* فصل توزیع های احتمال معروف :

بعضی از متغیرهای تصادفی دارای یک الگوی خاص هستند و می‌توان برای آن ها یک نوع توزیع احتمال را درنظر گرفت. در این فصل به چند مورد از توزیع های احتمال خاص هم در حالت پیوسته و هم گسسته اشاره می کنیم.

* گسسته :

۱) توزیع برنولی : $X \sim Ber(p)$ ← متغیر X دارای توزیع برنولی با پارامتر p است.

✓ آزمایش تصادفی دارای دو نتیجه ی موفقیت و شکست است.

✓ X : متغیر تصادفی تعداد موفقیت ها در یک بار انجام آزمایش برنولی.

✓ مقدارهای ممکن X : ۰ و ۱

✓ جدول توزیع احتمال X :

x	0	1
$p_X(x)$	$1-p$	p

که در آن p احتمال موفقیت است. (پس $1-p$ احتمال شکست است.)

✓ تابع احتمال X : $p_X(x) = p^x (1-p)^{1-x}$; $x = 0, 1$

از این پس قرار می دهیم $1-p = q$.

✓ اگر $X \sim Ber(p)$ باشد آنگاه $E(X) = p$ و $var(X) = pq$.

مثال : تاسی را یک مرتبه پرتاب می کنیم و متغیرتصادفی X را تعداد دفعاتی که شماره ی 6 ظاهر شود قرار می دهیم پس $X \sim Ber(\frac{1}{6})$ است و داریم :

$$p_X(x) = \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{1-x} \quad x = 0, 1$$

$$E(X) = \frac{1}{6} \quad , \quad var(X) = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)$$

۲) توزیع دوجمله‌ای $X \sim Bin(n,p)$

✓ آزمایش تصادفی: انجام n بار آزمایش برنولی به طور مستقل

✓ X: متغیر تصادفی تعداد موفقیت‌ها در این n بار آزمایش

✓ مقادیر ممکن X: $0, 1, \dots, n$

✓ تابع احتمال: $p_X(x) = \binom{n}{x} p^x q^{n-x} \quad x=0,\dots,n$

✓ اگر $X \sim Bin(n,p)$ باشد آنگاه $E(X)=np$ و $Var(X)=npq$.

مثال: سکه‌ای که شانس مشاهده شیر در آن دو برابر خط است را ۴ مرتبه پرتاب می‌کنیم. اگر X تعداد شیرهای مشاهده شده در ۴ مرتبه پرتاب سکه باشد $P(X=3)$ را بیابید:

$$X \sim Bin(4, \tfrac{2}{3})$$

$$P(X=3) = \binom{4}{3} (2/3)^3 (1/3)^{4-3}$$

* برای محاسبه‌ی احتمالات در توزیع دوجمله‌ای از جدول دوجمله‌ای نیز می‌توان استفاده کرد. این جدول شامل مقادیر احتمال تجمعی یعنی $P(X \le r)$ به ازای n و p های مختلف می‌باشد.

۳) توزیع هندسی $X \sim Ge(p)$

✓ آزمایش تصادفی: تکرار آزمایش تصادفی برنولی به طور مستقل

✓ X: متغیر تصادفی تعداد دفعات تکرار آزمایش تا رسیدن به اولین موفقیت

✓ مقادیر ممکن X: $1, 2, 3, \ldots$

✓ تابع احتمال X: $p_X(x) = p\,q^{x-1} \quad ; \quad x = 1, 2, \ldots$

✓ اگر $X \sim Ge(p)$ باشد آنگاه $E(X) = \frac{1}{p}$ و $var(X) = \frac{q}{p^2}$.

مثال: سکه‌ی سالم را پرتاب می‌کنیم تا به یک شیر برسیم. احتمال اینکه اولین شیر در پرتاب پنجم ظاهر شود چقدر است؟

$X \sim Ge(p) \quad ; \quad p = 1/2$

$$P(X=5) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^4 = \left(\frac{1}{2}\right)^5$$

۴) توزیع دوجمله‌ای منفی: $X \sim NB(r, p)$

✓ آزمایش تصادفی: انجام آزمایش برنولی به طور مستقل

✓ X: متغیر تصادفی تعداد آزمایش‌ها تا رسیدن به r امین موفقیت.

✓ مقادیر ممکن X: $r, r+1, r+2, \ldots$

✓ تابع احتمال X: $p_X(x) = \binom{x-1}{r-1} p^r q^{x-r} \quad ; \quad x = r, r+1, \ldots$

✓ اگر $X \sim NB(r, p)$ آنگاه $E(X) = \frac{r}{p}$ و $var(X) = \frac{rq}{p^2}$.

