توابع توزیع

Distribution Functions

تابع توزیع احتمالات مختلفی را توصیف می‌کند که یک متغیر دلخواه X به خود می‌گیرد. اگر احتمالات وضعیت‌های مختلف یک متغیر گسسته یا پیوسته در جدولی نشان داده شود یک تابع توزیع حاصل می‌شود (Evans et al, 2000: 6). جدول(۱)، که ردیف اول آن متغیر گسسته تعداد خودروی تولیدشده در هر فصل و ردیف دوم احتمال نقص در این خودروها را نشان می‌دهد، بیانگر یک تابع توزیع است.

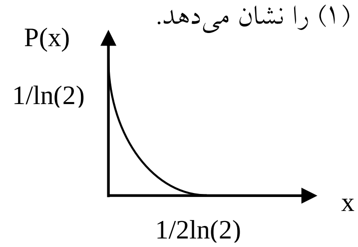
جدول(۱): تابع توزیع خودروهای معیوب

فصل	۱	۲	۳	۴
تعداد خودرو	۲۰۰۰	۵۰۰۰	۱۰۰۰۰	۲۰۰۰۰
احتمال نقص	۰،۲	۰،۲۵	۰،۲۵	۰،۳

تابع توزیع برای متغیر پیوسته دلخواه x به فرم ریاضی (۱) بیان شده است:

$$P(x)=1/[(x+1)\ln 2],\ \ 0\leq x\leq 1 \qquad (۱)$$

در این تابع، P(0) و P(1) هردو کوچک‌تر از یک هستند و مجموع احتمالات در بازه صفر تا یک برابر یک است. شکل(۱) تابع توزیع (۱) را نشان می‌دهد.



شکل(۱): تابع توزیع متناظر با رابطه (۱)

بسته به گسسته یا پیوسته بودن متغیر موردبحث، انواع مختلفی از توابع توزیع احتمال قابل‌شناسایی و تعریف هستند.

تابع توزیع احتمال (Probability Distribution Function)

یک تابع توزیع احتمال تابعی است که برای تعریف و تصریح توزیع احتمال مشخص به‌کار می‌رود. بسته به نوع صفت متغیر، عبارات تابع توزیع تجمعی تابع جِرم احتمال و تابع چگالی احتمال مورداستفاده می‌شود.

تابع توزیع تجمعی (Cumulative Distribution Function)

در نظریه احتمال،تابع توزیع تجمعی(CDF) یک متغیر تصادفی با مقادیر حقیقی، احتمال قبول مقادیر کمتر یا مساوی x است. اگر توزیع از نوع پیوسته باشد، این احتمال برابر سطح زیر منحنی مربوط به تابع توزیع تجمعی در دامنه منفی بی‌نهایت تا x است. به بیان ریاضی، تابع توزیع تجمعی برای متغیر تصادفی X با عبارت زیر بیان می‌شود (Hogg et al, 2019: 37):

$$F(X)=\Pr(X\leq x) \qquad (۲)$$

بر اساس تعریف تابع توزیع تجمعی، احتمال اینکه متغیر X در بازه بین دو مقدار حقیقی a و b قرار گیرد عبارت است از:

$$\Pr(a \prec X \prec b)=F(b)-F(a) \qquad (۳)$$

برای متغیر تصادفی پیوسته که تابع چگالی احتمال آن با نماد $f_X(x)$ داده می‌شود، تابع توزیع تجمعی به‌صورت زیر تعریف می‌شود:

$$F_X(x)=\int_{-\infty}^{x} f_X(t)dt \qquad (۴)$$

بر اساس رابطه فوق، تابع چگالی احتمال یعنی $f_X(x)$ مشتق تابع توزیع تجمعی است. هر تابع توزیع تجمعی در دامنه تعریف خود غیرکاهنده و از سمت راست پیوسته است. همچنین:

$$\lim_{x\to-\infty} F_X(x)=0, \lim_{x\to+\infty} F_X(x)=1 \qquad (۵)$$

برای نمونه، در صورتی که متغیر تصادفی X در بازه [0,1] به‌طور یکنواخت توزیع شده باشد، تابع توزیع تصادفی آن عبارت است از (ibid: 40):

$$F_X(x)=\begin{cases}0 & if\ x\prec 0\\ x & if\ 0\le x\le 1\\ 1 & if\ x\succ 1\end{cases} \qquad (۶)$$

اما اگر متغیر تصادفی X تنها مقادیر گسسته صفر و یک را با احتمالات برابر اختیار کند، تابع توزیع تجمعی با عبارت زیر بیان می‌شود:

$$F_X(x)=\begin{cases}0 & if\ x\prec 0\\ 0.5 & if\ 0\prec x\prec 1\\ 1 & if\ x\ge 1\end{cases} \qquad (۷)$$

تابع احتمال مشترک (توأم) (Joint Probability Function)

این تابع در واقع توزیع احتمال را برای دو یا چند متغیر تصادفی دلخواه نشان می‌دهد. برای نمونه، اگر دو کیسه هرکدام حاوی ۳ مهره آبی و ۲ مهره قرمز باشد و یک مهره به‌طور تصادفی از هر کیسه بیرون کشیده شود، پیشامدهای مستقل از هم هستند. یعنی اینکه مهره استخراج‌شده از کیسه اول آبی باشد، به مهره بیرون‌آمده از کیسه دوم ربط ندارد. با فرض اینکه A و B متغیرهای تصادفی گسسته باشند که A مربوط به پیشامد بیرون آمدن مهره از کیسه اول و B مربوط به پیشامد بیرون آمدن مهره از کیسه دوم باشد، احتمال بیرون آمدن یک مهره آبی از هر کیسه ۶۰ درصد است و احتمال بیرون آمدن یک مهره قرمز از هر کیسه، ۴۰ درصد است. توزیع احتمالات مشترک (توأم) را با جدول (۲) می‌توان نشان داد:

جدول(۲): توزیع احتمالات مشترک

۱ \ ۲	A: مهره آبی	B: مهره قرمز	P(B)
A: مهره آبی	۰/۶*۰/۶=۰/۳۶	۰/۶*۰/۴=۰/۲۴	۰/۶
B: مهره قرمز	۰/۴*۰/۶=۰/۲۴	۰/۴*۰/۴=۰/۱۶	۰/۴
P(A)	۰/۶	۰/۴	

خانه‌های سایه‌دار جدول ۲ احتمالات مشترک را نشان می‌دهند. لذا برای مثال، احتمال توأم بیرون آمدن مهره آبی از کیسه اول و بیرون آمدن مهره قرمز از کیسه دوم برابر ۲۴ درصد است.

برای دو متغیر تصادفی X و Y، تابع احتمال مشترک به فُرم زیر تعریف می‌شود (گجراتی، ۱۳۷۸: ۹۷۸):

$$F_{X,Y}(x,y)=\Pr(X\le x, Y\le y) \qquad (۸)$$

اگر دو متغیر تصادفی X و Y مستقل از هم باشند، تابع احتمال مشترک به فرم زیر تعریف می‌شود:

$$F_{X,Y}(x,y)=F_X(x).F_Y(y) \qquad (۹)$$

تابع احتمال مشترک برای دو متغیر تصادفی گسسته X و Y عبارت است از:

$$\Pr_{X,Y}(x,y)=\Pr(X=x, Y=y) \qquad (۱۰)$$

رابطه فوق برای n متغیر X_1 تا X_n قابل‌تعمیم است.

تابع چگالی احتمال (Probability Density Function)

در نظریه احتمال، تابع چگالی احتمال (PDF) یک متغیر تصادفی پیوسته تابعی است که مقدار آن در هر نقطه از فضای نمونه می‌تواند احتمال نسبی تعلق آن نقطه به فضای نمونه را به‌دست دهد. اگر نمونه‌گیری با حجم‌های نمونه

متفاوت صورت گیرد، مقادیر تابع چگالی احتمـال، متفـاوت بوده و دلالت دارد بر اینکـه احتمـال متنـاظر بـا یـک متغیـر تصادفی از یک نمونه با احتمال مربوط به متغیـر تصـادفی از نمونه دیگر برابر نیست.

تابع چگالی احتمالِ متغیر تصادفی گسسـته بـه فُـرم زیـر تعریف می‌شود (Mittelhammer, 2013: 52):

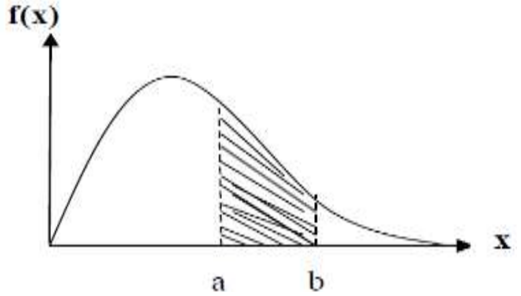
$$f(x) = \Pr(x), \forall x \in D_f \qquad (۱۱)$$
$$f(x) = 0, \forall x \notin D_f$$

مطابق رابطه (۱۱)، مقدار تـابع چگـالی احتمـال در دامنه تعریف (D_f) متغیر تصادفی، احتمال x را نشـان می‌دهد، در حالی کـه مقـدار f در سـایر نقـاط روی محور اعداد حقیقی برابر صفر است.

برای متغیر تصادفی پیوسته، دامنـه تعریـف تـابع نامتناهی و شمارش‌ناپذیر اسـت، لـذا یـک تـابع بـا مقدار غیرمنفی در بازه ∞- تا ∞+ می‌توان تعریـف کرد، به نحوی که در هـر زیرمجموعـه A از دامنـه تعریف (D_f) مقدار احتمال با تـابع چگـالی احتمـال بیان می‌شود و در خـارج از دامنـه تعریـف، مقـدار احتمال برابر صـفر اسـت (Larsen and Marx, 2018,: pp. 133-134):

$$A \subset D_f, P_X(A) = \int_{x\in A} f(x)dx \qquad (۱۲)$$
$$f(x) = 0, \forall x \notin D_f$$

به‌طور مشخص، احتمال اینکه متغیر x در بازه a و b قرار گیرد، با سطح زیر تابع چگالی احتمال در فاصله a تـا b داده می‌شود(Mittelhammer, 2013: 20):



شکل(۲): احتمال به‌مثابه سطح زیر منحنیPDF

در شکل (۲)، مجموعـه A کلیـه نقـاط بـین a و b را نشـان می‌دهد، لذا احتمال رخداد A برابر با سطح زیر منحنـی f(x) در فاصله a تا b (سطح هاشورخورده) است:

$$P_X(A) = \int_a^b f(x)dx \qquad (۱۳)$$

امید ریاضی (Expected Value)

