فاصله اطمینان برای پارامتر توزیع نمایی تحت سانسور بازه‌ای

محمدحسین پور سعید [۱]

تاریخ دریافت: ۹۹/۲/۲۷

تاریخ پذیرش: ۹۹/۱۱/۱

چکیده:

در این مقاله با در نظر گرفتن داده‌های تحت سانسور بازه‌ای که دارای توزیع نمایی با نرخ خطر θ هستند، روشی برای برآورد بازه‌ای توابعی از θ ارائه می‌شود. همچنین زمان نظارت بهینه و مطالعات شبیه‌سازی موردبررسی قرارگرفته و به کاربردی از مباحث نیز اشاره می‌گردد.

واژه‌های کلیدی: آماره‌های ترتیبی، توزیع نمایی، زمان نظارت بهینه، خاصیت مارکوفی، سانسور بازه‌ای.

۱ مقدمه

در تحلیل مواردی مانند زمان شکست یک واحد صنعتی یا زمان مرگ یک واحد زیستی، از مبحث آماری تحلیل بقاء استفاده می‌شود. اگر X زمان ازکارافتادگی یک پدیده فیزیکی یا زمان وقوع یک واقعه پزشکی باشد، آنگاه مقادیر ممکنه متغیر تصادفی X، نامنفی و داده‌های به‌دست‌آمده نیز موسوم به داده‌های بقاء خواهند بود. در تحلیل داده‌های بقاء، مجموعه‌ای از روش‌های آماری به کار می‌روند که در آن‌ها غالباً از انواع سانسور استفاده می‌شود. برای آشنایی با انواع سانسور می‌توان به لاولس [۵] و نلسون[۷] مراجعه کرد. در اینجا به بعضی از طرح‌های مرتبط با موضوع مقاله اشاره می‌گردد.

فرض کنید θ، پارامتر مجهول یک جامعه آماری است. استنباطی آماری پیرامون θ بر پایه آزمایش اعضای نمونه تصادفی صورت می‌گیرد که در آن، برحسب مورد از طرح‌های مختلف سانسور نیز استفاده می‌شود. ازاین‌رو فرض می‌شود که $X_1,\dots,X_n$ نمونه‌ای تصادفی از جامعه آماری با تابع چگالی $f(x;\theta)$ و تابع توزیع $F(x;\theta)$ است. یکی از رایج‌ترین طرح‌های سانسور، سانسور نوع اول [۲] است که برای کاهش زمان و کاهش هزینه انجام آزمایش مورداستفاده قرار می‌گیرد. در این نوع سانسور، آزمایش در زمان از پیش تعیین‌شده‌ای مانند t پایان می‌یابد و در آن، فقط زمان شکست اعضایی از نمونه که قبل از زمان t خراب می‌شوند، مشاهده و ثبت می‌گردد.

در سانسور نوع اول، با نظارت مداوم بر انجام آزمایش تا زمان t، زمان‌های خرابی مشاهده‌شده‌ای مانند $x_1,\dots,x_y$ ثبت می‌گردند. ازاین‌رو، t زمان پایان آزمایش مقداری معلوم و تعیین‌شده است ولی y، تعداد واحدهایی که در آزمایش خراب می‌شوند و همچنین زمان‌های خرابی، متغیرهای تصادفی هستند و ممکن است که تا پایان آزمایش، هیچ‌یک از واحدهای آزمایشی خراب نشوند. بنابراین در بعضی از موارد، از طرح دیگری که موسوم به سانسور نوع دوم [۳] است نیز استفاده می‌شود. در این طرح نیز با در نظر گرفتن تعداد از پیش تعیین‌شده‌ای مانند r و همچنین نظارت مداوم بر انجام آزمایش، هرگاه r مورد خرابی مشاهده شود، آزمایش پایان می‌یابد. در سانسور نوع دوم، r تعداد واحدهایی که خراب می‌شوند، مقداری معلوم است ولی مدت‌زمان انجام آزمایش و همچنین زمان‌های خرابی، متغیرهایی تصادفی هستند.

در صنعت، مطالعات پزشکی و داده‌هایی که به‌صورت دوره‌ای اندازه‌گیری می‌شوند، با شرایطی روبرو هستیم که دسترسی به زمان دقیق خرابی واحدهای تحت آزمایش مقدور نیست. در مواردی که زمان‌های دقیق خرابی واحدهای تحت آزمون در دسترس نیستند یا نظارت مداوم بر انجام آزمایش میسر نیست، از داده‌های تحت سانسور بازه‌ای [۴] استفاده می‌شود. برای مثال در مراجعات پزشکی، اگر شخصی در اولین ویزیت سالم و در زمان دومین ویزیت بیمار باشد، زمان دقیق بروز بیماری معلوم نخواهد بود و بدیهی است که ابتلا به بیماری در فاصله زمانی بین دو ویزیت رخداده است که در این حالت گفته می‌شود زمان واقعی بروز بیماری به‌صورت بازه‌ای یا فاصله‌ای سانسور شده است.

در حالت کلی، یک آزمایش تحت سانسور بازه‌ای با زمان‌های نظارت چندگانه به‌صورت زیر انجام می‌شود (اسدی، [۱]).

