فصل ۴ توزیع‌های گسسته و پیوسته

مقدمه

هر آزمایش آماری با فضای نمونه **گسسته** یا **پیوسته** وضعیتی را ایجاد می‌کند که می‌توانیم متغیر تصادفی (X) و تابع احتمال $(f(x))$ را بر اساس آن تعریف کنیم.

در این فصل سعی داریم ضمن بررسی آزمایش‌های مختلف گسسته و پیوسته، آن‌ها را دسته‌بندی کرده و برای هر کدام یک تابع احتمال معرفی کنیم و به تحلیل آن و یا محاسبه امید ریاضی، واریانس، تقریب و ... بپردازیم.

توابع گسسته	
۱) یکنواخت	$DU(x_1,x_2,\dots,x_N)$
۲) برنولی	$Bin(1,p)$ یا $B(1,p)$
۳) دوجمله‌ای	$Bin(n,p)$ یا $B(n,p)$
۴) چندجمله‌ای	$MN(n_1,n_2,\dots,p_1,p_2,\dots)$
۵) دوجمله‌ای منفی	$NB(r,p)$
۶) هندسی	$Ge(p)$ یا $G(p)$
۷) فوق هندسی	$HG(N,k,n)$
۸) پواسون	$P(\lambda)$

توابع پیوسته	
۱) یکنواخت	$U(a,b)$
۲) نمایی	$E(\lambda)$ یا $Exp(\lambda)$
۳) گاما	$G(r,\lambda)$ یا $\Gamma(r,\lambda)$
۴) نرمال	$N(\mu,\sigma^2)$
۵) (کای - دو یا خی دو)	$\chi^2_{(n)}$
۶) t ـ استیودنت	$t_{(n)}$
۷) کوشی	$C(\mu,\sigma)$
۸) F (فیشر)	F_{n_1,n_2}

توزیع یکنواخت گسسته (Discrete Uniform Distribution)

مسئله (درک مطلب): اگر متغیر تصادفی X نتیجه پرتاب یک تاس به صورت زیر باشد، امید ریاضی و واریانس X کدام است؟

$$f(x)=\frac{1}{6} \quad ; \quad x=1,2,\ldots,6$$

اگر نتیجه انجام آزمایشی مانند x بتواند به یکی از N وضعیت هم‌شانس $x_1,\ldots,x_N$ با احتمال یکسان منجر شود، آن‌گاه احتمال وقوع هر یک از وضعیت‌های x_i به صورت زیر برابر $\frac{1}{N}$ است:

$$\sum_{i=1}^{N} P(x_i)=1 \rightarrow P(x_1)+P(x_2)+\ldots+P(x_N)=1 \xrightarrow{P(x_1)=\ldots=P(x_N)} P(x_i)=\frac{1}{N} \;;\; i=1,2,..,N$$

تعریف: درصورتی‌که متغیر تصادفی گسسته X بتواند N مقدار مختلف $(x_N,\ldots,x_2,x_1)$ را با احتمال‌های برابر $\frac{1}{N}$ اختیار کند، آن‌گاه X دارای توزیع یکنواخت گسسته است.

به عبارت دیگر:

$$\begin{array}{c|cccc} X & x_1 & x_2 & \ldots & x_N \\ \hline f(x) & \frac{1}{N} & \frac{1}{N} & \ldots & \frac{1}{N} \end{array} \longrightarrow f(x)=\frac{1}{N} \;;\; x=x_1,x_2,\ldots,x_N$$

پارامترهای توزیع

اگر متغیر تصادفی X دارای توزیع یکنواخت گسسته با N مقدار از X باشد، پارامتر آن N است و به صورت $X \sim DU(x_1,x_2,\ldots,x_N)$ نمایش داده می‌شود.

تابع احتمال	$P(x)=\frac{1}{N}$ $x=x_1,\ldots,x_N$	احتمال وقوع هر یک از مقادیر x_i
امید ریاضی (میانگین)	$\mu_X=E(X)=\frac{\sum x_i}{N}$	متوسط مقادیر x_i
واریانس	$\sigma_X^2=\frac{\sum(x_i-\mu)^2}{N}=\frac{\sum x_i^2}{N}-\left(\frac{\sum x_i}{N}\right)^2$	واریانس مقادیر x_i
انحراف معیار	$\sigma_X=\sqrt{\frac{\sum(x_i-\mu)^2}{N}}$	
تابع مولد گشتاور	$M_X(t)=\frac{e^t(1-e^{Nt})}{N(1-e^t)}$	

حالات خاص

الف) اگر متغیر تصادفی یکنواخت گسسته $x = x_1,...,x_N$ به صورت $x = 1,2,...,N$ در نظر گرفته شود، آن‌گاه:

تابع احتمال	امید ریاضی (میانگین)	واریانس	انحراف معیار
$P(x)=\frac{1}{N}$ $x=1,2,...,N$	$\mu_X = E(X)=\frac{N+1}{2}$	$\sigma_X^2=\frac{N^2-1}{12}$	$\sigma_X=\sqrt{\frac{N^2-1}{12}}$

ب) اگر متغیر تصادفی یکنواخت گسسته $x = x_1,x_2,...,x_N$ به صورت یک تصاعد حسابی (با قدرنسبت d) به صورت $x = x_1, x_1+d,...,x_1+(N-1)d$ مطرح شود، آن‌گاه:

تابع احتمال	امید ریاضی (میانگین)	واریانس	انحراف معیار
$P(x)=\frac{1}{N}$ $x=x_1,x_1+d,..,x_1+(N-1)d$	$\mu_X=E(X)=\frac{x_1+x_N}{2}$	$\sigma_X^2=d^2\cdot\frac{N^2-1}{12}$	$\sigma_X=\sqrt{d^2\cdot\frac{N^2-1}{12}}$

نکته: تابع توزیع تجمعی متغیر تصادفی یکنواخت گسسته به صورت زیر است:

$$F_X(\alpha)=P(X\le\alpha)=\frac{\lfloor\alpha\rfloor}{N} \quad ; \quad \lfloor\ \rfloor: \text{جزء صحیح}$$

حل مسئله (درک مطلب):

با توجه به حالت خاص (الف) خواهیم داشت:

$$\mu=\frac{N+1}{2}=\frac{6+1}{2}=3.5$$

$$\sigma^2=\frac{N^2-1}{12}=\frac{6^2-1}{12}=\frac{35}{12}$$

مثال اگر تابع یکنواخت متغیر تصادفی گسسته X به صورت زیر باشد:

$$\begin{cases} f(x)=\frac{1}{5} \\ x=2,6,10,14,18 \end{cases}$$

میانگین و واریانس X کدام است؟

۱) 5 , 10   ۲) 32 , 10   ۳) 20 , 64   ۴) 16 , 8

حل: گزینه ۲ درست است.

همان‌طورکه دیده می‌شود مشاهدات، یک تصاعد حسابی با $x_1=2,\ x_N=18,\ N=5,\ d=4$ را تشکیل می‌دهند، بنابراین با توجه به حالت خاص (ب) داریم:

$$\mu=\frac{x_1+x_N}{2}=\frac{2+18}{2}=10$$

$$\sigma^2=d^2\cdot\frac{N^2-1}{12}=4^2\cdot\frac{5^2-1}{12}=32$$

آزمایش برنولی

اگر هر بار انجام آزمایش، مستقل از نتایج به‌دست‌آمده از دفعات قبلی، فقط و فقط به یکی از دو پیشامد «موفقیت» یا «شکست» با احتمال ثابت (p) یا $(q=1-p)$ منجر شود، آن‌گاه به هر یک از این آزمایش‌ها «آزمایش برنولی» و به متغیر تصادفی X، **تعداد موفقیت** در هر بار انجام آزمایش $(x=0,1)$، «توزیع برنولی» نسبت داده می‌شود.

نکته:

«موفقیت»: وقوع پیشامد مطلوب (مورد نظر سؤال) با احتمال ثابت p

«شکست»: عدم وقوع پیشامد مطلوب (مورد نظر سؤال) با احتمال ثابت $q=1-p$

نکته: آزمایش‌هایی که فقط دو پیامد دارند، مانند (برد و باخت)، (پسر و دختر)، (شیر و خط)، (زوج و فرد)، (معیوب و سالم) با توجه به پیشامد مطلوب (مورد نظر سؤال) ممکن است در هر زوج یکی «موفقیت» و دیگری «شکست» باشد. به عبارت دیگر:

«موفقیت» = (برد یا باخت)، (پسر یا دختر)، (شیر یا خط)، ...

«شکست» = (باخت یا برد)، (دختر یا پسر)، (خط یا شیر)، ...

توزیع‌های حاصل از آزمایش برنولی

هرگاه یک آزمایش برنولی با احتمال موفقیت p را انجام دهیم یا آن را با توجه به هدف مورد نظر به صورت مستقل تکرار کنیم، یکی از چهار توزیع زیر حاصل می‌شود:

۱- توزیع برنولی (دونقطه‌ای)

آزمایش برنولی را یک بار انجام می‌دهیم. هدف، بررسی تعداد موفقیت‌های حاصل است که می‌تواند 0 یا 1 باشد.

مانند، تعداد خط در 1 بار پرتاب سکه.

۲- توزیع دوجمله‌ای (باینم)

آزمایش برنولی را n بار به صورت مستقل تکرار می‌کنیم. هدف، بررسی تعداد موفقیت‌های حاصل است که می‌تواند $0, 1, \dots, n$ باشد.

مانند، تعداد خط در n بار پرتاب سکه.

* توزیع برنولی حالت خاصی از توزیع دوجمله‌ای است، وقتی $n=1$ باشد.

۳- توزیع دوجمله‌ای منفی (پاسکال)

آزمایش برنولی را به صورت مستقل آن‌قدر تکرار می‌کنیم تا به rامین موفقیت برسیم. هدف، بررسی تعداد آزمایش لازم است که می‌تواند $r, r+1, \dots$ باشد.

مانند، تعداد پرتاب سکه برای رسیدن به rامین خط.

۴- توزیع هندسی

آزمایش برنولی را به صورت مستقل آن‌قدر تکرار می‌کنیم تا به اولین موفقیت برسیم. هدف، بررسی تعداد آزمایش لازم است که می‌تواند $1, 2, \dots$ باشد.

مانند، تعداد پرتاب سکه برای رسیدن به اولین خط.

* توزیع هندسی حالت خاصی از توزیع دوجمله‌ای منفی است، وقتی $r=1$ باشد.

احتمال ثابت در هر آزمایش

شرط لازم برای هر آزمایش برنولی آن است که وقوع پیشامد موفقیت یا شکست مستقل از نتیجه آزمایش‌های قبلی، ثابت باشد.

در هر یک از شرایط زیر، احتمال وقوع پیشامد ثابت است و می‌توانند به عنوان یک آزمایش برنولی در نظر گرفته شوند.

۱ـ هرگاه احتمال یا نسبت ثابتی داده شود (نمونه‌گیری با جایگذاری یا بدون جایگذاری).

برای مثال:

ـ 0.20 کالاهای یک کارخانه معیوب است.

ـ از هر 100 کالا، 20 کالا معیوب است.

ـ با احتمال 0.20 پرتاب‌های یک بازیکن به هدف اصابت می‌کند.

۲ـ هرگاه از یک جامعه **محدود** N **تایی**، نمونه‌گیری **با جایگذاری** انجام شود.

برای مثال، از یک جامعه با 4 کالای سالم و 6 کالای معیوب، نمونه‌ای با جایگذاری انتخاب می‌کنیم (بعد از هر بار انتخاب نمونه، آن را دوباره به جعبه برمی‌گردانیم).

احتمال معیوب بودن در هر بار نمونه‌گیری $= \frac{6}{10}$ (ثابت)

احتمال سالم بودن در هر بار نمونه‌گیری $= \frac{4}{10}$ (ثابت)

✓ **دقت کنید!**

اولاً، در سؤالات به طور پیش‌فرض انتخاب بدون جایگذاری است، بنابراین برای جامعه محدود باید در صورت سؤال به‌گونه‌ای «با جایگذاری بودن انتخاب» ذکر شود، در غیر این صورت احتمال ثابت نخواهد بود.

ثانیاً، اگر از جامعه محدود مانند مثال بالا نمونه‌گیری بدون جایگذاری انجام شود، احتمال معیوب یا سالم بودن در هر بار آزمایش دیگر ثابت نیست و وابسته به آزمایش‌های قبلی است و دیگر نمی‌تواند به عنوان آزمایش برنولی در نظر گرفته شود.

انتخاب در بار اول (بدون جایگذاری)		انتخاب در بار دوم (بدون جایگذاری)
$\frac{6}{10} =$ احتمال معیوب بودن	$\xrightarrow[\text{در انتخاب اول}]{\text{خارج کردن معیوب}}$	$\frac{4}{9} =$ سالم و $\frac{5}{9} =$ معیوب
$\frac{4}{10} =$ احتمال سالم بودن	$\xrightarrow[\text{در انتخاب اول}]{\text{خارج کردن سالم}}$	$\frac{3}{9} =$ سالم و $\frac{6}{9} =$ معیوب

۳ـ در هر بار پرتاب سکه یا تاس، احتمال ثابتی داریم؛ برای مثال:

$\frac{1}{2} =$ خط آمدن سکه $\frac{1}{2} =$ شیر آمدن سکه

$\frac{3}{6} =$ فرد آمدن تاس $\frac{3}{6} =$ زوج آمدن تاس

۴ـ پیشامد به دنیا آمدن پسر یا دختر در هر بار زایمان مادر با احتمال ثابتی اتفاق می‌افتد.

پیش‌فرض : $\frac{1}{2} =$ پسر $\frac{1}{2} =$ دختر (دختر و پسر هم‌شانس)

$\frac{3}{4} =$ پسر $\frac{1}{4} =$ دختر (پیشامد پسر 3 برابر دختر)

توزیع برنولی (Bernoulli Distribution)

تعریف: هرگاه آزمایش برنولی با احتمال موفقیت p را یک بار انجام دهیم، آن‌گاه متغیر تصادفی « X : **تعداد موفقیت** در انجام 1 **بار** آزمایش برنولی» $(x = 0,1)$ دارای توزیع **برنولی** است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع برنولی باشد، پارامتر آن p است و به صورت $X \sim Bin(1,p)$ یا $X \sim B(1,p)$ نمایش داده می‌شود.

نکته: ازآنجاکه در توزیع برنولی، تعداد موفقیت (X) در انجام 1 بار آزمایش برنولی فقط می‌تواند یکی از دو وضعیت «صفر موفقیت» یا « 1 موفقیت» را داشته باشد، به آن توزیع دونقطه‌ای نیز می‌گویند:

X : تعداد موفقیت در 1 بار انجام آزمایش برنولی	0	1
$f(x) = P(x)$	$q = 1 - p$	p

احتمال x موفقیت در یک آزمایش برنولی	$f(x) = P(x) = p^x q^{1-x} \quad ; \quad x = 0,1$	**تابع احتمال**
میانگین(امید) تعداد موفقیت در یک آزمایش برنولی	$E(X) = \mu = p$	**امید ریاضی (میانگین)**
واریانس تعداد موفقیت در یک آزمایش برنولی	$\sigma_X^2 = pq$	**واریانس**
	$\sigma_X = \sqrt{pq}$	**انحراف معیار**
	$M_X(t) = pe^t + q$	**تابع مولد گشتاور**

نکته: تابع توزیع تجمعی متغیر تصادفی برنولی به صورت زیر است:

$$F_X(\alpha) = P(X \le \alpha) = \begin{cases} 0 & \alpha < 0 \\ 1 - p = q & 0 \le \alpha < 1 \\ 1 & \alpha \ge 1 \end{cases}$$

✓ دقت کنید!

۱- ازآنجاکه $p + q = 1$ است، می‌توان $f(x) = p^x (1-p)^{1-x}$ را در نظر گرفت.

۲- در بعضی منابع، احتمال موفقیت را با نماد θ نمایش می‌دهند، درنتیجه داریم: $f(x) = \theta^x (1-\theta)^{1-x}$.

نکات مهم توزیع برنولی

۱- اگر X دارای توزیع برنولی با احتمال موفقیت p باشد، آن‌گاه $Y = 1 - X$ دارای توزیع برنولی با احتمال موفقیت $1-p = q$ است.

۲- اگر X دارای توزیع برنولی با احتمال موفقیت p باشد، آن‌گاه $Y = X^k$ نیز دارای توزیع برنولی با همان احتمال موفقیت p است.

۳- اگر $X_1, X_2, \ldots, X_n$ متغیرهای مستقل و هم‌توزیع برنولی با احتمال موفقیت p باشند، آن‌گاه مینیمم آن‌ها دارای توزیع برنولی با احتمال موفقیت p^n است.

۴- اگر $X_1, X_2, \ldots, X_n$ متغیرهای مستقل و هم‌توزیع برنولی با احتمال موفقیت p باشند، آن‌گاه ماکزیمم آن‌ها دارای توزیع برنولی با احتمال موفقیت $1-(1-p)^n = 1-q^n$ است.

۵- اگر $X_1, X_2, \ldots, X_n$ متغیرهای مستقل و هم‌توزیع برنولی با احتمال موفقیت p باشند، آنگاه ضرب این n متغیر با هر توان دلخواه برای تک تک آن‌ها دارای توزیع برنولی با احتمال موفقیت p^n است.

	شرایط	متغیر	توزیع متغیر
۱	$X \sim Bin(1,p)$	$Y = 1 - X$	$Y \sim Bin(1,q)$
۲	$X \sim Bin(1,p)$	$Y = X^k$	$Y \sim Bin(1,p)$
۳	$X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} Bin(1,p)$	$Y = \min\{X_1, X_2, \ldots, X_n\}$	$Y = Bin(1, p^n)$
۴	$X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} Bin(1,p)$	$Y = \max\{X_1, X_2, \ldots, X_n\}$	$Y = Bin(1, 1-q^n)$
۵	$X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} Bin(1,p)$	$Y = X_1^a \times X_2^b \times \ldots \times X_n^z$	$Y \sim Bin(1, p^n)$

مثال فرض کنید $P(X_i = x_i) = \binom{1}{x_i}\theta^{x_i}(1-\theta)^{1-x_i}$; $i=1,2,\ldots,n$. در آن صورت $Var\left[\prod_{i=1}^{n} X_i\right]$ عبارت است از:

۱) $p(1-p)$ ۲) $p(1-p^n)$ ۳) $p^n(1-p^n)$ ۴) $p^n(1-p)^n$

حل: گزینه ۳ درست است.

با توجه به نکته ۵، ضرب X_iها $\left(\prod_{i=1}^{n} X_i = X_1.X_2\ldots X_n\right)$ دارای توزیع برنولی با احتمال موفقیت p^n است؛ بنابراین واریانس آن برابر است با:

$$Var\left(\prod_{i=1}^{n} X_i\right) = p^n\left(1-p^n\right)$$

توزیع دوجمله‌ای (Binominal Distribution)

مسئله ۱ (درک مطلب): در خانواده‌ای با 3 فرزند، احتمال آنکه حداقل 1 فرزند پسر باشد، چقدر است؟

مسئله ۲ (درک مطلب): در 3 بار پرتاب یک تاس سالم، احتمال آنکه حداکثر 1 بار عدد 4 یا 6 ظاهر شود، چقدر است؟

مسئله ۳ (درک مطلب): از جعبه‌ای با 8 کالای سالم و 2 کالای معیوب، 3 کالا **با جایگذاری** انتخاب می‌کنیم. احتمال آنکه 1 کالا معیوب باشد، چقدر است؟

مسئله ۴ (درک مطلب): اگر تابع چگالی احتمال $f(x)$ به صورت $f(x)=2x\ ;\ 0<x<1$ باشد، درصورتی‌که 3 بار نقطه‌ای از فاصله $0<x<1$ انتخاب کنیم، احتمال آنکه 2 بار نقطه در فاصله $\left(0<x<\frac{1}{2}\right)$ باشد، کدام است؟

مقدمه: در هنگام انجام آزمایش‌های مستقل برنولی که نتیجه هر آزمایش در آن به یکی از دو وضعیت موفقیت (با احتمال p) و یا شکست (با احتمال q) منجر می‌شود، اگر به دنبال محاسبه احتمال x **موفقیت** در n **بار تکرار** (نمونه) آزمایش باشیم، باید به صورت زیر عمل کنیم:

برای مثال، اگر بخواهیم احتمال وقوع $x=2$ موفقیت را در $n=3$ بار تکرار آزمایش مستقل برنولی بررسی کنیم:

آزمایش / وضعیت‌های ممکن	اول	دوم	سوم	
وضعیت اول	p ×	p ×	q	$=p^2q^1$
				+
وضعیت دوم	p ×	q ×	p	$=p^2q^1$
				+
وضعیت سوم	q ×	p ×	p	$=p^2q^1$

$$\rightarrow 3p^2q^1$$

همان‌طور که دیده می‌شود:

۱- وقوع 2 موفقیت در 3 آزمایش، 3 حالت دارد که باید آن را محاسبه کنیم، برای این کار می‌توانیم از $\binom{n}{x}=\binom{3}{2}=3$، که همان ترکیب x از n است، استفاده کنیم تا تعداد حالات وقوع x موفقیت در n آزمایش را به دست آوریم.

۲- احتمال وقوع $x=2$ موفقیت در هر حالت برابر است با $p^xq^{n-x}=p^2q^1$ که برای محاسبه احتمال کل باید در تعداد حالات ضرب شود، بنابراین:

$$P(\text{آزمایش } n=3 \text{ موفقیت در } X=2)=\binom{n}{x}p^xq^{n-x}=\binom{3}{2}p^2q^1=3p^2q$$

تعریف: هرگاه یک آزمایش مستقل برنولی را n **بار تکرار** کنیم، (یک نمونه مستقل n تایی از آزمایش برنولی انتخاب کنیم) آن‌گاه متغیر تصادفی« X : **تعداد موفقیت در n بار آزمایش مستقل برنولی**» $(x=0,1,\dots,n)$ دارای توزیع **دوجمله‌ای (باینم)** خواهد بود.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع دوجمله‌ای باشد، پارامترهای آن n و p است و به صورت $X\sim Bin(n,p)$ یا $X\sim B(n,p)$ نمایش داده می‌شود.

احتمال x بار موفقیت در n بار آزمایش برنولی	$P(x)=\binom{n}{x}p^x q^{n-x}$ $x=0\,,1,2,\dots,n$	**تابع احتمال**
متوسط تعداد موفقیت مورد انتظار در n بار آزمایش مستقل برنولی	$E(X)=np$	**امید ریاضی (میانگین)**
واریانس تعداد موفقیت در n بار آزمایش مستقل برنولی	$\sigma_X^2=npq$	**واریانس**
	$\sigma_X=\sqrt{npq}$	**انحراف معیار**
	$M_X(t)=\left(pe^t+q\right)^n$	**تابع مولد گشتاور**

نکته: همان‌طورکه در مقدمه مطرح شد، $C_n^x=\binom{n}{x}$ حالات وقوع x موفقیت در n آزمایش را محاسبه می‌کند که برابر است با:

$$C_n^x=\binom{n}{x}=\frac{n!}{x!(n-x)!}$$

✓ **دقت کنید!**

احتمال وقوع دقیقاً x موفقیت (هر موفقیت با احتمال p) در n بار آزمایش مستقل برنولی به مفهوم داشتن $n-x$ شکست (هر شکست با احتمال q) نیز است، به طوری که:

$$P(x)=\binom{n}{x}p^x q^{n-x}=\frac{n!}{x!(n-x)!}p^x q^{n-x}$$

مثال در یک توزیع دوجمله‌ای خاص با $p=0.4$، مقدار $\frac{7!}{3!4!}(0.4)^3(0.6)^4$ نشان‌دهنده احتمال کدام‌یک از این حالات است؟

۱) دقیقاً 3 موفقیت در 7 آزمایش ...

۱) دقیقاً 4 موفقیت در 7 آزمایش
۲) 3 موفقیت یا بیشتر در 7 آزمایش
۳) دقیقاً 3 موفقیت در 7 آزمایش
۴) 4 موفقیت یا کمتر در 7 آزمایش

حل: گزینه ۳ درست است.

$$\begin{cases}P(X=3)=\binom{7}{3}(0.4)^3(0.6)^4=\dfrac{7!}{3!4!}(0.4)^3(0.6)^4\\ n=7\ ,\ x=3\ ,\ p=0.4\ ,\ q=0.6\end{cases}$$

این احتمال نشان‌دهنده احتمال دقیقاً 3 موفقیت یا دقیقاً 4 شکست در 7 بار تکرار آزمایش است.