میانگین یک متغیر تصادفی X امید ریاضی نامیده می‌شـود و به‌طور معمول با نماد E(X) نشان داده می‌شود کـه در آن، E حرف اول از معادل انگلیسی امید ریاضی اسـت. اگـر متغیـر گسسته X مقادیر x_1، x_2 و...x_n را با احتمـالات p_1، p_2 و...p_n به خود بگیرد، آنگاه E(X) به فرم زیر نوشته می‌شـود (Gut, 2005: 47):

$$E(X) = \sum_{i=1}^{n} p_i x_i \qquad (۱۴)$$

برای متغیر پیوسته X با تابع چگالی f(x)، امید ریاضی با عبارت زیر داده می‌شود:

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx \qquad (۱۵)$$

امید ریاضی خواص جالبی دارد که برخی از آن‌ها عبارتند از (گجراتی، ۱۳۷۸: ۹۸۶–۹۸۵):

۱) امید ریاضی یک مقدار ثابت a با خـودش برابـر اسـت: $E(a) = a$

۲) امید ریاضی حاصل‌ضرب یک مقدار ثابت a در متغیر تصادفی X برابر است با حاصل‌ضرب مقدار ثابت a در امید ریاضی آن متغیر تصادفی: $E(aX) = aE(X)$

۳) امید ریاضی مجموع چند متغیر تصادفی برابر است با مجموع امید ریاضی آن متغیرها:

$$E(X + ... + Z) = E(X) + ... + E(Z)$$

۴) اگر متغیر تصادفی Y ترکیب خطی از متغیر تصادفی X باشد، امید ریاضی Y ترکیب خطی از امید ریاضی X خواهد بود. برای ترکیب خطی Y=a+bX داریم: E(Y)=a+bE(X)

۵) اگر X و Y دو متغیر تصادفی مستقل از هم باشند، آنگاه امید ریاضی حاصل‌ضرب آن‌ها برابر است با حاصل‌ضرب امید ریاضی آن‌ها:

$$E(XY)=E(X).E(Y)$$

۶) اگر X متغیر تصادفی دلخواهی باشد، امید ریاضی X^2 در دو فرم گسسته و پیوسته عبارت است از:

$$E(X^2) = \sum_{i=1}^{n} p_i x_i^2 \qquad (۱۶)$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx \qquad (۱۷)$$

توزیع برنولی (Bernoulli Distribution)

در نظریه احتمال، توزیع برنولی، به افتخار ریاضیدان سوئیسی ژاکوب برنولی[۱] یک توزیع احتمال گسسته به‌شمار می‌رود. در این توزیع متغیر تصادفی *X* با احتمال p مقدار یک و با احتمال q=1-p مقدار صفر را اختیار می‌کند. یعنی توزیع برنولی توزیع احتمال هر تجربه (آزمایش) مستقل است که پاسخ درست/غلط یا آری/نه را به‌دنبال دارد. به بیان دیگر، احتمال موفقیت برابر p و احتمال شکست برابر q است. برای مثال در پرتاب یک سکه اگر شیر و خط را با صفر و یک نشان دهیم و p احتمال شیر آمدن باشد، q احتمال خط آمدن خواهد بود. توزیع برنولی، یک حالت خاص از توزیع دو نقطه‌ای قلمداد می‌شود، با این تفاوت که در توزیع دونقطه‌ای[۲] رخدادها لزوماً صفر و یک نیستند. برای نمونه، توزیع

$$f(x) = \begin{cases} 1 & if \quad p = 0.5 \\ -1 & if \quad 1-p \end{cases} \qquad (۱۸)$$

یک توزیع دونقطه‌ای است. در توزیع دونقطه‌ای، امید ریاضی، صفر و واریانس یک بوده و شرط $Kurt \geq (Skew)^2 + 1$ که در آن *Skew* و *Kurt* به‌ترتیب چولگی و کشیدگی را نشان می‌دهند، به تساوی بدل می‌شود (Davidson et al, 2007: 3).

اگر متغیر تصادفی X دارای توزیع برنولی باشد، آنگاه (Mittelhammer, 2013: 350):

$$\Pr(X = 1) = p = 1 - \Pr(X = 0) = 1 - q \qquad (۱۹)$$

امید ریاضی (میانگین) توزیع برنولی برابر p و واریانس توزیع برابر pq است.

$$E(X) = p, \quad Var(X) = pq = p(1-p) \qquad (۲۰)$$

این توزیع متقارن نیست و میزان چولگی[۳] و کشیدگی[۴] تقریبی آن عبارتند از:

$$S = \frac{q-p}{\sqrt{pq}} \qquad (۲۱)$$

۱. Jacob Bernoulli (1654-1705)

۲. Two-point Distribution

۳. Skewness

۴. Kurtosis

$$K = \frac{1-6pq}{pq} \qquad (۲۲)$$

توزیع نتایج رفراندوم (همه‌پرسی)، نتایج پرتاب سکه و تاس و جنسیت فرزندان از مواردی هستند که با توزیع برنولی قابل‌بیان هستند.

