بسم الله الرحمن الرحيم

دانشکده ریاضی و رایانه

بخش ریاضی

رساله برای دریافت درجه دکتری

رشته ریاضی گرایش کاربردی

متغیر همراه در نظریه آماره‌های ترتیبی

مؤلف :

ایوب شیخی

استاد راهنما :

دکتر ماه بانو تاتا

شهریور ماه1391

چکیده

توزیع آماره های ترتیبی و متغیر های همراه آنها از اهمیت ویژه ای در مباحث آماری برخوردار است. در این رساله برخی از توزیع های مربوط به آماره های ترتیبی و متغیر های همراه آنها در یک نمونه تصادفی نرمال مورد بررسی قرار می گیرد. علاوه بر این توزیع یک ترکیب خطی از آماره های ترتیبی، توزیع یک ترکیب خطی از متغیرهای همراه آماره های ترتیبی ، توزیع توام یک ترکیب خطی از آماره های ترتیبی و یک ترکیب خطی از متغیرهای همراه و توزیع توام یک ترکیب خطی از متغیرهای همراه با یک ترکیب خطی از مشاهدات مرتب شده متغیرهای همراه نیز مورد مطالعه قرار می گیرند. نشان داده می شود که توزیع همه این متغیرها، متعلق به خانواده توزیع های چوله نرمال است.

توزیع های شرطی و حاشیه ای و توزیع توام این متغیرها در حضور یک یا چند متغیر میانجی را بدست آورده و یک مطالعه رگرسیونی بر اساس متغیرهای همراه و میانجی انجام می شود. استقلال میانگین و واریانس یک نمونه تصادفی با استفاده از آماره های ترتیبی، اثبات و با استفاده از رابطه توزیع های چوله نرمال و آماره های ترتیبی، توزیع چوله نرمال دو تکه ساخته می شود.

تمام محاسبات و رسم نمودار ها با استفاده از نرم افزار R انجام شده و دستورات آنها در ضمیمه ارائه شده است. اثبات های جدید در این رساله با علامت (*) مشخص شده اند.

تقدیر و تشکر

ستایش خدای راست که پیوند دهندهٔ ستایش به نعمت است و نعمت به ستایش. خداوند بلند مرتبه را از عمق وجود سپاس گذارم که توانایی طی این مسیر را برایم مهیا نمود.

وجود نازنین پدر و مادرم را غرق بوسه می‌کنم زیرا که معتقدم هر چه دارم نبوده است جز به دعای خیر ایشان. آنان که در تمامی لحظات با دعای سرشار از عطوفت خود راه را برایم هموار نمودند.

از استاد فرهیخته و فرزانه سرکارخانم دکتر تاتا که همواره راهنما و راه گشای حقیر در اتمام و اکمال رساله بوده است، تقدیر و تشکر شایسته به عمل می آورم.

و اگر نبود فداکاریهای همسری مهربان، چنین سرانجامی می‌بود؟ در این مجال یارای تشکر از زحمات طاقت فرسای وی نیست پس بهترین آرزوهایم را نثار وجودش می‌کنم.

فهرست مطالب

۱ مقدمه

۱.۱ آماره های ترتیبی و متغیرهای همراه

نظریه آماره های ترتیبی[۱] نقش مهمی در مباحث آمار نظری و کاربردی ایفا می کند. این مبحث از آمار با متغیرها و داده های مرتب شده سر و کار دارد و در مباحث آمار پارامتری و آمار ناپارامتری بسیار مورد توجه قرار می گیرد. توزیع آماره های ترتیبی در بسیاری از آزمون فرضها و فواصل اطمینان استفاده و توزیع یک ترکیب خطی از آماره های ترتیبی نیز توسط بسیاری از محققان بررسی شده است.

اگر $X_1,\ X_2,...,\ X_n$ یک نمونه تصادفی از تابع چگالی احتمال $f(x)$ با تابع تجمعی $F(x)$ باشد و بزرگترین مقدار نمونه را با $X_{n:n}$ و کوچکترین مقدار نمونه را با $X_{1:n}$ نمایش دهیم، آنگاه تابع چگالی احتمال آنها به ترتیب برابر است با

$$f_{X_{n:n}}(x) = nF^{n-1}(x)f(x) \quad x \in R$$

و

$$f_{X_{1:n}}(x) = n[1-F(x)]^{n-1}f(x) \quad x \in R.$$

به طور کلی تابع چگالی احتمال r امین آماره ترتیبی $X_{r:n}$ به صورت زیر است.

$$f_{X_{r:n}}(y) = \begin{pmatrix} n \\ r-1,\ n-r,\ 1 \end{pmatrix} F^{r-1}(x)[1-F(x)]^{n-r}f(x) \qquad x \in R.$$

برای $x_1 \leq x_2$، تابع چگالی احتمال توام آماره های ترتیبی rام و sام عبارتست از

[۱] Order Statistics

$$f_{X_{r:n},X_{s:n}}(x_1,\ x_2) = \begin{pmatrix} n \\ r-1, s-t-1,\ n-s,\ 1,\ 1 \end{pmatrix} F^{r-1}(x_1)$$
$$\times[F(x_2)-F(x_1)]^{s-t-1}[1-F(x_2)]^{n-s}f(x_1)f(x_2).$$

