متوسط اندازهٔ نمونهٔ لازم برای آزمون نسبت احتمال دنباله‌ای پارامتر مقیاس در یک زیرخانوادهٔ نمایی

ناصر داورزنی[۱] محمد جعفری جوزانی[۲]

چکیده

در این مقاله زیرخانواده‌ای نمایی از توزیع‌ها با پارامتر مقیاس نامعلوم θ را در نظر گرفته و آزمون نسبت احتمال دنباله‌ای برای فرض ساده در مقابل ساده دیگر در مورد پارامتر θ با احتمال‌های خطای نوع اول و دوم داده شدهٔ (α,β) به دست می‌آوریم. همچنین یک فرم کلی برای تقریب متوسط اندازهٔ نمونه لازم برای انجام آزمون به روش دنباله‌ای ارائه می‌کنیم. سپس روش دنباله‌ای را به آزمون فرض‌های یکطرفه تعمیم داده، یک فرم کلی برای تقریب اندازهٔ نمونهٔ لازم در انجام آزمون پارامتر مقیاس این خانواده از توزیع‌ها به دست می‌آوریم. در نهایت نتایج به دست آمده را به خانوادهٔ توزیع‌های کای- دو تبدیل یافته تعمیم می‌دهیم.

واژه‌های کلیدی: تحلیل دنباله‌ای، آزمون نسبت احتمال دنباله‌ای[۳]، خانوادهٔ نمایی، خانوادهٔ کای- دو تبدیل یافته.

۱. مقدمه

در بیشتر مسائل آماری پس از مشاهدهٔ نمونهٔ تصادفی به حجم n از توزیعی با تابع چگالی احتمال (یا تابع جرم احتمال) f_θ استنباط هایی دربارهٔ پارامتر مجهول θ انجام می‌شود که عموماً حجم نمونه قبل از بیان و انجام آزمایش تعیین شده است. اما در بسیاری از مسائل آماری لزومی به تعیین حجم نمونه قبل از انجام آزمایش نیست بلکه در عمل نتایج آزمایش تعیین کنندهٔ حجم نمونه است. چنین تحلیل‌هایی را تحلیل‌های دنباله‌ای گویند و می‌توان از این روش در برآورد نقطه‌ای یا آزمون فرض استفاده کرد. هرچند تلاش‌های اولیه در بکارگیری روش‌های دنباله‌ای در رابطه با آزمون فرض را می‌توان در مباحث و مقالات مربوط به کنترل کیفیت آماری بخصوص طرح بازرسی نمونه‌ای در دههٔ ۱۹۳۰ و در کارهای انجام شده توسط دوج[۴] و رومینگ[۵] [۵] در سال ۱۹۲۹ مشاهده کرد. اما فرمول‌بندی نظری عمومی آزمون‌های دنباله‌ای اولین‌بار در سال۱۹۴۳ توسط آماردان برجسته والد[۶] [۱۰] معرفی شد.

والد در سال ۱۹۴۷ نتایج خود را در کتابی به نام«تحلیل‌های دنباله‌ای» منتشر کرد [۹]. آزمون نسبت احتمال دنباله‌ای *(SPRT)* معرفی شده توسط والد[۱۰] که در سال ۱۹۴۴نیز توسط برنارد[۷] معرفی شد براساس روش نیمن پیرسون بنا شده است[۳]. در سال ۱۹۴۸ والد و ولف ویتز[۸] [۱۱] نشان دادند که تحت شرایط یکسان (از نظر خطای نوع اول و دوم و...) آزمون نسبت احتمال دنباله‌ای در مقایسه با سایر روش‌های آزمون فرض آماری برای تصمیم‌گیری نیاز به کمترین حجم نمونه دارد. کمی بعد از آنکه والد نظریهٔ معروف خود را دربارهٔ آزمون ثبت احتمال دنباله‌ای معرفی کرد، کاکس[۹] [۴] و بهادر[۱۰] [۲] توانستند تعمیم‌هایی از نظریهٔ او به دست آورند. آنها به ویژه توانستند نتایج جالبی در رابطه با آزمون نسبت احتمال دنباله‌ای دربارهٔ پارامتر مجهول مورد بررسی در صورت وجود سایر پارامترهای مزاحم به دست آورند. امروزه استفاده از آزمون‌های دنباله‌ای بسیار معمول شده است به طوری که تعمیم‌هایی از این روش تحت عنوان آزمون‌های نسبت احتمال دنباله‌ای تعمیم یافته[۱۱] توسط لی[۱۲] [۶]، آزمون‌های معنی‌داری تکرار شده. توسط سیموند[۱۳] [۸] و روش‌های دیگر ارائه گردیده است.

