ارتباط تابع نرخ خطر با اطلاع فیشر در داده‌های سانسور شده

جعفر احمدی[۱] - بهاره خطیب آستانه[۱] - مصطفی رزمخواه[۱]

چکیده

موارد زیادی وجود دارد که در آزمایش‌های دنباله‌ای همهٔ مشاهدات نمونه ثبت نشده‌اند و یا همهٔ واحدهای نمونه به نتیجه نرسیده‌اند، در واقع با یک نمونهٔ سانسور شده روبرو هستیم. این مقاله، سعی بر این دارد تا انواع سانسورهای معروف را به همراه کاربردهای آنها معرفی کند و موضوع اطلاع فیشر نهفته در داده‌های سانسور شده را مورد مطالعه قرار دهد. همچنین ارتباط تابع نرخ خطر با اطلاع فیشر نهفته در یک متغیر تصادفی را بیان کند و نهایتاً شرایطی را بر اساس اطلاع فیشر داده‌های سانسور شده نوع II ارائه دهد که منجر به تجزیهٔ تابع نرخ خطر به صورت دو تابع یکی برحسب مشاهده و دیگری بر حسب پارامتر شود.

واژه های کلیدی: داده‌های سانسور شده، اطلاع فیشر، تابع نرخ خطر، آنالیز بقا، نامساوی کرامر_ رائو، آماره‌های ترتیبی

۱. مقدمه

در آزمایش های دنباله‌ای موارد زیادی وجود دارد که آزمایشگر بنا به دلایلی به تمام مشاهدات دسترسی ندارد، به عبارت دیگر تعدادی از مشاهدات به نوعی حذف شده‌اند. بنابراین موضوع اطلاع فیشر نهفته در باقیماندهٔ مشاهدات حائز اهمیت است که امروزه توجه محققین زیادی را به خود جلب نموده است.

فرض کنید n واحد را تا زمان خرابی آنها تحت نظر داریم، این واحدها می‌توانند برخی از سیستم‌ها، مؤلفه‌ها یا تراشه‌های رایانه‌ای در آزمایش های مطالعهٔ قابلیت اعتماد و مانند آن باشند. طول عمر این n واحد را با $X_1, X_2, \cdots, X_n$ نشان می‌دهیم که $X_i(>0)$ ها متغیرهای تصادفی پیوسته، مستقل و هم توزیع با تابع توزیع مشترک $F(x;\theta)$ و تابع چگالی احتمال $f(x,\theta)$ می‌باشند، که در آن $\theta \in \Theta$ پارامتر مجهول است. بنابراین مشاهدات به طور طبیعی، به صورت بردار آماره‌های ترتیبی ظاهر می‌شوند. حال اگر بنا به دلایلی آزمایش را متوقف سازیم، در این صورت یک نمونهٔ سانسور شده خواهیم داشت که در آن آماره‌های ترتیبی نقش مهمی ایفا می‌کنند. بدیهی است که با سانسور کردن مشاهدات، مقداری از اطلاعات موجود را از دست می‌دهیم، با این وجود گاهی اوقات مایل (یا مجبور) به سانسور کردن هستیم.

امروزه پژوهش گران زیادی موضوع اطلاع فیشر در داده‌های سانسور شده و آماره‌های مرتب را مورد مطالعه قرار داده‌اند، از جمله می‌توان به پارک[۲] [۶]، زن[۳] [۸]، گرستبخ[۴] [۳]، آرنلد و همکاران[۵] [۲] اشاره نمود. پارک (۱۹۹۶)، تجزیهٔ اطلاع فیشر آماره‌های مرتب را بکار برده و با استفاده از روابط بازگشتی بین گشتاورهای آماره‌های مرتب، فرمولهایی ساده برای بدست آوردن اطلاع در قسمتهای مختلف آماره‌های مرتب بدست آورده است. همچنین نمودار اطلاع را برای یافتن اطلاع فیشر در زیر مجموعه‌ای از آماره‌های مرتب متوالی معرفی کرده است و زنگ [۸]، میزان اطلاع فیشر کاهش یافته را در نمونه‌ای که به روش تصادفی سانسور شده، بدست آورده است.

در این مقاله قضیه‌ای ارائه می‌شود که براساس آن می‌توانیم با استفاده از خاصیت خطی اطلاع فیشر در داده‌های سانسور شده نوع II، تابع نرخ خطر، $h(x;\theta)$، را به صورت حاصلضرب دو تابع یکی برحسب x و دیگری بر حسب θ بیان کنیم. در بخش دوم به معرفی سانسور و انواع آن می‌پردازیم، سپس دربارهٔ کاربرد هر یک از سانسورها صحبت می‌کنیم، در بخش سوم تابع نرخ خطر را معرفی کرده و کاربرد آن را در قابلیت اعتماد بیان می‌کنیم، در بخش چهارم اطلاع فیشر را

۱ -گروه آمار، دانشکدهٔ علوم ریاضی، دانشگاه فردوسی مشهد

۲- Park
۳- Zheng
۴- Gerstbakh
۵- Arnold et.al.

معرفی می‌کنیم و تعاریفی از اطلاع فیشر نهفته در داده‌های سانسور شده از سمت راست ارائه می‌دهیم، در بخش پنجم به بررسی ارتباط تابع نرخ خطر با اطلاع فیشر نهفته در یک متغیر تصادفی می‌پردازیم. در بخش ششم رابطهٔ تابع نرخ خطر را با اطلاع فیشر در داده‌های سانسور شدهٔ نوع II بیان می‌کنیم و در بخش هفتم چند مثال در خصوص نتایج ریاضی بخش های قبل ارائه می‌دهیم.