تابع چگالی احتمال توام آماره های های ترتیبی $X_{1:n},\ X_{2:n}, ..., X_{n:n}$ به صورت زیر است.

$$f_{X_{1:n},\ X_{2:n},...,X_{n:n}}(x_1,\ x_2, ..., x_n) = n! \prod_{i=1}^{n} f(x_i) \quad x_1 \le\ x_2 \le ... \le x_n.$$

کتابهای زیاد در این زمینه از جمله آرنولد و بالاکریشنان (۱۹۹۲) و دیوید و ناگاراجا (۲۰۰۳) به رشته تحریر در آمده است.

آماره های دیگری مرتبط با آماره های ترتیبی که در نمونه های زوجی نمود پیدا می کند، متغیر همراه[۲] آماره های ترتیبی نام دارد. فرض می کنیم برای $i = 1,\ 2, ..., n$ ، (X_i, Y_i) یک نمونه تصادفی از توزیع (X, Y) و مشاهدات $X_{1:n},\ X_{2:n}, ...,\ X_{n:n}$ آماره های ترتیبی مشاهدات نمونه X باشند. متغیر Y متناظر با آماره ترتیبی $X_{i:n}$ را متغیر همراه $X_{i:n}$ نامیده و آن را با نماد $Y_{[i:n]}$ نمایش می دهیم (دیوید ۱۹۷۳).

دیوید (۱۹۸۱) تابع چگالی احتمال متغیر همراه r امین آماره های ترتیبی را بر حسب تابع چگالی احتمال آماره های ترتیبی به صورت زیر بدست آورد.

$$f_{Y_{[r:n]}}(y) = \int f\ (y|x)\ f_{X_{r:n}}(x)dx,$$

که در آن، $f\ (y|x)$ تابع چگالی احتمال شرطی Y به شرط X است. تابع چگالی احتمال توام r امین آماره ترتیبی و متغیر همراه متناظر با آن به صورت زیر بدست می آید.

$$f_{X_{r:n},Y_{[r:n]}}(x, y) = f(y|x)f_{X_{r:n}}(x).$$

[۲] Concomitant Variable

در ضمن تابع چگالی احتمال توام rامین و sامین متغیرهای همراه و تابع چگالی احتمال توام k متغیر همراه به ترتیب به صورت زیر ارائه می شوند (دیوید و ناگاراجا ۲۰۰۳).

$$f_{Y_{[r:n,]}Y_{[s:n[}}(y_1,\ y_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{x_2} f\left(y_1|x_1\right) f\left(y_2|x_2\right) f_{X_{r:n},X_{s:n}}(x_1,\ x_2)dx_1dx_2$$

و

$$\begin{aligned} f_{Y_{[r_1:n]},Y_{[r_1:n,]}...,Y_{[r_k:n]}}(y_1,y_2...,y_k) &= \int_{-\infty}^{\infty}\int_{-\infty}^{x_k}...\int_{-\infty}^{x_2} f_{X_{r_1:n},...,X_{r_k:n}}(x_1,...,x_k) \\ &\quad \times\prod_{j=1}^{k} f\left(y_j|x_j\right)dx_j. \end{aligned}$$

بیشترین استفاده از نظریه متغیر همراه آماره های ترتیبی، در یک روند انتخاب[۳] است. در زیر دو مثال را بیان می کنیم.

۱- فرض کنید X امتیاز یک فرد در مرحله اول یک آزمون غربالگری و Y نمره مرحله دوم وی باشد. در واقع با مرتب کردن نمرات مرحله اول می توان اطلاعاتی از نمرات مرحله دوم افراد کسب نمود . به عبارت دیگر با داشتن یک مجموعه k عضوی از بهترین نمرات اولیه، می توان بهترین نمرات نهایی این مجموعه k عضوی را پیشگویی کرد.

۲- اگر متغیر X را میزان ضربان قلب یک ورزشکار در هنگام دویدن و متغیر Y را میزان اکسیژن مصرفی وی در نظر بگیریم، اندازه گیری ضربان قلب با یک دستگاه ضربان نگار کوچک قابل اندازه گیری است ولی اندازه گیری میزان اکسیژن مصرفی به راحتی امکان پذیر نیست. اینجاست که با داشتن رابطه بین آماره های ترتیبی و متغیرهای همراه آنها، می توان به سوال اینکه فردی با بیشترین (یا کمترین) ضربان قلب چه میزان اکسیژن مصرف می کند؟ پاسخ داد.

ناگاراجا (۱۹۸۲) تابعی را پیشنهاد کرده است که در آن بهترین مجموعه از متغیرهای همراه، برای استنباط درباره پارامترهای متغیر Y استفاده می شود. گاهی اوقات به دلیل در دسترس نبودن مقادیر متغیر Y برای انتخاب بهترین زیر مجموعه از مشاهدات آن، از مقادیر مرتب شده متغیر X استفاده می شود (یئو و دیوید ۱۹۸۴).

