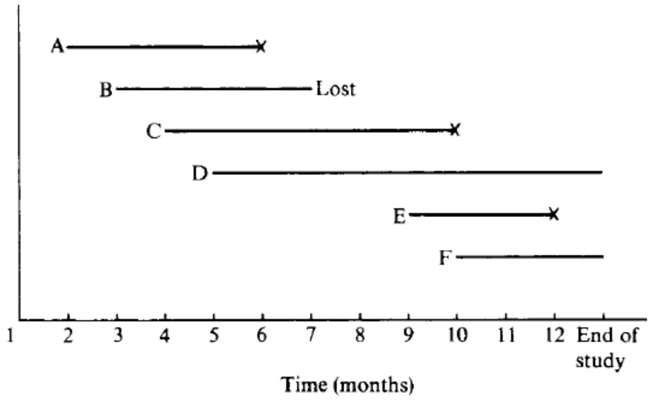
مثال ۱.۲. *فرض کنید شش بیمار مبتلا به سرطان خون حاد وارد یک مطالعه‌ی پزشکی با دوره‌ی یکساله می‌شوند. همچنین فرض کنید هر شش بیمار معالجه می‌شوند و بهبود می‌یابند. زمان‌های بهبودی آنها در شکل ۱ نشان داده شده است. بیماران A، C و E در شروع ماه‌های دوم، چهارم و نهم به بهبودی دست می‌یابند و به ترتیب بعد از چهار، شش و سه ماه به حال نخستین خود برمی‌گردند. بیمار B در شروع ماه سوم به بهبودی می‌رسد اما بعد از چهار ماه برای پیگیری مفقود می‌شود، بنابراین دوره‌ی بهبودی او حداقل چهار ماه است. بیماران D و F به ترتیب در شروع ماه‌های پنجم و دهم به بهبودی می‌رسند و تا پایان مطالعه در دوره‌ی بهبودی به سر می‌برند، پس زمان بهبودی آنها به ترتیب حداقل هشت و سه ماه است. بنابراین زمان‌های بهبودی این شش بیمار به ترتیب ۴، +۴، ۶، +۸، ۳ و +۳ است که علامت + نشان‌دهنده مشاهدات سانسور شده است.*

۲. سانسور از چپ: سانسور چپ زمانی رخ می‌دهد که زمان بقا از زمان ورود فرد به مطالعه کمتر است. به عبارتی در سانسور از چپ حادثه مورد نظر قبل از ورود فرد به مطالعه اتفاق افتاده و او را از دست داده باشیم. در سانسور چپ مشاهدات عبارتند از (Y_i, δ_i) که در آن $Y_i = \max(T_i, C_i)$ و نشانگر سانسور برابر است با،

$$\delta_i = \begin{cases} 0, & C_i \leq T_i, \\ 1, & C_i > T_i. \end{cases}$$

مثال ۲.۲. *جدی‌ترین عارضه‌ی دیابت برای چشم، پیشرفت شبکیه‌ی دیابتی است. در یک آزمایش، یک همه‌گیر شناس قصد دارد از معاینه‌ی شبکیه‌ی دیابتی، سن بیمار را در هنگام تشخیص اولیه‌ی بیماری دریابد. یک شرکت کننده‌ی ۵۰ ساله که مبتلا به شبکیه‌ی دیابتی پیشرفته‌ای است، وارد آزمایش می‌شود. چون ثبت دقیقی از زمان آغاز بیماری او در شواهد اولیه وجود ندارد سن او (یعنی ۵۰) یک مشاهده‌ی سانسور شده‌ی چپ برای زمان آغاز*

[1] *Epidemiologic*



شکل ۱: مثالی از داده‌های سانسورشده‌ی تصادفی

بیماری است. این بدان معناست که سن تشخیص اولیه‌ی بیماری او حداکثر ۵۰ است.

۳. سانسور بازه‌ای: سانسور بازه‌ای زمانی اتفاق می‌افتد که می‌دانیم پیشامد مورد نظر بین زمان‌های a و b رخ می‌دهد. برای مثال اگر ثبت‌های پزشکی نشان دهند که بیمار مثال قبل، در ۴۵ سالگی شبکیه‌ی دیابتی ندارد آنگاه سن تشخیص بیماری او بین ۴۵ و ۵۰ است. در واقع ترکیبی از هر دو نوع سانسور چپ و راست با هم سانسور بازه‌ای است. بدیهی است اگر $a = \infty$ سانسور از چپ و اگر $b = \infty$ سانسور از راست خواهد شد.

