بحثی پیرامون مفهوم تقارن و عدم تقارن

افشین فلاح [۱] ،حمیدرضا چاره [۲] ، عباس گرامی [۳]

چکیده:

در این مقاله مفهوم تقارن و انواع آن برای توزیع های یک متغیره و چند متغیره و برخی معیارهای معمول برای اندازه‌گیری عدم تقارن مورد بررسی قرار گرفته است. مروری بر روش هایی که براساس آنها بتوان بر پایه یک توزیع متقارن، توزیعی ساخت که دارای قابلیت مدل بندی جوامع نامتقارن باشد، از اهداف بعدی این مقاله است.

واژه‌های کلیدی: چولگی، تقارن، تقارن کروی، تقارن بیضوی ، تقارن مرکزی، تقارن زاویه‌ای.

۱ مقدمه

از زمان های بسیار قدیم تقارن [۴] مفهومی مهم در هنر و ریاضیات بوده و کاربردهای متفاوتی داشته است. تقارن در زیباشناسی یک قانون، در ریاضیات محصول ساختمان هندسه و در فلسفه، تعادل و هارمونی است. تقارن یکی از ویژگی های اصلی معماری اسلامی است. در نظریه احتمال تقارن توزیع های احتمال دارای تعابیر مختلفی است. تقارن را می‌توان به عنوان یک ویژگی ساختاری برای یک توزیع و بر اساس تابع توزیع، تابع مشخصه یا تابع چگالی آن توزیع تعریف کرد. از طرف دیگر می‌توان کلاسی از توزیع‌ها که تعمیم آنها توزیع نرمال چند متغیره باشد را جستجو کرد. معمولاً نحوهٔ توزیع متغیر تصادفی X حول θ را برای تعریف تقارن مورد استفاده قرار می‌دهند. ماردیا [۱۰] برای چولگی و کشیدگی در حالت چند متغیره معیارهایی پیشنهاد کرد. مالکوویچ و افیفی [۹] و آیزوگای [٦] معیارهایی برای آزمون تقارن و نرمال بودن در حالت چند متغیره ارائه نمودند. آوِروس و مِست [۲] روشهایی را برای اندازه‌گیری چولگی توزیع‌های چند متغیره ارائه دادند. بران و میلر [۴] مدلهای چند متغیره متقارن را مورد مطالعه قرار دادند. آزالینی [۳] تعمیمی از توزیع نرمال ارائه داد که دارای قابلیت مدل بندی چولگی نیز می‌باشد. این تعمیم بعدها مبنای مطالعات زیادی برای افزودن قابلیت مدل بندی چولگی به توزیع های متقارن کلاسیک شد. در این مقاله ابتدا در بخش ۲ مفهوم تقارن در توزیع‌های یک متغیره و چند متغیره مورد بحث قرار می‌گیرد. سپس در بخش ۳ برخی معیارهای اندازه‌گیری عدم تقارن یا چولگی [۵] معرفی می‌شوند. در بخش ۴ روش های توانا ساختن یک توزیع متقارن به اندازه‌گیری عدم تقارن، مورد مطالعه قرار

---

[۱] استادیار دانشگاه بین المللی امام خمینی
[۲] کارشناس ارشد آمار ریاضی، دانشگاه بین المللی امام خمینی
[۳] دانشیار دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه تهران
[۴] Symmetry
[۵] Skewness

می‌گیرد. در بخش ۵ توزیع‌هایی با قابلیّت مدلبندی عدم تقارن براساس برخی توزیع‌های متقارن کلاسیک معرفی و ویژگی های آنها بررسی می‌شود.

۲ تقارن و انواع آن

متغیر تصادفی X با توزیع تک مدی $f_X(x)$ را حول θ متقارن گویند، هرگاه رابطه

$$X-\theta \stackrel{D}{=} \theta - X$$

برقرار باشد، که در آن $\stackrel{D}{=}$ به معنی هم توزیعی است. معمولاً منحنی فراوانی نرمال استاندارد به دلیل تقارن حول مبدأ مبنای قضاوت در مورد تقارن و سنجش میزان چولگی توزیع های احتمالی تک مدی قرار می‌گیرد. اما در توزیع‌های چند متغیره، تقارن انواع متفاوتی دارد. در این قسمت انواع تقارن چند متغیره مانند تقارن کروی، بیضوی، مرکزی و زاویه‌ای مورد بحث قرار می‌گیرند.

تقارن کروی. بردار تصادفی $\boldsymbol{X}_{d\times 1}$ دارای توزیع متقارن کروی[۶] حول بردار $\boldsymbol{\theta}_{d\times 1}$ است، هرگاه چرخش $\boldsymbol{X}$ حول $\boldsymbol{\theta}$ به ازای هر ماتریس متعامد $\boldsymbol{A}_{d\times d}$ تغییری در توزیع ایجاد نکند. به عبارتی رابطهٔ

$$\boldsymbol{X}-\boldsymbol{\theta} \stackrel{D}{=} \boldsymbol{A}(\boldsymbol{X}-\boldsymbol{\theta})$$

برقرار باشد. تابع مشخصهٔ $\boldsymbol{X}$ به صورت $e^{i\boldsymbol{t}'\boldsymbol{\theta}}h(\boldsymbol{t}'\boldsymbol{t})$ و تابع چگالی احتمال آن در صورت وجود بصورت $g((\boldsymbol{x}-\boldsymbol{\theta})'(\boldsymbol{x}-\boldsymbol{\theta}))$ است، که در آن $g(\cdot)$ یک تابع نامنفی است.

