

دسترسی در سایت <http://jnrm.srbiau.ac.ir>

سال یازدهم، شماره پنجاه و هفتم، آذر و دی ۱۴۰۴

شماره شاپا: ۵۸۸۸-۲۵۸۸



پژوهش‌های نوین در ریاضی



دانشگاه آزاد اسلامی، واحد علوم و تحقیقات

برآورد حداکثر درست‌نمایی پارامترهای توزیع نمایی-پواسون تحت سانسور فاصله‌ای پیش‌رونده نوع I با الگوریتم‌های نیوتن-رافسون، EM و SEM

احمد محمدی^۱، علی شادرخ^{۲*}، مسعود یارمحمدی^۳

(^{۱و۲}) گروه آمار، دانشگاه پیام نور، تهران، ایران

تاریخ ارسال مقاله: ۱۴۰۲/۰۲/۲۴ تاریخ پذیرش مقاله: ۱۴۰۳/۰۸/۰۵

چکیده

توزیع نمایی-پواسون با استفاده از ترکیب توزیع نمایی با توزیع پواسون بریده شده در صفر به عنوان مدلی برای داده‌های طول عمر با نرخ شکست کاهشی، تولید می‌شود. این مقاله به مسئله برآورد کردن پارامترهای نامعلوم توزیع نمایی-پواسون به عنوان یک مدل برای طول عمر می‌پردازد که نمونه‌ها تحت سانسور فاصله‌ای پیش‌رونده نوع I مشاهده شده‌اند. ما از الگوریتم‌های نیوتن-رافسون، بیشینه‌سازی کلاسیک مقادیر مورد انتظار (EM) و بیشینه‌سازی تصادفی مقادیر مورد انتظار (SEM) برای یافتن برآوردهای حداکثر درست‌نمایی (MLE) پارامترهای نامعلوم استفاده می‌کنیم. عملکرد برآوردهای SEM پیشنهادی با استفاده از یک شبیه‌سازی مونت کارلو نشان داده شده است و نیز روی یک مجموعه داده واقعی استفاده شده است. شبیه‌سازی نشان داد که عملکرد الگوریتم SEM بر مبنای میانگین مربع خطا (MSE) کاملاً رضایت‌بخش است و با افزایش اندازه نمونه، کارایی برآوردها نیز افزایش می‌یابد.

واژه‌های کلیدی: برآورد حداکثر درست‌نمایی، میانگین مربع خطا، روش نیوتن-رافسون، الگوریتم EM، شبیه‌سازی مونت کارلو.

۱. مقدمه

آزمایش های بقاء و قابلیت اطمینان روش های مؤثری برای ارزیابی طول عمر محصول در شرایط عملیاتی مشخص شده هستند. چنین آزمایشاتی در کل بر روی واحدهای یکسان قبل از عرضه محصول، انجام می شوند و زمان های خرابی مشاهده شده ثبت می شوند و به منظور ارزیابی و پیش بینی ویژگی های محصول مورد بررسی قرار می گیرند. با توجه به محدودیت های زمانی، هزینه های مالی، یا خرابی سهوی، چنین آزمایشاتی قبل از اینکه همه اجزای آزمایش شده در طول عمر خراب شوند، قطع می شوند. در چنین شرایطی از تکنیک های سانسور کلاسیک استفاده می شود. سانسور نوع I و نوع II دو روش متداول هستند که در آن ها آزمایش به ترتیب پس از مدتی که از قبل تعیین شده و یا پس از مشاهده تعداد از قبل مشخص شده، خرابی ها متوقف می شود در نتیجه داده های جمع آوری شده اصطلاحاً ناقص هستند. با این حال، هیچ یک از این سیستم های سانسور یاد شده، به پژوهشگر اجازه نمی دهد تا قبل از پایان آزمایش، واحدی زنده را از آزمایش حذف کند. علاوه بر این در بسیاری از سناریوهای دنیای واقعی، برای مثال در تحقیقات پزشکی و بالینی، آزمایش کنندگان قادر به نظارت مستمر بر آزمونها نیستند. در چنین حالاتی از روش سانسور فاصله ای استفاده می شود که در آن، داده های طول عمر یا زمان های خرابی اغلب در فواصل زمانی از قبل مشخص شده، ثبت می شوند. این سانسور نیز اجازه حذف واحدهای بین آزمایش را نمی دهد. مفهوم سانسور پیش رونده، که شامل حذف واحدهای تحت آزمایش است، اخیراً مطرح و مورد توجه واقع شده است. آگاروالا با ترکیب مفاهیم سانسور پیش رونده، سانسور نوع I و سانسور فاصله ای، پیشگام سانسور فاصله ای پیش رونده نوع I شد [2]. این سانسور به صورت زیر مدل بندی آماری می شود.

فرض کنید در زمان $t_0 = 0$ یک نمونه تصادفی از n واحد تحت آزمایش بقاء قرار می گیرند و برای $i = 1, 2, \dots, m$ در زمان بازرسی t_i تعداد کل خرابی های مشاهده شده در بازه زمانی $(t_{i-1}, t_i]$ برابر D_i باشد. در این صورت تعداد واحدهای سالم (یا زنده) در زمان بازرسی t_i که با Y_i نشان خواهیم داد، یک متغیر تصادفی است. بعلاوه فرض کنید تعداد کل واحدهای حذف شده از آزمایش در زمان بازرسی t_i برابر R_i باشد بنابراین $R_i \leq Y_i$. برای تعیین R_i ها $i = 1, 2, \dots, m-1$ می توانیم نسبت معین شده از قبل مانند q_i از واحدهای سالم در زمان بازرسی t_i را R_i بگیریم؛ یعنی $R_i = q_i \times Y_i$ و یا کمترین مقدار بین تعداد مشخص شده از قبل و Y_i را می توانیم به عنوان R_i در نظر بگیریم. روشن است $R_m = Y_m$ ؛ یعنی در زمان t_m همه واحدهای سالم از آزمایش حذف و آزمایش تمام می شود. داده های حاصل از این نوع سانسور را می توان به صورت $(D_i, R_i, t_i)_{i=1}^m$ نشان داد.

بسیاری از مجموعه داده های طول عمر از نظر آماری با استفاده از سانسور فاصله ای پیش رونده نوع I تجزیه و تحلیل شده اند. به عنوان نمونه می توانیم به داده های مطالعه 112 بیمار مبتلا به میلوم سلول پلازما در [۵]، مجموعه داده های مطالعه HMO-HIV، در [۱۸] و داده های زمان بقاء از عمل جراحی گروهی متشکل از ۳۷۴ بیمار در [۱۳] اشاره کنیم. بعلاوه برای خانواده توزیع های وایبل نمای شده در [۳]، برای توزیع وایبل در [۱۲]، برای توزیع ریلی تعمیم یافته در [۱۱]، برای توزیع گاما در [۲۱]، برای توزیع نمایی تعمیم یافته در [۶] و [۱۴]، برای توزیع وایبل و توزیع های نمایی تعمیم یافته در [۱۵]، برای توزیع لاگ-نرمال در [۱۶]، و برای توزیع بور ۱۲ در [۱] روش های استنتاج آماری تحت این سانسور مطالعه شده است. این مقاله توزیع نمایی-پواسون (Exponential-Poisson) را تحت سانسور فاصله ای پیش رونده نوع I مورد بحث قرار می دهد. توزیع نمایی-پواسون (EP) به صورتی که در ادامه آمده است با ترکیب توزیع نمایی با توزیع پواسون بریده شده در صفر به عنوان مدلی برای داده های طول عمر با نرخ شکست کاهشی، در سال ۲۰۰۷ توسط جوشکون کوش معرفی شد [۱۰].

