مروری بر آزمون های نیکویی برازش بر اساس آنتروپی نمونه

ناصررضا ارقامی، احسان زمان زاده
گروه آمار، دانشگاه فردوسی مشهد

سادگی برآوردگر آنتروپی بر اساس فاصله داده‌ها(بخصوص برآوردگر واسیچک(۱۹۷۶)) باعث شد که انواع آزمون های نیکویی برازش بر مبنای ماکسیمم آنتروپی و می نیموم متقاطع توسط مولفین مختلف ابداع و معرفی شود. در این مقاله تاریخچه تحول و آخرین روش های ابداع شده مورد بررسی قرار می گیرد.

واژه‌های کلیدی: آزمون نیکوی برازش ، آنتروپی، فاصله کولبک- لایبلر.

۱ مقدمه

از آنجا که صحت نتایج بسیاری از استنباط‌های آماری پارامتری مبتنی بر صحت این فرض هستند که مشاهدات نمونه از چه توزیعی ناشی شده‌اند، آزمون‌های نیکویی برازش اهمیت فوق العاده‌ای دارند.
از این آزمون‌ها در جعبه ابزار آمارشناسان کاربردی بسیار وجود دارد. آزمون‌های χ^2(پیرسون)، $K-S$(کلموگروف-اسمیرنوف) و کرامر-وان میزس از معروفترین آن‌ها هستند و بعضی نیز مخصوص آزمون نرمال بودن توزیع می‌باشند.(مانند آزمون SW (شاپیرو-ویلک) و LS(لیلی-فورس)) پس از مقاله واسیچک(۱۹۷۶) که درآن یک آزمون برای نیکویی برازش توزیع نرمال برمبنای آنتروپی نمونه ارائه نمود و با شبیه سازی نشان داد که از توان خوبی‌برخوردار است، آزمون‌های دیگری بر مبنای ماکزیمم آنتروپی و یا مینیموم فاصله‌کولبک-لایبلر توسط مولفین دیگر معرفی شد.
در بخش دوم این مقاله آزمون‌های نیکویی برازش مبتنی بر ماکزیمم آنتروپی (یا مینیموم فاصله کولبک-لایبلر) فهرست شده، ویژگی‌های آن ها مقایسه می‌شود. دربخش سوم آزمون‌های نیکویی برازشی که اخیرا ابداع شده‌اند، ارائه می‌گردد.

۲ آزمون‌های نیکویی برازش بر مبنای ماکزیمم آنتروپی (یا مینیموم فاصله کولبک- لایبلر)

۲-۱ آزمون واسیچک

زمانی که واسیچک آزمون نرمالیتی خود را ارائه داد، از این واقعیت استفاده کرد که توزیع نرمال در بین توزیع‌های با واریانس معلوم، دارای بیشترین آنتروپی است.
وی ابتدا نشان داد که آنتروپی یک متغیر تصادفی پیوسته را که به صورت

$$H(f) = -\int f(x)\log(f(x))dx$$

تعریف می‌شود، به صورت زیر نیز می‌توان نوشت

$$H(f) = \int_{0}^{1} \log\left\{\frac{dF^{-1}(p)}{dp}\right\} dp$$

که درآن F تابع توزیع توزیع مربوطه است. آنگاه با جایگزین کردن تابع توزیع‌تجربی به‌جای F و عملگر تفاضل به جای مشتق، برآورد آنتروپی یک توزیع را برمبنای یک نمونه تصادفی $X_1, \cdots, X_n$ از آن توزیع به صورت

$$\hat{H}(f) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{2m}{n}\left(X_{(i+m)} - X_{(i-m)}\right)\right) \quad (1)$$