توزیع چند متغیره نرمال با ماتریس کوواریانس $\sigma^2 I_d$ تنها توزیع متقان کروی نیست، بلکه توزیع‌های دیگری نیز مانند توزیع t استاندارد چند متغیره و توزیع لوژستیک نیز چنین هستند. تقارن کروی ویژگی های جالبی دارد، از جمله اینکه اگر $\|\cdot\|$ نشان دهندهٔ نرم اقلیدسی باشد، در این صورت $\|\boldsymbol{X}-\boldsymbol{\theta}\|$ و بردار یکه تصادفی آن $\frac{\boldsymbol{X}-\boldsymbol{\theta}}{\|\boldsymbol{X}-\boldsymbol{\theta}\|}$، مستقلند.

تقارن بیضوی. بردار تصادفی $\boldsymbol{X}_{d\times 1}$ دارای توریع متقارن بیضوی[۷] با پارامترهای $\boldsymbol{\theta}_{d\times 1}$ و $\boldsymbol{\Sigma}_{d\times d}$ است، هرگاه با بردار تصادفی متقارن کروی $\boldsymbol{Y}_{d\times 1}$ متناسب باشد. به عبارتی رابطه

$$\boldsymbol{X} \stackrel{D}{=} \boldsymbol{A}'\boldsymbol{Y}+\boldsymbol{\theta}$$

برقرار باشد، که در آن $\boldsymbol{\Sigma}=\boldsymbol{A}'\times\boldsymbol{A}$ و رتبه $\boldsymbol{\Sigma}$ برابر $k(\le d)$ است. تابع مشخصهٔ $\boldsymbol{X}$ بصورت $e^{i\boldsymbol{t}'\boldsymbol{\theta}}h(\boldsymbol{t}'\Sigma\boldsymbol{t})$ و تابع چگالی احتمال آن در صورت وجود به صورت

$$|\Sigma|^{-\frac{1}{2}}g((\boldsymbol{x}-\boldsymbol{\theta})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\theta}))$$

است، که در آن $g(\cdot)$ تابعی نامنفی است. می‌توان اثبات کرد که خانواده چگالی‌های متقارن بیضوی تحت تبدیلات خطی و شرطی کردن بسته هستند. تقارن کروی و بیضوی برای ماتریس تصادفی نیز به صورت مشابه تعریف می‌شوند. ماتریس تصادفی $\boldsymbol{X}_{n\times m}$ دارای توزیع تقارن کروی است، هرگاه رابطهٔ

$$\boldsymbol{X} \stackrel{D}{=} \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$$

برقرار باشد، که در آن $\boldsymbol{A}_{n\times n}$ و $\boldsymbol{B}_{m\times m}$ ماتریسهای متعامد هستند. تقارن بیضوی نیز با یک تبدیل خطی از روی تقارن کروی به دست می‌آید.

[۶] Spherical Symmetry
[۷] Elliptical Symmetry

تقارن مرکزی و علامت. بردار تصادفی $\boldsymbol{X}_{d\times 1}$ دارای توزیع متقارن مرکزی[۸] حول $\boldsymbol{\theta}$ است، هرگاه رابطهٔ

$$\boldsymbol{X}-\boldsymbol{\theta} \stackrel{D}{=} \boldsymbol{\theta}-\boldsymbol{X}$$

برقرار باشد. چگالی آن در صورت وجود در رابطه $f(\boldsymbol{\theta}-\boldsymbol{x}) \stackrel{d}{=} f(\boldsymbol{x}-\boldsymbol{\theta})$ صدق می‌کند. این نوع تقارن گاهی تقارن ساده نامیده می‌شود. تقارن علامت حالتی بین تقارن کروی و مرکزی است [۴].

به نوعی می‌توان گفت $\boldsymbol{X}$ حول $\boldsymbol{\theta}$ متقارن مرکزی است، اگر و فقط اگر برای هر نیم فضای بسته $\boldsymbol{H}\in\mathcal{R}^d$ رابطه

$$P(\boldsymbol{X}-\boldsymbol{\theta}\in\boldsymbol{H})=P(\boldsymbol{X}-\boldsymbol{\theta}\in-\boldsymbol{H})$$

برقرار باشد. به عبارت دیگر به ازای هر بردار یکه $\boldsymbol{u}$ در $\mathcal{R}^d$ رابطه $\boldsymbol{u}'(\boldsymbol{X}-\boldsymbol{\theta}) \stackrel{D}{=} \boldsymbol{u}'(\boldsymbol{\theta}-\boldsymbol{X})$ برقرار باشد. برای نشان دادن تفاوت بین تقارن کروی و مرکزی، توزیع یکنواخت d بعدی روی $[-c,c]^d$ را در نظر بگیرید، که متقارن مرکزی است، اما متقارن کروی نیست.

تقارن زاویه‌ای. بردار تصادفی $\boldsymbol{X}_{d\times 1}$ دارای تقارن زاویه‌ای[۹] حول $\boldsymbol{\theta}_{d\times 1}$ است، هرگاه رابطهٔ

$$\frac{\boldsymbol{X}-\boldsymbol{\theta}}{\parallel \boldsymbol{X}-\boldsymbol{\theta}\parallel} \stackrel{D}{=} \frac{\boldsymbol{\theta}-\boldsymbol{X}}{\parallel \boldsymbol{X}-\boldsymbol{\theta}\parallel}$$

برقرار باشد، به عبارت دیگر هرگاه $\frac{\boldsymbol{X}-\boldsymbol{\theta}}{\|\boldsymbol{X}-\boldsymbol{\theta}\|}$ دارای توزیع متقارن مرکزی باشد.

