روش برآوردیابی دنباله‌ای و لزوم استفاده

عیسی محمودی[۱] ، مرضیه ترکی[۲]

چکیده:

در بسیاری از مسائل آماری نمی‌توان با یک اندازه نمونه ثابت به استنباط در مورد پارامتر مورد نظر پرداخت، اما استفاده از تحلیل دنباله‌ای می تواند در این زمینه راه گشا باشد. در دهه اخیر روش‌های دنباله‌ای مورد توجه بسیار قرار گرفته‌اند و کاربردهای آنها در زمینه‌های کشاورزی، آزمایش‌های بالینی، داده کاوی [۱]، دارایی، شبیه سازی کامپیوتری [۲] و مقایسه‌های چندگانه بیان شده است. طرح‌های دنباله‌ای روش‌های ارزشمندی را برای اجرای آزمایشات آماری و سهولت تحلیل داده‌های واقعی در کاربردهای آماری بیان می‌کنند. در این مقاله ابتدا مناسب نبودن روش با اندازه نمونه ثابت در حل مسائلی همچون تعیین فاصله اطمینان با طول ثابت و مخاطره کراندار شده بیان، و در ادامه روش دنباله‌ای را برای حل چنین مسائلی در توزیع‌های نرمال، نمایی منفی و نمایی شرح می‌دهیم. [۳]

واژه‌های کلیدی: برآورد دنباله‌ای، مخاطره کراندار، فاصله اطمینان، متغیر توقف.

۱ مقدمه

در تحلیل دنباله‌ای، یک آزمایشگر اطلاعات مربوط به پارامتر مجهول θ را بوسیله مشاهده‌ی نمونه تصادفی در مراحل موفقیت جمع آوری می کند. ممکن است این مشاهدات در یک زمان یا در یک مدت کوتاه جمع آوری شوند، اما یک ویژگی عمومی چنین طرح‌های نمونه‌گیری این است که تعداد کل مشاهداتی که در پایان جمع آوری 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 می آید 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.

۲ مسائل برآورد

در مباحث مربوط به برآورد پارامترها، مسائل زیادی وجود دارد که نمی‌توان با روش اندازه نمونه ثابت آنها را حل کرد. اما این مسائل با روش دنباله‌ای قابل حل هستند. قبل از بیان مثال، یک قضیه کلی که نشان دهنده مناسب نبودن روش مبتنی بر اندازه نمونه ثابت در حل

[۱] عضو هیئت علمي (استادیار)— دانشگاه یزد—دانشکده ریاضی— گروه آمار
[۲] کارشناسی ارشد آمار— دانشگاه یزد—دانشکده ریاضی— گروه آمار
[۳] عمده‌ی قسمت‌های مقاله برگرفته از مراجع [۱]، [۲]، [۳] و [۴] می‌باشد.

اینگونه مسائل می‌باشد، را ارائه می‌دهیم:

قضیه ۱ فرض کنید $X_1,\cdots X_n$ نمونه تصادفی با تابع چگالی احتمال $\frac{1}{\sigma}f(\frac{x-\theta}{\sigma})$ باشد، به طوری که $-\infty<\theta<\infty$ و $0<\sigma<\infty$ دو پارامتر مجهول باشند. برای برآورد θ، تابع زیان $L(\theta,\delta(\mathbf{x}))=H(|\delta(\mathbf{x})-\theta|)$ را در نظر بگیرید، به طوری که $\mathbf{x}=(x_1,\cdots,x_n)$ مقادیر نمونه تصادفی $\mathbf{X}=(X_1,\cdots,X_n)$ باشند. فرض کنید که $H(|u|)$ در $|u|$ صعودی باشد و قرار دهید $M=\sup_{-\infty<u<\infty}H(|u|)$ که ممکن است نامتناهی باشد. برای هر مقدار ثابت $K<M$ ، برآوردگر $\delta(\mathbf{X})$ وجود ندارد به طوری که $\sup_{\theta,\sigma}E_{\theta,\sigma}\{L(\theta,\delta(\mathbf{X}))\}\le K$ باشد. [۴]