۱- گروه آمار، دانشگاه پیام نور
۲- دانشکده علوم ریاضی، گروه آمار، دانشگاه شهید بهشتی

۳- SPRT
۴- Dodge
۵- Roming
۶- Wald
۷- Barnard
۸- Wolfwitz
۹- Cox
۱۰- Bahadur
۱۱- Generalized SPRT
۱۲- Lai
۱۳- Siegmund

این مقاله تعمیم نتایج فصل تحلیل دنباله‌ای کتاب مبانی آمار ریاضی پارسیان[۱] [۱]، به خانوادهٔ توزیع‌های مورد بررسی می‌باشد. در این مقاله و در بخش دوم، زیر خانواده‌ای نمایی از توزیع‌های با پارامتر مقیاس نامعلوم θ را در نظر گرفته و آزمون نسبت احتمال دنباله‌ای را برای فرض ساده در مقابل فرض ساده دیگر دربارهٔ پارامتر θ بدست آورده شده است. در بخش سوم فرم کلی‌ای برای تقریب حجم نمونهٔ لازم برای انجام "آزمون نسبت احتمال دنباله‌ای" دربارهٔ پارامتر مقیاس θ در این زیر خانواده نمایی از توزیع‌ها به دست می‌آوریم. در بخش چهارم روش آزمون نسبت احتمال دنباله‌ای را به آزمون فرض‌های یکطرفه تعمیم می‌دهیم، سپس یک فرم کلی برای تقریب اندازهٔ نمونهٔ لازم برای انجام آزمون نسبت احتمال دنباله‌ای در این حالت به دست می‌آوریم. نهایتاً در بخش پنجم نتایج به دست آمده را به خانوادهٔ توزیع‌های کای- دو تبدیل یافته که توسط رحمان و گوپتا[۷] معرفی شد، تعمیم می‌دهیم.

۲. آزمون نسبت احتمال دنباله‌ای در یک زیرخانوادهٔ نمایی

فرض کنید $X_1, X_2, \ldots$ دنباله‌ای از متغیرهای تصادفی مستقل از توزیعی با تابع چگالی احتمال به فرم $\frac{1}{\sigma} f(\frac{x}{\sigma})$ باشد که در آن σ پارامتر مقیاس نامعلوم است. گاهی اوقات می‌توان چگالی فوق را به صورت زیر نوشت:

$$f(x;\theta) = C(x,m)\theta^{-v} \exp\{-T(x)/\theta\} \qquad (۲.۱)$$

که در آن $C(x)$ تابعی از x و مقادیر ثابت، $\theta = \sigma^2$ و v نیز مقداری معلوم است. همچنین $T(X)$ یک آمارهٔ بسندهٔ کامل برای θ با توزیع $Gamma(v,\theta)$ است. مثال‌هایی از توزیع‌های به فرم (۲.۱) عبارتند از:

الف: توزیع نمایی: $Exp(\lambda)$،

$$\theta = \frac{1}{\lambda} \quad (\sigma = \lambda, r = 1),\ v = 1,\ T(X) = X,\ c(x) = 1$$

ب: توزیع گاما: $Gamma(\alpha,\beta)$ با αی معلوم،

$$\theta = \beta\ (\sigma = \beta, r = 1),\ v = \alpha,\ T(X),\ c(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}$$

ج: توزیع نرمال: $N(0,\sigma^2)$،

$$\theta = \sigma^2\ (\sigma = \sigma, r = 2), v = \frac{1}{2}, T(X) = \frac{1}{2}X^2, c(x) = (2\pi)^{-\frac{1}{2}}$$

د: توزیع گاوسی وارون: $IG(\infty,\lambda)$ ،

$$\theta = \frac{1}{\lambda}\ (\sigma = \lambda, r = -1), v = \frac{1}{2}, T(X) = \frac{1}{2X}, c(x) = (2\pi x^3)^{-\frac{1}{2}}$$

هـ: توزیع وایبل: $W(\eta,\beta)$ با η معلوم،

$$\theta = \eta^\beta \quad (\sigma = \eta, r = \beta),\ v = 1,\ T(X) = X^\beta,\ c(x) = x$$

ی: توزیع رایلی: $Rayleigh(\beta)$،

$$\theta = \beta^2 \quad (\sigma = \beta, r = 2),\ v = m,\ T(X) = \frac{X^2}{2},\ c(x) = x$$

حال اگر نمونهٔ $X_1, \ldots, X_m$ نمونه‌ای تصادفی به حجم m از خانوادهٔ توزیع‌های (۲.۱) باشد، توزیع توأم نمونهٔ تصادفی عبارت است از:

$$f(x,\theta) = C(\mathbf{x},m)\theta^{-mv} \exp\{-\sum_{i=1}^{m} T(x_i)/\theta,\ \theta\} > 0 \qquad (۲.۲)$$

که در آن

$$C(\mathbf{x},m) = \Pi_{i=1}^{m} C(x_i)$$

و

$$S(\mathbf{X}) = \sum_{i=1}^{m} T(X_i) \sim Gamma(mv,\theta)$$

یک آمارهٔ بسندهٔ کامل برای θ است. حال ابتدا روش انجام آزمون نسبت احتمال دنباله‌ای را در قالب تعریف زیر بیان می‌کنیم و سپس از آن برای انجام آزمون نسبت احتمال دنباله‌ای دربارهٔ پارامتر θ در خانواده توزیع‌های (۲.۱) استفاده می‌کنیم.