این نوع تقارن تعمیمی از تقارن کروی است و برخی ویژگی های آن عبارتند از [۷]:

۱) نقطه $\boldsymbol{\theta}$ در تقارن زاویه‌ای در صورت وجود یکتا است، مگر اینکه توزیع حول یک خط تمرکز یافته باشد، که در این صورت توزیع احتمال روی خط بیشتر از توزیع احتمال میانه ($\frac{1}{2}$) است.

۲) اگر نقطه $\boldsymbol{\theta}$ در تقارن زاویه‌ای وجود داشته باشد، آنگاه هر ابر صفحه گذرنده از $\boldsymbol{\theta}$، $\mathcal{R}^d$ را به دو نیم فضا با احتمال برابر تقسیم می‌کند. عکس این مطلب هم برقرار است.

۳) اگر نقطه $\boldsymbol{\theta}$ در تقارن زاویه‌ای وجود داشته باشد، آنگاه $\boldsymbol{\theta}$ میانه توزیع شرطی $\boldsymbol{X}$ روی $\boldsymbol{\theta}$ است.

از آنجا که توزیع متقارنی که اندازه مکانی و نقطه تقارن آن برابر باشند (دست کم با یک تبدیل خطی)، از جذابیت بیشتری برخوردار است، معقول است که در توزیع‌های متقارن، نقطه تقارن با میانه چندبعدی برابر باشد. نوعی از تقارن که این ویژگی را به خوبی نشان می‌دهد، تقارن نیم فضا نامیده می‌شود. بردار تصادفی $\boldsymbol{X}$ دارای تقارن نیم‌فضا حول $\boldsymbol{\theta}$ است، اگر به ازای هر نیم فضای بسته H شامل $\boldsymbol{\theta}$، رابطه $P(\boldsymbol{\theta}\in\boldsymbol{H})\geq\frac{1}{2}$ برقرار باشد. به عبارت دیگر هر ابر صفحه گذرنده از $\boldsymbol{\theta}$ باید R^d را به دو نیم فضای بسته که هر کدام دارای حداقل احتمال $\frac{1}{2}$ هستند، تقسیم کند.

۳ اندازه‌های عدم تقارن

وقتی تقارن به عنوان یک ویژگی توزیع در نظر گرفته می‌شود، چولگی یا عدم تقارن را می‌توان نوع و جهت انحراف از تقارن در نظر گرفت. به طور کلی اندازه چولگی برای توزیع‌های متقارن به صفر کاهش می‌یابد و علامت و اندازه چولگی به ترتیب جهت و مقدار عدم تقارن را نشان می‌دهند. اغلب معیارهایی که برای

---
[۸] Central Symmetry
[۹] Angular Symmetry

سنجش میزان چولگی یک توزیع احتمالی به کار می‌روند، بر اساس گشتاورهای توزیع تعریف می‌شوند. در ساده‌ترین حالت فرض کنید $\bar{X}$ میانگین، m میانه، M مد، S انحراف استاندارد و μ_3 گشتاور مرکزی سوم توزیع باشند. در این صورت کمیت‌های $(\bar{X}-M)/S$ و $3(\bar{X}-\mu)/S$ و μ_3/S^3 که به ترتیب ضرایب چولگی اول و دوم پیرسون و ضریب چولگی گشتاوری نامیده می‌شوند، کمیت‌های نمونه‌ای ساده و مفیدی برای سنجش میزان عدم تقارن یک توزیع هستند. در روابط بالا انحراف معیار S در جهت رفع وابستگی این ضرایب به واحد اندازه‌گیری به کار گرفته شده است. در صورتی که داده‌ها نسبت به میانگین متقارن باشند، تمام ضرایب بالا برابر صفر هستند. همانطور که ملاحظه می‌شود مقایسه میانگین و میانه یک توزیع، روش مناسبی برای ارزیابی تقارن آن توزیع است. وجود فاصله بین میانگین و میانه یک توزیع حاکی از چولگی آن توزیع و جهت اختلاف بین این دو کمیت، جهت چولگی را نشان می‌دهد. میانگین، میانه و مد یک توزیع متقارن بر هم منطبق هستند.

برای گسترش معیارهای سنجش میزان چولگی از حالت یک متغیره به توزیع‌های چندمتغیره، باید از بردارها برای تعیین مسیر و اندازه چولگی استفاده کرد. به این ترتیب اندازه چولگی یک توزیع چندمتغیره متقارن با بردار صفر نشان داده می‌شود. برای اندازه‌گیری چولگی توزیع‌های چند متغیره معیارهای مختلفی پیشنهاد شده است. ماردیا [۹] معیار $E\{[(\boldsymbol{X}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\mu})]^3\}$ را برای اندازه‌گیری میزان چولگی یک توزیع پیشنهاد کرد، که در آن $\boldsymbol{X}$ و $\boldsymbol{Y}$ مستقل و هم توزیع با میانگین $\boldsymbol{\mu}$ و ماتریس کوواریانس $\boldsymbol{\Sigma}$ هستند. مالکوویچ و افیفی [۸] معیار چولگی چند متغیره را به صورت

$$\sup_{\boldsymbol{u}\in S^{d-1}} \frac{[E(\boldsymbol{u}'\boldsymbol{X}-\boldsymbol{u}'\boldsymbol{\mu})^3]^2}{[Var(\boldsymbol{u}'\boldsymbol{X})]^3}$$

معرفی کردند. آیزوگای [۵] معیار $(\boldsymbol{\mu}-\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}-\boldsymbol{\theta})$ را پیشنهاد کرد، که در آن $\boldsymbol{\theta}$ مد توزیع است. اوجا[۱۰] معیار

$$\frac{E[\Delta(X_1,\cdots,X_{d-1},\mu_1,\mu_2)]}{(E[\Delta(X_1,\cdots,X_d,\mu_2)^2])^{\frac{1}{2}}}$$