نظریه متغیر همراه آماره های ترتیبی در استنباط پارامترها نیز مورد استفاده قرار می گیرد. بارتون و کاسلی (۱۹۵۸) با

[۳] Selection procedure

فرض توزیع نرمال دو متغیره، برآوردگر پارامتر β را در معادله رگرسیونی $E(Y|x) = \alpha + \beta x$ به صورت زیر بدست آوردند:

$$B = \frac{\bar{Y}'_{[k:n]} - \bar{Y}_{[k:n]}}{\bar{X}'_{k:n} - \bar{X}_{k:n}}$$

که در آن

$\bar{X}'_{k:n} = \frac{1}{k}\sum_{i=1}^{k} X_{n-i+1:n},\ \ \bar{X}_{k:n} = \frac{1}{k}\sum_{i=1}^{k} X_{i:n},$

$\bar{Y}'_{[k:n]} = \frac{1}{k}\sum_{i=1}^{k} Y_{[n-i+1:n]},\ \ \bar{Y}_{[k:n]} = \frac{1}{k}\sum_{i=1}^{k} Y_{[i:n]}.$

سوکیبایاشی (۱۹۶۴) برآوردگر ضریب همبستگی بین دو متغیر X و Y را به صورت زیر پیشنهاد کرده و خواص آن را مورد بررسی قرار داده است:

$$\hat{\rho} = \frac{\bar{Y}'_{[k:n]} - \bar{Y}_{[k:n]}}{\bar{Y}'_{k:n} - \bar{Y}_{k:n}}.$$

وی همچنین برآوردگر ناریب کواریانس بین دو متغیر X و Y را به صورت زیر مشخص کرده است:

$$\hat{\sigma}_{xy} = \frac{(X_{n:n} - X_{1:n})(Y_{[n:n]} - Y_{[1:n]})}{c_n^2}$$

که در آن $c_n^2 = E(\frac{W_x^2}{\sigma_x^2})$ و $W_x = X_{n:n} - X_{1:n}$ می باشد.

کاربرد مهم دیگر متغیر همراه در نظریه نمونه گیری مجموعه های رتبه دار [۴] است. این گونه نمونه گیری برای اولین بار توسط مکنتایر (۱۹۵۲) جهت برآورد پارامترهای یک توزیع با استفاده از آماره های ترتیبی پیشنهاد شده و طریقه اجرای آن به شرح زیر است.

در گام یا حلقه اول، m نمونه تصادفی را، هر یک به حجم m، از یک توزیع نرمال با میانگین μ و واریانس σ^2 می گیریم. فرض کنیم این نمونه ها به صورت زیر باشند:

[۴] Ranked Set Sampling

مشاهدات	نمونه
$X_{11},\ X_{12}, ..., X_{1m}$	۱
$X_{21},\ X_{22}, ..., X_{2m}$	۲
$\vdots \quad \vdots \quad \vdots$	$\vdots$
$X_{m1},\ X_{m2}, ...,\ X_{mm}.$	m

در نمونه تصادفی اول کوچکترین مشاهده ($X_{(1)1}$) ، در نمونه تصادفی دوم، دومین آماره ترتیبی ($X_{(2)1}$) ، در نمونه تصادفی سوم، سومین آماره ترتیبی ($X_{(3)1}$) ، ... و در نمونه تصادفی آخر، m امین آماره ترتیبی (بزرگترین مشاهده) ($X_{(m)1}$) را یادداشت می کنیم. برآوردگر میانگین بر اساس این مجموعه عبارتست از :

$$\bar{X}_{RSS,\,1} = \frac{1}{m}\sum_{i=1}^{m} X_{(i)1}.$$

اگر حلقه یاد شده را r بار تکرار کنیم، برآوردگر میانگین به صورت زیر بدست می آید.

$$\bar{X}_{RSS} = \frac{1}{mr}\sum_{j=1}^{r}\sum_{i=1}^{m} X_{(i)j},$$

که در آن $X_{(i)j}$ آماره ترتیبی iام در تکرار حلقه jام است.

استوک (۱۹۷۷) برای برآورد پارامتر ها در نمونه های زوجی، از متغیرهای همراه آماره های ترتیبی استفاده کرده است. فرض کنید m نمونه از توزیع نرمال دو متغیره $N_2(\mu_x,\ \mu_y,\ \sigma_x^2,\ \sigma_y^2,\ \rho)$ ،به صورت زیر باشند:

$$(X_{11}, Y_{11}),\ (X_{12}, Y_{12}), ..., (X_{1m}, Y_{1m})$$

$$(X_{21}, Y_{21}),\ (X_{22}, Y_{22}), ..., (X_{2m}, Y_{2m})$$

$$\vdots$$

$$(X_{m1}, Y_{m1}),\ (X_{m2}, Y_{m2}), ..., (X_{mm}, Y_{mm})\ .$$

برآوردگر پارامتر میانگین Y بر اساس متغیر همراه آماره های ترتیبی عبارتست از :

$$\bar{Y}_{IRSS} = \frac{1}{mr}\sum_{j=1}^{r}\sum_{i=1}^{m} Y_{[i]j}$$

که در آن، $Y_{[i]j}$ متغیر همراه متناظر با $X_{(i)j}$ است.