تعریف ۲.۱: فرض کنید $X_1, X_2, \ldots$ دنباله‌ای از متغیرهای تصادفی مستقل و هم توزیع با تابع چگالی احتمال یا تابع احتمال f باشند و بخواهیم آزمون

$$H_0 : f = f_0 \ v.s.\ H_1 : f = f_1 \qquad (۲.۳)$$

را انجام دهیم که در آن f_0 و f_1 دو چگالی ممکن برای X_i ها با

۱- پارسیان

تکیه‌گاه یکسان هستند. تعریف می‌کنیم

$$R_m(x_1,x_2,\dots,x_m)=\frac{f_1(x_1,x_2,\dots,x_m)}{f_0(x_1,x_2,\dots,x_m)}=\Pi_{i=1}^{m}\frac{f_1(x_i)}{f_0(x_i)}$$

دو مقدار A و B را که در شرط $0<B<1<A<\infty$ صدق کنند انتخاب می‌کنیم، در مرحلهٔ m ام پس از انتخاب

$$X_1=x_1,\dots,X_m=x_m$$

۱. اگر $R_m(x_1,x_2,\dots,x_m)\geq A$ نمونه‌گیری متوقف و فرض H_0 رد می‌شود.

۲. اگر $R_m(x_1,x_2,\dots,x_m)\geq B$ نمونه‌گیری متوقف و فرض H_1 رد می‌شود.

۳. اگر $B<R_m(x_1,x_2,\dots,x_m)\geq A$ مشاهدهٔ X_{m+1} را x_{m+1} اختیار می‌کنیم، مقدار R_{m+1} را محاسبه نموده و مراحل فوق را با R_{m+1} تکرار می‌کنیم.

آزمون فوق را که براساس لم نیمن پیرسون بنا شده است، آزمون نسبت احتمال دنباله‌ای (*SPRT*) می‌نامیم.

حال فرض کنید $X_1,X_2,\dots$ دنباله‌ای از متغیرهای تصادفی مستقل از خانوادهٔ توزیع‌های (۲.۱) باشد، برای انجام آزمون فرض

$$H_0:\theta=\theta_0\ v.s.\ H_1:\theta=\theta_1\ \ \theta_1>\theta_0 \qquad (۲.۴)$$

به روش *SPRT* با محاسبه R_m داریم

$$R_m(x_1,x_2,\dots,x_m)=\frac{C(\mathbf{x},m)\theta_1^{-mv}\exp\left\{-\frac{s(\mathbf{x})}{\theta_1}\right\}}{C(\mathbf{x},m)\theta_1^{-mv}\exp\left\{-\frac{s(\mathbf{x})}{\theta_0}\right\}}$$

$$=(\frac{\theta_0}{\theta_1})^{mv}\exp\left\{S(\mathbf{x})(\frac{1}{\theta_0}-\frac{1}{\theta_1})\right\} \qquad (۲.۵)$$

برای زوج داده شدهٔ احتمال خطای نوع اول و دوم (α,β) کران‌های تقریبی A و B را به صورت زیر در نظر می‌گیریم.

$$A=\frac{1-\beta}{\alpha}\quad \&\quad B=\frac{\beta}{1-\alpha} \qquad (۲.۶)$$

طبق تعریف۲. ۱ در مرحلهٔ m ام به نمونه‌گیری ادامه خواهیم داد اگر

$$B<(\frac{\theta_0}{\theta_1})^{mv}\exp\left\{S(\mathbf{x})(\frac{1}{\theta_0}-\frac{1}{\theta_1})\right\}<A$$

یا به طور معادل

$$b(m)=<\sum_{i=1}^{m}T(x_i)<a(m) \qquad (۲.۷)$$

که در آن

$$a(m)=\left(\frac{\theta_0\theta_1}{\theta_1-\theta_0}\right)\ln A+\left(\frac{\theta_0\theta_1}{\theta_1-\theta_0}\right)mv\ln\left(\frac{\theta_1}{\theta_0}\right)$$

$$b(m)=\left(\frac{\theta_0\theta_1}{\theta_1-\theta_0}\right)\ln B+\left(\frac{\theta_0\theta_1}{\theta_1-\theta_0}\right)mv\ln\left(\frac{\theta_1}{\theta_0}\right)$$

(۲.۸)

همچنین به سادگی می‌توان تحقیق کرد که $\left(\frac{\theta_1\theta_0}{\theta_1-\theta_0}\right)\ln\left(\frac{\theta_0}{\theta_1}\right)v>0$، لذا با توجه به رابطهٔ (۲.۸)، از آنجا که $a(m)$ و $b(m)$ توابعی خطی از m می‌باشند، نواحی سه‌گانه برای انجام آزمون (۲.۴) به روش *SPRT* به شکل ۱ می‌باشد.

مثال۲. ۱ : فرض کنید $X_1,X_2,\dots$ دنباله‌ای از متغیرهای تصادفی مستقل با توزیع یکسان $exp(\lambda)$ باشد. در انجام آزمون فرض $H_0:\lambda=\lambda_0\ \ v.s.\ \ H_1:\lambda=\lambda_1\ (\lambda_0>\lambda_1)$ به روش *SPRT* که در آن $f_\lambda(x)=\lambda e^{-\lambda x}\ ,\ \lambda>0$ برای زوج داده شدهٔ (α,β) و کران‌های تقریبی A و B با استفاده از رابطه (۲.۶) نمونه‌گیری در مرحلهٔ m ام ادامه می‌یابد اگر

$$b(m)<\sum_{i=1}^{m}X_i<a(m)$$

که در آن با توجه به رابطه (۲.۸) و با قرار دادن $\theta_0=\frac{1}{\lambda_0}$ و $\theta_1=\frac{1}{\lambda_1}$ داریم

$$a(m)=\frac{\ln(A)}{\lambda_0-\lambda_1}+m\frac{\ln(\frac{\lambda_0}{\lambda_1})}{\lambda_0-\lambda_1},b(m)=\frac{\ln(B)}{\lambda_0-\lambda_1}+m\frac{\ln(\frac{\lambda_0}{\lambda_1})}{\lambda_0-\lambda_1}$$