محاسبه احتمال

برای محاسبه احتمال x موفقیت در n بار آزمایش مستقل برنولی (احتمال موفقیت، p و احتمال شکست، q) به صورت زیر عمل می‌کنیم:

الف) n، p و x را مشخص می‌کنیم:

n: تعداد کل آزمایش

p: احتمال وقوع وضعیت مطلوب در هر آزمایش (موفقیت)

x: تعداد مورد نظر وقوع برای وضعیت مطلوب در n بار آزمایش

ب) با استفاده از رابطه زیر احتمال را محاسبه می‌کنیم:

$$P(x)=(p+q)^n=\binom{n}{x}p^x q^{n-x}$$

با توجه به آنکه همواره $\sum P(x)=1$ (مجموع مقادیر تابع احتمال متغیر تصادفی گسسته برابر با 1 است)، برای تابع احتمال دوجمله‌ای، رابطه $\sum_{x=0}^{n} P(x)=\sum_{x=0}^{n}\binom{n}{x}p^x q^{n-x}=1$ برقرار است و داریم:

$$\underbrace{\binom{n}{0}p^0q^n}_{P(x=0)}+\underbrace{\binom{n}{1}p^1q^{n-1}}_{P(x=1)}+\underbrace{\binom{n}{2}p^2q^{n-2}}_{P(x=2)}+\cdots+\underbrace{\binom{n}{n}p^nq^0}_{P(x=n)}=1$$

درنتیجه برای مثال:

$$P(X=0)=\binom{n}{0}p^0q^{n-0}=q^n$$

(احتمال عدم موفقیت در n بار آزمایش)

$$P(X=1)=\binom{n}{1}p^1q^{n-1}=npq^{n-1}$$

(احتمال وقوع 1 موفقیت در n بار آزمایش)

$$P(X=2)=\binom{n}{2}p^2q^{n-2}$$

(احتمال وقوع 2 بار موفقیت در n بار آزمایش)

$$P(X\le 1)=P(X=0)+P(X=1)$$

(احتمال وقوع حداکثر 1 موفقیت در n بار آزمایش)

$$P(X>1)=1-P(X\le 1)=1-P(X=0)-P(X=1)$$

(احتمال وقوع بیش از 1 موفقیت در n بار آزمایش)

$$P(X\ge 1)=1-P(X<1)=1-P(X=0)=1-q^n$$

(احتمال وقوع حداقل 1 موفقیت در n بار آزمایش)

حل مسئله۱ (درک مطلب):

در این سؤال به ترتیب $n=3$، $p=\frac{1}{2}$ (پسر) و $x\ge 1$ است و داریم:

$$P(X\ge 1)=1-P(X=0)=1-\binom{n}{0}p^0q^{n-0}=1-\binom{3}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3=1-\frac{1}{8}=\frac{7}{8}$$

حل مسئله۲ (درک مطلب):

در این سؤال به ترتیب $n=3$، $p=\frac{2}{6}=\frac{1}{3}$ و $x\le 1$ است و داریم:

$$P(X\le 1)=P(X=0)+P(X=1)=\binom{n}{0}p^0q^{n-0}+\binom{n}{1}p^1q^{n-1}=\binom{3}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^3+\binom{3}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2=\frac{8}{27}+\frac{12}{27}=\frac{20}{27}$$

حل مسئله ۳ (درک مطلب):

با توجه به نوع انتخاب (با جایگذاری) احتمال وقوع سالم یا معیوب بودن کالا در هر بار انتخاب ثابت و به ترتیب برابر $\frac{8}{10}$ و $\frac{2}{10}$ است، بنابراین شرایط آزمایش برنولی وجود دارد و داریم:

$$\begin{cases} P(X=1)=\binom{3}{1}\left(\frac{2}{10}\right)^1\left(\frac{8}{10}\right)^2=0.384 \\ p=P(\text{معیوب})=\frac{2}{10}\ ,\ q=\frac{8}{10}\ ,\ n=3 \end{cases}$$

حل مسئله ۴ (درک مطلب):

در این سؤال $n=3$ و $x=2$ است و داریم:

$$p=P\left(0<X<\frac{1}{2}\right)=\int_0^{\frac{1}{2}}2x\,dx=\left[x^2\right]_0^{\frac{1}{2}}=\frac{1}{4}\ \rightarrow\ q=\frac{3}{4}$$

در نتیجه:

$$P(X=2)=\binom{3}{2}p^2q^1=3\times\left(\frac{1}{4}\right)^2\times\frac{3}{4}=\frac{9}{64}$$

مثال ۱ امید و واریانس تعداد 4 یا 5 آمدن در 120 بار پرتاب یک تاس سالم چقدر است؟

۱) $\mu_X=40\ ,\ \sigma_X^2=\frac{80}{3}$
۲) $\mu_X=20\ ,\ \sigma_X^2=\frac{50}{3}$
۳) $\mu_X=20\ ,\ \sigma_X^2=\frac{80}{3}$
۴) $\mu_X=40\ ,\ \sigma_X^2=\frac{50}{3}$

حل: گزینه ۱ درست است.

ازآنجاکه آزمایش برنولی با دو نتیجه (ظهور یا عدم ظهور 4 یا 5) به تعداد 120 بار تکرار می‌شود، توزیع احتمال دوجمله‌ای است و درنتیجه:

$$\begin{cases} E(X)=np=120\times\frac{1}{3}=40 \\ \sigma_X^2=npq=120\times\frac{1}{3}\times\frac{2}{3}=\frac{80}{3} \\ n=120 \\ p=P(X=4\text{ or }5)=\frac{2}{6}=\frac{1}{3}\ ,\ q=1-p=\frac{2}{3} \end{cases}$$

مثال ۲ اگر در یک توزیع دوجمله‌ای، میانگین 4 و انحراف معیار $\sqrt{2}$ باشد، مقدار $P(X>0)$ کدام است؟

۱) 1
۲) $1-\left(\frac{1}{2}\right)^8$
۳) $\left(\frac{1}{2}\right)^8$
۴) $1-\left(\frac{1}{3}\right)^4$

حل: گزینه ۲ درست است.

اولاً، اگر $\sigma=\sqrt{2}$ باشد، آن‌گاه $\sigma^2=2$ است. ثانیاً، در توزیع دوجمله‌ای $\mu=np$ و $\sigma^2=npq$ است، بنابراین:

$$\sigma^2=npq=2\ \xrightarrow{\mu=np=4}\ 4q=2\ \rightarrow\ q=\frac{1}{2}\ ,\ p=\frac{1}{2}$$

$$\mu=np=4\ \xrightarrow{p=\frac{1}{2}}\ n\times\frac{1}{2}=4\ \rightarrow\ n=8$$

همواره در توزیع دوجمله‌ای داریم:

$$P(X=0)+P(X=1)+\ldots+P(X=n)=1$$

بنابراین:

$$P(X>0)=P(X\geq 1)=1-P(X=0)=1-\binom{n}{0}p^0q^n=1-q^n=1-\left(\frac{1}{2}\right)^8$$

دقت کنید اگر در صورت مسئله احتمال $P(X\geq 0)$ مد نظر بود، آن‌گاه مقدار احتمال برابر 1 و گزینه ۱ درست بود.

مثال ۳ کمیت تصادفی X توسط تابع توزیع (تابع تجمعی احتمال):

$$F_X(x)=\begin{cases}0 & x<0\\ x^2 & 0\le x\le 1\\ 1 & 1<x\end{cases}$$

بیان شده است. احتمال اینکه در 4 آزمایش مستقل، X درست سه بار مقادیر خود را در فاصله $(0.25,0.75)$ اختیار کند، کدام است؟

۱) 0.25 ۲) 0.35 ۳) 0.5 ۴) 0.75

حل: گزینه ۱ درست است.

ابتدا احتمال این را که X عددی در فاصله $(0.25,0.75)$ باشد، محاسبه می‌کنیم:

$$P(\alpha<X<\beta)=F(\beta)-F(\alpha)\to P(0.25<X<0.75)=F(0.75)-F(0.25)=\left(\frac{3}{4}\right)^2-\left(\frac{1}{4}\right)^2=\frac{1}{2}$$

حال X دارای توزیع دوجمله‌ای با $p=\frac{1}{2}$ و $n=4$ است.

$$P(X=3)=\binom{n}{x}p^xq^{n-x}=\binom{4}{3}\left(\frac{1}{2}\right)^4=\frac{4}{16}=\frac{1}{4}=0.25$$

مثال ۴ پنج بار عددی را به تصادف در بازه $(0,1)$ انتخاب می‌کنیم. احتمال آنکه دقیقاً سه بار عددی بیش از 0.6 انتخاب شود؟

۱) $10(0.4)^3(0.6)^2$ ۲) $(0.4)^3(0.6)^2$ ۳) $10(0.6)^3(0.4)^2$ ۴) $(0.6)^3(0.4)^2$

حل: گزینه ۱ درست است.

ابتدا احتمال این را که عددی بیش از 0.6 باشد، به دست می‌آوریم:

$$X\sim U(0,1)\to f(x)=\frac{1}{1-0}=1\ ;\ 0<x<1$$

$$p=P(X>0.6)=\int_{0.6}^{1}f(x)dx=\int_{0.6}^{1}1.dx=[x]_{0.6}^{1}=1-0.6=0.4$$

حال با احتمال $p=0.4$ عددی بیش از 0.6 انتخاب می‌شود؛ بنابراین توزیع تعداد موفقیت در $n=5$ انتخاب دوجمله‌ای است با $p=0.4, q=0.6, n=5$

$$P(Y=3)=\binom{n}{y}p^yq^{n-y}=\binom{5}{3}(0.4)^3(0.6)^2=10(0.4)^3(0.6)^2$$

مثال ۵ تابع چگالی احتمال $f(x)$ به صورت $f(x)=2x\ ;\ 0<x<1$ است. اگر 3 بار نقطه‌ای از فاصله $0<x<1$ انتخاب کنیم، احتمال آنکه 2 بار نقطه در فاصله $\left(0<X<\frac{1}{2}\right)$ باشد کدام است؟

۱) $\frac{9}{64}$ ۲) $\frac{3}{16}$ ۳) $\frac{1}{64}$ ۴) $\frac{1}{3}$

حل: گزینه ۱ درست است.

در این سؤال $n=3$ و $x=2$ و داریم:

$$p=P\left(0<X<\frac{1}{2}\right)=\int_0^{\frac{1}{2}}2x\,dx=\left[x^2\right]_0^{\frac{1}{2}}=\frac{1}{4}\ \to\ q=\frac{3}{4}$$

درنتیجه:

$$P(X=2)=\binom{3}{2}p^2q^1=3\times\left(\frac{1}{4}\right)^2\times\frac{3}{4}=\frac{9}{64}$$

مثال ۶ احتمال اینکه در 2 بار پرتاب، حداقل یک پرتاب به هدف اصابت کند، مساوی 0.36 است. احتمال اصابت در هر بار پرتاب چقدر است؟

۱) 0.2 ۲) 0.8 ۳) 0.5 ۴) 0.3

حل: گزینه ۱ درست است.

اولاً، ازآنجاکه احتمال اصابت در هر پرتاب (p) خواسته شده، احتمال ثابت و آزمایش برنولی است.

ثانیاً، «X: تعداد اصابت در 2 بار پرتاب» دارای توزیع دوجمله‌ای و $x = 0,1,2$ است.

حال با توجه به آنکه احتمال حداقل یک اصابت در 2 بار پرتاب 0.36 شده است، داریم:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{2}{0} p^0 q^2 = 0.36 \rightarrow q^2 = 0.64 \rightarrow q = 0.8, p = 0.2$$

نمایش خاص توزیع دوجمله‌ای

هرگاه متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامترهای n و $p = q = \frac{1}{2}$ باشد، تابع احتمال آن به صورت زیر است:

$$\boxed{f(x) = \binom{n}{x}\left(\frac{1}{2}\right)^n = \frac{\binom{n}{x}}{2^n} \quad ; \quad x = 0,1,2,\ldots,n}$$

اثبات:

$$f(x) = \binom{n}{x} p^x q^{n-x} \xrightarrow{p = q = \frac{1}{2}} \binom{n}{x}\left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} = \binom{n}{x}\left(\frac{1}{2}\right)^n = \frac{\binom{n}{x}}{2^n}$$

مثال ۷ شانه تخم‌مرغ‌های 6 تایی به قیمت 600 تومان و شانه‌های تاریخ‌گذشته آن به قیمت 400 تومان عرضه می‌شود. اگر تابع احتمال تخم‌مرغ‌های سالم در شانه تاریخ‌گذشته به صورت زیر باشد، احتمال آنکه خرید آن مقرون به صرفه باشد، چقدر است؟

$$\begin{cases} f(x) = \binom{6}{x}\frac{1}{64} & ; \quad x = 0, 1, \ldots, 6 \\ f(x) = 0 & ; \quad \text{برای سایر مقادیر } x \end{cases}$$

۱) $\frac{7}{64}$ ۲) $\frac{16}{64}$ ۳) $\frac{28}{64}$ ۴) $\frac{39}{64}$

حل: گزینه ۱ درست است.

دقت کنید که اگر هر شانه 6 تایی سالم را 600 تومان بخریم، برای هر عدد تخم مرغ، 100 تومان هزینه کرده‌ایم. حال اگر یک شانه 6 تایی تاریخ‌گذشته را 400 تومان بخریم، درصورتی‌که 4 تخم مرغ سالم داشته باشد، برای هر دانه تخم مرغ، 100 تومان هزینه کرده‌ایم که این هزینه فرقی با هزینه خرید یک شانه سالم ندارد و مقرون به صرفه نیست.

پس برای آنکه خرید یک شانه تاریخ گذشته مقرون به صرفه باشد، باید بیش از 4 تخم مرغ سالم (حداقل 5 سالم) داشته باشد:

$$\begin{cases} P(\text{مقرون به صرفه بودن}) = P(X \geq 5) = P(X=5) + P(X=6) = \binom{6}{5}\frac{1}{64} + \binom{6}{6}\frac{1}{64} = \frac{6}{64} + \frac{1}{64} = \frac{7}{64} \\ X: \text{تعداد تخم مرغ سالم در هر شانه تاریخ‌گذشته} \end{cases}$$

با کمی دقت متوجه می‌شوید که تابع $f(x)$ همان نمایش خاص از توزیع دوجمله‌ای با پارامترهای $n = 6$ و $p = \frac{1}{2}$ است که:

$$\begin{cases} f(x) = \binom{6}{x}\left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{6-x} = \binom{6}{x}\left(\frac{1}{2}\right)^6 = \binom{6}{x}\frac{1}{64} \quad ; \quad x = 0,1,2,\ldots,6 \\ X : \text{تعداد تخم مرغ سالم در هر شانه تاریخ‌گذشته} \end{cases}$$

مثال ۸ اگر $f(x) = \dfrac{\binom{10}{x}}{2^{10}} \; ; \; x = 0,1,\ldots,10$ باشد، ضریب تغییرات کمیت X چند درصد است؟

۱) 5 ۲) $\frac{\sqrt{5}}{2}$ ۳) $20\sqrt{2}$ ۴) $20\sqrt{\frac{5}{2}}$

حل: گزینه ۴ درست است.

با کمی دقت متوجه می‌شویم که تابع $f(x)$ همان نمایش خاص از توزیع دوجمله‌ای با پارامترهای $p = \frac{1}{2}$ و $n = 10$ است.

حال برای محاسبه ضریب پراکندگی توزیع به صورت زیر عمل می‌کنیم:

$$\begin{cases} CV_X \times 100 = \dfrac{\sigma}{\mu} \times 100 = \dfrac{\sqrt{\frac{5}{2}}}{5} \times 100 = 20\sqrt{\dfrac{5}{2}} \\ \mu = np = 10 \times \dfrac{1}{2} = 5 \\ \sigma = \sqrt{npq} = \sqrt{10 \times \dfrac{1}{2} \times \dfrac{1}{2}} = \sqrt{\dfrac{5}{2}} \end{cases}$$

محتمل‌ترین پیشامد

پیشامدی را که احتمال وقوع آن از سایر پیشامدها بیشتر باشد، محتمل‌ترین پیشامد گویند. حال در n بار تکرار آزمایش برنولی با احتمال ثابت p، می‌خواهیم بدانیم کدام‌یک از مقادیر X (تعداد موفقیت)، بزرگ‌ترین احتمال را دارند، برای این کار از رابطه زیر استفاده می‌کنیم:

$$\boxed{np - q \le x \le np + p}$$

الف) اگر مقدار $np + p$، عدد صحیح شود، آن‌گاه $np - q$ نیز عدد صحیح خواهد شد. در این حالت دو عدد برای مقادیر x وجود دارد که بزرگ‌ترین احتمال را دارند.

ب) اگر مقدار $np + p$، عدد اعشاری شود، آن‌گاه $np - q$ نیز عدد اعشاری خواهد شد. در این حالت تنها عدد صحیح بین این دو عدد اعشاری، بزرگ‌ترین احتمال را دارد.

نکته: محتمل‌ترین تعداد برای وقوع پیشامد X (موفقیت) $[p(n+1)]$ است، یعنی جزء صحیح (حد پایین) عدد اعشاری $p(n+1)$.

مثال در یک توزیع دوجمله‌ای احتمال وقوع S، سه برابر احتمال وقوع F است. اگر در این توزیع $n = 10$ باشد، محتمل‌ترین تعداد دفعاتی که S رخ می‌دهد برابر است با:

۱) 4 ۲) 8 ۳) 5 ۴) 2

حل: گزینه ۲ درست است.

راه حل اول:

$$\begin{cases} np-q\le x\le np+p \;\to\; 10\times\frac{3}{4}-\frac{1}{4}\le x\le 10\times\frac{3}{4}+\frac{3}{4} \;\to\; \frac{29}{4}\le x\le\frac{33}{4} \\ p=\frac{3}{4}\;,\; q=\frac{1}{4}\;,\; n=10 \end{cases}$$

بنا بر قسمت «ب» هر دو عدد $np+p$ و $np-q$ اعشاری هستند و تنها عدد صحیح بین آن دو $x=8$ است، بنابراین $x=8$ محتمل‌ترین وقوع پیشامد S (پیروزی) است.

راه حل دوم:

$$x \text{ محتمل‌ترین تعداد} = \left[p(n+1)\right]=\left[\frac{3}{4}(10+1)\right]=\left[\frac{33}{4}\right]=[8.25]=8$$

نکات مهم توزیع دو جمله‌ای

۱- اگر X دارای توزیع دوجمله ای با پارامترهای n و p باشد، آن‌گاه $Y=n-X$ دارای توزیع برنولی با پارامترهای n و q است.

۲- اگر $X_1,X_2,\dots,X_n$ متغیرهای تصادفی مستقل و هم‌توزیع برنولی با پارامتر p باشند، آن‌گاه مجموع آن‌ها، $Y=\sum_{i=1}^{n}X_i$، دارای توزیع دوجمله ای با پارامتر های n و p است.

۳- اگر X_1 دارای توزیع برنولی با پارامترهای n_1 و p و X_2 مستقل از X_1 دارای توزیع دوجمله‌ای با پارامترهای n_2 و p باشد، آن‌گاه مجموع آن‌ها، $Y=X_1+X_2$، دارای توزیع دوجمله‌ای با پارامترهای n_1+n_2 و همان p است.

	شرایط	متغیر	توزیع متغیر
۱	$X\sim Bin(n,p)$	$Y=n-X$	$Y\sim Bin(n,q)$
۲	$X_1,X_2,\dots,X_n \overset{i.i.d}{\sim} Bin(1,p)$	$Y=\sum_{i=1}^{n}X_i$	$Y\sim Bin(n,p)$
۳	$X_1\sim B(n_1,p)\,,\;X_2\sim B(n_2,p)$	$Y=X_1+X_2$	$Y\sim B(n_1+n_2,p)$

مثال ۹ اگر $X_1,X_2,\dots,X_n$ نمونه‌ای تصادفی از توزیعی با چگالی $f(x)=(0.2)^x(0.8)^{1-x}$ باشد. مقدار واریانس $S_n=\sum_{i=1}^{n}X_i$ کدام است؟

۱) 0.16 ۲) $(0.16)n$ ۳) $(0.2)n$ ۴) $(0.8)n$

حل: گزینه ۲ درست است.

با توجه به اینکه X دارای توزیع برنولی با پارامتر $p=0.2$ است و نیز بر اساس نکته (۲)، $S_n=\sum_{i=1}^{n}X_i$ دارای توزیع دوجمله‌ای با پارامترهای n و $p=0.2$ است؛ درنتیجه واریانس آن برابر است با:

$$Var(S_n)=npq=n\cdot(0.2)(0.8)=0.16n$$

مثال ۱۰ فرض کنید $X \sim Bin\left(4,\frac{1}{3}\right)$ و $Y \sim Bin\left(9,\frac{2}{3}\right)$ دو متغیر تصادفی مستقل از هم هستند. اگر $T = X + 9 - Y$ باشد، توزیع T کدام است؟

۱) $Ge\left(\frac{1}{3}\right)$

۲) $B\left(13,\frac{1}{3}\right)$

۳) $B\left(13,\frac{2}{3}\right)$

۴) توزیع استاندارد معلومی ندارد.

حل: گزینه ۲ درست است.

با توجه به نکته (۱)، می‌دانیم که $9 - Y$ دارای توزیع دوجمله‌ای با پارامترهای $n = 9$ و $p = \frac{2}{3}$ است.

$$9 - Y \sim Bin\left(9,\frac{1}{3}\right)$$

حال با توجه به نکته (۳)، مجموع X و $9 - Y$ دارای توزیع دوجمله‌ای با پارامترهای $n = 9 + 4 = 13$ و $p = \frac{1}{3}$ است.

$$X \sim Bin\left(4,\frac{1}{3}\right),\ 9 - Y \sim Bin\left(9,\frac{1}{3}\right) \rightarrow T = X + 9 - Y \sim Bin\left(4+9,\frac{1}{3}\right)$$

توزیع چندجمله‌ای (Multinomial Distribution)

مسئله (درک مطلب): نتیجه هر مسابقه برای یک تیم به ترتیب با **احتمال‌های** 0.1، 0.4 **و** 0.5 می‌تواند **برد، باخت و مساوی** باشد، احتمال آنکه این تیم در 5 مسابقه دارای 2 برد، 2 باخت و 1 مساوی شود کدام است؟

در توزیع دوجمله‌ای انجام هر آزمایش به یکی از دو نتیجه «موفقیت» یا «شکست» منجر می‌شد (برد یا باخت، شیر یا خط، پسر یا دختر). هرگاه انجام آزمایشی بتواند بیش از دو نتیجه ممکن داشته باشد، به طوری که احتمال هر یک از نتایج در هر آزمایش ثابت باشد و آزمایش‌ها از هم مستقل باشند، توزیع مربوط به این آزمایش، چندجمله‌ای خواهد بود (برد، باخت یا مساوی؛ خوب، عالی، متوسط یا ضعیف؛ شاغل، بازنشسته یا بیکار).

تعریف: هرگاه انجام آزمایشی منجر به یکی از k نتیجه مجزا با احتمال‌های ثابت $p_1, p_2, ..., p_k$ شود، به‌گونه‌ای که $p_1 + p_2 + ... + p_k = 1$ باشد، اگر آن آزمایش را n بار به طور مستقل انجام دهیم، احتمال وقوع x_1 بار از نتیجه اول و x_2 بار از نتیجه دوم و ... و x_k بار از نتیجه kام $(x_1 + x_2 + ... + x_k = n)$ دارای توزیع چندجمله‌ای است و به صورت زیر محاسبه می‌شود:

$$P(x_1, x_2, ..., x_k) = \binom{n}{x_1, x_2, ..., x_k} p_1^{x_1} p_2^{x_2} ... p_k^{x_k} = \frac{n!}{x_1! x_2! ... x_k!} \; p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$$

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع چندجمله‌ای باشد، پارامترهای آن $n_1, n_2, ...$ و $p_1, p_2, ...$ هستند و به صورت $X \sim MN(n_1, n_2, ..., p_1, p_2, ...)$ نمایش داده می‌شود.

حل مسئله درک مطلب:

در این مسئله با توجه به شرایط توزیع چندجمله‌ای، احتمال مورد نظر به صورت زیر محاسبه می‌شود:

$$\begin{cases} P(X_1 = 2, X_2 = 2, X_3 = 1) = \dfrac{5!}{2!2!1!} (0.1)^2 (0.4)^2 (0.5)^1 \\ p_1 = 0.1, \; p_2 = 0.4, \; p_3 = 0.5 \\ x_1 = 2, \; x_2 = 2, \; x_3 = 1 \\ n = 5 \end{cases}$$

$E(X)$ و σ_X^2 در توزیع چندجمله‌ای

برای محاسبه میانگین و واریانس «**تعداد دفعات تکرار مربوط به نتیجه iام در آزمایش چندجمله‌ای**» از **توزیع دوجمله‌ای** استفاده می‌کنیم.

x_1 بار از نتیجه اول و ... و x_k بار از نتیجه kام $(x_1 + ... + x_k = n)$	$P(x_1, ..., x_k) = \dfrac{n!}{x_1! ... x_k!} p_1^{x_1} ... p_k^{x_k}$	**تابع احتمال**
میانگین نتیجه iام در توزیع چندجمله‌ای	$E(X) = np$	**امید ریاضی (میانگین)**
واریانس نتیجه iام در توزیع چندجمله‌ای	$\sigma_X^2 = npq$	**واریانس**
	$\sigma_X = \sqrt{npq}$	**انحراف معیار**
	$p = p_i, \; q = 1 - p_i$	

مثال ۱ در یک شهر بزرگ به ترتیب 0.40، 0.30 و 0.30 از مردم، شاغل، بیکار و بازنشسته هستند. درصورتی‌که 6 نفر از مردم این شهر را انتخاب کنیم، احتمال اینکه به ترتیب 2 نفر شاغل، 2 نفر بیکار و 2 نفر بازنشسته باشند، چقدر است؟

۱) 0.6 ۲) 0.4 ۳) 0.1166 ۴) 0.2

حل: گزینه ۳ درست است.

$$P(X_1=2\ ,\ X_2=2\ ,\ X_3=2)=\frac{6!}{2!2!2!}(0.4)^2(0.3)^2(0.3)^2=0.1166$$

مثال ۲ از دانش‌آموزان یک مدرسه 0.20 کلاس اول، 0.30 کلاس دوم و 0.50 کلاس سوم هستند. اگر 4 نفر از دانش‌آموزان مدرسه انتخاب شده باشند، مطلوب است احتمال آنکه:

الف) 1 نفر کلاس اول، 2 نفر کلاس دوم و 1 نفر کلاس سوم باشند.

ب) 1 نفر کلاس اول و 3 نفر کلاس دوم باشند.

ج) 1 نفر کلاس اول و بقیه کلاس دوم یا سوم باشند.

حل:

الف)

$$\begin{cases}P(X_1=1\ ,\ X_2=2\ ,X_3=1)=\dfrac{4!}{1!2!1!}(0.2)^1(0.3)^2(0.5)^1=0.108\\ x_1=1,\ x_2=2, x_3=1\end{cases}$$

ب)

$$\begin{cases}P(X_1=1, X_2=3, X_3=0)=\dfrac{4!}{1!3!0!}(0.2)^1(0.3)^3(0.5)^0=0.0216\\ x_1=1, x_2=3, x_3=0\end{cases}$$

ج)

$$\begin{cases}P(X_1=1, X_2=3)=\dfrac{4!}{1!3!}(0.2)^1(0.8)^3=0.4096\\ \text{کلاس اول}: p_1=0.2\quad ,\quad \text{کلاس دوم یا سوم}: p_2=0.3+0.5=0.8\\ x_1=1\quad ,\quad x_2=3\end{cases}$$

مثال ۳ در یک کارخانه به ترتیب 0.2، 0.4 و 0.4 کالاها در شیفت‌های صبح، عصر و شب تولید می‌شوند. اگر 100 کالا انتخاب کنیم، امید و واریانس تعداد کالاهای شیفت صبح در این نمونه کدام است؟

۱) 20 , 16 ۲) 8 , 64 ۳) 4 , 16 ۴) 9 , 25

حل: گزینه ۱ درست است.

با در نظر گرفتن $p=0.2$ برای کالاهای شیفت صبح و با استفاده از توزیع دوجمله‌ای داریم:

$$\begin{cases}E(X)=np=100\times 0.2=20\\ \sigma_X^2=npq=100\times 0.2\times 0.8=16\\ n=100\ ,\ p=0.2\ ,\ q=1-p=0.8\end{cases}$$

توزیع دوجمله‌ای منفی (پاسکال) (Negative Binomial Distribution)

مسئله ۱ (درک مطلب): احتمال آنکه هر پرتاب بازیکنی به هدف بخورد، $\frac{2}{3}$ است. احتمال اینکه **سومین پرتابی که به هدف می‌خورد، پنجمین پرتاب** وی باشد، چقدر است؟

مسئله ۲ (درک مطلب): در یک ظرف، 10 توپ سفید و 5 توپ سیاه داریم. یک توپ به طور تصادفی با جایگذاری انتخاب می‌کنیم. این عمل را آن‌قدر تکرار می‌کنیم تا چهارمین توپ سیاه انتخاب شود. متوسط تعداد انتخاب (تکرار) چقدر است؟

مقدمه: هنگام انجام آزمایش‌های مستقل برنولی که نتیجه هر آزمایش در آن به یکی از دو وضعیت موفقیت (با احتمال p) و یا شکست (با احتمال q) منجر می‌شود، محاسبه احتمال وقوع x موفقیت در n بار تکرار آزمایش برنولی با استفاده از توزیع دوجمله‌ای برابر است با $\binom{n}{x}p^x q^{n-x}$.