را پیشنهاد کرد، که در آن $\Delta(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_{d+1})$، نشان دهندهٔ حجم در $\mathcal{R}^d$ است و به وسیله $d+1$ نقطه $x_1,\cdots,x_{d+1}$ و اندازه مکانی μ_α تعیین می‌شود و $E[\Delta(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_d,\boldsymbol{\lambda})^\alpha]$ را نسبت به $\boldsymbol{\lambda}$ مینیمم می‌کند. اَورِوس و مِست[۲] تابع

$$s_F(r)=\frac{\ell_F(r)-M_F}{\frac{1}{f(M_F)}}\quad,\quad 0<r<1$$

را در نظر گرفتند، که در آن f چگالی توزیع F، M_F میانه فضایی [۱۰] مینیمم کنندهٔ $E(\|\boldsymbol{X}-\boldsymbol{c}\|)$ نسبت به c است و $\ell_F(r)$ گلوله میانه [۱۱] تابع مکانی است. گلوله میانه، تعمیم دامنه میان چارکی در حالت یک متغیره است و بصورت

$$[F^{-1}(\frac{1}{2}-\frac{r}{2}),F^{-1}(\frac{1}{2}+\frac{r}{2})]\ ,\ 0<r<1$$

تعریف می‌شود. برای هر $r=r_0$، $s_F(r)$ اندازه و مسیر کلی بردار چولگی را نشان می‌دهد. می‌توان اندازه واقعی چولگی F را در هر مسیر خاص h از میانه M_F به وسیله اندازه بردار $<s_F(r),h>$ بدست آورد، که در

[۱۰] Spatial Median
[۱۱] Median Balls

آن $1 > r > 0$ است. توزیع F چوله ضعیف [۱۲] در مسیر h از M_F نامیده می‌شود، اگر $< s_F(r), h >$ برای هر r، نامنفی باشد و چوله قوی [۱۳] خوانده می‌شود، هرگاه $< s_F(r), h >$ در r صعودی باشد. اگر رابطه

$$< s_G(r) - s_F(r), h > \geq 0$$

برای هر r برقرار باشد، F نسبت به G در مسیر h از M_F چوله ضعیف‌تری است، که بصورت $F \prec hG$ نشان داده می‌شود. فرض کنید $A_F = \sup_{0<r<1} \parallel s_F(r) \parallel$، دراین صورت اگر $A_F \leq A_G$، آنگاه F کمتر از G نامتقارن است. در بخش بعدی برخی روش ها که به کمک آنها می‌توان از توزیع های متقارن توزیع‌های چوله ایجاد کرد، توضیح داده می‌شوند.

۴ ایجاد قابلیت مدل بندی عدم تقارن

توزیع نرمال در آمار از اهمیت خاصی برخوردار است. در بسیاری از تحلیل‌های آماری فرض نرمال بودن توزیع، زیربنای اصلی استنباط است. این در حالی است که جوامع متعددی را می‌توان برشمرد که توزیع آنها برخلاف توزیع نرمال چوله است. از گذشته دور همواره تلاش هایی برای دستیابی به توزیع‌هایی که قابلیت مدلبندی چولگی را داشته باشند، صورت پذیرفته است. بدین منظور سعی‌شده است توزیع‌هایی ساخته شوند، که علاوه بر ویژگی‌های توزیع نرمال قابلیت مدلبندی چولگی را نیز داشته باشند.

برای ایجاد چولگی راههای متفاوتی وجود دارد. آزالینی [۳] یک راه ساده برای ایجاد چولگی در یک توزیع متقارن ارائه نمود. وی از این مطلب استفاده کرد که اگر $f_0(y)$ یک تابع چگالی متقارن حول صفر، G یک تابع توزیع مطلقاً پیوسته که چگالی متناظر آن یعنی G'، حول صفر متقارن است و w یک تابع فرد باشد. در این صورت

$$f(y) = 2 f_0(y) G(w(y))$$

یک تابع چگالی است. با استفاده از این روش به سادگی می‌توان با افزودن پارامترهایی، قابلیت مدل بندی چولگی را در بسیاری از توزیع‌های متقارن بوجود آورد.
یک راه دیگر برای ایجاد چولگی در یک توزیع متقارن، به شکافتن پارامتر مقیاس [۱۴] معروف است [۷]. ایده اصلی در این روش به کار بردن پارامتر مقیاس متفاوت در قسمت مثبت و منفی یک چگالی متقارن است. هر چند در توزیع جدید نیمی از جرم احتمال در سمت چپ و نیم دیگر در سمت راست محور yها قرار می‌گیرد، اما اگر شکافتن پارامتر مقیاس به طور مناسب انجام شود، این ایراد از بین می‌رود. فرض کنید X یک متغیر تصادفی متقارن با چگالی f باشد. در این روش چگالی جدید به صورت

$$\begin{aligned} f_a(x;\nu) \quad \equiv \quad & a(\nu)\frac{1}{\nu} f(\frac{x}{\nu}) I_{(-\infty,0)}(x) + \\ & (2 - a(\nu))\nu f(x\nu) I_{(0,\infty)}(x) \end{aligned}$$

تعریف می شود، که در آن به ازای $\nu > 0$، $0 \leq a(\nu) \leq 2$ و $a(1) = 1$. وجه تسمیه این

---
[۱۲] *Weakly Skew*
[۱۳] *Strongly Skew*
[۱۴] Splitting Scale Parameter

روش آن است که برای دو قسمت مثبت و منفی توزیع دو پارامتر مقیاس متفاوت مورد استفاده قرار می‌گیرد. این چگالی به ازای $\nu = ۱$ با چگالی اولیه معادل است.