حال اگر $\lambda_0=2$ و $\lambda_1=1$ در انجام آزمون فرض

$$H_0:\lambda=2\ \ v.s.\ \ H_1:\lambda=1 \qquad (۲.۹)$$

به روش *SPRT* با احتمال خطای نوع اول و دوم $\alpha=\beta=0/05$، $a(m)$ و $b(m)$ عبارتند از:

$$b(m)=-2/94+m\ln 2\quad ,\quad a(m)=2/94+m\ln 2$$

نواحی سه‌گانه آزمون فرض(۲.۹) در شکل ۲ نشان داده شده است.

مثال ۲. ۲: فرض کنید $X_1, X_2, \ldots$ دنباله‌ای از متغیرهای تصادفی مستقل با توزیع یکسان $Rayleigh(\beta)$ با تابع چگالی احتمال $f_\beta(x) = \frac{x}{\beta^2}\exp\{-x^2/2\beta^2\}$ باشد، در انجام آزمون فرض (۲.۴) اگر قرار دهیم $\theta = \beta^2$ رابطه (۲.۴) به رابطه

$$H_0 : \beta = \beta_0 \quad v.s. \quad H_1 : \beta = \beta_1$$

تبدیل می‌شود، که در آن $\beta_1 > \beta_0 > 0$ است. برای زوج (α, β) داده شده در مرحلهٔ m ام نمونه‌گیری ادامه می‌یابد اگر

$$b(m) < \frac{1}{2}\sum_{i=1}^{m} x_i^2 < a(m)$$

که در آن

$$a(m) = \left(\frac{\beta_1^2\beta_0^2}{\beta_1^2 - \beta_0^2}\right)\left\{\ln A + 4m\ln\left(\frac{\beta_1}{\beta_0}\right)\right\}$$

$$b(m) = \left(\frac{\beta_1^2\beta_0^2}{\beta_1^2 - \beta_0^2}\right)\left\{\ln A + 4m\ln\left(\frac{\beta_1}{\beta_0}\right)\right\}$$

لذا می‌توان نواحی سه‌گانه مورد نظر در آزمون فرض را به ازای مقادیر مختلف β_0 و β_1 مشخص کرد و ترسیم نمود.

۳. اندازهٔ نمونه لازم برای *SPRT* پارامتر مقیاس توزیع‌های زیرخانواده نمایی

در این بخش متوسط اندازه نمونه لازم برای انجام آزمون (۲.۴) به روش *SPRT* درباره پارامتر مقیاس در خانواده توزیع‌های (۲.۱) را تحت فرض H_0 و H_1 محاسبه می‌کنیم. فرض کنید N تعداد مشاهدات لازم تا حصول نتیجه مورد نظر، یک متغیر تصادفی گسسته با مقادیر صحیح غیرمنفی باشد به طوریکه $P(N = \infty) = 0$. در خانواده توزیع‌های (۲.۱) قرار می‌دهیم

$$Z_i = \ln\frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)} = v\ln(\frac{\theta_0}{\theta_1}) + T(x_i)(\frac{1}{\theta_0} - \frac{1}{\theta_1}) \qquad (۳.۱)$$

و

$$S_m = \sum_{i=1}^{m} Z_i = mv\ln(\frac{\theta_0}{\theta_1}) + S(\mathbf{x})(\frac{1}{\theta_0} - \frac{1}{\theta_1}) \qquad (۳.۲)$$

که در آن $S(\mathbf{X}) = \sum_{i=1}^{m} T(X_i)$ دارای توزیع $Gamma(nv, \theta)$ می‌باشد. همچنین فرض می‌کنیم $S_0 = 0$ و از آنجا که همچنین چون $0 < B < 1 < A < \infty$، بنابراین $b = \ln B < 0 < \ln A = a$ و مقادیر b ، a کران‌های آزمون *SPRT* در به کارگیری S_N خواهند بود. برای اثبات قضیه اصلی این بخش ابتدا لم‌های زیر را بیان می‌کنیم.

لم ۱.۳: در آزمون فرض (۲.۴) به روش *SPRT* دربارهٔ پارامتر مقیاس θ در خانوادهٔ توزیع‌های (۲.۱) داریم

$$E(S_N) = E(N)E(Z_1) \qquad (۳.۳)$$

لم ۳. ۲: در خانواده توزیع‌های (۲.۱) با در نظر گرفتن رابطه (۳.۱) به ازای $i = 1$ داریم.

$$(i)\ E_{H_0}(Z_1) = v\ln(\frac{\theta_0}{\theta_1}) + v(\frac{\theta_1 - \theta_0}{\theta_1}) \qquad (۳.۴)$$

$$(ii)\ E_{H_1}(Z_1) = v\ln(\frac{\theta_0}{\theta_1}) + v(\frac{\theta_1 - \theta_0}{\theta_1}) \qquad (۳.۵)$$

