ریسک برآوردگر دنباله‌ای نرخ شکست توزیع نمایی تحت کرانداری محدب

عیسی محمودی [۱] و مرضیه ترکی [۲]

چکیده:

در این مقاله توزیع دقیق متغیر توقف، گشتاور و ریسک برآوردگر دنباله‌ای نرخ شکست توزیع نمایی، تحت کرانداری محدب به‌دست می‌آید. برای به‌دست آوردن توزیع دقیق متغیر توقف نظیر تابع نرخ شکست، از فرآیند پوآسون متناظر با آن استفاده می‌شود. در نهایت مقدار دقیق امید ریاضی و ریسک برآوردگر تابع نرخ شکست به صورت جدول ارائه می‌گردد.

واژه‌های کلیدی: برآوردگرهای دنباله‌ای، توزیع نمایی، روش‌های دنباله‌ای، کرانداری محدب، متغیر توقف.

۱ مقدمه

در تحلیل دنباله‌ای، یک آزمایشگر اطلاعات مربوط به پارامتر مجهول θ را بوسیله مشاهده‌ی نمونه تصادفی در مراحل موفقیت جمع‌آوری می‌کند. ممکن است این مشاهدات در یک زمان یا در یک مدت کوتاه جمع‌آوری شوند، اما یک ویژگی عمومی چنین طرح‌های نمونه‌گیری این است که تعداد کل مشاهداتی که در پایان جمع‌آوری می‌شود یک متغیر تصادفی با مقدار مثبت می‌باشد که معمولا با N نمایش داده می‌شود. ممکن است پارامتر θ بوسیله آماره T_n، که از یک نمونه تصادفی $X_1, X_2, \cdots, X_n$ با اندازه نمونه ثابت n بدست می‌آید و این اندازه برای آزمایشگر معلوم است، برآورد شود. این همان روش با اندازه نمونه ثابت است. در مقابل یک آزمایش دنباله‌ای، مرتبط با یک متغیر تصادفی N و نمونه تصادفی $X_1, X_2, \cdots X_N$ است. سپس، در این روش θ بوسیله برآوردگر T_N، که تابعی از متغیر توقف N است، برآورد می‌شود.

در بسیاری از مسایل آماری نمی‌توان با یک اندازه‌ی نمونه‌ی ثابت به استنباط در مورد پارامتر مجهول پرداخت در چنین مواردی استفاده از روش دنباله‌ای ممکن است کارگشا باشد. به عنوان مثال در برآورد $1/p$ در دنباله‌ی آزمایشهای برنولی نمی‌توان با اندازه نمونه‌ی ثابت به برآورد ناریب دست‌یافت. اما استفاده از روش دنباله‌ای برای دست‌یابی به چنین برآورد نااریبی می‌تواند مفید واقع شود.

یک مسئله‌ی تحلیل دنباله‌ای دارای دو عنصر است: عنصر اول قاعده‌ی توقف است که مشخص می‌کند یک آزمایش با $(X_1, X_2, \cdots, X_n)$ نمونه‌ی تصادفی باید متوقف شود یا باید X_{n+1} به آن اضافه شود. عنصر دوم قاعده‌ی تصمیم است که پس از تعیین قاعده‌ی توقف N، تابعی از یافته‌های $(X_1, X_2, \cdots, X_N)$ را برای

[۱] دانشگاه یزد– دانشکده ریاضی– گروه آمار
[۲] دانشگاه یزد– دانشکده ریاضی– گروه آمار

تصمیم‌گیری در مورد پارامتر مورد نظر تعیین می‌کند. به طور کلی هدف روش دنباله‌ای ارایه یک قاعده‌ی توقف و قاعده‌ی تصمیم مناسب است به طوری که بتواند در شرایطی که با توجه به طبیعت داده‌های مورد بررسی و مسئله‌ی استنباط مورد نظر تعیین می‌شود، صدق کند.

