

برآوردیاب سازگار پارامتر توزیع یکنواخت تحت سانسور بازه‌ای

محمدحسین بورسعيد^۱

تاریخ دریافت: ۱۳۹۷/۵/۲۵

تاریخ پذیرش: ۱۳۹۷/۱۱/۷

چکیده:

داده‌های سانسور شده در آزمون‌های آماری و برآورد پارامترها کاربرد گسترده‌ای دارند. در نوعی از داده‌های بقا که امکان اندازه‌گیری زمان وقوع پیشامد وجود ندارد، از داده‌های تحت سانسور بازه‌ای استفاده می‌شود. در این مقاله با فرض آن که داده‌های تحت سانسور بازه‌ای در فاصله (θ, ∞) به‌طور یکنواخت توزیع شده‌اند، برآوردگری سازگار برای θ معرفی می‌شود. سپس، یک فاصله اطمینان مجانبی برای میانگین طول عمر سیستمی موازی با مؤلفه‌هایی که دارای توزیع یکنواخت هستند، ارائه می‌گردد. **واژه‌های کلیدی:** آماره‌های ترتیبی، سازگاری، سانسور بازه‌ای، سیستم‌های k از n ، میانگین باقیمانده عمر.

۱ مقدمه

درست‌نمایی به‌صورت زیر خواهد بود:

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

نظر به این که t_i ها، زمان‌های نظارت هستند و از زمان دقیق شکست واحدهای آزمایشی اطلاعی نداریم، لذا نمایشی دیگر از رابطه (۱) که به تابع شبه‌درست‌نمایی^۲ موسوم است، به‌صورت زیر در نظر گرفته می‌شود:

$$L^*(\theta|y_1, y_2, \dots, y_k) \propto (F_\theta(\tilde{t}_1))^{y_1} (F_\theta(\tilde{t}_2))^{y_2} \dots \\ \times (F_\theta(\tilde{t}_{k-1}))^{y_k} (1 - F_\theta(t_k))^{n - \sum_{i=1}^k y_i}$$

که در آن $\tilde{t}_i$ یکی از نقاط کرانی بازه (t_{i-1}, t_i) یا نقطه میانی آن است.

کاربرد وسیع توزیع‌های نمایی، گاما و وایبل، سبب شده که عموم مطالعات پارامتری مرتبط با داده‌های تحت سانسور بازه‌ای، مربوط به این توزیع‌ها باشند. اودل و دیگران [۷] مطالعاتی مرتبط با توزیع وایبل انجام دادند و پتو [۸] روشی را برای برآورد تابع توزیع تجمعی و فینکلشتاین و ولف [۳] روش رگرسیونی را

در بررسی پدیده‌های تصادفی، به‌کارگیری روش‌های آماری اجتناب‌ناپذیر است و در تحلیل داده‌های بقا که امکان اندازه‌گیری زمان وقوع پیشامد وجود ندارد و یا این که ثبت زمان وقوع آنها وقت گیر و یا هزینه‌بردار است، از داده‌های تحت سانسور بازه‌ای استفاده می‌شود.

فرض کنید $T_1, \dots, T_n$ نمونه تصادفی از جمعیتی با پارامتر مجهول θ ، تابع چگالی $f_\theta(t)$ و تابع توزیع $F_\theta(t)$ است. جهت استنباطی آماری از θ ، آنها در آزمونی وارد می‌شوند که در انجام آزمایش تحت سانسور بازه‌ای، به‌صورت زیر عمل می‌شود [۱].

در زمان‌های از پیش تعیین‌شده‌ای مانند $t_0 < t_1 < \dots < t_k$ ، واحدهای آزمایشی را مورد بازرسی قرار داده و تعداد واحدهای خراب‌شده در فواصل $(t_0 = 0, t_1), (t_1, t_2), \dots, (t_{k-1}, t_k)$ را ثبت می‌شود که در این حالت، فقط از تعداد واحدهای خراب‌شده در بازه‌ها اطلاع داریم. بنا بر این اگر $i = 1, 2, \dots, k$: y_i تعداد واحدهای خراب‌شده در بازه (t_{i-1}, t_i) باشد، آن‌گاه تابع

^۱ هیأت علمی گروه آمار، دانشگاه لرستان، ایران

^۲ Pseudo-Likelihood

^۳ Under Estimate

۱.۲ برآورد ماکسیمم درست‌نمایی θ ، زمانی که

$$1 \leq \sum_{i=1}^k y_i \leq n$$

فرض کنید برآورد ماکسیمم درست‌نمایی θ را با $\hat{\theta}$ نشان دهیم،

لم ۱.۲ الف) اگر $\sum_{i=1}^{k-1} y_i = 0$ و $y_k = n$ ، آن‌گاه $\hat{\theta} = t_k$.

ب) اگر $1 \leq \sum_{i=1}^k y_i < n$ ، آن‌گاه $\hat{\theta} = \frac{nt_k}{\sum_{i=1}^k y_i}$.