در زمان‌های نظارت از پیش تعیین‌شده‌ای مانند $0 < t_1 < \dots < t_k$

[۱] گروه آمار – دانشگاه لرستان *Poursaeed.m@lu.ac.ir*

[2] Type I censoring

[3] Type II censoring

[4] Interval censoring

واحدهای آزمایشی $X_1,\dots,X_n$ مورد بازرسی قرار می‌گیرند و تعداد واحدهای خراب‌شده در فاصله‌های زمانی $(t_0=0,t_1],(t_1,t_2],\dots,(t_{k-1},t_k]$ ثبت می‌شوند. در این حالت، بدون دانستن زمان خرابی واحدهای آزمایشی، تنها از تعداد واحدهایی که در فاصله‌های زمانی مذکور خراب می‌شوند، اطلاع داریم. بنابراین اگر $y_i,\ i=1,2,\dots,k$ تعداد واحدهای خراب‌شده در بازه $(t_{i-1},t_i]$ باشد، تابع درستنمایی به‌صورت زیر خواهد بود

$$L(\theta\mid y_1,y_2,\dots,y_k)$$
$$\propto (F(t_1;\theta))^{y_1}(F(t_2;\theta)-F(t_1;\theta))^{y_2}$$
$$\cdots(F(t_k;\theta)-F(t_{k-1};\theta))^{y_k}(1-F(t_k;\theta))^{n-\sum_{i=1}^{k}y_i}.$$

توزیع نمایی یکی از توزیع‌هایی است که در مدل بندی داده‌های بقاء کاربرد بسیار گسترده‌ای دارد. فرض کنید $X_1,\dots,X_n$ زمان‌های خرابی n واحد آزمایشی دارای توزیع نمایی با تابع چگالی

$$f(x;\theta)=\theta\exp(-\theta x),\quad x>0,$$

هستند که در آن پارامتر θ نرخ خطر [5] توزیع است. بر پایه داده‌های تحت سانسور نوع اول، تابع درستنمایی به‌صورت

$$L(\theta\mid y_1,y_2,\dots,y_k)$$
$$\propto \theta\exp(-\theta x_1)\times\dots\times\theta\exp(-\theta x_y)\times(\exp(-\theta t))^{n-y}$$

و برآورد ماکسیمم درستنمایی θ نیز به‌صورت

$$\hat{\theta}=\frac{y}{\sum_{i=1}^{y}x_i+(n-y)t},\tag{۱}$$

است که در آن y تعداد خرابی‌ها و $x_1,\dots,x_y$ زمان‌های خرابی مشاهده‌شده تا زمان t هستند. بر پایه داده‌های تحت سانسور نوع دوم نیز برآورد ماکسیمم درستنمایی θ به‌صورت

$$\hat{\theta}=\frac{r}{\sum_{i=1}^{r}x_{(i:n)}+(n-y)x_{(r:n)}},\tag{۲}$$

است که در آن $x_{(1:n)},\dots,x_{(r:n)}$، r زمان خرابی مرتب‌شده هستند.

اگر توزیع زمان‌های خرابی n واحد آزمایشی نمایی باشد و آزمایش نیز تحت سانسور بازه‌ای با زمان‌های نظارت $t_1<t_2<\dots<t_k$ انجام شود، تابع درستنمایی به‌صورت

$$L(\theta\mid y_1,y_2,\dots,y_k)$$
$$\propto(1-\exp(-\theta t_1))^{y_1}(\exp(-\theta t_1)-\exp(-\theta t_2))^{y_2}$$
$$\times\dots\times(\exp(-\theta t_{k-1})-\exp(-\theta t_k))^{y_k}(\exp(-\theta t_k))^{n-\sum_{i=1}^{k}y_i}$$

است که در این صورت، برآورد ماکسیمم درستنمایی θ فرم صریح و بسته‌ای ندارد و برآورد آن بایستی به روش‌های عددی محاسبه شود.

ژانگ و همکاران [۱۰]، حالتی خاص از سانسور بازه‌ای $(k=1)$ را برای داده‌های با توزیع نمایی بررسی کردند که در آن، واحدهای تحت آزمایش در یک زمان از پیش تعیین‌شده مانند $t_1=t$ مورد بازرسی قرارگرفته و فقط تعداد واحدهای خراب‌شده در فاصله $[0,t]$ ثبت می‌شود. در این حالت، تابع درستنمایی به‌صورت

$$L(\theta\mid y_1,y_2,\dots,y_k)\propto(F(t;\theta))^{y}(1-F(t;\theta))^{n-y}$$
$$=(1-\exp(-\theta t))^{y}(\exp(-\theta t))^{n-y}$$

و برآوردگر ماکسیمم درستنمایی θ نیز به‌صورت

$$\hat{\theta}=\frac{-\ln(1-\frac{Y}{n})}{t}.\tag{۳}$$

است که در آن، Y تعداد واحدهایی هستند که در فاصله $[0,t]$ خراب می‌شوند. همان‌گونه که ملاحظه می‌شود، برآوردگر معرفی‌شده در این حالت خاص، فرم بسته‌ای دارد. همچنین، لازم به ذکر است که روش‌های معرفی‌شده در (۱) و (۳)، برآوردهای متفاوتی را ارائه می‌دهند زیرا در (۳) فقط تعداد واحدهای خراب‌شده تا زمان t در دسترس است ولی در (۱)، هم تعداد واحدهای خراب‌شده تا زمان t و هم، زمان‌های خرابی آن‌ها معلوم و مشخص می‌شوند.

در مطالعات مرتبط با سانسور بازه‌ای، تعداد واحدهای آزمایشی، تعداد دفعات نظارت و زمان‌های نظارت بهینه موردبررسی قرارگرفته است که برای کسب اطلاع می‌توان به لاولس [۵] و سان [۸] مراجعه کرد. در [۹]، فاصله‌های زمانی بهینه، بررسی‌شده و برای حالتی که توزیع داده‌ها به‌صورت وایبول باشد نیز مطالعاتی توسط نلسون [۶] انجام‌شده است.