بدست آوردن توزیع متغیرهای همراه و خواص توزیعی آنها نیز توسط بسیاری از محققان بررسی شده است و توزیع این متغیرها یا توزیع تقریبی آنها بدست آمده است. سوکیبایاشی (۱۹۹۸) توزیع توام مقدار غایی و متغیر همراه آن را برحسب یک رابطه انتگرالی نوشته و گشتاورهای این توزیع توام را بدست آورده است. جوئل و هال (۱۹۹۴) با بررسی این توزیع، تابعی به شکل $\sum_{i=1}^{n} h(Y_{i:n} - Y_{[i:n]})$ که در آن h یک تابع هموار است را برای بحث درباره خطای عدم تطابق آماره های ترتیبی و متغیرهای همراه آنها استفاده کرد. هی و ناگاراجا (۲۰۰۹) نیز با الهام از تحقیقات سوکیبایاشی (۱۹۹۸) و جوئل و هال (۱۹۹۴)، سعی کرده اند توزیع توام آماره های ترتیبی متغیر همراه و متغیر همراه آماره های ترتیبی X را بدست آورند و به ویژه توزیع توام $(Y_{i:n}, Y_{[j:n]})$ را بر حسب یک سری انتگرال مشخص کرده اند. اریلماز (۲۰۰۵) در نمونه های تصادفی زوجی، توزیع متغیرهای همراه را مطالعه کرده و توزیع توام $(Y_{[1:n]}, Y_{[n:n]})$ را بدست آورده است. ناگاراجا و دیوید (۱۹۹۴) توزیع تقریبی متغیرهای $V_{k,n} = \max(Y_{[n-k+1:n]}, ..., Y_{[n:n]})$ را برای $k = 1, 2, ..., n$ بدست آورده اند. در واقع این متغیرها ماکزیمم k متغیرهای همراهی را که مربوط به بیشترین آماره های ترتیبی X را دارند در نظر می گیرد. جوشی و ناگاراجا (۱۹۹۵) توزیع توام $V_{k,n}$ و $V^*_{k,n}=\max(Y_{[j:n]}, ..., Y_{[n-k:n]})$ را برای $j = 1, 2, ..., n-k-1$ و همچنین توزیع توام $V_{k,n}$ و $\max(V_{k,n}, V^*_{k,n})$ را بدست آورده اند. بعضی کارهای دیگر در زمینه توزیع متغیرهای همراه را می توان در باتاچاریا (۱۹۸۴) دیوید و ناگاراجا (۱۹۹۸ و ۲۰۰۳) و وینا و توماس (۲۰۰۸) مشاهده کرد.

در نظریه ترکیب خطی، از آماره های ترتیبی و ترکیب خطی از متغیرهای همراه نیز تحقیقات بسیار زیادی انجام شده است. ینگ (۱۹۸۱) دو ترکیب خطی زیر از متغیرهای همراه را مورد مطالعه قرار داده است:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n} J(\frac{i}{n})Y_{[i:n]}$$

و

$$L_2 = \frac{1}{n}\sum_{i=1}^{n} J(\frac{i}{n})\eta(Y_{i:n}, Y_{[i:n]})$$

که در آن J یک تابع درجه دوم و η یک تابع حقیقی است. وی نشان داده است که این ترکیبهای خطی دارای توزیع تقریبی نرمال اند و می توانند به ترتیب به عنوان برآوردگرهای نااریب و سازگار برای $E(Y|X=x)$ و $Var(Y|X=x)$ مورد استفاده قرار گیرند.

دیوید (۱۹۹۴) توزیع حدی $Y_{[i:n]}$ را در یک توزیع دومتغیره دلخواه $F(x,y)$ بدست آورده است. ناگاراجا و دیوید (۱۹۹۴) توزیع حدی $V_{k,n}$ را حدی مورد بررسی قرار داده اند. سورش (۱۹۹۳) نشان داده است که متغیرهای همراه غایی و مرکزی به صورت مجانبی از هم مستقلند. اخیرا وانگ و ناگاراجا (۲۰۰۹) توزیع حدی متغیر همراه را بدست آورده اند.

