اندیشه آماری، بهار و تابستان ۱۳۹۱، شماره پیاپی ۳۳

سال هفدهم شماره اول، ص ۱۵-۲۸

پیش‌بینی بیزی مشاهده آینده بر اساس نمونه سانسور شده دوطرفه در مدل نمایی

جعفر احمدی [۱] ، منصور کرم زاده [۲]

چکیده:

در بسیاری تحقیقات مربوط به آزمونهای طول عمر، با محدودیتهایی مانند زمان و تعداد واحد نمونه مواجه هستیم، که این عوامل باعث می شود پژوهشگر دسترسی به کل داده‌ها نداشته باشد. بنابراین پیش بینی مقادیر آینده بر اساس اطلاعات واحدهایی از نمونه که در دسترس می‌باشند، ارزش مطالعه دارد. در این مقاله با فرض اینکه مشاهدات اصلی از مدل نمایی پیروی می‌کنند و محدودیت اعمال شده بر داده‌ها سانسور دوطرفه است، پیش بینی مقادیر آینده را از دیدگاه آمار بیز در دو حالت یک نمونه ای و دو نمونه ای مورد بررسی قرار داده ایم. در هر حالت پیش بینی فاصله ای با احتمال پوشش داده شده به‌دست می‌آید. در پایان نتایج روی یک مثال عددی اعمال شده است.

واژه‌های کلیدی: آماره های ترتیبی، پیش بینی بیزی، توزیع پیشین مزدوج، مدل نمایی، سانسور دوطرفه.

[۱] گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

[۲] گروه ریاضی، دانشکده علوم، دانشگاه یاسوج

۱ مقدمه

گوییم متغیر تصادفی X دارای توزیع نمایی با پارامتر θ است ، هرگاه تابع توزیع آن به صورت زیر باشد.

$$F_X(t;\theta) = 1 - \exp(-\theta t)\ , t > 0 \quad (۱)$$

که در آن θ پارامتر مقیاس می‌باشد. بنابراین تابع چگالی احتمال X عبارت است از:

$$f_X(t;\theta) = \theta \exp(-\theta t)\ , t > 0. \quad (۲)$$

این توزیع یکی از مدلهای مهم در زمینه مطالعات مربوط به طول عمر سیستمها می‌باشد. از جمله کاربردهای معروف آن فاصله زمانی بین رویدادها در فرآیند پواسن می‌باشد، بطوری‌که برای مدل بندی این فرآیند می‌توان از آزمون نیکویی برازش توزیع نمایی استفاده کرد. معروف ترین ویژگی این توزیع، خاصیت فقدان حافظه است که عامل تشخیص و توصیف مدل نمایی می‌باشد. بالاکریشنان و باسو در سال (۱۹۹۵) [۳] مجموعه تحقیقات انجام شده در خصوص این توزیع را تا آن زمان جمع آوری نموده‌اند که می‌تواند منبع مفیدی برای علاقمندان به تحقیق درباره خواص توزیع نمایی باشد. در بسیاری اوقات درآزمونهای طول عمر، آزمایشهای کلینیکی، تحلیل بقا و دیگر زمینه های کاربردی علم آمار با نمونه‌هایی مواجه هستیم که نمونه مورد مطالعه محدود شده و همه مشاهدات ثبت نشده‌اند یا آزمایش تا به نتیجه رسیدن همه واحدهای نمونه جریان نداشته است. این محدودیت ممکن است به صورت اختیاری توسط محقق اعمال شود یا ماهیت آزمایشها طوری باشند که خود بخود محدودیتی را در مشاهدات به وجود آورند. بعضی از این محدودیتها عبارتند از: فرصت کم برای اعلام نتایج، طولانی شدن مدت آزمایش، عدم دسترسی به همه واحدها و یا مایوس شدن از نتیجه دادن همه واحدهای نمونه. به این محدودیت‌ها در متون آماری سانسور مشاهدات گویند که به صورتهای مختلف اعمال می‌شوند. برای بررسی جزئیات بیشتر در مورد انواع سانسور، نظریه و کاربرد آن‌ها می‌توان به لاوولس(۲۰۰۳) [۴] ، نلسن(۲۰۰۴) [۵] و همچنین بالاکریشنان(۲۰۰۷) مراجعه نمود.