توزیع برنولی حالت خاصی از توزیع دوجمله‌ای[1] است. در توزیع دوجمله‌ای احتمال رخداد حالت k از میان n حالت با عبارت زیر بیان می‌شود (Larsen and Marx, 2018, pp. 243, 353)

$$b(n,k,p) = \binom{n}{k} p^k (1-p)^{n-k} \qquad (۲۳)$$

که در آن $\binom{n}{k}$ ترکیب k از n حالت است. اگر n برابر یک فرض شود، توزیع برنولی حاصل می‌شود. لذا k باید صفر یا یک باشد، در این صورت:

$$\begin{aligned} b(1,0,p) &= (1-p) = q \\ b(1,1,p) &= p \end{aligned} \qquad (۲۴)$$

توزیع پوآسون (Poisson Distribution)

در نظریه احتمال، توزیع پوآسون- به افتخار ریاضیدان فرانسوی سیمون دنیس پوآسون[2]- یک توزیع احتمال گسسته است که بیانگر احتمال وقوع تعدادی از پیشامدها است که در بازه‌های ثابتی از زمان یا مکان رخ می‌دهند، با این شرط که این پشامدها با آهنگ ثابت و معلوم (فراوانی ثابت) و به‌طور مستقل از زمانِ آخرین رخداد رخ دهند. برای مثال، تعداد دفعات تماس تلفنی با دفتر مدیرعامل یک شرکت در یک ساعت، تعداد دفعات طغیان رودخانه در یک سال، تعداد گل‌ها در یک مسابقه فوتبال، تعداد برخورد شهاب‌سنگ‌ها به کره زمین می‌تواند دارای توزیع پوآسون باشد.

توزیع پوآسون برای رخدادهای نادر (با احتمال کم) و با دفعات رخداد (فراوانی) زیاد به کار می‌رود. در این توزیع، تعداد دفعات یا فراوانی رخداد، n، زیاد بوده (در حد به بی‌نهایت میل می‌کند) و احتمال رخداد، p، کوچک است (در حد به صفر میل می‌کند)، به‌طوری که $np=\lambda$ به صفر نزدیک می‌شود (کمتر از ۰/۰۵ است). پارامتر λ میانگین یا متوسط تعداد رخدادها در یک بازه زمانی را نشان می‌دهد (Linde, 2016: 27-28). اگر k تعداد رخداد یا فراوانی احتمال باشد، احتمال رخداد k با فرمول (۲۵) داده می‌شود:

$$P(k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad (۲۵)$$

در فرمول مذکور، e عدد اویلر و برابر ۲/۷۱۸۲۸۱ بوده و داریم:

$$k! = k \times (k-1) \times (k-2) \ldots \times 3 \times 2 \times 1$$

برای مثال، اگر متوسط تعداد تماس‌های تلفنی در یک ساعت ۵ بار باشد، احتمال ۴ بار تماس تلفنی در ساعت حدود ۴۸ درصد است:

$$P(4) = \frac{e^{-5}5^4}{4!} = 0.47697$$

برای توزیع پوآسون با تابع توزیع احتمال $P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$، تابع مولد گشتاور[3] به فرم زیر تعریف می‌شود:

---

[1] binomial distribution

[2] Siméon Denis Poisson

[3] Moment Generating Function

$$M(t)=E(e^{tx})=\sum_{x=0}^{\infty} e^{tx}p(x)= \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x e^{-\lambda}}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda e^t)^x}{x!} \qquad (۲۶)$$

با ساده‌سازی، رابطه (۲۷) حاصل می‌شود:

$$M(t)=e^{\lambda(e^t-1)} \qquad (۲۷)$$

اگر از تابع فوق نسبت به t مشتقات مرتبه اول و دوم گرفته شود:

$$M'(t)=(\lambda e^t)e^{\lambda(e^t-1)} \qquad (۲۸)$$
$$M''(t)=(\lambda e^t)e^{\lambda(e^t-1)}+(\lambda e^t)[\lambda e^t e^{\lambda(e^t-1)}]$$

بر اساس رابطه (۲۸) می‌توان نوشت:

$$E(x)=M'(0)=\lambda \qquad (۲۹)$$
$$Var(x)=M''(0)-[E(x)]^2=\lambda$$

بنابراین، امید ریاضی و واریانس توزیع مساوی هم و برابر λ هستند (Hogg et al, 2019: 169-170).

توزیع نرمال (Normal Distribution)

در نظریه احتمال، توزیع نرمال یا توزیع گاوس یک توزیع بسیار معروف احتمالات برای صفات پیوسته است.

زمانی که توزیع جامعه آماری در علوم طبیعی یا اجتماعی نامعلوم باشد، فقط در نمونه‌های بزرگ فرض می‌شود که توزیع جامعه آماری به‌سمت توزیع نرمال میل می‌کند.

توزیع نرمال با قضیه حد مرکزی[۱] ارتباط قوی دارد. بر اساس این قضیه، اگر X_1، X_2 و...تا X_n دنباله‌ای از متغیرهای σ^2 تصادفی با توزیع مستقل از هم و یکنواخت و دارای واریانس محدود باشند. چنانچه μ امید ریاضی X_iها و σ^2 واریانس آن‌ها باشد، نسبت زیر دارای توزیع نرمال استاندارد خواهد بود (Linde, 2016: 285-286):

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\approx z(0,1) \qquad (۳۰)$$

به توزیع نرمال گاهی توزیع زنگوله‌ای هم گفته می‌شود اما بسیاری از توزیع‌های دیگر ازجمله توزیع کوشی،[۲] توزیع t استیودنت و توزیع‌های لگاریتمی نیز به‌شکل زنگوله هستند. تابع چگالی احتمال برای توزیع نرمال با رابطه (۳۱) بیان می‌شود (Borovkov, 2013: 37):

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (۳۱)$$

توزیع نرمال با دو پارامتر اصلی μ که امید ریاضی و σ^2 که واریانس توزیع است، شناخته می‌شود. همچنین، این توزیع حاوی دو ثابت گنگ $\pi=3.141619...$ و $e=2.718281...$ است. این توزیع تنها توزیعی است که میانگین، نما و میانه بر هم منطبق و برابر با μ هستند.