اثبات. با توجه به توابع درست‌نمایی زیر، درستی قسمت‌های الف و ب بدیهی است.

$$L(\theta | \sum_{i=1}^{k-1} y_i = 0, y_k = n) \propto \left(\frac{\theta - t_{k-1}}{\theta}\right)^n I(t_{k-1} \leq \theta < t_k) + \left(\frac{t_k - t_{k-1}}{\theta}\right)^n I(t_k \leq \theta),$$

$$L(\theta | 1 \leq \sum_{i=1}^k y_i < n) \propto \left(\frac{t_k}{\theta}\right)^{\sum_{i=1}^k y_i} \left(1 - \frac{t_k}{\theta}\right)^{n - \sum_{i=1}^k y_i} \times I(t_k \leq \theta).$$

□

لم ۲.۲ اگر $\sum_{i=1}^k y_i = n$ ، $1 \leq y_k < n$ ، $1 \leq \sum_{i=1}^{k-1} y_i < n$ ، آن‌گاه $\hat{\theta} = \min \left\{ t_k, \frac{nt_{k-1}}{\sum_{i=1}^{k-1} y_i} \right\}$.

اثبات.

$$L(\theta | \sum_{i=1}^k y_i = n, 1 \leq \sum_{i=1}^{k-1} y_i < n, 1 \leq y_k < n) \propto \begin{cases} \left(\frac{t_{k-1}}{\theta}\right)^{\sum_{i=1}^{k-1} y_i} \left(\frac{\theta - t_{k-1}}{\theta}\right)^{y_k} & t_{k-1} \leq \theta < t_k \\ \left(\frac{t_{k-1}}{\theta}\right)^{\sum_{i=1}^{k-1} y_i} \left(\frac{t_k - t_{k-1}}{\theta}\right)^{y_k} & t_k \leq \theta \end{cases}$$

نظر به رفتار ضابطه اول و نزولی بودن ضابطه دوم، نتیجه مطلوب

□

به دست می‌آید.

۲.۲ برآورد θ ، زمانی که $\sum_{i=1}^k y_i = 0$

با توجه به رابطه (۱)، تابع درست‌نمایی به‌ازای $\sum_{i=1}^k y_i = 0$ به‌صورت زیر است:

$$L(\theta | y_1, y_2, \dots, y_k) = L(\theta | 0, 0, \dots, 0) = (1 - F_\theta(t_k))^n = \left(1 - \frac{t_k}{\theta}\right)^n I(t_k < \theta),$$

جهت برآورد پارامترها معرفی کردند. همچنین، لیندزی و رایان [۶] نشان دادند که برآوردیاب حاصل از به‌کارگیری تابع شبه‌درست‌نمایی، تمایل به کم‌برآوردی^۳ پارامترها دارد. جهت اطلاع بیشتر می‌توان به کتابی که اخیراً توسط سان [۱۰] تألیف شده، رجوع کرد که در آن، مدل‌ها و روش‌های آماری مختلفی برای تحلیل داده‌های تحت سانسور پیشنهاد شده است.

در تحلیل داده‌های بقا، توزیع نمایی کاربرد بسیار گسترده و توزیع یکنواخت به‌عنوان یک توزیع طول عمر، کاربرد کمتری دارد. با این حال در بعضی از موارد، طول عمر مؤلفه‌های یک سیستم به‌صورت یکنواخت در نظر گرفته می‌شود. جهت اطلاع بیشتر، مبحث توزیع طول عمر سیستم‌های منسجم و تمرینات مربوط به آن در [۱] ملاحظه شوند.

لذا در این مقاله فرض می‌شود که داده‌های تحت سانسور در فاصله $(0, \theta)$ به‌طور یکنواخت توزیع شده‌اند. در بعضی از حالات، برآورد θ را به روش ماکسیمم درست‌نمایی به دست می‌آید و در حالتی که آن روش کارآیی ندارد، با معرفی یک برآورد و تلفیق آن با برآورد ماکسیمم درست‌نمایی، برآوردگر نهایی پیشنهاد می‌شود.

بنا بر این، برآوردگر θ در بخش ۲ معرفی و در بخش ۳ سازگاری^۴ آن ثابت می‌شود. در بخش ۴ زمان‌های نظارت و مطالعات شبیه‌سازی بررسی می‌شوند. در بخش پایانی نیز به کاربردی از روش پیشنهادی در برآورد میانگین باقیمانده عمر یک سیستم موازی اشاره می‌شود.

۲ معرفی برآوردگر

فرض کنید $T_1, \dots, T_n$ زمان‌های خرابی n واحد آزمایشی هستند که در فاصله $(0, \theta)$ به‌طور یکنواخت توزیع شده‌اند و در زمان‌های از پیش تعیین‌شده‌ای مانند $t_1 < t_2 < \dots < t_k$ مورد بازرسی قرار می‌گیرند به‌طوری که $y_i : i = 1, 2, \dots, k$ تعداد واحدهای خراب‌شده در بازه (t_{i-1}, t_i) است. برآوردگر در حالت‌های مختلف و در قالب لم‌های زیر به دست می‌آید.

^۴ Consistency

که در این حالت برآورد ماکسیمم درست‌نمایی θ وجود ندارد.

با توجه به این که برای داده‌های کامل و سانسورنشده، روابط $E(T_{n:n}) = nE(T_{1:n})$ و $\hat{\theta} = T_{n:n}$ برقرار است و همچنین،

قسمت ب لم ۱.۲ متناسب بودن تعداد خرابی‌ها را با فاصله زمانی نشان می‌دهد، می‌توان با کمی اغماض، $nT_{1:n}$ را نیز برآوردگری مناسب برای θ دانست. ولی با توجه به این که در سانسور بازه‌ای، مشاهده زمان خرابی واحدهای آزمایشی مقدور نیست لذا هرگاه $\sum_{i=1}^k y_i = 0$ ، اولین زمان خرابی با محتمل‌ترین زمان، تخمین زده می‌شود. بنا بر این، با توجه به

$$\text{Mode}(T_{1:n}|Y_1 = 0, \dots, Y_k = 0)$$

$$= \text{Mode}(T_{1:n}|t_k < T_{1:n}) = t_k$$

مقدار $T_{1:n}$ را توسط t_k تخمین زده که در این حالت nt_k نیز برآوردی از θ خواهد بود. با تلفیق نتایج به دست آمده، برآوردگر θ به صورت زیر معرفی می‌شود:

(۲)

$$\tilde{\theta} = \begin{cases} nt_k & \sum_{i=1}^k Y_i = 0 \\ \frac{nt_k}{\sum_{i=1}^k Y_i} & 1 \leq \sum_{i=1}^k Y_i \leq n-1 \\ t_k & \sum_{i=1}^{k-1} Y_i = 0, Y_k = n \\ \min \left\{ t_k, \frac{nt_{k-1}}{\sum_{i=1}^{k-1} Y_i} \right\} & \sum_{i=1}^k Y_i = n, 1 \leq Y_k < n, \\ & 1 \leq \sum_{i=1}^{k-1} Y_i < n \end{cases}$$

۳ سازگاری $\tilde{\theta}$

با توجه به منابع موجود، برآوردگر ماکسیمم درست‌نمایی در صورت وجود و تحت شرایط خاص، سازگار است. جهت اطلاع بیشتر به [۴] مراجعه شود. همان‌طور که در بخش پیشینی ملاحظه شد، $\tilde{\theta}$ همان برآوردگر ماکسیمم درست‌نمایی است، هرگاه $\sum_{i=1}^k y_i \neq 0$ و اگر $\sum_{i=1}^k y_i = 0$ ، برآورد ماکسیمم درست‌نمایی θ وجود ندارد، لذا استناد به سازگاری برآوردیاب ماکسیمم درست‌نمایی موضوعیتی نخواهد داشت. بنا بر این، با توجه به لم‌های زیر و انجام محاسبات، ابتدا ناریبی مجانبی و سپس سازگاری $\tilde{\theta}$ ثابت می‌شود.

لم ۱.۳

$$\lim_{n \rightarrow \infty} E\left(nt_k I\left(\sum_{i=1}^k Y_i = 0\right)\right) = \lim_{n \rightarrow \infty} nt_k \left(1 - \frac{t_k}{\theta}\right)^n = 0$$

لم ۲.۳

$$\lim_{n \rightarrow \infty} E\left(\frac{nt_k}{\sum_{i=1}^k Y_i} I\left(1 \leq \sum_{i=1}^k Y_i \leq n-1\right)\right) = \theta$$

اثبات.

$$\frac{nt_k}{\sum_{i=1}^k Y_i} I\left(1 \leq \sum_{i=1}^k Y_i \leq n-1\right) \sim \frac{nt_k}{W} I\left(1 \leq W \leq n-1\right) :$$

$$W \sim \text{Binomial}(n, p), p = \frac{t_k}{\theta}$$

که در آن $X \sim Y$ هم توزیع بودن X و Y را نشان می‌دهد.

$$E\left(\frac{nt_k}{W} I\left(1 \leq W \leq n-1\right)\right) \leq nt_k E\left(\frac{(W+\delta) I\left(1 \leq W \leq n-1\right)}{(W+1)(W+2)}\right) \quad (3)$$

$$\begin{aligned} E\left(\frac{(W+\delta) I\left(1 \leq W \leq n-1\right)}{(W+1)(W+2)}\right) &= E\left(\frac{I\left(1 \leq W \leq n-1\right)}{(W+1)}\right) \\ &\quad + 2E\left(\frac{I\left(1 \leq W \leq n-1\right)}{(W+1)(W+2)}\right) \\ E\left(\frac{I\left(1 \leq W \leq n-1\right)}{(W+1)}\right) &= \frac{1 - (1-p)^{n+1}}{(n+1)p} \\ &\quad - \frac{(n+1)p(1-p)^n - p^{n+1}}{(n+1)p} \end{aligned} \quad (4)$$

در نتیجه

$$\lim_{n \rightarrow \infty} nt_k E\left(\frac{I\left(1 \leq W \leq n-1\right)}{(W+1)}\right) = \frac{t_k}{p} = \theta \quad (5)$$

$$\begin{aligned} E\left(\frac{I\left(1 \leq W \leq n-1\right)}{(W+1)(W+2)}\right) &= \frac{1}{(n+1)(n+2)p^2} \left(1 - (1-p)^{n+2}\right. \\ &\quad \left. - (n+2)p(1-p)^{n+1}\right. \\ &\quad \left. - \binom{n+2}{2} p^2 (1-p)^n - p^{n+2}\right) \end{aligned}$$

در نتیجه

$$\lim_{n \rightarrow \infty} nt_k E\left(\frac{I\left(1 \leq W \leq n-1\right)}{(W+1)(W+2)}\right) = 0 \quad (6)$$

با توجه به رابطه‌های (۳)، (۴)، (۵) و (۶) داریم:

$$\lim_{n \rightarrow \infty} E\left(\frac{nt_k}{\sum_{i=1}^k Y_i} I\left(1 \leq \sum_{i=1}^k Y_i \leq n-1\right)\right) \leq \theta \quad (7)$$

از طرفی، داریم:

$$\begin{aligned} & E\left(nt_k E\left(\frac{1}{W} I(1 \leq W \leq n-1)\right)\right) \\ & \geq nt_k E\left(\frac{1}{W+1} I(1 \leq W \leq n-1)\right) \end{aligned} \quad (8)$$