حال اگر **«پیشامد وقوع r امین موفقیت در x امین آزمایش»** مورد نظر باشد $(x \geq r)$، برای محاسبه احتمال این پیشامد:

اولاً، باید در $x-1$ آزمایش قبلی دقیقاً به $r-1$ موفقیت برسیم: $\binom{x-1}{r-1}p^{r-1}q^{x-r}$

ثانیاً، در آزمایش xام به موفقیت r ام برسیم: p

به عبارت دیگر:

$$\underbrace{P(x)}_{\text{r امین موفقیت در x امین آزمایش}} = \overbrace{\underbrace{\binom{x-1}{r-1}p^{r-1}q^{x-r}}_{\text{r-1 موفقیت}}}^{\text{x-1 آزمایش قبلی}} \cdot \overbrace{\underbrace{p}_{\text{موفقیت r ام}}}^{\text{آزمایش x ام}} = \binom{x-1}{r-1}p^r q^{x-r}$$

تعریف: هرگاه یک آزمایش برنولی با احتمال موفقیت p را به طور مستقل، آن‌قدر تکرار کنیم تا به r **امین موفقیت** برسیم، آن‌گاه متغیر تصادفی«X: تعداد آزمایشات لازم برای رسیدن به r امین موفقیت» $(x = r, r+1, \ldots)$ دارای توزیع «**دوجمله‌ای منفی (پاسکال)**» یا توزیع «زمان انتظار» خواهد بود.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع دوجمله‌ای منفی باشد، پارامترهای آن r و p هستند و به صورت $X \sim NB(r,p)$ نمایش داده می‌شود.

تابع احتمال	$P(x) = \binom{x-1}{r-1}p^r q^{x-r}$ $x = r, r+1, \ldots$	احتمال وقوع r امین موفقیت در x امین آزمایش (احتمال وقوع موفقیت r ام در آزمایش x ام)
امید ریاضی (میانگین)	$E(X) = \frac{r}{p}$	میانگین تعداد آزمایش مورد انتظار برای رسیدن به r امین موفقیت
واریانس	$\sigma_X^2 = \frac{rq}{p^2}$	واریانس تعداد آزمایش لازم برای رسیدن به r امین موفقیت
انحراف معیار	$\sigma_X = \sqrt{\frac{rq}{p^2}}$	
تابع مولد گشتاور	$M_X(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$	

محاسبه احتمال

الف) ابتدا وضعیتی را که به دنبال r امین وقوع آن هستیم، به عنوان موفقیت (با احتمال p) در نظر می‌گیریم.

ب) $\binom{x-1}{r-1}p^r q^{x-r}$؛ احتمال وقوع r امین موفقیت در x امین آزمایش است.

حل مسئله ۱ (درک مطلب):

الف) در این مسئله $p=\frac{2}{3}$ (احتمال به هدف خوردن) است.

ب)

$$\begin{cases} P(\text{سومین هدف، پنجمین پرتاب}) = \binom{x-1}{r-1}p^r q^{x-r} = \binom{5-1}{3-1}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2 = \binom{4}{2}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2 \\ r=3\ ,\ x=5 \\ p=\frac{2}{3}\ ,\ q=\frac{1}{3} \end{cases}$$

حل مسئله ۲ (درک مطلب):

با توجه به اینکه انتخاب با جایگذاری است، احتمال موفقیت (سیاه بودن) ثابت بوده و توزیع تعداد تکرار برای رسیدن به چهارمین موفقیت، دوجمله‌ای منفی است.

$$\begin{cases} E(X)=\frac{r}{p}=\frac{4}{\frac{5}{15}}=12 \\ p=P(\text{سیاه})=\frac{5}{15}\quad ,\quad q=P(\text{سفید})=\frac{10}{15} \end{cases}$$

مثال ۱ 0.20 تولیدات یک کارخانه معیوب است. متوسط و واریانس تعداد کالاهای انتخاب‌شده برای رسیدن به دهمین کالای معیوب کدام است؟

۱) $\mu=0.02\ ,\ \sigma^2=200$ ۲) $\mu=50\ ,\ \sigma^2=4$ ۳) $\mu=0.02\ ,\ \sigma^2=4$ ۴) $\mu=50\ ,\ \sigma^2=200$

حل: گزینه ۴ درست است.

با توجه به $p=0.2$ و $r=10$ (دهمین کالای معیوب) داریم:

$$\begin{cases} E(X)=\frac{r}{p}=\frac{10}{0.2}=50 \\ \sigma_X^2=\frac{rq}{p^2}=\frac{10\times 0.8}{(0.2)^2}=200 \\ r=10\ ,\ p=0.2\ ,\ q=0.8 \end{cases}$$

مثال ۲ در یک تیراندازی تمرینی به یک هدف آن‌قدر شلیک می‌شود تا این‌که هدف دقیقاً n بار مورد اصابت قرار گیرد. اگر هر بار تیراندازی به‌صورت مستقل انجام گیرد و احتمال اصابت به هدف در هر بار تیراندازی p باشد، به‌طور متوسط چند گلوله مصرف می‌شود؟

۱) $np(1-p)$ ۲) np ۳) $\frac{n}{p}$ ۴) $\frac{n(1-p)}{p^2}$

حل: گزینه ۳ درست است.

اگر X تعداد دفعات تیراندازی تا اصابت n امین گلوله به هدف باشد، X دارای توزیع دو جمله‌ای منفی با پارامترهای n و p است

و ازاین‌رو: $E(X)=\frac{n}{p}$

توزیع هندسی (Geometric Distribution)

مسئله۱ (درک مطلب): احتمال اصابت موشکی به یک جنگنده 0.4 است. با اصابت یک موشک، جنگنده سقوط می‌کند. احتمال آنکه **جنگنده در پرتاب پنجمین موشک سقوط کند** چقدر است؟

مسئله۲ (درک مطلب): 10 درصد تولیدات کارخانه‌ای معیوب‌اند، احتمال اینکه **سومین کالای** کنترل‌شده، **اولین کالای معیوب باشد**، چقدر است؟

مسئله۳ (درک مطلب): در یک ظرف، 10 توپ سفید و 5 توپ سیاه داریم. یک توپ به طور تصادفی با جایگذاری انتخاب می‌کنیم. این عمل را **آن‌قدر تکرار می‌کنیم تا توپ سیاه انتخاب شود**. متوسط تعداد انتخاب (تکرار) کدام است؟

مقدمه: در انجام آزمایش‌های مستقل برنولی که نتیجه هر آزمایش در آن به یکی از دو وضعیت موفقیت (با احتمال p) و یا شکست (با احتمال q) منجر می‌شود، اگر «پیشامد وقوع اولین موفقیت در xامین آزمایش» مورد نظر باشد $(x \geq 1)$، برای محاسبه احتمال این پیشامد:

اولاً، باید در تمام $x-1$ آزمایش قبلی با شکست مواجه شده باشیم: q^{x-1}

ثانیاً، در آزمایش xام به موفقیت برسیم (اولین موفقیت): p

به عبارت دیگر:

$$\text{احتمال وقوع اولین موفقیت در } x \text{ امین آزمایش} = \overbrace{\underbrace{q^{x-1}}_{x-1\ \text{شکست}}}^{x-1\ \text{آزمایش قبلی}} \cdot \overbrace{\underbrace{p}_{\text{اولین موفقیت}}}^{\text{آزمایش } x \text{ ام}} = q^{x-1}p$$

تعریف: هرگاه یک آزمایش برنولی با احتمال موفقیت p را به طور مستقل آن‌قدر تکرار کنیم تا به **اولین موفقیت** برسیم (بالاخره موفق شویم)، آن‌گاه متغیر تصادفی «X : **تعداد آزمایشات** لازم برای رسیدن به **اولین موفقیت**» $(x = 1,2,...)$ دارای توزیع **هندسی** خواهد بود.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع هندسی باشد، پارامتر آن p است و به صورت $X \sim G(p)$ یا $X \sim Ge(p)$ نمایش داده می‌شود.

تابع احتمال	$P(x) = q^{x-1} \cdot p$ $x = 1,2,...$	احتمال وقوع اولین موفقیت در xامین آزمایش (احتمال آنکه در آزمایش xام بالاخره موفق شویم)
امید ریاضی (میانگین)	$E(X) = \frac{1}{p}$	میانگین تعداد آزمایش مورد انتظار برای رسیدن به اولین موفقیت
واریانس	$\sigma_X^2 = \frac{q}{p^2}$	واریانس تعداد آزمایش لازم برای رسیدن به اولین موفقیت
انحراف معیار	$\sigma_X = \sqrt{\frac{q}{p^2}}$	
تابع مولد گشتاور	$M_X(t) = \frac{pe^t}{1-qe^t}$	

محاسبه احتمال

$$\begin{cases}\text{الف) ابتدا وضعیتی را که به دنبال } r \text{ امین وقوع آن هستیم، به عنوان موفقیت (با احتمال } p\text{) در نظر می‌گیریم.}\\ \text{ب) } \binom{x-1}{r-1}p^r q^{x-r} \text{ ؛ احتمال وقوع } r \text{ امین موفقیت در } x \text{ امین آزمایش است.}\end{cases}$$

حل مسئله ۱ (درک مطلب):

الف) در این مسئله $p=\frac{2}{3}$ (احتمال به هدف خوردن) است.

ب)

$$\begin{cases}P(\text{سومین هدف، پنجمین پرتاب})=\binom{x-1}{r-1}p^r q^{x-r}=\binom{5-1}{3-1}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2=\binom{4}{2}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2\\ r=3\ ,\ x=5\\ p=\frac{2}{3}\ ,\ q=\frac{1}{3}\end{cases}$$

حل مسئله ۲ (درک مطلب):

با توجه به اینکه انتخاب با جایگذاری است، احتمال موفقیت (سیاه بودن) ثابت بوده و توزیع تعداد تکرار برای رسیدن به چهارمین موفقیت، دوجمله‌ای منفی است.

$$\begin{cases}E(X)=\frac{r}{p}=\frac{4}{\frac{5}{15}}=12\\ p=P(\text{سیاه})=\frac{5}{15}\quad,\quad q=P(\text{سفید})=\frac{10}{15}\end{cases}$$

مثال ۱ 0.20 تولیدات یک کارخانه معیوب است. متوسط و واریانس تعداد کالاهای انتخاب‌شده برای رسیدن به دهمین کالای معیوب کدام است؟

۱) $\mu=0.02\ ,\ \sigma^2=200$ ۲) $\mu=50\ ,\ \sigma^2=4$ ۳) $\mu=0.02\ ,\ \sigma^2=4$ ۴) $\mu=50\ ,\ \sigma^2=200$

حل: گزینه ۴ درست است.

با توجه به $p=0.2$ و $r=10$ (دهمین کالای معیوب) داریم:

$$\begin{cases}E(X)=\frac{r}{p}=\frac{10}{0.2}=50\\ \sigma_X^2=\frac{rq}{p^2}=\frac{10\times 0.8}{(0.2)^2}=200\\ r=10\ ,\ p=0.2\ ,\ q=0.8\end{cases}$$

مثال ۲ در یک تیراندازی تمرینی به یک هدف آن‌قدر شلیک می‌شود تا این‌که هدف دقیقاً n بار مورد اصابت قرار گیرد. اگر هر بار تیراندازی به‌صورت مستقل انجام گیرد و احتمال اصابت به هدف در هر بار تیراندازی p باشد، به‌طور متوسط چند گلوله مصرف می‌شود؟

۱) $np(1-p)$ ۲) np ۳) $\frac{n}{p}$ ۴) $\frac{n(1-p)}{p^2}$

حل: گزینه ۳ درست است.

اگر X تعداد دفعات تیراندازی تا اصابت n امین گلوله به هدف باشد، X دارای توزیع دو جمله‌ای منفی با پارامترهای n و p است و ازاین‌رو: $E(X)=\frac{n}{p}$

توزیع هندسی (Geometric Distribution)

مسئله۱ (درک مطلب): احتمال اصابت موشکی به یک جنگنده 0.4 است. با اصابت یک موشک، جنگنده سقوط می‌کند. احتمال آنکه **جنگنده در پرتاب پنجمین موشک سقوط کند** چقدر است؟

مسئله۲ (درک مطلب): 10 درصد تولیدات کارخانه‌ای معیوب‌اند، احتمال اینکه **سومین کالای** کنترل‌شده، **اولین کالای معیوب باشد**، چقدر است؟

مسئله۳ (درک مطلب): در یک ظرف، 10 توپ سفید و 5 توپ سیاه داریم. یک توپ به طور تصادفی با جایگذاری انتخاب می‌کنیم. این عمل را **آن‌قدر تکرار می‌کنیم تا توپ سیاه انتخاب شود**. متوسط تعداد انتخاب (تکرار) کدام است؟

مقدمه: در انجام آزمایش‌های مستقل برنولی که نتیجه هر آزمایش در آن به یکی از دو وضعیت موفقیت (با احتمال p) و یا شکست (با احتمال q) منجر می‌شود، اگر «پیشامد وقوع اولین موفقیت در xامین آزمایش» مورد نظر باشد $(x \geq 1)$، برای محاسبه احتمال این پیشامد:

اولاً، باید در تمام $x-1$ آزمایش قبلی با شکست مواجه شده باشیم: q^{x-1}

ثانیاً، در آزمایش xام به موفقیت برسیم (اولین موفقیت): p

به عبارت دیگر:

$$\text{احتمال وقوع اولین موفقیت در } x \text{امین آزمایش} = \overbrace{\underbrace{q^{x-1}}_{x-1\ \text{شکست}}}^{x-1\ \text{آزمایش قبلی}} \cdot \overbrace{\underbrace{p}_{\text{اولین موفقیت}}}^{\text{آزمایش } x \text{ام}} = q^{x-1}p$$

تعریف: هرگاه یک آزمایش برنولی با احتمال موفقیت p را به طور مستقل آن‌قدر تکرار کنیم تا به **اولین موفقیت** برسیم (بالاخره موفق شویم)، آن‌گاه متغیر تصادفی« X : **تعداد آزمایشات** لازم برای رسیدن به **اولین موفقیت**» $(x = 1,2,...)$ دارای توزیع **هندسی** خواهد بود.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع هندسی باشد، پارامتر آن p است و به صورت $X \sim G(p)$ یا $X \sim Ge(p)$ نمایش داده می‌شود.

تابع احتمال	$P(x) = q^{x-1} \cdot p$ $x = 1,2,...$	احتمال وقوع اولین موفقیت در xامین آزمایش (احتمال آنکه در آزمایش xام بالاخره موفق شویم)
امید ریاضی (میانگین)	$E(X) = \frac{1}{p}$	میانگین تعداد آزمایش مورد انتظار برای رسیدن به اولین موفقیت
واریانس	$\sigma_X^2 = \frac{q}{p^2}$	واریانس تعداد آزمایش لازم برای رسیدن به اولین موفقیت
انحراف معیار	$\sigma_X = \sqrt{\frac{q}{p^2}}$	
تابع مولد گشتاور	$M_X(t) = \frac{pe^t}{1-qe^t}$	

نکته: توزیع دوجمله‌ای منفی در حالت خاص $(r=1)$ همان توزیع هندسی است، بنابراین توزیع هندسی حالت خاصی از توزیع دوجمله‌ای منفی است زمانی که به دنبال اولین موفقیت $(r=1)$ هستیم.

نکته: تابع توزیع تجمعی متغیر تصادفی هندسی به صورت زیر است:

$$F_X(\alpha)=P(X\le\alpha)=1-q^{\alpha}$$

محاسبه احتمال

الف) ابتدا وضعیتی را که به دنبال اولین وقوع آن هستیم، به عنوان موفقیت (با احتمال p) در نظر می‌گیریم.

ب) $P(x)=q^{x-1}p$؛ احتمال وقوع اولین موفقیت در xامین آزمایش است.

حل مسئله ۱ (درک مطلب):

الف) در این مسئله $p=0.4$ (به هدف خوردن)

ب)

$$\begin{cases}P(x)=q^{x-1}p=(0.6)^4(0.4)^1=0.05184\\ x=5\ ,\ \ p=0.4\ ,\ q=0.6\end{cases}$$

حل مسئله ۲ (درک مطلب):

الف) در این مسئله $p=0.1$ (معیوب بودن)

ب)

$$\begin{cases}P(x)=q^{x-1}p=(0.9)^2(0.1)^1=0.081\\ x=3\ ,\ \ p=0.1\ ,\ q=0.9\end{cases}$$

حل مسئله ۳ (درک مطلب):

با توجه به اینکه انتخاب با جایگذاری است، احتمال موفقیت (سیاه بودن) ثابت بوده و توزیع تعداد تکرار برای رسیدن به اولین موفقیت، هندسی است.

$$\begin{cases}E(X)=\dfrac{1}{p}=\dfrac{1}{\frac{5}{15}}=3\\[2ex] p=P(\text{سیاه})=\dfrac{5}{15}\quad,\quad q=P(\text{سفید})=\dfrac{10}{15}\end{cases}$$

یادآوری:

هر تصاعد هندسی به صورت زیر است:

$$\begin{cases}a_1,a_2,\dots,a_n\ \rightarrow\ a_1\ ,\ a_1q\ ,\ a_1q^2,\dots a_1q^{n-1}\\ q:\text{قدرنسبت}\end{cases}$$

و مجموع هر تصاعد هندسی را می‌توان به صورت زیر در نظر گرفت:

$$a_1+a_1q+a_1q^2+\dots+a_1q^{n-1}=\frac{a_1\left(1-q^n\right)}{1-q}\ \xrightarrow[0<q<1]{n\to\infty}\ =\frac{a_1}{1-q}=\frac{\text{جمله اول}}{1-\text{قدر نسبت}}$$

نکته: با توجه به آنکه همواره $\sum P(x)=1$ (مجموع مقادیر تابع احتمال متغیر تصادفی گسسته برابر با 1 است)، برای تابع احتمال هندسی رابطه $\sum_{x=1}^{\infty} P(x)=\sum_{x=1}^{\infty} p\,q^{x-1}=1$ برقرار است و داریم:

$$\underbrace{p}_{P(X=1)}+\underbrace{qp}_{P(X=2)}+\underbrace{q^2p}_{P(X=3)}+\underbrace{q^3p}_{P(X=4)}\cdots=1$$

برای مثال:

۱) سه آزمایش برای اولین موفقیت لازم باشد (سومین آزمایش، اولین موفقیت). $P(X=3)=q^2p$

۲) حداکثر 2 آزمایش برای اولین موفقیت لازم باشد. $P(X\le 2)=P(X=1)+P(X=2)=p+qp$

۳) حداقل 2 آزمایش برای اولین موفقیت لازم باشد.

روش اول:

$$P(X\ge 2)=1-P(X=1)=1-p=q$$

روش دوم:

$$P(X\ge 2)=P(X=2)+P(X=3)+\ldots=qp+q^2p+\ldots=\frac{\text{جمله اول}}{1-\text{قدر نسبت}}=\frac{qp}{1-q}=q$$

۴) بیش از 2 آزمایش برای اولین موفقیت لازم باشد.

$$P(X>2)=P(X=3)+P(X=4)+\ldots=q^2p+q^3p+\ldots=\frac{\text{جمله اول}}{1-\text{قدر نسبت}}=\frac{q^2p}{1-q}=q^2$$

۵) تعداد آزمایش لازم برای رسیدن به اولین موفقیت، فرد باشد.

$$P(\text{فرد } X)=P(X=1)+P(X=3)+P(X=5)+\ldots=p+q^2p+q^4p+\ldots=\frac{\text{جمله اول}}{1-\text{قدر نسبت}}=\frac{p}{1-q^2}$$

مثال ۱ در یک جعبه با 5 مهره قرمز و 15 مهره آبی، مهره‌ها را با جایگذاری خارج می‌کنیم تا به اولین مهره قرمز برسیم. امید و واریانس تعداد مهره‌های انتخابی کدام است؟

۱) $3\,,\,4$ ۲) $12\,,\,4$ ۳) $3\,,\,\frac{4}{3}$ ۴) $12\,,\,\frac{4}{3}$

حل: گزینه ۲ درست است.

ازآنجاکه نمونه‌گیری با جایگذاری از جامعه محدود انجام می‌شود، شرایط قانون برنولی (داشتن احتمال ثابت) برقرار است ($\frac{5}{20}=$احتمال قرمز بودن و $\frac{15}{20}=$احتمال آبی بودن).

از طرفی تعداد انتخاب برای رسیدن به اولین مهره قرمز دارای توزیع هندسی با پارامتر $p=\frac{5}{20}$ است.

$$\begin{cases} E(X)=\dfrac{1}{p}=\dfrac{1}{\frac{1}{4}}=4 \\[2ex] \sigma_X^2=\dfrac{q}{p^2}=\dfrac{\frac{3}{4}}{\left(\frac{1}{4}\right)^2}=12 \\[2ex] p=P(\text{قرمز})=\dfrac{5}{20}=\dfrac{1}{4}=0.25\quad,\quad q=P(\text{آبی})=\dfrac{15}{20}=\dfrac{3}{4}=0.75 \end{cases}$$

مثال ۲ یک تاس سالم را آن‌قدر پرتاب می‌کنیم تا برای اولین بار نتیجه زوج ظاهر شود $(6,4,2)$. مطلوب است احتمال آنکه:

الف) حداقل 2 پرتاب برای رسیدن به اولین نتیجه زوج لازم باشد.

ب) تعداد پرتاب‌های لازم برای رسیدن به اولین نتیجه زوج مضربی از 3 باشد.

ج) تعداد پرتاب‌های لازم برای رسیدن به اولین نتیجه زوج، عددی فرد باشد.

حل:
در این مثال « X : تعداد پرتاب لازم برای رسیدن به اولین نتیجه زوج» دارای توزیع هندسی است و داریم:

$$\begin{cases} f(x)=P(x)=q^{x-1}p \quad ; \quad x=1,2,\ldots \\ p=P(\text{نتیجه زوج})=\dfrac{3}{6}=\dfrac{1}{2}\,,\ q=\dfrac{1}{2} \end{cases}$$

الف)

$$P(X\geq 2)=P(X=2)+P(X=3)+\ldots=qp+q^2p+\ldots=\frac{qp}{1-q}=\frac{qp}{p}=q=\frac{1}{2}$$

ب)

$$P(\text{مضرب 3}\ X)=P(X=3)+P(X=6)+P(X=9)+\ldots=q^2p+q^5p+q^8p+\ldots=\frac{q^2p}{1-q^3}=\frac{\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)}{1-\left(\frac{1}{2}\right)^3}=\frac{1}{7}$$

ج)

$$P(\text{فرد}\ X)=P(X=1)+P(X=3)+P(X=5)+\ldots=\underbrace{\cancel{q^0}}_{1}p+q^2p+q^4p+\ldots=\frac{p}{1-q^2}=\frac{\frac{1}{2}}{1-\left(\frac{1}{2}\right)^2}=\frac{2}{3}$$

مثال ۳ یک تاس سالم پی در پی پرتاب می شود. اگر X و Y به ترتیب نشان‌دهنده تعداد پرتاب‌های لازم تا مشاهده یک 6 و یک 5 باشند، مطلوب است محاسبه:

الف) $E(X)$ ب) $E(X|Y=1)$

حل: X و Y هر دو دارای توزیع هندسی با پارامتر $p=\frac{1}{6}$ هستند.

X : تعداد پرتاب‌های لازم برای رسیدن به اولین 6

Y : تعداد پرتاب‌های لازم برای رسیدن به اولین 5

احتمال موفقیت در هر دو متغیر X و Y، $\frac{1}{6}$ است زیرا احتمال آمدن یک عدد خاص در تاس $\frac{1}{6}$ است.

$$f(x)=pq^{x-1}=\frac{1}{6}\left(\frac{5}{6}\right)^{x-1} \quad ; \quad x=1,2,3,\ldots$$

$$f(y)=pq^{y-1}=\frac{1}{6}\left(\frac{5}{6}\right)^{y-1} \quad ; \quad y=1,2,3,\ldots$$

الف)

$$E(X)=\frac{1}{p}=\frac{1}{\frac{1}{6}}=6$$

ب) $E(X|Y=1)=E(X)+1=6+1=7$

دقت کنید که چون $Y=1$ است، یعنی در پرتاب اول عدد 5 ظاهر شده است، پس یک پرتاب به پرتاب‌های ما اضافه می‌شود.

نکات مهم توزیع هندسی

۱- اگر $X_1,\dots,X_r$ ، r متغیر تصادفی مستقل و هم‌توزیع هندسی با پارامتر p باشند، آن‌گاه مجموع این r متغیر، $Y=\sum_{i=1}^{r} X_i$ ، دارای توزیع دوجمله‌ای منفی با پارامترهای r و p است.

۲- اگر $X_1,\dots,X_n$ ، n متغیر تصادفی مستقل و هم‌توزیع هندسی با پارامتر p باشند، آن‌گاه مینیمم آن‌ها دارای توزیع هندسی با پارامتر $1-q^n$ است.

۳- اگر X و Y دو متغیر تصادفی مستقل و هم‌توزیع هندسی با پارامتر p باشند، آن‌گاه توزیع یکی از آن‌ها به شرط اینکه مجموعشان برابر عدد ثابت باشد، یکنواخت گسسته از یک تا یک منهای مجموعشان است.

	شرایط	متغیر	توزیع متغیر
۱	$X_1,\dots,X_r \overset{i.i.d}{\sim} Ge(p)$	$Y=\sum_{i=1}^{r} X_i$	$Y=NB(r\,,\,p)$
۲	$X_1,\dots,X_n \overset{i.i.d}{\sim} Ge(p)$	$Y=\min\{X_1,X_2,\dots,X_n\}$	$Y\sim Ge(1-q^n)$
۳	$X,Y \overset{i.i.d}{\sim} Ge(p)$	$Z=(X=k\,|\,X+Y=n)$	یکنواخت گسسته $f(z)=\frac{1}{n-1}\ ;\ z=1,\dots,n-1$

مثال ۴ فرض کنید X و Y دو متغیر تصادفی مستقل با توزیع یکسان از تابع احتمال زیر باشند:

$$f_p(x)=p(1-p)^{x-1} \quad ; \quad x=1,2,\dots$$

توزیع $Z=X+Y$ کدام است؟

۱) فوق هندسی

۲) هندسی با پارامتر $1-2p$

۳) هندسی با پارامتر $2p$

۴) دوجمله‌ای منفی با پارامترهای 2 و p

حل: گزینه ۴ درست است.

با توجه به نکته (۱)، توزیع مجموع دو متغیر تصادفی مستقل و هم‌توزیع هندسی با پارامتر p دارای توزیع دوجمله‌ای منفی با پارامترهای 2 و p است.