فرض کنید $T_1, T_2, \dots, T_W$ یک نمونه تصادفی از توزیع نمایی با پارامتر $\beta > 0$ با تابع چگالی احتمال $f(t; \beta) = \beta \exp(-\beta t)$ ، $t \geq 0$ و W یک متغیر تصادفی با توزیع پواسن بریده شده در صفر با پارامتر $\lambda > 0$ باشد؛ یعنی $p(w; \lambda) = (w! (1 - e^{-\lambda}))^{-1} \lambda^w e^{-\lambda}$ ، $w \in \mathbb{N}$ و $X = \min(T_1, T_2, \dots, T_W)$ که در آن W و $T_1, T_2, \dots, T_W$ مستقل از همند. در این صورت داریم $f(x|w; \beta) = \beta w \exp(-\beta w x)$ ، $x \geq 0$ ، $w \in \mathbb{N}$ و تابع چگالی احتمال و تابع توزیع تجمعی کناری (حاشیه ای) X به ترتیب به صورت زیر است.

$$f(x; \lambda, \beta) = \frac{\lambda \beta e^{-\lambda - \beta x + \lambda \exp(-\beta x)}}{(1 - e^{-\lambda})}, x \geq 0 \quad (1)$$

$$F(x; \lambda, \beta) = (e^{\lambda \exp(-\beta x)} - e^{\lambda})(1 - e^{\lambda})^{-1} \quad (2)$$

تعریف ۱. یک متغیر تصادفی با تابع چگالی احتمال (1) و تابع توزیع تجمعی (2) را دارای توزیع نمایی-پواسون می‌نامیم و با $EP(\lambda, \beta)$ نشان می‌دهیم که در آن پارامترهای λ و β به ترتیب پارامتر مقیاس و پارامتر شکل توزیع هستند. در اینجا باید به خوانندگان این مقاله خاطر نشان کرد که توزیع دیگری برای طول عمر، موسوم به توزیع پواسن-نمایی (Poisson-Exponential) در سال ۲۰۱۱ توسط کانچو و همکارانش با تعریف $X = \max(T_1, T_2, \dots, T_W)$ ارایه شده است [4] که مورد بحث این مقاله نیست. توزیع نمایی-پواسون کاربرد گسترده‌ای در اقتصاد، پزشکی، تجارت و علوم مالی یا اکچواری پیدا کرده است. احتمال مشاهده مقدار صفر در توزیع‌هایی مانند گاما، وایبل و لاگ-نرمال که اغلب در آزمایش‌های طول عمر یا بقاء استفاده می‌شوند، صفر است اما در توزیع EP برای هر λ و β مشاهده مقدار صفر، احتمال غیر صفر دارد؛ یعنی $f(0; \lambda, \beta) > 0$. این خاصیت اجازه می‌دهد تا این توزیع به یک گزینه مناسب برای استفاده در مسائلی تبدیل شود که شامل مقادیر صفر است؛ مثلاً در مطالعات محیطی وقوع مقادیر صفر در طول مطالعه بارش، به ویژه در فصول خشک معمول است. چنین رفتاری در مطالعات قابلیت اطمینان نیز یافت می‌شود که خرابی‌های آنی ممکن است به دلیل کیفیت پایین تر یا مشکلات در طول ساخت قطعات رخ دهد. برای شرح جزئیات این توزیع و خواص آن خواننده را به [۱۰] ارجاع می‌دهیم. این مقاله با فرض اینکه که داده‌ها با استفاده از سانسور فاصله‌ای پیش‌رونده نوع I جمع‌آوری شده‌اند، بر مساله روش-های برآورد پارامترهای ناشناخته توزیع EP می‌پردازد. بخش ۲ به روش برآورد حداکثر درست‌نمایی (MLE) و مشکلات پیش رو می‌پردازد. بخش ۳ روش نیوتن رافسون را مورد بحث قرار می‌دهد. در بخش ۴ برآوردها را با استفاده از روش مشهور بیشینه‌سازی مقادیر مورد انتظار (EM) و بیشینه‌سازی تصادفی مقادیر مورد انتظار (SEM) به دست می‌آوریم. در بخش ۵ تجزیه و تحلیل بر مبنای یک مجموعه داده واقعی و شبیه‌سازی شده انجام و نتایج در بخش ۶ ذکر شده است.

۲. برآورد حداکثر درست‌نمایی

هدف این بخش یافتن برآوردهای حداکثر درست‌نمایی (MLE) برای پارامترهای نامعلوم توزیع EP است، هنگامی که داده‌ها با سانسور فاصله‌ای پیش‌رونده نوع I مشاهده می‌شوند. فرض کنید n واحد که طول عمر آنها توزیع $EP(\lambda, \beta)$ دارند در زمان $t_0 = 0$ ، برای آزمایش بقاء استفاده می‌شوند. علاوه بر قراردادهای بخش ۱ فرض کنید $(D_i, R_i, t_i)_{i=1}^m$ مشاهدات ما تحت سانسور فاصله‌ای پیش‌رونده نوع I در زمان‌های از پیش تعیین شده t_i باشند. در این صورت مشاهدات، دارای تابع درست‌نمایی (3) خواهند بود که در آن $F(\cdot; \lambda, \beta)$ تابع توزیع تجمعی (2) است.

$$\mathcal{L}(\lambda, \beta) = \prod_{i=1}^m \{ [F(t_i; \lambda, \beta) - F(t_{i-1}; \lambda, \beta)]^{D_i} \times [1 - F(t_i; \lambda, \beta)]^{R_i} \} \quad (3)$$

با استفاده از $l(\lambda, \beta) = \ln \mathcal{L}(\lambda, \beta)$ ، برآوردهای حداکثر درست‌نمایی (λ, β) ، جواب معادلات (4) و (5) زیر است.

$$\frac{\partial l(\lambda, \beta)}{\partial \lambda} = \frac{e^{\lambda}}{1 - e^{\lambda}} \sum_{i=1}^m (D_i + R_i) + \sum_{i=1}^m D_i \left(\frac{e^{-\beta t_i} e^{\lambda \exp(-\beta t_i)} - e^{-\beta t_{i-1}} e^{\lambda \exp(-\beta t_{i-1})}}{e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})}} \right) + \sum_{i=1}^m R_i \left(\frac{-e^{-\beta t_i} e^{\lambda \exp(-\beta t_i)}}{1 - e^{\lambda \exp(-\beta t_i)}} \right) = 0 \quad (4)$$

$$\frac{\partial l(\lambda, \beta)}{\partial \beta} = \sum_{i=1}^m D_i \left(\frac{-\lambda t_i e^{-\beta t_i} e^{\lambda \exp(-\beta t_i)} + \lambda t_{i-1} e^{-\beta t_{i-1}} e^{\lambda \exp(-\beta t_{i-1})}}{e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})}} \right) + \sum_{i=1}^m R_i \left(\frac{\lambda t_i e^{-\beta t_i} e^{\lambda \exp(-\beta t_i)}}{1 - e^{\lambda \exp(-\beta t_i)}} \right) = 0 \quad (5)$$

با این حال، از آنجایی که معادلات بالا جواب با فرم بسته ندارند، روش های عددی باید اعمال شود. روش نیوتن-رافسون (NR) رایج ترین رویکردی است که در حال حاضر مورد استفاده قرار می گیرد. عیب اصلی این تکنیک این است که به مشتق دوم برای هر تکرار نیاز دارد که به دلیل ساختار پیچیده تابع درست نمایی می تواند محاسباتی سخت و خسته کننده داشته باشد.