مثال ۱ فرض کنید $X_1,\cdots,X_n$ یک نمونه تصادفی $n(\ge 2)$ تایی از توزیع $N(\mu,\sigma^2)$ باشند، وقتی که $-\infty<\mu<\infty$ و $0<\sigma<\infty$ هر دو پارامتر مجهول‌اند. با تعریف $\bar{X}_n=1/n\sum_{i=1}^n X_i$ و $S_n^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X}_n)^2$، برای $0<\alpha<1$، یک فاصله اطمینان $100(1-\alpha)\%$ برای μ برابر است با:

$$J_n=[\bar{X}_n\pm t_{\alpha/2}(n-1)\frac{S_n}{\sqrt{n}}],$$

که در آن طول J_n برابر است با:

$$K_n=2t_{\alpha/2}(n-1)\frac{S_n}{\sqrt{n}}.$$

K_n یک متغیر تصادفی است و ممکن است برای n ثابت با احتمال مثبت مقدار بزرگی باشد. تلاش می‌کنیم تا یک فاصله اطمینان $100(1-\alpha)\%$ برای μ بسازیم به طوری که طول فاصله اطمینان، مقدار از قبل تعیین شده $2d$ باشد. ساختن چنین فاصله اطمینانی با طول ثابت، متناظر با پیدا کردن یک برآوردگر برای پارامتر μ تحت تابع زیان $1-0$ است. فرض کنید $\delta(\mathbf{X})$ یک برآوردگر μ باشد و

$$L(\mu,\delta(\mathbf{x}))=\begin{cases}0 & |\delta(\mathbf{x})-\mu|\le d\\ 1 & |\delta(\mathbf{x})-\mu|>d,\end{cases}\qquad (1)$$

به طوری که $d(>0)$ از قبل تعیین شده است. تحت تابع زیان فوق تابع مخاطره به فرم زیر است:

$$E_{\mu,\sigma}\{L(\mu,\delta(\mathbf{X}))\}=P_{\mu,\sigma}\{|\delta(\mathbf{X})-\mu|>d\}.$$

در این مثال توجه کنید که $M=1$ است و با توجه به قضیه ۱ برای $0<\alpha<1$ ثابت، هیچ برآوردگر $\delta(\mathbf{X})$ ی با شرط $P_{\mu,\sigma}\{|\delta(\mathbf{X})-\mu|>d\}\le\alpha$ وجود ندارد.
بنابراین مسئله فاصله اطمینان با طول ثابت نمی‌تواند بوسیله روش با اندازه نمونه ثابت حل شود. به عبارت دیگر، روش با اندازه نمونه ثابت نمی تواند یک فاصله اطمینان با طول ثابت برای θ وقتی σ^2 نامعلوم است بسازد، به طوری که ضریب اطمینان حداقل $(1-\alpha)$ باشد. در ادامه روش دنباله‌ای را برای حل این مساله بیان می‌کنیم.

مثال ۲ توزیع $N(\mu,\sigma^2)$ را در مثال قبل در نظر بگیرید و فرض کنید می‌خواهیم برای اندازه نمونه ثابت n یک برآوردگر نقطه‌ای برای μ بر اساس $\mathbf{X}=(X_1,\cdots,X_n)$

[۴] برگرفته شده از مرجع [۳].

بسازیم. برای یک برآوردگر $\delta(\mathbf{X})$ از μ، فرض کنید که تابع زیان درجه دوم به صورت زیر باشد:

$$L(\mu, \delta(\mathbf{X})) = (\delta(\mathbf{X}) - \mu)^{2}.$$

تحت تابع زیان فوق، تابع مخاطره برابر است با:

$$E_{\mu,\sigma}\{L(\mu, \delta(\mathbf{X}))\} = E_{\mu,\sigma}\{(\delta(\mathbf{X}) - \mu)^{2}\}.$$

دیده می‌شود که در این جا، $M = \infty$ است و برای هر K ثابت، $0 < K < \infty$ است. اما در این‌جا برآوردگر $\delta(\mathbf{X})$ ی با شرط آنکه برای هر μ و σ، داشته باشیم $E_{\mu,\sigma}\{(\delta(\mathbf{X}) - \mu)^{2}\} \leq K$ وجود ندارد. به عبارت دیگر نمی‌توان یک کران بالای K برای میانگین مربع خطا از یک برآوردگر برای هر μ و σ از قبل تعیین کرد، این همان هدف غیر قابل دسترسی است. این مسئله اشاره به مسئله برآورد نقطه‌ای با مخاطره کراندار برای μ دارد. در دو مثال قبل قضیه به طور مستقیم نشان می‌دهد که روش با اندازه نمونه ثابت برای حل چنین مسائلی وجود ندارد. اغلب مسائل مربوط به مینیمم مخاطره و برآورد ناحیه اطمینان با طول ثابت، در حالت چند پارامتری، بوسیله روش با اندازه نمونه ثابت قابل حل نیستند.

مثال ۳ توزیع نمایی منفی با تابع چگالی

$$f(x,\theta,\sigma) = \begin{cases} \frac{1}{\sigma}\exp(-(x-\theta)/\sigma) & x > \theta \\ 0 & x \leq \theta \end{cases} \qquad (2)$$

را در نظر بگیرید که به‌صورت $NE(\theta,\sigma)$ نمایش داده می‌شود، به‌طوری که $-\infty < \theta < \infty$ و $0 < \sigma < \infty$ پارامترهای نامعلوم می‌باشند.

