آزمون‌های نیکویی برازش توزیع‌های کروی و بیضوی

محمد بلبلیان [۱] ، ناصر رضا ارقامی [۲]

چکیده:

توزیع‌های نرمال تک متغیره و چند متغیره در علم آمار نقش کلیدی دارند. با این وجود بسیاری از پدیده‌ها وجود دارند که از توزیع نرمال پیروی نمی‌کنند، پس باید امکان مدل بندی آماری با استفاده از توزیع‌های غیر نرمال چند متغیره نیز فراهم شود. در دهه‌های اخیر، از میان توزیع‌های غیر نرمال چند متغیره، توزیع‌های کروی و بیضوی [۳]، به علت دارا بودن خواص مطلوبی مشابه توزیع‌های نرمال چند متغیره، بیشتر از سایر توزیع‌ها مورد توجه قرار گرفته‌اند.

دراین مقاله به معرفی توزیع‌های بیضوی، خواص آنها و آزمون‌های یکنواخت برای نیکویی برازش [۴] توزیع‌های بیضوی می‌پردازیم.

واژه‌های کلیدی: توزیع بیضوی، توزیع کروی، آزمون نیکویی برازش، آماره‌های یکنواخت.

۱ مقدمه

توزیع‌های کروی و بیضوی به ترتیب تعمیم توزیع نرمال استاندارد چند متغیره $N_d(0, I_d)$ و توزیع نرمال چند متغیره $N_d(\mu, \Sigma)$ می‌باشند و می‌توان بسیاری از روش‌های قدیمی تحلیل‌های آماری را مستقیماً یا با اندکی تغییر در مورد جوامع با توزیع بیضوی بکار برد[۱،۱٦،۲۱،۲۵،۲٦].

در ادامه این مقاله در بخش ۲ توزیع‌های بیضوی معرفی خواهند شد. در بخش ۳ برخی از خواص توزیع‌های بیضوی را بررسی خواهیم کرد. در بخش ۴ مقدمه‌ای از آزمون‌های نیکویی برازش مطرح می‌کنیم. در بخش ۵ به معرفی آماره‌های یکنواخت خواهیم پرداخت، در بخش ٦ از آزمون‌های یکنواخت برای آزمون نیکویی برازش توزیع کروی استفاده خواهیم کرد. در بخش ۷ آزمون‌های یکنواخت را برای نیکویی برازش توزیع بیضوی بکار خواهیم برد و در بخش ۸ کارایی این روش‌ها را با ارائه یک مثال شبیه سازی شده بررسی خواهیم نمود.

فرض کنید ϕ یک ماتریس $p \times q$ تمام رتبه باشد $(q \leq p)$ و زیرفضای تولید شده توسط پایه ϕ با $S(\phi)$ نمایش داده شود. در این مقاله نماد $(.)^T$، برای نمایش ترانهاده یک ماتریس مورد استفاده قرار می‌گیرد. به عنوان مثال ϕ^T ترانهاده ماتریس ϕ می‌باشد، نماد $P_{\phi(\Sigma)}$ مبین عملگر تصویر متعامد بر زیر فضای $S(\phi)$ نسبت به حاصلضرب

[۱] دانشجوی کارشناسی ارشد آمار– دانشگاه فردوسی مشهد
[۲] استاد گروه آمار– دانشگاه فردوسی مشهد
[۳] Spherical and Elliptical Distribution
[۴] Goodness-of-Fit Test

داخلی $(a,b)=a^T\Sigma b$ می‌باشد،

$$P_{\phi(\Sigma)}=\phi(\phi^T\Sigma\phi)^{-1}\phi^T\Sigma$$

و از نماد $Q_{\phi(\Sigma)}=I-P_{\phi(\Sigma)}$ به عنوان عملگر تصویر متعامد بر متمم زیر فضای $S(\phi)$ نسبت به این حاصلضرب داخلی استفاده می‌شود. لازم به ذکر است که عملگرهای تصویر نسبت به حاصلضرب داخلی معمولی، بدون اندیس دوم، با نمادهای $P_\phi=P_{\phi(I)}$ و $Q_\phi=I-P_\phi$ نمایش داده می‌شوند.

۲ معرفی توزیع‌های بیضوی

یک بردار تصادفی بیضوی، مانند $\mathbf{x}$، بوسیله یک سه تایی (μ,D,g) مشخص می‌شود و بصورت $\mathbf{x}\sim \mathbf{ECp}(\mu,\mathbf{D},\mathbf{g})$ نمایش داده می‌شود که در آن $E(\mathbf{x})=\mu$، $Var(\mathbf{x})=\mathbf{\Sigma}=\mathbf{aD}$، که a یک مقدار ثابت مثبت است، و g تابعی است که یک خانواده خاص از توزیع‌های بیضوی را مشخص می‌نماید. اگر تابع چگالی بردار تصادفی بیضوی وجود داشته باشد بصورت زیر نوشته می‌شود:

$$f_{\mathbf{x}}(x)=|D|^{-\frac{1}{2}}g\left[(x-\mu)^TD^{-1}(x-\mu)\right]$$

به عنوان مثال خانواده نرمال چند متغیره ناویژه به این کلاس تعلق دارد، که در آن $a=1$ و $g(t)=exp\{-\frac{t}{2}\}$ در نظر گرفته می‌شود.