۳. نیوتن-رافسون

با روش نیوتن-رافسون (NR) می توانیم مقدار برآورد حداکثر درست نمایی پارامترهای نامعلوم را به طور تقریبی پیدا کنیم. فرض کنید $\theta = (\lambda, \beta)^T$ اگر $\hat{\theta}^0 = (\hat{\lambda}^0, \hat{\beta}^0)^T$ را به عنوان یک تخمین اولیه در این روش در نظر بگیریم، برآورد مرحله

بعد عبارت است از $\hat{\theta}_1 = \hat{\theta}_0 - \left[\frac{\partial l(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}_0} \right]^T \left[\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}_0} \right]^{-1}$ که در آن

$$\frac{\partial l(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial l(\lambda, \beta)}{\partial \lambda} \\ \frac{\partial l(\lambda, \beta)}{\partial \beta} \end{pmatrix}$$

و نیز $\partial l(\lambda, \beta) / \partial \lambda$ و $\partial l(\lambda, \beta) / \partial \beta$ که در معادلات (4) و (5) آمده اند. علاوه بر آن ها

$$\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T} = \begin{pmatrix} \frac{\partial^2 l(\lambda, \beta)}{\partial \lambda^2} & \frac{\partial^2 l(\lambda, \beta)}{\partial \lambda \partial \beta} \\ \frac{\partial^2 l(\lambda, \beta)}{\partial \lambda \partial \beta} & \frac{\partial^2 l(\lambda, \beta)}{\partial \beta^2} \end{pmatrix}$$

که مشتق های جزئی در آن به صورت (6) و (7) و (8) است که در زیر به دست می آیند. $\hat{\theta}_1$ برای تکرار بعدی، به جای $\hat{\theta}_0$ قرار می گیرد و این چرخه ادامه پیدا می کند تا $(\hat{\lambda}, \hat{\beta})$ یعنی جواب های تقریبی معادلات (4) و (5) به دست آید.

$$\begin{aligned} \frac{\partial^2 l(\lambda, \beta)}{\partial \lambda^2} = & \sum_{i=1}^m D_i \frac{\left\{ \left[\exp(-2\beta t_i) e^{\lambda \exp(-\beta t_i)} - \exp(-2\beta t_{i-1}) e^{\lambda \exp(-\beta t_{i-1})} \right] \left[e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})} \right] \right.}{\left[e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})} \right]^2} \\ & \left. - \left[\exp(-\beta t_i) e^{\lambda \exp(-\beta t_i)} - \exp(-\beta t_{i-1}) e^{\lambda \exp(-\beta t_{i-1})} \right] \left[e^{-\beta t_i} e^{\lambda \exp(-\beta t_i)} - e^{-\beta t_{i-1}} e^{\lambda \exp(-\beta t_{i-1})} \right] \right\} \\ & + \sum_{i=1}^m R_i \frac{\left[-\exp(-2\beta t_{i-1}) e^{\lambda \exp(-\beta t_{i-1})} (1 - e^{\lambda \exp(-\beta t_{i-1})}) - \exp(-2\beta t_i) e^{2\lambda \exp(-\beta t_{i-1})} \right]}{(1 - e^{\lambda \exp(-\beta t_{i-1})})^2} - \\ & \frac{e^{\lambda}}{(1 - e^{\lambda})^2} \sum_{i=1}^m (D_i + R_i) \end{aligned} \quad (6)$$

$$\begin{aligned} \frac{\partial^2 l(\lambda, \beta)}{\partial \beta^2} = & \sum_{i=1}^m D_i \frac{\left\{ (1 + \lambda) (t_{i-1} \exp(-\beta t_{i-1}) e^{\lambda \exp(-\beta t_{i-1})} - t_i \exp(-\beta t_i) e^{\lambda \exp(-\beta t_i)}) [e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})}] \right.}{\left[e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})} \right]^2} \\ & \left. + [t_{i-1} \lambda \exp(-\beta t_{i-1}) e^{\lambda \exp(-\beta t_i)} + \lambda t_i \exp(-\beta t_{i-1}) e^{\lambda \exp(-\beta t_{i-1})}] \right\} \\ & \left[\exp(-\beta t_i) e^{\lambda \exp(-\beta t_i)} - \exp(-\beta t_{i-1}) e^{\lambda \exp(-\beta t_{i-1})} \right] \\ & + \sum_{i=1}^m R_i \frac{(1 + \lambda) (t_i \exp(-\beta t_i) e^{\lambda \exp(-\beta t_i)}) (1 - e^{\lambda \exp(-\beta t_{i-1})}) + \lambda t_i \exp(-2\beta t_i) e^{2\lambda \exp(-\beta t_i)}}{(1 - e^{\lambda \exp(-\beta t_{i-1})})^2} \end{aligned} \quad (7)$$

$$\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda \partial \beta} = \sum_{i=1}^m D_i \frac{\left\{ \begin{aligned} &[-t_i \exp(-\beta t_i) e^{\lambda \exp(-\beta t_i)} (1 + \lambda \exp(-\beta t_i)) + t_{i-1} \exp(-\beta t_{i-1}) e^{\lambda \exp(-\beta t_{i-1})} (1 + \lambda \exp(-\beta t_{i-1}))] \\ &[e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})}] + [t_i \lambda \exp(-\beta t_i) e^{\lambda \exp(-\beta t_i)} (1 - \lambda \exp(-\beta t_i))] \\ &[e^{-\beta t_i} e^{\lambda \exp(-\beta t_i)} - e^{-\beta t_{i-1}} e^{\lambda \exp(-\beta t_{i-1})}] \end{aligned} \right\}}{[e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})}]^2} + \sum_{i=1}^m R_i \left[\frac{t_i \exp(-\beta t_i) e^{\lambda \exp(-\beta t_i)} (1 - \lambda \exp(-\beta t_i)) (1 - e^{\lambda \exp(-\beta t_i)}) - (t_i \exp(-\beta t_i) e^{\lambda \exp(-\beta t_i)})}{(1 - \lambda \exp(-\beta t_i))^2} \right] \quad (8)$$