توزیع فوق هندسی (Hypergeometric Distribution)

مسئله (درک مطلب): از جعبه‌ای با 12 کالا که 4 تای آن سالم و 8 تا معیوب است، 3 کالا به‌تصادف انتخاب شده است. مطلوب است محاسبه احتمال آنکه

الف) 2 کالا سالم باشد.

ب) حداکثر 1 کالا سالم باشد.

ج) حداقل 1 کالا سالم باشد.

مقدمه: جامعه محدود N تایی را در نظر بگیرید که k تای آن موفقیت و $N-k$ تای دیگر شکست تلقی شود. حال اگر یک نمونه n تایی را از جامعه انتخاب کنیم، توزیع «تعداد موفقیت در نمونه» بسته به آنکه نمونه‌گیری «بدون جایگذاری» یا «با جایگذاری» باشد با هم متفاوت است.

۱- با جایگذاری: در این وضعیت احتمال موفقیت در هر بار نمونه‌گیری، مستقل، ثابت و برابر با $p=\frac{k}{N}$ است، درنتیجه آزمایش، برنولی و توزیع «تعداد موفقیت در نمونه» همان‌طورکه قبلاً بررسی شده بود، دوجمله‌ای است.

$$\text{احتمال موفقیت} = \underbrace{\overset{\text{بار اول}}{\frac{k}{N}}}_{p} \times \underbrace{\overset{\text{بار دوم}}{\frac{k}{N}}}_{p} \times \cdots$$

۲- بدون جایگذاری (پیش‌فرض): در این وضعیت احتمال موفقیت در هر بار نمونه‌گیری، وابسته به دفعات قبل است، زیرا در هر بار انتخاب، حجم جامعه 1 واحد کم می‌شود و احتمال موفقیت، دیگر ثابت نیست.

$$\text{احتمال موفقیت} = \overset{\text{بار اول}}{\frac{k}{N}} \times \overset{\text{بار دوم}}{\frac{k \text{ یا } k-1}{N-1}} \times \cdots$$

در چنین شرایطی برای محاسبه احتمال از همان رابطه کلاسیک فصل احتمال یعنی:

$$\text{احتمال} = \frac{\text{حالات مساعد}}{\text{حالات ممکن}}$$

و برای محاسبه حالات مساعد و ممکن، از ترکیب استفاده می‌کنیم.

یادآوری:

۱- در صورت عدم بیان، نوع انتخاب به طور **پیش‌فرض بدون جایگذاری** در نظر گرفته می‌شود.

۲- ترکیب $\binom{N}{n}$ حالات انتخاب یک نمونه n تایی (بدون جایگذاری) از یک جامعه N تایی است.

برای مثال در مسئله بالا حالات ممکن برای انتخاب 3 کالا از بین 12 کالا برابر است با $\binom{12}{3}$ که با توجه به 4 کالای سالم و 8 کالای معیوب، وضعیت‌های متفاوتی برای انتخاب به شرح زیر وجود دارد:

$$\underbrace{\binom{12}{3}}_{\text{حالات ممکن}} = \underbrace{\binom{4}{3}\binom{8}{0}}_{\text{هر 3 کالا سالم}} + \underbrace{\binom{4}{2}\binom{8}{1}}_{\text{2 سالم و 1 معیوب}} + \underbrace{\binom{4}{1}\binom{8}{2}}_{\text{1 سالم و 2 معیوب}} + \underbrace{\binom{4}{0}\binom{8}{3}}_{\text{هر 3 کالا معیوب}}$$

تعریف: هرگاه از یک **جامعه محدود** N تایی که K تای آن موفقیت و $N-K$ تای آن شکست است، یک نمونه n تایی **بدون جایگذاری** انتخاب کنیم، آن‌گاه «X: تعداد موفقیت در نمونه n تایی» دارای توزیع **فوق هندسی** است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع فوق هندسی باشد، پارامترهای آن N، n و k هستند و به صورت $X \sim HG(N,k,n)$ نمایش داده می‌شود.

احتمال آنکه در یک نمونه n تایی (بدون جایگذاری)، x تا متعلق به مجموعه k (موفقیت) باشد.	$P(x)=\dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ $x=0,1,2,\ldots,\min(n,k)$	**تابع احتمال**
متوسط تعداد موفقیت در نمونه n تایی با احتمال موفقیت $p=\dfrac{K}{N}$	$E(X)=np$	**امید ریاضی(میانگین)**
واریانس تعداد موفقیت در نمونه n تایی با احتمال موفقیت $p=\dfrac{K}{N}$ ضریب تصحیح واریانس: $\dfrac{N-n}{N-1}$	$\sigma_X^2=\dfrac{N-n}{N-1}npq$	**واریانس**
	$\sigma_X=\sqrt{\dfrac{N-n}{N-1}npq}$	**انحراف معیار**

ضریب تصحیح واریانس $\left(\dfrac{N-n}{N-1}\right)$

کمیت $\dfrac{N-n}{N-1}$ به عنوان ضریب تصحیح برای واریانس توزیع فوق هندسی به کار برده می‌شود و همان‌طورکه در ادامه خواهیم دید، در بعضی شرایط $\left(\dfrac{n}{N}\le 0.05\right)$ از این ضریب چشم‌پوشی می‌شود.

حل مسئله (درک مطلب)

الف)

$$P(X=2)=\frac{\binom{4}{2}\binom{8}{1}}{\binom{12}{3}}$$

ب)

$$P(X\le 1)=P(X=0)+P(X=1)=\frac{\binom{4}{0}\binom{8}{3}}{\binom{12}{3}}+\frac{\binom{4}{1}\binom{8}{2}}{\binom{12}{3}}$$

ج) با توجه به رابطه زیر دو روش برای محاسبهٔ احتمال حداقل 1 کالای سالم وجود دارد:

$$1=\underbrace{\frac{\binom{4}{0}\binom{8}{3}}{\binom{12}{3}}}_{P(X=0)\text{: هیچ کالای سالم}}+\underbrace{\frac{\binom{4}{1}\binom{8}{2}}{\binom{12}{3}}+\frac{\binom{4}{2}\binom{8}{1}}{\binom{12}{3}}+\frac{\binom{4}{3}\binom{8}{0}}{\binom{12}{3}}}_{P(X\geq 1)\text{: حداقل 1 کالای سالم}}$$

روش اول:

$$P(X\geq 1)=\frac{\binom{4}{1}\binom{8}{2}+\binom{4}{2}\binom{8}{1}+\binom{4}{3}\binom{8}{0}}{\binom{12}{3}}$$

روش دوم:

$$P(X\geq 1)=1-P(X=0)=1-\frac{\binom{4}{0}\binom{8}{3}}{\binom{12}{3}}$$

روش دوم که برای محاسبه سریع‌تر و آسان‌تر است، به روش مکمل نیز معروف است.

مثال ۱ از 9 عدد کالای یکسان موجود در یک کارتن، 3 عدد معیوب است. 4 کالا به طور تصادفی از بین آن‌ها برداشته می‌شود. با کدام احتمال لااقل سه کالای برداشته‌شده، سالم است؟

۱) $\frac{8}{21}$ ۲) $\frac{13}{21}$ ۳) $\frac{17}{42}$ ۴) $\frac{25}{42}$

حل: گزینه ۴ درست است.

در این مثال جامعه محدود است $(N=9)$ و داریم $k=6$ (کالای سالم) و $n=4$ (نمونه پیش‌فرض بدون جایگذاری)، درنتیجه توزیع فوق هندسی است و داریم:

$$\begin{cases} P(X\geq 3)=P(X=3)+P(X=4)=\dfrac{\binom{6}{3}\binom{3}{1}}{\binom{9}{4}}+\dfrac{\binom{6}{4}\binom{3}{0}}{\binom{9}{4}}=\dfrac{60+15}{126}=\dfrac{25}{42} \\ N=9\ ,\ n=4\ ,\ k=6 \end{cases}$$

مثال ۲ از جوراب‌های بسته‌بندی‌شده در جعبه‌ای 9 عدد سالم و 3 عدد معیوب است. یک مشتری به طور تصادفی 4 عدد را خریداری می‌کند. میانگین و واریانس تعداد جوراب‌های معیوب در این خرید به ترتیب از چپ به راست چقدر است؟

۱) $1\,,\frac{6}{11}$ ۲) $1\,,\frac{3}{4}$ ۳) $1\,,\frac{27}{12}$ ۴) $2\,,\frac{9}{12}$

حل: گزینه ۱ درست است.

در این مثال جامعه محدود است $(N=12)$ و داریم $k=3$ (کالای معیوب) و $n=4$ (نمونه پیش‌فرض بدون جایگذاری)، درنتیجه توزیع فوق هندسی است و داریم:

$$\begin{cases} E(X)=np=4\times\frac{3}{12}=1 \\ Var(X)=npq\left(\frac{N-n}{N-1}\right)=4\times\frac{3}{12}\times\frac{9}{12}\times\frac{12-4}{12-1}=\frac{6}{11} \\ N=12\ ,\ n=4\ ,\ k=3\ ,\ p=\frac{k}{N}=\frac{3}{12} \end{cases}$$

مثال ۳ از یک مجموعه شامل 3 کالای معیوب و 7 کالای سالم، یک نمونه 4 تایی انتخاب کرده‌ایم. احتمال آنکه حداقل 3 کالا سالم باشد، چقدر است؟

حل:
دقت کنید کل حالات انتخاب به شرح زیر است:

$$\binom{10}{4}=\underbrace{\cancel{\binom{7}{0}\binom{3}{4}}}_{\text{غیرممکن}}+\binom{7}{1}\binom{3}{3}+\binom{7}{2}\binom{3}{2}+\underbrace{\binom{7}{3}\binom{3}{1}+\binom{7}{4}\binom{3}{0}}_{\text{حداقل 3 سالم}}$$

همان‌طورکه مشاهده می‌شود به هیچ وجه امکان انتخاب $\binom{7}{0}\binom{3}{4}$ وجود ندارد و جواب سؤال $\dfrac{\binom{7}{3}\binom{3}{1}+\binom{7}{4}\binom{3}{0}}{\binom{10}{4}}$ است.

تقریب توزیع فوق هندسی به دوجمله‌ای

زمانی که N (حجم جامعه) بزرگ و n (حجم نمونه) کوچک شود (بنا بر قانون سرانگشتی n از 5 درصد N تجاوز نکند)، تفاوت چندانی بین نمونه‌گیری بدون جایگذاری و نمونه‌گیری با جایگذاری وجود ندارد؛ در این وضعیت می‌توان برای تقریب احتمال‌های توزیع فوق هندسی از توزیع دوجمله‌ای پارامترهای n و $p=\frac{K}{N}$ استفاده کرد. بدیهی است در این شرایط از ضریب تصحیح $\frac{N-n}{N-1}$ برای واریانس نیز چشم‌پوشی می‌شود؛ به عبارت دیگر:

$$\begin{cases} E(X)=np=n\cdot\frac{k}{N} \\ \sigma_X^2=npq=n\cdot\frac{k}{N}\left(1-\frac{k}{N}\right) \end{cases}$$

هرگاه در یک توزیع فوق هندسی، حجم نمونه (n) را کاهش و حجم جامعه (N) را افزایش دهیم، به طوری که $\frac{n}{N}\le 0.05$، آن‌گاه توزیع دوجمله‌ای تقریب مناسبی برای توزیع فوق هندسی است و از ضریب تصحیح واریانس $\frac{N-n}{N-1}$ چشم‌پوشی می‌کنیم.
واضح است در این شرایط روابط زیر برقرار است:

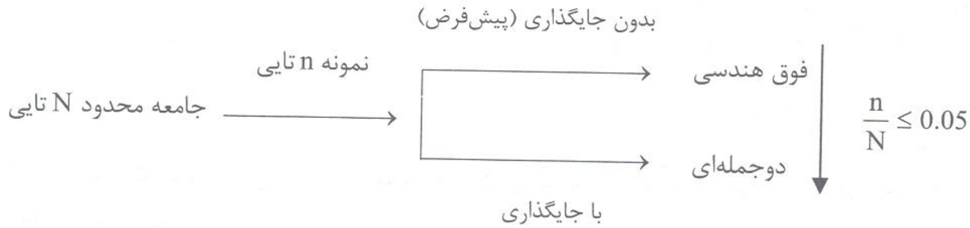
$$\frac{n}{N}\le\frac{5}{100}\ \rightarrow\ \begin{cases} n\le \%5N & \textbf{نمونه از 5\% جامعه تجاوز نمی‌کند.} \\ 20n\le N & \textbf{جامعه حداقل 20 برابر نمونه است.} \end{cases}$$

✓ دقت کنید!

هرگاه $\frac{n}{N} > 0.05$ یا $n > 0.05N$ باشد، تقریب دوجمله‌ای دیگر مناسب نیست و از ضریب تصحیح $\frac{N-n}{N-1}$ در واریانس چشم‌پوشی نمی‌شود.

نتیجه:

برای تعیین توزیع **«تعداد موفقیت در نمونه»** همواره:



مثال ۴ از یک جامعه 4000 نفره یک نمونه تصادفی 40 تایی انتخاب شده است. در این حالت تابع احتمال متغیر تصادفی X، کدام است؟

۱) هندسی
۲) دوجمله‌ای
۳) فوق هندسی
۴) هم فوق هندسی و هم دوجمله‌ای

حل: گزینه ۴ درست است.

اولاً، ازآنجاکه جامعه محدود است $(N = 4000)$ و نمونه $(n = 40)$ به طور پیش‌فرض بدون‌جایگذاری انتخاب شده، توزیع فوق هندسی است.

ثانیاً، با توجه به آنکه $0.05 \ge 0.01 = \frac{40}{4000} = \frac{n}{N}$ است، توزیع دوجمله‌ای تقریب مناسبی برای توزیع فوق هندسی است.

ثالثاً، اگر گزینه ۴ وجود نداشت گزینه ۲ را انتخاب می‌کردیم زیرا دوجمله‌ای تقریب فوق هندسی است و در شرایط تقریب، محاسبه تقریب توزیع آسان‌تر است.

دقت کنید که اگر $\frac{n}{N} > 0.05$ بود، آن‌گاه تقریب دوجمله‌ای مناسب نبود و گزینه ۳ پاسخ درست بود.

نکات مهم توزیع فوق هندسی

اگر X دارای توزیع دوجمله‌ای با پارامترهای« m_1 و p» و Y دارای توزیع دوجمله‌ای با پارامترهای« m_2 و p» باشد، آن‌گاه $Z = (X = v | X + Y = t)$ که در آن v و t اعداد ثابت هستند، دارای توزیع فوق هندسی با پارامترهای $k = v$ ،$n = t$ و $N = m_1 + m_2$ است.

توزیع متغیر	متغیر	شرایط
$Z \sim HG(N = m_1 + m_2 , k = v , n = t)$	$Z = (X = v \| X + Y = t)$	$X \sim Bin(m_1,p), Y \sim Bin(m_2,p)$ X و Y مستقل‌اند

مثال ۵ اگر X و Y متغیرهای تصادفی دوجمله‌ای مستقل با پارامترهای یکسان n و p باشند، تابع چگالی احتمال شرطی X، به شرط $X+Y=m$، یعنی $P(X=v | X+Y=m)$ برابر کدام است؟

۱) $\frac{\binom{n}{v}}{\binom{2n}{m}}$
۲) $\frac{\binom{n}{m-v}}{\binom{2n}{m}}$
۳) $\frac{\binom{n}{v}\binom{n}{m-v}}{\binom{2n}{m}}$
۴) $\frac{v}{m}$

حل: گزینه ۳ درست است.

با توجه به نکته گفته شده توزیع $(X = v | X + Y = m)$ فوق هندسی است با $N = n + n = 2n , k = v , n = m$.

توزیع پواسون (Poisson Distribution)

مسئله۱ (درک مطلب): به طور متوسط در هر ساعت، 12 اتومبیل به یک پمپ بنزین مراجعه می‌کنند. احتمال اینکه در 15 دقیقه 3 اتومبیل مراجعه کنند، چقدر است؟

مسئله۲ (درک مطلب): به طور متوسط در هر 200 متر پارچه، 3 زدگی وجود دارد. احتمال آنکه در یک بسته 600 متری 1 زدگی وجود داشته باشد، چقدر است؟

مقدمه: گاهی **تعداد اتفاقات** (رویداد، رخداد) در یک بازه **زمانی یا مکانی** مطرح است، مانند:

تعداد اتفاقات در واحد زمان:
- **تعداد** تماس‌های تلفنی در **یک ساعت**
- **تعداد** مشتریان در **یک روز**
- **تعداد** خرابی‌های اتومبیل در **یک سال**
- **تعداد** مسافران هواپیما در **شش ماه**

تعداد اتفاقات در واحد مکان:
- **تعداد** اتومبیل‌ها در 2 **کیلومتر**
- **تعداد** زدگی‌ها در 100 **متر** پارچه
- **تعداد** غلط‌های املایی در 10 **صفحه**
- **تعداد** نقطه‌ها در 100 **سانتی‌متر**

در چنین حالاتی با استفاده از توزیعی به نام پواسون به تشریح احتمال مسایل می‌پردازیم.

تعریف: فرض کنید تعداد اتفاقات در یک فاصله مشخص از زمان یا مکان مورد نظر باشد؛ در این صورت «X : **تعداد اتفاقات** در یک بازه **زمانی** یا **مکانی**» $(X = 0,1,2,...)$ دارای **توزیع پواسون** است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع پواسون باشد، پارامتر آن λ است و به صورت $X \sim P(\lambda)$ نمایش داده می‌شود.

تابع احتمال	$P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $x = 0,1,2,...$	احتمال x اتفاق در یک بازه زمانی یا مکانی با متوسط λ
امید ریاضی (میانگین)	$E(X) = \lambda$	متوسط تعداد اتفاقات در بازه زمانی یا مکانی
واریانس	$\sigma_X^2 = \lambda$	واریانس تعداد اتفاقات در بازه زمانی یا مکانی
انحراف معیار	$\sigma_X = \sqrt{\lambda}$	
تابع مولد گشتاور	$M_X(t) = e^{-\lambda(1-e^t)}$	

توجه: توزیع پواسون تنها توزیعی است که در آن میانگین و واریانس توزیع با هم برابرند و هر دو مساوی پارامتر توزیع (λ) هستند.

پارامتر پواسون

متوسط تعداد اتفاقات در هر **بازه زمانی یا مکانی** به عنوان پارامتر توزیع پواسون شناخته شده و با نماد λ نمایش داده می‌شود.

اگر $0 \le \lambda \le 10$ باشد، امکان استفاده از توزیع پواسون برای حل مسایل مناسب است، در غیر این صورت $(\lambda > 10)$ بهتر است از تقریب نرمال که بعداً در توزیع نرمال بررسی می‌شود، استفاده شود.

محاسبه احتمال در پواسون

الف) مقدار λ (پارامتر پواسون) را با توجه به زمان یا مکان مشخص می‌کنیم؛ درصورتی‌که زمان یا مکان تغییر کند، با استفاده از تناسب مقدار λ را به دست می‌آوریم.

برای مثال، اگر $\lambda = 2$ مشتری در دقیقه باشد، در 20 ثانیه داریم:

زمان	متوسط تعداد مشتری (λ)
1 دقیقه = 60 ثانیه	2
20 ثانیه	?

$$\Rightarrow \lambda = \frac{2 \times 20}{60} = \frac{2}{3}$$

و در 5 دقیقه خواهیم داشت:

زمان	متوسط تعداد مشتری (λ)
1 دقیقه	2
5 دقیقه	?

$$\Rightarrow \lambda = \frac{2 \times 5}{1} = 10$$

ب) $\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} \\ x = 0,1,2,... \end{cases}$، احتمال وقوع x اتفاق در واحد زمان یا مکان است.

با توجه به آنکه همواره $\sum P(x) = 1$ (مجموع مقادیر تابع احتمال متغیر تصادفی گسسته برابر با 1 است)، برای تابع احتمال پواسون رابطه $\sum_{x=0}^{\infty} P(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = 1$ برقرار است و داریم:

$$\underbrace{\frac{e^{-\lambda}\lambda^0}{0!}}_{P(X=0)} + \underbrace{\frac{e^{-\lambda}\lambda^1}{1!}}_{P(X=1)} + \underbrace{\frac{e^{-\lambda}\lambda^2}{2!}}_{P(X=2)} + \cdots = 1$$

$$\begin{cases} \text{احتمال عدم وقوع اتفاق} = P(X=0) = \dfrac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda} \\ \text{احتمال وقوع 1 اتفاق} = P(X=1) = \dfrac{e^{-\lambda}\lambda^1}{1!} = \lambda e^{-\lambda} \\ \text{احتمال وقوع حداکثر 1 اتفاق} = P(X \le 1) = P(X=0) + P(X=1) = (\lambda+1)e^{-\lambda} \\ \text{احتمال وقوع حداقل 1 اتفاق} = P(X \ge 1) = 1 - P(X=0) = 1 - e^{-\lambda} \\ \text{احتمال وقوع بیش از 1 اتفاق} = P(X > 1) = 1 - P(X=0) - P(X=1) = 1 - (\lambda+1)e^{-\lambda} \\ \vdots \end{cases}$$

حل مسئله ۱ (درک مطلب):

به طور متوسط $\lambda = 12$ اتومبیل در هر ساعت به پمپ بنزین مراجعه می‌کنند، ولی احتمال در 15 دقیقه خواسته شده است؛ بنابراین:

زمان	متوسط تعداد اتومبیل (λ)
1 ساعت = 60 دقیقه	12
15 دقیقه	?

$$\Rightarrow \lambda = \frac{12 \times 15}{60} = 3$$

یعنی به طور متوسط $\lambda = 3$ اتومبیل در 15 دقیقه به پمپ بنزین مراجعه می‌کنند؛ درنتیجه:

$$\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} \xrightarrow[\lambda=3]{x=3} P(X=3) = \dfrac{e^{-3}3^3}{3!} = \dfrac{27e^{-3}}{6} = \dfrac{9}{2}e^{-3} = 4.5e^{-3} \\ X: \text{تعداد اتومبیل‌هایی که در 15 دقیقه به پمپ بنزین مراجعه می‌کنند} \quad , \quad \lambda = 3 \end{cases}$$

حل مسئله ۲ (درک مطلب):

به طور متوسط $\lambda = 3$ زدگی در هر 200 متر پارچه دیده می‌شود، ولی احتمال در بسته‌های 600 متری خواسته شده است؛ بنابراین:

زمان	متوسط تعداد زدگی (λ)
200 متر	3
600 متر	?

$$\Rightarrow \lambda = \frac{3\times 600}{200} = 9$$

یعنی به طور متوسط $\lambda = 9$ زدگی در 600 متر پارچه وجود دارد؛ درنتیجه:

$$\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} \xrightarrow[\lambda=9]{x=1} P(X=1) = \dfrac{e^{-9}9^1}{1!} = 9e^{-9} \\ X: \text{تعداد زدگی‌ها در یک بسته 600 متری} \quad , \quad \lambda = 9 \end{cases}$$

مثال ۱ تعداد مشتریان یک فروشگاه در ساعت خاصی از روز دارای توزیع پواسون با متوسط 4 نفر در ساعت است. مطلوب است محاسبه:

الف) احتمال آنکه در یک ساعت هیچ مشتری وارد نشود.

ب) احتمال آنکه در نیم ساعت 3 مشتری وارد فروشگاه شوند.

ج) احتمال آنکه در 5 ساعت حداکثر یک مشتری وارد فروشگاه شود.

د) احتمال آنکه در یک ربع ساعت حداقل یک مشتری وارد فروشگاه شود.

ه) تابع احتمال تعداد مشتریان در 10 ساعت.

حل:

الف) $\lambda = 4$ نفر در ساعت است و با توجه به آنکه احتمال در ساعت خواسته شده، λ تغییری نمی‌کند؛ بنابراین:

$$\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} \xrightarrow[\lambda=4]{x=0} P(X=0) = \dfrac{e^{-4}4^0}{0!} = e^{-4} \\ X: \text{تعداد مشتری‌هایی که در یک ساعت وارد می‌شوند} \quad , \quad \lambda = 4 \end{cases}$$

ب) $\lambda = 4$ نفر در ساعت است ولی احتمال در نیم ساعت خواسته شده؛ بنابراین:

زمان	متوسط تعداد مشتری (λ)
1 ساعت	4
$\frac{1}{2}$ ساعت	?

$$\Rightarrow \lambda = \frac{4\times\frac{1}{2}}{1} = 2$$

یعنی به طور متوسط $\lambda = 2$ نفر در نیم‌ساعت وارد می‌شود؛ درنتیجه:

$$\begin{cases} P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} \xrightarrow[\lambda=2]{x=3} P(X=3) = \dfrac{e^{-2}2^3}{3!} = \dfrac{8}{6}e^{-2} = \dfrac{4}{3}e^{-2} \\ X: \text{تعداد مشتری‌هایی که در نیم‌ساعت وارد می‌شوند} \quad , \quad \lambda = 2 \end{cases}$$

ج) $\lambda = 4$ نفر در ساعت است اما احتمال در 5 ساعت خواسته شده؛ بنابراین:

زمان	متوسط تعداد مشتری (λ)
1 ساعت	4
5 ساعت	?

$$\Rightarrow \lambda = \frac{4\times5}{1} = 20$$

یعنی به طور متوسط $\lambda = 20$ نفر در 5 ساعت وارد می‌شود؛ درنتیجه:

$$\begin{cases} P(X \le 1) = P(X=0)+P(X=1) = \dfrac{e^{-\lambda}\lambda^0}{0!}+\dfrac{e^{-\lambda}\lambda^1}{1!} = e^{-20}+20e^{-20} = 21e^{-20} \\ X: \text{تعداد مشتری‌هایی که در 5 ساعت وارد می‌شوند} \quad , \quad \lambda = 20 \end{cases}$$

د) $\lambda = 4$ نفر در ساعت است اما احتمال در یک ربع ساعت خواسته شده؛ بنابراین:

زمان	متوسط تعداد مشتری (λ)
1 ساعت	4
$\frac{1}{4}$ ساعت	?

$$\Rightarrow \lambda = \frac{4\times\frac{1}{4}}{1} = 1$$

یعنی به طور متوسط $\lambda = 1$ نفر در یک ربع ساعت وارد می‌شود؛ در نتیجه:

$$\begin{cases} P(X \ge 1) = 1-P(X=0) = 1-\dfrac{e^{-\lambda}\lambda^0}{0!} = 1-e^{-1} \\ X: \text{تعداد مشتری‌هایی که در یک ربع ساعت وارد می‌شوند} \quad , \quad \lambda = 1 \end{cases}$$

ه) $\lambda = 4$ نفر در ساعت است و تابع احتمال در 10 ساعت خواسته شده؛ بنابراین:

زمان	متوسط تعداد مشتری (λ)
1 ساعت	4
10 ساعت	?

$$\Rightarrow \lambda = \frac{4\times10}{1} = 40$$

یعنی به طور متوسط $\lambda = 40$ نفر در 10 ساعت وارد می‌شود؛ در نتیجه:

$$\begin{cases} f(x) = P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!} = \dfrac{e^{-40}40^x}{x!} \\ X: \text{تعداد مشتری‌هایی که در 10 ساعت وارد می‌شوند} \quad , \quad \lambda = 40 \end{cases}$$

مثال ۲ از یک کانال مخابراتی پالس‌هایی با توزیع پواسون با متوسط 3 پالس در دقیقه عبور می‌کنند. می‌خواهیم پیامی به طول 10 ثانیه را از همین کانال مخابره کنیم. احتمال اینکه این پیام با پالس‌های مزبور تداخل پیدا نکند، چقدر است؟

۱) $\frac{1}{2}$ ۲) $2e^{-2}$ ۳) e^{-2} ۴) $e^{-\frac{1}{2}}$

حل: گزینه ۴ درست است.