در این مقاله برای حالتی که تعداد دفعات نظارت بیشتر از یک مورد باشد $(k>1)$، با ایده گرفتن از فرم برآوردگر معرفی‌شده در (۳)، نسخه‌ای مشابه، به‌صورت

$$\hat{\theta}=\frac{-\ln(1-\frac{\sum_{i=1}^{k}Y_i}{n})}{t_k},\tag{۴}$$

در نظر گرفته می‌شود که در آن $Y_i: i=1,2,\dots,k$ تعداد واحدهایی هستند که در بازه $(t_{i-1},t_i]$ خراب می‌شوند. برآوردگر معرفی‌شده در (۴)، برآوردگر ماکسیمم درستنمایی θ نیست ولی تابعی از آماره بسنده $(Y_1,\dots,Y_k)$ است که برخلاف برآورد ماکسیمم درستنمایی θ، فرم صریح و بسته‌ای دارد و همان‌گونه که نشان داده می‌شود، برآوردگری سازگار [6] و مجانباً نااریب [7] برای پارامتر است. ازاین‌رو، با توجه به مزایای ذکرشده، به نظر می‌رسد که با استفاده از آن بتوان نسبت به انجام یک استنباط آماری مناسب اقدام نمود.

[5] Failure rate

[6] Consistent

[7] Asymptotically unbiased

بنابراین، در این مقاله برای داده‌های نمایی تحت سانسور بازه‌ای که تعداد دفعات نظارت بیشتر از یک مورد است، با بهره‌گیری از برآوردگر معرفی‌شده در (۴) و تعیین توزیع مجانبی آن، زمان نظارت بهینه و بازه اطمینان مجانبی تعیین می‌شود. همچنین، برآورد بازه‌ای دقیق توابعی از پارامتر نیز موردبررسی قرار می‌گیرد. لازم به ذکر است که روش برآورد زمان نظارت بهینه و همچنین توزیع آماره‌های مورداستفاده در این مقاله و مقاله ژانگ و همکاران متفاوت هستند.

در بخش دوم مقاله، برآورد بازه‌ای توابعی از پارامتر و در بخش سوم، زمان نظارت بهینه و مطالعات شبیه‌سازی بررسی می‌شوند. در بخش پایانی به کاربردی از روش پیشنهادی در برآورد میانگین عمر باقیمانده یک سیستم موازی نیز اشاره می‌شود.

۲ برآورد بازه‌ای توابعی از پارامتر

توزیع نمایی یکی از معروف‌ترین توزیع‌هاست که ثابت بودن نرخ خطر و کاربرد گسترده آن در صنعت و علوم پزشکی، از ویژگی‌های شاخص آن به شمار می‌روند. توزیع نمایی برای توصیف متغیرهای مربوط به طول عمر به کار می‌رود و نرخ خطر، میانگین و چندک‌های [8] آن از شاخص‌های مهم در مباحث آماری هستند که برآورد آن‌ها نیز از اهمیت بالایی برخوردار است. برای برآورد نرخ خطر، میانگین و چندک‌ها، طرح‌ها و روش‌های مختلفی ارائه‌شده است و در کارهای عملی، برای آن‌ها معمولاً از فواصل اطمینان یک‌طرفه استفاده می‌شود (ژانگ و همکاران،[۱۰]).

در تعیین بازه اطمینان برای یک پارامتر، روش‌های مختلفی وجود دارد که برای آشنایی با آن‌ها می‌توان به کسلا و برگر [۴] مراجعه کرد. یکی از روش‌های موجود، روش کمیت محوری [9] است که در آن، تابعی از یک آماره و پارامتر مجهول در نظر گرفته می‌شود به‌گونه‌ای که توزیع آن تابع به پارامتر بستگی نداشته باشد. اگر تابعی از برآوردگر معرفی‌شده در (۴) و θ به‌صورت $\sqrt{n}(\hat{\theta}-\theta)$ در نظر گرفته شود، تعیین توزیع دقیق این تابع میسر نیست ولی می‌توان نشان داد که توزیع مجانبی مضربی از آن به پارامتر مجهول بستگی ندارد.

روش دلتا، یکی از روش‌های تعیین توزیع مجانبی دنباله‌ای از متغیرهای تصادفی است. فرض کنید $\{U_n\}$ دنباله‌ای از متغیرهای تصادفی و همچنین $\sqrt{n}(U_n-\vartheta)$ به‌طور مجانبی دارای توزیع نرمال با میانگین صفر و واریانس σ^{2} است. اگر $h(.)$ تابعی پیوسته و $h'(\vartheta)$ نیز متناهی و غیر صفر باشد آنگاه طبق روش دلتا، توزیع مجانبی $\sqrt{n}(h(U_n)-h(\vartheta))$ به‌صورت نرمال با میانگین صفر و واریانس $(\sigma h'(\vartheta))^{2}$ خواهد بود [۲].