توزیع نرمال حول $x=\mu$ متقارن است و در این نقطه بیشینه تابع برابر $\frac{1}{\sigma\sqrt{2\pi}}$ است. این تابع تک‌نمایی است، یعنی فقط یک نما دارد. نقاط عطف این تابع در نقاطی به طول‌های $\mu\pm\sigma$ قرار دارند.

اگر در تابع (۳۱)، μ برابر صفر و σ^2 برابر یک باشد، آنگاه

$$z(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \qquad (۳۲)$$

---

۱ Central limit theorem

۲ Cauchy distribution

دارای توزیع نرمال استاندارد خواهد بود. در تابع (۳۲)، سطح زیر منحنی در بازه $[-\infty,+\infty]$ برابر یک است. افزون بر این، تابع فوق حول x=0 متقارن است و نقاط عطف آن در نقاطی به طول‌های x=-1 و x=+1 قرار دارند و عرض از مبدأ آن‌ها برابر $\frac{1}{\sqrt{2e\pi}}$ است. حداکثر تابع نیز به‌ازای x=0 برابر $\frac{1}{\sqrt{2\pi}}$ است (Stigler, 1982: 137-138).

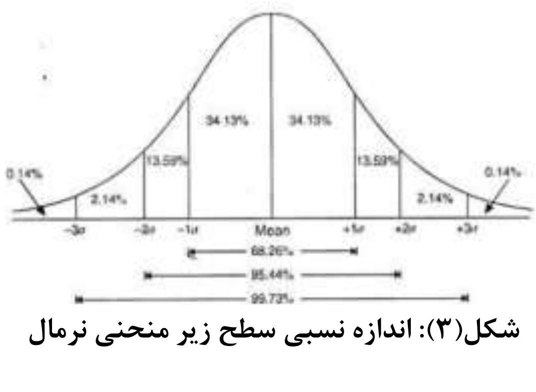
با فرض اینکه متغیر تصادفی X دارای واریانس محدود σ^2 و امید ریاضی μ باشد، به‌ازای هر k مثبت،

$$P(|X-\mu| \ge k\sigma) \le 1/k^2 \qquad (۳۳)$$

یا

$$P(|X-\mu| < k\sigma) \ge 1-\frac{1}{k^2} \qquad (۳۴)$$

عبارت (۳۳) یا (۳۴) به نامساوی چبیشف[1] معروف است. با قرار دادن k=1,2, 3 در توزیع نرمال، تقریباً ۶۸ درصد سطح زیرمنحنی نرمال بین دو مقدار $\mu\pm\sigma$، ۹۵ درصد بین دو مقدار $\mu\pm2\sigma$ و ۹۹/۷ درصد بین دو مقدار $\mu\pm3\sigma$ قرار می‌گیرند (Mittelhammer, 2013: 135-136).



شکل(۳): اندازه نسبی سطح زیر منحنی نرمال

1. Chebyshev's inequality

در صورتی که متغیر X دارای توزیع نرمال با میانگین μ_X و واریانس σ_X^2 و متغیر Y دارای توزیع نرمال با میانگین μ_Y و واریانس σ_Y^2 باشند و این دو متغیر مستقل از هم توزیع شده باشند، آنگاه برحسب قضیه حد مرکزی و قضایای لیندبرگ و لوی (Lindeberg–Lévy)، ترکیب خطی دو متغیر نیز دارای توزیع نرمال خواهد بود (Marx, Larsen and 2018: 254):

$$Z = aX + bY \approx N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2) \qquad (۳۵)$$

توزیع کای‌دو (Chi-Square Distribution)

در نظریه احتمال، توزیع کای‌دو (چی دویا χ^2) با k درجه آزادی، توزیع مجموع مربعات k متغیر تصادفی مستقل دارای توزیع نرمال استاندارد است (Borovkov, 2013:177). توزیع χ^2 یک حالت خاص از توزیع احتمال گاما است که در آمار استنباطی، به‌ویژه در آزمون فرضیه‌ها یا ساخت فاصله اطمینان، زمانی که واریانس جامعه آماری معلوم نیست، مورداستفاده قرار می‌گیرد.

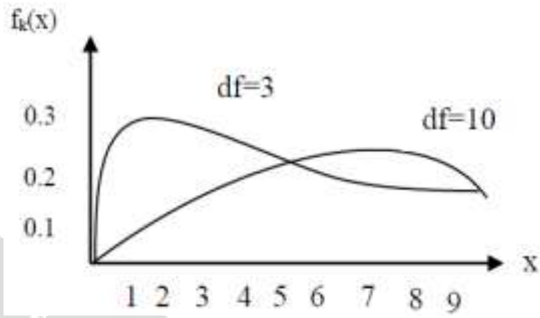
توزیع $\chi2$ برای برازش توزیع مناسب با مشاهدات تجربی، در بررسی استقلال معیارهای طبقه‌بندی داده‌های کیفی و برآورد فاصله اطمینان برای واریانس جامعه بر حسب واریانس نمونه دارای توزیع نرمال هم به‌کار می‌رود.