تعریف کرد که در آن $X_{(1)} \leq \cdots \leq X_{(n)}$ آماره‌های مرتب نمونه است و $X_{(i)} = X_{(1)}$ به ازای $i < 1$ و $X_{(i)} = X_{(n)}$ به‌ازای $i > n$. در اینجا m ($m \leq \frac{n}{2}$) عددی است صحیح و مثبت و آن‌را اندازه پنجره نامند و بزرگتر بودن آن از یک به هموارتربودن برآورد چگالی کمک می‌کند.
از طریق زیر نیز می‌توان به برآورد فوق رسید. چون

$$H(f) = -E_f(\log f(X))$$

است می‌توان آن را با میانگین نمونه یعنی $\frac{1}{n}\sum\limits_{i=1}^{n}\log f(X_i)$ برآورد کرد. حال اگر به جای تابع‌چگالی مجهول f (که مشتق F نسبت به x است) از

$$f\left(x_i\right)=\frac{F\left(X_{(i+m)}\right)-F\left(X_{(i-m)}\right)}{X_{(i+m)}-X_{(i-m)}}$$

استفاده کرده و بجای F نیز تابع توزیع تجربی $F_n\left(x_{(i)}\right)=\frac{i}{n}$ را قرار دهیم، همان برآورد(۱) بدست می آید.
حال فرض کنید می‌خواهیم فرضیه $H_0: f \in N\left(\mu, \sigma^2\right)$ را در مقابل نقیض آن آزمون کنیم و μ, σ^2 هردومجهول‌اند. $\hat{H}\left(f\right)$ نسبت به تبدیلات مکانی پایاست ولی نسبت به تبدیلات مقیاس پایانیست. برای پایا بودن آن تحت تبدیلات مقیاسی از آماره

$$T_v=\hat{H}\left(f\right)-\log\left(S\right)$$

که $S^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ است، استفاده می‌کنیم و در نتیجه توزیع T_v به μ, σ^2 بستگی ندارد و مقادیربحرانی آن را می‌توان به ازای $\mu=0$ و $\sigma^2=1$ به کمک شبیه سازی به‌دست آورد. چون توزیع نرمال در بین همه توزیع‌های با واریانس معلوم، بیشترین آنتروپی را دارد، H_0 را به‌ازای مقادیر کوچک T_v (کوچکتر از مقدار بحرانی) رد می‌کنیم. واسیچک (۱۹۷۶) به کمک شبیه‌سازی نشان داد که آزمون نرمالیتی وی از توان خوبی نیز برخوردار است و نه تنها تقریبا به ازای همه توزیع‌های جانشین از آزمون $K-S$ (کلموگروف-اسمیرنوف) توان بیشتری دارد، بلکه، به‌ازای بعضی از توزیع‌های جانشین از آزمون پرتوان SW(شاپیرو-ویلک) نیز پرتوان‌تر است. خوبی عمده آزمون واسیچک (۱۹۷۶) در مقایسه با آزمون SW سادگی محاسبات آن است و برای محاسبه آماره آزمون نیازی به جداول مخصوص برای ضرایب لازم به ازای حجم نمونه‌های متفاوت ندارد.

۲-۲ آزمون نیکویی برازش بر اساس مینیموم فاصله کولبک لایبلر

ذیلا نشان می‌دهیم که آزمون واسیچک حالت خاصی از یک آزمون بسیار کلی‌تر است.(ارقامی و علیزاده ۲۰۱۰)
فرض کنید می‌خواهیم فرضیه $H_0: f \in G_\theta$ را در مقابل نقیض آن آزمون کنیم. فاصله کولبک-لایبلر بین f و هر عضو از خانواده G_θ عبارتست از

$$D\left(f||g_\theta\right)=\int\left[\log\frac{f\left(x\right)}{g_\theta\left(x\right)}\right]f\left(x\right)dx=E_f\left[\log\frac{f\left(x\right)}{g_\theta\left(x\right)}\right]=E_f\left[\log f\left(x\right)\right]-E_f\left[\log g_\theta\left(x\right)\right].$$