برای یک متغیر میانجی Z، الکین و ویانا (۱۹۹۵) با فرض آنکه بردار تصادفی $(Z, X_1, X_2)^T$ دارای توزیع نرمال سه متغیره جابجا پذیر است، ساختار ماتریس کواریانس $(Z, X_{1:2}, X_{2:2})^T$ را بدست آورده اند. ویانا و لی (۲۰۰۶) نیز ساختار ماتریس کواریانس دو بردار تصادفی $\mathbf{X}_{(n)}$ و $\mathbf{Y}_{[n]}$ را در حضور یک متغیر میانجی بدست آورده اند که در آن $\mathbf{X}_{(n)} = (X_{1:n},\ X_{1:n}, ...,\ X_{n:n})^T$ بردار آماره های ترتیبی و $\mathbf{Y}_{[n]} = (Y_{[1:n]},\ Y_{[2:n]}, ..., Y_{[n:n]})^T$ بردار متغیرهای همراه آن است. آنها همچنین چند رابطه رگرسیونی را بین آماره های ترتیبی، متغیرهای همراه و متغیر میانجی بدست آورده اند. لوپرفیدو (۲۰۰۸b) با در نظر گرفتن فرضیات الکین و ویانا (۱۹۹۵)، توزیع توام $(Z, X_{2:2})^T$ را بدست آورده است. شیخی و جمالیزاده (۲۰۱۱) نیز با همین فرض ها توزیع توام دو ترکیب خطی از آماره های ترتیبی و متغیر میانجی یعنی توزیع بردار تصادفی $\begin{pmatrix} a_1X_{1:n} + a_2X_{2:n} \\ b_1X_{1:n} + b_2X_{2:n} \\ Z \end{pmatrix}$ را بدست آورده و نشان داده اند که توزیع توام این بردار متعلق به خانوده توزیع های چوله نرمال است. ناگاراجا (۱۹۸۲) توزیع یک ترکیب خطی از آماره های ترتیبی را در توزیع نرمال دو متغیره بدست آورده و لوپرفیدو (۲۰۰۸a) نتایج ناگاراجا (۱۹۸۲) را به توزیع های بیضوی بسط داده است. آرلانو

والی و جنتون (۲۰۰۷) توزیع ترکیب خطی از آماره های ترتیبی و جمالیزاده و بالاکریشنان (۲۰۱۱) توزیع ترکیب خطی از متغیرهای همراه را برای توزیع های بیضوی بدست آورده اند.

در این رساله توزیع متغیرهای همراه، ترکیب خطی آنها و توزیع توام ترکیبهای خطی آماره های ترتیبی و ترکیبهای خطی متغیرهای همراه آنها و سایر توزیع های دیگر در این زمینه مورد بررسی قرار می گیرد. با فرض این که جامعه مورد نظر نرمال چند متغیره است، این توزیع ها را بر حسب توزیع های چوله نرمال بدست می آوریم.

حالتی دیگر از متغیر همراه، زمانی رخ می دهد که در یک نمونه تصادفی، داده ها براساس قدر مطلق مرتب شوند. اکنون فرض کنید برای نمونه تصادفی $Y_1, Y_2, ..., Y_n$ زوج های $(|Y_i|,\ Y_i),\ i = 1, 2, ..., n$ را تشکیل دهیم. آنگاه متغیر Y_i متناظر با آماره ترتیبی r ام یعنی آماره ترتیبی r ام در مجموعه $\{|Y_1|,\ |Y_2|, ...,\ |Y_n|\}$ را r امین آماره ترتیبی قدرمطلق می نامند و با $Y_{[r]}$ نمایش می دهند. به عنوان مثال فرض کنید مشاهدات متغیر Y بصورت $\{-4, -10, 3, 7, 5\}$ و در نتیجه قدر مطلق آنها بصورت $\{4, 10, 3, 7, 5\}$ باشند. حال آماره های ترتیبی قدرمطلق بصورت $\{3, -4, 5, 7, -10\}$ است. روشن است که برای نمونه تصادفی $Y_1,\ Y_2, ..., Y_n$

$$|Y_{[1]}| \leq |Y_{[2]}| \leq ... \leq |Y_{[n]}|.$$

در زمینه ترکیب خطی از آماره های ترتیبی قدر مطلق نیز تحقیقاتی انجام شده است. اگورو و نوزورو (۱۹۷۵) مجموع آماره های ترتیبی قدر مطلق را به صورت $T = \sum_{i=1}^{n} a_i Y_{[i]}$ نشان داده و توزیع حدی آن را بدست آورده اند. نوزوروا و نوزورو (۱۹۹۹) توزیع آماره های ترتیبی قدر مطلق را بدست آورده اند. بخش (۵.۱) از کتاب احسان الله و نوزورو (۲۰۰۱) بطور جامع به این موضوع پرداخته است.

در این رساله پس از معرفی توزیع چوله نرمال دوتکه متقارن، رابطه آن با توزیع نرمال و توزیع آماره های ترتیبی قدر مطلق را بدست می آوریم.

۲.۱ توزیع چوله نرمال

توزیع چوله نرمال[۵] زمانی جای خود را پیدا کرد که آن را آزالینی (۱۹۸۵) برای مدل توزیع داده هایی که توزیع آن ها دقیقا نرمال نیست بکار برد. قبل از این نیز مفاهیم مربوط به توزیع چوله نرمال توسط رابرت (۱۹۶۶) و اینگر و همکاران (۱۹۷۷) و آندل و همکاران (۱۹۸۴) معرفی و مورد بررسی قرار گرفته است.