۲.۲ آگاهندگی سانسور

صرفنظر از نوع سانسور، نوعاً فرض می‌شود که سانسور ناآگاهنده[۲] است بدین معنا که زمان سانسور C_i مستقل از زمان بقای T_i است. به عبارتی اطلاعات زمان سانسور برای یک فرد، هیچ اطلاعات اضافه‌ای را درباره‌ی درستنمایی بقای این شخص در آینده‌ای که فرد مورد پیگیری قرار گیرد، بدست نمی‌دهد. اکثر روش‌های تحلیل بقا به این فرض کلیدی نیاز دارند، یعنی موضوعات یا افراد سانسور شده نسبت به بقیه در مخاطره‌ی بالاتری برای شکست قرار ندارند. بنابراین یک فرد سانسور شده در زمان t باید نماینده همه‌ی افرادی باشد که تا زمان t زنده می‌مانند [۱۰].

تعریف ۱.۲. *سانسور را* **ناآگاهنده (مستقل)** *می‌گویند اگر احتمال سانسور شدن در زمان t به احتمال شکست در زمان t بستگی نداشته باشد.*

بنابراین در سانسور ناآگاهنده فرض بر آن است که خروج شرکت کنندگان از مطالعه به دلایلی است که به مطالعه ارتباط ندارد.

اما سانسور آگاهنده[۳] زمانی اتفاق می‌افتد که شرکت‌کنندگان، فرآیند پیگیری مطالعه را بخاطر دلایلی مرتبط با مطالعه از دست دهند.

تعریف ۲.۲. *سانسور را* **آگاهنده (وابسته)** *می‌گویند اگر توزیع T_i شامل اطلاعاتی درباره‌ی پارامترهای مشخص کننده‌ی توزیع C_i باشد.*

[۲] $Non-informative$

[۳] $Informative$

متأسفانه سانسور آگاهنده می‌تواند منجر به یک اریبی در نتایج تحلیل بقا شود. به عنوان مثال در یک مطالعه به منظور مقایسه‌ی بقای بیماران بعد از دو درمان متفاوت، درمان برای عده‌ای ممکن است نامؤثر باشد و بیمار را ناخوش‌تر کند و فرد دیگر قادر به پیگیری مطالعه نباشد و یا بر عکس برای عده‌ای دیگر ممکن است درمان مؤثر واقع شود و فرد دیگر نیاز به ادامه‌ی پیگیری را احساس نکند. اگر این شرکت‌کنندگان به طور منظم سانسور شوند آنگاه تأثیر درمان واقعی قابل دریافت نخواهد بود و نتایج مطالعه اریب خواهد شد [۱].

۳ برآورد تابع بقا

در تحلیل بقا، توزیع‌های مختلفی برای زمان وقوع حادثه (زمان بقا) به کار می‌رود که یکی از آنها تابع بقاست.

تعریف ۱.۳. *احتمال این که فرد بیشتر از t زنده بماند، تابع بقا نامیده می‌شود. این تابع را با $\bar{F}(t)$ نمایش داده و برابر است با،*

$$\begin{aligned}\bar{F}(t) &= 1 - F(t)\\ &= P(T > t),\end{aligned}$$

که در آن $F(.)$ تابع توزیع متغیر تصادفی T است.

فرض کنید $T_1,\ldots,T_n$ متغیرهای تصادفی مستقل و هم‌توزیع با تابع توزیع پیوسته‌ی $F(t) = P(T \leq t)$ باشند. گیریم دنباله‌ای مستقل از متغیرهای تصادفی $C_1,\ldots,C_n$ با تابع توزیع مشترک $G(t) = P(C \leq c)$ از راست آنها را سانسور کنند. می‌خواهیم تابع بقای T را برآورد کنیم. به علت سانسور راست تنها Y_i و δ_i ($i = 1, ..., n$) قابل مشاهده هستند.

برآورگر کاپلان مه‌یر(KM)[۴] روش متداول برای برآورد تابع بقای $\bar{F}(t)$ است [۶]. فرض کنید $Y_{(1)} < ... < Y_{(n)}$، $Y_{(i)}$های مرتب و $\delta_{[i]}$ نشانگر سانسور همراه با $Y_{(i)}$ را نشان می‌دهد، یعنی $\delta_{[i]} = \delta_j$ اگر $Y_{(i)} = Y_j$. با این نماد گذاری برآوردگر KM به صورت زیر است،

$$\bar{F}_n^{KM}(t) = \prod_{i=1}^{n} \left(1 - \frac{\delta_{[i]}}{n-i+1}\right)^{I[Y_{(i)} \leq t]} I[Y_{(n)} > t],$$