مثال: سکه‌ی سالم را پرتاب می‌کنیم. احتمال اینکه سومین شیر در چهارمین پرتاب ظاهر شود چقدر است؟

$X \sim NB(3, \frac{1}{2})$

$$P(X=4) = \binom{3}{2}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1$$

$X \sim HG(N, M, n)$

(5) توزیع فوق هندسی

✓ آزمایش تصادفی: انجام آزمایش برنولی به طوری که آزمایشات مستقل از هم نباشد.

(انتخاب بدون جایگذاری)

✓ X: متغیر تصادفی تعداد موفقیت ها در یک آزمایش فوق هندسی (تعداد موفقیت ها در n آزمایش غیرمستقل)

✓ مقادیر ممکن X: $\max(0, n-N+M) \leq x \leq \min(n, M)$

✓ تابع احتمال X:

$$p_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

✓ اگر $X \sim HG(N, M, n)$ باشد آنگاه:

$$E(X) = \frac{nM}{N} \quad , \quad var(X) = \frac{nM}{N}\left(1 - \frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

مثال: می خواهیم از دانشجویان یک کلاس که شامل 10 مرد و 7 زن می باشد سه نفر به طور تصادفی انتخاب کنیم. احتمال اینکه در بین این سه نفر حداکثر 2 مرد انتخاب شود چقدر است؟

$$X \sim HG(17, 10, 3)$$

روش اول

$$p(X \leq 2) = 1 - p(X > 2) = 1 - p(X = 3)$$

$$= 1 - \frac{\binom{10}{3}\binom{7}{0}}{\binom{17}{3}}$$

روش دوم

$$p(X \leq 2) = p(X=0) + p(X=1) + p(X=2)$$

$$= \frac{\binom{10}{0}\binom{7}{3} + \binom{10}{1}\binom{7}{2} + \binom{10}{2}\binom{7}{1}}{\binom{17}{3}}$$

(6) توزیع پواسن : $X \sim Poi(\lambda)$

✓ آزمایش تصادفی : آزمایشی که تعداد موفقیت ها در یک فاصله زمانی یا ناحیه ای مشخص را به دست دهد

✓ X : متغیر تصادفی تعداد موفقیت ها در یک فاصله زمانی یا ناحیه مشخص

✓ مقادیر ممکن X : $0, 1, 2, \ldots$

✓ تابع احتمال X : $p_X(x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} \quad ; \quad x = 0, 1, \ldots$

که در آن λ میانگین تعداد موفقیت ها در فاصله زمانی یا مکانی مورد نظر است.

✓ اگر $X \sim Poi(\lambda)$ باشد آنگاه :

$E(X) = var(X) = \lambda$

مثال : به طور متوسط در هر ده دقیقه 6 نفر وارد یک بانک می شوند. احتمال اینکه در پنج دقیقه حداقل 2 نفر وارد بانک شوند چقدر است؟

$$\frac{10\ \text{دقیقه}}{5\ \text{دقیقه}} = \frac{6\ \text{نفر}}{\lambda\ \text{نفر}} \Rightarrow \lambda = 3$$

$$\Rightarrow X \sim Poi(3) \Rightarrow p(X \geq 2) = 1 - p(X \leq 1)$$

$$= 1 - [p(X=0) + p(X=1)]$$

$$= 1 - \left[\frac{e^{-3}3^{0}}{0!} + \frac{e^{-3}3^{1}}{1!}\right]$$

$$= 1 - [e^{-3}(1+3)] = 1 - 4e^{-3}$$

* برای محاسبه ی احتمالات در توزیع پواسن می توان از جداول این توزیع نیز استفاده کرد. در این جداول مقادیر احتمال تجمعی $p(X \leq r)$ را به ازای λ های مختلف وجود دارد.

* تقریب توزیع دوجمله‌ای به وسیله‌ی پواسن :

در آزمایش دوجمله‌ای هرگاه $n \to +\infty$ و $p \to 0$ آنگاه :

$$\lambda \simeq np \Rightarrow \lim_{n \to +\infty} \binom{n}{x} p^{x} (1-p)^{n-x} = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

مثال ۶ : در ظرفی 99 مهره سفید و یک مهره سیاه داریم. از این ظرف 200 مرتبه با جایگذاری مهره بیرون می‌آوریم. احتمال اینکه 7 مرتبه مهره سیاه دیده شود را بیابید.