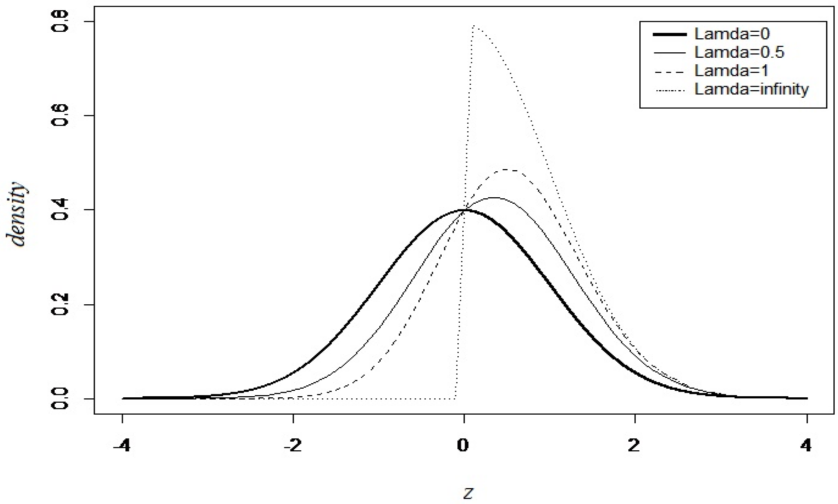
تعریف ۱: متغیر تصادفی Y دارای توزیع چوله نرمال با پارامترهای μ، σ^2 و λ است (با نماد $(Y{\sim}SN(\mu,\sigma^2,\lambda)$) اگر تابع چگالی احتمال آن به صورت زیر باشد:

$$f_Y(y) = 2\varphi\left(y;\ \mu,\ \sigma^2\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right) \qquad y \in R \tag{۱}$$

در اینجا $\varphi\left(.;\ \mu,\ \sigma^2\right)$ تابع چگالی احتمال نرمال با میانگین μ و واریانس σ^2 و $\Phi(.)$ تابع توزیع تجمعی نرمال استاندارد می باشد. در این توزیع، پارامتر λ مربوط به چولگی است. اگر $\lambda = 0$، آنگاه توزیع چوله نرمال به توزیع نرمال با میانگین μ و واریانس σ^2 تبدست می آید. اگر $\mu = 0$ و $\sigma = 1$ باشد می گویند توزیع چوله نرمال استاندارد است. به وضوح می توان دریافت که $Z = \frac{Y-\mu}{\sigma}$ دارای توزیع چوله نرمال استاندارد است و اگر $\lambda > 0$، آنگاه توزیع چوله نرمال چوله به راست و برای مقادیر $\lambda < 0$، چوله به چپ خواهد بود. برای مقدار $\lambda = \infty$، توزیع چوله نرمال استاندارد تبدیل به توزیع نیم نرمال مثبت و برای $\lambda = -\infty$ ، تبدیل به توزیع نیم نرمال منفی می شود.

---

[۵] Skew Normal



شکل ۱. نمودار تابع چگالی احتمال توزیع چوله نرمال استاندارد

شکل (۱) منحنی تابع چگالی احتمال توزیع چوله نرمال استاندارد را برای مقادیر مختلف λ ارائه می دهد. منحنی پر رنگ نشان دهنده تابع چگالی نرمال استاندارد و سایر منحنی ها به ترتیب نمایانگر توزیع های چوله نرمال استاندارد برای مقادیر $\lambda = 0.5$ ، $\lambda = 1$ ، $\lambda = 10$ و $\lambda = \infty$ می باشند.

توزیع چوله نرمال دو متغیره [۶] نیز توسط آزالینی و دالاوالی (۱۹۹۶) معرفی شده و خواص آن مورد بررسی قرار گرفته است. همچنین این توزیع توسط دیگر محققان از جمله گوپتا و چن (۲۰۰۴) و گوپتا و همکاران (۲۰۰۴) معرفی و مطالعه شده است.

تعریف ۲: بردار تصادفی $\mathbf{Y} = (Y_1, Y_2)^T$ دارای توزیع چوله نرمال دو متغیره است ($SN_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$) اگر تابع چگالی احتمال توام آن به صورت زیر باشد،

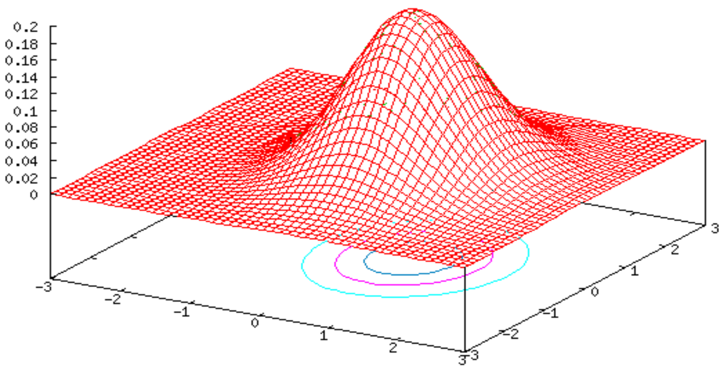
$$f_{\mathbf{Y}}(\mathbf{y}) = 2\varphi_2\left(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \Phi\left(\lambda_1 \frac{y_1 - \mu_1}{\sigma_1} + \lambda_2 \frac{y_2 - \mu_2}{\sigma_2}\right) \qquad \mathbf{y} \in \mathbf{R^2} \tag{۲}$$