که در آن $I(.)$ تابع نشانگر است. برآورگر کاپلان مه‌یر روش متداول برای برآورد تابع بقای $S_T(t)$ تحت فرض سانسور مستقل است. با وجود این نشان داده شده است که استفاده از این برآوردگر هنگامی که T و C وابسته هستند، ممکن است منجر به استنباط‌های نادرست در تحلیل بقا شود. لاگاکووس [۱۰] چندین مثال ارائه داده که فرض سانسور ناآگاهنده را مورد بحث قرار داده است و نشان داده که اگر چنین وابستگی نادیده گرفته شود نتایج استنباط اریب خواهد بود. به منظور برآورد یکتای تابع توزیع زمان بقا ناگزیر به در نظر گرفتن فرض وابستگی بین T_i و C_i به ازای هر i هستیم [۱۴]. بنابراین باید این وابستگی را مدل‌بندی کنیم که تا کنون روش‌های زیادی بدین منظور پیشنهاد شده‌اند.

۱.۳ برآورد تابع بقا تحت تابع مفصل

یکی از روش‌های متداول در مدل‌بندی الگوی وابستگی بین زمان بقا و زمان سانسور استفاده از تابع مفصل[۵] است. در نظریه احتمال و آمار تابع مفصل برای بررسی الگوی وابستگی بین دو متغیر تصادفی به‌کار می‌رود. در

[۴] $Kaplan - Meier$
[۵] $Copula$

واقع تابع مفصل، تابعی است که تابع توزیع چند متغیره چند بعدی را به توابع حاشیه‌ای یک بعدی آن مرتبط می‌سازد [۱۱، ۳]. ویژگی خاص کلاس مفصل این است که الگوی وابستگی از اثرهای حاشیه‌ای جدا می‌شود به طوری که رابطه‌ی وابستگی می‌تواند بدون مشخص کردن توزیع‌های حاشیه‌ای مطالعه شود. برای اطلاعات بیشتر درباره‌ی مفصل‌ها به نیلسین [۱۱] مراجعه کنید. ژینگ و کلیین [۱۷] به طور کلی یک تابع مفصل معلوم U از T و C را به صورت زیر معرفی کردند،

$$\begin{aligned} S(t_1, t_2) &= P(T > t_1, C > t_2) \\ &= U(\bar{F}(t_1), \bar{G}(t_2)), \end{aligned}$$

که در آن $\bar{F}$، $\bar{G}$ به ترتیب تابع‌های بقای حاشیه‌ای متغیرهای تصادفی T، C هستند و S تابع بقای توأم از (T, C) است. همچنین تابع مفصل U یک تابع توزیع دو متغیره روی مربع واحد با توزیع‌های حاشیه‌ای یکنواخت است.

ژینگ و کلیین [۱۷] تحت این تابع مفصل یک برآوردگر برای تابع بقا به نام برآوردگر مفصل-گرافیک[۶] معرفی کردند. با وجود این برآوردگر آنها شکل بسته‌ای نداشت و بدین منظور ری ویست و ویلز [۱۲] برای مدل‌بندی وابستگی از کلاس مفصل ارشمیدسی[۷] زیر استفاده کردند،

$$S(t_1, t_2) = \varphi^{-1}(\varphi(\bar{F}(t_1)) + \varphi(\bar{G}(t_2))), \qquad (۱)$$

که در آن $\varphi : [0, 1] \to [0, \infty)$ یک تابع معلوم پیوسته، محدب و اکیدا نزولی است و $\varphi(1) = 0$. آنها تحت این کلاس برآوردگر زیر را برای تابع بقا پیشنهاد دادند،

$$\bar{F}_n^{RW}(t) = \varphi^{-1}[-\sum_{\substack{Y_i \le t \\ \delta_i = 1}} \varphi(\bar{H}_n(Y_i)) - \varphi(\bar{H}_n(Y_i) - \frac{1}{n})],$$

که در آن H_n تابع توزیع تجربی Y است.

کلیین و مویسبرگر [۷] نیز تحت مفصل کلایتون وابستگی بین زمان سانسور و زمان بقا را بررسی کردند.