فرض کنید $X_1, \cdots, X_n$ یک نمونه تصادفی n تایی از $NE(\theta,\sigma)$ باشد، در این‌صورت برآورد حداکثر درستنمایی پارامترهای θ و σ عبارتند از:

$$\hat{\theta} = X_{(1)}.$$

و

$$\hat{\sigma} = U_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - X_{(1)}), \quad n \geq 2.$$

فاصله اطمینان زیر را برای θ در نظر بگیرید:

$$J_n = (X_{(1)} - d, X_{(1)}),$$

به طوری که d مقداری از قبل تعیین شده است.

برای آنکه ضریب اطمینان مربوط به این فاصله اطمینان حداقل $1-\alpha$ ($0 < \alpha < 1$) شود، باید داشته باشیم:

$$\begin{aligned} &P_{\theta,\sigma}\{X_{(1)} - d < \theta < X_{(1)}\} \\ &= P_{\theta,\sigma}\{0 < \frac{n(X_{(1)} - \theta)}{\sigma} < \frac{nd}{\sigma}\} \\ &= 1 - \exp(-nd/\sigma) \geq 1 - \alpha, \end{aligned}$$

به طوری که $n(X_{(1)} - \theta)/\sigma$ دارای توزیع $NE(0,1)$ می‌باشد.

رابطه فوق مستلزم آن است که اندازه نمونه کوچکترین عدد صحیح بزرگتر یا مساوی $C = \frac{a\sigma}{d}$ $(a = \log(1/\alpha))$ باشد. C به اندازه نمونه ثابت بهینه اشاره دارد، وقتی که σ معلوم است. دقت شود که وقتی مقدار پارامتر σ نامعلوم باشد، حجم نمونه C هم نامعلوم می‌ماند، بنابراین باید برای بهینه کردن، روش برآورد دنباله‌ای را به کار گرفت.

مثال ۴ فرض کنید $X_1, \cdots, X_n$ یک نمونه تصادفی $n(\geq ۲)$ تایی از توزیع نمایی با میانگین θ باشد وقتی که $\theta > ۰$ پارامتر مجهول است. تابع زیان حاصل از برآورد θ بوسیله $\bar{X}_n$ را به‌صورت زیر در نظر می گیریم:

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

که در آن مقادیر $A > ۰$ و $c > ۰$ معلوم هستند. تابع مخاطره متناظر با این تابع زیان عبارتست از :

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

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:

$$g(n) = A\theta^۲ n^{-۱} + cn.$$

اندازه‌ی نمونه‌ای که تابع فوق را مینیمم می کند برابر است با:

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

$n^*(c)$ اندازه نمونه ثابت بهینه است زمانی که θ معلوم باشد. بنابراین زمانی که θ مجهول باشد مقدار n^* نیز نامعلوم خواهد بود. در واقع روش با اندازه نمونه ثابت برای مینیمم کردن تابع مخاطره زمانی که مقدار θ نامعلوم است وجود ندارد. بنابراین روش دنباله‌ای برای حل این مساله مفید به نظر می‌رسد.

۳　برآورد میانگین در یک جامعه نرمال

۱.۳　فاصله اطمینان با طول ثابت

در مثال ۱ مسئله‌ی برآورد فاصله اطمینان با طول ثابت برای پارامتر مجهول μ بحث شد. دو مقدار $۰ < \alpha < ۱$ و $d(> ۰)$ از قبل تعیین شده‌اند و می‌خواهیم یک فاصله اطمینان J برای μ بسازیم به طوری که:
برای هر μ و σ ، $P_{\mu,\sigma}\{\mu \in J\} \geq ۱ - \alpha$ و طول J برابر $۲d$ باشد.
درباره این مسئله در مثال ۱ نشان دادیم که روش با اندازه نمونه ثابت پاسخگو نیست.
فرض کنید $X_1, \cdots, X_n$ یک نمونه تصادفی n تایی از توزیع $N(\mu, \sigma^۲)$ باشند. فاصله اطمینان با طول ثابت $J_n = [\bar{X}_n - d, \bar{X}_n + d]$ را برای میانگین مجهول μ در نظر بگیرید. ضریب اطمینان J_n به صورت زیر محاسبه می‌شود:

$$\begin{aligned} P_{\mu,\sigma}\{\mu \in J_n\} &= P_{\mu,\sigma}\{|\bar{X}_n - \mu| \leq d\} \\ &= P_{\mu,\sigma}\{\frac{\sqrt{n}|\bar{X}_n - \mu|}{\sigma} \leq \frac{\sqrt{n}d}{\sigma}\} \\ &= ۲\Phi(\sqrt{n}d/\sigma) - ۱, \end{aligned}$$