۲. داده‌های سانسور شده

در بسیاری اوقات در آزمون های طول عمر، آزمایش های کلینیکی مطالعات تأثیر دوز سم‌ها، تحقیقات زیست‌شناسی، آنالیز بقا و دیگر زمینه‌های کاربردی علم آمار با نمونه‌هایی مواجه هستیم که در قسمت‌هایی از مقادیر ممکن متغیر تصادفی مورد مطالعه محدود شده‌اند و همهٔ مشاهدات نمونه ثبت نشده‌اند یا همهٔ واحدهای نمونه به نتیجه نرسیده‌اند. به‌عنوان مثال در یک نمونه به حجم n ممکن است تنها n_1 واحد نمونه قابل مشاهده باشند و لذا برای $n-n_1$ واحد نمونه مشاهده‌ای در دست نیست. این محدودیت ممکن است به صورت اختیاری توسط مشاهده‌گر (آمارگر) اعمال شود یا ماهیت آزمایش‌ها طوری باشند که خودبخود محدودیتی را در مشاهدات بوجود آورده‌اند. بعضی از محدودیت‌ها عبارتند از فرصت کم برای اعلام نتایج، طولانی شدن مدت آزمایش، عدم دسترسی به همهٔ واحدها و یا مأیوس شدن از نتیجه دادن همهٔ واحدها. این محدودیت پیش‌آمده در نمونه‌ها را سانسور تعریف می‌کنیم.

سانسور معمولاً به صورت‌های مختلف اعمال می‌شود و بعضی از انواع آن در ذیل آمده است. برای جزئیات بیشتر دربارهٔ انواع سانسورها و کاربرد آنها می‌توان به کتاب های آرنلد و همکاران [۲]، لاوولس[۱] [۴] و نلسن[۲] [۵] مراجعه کرد.

۲. ۱ سانسور نوع I

گاهی اوقات به دلایل ذکر شده در بالا، یک آزمون طول عمر باید در زمان مشخص $t_\circ$ به‌پایان رسانده شود، در این‌صورت همهٔ واحدها در زمان $t=\circ$ فعال می‌شوند و شروع به کار می‌کنند تا وقتی که یکی از دو حالت زیر رخ دهد.

الف) واحد قبل از زمان $t_\circ$ از کار بیافتد که در این‌صورت طول عمر این واحد به طور دقیق معلوم خواهد بود.

ب) واحد تا رسیدن به زمان $t_\circ$ کار کند. در این‌صورت در لحظهٔ $t_\circ$، به علت پایان یافتن آزمایش واحد از کار می‌افتد و طول عمر $t_\circ$ برای این واحد ثبت می‌شود. همانطور که ذکر شد این یک طول عمر سانسور شده است و در واقع طول عمر واقعی آن بزرگتر یا مساوی این مقدار است. این نوع از سانسور، سانسور نوع اول نامیده می‌شود. این حالت اغلب در تحقیقات پزشکی رخ می‌دهد.

حال فرض کنید تعداد واحدها n باشد و r واحد تا قبل از زمان $t_\circ$ از کار بیافتد، بنابراین فقط اطلاعات کاملی از $r(\le n)$ مشاهدهٔ مرتب اول تا زمان $t_\circ$ در اختیار داریم، طول عمر مرتب شدهٔ این r واحد به صورت زیر است

$$x_{1:n} < x_{2:n} < \cdots < x_{r:n}$$

که در آن $x_{r:n} < t_\circ < x_{r+1:n}$.

علاوه بر این می‌دانیم که $n-r$ واحد نیز در زمان $t_\circ$ بقاء داشته‌اند، به عبارت دیگر طول عمر آنها بزرگتر یا مساوی $t_\circ$ است اما طول عمر سانسور شدهٔ $t_\circ$ برای هر کدام از آنها در نظر گرفته شده است.

تابع درستنمایی این نوع سانسور را می‌توان به صورت زیر نوشت

$$L(\theta \mid r, y) = \frac{n!}{(n-r)!}\prod_{i=1}^{r} f(y_i;\theta)\times \{1-F(t_\circ;\theta)\}^{n-r}$$
$$, y_1 < y_2 < \cdots < y_r < t_\circ \;, \quad \circ < r \le n$$

و

$$L(\theta \mid r, y) = \{1-F(t_\circ;\theta)\}^{n}$$
$$, \; t_\circ < y_1 \;, \quad r = \circ$$

که در آن y بردار آماره‌های ترتیبی است.

از آنجا که r، تعداد واحدیی که تا قبل از زمان $t_\circ$ از کار افتاده‌اند، یک متغیر تصادفی است، ممکن است یک عدد کوچک یا حتی صفر باشد، بنابراین می‌توان این موضوع را به عنوان یک ضعف در ساخت این نوع از داده‌ها (سانسور نوع اول) به حساب آورد.

مثال: از میان ترانزیستورهایی از یک نوع خاص، ۱۰۰ ترانزیستور انتخاب می‌شوند و در مدارهای الکترونیکی یکسانی قرار می‌گیرند. هدف ثبت مدت عمر ترانزیستورها به روز می‌باشد. چون احتمال می‌رود بعضی از ترانزیستورها مدت زمان خیلی زیادی عمر کنند و اعلام نتایج در اسرع وقت لازم باشد، آزمایش را بعد از زمانی که از قبل مشخص شده، خاتمه می‌دهیم. در این‌صورت یک نمونهٔ سانسور شدهٔ خواهیم داشت.