۲.۳ برآورد تابع بقا تحت مدل KG[۸]

همچنین در مطالعه‌ی سانسور آگاهنده، کوزیول-گرین [۹] یک زیر مدل برای برآوردگر کاپلان مه‌یر در نظر گرفت که در آن فرض می‌کند تابع بقای طول عمر سانسور، توانی از تابع بقای متناظر طول عمر واقعی است و در عمل نشان داده شده این فرض اصلاً غیرواقعی نیست. در تحلیل‌های بقا مدل کلاسیک KG تحت سانسور تصادفی آگاهنده، به‌کار می‌رود [۱۵]. در حقیقت فرض تناسب در مدل KG منجر به یک برآوردگر کاراتر برای تابع بقا نسبت به برآوردگر کاپلان-مه‌یر می‌شود. امتیاز اصلی مدل KG این است که برآوردگر تابع توزیع زمان بقا از شکل ساده‌ای برخوردار است. فرض کنید تابع بقای متغیر زمان بقا را با $\bar{F}(t)$ و تابع بقای متغیر زمان سانسور تصادفی را با $\bar{G}(t)$ نشان می‌دهیم. بنابراین تحت مفروضات مدل KG داریم،

$$\bar{G}(t) = \bar{F}(t)^{\beta}, \qquad (۲)$$

که در آن $\beta > 0$ و از آن به عنوان پارامتر سانسور تعبیر می‌شود. سوروگو [۵] مرور مفصلی بر مدل KG داشته و

[۶] $Copula - graphic$

[۷] $Archimedean \quad copula$

[۸] $Koziol - Green$

برآوردهای مربوطه و روش‌های آزمون آن را بررسی کرده است. وی بر اساس مدل KG برآورد تابع بقا را به صورت زیر ارائه داد،

$$\bar{F}_n^{(KG)}(t) = [1 - H_n(t)]^{\alpha_n},$$

که $\alpha_n = \frac{1}{n}\sum_{j=1}^{n}\delta_j$ و $H_n(t) = \frac{1}{n}\sum_{j=1}^{n} I\{Y_j > t\}$. این برآوردگر در غیاب متغیرهای کمکی به‌کار می‌رود.

اگر بتوان پارامتر سانسور β در مدل KG را به متغیرهای کمکی از طریق مدل KG شرطی، وابسته کرد در این صورت مدل جدید واقع بینانه‌تر خواهد شد و قابلیت مدل به شدت افزایش می‌یابد. مدل متناسب KG شرطی که توسط ویراویربیکیا و سوواریز [۱۶] به عنوان تعمیمی از مدل KG معرفی شد، فرض می‌کند β تابعی از متغیر کمکی $\mathbf{X}$ باشد. آنها نشان دادند که امکان وابستگی بین پارامتر β از مدل KG و متغیر کمکی $\mathbf{X}$ وجود دارد و در مدل متناسب شرطی این وابستگی را لحاظ کردند. بنابراین مدل مخاطره‌ی متناسب شرطی به جای مدل (۲)، مدل زیر را پیشنهاد می کند،

$$\bar{G}_x(t) = \bar{F}_x(t)^{\beta_x},$$

که در آن $\bar{F}_x(t)$ و $\bar{G}_x(t)$ به ترتیب توابع توزیع شرطی T و C به شرط x هستند.

براییکرز و ویراویربیکیا [۴] با استفاده از ایده‌ی به‌کارگیری تابع مفصل در مدل KG این مدل را به صورت زیر بسط دادند،

$$\bar{G}_x(t) = \varphi_x^{-1}(\beta_x\varphi_x(\bar{F}_x(t))).$$

همچنین از رابطه‌ی (۱) داریم،

$$\begin{aligned}\bar{H}_x(t) &= S_x(t,t)\\ &= \varphi_x^{-1}(\varphi_x(\bar{F}_x(t)) + \varphi_x(\bar{G}_x(t))),\end{aligned}$$

در این صورت،

$$\begin{aligned}\bar{H}_x(t) &= \varphi_x^{-1}(\varphi_x(\bar{F}_x(t)) + \beta_x\varphi_x(\bar{F}_x(t)))\\ &= \varphi_x^{-1}((1+\beta_x)\varphi_x(\bar{F}_x(t))),\end{aligned}$$

در نتیجه یک برآوردگر برای تابع بقا به صورت زیر حاصل می‌شود،

$$\bar{F}_x^{(BV)}(t) = \varphi_x^{-1}\left(\frac{1}{(1+\beta_x)}\varphi_x(\bar{H}_x(t))\right),$$

که با فرض $\gamma_x = \frac{1}{1+\beta_x} = P(\delta_x = 1)$ داریم:

$$\bar{F}_x^{(BV)}(t) = \varphi_x^{-1}\left(\gamma_x\varphi_x(\bar{H}_x(t))\right). \qquad (۳)$$