مثال ۱ فرض کنید $X_۱, \cdots, X_n$ یک نمونه تصادفی $n(\geq ۲)$—تایی از توزیع نمایی با میانگین θ باشد. تابع زیان در برآورد θ بوسیله $\bar{X}_n$ عبارت از مقدار زیر است:

$$L_n(\theta, \bar{X}_n) = A(\bar{X}_n - \theta)^۲ + cn,$$

که در آن مقادیر $A > ۰$ و $c > ۰$ مقادیر معلومی هستند. تابع مخاطره (ریسک) متناظر با این تابع زیان به صورت زیر می‌باشد:

$$\begin{aligned} R_n(c) &= E_\theta[L_n(\theta, \bar{X}_n)] \\ &= AE_\theta\left[(\bar{X}_n - \theta)^۲\right] + cn \\ &= A\theta_۲ n^{-۱} + cn. \end{aligned}$$

هدف آن است که برای هر $۰ < \theta < \infty$ تابع مخاطره را مینیمم کنیم. بنابراین تابع $g(n)$ را برای $n > ۰$ به صورت زیر در نظر می گیریم:

$$g(n) = A\theta^۲ n^{-۱} + cn.$$

اندازه نمونه‌ای که تابع فوق را مینیمم می کند برابر مقدار زیر است:

$$n \equiv n^*(c) = \left(\frac{A}{c}\right)^{۱/۲}\theta.$$

$n^*(c)$ اندازه‌ی نمونه‌ی ثابت بهینه است زمانی که θ معلوم باشد. بنابراین زمانی که θ مجهول باشد مقدار n^* نیز نامعلوم خواهد بود. در واقع روش بااندازه‌ی نمونه‌ی ثابت برای مینیمم‌کردن تابع مخاطره زمانی که مقدار θ نامعلوم است وجود ندارد. بنابراین روش دنباله‌ای را برای حل آن در نظر می گیریم.

موضوع اصلی این مقاله تعیین توزیع دقیق متغیر توقف و به دست آوردن گشتاورها و ریسک (مخاطره‌ی) برآوردگر دنباله‌ای نرخ شکست توزیع نمایی است.
فرض کنید $X_۱, X_۲, \ldots$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع نمایی با میانگین β باشد، در این مقاله تابع نرخ شکست توزیع نمایی $(\lambda = ۱/\beta)$ براساس یک طرح دنباله‌ای کامل برآورد می‌شود.
برای برآورد پارامتر β دو متغیر توقف به صورت‌های زیر در نظر گرفته می‌شود:

$$M^{(۱)} = \min\{m \geq k(\geq ۱) : m \geq (A/c)^{۱/۲}\bar{X}_m\}. \quad (۱)$$

این متغیر اولین بار توسط استار و وودروف[3] [۴]، وودروف[4] ([۵]، [٦]) و گوش و همکاران[5] [۲] مورد استفاده قرار گرفت که در آن $\bar{X}_m = \frac{۱}{m}\sum_{i=۱}^m X_i$ برای $m \geq k$، A و c مقادیر ثابت و مثبت و k اندازه نمونه مقدماتی هستند. این متغیر تحت تابع زیان

$$L^{(۱)}(\hat{\beta}_m, \beta) = A(\hat{\beta}_m - \beta)^۲ + cm \qquad (۲)$$

[3] Starr and Woodroofe
[4] Woodroofe
[5] Ghosh et al.

به‌دست می‌آید که در آن c هزینه مربوط به هر مشاهده است. داتا و موخوپادهای [۶] [۳] متغیر توقف

$$M^{(۲)} = \min\{m \geq k(\geq ۱) : m \geq (\frac{A}{w})\bar{X}_m^{۲}\} \quad (۳)$$

را معرفی نمودند. این متغیر توقف برای به‌دست آوردن برآوردگر $\hat{\beta}_m$ وقتی که تابع ریسک توسط ثابت$(w(> ۰$ کران‌دار شده است، طرح‌ریزی شد که با تابع زیان درجه دوم

$$L^{(۲)}(\hat{\beta}_m, \beta) = A(\hat{\beta}_m - \beta)^{۲} \quad (۴)$$

در ارتباط است . برای برآورد پارامتر $\lambda = \frac{۱}{\beta}$ (نرخ شکست) متغیر توقف به صورت

$$M = min\{m \geq k(\geq ۳) : m \geq (A/c)\bar{X}_m^{-۲}\} \quad (۵)$$

تحت تابع زیان

$$L(\hat{\lambda}_m, \lambda) = A(\hat{\lambda}_m - \lambda)^{۲} + cm \quad (۶)$$

در نظر گرفته می‌شود. برای یک متغیر توقف M، $S_M = \sum_{i=۱}^{M} X_i$ را در نظر بگیرید. تحت اندازه نمونه ثابت m، آماره S_m برای خانواده $F = \{Exp(\beta) : ۰ < \beta < \infty\}$ بسنده مینیمال است. بنابراین، (M, S_M) (گوش و همکاران [۲]) برای خانواده F بسنده بوده و در نتیجه اکثر برآوردگرها برای توابعی از پارامتر β بعد از توقف تابعی از (M, S_M) می‌باشند.