وقتی صحبت از داده های سانسورشده می‌شود سوال مهم چگونگی به‌دست آوردن اطلاعات در مورد مقادیر مشاهده نشده از این داده‌ها است. یکی از راه حلهای مفید، می‌تواند پیش بینی این مقادیر باشد. امروزه پیش بینی در اکثر زمینه ها از جمله پدیده های طبیعی مانند زلزله و سیل، فعال شدن آتشفشان، تغییرات جوی و حوادثی نظیر سونامی (که در دی ماه سال ۱۳۸۳ خسارت جانی و مالی زیادی در آسیای جنوب شرقی وارد نمود) نقش مهمی‌ایفا می‌کند. از دیگر موارد استفاده از نظریه

[۳] Balakrishnan and Basu

[۴] Lawless

[۵] Nelson

پیش بینی، مطالعه طول عمر در سیستمهای مهندسی می‌باشد. برای مثال در یک سیستم مهندسی مو لفه‌های سیستم به نحوی نصب شده‌اند که درصورت خراب شدن یک مولفه، مولفه یدک به جای آن قرار می‌گیرد. بنابراین پیش بینی طول عمر مو لفه یدک تا زمان تعمیر مولفه اول ارزش مطالعه دارد. با توجه به اینکه پارامتر جامعه ممکن است تحت تاثیر عوامل آزمایش تغییر کند لذا همیشه نمی‌توان آنرا مقداری ثابت تلقی کرد. بنابراین در چنین شرایطی استفاده از اطلاعات پیشین پارامتر در این کار و در نتیجه پیش بینی بیزی اهمیت پیدا می‌کند. محققین زیادی مسئله پیش بینی را از دیدگاه آمار بیز مطالعه نموده‌اند. از جمله گیسر (۱۹۸۵; ۱۹۸۴) [۶] نیگم و همدی (۱۹۸۷) [۷] ، آرنولد و پرس (۱۹۸۹) [۸] ، دنسمر و امین(۱۹۹۸) [۹] ، در حالتهای مختلف پیش بینی بیزی را در مدل پارتو انجام داده‌اند. الحسینی و جاهین (۱۹۹۵) [۱۰] و نیگم(۱۹۸۸) ، حدود پیش بینی را برای مشاهداتی از مدل بر نوع ۱۲ بر اساس چگالیهای پیشین مختلف به‌دست آورده‌اند. نیگم و همکاران(۲۰۰۳) [۱۱] ، کران بالا و پایین را برای مشاهده آینده بر اساس مقادیر مشاهده شده از توزیع پارتو نوع ۲ با استفاده از تکنیکهای بیزی به‌دست آورده‌اند. در این مقاله فرض می‌کنیم مشاهدات اصلی از مدل نمایی تبعیت می‌کنند و یک نوع سانسور روی داده‌ها اعمال شده است. با استفاده از روشهای بیزی سعی داریم مقادیر آینده را پیش‌بینی نمائیم. بخشهای بعدی مقاله بدین شرح است که بخش دو شامل تعریفها و مقدمات لازم از آماره‌های ترتیبی، سانسور دوطرفه، پیش بینی بیزی و طرحهای نمونه گیری است. در بخش سه حدود پیش بینی بیزی را تحت مدل نمایی در دو حالت یک نمونه ای و دو نمونه‌ای برای مقادیر آینده محاسبه می‌کنیم. در بخش چهار نتایج به‌دست آمده تشریح شده است.

[۶] Geisser

[۷] Nigm and Hamdy

[۸] Arnold and Press

[۹] Dunsmore and Amin

[۱۰] Al-Hussaini and Jaheen

[۱۱] Nigm et.al.

۲ تعاریف و مقدمات

دراین بخش، آماره‌های ترتیبی، سانسور دوطرفه، روش پیش بینی بیزی و طرحهای نمونه‌گیری بطور مختصر معرفی میشوند که در قسمتهای بعدی مورد استفاده قرار می‌گیرند.

۱.۲ آماره‌های ترتیبی

فرض کنید $T_n, \cdots, T_2, T_1$ یک نمونه تصادفی مستقل و هم توزیع از جامعه ای با تابع توزیع $F(t;\theta)$ و تابع چگالی احتمال $f(t;\theta)$ باشد و متغیرهای

$T_{\text{-}1}, T_{\text{-}2}, \cdots, T_{\text{-}n}$ آماره‌های ترتیبی حاصل از این نمونه باشد که در آن $T_{\text{-}i}$ ، i -امین آماره ترتیبی در نمونه ای به حجم n می‌باشد، در این صورت تابع چگالی احتمال حاشیه ای i -امین آماره $(T_{\text{-}i})$ عبارتست از

$$f_{\text{-}i:n}(t) = \frac{n!}{(i-1)!(n-i)!} \times [F(t)]^{i-1}[1-F(t)]^{n-i}f(t) \qquad (۳)$$

همچنین تابع چگالی احتمال شرطی j -امین آماره ترتیبی به شرط i -امین آماره ترتیبی به صورت

$$\begin{aligned} & f_{\text{-}j:n|i:n}(t_j|t_i) \qquad (۴) \\ = \quad & \frac{(n-i)!}{(j-i-1)!(n-j)!} \\ \times \quad & [(F(t_j)-F(t_i))/(1-F(t_i))]^{j-i-1} \\ \times \quad & [(1-F(t_j))/(1-F(t_i))]^{n-j} \\ \times \quad & [f(t_j)/(1-F(t_i))], \\ & t_i \leq t_j < \infty, \;\; i < j \leq n \end{aligned}$$

است. برای بررسی جزئیات بیشتر در خصوص ویژگی آماره‌های ترتیبی می‌توان به کتابهای آرنولد و همکاران (۱۹۹۲) [۱۲] یا دیوید و ناگاراجا (۲۰۰۳) [۱۳] مراجعه نمود.