۲. ۲ سانسور نوع II

همانطور که دیدیم در سانسور نوع اول، r، تعداد واحدهای از کار افتاده تصادفی بود، حال وضعیتی را در نظر بگیرید که در آن شخص

۱- Lawless
۲- Nelson

بخواهد حتماً مقدار r طول عمر واقعی را مشاهده کند در این‌صورت زمان لازم برای رسیدن به rامین از کار افتادگی یعنی $x_{(r)}$ یک متغیر تصادفی است. بنابراین با فرض اینکه تمام n واحد در لحظهٔ $t=\circ$ فعال شده باشند، اطلاع بدست آمده در طول آزمایش شامل مجموعهٔ داده‌های زیر است

$$x_{1:n} < x_{2:n} < \cdots < x_{r:n}$$

تابع درستنمایی، عبارت است از

$$L(\theta \mid \underline{y}) = \frac{n!}{(n-r)!}\prod_{i=1}^{r} f(y_i;\theta)\times$$
$$\{1-F(y_r;\theta)\}^{n-r}, \quad y_1 < y_2 < \cdots < y_r .$$

در این نوع سانسور شخص نمی‌تواند از قبل، زمان طول کشیدن آزمایش را مشخص کند: این موضوع را می‌توان به عنوان یکی از ضعف های سانسور نوع II، به حساب آورد.

مثال: از میان گونه‌ای خاص از موش ها، ۱۰ موش به تصادف انتخاب و به آنها میکرب سل انسانی یکسانی را تزریق می‌کنیم و منتظر مرگ آنها می‌شویم، در صورتی که ۷ موش بمیرند آزمایش را خاتمه می‌دهیم و مدت عمر موش ها را به روز ثبت می‌کنیم. در اینجا تعداد روزی (مدت زمانی) که بعد از آن، سانسور انجام شده است متغیر تصادفی است.

همانگونه که ملاحظه می‌کنید تابع درستنمایی در سانسور نوع I و نوع II مشابه نمونهٔ سانسور شده از راست است، یعنی مقادیر بزرگ سانسور شده‌اند. سانسور از چپ نیز ممکن است داشته باشیم که در آن مقادیر کوچک‌تر سانسور شده‌اند، همچنین می‌توان سانسور از نواحی مختلف داشت. زمانی که سانسور از چپ و راست صورت می‌گیرد اغلب سانسور دو طرفه نامیده می‌شود.

۲. ۳ سانسور تصادفی

فرض کنید X و Y متغیرهای مستقل وهمتوزیع با تابع توزیع $F_X(x;\theta)$ و $G_Y(x;\theta)$ باشند، تحت سانسور تصادفی از سمت راست مشاهدات تنها شامل $(Z_1,\delta_1),\cdots,(Z_n,\delta_n)$ هستند که در آن $Z_i=\min(X_i,Y_i)$ و $\delta_i = I(X_i \le Y_i)$، ($I(A)$ یک تابع نشانگر است) به طوری که

$$\delta_i = 1 \quad if \;\; X_i \le Y_i \;, \quad \text{بدون سانسور}$$
$$= \circ \quad if \;\; X_i > Y_i \;, \quad \text{با سانسور}$$

لازم بـه ذکـر اسـت، Z_iهـا و δ_iهـا مـستقل از یکدیگرنـد. همچنـین $Z_1,Z_2,\cdots,Z_n$مستقل و همتوزیع می‌باشند و δ_i نیز بـرای نـشان دادن اطلاعات سانسور شده به کار رفته است.

تابع درستنمایی $(\mathrm{Z},\boldsymbol{\delta})$ عبارت است از

$$L(\theta \mid \mathbf{z},\boldsymbol{\delta}) = \prod_{i=1}^{n}\left\{f(z_i;\theta)\overline{G}(z_i;\theta)\right\}^{\delta_i}\times$$
$$\left\{g(z_i;\theta)\overline{F}(z_i;\theta)\right\}^{1-\delta_i}$$

که در آن $\overline{F}=1-F$ و $\overline{G}=1-G$.

حالت ساده‌تر سانسور تصادفی زمانی اتفاق می‌افتد که توزیع Y به θ بستگی ندارد. همچنین هنگامیکه $Y_i=t_c$، یعنی Y_iها مقداری ثابت باشند، داده‌های سانسور شدهٔ نوع اول را خواهیم داشت.

سانسور تصادفی در مطالعات پزشکی، اپیدمیولوژی و مهندسی کاربرد فراوان دارد. چرا که در این مطالعات حالت های مختلفی از داده‌های سانسور شده را می‌توانیم داشته باشیم.

مثال: در یک بررسی طول عمر، در مورد یک بیماری یا یک دستگاه، ممکن است برخی از واحدها تا پایان دوره زنده بمانند و یا به علتی غیر از عامل مورد نظر از مطالعه خارج شوند.