همچنین برای $i = 1,2$ به راحتی می‌توان نشان داد

$$(iii) E_{H_i}(Z_1^2)$$
$$= \left\{v\ln(\frac{\theta_0}{\theta_1})\right\}^2 + v(v+1)\theta_i^2(\frac{1}{\theta_0} - \frac{1}{\theta_1})^2$$
$$+ 2v^2\theta_i\ln(\frac{\theta_0}{\theta_1})(\frac{1}{\theta_0} - \frac{1}{\theta_1}) \qquad (۳.۶)$$

قضیه ۳. ۱: فرض کنید $X_1, X_2, \ldots$ دنباله‌ای از متغیرهای تصادفی مستقل از خانواده توزیع‌های (۲.۱) باشد. در انجام آزمون (۲.۴) به روش *SPRT* برای زوج (α, β) داده شده، با فرض

$$E_{H_i}(Z_1) \neq 0 \ ; i = 1,2$$

$$E_{H_0}(N) = \frac{\alpha a + b(1-\alpha)}{v\ln(\frac{\theta_0}{\theta_1}) + v(\frac{\theta_1 - \theta_0}{\theta_0})},$$

$$E_{H_1}(N) = \frac{a(1-\alpha) + b\beta}{v\ln(\frac{\theta_0}{\theta_1}) + v(\frac{\theta_1 - \theta_0}{\theta_1})}, \qquad (۳.۷)$$

و در حالتی که $E_{H_0}(Z_1)=0$ یا $E_{H_1}(Z_1)=0$ باشد آنگاه

$$E_{H_i}(N)=\frac{a.b}{E_{H_i}(Z_1^2)} \qquad (۳.۸)$$

برهان: برای زوج داده شدهٔ $A\approx\frac{1-\beta}{\alpha}$ و $B\approx\frac{\beta}{1-\alpha}$، در مرحلهٔ m ام به نمونه‌گیری ادامه می‌دهیم اگر (۲.۷) برقرار باشد، لذا نمونه‌گیری ادامه می‌یابد اگر

$$b=\ln(\frac{\beta}{1-\alpha})<S_m=\ln(R_m)<\ln(\frac{1-\beta}{\alpha})=a \qquad (۳.۹)$$

با توجه به تعریف متغیر ایست N و S_N می‌دانیم که S_N فقط دو مقدار بزرگتر یا مساوی a و یا کوچکتر یا مساوی b را اختیار می‌کند. حال با صرفنظر کردن انحراف S_N از کران‌های a و b داریم

$$E_{H_0}\approx aP_{H_0}(S_N=a)+bP_{H_0}(S_N=b)\approx a\alpha+b(1-\beta) \qquad (۳.۱۰)$$

و

$$E_{H_1}\approx aP_{H_1}(S_N=a)+bP_{H_1}(S_N=b)\approx a(1-\beta)+b\beta \qquad (۳.۱۱)$$

و اگر $E_{H_1}(Z_1)\neq 0; i=1,2$ باشد با استفاده از روابط (۳.۴) و (۳.۵) و لم (۳.۱) روابط (۳.۷) به راحتی اثبات خواهد شد، همچنین در حالتی که $E_{H_0}[Z_1]=0$ یا $E_{H_1}[Z_1]=0$ باشد با استفاده از لم (۳.۱) و قاعده مشتق‌گیری هوپیتال رابطه (۳.۸) به راحتی اثبات می‌شود که در آن $E_{H_i}(Z_1^2); i=1,2$ به صورت (۳.۶) تعریف می‌شود.

مثال ۱.۳: در مثال (۲.۱) برای انجام آزمون فرض (۲.۹) با خطای نوع اول و دوم $(\alpha,\beta)=(0/05$ و $0/05)$

$$a=-b=2/9444,\quad \theta_0=\frac{1}{2}\quad,\quad \theta_1=1\ ,\ v=1$$

لذا متوسط اندازه نمونه لازم برای انجام آزمون (۲.۹) به روش $SPRT$ تحت فرض‌های H_0,H_1 به ترتیب ۱۴ و ۷ می‌باشد.

٤. متوسط اندازه نمونه لازم برای انجام آزمون‌های یکطرفه به روش SPRT

در این بخش فرض می‌کنیم $X_1,X_2,\ldots$ دنباله‌ای از متغیرهای تصادفی مستقل از خانواده توزیع‌های (۲.۱) باشد، می‌خواهیم آزمون فرض

$$H_0:\theta\le\theta_0\quad v.s.\quad H_1:\theta\ge\theta_1 \qquad (۴.۱)$$

را به روش $SPRT$ برای زوج داده شدهٔ احتمال خطای نوع اول و دوم (α,β) با فرض $\theta_1>\theta_0$ انجام دهیم. در این حالت فاصلهٔ (θ_0,θ_1) فاصلهٔ بی‌تفاوتی نامیده می‌شود. ابتدا تابع مشخصه عملگر را تعریف می‌کنیم.

تعریف ۱.۴: فرض کنید $\overline{C}$ ناحیه پذیرش فرض H_0 در یک آزمون فرض باشد، تابع مشخصهٔ عملگر آزمون که آن را با $OC(\theta)$ نمایش می‌دهیم عبارت است از احتمال پذیرفته شدن فرض H_0 تحت مقدار واقعی θ، یعنی

$$OC(\theta)=P_\theta(X\in\overline{C})=1-\Pi(\theta) \qquad (۴.۲)$$

که در آن $\Pi(\theta)$ تابع توان آزمون است.