[۶] Bivariate skew normal

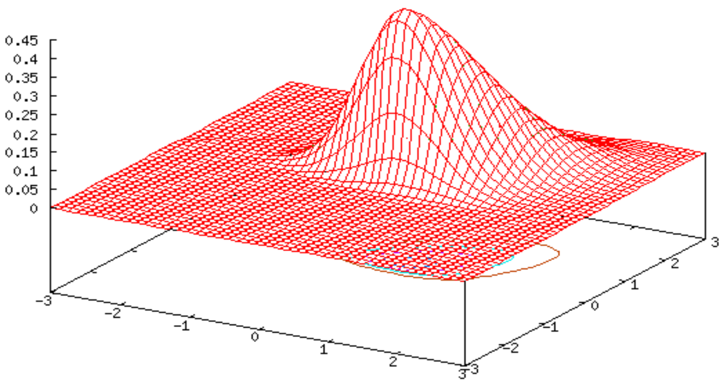
که در آن $\varphi_2\left(.;\boldsymbol{\mu},\Sigma\right)$ تابع چگالی نرمال دو متغیره با بردار میانگین $\boldsymbol{\mu}$ و ماتریس کواریانس Σ است.

در شکل (۲) رویه تابع چگالی احتمال توزیع نرمال دو متغیره $N_2(0,0,1,1,0.75)$ رسم شده است و شکل (۳) رویه تابع چگالی احتمال توزیع چوله نرمال دو متغیره را برای پارامترهای زیر نمایش می دهد.

$$\boldsymbol{\mu}=\begin{pmatrix}0\\0\end{pmatrix},\ \boldsymbol{\Sigma}=\begin{pmatrix}1&0.2\\0.2&1\end{pmatrix},\boldsymbol{\lambda}=\begin{pmatrix}0.5\\1\end{pmatrix}$$



شکل ۲. رویه تابع چگالی احتمال توزیع نرمال دومتغیره



شکل ۳. رویه تابع چگالی احتمال توزیع چوله نرمال دومتغیره

از آنجا که حالت چند متغیره توزیع چوله نرمال اهمیت ویژه ای دارد، محققان زیادی توزیع چوله نرمال چند متغیره [۷] را معرفی و خواص آن را بررسی نموده و برای این توزیع نامهای متفاوتی ارائه کرده اند. گوپتا و همکاران (۲۰۰۴) با معرفی نوعی از توزیع چوله نرمال چند متغیره به نام توزیع چوله نرمال چند متغیره ساختار بسته [۸] اثبات کرده اند که ترکیب خطی از توزیع های چوله نرمال چند متغیره نیز یک توزیع چوله نرمال چند متغیره است و همچنین نشان داده اند که توزیع های شرطی مربوط به توزیع چوله نرمال چند متغیره نیز چوله نرمال چند متغیره خواهد بود. آرلانو والی و جنتون (۲۰۰۵) نوع دیگری از توزیع چوله نرمال چند متغیره به نام توزیع چوله نرمال چند متغیره اساسی [۹] را معرفی کرده اند.

آرلانو والی و آزالینی (۲۰۰۶) با تک ساختار کردن تمام توزیع های چوله نرمال چند متغیره، یک توزیع کلی با نام

[۷] Multvariate skew normal

[۸] Multvariate closed skew normal

[۹] Multvariate fundamental skew normal

توزیع چوله نرمال چند متغیره تک ساختار [10] معرفی کرده و آن را با نماد SUN نشان داده اند. آنها همچنین تابع مولد گشتاور، توابع چگالی احتمال حاشیه ای و توابع چگالی احتمال شرطی را برای این توزیع بدست آورده اند.

فرض کنید

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim\ N_{m+d}\left(\begin{pmatrix} \boldsymbol{\delta} \\ \boldsymbol{\xi} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Omega} \end{pmatrix}\right)$$

که در آن $\boldsymbol{\xi} \in R^d$ ، $\boldsymbol{\delta} \in R^m$ ، $\boldsymbol{\Gamma} \in R^{m\times m}$ و $\boldsymbol{\Omega} \in R^{d\times d}$ ماتریسهای معین مثبت باشند و $\boldsymbol{\Lambda} \in R^{m\times d}$.

تعریف ۳: بردار تصادفی d بعدی $\mathbf{Y}$ را دارای توزیع چوله نرمال چند متغیره تک ساختار می گویند و با نماد $\mathbf{Y} \sim SUN_{d,m}\ (\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega},\ \boldsymbol{\Gamma},\ \boldsymbol{\Lambda})$ نمایش داده می شود اگر

$$\mathbf{Y}\sim\mathbf{V} \mid \mathbf{U} > 0. \tag{۳}$$

به عبارت دیگر تابع چگالی احتمال توام متغیر تصادفی $\mathbf{Y}$ را بتوان به صورت زیر نوشت:

$$f_{\mathbf{Y}}\left(\mathbf{y}\right) = \varphi_d\left(\mathbf{y};\ \boldsymbol{\xi},\ \boldsymbol{\Omega}\right) \frac{\Phi_m\left(\boldsymbol{\delta} + \boldsymbol{\Lambda}^T\boldsymbol{\Omega}^{-1}\left(\mathbf{y}-\boldsymbol{\xi}\right);\ \boldsymbol{\Gamma} - \boldsymbol{\Lambda}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\Lambda}\right)}{\Phi_m\left(\boldsymbol{\delta};\ \boldsymbol{\Gamma}\right)} \quad \mathbf{y} \in R^d \tag{۴}$$