که با توجه به رابطه (۵) نیز داریم:

$$\begin{aligned} \lim_{n \rightarrow \infty} E\left(nt_k E\left(\frac{1}{W} I(1 \leq W \leq n-1)\right)\right) & \geq \\ \lim_{n \rightarrow \infty} nt_k E\left(\frac{I(1 \leq W \leq n-1)}{W+1}\right) & = \theta \end{aligned} \quad (9)$$

با تلفیق روابط (۷) و (۹)، خواهیم داشت:

$$\lim_{n \rightarrow \infty} E\left(nt_k E\left(\frac{1}{W} I(1 \leq W \leq n-1)\right)\right) = \theta. \quad (10)$$

□

لم ۳.۳. $\lim_{n \rightarrow \infty} E(t_k I(\sum_{i=1}^{k-1} Y_i = \circ, Y_k = n)) = \circ$.
اثبات.

$$\begin{aligned} & \circ \leq E(t_k I(\sum_{i=1}^{k-1} Y_i = \circ, Y_k = n)) \\ & = t_k P(\sum_{i=1}^{k-1} Y_i = \circ, Y_k = n) \\ & \leq t_k P(Y_k = n) \\ & = t_k \left((1 - \frac{t_{k-1}}{\theta})^n I(t_{k-1} \leq \theta < t_k) + (\frac{t_k - t_{k-1}}{\theta})^n I(t_k \leq \theta) \right) \rightarrow \circ \end{aligned}$$

□

لم ۴.۳

$$\begin{aligned} \lim_{n \rightarrow \infty} E\left(\min\left\{t_k, \frac{nt_{k-1}}{\sum_{i=1}^{k-1} Y_i}\right\}\right) \\ \times I\left(\sum_{i=1}^k Y_i = n, 1 \leq Y_k < n, 1 \leq \sum_{i=1}^{k-1} Y_i < n\right) = \circ. \end{aligned}$$

اثبات.

$$\begin{aligned} & \circ \leq E\left(\min\left\{t_k, \frac{nt_{k-1}}{\sum_{i=1}^{k-1} Y_i}\right\} I\left(\sum_{i=1}^k Y_i = n, 1 \leq Y_k < n, 1 \leq \sum_{i=1}^{k-1} Y_i < n\right)\right) \\ & \leq t_k P\left(\sum_{i=1}^k Y_i = n, 1 \leq Y_k < n, 1 \leq \sum_{i=1}^{k-1} Y_i < n\right) \\ & \leq t_k P\left(\sum_{i=1}^k Y_i = n\right) \\ & = t_k \left((1 - \frac{t_{k-1}}{\theta})^n I(t_{k-1} \leq \theta < t_k) + (\frac{t_k - t_{k-1}}{\theta})^n I(t_k \leq \theta) \right) \rightarrow \circ \end{aligned}$$

□

بنا بر این با توجه به لم‌های ۲.۲ تا ۴.۳، قضیه زیر برقرار است:

قضیه ۵.۳. $\tilde{\theta}$ برآوردگری مجاناً نااریب^۵ برای θ است.

با توجه به قضیه ۵.۳، نامساوی زیر

$$\begin{aligned} \frac{I(1 \leq x \leq n-1)}{(x+1)(x+2)} & \leq \frac{I(1 \leq x \leq n-1)}{x^2} \\ & \leq \frac{(x+5)(x+19)I(1 \leq x \leq n-1)}{(x+1)(x+2)(x+3)(x+4)}, \end{aligned}$$

همچنین نحوه اثبات لم‌های ۲.۲ تا ۴.۳ و درستی $\lim_{n \rightarrow \infty} E(\tilde{\theta})^2 = \theta^2$ ، قضیه زیر برقرار است:

قضیه ۶.۳. $\tilde{\theta}$ برآوردگری سازگار برای θ است.

۴ بررسی زمان‌های نظارت و مطالعات شبیه‌سازی

در سانسور بازه‌ای، $t_1 < t_2 < \dots < t_k$ زمان‌های از پیش تعیین‌شده‌ای هستند که در آنها واحدهای آزمایشی مورد بازرسی قرار می‌گیرند. با این حال، جهت بررسی تأثیر تعداد دفعات نظارت و همچنین زمان‌های نظارت بر رفتار برآوردگر پیشنهادی، به‌ازای برخی مقادیر θ ، نمونه‌های تصادفی n تایی از توزیع یکنواخت در فاصله $(\circ, \theta)$ تولید و با در نظر گرفتن زمان‌های نظارت مختلف، $\tilde{\theta}$ برآورد می‌شود. با $N = 100000$ بار تکرار، $\text{Mean} = \frac{\sum_{i=1}^{100000} \tilde{\theta}_i}{100000}$ و $\text{MSE} = \frac{\sum_{i=1}^{100000} (\tilde{\theta}_i - \theta)^2}{100000}$ محاسبه شده که نتایج در جدول (۱) و (۲) آمده است.