۴. الگوریتم‌های بیشینه‌سازی مقادیر مورد انتظار

۴.۱. الگوریتم EM

دمپستر و همکاران در سال ۱۹۷۷ تکنیک بیشینه‌سازی مقادیر مورد انتظار شرطی (الگوریتم EM) را برای پیدا کردن برآورد حداکثر درست‌نمایی پارامترهای نامعلوم ارائه کردند [۸]. از این روش وقتی استفاده می‌شود که با مشاهدات ناقص (داده‌های گم شده) کار می‌کنیم مثل داده‌های سانسور فاصله‌ای که ناقص ثبت شده‌اند. الگوریتم EM شامل دو مرحله است؛ مرحله E یا پیدا کردن مقادیر مورد انتظار (Expectation) و مرحله M یا بیشینه‌سازی (Maximization). مرحله E در هر تکرار، داده‌های گم شده را با مقادیر مورد انتظار شرطی (توسط مشاهدات ثبت شده و پارامترهای برآورد شده تکرار قبل) در تابع درست‌نمایی مشاهدات کامل جایگزین می‌کند و مرحله M، شامل پیدا کردن برآوردهای حداکثر درست‌نمایی پارامترها از روی لگاریتم تابع درست‌نمایی حاصل از محاسبات مرحله E. در واقع برآوردهای الگوریتم EM به شکل برآوردهای حداکثر درست‌نمایی پارامترهای داده‌های کامل است با این تفاوت که به جای مقادیر ثبت نشده یا ناقص ثبت شده، از مقدار برآورد شده آنها در تکرار قبل استفاده می‌کنیم.

فرض کنید فرض کنید n واحد که طول عمر آنها توزیع $EP(\lambda, \beta)$ دارند در زمان $t_0 = 0$ ، برای آزمایش بقاء استفاده می‌شوند. بعلاوه با قراردادهای بخش ۱ فرض کنید $(D_i, R_i, t_i)_{i=1}^m$ مشاهدات ما تحت سانسور فاصله‌ای پیش‌رونده نوع I در بازه زمان‌های از پیش تعیین شده $[t_{i-1}, t_i]$ باشند. یادآوری می‌شود که در این سانسور به جای زمان دقیق خرابی‌ها، D_i ها که تعداد خرابی‌ها در بازه زمانی $[t_{i-1}, t_i]$ هستند، ثبت می‌شود و بنابراین داده‌های جمع‌آوری شده به نوعی ناقص هستند. حال فرض کنید برای $i = 1, 2, \dots, m$ و $j = 1, 2, \dots, D_i$ طول عمر دقیق کل خرابی‌های مشاهده شده را با x_{ij} ها نشان دهیم همچنین z_{ij} ها برای $i = 1, 2, \dots, m$ و $j = 1, 2, \dots, R_i$ طول عمر دقیق کل واحدهای سانسور شده باشند. بنابراین داده‌های کامل را می‌توان ترکیبی از x_{ij} های ناقص ثبت شده و z_{ij} های ثبت نشده در نظر گرفت. با استفاده از (2)، لگاریتم تابع درست‌نمایی برای نمونه کامل را به شکل زیر داریم:

$$\begin{aligned} \ln \mathcal{L}_c(\lambda, \beta) = & n \ln \lambda + n \ln \beta - n \lambda - n \ln(1 - e^{-\lambda}) - \\ & \beta \sum_{i=1}^m \sum_{j=1}^{D_i} x_{ij} + \lambda \sum_{i=1}^m \sum_{j=1}^{D_i} \exp(-\beta x_{ij}) - \\ & \beta \sum_{i=1}^m \sum_{j=1}^{R_i} z_{ij} + \lambda \sum_{i=1}^m \sum_{j=1}^{R_i} \exp(-\beta z_{ij}) \end{aligned}$$

با مساوی صفر قرار دادن معادلات حاصل از مشتقات جزئی $\ln \mathcal{L}_c(\lambda, \beta)$ نسبت به λ و β به ترتیب روابط (9) و (10) زیر نتیجه می‌شوند.

$$\frac{n}{\lambda} = \frac{n(e^{\lambda} + 1)}{e^{\lambda} - 1} - \sum_{i=1}^m \left[\sum_{j=1}^{D_i} \exp(-\beta x_{ij}) + \sum_{j=1}^{R_i} \exp(-\beta z_{ij}) \right] \quad (9)$$

$$\frac{n}{\beta} = \sum_{i=1}^m \left(\sum_{j=1}^{D_i} x_{ij} + \sum_{j=1}^{R_i} z_{ij} \right) + \lambda \sum_{i=1}^m \left[\sum_{j=1}^{D_i} x_{ij} \exp(-\beta x_{ij}) + \sum_{j=1}^{R_i} z_{ij} \exp(-\beta z_{ij}) \right] \quad (10)$$

بنابراین در مرحله E ابتدا مقادیر مورد انتظار شرطی زیر برای تخمین مقادیر به اصطلاح گم شده محاسبه می شوند.

$$\begin{aligned}
 &= \frac{1-e^{\lambda}}{e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})}} \int_{t_{i-1}}^{t_i} e^{-\beta x} f(x) dx \\
 E_{2i}(\lambda, \beta) &= E(\exp(-\beta z_{ij}) | z_{ij} \in (t_i, \infty)) = \frac{1-e^{\lambda}}{1-e^{\lambda \exp(-\beta t_i)}} \int_{t_i}^{\infty} e^{-\beta x} f(x) dx \\
 E_{3i}(\lambda, \beta) &= E(x_{ij} | x_{ij} \in (t_{i-1}, t_i]) = \frac{1-e^{\lambda}}{e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})}} \int_{t_{i-1}}^{t_i} x f(x) dx \\
 E_{4i}(\lambda, \beta) &= E(z_{ij} | z_{ij} \in (t_i, \infty)) \\
 &= \frac{1-e^{\lambda}}{1-e^{\lambda \exp(-\beta t_i)}} \int_{t_i}^{\infty} x f(x) dx \\
 E_{5i}(\lambda, \beta) &= E(x_{ij} \exp(-\beta x_{ij}) | x_{ij} \in (t_{i-1}, t_i]) = \frac{1-e^{\lambda}}{e^{\lambda \exp(-\beta t_i)} - e^{\lambda \exp(-\beta t_{i-1})}} \int_{t_{i-1}}^{t_i} x e^{-\beta x} f(x) dx \\
 E_{6i}(\lambda, \beta) &= E(z_{ij} \exp(-\beta z_{ij}) | z_{ij} \in (t_i, \infty)) = \frac{1-e^{\lambda}}{1-e^{\lambda \exp(-\beta t_i)}} \int_{t_i}^{\infty} x e^{-\beta x} f(x) dx
 \end{aligned}$$

حال اگر $(\lambda^{(0)}, \beta^{(0)})$ را به عنوان یک تخمین اولیه برای مقدار واقعی (λ, β) در نظر بگیریم، برآورد (λ, β) در تکرار (k) الگوریتم EM را با $(\lambda^{(k)}, \beta^{(k)})$ نشان دهیم، مرحله M با استفاده از روابط (9) و (10) برآورد پارامترها در تکرار $(k+1)$ به صورت (11) و (12) زیر خواهد بود.

زمانی که $|\lambda^{(k+1)} - \lambda^{(k)}| + |\beta^{(k+1)} - \beta^{(k)}| < \epsilon$ برای یک مقدار خیلی کوچک و مثبت ϵ صدق کند، همگرایی مطلوب وجود دارد و این الگوریتم می تواند خاتمه یابد.