۱.۴ برخی توزیع‌های چوله – متقارن

گوپتا و سایرین [۵] نشان دادند که چگونه می‌توان با استفاده از روش آزالینی [۳] و با افزودن یک پارامتر به بسیاری از توزیعهای متقارن و شناخته شده در نظریهٔ احتمال، قابلیت مدل بندی چولگی را در این توزیعها بوجود آورد. در تمام موارد زیر تابع فرد $w(x)$ برابر λx در نظر گرفته شده است.

توزیع چوله یکنواخت. متغیر تصادفی X دارای توزیع چوله یکنواخت[۱۵] دربازه $(-\theta,\theta)$ است، اگر تابع چگالی آن به شکل

$$\begin{aligned} F_X(x) &= \frac{۱}{\theta^۲}[\max(\min(\lambda x,\theta),-\theta)+\theta] \\ & -\theta < x < \theta \ , \ \theta > ۰ \end{aligned}$$

باشد، که در آنλ عددی حقیقی است. تابع چگالی احتمال و تابع توزیع تجمعی توزیع یکنواخت که در روش آزالینی مورد استفاده هستند را می‌توان به صورت

$$\begin{aligned} f(x) &= \frac{۱}{۲\theta} \\ F(x) &= \frac{[\max(\min(x,\theta),-\theta)+\theta]}{۲\theta} \end{aligned}$$

نوشت. تابع مولد گشتاورهای متغیر تصادفی X عبارت است از

$$M(t) = (\frac{\lambda}{t})\cosh(\theta t) - (\frac{\lambda-\theta t}{\theta t^۲})\sinh(\theta t)$$

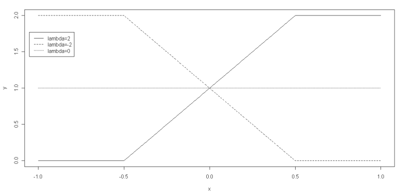
و $Y = X^۲$ دارای توزیعی بصورت

$$f_Y(y) = \frac{۱}{۲\theta\sqrt{y}} \qquad ۰ \le y < \theta^۲$$

است. با استفاده از تابع مولد گشتاورهای این توزیع داریم

$$\begin{aligned} E(X) &= \frac{\lambda\theta}{۳} \\ Var(X) &= (۳-\lambda^۲)\frac{\theta^۲}{۹} \\ \gamma_۱ &= \frac{۲}{۵}\frac{\lambda(۵\lambda^۲-۹)}{(۳-\lambda^۲)^{\frac{۳}{۲}}} \\ \gamma_۲ &= \frac{۳}{۵}\frac{(۹-۵\lambda^۲)(\lambda^۲+۳)}{(۳-\lambda^۲)^۲} \end{aligned}$$

که $\gamma_۱$ و$\gamma_۲$ به ترتیب ضرایب چولگی و کشیدگی این توزیع را نشان می‌دهند. شکل ۱ توزیع چوله یکنواخت را به ازای مقادیر ۲، ۰ و ۲− برای پارامتر چولگی λ نشان می‌دهد.



شکل ۱. توزیع چوله یکنواخت به ازای مقادیر ۲، ۰ و ۲− برای λ.

توزیع چوله نرمال. متغیر تصادفی X دارای توزیع چوله نرمال است، اگر تابع چگالی آن به صورت

$$f_X(x;\lambda,\mu,\sigma^۲) = ۲\phi(x;\mu,\sigma^۲)\Phi(\lambda(\frac{x-\mu}{\sigma}))$$

باشد، که در آن $\phi(\cdot)$ و $\Phi(\cdot)$ به ترتیب تابع چگالی احتمال و تابع توزیع تجمعی توزیع نرمال استاندارد هستند. تابع

[۱۵] Skew Uniform Distribution

مولد گشتاورهای این توزیع بصورت

$$M_X(t) = ۲ \ \exp\{\mu t + \frac{t^۲\sigma^۲}{۲}\}\Phi(\delta t\sigma)$$

است، که در آن $\delta = \frac{\lambda}{\sqrt{۱+\lambda^۲}}$ است. شکل تابع مولد گشتاورها نشان می‌دهد خانواده توزیع های چوله نرمال تحت تبدیلات خطی متغیرها بسته است. شکل استاندارد این توزیع به ازای $\mu = ۰$ و $\sigma = ۱$ به‌دست می آید و با نماد $Z \sim SN(\lambda)$ نشان داده می‌شود.

توزیع چوله نرمالی که با استفاده از این روش ساخته می‌شود، دارای ویژگی های مطلوبی است، که برخی از آنها عبارتند از

۱) $-Z$ دارای چگالی چوله نرمال با پارامتر $-\lambda$ است. زیرا

$$\begin{aligned} f_{-Z}(y) = f_Z(-y) &= ۲\phi(-y)\Phi(\lambda(-y)) \\ &= ۲\phi(y)\Phi((-\lambda)y) \end{aligned}$$

۲) اگر $\lambda = ۰$ باشد، آنگاه توزیع Z نرمال استاندارد است.