با توجه به جدول‌های (۱) و (۲) و همچنین ضابطه (۲)، ملاحظه می‌شود که $\tilde{\theta}$ فقط تابعی از $\sum_{i=1}^{k-1} y_i$ و $\sum_{i=1}^k y_i$ است و لذا فقط اطلاعات حاصل از دو زمان نظارت آخر، بر تخمین پارامتر مؤثرند. همچنین، برای $k = 2$ ، سرعت همگرایی در Mean و MSE کاهش می‌یابد، هرگاه دومین (و یا آخرین) زمان نظارت بزرگ‌تر از پارامتر انتخاب شود، لذا بایستی دومین زمان نظارت،

^۵ Asymptotically Unbiased

با $m(t) = E(T - t | T > t)$ در این زمینه، مطالعات بسیاری انجام شده است که جهت اطلاع می‌توان به [۲، ۵] مراجعه کرد. سیستمی با n مؤلفه که لازمه فعالیت آن، سالم بودن حد اقل r مؤلفه است را یک سیستم r از n گویند. اگر $T_1, \dots, T_n$ طول عمر مؤلفه‌های مستقل از هم سیستم، با تابع چگالی و تابع توزیع مشترک f و F باشند، آن‌گاه بدیهی است که طول عمر آن برابر با $T_{n-r+1:n}, \dots, T_n$ -امین آماره ترتیبی^۸، خواهد بود.

فرض کنید که در k زمان مختلف $t_1 < t_2 < \dots < t_k$ بر عملکرد سیستم نظارت داشته به‌طوری که $y_i : i = 1, 2, \dots, k$ تعداد واحدهای خراب‌شده در فاصله بین زمان‌های نظارت t_{i-1} و t_i بوده و سیستم نیز در زمان t_k فعال است. با توجه به باقیمانده عمر یک سیستم تحت نظارت چندگانه که در [۹] بررسی شده و معادل بودن پیشامدهای $(Y_1 = y_1, Y_2 = y_2, \dots, Y_k = y_k)$ و

$$(T_{y_1:n} < t_1 < T_{y_1+y_2:n} < t_2 < \dots$$

$$< t_{k-1} < T_{\sum_{i=1}^k y_i:n} < t_k < T_{\sum_{i=1}^k y_i+1:n})$$

برای $\sum_{i=1}^k y_i + 1 \leq m \leq n$ داریم:

$$f_{T_{m:n}}(x | T_{y_1:n} < t_1 < \dots < t_{k-1} < T_{\sum_{i=1}^k y_i:n} < t_k < T_{\sum_{i=1}^k y_i+1:n}) = \frac{\left((n - \sum_{i=1}^k y_i)! (F(x) - F(t_k))^{m-1-\sum_{i=1}^k y_i} f(x) (1 - F(x))^{n-m} \right)}{(m-1 - \sum_{i=1}^k y_i)! (n-m)! (1 - F(t_k))^{n-\sum_{i=1}^k y_i}}$$

اگر طول عمر قطعات در فاصله (θ, ∞) به‌طور مستقل و یکنواخت توزیع شده باشند، آن‌گاه

$$f_{T_{m:n}}(x | T_{y_1:n} < t_1 < \dots < t_{k-1} < T_{\sum_{i=1}^k y_i:n} < t_k < T_{\sum_{i=1}^k y_i+1:n}) = \frac{(n - \sum_{i=1}^k y_i)! (x - t_k)^{m-1-\sum_{i=1}^k y_i} (\theta - x)^{n-m}}{(m-1 - \sum_{i=1}^k y_i)! (n-m)! (\theta - t_k)^{n-\sum_{i=1}^k y_i}} I(t_k < x < \theta)$$

و می‌توان نشان داد:

$$\frac{T_{m:n} - t_k}{\theta - t_k} | T_{y_1:n} < t_1 < \dots < t_k < T_{\sum_{i=1}^k y_i+1:n} \sim \text{Beta}(m - \sum_{i=1}^k y_i, n - m + 1)$$

بنا بر این، خواهیم داشت:

کوچک‌تر از پارامتر در نظر گرفته شود. علاوه بر این که

$$P(\tilde{\theta} = x | 1 \leq \sum_{i=1}^k Y_i \leq n-1) = \frac{P(\sum_{i=1}^k Y_i = \frac{nt_k}{x})}{P(1 \leq \sum_{i=1}^k Y_i \leq n-1)},$$

و $\lim_{n \rightarrow \infty} P(1 \leq \sum_{i=1}^k Y_i \leq n-1) = 1$ و لذا توزیع مجانبی $1 \leq \sum_{i=1}^k Y_i \leq n-1$ به‌صورت نرمال با میانگین θ و واریانس $\frac{(\theta - t_k)}{nt_k \theta^2}$ است. که در این حالت نیز، واریانس $\tilde{\theta}$ مقدار کمتری را خواهد داشت، هرگاه مقدار t_k به‌عنوان زمان نظارت k -ام، کوچک‌تر از θ و نزدیک‌تر به آن انتخاب شود که با نتایج به‌دست‌آمده در جدول (۱) و (۲) نیز همخوانی دارد.

بنا بر این، θ را با توجه به مجهول بودن آن، می‌توان بر اساس اطلاعات به‌دست‌آمده از اولین نظارت و توسط $\frac{nt_1}{y_1}$ برآورد نمود. در جدول (۳) و (۴)، دومین زمان نظارت را به‌صورت ضرایب مختلفی از $\frac{nt_1}{y_1}$ در نظر گرفته و رفتار $\tilde{\theta}$ بررسی شده است. با توجه به موارد فوق‌الذکر و نتایج به‌دست‌آمده در جدول‌های (۱) تا (۴)، به نظر می‌رسد که با انتخاب اولیه t_1 (به‌طوری که تا آن زمان، بعضی از واحدهای آزمایشی و نه همه آنها، خراب‌شده باشند) و برآورد پارامتر، بایستی دومین زمان نظارت را کوچک‌تر از مقدار برآوردشده پارامتر و نزدیک به آن، در نظر گرفت. برای مثال، با انتخاب $t_2 = 0.95 \frac{nt_1}{y_1}$ می‌توان انتظار برقراری شرط $1 \leq \sum_{i=1}^k Y_i \leq n-1$ و در نتیجه، کوچک بودن واریانس $\tilde{\theta}$ را داشت.