[۱۲] Arnold et.al.

[۱۳] David and Nagaraja

۲.۲ سانسور دوطرفه

اغلب در آزمونهای طول عمر به دلیل محدودیتهای ذکر شده در مقدمه، داده‌ها به صورت سانسور شده در اختیار ما هستند. اگر یک نمونه تصادفی به حجم n در معرض آزمون بقا قرار گیرد در این صورت اولین زمان خرابی مشاهده شده، خود به خود کوچک‌ترین آماره ترتیبی است. به طور مشابهی دومین خرابی ثبت شده، دومین آماره ترتیبی و الی آخر. اگر به دلایلی $s-1$ مشاهده مرتب شده اول در دسترس نباشند و تنها زمانهای خرابی s -امین تا r -امین مو لفه در دسترس باشند، $(1 \leq s < r \leq n)$ آن‌گاه نمونه در دسترس را نمونه سانسور شده دوطرفه می‌نامیم. در این مقاله از این نوع سانسور استفاده می‌کنیم. با فرض اینکه تمام n واحد درلحظه $t=0$ فعال شده باشند، مشاهدات به‌دست آمده در طول آزمایش شامل مجموعه داده‌های زیر است.

$$t_{\text{-}s}, t_{\text{-}s+1}, \cdots t_{\text{-}r}$$

می‌توان نشان داد که تابع درستنمایی سانسور دوطرفه به صورت زیر است.

$$L(\underline{t}) = \frac{n!}{(s-1)!(n-r)!}\prod_{i=s}^{r} f(t_i) \times \{1-F(t_r)\}^{n-r}\{F(t_s)\}^{s-1} \qquad (۵)$$

که در آن $\underline{t} = (t_{\text{-}s}, \cdots t_{\text{-}r})$.

برای مطالعه جزئیات بیشتر در خصوص نظریه و

کاربرد سانسورها می‌توان به لاوولس(۲۰۰۳) مراجعه نمود.

۳.۲ پیش بینی بیزی

فرض کنید $T_1, T_2, \cdots, T_n$ نمونه تصادفی مشاهده شده از جامعه $f(t;\theta)$ باشد و همچنین فرض کنید داده‌های حاصل از سانسور دو طرفه را در اختیار داریم، هدف پیش بینی $n-r$ مشاهده آینده می‌باشد. در مباحث بیزی θ یک متغیرتصادفی تلقی می‌شود، با فرض اینکه $\pi(\theta)$ توزیع پیشین و $\pi(\theta|\underline{t})$ توزیع پسین باشد، اطلاعات درباره مقدار آینده $\{Y=y\}$ را می‌توان از طریق تابع چگالی احتمال زیر به‌دست آورد.

$$f(y|\underline{T}=\underline{t}) = \int_{\Theta} f(y|\underline{t},\theta)\pi(\theta|\underline{t})\ \mathrm{d}\theta, \quad (۶)$$

$$\underline{T}=(T_{\text{-}s},\cdots T_{\text{-}r}),\ \underline{t}=(t_{\text{-}s},\cdots t_{\text{-}r})$$

این تابع به تابع چگالی پیش بینی بیزی معروف است.

برای مطالعه بیشتر در این خصوص به بالاکریشنان و باسو (۱۹۹۵) فصل ۹ مراجعه شود.

۴.۲ طرحهای نمونه گیری

الف. **طرح یک نمونه ای:** اگر یک نمونه تصادفی از جامعه $f(t;\theta)$ در یک آزمون بقا قرارگیرد، آن‌گاه نمونه سانسور شده دوطرفه شامل s - امین تا r - امین زمان خرابی می‌باشد ($1 \leq s < r \leq n$). زمان خرابی j - امین مولفه را با t_j نشان می‌دهیم. هدف ما پیش بینی زمان‌های خرابی $t_{\text{-}r+1}, t_{\text{-}r+2}, \cdots t_{\text{-}n}$ بر اساس نمونه سانسور شده دوطرفه می‌باشد. گیسر (۱۹۹۳) جزئیات بیشتر را در مورد این طرح مورد بررسی قرار داده است.