تعریف می‌کنیم

$$D_{Inf} = Inf_{\theta} D\left(f || g_{\theta}\right)$$

با توجه به ویژگی فاصله کولبک-لایبلر داریم

$$D_{Inf} = 0 \Leftrightarrow f \in G_{\theta}$$

بنابراین اگر $\hat{D}_{Inf}$ برآورد D_{Inf} باشد و H_0 را به ازای مقادیر بزرگ $\hat{D}_{Inf}$ رد کنیم، آزمون معقولی خواهیم داشت. چون

$$D_{Inf} = E_f\left[\log f\left(x\right)\right] - Sup_{\theta} E_f\left[\log g_{\theta}\left(x\right)\right]$$

می‌توان نوشت

$$\hat{D}_{Inf} = -\hat{H}\left(f\right) - Sup_{\theta}\left[\frac{1}{n}\sum_{i=1}^{n}\log g_{\theta}\left(x_i\right)\right]$$

که در آن $\hat{H}\left(f\right)$ برآورد آنتروپی واسیچک است. بدیهی است که جمله دوم عبارت فوق به‌ازای $\theta = \hat{\theta}$ که $\hat{\theta}$ MLEی θ است، ماکزیمم می‌شود. پس

$$\begin{aligned}\hat{D}_{Inf} &= \frac{1}{n}\sum_{i=1}^{n}\frac{n}{2m}\left(X_{(i+m)} - X_{(i-m)}\right) - \frac{1}{n}\sum_{i=1}^{n}\log g_{\hat{\theta}}\left(X_i\right) \\ &= \frac{1}{n}\left[\sum_{i=1}^{n}\hat{f}\left(X_i\right) - \sum_{i=1}^{n}\log g_{\hat{\theta}}\left(X_i\right)\right] = \frac{1}{n}\log\frac{\prod_{i=1}^{n}\hat{f}(X_i)}{\prod_{i=1}^{n}g_{\hat{\theta}}(X_i)} = \frac{1}{n}\log\lambda\end{aligned} \quad (2)$$

که $\lambda = \frac{\prod_{i=1}^{n}\hat{f}(X_i)}{\prod_{i=1}^{n}g_{\hat{\theta}}(X_i)}$. چون $\hat{f}$ (برآورد تابع چگالی واسیچک) تابعی از تابع توزیع تجربی (F_n) است و F_n برآورد درستنمایی ماکزیمم F می‌باشد، نتیجه می‌گیریم که λ آماره آزمون نسبت درستنمایی است. در حالت خاص که G_{θ} یک خانواده مکانی، مقیاسی یا مکانی-مقیاسی باشد، به راحتی می‌توان نشان داد که آماره آزمون فوق $(\hat{D}_{Inf})$ نسبت به تبدیلات به ترتیب مکانی، مقیاسی و مکانی-مقیاسی پایا است. لذا توزیع آماره به هیچ پارامتر مجهولی بستگی ندارد و مقادیر بحرانی را می‌توان به کمک شبیه‌سازی به‌ازای $\theta = \theta_0$ (که θ_0 یکی از مقادیر

فضای پارامتر تحت $H_{\circ}$ است.) به‌دست آورد وفرضیه $H_{\circ}$ به ازای مقادیر $\hat{D}_{Inf} > c$ رد می‌شود. اگر G_{θ} خانواده $N(\mu, \sigma^{۲})$ ($\mu, \sigma^{۲}$ هردو مجهول) باشد، داریم

$$\log g_{\theta}(x) = -\frac{۱}{۲}\log ۲\pi - \frac{۱}{۲}\log \sigma^{۲} - \frac{۱}{۲\sigma^{۲}}(x-\mu)^{۲}$$

حال اگر به جای $\sigma^{۲}$ برآورد درستنمایی ماکزیمم آن، یعنی $S^{۲} = \frac{۱}{n}\sum_{i=۱}^{n}(x_i - \bar{x})^{۲}$ را قرار دهیم، داریم

$$\frac{۱}{n}\log g_{\hat{\theta}}(x_i) = -\frac{۱}{۲}\log ۲\pi - \frac{۱}{۲}\log S^{۲} - \frac{۱}{۲}$$