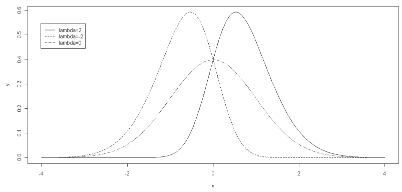
۳) اگر $\lambda \to \infty$، آنگاه توزیع Z به توزیع نیم نرمال[۱۶] میل می‌کند، یعنی

$$\begin{aligned} \lim_{\lambda\to\infty} ۲\phi(z)\Phi(\lambda z) &= ۲\phi(z)\int_{-\infty}^{\infty}\phi(t)\ dt.I(Z \geq ۰) \\ &= \sqrt{\frac{۲}{\pi}}e^{-\frac{z^۲}{۲}}.I(Z \geq ۰) \end{aligned}$$

۴) تابع توزیع چوله نرمال به ازای یک λ ثابت، قویاً تک مدی[۱۷] است.

۵) اگر $Z \sim SN(\lambda)$ و $Y \sim N(۰,۱)$ باشد، آنگاه $|Z|$ و $|Y|$ همتوزیع هستند.

شکل ۲ توزیع چوله نرمال را به ازای مقادیر ۲، ۰ و ۲− برای پارامتر چولگی λ نشان می‌دهد.



شکل ۲. توزیع چوله نرمال به ازای مقادیر ۲، ۰ و ۲− برای λ.

با توجه به ویژگی (۳) این توزیع دارای این نقیصه است، که برای λهای بزرگ احتمال در یک طرف محور yها تقریباً برابر صفر است. کلاس جدیدی از این توزیع‌ها توسط ارلانوواله و همکاران[۱] بصورت

$$\begin{gathered} f(x|\lambda_۱,\lambda_۲) = ۲\phi(x)\Phi(\frac{\lambda_۱ x}{\sqrt{۱+\lambda_۲ x^۲}}) \\ x,\lambda_۱ \in \mathcal{R} \ , \ \ \lambda_۲ \geq ۰ \end{gathered}$$

معرفی شد، که با دارا بودن دو پارامتر برای مدلبندی چولگی، از انعطاف پذیری بیشتری برخوردار است. این توزیع با نماد $SGN(\lambda_۱,\lambda_۲)$ نشان داده می‌شود و برخی از ویژگی های مهم این توزیع عبارتند از

۱) به ازای هر $x \in R$ و $\lambda_۲ \geq ۰$، $SGN(\lambda_۱ = ۰, \lambda_۲) = N(۰,۱)$.

۲) به ازای هر $x \in R$، $SGN(\lambda_۱, \lambda_۲ = ۰) = SN(\lambda_۱)$.

۳) اگر $X \sim SGN(\lambda_۱,\lambda_۲)$، آنگاه $-X \sim SGN(-\lambda_۱,\lambda_۲)$.

---

[۱۶] Half Normal Distribution

[۱۷] Strongly Unimodal

۴) به ازای هر $x \in R$،

$$f(x|-\lambda_1,\lambda_2)+f(x|\lambda_1,\lambda_2)=2\phi(x)$$

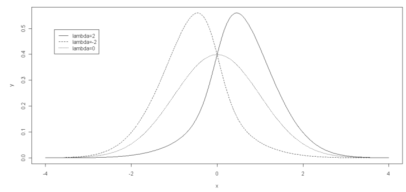
۵) به ازای هر $x \in R$،

$$\lim_{\lambda_1\to\pm\infty} f(x|\lambda_1,\lambda_1^2)=2\phi(x)\Phi(\pm sgn(x))$$

٦) به ازای هر λ_2 رابطه زیر برقرار است

$$\begin{aligned}\lim_{\lambda_1\to\infty} f(x|\lambda_1,\lambda_2) &= 2\phi(x)I_{\{x\geq 0\}}\\ \lim_{\lambda_1\to-\infty} f(x|\lambda_1,\lambda_2) &= 2\phi(x)I_{\{x\leq 0\}}\end{aligned}$$

شکل ۳ توزیع SGN را به ازای مقادیر ۲، ۰ و ۲− برای پارامترهای چولگی λ_1 و λ_2 نشان می‌دهد.



شکل ۳. توزیع SGN. به ازای مقادیر مختلف λ_1 و λ_2 .

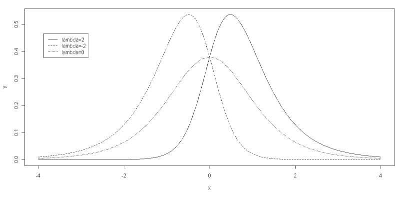
توزیع چوله t. چنانچه W یک متغیر تصادفی با توزیع چوله نرمال و V یک متغیر تصادفی کای دو با ν درجه آزادی و مستقل از W باشند، آنگاه توزیع $X=\frac{W}{\sqrt{V/\nu}}$ چوله t با ν درجه آزادی است و تابع چگالی آن به صورت

$$\begin{aligned}f_x(x) &= 2f(x)F(\lambda x)\\ &= 2\frac{\Gamma(\frac{(\nu+1)}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})}(1+\frac{x^2}{\nu})^{-\frac{\nu+1}{2}}\\ &\times\ F_{T_{\nu+1}}\left(\lambda x\sqrt{\frac{1+\nu}{\nu(\nu+x^2)}}\right)\end{aligned}$$

است. در این توزیع

$$\begin{aligned}E(X) &= b\delta\sqrt{\frac{\nu}{2}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\\ Var(X) &= \frac{\nu}{\nu-2}-\frac{\nu}{\pi}\frac{\lambda^2}{1+\lambda^2}[\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}]\\ E(X^3) &= b\delta(3-\delta^2)(\frac{3}{2})^{\frac{3}{2}}\frac{\Gamma(\frac{\nu-3}{2})}{\gamma(\frac{\nu}{2})}\\ E(X^4) &= \frac{3\nu^2}{(\nu-2)(\nu-4)}\\ \gamma_1 &= [E(X^3)-3E(X^2)E(X)+\\ &\quad 2E^3(X)]/[Var(X)]^{\frac{3}{2}}\\ \gamma_2 &= [E(X^4)-4E(X^3)E(X)+\\ &\quad 6E(X^2)E^2(X)-3E^4(X)]\\ &\quad /[Var(X)]^2\end{aligned}$$