۲ فرآیند پوآسون متناظر با زمان‌های توقف

یک فرآیند پوآسون شمارشی $\{N(t), t \geq ۰\}$ با $N(۰) \equiv ۰$ را در نظر بگیرید. این یک فرآیند همگن با نرخ ورود $\lambda = ۱/\beta$ است. $\tau_۰ < \tau_۱ < \cdots$ با $\tau_۰ \equiv ۰$ را به عنوان زمان‌های ورود در نظر بگیرید. فواصل زمانی بین زمان‌های ورود $X_i = \tau_i - \tau_{i-۱}(i \geq ۱)$ دارای توزیع $Exp(\beta)$ بوده و $X_۱, X_۲, \cdots$ از هم مستقل می‌باشند. بنابراین زوج (m, S_m) با $(N(\tau_n), \tau_n)$ متناظر می‌شود. با جایگزین کردن m و $m\bar{X}_m$ در تعریف متغیر توقف به ترتیب با $N(t)$ و t، زمان توقف به صورت زیر به‌دست می‌آید:

$$T = inf\{t \geq t_k : N(t) \leq \gamma t^{۲}\} \quad (۷)$$

که در آن $\gamma = c/A$ و $t_k = (k/\gamma)^{۱/۲}$ می‌باشند.
ملاحظه می‌شود که T اولین زمانی است که در آن $N(t)$ کمتر یا مساوی $B(t) = \gamma t^{۲}$ است. $B(t)$ یک تابع کراندار محدب است، زیرا $B''(t) = ۲\gamma > ۰$ که در آن B'' نمایانگر مشتق مرتبه‌ی دوم B است.

۳ توزیع T و M تحت کرانداری محدب

با توجه به (۳) زمان توقف را به صورت زیر در نظر می‌گیریم:

$$T = \inf\{t \geq t_k : N_t \leq \gamma t^{۲}\}, \quad ۰ < \gamma < \infty, \quad (۸)$$

[۶] Datta and Mukhopadhyay

و دنباله‌ی $\{t_l : t_l = (1/\gamma)^{1/2}, l \geq k\}$ را تعریف می‌کنیم.

ازآن‌جایی‌که فرآیند $\{N(t), t \geq 0\}$ یک فرآیند افزایشی است، اگر $N(t_k) > \gamma t_k^2$، آن‌گاه فرآیند فقط می‌تواند از بالا در مقادیر صحیح مثبت، از کران عبور کند. به عبارت دیگر با تعریف $l = \gamma t_l^2$، T فقط می‌تواند مقادیر دنباله‌ی $\{t_l, l \geq k\}$ را اختیار کند. T یک توزیع گسسته دارد. یادآور می‌شویم که $M = N(T) + 1$. بنابراین $\{T = t_l\} = \{N(T) = l\} = \{M = l + 1\}$ برقرار است. علاوه بر این داریم

$$P_\lambda\{T = t_l\} = P_\lambda\{M = l + 1\}, \quad l \geq k. \qquad (9)$$

فرض کنید $\psi_D(l;\lambda) = P_\lambda\{T = t_l\}$. در این صورت لم زیر را داریم :

لم ۱ برای هر $0 < \gamma < \infty$ و $0 < \lambda < \infty$، داریم $P_\lambda\{T < \infty\} = 1$.