به طوری که $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ دارای توزیع نرمال استاندارد است. می‌دانیم که J_n دارای طول ثابت $۲d$ است. حال می‌خواهیم ضریب اطمینان حداقل $۱ - \alpha$ باشد، بنابراین بایستی

$$۲\Phi(\sqrt{n}d/\sigma) - ۱ \geq ۱ - \alpha = ۲\Phi(-z_{\alpha/۲}) - ۱.$$

اما با توجه به صعودی بودن $\Phi(z)$ داریم:

$$\frac{\sqrt{n}d}{\sigma} \geq -z_{\alpha/2} \Longrightarrow n \geq \frac{z^2_{\alpha/2}\sigma^2}{d^2} = C,$$

به عبارت دیگر ضریب اطمینان فاصله J_n حداقل برابر $1-\alpha$ خواهد بود اگر و فقط اگر n کوچکترین عدد صحیح بزرگتر یا مساوی C باشد. دراین جا C بیان کننده یک اندازه نمونه ثابت بهینه برای ساختن یک فاصله اطمینان با طول ثابت برای μ وقتی σ معلوم است، می‌باشد. اعداد $z_{\alpha/2}$ و d مقادیر ثابت هستند. بنابراین اگر σ معلوم باشد می‌توان C را به راحتی تعیین کرد.
سوال مهم این است که وقتی σ نامعلوم است، اندازه نمونه لازم n برای ساختن یک فاصله اطمینان J_n با ضریب اطمینان بزرگتر یا مساوی $1-\alpha$ و طول ثابت $2d$ چه قدر باید باشد؟

۱.۱.۳ روش دو مرحله‌ای

اشتاین [۴] پیشنهاد کرد مشاهدات را در دو مرحله جمع آوری کنند. مشاهداتی که در مرحله اول با اندازه $m(\geq 2)$ بودند، که نمونه مقدماتی نامیده می‌شوند. حال بر اساس $X_1, \cdots, X_m$ برآورد پارامتر نامعلوم σ^2 بوسیله S^2_m حاصل می شود و در نتیجه C هم برآورد می‌شود. کوچکترین عدد صحیح بزرگتر یا مساوی $\hat{C}(N \geq \hat{C})$ اندازه نمونه نهایی خواهد بود و $J_N = [\bar{X}_N \pm d]$ فاصله پیشنهاد شده برای μ است.
اگر $\lfloor u \rfloor$ نشان‌دهنده بزرگترین عدد صحیح کوچکتر از u باشد و $a_{m-1} = t_{\alpha/2}(m-1)$ اختیار کنیم، آنگاه اندازه نمونه نهایی را به صورت زیر تعریف می‌کنیم:

$$N \equiv N(d) = \max\{m, \lfloor \frac{a^2_{m-1}S^2_m}{d^2} \rfloor + 1\} \qquad (3)$$

ثابت می‌شود که N با احتمال یک متناهی است. یادآوری می‌کنیم که $C = \frac{z^2_{\alpha/2}\sigma^2}{d^2}$ است و برای σ^2 مجهول C بوسیله $\frac{a^2_{m-1}S^2_m}{d^2}$ برآورد می‌شود. مقدار عدد صحیح متغیر تصادفی N اشاره به زمان توقف یا متغیر توقف دارد. روش دو مرحله ای نتایج زیر را در بر دارد:
۱) اگر $N = m$ باشد، یعنی m بزرگتر از برآوردگر C است و این نشان می‌دهد که مشاهدات زیادی را در مرحله مقدماتی داریم بنابراین احتیاج به مشاهدات بیشتر در مرحله دوم نداریم.
۲) اگر $N > m$ باشد، یعنی m کمتر از برآوردگر C است و این نشان‌دهنده آن است که با یک تعداد کمی از مشاهدات در مرحله مقدماتی شروع می‌کنیم. بنابراین، این اختلاف اندازه نمونه را در مرحله دوم نمونه‌گیری می‌کنیم، یعنی مشاهدات $X_{m+1}, \cdots, X_N$ را در مرحله دوم بدست می آوریم.
مورد اول : اگر $N = m$ باشد، مجموعه داده‌های نهایی به صورت $X_1, \cdots, X_m$ است. مورد دوم : اگر $N > m$ باشد، مجموعه داده‌های نهایی به صورت $X_1, \cdots, X_m, X_{m+1}, \cdots, X_N$ است. با ترکیب کردن دو مورد بالا می توان گفت، مجموعه داده‌های نهایی ترکیبی از N و $X_1, \cdots, X_N$ است. دیده می‌شود که فاصله اطمینان $J_N = [\bar{X}_N \pm d]$ از یک نمونه تصادفی $X_1, \cdots, X_N$ با اندازه تصادفی N تعریف شده‌است. روشن است که
(i) پیشامد $\{N = n\}$ فقط به متغیر تصادفی S^2_m وابسته

است.