و در نتیجه

$$\hat{D}_{Inf} = -\hat{H}(f) + \log S + \left(\frac{۱}{۲}\log ۲\pi + ۱\right)$$

که قسمت غیرثابت آن قرینه لگاریتم آماره آزمون واسیچک است. یعنی آزمون بر مبنای $\hat{D}_{Inf}$ و آزمون واسیچک (که بر اساس ماکزیمم آنتروپی است) درواقع یک آزمون هستند. لازم به ذکر است این مطلب در مورد خاص آزمون نیکویی برازش توزیع نرمال توسط آریزونو و اوهتا (۱۹۸۱) ارائه شد.

۲-۳ آزمون نمایی بودن بر اساس مینیموم فاصله کولبک-لایبلر

ابراهیمی و همکاران (۱۹۹۲) آزمونی برای نیکویی برازش توزیع نمایی بر مبنای آنتروپی ارائه نمودند که آن نیز حالت خاصی از آزمون کلی فوق است.

$$H_{\circ} : f \sim Exp(\theta)$$
$$H_{۱} : \sim H_{\circ}$$

آماره آزمون پیشنهادی آن‌ها به صورت زیر بود

$$-\hat{H}(f) - \log \bar{x} + ۱ \qquad (۳)$$

حال اگر در فرمول (۲) به‌جای $g_{\theta}(x)$ عبارت $\frac{۱}{\theta}e^{-\frac{x}{\theta}}$ را قرار دهیم، داریم

$$\hat{D}_{Inf} = -\hat{H}(f) - \log\hat{\theta} + \frac{\bar{x}}{\hat{\theta}}$$

که $\hat{\theta}$ برآورد درستنمایی ماکزیممθ یعنی $\bar{x}$ است. در نتیجه همان آماره (۳) به‌دست می‌آید.

۲-۴ آزمون یکنواختی

دادویچ و وندرمولن (۱۹۸۱) با استفاده از اینکه توزیع یکنواخت در بین همه توزیع‌های با تکیه‌گاه[۰, ۱] دارای بیشترین آنتروپی است، آزمونی با آماره

$$T_{DV} = -\hat{H}(f)$$

معرفی کردند که به ازای مقادیر کوچک T_{DV} فرضیه $H_0 : f \sim U(0,1)$ را رد می‌کند.

۲-۵ آزمون نیکویی برازش نرمال معکوس

مادهالکار و تیال (۲۰۰۲) با استفاده از این واقعیت که آزمون گوسی معکوس$IG(\mu,\lambda)$ در بین همه توزیع‌ها با تکیه‌گاه[۰,∞[، تحت دوشرط $E\left(X^{-2}\right) = \mu$ و $E\left(X^{2}\right) = \frac{1}{\mu} + \frac{1}{\lambda}$ دارای بیشترین آنتروپی است، آزمونی با آماره

$$T_{MT} = \hat{H}(f) - \log\left(\frac{w}{2}\right)$$

معرفی کردند که در آن

$$w^2 = \frac{\sum_{i=1}^{n}\left(\frac{1}{x_i^2} - \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i^2}}\right)}{n-1}$$

و H_0را به ازای مقادیر کوچک T_{MT} رد می‌کنند.