توزیع‌های کروی حالت خاص توزیع‌های بیضوی هستند و اگر در یک توزیع بیضوی $\mu=0$ و $D=I_p$ در نظر گرفته شود، توزیع حاصل توزیع کروی خواهد بود. در نتیجه اگر $\mathbf{y}$ دارای توزیع کروی باشد آنگاه، با انتخاب یک تبدیل مناسب، توزیع $\mathbf{x}=\mu+B\mathbf{y}$ بیضوی خواهد بود، که در آن $D=BB^T$ و در واقع $Var(\mathbf{x})\propto \mathbf{BB^T}$.

می‌توان بردار تصادفی بیضوی $\mathbf{x}_{p\times 1}$ را بصورت $\mathbf{x}=\mathbf{RBU}+\mu$ نیز نمایش داد، که در آن U یک بردار تصادفی است که بطور یکنواخت روی سطح کره واحد در $\mathbb{R}^p$ توزیع شده است و R یک متغیر تصادفی مثبت و مستقل از U می‌باشد که با استفاده از آن می‌توان به تابع g دست پیدا کرد. به عنوان مثال اگر R^2 دارای توزیع کای دو با p درجه آزادی باشد آنگاه $\mathbf{x}$ دارای توزیع نرمال p متغیره است.

اگر $\mathbf{x}\sim EC_p(\mu,D,g)$، و $f(.)$ تابع چگالی متغیر تصادفی R باشد، آنگاه رابطه میان $g(.)$ و $f(.)$ بصورت

$$f(r)=\frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}r^{n-1}g(r^2)$$

می‌باشد[۷].

۳ خواص توزیع‌های بیضوی

اکنون به اختصار برخی از خواص توزیع‌های بیضوی را معرفی می‌کنیم.

۱) اگر $\mathbf{x}$ یک بردار تصادفی $p\times 1$ از توزیع بیضوی، یعنی $\mathbf{x}\sim \mathbf{ECp}(\mu,\mathbf{D},\mathbf{g})$، باشد آنگاه $\phi^T\mathbf{x}$ نیز دارای توزیع بیضوی است و $\phi^T\mathbf{x}\sim EC_q(\phi^T\mu,\phi^TD\phi,g)$.

همچنین با شرطی کردن نیز یک توزیع بیضوی از همان خانواده بدست می‌آید. برای توضیحات بیشتر در این زمینه به [۲] و [۷] مراجعه کنید.

۲) $\mathbf{x}$ دارای توزیع بیضوی است اگر و تنها اگر به ازای هر ماتریس ϕی دلخواه، $E(\mathbf{x}|\phi^T\mathbf{x})$ یک تابع خطی از $\phi^T\mathbf{x}$ باشد[٦].

۳) فرض کنید $\mathbf{x}$ دارای توزیع بیضوی باشد، بطوری که $E(\mathbf{x})=\circ$ و ماتریس کوواریانس Σ مثبت معین باشد. اگر $E(\mathbf{x}|\phi^T\mathbf{x})$ تابعی خطی از $\phi^T\mathbf{x}$ باشد آنگاه $E(\mathbf{x}|\phi^T\mathbf{x}) = \Sigma\phi(\phi^T\Sigma\phi)^{-1}\phi^T\mathbf{x}$ [۴].

۴) با توجه به خاصیت‌های (۲) و (۳) می‌توان نتیجه گرفت:

اگر $\mathbf{x} \sim EC_p(\mu, D, g)$ باشد آنگاه:

$$E(\mathbf{x}|\phi^{\mathbf{T}}\mathbf{x}) = \mu + \mathbf{P}^{\mathbf{T}}_{\phi(\mathbf{\Sigma})}(\mathbf{x}-\mu)$$

۵) با انتخاب یک تابع تناسب، مانند ν، واریانس شرطی $Var(\mathbf{x}|\phi^T\mathbf{x})$ بصورت

$$\begin{aligned} Var(x|\phi^T x) &= \nu(\phi^T x)\left[\Sigma - \Sigma\phi(\phi^T\Sigma\phi)^{-1}\phi^T\Sigma\right] \\ &= \nu(\phi^T x)\Sigma Q_{\phi(\Sigma)} \\ &= \nu(\phi^T x) Q^T_{\phi(\Sigma)}\Sigma Q_{\phi(\Sigma)} \\ &= \nu(\phi^T x)\Sigma^{\frac{1}{2}} Q_{\Sigma^{\frac{1}{2}}\phi}\Sigma^{\frac{1}{2}}. \end{aligned}$$

در می‌آید، که در آن تابع تناسب $\nu(\phi^T\mathbf{x})$ تنها به فرم درجه دوم

$$(\mathbf{x}-\mu)^T\phi\left[Var(\phi^T\mathbf{x})\right]^{-1}\phi^T(\mathbf{x}-\mu)$$

وابسته است و ν ثابت است اگر و تنها اگر x دارای توزیع نرمال باشد[۱۵].