ب. **طرح دو نمونه ای:** فرض کنید یک نمونه سانسور شده دوطرفه از جامعه $f(t;\theta)$ داشته باشیم، بر اساس این سانسور علاقمند به پیش بینی b - امین زمان خرابی، $b=1,2,\cdots,m$ از نمونه $z_1, z_2, \cdots, z_m$ می‌باشیم که مستقل و هم توزیع نمونه اول است. ایتچیسن و دنسمر (۱۹۷۵) [۱۴] از روش فوق استفاده نموده‌اند.

۵.۲ لم تکنیکی

در این بخش یک لم را بیان و اثبات می‌کنیم، که به عنوان ابزار پایه ای در اثبات نتایج مقاله مورد استفاده قرار می‌گیرد.

لم ۱.۲. *فرض کنید ε و N اعداد طبیعی باشند، آن‌گاه*

$$\sum_{i=0}^{N}(-1)^i\binom{N}{i}(i+\varepsilon)^{-1} = [(N+1)\times\binom{N+\varepsilon}{N+1}]^{-1}.$$

[۱۴] Aitchison and Dunsmore

اثبات. به بسط دو جمله‌ای خیام داریم

$$(1-t)^N = \sum_{i=0}^{N} (-1)^i \binom{N}{i} t^i.$$

طرفین تساوی بالا را در $t^{\varepsilon-1}$ ضرب می‌کنیم و از طرفین نسبت به t با فرض $0 \le t \le 1$ ، انتگرال می‌گیریم، یعنی

$$\int_0^1 (1-t)^N t^{\varepsilon-1}\ \mathrm{d}t = \sum_{i=0}^{N} (-1)^i \binom{N}{i} \int_0^1 t^{i+\varepsilon-1}\ \mathrm{d}t.$$

با محاسبه انتگرال در دو طرف تساوی بالا در نهایت داریم

$$\frac{N!(\varepsilon-1)!}{(N+\varepsilon)!} = \sum_{i=0}^{N} (-1)^i \binom{N}{i} (1/i+\varepsilon).$$

از طرفی می‌دانیم

$$\frac{N!(\varepsilon-1)!}{(N+\varepsilon)!} = [(N+1)\binom{N+\varepsilon}{N+1}]^{-1}.$$

با مقایسه دو رابطه اخیر اثبات کامل می‌شود. □

۶.۲ احتمال پوشش

اگر A یک فاصله پیش بینی باشد آن‌گاه ϑ احتمال پوشش این فاصله است اگر و فقط اگر رابطه زیر برقرار باشد

$$\int_A f(y\,|\underline{t})\ \mathrm{d}y = \vartheta$$

که در آن $f(y\,|\underline{t})$ تابع چگالی پیش بینی است. برای کسب اطلاعات بیشتر در این خصوص به کتاب ایتچیسن و دنسمر (۱۹۷۵) مراجعه شود.

۳ پیش‌بینی بیزی بر اساس مشاهدات سانسور دو طرفه تحت مدل نمایی

در این قسمت سعی می کنیم با توجه به مطالب ارائه شده در بخش‌های قبل، برای مقادیر آینده از مدل نمایی بر اساس مقادیر مشاهده شده سانسور دوطرفه فاصله پیش بینی به‌دست آوریم. این کار را در دو حالت یک نمونه ای و دو نمونه ای انجام می دهیم.

۱.۳ طرح یک نمونه ای

فرض کنید $t_{\text{-}1}, t_{\text{-}2}, \cdots t_{\text{-}n}$ مقادیر مشاهده شده زمان خرابی از یک نمونه شامل n مو لفه در یک آزمون بقا باشد. بر اساس نمونه سانسور شده دوطرفه از یک جامعه با تابع توزیع و تابع چگالی به ترتیب (۱) و (۲)، تابع درستنمایی عبارتست از :

$$L(\theta;\underline{t}) \propto \theta^{r-s+1} \sum_{k=0}^{s-1} Q_k(s) \exp\left[-\theta T_k(\underline{t})\right]$$

که در آن

$$\begin{aligned} \underline{t} &= (t_{\text{-}s}, \cdots, t_{\text{-}r}), \\ Q_k(s) &= (-1)^k \binom{s-1}{k}, \\ T_k(\underline{t}) &= \sum_{i=s}^{r} t_i + (n-r)t_{\text{-}r} + kt_{\text{-}s}. \end{aligned}$$

در آمار بیز پارامتر θ در حکم یافته متغیر تصادفی θ تلقی می شود لذا از یک توزیع مشخص پیروی می کند که به توزیع پیشین معروف است. در این بخش برای پارامتر θ توزیع پیشین مزدوج گاما را به صورت زیر در نظرمی گیریم

$$g(\theta|\alpha,\beta) \propto \theta^{\alpha-1}\exp(-\beta\theta)$$

که در آن α و β مقادیر حقیقی مثبت، پارامترهای توزیع پیشین می باشند. با ضرب توزیع پیشین و تابع درستنمائی سانسور دوطرفه تابع چگالی احتمال پسین به صورت زیر به‌دست می آید.