لم ۱.۴: در انجام آزمون فرض (۴.۱) با احتمال خطای نوع اول و دوم داده شده (α,β)، می‌توان $OC(\theta)$ را با استفاده از تقریب والد به صورت زیر محاسبه کرد

$$OC(\theta)\approx\frac{(\frac{1-\beta}{\alpha})^{h(\theta)}-1}{(\frac{1-\beta}{\alpha})^{h(\theta)}-(\frac{\beta}{1-\alpha})^{h(\theta)}} \qquad (۴.۳)$$

که در آن $h(\theta)$ تابع غیر صفری است که برای هر $\theta\in\Theta$

$$E_\theta\left[\left\{\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}\right\}^{h(\theta)}\right]=1 \qquad (۴.۴)$$

همچنین مشابه قضیه (۳.۱) می‌توان $E_\theta(N)$ را به صورت زیر تقریب زد. اگر $E_\theta(Z_1)\neq 0$ باشد.

$$E_\theta(N)\approx\frac{OC(\theta)\ln B+(1-OC(\theta)\ln A}{E_\theta(Z_1)} \qquad (۴.۵)$$

که در آن A و B به وسیله رابطه (۲.۶) تعریف می‌شوند و Z_1 در (۳.۱) تعریف شده است. همچنین هنگامی که $E_\theta(Z_1)=0$ باشد آنگاه $E_\theta(N)$ را به کمک رابطه (۳.۳) محاسبه می‌کنیم. بنابراین همانطور که ملاحظه می‌شود در دستیابی به تقریب‌های $OC(\theta)$

و $E_\theta(N)$ در به کارگیری روش $SPRT$ با زوج (α,β) داده شده، اساس کار پیدا کردن تابع $h(\theta)$ می‌باشد به طوری که در رابطهٔ (۴.۴) صدق کند. با استفاده از روابط (۴.۴) و (۲.۱) داریم.

$$1 = E_\theta\left[\left\{\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)}\right\}^{h(\theta)}\right]$$

$$= \int\left(\frac{\theta_0}{\theta_1}\right)^{vh(\theta)}\exp\left\{-T(x)\left[\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)h(\theta)+\frac{1}{\theta}\right]\right\}C(x)\theta^{-v}dx$$

$$= \int C(x)\exp\{-T(x)/\theta^*\}\theta^{*-v}\left\{\left(\frac{\theta^*}{\theta}\right)v\left(\frac{\theta_0}{\theta_1}\right)^{vh(\theta)}\right\}dx \quad (۴.۶)$$

که در آن

$$\frac{1}{\theta^*} = \left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)h(\theta)+\frac{1}{\theta} \quad (۴.۷)$$

تساوی (۴.۶) صحیح است اگر و تنها اگر

$$\left(\frac{\theta^*}{\theta}\right)^v\left(\frac{\theta_0}{\theta_1}\right)^{vh(\theta)} = 1$$

یا به طور معادل

$$\theta = \frac{\left\{\left(\frac{\theta_0}{\theta_1}\right)^{h(\theta)}-1\right\}\theta_0\theta_1}{(\theta_0-\theta_1)h(\theta)} \quad (۴.۸)$$

در اینجا برخلاف معمول به جای اینکه $h(\theta)$ را برحسب θ بیان کنیم، θ را برحسب $h(\theta)$ پیدا کرده‌ایم. اگر $h(\theta)\to 0$ آنگاه

$$\theta \to \frac{\theta_0\theta_1\ln\left(\frac{\theta_0}{\theta_1}\right)}{\theta_0-\theta_1} = \theta^{**} \quad (۴.۹)$$

حال برای محاسبهٔ $E_\theta(N)$ به کمک روابط (۳.۴) و (۳.۵) داریم

$$E_\theta(Z_1) = v\ln\left(\frac{\theta_0}{\theta_1}\right)+v\theta\left(\frac{\theta_1-\theta_0}{\theta_0\theta_1}\right) \quad (۴.۱۰)$$

و لذا $E_\theta(Z_1)=0$ اگر و تنها اگر $\theta=\theta^{**}$. بنابراین با انتخاب $A\approx\frac{1-\beta}{\alpha}$ ، $\alpha=\ln A$ ، $B\approx\frac{\beta}{1-\alpha}$ و $b=\ln B$ با استفاده از رابطه (۴.۵) متوسط اندازه نمونه لازم برای انجام آزمون (۴.۱) به روش SPRT در خانواده توزیع‌های (۲.۱) عبارت است از:

$$E_\theta(N) = \frac{bOC(\theta)+a(1-OC(\theta))}{v\ln\left(\frac{\theta_0}{\theta_1}\right)+v\theta\left(\frac{\theta_1-\theta_0}{\theta_0\theta_1}\right)},\ \theta\to\theta^{**} \quad (۴.۱۱)$$

همچنین برای حالتی که $\theta=\theta^{**}$، به کمک قاعده هوپیتال و رابطه (۳.۳) $E_\theta(N) = -\frac{ab}{v\left\{\ln\left(\frac{\theta_0}{\theta_1}\right)\right\}^2}$ است(جدول۱). برخی از مقادیر $h(\theta)$، $OC(\theta)$ و $E_\theta(N)$ را برای آزمون فرض (۴.۱) در خانواده توزیع‌های (۲.۱) را به ازای مقادیر مختلف θ نشان می‌دهد.