۲-۶ آزمون نیکویی برازش توزیع لاپلاس

چوی و کیم (۲۰۰۶) با استفاده از اینکه توزیع لاپلاس در بین همه توزیع‌های با تکیه‌گاه$\Re$ با میانگین معلوم دارای بیشترین آنتروپی است، آزمونی برای نیکویی برازش توزیع لاپلاس با آماره آزمون

$$T_{CK}=\frac{n}{۲m\hat{\theta}}\prod_{i=۱}^{n}\left(\tilde{Y}_{(i+m)}-\tilde{Y}_{(i-m)}\right)$$

که در آن $\hat{\theta}=\frac{۱}{n}\sum_{i=۱}^{n}\left|\tilde{Y}_{(i)}\right|$ و $\tilde{Y}_{(i)}=X_{(i)}-median\left(X_{(i)}\right)$ معرفی کردند. این آزمون فرضیه$H_{\circ}:f\sim Laplace$ را به ازای مقادیر کوچک T_{CK} رد می‌کند.
لازم به ذکر است که کلیه آزمون های نامبرده شده در بالا از طریق آزمون کلی براساس $\hat{D_{Inf}}$ قابل حصول است.

۲-۷ آزمون نمایی بودن بر اساس داده‌های سانسور شده

پارک (۲۰۰۳) نشان داد که می‌توان آنتروپی آماره‌های مرتب $X_{(۱)},\cdots,X_{(r)}$ $(r<n)$، از یک نمونه تصادفی به حجم n، را برحسب تابع مخاطره $h(x)$ به صورت زیر نوشت

$$\begin{aligned}H_{۱,۲,\cdots,r;n} &= -\left(\log n+\cdots+\log\left(n-r+۱\right)\right)\\ &\quad +r-n\int_{-\infty}^{+\infty}\left(۱-F_{r;n-۱}\left(x\right)\right)f\left(x\right)\log h\left(x\right)dx\end{aligned}$$

و بر این اساس آزمون نمایی بودن بر اساس داده‌های سانسورشده نوع II را، با آماره

$$T\left(m,n,r\right)=-\bar{H}\left(m,n,r\right)+\frac{r}{n}\left(\log\left(\frac{۱}{r}\left(\sum_{i=۱}^{r}X_{(i)}\left(n-r\right)X_{(r)}\right)\right)+۱\right)$$

که در آن$\bar{H}\left(m,n,r\right)$ برآورد آنتروپی داده‌های سانسورشده به صورت

$$\bar{H}\left(m,n,r\right)=\frac{۱}{n}\sum_{i=۱}^{r}\log\left\{\frac{n}{۲m}\left(X_{(i+m)}-X_{(i-m)}\right)\right\}-\left(۱-\frac{r}{n}\right)\log\left(۱-\frac{r}{n}\right)$$

است، معرفی کرد.

۲-۸ آزمون براساس داده‌های سانسورشده پیشرونده

بالاکریشنان، حبیبی و ارقامی (۲۰۰۹) آزمون فوق را به داده‌های سانسورشده پیشرونده تعمیم دادند. آماره این آزمون به صورت زیر است

$$T(w,n,m) = -H(w,n,m) + \frac{m}{n}\left[\log\left(\frac{1}{m}\sum_{i=1}^{m}(R_i+1)X_{(i)}\right)+1\right]$$

۳ آزمون‌های جدید

۳-۱

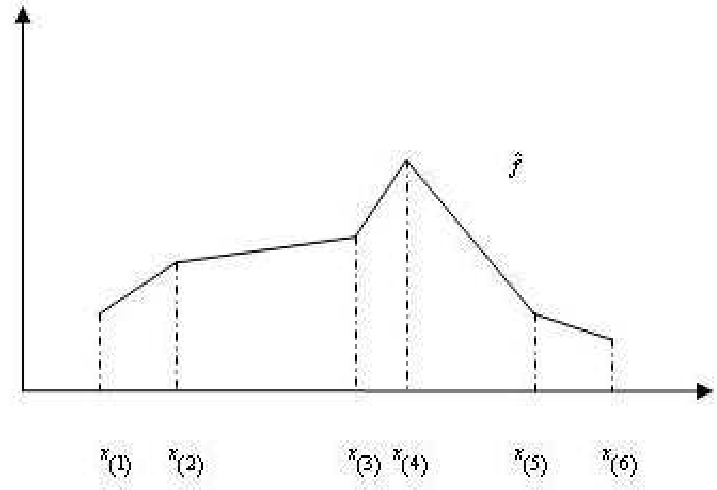
علیزاده (۲۰۱۰) برآوردی شبیه برآورد واسیچک ولی متفاوت از آن برای آنتروپی f به صورت زیر معرفی نمود

$$\hat{H}(f) = -\frac{1}{n}\sum_{i=1}^{n}\log\frac{\hat{f}\left(X_{(i+m)}\right)-\hat{f}\left(X_{(i-m)}\right)}{2}$$