نکته قابل توجه از خاصیت‌های (۴) و (۵) این است که $E(\mathbf{x}|\phi^T\mathbf{x})$ و $Var(\mathbf{x}|\phi^T\mathbf{x})$، برای یک توزیع بیضوی دلخواه، همان فرم توزیع نرمال چند متغیره را دارا می‌باشد[۳].

در جدول (۱) تعدادی از توزیع‌های کروی به همراه تابع چگالی و یا تابع مشخصه شان معرفی شده‌اند.

همانگونه که عنوان شد متغیر تصادفی $\mathbf{x}$، با توزیع بیضوی، می‌تواند بصورت $\mathbf{x}= RBU + \mu$ نمایش داده شود. پس برای مشخص شدن یک توزیع بیضوی باید $E(\mathbf{x})$، $Var(\mathbf{x})$ و توزیع R معلوم باشد. میانگین و واریانس را می‌توان با استفاده از میانگین و واریانس نمونه‌ای کل داده‌ها یا میانگین و واریانس نمونه‌ای داده‌های پیراسته[۵]، که با حذف درصدی از نقاط دور افتاده بدست می‌آید، و یا روش‌های استوار بر مبنای داده‌های پیراسته برآورد نمود[۵]. انتخاب توزیع R، اختیاری است، بدین معنا که با انتخاب‌های مختلف برای توزیع R می‌توان توزیع‌های مختلفی بدست آورد، مثلاً برای دستیابی به توزیع نرمال، R^2 باید دارای توزیع کای دو باشد. در عمل برای دستیابی به یک انتخاب مناسب برای توزیع R می‌توان از توزیع تجربی شعاع‌های نمونه[٦]، که از رابطه

$$r_i = \left[(x_i-\bar{x})^T\hat{\Sigma}^{-1}(x_i-\bar{x})\right]^{\frac{1}{2}}$$

بدست می‌آیند، استفاده نمود[۴].

[۵] Trimmed Data
[٦] Empirical Distribution of the Sample Radii

۴ آزمون‌های نیکویی برازش

مسأله اصلی زمانی پیش می‌آید که یک نمونه تصادفی از یک توزیع نامشخص در اختیار آماردان قرار می‌گیرد، در این صورت برای تشخیص توزیع این داده‌ها باید از آزمون‌های نیکویی برازش استفاده نمود. کاریا و ایتُن [۱۴] و گوپتا و کابِ[۹] برخی از آزمون‌های استوار را برای نیکویی برازش آزمون‌های کروی براساس نمونه‌های غیر مستقل پیشنهاد نموده‌اند. اما زمانی که نمونه انتخابی $x_1, ..., x_n$، مستقل و هم‌توزیع[۷] با تابع توزیع $F(x)$، $(x \in \mathbb{R}^p)$ باشند، فرضیات آزمون بصورت زیر تعریف می‌شود.

H_0 : تابع توزیع $F(x)$ یک توزیع کروی است.

H_1 : تابع توزیع $F(x)$ یک توزیع غیر کروی است.

این فرضیات را می‌توان بصورت مشابه برای آزمون نیکویی برازش توزیع بیضوی نیز نوشت. برخی روش‌ها و آماره‌های موجود برای آزمون نیکویی برازش توزیع‌های کروی و بیضوی براساس نمونه‌های مستقل و هم‌توزیع، توسط فَنگ و لیانگ[۸] جمع آوری شده است و به برخی روش‌ها که در سال‌های اخیر برای این امر پیشنهاد شده است در لیانگ و همکاران [۱۷] اشاره شده است، آنها با استفاده از آماره‌های یکنواخت یک متغیره و چند متغیره چند روش برای آزمون توزیع کروی و بیضوی معرفی می‌نمایند. آماره‌های یکنواخت یک متغیره از کوسِنبِری و میلِر[۲۲] و میلِر و کوسِنبِری[۱۹] و آماره‌های یکنواخت چند متغیره از لیانگ، فَنگ و همکاران [۱۷] انتخاب شده است و با استفاده از یک مطالعه مونت کارلو نرخ خطای نوع اول و توان این آزمون‌ها مورد بررسی قرار گرفته است.

۵ معرفی آماره‌های یکنواخت

۱.۵ آماره‌های یکنواخت یک متغیره

ابتدا به معرفی آماره‌های یکنواخت یک متغیره که توسط کوسِنبِری و میلِر[۲۲] و میلِر و کوسِنبِری[۱۹] معرفی شده‌اند خواهیم پرداخت.