X : تعداد مهره‌های سیاه مشاهده شده $\quad X \sim Bin(200, 0.01)$

هدف : $P(X=7) = ?$

چون $n \to +\infty$ و $p \to 0$ پس

$$\lambda \simeq (200)(0.01) = 2$$

$$P(X=7) = \frac{e^{-2}\, 2^{7}}{7!} = 0.0034$$

7) توزیع یکنواخت گسسته : $X \sim DU(\theta)$

توزیعی است که در آن متغیر تصادفی گسسته‌ی X تمام مقادیرش را با احتمالات یکسان اختیار می‌کند.

$$p_X(x) = \frac{1}{\theta} \quad ; \quad x = x_1, x_2, \dots, x_\theta$$

در این حالت : $E(X) = \frac{\sum_{i=1}^{\theta} x_i}{\theta}$, $Var(X) = \frac{\sum_{i=1}^{\theta} (x_i - \mu)^2}{\theta}$

مثال : جعبه‌ای شامل n مهره به شماره‌های 1 تا n است. از این جعبه مهره‌ای به تصادف انتخاب می‌کنیم. تابع احتمال شماره‌ی مهره بدست آمده را بنویسید.

$$p_X(x) = \frac{1}{n} \quad x = 1, \dots, n$$

* پیوسته :

۱) توزیع یکنواخت پیوسته : $X \sim U(a,b)$

✓ گوییم متغیر تصادفی X دارای توزیع یکنواخت در بازه‌ی (a,b) است هرگاه تابع چگالی احتمال آن به صورت زیر باشد:

$$f(x) = \begin{cases} \frac{1}{b-a} & a<x<b \\ 0 & \text{else} \end{cases}$$

✓ اگر $X \sim U(a,b)$ باشد آنگاه

$$E(X) = \frac{a+b}{2} \quad , \quad var(X) = \frac{(a-b)^2}{12}$$

مثال : مدت زمانی که طول می‌کشد تا شخصی فاصله‌ی خانه تا ایستگاه قطار را پیاده طی کند دارای توزیع یکنواخت بین 15 تا 20 دقیقه است. اگر شخص ساعت $7:30'$ از منزل خارج شود احتمال اینکه او سوار قطاری شود که $7:48'$ به ایستگاه می‌رسد را بیابید.

X : زمان طی کردن فاصله‌ی خانه تا ایستگاه توسط شخص $\quad X \sim U(15, 20)$

$$P(X<18) = \int_{15}^{18} \frac{1}{20-15}\, dx = \frac{3}{5}$$

۲) توزیع نمایی : $X \sim exp(\theta)$

✓ گوییم متغیر تصادفی X دارای توزیع نمایی با پارامتر θ است هرگاه تابع چگالی احتمال آن به صورت زیر باشد:

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x>0 \\ 0 & \text{else} \end{cases}$$

✓ اگر $X \sim exp(\theta)$ باشد، آنگاه $E(X) = \theta$ و $var(X) = \theta^2$ است.

مثال: مدت زمان برحسب دقیقه که قطار تهران مشهد تاخیر دارد متغیر تصادفی نمایی X با میانگین ۱۰ دقیقه است. مطلوبست تعیین $p(X > E(X))$.

$$X \sim exp(10) \Rightarrow E(X) = 10$$

$$p(X > 10) = \int_{10}^{+\infty} \frac{1}{10} e^{-x/10} dx = -e^{-x/10} \Big|_{10}^{+\infty} = e^{-1}$$

* رابطه‌ی توزیع نمایی و پواسن:

در یک آزمایش پواسن با پارامتر λ که در آن λ میانگین تعداد رخدادهای پیش آمده در یک واحد زمان است، فرض می‌دهیم:

X: تعداد رخدادها در بازه‌ی زمانی $[0, t]$

Y: زمان تا رسیدن به اولین رخداد

در این صورت $X \sim Poi(\lambda t)$ است و داریم:

$$P(Y > t) = p(X = 0) = e^{-\lambda t}$$

$$F_Y(t) = p(Y \leq t) = 1 - e^{-\lambda t} \Rightarrow f_Y(t) = \lambda e^{-\lambda t} \qquad t > 0$$

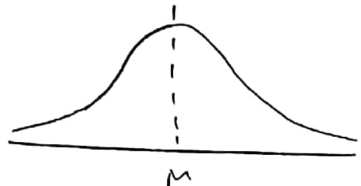
پس $Y \sim exp(\frac{1}{\lambda})$.

مثال: فرض کنید متغیر تصادفی X مدت زمان تا توقف یک دستگاه برحسب 50 روز باشد که دارای توزیع نمایی با پارامتر یک است. احتمال آن که دستگاه در 150 روز حداکثر 2 بار توقف کند را بیابید.

Y: تعداد دفعاتی که دستگاه در 150 روز توقف می‌کند.