۵ برآورد میانگین باقیمانده عمر سیستم

موازی^۶

یکی از مفاهیم مطرح در نظریه قابلیت اعتماد، میانگین باقیمانده عمر است که نقش مهمی را در تحلیل داده‌های طول عمر ایفا می‌کند. اگر طول عمر یک سیستم یا قطعه‌ای از آن را با متغیر تصادفی T نشان دهیم و فرض شود که طول عمر آن بیشتر از عددی مانند t است، آن‌گاه طبق تعریف، تابع میانگین باقیمانده عمر دستگاه و یا آن قطعه در t برابر است

^۶ Parallel system

^۷ R-out of-n system

^۸ Order statistics

$$\hat{p} > \frac{(z_{\frac{\alpha}{r}})^2}{n + (z_{\frac{\alpha}{r}})^2} \quad \text{هرگاه} \quad (ب)$$

$$\left(\frac{t_k}{\hat{p} + z_{\frac{\alpha}{r}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}, +\infty \right), \quad (۱۴)$$

$$\hat{p} < \frac{(z_{\frac{\alpha}{r}})^2}{n + (z_{\frac{\alpha}{r}})^2} \quad \text{هرگاه}$$

مثال. فرض کنید که سیستمی موازی شامل ۹۸ قطعه است که به‌طور مستقل و یکنواخت در فاصله $(0, \theta)$ توزیع شده‌اند. اگر $X_1, \dots, X_{98}$ را زمان خرابی قطعات بدانیم و زمان‌های نظارت به‌صورت $t_1 = 7, t_2 = 11$ انتخاب شوند و بعد از گذشت ۷ و ۱۱ ساعت، به‌ترتیب ۳۰ و ۵۱ عدد از قطعات خراب شده باشند، خواهیم داشت:

$$n = 98, k = 2, t_1 = 7, t_2 = 11, y_1 = 30, y_2 = 21,$$

$$\sum_{i=1}^2 y_i = 51, \hat{p} = 0.52$$

$$\hat{\theta} = \frac{nt_k}{\sum_{i=1}^2 y_i} = \frac{(98)(11)}{51} = 21.14 \quad (\text{الف})$$

(ب) با توجه به رابطه (۱۲)، متوسط زمان باقیمانده عمر سیستم در $t_2 = 11$ را می‌توان ۹.۹۳ ساعت تخمین زد.

(پ) با توجه به

$$\hat{p} = 0.52 > \frac{(z_{\frac{\alpha}{r}})^2}{n + (z_{\frac{\alpha}{r}})^2} = \frac{(1.645)^2}{(98) + (1.645)^2} = 0.027, 5 < n\hat{p} = 51$$

و $5 < n(1-\hat{p}) = 47$ به نظر می‌رسد که استفاده از تقریب نرمال و (۱۳) مناسب است. بنا بر این، یک فاصله اطمینان تقریبی ۹۰ درصد برای θ به‌صورت $(18.23, 25.15)$ و یک فاصله اطمینان تقریبی ۹۰ درصد نیز برای متوسط زمان باقیمانده عمر سیستم در $t_2 = 11$ به‌صورت $(7.08, 13.86)$ است.

$$\begin{aligned} E(T_{m:n}|T_{y_1:n} < t_1 < \dots < t_{k-1} < T_{\sum_{i=1}^k y_{i:n}} < t_k < T_{\sum_{i=1}^k y_{i+1:n}}) &= t_k + (\theta - t_k) \frac{m - \sum_{i=1}^k y_i}{n - \sum_{i=1}^k y_i + 1} \\ &= \frac{m - \sum_{i=1}^k y_i}{n - \sum_{i=1}^k y_i + 1} \theta \\ &\quad + \left(1 - \frac{m - \sum_{i=1}^k y_i}{n - \sum_{i=1}^k y_i + 1}\right) t_k \end{aligned} \quad (۱۱)$$

نکته. رابطه (۱۱) نشان می‌دهد که امید شرطی، ترکیبی محذب از θ و t_k است. با توجه به رابطه (۱۱)، میانگین باقیمانده عمر یک سیستم موازی در t_k که مؤلفه‌های آن در فاصله $(0, \theta)$ به‌طور مستقل و یکنواخت توزیع شده باشند، برابر است با:

$$\begin{aligned} E(T_{n:n} - t_k | T_{y_1:n} < t_1 < \dots < t_{k-1} < T_{\sum_{i=1}^k y_{i:n}} < t_k < T_{n:n}) \\ &= (\theta - t_k) \frac{n - \sum_{i=1}^k y_i}{n - \sum_{i=1}^k y_i + 1} \end{aligned}$$

از طرفی، با توجه به این که فرض می‌شود که سیستم در زمان t_k فعال است لذا با توجه به دومین ضابطه در (۲)، برآوردی نقطه‌ای از میانگین باقیمانده عمر سیستم موازی در t_k به‌صورت زیر است:

$$\begin{aligned} \hat{E}(T_{n:n} - t_k | T_{y_1:n} < t_1 < \dots < t_{k-1} < T_{\sum_{i=1}^k y_{i:n}} < t_k < T_{n:n}) \\ &= \frac{(n - \sum_{i=1}^k y_i) t_k}{(n - \sum_{i=1}^k y_i + 1) \sum_{i=1}^k y_i} \end{aligned} \quad (۱۲)$$