اگر Z_1، ...، Z_k متغیرهای با توزیع نرمال استاندارد و مستقل از هم باشند، متغیر تصادفی Q دارای توزیع کای‌دو با k درجه آزادی خواهد بود:

$$Q=\sum_{i=1}^{k} z_i^2 \approx \chi_k^2 \qquad (۳۶)$$

توزیع کای‌دو دارای چولگی به راست است و با افزایش درجه آزادی بالای ۱۰۰ توزیع کای‌دو به توزیع نرمال

استاندارد میل می‌کند. در این توزیع پارامتر k یک عدد صحیح مثبت بوده و درجه آزادی توزیع (تعداد z_i ها) را مشخص می‌کند. شکل(۴) تابع چگالی احتمال(PDF)[1] را برای توزیع تقریبی χ^2 در درجات آزادی ۳ و ۱۰ نشان می‌دهد.



شکل(۴): تابع چگالی احتمال در توزیع تقریبی کای‌دو با درجات آزادی ۳ و ۱۰

در صورتی که متغیر تصادفی X دارای توزیع کای دو با درجه آزادی k باشد و k به‌سمت بی‌نهایت میل کند، آنگاه متغیر z به توزیع نرمال استاندارد همگرا می‌شود (Johnson et al, 1994: 415-420):

$$z = \frac{\chi_k^2 - k}{\sqrt{2k}} \to z(0,1) \qquad (۳۷)$$

به‌طور خاص، به‌ازای درجهٔ آزادی بیش از ۱۰۰ متغیر زیر دارای توزیع نرمال استاندارد است:

$$\sqrt{2\chi^2} - \sqrt{2k-1} \approx z(0,1) \qquad (۳۸)$$

همچنین، اگر متغیر تصادفی X دارای توزیع کای‌دو با درجه آزادی k باشد دو رابطه زیر هم برقرار است (Hogg et al, 2019: 178):

$$E(X) = k, \quad Var(X) = 2k \qquad (۳۹)$$

افزون بر این، اگر a عدد ثابت مثبتی باشد، آنگاه aX دارای توزیع گاماست و در آن

$$X \approx \chi_k^2 \to aX \approx \Gamma(k/2, 2a) \qquad (۴۰)$$

اگر دو متغیر X و Y دارای توزیع های کای دو مستقل از هم با درجات آزادی K_1 و K_2 باشند، مجموع آن‌ها نیز دارای توزیع کای دو خواهد بود که درجه آزادی آن مجموع درجات آزادی دو توزیع قبلی خواهد بود (Linde, 2016: 187):

$$X + Y \approx \chi^2_{K1+K2} \qquad (۴۱)$$

توزیع F (F- Distribution)

توزیع F که در آمار و احتمال به توزیع فیشر–اسندکور[2]هم معروف است، از نوع توزیع احتمالات پیوسته است. این توزیع در تحلیل واریانس[3] آزمون برابری واریانس‌های دو یا چند جامعه آماری و سایر آزمون‌های آماری و اقتصادسنجی کاربرد دارد.

این توزیع از نسبت دو توزیع کای‌دو، که بر درجات آزادی‌شان تقسیم شده، به‌دست می‌آید. به‌عبارت دیگر، اگر X دارای توزیع کای‌دو با درجه آزادی m و متغیر Y دارای توزیع کای‌دو با درجه آزادی n باشد، نسبت دو توزیع با توزیع F نشان داده می‌شود (Linde, 2016: 195):

$$F = \frac{\chi_m^2}{\chi_n^2} \approx F(m,n) \qquad (۴۲)$$

---

1. Probability Density Function

2. Ronald Fisher and George W. Snedecor

3. Analysis of Variance (ANOVA)

در رابطه (۴۲)، m و n به ترتیب درجات آزادی صورت و مخرج آماره F هستند. در توزیع F بر اساس رابطه (۱۲)، میانگین توزیع برابر است با $n/(n-2)$ که به‌ازای n بزرگ‌تر از ۲ صادق است. همچنین واریانس توزیع F با عبارت زیر برابر است:

$$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

که به‌ازای n بزرگ‌تر از ۴ صادق است (Mittelhammer, 2013: 344-345).

اگر X دارای توزیع F با درجه آزادی df_1 درصورت و df_2 در مخرج کسر باشد،آنگاه 1/X نیز دارای توزیع F با درجه آزادی df_2 درصورت و df_1 در مخرج کسر خواهد بود. اگر X دارای توزیع t با درجه آزادی n باشد، آنگاه مربع (مجذور X) دارای توزیع F با درجه آزادی یک درصورت و n در مخرج خواهد بود (Mittelhammer, 2013: 346).

توزیع F در درجات آزادی کم به‌سمت راست چولگی دارد اما در صورتی که درجه آزادی صورت و مخرج در توزیع F بیشتر و بیشتر شوند، توزیع F به توزیع نرمال نزدیک می‌شود.

توزیع t (t-student Distribution)

در احتمال و آمار، توزیعt یا توزیع t-استیودنت در گروه یا خانواده توزیع‌های احتمالی پیوسته قرار دارد. اگر حجم یا اندازه نمونه آماری کوچک بوده ($n\leq30$) و انحراف معیار جامعه آماری (σ) نامعلوم باشد، با تکرار نمونه‌ها، توزیع t حاصل می‌شود که توسط ویلیام سیلی گوست[1] با نام مستعار Student- در مجله بیومتریکا[2] معرفی شد (Gosset, 1908).