که در آن $b=\sqrt{\frac{2}{\pi}}$ و $\delta=\frac{\lambda}{\sqrt{1+\lambda^2}}$. γ_1 و γ_2 به ترتیب ضرایب چولگی و کشیدگی این توزیع را نشان می‌دهند. اگر در توزیع بالا $\lambda=0$ باشد، این توزیع به توزیع t با ν درجه آزادی و اگر $\nu\to\infty$ به توزیع چوله نرمال تبدیل می‌شود و X^2 نیز دارای توزیع $F(1,\nu)$ است. شکل ۴ توزیع چوله t را به ازای مقادیر ۲، ۰ و ۲− برای پارامتر چولگی λ نشان می‌دهد.



شکل ۴. توزیع چوله t. به ازای مقادیر ۲، ۰ و ۲− برای λ.

توزیع چوله کوشی. متغیر تصادفی X دارای توزیع چوله کوشی است، اگر تابع چگالی آن به صورت

$$f_X(x) = \frac{\sigma[1 + \frac{2\arctan(\frac{\lambda x}{\sigma})}{\pi}]}{\pi(\sigma^2 + x^2)} \ , \ x \in \mathcal{R}$$

باشد، که در آن

$$\begin{aligned} f(x) &= \frac{\sigma}{\pi(\sigma^2 + x^2)} \\ F(x) &= \frac{1}{2} + \frac{\arctan(\frac{x}{\sigma})}{\pi} \end{aligned}$$

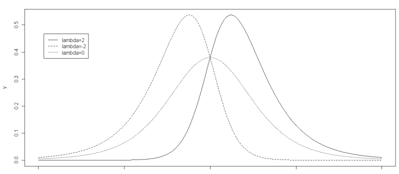
به ترتیب تابع چگالی و تابع توزیع تجمعی توزیع کوشی هستند. تابع مولد گشتاورهای این توزیع موجود نیست و تابع مشخصه آن شکل بسته‌ای ندارد.

روش دوم تعریف توزیع چوله کوشی قرار دادن $\nu = 1$ در توزیع چوله t است. بدین ترتیب چگالی آن به شکل زیر بدست خواهد آمد

$$\begin{aligned} f_X(x) &= \frac{1}{\pi(1 + x^2)} \\ &\times \ [1 + \frac{\lambda x}{\sqrt{1 + (1 + \lambda^2)x^2}}], x \in R \end{aligned}$$

ویژگیهای این چگالی مشابه چگالی اول است.

سومین تعریف چوله کوشی نیز به این قرار است که اگر متغیرهای تصادفی و مستقل Z و Y دارای چگالی $SN(\lambda)$ باشند، آنگاه $X = \frac{Y}{Z}$ نیز دارای توزیع چوله کوشی است. اگرچه این چگالی دارای شکل ساده‌ای نمی‌باشد، با این حال در ویژگی‌های چوله کوشی بالا سهیم است. شکل ۵ توزیع چوله کوشی را به ازای مقادیر ۲، ۰ و ۲− برای پارامتر چولگی λ نشان می‌دهد.



شکل ۵. توزیع چوله کوشی به ازای مقادیر ۲، ۰ و ۲− برای λ.

توزیع چوله لاپلاس. متغیر تصادفی X دارای توزیع چوله لاپلاس است، اگر چگالی آن به صورت

$$\begin{aligned} f_X(x) &= 2f(x)F(\lambda x) \\ &= \frac{e^{-\frac{|x|}{\sigma}}[1 + sgn(\lambda x)(1 - e^{-\frac{|\lambda x|}{\sigma}})]}{2\sigma} \end{aligned}$$

باشد، که در آن

$$\begin{aligned} f(x) &= \frac{e^{\frac{-|x|}{\sigma}}}{2\sigma} \\ F(x) &= \frac{1}{2}[1 + sign(x)(1 - e^{-\frac{|x|}{\sigma}})] \end{aligned}$$

به ترتیب تابع چگالی و تابع توزیع توزیع لاپلاس هستند. تابع چگالی $Y = X^2$ به صورت

$$f_Y(y) = \frac{e^{-\frac{\sqrt{y}}{\sigma}}}{2\sigma\sqrt{y}}$$

است. برخی ویژگی های این توزیع عبارتند از

$$\begin{aligned} E(X) &= \sigma[1 - \frac{1}{(\lambda + 1)^2}] \\ Var(X) &= \sigma^2[2 - \frac{\lambda^2(\lambda + 2)^2}{(\lambda + 1)^4}] \\ \gamma_1 &= [2\lambda(\lambda + 2)(\lambda^2 + \lambda + 1) \\ &\quad (\lambda^2 + 3\lambda + 3)]/[2(\lambda + 1)^4 - \\ &\quad \lambda^2(\lambda + 2)^2]^{\frac{3}{2}} \\ \gamma_2 &= [3(3\lambda^8 + 24\lambda^7 + 88\lambda^6 + \\ &\quad 192\lambda^5 + 276\lambda^4 + 272\lambda^3 + \\ &\quad 176\lambda^2 + 64\lambda + 8)]/[(\lambda^4 + \\ &\quad 4\lambda^3 + 8\lambda^2 + 8\lambda + 2)^2 \end{aligned}$$