طبق رابطه توزیع نمایی و پواسن $Y \sim Poi(3)$. پس $P(Y \leq 2) \overset{\text{طبق جدول پواسن}}{=} 0.4232$.

۳) توزیع نرمال: مهم‌ترین توزیع پیوسته در آمار.



نمودار توزیع نرمال به صورت مقابل است:

$x=\mu$ محور تقارن منحنی نرمال

توزیع نرمال دو پارامتر دارد: میانگین (μ) و واریانس (σ^2).

گوییم $X\sim N(\mu,\sigma^2)$ هرگاه تابع چگالی آن به صورت زیر باشد:

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\qquad x\in\mathbb{R}$$

که در آن $E(X)=\mu$ و $var(X)=\sigma^2$.

* توزیع نرمالی که میانگین آن صفر و واریانس آن یک باشد ← توزیع نرمال استاندارد.

متغیری با توزیع نرمال استاندارد را به صورت $Z\sim N(0,1)$ نمایش می‌دهیم.

* اگر $X\sim N(\mu,\sigma^2)$ آنگاه $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$.

* برای محاسبه‌ی احتمالات توزیع نرمال باید ابتدا متغیر را به نرمال استاندارد تبدیل کنیم و سپس از جدول نرمال استاندارد استفاده کنیم. در این جدول مقادیر احتمال تجمعی یعنی $p(Z\le z)$ قرار دارد.

مثال: فرض کنید متغیر تصادفی X نمرات دانش‌آموزان یک کلاس در امتحان آمار باشد، که دارای توزیع نرمال با میانگین 70 و انحراف معیار 10 است.

الف) اگر 12.3% از دانش‌آموزان این کلاس نمره A و 28.6% از آن‌ها نمره B بیاورند، پایین‌ترین نمره A و پایین‌ترین نمره B را در این کلاس تعیین کنید.

$$0.123=p(X>a)=p\left(\frac{X-\mu}{\sigma}>\frac{a-70}{10}\right)=1-p\left(Z<\frac{a-70}{10}\right)$$

$\Leftarrow$ a ← پایین‌ترین نمره A

b ← " " B

$\xRightarrow{\text{جدول نرمال استاندارد}}$ $\frac{a-70}{10}=1.16 \Rightarrow a=81.6$

b به روشی مشابه به دست می‌آید.

* تقریب توزیع دوجمله‌ای به وسیله‌ی توزیع نرمال:

✓ اگر X یک متغیر تصادفی دوجمله‌ای با میانگین $\mu = np$ و واریانس $\sigma^2 = npq$ باشد آنگاه توزیع حدی $Z = \frac{X-np}{\sqrt{npq}}$ وقتی $n \to +\infty$ یک توزیع نرمال استاندارد است.

✓ موقع تقریب زدن باید در محاسبه احتمالات از عمل تصحیح پیوستگی استفاده کرد. (کران پایین منهای $\frac{1}{2}$ کران بالا به علاوه $\frac{1}{2}$ می‌شود.)

* توزیع مجموع متغیرهای تصادفی نرمال:

$X_1, \dots, X_n \Rightarrow$ n متغیر تصادفی مستقل

$$X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

a_i مقادیر ثابت

مثال: نمرات درس های 4 واحدی ریاضی، 3 واحدی آمار و 3 واحدی زبان به ترتیب دارای توزیع نرمال با میانگین‌های 12، 14 و 15 و انحراف معیارهای 2، $\sqrt{2}$ و $\sqrt{2}$ می‌باشد.

الف) اگر 50 نفر این 3 درس را انتخاب کرده باشند، معدل چند نفرشان در این 3 درس از 14.5 بیشتر است؟

X_1 و X_2 و X_3 به ترتیب نمرات دروس ریاضی، آمار و زبان

Y معدل یک نفر در این 3 درس

$\Leftarrow$

$X_1 \sim N(12, 4)$

$X_2 \sim N(14, 2)$

$X_3 \sim N(15, 2)$

$$Y = \frac{4X_1 + 3X_2 + 3X_3}{10} \sim N(13.5, 1) \longrightarrow$$ توزیع ترکیب خطی از نرمال ها

$$P(Y > 14.5) = p\left(\frac{Y-\mu}{\sigma} > \frac{14.5-13.5}{1}\right) = p(Z > 1) = 1 - p(Z < 1)$$

$$= 1 - 0.8413 = 0.1587$$

پس تعداد $8 \simeq 7.935 = (50)(0.1587)$ نفر از 50 نفر معدل بیشتر از 14.5 دارند.