با توجه به این‌که، $\sum_{i=1}^{k} Y_i \equiv R$ معرفی‌شده در (۴) دارای توزیع دوجمله‌ای با پارامترهای n و $\vartheta = 1-e^{-\theta t_k}$ است، توزیع مجانبی $\sqrt{n}(\frac{R}{n}-(1-e^{-\theta t_k}))$ به‌صورت نرمال با میانگین صفر و واریانس $(1-e^{-\theta t_k})e^{-\theta t_k}$ خواهد بود. ازاین‌رو با در نظر گرفتن $U_n=\frac{R}{n}$، $\hat{\theta}=h(U_n)=\frac{-\ln(1-U_n)}{t_k}$ و همچنین استفاده از روش دلتا، توزیع مجانبی $\sqrt{n}(\hat{\theta}-\theta)$ به‌صورت نرمال با میانگین صفر و واریانس $\frac{e^{\theta t_k}-1}{(t_k)^{2}}$ است و توزیع استانداردشده آن نیز به پارامتر θ بستگی ندارد. بنابراین $\hat{\theta}$ برآوردگری سازگار و مجانباً نااریب است که با توجه به توزیع مجانبی به‌دست‌آمده، برای پارامتر مجهول می‌توان یک بازه اطمینان مجانبی معرفی نمود.

قضیهٔ ۱.۲. یک بازه اطمینان مجانبی [10] با ضریب اطمینان $(1-\alpha)100\%$ برای پارامتر θ به‌صورت

$$(0, \hat{\theta}+\frac{z_{1-\alpha}}{t_k}\sqrt{\frac{e^{\hat{\theta} t_k}-1}{n}}), \qquad (5)$$

است که در آن z_α، چندک مرتبه α توزیع نرمال استاندارد می‌باشد.

با استفاده از محوری کردن تابع توزیع [11] نیز می‌توان یک بازه اطمینان دقیق به دست آورد. فرض کنید که متغیر تصادفی گسسته T آماره‌ای است که تابع توزیع آن، $F_T(t;\vartheta)=P(T\le t;\vartheta)$ برای هر مقدار t تابعی نزولی برحسب ϑ می‌باشد. اگر $P(T\le t;\vartheta_{U(t)})=\alpha$ آنگاه طبق روش محوری کردن تابع توزیع، فاصله $(-\infty,\vartheta_{U(t)})$ برای ϑ یک بازه اطمینان با ضریب اطمینان $1-\alpha$ خواهد بود.

از طرفی، اگر متغیر تصادفی S دارای توزیع دوجمله‌ای با پارامترهای n و ϑ باشد آنگاه تابع توزیع آن، $F_S(s;\vartheta)$، برای هر مقدار s تابعی نزولی برحسب ϑ می‌باشد (کسلا و برگر[۴]). ازاین‌رو، تابع توزیع آماره R نیز تابعی نزولی برحسب $\vartheta = 1-e^{-\theta t_k}$ و ϑ نیز تابعی صعودی برحسب θ است. در این صورت $F_R(r;\theta)$، تابع توزیع آماره R برای هر مقدار r تابعی نزولی برحسب θ خواهد بود.

بنابراین، اگر $R=r$ مشاهده شود آنگاه با حل معادله $P(R\le r;\theta_{U(r)})=P(R<r+1;\theta_{U(r)})=\alpha$ و تعیین $\theta_{U(r)}$، بازه اطمینان مطلوب به دست می‌آید. برای تعیین $\theta_{U(r)}$، آماره مرتب mام از نمونه تصادفی n تایی $X_1,\ldots,X_n$ با $X_{(m:n)}$ نشان داده می‌شود که در این صورت، دو پیشامد

[8] Quantiles
[9] Pivotal quantity
[10] Asymptotic confidence interval
[11] Pivoting the CDF

$(R<r+۱)$ و $(X_{(r+۱:n)}>t_k)$ باهم معادل هستند. بنابراین

$$\begin{aligned}\alpha &= P(R\leq r;\theta_U(r)) = P(X_{(r+۱:n)}>t_k;\theta_U(r))\\ &= P(F(X_{(r+۱:n)})>F(t_k;\theta_U(r)))\\ &= P(U_{(r+۱:n)}>F(t_k;\theta_U(r)))\end{aligned}$$

که در آن $U_{(m:n)}$ نیز آماره مرتب m ام از یک نمونه تصادفی n تایی با توزیع یکنواخت روی $(۰,۱)$ است.

با توجه به رابطه بین توزیع‌های بتا و F و این‌که $U_{(r+۱:n)}$ دارای توزیع بتا با پارامترهای $r+۱$ و $n-r$ است، روابط

$$\begin{aligned}&U_{(r+۱:n)}\sim Beta(r+۱,n-r)\\ &\Longrightarrow \frac{(n-r)U_{(r+۱:n)}}{(r+۱)(۱-U_{(r+۱:n)})}\sim F(۲(r+۱),۲(n-r))\end{aligned}$$

و

$$\begin{aligned}۱-\alpha &= P(U_{(r+۱:n)}\leq F(t_k;\theta_U(r)) = ۱-e^{-\theta_U(r)t_k})\\ &= P(\frac{(n-r)U_{(r+۱:n)}}{(r+۱)(۱-U_{(r+۱:n)})}\leq \frac{(n-r)(۱-e^{-\theta_U(r)t_k})}{(r+۱)(e^{-\theta_U(r)t_k})})\\ &= P(F(۲(r+۱),۲(n-r))\leq \frac{(n-r)(۱-e^{-\theta_U(r)t_k})}{(r+۱)(e^{-\theta_U(r)t_k})})\\ &\Rightarrow \frac{(n-r)(۱-e^{-\theta_U(r)t_k})}{(r+۱)(e^{-\theta_U(r)t_k})} = f_{۲(r+۱),۲(n-r),۱-\alpha}\\ &\Rightarrow \theta_U(r) = \frac{۱}{t_k}\ln\left(۱+\frac{(r+۱)f_{۲(r+۱),۲(n-r),۱-\alpha}}{(n-r)}\right)\end{aligned}$$

برقرار هستند. بنابراین با توجه به روش محوری کردن تابع توزیع، نتیجه به‌دست‌آمده و همچنین مثبت بودن نرخ خطر θ، بازه اطمینان مطلوب معرفی می‌شود.