با در نظر گرفتن $\hat{p} = \frac{\sum_{i=1}^k y_i}{n}, p = F(t_k) = \frac{t_k}{\theta}$ و فرمول فاصله اطمینان تقریبی برای p می‌توان نشان داد:

قضیه ۱.۵. یک فاصله اطمینان تقریبی با ضریب اطمینان $(1-\alpha)$ برای θ درصد به‌صورت زیر است
(الف):

$$\left(\frac{t_k}{\hat{p} + z_{\frac{\alpha}{r}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}, \frac{t_k}{\hat{p} - z_{\frac{\alpha}{r}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \right), \quad (۱۳)$$

جدول ۱. محاسبه Mean و MSE به ازای زمانهای نظارت و n های مختلف، برای $\theta = 100$

t	Mean MSE	$n = 50$	$n = 100$	$n = 500$	$n = 10^3$	$n = 10^4$	$n = 2(10^4)$
$t_1 = 50, t_r = 80$	Mean	100/49	100/25	100/05	100/03	100/004	100/002
	MSE	53/83	25/99	5/04	2/51	0/25	0/12
$t_1 = 50, t_r = 90$	Mean	100/22	100/12	100/02	100/01	100/003	100/000
	MSE	23/99	11/44	2/22	0/111	0/110	0/055
$t_1 = 50, t_r = 95$	Mean	99/78	100/03	100/00	100/00	100/001	100/000
	MSE	18/65	5/76	1/06	0/53	0/052	0/026
$t_1 = 50, t_r = 97$	Mean	99/06	99/87	100/00	100/00	100/000	100/000
	MSE	25/9	5/68	0/62	0/31	0/031	0/015
$t_1 = 50, t_r = 99$	Mean	97/12	98/71	99/99	100/00	100/000	100/000
	MSE	48/18	17/03	0/283	0/102	0/01	0/005
$t_1 = 50, t_r = 100$	Mean	95/15	96/44	98/3	98/77	99/25	99/72
	MSE	68/66	37/64	8/79	4/6	0/482	0/246
$t_1 = 50, t_r = 101$	Mean	95/59	96/89	98/75	99/21	99/92	99/98
	MSE	69/1	38/09	9/22	5/01	0/74	0/43
$t_1 = 50, t_r = 103$	Mean	99/47	97/73	99/41	99/75	100/001	100/005
	MSE	72/63	41/37	11/76	7/02	0/991	0/501
$t_1 = 50, t_r = 105$	Mean	97/27	98/43	99/8	99/97	100/01	100/004
	MSE	78/86	46/88	14/81	8/78	0/99	0/5
$t_1 = 50, t_r = 110$	Mean	98/8	99/64	100/13	100/07	100/01	100/005
	MSE	101/48	64/48	19/28	10/04	0/995	0/5
$t_1 = 50, t_r = 120$	Mean	100/62	100/65	100/17	100/08	100/01	100/002
	MSE	153/68	92/42	20/27	10/09	0/999	0/51
$t_1 = 20, t_r = 50$ $t_r = 80$	Mean	100/49	100/25	100/05	100/03	100/04	100/002
	MSE	53/83	25/99	5/04	2/51	0/25	0/12
$t_1 = 10, t_r = 20$ $t_r = 50, t_r = 80$	Mean	100/49	100/25	100/05	100/03	100/04	100/002
	MSE	53/83	25/99	5/04	2/51	0/25	0/12

جدول ۲. محاسبه Mean و MSE به‌ازای زمان‌های نظارت و n ‌های مختلف، برای $\theta = 300$

t	Mean MSE	$n = 50$	$n = 100$	$n = 500$	$n = 10^3$	$n = 10^4$	$n = 2(10^4)$
$t_1 = 150, t_2 = 280$	Mean	300/05	300/2	300/02	300/01	300/006	300/002
	MSE	116/26	67/05	12/91	6/47	0/639	0/32
$t_1 = 150, t_2 = 290$	Mean	297/74	299/79	300/02	300/01	300/002	300/000
	MSE	214/3	48/14	6/24	3/12	0/310	0/155
$t_1 = 150, t_2 = 295$	Mean	293/96	298/09	300/00	300/00	300/1	100/000
	MSE	344/49	95/13	3/11	1/53	0/152	0/076
$t_1 = 150, t_2 = 297$	Mean	291/34	296/13	299/97	300/00	300/000	300/000
	MSE	433/61	153/24	2/54	0/917	0/091	0/045
$t_1 = 150, t_2 = 299$	Mean	287/78	292/37	299/06	299/87	300/000	300/000
	MSE	546/9	258/2	16/87	2/02	0/030	0/015
$t_1 = 150, t_2 = 300$	Mean	285/45	289/31	296/89	296/32	298/82	299/16
	MSE	617/95	338/74	79/08	41/38	4/34	2/21
$t_1 = 150, t_2 = 301$	Mean	285/89	289/77	295/37	296/8	299/26	299/57
	MSE	618/4	339/19	79/56	41/85	4/751	2/591
$t_1 = 150, t_2 = 303$	Mean	286/78	290/68	296/25	297/63	299/77	299/94
	MSE	621/93	342/84	83/01	45/11	6/66	3/9
$t_1 = 150, t_2 = 305$	Mean	287/66	291/59	297/01	298/3	299/96	300/006
	MSE	628/99	350/14	89/11	50/43	8/15	4/421
$t_1 = 150, t_2 = 310$	Mean	289/87	293/57	298/47	299/4	300/03	300/014
	MSE	662/08	379/51	110/56	66/35	8/943	4/51
$t_1 = 150, t_2 = 320$	Mean	293/46	296/76	299/94	300/13	300/03	300/014
	MSE	767/92	473/32	152/12	86/11	8/956	4/51
$t_1 = 120, t_2 = 150$ $t_3 = 280$	Mean	300/05	300/2	300/02	300/01	300/006	00/002
	MSE	166/26	67/05	12/91	6/47	0/639	0/32
$t_1 = 100, t_2 = 120$ $t_3 = 150, t_4 = 280$	Mean	300/05	300/2	300/02	300/01	300/006	00/002
	MSE	166/26	67/05	12/91	6/47	0/639	0/32