۵. تعمیم نتایج به دست آمده به خانوادهٔ توزیع‌های کای- دو تبدیل یافته

در این بخش زیرخانواده‌ای از خانوادهٔ توزیع‌های نمایی یک پارامتری موسوم به خانوادهٔ کای- دو تبدیل یافته که به وسیلهٔ رحمان و گوپتا[۷] معرفی شده است را معرفی کرده و ارتباط آن را با خانوادهٔ چگالی‌های (۲.۱) به دست می‌آوریم.

فرض کنید متغیر تصادفی X دارای توزیعی با تابع چگالی احتمال، متعلق به خانوادهٔ نمایی یک پارامتری زیر باشد

$$f(x,\eta) = \exp\{a(x)b(\eta)+c(\eta)+h(x)\} \quad (۵.۱)$$

رحمان و گوپتا [۷] قضیه زیر را برای خانوادهٔ توزیع‌های(۵.۱) بیان و ثابت کردند.

قضیه ۱.۵: در خانوادهٔ توزیع‌های نمایی (۵.۱)، $-2a(X)b(\eta)$ دارای توزیع $Gamma(\frac{k}{2},2)$ است اگر و تنها اگر

$$\frac{2c'(\eta)b(\eta)}{b'(\eta)} = k \quad (۵.۲)$$

که در آن k مثبت است و به η بستگی ندارد. در حالتیکه k عدد صحیح باشد $-2a(X)b(\eta)$ دارای توزیع کای- دو با k درجهٔ آزادی است.حال اگر قرار دهیم $\theta=-\frac{1}{b(\eta)}$ خانوادهٔ فوق را می‌توان به صورت

$$f(x,\theta) = C(x,m)\theta^{-\frac{k}{2}}\exp\{-\frac{a(x)}{\theta}\} \quad (۵.۳)$$

نوشت که در آن $\frac{2a(X)}{\theta} \sim Gamma(\frac{k}{2},2)$ یا $a(X) \sim Gamma(\frac{k}{2},\theta)$ است. واضح است این خانواده مشابه خانواده توزیع‌های (۲.۱) با

$$\nu = \frac{k}{2}$$

و

$$T(X) = a(X)$$

و

$$\theta = -\frac{1}{b(\eta)}$$

است (جدول۲).

تعدادی از توزیع‌های متعلق به خانواده توزیع‌های کای- دو تبدیل یافته را نشان می‌دهد.

حال اگر فرض کنیم $X_1, X_2, \ldots$ دنباله‌ای از متغیرهای تصادفی مستقل از خانواده توزیع‌های کای- دو تبدیل یافته باشند، توزیع توأم $X_1, X_2, \ldots, X_m$ به فرم (۲.۲) است که در آن

$$S(\mathbf{X}) = \sum_{i=1}^{m} a(X_i) \sim Gamma(\frac{mk}{2},\theta)$$

و

$$\theta = -\frac{1}{b(\eta)}$$

و

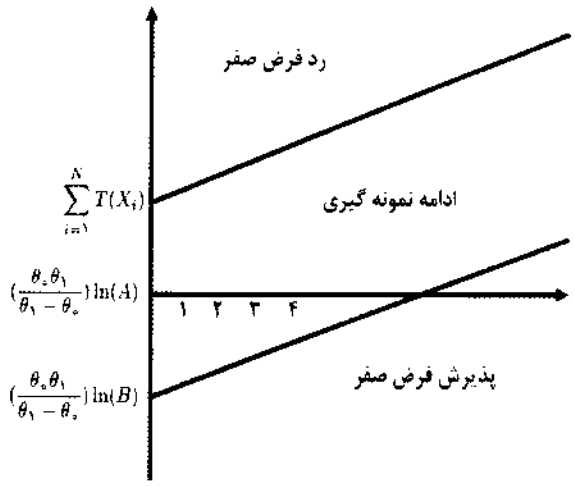
$$\nu = \frac{k}{2}$$

حال می‌توان تمامی نتایج بخشهای قبل را با جایگذاری $\frac{k}{2}$ بجای y، $S(\mathbf{X}) = \sum_{i=1}^{m} a(X_i)$ بجای $S(\mathbf{X}) = \sum_{i=1}^{m} T(X_i)$ و $-\frac{1}{b(\eta)}$ بجای θ، به خانواده توزیع‌های کای- دو تبدیل یافته تعمیم داد. برای مثال در انجام آزمون فرض (۲.۴) درباره پارامتر $\theta = -\frac{1}{b(\eta)}$ به روش $SPRT$ در خانواده توزیع‌های کای- دو تبدیل یافته با فرض $E_{H_i}(Z_1) \neq 0; i = 1,2$ متوسط اندازه نمونه لازم عبارت است از

$$E_{H_0}(N) = \frac{2\alpha a + 2b(1-\alpha)}{k\ln(\frac{\theta_0}{\theta_1}) + k(\frac{\theta_1 - \theta_0}{\theta_0})},$$

$$E_{H_1}(N) = \frac{2a(1-\alpha) + 2b\beta}{k\ln(\frac{\theta_0}{\theta_1}) + k(\frac{\theta_1 - \theta_0}{\theta_1})} \qquad (۵.۴)$$

به همین ترتیب می‌توان سایر نتایج به دست آمده در بخش‌های قبل را به خانواده توزیع‌های کای- دو تبدیل یافته تعمیم داد.