ملاحظه می شود که مقادیر مورد انتظار بالا، فرم بسته ندارند و روش های عددی لازم است اما به هر حال با نرم افزارهایی مانند R قابل محاسبه هستند. در زیربخش بعد به عنوان راه حل دیگر، الگوریتم بیشینه سازی تصادفی مقادیر مورد انتظار (SEM) را استفاده می کنیم.

$$\lambda^{(k+1)} = n \left[\frac{n(e^{\lambda^{(k)}} + 1)}{e^{\lambda^{(k)}} - 1} - \sum_{i=1}^m (D_i E_{1i}(\lambda^{(k)}, \beta^{(k)}) + R_i E_{2i}(\lambda^{(k)}, \beta^{(k)})) \right]^{-1} \quad (11)$$

$$\beta^{(k+1)} = n \left[\sum_{i=1}^m (D_i E_{3i}(\lambda^{(k)}, \beta^{(k)}) + R_i E_{4i}(\lambda^{(k)}, \beta^{(k)})) + \lambda^{(k)} \sum_{i=1}^m (D_i E_{5i}(\lambda^{(k)}, \beta^{(k)}) + R_i E_{6i}(\lambda^{(k)}, \beta^{(k)})) \right]^{-1} \quad (12)$$

۲.۴. الگوریتم SEM

در اجرای مرحله E الگوریتم EM ممکن است محاسبات پیچیده و غیرقابل حل دستی وجود داشته باشد. روش های مختلفی از طرف پژوهشگران برای کم کردن هزینه های محاسباتی موجود پیشنهاد شده است. به عنوان نمونه می توانیم به استفاده از روش انتگرال گیری مونت کارلو در [19] اشاره کنیم. با این وجود، اغلب در مرحله M، مساله بیشینه کردن همچنان پیچیده و زمان بر باقی می ماند [20]. در [9] خواص مجانبی برآوردهای حاصل از الگوریتم SEM با ایده جایگزینی مرحله E با یک مرحله تصادفی (Stochastic) و اجرای آن از طریق شبیه سازی ارایه شد. در این الگوریتم به جای ماکزیمم کردن امید ریاضی شرطی لگاریتم تابع درست نمایی مشاهدات کامل، ابتدا داده های گم شده را از توزیع شرطی داده های گم شده (به شرط مشاهدات ثبت شده و پارامترهای برآورد شده تکرار قبل) تولید می کنیم سپس مرحله M اجرا می شود. نشان

داده شده است که روش SEM از نظر بار محاسباتی از روش EM مناسب‌تر است [۲۲] و [۲۳]. در اینجا از این ایده استفاده خواهیم کرد.

ابتدا فرض کنید تعداد D_i مقدار مستقل x_{ij} را برای $i = 1, 2, \dots, m$ و $j = 1, 2, \dots, D_i$ از توزیع شرطی زیر داریم:

$$F_X(x_{ij}; \lambda, \beta | x_{ij} \in (t_{i-1}, t_i]) \\ = \frac{F(x_{ij}; \lambda, \beta) - F(t_{i-1}; \lambda, \beta)}{F(t_i; \lambda, \beta) - F(t_{i-1}; \lambda, \beta)}, \quad x_{ij} \in (t_{i-1}, t_i]$$

بعلاوه تعداد R_i مقدار مستقل z_{ij} را برای $i = 1, 2, \dots, m$ و $j = 1, 2, \dots, D_i$ از توزیع شرطی زیر داریم:

$$F_Z(z_{ij}; \lambda, \beta | z_{ij} > t_i) \\ = \frac{F(z_{ij}; \lambda, \beta) - F(t_i; \lambda, \beta)}{1 - F(t_i; \lambda, \beta)}, \quad z_{ij} > t_i$$

حال با استفاده از روابط (9) و (10) برآورد MLE پارامترها در تکرار $(k+1)$ در روش SEM از روابط (13) و (14) به دست خواهد آمد. ما در بخش ۱،۵ ابتدا با استفاده از ۱۰۰۰۰ بار شبیه‌سازی مقادیر گم شده از توزیع‌های شرطی بالا، مقادیر مورد انتظار را با روش میانگین‌گیری مونت کارلو حساب و سپس برآورد پارامترها را به دست آورده‌ایم. در ضمن این چرخه را ۱۰۰ بار تکرار کرده‌ایم و میانگین برآوردها را گزارش کرده‌ایم. اگر برآورد پارامترها را بعد از خاتمه یافتن الگوریتم‌ها با $(\hat{\lambda}, \hat{\beta})$ نشان دهیم، ماتریس اطلاع فیشر مشاهده شده را به صورت زیر می‌توان به دست آورد که درایه‌های آن با روابط (6) تا (8) که در بخش ۳ داده شده است، محاسبه می‌شوند.

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \left[\begin{array}{cc} -\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda^2} & -\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda \partial \beta} \\ -\frac{\partial^2 l(\lambda, \beta)}{\partial \lambda \partial \beta} & -\frac{\partial^2 l(\lambda, \beta)}{\partial \beta^2} \end{array} \right]_{\lambda=\hat{\lambda}, \beta=\hat{\beta}}^{-1}$$

فاصله‌های اطمینان تقریبی $100(1 - \alpha)\%$ برای λ و β به ترتیب $\hat{\lambda} \pm Z_{\alpha/2} \sqrt{\sigma_{11}}$ و $\hat{\beta} \pm Z_{\alpha/2} \sqrt{\sigma_{22}}$ خواهند بود که منظور از Z_γ صدک $100(1 - \gamma)$ توزیع نرمال استاندارد است.

$$\lambda^{(k+1)} = n \left[\frac{n(e^{\lambda^{(k)}} + 1)}{e^{\lambda^{(k)}} - 1} - \sum_{i=1}^m \left(\sum_{j=1}^{D_i} E[\exp(-\beta^{(k)} x_{ij})] + \sum_{j=1}^{R_i} E[\exp(-\beta^{(k)} z_{ij})] \right) \right]^{-1} \quad (13)$$

$$\beta^{(k+1)} = n \left[\sum_{i=1}^m \left(\sum_{j=1}^{D_i} E[x_{ij}] + \sum_{j=1}^{R_i} E[z_{ij}] \right) + \lambda^{(k)} \sum_{i=1}^m \left(\sum_{j=1}^{D_i} E[x_{ij} \exp(-\beta^{(k)} x_{ij})] + \sum_{j=1}^{R_i} E[z_{ij} \exp(-\beta^{(k)} z_{ij})] \right) \right]^{-1} \quad (14)$$

۵. تحلیل داده‌های شبیه‌سازی و واقعی

۱،۵. شبیه‌سازی

هدف این بخش بررسی عملکرد برآوردگرهای پیشنهادی از طریق یک مطالعه شبیه‌سازی است. ابتدا برای m و n معلوم، زمان‌های بازرسی t_i از پیش تعیین شده و q_i های از پیش تعیین شده، داده‌های $(D_i, R_i, t_i)_{i=1}^m$ را تحت سانسور فاصله‌ای نوع I با استفاده از الگوریتمی که در [2] پیشنهاد شده است، به شرح زیر تولید می‌کنیم.