توزیع t برای ارزیابی معناداری آماری تفاوت بین میانگین‌های دو جامعه آماری، ساخت فاصله اطمینان برای تفاوت بین دو میانگین‌های دو جامعه آماری و در تحلیل رگرسیون خطی به‌کار می‌رود. در آزمون تفاوت میانگین‌های دو جامعه آماری با میانگین‌های μ_1 و μ_2 فرضیه صفر (H_0) و فرضیه مقابل (H_A) به فرم زیر نوشته می‌شود:

$$H_0: \mu_1=\mu_2 \quad H_A: \mu_1\neq\mu_2 \tag{۴۳}$$

اگر یک نمونه تصادفی با مقادیر X_1، X_2، ...وX_n به از یک توزیع $N(\mu, \sigma^2)$ در دسترس باشد و واریانس جامعه آماری (σ^2) معلوم نباشد، پارامتر μ با میانگین نمونه‌ای و پارامترσ^2 توسط واریانس نمونه‌ای S^2 به فرم زیر برآورد می‌شوند:

$$\overline{X}=\sum_{i=1}^{n}X_i \tag{۴۴}$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2 \tag{۴۵}$$

با معلوم شدن میانگین و واریانس نمونه، متغیر تصادفی

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \tag{۴۶}$$

از توزیع t با درجه آزادی n-1 برخوردار است (رنچر و کریستنسن، ۱۳۹۳: ۱۵۶-۱۵۵).

توزیع t با توزیع نرمال و توزیع کای‌دو مرتبط است (Mittelhammer, 2013: 339). در واقع، اگر Z_1 دارای توزیع نرمال استاندارد با میانگین صفر و واریانس یک و متغیر Z_2 دارای توزیع کای‌دو با درجه آزادی k باشد و Z_1 مستقل از

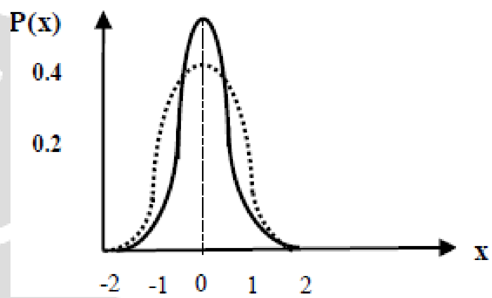
[1] William Sealy Gosset

[2] Biometrika

Z_2 توزیع شود، نسبت زیر یک توزیع t با درجه آزادی k خواهد بود:

$$Z_1 \approx N(0,1),\ Z_2 \approx \chi^2_k \rightarrow t = \frac{Z_1}{\sqrt{Z_2 / k}} \qquad (۴۷)$$

با افزایش درجه آزادی، توزیع t به توزیع نرمال نزدیک می‌شود (منحنی پیوسته در شکل ۲) و با کاهش درجه آزادی، برآمدگی[1]توزیع کمتر می‌شود (منحنی نقطه‌چین در شکل ۵).



شکل(۵): توزیع t با درجات آزادی متفاوت

(منحنی پیوسته، درجه آزادی بیشتر و منحنی نقطه چین، درجه آزادی کمتر)

توزیع t دارای میانگین، میانه و مُد (نما) برابر صفر است، اما واریانس توزیع t برابر k/(k-2) به‌ازای k بزرگ‌تر از ۲ و بی‌نهایت به‌ازای درجه آزادی ۱ تا ۲ است. همچنین، با k بیشتر از ۳ میزان چولگی برابر صفر و با k بیشتر از ۴ میزان کشیدگی برابر 6/(k-4) است. بر حسب درجه آزادی و سطوح معنی‌داری آماری ۱٪و ۵٪و ۱۰٪ و بسته به یک دنباله یا دو دنباله گرفتن توزیع، جدول‌های توزیع t تنظیم و در کتاب‌های مرجع آماری درج شده است.

توزیع لگاریتم نرمال (Log-Normal Distribution)

در نظریه احتمال، توزیع لوگ–نرمال (لگاریتمی نرمال)، یک توزیع احتمال پیوسته از یک متغیر تصادفی است که لگاریتم آن دارای توزیع نرمال است. بنابراین، در صورتی که متغیر تصادفی X دارای توزیع لوگ–نرمال باشد، آنگاه Ln(X) دارای توزیع نرمال خواهد بود. همین‌طور اگر Y دارای توزیع نرمال باشد، آنگاه e^Yدارای توزیع لوگ–نرمال خواهد بود (Shimizu andCrow, 1988: 2-3). متغیر تصادفی دارای توزیع لگاریتم نرمال، همواره مقادیر حقیقی مثبت را به خود می‌گیرد. این توزیع به توزیع کاب–داگلاس[2] توزیع گالتون[3]و توزیع گیبرات[4]نیز معروف است.

به‌طور کلی، اگر Z متغیر نرمال استاندارد باشد آنگاه لگاریتم طبیعی متغیر X به فرم (۴۸) دارای توزیع نرمال است:

$$Z \approx N(0,1) \rightarrow X = e^{\mu+\sigma Z} \qquad (۴۸)$$

که در آن μ میانگین و σ انحراف معیار لگاریتم طبیعی متغیر X هستند.