(ii) برای هر $n(\geq m)$ ثابت $\bar{X}_n$ و S_m^2 از هم مستقل هستند. بنابراین هر پیشامدی که با $\bar{X}_n$ تعریف می‌شود باید مستقل از پیشامد $\{N=n\}$ باشد.

بنابر استقلال $\bar{X}_n$ و $I(N=n)$، ضریب اطمینان مرتبط با J_N به صورت زیر محاسبه می‌شود:

$$\begin{aligned}
&P_{\mu,\sigma}\{\mu\in J_N\}\\
&=\sum_{n=m}^{\infty}P_{\mu,\sigma}\{|\bar{X}_N-\mu|\leq d\cap\{N=n\}\}\\
&=\sum_{n=m}^{\infty}P_{\mu,\sigma}\{|\bar{X}_n-\mu|\leq d\mid\{N=n\}\}P(N=n)\\
&=\sum_{n=m}^{\infty}P_{\mu,\sigma}\{|\bar{X}_n-\mu|\leq d\}P_{\mu,\sigma}(N=n)\\
&=\sum_{n=m}^{\infty}[2\Phi(\sqrt{n}d/\sigma)-1]P_{\mu,\sigma}(N=n)\\
&=E_{\mu,\sigma}[2\Phi(\sqrt{N}d/\sigma)-1]. \qquad (4)
\end{aligned}$$

قضیه ۲ برای متغیر توقف N که بوسیله رابطه (۲) تعریف شده، برای هر μ، σ، d و αی ثابت داریم:

(i)

$$a_{m-1}^2\sigma^2d^{-2}\leq E_{\mu,\sigma}(N)\leq m+a_{m-1}^2\sigma^2d^{-2}$$

(ii) $\bar{X}_N$ یک برآوردگر نااریب از μ است و دارای واریانس $\sigma^2E_{\mu,\sigma}(N^{-1})$ است.

(iii) $Q\equiv\sqrt{N}(\bar{X}_N-\mu)/\sigma$ دارای توزیع نرمال استاندارد است.

(iv) $P_{\mu,\sigma}\{\mu\in J_N\}\geq 1-\alpha$.

نتایج فوق نشان می‌دهد وقتی که راه حل با اندازه نمونه ثابت وجود ندارد روش برآورد دو مرحله‌ای اشتاین می‌تواند یک روش برای یک مسئله آماری ارائه دهد. [5]

۲.۱.۳ توزیع احتمالی اندازه نمونه نهایی N

اندازه نمونه N که بوسیله رابطه (۲) بدست می آید، یک متغیر تصادفی است، بنابراین می‌توان توزیع دقیق آن را یافت. مشاهده می‌شود که توزیع دقیق N به μ بستگی ندارد.

از آن جایی که $Y=\frac{(m-1)S_m^2}{\sigma^2}$ دارای توزیع χ_{m-1}^2 است، به راحتی می توان نشان داد که:

$$\begin{aligned}
P_\sigma(N=m)&=P_\sigma\{\lfloor\frac{a_{m-1}^2S_m^2}{d^2}\rfloor+1\leq m\}\\
&=P_\sigma\{0<\frac{a_{m-1}^2S_m^2}{d^2}\leq m\}\\
&=P_\sigma\{0<Y\leq\frac{m(m-1)d^2}{\sigma^2a_{m-1}^2}\}.
\end{aligned}$$

به طور مشابه برای هر عدد صحیح k:

$$\begin{aligned}
&P_\sigma(N=m+k)\\
&=P_\sigma\{\lfloor\frac{a_{m-1}^2S_m^2}{d^2}\rfloor=m+k-1\}\\
&=P_\sigma\{m+k-1<\frac{a_{m-1}^2S_m^2}{d^2}\leq m+k\}\\
&=P_\sigma\{\frac{(m+k-1)(m-1)d^2}{\sigma^2a_{m-1}^2}<Y\\
&\qquad\leq\frac{(m+k)(m-1)d^2}{\sigma^2a_{m-1}^2}\}.
\end{aligned}$$

قرار دهید:

$$g_m(y)=(2^{\frac{(m-1)}{2}}\Gamma(\frac{m-1}{2}))^{-1}\exp[-y/2]y^{\frac{(m-3)}{2}},$$

و

$$A_k=\frac{(m+k)(m-1)d^2}{\sigma^2a_{m-1}^2}.$$

[5]مطالب این قسمت از مراجع [۲]، [۳] و [۴] گرفته شده است.