قضیهٔ ۲.۲. یک بازه اطمینان $(۱-\alpha)۱۰۰\%$ یک‌طرفه برای پارامتر θ به‌صورت

$$(۰,\frac{۱}{t_k}\ln\left(۱+\frac{(r+۱)f_{۲(r+۱),۲(n-r),۱-\alpha}}{(n-r)}\right)), \tag{۶}$$

است که در آن $r=\sum_{i=۱}^{k} y_i$ و $f_{m_۱,m_۲,\alpha}$ چندک مرتبه α توزیع F با درجات آزادی $m_۱$ و $m_۲$ می‌باشد.

نتیجه ۳.۲. فاصله‌های اطمینان $(۱-\alpha)۱۰۰\%$ یک‌طرفه برای میانگین و چندک مرتبه γ توزیع، به ترتیب به‌صورت

$$(\frac{t_k}{\ln\left(۱+\frac{rf_{۲r,۲(n+۱-r),۱-\alpha}}{(n+۱-r)}\right)},+\infty), \tag{۷}$$

$$(\frac{-t_k\ln(۱-\gamma)}{\ln\left(۱+\frac{rf_{۲r,۲(n+۱-r),۱-\alpha}}{(n+۱-r)}\right)},+\infty) \tag{۸}$$

هستند.

۳ زمان نظارت بهینه

تعیین بهترین زمان‌های بازرسی، یکی از موضوع‌های اساسی در مطالعه داده‌های تحت سانسور بازه‌ای است. ازاین‌رو با توجه به ویژگی‌های $\hat{\theta}$ معرفی‌شده در (۴)، در این بخش، تعیین آخرین زمان نظارت بهینه بررسی می‌شود. نظر به نااریبی مجانبی $\hat{\theta}$ و واریانس مجانبی آن‌که برابر با $\frac{e^{\theta t_k}-۱}{n(t_k)^۲}$ است، مقدار t_k به‌گونه‌ای تعیین می‌شود تا واریانس $\hat{\theta}$ مینیمم شود. بنابراین با در نظر گرفتن $g(t_k):=\frac{e^{\theta t_k}-۱}{n(t_k)^۲}$،

$$\frac{dg(t_k)}{dt_k}\propto t_k e^{-\theta t_k}[\theta t_k+۲e^{-\theta t_k}-۲], \tag{۹}$$

و با انجام محاسبات می‌توان نشان داد که ریشه تقریبی (۹) برابر با $t_k\cong\frac{\pi}{۲\theta}$ است. ازاین‌رو برای نمونه‌های بزرگ، اگر آخرین زمان نظارت به‌صورت $t_k\cong\frac{\pi}{۲\theta}$ انتخاب شود آنگاه برآوردگر، نااریب و نیز دارای کمترین واریانس خواهد بود.

برای بررسی تأثیر مقدار آخرین زمان نظارت بر رفتار برآوردگر، به ازای برخی مقادیر θ، نمونه‌های تصادفی n تایی از توزیع نمایی تولید و با در نظر گرفتن زمان‌های مختلف برای آخرین زمان نظارت، θ برآورد می‌شود (همان‌گونه که ملاحظه می‌شود، ضابطه (۴) فقط به آخرین زمان نظارت بستگی دارد. ازاین‌رو، در نظر گرفتن مقادیر مختلفی برای سایر زمان‌های نظارت ضروری نیست). با انجام شبیه‌سازی و $N=۱۰۰۰۰$ بار تکرار، $Mean=\frac{\sum_{i=۱}^{۱۰۰۰۰}\hat{\theta}_i}{۱۰۰۰۰}$ و $SSE=\sum_{i=۱}^{۱۰۰۰۰}(\hat{\theta}_i-\theta)^۲$ محاسبه‌شده که نتایج در جدول ۱ آمده است.

نکته ۱. نتایج به‌دست‌آمده در جدول ۱ نیز نشان می‌دهند که با انتخاب $\frac{\pi}{۲\theta}$ به‌عنوان آخرین زمان بازرسی، برآوردگر کمترین واریانس را دارد. لازم به ذکر است که در کارهای عملی، به دلیل مجهول بودن پارامتر، تعیین $\frac{\pi}{۲\theta}$ مقدور نیست. ازاین‌رو می‌توان ابتدا پارامتر θ را توسط $\widehat{\theta^*}=\frac{-\ln(۱-\frac{\sum_{j=۱}^{k-۱}Y_j}{n})}{t_{k-۱}}$ برآورد کرد و سپس تخمینی از آخرین زمان بازرسی بهینه را به‌صورت $\widehat{t_k}=\frac{\pi}{۲\widehat{\theta^*}}$ در نظر گرفت. در این صورت، با بهره‌گیری از تخمین آخرین زمان بازرسی بهینه،

$$\widehat{\widehat{\theta^*}}=\frac{-\ln(۱-\frac{\sum_{j=۱}^{k}Y_j}{n})}{\widehat{t^*}}=\frac{۲\ln(۱-\frac{\sum_{j=۱}^{k-۱}Y_j}{n})\ln(۱-\frac{\sum_{j=۱}^{k}Y_j}{n})}{\pi t_{k-۱}} \tag{۱۰}$$

برآورد مجددی از پارامتر را ارائه می‌دهد.