I) **آماره U^2 واتسون.**

فرض کنید

$$W^2 = 1/n + \sum_{i=1}^{n} \left[(2i-1)/2n - u_{(i)}\right]^2$$

باشد، که در آن $u_{(i)}$ ها آماره‌های ترتیبی نمونه مستقل و هم‌توزیع، $u_1, ..., u_n$ می‌باشند. واتسون[۲۴] آماره زیر را برای آزمون یکنواختی در بازه (۱و۰) پیشنهاد نموده است.

$$WU^2 = W^2 - n(\bar{u} - 0/5)^2,$$

که در آن $\bar{u}$ میانگین نمونه‌ای می‌باشد. جدول مقادیر بحرانی برای آماره تبدیل یافته زیر محاسبه شده است.

$$MU^2 = \left[WU^2 - \frac{1}{10n} + \frac{1}{(10n)^2}\right]\left(1 + \frac{0/8}{n}\right).$$

مقادیر بزرگ MU^2 مبین غیر یکنواختی توزیع نمونه می‌باشد. نکته قابل توجه در این آماره این است که وابستگی کمی به مقدار n دارد، این مقادیر بحرانی در[۲۳] بصورت $0/152$ $(\alpha = 0/01)$،

---
[۷] $i.i.d$

$۰/۱۸۵ (\alpha = ۰/۰۵)$ و $۰/۲۶۷ (\alpha = ۰/۱)$ محاسبه شده است.

II) **آزمون هموار نیمن**[۸].

چند جمله‌ای‌های لژاندر را در نظر بگیرید.

$$\begin{aligned}
\pi_۰(y) &= ۱,\\
\pi_۱(y) &= \sqrt{۱۲}(y - ۱/۲),\\
\pi_۲(y) &= \sqrt{۵}[۶(y - ۱/۲)^۲ - ۱/۲],\\
\pi_۳(y) &= \sqrt{۷}[۲۰(y - ۱/۲)^۳ - ۳(y - ۱/۲)],\\
\pi_۴(y) &= ۲۱۰(y - ۱/۲)^۴ - ۴۵(y - ۱/۲)^۲ + ۹/۸,
\end{aligned}$$

که در آن $y \in [۰, ۱]$ و با استفاده از آنها مقادیر t_r را بصورت زیر بسازید.

$$t_r = \sum_{i=۱}^{n} \pi_r(u_i), \quad r = ۱, ۲, ۳, ۴.$$

که در آن $u_۱, ..., u_n$ نمونه مستقل و هم‌توزیع می‌باشد. در این صورت آزمون هموار نیمن[۲۰] با چند جمله‌ای‌های درجه ۴ بصورت زیر تعریف می‌شود.

$$P_۴^۲ = \frac{۱}{n} \sum_{r=۱}^{۴} t_r^۲$$

مقادیر بزرگ $P_۴^۲$ مبین غیر یکنواختی توزیع نمونه می باشد. مقادیر بحرانی این آماره در[۱۹] و [۲۲] محاسبه شده است.

۲.۵ آماره‌های یکنواخت چند متغیره

حال به معرفی آماره‌های یکنواخت چند متغیره خواهیم پرداخت که با استفاده از آنها این فرضیه که نمونه مستقل و هم‌توزیع و d بعدی $z_۱, ..., z_n$ در فضای d بعدی واحد $C^d = [۰, ۱]^d$ دارای توزیع یکنواخت است، آزمون می‌شود.

$H_۰$: z_iها بطور یکنواخت در فضای C^d توزیع شده‌اند:

$H_۱ : H_۰$ نقیض

لیانگ و همکاران [۱۷] دو آماره یکنواخت چند متغیره برای آزمون یکنواختی در فضای C^d پیشنهاد می کنند، که برای محاسبه این آماره ها، هیکرنِل[۱۰] از سه اندازه انحراف[۹] در روش‌های مونت کارلوی مشابه[۱۰]، برای حل انتگرال‌های چندگانه[۱۱]، استفاده نموده است. فرض کنید $z_۱, ..., z_n$ یک نمونه تصادفی d بعدی در فضای C^d باشد، و

$$z_k = (z_{k۱}, ..., z_{kd})^T, \quad k = ۱, ..., n.$$

در این صورت با توجه به کار هیکرنِل سه انحراف زیر قابل تعریف خواهند بود.

۱) انحراف متقارن[۱۱]، که در آن

$$\begin{aligned}
U_۱ &= \frac{۱}{n} \sum_{k=۱}^{n} \prod_{j=۱}^{d} \left(۱ + ۲z_{kj} - ۲z_{kj}^۲ \right),\\
U_۲ &= \frac{۲^{d+۱}}{n(n-۱)} \sum_{k<l}^{n} \prod_{j=۱}^{d} \left(۱- \mid z_{kj} - z_{lj} \mid \right),
\end{aligned}$$

و $M = ۴/۳$، $\zeta_۱ = (۹/۵)^d - (۶/۹)^d$ و $\zeta_۲ = ۲^d - (۱۶/۹)^d$ می‌باشد.