$$\begin{aligned}\pi^*(\theta|\underline{t}) \quad &\propto \quad \theta^{r-s+\alpha}\sum_{k=0}^{s-1} Q_k(s) \\ &\times \exp\{-\theta[T_k(\underline{t})+\beta]\},\end{aligned} \qquad (۷)$$

فرض کنید $y_a = t_{-r+a}$ مقادیر آینده نمونه باشند که در آن $a = 1, \cdots, n-r$ تابع چگالی احتمال y_a با شرط θ و مشاهدات قبلی با توجه به رابطه (۴) به صورت زیر نوشته می شود:

$$\begin{aligned}h(y_a|\theta,\underline{t}) \quad &\propto \quad [F(y_a)-F(t_{-r})]^{a-1} \\ &\times[1-F(y_a)]^{n-r-a} \\ &\times[1-F(t_{-r})]^{-(n-r)} \\ &\times f(y_a), \quad y_a > t.\end{aligned} \qquad (۸)$$

با جایگذاری روابط (۱) و (۲) در رابطه (۸) داریم

$$\begin{aligned}h(y_a|\theta,\underline{t}) \quad &\propto \quad \theta\sum_{j=0}^{a-1} Q_j(a) \\ &\times \exp[-\theta c_j(y_a - t_{-r})] \\ &, y_a > t_{-r}\end{aligned} \qquad (۹)$$

که در آن $c_j = n-r-a+j+1$.

هدف از انجام محاسبات فوق به‌دست آوردن تابع چگالی احتمال پیش بینی بیزی است که با استفاده از روابط (۷) و (۹) به صورت زیر به‌دست می آید

$$\begin{aligned}f(y_a|\underline{t}) &= \int_0^\infty h(y_a|\theta,\underline{t})\pi^*(\theta|\underline{t})\ \mathrm{d}\theta \qquad (۱۰) \\ &= K^*\sum_{j=0}^{a-1}\sum_{k=0}^{s-1} Q_j(a)Q_k(s) \\ &\times[T_k(\underline{t})+\beta+c_j(y_a-t_{-r})]^{-(d_0+1)} \\ &, y_a > t\end{aligned}$$

که در آن $d_0 = r-s+\alpha+1$ و ضریب K^* طوری اختیار می شود که $f(y_a|\underline{t})$ یک تابع چگالی احتمال باشد. برای محاسبه کرانهای پیش بینی نیاز به محاسبه مقدار احتمال $P(y_a \geq b_1|\underline{t})$ داریم که بر حسب b_1 به صورت زیر به‌دست می آید.

$$P(y_a \geq b_1|\underline{t}) = \int_{b_1}^\infty f(y_a|\underline{t})\ \mathrm{d}y_a$$

این انتگرال با استفاده از رابطه (۱۰)، لم ۱.۲ و با فرض $N = a-1$ و $\varepsilon = n-r-a+1$ به

صورت زیر نوشته می شود.

$$\begin{aligned} & P(y_a \geq b_1|\underline{t}) \\ = \; & A^{-1} a \binom{n-r}{a} \\ & \times \sum_{j=0}^{a-1} \sum_{k=0}^{s-1} Q_j(a) Q_k(s) \\ & \times \frac{[T_k(\underline{t}) + \beta + c_j(b_1 - t_{\text{-}r})]^{-d_0}}{c_j} \end{aligned}$$

که در آن

$$A = \sum_{k=0}^{s-1} Q_k(s) [T_k(\underline{t}) + \beta]^{-d_0}$$

با توجه به احتمال $P(y_a \geq b_1|\underline{t})$ ، کافی است برای محاسبه کران‌های پیش بینی فاصله (L, U) که در آن U کران بالای پیش بینی و L کران پایین پیش بینی است، معادلات زیر به طور توام حل شوند.

$$(1-\vartheta)/2 = P(y_a > U|\underline{t}), \qquad (۱۱)$$

$$(1+\vartheta)/2 = P(y_a > L|\underline{t}). \qquad (۱۲)$$

که در آن ϑ احتمال پوشش فاصله (L, U) می باشد و در بخش ۶-۲ توضیح داده شده است. جواب معادلات (۱۱) و (۱۲) به صورت بسته به‌دست نمی آید، اما با استفاده از روشهای عددی و نرم‌افزارهای ریاضی قابل حل می باشند. فاصله (L, U) که به این طریق به‌دست می آید را فاصله پیش بینی بیزی با احتمال پوشش ϑ برای y_a می گویند.