لازم به ذکر است که اگر شکست همه واحدهای آزمایشی تا زمان $t_{k-۱}$ مشاهده شود، پارامتر را نمی‌توان با استفاده از $\widehat{\theta^*}=\frac{-\ln(۱-\frac{\sum_{j=۱}^{k-۱}Y_j}{n})}{t_{k-۱}}$ برآورد کرد و از طرفی اگر $\widehat{t_k}\leq t_{k-۱}$ باشد، نظارت بعدی نیز منتفی خواهد بود. ازاین‌رو، اصلاح‌شده (۱۰) به‌صورت زیر در نظر گرفته می‌شود.

$$\widehat{\widehat{\theta^*}} = \begin{cases} \frac{\ln(1-\frac{\sum_{j=1}^{k-1} Y_j}{n+1})}{t_{k-1}}, & 1 \leq \sum_{j=1}^{k-1} \leq (n+1)(1-e^{-\frac{\pi}{2}}) \\ \frac{2\ln(1-\frac{\sum_{j=1}^{k-1} Y_j}{n})\ln(1-\frac{\sum_{j=1}^{k} Y_j}{n})}{\pi t_{k-1}}, & (n+1)(1-e^{-\frac{\pi}{2}}) < \sum_{j=1}^{k-1} Y_j \end{cases}$$

نکته ۲. لازم به ذکر است که در مبحث سانسور بازه‌ای، زمان‌های نظارت $t_1, t_2, \ldots, t_k$ مقادیری ثابت و از پیش تعیین‌شده هستند، ازاین‌رو در فواصل اطمینان معرفی‌شده (۶)، (۷) و (۸)، نمی‌توان به‌جای $\hat{\theta}$ از $\widehat{\widehat{\theta^*}}$ استفاده کرد.

۴ کاربردی از مباحث

قابلیت اعتماد در بالا بردن سطح استاندارد و کیفیت تجهیزات و تولیدات صنعتی سهم بسزایی دارد که یکی از مباحث مهم در آن، میانگین عمر باقیمانده [12] است. آماره‌های ترتیبی [13] نیز جایگاه ویژه‌ای در مطالعات طول عمر و تحلیل‌های آماری دارند و در قابلیت اعتماد نقشی اساسی ایفا می‌کنند. فرض کنید که سیستمی با طول عمر T در زمان $t = 0$ مورداستفاده قرارگرفته و در زمان $t > 0$ سیستم هنوز فعال است. میانگین عمر باقیمانده سیستم را، به‌عنوان تابعی از t با $m_T(t)$ نشان می‌دهند که به‌صورت $m_T(t) = E(T - t \mid T > t)$ تعریف می‌شود.

در مهندسی قابلیت اعتماد، سیستم‌هایی با ساختارهای مختلف در نظر گرفته می‌شوند که سیستم‌های متوالی [14]، موازی [15] و m از n [16] از مهم‌ترین آن‌ها به شمار می‌روند. برای اطلاع بیشتر می‌توان به اسدی [۱] و بارلو و پروشان [۳] مراجعه کرد.

یک سیستم شامل n مؤلفه را یک سیستم m از n گوییم، هرگاه فعال بودن آن مستلزم فعال بودن حداقل m مؤلفه آن باشد $(m \leq n)$. سیستم‌های موازی و متوالی حالت‌هایی خاص از سیستم m از n هستند که در آن‌ها m به ترتیب برابر با ۱ و n است.

فرض کنید $X_1, \ldots, X_n$ طول عمر مؤلفه‌های یک سیستم موازی هستند، در این صورت طول عمر آن برابر با $X_{n:n}$، یعنی بزرگ‌ترین آماره مرتب متناظر با طول عمر مؤلفه‌ها است و اگر سیستم در زمانی مانند t هنوز فعال باشد، آنگاه $m_{X_{(n:n)}}(t) = E(X_{(n:n)} - t \mid X_{(n:n)} > t)$ تابع میانگین عمر باقیمانده این سیستم در زمان t خواهد بود. بنابراین در زمان نظارت t فقط می‌دانیم که بعضی از مؤلفه‌های سیستم هنوز سالم هستند و از زمان‌های شکست مؤلفه‌های خراب‌شده نیز هیچ‌گونه اطلاعی نداریم. ازاین‌رو، میانگین عمر باقیمانده سیستم نیز به‌نوعی، مرتبط با طرح سانسور بازه‌ای محسوب می‌شود.

فرض کنید n مؤلفه یک سیستم موازی دارای توزیع نمایی با نرخ خطر θ هستند که مستقل از هم کار می‌کنند و مشابه طرح سانسور بازه‌ای، فقط در زمان‌هایی مانند $t_1, \ldots, t_k$ بر عملکرد سیستم نظارت کرده و فرض می‌شود که در آخرین زمان نظارت نیز سیستم فعال است. بنابراین، برای میانگین عمر باقیمانده سیستم در زمان t_k یک کران پایین با ضریب اطمینان $\%100(1-\alpha)$ معرفی می‌شود.