بر طبق این الگوریتم ابتدا $D_1 \sim \text{Bin}(n, F(t_1; \lambda, \beta))$ را تولید می‌کنیم. سپس $R_1 = [q_1 \times (n - D_1)]$ تولید می‌شود. بعلاوه با $R_i = [q_i \times (n - \sum_{j=1}^{t_i-1} (D_j + R_j))]$ ها D_i ها برای $i = 2, 3, \dots, m$ از توزیع دو جمله‌ای زیر تولید می‌شوند.

$$D_i | (D_{i-1}, R_{i-1}, \dots, D_1, R_1) \sim$$

$$Bin\left(n - \sum_{j=1}^{i-1} (D_j + R_j), \frac{F(t_i; \lambda, \beta) - F(t_{i-1}; \lambda, \beta)}{1 - F(t_{i-1}; \lambda, \beta)}\right)$$

در شبیه سازی، برای تعداد واحدهای تحت آزمایش مقادیر $n = \{20, 30, 40, 50, 100, 150, 200\}$ و زمان های $t_1 = 0.1, t_2 = 0.3, t_3 = 0.5, t_4 = 0.7, t_5 = 0.9$ را در نظر گرفتیم. همچنین سه طرح زیر را برای سانسور استفاده کردیم.

$$p1 = (q_1 = 0.25, q_2 = 0.25, q_3 = 0, q_4 = 0, q_5 = 1)$$

$$p2 = (q_1 = 0, q_2 = 0, q_3 = 0.25, q_4 = 0.25, q_5 = 1)$$

$$p3 = (q_1 = 0, q_2 = 0, q_3 = 0, q_4 = 0, q_5 = 1)$$

توجه کنید که طرح $p3$ همان سانسور فاصله ای معمولی نوع I است. با استفاده از این مقادیر بالا و توزیع های $EP(0.5, 2.5)$ و $EP(2, 1)$ ، داده های مورد نیاز را تولید کردیم. میانگین برآوردهای حداکثر درست نمایی پارامترها و مقادیر میانگین مربعات خطای (MSE) آن ها به دست آمده بر اساس ۱۰۰ تکرار با استفاده از الگوریتم های NR، EM و SEM در جدول (1) و جدول (2) به ترتیب برای هر دو توزیع گزارش شده است.

جدول (1) برآوردهای MLE برای پارامترهای توزیع $EP(0.5, 2.5)$ از سه الگوریتم با ۱۰۰ تکرار

	الگوریتم	SEM						EM						NR					
		n	20	30	50	100	150	200	20	30	50	100	150	200	20	30	50	100	150
طرح p1	$\hat{\lambda}$	0.5179	0.5127	0.5167	0.4888	0.4931	0.4954	0.8395	0.6997	0.7974	0.5555	0.7009	0.6309	2.0787	1.2514	1.0459	0.9235	0.9064	0.8832
	$MSE(\hat{\lambda})$	0.0629	0.0555	0.0573	0.0006	0.0014	0.0003	0.9813	0.5128	0.6833	0.1573	0.3743	0.2571	0.8513	0.0611	0.0324	0.3935	0.7058	0.0678
	$\hat{\beta}$	2.6906	2.5658	2.4847	2.553	2.4802	2.5075	2.5379	2.4718	2.3423	2.5187	2.3621	2.4478	3.0129	2.8877	2.7774	2.7297	2.6565	2.6191
	$MSE(\hat{\beta})$	0.834	0.2694	0.151	0.0897	0.0495	0.0374	1.1434	0.4359	0.3967	0.1607	0.1983	0.1451	10.08	5.2579	1.4416	4.3872	7.7222	7.7135
	$\hat{\lambda}$	0.5637	0.4906	0.4945	0.4848	0.4928	0.4924	0.8632	0.8252	0.7415	0.7312	0.5913	0.6074	0.8262	0.81	0.8223	0.8056	0.8055	0.8096
طرح p2	$MSE(\hat{\lambda})$	0.1816	0.0046	0.002	0.0009	0.0005	0.0003	0.9317	0.7732	0.6126	0.515	0.2012	0.1978	7.3298	2.4317	0.0086	0.6628	1.7693	12.689
	$\hat{\beta}$	2.5552	2.544	2.583	2.5066	2.506	2.5012	2.4328	2.3778	2.464	2.3757	2.5051	2.4952	2.6785	2.683	2.6695	2.6621	2.6437	2.6275
	$MSE(\hat{\beta})$	0.6085	0.2647	0.214	0.0683	0.0617	0.0513	0.9754	0.5834	0.4663	0.3018	0.1768	0.1513	0.2125	13.773	12.546	0.5896	1.5333	0.3366
	$\hat{\lambda}$	0.6001	0.5989	0.4877	0.4871	0.4918	0.4991	0.849	1.0981	0.6703	0.7397	0.6319	0.5678	0.7641	0.7854	0.7728	0.7636	0.7639	0.7544
	$MSE(\hat{\lambda})$	0.2803	0.2499	0.0012	0.0005	0.0003	0.0002	1.092	1.3513	0.4407	0.5285	0.2597	0.121	12.916	2.5614	0.0468	0.8437	0.309	0.315
طرح p3	$\hat{\beta}$	2.5304	2.5371	2.4851	2.4982	2.5296	2.5074	2.4338	2.2784	2.396	2.3613	2.4501	2.4604	2.6629	2.6471	2.6405	2.6332	2.621	2.6149
	$MSE(\hat{\beta})$	0.4267	0.3517	0.1666	0.0752	0.0427	0.038	0.7422	0.7607	0.3565	0.2821	0.155	0.0956	12.325	0.0103	6.3219	8.656	1.8806	5.561

جدول (2) برآوردهای MLE برای پارامترهای توزیع $EP(2, 1)$ از سه الگوریتم با ۱۰۰ تکرار

		SEM						EM						NR					
		n	20	30	50	100	150	200	20	30	50	100	150	200	20	30	50	100	150
طرح p1	$\hat{\lambda}$	1.8106	1.9215	1.9835	2.0132	2.0169	2.0141	1.5564	1.7435	1.7833	1.811	1.828	1.8853	1.5677	2.1743	2.2905	2.2168	2.0856	2.0142
	$MSE(\hat{\lambda})$	0.3551	0.169	0.0834	0.0251	0.0095	0.0079	1.4338	1.2677	0.4472	0.3446	0.2545	0.1736	2.4551	0.4235	2.6584	7.7856	1.4356	0.6324
	$\hat{\beta}$	1.1255	1.0688	1.0317	1.0011	0.9907	0.9865	1.2884	1.1853	1.1265	1.1014	1.0756	1.0408	1.7261	1.5776	1.4859	1.4324	1.3954	1.3608
	$MSE(\hat{\beta})$	0.1738	0.1245	0.043	0.0143	0.0053	0.0044	0.4093	0.2649	0.1169	0.1105	0.0618	0.0303	0.9809	2.1193	11.139	0.2507	1.5681	0.1182
طرح p2	$\hat{\lambda}$	1.9307	1.9586	2.0308	1.9936	2.0281	2.002	1.7699	1.7516	1.7913	1.8357	1.8973	1.8784	2.6177	2.7228	2.7294	2.7525	2.6377	2.5575
	$MSE(\hat{\lambda})$	0.1989	0.196	0.0291	0.0189	0.0129	0.0093	1.2597	1.0446	0.3735	0.2448	0.1806	0.175	0.149	0.0001	0.6559	1.3266	0.0065	0.1357
	$\hat{\beta}$	1.0569	1.0531	1.0036	0.9731	0.9998	0.9772	1.1901	1.1931	1.1235	1.0527	1.061	1.0349	1.3993	1.434	1.438	1.4197	1.4105	1.3938
	$MSE(\hat{\beta})$	0.1359	0.1218	0.0161	0.0116	0.0069	0.0057	0.3492	0.3214	0.1277	0.0886	0.0502	0.0477	12.575	8.4211	1.1836	1.3706	0.4117	0.6881
طرح p3	$\hat{\lambda}$	1.9138	1.9686	1.9776	2.0162	2.025	2.0334	1.793	1.9248	1.7016	1.8172	1.8136	1.9303	2.965	3.0505	3.0197	2.9586	2.9141	2.8663
	$MSE(\hat{\lambda})$	0.2723	0.1668	0.0406	0.0139	0.0121	0.0093	1.2394	0.6003	0.4645	0.2814	0.2858	0.1403	15.366	0.9574	0.4108	2.6016	0.6358	16.606
	$\hat{\beta}$	1.098	1.0756	1.0061	0.9894	0.9939	1.0032	1.1871	1.1217	1.1503	1.0831	1.0916	1.0498	1.4026	1.4013	1.4011	1.3934	1.3846	1.3729
	$MSE(\hat{\beta})$	0.1321	0.0802	0.0323	0.0081	0.0066	0.005	0.2417	0.1631	0.1878	0.0798	0.0772	0.0366	2.1271	3.6771	1.1819	0.0082	3.2348	0.4609