۲.۳ طرح دو نمونه ای

فرض کنید $t_{\text{-}s}, t_{\text{-}s+1}, \cdots, t_{\text{-}r}$ مقادیر مشاهده شده $r - s + 1$ زمان خرابی نمونه سانسور شده دوطرفه از یک نمونه n تایی از توزیع نمایی باشد. اگر z_b $(b = 1, 2, \cdots, m)$ ، b -امین زمان خرابی مرتب شده در یک نمونه m تایی (مستقل و هم توزیع نمونه اول) باشد که در آینده اتفاق خواهد افتاد، می خواهیم با استفاده از اطلاعات نمونه سانسور شده دوطرفه، برای z_b فاصله پیش بینی به‌دست آوریم. برای این کار روندی مشابه بخش قبل را انجام می دهیم. ابتدا تابع چگالی احتمال z_b به شرط θ ، که عبارتست از:

$$\begin{aligned} h(z_b|\theta) \;\; \propto \;\; & [1 - F(z_b)]^{m-b} [F(z_b)]^{b-1} \\ & \times f(z_b), \;\; z_b > 0. \qquad (۱۳) \end{aligned}$$

که با جایگذاری روابط (۱) و (۲) درآن به صورت زیر نوشته می شود .

$$h(z_b|\theta) \propto \theta \sum_{i=0}^{b-1} Q_i(b) \exp(-\theta z_b e_i), (۱۴)$$

که در آن $e_i = m - b + i + 1$.

از طرفی تابع چگالی احتمال پیش بینی بیزی به

کمک روابط (۷) و (۱۴) عبارتست از:

$$f(z_b|\underline{t}) = K^* \sum_{i=0}^{b-1} \sum_{k=0}^{s-1} Q_i(b) Q_k(s) \times (T_k(\underline{t}) + \beta + e_i z_b)^{-(d_0+1)}, \quad z_b > 0. \qquad (۱۵)$$

برای محاسبه فاصله پیش بینی لازم است مقدار احتمال $P(z_b > b_2|\underline{t})$ را بر حسب b_2 حل کنیم که با استفاده از تابع چگالی احتمال پیش بینی بیزی در (۱۵)، لم ۱.۲ و با فرض $N = b - 1$ و $\varepsilon = m - b + 1$ به صورت زیر نوشته می شود.

$$P(z_b > b_2|\underline{t}) = B^{-1} b \binom{m}{b} \times \sum_{i=0}^{b-1} \sum_{k=0}^{s-1} Q_i(b) Q_k(s) \times \frac{[T_k(\underline{t}) + \beta + b_2 e_i]^{-d_0}}{e_i}.$$

و در آن

$$B = \sum_{k=0}^{s-1} Q_k(s) [T_k(\underline{t}) + \beta]^{-d_0}$$

اکنون برای محاسبه فاصله پیش بینی (L, U) مانند حالت قبل باید معادلات زیر بطور توام حل شوند.

$$(1-\vartheta)/2 = P(z_b \geq U|\underline{t}), \qquad (۱۶)$$

$$(1-\vartheta)/2 = P(z_b \geq U|\underline{t}), \qquad (۱۷)$$

که در آن ϑ احتمال پوشش فاصله (L, U) می باشد. اما مانند طرح یک نمونه ای این معادلات نیز جواب بسته ای ندارند و تنها راه حل آن‌ها استفاده از روشهای عددی است. فاصله (L, U) که با حل معادلات (۱۶) و (۱۷) به‌دست می آید کران‌های بالا و پایین پیش بینی بیزی با احتمال پوشش ϑ را برای z_b می دهد.

۴ مثال عددی

۱.۴ پیش‌بینی در توزیع نمایی بر اساس طرح یک نمونه‌ای

مشاهدات زیر نشان دهنده اندازه سنگ‌هایی است که تحت یک عملیات خاص شکسته می شوند. این داده‌ها از توزیع نمایی (۱) پیروی می کنند (داده‌ها از دنسمر (۱۹۸۳) گرفته شده‌اند).

$$0.6,\ 2.4,\ 5.6,\ 6.6,\ 6.6,\ 9,\ 9.3,$$
$$13,\ 14.3,\ 18.1,\ 24.4,\ 33.8$$

برای تولید یک نمونه سانسور شده دوطرفه، فرض کنید دو مشاهده مرتب شده اول و دوم به دلایلی در دسترس نباشند $(s = 3)$ با انتخاب دو حالت $(r = 6)$ و $(r = 9)$ دو نمونه سانسور شده دوطرفه از مشاهده s ام تا r ام در اختیار داریم. معادلات (۱۱) و (۱۲) را با استفاده از نرم افزار MAPLE و با فرض اینکه $\vartheta = 0/95$ و $s = 3$ باشد حل نموده و فاصله پیش بینی را برای $t_{\text{-}7}$ $(r = 6)$ و $t_{\text{-}10}$ $(r = 9)$ در حالتهای مختلف پارامترهای