اگر $(y_1, \ldots, y_k)$ تعداد واحدهای خراب‌شده در فاصله‌های زمانی $(t_0 = 0, t_1], (t_1, t_2], \ldots, (t_{k-1}, t_k]$ باشند، آنگاه پیشامدهای

$$A = (X_{(y_1:n)} < t_1 < X_{(y_1+1:n)} < \cdots < t_{k-1} < X_{(\sum_{i=1}^{k} y_i:n)} < t_k < X_{(\sum_{i=1}^{k} y_i+1:n)} \leq X_{(n:n)})$$

و

$$B = (Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k)$$

باهم معادل هستند و رابطه

$$m_{X_{(n:n)}}(t_k) = E((X_{(n:n)} - t_k) \mid A)$$

برقرار است. با توجه به خاصیت مارکوفی [17] آماره‌های ترتیبی و خاصیت عدم حافظه [18] در توزیع نمایی می‌توان نشان داد که $(X_{(n:n)} - t_k) \mid A$ و $Max\{X_1, \ldots, X_{n-\sum_{i=1}^{k} y_i}\}$ دارای توزیع یکسان هستند. ازاین‌رو، با توجه به [۳]،

$$m_{X_{(n:n)}}(t_k) = E\left(X_{(n-\sum_{i=1}^{k} y_i:n-\sum_{i=1}^{k} y_i)}\right) = \frac{c}{\theta} \qquad (11)$$

برقرار است که در آن $c = \left(\frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{n - \sum_{i=1}^{k} y_i}\right)$.
بنابراین با ضرب کردن مقدار c در کران پایین به‌دست‌آمده در (۷)، کران پایین مطلوب به دست می‌آید.

[12] Mean residual life
[13] Order statistics
[14] Series system
[15] Parallel system
[16] m out n system
[17] Markovian property
[18] Memoryless property

برای مثال، فرض کنید که یک نمونه ۵۰ تایی از لامپ‌هایی که دارای توزیع نمایی هستند را مورد آزمایش قرار داده و بر پایه سه بار نظارت در زمان‌های $t_1 = 100$، $t_2 = 300$، $t_3 = 540$ (برحسب ساعت)، تعداد لامپ‌های خراب‌شده در فاصله‌های $[0,100)$، $[100,300)$ و $[300,540)$ به‌صورت

$$y_1 = 17, \quad y_2 = 21, \quad y_3 = 8.$$

به‌دست‌آمده است (با توجه به داده‌های صفحه ۱۶۲، اسدی [۱]). بنابراین،

الف) با توجه به (۴) و

$$k = 3, t_k = 540, r = \sum_{i=1}^{3} y_i = 46,$$

$$\hat{\theta} = \frac{-\ln(1 - \frac{46}{50})}{540} = 0.00468, \quad \frac{1}{\hat{\theta}} = 213.8$$

برآوردی از میانگین طول عمر لامپ‌ها، به‌صورت ۲۱۳٫۸ ساعت است.

ب) با توجه به (۷) و

$$f_{2r,2(n+1-r),1-\alpha} = f_{92,10,0.95} \cong 2.6,$$

$$\Big(\frac{540}{\ln(1 + \frac{46 \times 2.6}{(50+1-46)})}, +\infty\Big) = (167.93, +\infty)$$

یک بازه اطمینان ۹۵٪ برای میانگین طول عمر لامپ‌ها به‌صورت $(167.93, +\infty)$ خواهد بود. ضمناً با توجه به (۵)، بازه اطمینان مجانبی برای میانگین طول عمر لامپ‌ها به‌صورت $(162.81, +\infty)$ است.

پ) نظر به این‌که ۵۰ لامپ را می‌توان به‌عنوان مؤلفه‌های یک سیستم در نظر گرفت، طول عمر آخرین لامپ را نیز می‌توان متناظر با طول عمر سیستم موازی با ۵۰ مؤلفه دانست. ازاین‌رو با توجه به (۱۱) که برای تعیین بازه اطمینان میانگین عمر باقیمانده یک سیستم موازی معرفی‌شده است و همچنین با توجه به نتایج به‌دست‌آمده در قسمت ب داریم:

$$\begin{aligned} c &= \Big(\frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{n - \sum_{i=1}^{k} y_i}\Big) \\ &= \Big(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\Big) = \frac{25}{12}, \\ \frac{25}{12} &\times (167.93, +\infty) = (350, +\infty), \end{aligned}$$

یک فاصله پیش‌بینی ۹۵٪ برای باقیمانده عمر آخرین لامپ به‌صورت $(350, +\infty)$ است. در این صورت، با ۹۵٪ اطمینان، انتظار می‌رود که آخرین لامپ لااقل ۳۵۰ ساعت دیگر کار کند.

جدول ۱. تأثیر زمان‌های نظارت مختلف بر رفتار برآوردگر معرفی‌شده در (۴)