$\lambda = 3$ پالس در دقیقه است و احتمال در 10 ثانیه خواسته شده؛ بنابراین:

زمان	متوسط تعداد پالس (λ)
1 دقیقه = 60 ثانیه	3
10 ثانیه	?

$$\Rightarrow \lambda = \frac{3\times10}{60} = \frac{1}{2}$$

یعنی به طور متوسط $\lambda=\frac{1}{2}$ پالس در 10 ثانیه ارسال می‌شود.

برای اینکه پیام با پالس‌ها تداخل پیدا نکند باید هنگام مخابره پیام در طول 10 ثانیه هیچ پالسی ارسال نشود؛ درنتیجه:

$$\begin{cases} P(x)=\dfrac{e^{-\lambda}\lambda^x}{x!} \xrightarrow[\lambda=\frac{1}{2}]{x=0} P(X=0)=\dfrac{e^{-\frac{1}{2}}\left(\frac{1}{2}\right)^0}{0!}=e^{-\frac{1}{2}} \\ X: \text{تعداد پالس‌هایی که در 10 ثانیه ارسال می‌شود} \quad , \quad \lambda=\dfrac{1}{2} \end{cases}$$

مثال ۳ به طور متوسط 6 ماشین در دقیقه از یک جاده می‌گذرد. عرض جاده به اندازه عبور یک اتومبیل است. شخصی بدون توجه به عبور ماشین‌ها، عرض جاده را در 10 ثانیه می‌پیماید. احتمال سالم ماندن او چقدر است؟

۱) e^{-1} ۲) $1-e^{-1}$ ۳) e^{-6} ۴) $1-e^{-6}$

حل: گزینه ۱ درست است.

$$X\sim P(\lambda) \rightarrow E(X)=Var(X)=\lambda$$

$\lambda=6$ اتومبیل در دقیقه است و احتمال در 10 ثانیه خواسته شده؛ بنابراین:

زمان	متوسط تعداد اتومبیل (λ)
1 دقیقه = 60 ثانیه	6
10 ثانیه	?

$$\Rightarrow \lambda=\frac{6\times 10}{60}=1$$

یعنی به طور متوسط $\lambda=1$ اتومبیل در 10 ثانیه از جاده می‌گذرد.

چون عرض جاده به اندازه یک اتومبیل است، اگر اتومبیل در این 10 ثانیه رد شود حتماً به شخص برخورد می‌کند، پس برای اینکه شخص سالم بماند، در مدت 10 ثانیه نباید اتومبیلی رد شود؛ در نتیجه:

$$\begin{cases} P(x)=\dfrac{e^{-\lambda}\lambda^x}{x!} \xrightarrow[\lambda=1]{x=0} P(X=0)=\dfrac{e^{-1}\times 1^0}{0!}=e^{-1} \\ X: \text{تعداد اتومبیل‌هایی که در 10 ثانیه از جاده می‌گذرد} \quad , \quad \lambda=1 \end{cases}$$

مثال ۴ در یک توزیع پواسون $P(X=1)=P(X=2)$ برقرار است. ضریب تغییرات متغیر تصادفی X چند درصد است؟

۱) 50 ۲) $50\sqrt{2}$ ۳) $200\sqrt{2}$ ۴) 100

حل: گزینه ۲ درست است.

ابتدا با توجه به تابع احتمال پواسون $P(x)=\dfrac{e^{-\lambda}\lambda^x}{x!}$ داریم:

$$P(X=1)=P(X=2) \rightarrow \frac{\cancel{e^{-\lambda}}\lambda^1}{1!}=\frac{\cancel{e^{-\lambda}}\lambda^2}{2!} \rightarrow \lambda=2$$

حال برای محاسبه درصد ضریب تغییرات (پراکندگی) به صورت زیر عمل می‌کنیم:

$$\begin{cases} CV\times 100=\dfrac{\sigma}{\mu}\times 100=\dfrac{\sqrt{2}}{2}\times 100=50\sqrt{2} \\ \mu=\lambda=2 \\ \sigma=\sqrt{\lambda}=\sqrt{2} \end{cases}$$

تقریب توزیع دوجمله‌ای به پواسون

گاهی در یک توزیع دوجمله‌ای با پارامترهای n و p شرایطی پیش می‌آید که در آن تعداد تکرار آزمایش برنولی (n) زیاد و احتمال موفقیت در هر آزمایش (p) کم است، به طوری که یکی از حالات زیر اتفاق می‌افتد:

$$\begin{cases} \text{I)} \quad n \geq 20\ ,\ p \leq 0.05 \\ \text{II)} \quad n \geq 100\ ,\ np \leq 10 \end{cases}$$

بدیهی است در این شرایط محاسبه احتمال مشکل است. در این وضعیت می‌توانیم به‌جای استفاده از توزیع دوجمله‌ای از توزیع پواسون با پارامتر $\lambda = np$ بهره گیریم تا محاسبات ساده‌تر انجام شود.

هرگاه در یک توزیع دوجمله‌ای با پارامترهای n و p، یکی از شرایط زیر برقرار باشد آن‌گاه توزیع پواسون با پارامتر $\lambda = np$ تقریب مناسبی برای توزیع دوجمله‌ای است.

$$X \sim Bin(n,p) \xrightarrow[(\text{II}):\, n\geq 100\, ,\, np\leq 10]{(\text{I}):\, n\geq 20,\ p\leq 0.05} X \sim P(\lambda = np)$$

مثال ۵ نسبت خرابی کالا در یک کارخانه برابر 0.01 است. احتمال آنکه در 100 کالا حداکثر یک کالای خراب وجود داشته باشد، چقدر است؟

۱) e^2 ۲) e^{-2} ۳) $2e$ ۴) $2e^{-1}$

حل: گزینه ۴ درست است.

با توجه به آنکه احتمال معیوب بودن در هر انتخاب ثابت است $(p = 0.01)$ و نمونه 100تایی از جامعه انتخاب شده است، توزیع تعداد کالای معیوب در نمونه دوجمله‌ای است.

اما شرایط تقریب پواسون با توجه به شرط دوم برقرار است، بنابراین محاسبه احتمال از طریق آن ساده‌تر است.

$$n = 100 \geq 100\ ,\ np = 100 \times 0.01 = 1 \leq 10 \rightarrow \text{تقریب پواسون مناسب است} \rightarrow \lambda = np = 100 \times 0.01 = 1$$

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-1} \times 1^0}{0!} + \frac{e^{-1} \times 1^1}{1!} = 2e^{-1} = 0.736$$

در این مسئله اگر به طور مستقیم برای محاسبه احتمال از توزیع دوجمله‌ای استفاده می‌کردیم، محاسبه بسیار مشکل‌تر بود زیرا:

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \binom{100}{0}(0.01)^0(0.99)^{100} + \binom{100}{1}(0.01)^1(0.99)^{99} = 0.736$$

با این حال پاسخ هر دوی آن‌ها یکسان است.

مثال ۶ طبق آمار سالیانه‌ای که اداره راهنمایی و رانندگی منتشر کرده است، از هر 100 هزار نفر به طور متوسط 3 نفر در اثر حوادث رانندگی کشته می‌شوند. در شهری با 200 هزار نفر جمعیت مطلوب است محاسبه احتمال آنکه:

الف) 4 نفر کشته شوند.

ب) کمتر از 3 نفر کشته شوند.

حل:

با توجه به شرط دوم:

$$n = 200000 \geq 100\ ,\ np = 200000 \times 0.00003 = 6 \leq 10 \rightarrow \text{تقریب پواسون مناسب است.} \rightarrow \lambda = np = 6$$

الف) $P(X = 4) = \dfrac{e^{-\lambda}\lambda^x}{x!} = \dfrac{e^{-6}6^4}{4!}$

ب) $P(X<3)=P(X=0)+P(X=1)+P(X=2)=\sum_{x=0}^{2}\frac{e^{-\lambda}\lambda^{x}}{x!}=e^{-6}+6e^{-6}+\frac{6^2e^{-6}}{2!}=25e^{-6}$

محتمل‌ترین پیشامد

محتمل‌ترین تعداد برای وقوع پیشامد X (تعداد اتفاقات در واحد زمان یا مکان)، $[\lambda]$ است؛ یعنی جزء صحیح (حد پایین) λ است و درصورتی‌که λ عددی صحیح باشد، محتمل‌ترین پیشامد در دو نقطه λ , $\lambda-1$ است.

نکات مهم توزیع پواسون

۱- اگر $X_1, X_2, \dots, X_n$، n متغیر تصادفی مستقل و هم‌توزیع پواسون با پارامتر λ باشند، آن‌گاه مجموع آن‌ها، $Y=\sum_{i=1}^{n}X_i$، دارای توزیع پواسون با پارامتر $n\lambda$ است.

۲- اگر X دارای توزیع پواسون با پارامتر λ_1 و Y مستقل از آن دارای توزیع پواسون با پارامتر λ_2 باشد، آن‌گاه توزیع یکی از آن‌ها (مثلاً X) به شرط اینکه مجموع آن‌ها برابر عددی ثابت باشد، دارای توزیع دوجمله‌ای با پارامترهای $\frac{\lambda_1}{\lambda_1+\lambda_2}$ و n است.

	شرایط	متغیر	توزیع متغیر
۱	$X_1, X_2, \dots, X_n \overset{\text{i.i.d}}{\sim} P(\lambda)$	$Y=\sum_{i=1}^{n}X_i$	$y=P(n\lambda)$
۲	X و Y مستقل $X\sim P(\lambda_1)$, $Y\sim P(\lambda_2)$	$Z=(X=k \mid X+Y=n)$	$Z\sim Bin\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$

در تمام توزیع‌های گسسته $P(X=a)=f(X=a)\geq 0$ است.

توزیع یکنواخت پیوسته (Continuous Uniform Distribution)

مسئله (درک مطلب): درصورتی‌که تابع چگالی $f(x)=\frac{2}{3} \quad ; \quad -1 \le x < \frac{1}{2}$ در نظر گرفته شود:

الف) احتمال آنکه X در فاصله 0 تا $\frac{1}{2}$ باشد چقدر است؟

ب) امید ریاضی و واریانس X چقدر است؟

مقدمه: فاصله پیوسته‌ای را به صورت $\alpha < x < \beta$ در نظر بگیرید که وقوع احتمال در هر نقطه از آن با هر نقطه دیگر در آن برابر باشد؛ در این صورت احتمال وقوع در هر نقطه یا فاصله را می‌توانیم به صورت زیر نمایش دهیم:

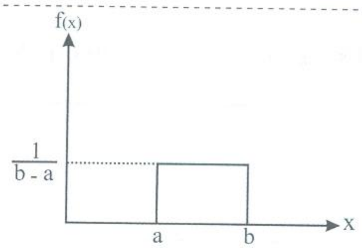
$$P(x)=\frac{\text{طول بازه } x}{\beta-\alpha}$$

استفاده از رابطه بالا به آن علت است که فاصله $\alpha < X < \beta$ به طور یکنواخت بین تمام نقاط توزیع شده است تا احتمال در تمام نقاط با هم برابر باشد در چنین حالتی طبیعی است که:

$$P(c<X<d)=\frac{d-c}{\beta-\alpha}$$

$$P(X=e)=\frac{e-e}{\beta-\alpha}=0$$

تعریف: متغیر تصادفی پیوسته X در بازه a تا b دارای **چگالی یکنواخت** است اگر و فقط اگر تابع چگالی آن به ازای هر مقدار از x برابر با عدد ثابت $\frac{1}{b-a}$ باشد.



پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع یکنواخت پیوسته باشد، پارامترهای آن a و b است و به صورت $X \sim u(a,b)$ نمایش داده می‌شود.

تابع احتمال	$f(x)=\frac{1}{b-a} \quad ; \quad a<x<b$	مقدار ثابت $\frac{1}{b-a}$ به ازای هر مقداری از x
امید ریاضی (میانگین)	$E(X)=\mu=\frac{a+b}{2}$	متوسط مقادیر x در بازه a تا b
واریانس	$\sigma_X^2=\frac{(b-a)^2}{12}$	واریانس مقادیر x در بازه a تا b
انحراف معیار	$\sigma_X=\sqrt{\frac{(b-a)^2}{12}}$	
تابع مولد گشتاور	$M_X(t)=\frac{e^{tb}-e^{ta}}{t(b-a)}$	

نکته: هرگاه تابع چگالی $f(x)$ در بازه $\alpha < x < \beta$ به صورت «$f(x) =$ **ثابت**» باشد، آن‌گاه حتماً X دارای توزیع یکنواخت پیوسته با تابع چگالی $f(x) = \dfrac{1}{\beta - \alpha}$ است و برعکس؛ به عبارت دیگر:

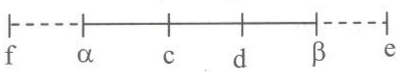
$$\begin{cases} f(x) = \text{ثابت} \\ \alpha < x < \beta \end{cases} \Longleftrightarrow \begin{cases} f(x) = \dfrac{1}{\beta - \alpha} \\ \alpha < x < \beta \end{cases}$$

نکته: تابع توزیع تجمعی متغیر تصادفی یکنواخت پیوسته بر بازه (α, β) به صورت زیر است:

$$F_X(t) = P(X \le t) = \begin{cases} 0 & t < \alpha \\ \dfrac{t - \alpha}{\beta - \alpha} & \alpha \le t < \beta \\ 1 & t \ge \beta \end{cases}$$

محاسبه احتمال

هرگاه $\alpha < x < \beta$ دارای توزیع چگالی یکنواخت باشد، برای محاسبه احتمال در بازه دلخواه $c < x < d$، درصورتی‌که در فاصله مورد نظر باشد، به صورت زیر عمل می‌کنیم:



$$P(c < X < d) = \int_c^d f(x)\,dx = \int_c^d \frac{1}{\beta - \alpha}\,dx = \left[\frac{x}{\beta - \alpha}\right]_c^d = \frac{d - c}{\beta - \alpha}$$

بنابراین:

$$P(\text{بازه}) = \frac{\text{طول بازه}}{\beta - \alpha}$$

$$\begin{cases} P(c < X < d) = \dfrac{d - c}{\beta - \alpha} \quad , \quad P(X < c) = \dfrac{c - \alpha}{\beta - \alpha} \\ P(X > d) = \dfrac{\beta - d}{\beta - \alpha} \quad , \quad P(X = c) = \dfrac{c - c}{\beta - \alpha} = 0 \end{cases}$$

✓ دقت کنید!

اگر بازه احتمال، خارج از بازه $\alpha < x < \beta$ یعنی بازه تابع یکنواخت پیوسته باشد (مانند نقاط f, e)، آن‌گاه:

$$\begin{cases} P(f < X < \alpha) = 0 \\ P(f < X < c) = P(f < X < \alpha) + P(\alpha < X < c) = \dfrac{c - \alpha}{\beta - \alpha} \\ P(d < X < e) = P(d < X < \beta) + P(\beta < X < e) = \dfrac{\beta - d}{\beta - \alpha} \end{cases}$$

حل مسئله (درک مطلب):

با توجه به نکته، ازآنجاکه (ثابت) $f(x) = \dfrac{2}{3}$ است، حتماً تابع یکنواخت پیوسته است، یعنی:

$$f(x) = \frac{1}{\beta - \alpha} = \frac{1}{\frac{1}{2} - (-1)} = \frac{1}{\frac{3}{2}} = \frac{2}{3}$$

الف)

$$P\left(0<X<\frac{1}{2}\right)=\frac{\frac{1}{2}-0}{\frac{1}{2}-(-1)}=\frac{1}{3}$$

ب)

$$\begin{cases} E(X)=\dfrac{\alpha+\beta}{2}=\dfrac{-1+\frac{1}{2}}{2}=-\dfrac{1}{4} \\ \sigma_X^2=\dfrac{(\beta-\alpha)^2}{12}=\dfrac{\left(\frac{1}{2}-(-1)\right)^2}{12}=\dfrac{\frac{9}{4}}{12}=\dfrac{3}{16} \end{cases}$$

مثال ۱ متغیر تصادفی X با چگالی یکنواخت زیر را در نظر بگیرید:

$$f(x)=\begin{cases} \dfrac{1}{8} & ; \quad \alpha<x<\beta \\ 0 & ; \quad \text{در غیر این صورت} \end{cases}$$

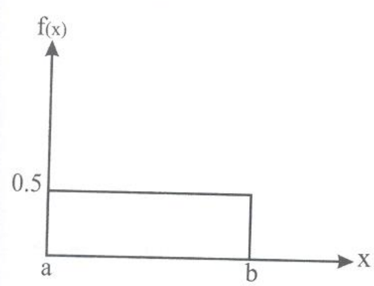
اگر امید ریاضی متغیر تصادفی X برابر 50 باشد، آن‌گاه مقادیر α ,β کدام است؟

۱) $\beta=56\ ,\ \alpha=44$ ۲) $\beta=54\ ,\ \alpha=46$ ۳) $\beta=40\ ,\ \alpha=22.5$ ۴) $\beta=50\ ,\ \alpha=46$

حل: گزینه ۲ درست است.

$$\begin{cases} E(X)=\dfrac{\alpha+\beta}{2}=50 \ \rightarrow\ \alpha+\beta=100 \\ f(x)=\dfrac{1}{\beta-\alpha}=\dfrac{1}{8} \ \rightarrow\ \beta-\alpha=8 \end{cases} \rightarrow \begin{matrix} \beta=54 \\ \alpha=46 \end{matrix}$$

مثال ۲ تابع چگالی یکنواخت پیوسته زیر را در نظر بگیرید:



الف) تابع چگالی $f(x)$ کدام است؟

ب) درصورتی‌که $P(X<C)=0.15$ باشد، مقدار C کدام است؟

ج) امید ریاضی (میانگین) و واریانس X کدام است؟

حل:

الف) با توجه به آنکه a در مبدأ مختصات قرار دارد، $a=0$ است:

$$f(x)=\frac{1}{b-a} \xrightarrow[f(x)=\frac{1}{2}]{a=0} \frac{1}{2}=\frac{1}{b-0} \ \rightarrow\ b=2$$

در نتیجه:

$$f(x)=\frac{1}{2} \quad ; \quad 0\le x\le 2$$

ب) **راه حل اول:**

$$P(X<C)=0.15 \ \rightarrow\ \int_0^C \frac{1}{2}dx=0.15 \ \rightarrow\ \left[\frac{x}{2}\right]_0^C=0.15 \ \rightarrow\ \frac{C}{2}=0.15 \rightarrow C=0.3$$

راه حل دوم:

$$P(X<C)=0.15 \rightarrow \frac{C-\alpha}{\beta-\alpha}=0.15 \rightarrow \frac{C-0}{2-0}=0.15 \rightarrow C=0.3$$

ج)

$$\begin{cases} E(X)=\dfrac{a+b}{2}=\dfrac{0+2}{2}=1 \\ \sigma_X^2=\dfrac{(b-a)^2}{12}=\dfrac{(2-0)^2}{12}=\dfrac{1}{3} \end{cases}$$

مثال ۳ تابع توزیع تجمعی مربوط به تابع چگالی یکنواخت پیوسته $\alpha \le x \le \beta$; $f(x)=\frac{1}{\beta-\alpha}$ کدام است؟

۱) $\frac{x-\alpha}{\beta-\alpha}$ ۲) $\frac{\alpha}{\beta-\alpha}$ ۳) $\frac{x}{\beta-\alpha}$ ۴) هیچ‌کدام

حل: گزینه ۱ درست است.

$$F_X(x)=\int_{\text{حد پایین}}^{x} f(x)dx=\int_{\alpha}^{x}\frac{1}{\beta-\alpha}dx=\left[\frac{x}{\beta-\alpha}\right]_{\alpha}^{x}=\frac{x-\alpha}{\beta-\alpha}$$

مثال ۴ فرض کنید X یک متغیر تصادفی پیوسته یکنواخت با میانگین 1 و واریانس $\frac{4}{3}$ باشد، $P(X<0)$ چقدر است؟

۱) $\frac{1}{2}$ ۲) $\frac{1}{4}$ ۳) $\frac{1}{3}$ ۴) $\frac{1}{5}$

حل: گزینه ۲ درست است.

$$\begin{cases} \mu=\dfrac{\alpha+\beta}{2}=1 \rightarrow \alpha+\beta=2 \\ \sigma^2=\dfrac{(\beta-\alpha)^2}{12}=\dfrac{4}{3} \rightarrow \beta-\alpha=4 \end{cases} \rightarrow \beta=3 \quad , \quad \alpha=-1$$

حال تابع چگالی X عبارت است از:

$$f(x)=\frac{1}{\beta-\alpha}=\frac{1}{3-(-1)}=\frac{1}{4} \quad ; \quad -1<x<4$$

$$P(X<0)=\int_{-1}^{0}\frac{1}{4}dx=\left[\frac{x}{4}\right]_{-1}^{0}=\frac{0-(-1)}{4}=\frac{1}{4}$$

مثال ۵ اگر Y به طور یکنواخت روی فاصله $(0,5)$ توزیع شده باشد. احتمال اینکه هر دو ریشه معادله $4x^2+4xy+y+2=0$ حقیقی باشند چقدر است؟

حل:

برای اینکه معادله درجه دو دارای دو ریشه حقیقی باشد باید Δ آن بزرگ‌تر یا مساوی صفر باشد:

$$\Delta=b^2-4ac=(4y)^2-4(4)(y+2)\ge 0 \rightarrow 16y^2-16y-32\ge 0 \rightarrow y^2-y-2\ge 0$$

$$\rightarrow (y+1)(y-2)\ge 0 \rightarrow \begin{cases} y+1\ge 0 \ , \ y-2\ge 0 \Rightarrow y\ge -1 \ , \ y\ge 2 \quad (I) \\ y+1\le 0 \ , \ y-2\le 0 \Rightarrow y\le -1 \ , \ y\le 2 \quad (II) \end{cases} \xrightarrow{(I)\cap(II)} \begin{cases} y\ge 2 \\ y\le -1 \end{cases}$$

حال می‌دانیم که Y دارای توزیع یکنواخت روی فاصله $(0,5)$ است؛ یعنی:

$$Y \sim \text{یکنواخت}\ (0,5) \rightarrow f(y)=\frac{1}{5-0}=\frac{1}{5} \ ; \ 0<y<5$$

بنابراین احتمال اینکه $Y \le -1$ یا $Y \ge 2$ باشد، برابر است با:

$$P(Y \le -1) + P(Y \ge 2) = 0 + \int_2^5 f(x)\,dx = \int_2^5 \frac{1}{5}dx = \left[\frac{x}{5}\right]_2^5 = \frac{5-2}{5} = \frac{3}{5}$$

مثال ۶ شما ساعت 10 صبح به یک ایستگاه اتوبوس می‌رسید و می‌دانید که اتوبوس در زمانی که به طور یکنواخت بین 10 و 10:30 است، به ایستگاه خواهد رسید.

الف) احتمال اینکه بیش از 10 دقیقه منتظر بمانید چقدر است؟

ب) اگر در ساعت 10:15 هنوز اتوبوس به ایستگاه نرسیده باشد، احتمال اینکه شما حداقل 10 دقیقه دیگر نیز منتظر بمانید چقدر است؟

حل:

X : زمان رسیدن اتوبوس به ایستگاه، دارای توزیع یکنواخت پیوسته در فاصله 0 تا 30 دقیقه است.

$$X \sim \text{یکنواخت}(0,30) \rightarrow f(x) = \frac{1}{30-0} = \frac{1}{30} \quad ; \quad 0 < x < 30$$

شما ساعت 10 به ایستگاه رسیده‌اید؛ یعنی ابتدای فاصله زمانی‌ای که اتوبوس در آن به ایستگاه می‌رسد، پس:

$$\text{(الف)} \quad P(X > 10) = \int_{10}^{30} f(x)\,dx = \int_{10}^{30} \frac{1}{30}dx = \left[\frac{x}{30}\right]_{10}^{30} = \frac{30-10}{30} = \frac{2}{3}$$

$$\text{(ب)} \quad P(X > 15+10 \mid X > 15) = \frac{P(X>25\,,\ X>15)}{P(X>15)} = \frac{P(X>25)}{P(X>15)} = \frac{\int_{25}^{30}\frac{1}{30}dx}{\int_{15}^{30}\frac{1}{30}dx} = \frac{\left[\frac{x}{30}\right]_{25}^{30}}{\left[\frac{x}{30}\right]_{15}^{30}} = \frac{\frac{30-25}{30}}{\frac{30-15}{30}} = \frac{5}{15}$$

مثال ۷ قرار است یک ایستگاه آتش‌نشانی در محلی، کنار جاده‌ای به طول A مستقر شود. اگر حریق در نقطه‌ای که به طور یکنواخت روی فاصله $(0,A)$ است رخ دهد، ایستگاه آتش‌نشانی را باید در چه محلی مستقر کرد تا متوسط فاصله از حریق حداقل شود؟ یعنی a را طوری انتخاب کنید که $E(|X-a|)$ وقتی که X دارای توزیع یکنواخت روی فاصله $(0,A)$ است، حداقل شود.

حل:

a : محل استقرار ایستگاه آتش‌نشانی

$$|x-a| = \begin{cases} x-a & x \ge a \\ a-x & x < a \end{cases}$$

$$X \sim \text{یکنواخت}(0,A) \rightarrow f(x) = \frac{1}{A-0} = \frac{1}{A} \ ; \ 0 < x < A$$

$$E(|X-a|) = \int_0^A |x-a| f(x)\,dx = \int_0^a (a-x)\frac{1}{A}dx + \int_a^A (x-a)\frac{1}{A}dx$$

$$= \frac{1}{A}\left[ax - \frac{x^2}{2}\right]_0^a + \frac{1}{A}\left[\frac{1}{2}x^2 - ax\right]_a^A = \frac{1}{A}\left(a^2 - \frac{a^2}{2} - 0\right) + \frac{1}{A}\left(\frac{1}{2}A^2 - aA - \frac{a^2}{2} + a^2\right) = \frac{a^2}{A} - a + \frac{A}{2}$$

حال می‌خواهیم نقطه a را به‌گونه‌ای بیابیم که $E(|X-a|)$ حداقل شود؛ بنابراین باید از تابع به‌دست‌آمده از $E(|X-a|)$ نسبت به a مشتق گرفته و مساوی صفر قرار دهیم:

$$\frac{\partial E(|X-a|)}{\partial a} = \frac{2}{A}a - 1 = 0 \rightarrow \frac{2a}{A} = 1 \rightarrow a = \frac{A}{2}$$

بنابراین باید محل استقرار ایستگاه آتش‌نشانی درست در وسط جاده باشد.