توزیع پیشین در جدول ۱ آورده ایم. با نگاهی به جدول۱ ملاحظه می شود که به عنوان مثال در حالت $r = 6$ وقتی پارامترهای توزیع پیشین $\alpha = 1$ و $\beta = 2$ باشد، فاصله پیش بینی برای $t_{\text{-}7}$ ، به صورت $(9/054, 19/3)$ است که مقدار واقعی آن یعنی $9/3$ را شامل می شود. به طور مشابه با تغییر پارامترهای توزیع پیشین، بقیه فاصله‌های پیش بینی به‌دست آمده نیز شامل $t_{\text{-}7}$ می باشند. این نتیجه برای مشاهده آینده $t_{\text{-}10}$ وقتی $r = 9$ باشد نیز برقرار است. در این مثال ملاحظه می شود که:

۱. طول فاصله با افزایش β (α ثابت) افزایش می یابد.

۲. طول فاصله با افزایش α (β ثابت) کاهش می یابد.

۳. با افزایش r ، طول فاصله افزایش پیدا کرده است.

۲.۴ پیش بینی در توزیع نمایی بر اساس طرح دو نمونه ای

همان‌طور که در بخش۲-۴ اشاره شد در طرح دو نمونه‌ای فرض می‌کنیم یک نمونه سانسور شده دوطرفه در دسترس می باشد و می خواهیم بر اساس این نمونه برای b -امین زمان خرابی ($b = 1, 2, \cdots, m$) از نمونه دوم فاصله پیش‌بینی به‌دست آوریم. ابتدا مشاهدات زیر را به عنوان نمونه اول در نظر می گیریم

$$0.6,\ 2.4,\ 6.6,\ 6.6,\ 9.3,\ 13,\ 14.3,\ 18.1\ .$$

همچنین مقادیر نمونه دوم عبارتند از

$$5.6,\ 9,\ 24.4,\ 33.8$$

(داده‌ها از دنسمر (۱۹۸۳) گرفته شده اند). اکنون برای محاسبه فاصله پیش بینی برای z_1 $(b = 1)$ معادلات (۱۶) و (۱۷) را با فرض $\vartheta = 0/95$ و $s = 3$ در دو حالت $r = 6$ و $r = 7$ با استفاده از نرم افزار MAPLE حل نموده و کرانهای پیش بینی به‌دست آمده است(جدول ۲). با نگاهی به جدول به عنوان مثال در حالت $r = 7$ وقتی $\alpha = 1$ و $\beta = 5$ باشد، فاصله پیش بینی برای z_1 ، به صورت $(0/086, 15/82)$ است که مقدار واقعی آن یعنی $5/6$ را شامل می شود. به وضوح بقیه فاصله‌ها نیز در هر دو حالت $r = 6$ و $r = 7$ مقدار واقعی z_1 یعنی $5/6$ را در بر می گیرند. در این مثال جدول ۲ موارد زیر را نیز نشان می دهد:

۱. در حالت دو نمونه ای نیز مانند بخش قبل طول فاصله با افزایش α (β ثابت) کاهش می یابد.

۲. واضح است که طول فاصله با افزایش β (α ثابت) افزایش می یابد.

۳. در حالت دو نمونه ای با افزایش r ، طول فاصله کاهش پیدا کرده است.

جدول ۱: فاصله پیش بینی برای t_{-7} و t_{-10} در حالت یک نمونه ای

پارامترهای توزیع پیشین		$r=6$			$r=9$		
α	β	کران پایین t_{-7}	کران بالای t_{-7}	طول فاصله	کران پایین t_{-10}	کران بالای t_{-10}	طول فاصله
۱	۲	9/054	19/3	10/25	14/366	31/4	17/034
۱	۵	9/056	19/6	10/54	14/369	31/8	17/43
۳	۱	9/04	16/4	7/36	14/38	27/9	13/52
۵	۱	9/03	14/8	5/77	14/38	25/7	11/32

جدول ۲: فاصله پیش بینی برای z_1 در حالت دو نمونه ای

پارامترهای توزیع پیشین		$r=6$			$r=7$		
α	β	کران پایین	کران بالا	طول فاصله	کران پایین	کران بالا	طول فاصله
۱	۲	0/095	18/22	18/125	0/083	15/38	15/297
۱	۵	0/098	18/75	18/652	0/086	15/82	15/734
۳	۱	0/073	13/16	13/09	0/066	11/58	11/514
۵	۱	0/06	10/34	10/28	0/055	9/33	9/275

پیوست

برای حل معادلات به روش عددی از نرم افزار MAPLE استفاده شده است که در هر یک از حالت‌های یک نمونه ای و دو نمونه ای برنامه مربوطه بعد از جایگذاری مقادیر عددی معلوم در زیر آورده شده است.