۳. تعریف تابع نرخ خطر و کاربرد آن

یکی از توابع مهم در مبحث قابلیت اعتماد تابع نرخ شکست است که گاهی تحت عناوینی مانند تابع نرخ خطر، تابع نرخ زمانی تعمیر و یا نرخ ویژهٔ از کار افتادن نیز به کار می‌رود. به بیان ریاضی معیار تابع نرخ خطر به عنوان یک نسبت، عبارت است از احتمال خراب شدن سیستم در فاصلهٔ زمانی $(t,t+\Delta t)$ نسبت به Δt به شرط اینکه تا زمان t خراب نشده باشد. بنابراین

$$h(x;\theta) = \lim_{\Delta t\to \circ}\frac{P(t<T\le t+\Delta t \mid T>t)}{\Delta t}. \qquad (۱)$$

در تعریف فوق متغیر پیوستهٔ T بیانگر طول عمر سیستم می‌باشد. اگر $f(t)$ و $F(t)$ به ترتیب تابع چگالی احتمال و تابع توزیع تجمعی متغیر تصادفی T باشند، آنگاه با توجه به (۱) داریم

$$h(x;\theta) = \lim_{\Delta t\to \circ}\frac{P(t<T\le t+\Delta t \mid T>t)}{\Delta t}$$
$$= \lim_{\Delta t\to \circ}\frac{P(t<T\le t+\Delta t \;\; and \;\; T>t)}{\Delta t . P(T>t)}.$$

از طرفی می‌دانیم $\overline{F}(t)=P(T>t)$ است، بنابراین

$$h(x;\theta) = \lim_{\Delta t\to \circ}\frac{P(t<T\le t+\Delta t)}{\Delta t \;\overline{F}(t;\theta)}$$
$$= \frac{1}{\overline{F}(t;\theta)}\lim_{\Delta t\to \circ}\frac{P(t<T\le t+\Delta t)}{\Delta t}$$
$$= \frac{f(t;\theta)}{\overline{F}(t;\theta)}. \qquad (۲)$$

سیستمی که دارای نرخ خطر صعودی(IFR)[۱] باشد با افزایش زمان فرسوده می‌شود. به عنوان مثال احتمال خرابی چنین سیستمی، درست پس از هزار ساعت با شرط اینکه سیستم هزار ساعت دوام آورده باشد، از احتمال خرابی درست پانصد ساعت با شرط اینکه پانصد ساعت دوام آورده باشد، بزرگتر است. اگر تابع نرخ خطر نزولی باشد، قابلیت اعتماد سیستم با افزایش زمان بهبود پیدا می‌کند. به عبارت دیگر سیستم دارای نرخ خطر نزولی(DFR)[۲] است. برخی سیستم ها دارای نرخ خطری هستند که در ابتدا نزولی، سپس به طور نسبی ثابت(CFR)[۳] و در نهایت صعودی می‌شود. به عنوان مثال، یک خودرو در ابتدای تولید دارای معایبی است که توسط سازندهٔ خودرو رفع می‌شود، در این دوره، دورهٔ رفع ایرادات، خودرو دارای نرخ خطر نزولی می‌باشد، یعنی با افزایش زمان قابلیت اعتماد افزایش پیدا می‌کند. سپس برای مدتی نرخ خطر کم و بیش ثابت می‌ماند (دورهٔ عمر مفید) و سرانجام زمانی که خودرو (سیستم) به مرحلهٔ فرسودگی می‌رسد، نرخ خطر صعود می‌کند.

٤. تعریف اطلاع فیشر و کاربرد آن

تعریف۱: برای یک متغیر تصادفی پیوسته (گسسته) X با تابع چگالی احتمال(تابع جرم احتمال) $f(x,\theta)$, $\theta\in\Omega$، تحت شرایط نظم اطلاع فیشر نهفته در X به صورت زیر تعریف می‌شود (بهبودیان، [۱])

$$I_{۱}(\theta)=$$

$$I_X(\theta)=E\left(\left[\frac{\partial}{\partial\theta}\log f_X(X;\theta)\right]^{۲}\right) \qquad (۳)$$

$$=-E\left(\frac{\partial^{۲}}{\partial\theta^{۲}}\log f_X(X;\theta)\right).$$

امید ریاضی فوق را نسبت به چگالی $f(x,\theta)$ حساب می‌کنیم و فرض می‌کنیم که وجود داشته باشد. زمانی که $\mathrm{X}=(X_{۱},X_{۲},\cdots,X_n)$ (n متغیر تصادفی مستقل از هم و همتوزیع) اطلاع فیشر به مانند تعریف بالا است با این تفاوت که در این حالت $f(x,\theta)$ به $f_{\mathrm{X}}(\mathrm{X};\theta)$ (تابع درستنمایی X) تبدیل می‌شود.

یکی از کاربردهای اطلاع فیشر در نامساوی کرامر رائو برای بدست آوردن کران پایین واریانس برآوردگرهای ناااریب می‌باشد؛ همچنین واریانس مجانبی توزیع حدی برآوردگرهای ماکزیمم درستنمایی (MLE) به اطلاع فیشر بستگی دارد. در مسائل آزمون فرض، اطلاع فیشر می‌تواند جهت ساختن آزمون امتیاز رائو استفاده شود. همچنین اطلاع فیشر با بسندگی و کارایی رابطهٔ نزدیکی دارد و نیز می‌تواند در انتخاب تعداد مشاهدات لازم برای سانسور، در نمونهٔ سانسور شدهٔ نوع II و ارزیابی اهمیت برآوردگرهای قرار گرفته شده روی داده‌های سانسور شده، در نظر گرفته شود.

٤. ۱ اطلاع فیشر در در داده‌های سانسور شدهٔ نوع II

در این بخش میزان اطلاع فیشر نهفته در نخستین r آمارهٔ مرتب را به طرق مختلف بیان می‌کنیم.