θ	t	n	$Mean$	SSE	θ	t	n	$Mean$	SSE
۰٫۰۱	۱۳۰	۲۰	۰٫۰۰۹۴۹	۰٫۰۷۳۷۴	۰٫۰۵	۲۰	۶۰	۰٫۰۴۹۲۵	۰٫۷۰۴۸۰
”	$\frac{\pi}{2\theta} \cong 157$	”	۰٫۰۰۹۳۶	۰٫۰۶۶۳۹	”	$\frac{\pi}{2\theta} \cong 31$	”	۰٫۰۴۹۰۶	۰٫۶۳۴۶۸
”	۱۸۵	”	۰٫۰۰۹۲۱	۰٫۰۶۰۳۷	”	۴۰	”	۰٫۰۴۸۶۸	۰٫۶۲۶۰۸
”	۱۳۰	۴۰	۰٫۰۰۹۷۷	۰٫۰۳۶۴۸	”	۲۰	۸۰	۰٫۰۴۹۴۶	۰٫۵۳۷۰۵
”	$\frac{\pi}{2\theta} \cong 157$	”	۰٫۰۰۹۷۱	۰٫۰۳۵۷۰	”	$\frac{\pi}{2\theta} \cong 31$	”	۰٫۰۴۹۴۸	۰٫۴۷۷۴۷
”	۱۸۵	”	۰٫۰۰۹۶۰	۰٫۰۳۶۸۳	”	۴۰	”	۰٫۰۴۸۹۵	۰٫۴۸۸۸۳
”	۱۳۰	۶۰	۰٫۰۰۹۸۴	۰٫۰۲۵۷۹	۰٫۰۷	۲۰	۲۰	۰٫۰۶۶۱۶	۰٫۴۹۷۵۲
”	$\frac{\pi}{2\theta} \cong 157$	”	۰٫۰۰۹۷۸	۰٫۰۲۵۲۲	”	$\frac{\pi}{2\theta} \cong 22.5$	”	۰٫۰۶۵۵۸	۳٫۲۱۷۰۰
”	۱۸۵	”	۰٫۰۰۹۷۶	۰٫۰۲۵۴۷	”	۲۵	”	۰٫۰۶۵۲۷	۳٫۱۷۰۱۴
”	۱۳۰	۸۰	۰٫۰۰۹۸۸	۰۱۹۱۷	”	۲۰	۴۰	۰٫۰۶۸۲۳	۱٫۸۴۰۹۴
”	$\frac{\pi}{2\theta} \cong 157$	”	۰٫۰۰۹۹۱	۰٫۰۱۸۷۶	”	$\frac{\pi}{2\theta} \cong 22.5$	”	۰٫۰۶۷۵۴	۱٫۷۹۰۸۲
”	۱۸۵	”	۰٫۰۰۹۸۱	۰٫۰۱۸۷۹	”	۲۵	”	۰٫۰۶۷۷۴	۱٫۷۸۳۴۶
۰٫۰۵	۲۰	۲۰	۰٫۰۴۷۹۶	۰٫۰۴۲۱۶	”	۲۰	۶۰	۰٫۰۶۸۸۴	۱٫۲۶۳۰۴
”	$\frac{\pi}{2\theta} \cong 31$	”	۰٫۰۴۷۲۰	۱٫۷۲۳۰۳	”	$\frac{\pi}{2\theta} \cong 22.5$	”	۰٫۰۶۸۳۹	۱٫۲۳۴۴۹
”	۴۰	”	۰٫۰۴۵۷۷	۱٫۵۰۷۱	”	۲۵	”	۰٫۰۶۸۳۸	۱٫۱۹۸۸۴
”	۲۰	۴۰	۰٫۰۴۸۹۳	۱٫۰۰۷۸	”	۲۰	۸۰	۰٫۰۶۹۰۲	۰٫۹۲۰۱۹
”	$\frac{\pi}{2\theta} \cong 31$	”	۰٫۰۴۸۳۴	۰٫۸۹۳۹۰	”	$\frac{\pi}{2\theta} \cong 22.5$	”	۰٫۰۶۹۰۶	۰٫۹۱۶۸۰
”	۴۰	”	۰٫۰۴۸۰۳	۰٫۹۰۳۷۷	”	۲۵	”	۰٫۰۶۸۶۸	۰٫۹۴۶۶۳

۵ نتیجه‌گیری

بر پایه داده‌های تحت سانسور بازه‌ای، برآورد نقطه‌ای و فاصله‌ای پارامتر توزیع نمایی موردبررسی قرار گرفت. برآوردگر نقطه‌ای معرفی‌شده در این مقاله، تابعی از آماره بسنده است و فرم بسته‌ای دارد به‌گونه‌ای که تعیین توزیع مجانبی آن و تعیین فاصله‌های اطمینان مجانبی و دقیق مناسب برای پارامتر مجهول و همچنین تعیین زمان نظارت بهینه را میسر کرده است. با توجه به نتایج به‌دست‌آمده از مطالعات شبیه‌سازی، بهینه بودن زمان نظارت پیشنهادی مستلزم بزرگ بودن اندازه نمونه است که در این صورت، تحت زمان نظارت بهینه، اریبی و واریانس برآوردگر کمتر می‌شود و در واقع با انتخاب زمان نظارت بهینه، کم برآوردی نرخ خطر یا بیش برآوردی میانگین توزیع کاهش می‌یابد.

مراجع

[۱] مجید اسدی، (۱۳۹۲)، آشنایی با نظریه قابلیت اعتماد، مرکز نشر دانشگاهی، تهران.

[2] Agresti, A., (1990), *Categorical data analysis*, John Wiley & Sons, New York.

[3] Barlow, R. E. and Proschan, F., (1975), *Statistical Theory of Reliability and Life Testing*, Holt, Rinehart and Winston, New York.

[4] Casella, G. and Berger, R. L. (1990), *Statistical inference*, Wadsworth & Brooks/Cole.

[5] Lawless, J. F., (2003), *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.

[6] Nelson, W. (1985), *Weibull analysis of reliability data with few or no failures, Journal of Quality Technology*, **17**, 140–146.

[7] Nelson, W. (2004), *Applied life data analysis*, John Wiley & Sons, New York.

[8] Sun J., (2007), *The Statistical Analysis of interval-censored failure time data*, Springer, Science & Business Media.

[9] William, Q. and Meeker, Jr., (1986), Planning life tests in which units are inspected for failure, *IEEE Trans.* Reliability, **35**, 571-578.

[10] Zhang, C.W., Zhang, T., Xu, D. and Xie, M., (2013), Analyzing highly censored reliability data without exact failure times: An efficient tool for practitioners, *Quality Engineering*, **25**, 392-400.