---
[۸] Neyman Smooth Test
[۹] Measures of Discrepancies
[۱۰] Quasi Monte Carlo Methods
[۱۱] Symmetric Discrepancy

۲) انحراف مرکزی[۱۲]، که در آن

$$U_1 = \frac{1}{n}\sum_{k=1}^{n}\prod_{j=1}^{d}\left(1+\frac{1}{2}\mid z_{kj}-\frac{1}{2}\mid - \frac{1}{2}\mid z_{kj}-\frac{1}{2}\mid^2\right),$$

$$U_2 = \frac{2}{n(n-1)}\sum_{k<l}^{n}\prod_{j=1}^{d}\left(1+\frac{1}{2}\mid z_{kj}-\frac{1}{2}\mid + \frac{1}{2}\mid z_{lj}-\frac{1}{2}\mid - \frac{1}{2}\mid z_{kj}-z_{lj}\mid\right),$$

و $M = 13/12$، $\zeta_1 = (47/40)^d - (13/12)^{2d}$ و $\zeta_2 = (57/48)^d - (13/12)^{2d}$ می‌باشد.

۳) انحراف ستاره[۱۳]، که در آن

$$U_1 = \frac{1}{n}\sum_{k=1}^{n}\prod_{j=1}^{d}\left(\frac{3-z_{kj}}{2}\right),$$

$$U_2 = \frac{2}{n(n-1)}\sum_{k<l}^{n}\prod_{j=1}^{d}\left[2-max(z_{kj},z_{lj})\right],$$

و $M = 4/3$، $\zeta_1 = (9/5)^d - (16/9)^d$ و $\zeta_2 = (11/6)^d - (16/9)^d$ می‌باشد.

حال به معرفی دو روش برای آزمون یکنواختی در فضای C^d که در[۱۷] معرفی شده است می‌پردازیم.

I) **آماره تقریبی** $N(0,1)$.

تحت فرضیه H_0

$$A_n = \sqrt{n}\left[(U_1 - M^d) + 2(U_2 - M^d)\right] / \left(5\sqrt{\zeta_1}\right) \xrightarrow{\mathcal{D}} N(0,1),\ (n\to\infty)$$

که در آن " $\xrightarrow{\mathcal{D}}$ " به منزله همگرایی در توزیع احتمال می‌باشد و براساس سه اندازه انحراف متقارن، مرکزی و ستاره، سه انتخاب ممکن برای A_n داریم.

II) **آماره تقریبی** χ^2.

تحت فرضیه H_0

$$T_n = n\left[(U_1 - M^d),\ (U_2 - M^d)\right]\Sigma_n^{-1}\left[(U_1 - M^d),\ (U_2 - M^d)\right]^T \xrightarrow{\mathcal{D}} \chi^2(2),\ (n\to\infty)$$

که در آن

$$\Sigma_n = \begin{pmatrix} \zeta_1 & 2\zeta_1 \\ 2\zeta_1 & \frac{4(n-2)}{n-1}\zeta_1 + \frac{2}{n-1}\zeta_2 \end{pmatrix}$$

و مشابه آماره A_n براساس سه اندازه انحراف متقارن، مرکزی و ستاره، سه انتخاب ممکن برای T_n داریم.

مقادیر بزرگ $|A_n|$ یا T_n مبین غیر یکنواخت بودن توزیع نمونه مورد بررسی در فضای C^d می‌باشد.

مطالعه مونت کارلو بر روی نرخ خطای نوع اول، با استفاده از مقادیر بحرانی $N(0,1)$ (برای A_n) و χ^2 (برای T_n)، بیانگر این واقعیت است که این تقریب‌ها، حتی برای حجم نمونه‌های کوچک($n = 25$)، نتایج قابل قبولی دارند[۱۷].

۶ آزمون‌های یکنواخت برای نیکویی برازش توزیع کروی

فرض کنید $x_1,\ldots,x_n$ مشاهدات مستقل و هم‌توزیع، از توزیعی دلخواه در فضای d بعدی باشند و فرضیات آزمون را بصورت زیر در نظر بگیرید:

H_0 : x_iها دارای توزیع کروی هستند

H_1 : x_iها دارای توزیع غیر کروی هستند

---
[۱۲] Centered Discrepancy
[۱۳] Star Discrepancy

آزمون نیکویی برازش توزیع کروی به این داده‌ها بصورت زیر انجام می‌شود.

- بردارهای u_i را با استفاده از x_i ها بصورت زیر بسازید.

$$u_1 = x_1/||x_1||,\ \dots\ ,\ u_n = x_n/||x_n||$$

که در آن $u_i = (u_{i1}, ..., u_{id})^T$، $i = 1, ..., n$، و $||.||$ بیانگر نرم اقلیدسی یک بردار دلخواه می باشد.