تعریف ۲: میزان اطلاع فیشر پارامتر θ در r آمارهٔ مرتب اوّل، $(X_{۱:n},\cdots,X_{r:n})$، را با $I_{۱,\cdots,r:n}(\theta)$ نشان می‌دهیم وتحت شرایط نظم از رابطهٔ زیر بدست می‌آید

$$I_{۱,\cdots,r:n}(\theta)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{x_{۲:n}}\left(\frac{\partial}{\partial\theta}f_{۱,\cdots,r:n}\right)^{۲}dF_{۱,\cdots,r:n}.$$

که در آن $f_{1,\dots,r:n}$ تابع چگالی احتمال توأم r آمارهٔ مرتب اوّل است.

تعریف ۳: فرض کنید $r/n\to p$، هنگامیکه $n\to\infty$ ، تاکاهاشی[۴] و سوگییورا[۵] [۷] نشان دادند که

$$\lim_{n\to\infty}\frac{۱}{n}I_{۱,\cdots,r:n}(\theta)=\int_{-\infty}^{\lambda_p}\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^{۲}f(x;\theta)\,dx+\frac{۱}{۱-p}\left(\int_{-\infty}^{\lambda_p}\frac{\partial}{\partial\theta}f(x;\theta)\,dx\right)^{۲}. \qquad (۴)$$

که در آن λ_p چندک p ام توزیع $F(x;\theta)$ می‌باشد. آنها سمت راست (۴) را با $I_{[\cdot,p]}(\theta)$ نشان داده و آن را اطلاع فیشر مجانبی سانسور نوع II نامیده‌اند.

حالت خاص تعریف بالا زمانی اتفاق می‌افتد که $r=n$ است، در این حالت

$$I_{[\cdot,۱]}(\theta)=\lim_{n\to\infty}\frac{۱}{n}I_{۱,\cdots,n:n}(\theta)$$

۱- Increasing Failure Rate
۲- Decreasing Failure Rate
۳- Constant Failure Rate
۴- Takahashi
۵- Sugiura

می‌باشد و اطلاع فیشر مجانبی در یک متغیر تصادفی نامیده می‌شود.

تعریف ۴: پارک (۱۹۹۶) با استفاده از خاصیت آماره‌های مرتب رابطهٔ زیر را برای اطلاع فیشر در داده‌های سانسور شدهٔ نوع II بیان نمود.

$$I_{1,\cdots,r:n}(\theta) = n\,I_1(\theta) - (n-r)\int_{-\infty}^{\infty} g(\omega;\theta)\, f_{r:n}(\omega;\theta)d\omega \qquad (۵)$$

که در آن، $I_1(\theta)$ در (۳) آمده است و $f_{r:n}(\omega;\theta)$، تابع چگالی $X_{r:n}$ می‌باشد و

$$g(\omega;\theta) = \int_{\omega}^{\infty}\left(\frac{\partial}{\partial\theta}\log\frac{f(x;\theta)}{1-F(\omega;\theta)}\right)^2 \times\frac{f(x;\theta)}{1-F(\omega;\theta)}dx. \qquad (۶)$$

۵. اطلاع فیشر برحسب تابع نرخ خطر

در این بخش نشان می‌دهیم که می‌توان اطلاع فیشر نهفته در یک متغیر تصادفی را برحسب تابع نرخ خطر بیان نمود.

لم۱: فرض کنید $h(x;\theta)=\frac{f(x;\theta)}{\overline{F}(x;\theta)}$، تابع نرخ خطر متغیر تصادفی X باشد، در این‌صورت

$$I_X(\theta) = \int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log h(x;\theta)\right)^2 f(x;\theta)\,dx.$$

برهان: طرف راست تساوی بالا را می‌توان به صورت زیر نوشت

$$\int_{-\infty}^{\infty}\left\{\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right\} f(x;\theta)dx + \int_{-\infty}^{\infty}\left\{\left(\frac{\partial}{\partial\theta}\log \overline{F}(x;\theta)\right)^2\right\} f(x;\theta)dx - 2\int_{-\infty}^{\infty}\left\{\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)\left(\frac{\partial}{\partial\theta}\log \overline{F}(x;\theta)\right)\right\} f(x;\theta)dx \qquad (۷)$$

آخرین انتگرال در رابطهٔ بالا برابر است با

$$\int\left(1/\overline{F}(x;\theta)\right)\,2\left(\frac{\partial}{\partial\theta}\frac{d}{dx}\overline{F}(x;\theta)\right)\times\left(\frac{\partial}{\partial\theta}\overline{F}(x;\theta)\right)dx = \int\left(1/\overline{F}(x;\theta)\right)d\left(\frac{\partial}{\partial\theta}\overline{F}(x;\theta)\right)^2$$

حال با روش انتگرال‌گیری جزء به جزء داریم

$$-2\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)\left(\frac{\partial}{\partial\theta}\log \overline{F}(x;\theta)\right) f(x;\theta)\,dx = -\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log \overline{F}(x;\theta)\right)^2 f(x;\theta)dx$$

با جایگذاری عبارت فوق در (۷) داریم

$$\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log h(x;\theta)\right)^2 f(x;\theta)\,dx = \int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 f(x;\theta)dx,$$

و با در نظر گرفتن (۳) برهان کامل می‌شود.