اثبات ۱ اگر $N(a,b) = N(b) - N(a)$ تعداد ورودی‌ها در فاصله‌ی زمانی $[a, b)$ باشد، بنابراین برای $n < t \leq n + 1$،

$$N(n) = \sum_{i=1}^{n} N(i-1, i) \leq N(t) \leq N(n+1).$$

با توجه به این‌که $\frac{1}{n+1} < \frac{1}{t} < \frac{1}{n}$ پس

$$\frac{N(n)}{n+1} < \frac{N(t)}{t} < \frac{N(n+1)}{n}.$$

می‌توان رابطه‌ی زیر را نتیجه گرفت:

$$\frac{Nn}{n}\frac{n}{n+1} < \frac{N(t)}{t} < \frac{N(n+1)}{n+1}\frac{n+1}{n}.$$

طبق قانون قوی اعداد بزرگ $\frac{N(n)}{n} \longrightarrow \lambda$ و $\frac{N(n+1)}{n+1} \longrightarrow \lambda$ برقرار است و همچنین وقتی $n \longrightarrow \infty$ داریم $\frac{n+1}{n} \longrightarrow 1$ و $\frac{n}{n+1} \longrightarrow 1$. بنابراین طبق قضیه فشردگی

$$\lim_{t \longrightarrow \infty} \frac{N(t)}{t} = \lambda, \quad a.s.$$

بنابراین

$$\begin{aligned} P_\lambda\{T = \infty\} &= \lim_{t \longrightarrow \infty} P_\lambda\{T > t\} \\ &= \lim_{t \longrightarrow \infty} P_\lambda\{\cap_{s \leq t} N(s) < \gamma s^\alpha\} \\ &\leq \lim_{t \longrightarrow \infty} P_\lambda\{\frac{N(t)}{t} < \frac{\gamma t^\alpha}{t}\} \\ &= P_\lambda\Big(\lim_{t \longrightarrow \infty}\{\frac{N(t)}{t} < \frac{\gamma t^\alpha}{t}\}\Big). \end{aligned}$$

از آنجا که دنباله‌ی پیشامدهای

$$A_t = \{w : \frac{N_t(w)}{t} < \gamma t^{\alpha - 1}\},$$

دنباله‌ای نزولی است، هرگاه $t \longrightarrow \infty$ آن‌گاه $A_t \longrightarrow \emptyset$، بنابراین

$$P_\lambda\{T = \infty\} = P_\lambda(\emptyset) = 0.$$

در نتیجه داریم $P_\lambda(T < \infty) = 1$.

قضیه ۱ برای $l = k$ داریم:

$$\begin{aligned} \psi_D(k;\lambda) &= P_\lambda\{N(t_k) \leq \gamma t_k^2\} \\ &= P_\lambda\{N(t_k) \leq k\} \\ &= P(k; \lambda t_k), \end{aligned} \qquad (10)$$

و برای هر $l > k$ داریم:

$$\begin{aligned}\psi_D(l;\lambda) &= p(l;\lambda t_l) \\ &- \sum_{j=k}^{l-1}\psi_D(j;\lambda) \\ &\times p(l-j;\lambda(t_l-t_j)). \end{aligned} \tag{11}$$

بنابراین r–امین گشتاور $M = N(t) + 1$ به صورت زیر است:

$$\begin{aligned}E_\lambda\{M^r\} &= \sum_{j=k}^{\infty}(j+1)^r P_\lambda\{N(T)=j\} \\ &= \sum_{j=k}^{\infty}(j+1)^r \psi_D(j;\lambda). \end{aligned} \tag{12}$$

متغیر تصادفی $S_M = T + R$ یک توزیع به طور مطلق پیوسته روی بازه (t_k, ∞) دارد. فرض کنید $\tilde{\psi}(t;\lambda)$ نشان‌دهنده تابع چگالی S_M باشد، برای $t > t_k$ داریم:

$$\begin{aligned}\tilde{\psi}(t;\lambda) &= \lambda \sum_{l=k+1}^{\infty} I\{t_{l-1} < t \le t_l\} \\ &\times \sum_{j=k}^{l-1}\psi_D(j;\lambda)e^{-\lambda(t-t_j)}. \end{aligned} \tag{13}$$

دیده می‌شود که تابع چگالی S_M مستقل از T و R است.