- متغیرهای تصادفی مستقل $B_k(i)$ را بصورت زیر بسازید.

$$\begin{aligned}
B_1(i) &= u_{i1}^2 \sim \beta(1/2, (d-1)/2),\\
B_2(i) &= \left\{ \left(1 - u_{i1}^2\right)^{-1} u_{i2}^2 | u_{i1} \right\} \sim \beta(1/2, (d-2)/2),\\
\vdots \quad & \quad \vdots \qquad\qquad \vdots\\
B_k(i) &= \left\{ \left(1 - \sum_{j=1}^{k-1} u_{ij}^2\right)^{-1} u_{ik}^2 | (u_{i1}, ..., u_{i,k-1}) \right\}\\
&\sim \beta(1/2, (d-k)/2),\\
& \quad i = 1, ..., n; \quad k = 2, ..., d-1
\end{aligned}$$

که در آنها $\beta(1/2, (d-k)/2)$ مبین توزیع بتای یک متغیره می‌باشد($k = 2, ..., d-1$).

- حال فرض کنید $F_{b_j}(.)$ تابع توزیع، چگالی $\beta(1/2, (d-k)/2)$، ($j = 1, ..., d-1$)، باشد و بردارهای $v_i = (v_{i1}, ..., v_{i,d-1})^T$ را بصورت زیر بسازید.

$$v_{ij} = F_{b_j}\left(B_j(i)\right),$$
$$i = 1, ..., n; \quad j = 1, ..., d-1$$

اگر نمونه مستقل و هم‌توزیع، $x_1, ..., x_n$ دارای توزیع کروی باشند، آنگاه متغیرهای تصادفی v_{ij} ها متقابلاً مستقل و دارای توزیع یکنواخت $U(0, 1)$ می‌باشند و نقاط تصادفی $v_i = (v_{i1}, ..., v_{i,d-1})^T$ در فضای چند بعدی $[0, 1]^{d-1}$ بصورت یکنواخت توزیع می‌شوند(برای جزئیات بیشتر به [۱۷] مراجعه کنید). در این صورت آزمون نیکویی برازش توزیع کروی، با یک آزمون یکنواخت چند متغیره بصورت

H_0: v_i ها بطور یکنواخت در فضای $[0, 1]^{d-1}$توزیع شده‌اند:

H_1 : H_0 نقیض

و یا با یک آزمون یکنواخت یک متغیره بصورت

H'_0 : v_{ij} ها بطور یکنواخت در $(0, 1)$ توزیع شده‌اند

H'_1 : H_0 نقیض

جایگزین می‌شود و اگر H_0 یا H'_0 رد شد، آنگاه نتیجه می‌شود که نمونه $x_1, ..., x_n$ دارای توزیع کروی نیست. می‌توان از آماره‌های A_n و T_n برای آزمون فرضیه H_0 و از آماره‌های MU^2 و P_4^2 برای آزمون فرضیه H'_0 استفاده نمود. لازم به ذکر است که پذیرش H_0 معمولاً منجر به پذیرش H'_0 می‌شود، ولی عکس آن درست نیست و این نتیجه دلیلی بر این مدعا است که، توزیع یکنواخت همه توزیع‌های حاشیه‌ای یک بعدی منجر به تأیید توزیع یکنواخت چند متغیره توأم نمی‌شود[۱۷].

۷ آزمون‌های یکنواخت برای نیکویی برازش توزیع بیضوی

یک روش معمول برای ساخت آزمون‌هایی برای نیکویی برازش توزیع بیضوی، تبدیل آزمون‌های توزیع بیضوی به آزمون‌های توزیع کروی می‌باشد. فرض کنید $x_1, ..., x_n$ یک نمونه‌مستقل و هم‌توزیع از جامعه‌ای با تابع توزیع $F(x)$، $(x \in \mathbb{R}^d)$ و با گشتاورهای مرتبه دوم متناهی باشد و میانگین و واریانس نمونه‌ای را با استفاده از روابط زیر محاسبه کنید:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad S_n = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T.$$

برای کروی نمودن x_i ها از تبدیل

$$y_j = S_n^{-\frac{1}{2}}(x_j - \bar{x}), \quad j = 1, ..., n. \ (n > d)$$

استفاده می‌کنیم. براساس قضایای حدی آمار[۱۴] اگر x_iها دارای توزیع بیضوی باشند، آنگاه بردارهای تصادفی تبدیل یافته y_jها، به ازای حجم نمونه به اندازه کافی بزرگ، مستقل و دارای توزیع کروی می‌باشند[۱۷].

یکی از اشکالاتی که بر این روش وارد است این است که، حتی اگر داده‌ها دارای توزیع بیضوی باشند، y_jها به کندی به توزیع کروی همگرا می‌شوند، که این نقیصه را می‌توان با استفاده از استاندارد سازی (کروی نمودن) با تبدیل گرام—اشمیت[۱۵] از بین برد[۱۲]. در این روش

$$y_j = R(x_j - \bar{x}), \quad j = 1, ..., n.$$

که در آن ماتریس $R = R(S_n)$، یک ماتریس پایین مثلثی، با عناصر روی قطر اصلی مثبت می‌باشد و $RS_nR^T = I$.