که در آن $\hat{f}$ برآورد هسته تابع چگالی احتمال f است. وی نشان داد که با این برآورد آزمون نرمالیتی و نمایی بودن از توان بیشتری برخوردارند.

۳-۲

یوسف زاده و ارقامی (۲۰۰۸) برآورد جدیدی برای تابع چگالی احتمال به نام برآورد درستنمایی ماکزیمم مقید شده معرفی کردند که در آن $\prod_{i=1}^{n} f(x_i)$ با شرط پیوسته‌بودن توزیع (با تابع چگالی احتمال خطی قطعه‌ای) ماکزیمم می‌شود.
برآوردهای f و F به ترتیب عبارتند از



$$\hat{F}_n(x) = \begin{cases} 0 & x \le x_0 \\ \frac{1}{n+1} - \frac{1}{n+1}\frac{x_{(1)}-x}{x_{(1)}-x_{(0)}} & x_0 \le x \le x_{(1)} \\ \frac{n-1}{n(n+1)}\left(i + \frac{1}{n-1} + \frac{x-x_{(i-1)}}{x_{(i+1)}-x_{(i-1)}} + \frac{x-x_{(i)}}{x_{(i+2)}-x_{(i)}}\right) & x_{(i)} \le x \le x_{(i+1)} \\ \frac{n}{n+1} - \frac{1}{n+1}\frac{x-x_{(n)}}{x_{n+1}-x_{(n)}} & x_{(n)} \le x \le x_{n+1} \\ 1 & x \ge x_{n+1} \end{cases}$$

که در آن $x_0 = x_{(1)} - \frac{n}{n-1}\left(x_{(2)} - x_{(1)}\right), x_{n+1} = x_{(n)} + \frac{n}{n-1}\left(x_{(n)} - x_{(n-1)}\right)$.

$$\hat{f}_i = \frac{\hat{F}_n\left(X_{(i+m)}\right) - \hat{F}_n\left(X_{(i+m)}\right)}{X_{(i+m)} - X_{(i-m)}}$$

که در آن $X_{(1)} \le \cdots \le X_{(n)}$ آماره‌های مرتب نمونه است و $X_{(i)} = X_{(1)}$ به ازای $i < 1$ و $X_{(i)} = X_{(n)}$ به ازای $i > n$.

آن‌ها بر اساس این برآوردها، آزمون‌های نرمال بودن و نمایی بودن را انجام داده و با شبیه‌سازی نشان دادند که این آزمون‌ها از توان خوبی نسبت به آزمون‌های قبلی برخوردارند.

۳-۳ یک آزمون کلی‌تر

علیزاده و ارقامی (۲۰۱۰) با تغییر کوچکی در آماره $\hat{D_{Inf}}$ آزمون کلی جدیدی معرفی کردند. از (۲) داریم

$$\hat{D}_{Inf}=\frac{1}{n}\sum_{i=1}^{n}\frac{n}{2m}\left(X_{(i+m)}-X_{(i-m)}\right)-\frac{1}{n}\sum_{i=1}^{n}\log g_{\hat{\theta}}\left(X_i\right)$$