بنابراین داریم:

$$\begin{aligned} P_\sigma(N=m) &= P_\sigma\{0 < \chi^2_{m-1} \le A_0\} \\ &= \int_0^{A_0} g_m(y)dy, \end{aligned}$$

و برای هر $k = 1, 2, \cdots$

$$\begin{aligned} P_\sigma(N=m+k) &= P_\sigma\{A_{k-1} < \chi^2_{m-1} \le A_k\} \\ &= \int_{A_{k-1}}^{A_k} g_m(y)dy. \end{aligned}$$

۳.۱.۳ روش دنباله‌ای کامل

به منظور کاهش دادن حجم نمونه در روش دو مرحله‌ای، ممکن است σ^2 را به صورت متوالی در یک طرح دنباله‌ای برآورد کنیم. با $m(\ge 2)$ مشاهده شروع می‌کنیم و سپس با اضافه کردن یک مشاهده در واحد زمان این روند را ادامه می‌دهیم. اما این نوع نمونه‌گیری وقتی تعداد مشاهدات کافی را جمع‌آوری کردیم خاتمه می‌یابد. ولی چگونه می‌دانیم که به این مرحله رسیده‌ایم؟

به طور پی در پی σ^2 را برآورد می‌کنیم و به برآورد دنباله‌ای C انتقال می‌دهیم، و کنترل می‌کنیم که در هر مرحله آیا اندازه نمونه از برآورد C در هر مرحله تجاوز می‌کند؟ زمانی که این اندازه از برآورد C تجاوز کرد، نمونه گیری خاتمه می‌یابد و می‌توان فاصله اطمینان را ساخت. ما با $\mathbf{X} = (X_1, \cdots, X_m)$ شروع می‌کنیم که m اندازه نمونه مقدماتی است، سپس یک نمونه دیگر X را در واحد زمان اضافه می‌کنیم. زمان توقف به صورت زیر تعریف می‌شود: کوچکترین عدد صحیح $n(\ge m)$ که در شرط $n \ge \frac{a^2 S_n^2}{d^2}$ صدق کند.

این روش به صورت زیر اجرا می‌شود:

در مرحله ابتدایی، S_m^2 را بر اساس $X = (X_1, \cdots, X_m)$ بدست می‌آوریم، و کنترل می‌کنیم که آیا $m \ge \frac{a^2 S_m^2}{d^2}$ یا نه؟ اگر $m \ge \frac{a^2 S_m^2}{d^2}$ باشد در این مرحله نمونه‌گیری خاتمه می‌یابد و اندازه نمونه نهایی m است. اما اگر $m < \frac{a^2 S_m^2}{d^2}$ باشد مشاهده‌ی، X_{m+1} را اضافه می‌کنیم و این بار S_{m+1}^2 را از روی $X_1, \cdots, X_m, X_{m+1}$ بدست می‌آوریم. دوباره کنترل می‌کنیم آیا $m+1 \ge \frac{a^2 S_{m+1}^2}{d^2}$ یا نه؟ اگر نمونه‌گیری در این مرحله خاتمه یابد، اندازه نمونه نهایی برابر $m+1$ خواهد بود. اما اگر $m+1 < \frac{a^2 S_{m+1}^2}{d^2}$ باشد مشاهده‌ی دیگر X_{m+2} را اضافه کرده و دوباره مراحل قبل را تکرار می‌کنیم. این کار ادامه می‌یابد تا زمانی که برای اولین بار به اندازه نمونه n برسیم، که این اندازه حداقل بزرگتر از برآوردگر مربوط به C که $\frac{a^2 S_n^2}{d^2}$ است، باشد.

با μ، σ، d و αی ثابت برای اندازه نمونه نهایی نتایج زیر را داریم:

(i) $P_{\mu,\sigma}(N < \infty) = 1$.

(ii) جایی که در نمونه‌گیری دنباله‌ای کامل، نمونه‌گیری متوقف می‌شود، فاصله اطمینان با طول ثابت $J_N = [\bar{X}_N \pm d]$ برای μ بدست می‌آید، به طوری که $\bar{X}_N$ میانگین نمونه بر اساس N مشاهده است.