جدول ۳. محاسبه Mean و MSE به ازای زمانهای نظارت و n های مختلف، برای $\theta = 100$

t	Mean MSE	$n = 50$	$n = 100$	$n = 500$	$n = 10^3$	$n = 10^4$	$n = 2(10^4)$
$t_1 = 50, t_2 = \frac{0.8nt_1}{y_1}$	Mean MSE	101/24 82/6	100/19 24/28	99/82 2/31	99/9 1/12	99/9 0/11	99/99 0/06
$t_1 = 50, t_2 = \frac{0.85nt_1}{y_1}$	Mean MSE	103/26 119/58	101/67 39/05	100/14 2/54	99/99 0/76	99/99 0/05	99/99 0/03
$t_1 = 50, t_2 = \frac{0.97nt_1}{y_1}$	Mean MSE	104/28 141/61	102/51 50/27	100/57 4/35	100/22 1/3	99/99 0/03	99/99 0/02
$t_1 = 50, t_2 = \frac{1.02nt_1}{y_1}$	Mean MSE	105/9 182/27	103/91 74/21	101/44 12/42	100/8 6/45	100/01 0/98	100 0/5
$t_1 = 50, t_2 = \frac{1.05nt_1}{y_1}$	Mean MSE	105/67 182/41	103/64 75/12	101 14/2	100/39 8/08	100/01 0/99	100 0/5
$t_1 = 50, t_2 = \frac{1.1nt_1}{y_1}$	Mean MSE	104/93 188/18	102/59 83/36	100/28 19/11	100/08 10/1	100/01 1	100 0/5

جدول ۴. محاسبه Mean و MSE به ازای زمانهای نظارت و n های مختلف، برای $\theta = 300$

t	Mean MSE	$n = 50$	$n = 100$	$n = 500$	$n = 10^3$	$n = 10^4$	$n = 2(10^4)$
$t_1 = 150, t_2 = \frac{0.8nt_1}{y_1}$	Mean MSE	303/72 743/43	300/56 218/55	299/46 20/82	299/7 10/09	299/97 0/99	299/99 0/49
$t_1 = 150, t_2 = \frac{0.85nt_1}{y_1}$	Mean MSE	309/78 1076/2	305 351/44	300/43 22/88	299/96 6/83	299/97 0/47	299/99 0/24
$t_1 = 150, t_2 = \frac{0.97nt_1}{y_1}$	Mean MSE	312/84 1274/5	307/53 452/41	301/71 39/15	300/66 11/74	299/97 0/28	299/99 0/14
$t_1 = 150, t_2 = \frac{1.02nt_1}{y_1}$	Mean MSE	317/69 1640/4	311/74 667/85	304/31 111/81	302/41 58/06	300/04 8/83	300/01 4/51
$t_1 = 150, t_2 = \frac{1.05nt_1}{y_1}$	Mean MSE	317/01 1641/7	310/92 676/12	303 127/82	301/18 72/72	300/03 8/95	300/01 4/51
$t_1 = 150, t_2 = \frac{1.1nt_1}{y_1}$	Mean MSE	314/8 1693/6	307/78 750/26	300/84 172	300/25 90/15	300/03 8/96	300/01 4/52

مراجع

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

- [2] Asadi, M. and Bayramoglu, I. (2005). A note on the mean residual life function of a parallel system. *Communications in Statistics - Theory and Methods*, **34**(2), 475–484.

- [3] Finkelstein, D. M. and Wolfe, R. A. (1985). A semi-parametric model for regression analysis of interval censored failure time data, *Biometrics*, **41**(4), 933- 945.
- [4] Gentleman, R. and Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation, *Biometrika*, **81**(3), 618-623.
- [5] Khanjari, S.M. (2008). Mean past and mean residual life functions of a parallel system with nonidentical components, *Communications in Statistics - Theory and Methods*, **37**(7), 1134–1145.
- [6] Lindsey, J. C. and Ryan, L. M. (1998). Methods for interval-censored data, *Statistics in Medicine*, **17**(2), 219-238.
- [7] Odell, P. M. Anderson, K.M. and Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a weibull -based accelerated failure time model, *Biometrics*, **48**(3), 951-959.
- [8] Peto, R. (1973). Experimental survival curves for interval-censored data, *Applied Statistics*, **22**(1), 86-91.
- [9] Poursaeed, M. H. (2010), A note on the mean past and the mean residual life of a $(n-k+1)$ -out-of- n system under multi monitoring, *Statistical Papers*, **51**(2), 409-419.
- [10] Sun J. (2007). *The Statistical Analysis of Interval-Censored Failure Time Data*, Springer Science and Business Media.