الف. یک نمونه ای

$$> t := \frac{6}{(83.8)^{-5} - 2\cdot(89.4)^{-5} + (95)^{-5}} \cdot\left(\left(\frac{1}{(6\cdot(83.8 + 6\cdot(x-9))^5)}\right) - \left(\frac{2}{(6\cdot(89.4 + 6\cdot(x-9))^5)}\right) + \left(\frac{1}{(6\cdot(95 + 6\cdot(x-9))^5)}\right)\right) - .975 = 0:$$

> fsolve(t, x)

9.053706397

$$> t := \frac{6}{(83.8)^{-5} - 2\cdot(89.4)^{-5} + (95)^{-5}} \cdot\left(\left(\frac{1}{(6\cdot(83.8 + 6\cdot(x-9))^5)}\right) - \left(\frac{2}{(6\cdot(89.4 + 6\cdot(x-9))^5)}\right) + \left(\frac{1}{(6\cdot(95 + 6\cdot(x-9))^5)}\right)\right) - .025 = 0:$$

> fsolve(t, x)

19.29478466

ب. دو نمونه ای

$$> t := \frac{4}{(97.9)^{-10} - 2\cdot(104.5)^{-10} + (111.1)^{-10}} \cdot\left(\left(\frac{1}{(4\cdot(97.9 + 4\cdot x)^{10})}\right) - \left(\frac{2}{(4\cdot(104.5 + 4\cdot x)^{10})}\right) + \left(\frac{1}{(4\cdot(111.1 + 4\cdot x)^{10})}\right)\right) - .975 = 0:$$

> fsolve(t, x)

0.05470300826

$$> t := \frac{4}{(97.9)^{-10} - 2\cdot(104.5)^{-10} + (111.1)^{-10}} \cdot\left(\left(\frac{1}{(4\cdot(97.9 + 4\cdot x)^{10})}\right) - \left(\frac{2}{(4\cdot(104.5 + 4\cdot x)^{10})}\right) + \left(\frac{1}{(4\cdot(111.1 + 4\cdot x)^{10})}\right)\right) - .025 = 0:$$

> fsolve(t, x)

9.331903149

مراجع

[1] Aitchison, J. and Dunsmore, . R. (1975). Statistical Prediction Analysis, Cambridge University Press.

[2] AL-Hussaini, E. K., Jaheen, Z. F. (1995). *Bayesian prediction bounds for the Burr type XII failure model*, Comm. Statist. (Theory & Meth).**24** , 1829- 1842.

[3] Arnold, B. C., Balakrishnan, N. and Nagaraja, H.N. (1992). *A First Course in Order statistics* , John wiley New York.

[4] Arnold, B. C., and Press, S. J. (1989).*Bayesian estimation and prediction for Pareto data.* J. Amer. Statist. Assoc.**84** ,1079-1084.

[5] Balakrishnan, N. (2007),*Progressive censoring methodology: an appraisal. Test* .**16**(2), 211-259

[6] Balakrishnan, N. and Basu, A. P. (1995).*The exponential distribution: Theory and methods and application* . Gordon and Breach publishers, USA.

[7] Dunsmore, I. R. (1983). *The Future Occurrence of Records,* Ann. Inst. Statist. Math. **35** , 267-277.

[8] Dunsmore, I. R. and Amin, Z. H. (1998).*Some prediction problems censoring samples from the Pareto distribution.* Comm. Statist. (Theory & Math.). **27**(5), 1221-1238.

[9] David, H. A. and Nagaraja, H. N. (2003). Order Statistics, third edition, John Wiley & Sons, New York.

[10] Geisser, S. (1984). *Predicting Pareto and Exponential observables.* Canadian J. Statist. **12**, 143-152.

[11] Geisser, S. (1985) .*Interval prediction for Pareto and Exponential.* J.Econometrics.**29**, 173-185.

[12] Geisser, S. (1993). *Predictive Inference: An Introduction.* Chapman & Hall, New York

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

[14] Nagaraja, H. N. (1995). *Prediction problems, In: Basu, A. P. and Balakrishnan , N(Eds).* The exponential distribution: Theory and methods and application. Gordon and Breach publishers, USA.

[15] Nelson, W. B. (2004). Applied Life Data Analysis. John Wiley, New York.

[16] Nigm, A. M. (1988). *Prediction bounds for the Burr model.* Comm. Statist. **17**, 289-297.

[17] Nigm, A. M., Al-Hussaini, E. K. and Jaheen, Z. F. (2003). *Bayesian one sample prediction of future observations under Pareto distribution.* Statistics, **37**, 527-536.

[18] Nigm, A. M. and Hamdy, H. I. (1987). *Bayesian prediction bounds for the Pareto lifetime model.* Comm. Statist. (Theory & Meth.),**16** ,1761-1772.