اگر در فرمول فوق به‌جای $g_{\hat{\theta}}(x)$ که مشتق $G_{\hat{\theta}}$ است از تفاضل $G_{\hat{\theta}}$ استفاده کنیم، یعنی اگر $g_{\hat{\theta}}(x)$ را با $\frac{G_{\hat{\theta}}(X_{(i+m)})-G_{\hat{\theta}}(X_{(i-m)})}{X_{(i+m)}-X_{(i-m)}}$ برآورد کنیم، برآورد دیگری به صورت زیر برای $\hat{D_{inf}}$ بدست می آید

$$\hat{\hat{D}}_{Inf}=-\frac{1}{n}\sum_{i=1}^{n}\log\left[\frac{G_{\hat{\theta}}\left(X_{(i+m)}\right)-G_{\hat{\theta}}\left(X_{(i-m)}\right)}{n/2m}\right]$$

بدست می آید. برآورد جدید D_{Inf} به نظر می‌رسد که از برآورد قبلی بهتر عمل کند زیرا g_θ همانند f از یک رابطه تفاضلی به‌دست می آید و این دو برآورد بیشتر هم‌آهنگ هستند، مضافا به اینکه در برآورد جدید $X_{(i+m)}-X_{(i-m)}$ که عامل بی‌ثباتی $\hat{f}$ است، نیز وجود ندارد.
نتایج شبیه سازی این حدس را تایید می‌کند، یعنی آزمون نیکویی برازش بر مبنای $\hat{\hat{D}}_{Inf}$ عموما از توان‌های بهتری (به خصوص در مقایسه با آزمون نمایی بودن ابراهیمی وهمکاران (۱۹۹۲)) برخوردارند. امتیاز دیگر آزمون‌های بر مبنای $\hat{\hat{D}}_{Inf}$ در قضیه زیربیان می‌شود.

قضیه ۱ اگر خانواده G_θ تحت گروه تبدیلات H پایا باشد، آنگاه $\hat{\hat{D}}_{Inf}$ نیز تحت همین گروه از تبدیلات پایا است.
(برای اثبات به علیزاده و ارقامی (۲۰۱۰) مراجعه شود)

معنی قضیه فوق این است که برخلاف آزمون‌های بر مبنای $\hat{D}_{Inf}$ که فقط در خانواده‌های مکانی-مقاسی کاربرد دارند، آزمون‌های نیکویی برازش بر مبنای $\hat{\hat{D}}_{Inf}$ در هر خانواده گروهی کاربرد دارند.

آزمون نیکویی برازش توزیع وایبل

خانواده توزیع‌ها با تابع چگالی $f_\theta(x) = \theta x^{\theta-1}\exp\left(-x^\theta\right)$ معروف به خانواده وایبل است. به سادگی معلوم می‌شود که خانواده فوق تحت تبدیلات $Y = X^c$ پایا است. بنابراین اگر آزمون فرضیه $H_0 : f \in W(\theta)$ را در مقابل نقیض آن به ازای مقادیر بزرگ $\hat{\hat{D}}_{Inf}$ رد کند، آماره آزمون به هیچ پارامتر مجهولی بستگی ندارد و می‌توان مقادیر بحرانی آزمون را به ازای $\theta = \theta_0$ (که θ_0 هر مقدار خاصی از θ می‌تواند باشد) به وسیله شبیه‌سازی بدست آورد.

۳-۴ آزمون نیکویی برازش بر اساس تبدیل متغیرها

علیزاده و ارقامی (۲۰۱۰) با توجه به اینکه تابع چگالی و تابع توزیع $U(0,1)$ به‌ترتیب برابر $f_0(x) = 1$ و $F_0(x) = x$ است، آماره

$$T_v = \frac{1}{n}\sum_{i=1}^{n}\left|x_i\hat{f}(x_i) - F_0(x_i)\right|$$

را، که در آن

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x - x_i}{h}\right)$$

برآورد هسته‌ای تابع چگالی احتمال است، برای آزمون یکنواختی توزیع با تکیه‌گاه[۰,۱] مورد استفاده قرار دادند. شبیه‌سازی ها نشان دادند که گرچه آزمون فوق به‌ازای بعضی از توزیع‌های جانشین برسایر آزمون‌های یکنواختی از نظر توان، ترجیح دارد اما به خوبی آزمون یکنواختی دادویچ و وندرمولن(۱۹۸۱) نیست. ولی آزمون نمایی بودن که براساس آزمون فوق و به شرح زیر معرفی می‌شود، نسبت به آزمون‌های رقیب به خصوص آزمون نمایی بودن ابراهیمی و همکاران(۱۹۹۲) از توان بسیار بالاتری برخوردار است.