مشاهده می‌شود که برای همه $n(\ge m)$ های ثابت، پیشامد $\{N = n\}$ فقط به بردار تصادفی $(S_m^2, \cdots, S_n^2)$ وابسته است. اما میانگین نمونه $\bar{X}_n$ و $(S_m^2, \cdots, S_n^2)$ مستقل هستند. بنابراین هر پیشامد تعریف شده توسط $\bar{X}_n$ باید از پیشامد $\{N = n\}$ مستقل باشد. بنابراین ضریب

اطمینان مربوط به J_N به صورت رابطه (۳) است که در اینجا N زمان توقف در روش دنباله‌ای کامل می‌باشد. [۶]

۲.۳ برآورد دنباله‌ای پارامتر میانگین توزیع نرمال تحت مخاطره کراندار شده در نمونه‌گیری دو مرحله‌ای

در این بخش روش دو مرحله‌ای برای مخاطره کراندار شده برآورد نقطه‌ای پارامتر مجهول μ از توزیع $N(\mu,\sigma^۲)$ با $\sigma^۲$ مجهول بیان می‌شود. فرض کنید که $X_۱,\cdots,X_n$ یک نمونه تصادفی $n(\geq ۲)$ تایی از توزیع فوق باشد. تابع زیان زیر را در نظر بگیرید:

$$L_n(\mu,\bar{X}_n)=A|\bar{X}_n-\mu|^k \quad A>۰, \quad k>۰,$$

وقتی $k=۱$ یا ۲ باشد تابع زیان به ترتیب همان قدرمطلق خطا و مربع خطا است.
هدفمان این است که مخاطره $\bar{X}_n$ به عنوان برآوردی برای پارامتر μ، از یک مقدار از قبل تعیین شده w (برای همه‌ی مقادیر $۰<\sigma<\infty$)، تجاوز نکند، یعنی،

$$\begin{aligned}E_{\mu,\sigma}\{L_n(\mu,\bar{X}_n)\} &= E_{\mu,\sigma}[A|\bar{X}_n-\mu|^k]\\ &= B\sigma^k n^{-k/۲}\leq w,\end{aligned}$$

که در آن $B=\frac{۲^{k/۲}A}{\sqrt{\pi}}\Gamma(\frac{k+۱}{۲})$ می‌باشد.
اما برای هر اندازه نمونه از قبل تعیین شده، این هدف (کراندار شدن تابع مخاطره توسط w) دست نیافتنی است. توجه کنید که مخاطره مربوط به $\bar{X}_n$ از مقدار $w(>۰)$ تجاوز نخواهد کرد اگر و فقط اگر

$$n\geq(\frac{B}{w})^{۲/k}\sigma^۲=n^*. \qquad (۵)$$

در اینجا n^* همان‌اندازه نمونه ثابت بهینه لازم وقتی که σ معلوم است، می‌باشد.
در صورتی که σ مقداری نامعلوم باشد حجم n^* مجهول باقی می‌ماند. بنابراین روش دو مرحله‌ای برای برآورد μ لازم خواهد بود.

۱.۲.۳ روش دو مرحله‌ای برای برآورد تحت مخاطره کراندار

با مشاهدات مقدماتی $X_۱,\cdots,X_m (m>k+۱)$ شروع می‌کنیم و $S_m^۲$ را به‌دست می‌آوریم. اندازه نمونه نهایی را به صورت زیر تعریف می‌کنیم:

$$N\equiv N(A)=\max\{m,\lfloor(\frac{b_mB}{w})^{۲/k}S_m^۲\rfloor+۱\}, \quad (۶)$$

به طوری که

$$\begin{aligned}b_m=\{\frac{۱}{۲}(m-۱)\}^{k/۲}\Gamma(\frac{۱}{۲}(m-k-۱))\\ \times\{\Gamma(\frac{۱}{۲}(m-۱))\}^{-۱}.\end{aligned}$$

در تعریف فوق برای N، ملاحظه می‌گردد که n^* از رابطه (۴) بوسیله $(\frac{b_mB}{w})^{۲/k}S_m^۲$ برآورد می‌شود که در آن $\sigma^۲$ و B به ترتیب بوسیله $S_m^۲$ و b_mB جایگزین می‌شوند. بنابراین N خودش یک برآوردگر برای n^* است.
طرح برآورد دو مرحله‌ای مشابه با آنچه توضیح داده شد، انجام می‌شود.
مخاطره مربوط به برآوردگر $\bar{X}_N$ به صورت زیر به‌دست می‌آید:

$$E_{\mu,\sigma}\{L_N(\mu,\bar{X}_N)\}$$

[۶] در بیان مطالب این بخش از مراجع [۱] و [۳] استفاده شده است.