که در آن $\varphi_d(.,\ \boldsymbol{\xi},\ \boldsymbol{\Omega})$ تابع چگالی نرمال چند متغیره با بردار میانگین $\boldsymbol{\xi}$ و ماتریس کواریانس $\boldsymbol{\Omega}$ است و $\Phi_m(.;\ \Sigma)$ تابع توزیع تجمعی نرمال استاندارد m متغیره با بردار میانگین صفر و ماتریس کواریانس می باشد.

در صورتی که ماتریس Σ رتبه کامل نباشد، این توزیع را چوله نرمال منفرد [11] چند متغیره می نامند و با نماد $SSUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega},\ \boldsymbol{\Gamma},\ \boldsymbol{\Lambda})$ نمایش می دهند.

لیزو و لوپرفیدو (۲۰۰۳) کاربرد توزیع چوله نرمال چند متغیره را در آمار بیزی مطالعه نموده و یک مدل چوله نرمال سلسله مراتبی [12] ارائه کرده اند. مفهوم ساخت توزیع چوله نرمال چند متغیره با انتخاب بر حسب یک بازه از زیر بردار

---

[10] Multvariate unified skew normal

[11] Singular unified skew normal

[12] Hierarchical skew normal

تصادفی توسط آرلانو والی و همکاران (۲۰۰۶) پیشنهاد شده است. آن ها نشان داده اند که توزیع شرطی یک بردار به شرط این که بردار دیگری متعلق به یک بازه مشخص باشد نیز یک توزیع چوله نرمال چند متغیره است. شکل های درجه دوم توزیع چوله نرمال چند متغیره توسط جنتون و همکاران (۲۰۰۱) ، هوانگ و چن (۲۰۰۶) و اخیرا توسط وانگ و گوپتا (۲۰۰۹) مورد بررسی قرار گرفته و رابطه توزیع چوله نرمال چند متغیره با توزیع کای دو بررسی شده است.

آزالینی و کاپیتانیو (۱۹۹۹) ضمن مطالعه بعضی از کاربردهای آماری توزیع چوله نرمال چند متغیره، مثالی از کاربرد توزیع چوله نرمال چند متغیره در داده های واقعی ارائه نموده و از خواص این توزیع در اثبات قضیه ککران استفاده کرده اند. چن و همکاران (۱۹۹۹) رابطه بین توزیع چوله نرمال چند متغیره را با داده های دو پاسخی بررسی کرده و در نتیجه دسته جدیدی از توزیع های چوله نرمال ارائه کرده اند. جنتون (۲۰۰۴) مطالبی کلی از نظریه و کاربرد توزیع های چوله نرمال یک و چند متغیره ارائه کرده است. خواننده می تواند برای کسب اطلاعات بیشتر راجع به توزیع چوله نرمال چند متغیره، به گنزالز و همکاران (۲۰۰۳) و گوپتا و همکاران (۲۰۰۴) و مقاله مروری آزالینی (۲۰۰۵) مراجعه کند. اما اخیرا باز مدلهای جدید توزیع چوله نرمال چند متغیره توسط آزالینی (۲۰۱۱) معرفی شده است. آرلانو والی و آزالینی (۲۰۰۶) سه حالت ممکن را برای توزیع چوله نرمال منفرد در نظر گرفته اند. در این رساله حالتی را در نظر می گیریم که $rank(\mathbf{\Gamma}) = m$ و $rank(\mathbf{\Omega}) = d$ اما $rank(\Sigma^*) < m+d$. آرلانو والی و آزالینی (۲۰۰۶) نشان داده اند که در این حالت توزیع $\mathbf{Y}$ با کمی تفاوت همان توزیع (۴) است. در واقع توزیع $\mathbf{Y}$، همان توزیع چوله نرمال چند متغیره است که در آن:

$$\Phi_m(\mathbf{a};\ \mathbf{\Gamma}) = P(\mathbf{C}_0\mathbf{W} \le \mathbf{a}) = \int_{\{\omega:\ \mathbf{C}_0\omega \le a\}} \varphi_d(\omega;\ \mathbf{I}_r)\, d\omega$$

که در آن ماتریس $\mathbf{C}_0$ در رابطه $\mathbf{C}_0\mathbf{C}_0^T = \mathbf{\Gamma} - \mathbf{\Lambda}^T\mathbf{\Omega}^{-1}\mathbf{\Lambda}$ صدق می کند. در این حالت می گوئیم $\mathbf{Y}$ دارای توزیع چوله نرمال منفرد چند متغیره است. برای مطالعه و بررسی جزئیات بیشتر، خواننده می تواند به شیخی و جمالیزاده (۲۰۱۱) و بخش ضمیمه آرلانو والی و آزالینی (۲۰۰۶) مراجعه کند.