۸ شبیه سازی

در این بخش می‌خواهیم مطالب عنوان شده در بخش‌های قبلی را بر روی یک مثال شبیه سازی شده اجرا نموده و کارایی آماره‌های A_n و T_n را بررسی کنیم. برای این کار یک نمونه ۲۵ (۵۰ و ۱۰۰) تایی از توزیع نرمال ۳ متغیره (۵ متغیره) استخراج می‌کنیم و با استفاده از آزمون‌های ارائه شده، این فرضیه را که این داده‌ها دارای توزیع کروی هستند، آزمون می‌نماییم. این عمل را ۱۰۰۰ بار تکرار می‌کنیم، نسبت دفعاتی را که این آزمون‌ها منجر به رد فرضیه کروی بودن می‌شوند در جدول ۲ و ۳ ارائه شده است.

همانگونه که در جداول فوق مشاهده می‌شود، نتایج شبیه سازی برای نمونه‌های کوچک نیز نتایج قابل قبولی هستند. لیانگ و همکاران [۱۷] نتایج شبیه سازی این روش‌ها را، بر روی چند توزیع کروی دیگر نیز ارائه نموده‌اند، که همگی کارایی این روش‌ها را تأیید می‌نماید.

---
[۱۴] Large Sample Theory
[۱۵] Gram Schmidt Transformation

جدول ۱: زیر گروه‌هایی از توزیع‌های کروی p بعدی

نوع	تابع چگالی $f(x)$ یا تابع مشخصه $\psi(t)$
نوع کوتز	$f(x)=c(x^Tx)^{N-1}exp(-r(x^Tx)^s)\quad , r,s>0, 2N+p>2$
نرمال چند متغیره	$f(x)=c\,exp(-\frac{1}{2}\,x^Tx)$
پیرسن نوع ۷	$f(x)=c(1+\frac{x^Tx}{s})^{-N}\quad , s>0, N>\frac{p}{2}$
t چند متغیره	$f(x)=c(1+\frac{x^Tx}{s})^{-\frac{(p+m)}{2}}\quad , m>0\quad , m\in Z$
کوشی چند متغیره	$f(x)=c(1+\frac{x^Tx}{s})^{-\frac{(p+1)}{2}}\quad , s>0$
پیرسن نوع ۲	$f(x)=c(1-x^Tx)^m\quad , m>0$
لجستیک	$f(x)=c\frac{exp(-x^Tx)}{[1+exp(-x^Tx)]^2}$
بسل چند متغیره	$f(x)=c(\frac{\|x\|}{\beta})^aK_a(\frac{\|x\|}{\beta})\quad , a>-\frac{p}{2}, \beta>0,$ که در آن $K_a(.)$ مبین تابع بسل نوع سوم است.
ترکیب مقیاس	$f(x)=c\int_0^\infty t^{-\frac{p}{2}}exp(-\frac{x^Tx}{2t})dG(t),$ که در آن $G(t)$ یک تابع توزیع پیوسته است.
قوانین پایا	$\psi(t)=exp\{r(t^Tt)^{\frac{\alpha}{2}}\}\quad , 0<\alpha\le 2, r<0$
یکنواخت چند متغیره	$\psi(t)={}_0F_1(\frac{p}{2};-\frac{1}{4}\|t\|^2),$ که در آن ${}_0F_1(.;.)$ یک تابع فوق هندسی تعمیم یافته است.

جدول ۲ نتایج آزمون کروی بودن بر روی داده‌های شبیه سازی شده از توزیع نرمال ۳ متغیره

آماره	A_n			T_n		
انحراف	متقارن	مرکزی	ستاره	متقارن	مرکزی	ستاره
$n=25$	۰/۰۶۲	۰/۰۵۹	۰/۰۵۱	۰/۰۵۷	۰/۰۴۹	۰/۰۵۵
$n=50$	۰/۰۵۴	۰/۰۶۲	۰/۰۵۸	۰/۰۵۰	۰/۰۵۳	۰/۰۵۳
$n=100$	۰/۰۵۰	۰/۰۵۴	۰/۰۵۲	۰/۰۵۲	۰/۰۵۸	۰/۰۶۰

جدول ۳ نتایج آزمون کروی بودن بر روی داده‌های شبیه سازی شده از توزیع نرمال ۵ متغیره

آماره	A_n			T_n		
انحراف	متقارن	مرکزی	ستاره	متقارن	مرکزی	ستاره
$n=25$	۰/۰۶۵	۰/۰۶۰	۰/۰۵۴	۰/۰۵۲	۰/۰۴۸	۰/۰۵۴
$n=50$	۰/۰۵۹	۰/۰۶۲	۰/۰۵۹	۰/۰۵۱	۰/۰۵۰	۰/۰۵۹
$n=100$	۰/۰۵۶	۰/۰۵۷	۰/۰۵۲	۰/۰۵۵	۰/۰۵۶	۰/۰۶۲

مراجع

[1] Anderson, T. W. ,1993, Nonnormal Multivariate Distribution: Inference Based on Elliptically Contoured Distributions, in *Multivariate Analysis: Future Directions*, (ed. C. R. Rao), Elsevier Science Publishers, 1-24.