ب) اگر 40 نفر از دانشجویانی که این 3 درس را انتخاب کرده اند به طور مستقل انتخاب کنیم احتمال اینکه حداقل 15 نفر

از آن‌ها دارای معدلی کمتر از 13.5 در این 3 درس باشند را بیابید.

M : تعداد دانشجویان در بین ۴۰ نفر که دارای معدل کمتر از 13.5 هستند.

$$M \sim Bin(40, p)$$

$$p = P(Y < 13.5) = P\left(\frac{Y-\mu}{\sigma} < \frac{13.5-13.5}{1}\right) = P(Z<0) = \frac{1}{2}$$

از تقریب نرمال

$$\Rightarrow P(M \geq 15) = P(M > 14.5) = P\left(\frac{M-np}{\sqrt{npq}} > \frac{14.5-20}{\sqrt{10}}\right)$$

$$\cong P(Z > -1.74)$$

$$= 1 - P(Z < -1.74) = 1 - 0.0409$$

$$= 0.9591$$

خلاصه مطالب فصل توزیع های معروف

گسسته ها :

(۱) برنولی $X \sim Ber(p)$

$$P_X(x) = p^x (1-p)^{1-x} \quad x = 0, 1$$

$$E(X) = p$$

$$var(X) = pq \; ; \; q = 1-p$$

(۲) دوجمله‌ای $X \sim Bin(n, p)$

$$P_X(x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, \dots, n$$

$$E(X) = np$$

$$var(X) = npq$$

(۳) هندسی $X \sim Ge(p)$

$$P_X(x) = pq^{x-1} \quad x = 1, 2, \dots$$

$$E(X) = \frac{1}{p}$$

$$var(X) = \frac{q}{p^2}$$

(۴) دوجمله‌ای منفی $X \sim NB(r, p)$

$$P_X(x) = \binom{x-1}{r-1} p^r q^{x-r} \quad x = r, r+1, \dots$$

$$E(X) = r/p$$

$$var(X) = \frac{rq}{p^2}$$

(۵) فوق هندسی $X \sim HG(N, M, n)$

$$P_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$\max(0, n-N+M) \leq x \leq \min(n, M)$$

$$E(X) = \frac{nM}{N}$$

$$var(X) = \frac{nM}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

(۶) $X \sim Poi(\lambda)$

$$P_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \; ; \; x = 0, 1, \dots$$

$$E(X) = \lambda$$

$$var(X) = \lambda$$

(۷) یکنواخت گسسته $X \sim DU\{\theta\}$

$$P_X(x) = \frac{1}{\theta} \; ; \; x = x_1, \dots, x_\theta$$

$$E(X) = \frac{\sum x_i}{\theta}$$

$$var(X) = \frac{\sum (x_i - \mu)^2}{\theta}$$

نکات گسسته ها :

(۱) برای توزیع دوجمله‌ای و پواسن جداولی شامل احتمال های تجمعی ($P(X \leq r)$) وجود دارد.

(۲) اگر در توزیع دوجمله‌ای $n \to +\infty$ و $p \to 0$ آنگاه می‌توان آن را با توزیع پواسن تقریب زد یعنی :

$$\lambda = np \Rightarrow P_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

پیوسته :

(۱) یکنواخت پیوسته $X \sim U(a, b)$

$$f(x) = \frac{1}{b-a} \quad a < x < b$$

$$E(X) = \frac{a+b}{2}$$

$$var(X) = \frac{(a-b)^2}{12}$$

(۲) نمایی $X \sim exp(\theta)$

$$f(x) = \frac{1}{\theta} e^{-x/\theta} \quad x > 0$$

$$E(X) = \theta$$

$$var(X) = \theta^2$$

(۳) نرمال $X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad x \in R$$

$$E(X) = \mu$$

$$var(X) = \sigma^2$$

نکات پیوسته ها :

(۱) رابطه نمایی و پواسن ← اگر λ میانگین تعداد رخدادها در یک واحد زمان باشه ⇐

X: تعداد رخدادها در بازه زمانی $[0, t]$

Y: زمان رسیدن به اولین رخداد

$X \sim Poi(\lambda t)$ و $Y \sim exp(1/\lambda)$

(۲) برای نرمال در حالت $Z \sim N(0, 1)$ جدول احتمالات تجمعی ($P(Z < z)$) وجود دارد.

(۳) تقریب دوجمله‌ای توسط نرمال ← اگر $n \to +\infty$ و $p \to 1/2$ آنگاه $Z = \frac{X - np}{\sqrt{npq}} \sim N(0, 1)$ (دقت $X \sim Bin(n, p)$).

(۴) هر ترکیب خطی از نرمال ها خودش نرمال است.