۶. تابع نرخ خطر و اطلاع فیشر داده‌های سانسور شدهٔ نوع II

در این بخش ابتدا ارتباط اطلاع فیشر داده‌های سانسور شدهٔ نوع II را با تابع نرخ خطر بیان می‌کنیم، سپس تحت قضیه‌ای نشان خواهیم داد که رابطهٔ $h(x;\theta)=u(x)v(\theta)$ برقرار است اگر و تنها اگر برای هر $n\geq 1$ و هر $1\leq r\leq n$ تساوی $I_{1,\cdots,r:n}(\theta)=rI_1(\theta)$ برقرار باشد. لازم به ذکر است که لم۲ و قضیهٔ۱ از مقالهٔ زنگ (۲۰۰۱) اقتباس شده است.

لم۲: زنگ (۲۰۰۱)، اگر $h(x;\theta)$ تابع نرخ خطر X باشد، در این‌صورت اطلاع فیشر مجانبی سانسور نوع II را به‌صورت زیر می‌توان محاسبه نمود

$$I_{[\cdot,p]} = \int_{-\infty}^{\lambda_p}\left(\frac{\partial}{\partial\theta}\log h(x;\theta)\right)^2 f(x;\theta)\,dx. \qquad (۸)$$

که در آن λ_p چندک p ام توزیع $F(x;\theta)$ می‌باشد.

برهان: همانند اثبات لم۱، طرف راست تساوی (۸) عبارت است از

$$\int_{-\infty}^{\lambda_p}(\frac{\partial}{\partial\theta}\log f)^{۲}\ fdx$$
$$-۲\int_{-\infty}^{\lambda_p}(\frac{\partial}{\partial\theta}\log f)(\frac{\partial}{\partial\theta}\log\overline{F})fdx$$
$$+\ \int_{-\infty}^{\lambda_p}(\frac{\partial}{\partial\theta}\log\overline{F})^{۲}\ fdx,$$

در محاسبهٔ انتگرال میانی به روش جزء به جزء خواهیم داشت

$$-۲\int_{-\infty}^{\lambda_p}(\frac{\partial}{\partial\theta}\log f)(\frac{\partial}{\partial\theta}\log\overline{F})f\,dx$$
$$=\frac{۱}{۱-p}\left(\int_{-\infty}^{\lambda_p}\frac{\partial}{\partial\theta}f(x;\theta)\,dx\right)^{۲}$$
$$-\int_{-\infty}^{\lambda_p}(\frac{\partial}{\partial\theta}\log\overline{F})^{۲}\ f\,dx.$$

و با جایگذاری در (۸) نتیجهٔ مطلوب حاصل می‌شود.

لم۳: هرگاه تابع نرخ خطر به صورت $h(x;\theta)=u(x)v(\theta)$ باشد، که در آن $u(x)$ و $v(\theta)$ توابعی مثبت هستند، آنگاه

$$I_{۱}(\theta)=(v'(\theta)/v(\theta))^{۲}. \qquad (۹)$$

برهان:

با در نظر گرفتن لم۱ داریم

$$I_1(\theta)=\int_{-\infty}^{\infty}(\frac{\partial}{\partial\theta}\log h(x;\theta))^2 f(x;\theta)\,dx$$
$$=\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}(\log u(x)+\log v(\theta))\right)^{۲} f(x;\theta)dx$$
$$=\int_{-\infty}^{\infty}\left(\frac{v'(\theta)}{v(\theta)}\right)^{۲} f(x;\theta)\,dx$$
$$=\left(\frac{v'(\theta)}{v(\theta)}\right)^{۲}.$$

در لم زیر نشان می‌دهیم که اگر تابع نرخ خطر را بتوان به صورت دو تابع، یکی برحسب مشاهده و دیگری بر حسب پارامتر تجزیه نمود آنگاه X متعلق به خانوادهٔ نمایی می‌باشد.

لم ۴: هرگاه تابع نرخ خطر به صورت $h(x;\theta)=u(x)v(\theta)$ باشد، که در آن $u(x)$ و $v(\theta)$ توابعی مثبت هستند، آنگاه

$$\overline{F}(x;\theta)=\exp[-(U(x)v(\theta)+d(x))].$$

برهان: می‌دانیم

$$h(x;\theta)=\frac{f(x;\theta)}{\overline{F}(x;\theta)}.$$

ازطرفی داریم

$$h(x;\theta)=\frac{f(x;\theta)}{\overline{F}(x;\theta)}=-\frac{d}{dx}\log\overline{F}(x;\theta).$$

و طبق فرض خواهیم داشت

$$-\frac{d}{dx}\log\overline{F}(x;\theta)=u(x)v(\theta).$$

حال از طرفین نسبت به x انتگرال می‌گیریم

$$\log\overline{F}(x;\theta)=-[U(x)v(\theta)+d(x)],$$

و از اینجا نتیجهٔ مطلوب حاصل می‌شود.

نتیجه: با در نظر گرفتن لم بالا تابع چگالی احتمال X برابر است با

$$f(x;\theta)=u(x)v(\theta)$$
$$\times\exp[-(U(x)v(\theta)+d(x))].$$

همانگونه که در تعریف۴ ذکر شد برای محاسبهٔ اطلاع فیشر نهفته در نخستین r آمارهٔ مرتب، نیاز به محاسبهٔ $g(\omega;\theta)$ داریم و از آنجا که به‌دست آوردن $g(\omega;\theta)$ با استفاده از تعریف آن همیشه ساده نیست، تحت شرایط لم زیر می‌توان روش ساده‌تری را برای محاسبهٔ اطلاع فیشر نهفته در داده‌های سانسور شده از راست ارائه داد.