شکل ۱: نواحی سه‌گانه آزمون فرض (۲.۴) به روش دنباله‌ای

متوسط اندازه نمونه لازم برای آزمون ... ـــــــ

Name of distribution with p.d.f	a(X)	$b(\circ)$	$c(\circ)$	$-2a(X)b(\circ)$	k	$\theta=\frac{1}{b(\eta)}$
Normal $\frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{(2\pi)}}$	$\frac{1}{2}X^2$	$-\frac{1}{\sigma^2}$	$-\mathrm{Ln}\,\sigma$	$\frac{X^2}{\sigma^2}$	۱	σ^2
Lognormal $\frac{e^{-\frac{(\mathrm{Ln}\,x)^2}{2\sigma^2}}}{\sigma x\sqrt{(2\pi)}}$	$\frac{1}{2}\mathrm{Ln}\,X^2$	$-\frac{1}{\sigma^2}$	$-\mathrm{Ln}\,\sigma$	$\frac{\mathrm{Ln}\,X^2}{\sigma^2}$	۱	σ^2
Exponential $\frac{e^{-\frac{\lambda x}{}}}{\lambda}$	X	$-\frac{1}{\lambda}$	$-\mathrm{Ln}\,\lambda$	$\frac{2X}{\lambda}$	۲	λ
Gamma $x^{\alpha-1}\frac{e^{-\frac{x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}$ *with known* α	X	$-\frac{1}{\beta}$	$-\alpha\,\mathrm{Ln}\,\beta$	$\frac{2X}{\beta}$	2α	β
Rayleigh $x\frac{e^{-\frac{x^2}{2\beta^2}}}{\beta^2}$	$\frac{1}{2}X^2$	$-\frac{1}{\beta^2}$	$-2\mathrm{Ln}\,\beta$	$\frac{X^2}{\beta^2}$	۲	β^2
Pareto $\alpha x^{-(\alpha+1)}$	$\mathrm{Ln}\,X$	$-\alpha$	$\mathrm{Ln}\,\alpha$	$2a\mathrm{Ln}\,X$	۲	$\frac{1}{\alpha}$
Weibull $\beta\alpha x^{\alpha-1}e^{-\beta x^\alpha}$ *With known* α	X^α	$-\beta$	$\mathrm{Ln}\,\beta$	$2\beta X^\alpha$	۲	$\frac{1}{\beta}$
Maxwell $\sqrt{\frac{2}{\pi}}\beta^{\frac{3}{2}}x^2e^{-\frac{\beta}{2}x^2}$	$\frac{1}{2}X^2$	$-\beta$	$\frac{3}{2}\mathrm{Ln}\,\beta$	βX^2	۳	$\frac{1}{\beta}$
InverseGaussian $\sqrt{\frac{\lambda}{2\pi x^2}}e^{-\frac{\lambda(x-\mu)^2}{2\mu^2x}}$ *With known* μ	$\frac{1}{2}\frac{(X-\mu)^2}{\mu^2X}$	$-\lambda$	$\frac{\mathrm{Ln}\,\lambda}{2}$	$\frac{1}{2}\frac{\lambda(X-\mu)^2}{\mu^2X}$	۱	$\frac{1}{\lambda}$

جدول۲: بعضی از توزیع‌های متعلق به خانواده توزیع‌های کای- دو تبدیل یافته

مراجع

[۱] پارسیان، احمد، ۱۳۷۷، مبانی آمار ریاضی، مرکز نشر دانشگاه صنعتی اصفهان.

[2] Bahadur, R.R.,1954, Sufficiency and statistical decision functions, *Ann. Math. Statist., 25, 423-462.*

[3] Barnard, G. A.,1944, *Statistics, Mathematics and Quality Control, No QC/R/7, British Ministry of Supply*, London, England.

[4] Cox, D. R.,1952, A Note on the sequential estimation of means, *Proc. Comb. Philos. Soc, 48, 290-299.*

[5] Dodge, H. F., and Roming, H. G., 1929, A Method of Sampling Inspection, *Bell. Syst. Tech. J., 8, 613-631.*

[6] Lai, T. L.,2000, Sequential Analysis: Some classical problems and new challenges, *Statistica Sinica, 11, 303-408.*

[7] Rahman, M. S. and Gupta, R. R., 1993, Family of transformed chi-square distributions, *Comm. Statis. Theory Meth., 22, 135-146.*

[8] Siegmound, D., 1985, *Sequential Analysis*, Springer Verlage, New York.

[9] Wald, A., 1947, *Sequential Analysis*, Wiley, New York.

[10] Wald, A., 1943, *Statistical Research Group Report, NO. 75*, Columbia Univ., New York.

[11] Wolfowitz, J., 1947, The efficiency sequential estimates and Wald's equation for sequential processes, *Ann. Math. Statist., 18, 215-230.*