[2] Cambanis, S., Huang, S. and Simons, G. ,1981, On the Theory of Elliptically Contoured Distribution, *Journal of Multivariate Analysis*, 11, 368-385.

[3] Cook, R. D. ,1992, Regression Plotting Based on Quadratic Predictors, *Statistical Analysis and Related Methods*, 115-128.

[4] Cook, R. D. ,1998, *Regression Graphics: Ideas for Studying Regressions through Graphics,* Wiley, New York.

[5] Cook, R. D. and Nachtsheim, C. J. ,1994, Re-weighting to Achieve Elliptically Contoured Covariates in Regression, *J. Amer. Statist. Assoc.*, 89, 592-600.

[6] Eaton, M. L. ,1986, A Characterization of Spherical Distribution, *Journal of Multivariate Analysis*, 20, 272-276.

[7] Fang, K. T., Kotz, S. and Ng, K. W. ,1990, *Symmetric Multivariate and Related Distributions,* Chapman and Hall, London.

[8] Fang, K. T. and Liang, J. ,1999, Testing Spherical and Elliptical Symmetry, *Encyclopedia of Statistical Sciences,* (ed. Kotz S., Read, C. B., Banks, D. L.), Vol. 3, 686-691.

[9] Gupta, A. K. and Kabe, D. G. ,1993, Multivariate Robust Tests for Spherical Symmetry with Applications to Multivariate Least Squares Regression, *Journal of Applied Statistical Science*, 1, 2, 159-168.

[10] Hickernell, F. J. ,1998, A Generalized Discrepancy and Quadrature Error Bound, *Mathematics of Computation*, 67, 299-322.

[11] Hua, L. G. and Wang, Y. ,1987, *The Application of Number Theory to Approximate Analysis*, Science Press, Beijing, China.

[12] Huffer, F. W. and Park, C. ,2005, A Test for Elliptical Symmetry. *Accepted from Journal of Multivariate Analysis.*

[13] Johnson, M. ,1987, *Multivariate Statistical Simulation,* Wiley, New York.

[14] Kariya, T. and Eaton, M. L. ,1977, Robust Tests for Spherical Symmetry, *Annals of Statistics*, 5, 206-215.

[15] Kelker, D. ,1970, Distribution Theory of Spherical Distributions an a Location-Scale Parameter Generalization. *Sankhya*, Series A, 32, 419-430.

[16] Lange, K. L., Little, R. J. A. and Taylor, J. M. G. ,1989, Robust Statistical Modeling Using the t-Distribution, *J. Amer. Statist. Assoc.*, 84, 881-896.

[17] Liang, J., Fang K. T. and Hickernell, F. J. ,2007, Some Necessary Uniform Test for Spherical and Elliptical Symmetry, *Ann. Inst. Statist. Math.*, In Press.

[18] Liang, J., Fang, K. T., Hickernell, F. J. and Li, R. ,2001, Testing Multivariate Uniformity and Its Applications, *Mathematics of Computation*, 70, 337-355.

[19] Miller, F. L. J. and Quesenberry, C. P. ,1979, Power Studies of Tests for Uniformity, II. *Communications in Statistics-Simulation and Computation*, B8(3), 271-290.

[20] Neyman, J. ,1937, Smooth Test for Goodness of Fit.*J. Amer. Statist. Assoc.*, 20, 149-199.

[21] Olkin, I. ,1992, Multivariate Non-Normal Distributions and Models of Dependency, *Multivariate Analysis and Its Applications* 1994, 24, 37-53.

[22] Quesenberry, C. P. and Miller, F. L. J. ,1977, Power Studies of Tests for Uniformity, *Journal of Statistical Computation and Simulation*, 5, 169-191.

[23] Stephens, M. A. ,1970, Use of the Kolmogorov Smirnov, Cramer-von Mises and Related Statistics Without Extensive Tables. *J. R. Statist. Soc. Ser. B*, 32, 115-122.

[24] Watson, G. S. ,1962, Goodness-of-Fit Tests on a Circle, *Biometrika*, 49, 57-63.

[25] Yuan, K. H. and Bentler, P. M. ,2004, On the Asymptotic Distributions of Two Statistics for Two-Level Covariance Structure Models Within the Class Of Elliptical Distribution, *Psychometrica*, In Press.

[26] Zellner, A. ,1976, Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms,*J. Amer. Statist. Assoc.*, 71, 400-405.