قضیه ۲ r–امین گشتاور $\bar{X}_M (r \ge 1)$ از رابطه زیر به‌دست می آید:

$$\begin{aligned}E_\lambda\{\bar{X}_M^r\} &= \frac{1}{\lambda^r}\prod_{j=0}^{r-1}(1+\frac{j}{k}) \\ &\times P(k+r-1;\lambda t_k) \\ &+ \frac{r!}{\lambda^r}\sum_{j=k}^{\infty}e^{\lambda t_j}\psi_D(j;\lambda) \\ &\times \sum_{l=j+1}^{\infty}\frac{1}{l^r}[P(r;\lambda t_{l-1}) \\ &- P(r;\lambda t_l)]. \end{aligned} \tag{14}$$

[7] Abramowitz and Stegun

۴ برآورد تابع نرخ شکست $\lambda = 1/\beta$ تحت متغیر توقف M

متغیر توقف M را در نظر گرفته و دو برآوردگر متفاوت λ را به صورت‌های

$$\hat{\lambda}_{1,M} = \frac{1}{\bar{X}_M} \tag{15}$$

و

$$\hat{\lambda}_{2,M} = \frac{A}{c}(M^{-1}) \tag{16}$$

در نظر می‌گیریم. گشتاور مرتبه r ام $\hat{\lambda}_{1,M}$ به صورت زیر محاسبه می‌شود:

$$\begin{aligned}&E_\lambda\{\bar{X}_M^{-r}\} \\ &= k^r\int_{t_k}^{\infty}\frac{1}{t^r}g(t;k,\lambda)dt \\ &+ \lambda\sum_{j=k}^{\infty}e^{\lambda t_j}\psi_D(j;\lambda) \\ &\times \sum_{l=j+1}^{\infty} l^r\int_{t_{l-1}}^{t_l}\frac{1}{t_{l-1}}e^{-\lambda t}dt. \end{aligned} \tag{17}$$

انتگرال

$$Ei_r(\zeta) = \int_{\zeta}^{\infty}\frac{1}{x^r}e^{-x}dx \tag{18}$$

انتگرال نمایی از مرتبه r خوانده می‌شود (آبراموبتز و استیگن [7][1]). به وسیله انتگرال‌گیری جزء به جزء رابطه زیر برای $r \ge 2$ به‌دست می آید:

$$Ei_r(\zeta) = \frac{1}{r-1}\left[\frac{e^{-\zeta}}{\zeta^r - 1} - Ei_{r-1}(\zeta)\right]. \tag{19}$$

قضیه ۳ گشتاورهای مرتبه اول و دوم $\hat{\lambda}_{1,M}$ به صورت زیر هستند:

$$\begin{aligned} E_\lambda\{\bar{X}_M^{-1}\} &= \lambda\frac{k}{k-1}P(k-2,\lambda t_k) \\ &+ \lambda\sum_{j=k}^{\infty}e^{\lambda t_j}\psi_D(j;\lambda) \\ &\times \sum_{l=j+1}^{\infty}l(Ei_l(\lambda t_{l-1}) \\ &- Ei_l(\lambda t_l)), \qquad (20) \end{aligned}$$

و

$$\begin{aligned} E_\lambda\{\bar{X}_M^{-2}\} &= \lambda^2\frac{k^2}{(k-1)(k-2)} \\ &\times P(k-3;\lambda t_k)+\lambda^2\sum_{j=k}^{\infty}e^{\lambda t_j}\psi_D(j;\lambda) \\ &\times \sum_{l=j+1}^{\infty}l^2(Ei_2(\lambda t_{l-1}) \\ &- Ei_2(\lambda t_l)). \qquad (21) \end{aligned}$$

ریسک $\hat{\lambda}_{1,M}$ به کمک رابطه‌ی زیر حاصل می‌شود:

$$\begin{aligned} R_{1,M} &= AE_\lambda\{\bar{X}_M^{-2}\}-2\lambda AE_\lambda\{\bar{X}_M^{-1}\} \\ &+ A\lambda^2+cE_\lambda\{M\}. \qquad (22) \end{aligned}$$

گشتاور مرتبه rام $\hat{\lambda}_{2,M}$ عبارت‌ست از:

$$E_\lambda\{\hat{\lambda}_{2,M}^r\}=(\frac{A}{c})^r\sum_{l=k}^{\infty}\frac{1}{(l+1)^r}\psi_D(l;\lambda). \qquad (23)$$