توزیع لوگ–نرمال در توصیف پدیده‌های طبیعی اهمیت دارد. در این پدیده‌ها فرایندهای طبیعی رشد حاصل انباشت تغییرات نسبی بسیار اندک است. برای مثال، از دیدگاه گیبرات اگر نرخ انباشت این تغییرات جزئی در طول زمان تغییر نکند، رشد متغیرها مستقل از اندازه آن‌ها خواهد بود. مدت زمان بازی شطرنج، مقدار محتوای ارسال‌شده در شبکه‌های اجتماعی و اینترنت، اندازه بافت‌های زنده بدن، ویژگی‌های فیزیولوژیک مثل فشار خون، اندازه ذرات، جرم

---

۱. Kurtosis

۲. Charles W. Cobb (1875–1949) &Paul H. Douglas (1892 1976)

۳. Galton, Francis (1822-1911)

۴. Gibrat, Robertt (1904–1980)

مولی، بیشینه بارندگی ماهانه (سالانه)، درآمد اکثریت جمعیت، لگاریتم نرخ ارز، لگاریتم قیمت سهام، تعداد ارجاعات به مقالات، اندازه شهرها، زمان تعمیر در سیستم‌های قابل‌تعمیر و مانند آن‌ها مواردی هستند که توزیع لگاریتمی نرمال در مورد آن‌ها مصداق دارد.

اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد، آنگاه e^X دارای توزیع لوگ-نرمال با میانگین μ و واریانس σ^2 خواهد بود. بالعکس، اگر متغیر تصادفی X دارای توزیع لوگ-نرمال با میانگین μ و واریانس σ^2 باشد، آنگاه Ln(X) دارای توزیع نرمال با میانگین μ و واریانس σ^2 خواهد بود.

اگر X_1 دارای توزیع لوگ-نرمال با میانگین μ_1 و واریانس σ_1^2 و ....و متغیر X_n دارای توزیع لوگ-نرمال با میانگین μ_n و واریانس σ_n^2 باشند، آنگاه حاصل‌ضرب آن‌ها نیز دارای توزیع لوگ-نرمال خواهد بود که میانگین آن مجموع میانگین متغیرهای X_1 تا X_n و واریانس آن مجموع واریانس متغیرهای X_1 تا X_n خواهد بود (Mittelhammer, 2013: 355).

$$X_1 \ldots X_n \approx \log normal\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right) \qquad (۴۹)$$

اگر متغیر تصادفی X دارای توزیع لوگ-نرمال با میانگین μ و واریانس σ^2 باشد، آنگاه aX دارای توزیع لوگ-نرمال با میانگین $\mu+\ln a$ و واریانس σ^2، متغیر X+a دارای توزیع لوگ-نرمال با میانگین $\mu+a$ و واریانس σ^2، متغیر 1/X دارای توزیع لوگ-نرمال با میانگین $-\mu$ و واریانس σ^2 و متغیر X^k دارای توزیع لوگ-نرمال با میانگین $k\mu$ و واریانس $k^2\sigma^2$ است.

کتاب‌شناسی

گجراتی، دامودار (۱۳۷۸). *مبانی اقتصادسنجی*، ترجمه حمید ابریشمی، جلد دوم، چاپ دوم، تهران: انتشارات دانشگاه تهران.

رنچر، آلوین سی و ویلیام اف. کریستنسن (۱۳۹۳). *روش‌های تحلیل چندمتغیره آماری*، ترجمهٔ امیرافشین فتاحی، چاپ اول، تهران: مؤسسه عالی آموزش و پژوهش مدیریت و برنامه‌ریزی.

Borovkov, Alexandr A. (2013). *Probability Theory*.London: Springer-Verlag.

Davidson, J., Monticini, A., & Peel, D. (2007). "Implementing the wild bootstrap using a two-point distribution", *Economics Letters*, 96(3), 309-315.

Evans, Merran, Nicholas Hastings, and Brian Peacock. (2000). *Statistical Distributions*, 3rd ed, New York: Wiley.

Gut, Allan (2005). *Probability: a graduate course*, First ed, NewYork: Springer Science & Business Media.

Hogg, Robert V.; Joseph W. McKean and Allen T. Craig (2019). *IntroductiontoMathematical Statistics*, Eighth ed, US: Pearson Education, Inc.

Johnson, Norman L., Adrienne W. Kemp, and Samuel Kotz.(2005). *Univariate discrete distributions*, Vol. 444. John Wiley & Sons.

Larsen, Richard J. and Morris L. Marx (2018). *Introduction to Mathematical Statistics and Its Applications*, Sixth ed, US: Pearson.

Linde, Werner (2016). *Probability Theory: A First Course in Probability Theory and Statistics*, First ed, Berlin: Walter de Gruyter.

Mittelhammer, Ron C. (2013). *Mathematical statistics for economics and business*, 2nd ed, New York: Springer.

Saw, John G.; Yang, Mark C. K.; Mo, Tse Chin (1984). "Chebyshev Inequality with Estimated Mean and Variance", *The American Statistician*, 38 (2): 130–2.

Shimizu, Kunio, and Edwin L. Crow. History, genesis, and properties, Pp.1-25, In: Crow, E. L. and Shimizu, K. (Ed.) (1988). *Lognormal Distributions: Theory and Applications*, New York: Dekker.

Stigler, Stephen M. (1982). "A Modest Proposal: A New Standard for the Normal", *The American Statistician*, 36 (2): 137–138.

Student [William Sealy Gosset] (1908). "The probable error of a mean", *Biometrika*, 6(1): 1–25.

لطفعلی عاقلی

پژوهشکده اقتصاد دانشگاه تربیت مدرس، تهران، ایران