نکات مهم توزیع یکنواخت پیوسته

۱- اگر متغیر تصادفی X دارای توزیع یکنواخت در فاصله (a,b) باشد، آن‌گاه $Y=\frac{X-a}{b-a}$ دارای توزیع یکنواخت در فاصله $(0,1)$ است.

۲- اگر متغیر تصادفی X دارای توزیع یکنواخت در فاصله $(0,1)$ باشد، آن‌گاه $Y=a+(b-a)X$ دارای توزیع یکنواخت در فاصله (a,b) است.

۳- اگر متغیر تصادفی X هر توزیع دلخواهی داشته باشد، آن‌گاه توزیع تابع توزیع تجمعی آن یکنواخت پیوسته در فاصله $(0,1)$ است.

	شرایط	متغیر	توزیع متغیر
۱	$X\sim U(a,b)$	$Y=\frac{X-a}{b-a}$	$Y\sim U(0,1)$
۲	$X\sim U(0,1)$	$Y=a+(b-a)X$	$Y\sim U(a,b)$
۳	توزیع دلخواه با تابع چگالی $X\sim f(x)$	$Y=F_X(x)$	$Y\sim U(0,1)$

مثال ۸ متغیر تصادفی X دارای توزیع یکنواخت در فاصله $(0,1)$ است. اگر $Y=a+(b-a)X$ باشد که a و b ثابت‌های دلخواه‌اند، کدام عبارت درست است؟

۱) X و Y میانگین‌های مساوی دارند.
۲) Y دارای توزیع یکنواخت در (a,b) است
۳) X و Y توزیع‌های یکسان دارند.
۴) X و Y مستقل‌اند.

حل: گزینه ۲ درست است.

راه حل اول: با توجه به نکته (۲)، توزیع Y یکنواخت در فاصله (a,b) است.

راه حل دوم: با استفاده از امید و واریانس Y به آسانی می‌توان توزیع Y را به دست آورد:

$$E(Y)=E(a+(b-a)X)=a+(b-a)E(X)=a+(b-a)\times\frac{0+1}{2}=\frac{a+b}{2}$$

$$Var(Y)=Var(a+(b-a)X)=(b-a)^2\,Var(X)=(b-a)^2\times\frac{(1-0)^2}{12}=\frac{(b-a)^2}{12}$$

حال واضح است که توزیع Y یکنواخت در فاصله (a,b) است.

توزیع نمایی (Exponential Distribution)

مسئله ۱ (درک مطلب): تعداد مشتریان یک فروشگاه در ساعت خاصی از روز دارای توزیع پواسون با میانگین 4 نفر است. متوسط زمان بین ورود دو مشتری یا ورود اولین مشتری کدام است؟

مسئله ۲ (درک مطلب): زمان ورود مشتریان به یک بانک دارای توزیع نمایی با میانگین $\frac{1}{2}$ ساعت است. مطلوب است احتمال آنکه:

الف) مشتری بعدی دقیقاً رأس 1 ساعت وارد شود.

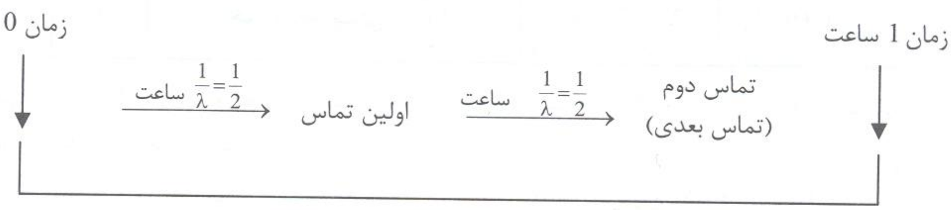
ب) مشتری بعدی تا 3 ساعت بعد وارد شود.

ج) صندوق‌دار بیش از 20 دقیقه منتظر اولین مشتری شود.

د) مشتری بعدی بین 2 تا 3 ساعت وارد شود.

مقدمه: در بخش توزیع‌های گسسته دیدیم که **تعداد وقایعی** که در یک «**فاصله زمانی یا مکانی**» رخ می‌دهند، دارای توزیع پواسون هستند. حال اگر بخواهیم «**فاصله زمانی** بین دو اتفاق متوالی یا **زمان لازم** برای وقوع اولین اتفاق» در توزیع پواسون را بررسی کنیم، با یک متغیر تصادفی پیوسته (زمان) روبه‌رو می‌شویم که به آن **نمایی** گفته می‌شود. به مثال زیر توجه کنید:

فرض کنید تعداد تماس‌های تلفنی در ساعت به یک بانک دارای توزیع پواسون با متوسط $\lambda = 2$ تلفن در ساعت باشد، آن‌گاه:



متوسط $\lambda = 2$ تماس در ساعت

درنتیجه، اگر به طور متوسط 2 تماس تلفنی در ساعت برقرار شود، انتظار داریم در هر $\frac{1}{2}$ ساعت، 1 تماس داشته باشیم.

> **تعریف:** درصورتی‌که تعداد اتفاقات در واحد زمان دارای توزیع پواسون با میانگین λ اتفاق باشد، آن‌گاه متغیر تصادفی« X : **زمان** بین دو اتفاق متوالی یا **زمان** لازم برای اولین اتفاق» $(x \geq 0)$ دارای **توزیع نمایی** با میانگین زمان $\frac{1}{\lambda}$ است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع نمایی باشد، پارامتر آن λ است و به صورت $X \sim E(\lambda)$ نمایش داده می‌شود.

X زمان لازم برای وقوع اتفاق بعدی زمان لازم برای وقوع اولین اتفاق زمان لازم بین دو اتفاق متوالی	$f(x) = \lambda e^{-\lambda x} \quad ; \quad 0 \leq x < \infty$	**تابع احتمال**
متوسط زمان لازم برای وقوع اولین اتفاق یا دو اتفاق متوالی	$E(X) = \mu = \frac{1}{\lambda}$	**امید ریاضی (میانگین)**
واریانس زمان لازم برای وقوع اولین اتفاق یا دو اتفاق متوالی	$\sigma_X^2 = \frac{1}{\lambda^2}$	**واریانس**
	$\sigma_X = \frac{1}{\lambda}$	**انحراف معیار**
	$M_X(t) = \frac{\lambda}{\lambda - t}$	**تابع مولد گشتاور**

نکته: توزیع نمایی تنها توزیعی است که در آن میانگین توزیع با انحراف معیار آن برابر است و درنتیجه همواره ضریب تغییرات (ضریب پراکندگی) آن یک است.

نکته: تابع توزیع تجمعی متغیر تصادفی نمایی به صورت زیر است:

$$F_X(\alpha)=P(X\le\alpha)=1-e^{-\lambda\alpha}$$

رابطه بین λ و $\frac{1}{\lambda}$

$\frac{1}{\lambda}=$ متوسط **زمان** لازم برای وقوع اولین اتفاق یا دو اتفاق متوالی $\longleftrightarrow$ $\boldsymbol{\lambda}=$ متوسط **تعداد** اتفاقات در واحد **زمان**

حل مسئله ۱ (درک مطلب):

با توجه به رابطه λ و $\frac{1}{\lambda}$ داریم:

$\frac{1}{\lambda}=\frac{1}{4}$ متوسط ساعت (15 دقیقه) برای ورود مشتری بعدی یا اولین مشتری $\longrightarrow$ $(\lambda=4)$ متوسط نفر در ساعت

حل مسئله ۲ (درک مطلب):

$\frac{1}{\lambda}=\frac{1}{2}$ متوسط ساعت برای ورود مشتری بعدی یا اولین مشتری $\rightarrow$ $\lambda=2$ متوسط نفر در ساعت

$$f(x)=\lambda e^{-\lambda x}\ ;\ x<0 \xrightarrow{\lambda=2} f(x)=2e^{-2x}\ \ ;\ \ x>0$$

الف) $P(X=1)=0$

ب) $P(X<3)=\int_0^3 2e^{-2x}\,dx=\left[-e^{-2x}\right]_0^3=1-e^{-6}$

ج) $P(X>20\,\text{min})=P\left(X>\frac{1}{3}\,\text{hour}\right)=\int_{\frac{1}{3}}^{\infty}2e^{-2x}\,dx=\left[-e^{-2x}\right]_{\frac{1}{3}}^{\infty}=e^{-\frac{2}{3}}$

د) $P(2<X<3)=\int_2^3 2e^{-2x}\,dx=\left[-e^{-2x}\right]_2^3=e^{-4}-e^{-6}$

مثال ۱ تابع چگالی نمایی زیر را در نظر بگیرید:

$$f(x)=\begin{cases}\frac{1}{2}e^{-\frac{x}{2}} & ;\quad x>0\\ 0 & ;\quad \text{در غیر این صورت}\end{cases}$$

الف) میانگین و واریانس X چقدر است؟

ب) احتمال اینکه X مقداری بین $\frac{1}{2}$ و $\frac{3}{2}$ را اختیار کند، چقدر است؟

حل:

الف) با توجه به آنکه در توزیع نمایی $E(X)=\frac{1}{\lambda}$ و $\sigma_X^2=\frac{1}{\lambda^2}$ است، داریم:

$$f(x)=\lambda e^{-\lambda x}=\frac{1}{2}e^{-\frac{x}{2}}\rightarrow\lambda=\frac{1}{2}\rightarrow\begin{cases}E(X)=\frac{1}{\lambda}=\frac{1}{\frac{1}{2}}=2\\ \sigma^2(X)=\frac{1}{\lambda^2}=\frac{1}{\frac{1}{4}}=4\end{cases}$$

ب)

$$P\left(\frac{1}{2}<X<\frac{3}{2}\right)=\int_{\frac{1}{2}}^{\frac{3}{2}}\frac{1}{2}e^{-\frac{x}{2}}dx=\left[-e^{-\frac{x}{2}}\right]_{\frac{1}{2}}^{\frac{3}{2}}=-e^{-\frac{3}{4}}-\left(-e^{-\frac{1}{4}}\right)=e^{-\frac{1}{4}}-e^{-\frac{3}{4}}$$

مثال ۲ تعداد خرابی‌های ماشین در ماه (30 روز) دارای توزیع پواسون با میانگین 3 خرابی است.

الف) متوسط زمان بین دو خرابی چند روز است؟

ب) احتمال اینکه در 5 روز اول پس از سرویس، ماشین خراب شود، چقدر است؟

حل:

الف) با توجه به رابطه λ و $\frac{1}{\lambda}$:

$\lambda=3$ متوسط خرابی در ماه $\longrightarrow$ $\frac{1}{\lambda}=\frac{1}{3}$ متوسط ماه بین دو خرابی یا برای اولین خرابی

بنابراین به طور متوسط $\frac{1}{\lambda}=\frac{1}{3}$ ماه (10 روز) بین دو خرابی یا برای اولین خرابی زمان لازم است.

ب) ازآنجاکه به طور متوسط $\lambda=3$ خرابی در ماه است، داریم:

$$f(x)=\lambda e^{-\lambda x}=3e^{-3x}\quad,\quad 0<x<\infty$$

حال احتمال مورد نظر را محاسبه می‌کنیم:

$$P(X<\text{روز } 5)=P(X<\text{ماه } \tfrac{1}{6})=\int_0^{\frac{1}{6}}3e^{-3x}dx=\left[-e^{-3x}\right]_0^{\frac{1}{6}}=1-e^{-\frac{1}{2}}$$

دقت کنید که حتماً باید واحد زمانی احتمال خواسته‌شده، منطبق بر واحد زمانی λ باشد. در این سؤال ازآنجاکه واحد زمانی ($\lambda=3$) ماه است، بنابراین واحد زمانی 5 روز برحسب ماه $\frac{5}{30}=\frac{1}{6}$ ماه می‌شود.

مثال ۳ مدت‌زمان تعمیر ماشینی دارای توزیع نمایی با میانگین 3 ساعت است. مطلوب است احتمال آنکه:

الف) ماشین کمتر از 1 ساعت تعمیر شود.

ب) مدت تعمیر بین 1 تا 3 ساعت طول بکشد.

حل:

$\frac{1}{\lambda}=3$ (متوسط در ساعت) $\rightarrow$ $\lambda=\frac{1}{3}$ (متوسط $\frac{1}{3}$ ماشین در ساعت) $\rightarrow$ $f(x)=\frac{1}{3}e^{-\frac{1}{3}x}$

الف) $$P(X<1)=\int_0^1\frac{1}{3}e^{-\frac{1}{3}x}dx=\left[-e^{-3x}\right]_0^1=1-e^{-\frac{1}{3}}$$

ب) $$P(1<X<3)=\int_1^3\frac{1}{3}e^{-\frac{1}{3}x}dx=\left[-e^{-\frac{1}{3}x}\right]_1^3=e^{-\frac{1}{3}}-e^{-1}$$

مثال ۴ به طور متوسط هر 0.5 دقیقه 2 مشتری با توزیع پواسون به گیشه پرداخت بانکی مراجعه می‌کند. احتمال اینکه اولین مشتری بعد از 2 دقیقه وارد شود چقدر است؟

حل:

$\lambda=2$ (مشتری در $\frac{1}{2}$ دقیقه) $\rightarrow$ $\lambda=2\times2=4$ (مشتری در 1 دقیقه) $\rightarrow$ $f(x)=4e^{-4x}\ ;\ x>0$

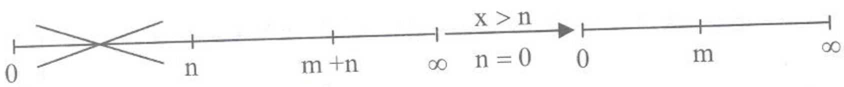
$$P(X>2)=\int_2^{\infty}4e^{-4x}dx=\left[-e^{-4x}\right]_2^{\infty}=e^{-8}$$

خاصیت عدم حافظه (بی‌حافظه بودن)

یکی از مهم‌ترین خصوصیات توزیع نمایی، خاصیت «بی‌حافظه بودن» است:

$$P(X>m+n \mid x>n)=P(X>m)$$

بی‌حافظه بودن به آن مفهوم است که اگر متغیر X با توزیع نمایی تا زمان n اتفاق نیفتاده باشد $(X>n)$، آن‌گاه وقوع آن در m واحد زمان بعدی مستقل از n است،؛به عبارت دیگر وقوع X بعد از m واحد زمانی را می‌توان بدون در نظر گرفتن n، یک توزیع نمایی مستقل دانست که از زمان صفر در نظر گرفته می‌شود:



کاربرد توزیع نمایی

یکی از کاربردهای مهم توزیع نمایی با استفاده از خاصیت «بی‌حافظه بودن»، محاسبه طول عمر قطعات برقی است، زیرا یک دستگاه برقی هر چقدر هم که عمر کرده باشد، طول عمر باقی‌مانده آن ربطی به مدت‌زمان کارکرد قبلی نداشته و دوباره از 0 در نظر گرفته می‌شود؛ به عبارت دیگر هر دستگاه برقی ممکن است هر لحظه خراب شود، از این رو معمولاً طول عمر قطعات برقی شامل گارانتی نمی‌شود!

نکته: در توزیع‌های گسسته، تنها توزیع هندسی دارای خاصیت عدم حافظه است.

مثال ۵ اگر توزیع طول عمر یک مؤلفه صنعتی توزیع نمایی با پارامتر λ باشد و بدانیم که متوسط عمر این مؤلفه 10 سال است، آن‌گاه احتمال اینکه این مؤلفه 5 سال دیگر هم کار کند در حالی که بدانیم 7 سال کار کرده است، چقدر است؟

۱) $\frac{2}{3}$ ۲) $\frac{1}{2}$ ۳) $e^{-\frac{1}{2}}$ ۴) e^{-1}

حل: گزینه ۳ درست است.

طول عمر یک مؤلفه: $X \sim Exp(\lambda)\ ,\ E(X)=\frac{1}{\lambda}=10 \to \lambda=\frac{1}{10} \to f(x)=\lambda e^{-\lambda x}=\frac{1}{10}e^{-\frac{x}{10}}\ ;\ x>0$

می‌دانیم که در توزیع نمایی با توجه به خاصیت بی‌حافظه بودن آن، طول عمر دستگاه مستقل از سال‌های عمر گذشته است؛ بنابراین داریم:

$$P(X\ge 7+5 \mid X\ge 7)=P(X\ge 5)=\int_5^{\infty} f(x)dx=\int_5^{\infty}\frac{1}{10}e^{-\frac{x}{10}}dx=\left[-e^{-\frac{x}{10}}\right]_5^{\infty}=e^{-\frac{1}{2}}$$

نکات مهم توزیع نمایی

نکته: اگر X دارای توزیع نمایی با پارامتر λ باشد $E(X^k)=\frac{k!}{\lambda^k}$.

۱- اگر متغیرهای تصادفی $X_1, X_2, \dots, X_n$ مستقل از هم به ترتیب دارای توزیع نمایی با پارامترهای $\lambda_1, \lambda_2, \dots, \lambda_n$ باشند، آن‌گاه مینیمم آن‌ها دارای توزیع نمایی با پارامتر مجموع λهایشان است.

۲- اگر متغیر تصادفی X یکنواخت پیوسته در فاصله $(0,1)$ باشد، آن‌گاه $Y=-\frac{\ln X}{\lambda}$ دارای توزیع نمایی با پارامتر λ است.

	شرایط	متغیر	توزیع متغیر
۱	$X_1 \sim Exp(\lambda_1),\dots,X_n \sim Exp(\lambda_n)$	$Y=\min\{X_1,X_2,\dots,X_n\}$	$Y \sim Exp(\lambda_1+\dots+\lambda_n)$
۲	$X \sim U(0\,,1)$	$Y=-\frac{\ln X}{\lambda}$	$Y \sim Exp(\lambda)$

مثال ۶ اگر $Y=\min(X_1,X_2,X_3)$ باشد و X_iها $(i=1,2,3)$ مستقل از یکدیگر و دارای توزیع نمایی با پارامتر 2 باشند، $E(Y)$ کدام است؟

۱) 6 ۲) 2 ۳) $\frac{1}{6}$ ۴) $\frac{1}{2}$

حل: گزینه ۳ درست است.

با توجه به نکته (۱) داریم:

$$X_i \sim \text{نمایی}(\lambda=2)\ ,\ Y=\min(x_i) \ \to\ Y \sim \text{نمایی}(\lambda=2+2+2=6) \to E(Y)=\frac{1}{\lambda}=\frac{1}{6}$$

اثبات: ابتدا تابع توزیع تجمعی Y را به دست آورده و از آن مشتق می‌گیریم تا به تابع چگالی Y برسیم:

$$F_Y(\alpha)=P(Y\le\alpha)=P(\min(X_i)\le\alpha)=1-P(\min(X_i)\ge\alpha)=1-P(X_1>\alpha,X_2>\alpha,X_3>\alpha)$$

$$\underset{\text{ها مستقل}\ X_i}{=\!=}\ 1-P(X_1>\alpha)P(X_2>\alpha)P(X_3>\alpha)=1-P(X>\alpha)^3=1-\left(\int_\alpha^\infty f(x)dx\right)^3$$

$$=1-\left(\int_\alpha^\infty 2e^{-2x}dx\right)^3=1-\left(\left[-e^{-2x}\right]_\alpha^\infty\right)^3=1-\left(e^{-2\alpha}\right)^3=1-e^{-6\alpha}$$

$$f_Y(y)=F'_Y(y)=6e^{-6y}\ \to\ Y \sim \text{نمایی}(\lambda=6)$$

یادآوری: هرگاه مینیمم یک مجموعه از عددی بزرگ‌تر باشد، تک تک اعضای مجموعه نیز از آن عدد بزرگ‌تر هستند.

توزیع گاما (Gamma)

تعریف: توزیع گاما به عنوان مدت‌زمان انتظار برای r اتفاق در نظر گرفته می‌شود، به این صورت که اگر اتفاقات به طور تصادفی در طول زمان رخ دهند، آن‌گاه مدت‌زمانی که فرد باید منتظر باشد تا r اتفاق رخ دهد، دارای توزیع گاما با پارامترهای (r,λ) است.

پارامترهای توزیع

هرگاه متغیر تصادفی X دارای توزیع گاما باشد، پارامترهای آن λ, r است و به صورت $X \sim G(r,\lambda)$ یا $X \sim \Gamma(r,\lambda)$ نمایش داده می‌شود.

X : مدت‌زمان انتظار برای r اتفاق	$f(x)=\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x} \quad ; \quad 0<x<\infty \quad , \quad r\geq 1$ $\Gamma(r)=(r-1)!$	تابع احتمال
متوسط مدت‌زمان انتظار برای r اتفاق	$E(X)=\dfrac{r}{\lambda}$	امید ریاضی (میانگین)
واریانس مدت‌زمان انتظار برای r اتفاق	$\sigma_X^2=\dfrac{r}{\lambda^2}$	واریانس
	$\sigma_X=\dfrac{\sqrt{r}}{\lambda}$	انحراف معیار
	$M_X(t)=\left(\dfrac{\lambda}{\lambda-t}\right)^r$	تابع مولد گشتاور

یادآوری: $\Gamma(r)=\int_0^\infty x^{r-1}e^{-x}\,dx=(r-1)!$

مثال ۱ اگر $f(x)=\dfrac{\lambda e^{-\lambda x}}{\Gamma(\alpha)}(\lambda x)^{\alpha-1}\ ;\ x\geq 0$ تابع مولد گشتاور آن برابر است با:

۱) $\dfrac{\lambda}{(\lambda-t)^\alpha}$ ۲) $\dfrac{\lambda^\alpha}{\lambda+t}$ ۳) $\left(\dfrac{\lambda}{\lambda-t}\right)^\alpha$ ۴) $\dfrac{1}{(\lambda-t)^\alpha}$

حل: گزینه ۳ درست است.

راه حل اول: با توجه به اینکه $f(x)$ تابع چگالی توزیع گاما با پارامترهای λ,α است، تابع مولد گشتاور آن برابر است با:

$$\mu_X(t)=\left(\frac{\lambda}{\lambda-t}\right)^\alpha$$

راه حل دوم:

$$M_X(t)=E\left(e^{tX}\right)=\int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{tx}e^{-\lambda x}\,dx=\frac{\lambda^\alpha}{(\lambda-t)^\alpha}\underbrace{\int_0^\infty \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-(\lambda-t)x}\,dx}_{\text{توزیع گاما با } \alpha,\ \lambda-t}=\frac{\lambda^\alpha}{(\lambda-t)^\alpha}\times 1$$

$$\rightarrow\ M_X(t)=\left(\frac{\lambda}{(\lambda-t)}\right)^\alpha$$

نکات مهم توزیع گاما

نکته: اگر X دارای توزیع گاما با پارامتر (r,λ) باشد $E\left(X^k\right)=\dfrac{\Gamma(r+k)}{\lambda^k\ \Gamma(r)}$.

۱- اگر متغیر تصادفی X دارای توزیع نمایی با پارامتر λ باشد، در حقیقت دارای توزیع گاما با پارامترهای λ و 1 نیز هست.

۲- اگر $X_1,X_2,\dots,X_n$، n متغیر تصادفی مستقل و هم‌توزیع نمایی با پارامتر λ باشند، آن‌گاه مجموع آن‌ها دارای توزیع گاما با پارامترهای λ و n است.

۳- اگر متغیر تصادفی X دارای توزیع گاما با پارامترهای λ و r باشد، آن‌گاه هر مضربی از X نیز دارای توزیع گاما با پارامترهای $\dfrac{\lambda}{c}$ و r است.

	شرایط	متغیر	توزیع متغیر
۱	$X \sim Exp(\lambda)$	$Y=X$	$Y \sim G(1,\lambda)$
۲	$X_1,X_2,\dots,X_n \underset{}{\overset{i.i.d}{\sim}} Exp(\lambda)$	$Y=\sum_{i=1}^{n} X_i$	$Y \sim G(n,\lambda)$
۳	$X \sim G(r,\lambda)$	$y=cx$	$Y \sim G\left(r,\dfrac{\lambda}{c}\right)$

مثال ۲ تابع چگالی توأم $f(x,y)=e^{-(x+y)}\ ;\ x>0\ ,\ y>0$ مفروض است. متغیر تصادفی $Z=X+Y$ است. در این صورت $E(Z)$ کدام است؟

۱) 1 ۲) 2 ۳) 3 ۴) 4

حل: گزینه ۲ درست است.

$$f(x,y)=e^{-(x+y)}=e^{-x}e^{-y} \quad ; \quad x>0\ ,\ y>0$$

$$f(x)=e^{-x}\ ;\ x>0 \qquad , \qquad f(y)=e^{-y}\ ;\ y>0$$

ازآنجاکه حدود X و Y به هم وابسته نیست و تابع چگالی توأم را می‌توان به صورت دو تابع مجزا از X و Y نوشت، بنابراین X و Y مستقل بوده و با کمی دقت متوجه می‌شویم که هر دو متغیر دارای توزیع نمایی با پارامتر $\lambda=1$ هستند. همچنین Z، مجموع دو متغیر تصادفی مستقل نمایی با پارامتر مشابه λ، دارای توزیع گاما با پارامترهای λ و $r=2$ است و امید ریاضی توزیع گاما برابر است با:

$$E(Z)=\frac{r}{\lambda}=\frac{2}{1}=2$$

توزیع نرمال (Normal Distribution)

مقدمه: توزیع نرمال یا توزیع زنگی (به‌هنجار) به عنوان یک توزیع متقارن، مهم‌ترین توزیع پیوسته است و دارای کاربردهای فراوانی از جمله موارد زیر است:

۱- بسیاری از پدیده‌های طبیعی دارای توزیع نرمال هستند.

۲- بسیاری از توزیع‌ها در شکل حدی دارای تقریب نرمال هستند (دوجمله‌ای، پواسون، کای دو).