لم۵: هرگاه تابع نرخ خطر به صورت $h(x;\theta)=u(x)v(\theta)$ باشد، که در آن $u(x)$ و $v(\theta)$ توابعی مثبت هستند، در این‌صورت

$$g(\omega;\theta)=I_X(\theta).$$

که $g(\omega;\theta)$ در (۶) تعریف شده است.

برهان: با توجه به تعریف $g(\omega;\theta)$، داریم

$$\log\frac{f(x;\theta)}{۱-F(\omega;\theta)}=\log\left(\frac{f(x;\theta)}{\overline{F}(\omega;\theta)}\frac{\overline{F}(x;\theta)}{\overline{F}(x;\theta)}\right)$$
$$=\log\left(h(x;\theta)\frac{\overline{F}(x;\theta)}{\overline{F}(\omega;\theta)}\right)$$
$$=\log h(x;\theta)\log\left(\frac{\overline{F}(x;\theta)}{\overline{F}(\omega;\theta)}\right).$$

با توجه به لم ۴ داریم

$$\log\left(\frac{\overline{F}(x;\theta)}{\overline{F}(\omega;\theta)}\right)$$

$$= \log \frac{\exp[-(U(x)v(\theta)+d(x))]}{\exp[-(U(\omega)v(\theta)+d(x))]}$$
$$= -(U(x)-U(\omega))v(\theta).$$

لذا با توجه به فرض و لم ۳ خواهیم داشت

$$\frac{\partial}{\partial\theta}\log\frac{f(x;\theta)}{1-F(\omega;\theta)} = \left(\frac{v'(\theta)}{v(\theta)}\right) - v'(\theta)(U(x)-U(\omega)).$$

از طرفی

$$\frac{f(x;\theta)}{1-F(\omega;\theta)} = \frac{u(x)v(\theta)\exp[-(U(x)v(\theta)+d(x))]}{\exp[-(U(\omega)v(\theta)+d(x))]}$$
$$= u(x)v(\theta)\exp[-(U(x)-U(\omega))v(\theta)].$$

و با جایگذاری در (۶) داریم

$$g(\omega;\theta) = \int_{-\infty}^{\infty}\Big\{\left(\frac{v'(\theta)}{v(\theta)}\right)^2 - 2\frac{v'(\theta)}{v(\theta)}v'(\theta)U(x,\omega) + (v'(\theta))^2(U(x,\omega))^2\Big\} \times v(\theta)u(x)\exp[-U(x,\omega)v(\theta)]\,dx,$$

که در آن $U(x,\omega)=U(x)-U(\omega)$. با محاسبهٔ انتگرال بالا خواهیم داشت

$$g(\omega;\theta) = \left(\frac{v'(\theta)}{v(\theta)}\right)^2,$$

و با در نظر گرفتن(۹) برهان کامل است.

لم۶: هرگاه برای هر $n\ge 1$ و هر $1\le r\le n$ ، رابطهٔ $I_{1,\dots,r:n}(\theta)=rI_1(\theta)$، برقرار باشد، آنگاه برای هر $0<p<1$، $I_{[\cdot,p]}=pI_{[\cdot,1]}$.

برهان: با در نظر گرفتن فرض داریم

$$\frac{1}{n}I_{1,\dots,r:n}(\theta)=\frac{r}{n}I_1(\theta).$$

با گرفتن حد از طرفین زمانی که $n\to\infty$ و در نظر گرفتن تعریف ۴ برهان کامل است.

لم ۷: اگر λ_p چندک p ام توزیع $F(x;\theta)$ باشد در این‌صورت

$$\frac{\partial}{\partial p}\lambda_p = \frac{1}{f(\lambda_p)}. \qquad (10)$$

برهان: از آنجایی که λ_p چندک p ام توزیع $F(x;\theta)$ می‌باشد داریم

$$F^{-1}(p)=\lambda_p$$
$$F(\lambda_p)=p \;\Rightarrow\; \frac{\partial}{\partial p}F(\lambda_p)=1$$

بنابراین

$$\frac{\partial}{\partial p}\lambda_p = \frac{\partial\lambda_p}{\partial F(\lambda_p)}\frac{\partial F(\lambda_p)}{\partial p} = \frac{1}{\frac{\partial F(\lambda_p)}{\partial\lambda_p}}\frac{\partial F(\lambda_p)}{\partial p} = \frac{1}{f(\lambda_p)}.$$

قضیه۱: زنگ [۸]، دو رابطهٔ زیر با یکدیگر معادلند.

الف) $h(x;\theta)=u(x)v(\theta)$، که در آن $u(x)$ و $v(\theta)$ توابع مثبت هستند.

ب) $I_{1,\dots,r:n}(\theta)=rI_1(\theta)$، به ازای هر $n\ge 1$ و هر $1\le r\le n$.

برهان الف به **ب** :

در لم ۵ نشان دادیم که اگر تابع نرخ خطر به صورت $h(x;\theta)=u(x)v(\theta)$ باشد، آنگاه $g(\omega;\theta)=I_X(\theta)$ می‌باشد، حال با جایگذاری در (۵) به نتیجهٔ مطلوب خواهیم رسید.