فرض کنید که $X_1, \cdots, X_n$ یک نمونه تصادفی از توزیع پیوسته با تکیه گاه $(0, \infty)$ است. در این صورت الزیاد و الاوش (۱۹۹۲) نشان دادند که هریک از متغیرهای تصادفی

$$Y_{ij} = \frac{X_i}{X_i + X_j} i \neq j$$

دارای توزیع یکنواخت است اگر و تنها اگر توزیع مشترک X_iها نمایی باشد. علیزاده و ارقامی (۲۰۱۰) از قضیه فوق استفاده کرده و آماره زیر را برای آزمون فرضیه صفر $H_0 : X_i \sim \exp(\theta)\, i = 1, \cdots, n$ بر اساس آماره

$$T_E = \frac{۱}{n\,(n-۱)} \sum \left| y_{ij} \hat{f}\,(y_{ij}) - F_{\circ}\,(y_{ij}) \right|$$

معرفی نمودند و با شبیه‌سازی نشان دادند که این آزمون سازگار است و توان این آزمون به میزان قابل ملاحظه‌ای از توان آزمون نمایی بودن ابراهیمی و همکاران (۱۹۹۲) بیشتر است.

مراجع

[1] Alizadeh Noughabi, H. and Arghami, N.R. (2010). Testing exponentiality using transformed data, *To appear in Journal of Statistical Computation and Simulation*.

[2] Alizadeh Noughabi, H. (2010). A new entropy estimator and its application in testing normality, *To appear in Journal of Statistical Computation and Simulation*.

[3] Alizadeh Noughabi, H. and Arghami, N.R. (2010). General treatment of goodness of fit tests based on kullback-liebler information,*Submitted*.

[4] Alzaid, A.A. and Al-Osh, M.A (1992). Characterization of probability distributions based on the relation $Y = U(X_1 + X_2)$,*Sankhya*,Ser B **53**,188-190.

[5] Arizono, I. and Ohta, H. (1989).A test for normality based on kullback-liebler information, *The American Statistician*, **43**,20-22

[6] Balakrishnan, N. and Habibi Rad, A. and Arghami, N.R, (2007). Testing exponentiality Based on Kullback-Leibler information with progressively type II censored data, *IEEE Trans on Rel* **53**, 349-356.

[7] Choi,B. and Ki,K (2006). Goodness of fit test for laplace distribution based on maximum entropy, *Statistics* **40**, 517-531.

[8] Dudewicz, E.J and Van der Meulen, E.C. (1981),Entropy-Based Tests of Uniformity , *JASA* **76**, 967-974.

[9] Ebrahimi, N and Soofi., E.S. and Habibullah, M., (1992), Testing Exponentiality Based on Kullback-Leibler Information , *JRSS*, Ser B, **54**, 739-748.

[10] Mudholkar, G. S. and Tian, L., (2002) An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit test , *J. Statist. Plan. Infer*, **102**, 211-221.

[11] Park, S. (2005),Testing Expoentiality Based on the Kullback-Leibler Information With the type II Censored Data , *IEEE Transaction on Reliability*, **54**, 22-26.

[12] Vasicek, O. (1976), A test for Normality Based on Sample Entropy , *JRSS*,Ser B **38**, 54-59.

[13] Yousefzadeh F and Arghami, N.R (2008), Testing Exponentiality Based on Type II Censored Data and a New cdf Estimator, *Communications in Statistics-Simulation and Computation* **37**, 1479-1499.