تعریف: اگر متغیر تصادفی پیوسته X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد، تابع چگالی و تابع مولد گشتاور آن به شرح زیر است:

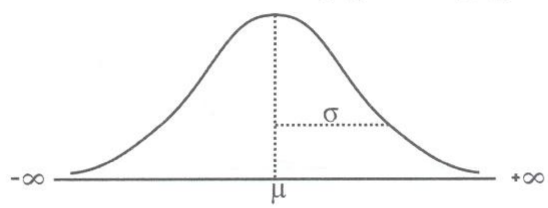
$$f(x)=\frac{1}{\sqrt{2\pi}\ \sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\quad ;\quad -\infty<x<+\infty$$

$$M_X(t)=e^{\mu t+\frac{1}{2}\sigma^2t^2}$$

پارامترهای توزیع

اگر متغیر تصادفی X دارای توزیع نرمال باشد، پارامترهای آن μ (میانگین) و σ^2 (واریانس) است و از شکل نمادین زیر برای نمایش آن استفاده می‌شود:

$$\boxed{X\sim N\left(\mu,\sigma^2\right)}$$



نکته: با توجه به تابع مولد گشتاور توزیع نرمال، نتیجه زیر به دست می‌آید:

$$\mu=E(X)=0\quad\longrightarrow\quad E\left(X^3\right)=E\left(X^5\right)=\cdots=0$$

به عبارت بهتر هر گاه میانگین $\mu=E(X)$ در یک توزیع نرمال برابر صفر باشد، میانگین (امید ریاضی) تمام توان‌های فرد نیز برابر صفر می‌شود.

مثال اگرمتغیر تصادفی X نرمال به فرم $X\sim N\left(4,\frac{1}{2}\right)$ باشد، $E\left(e^{2X}\right)$ کدام است؟

۱) $e^{\frac{1}{2}}$ ۲) e^4 ۳) e^9 ۴) e^8

حل: گزینه ۳ درست است.

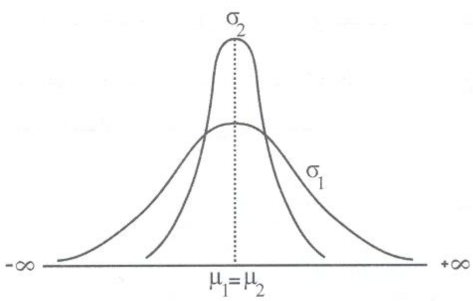
با توجه به تابع مولد گشتاور نرمال داریم:

$$E\left(e^{2X}\right)\xlongequal{t=2}e^{\mu\times2+\frac{1}{2}\sigma^2(2)^2}\xlongequal{\mu=4\ ,\ \sigma^2=\frac{1}{2}}e^{4\times2+\frac{1}{2}\times2}=e^9$$

تأثیرات μ و σ^2 روی منحنی نرمال

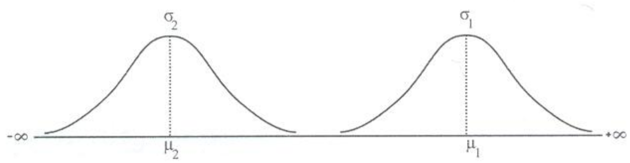
در صورت معلوم بودن پارامترهای μ و σ^2، به راحتی می‌توانیم توزیع را مشخص و منحنی آن را ترسیم کنیم. وضعیت‌های زیر تأثیرات μ و σ^2 را روی منحنی نرمال نشان می‌دهند.

دو منحنی نرمال به صورت $N\left(\mu_1, \sigma_1^2\right)$ و $N\left(\mu_2, \sigma_2^2\right)$ را در نظر می‌گیریم.



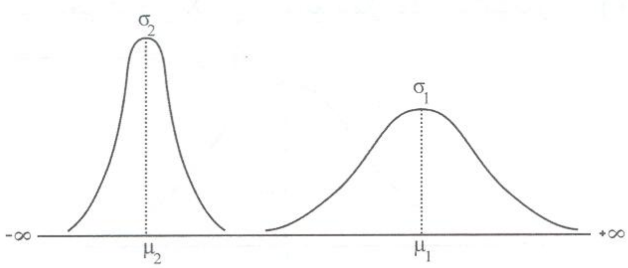
الف) $\sigma_1 > \sigma_2 \quad , \quad \mu_1 = \mu_2$

(میانگین‌ها برابر و انحراف معیارها متفاوت)



ب) $\sigma_1 = \sigma_2 \quad , \quad \mu_1 > \mu_2$

(میانگین‌ها متفاوت و انحراف معیارها برابر)



ج) $\sigma_1 > \sigma_2 \quad , \quad \mu_1 > \mu_2$

(میانگین‌ها متفاوت و انحراف معیارها متفاوت)

نتیجه:

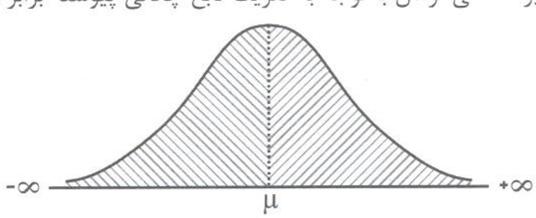
۱) افزایش میانگین، منحنی را به سمت راست و کاهش میانگین، منحنی را به سمت چپ انتقال می‌دهد.

۲) افزایش انحراف معیار، ارتفاع منحنی را کوتاه‌تر (پخ‌تر)، پراکندگی را بیشتر و تمرکز حول میانگین را کمتر می‌کند و کاهش انحراف معیار، ارتفاع منحنی را بلندتر (کشیده‌تر)، پراکندگی را کمتر و تمرکز حول میانگین را بیشتر می‌کند.

خصوصیات توزیع نرمال

۱- سطح زیر منحنی نرمال با توجه به تعریف تابع چگالی پیوسته برابر 1 است.

$$\boxed{\int_{-\infty}^{+\infty} f(x)\,dx = 1}$$



۲- پارامترهای میانگین (μ)، میانه (Md) و مد (نما) (Mo) در توزیع نرمال با هم برابر هستند.

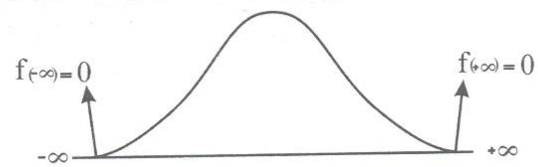
$$\boxed{\mu = Md = Mo}$$

۳- با توجه به برابری $\mu = Mo$ حداکثر مقدار تابع در نقطه $x = \mu$ به دست می‌آید؛ به عبارت دیگر:

$$\boxed{f'_X(\mu) = 0 \quad \rightarrow f(\mu) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{1}{2}\left(\frac{\mu-\mu}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}\ \sigma}}$$

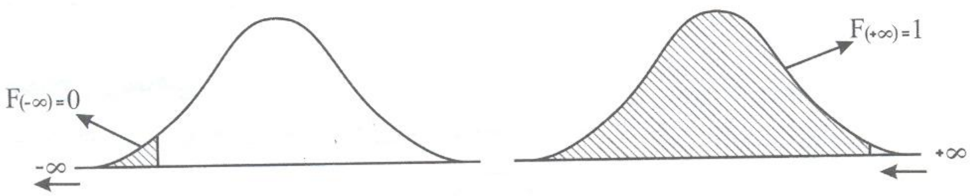
۴- در منحنی توزیع نرمال، با فاصله گرفتن از میانگین (μ) در هر دو سمت منحنی، به محور xها نزدیک می‌شویم به طوری که:

$$\lim_{x \to +\infty} f(x) = \lim_{x \to -\infty} f(x) = 0 \rightarrow f(-\infty) = f(+\infty) = 0$$



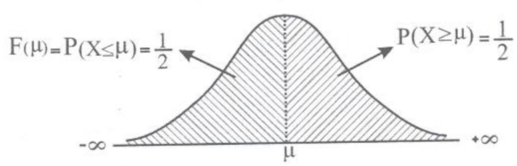
نتیجه:

$$\begin{cases} \lim\limits_{x \to +\infty} F_X(x) = P(X \le +\infty) = 1 \\ \lim\limits_{x \to -\infty} F_X(x) = P(X \le -\infty) = 0 \end{cases} \rightarrow \begin{cases} F(+\infty) = 1 \\ F(-\infty) = 0 \end{cases}$$



۵- خط $x = \mu$ محور تقارن منحنی است (با توجه به برابری $Md = \mu$) و درنتیجه:

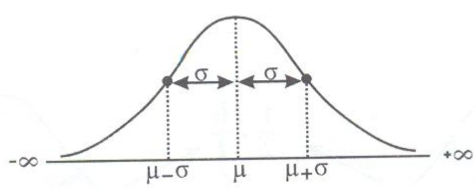
$$P(X \le \mu) = P(X \ge \mu) = \frac{1}{2} \quad \rightarrow \quad F(\mu) = \frac{1}{2}$$



یادآوری: با توجه به پیوسته بودن توزیع نرمال $P(X = \mu) = 0$ است، بنابراین:

$$P(X \ge \mu) = P(X > \mu) \quad , \quad P(X \le \mu) = P(X < \mu)$$

۶- نقاط $\mu \pm \sigma$ تنها دو نقطه عطف منحنی نرمال هستند به طوری که $f''(\mu \pm \sigma) = 0$.



مثال ۱ اگر متغیر تصادفی X دارای توزیع نرمال با میانگین 50 باشد، $P(X \le 50)$ کدام است؟

۱) 0 ۲) $\frac{3}{4}$ ۳) $\frac{1}{2}$ ۴) 1

حل: گزینه ۳ درست است.

$$\begin{cases} P(X \le \mu) = P(X \ge \mu) = \dfrac{1}{2} \\ \mu = 50 \end{cases}$$

مثال ۲ اگر متغیر تصادفی X دارای توزیع نرمال به فرم $X \sim N(60,25)$ باشد، مقدار مد توزیع کدام است؟

۱) 60 ۲) 30 ۳) 25 ۴) 5

حل: گزینه ۱ درست است.

در توزیع نرمال به صورت $X \sim N(60,25)$، مقدار $\mu = 60$ و $\sigma^2 = 25$ است، بنابراین چون در توزیع نرمال میانگین، میانه و مد بر هم منطبق هستند، مقدار مد نیز برابر با 60 است.

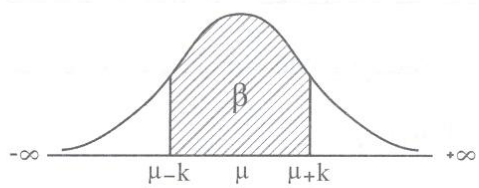
مثال ۳ اگر متغیر تصادفی X، دارای توزیع نرمال باشد، حداکثر مقدار تابع کدام است؟

۱) μ ۲) $\dfrac{1}{\sqrt{2\pi}\,\sigma}$ ۳) $\dfrac{1}{\sqrt{2\pi\sigma}}$ ۴) σ

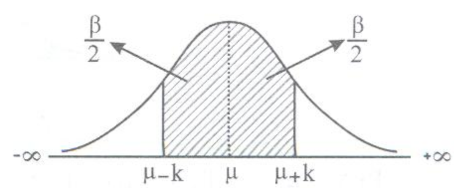
حل: گزینه ۲ درست است.

تقارن و سطح زیرمنحنی

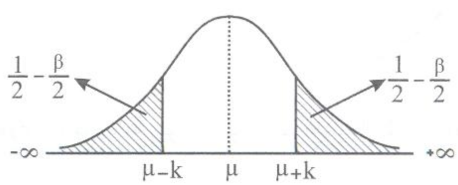
به ازای هر نقطه دلخواه k در روابط زیر، به علت تقارن منحنی نسبت به خط $x = \mu$ داریم:



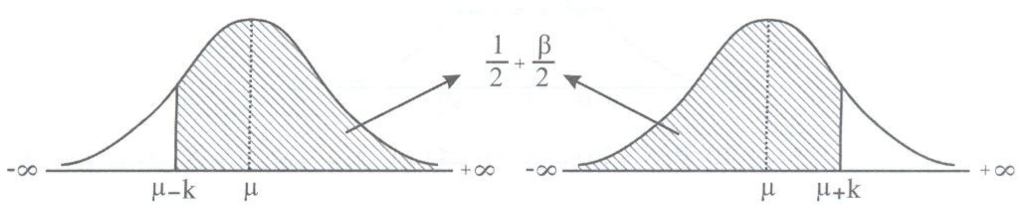
$$P(\mu - k \le X \le \mu + k) = \beta$$



$$1)\ P(\mu - k \le X \le \mu) = P(\mu \le X \le \mu + k) = \frac{\beta}{2}$$



$$2)\ P(X \le \mu - k) = P(X \ge \mu + k) = \frac{1}{2} - \frac{\beta}{2}$$

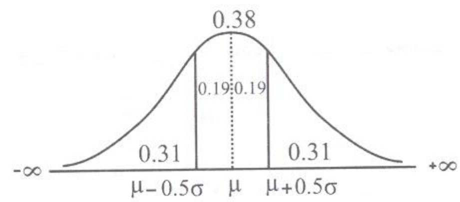


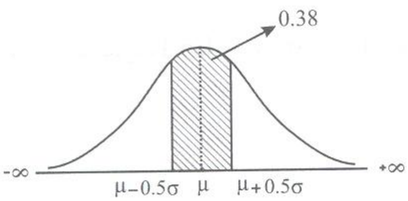
$$3)\ P(X \ge \mu - k) = P(X \le \mu + k) = \frac{1}{2} + \frac{\beta}{2}$$

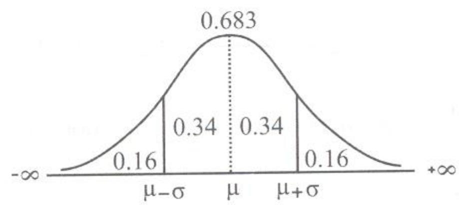
انحرافات حول میانگین

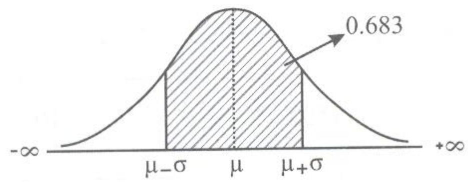
در توزیع نرمال، مقدار احتمال برای 0.5, 1, 2 و 3 انحراف معیار حول میانگین به صورت زیر است:

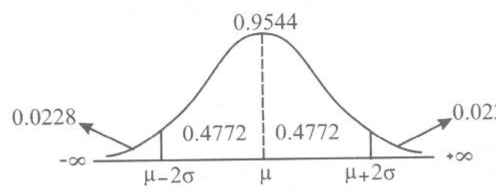
احتمال در فاصله 0.5 انحراف معیار (0.5σ) حول میانگین:	$P(\mu-0.5\sigma \le X \le \mu+0.5\sigma) \simeq 0.38$
احتمال در فاصله 1 انحراف معیار (σ) حول میانگین:	$P(\mu-\sigma \le X \le \mu+\sigma) = 0.683 \simeq 0.68$
احتمال در فاصله 2 انحراف معیار (2σ) حول میانگین:	$P(\mu-2\sigma \le X \le \mu+2\sigma) = 0.9544 \simeq 0.95$
احتمال در فاصله 3 انحراف معیار (3σ) حول میانگین:	$P(\mu-3\sigma \le X \le \mu+3\sigma) = 0.997$

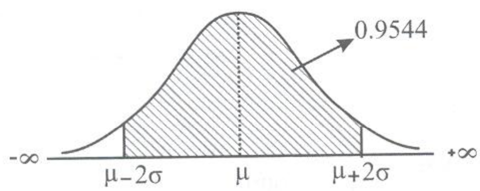


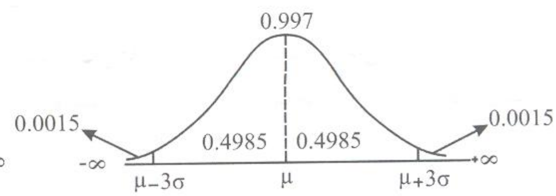


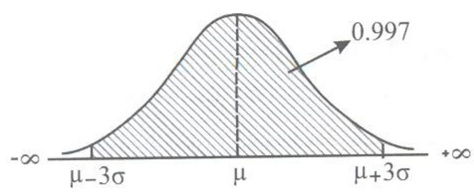




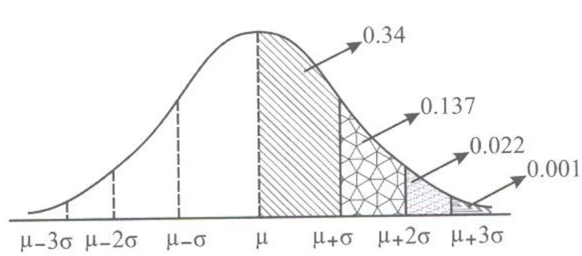








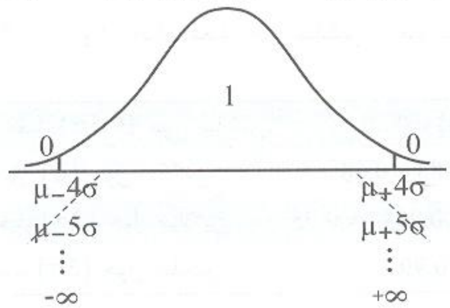
در حالت کلی مقدار احتمال در فواصل 1, 2 و 3 انحراف معیار بالای میانگین به طور تقریبی به صورت زیر است:



نکته: با توجه به احتمال مربوط به 3 انحراف معیار حول میانگین (0.997) و با در نظر گرفتن این موضوع که سطح کل زیر منحنی نرمال 1 است، می‌توان نتیجه گرفت که:

اولاً: احتمال یا سطح زیر منحنی خارج از فاصله $(\mu - 3\sigma\ ,\ \mu + 3\sigma)$ تقریباً برابر 0 است.

ثانیاً: احتمال یا سطح زیر منحنی، درون فواصل بیش از 3 انحراف معیار یعنی $(\mu \pm 4\sigma),(\mu \pm 5\sigma)$,... تقریباً برابر 1 است.



نکته: برای هر بازه به صورت $(\mu - k\sigma\ ,\ \mu + k\sigma)$ همواره می‌توان μ,σ و k را با استفاده از روابط زیر محاسبه کرد:

$$\left(\underbrace{\mu - k\sigma}_{a}\ ,\ \underbrace{\mu + k\sigma}_{b}\right) \longrightarrow \begin{cases} \mu = \dfrac{a+b}{2} \\ k\sigma = \dfrac{b-a}{2} \end{cases}$$

مثال ۱ درآمد حدود 95% از رانندگان تاکسی در روز بین 1000 تا 5000 تومان است. با فرض نرمال بودن توزیع درآمد، انحراف معیار درآمد این صنف کدام است؟

۱) 1000 ۲) 1500 ۳) 2000 ۴) 666.6

حل: گزینه ۱ درست است.

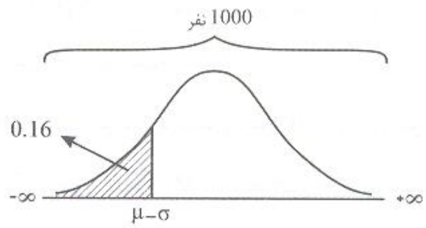
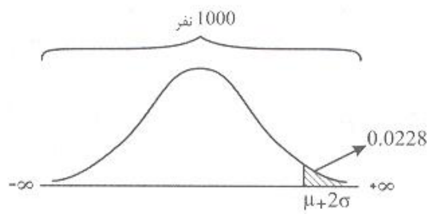
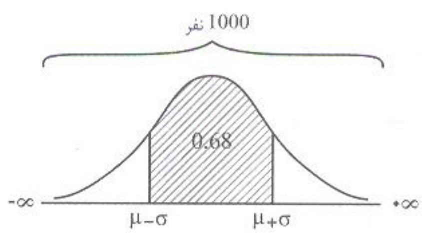
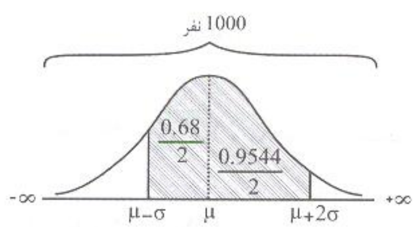
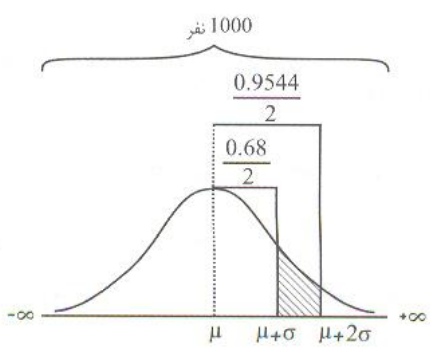
می‌دانیم در توزیع نرمال: $P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544 \simeq 0.95 \rightarrow k = 2$

بنابراین با توجه به نکته ۲ داریم:

$$\left(\underbrace{1000}_{a}\ ,\ \underbrace{5000}_{b}\right) \longrightarrow \begin{cases} \mu = \dfrac{a+b}{2} = \dfrac{1000+5000}{2} = 3000 \\ k\sigma = \dfrac{b-a}{2} \longrightarrow 2\sigma = \dfrac{5000-1000}{2} \longrightarrow \sigma = 1000 \end{cases}$$

مثال ۲ برای بررسی وضعیت تحصیلی 1000 دانش‌آموز، نمرات آن‌ها را بر روی محور نرمال برده‌اند. برای این‌کار ابتدا میانگین (μ) و انحراف معیار (σ) نمرات را به دست آورده و سپس نمرات را روی منحنی نرمال در نظر گرفته‌اند. مشخص کنید چند نفر از دانش‌آموزان در محدوده نمرات مشخص شده قرار دارند؟

حل:

محدوده نمرات		
$x < \mu - \sigma$		$\rightarrow = 1000 \times 0.16 = 160$ نفر $\rightarrow$
$x > \mu + 2\sigma$		$\rightarrow = 1000 \times 0.0228 = 22.8 \simeq 23$ نفر $\rightarrow$
$\mu - \sigma < x < \mu + \sigma$		$\rightarrow = 1000 \times (0.68) = 680$ نفر
$\mu - \sigma < x < \mu + 2\sigma$		$\rightarrow = 1000 \times \left(\frac{0.9544}{2} + \frac{0.68}{2}\right) = 1000 \times 0.8172 = 817.2 \simeq 818$ نفر
$\mu + \sigma < x < \mu + 2\sigma$		$\rightarrow = 1000 \times \left(\frac{0.9544}{2} - \frac{0.68}{2}\right) = 1000 \times 0.1372 = 137.2 \simeq 138$ نفر

ترکیب خطی از توزیع نرمال

هرگاه X دارای توزیع نرمال $X \sim N\left(\mu, \sigma^2\right)$ باشد، هر ترکیب خطی از آن به صورت $Y = bX + a$ نیز نرمال است؛ در این وضعیت کافی است امید و واریانس Y را محاسبه کنیم.

مثال ۱ متغیر تصادفی X بر طبق قانون نرمال با امید ریاضی 150 و واریانس 64 توزیع شده است. اگر متغیر تصادفی Y براساس معادله $Y = \frac{1}{2}X + 25$ از X پیروی کند، آن‌گاه تابع چگالی احتمال متغیر تصادفی Y عبارت است از:

۱) $f(y) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{(y-100)^2}{32}}$

۲) $f(y) = \frac{1}{4\sqrt{2\pi}} e^{-\frac{(y-100)^2}{16}}$

۳) $f(y) = \frac{1}{16\sqrt{2\pi}} e^{-\frac{(y-150)^2}{64}}$

۴) $f(y) = \frac{1}{16\sqrt{2\pi}} e^{-\frac{(y-75)^2}{32}}$

حل: گزینه ۱ درست است.

$$\begin{cases} X \sim N\left(\mu_X = 150, \sigma_X^2 = 64\right) \\ Y = \frac{1}{2}X + 25 \rightarrow \begin{cases} E(Y) = \frac{1}{2}E(X) + 25 = 100 \rightarrow \mu_Y = 100 \\ \sigma^2(Y) = \frac{1}{4}\sigma_X^2 = 16 \rightarrow \sigma_Y = 4 \end{cases} \end{cases}$$

$$f(y) = \frac{1}{\sqrt{2\pi}\ \sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma}\right)^2} = \frac{1}{\sqrt{2\pi}\ 4} e^{-\frac{1}{2}\left(\frac{y-100}{4}\right)^2} = \frac{1}{4\sqrt{2\pi}} e^{-\frac{(y-100)^2}{32}}$$

مثال ۲ اگر $X_1 \sim N(4, 64)$ و $X_2 \sim N(9, 36)$ باشد و X_1 و X_2 از هم مستقل باشند، توزیع $X_1 \pm X_2$ کدام است؟

حل:

باتوجه به آنکه $E(X_1)=4$، $\sigma_{x_1}^2=64$، و $E(X_2)=9$ و X_1 و X_2 مستقل‌اند، داریم:

$$\begin{cases} E(X_1 + X_2) = E(X_1) + E(X_2) = 9 + 4 = 13 \\ \sigma^2(X_1 + X_2) = \sigma_{x_1}^2 + \sigma_{x_2}^2 = 64 + 36 = 100 \end{cases} \rightarrow X_1 + X_2 \sim N(13, 100)$$

$$\begin{cases} E(X_1 - X_2) = E(X_1) - E(X_2) = 4 - 9 = -5 \\ \sigma^2(X_1 - X_2) = \sigma_{x_1}^2 + \sigma_{x_2}^2 = 64 + 36 = 100 \end{cases} \rightarrow X_1 - X_2 \sim N(-5, 100)$$

مثال ۳ اگر T_1 دارای توزیع نرمال با میانگین θ و واریانس σ^2 (معلوم)، T_2 مستقل از T_1 و دارای توزیع نرمال با میانگین θ و واریانس $2\sigma^2$ باشد، واریانس $\frac{1}{\sigma}\left(T_2 - T_1\right)$ برابر است با:

۱) 1 ۲) 3 ۳) σ^2 ۴) $3\sigma^2$

حل: گزینه ۲ درست است.

چون T_1 و T_2 مستقل‌اند، داریم:

$$Var\left[\frac{1}{\sigma}\left(T_2 - T_1\right)\right] = \frac{1}{\sigma^2} Var\left(T_2 - T_1\right) = \frac{1}{\sigma^2}\left[Var(T_2) + Var(T_1)\right] = \frac{1}{\sigma^2}\left(2\sigma^2 + \sigma^2\right) = 3$$

توزیع $\sum_{i=1}^{n} X_i$

اگر $X_1, X_2, \ldots, X_n$ مستقل از هم و دارای توزیع نرمال $X_i \sim N\left(\mu, \sigma^2\right)$ باشند، آن‌گاه $\sum_{i=1}^{n} X_i$ دارای توزیع نرمال $N\left(n\mu, n\sigma^2\right)$ است:

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$$

نتیجه:

مجموع متغیرهای مستقل نرمال، همواره نرمال است و میانگین و واریانس مجموع متغیرهای مستقل نرمال برابر است با مجموع میانگین‌ها و مجموع واریانس‌ها.

توزیع نرمال استاندارد (Standard Normal Distribution)

ازآنجاکه انتگرال‌گیری از تابع چگالی نرمال برای محاسبه احتمال در فاصله محدود غیرممکن است، با استفاده از روش‌های عددی، جدولی به دست آمده است که مقادیر احتمال مربوط به توزیع نرمال با میانگین 0 و واریانس 1 در آن قابل محاسبه است؛ بنابراین برای محاسبه احتمال در هر توزیع نرمال به صورت $X \sim N\left(\mu , \sigma^2\right)$ ابتدا باید توزیع نرمال را به صورت $N(0,1)$ تبدیل کنیم، سپس از روی جدول نرمال استاندارد مقدار عددی احتمال را به دست آوریم.