در این جدول‌ها برآوردهای به دست آمده با استفاده از الگوریتم‌های EM و SEM تقریباً برای همه طرح‌های در نظر گرفته شده به هم نزدیک هستند. با این حال، برآوردهای روش NR اندکی بیشتر از برآوردهای روش‌های دیگر است. MSE برآوردها با استفاده از روش SEM در مقایسه با روش‌های دیگر کوچکتر است. از سوی دیگر، روش NR دارای کمترین کارایی بین روش‌های دیگر است. نتایج نشان می‌دهد که با افزایش حجم نمونه، مقادیر برآورد شده پارامترها به مقادیر واقعی نزدیکتر شده و کارایی نیز افزایش می‌یابد.

جدول (3) و جدول (4) که به ترتیب از جدول (1) و جدول (2) به دست آمده‌اند، میانگین مقادیر برآورد پارامترها و میانگین MSE های آن‌ها را برای سه الگوریتم را نشان می‌دهند. جدول (3) نشان می‌دهد که نتایج روش SEM نزدیک‌ترین تخمین‌ها (یعنی ۵۱۱۵۶۱، ۰، ۵۳۱۴۲۲) را به مقدار واقعی پارامترها (یعنی ۵، ۰، ۲، ۵) دارد و میانگین مقادیر MSE برای این الگوریتم، حداقل است. جدول (4) نیز نشان می‌دهد که نتایج روش SEM نزدیک‌ترین تخمین‌ها (یعنی ۱، ۰۲۴۱۲۲، ۱۰، ۲) را به مقدار واقعی پارامترها (یعنی ۲، ۱۰) دارد همچنین میانگین مقادیر MSE نیز برای این الگوریتم، حداقل است.

جدول (3) میانگین برآوردهای پارامترهای توزیع EP(0.5, 2.5) از سه الگوریتم با ۱۰۰ تکرار

الگوریتم	SEM	EM	NR
میانگین $\hat{\lambda}$ ها	0.511561	0.730028	0.920694
میانگین $MSE(\hat{\lambda})$ ها	0.050017	0.555044	2.444167
میانگین $\hat{\beta}$ ها	2.531422	2.428400	2.692622
میانگین $MSE(\hat{\beta})$ ها	0.211133	0.418178	5.574900

جدول (4) میانگین برآوردهای پارامترهای توزیع EP(2, 1) از سه الگوریتم با ۱۰۰ تکرار

الگوریتم	SEM	EM	NR
میانگین $\hat{\lambda}$ ها	1.979900	1.806233	2.563383
میانگین $MSE(\hat{\lambda})$ ها	0.090561	0.567294	3.013444
میانگین $\hat{\beta}$ ها	1.024122	1.119828	1.434967
میانگین $MSE(\hat{\beta})$ ها	0.051533	0.153594	2.862017

۲.۵. تحلیل یک مجموعه داده واقعی

برای نشان دادن عملکرد روش‌های مورد استفاده در این مقاله، داده‌های جمع‌آوری شده در تحقیقی را بررسی می‌کنیم که شامل زمان بقاء برای ۱۱۲ بیمار مبتلا به سرطان پوست بدخیم با منشا سلول‌های خونی است که در موسسه ملی سرطان (N.C.I) درمان شده بودند [7]. این مجموعه داده در جدول (۵) ارائه شده است که در آن تعداد بیمارانی را که در پایان بازه‌های زمانی $[t_{i-1}, t_i]$ به علت زنده ماندن از مطالعه کنار گذاشته شده‌اند را با D_i ها نشان داده‌ایم. همچنین تعداد بیمارانی که به میل خودشان از ادامه درمان انصراف داده‌اند را با R_i ها نشان داده‌ایم. اگر چه این اعداد سانسور پیش‌رونده قبل از تحقیق تعیین نشده‌اند، این داده‌ها برای برازش توزیع وایبل با پارامتر شکل λ و پارامتر مقیاس β در [12] استفاده شده است. ما نیز در این مقاله استنباط‌های آماری را بر اساس آن‌ها انجام داده‌ایم. نتایج برازش توزیع $EP(\lambda, \beta)$ به داده‌ها شرح زیر است.

برآوردهای پارامترها در جدول (6) داده شده‌اند. آماره آزمون نیکویی برازش کلموگروف-اسمیرنوف (K-S) را در اینجا به صورت $D_n(F) = \sup_{0 \leq t_i \leq \infty} |\hat{F}(t_i; \lambda, \beta) - F(t_i; \hat{\lambda}, \hat{\beta})|$ گرفته‌ایم که در آن $\hat{\lambda}$ و $\hat{\beta}$ برآورد پارامترها از روش SEM است و نیز توزیع تجربی برای مشاهدات را به صورت زیر گرفته‌ایم.

$$\hat{F}(t_i; \lambda, \beta) = 1 - \prod_{j=1}^i (1 - \hat{p}_j) \quad , \quad i = 1, 2, \dots, m$$

که در آن $\hat{p}_1 = \frac{D_1}{n}$ و همچنین $\hat{p}_j = \frac{D_j}{n - \sum_{k=1}^{j-1} (D_k + R_k)}$ برای $i = 2, 3, \dots, m$. مقدار آماره آزمون کلموگروف-اسمیرنوف برای برازش توزیع EP به مشاهدات 0.0961 و همچنین مقدار شاخص منفی لگاریتم تابع درست‌نمایی مشاهدات 194.55 به دست آمد که این مقادیر برای برازش توزیع وایبل در [12] به ترتیب 0.0972 و 198.43 گزارش شده است. همچنین بر اساس مشاهدات جمع‌آوری شده جدول (5) بازه‌های اطمینان 95% تقریبی برای λ و β به ترتیب (0.8153, 0.8578) و (0.0334, 0.0368) به دست آمد.