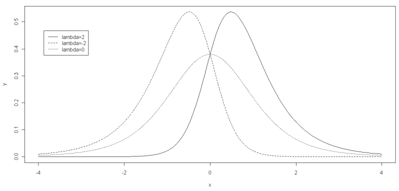
که γ_1 وγ_2 به ترتیب ضرایب چولگی و کشیدگی این توزیع را نشان می‌دهند. تابع مشخصه این توزیع به صورت

$$\psi(t)=\frac{\sigma t+(\lambda+1)^2 i}{(\sigma t+i)((\sigma t)^2+(\lambda+1)^2)}$$

است. به علاوه روابط

$$\begin{aligned}\lim_{\lambda\to\pm\infty}\gamma_2 &= 9\\ \lim_{\lambda\to\pm-1\pm\sqrt{\sqrt{2}-1}}\gamma_1 &= -\infty\\ \lim_{\lambda\to\pm\infty}\gamma_1 &= 2\end{aligned}$$

نیز برقرارند.



شکل ۶. توزیع چوله لاپلاس به ازای مقادیر ۲، ۰ و ۲− برای λ.

شکل ۶ توزیع چوله لاپلاس را به ازای مقادیر ۲، ۰ و ۲− برای پارامتر چولگی λ نشان می‌دهد.

توزیع چوله لوژستیک. متغیر تصادفی X دارای توزیع چوله لوژستیک است، اگر تابع چگالی آن به صورت

$$\begin{aligned}f_X(x) &= 2f(x)F(\lambda x)\\ &= \frac{2e^{-\frac{x}{\sigma}}}{\sigma(1+e^{-\frac{x}{\sigma}})^2(1+e^{-\frac{\lambda x}{\sigma}})},x\in R,\end{aligned}$$

باشد، که در آن

$$\begin{aligned}f(x) &= \frac{e^{-\frac{x}{\sigma}}}{\sigma(1+e^{-\frac{x}{\sigma}})}\\ F(x) &= \frac{1}{1+e^{-\frac{x}{\sigma}}},\end{aligned}$$

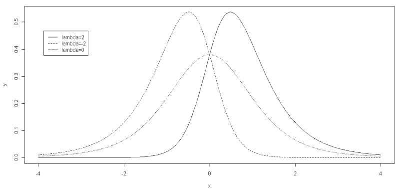
به ترتیب تابع چگالی و تابع توزیع تجمعی توزیع لوژستیک هستند. تابع چگالی $Y=X^2$ بصورت

$$F_Y(y)=\frac{e^{-\frac{\sqrt{y}}{\sigma}}}{[\sigma(1+e^{-\frac{\sqrt{y}}{\sigma}})^2\sqrt{y}]},y\geq 0,$$

است. تابع مولد گشتاور و تابع مشخصه این چگالی شکل بسته‌ای ندارد و چهار گشتاور اول توزیع عبارتند از

$$\begin{aligned}E(X) &= 2\sigma A_1\ \ ,\ \ E(X^2)=\frac{1}{3}(\pi\sigma)^2\\ E(X^3) &= 2\sigma^3 A_2\ \ ,\ \ E(X^4)=\frac{7}{15}(\pi\sigma)^4,\end{aligned}$$

که در آن $A_j=\int_0^\infty\frac{(\ln z)^j}{[(1+z)^2(1+z^\lambda)]}\,dz, j=1,2$.



شکل ۷. توزیع چوله لوژستیک به ازای مقادیر ۲، ۰ و ۲− برای λ.

شکل ۷ توزیع چوله لوژستیک را به ازای مقادیر ۲، ۰ و ۲− برای پارامتر چولگی λ نشان می‌دهد.

مراجع

[1] Arellano-Valle, R. B. and Gomez, H. W. and Quintana, F. A. (2003), A New Class of Skew-Normal Distributions. Communication in Statistics-Theory and Methods, **33**. 1975-1990.

[2] Averous, J. and Meste, M. (1997), Skewness for Multivariate Distribution: Two Approaches. Ann. Stat.,**25**, 1984-1997.

[3] Azzalini, A. (1985), A Class of Distributions which Includes the Normal Ones. Scand. J. Statist. **12**, 171-178.

[4] Beran, R. J. and Miller, P. W. (1997), Multivariate Symmetry Models. In Festschrift for Lucien le Cam-Reserch Papers in Probability and Statistics, D. Pollard, E. Torgersen and G. L. Yang, eds. Springer, New York, 13-42.

[5] Gupta, A. K. ,Chang, F. C. ,Hung, W. J. (2002), Some Skew-Symmetric Models. Random Oper. And Stoch. Equ, Vol **10**, No 2, 133-140.

[6] Isogai, T. (1982), On a Measure of Multivariate Skewness and a Test for Multivariate Normality. Ann. Inst. Stat. Math., **34**, 531-541.

[7] Klein, I. and Fischer, M. (2006). Skewness by Splitting the Scale Parameter. Communications in Statistics-Theory and Methods, **35**, 7, 1159-1171.

[8] Liu, R. Y. (1990) On a Notion of Data Depth Based on Random Simplices. Ann. Stat, **18**, 405-414.

[9] Malkovich, L. F. and Afifi, A. A. (1973), On Tests for Multivariate Normality. J. Am. Stat. Assoc., **68**, 176-179.

[10] Mardia, K. V.(1970), Measures of Multivariate Skewness and Kurtosis with Applications. Biometrika, **57**, 519-530.

[11] Oja, H. (1983), Descriptive Statistics for Multivariate Distribution. Stat. Probab. Lett, **1**, 327-333.

[12] Serflig, R. J. (2006), Multivariate Symmetry and Asymmetry. In the Encyclopedia of Statistics Sciences, Second Edition, Wiley, Vol. 8, 5338-5345.