$$
\begin{aligned}
&= E_{\mu,\sigma}[A|\bar{X}_N - \mu|^k] \\
&= \sum_{n=m}^{\infty} E_{\mu,\sigma}[A|\bar{X}_n - \mu|^k \cap \{N=n\}] \\
&= \sum_{n=m}^{\infty} E_{\mu,\sigma}[A|\bar{X}_n - \mu|^k] \times E_{\mu,\sigma}[I(N=n)] \\
&= \sum_{n=m}^{\infty} [B\sigma^k n^{-k/2}] P_{\mu,\sigma}(N=n) \\
&= B\sigma^k E_{\mu,\sigma}[N^{-k/2}].
\end{aligned}
$$

به طوری که $B = \frac{2^{k/2}A}{\sqrt{\pi}}\Gamma(\frac{k+1}{2})$ می‌باشد. بعضی از ویژگی‌های این برآورد در قضیه زیر بیان شده است:

قضیه ۳ در روش برآورد دنباله‌ای دو مرحله‌ای $(N, \bar{X}_N)$ که بوسیله رابطه (۵) تعریف شد، برای همه‌ی مقادیر ثابت μ، σ، A و w ویژگی‌های زیر را داریم:

(i)

$(b_m B/w)^{2/k}\sigma^2 \leq E_{\mu,\sigma}[N] \leq m + (b_m B/w)^{2/k}\sigma^2$

(ii) $\bar{X}_N$ یک برآوردگر نااریب برای μ با واریانس $\sigma^2 E_{\mu,\sigma}[N^{-1}]$ می‌باشد.

(iii) $E_{\mu,\sigma}\{L_N(\mu, \bar{X}_N)\} \leq w$.[7]

۴ روش دو مرحله‌ای برای برآورد فاصله اطمینان با طول ثابت در توزیع نمایی منفی

با مشاهدات مقدماتی $X_1, \cdots, X_m$ شروع کرده به طوری که $m(\geq 2)$ اندازه نمونه مقدماتی می‌باشد. اندازه نمونه نهایی را به صورت زیر تعریف می‌کنیم:

$$N \equiv N(d) = \max\{m, \lfloor \frac{b_m U_m}{d} \rfloor + 1\}. \qquad (7)$$

به طوری که $b_m = F_\alpha(2, 2m-2)$.

در روش دو مرحله‌ای اگر $N > m$ باشد با نمونه‌گیری $N - m$ واحد اضافی در مرحله دوم و در نظر گرفتن نمونه نهایی $X_1, \cdots, X_N$، فاصله اطمینان $J_N \equiv (X_{(1)} - d, X_{(1)})$، با طول ثابت d در نظر گرفته می‌شود.

روشن است که متغیرهای $I(N=n)$ و $X_{(1)}$ برای هر $n(\geq m)$ مستقل هستند. بنابراین ضریب اطمینان مربوط به J_N به صورت زیر است:

$$P_{\theta,\sigma}\{X_{(1)} - d < \theta < X_{(1)}\} = E_{\theta,\sigma}[1 - e^{-Nd/\sigma}].$$

بعضی از ویژگی‌های روش دو مرحله‌ای فوق را می‌توان در قضیه زیر بیان کرد:

قضیه ۴ برای متغیر توقف N که بوسیله رابطه (۷) تعریف شده، برای همه مقادیر ثابت θ، σ، d، m و α ویژگی‌های زیر را داریم:

(i) $b_m \sigma d^{-1} \leq E_{\theta,\sigma}(N) \leq m + b_m \sigma d^{-1}$

(ii) $Q = N(X_{(1)} - \theta)/\sigma$، دارای توزیع نمایی استاندارد است.

(iii) $P_{\theta,\sigma}\{\theta \in J_N\} \geq 1 - \alpha$،

وقتی که b_m توزیع F درجات آزادی ۲ و $2m - 2$ دارد.[8]

[7] برگرفته از مرجع [۳].
[8] برگرفته از مرجع [۳].

مراجع

[1] Chang, Y.I. and Martinsek, A.T. (2004). Sequential approaches to data mining. In: *Applied Sequential Methodologies*, N. Mukhopadhyay, S. Datta, and S. Chattopadhyay,(eds.), pp. 85-103. Marcel Dekker, New York.

[2] Mukhopadhyay, N. and Cicconetti, G. (2004b). How many simulations should one run? In: *Applied Sequential Methodologies*, N.Mukhopadhyay, S. Datta, and S. Chattopahyay, (eds.), pp. 261-292. Marcel Dekker, New York.

[3] Mukhopadhyay, N., de Silva, B.M. (2009).*Sequential Methods and Their Applications.* Chapman and Hall, New York.

[4] Stein, C. (1949). Some problems in sequential estimation (abstract). *Econometrica*, 17, 77-78.