در این رابطه $\bar{H}_x$ و γ_x را بوسیله برآوردگرهای زیر جایگزین می‌کنیم،

$$H_x(t) = \sum_{i=1}^{n} w_{ni}(x,h_n)I(Y_i \le t),$$

و

$$\gamma_{xh} = \sum_{i=1}^{n} w_{ni}(x,h_n)I(\delta_i = 1),$$

که در آن داریم،

$$w_{ni}(x,h_n)=\frac{1}{c_n(x,h_n)}\int_{x_{i-1}}^{x_i}\frac{1}{h_n}K(\frac{x-z}{h_n})dz,$$

$$c_n(x,h_n)=\frac{1}{c_n(x,h_n)}\int_{x_0}^{x_n}\frac{1}{h_n}K(\frac{x-z}{h_n})dz.$$

همچنین تابع هسته $K(.)$ یک تابع چگالی احتمال معلوم است و پهنای باند h_n یک پارامتر هموارسازی در هر یک از مختصات آن است. با ترکیب این برآوردگرها در رابطه (۳) ، برآوردگر زیر برای تابع بقای $\bar{F}_x^{(BV)}(t)$ به صورت زیر حاصل می‌شود،

$$\bar{F}_{xh}^{(BV)}(t)=\varphi_x^{-1}\left(\gamma_{xh}\varphi_x\bar{H}_{xh}(t)\right).$$

ملاحظه می‌شود که این برآوردگر الگویی ساده‌تر از برآوردگر مفصل-گرافیک تحت سانسور آگاهنده دارد [۲].

۴ بحث و نتیجه‌گیری

در این مقاله موضوع سانسور آگاهنده یا وابستگی بین زمان پیشامد و زمان سانسور در تحلیل بقا مورد بحث قرار گرفت. تحت تابع مفصل معلوم ارشمیدسی، به معرفی چندین روش برای مدل‌بندی الگوی وابستگی بین زمان پیشامد و زمان سانسور پرداختیم. همچنین تحت سانسور آگاهنده مدل KG کلاسیک را معرفی کردیم. سپس با استفاده از توابع مفصل، مدل KG را به نحوی تعمیم دادیم که امکان در نظر گرفتن ارتباط میان زمان سانسور و زمان پیشامد در برآورد تابع بقا وجود داشته باشد.

مراجع

[۱] صادپاپی، ا. (۱۳۹۳). کاهش بعد بسنده برای داده‌های بقای سانسوریده. پایان نامه کارشناسی ارشد، دانشگاه شهید بهشتی تهران.

[2] Beran, R. (1981). Nonparametric regression with randomly censored survival data. Technical Report, Univ. California, Berkeley.

[3] Braekers, R. and Veraverbeke, N. (2005). A copula-graphic estimator for the conditional survival function under dependent censoring. The Canadian Journal of Statistics, 33, 429-447.

[4] Braekers, R. and Veraverbeke, N. (2007). A conditional koziol-green model under dependent censoring. Statistics and Probability Letters, 78, 927-937.

[5] Csorgo, S. (1988). Estimation in the proportional hazards model of random censorship. Statistics, 19, 437-463.

[6] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc., 53, 457-481.

[7] Klein, J. P. and Moeschberger, M. L. (1988). Bounds on net survival probabilities for dependent competing risks. Biometrics, 44, 529-538.

[8] Klein, J. P. and Moeschberger, M. L. (1997). Survival analysis: techniques for censored and truncated data. Springer-Verlag, New York.

[9] Koziol, J. A. and Green, S. B. (1976). A Cramer-von mises statistic for randomly censored data. Biometrika, 63, 465-474.

[10] Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics, 35, 139-156.

[11] Nelsen, R. B. (2006). An introduction to copulas. Springer-Verlag, New York.

[12] Rivest, L. and Wells, M. T. (2001). A martingale approach to the copulagraphic estimator for the survival function under dependent censoring. J. Multivariate Anal, 79, 138-155.

[13] Sklar, A. (1973). Random variables, joint distribution functions, and copulas. Kybernetika.

[14] Tsiatis, A. (1972). A nonidentifiability aspect of the problem of competing risks. Proceedings of the National Academy of Sciences, USA, 72, 20-22.

[15] Veraverbeke, N. (2006). Regression quantiles under dependent censoring. Statistics, 40, 117-128.

[16] Veraverbeke, N. and Cadarso Suarez, C. (2000). Estimation of the conditional distribution in a conditional Koziol-Green model. Test, 9, 97-122.

[17] Zheng, M. and Klein, J. P. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. Biometrika, 82, 127-138.