جدول (5) داده‌های جمع‌آوری شده بیماران در موسسه ملی سرطان (N.C.I)

R_i	D_i	تعداد بیماران تحت درمان	بازه زمانی درمان (برحسب ماه)
1	۱۸	112	[0, 5.50]
1	۱۶	93	[5.5, 10.5]
3	۱۸	76	[10.5, 15.5]
0	۱۰	55	[15.5, 20.5]
0	۱۱	45	[20.5, 25.5]
1	۸	34	[25.5, 30.5]
2	۱۳	25	[30.5, 40.5]
3	۴	10	[40.5, 50.5]
2	۱	3	[50.5, 60.5]
0	۰	0	[60.5, ∞)

جدول (6) برآوردهای پارامترهای توزیع $EP(\lambda, \beta)$ برای داده‌های جدول (5) از سه الگوریتم

الگوریتم	SEM	EM	NR
$\hat{\lambda}$	0.83657	0.83664	0.89872
$\hat{\beta}$	0.03512	0.03511	0.034314

۶. نتیجه‌گیری

در این مقاله ما داده‌های طول عمر را که دارای توزیع EP هستند، تحت سانسور بازه‌ای پیش رونده نوع I در نظر گرفتیم. مشاهده کردیم که MLEهای پارامترهای ناشناخته توزیع، فرم بسته‌ای ندارند و اجرای الگوریتم EM در مرحله E نیاز به روش‌های عددی دارد. به عنوان یک راه حل این مشکلات، الگوریتم SEM را با استفاده از روش مونت کارلو برای بدست آوردن MLEها در نظر گرفتیم. شبیه‌سازی نشان داد که در مقایسه بین برآوردهای به دست آمده توسط الگوریتم SEM و برآوردهای الگوریتم‌های NR و EM، عملکرد الگوریتم SEM براساس میانگین مربعات خطا کاملاً رضایت‌بخش است و با افزایش حجم نمونه، مقادیر برآورد شده پارامترها به مقدار واقعی آن‌ها نزدیکتر می‌شود و کارایی برآوردکننده‌ها نیز افزایش می‌یابد. یک مجموعه داده واقعی نیز به عنوان یک مثال در نظر گرفتیم و مشاهده کردیم با برآوردهای روش SEM، توزیع EP به داده‌های جمع‌آوری شده برازش قابل قبول دارد.

COPYRIGHTS

© 2022 by the authors. Licensee PNU, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY4.0) (<http://creativecommons.org/licenses/by/4.0>)

- [1] Arabi Belaghi, R. & Noori Asl, M. & Sukhdev Singh. (2017). On estimating the parameters of the Burr XII model under progressive type-I interval censoring. *Journal of Statistical Computation and Simulation*, 87(16), 3132-3151.
- [2] Aggarwala, R. (2001). Progressive interval censoring: Some mathematical results with applications to inference. *Communications in Statistics-Theory and Methods*, 30, 1921-1935.
- [3] Ashour, S.K., & Afify, W.M. (2007). Statistical analysis of exponentiated Weibull family under type I progressive interval censoring with random removals. *Journal of Applied Sciences Research* 3(12), 1851-1863.
- [4] Cancho, V.G., & Louzada-Neto, F., & Barriga, G.D.C. (2011). The Poisson-Exponential lifetime model. *Computational Statistics and Data Analysis*, 55(1), 677-686.
- [5] Chen, D.G, & Lio Y.L. (2010). Parameter estimations for generalized Exponential distribution under progressive type-I interval censoring. *Computational Statistics and Data Analysis*, 54(6), 1581-1591.
- [6] Chen, D.G, & Lio Y.L, & Jiang, N. (2013). Lower confidence limits on the generalized Exponential distribution percentiles under progressive type-I interval censoring. *Communications in Statistics-Simulation and Computation*, 42(9), 2106-2117.
- [7] Carbone, P.P., & Kellerhouse, L.E. & Gehan, E.A. (1967). Plasmacytic myeloma: A study of the relationship of survival to various clinical manifestations and anomalous protein type in 112 patients. *The American Journal of Medicine*, 42(6), 937-948.
- [8] Dempster, A.P., & Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society series B (Methodological)*, 39(1), 1-22.
- [9] Diebolt, J., & Celeux, G. (1993). Asymptotic properties of a stochastic EM algorithm for estimating mixing proportions. *Communications in Statistics-Stochastic Models*, 9(4), 599-613.
- [10] Kus, C. (2007). A new lifetime distribution. *Computational Statistics and Data Analysis*, 51(9), 4497-4509.
- [11] Lio, Y.L, & Chen, D.G., & Tsai, T.R. (2011). Parameter estimations for generalized Rayleigh distribution under progressively type-I interval censored data. *Open Journal of Statistics*, 1(2), 46- 57.
- [12] Ng, H.K.T, & Wang, Z. (2009). Statistical estimation for the parameters of Weibull distribution based on progressively type-I interval censored sample. *Journal of Statistical Computation and Simulation*, 79(2), 145-159.
- [13] Pradhan, B., Gijo, E. (2013). Parameter estimation of lognormal distribution under progressive type-I interval censoring. Technical report no. sqcor-2013-02.
- [14] Peng, X.Y., & Yan, Z.Z. (2013). Bayesian estimation for generalized Exponential distribution based on progressive type-I interval censoring. *Acta Mathematicae Applicatae Sinica, English Series* , 29, 391-402.
- [15] Lin, Y.J., & Lio, Y.L. (2012) Bayesian inference under progressive type-I interval censoring. *Journal of Applied Statistics*, 39, 1811-1824.

- [16] Lin, C.T., & Wu, S.J.S., & Balakrishnan, N. (2009). Planning life tests with progressively type-I interval censored data from the lognormal distribution. *Journal of Statistical Planning and Inference*, 139(1), 54–61.
- [17] Wingo, D.R. (1993). Maximum likelihood estimation of burr xii distribution parameters under type II censoring. *Microelectronics Reliability*, 33(9), 1251–1257.
- [18] Singh, S., & Tripathi, Y.M. (2018). Estimating the parameters of an inverse Weibull distribution under progressive type-I interval censoring. *Statistical Papers*, 59, 21–56.
- [19] Wei, G.C.G., & Tanner, M.A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85, 699–704.
- [20] Wang, F.K., & Cheng, Y.F. (2010). EM algorithm for estimating the burr XII parameters with multiple censored data. *Quality and Reliability Engineering*, 26(6), 615–630.
- [21] Xiuyun, P., & Zaizai, Y.L. (2011). Parameter estimations with gamma distribution based on progressive type-I interval censoring. 2011 IEEE International Conference on Computer Science and Automation Engineering, Shanghai, China, 1, 449–453.
- [22] Tregouet, D.A., & Escolano, S., & Tired L, et al. (2004). A new algorithm for haplotype-based association analysis: the stochastic EM algorithm. *Annals of Human Genetics*, 68(2), 165–177.
- [23] Zhang, M., & Ye, Z., & Xie, M. (2014). A stochastic EM algorithm for progressively censored data analysis. *Quality and Reliability Engineering International*, 30(5), 711–722.