در نهایت ریسک برآوردگر $\hat{\lambda}_{2,M}$ برابر است با:

$$\begin{aligned} R_{2,M} &= \frac{A^3}{c^2}\sum_{l=k}^{\infty}\frac{1}{(l+1)^2}\psi_D(l;\lambda) \\ &- 2\lambda\frac{A^2}{c}\sum_{l=k}^{\infty}\frac{1}{l+1}\psi_D(l;\lambda) \\ &+ A\lambda^2+cE_\lambda\{M\}. \qquad (24) \end{aligned}$$

در جدول ۱ مقادیر دقیق امید ریاضی و ریسک برآوردگرهای $\hat{\lambda}_{1,M}$ و $\hat{\lambda}_{2,M}$ ارائه شده است. با توجه به مقادیر جدول ۱ به نظر می‌رسد تحت متغیر توقف M، برآوردگر $\hat{\lambda}_{1,M}$ عملکرد بهتری نسبت به $\hat{\lambda}_{2,M}$ دارد.

جدول ۱. مقادیر $E_\lambda\{M\}$، $E_\lambda\{\hat{\lambda}_{i,M}\}$ و $R_{i,M}$ برای $i=1,2$ و $\lambda=1$ برحسب مقادیر مختلف A و k.

A	k	c	$E_\lambda\{M\}$	$E_\lambda\{\hat{\lambda}_{1,M}\}$	$R_{1,M}$	$E_\lambda\{\hat{\lambda}_{2,M}\}$	$R_{2,M}$
۵	۳	۰/۵	۱۰	۱/۰۸۹۸	۵/۳۳۷۰	۱/۲۵۹۸	۸/۲۵۳۰
		۰/۱	۵۰	۱/۰۰۶۴	۵/۰۸۱۴	۱/۱۴۹۱	۷/۸۸۲۵
		۰/۰۵	۱۰۰	۱/۰۰۴۸	۵/۰۳۳۱	۱/۰۴۸۲	۵/۴۲۳۹
		۰/۰۱	۵۰۰	۱/۰۰۱۱	۴/۹۹۰۵	۱/۰۰۸۲	۵/۰۴۲۸
۱۰	۳	۰/۵	۲۰	۱/۰۰۹۶	۱۰/۳۵۱۴	۱/۴۰۸۸	۲۳/۰۷۲۴
		۰/۱	۱۰۰	۱/۰۰۴۷	۱۰/۰۶۶۲	۱/۰۴۸۲	۱۰/۸۴۷۸
		۰/۰۵	۲۰۰	۱/۰۰۲۱	۱۰/۰۱۲۵	۱/۰۲۱۳	۱۰/۲۳۸۸
		۰/۰۱	۴۰۰	۱/۰۰۱۳	۱۰/۰۰۰۰	۱/۰۱۳۵	۱۰/۱۰۸۸
۱۰	۵	۰/۵	۲۱	۰/۹۶۷۴	۱۰/۳۴۱۱	۱/۱۶۸۱	۱۶/۳۷۹۴
		۰/۱	۱۰۱	۰/۹۹۲۲	۱۰/۰۶۶۱	۱/۰۲۹۵	۱۰/۶۵۹۲
		۰/۰۵	۲۰۲	۰/۹۹۴۴	۱۰/۰۱۲۴	۱/۰۱۷۷	۱۰/۲۳۸۷
		۰/۰۱	۴۰۱	۰/۹۹۶۶	۹/۹۷۵۹	۱/۰۰۹۳	۱۰/۰۸۷۳

مراجع

[1] Abramowitz, M. and Stegun, I. A. (1968). *Handbook of Mathematical Functions*, Seventh printing, edited volume, New York, Dover.

[2] Ghosh, M. Mukhopadhyay, N. and Sen, P.K. (1997). *Sequential Estimation*, New York, Wiley.

[3] Datta, S. and Mukhopadhyay, N. (1995). On Fine-Tuned Bounded Risk Sequential Point Estimation of the Mean of an Exponential Distribution, *South African Statistical Journal* 29: 9-27.

[4] Starr, N. and Woodroofe, M. (1972). Further Remarks on Sequential Estimatial: The Exponentiol Case, *Annals of Mathematical Statistics* 43: 1147-1154.

[5] Woodroofe, M. (1977). Second-Order Approximations for Sequential Point and Interval Estimation, *Annals of Statistics* 5: 984-995.

[6] Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, Philadelphia, SIAM, 20: 789-811.