برهان ب به **الف** :

در این قسمت باید نشان دهیم که اگر تساوی $I_{1,\dots,r:n}(\theta)=rI_1(\theta)$ برقرار باشد، تابع نرخ خطر به صورت حاصلضرب دو تابع مثبت، یکی فقط برحسب x و دیگری تنها برحسب θ است. برای این منظور، در لم ۶ نشان دادیم که اگر تساوی بالا برقرار باشد در این‌صورت برای هر $0<p<1$، رابطهٔ زیر برقرار است

$$I_{[\cdot,p]} = pI_{[\cdot,1]}$$

از لم ۲ برای هر $0<p<1$، داریم

$$\int_{-\infty}^{\lambda_p}\left(\frac{\partial}{\partial\theta}\log h(x;\theta)\right)^2 f(x;\theta)\,dx = pI_{[\cdot,1]}.$$

حال با در نظر گرفتن (۱۰)، از طرفین رابطهٔ بالا نسبت به p مشتق می‌گیریم

$$\frac{\partial}{\partial p}\lambda_p\left[\frac{\partial}{\partial\theta}\log h(x;\theta)\right]^{2}\Bigg|_{x=\lambda_p} f(\lambda_p;\theta)$$
$$=I_{[\cdot,1]}(\theta).$$

بنابراین

$$\left(\frac{h'_\theta(\lambda_p;\theta)}{h(\lambda_p;\theta)}\right)^{2}=I_{[\cdot,1]}(\theta).$$

و چون λ_p به ازای هر $\theta\in\Theta$، نسبت به p صعودی است، بـه ازای هر $\theta\in\Theta$ و هر $x\in R^1$، داریم

$$\left[\frac{\partial}{\partial\theta}\log h(x;\theta)\right]^{2}=I_{[\cdot,1]}(\theta).$$

و در نتیجه

$$\left(\frac{h'_\theta(x;\theta)}{h(x;\theta)}\right)^{2}=I_{[\cdot,1]}(\theta),$$

و از آنجایی که سمت راست تساوی بالا تابعی از θ است

$$\frac{\partial}{\partial\theta}\log h(x;\theta)=a(\theta),$$

بنابراین

$$h(x;\theta)=\exp(A(\theta)+c(x))$$
$$=v(\theta)u(x).$$

۷. بحث و نتیجه‌گیری

همانطوری که در بخش ۳ اشاره شد، رفتار تابع نرخ خطر در مباحث قابلیت اعتماد اهمیت ویژه‌ای دارد. بنابراین براساس تجزیهٔ تابع نرخ خطر به‌صورت فوق می‌توانیم با توجه به رفتار $u(x)$، در خصوص IFR، CFR و DFR بودن سیستم بحث کنیم. بدیهی است که بررسی خواص $u(x)$ ساده‌تر از $h(x;\theta)$ می‌باشد.

مثال۱: گوییم X دارای توزیع وایبل با پارامترهای (α,β) است هرگاه دارای تابع توزیع زیر باشد:

$$F(x;\alpha,\beta)=1-\exp(-\alpha x^{\beta}),\quad x>0.$$

اکنون با فرض معلوم بودن β، قرار می‌دهیم $\theta=\alpha$. در این‌صورت X دارای تابع نرخ خطری زیر است:

$$h(x,\theta)=\theta\beta x^{\beta-1},$$

در اینجا $u(x)=\beta x^{\beta-1}$ به ازای $\beta<1$، $\beta=1$ و $\beta>1$ به ترتیب نزولی، ثابت و صعودی است؛ در نتیجه F به ترتیب دارای خواص DFR، CFR و IFR می‌باشد.

مثال۲: گوییم X دارای توزیع بر نوع دوازده با پارامترهای (α,β) است هرگاه دارای تابع توزیع زیر باشد:

$$F(x;\alpha,\beta)=1-(1+x^{\beta})^{-\alpha},\quad x>0$$

اکنون با فرض معلوم بودن β، قرار می‌دهیم $\theta=\alpha$. در این صورت X دارای تابع نرخ خطر زیر است:

$$h(x,\theta)=\beta\theta x^{\beta-1}(1+x^{\beta})^{-1},$$

در اینجا $u(x)=\beta x^{\beta-1}(1+x^{\beta})^{-1}$ به ازای $\beta=1$ نزولی است؛ درنتیجه F دارای خاصیت DFR می‌باشد. همچنین به ازای $\beta=2$ ، F نه DFR است و نه IFR.

۸. تقدیر و تشکر

نویسندگان مقاله از پیشنهادات داوران محترم که باعث بهبود مقاله گردید، تقدیر و تشکر می کنند

مراجع

[۱] بهبودیان، ج. (۱۳۷۰)، آمار ریاضی، مؤسسهٔ انتشارات امیرکبیر.

[2] Arnold, B.C., Balakrishnan, N. and Nagaraja, H. N., 1992, *A first Course in Order Statistics*, John Wiley New York.

[3] Gerstbakh, I., Kagan, A., 1998, Characterization of the weibull distribution by properties of the Fisher information under type-I censoring. *Statist*. *Prob. Lett*. 42, 99-105.

[4] Lawless, J. F., 2003, *Statistical Models and Methods for lifetime Data*, John Wiley. New York.

[5] Nelson, W. B., 2004, *Applied life Data Analysis,* John Wiley, New York.
[6] Park, S., 1996, Fisher Information in Order Statistics. *J. Amer. Stat. Assoc*. 91, 385-390.
[7] Takahashi, H., Sugiura, N., 1989, The rate of convergence of Fisher information for Type II censored sample. *J. Japan. Statist. Soc*. 19, 139-144.
[8] Zheng, G., 2001, A characterization of the factorization of hazard function by the Fisher information under type-II censoring with application to the weibull family. *Statist. Prob. Lett*. 52, 249-253.