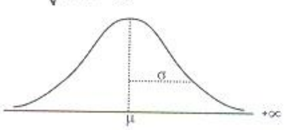
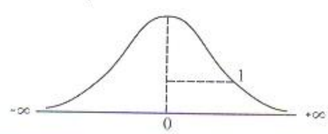
تبدیل نرمال $N\left(\mu , \sigma^2\right)$ به نرمال استاندارد $N(0,1)$

با استفاده از تغییر متغیر $Z = \frac{X-\mu}{\sigma}$، تابع نرمال $X \sim N\left(\mu , \sigma^2\right)$ به صورت زیر به تابع نرمال استاندارد $Z \sim N(0,1)$ تبدیل می‌شود:

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = E\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{E(X)}{\sigma} - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$\sigma^2(Z) = \sigma^2\left(\frac{X-\mu}{\sigma}\right) = \sigma^2\left(\frac{X}{\sigma} - \underbrace{\cancel{\frac{\mu}{\sigma}}}_{0}\right) = \left(\frac{1}{\sigma}\right)^2 \sigma_X^2 = \frac{\sigma^2}{\sigma^2} = 1$$

درنتیجه:

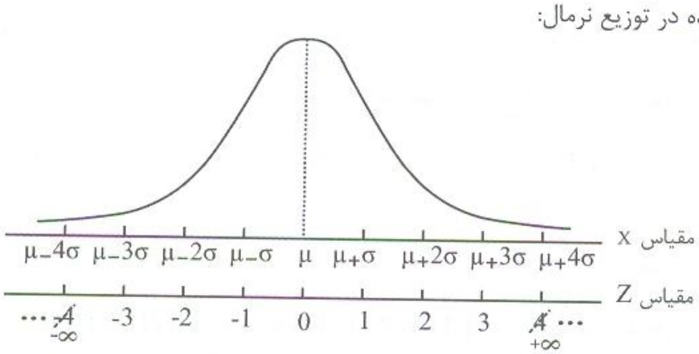
نرمال		نرمال استاندارد
$X \sim N\left(\mu , \sigma^2\right)$		$Z \sim N(0,1)$
$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\xrightarrow{Z = \frac{X-\mu}{\sigma}}$	$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
		

$$X \sim N\left(\mu , \sigma^2\right) \xrightarrow{Z = \frac{X-\mu}{\sigma}} Z \sim N(0,1)$$

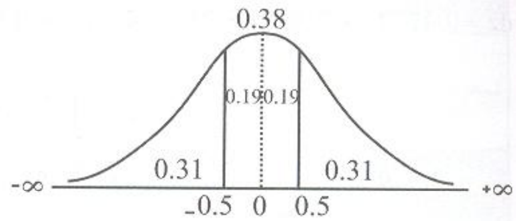
تغییر مقیاس از $X \sim N\left(\mu , \sigma^2\right)$ به $Z \sim N(0,1)$

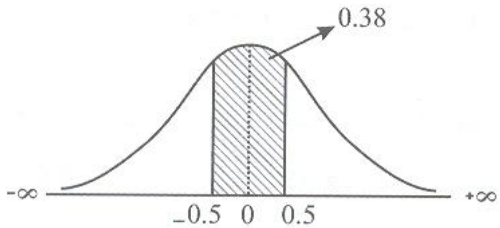
با تبدیل متغیر $Z = \frac{X-\mu}{\sigma}$، مقیاس از X به Z به صورت زیر تغییر می‌کند:

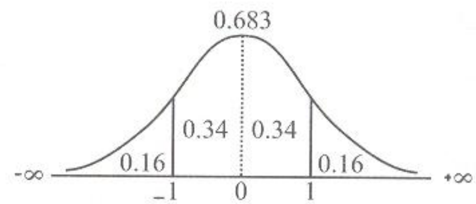
با توجه به سطوح مطرح‌شده در توزیع نرمال:

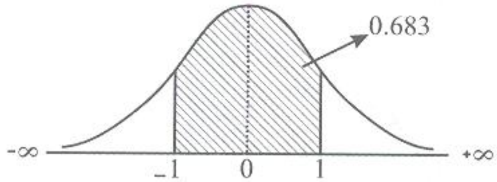


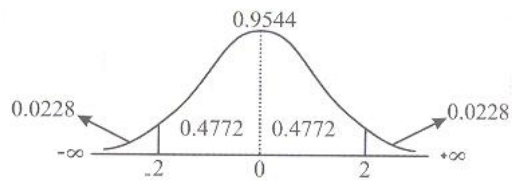
و با در نظر گرفتن محور تقارن $Z = 0$ در نرمال استاندارد خواهیم داشت:

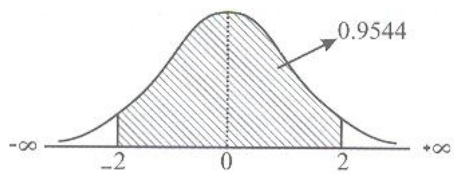


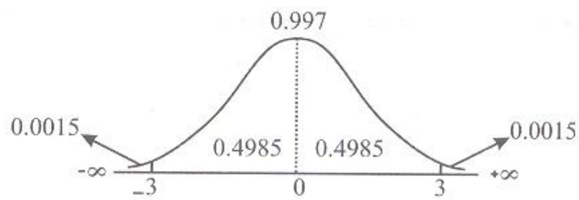


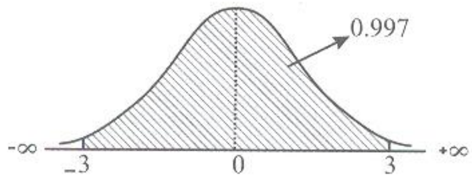




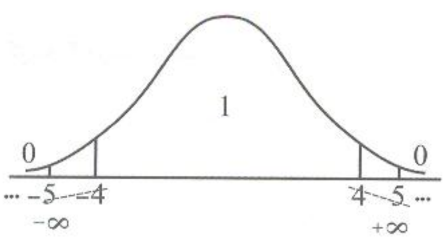








نکته: با توجه به احتمال مربوط به $P(-3 < Z < 3) = 0.997$ و با در نظر گرفتن این موضوع که سطح کل زیرمنحنی نرمال استاندارد 1 است، می‌توان نتیجه گرفت که:



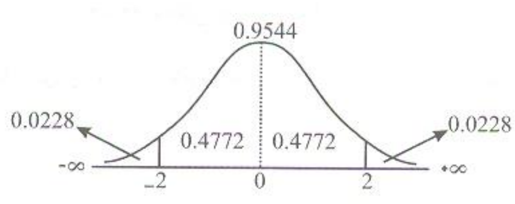
اولاً: احتمال یا سطح زیرمنحنی خارج از فاصله $(-3 < Z < 3)$ تقریباً برابر صفر است.

ثانیاً: احتمال یا سطح زیرمنحنی، درون فواصل بیش از $Z = \pm 3$ یعنی $(-4 < Z < 4)$ یا $(-5 < Z < 5)$ و ... تقریباً برابر 1 است.

مثال ۱ فرض کنید Z متغیر استاندارد و $\int_0^2 f(z)\,dz = 0.4772$ است. $P(Z > 2)$ کدام است؟

۱) 0.0228 ۲) 0.9772 ۳) 0.4772 ۴) 0.5228

حل: گزینه ۱ درست است.



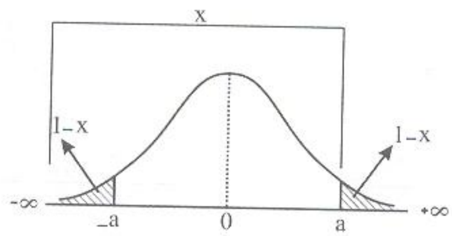
با توجه به شکل:

$$\int_0^2 f(z)\,dz = 0.4772 \rightarrow P(0<Z<2)=0.4772 \rightarrow P(Z>2)=0.5-0.4772=0.0228$$

مثال ۲ اگر در یک توزیع نرمال استاندارد $x=\int_{-\infty}^{a}$ باشد، آن‌گاه $\int_{-a}^{+a}$ کدام است؟

۱) $x-0.5$ ۲) $2-2x$ ۳) $2x-1$ ۴) $2x-0.5$

حل: گزینه ۳ درست است.



با توجه به محور تقارن $z=0$، داریم $P(Z>a)=P(Z<-a)$، بنابراین:

$$\int_{-\infty}^{a} = P(-\infty<Z<a)=P(Z<a)=x \quad \rightarrow \quad \begin{cases} P(Z>a)=1-x \\ P(Z<-a)=1-x \end{cases}$$

حال با توجه به شکل ازآنجاکه سطح کل زیرمنحنی برابر 1 است، داریم:

$$P(-a<Z<a)=1-\left[\underbrace{P(Z>a)}_{1-x}+\underbrace{P(Z<-a)}_{1-x}\right]=1-2(1-x)=2x-1$$

یا

$$P(-a<Z<a)=\underbrace{P(Z<a)}_{x}-\underbrace{P(Z<-a)}_{1-x}=2x-1$$

مثال ۳ X و Y دو متغیر تصادفی و هر یک دارای توزیع نرمال استاندارد هستند. ضریب همبستگی $X+Y$ و $X-Y$ کدام است؟

۱) -0.5 ۲) 0 ۳) 0.5 ۴) 1

حل: گزینه ۲ درست است.

$$X,Y \sim N(\mu=0,\sigma^2=1)$$

$$\rho_{X-Y,X+Y}=\frac{Cov(X-Y,X+Y)}{\sigma_{x-y}\,\sigma_{x+y}}=\frac{0}{\sigma_{x-y}\,\sigma_{x+y}}=0$$

$$Cov(X-Y,X+Y)=Cov(X,X)-Cov(X,Y)+Cov(X,Y)-Cov(Y,Y)=\sigma_X^2-\sigma_Y^2=1-1=0$$

تأثیر تغییرات در جامعه نرمال بر نرمال استاندارد

هرگاه Z، اندازه استاندارد شده متغیر X از جامعه نرمالی به صورت $X\sim N(\mu,\sigma^2)$ با استفاده از رابطه $Z_X=\frac{X-\mu}{\sigma}$ باشد، آن‌گاه هرگونه **تغییرات در** X **هیچ تأثیری بر** Z **(استاندارد) ندارد؛** تنها در بعضی شرایط علامت آن را عوض می‌کند.

$1)\ Z_{x\pm a}=Z_x$	به تمام داده‌ها ثابت a اضافه یا از آن کم شود.
$2)\ Z_{\frac{x}{a}}=Z_{ax}=\begin{cases} Z_x & (\text{مثبت } a) \\ -Z_x & (\text{منفی } a) \end{cases}$	تمام داده‌ها در ثابت a ضرب یا بر آن تقسیم شود
$3)\ Z_{x\pm\frac{a}{100}x}=Z_x$	a درصد هر داده به آن اضافه یا از آن کم شود.

مثال اگر اندازه X از جامعه نرمالی برحسب متغیر استاندارد، $Z_x=10$ باشد، در صورتی‌که هر یک از تغییرات زیر در جامعه نرمال اعمال شود، مقدار استاندارد شده Z چه تغییری می‌کند؟

الف) به تمام داده‌ها 2 واحد اضافه یا کم کنیم.

ب) به تمام داده‌ها 0.20 هر داده را اضافه یا کم کنیم.

ج) تمام داده‌ها را در 2 ضرب یا برآن تقسیم کنیم.

د) تمام داده‌ها را در 2− ضرب یا بر آن تقسیم می‌کنیم.

حل:

الف) $Z_{x\pm2}=Z_x=10$ (هیچ تغییری نمی‌کند)

دقت کنید که:

$$\begin{cases} Z_{x\pm2}=\dfrac{(X\pm2)-\mu(X\pm2)}{\sigma(X\pm2)}=\dfrac{(X\pm2)-(\mu\pm2)}{\sigma}=\dfrac{X-\mu}{\sigma}=Z_x=10 \\ \mu(X\pm2)=\mu\pm2 \\ \sigma(X\cancel{\pm2}^{\,0})=\sigma \end{cases}$$

ب) $Z_{x\pm0.20x}=Z_x=10$ (هیچ تغییری نمی‌کند)

برای مثال:

$$Z_{x+0.2x}=Z_{1.2x}=\frac{1.2X-\mu(1.2X)}{\sigma(1.2X)}=\frac{1.2X-1.2\mu}{1.2\sigma}=\frac{X-\mu}{\sigma}=z_x=10$$

$$Z_{x-0.2x}=Z_{0.8x}=\frac{0.8X-\mu(0.8X)}{\sigma(0.8X)}=\frac{0.8X-0.8\mu}{0.8\sigma}=\frac{X-\mu}{\sigma}=z_x=10$$

ج) $Z_{2x}=Z_{\frac{x}{2}}=Z_x=10$ (تغییری نمی‌کند.)

د) $Z_{-2x}=Z_{\frac{x}{-2}}=-Z_x=-10$ (بدون تغییر منفی می‌شود.)

برای مثال:

$$Z_{2x}=\frac{2X-\mu(2X)}{\sigma(2X)}=\frac{2X-2\mu}{2\sigma}=\frac{X-\mu}{\sigma}=z_x=10$$

$$Z_{-2x}=\frac{-2X-\mu(-2X)}{\sigma(-2X)}=\frac{-2X-(-2\mu)}{|-2|\sigma}=-\frac{X-\mu}{\sigma}=-z_x=-10$$

توزیع $\sum_{i=1}^{n} z_i$

درصورتی‌که $Z_1, Z_2, \ldots, Z_n$ مستقل از هم و دارای توزیع نرمال استاندارد $Z_i \sim N(0,1)$ باشند، آن‌گاه $\sum_{i=1}^{n} z_i$ دارای توزیع نرمال $N(0,n)$ است و دیگر نرمال استاندارد نیست:

$$\boxed{\sum_{i=1}^{n} z_i \sim N(0,n)}$$

نتیجه: مجموع متغیرهای مستقل نرمال استاندارد، دارای توزیع نرمال است ولی استاندارد نیست.

مثال ۱ درصورتی‌که Z_1 و Z_2 مستقل و دارای توزیع نرمال استاندارد باشند، آن‌گاه $Z_1 + Z_2$ دارای توزیع ....

۱) نرمال استاندارد است. ۲) $N(0,2)$ است. ۳) نرمال نیست. ۴) $N(0,3)$ است.

حل: گزینه ۲ درست است.

$$Z_1 + Z_2 \sim N(0,2)$$

نکته: هرگاه X دارای توزیع نرمال استاندارد $Z \sim N(0,1)$ باشد، دو قاعده زیر برقرار است:

1) $E\left(Z^{2m+1}\right) = 0 \;\; ; \;\; m = 0,1,2,\ldots \rightarrow$ امید توان فرد Z برابر صفر است.

2) $E\left(Z^{2m}\right) = \dfrac{(2m)!}{2^m m!} \;\; ; \;\; m = 0,1,2,\ldots$

مثال ۲ متغیر تصادفی X دارای توزیع نرمال با میانگین صفر و واریانس 2 است. $E\left(X^4\right)$ برابر است با:

۱) 18 ۲) 16 ۳) 12 ۴) 6

حل: گزینه ۳ درست است.

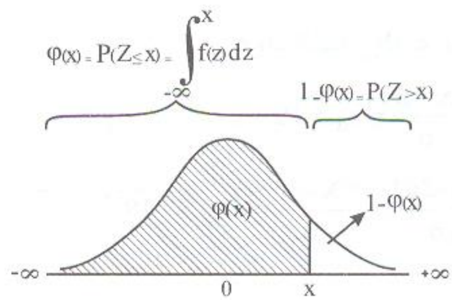
ابتدا X را به Z تبدیل می‌کنیم، سپس با توجه به قاعده دوم داریم:

$$Z = \frac{X-\mu}{\sigma} = \frac{X-0}{\sqrt{2}} \rightarrow Z^4 = \frac{X^4}{4} \rightarrow X^4 = 4Z^4 \rightarrow E\left(X^4\right) = 4E\left(Z^4\right) \xlongequal{m=2} 4\frac{(2\times2)!}{2!2^2} = 3\times4 = 12$$

تابع توزیع تجمعی (Cumulative Distribution Function)

اگر $Z \sim N(0,1)$ متغیر نرمال استاندارد باشد، تابع توزیع تجمعی آن $F_Z(x) = P(Z \le x)$ به صورت زیر خواهد بود:

$$P(Z \le x) = \int_{-\infty}^{x} f(z)\,dz = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz = \varphi(x)$$



درنتیجه:

$$\boxed{\begin{array}{l} P(Z \le x) = \varphi(x) \\ P(Z > x) = 1 - \varphi(x) \end{array}}$$

نکته: با توجه به تقارن توزیع نرمال استاندارد داریم:

$$P(Z \le -x) = P(Z \ge +x) \quad \rightarrow \quad \boxed{\varphi(-x) = 1 - \varphi(x)}$$

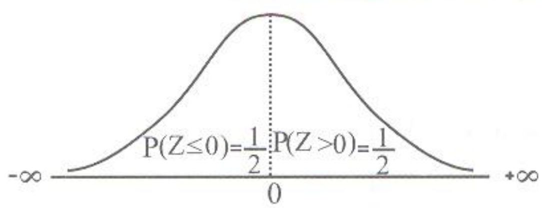
یادآوری:

۱) با استفاده از جدول نرمال، مقدار $\varphi(x)$ قابل محاسبه است.

۲) با توجه به پیوسته بودن توزیع نرمال، $P(Z = x) = 0$ است.

تابع توزیع تجمعی $\varphi(0)$

اگر $Z \sim N(0,1)$ استاندارد باشد، آن‌گاه:



$$\begin{cases} P(Z \le 0) = \varphi(0) = \dfrac{1}{2} \\ P(Z > 0) = 1 - \varphi(0) = \dfrac{1}{2} \end{cases}$$

نکته: اگر k مقدار ثابتی باشد (مثبت یا منفی)، آن‌گاه همواره:

$$P(kZ \le 0) = P(kZ \ge 0) = \frac{1}{2}$$

مثال اگر $Z \sim N(0,1)$ استاندارد باشد، مقدار $P(-6Z \le 0)$ کدام است؟

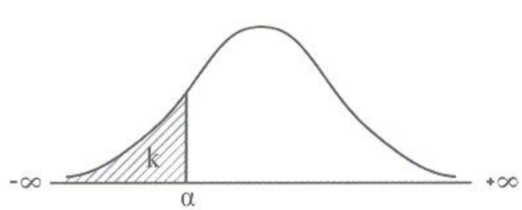
۱) 0.5 ۲) 0.25 ۳) $\frac{1}{6}$ ۴) 0

حل: گزینه ۱ درست است.

$$P(-6Z \le 0) = P\left(\frac{-6Z}{-6} \ge \frac{0}{-6}\right) = P(Z \ge 0) = \frac{1}{2}$$

نمایش سطح در نرمال استاندارد

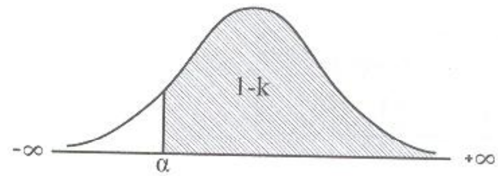
اگر $Z \sim N(0,1)$ باشد، سطح زیر منحنی به صورت‌های زیر نمایش داده می‌شود:



نمایش اول: $\int_{-\infty}^{\alpha} = k$

نمایش دوم: $F(\alpha) = \varphi(\alpha) = P(Z \le \alpha) = k$

نمایش سوم: $Z_k = \alpha$

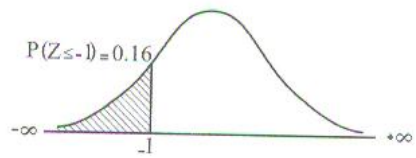


$\int_{\alpha}^{+\infty} \quad = 1 - k$: نمایش اول

$P(Z > \alpha) = 1 - k$: نمایش دوم

$Z_{1-k} = \alpha$: نمایش سوم

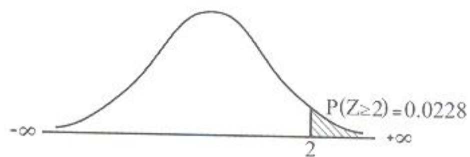
مثال



$\int_{-\infty}^{-1} = 0.16$: نمایش اول

$F(-1) = \varphi(-1) = P(Z \leq -1) = 0.16$: نمایش دوم

$Z_{0.16} = -1$: نمایش سوم



$\int_{2}^{+\infty} = 0.0228$: نمایش اول

$P(Z > 2) = 0.0228$: نمایش دوم

$Z_{0.0228} = 2$: نمایش سوم

انواع مسایل نرمال

۱- تبدیل نرمال به نرمال استاندارد

بین هر متغیر X با توزیع نرمال $X \sim N(\mu, \sigma^2)$ و متغیر استاندارد Z با توزیع $Z \sim N(0,1)$ همواره تبدیل متغیر زیر وجود دارد:

$$Z = \frac{X-\mu}{\sigma}$$

مثال ۱ اگر کمیت $X \sim N(9,16)$ توزیع شده باشد، مقدار استانداردشده $z = 1.5$ با کدام مقدار X برابر است؟

۱) 20 ۲) 15 ۳) 30 ۴) 7

حل: گزینه ۲ درست است.

$$Z = \frac{X-\mu}{\sigma} \xrightarrow[\mu=9,\ \sigma=4]{z=1.5} 1.5 = \frac{X-9}{4} \rightarrow x = 15$$

مثال ۲ نمرات دو داوطلب در آزمون درس آمار، نرمال و برابر با 14 و 12 است. این نمرات برحسب واحدهای استاندارد به ترتیب 0.25 و 0.25− شده، میانگین و واریانس نمرات آزمون چقدر است؟

۱) 16 , 13 ۲) 1 , 0 ۳) 4 , 12 ۴) 4 , 13

حل: گزینه ۱ درست است.

با توجه به رابطه $Z_i = \frac{X_i-\mu}{\sigma}$ که بین یک متغیر (X) نرمال و یک متغیر (Z) نرمال استاندارد وجود دارد و با توجه به مقادیر $x_1 = 14$, $x_2 = 12$ و $z_1 = 0.25, z_2 = -0.25$ داریم:

$$\begin{cases} Z_1 = \dfrac{X_1-\mu}{\sigma} \\ Z_2 = \dfrac{X_2-\mu}{\sigma} \end{cases} \rightarrow \begin{cases} 0.25 = \dfrac{14-\mu}{\sigma} \\ -0.25 = \dfrac{12-\mu}{\sigma} \end{cases} \rightarrow \begin{cases} 0.25\sigma = 14-\mu \\ -0.25\sigma = 12-\mu \end{cases} \rightarrow 0 = 26-2\mu \rightarrow \mu = 13$$

حال با توجه به $\mu = 13$، مقدار σ را می‌توانیم محاسبه کنیم:

$$0.25 = \frac{14-\mu}{\sigma} \xrightarrow{\mu=13} \frac{1}{4} = \frac{14-13}{\sigma} \rightarrow \sigma = 4 \rightarrow \sigma^2 = 16$$

مثال ۳ اگراندازه دو نفر از جامعه نرمالی 13 و 19 و اندازه این دو برحسب متغیر استاندارد، صفر و 3 باشد، میانگین و انحراف معیار به ترتیب (از چپ به راست) کدام‌اند؟

۱) 19 و 2 ۲) 3 و 6 ۳) 2 و 13 ۴) 3 و 19

حل: گزینه ۳ درست است.

$$\begin{cases} Z_1 = \dfrac{X_1-\mu}{\sigma} \rightarrow 0 = \dfrac{13-\mu}{\sigma} \\ Z_2 = \dfrac{X_2-\mu}{\sigma} \rightarrow 3 = \dfrac{19-\mu}{\sigma} \end{cases} \rightarrow \mu = 13 , \sigma = 2$$

مثال ۴ اندازه قد دانش‌آموزان کلاس اول دارای توزیع نرمال با میانگین 100 و انحراف معیار 5 سانتی‌متر است. متغیر نرمال استاندارد برای قد دانش‌آموزان بین 110 تا 115 سانتی‌متر در کدام فاصله است؟

۱) $1 < z < 2$ ۲) $2 < z < 3$ ۳) $1.5 < z < 2$ ۴) $2 < z < 2.5$

حل: گزینه ۲ درست است.

$$\begin{cases} 110 < x < 115 \\ \mu = 100 \ , \ \sigma = 5 \end{cases} \xrightarrow{Z=\frac{X-\mu}{\sigma}} \frac{110-100}{5} < \frac{x-\mu}{\sigma} < \frac{115-100}{5} \rightarrow 2 < z < 3$$

۲- محاسبه احتمال در توزیع نرمال

اگر متغیر تصادفی X دارای توزیع نرمال به صورت $X \sim N\left(\mu, \sigma^2\right)$ باشد، برای محاسبه احتمال در هر بازه دلخواه به صورت زیر عمل می‌کنیم:

الف) با استفاده از رابطه $Z = \frac{X-\mu}{\sigma}$، احتمال مورد نظر را به فرم استاندارد تبدیل می‌کنیم.

ب) به کمک جدول نرمال استاندارد، مقدار احتمال مورد نظر را به دست می‌آوریم:

$$P(X < a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(Z < \frac{a-\mu}{\sigma}\right) = \varphi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X > a) = P\left(\frac{X-\mu}{\sigma} > \frac{a-\mu}{\sigma}\right) = P\left(Z > \frac{a-\mu}{\sigma}\right) = 1 - \varphi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \varphi\left(\frac{b-\mu}{\sigma}\right) - \varphi\left(\frac{a-\mu}{\sigma}\right)$$

مثال ۱ اگر X دارای توزیع نرمال با $\mu = 20$ و $\sigma^2 = 9$ باشد، احتمال آنکه کمیت تصادفی 2X مقادیری کمتر از 42 اختیار کند ($P(2X < 42)$)، برابر است با:

۱) $\varphi(0.33)$ ۲) $\varphi(0.2)$ ۳) $\varphi(2)$ ۴) $\varphi(0.5)$

حل: گزینه ۱ درست است.

$$\begin{cases} P(2X < 42) = P(X < 21) = P\left(\frac{X-\mu}{\sigma} < \frac{21-20}{3}\right) = P\left(Z < \frac{1}{3}\right) = \varphi\left(\frac{1}{3}\right) = \varphi(0.33) \\ \mu = 20 \\ \sigma^2 = 9 \ \rightarrow \ \sigma = 3 \end{cases}$$

مثال ۲ اگر $X \sim N(1,1)$ باشد، $P(X \geq 2 \mid X \geq 1)$ برابر کدام گزینه است؟ ($\phi(x)$ معرف تابع توزیع تجمعی متغیرتصادفی نرمال استانداردااست.)

۱) $2\phi(1)$ ۲) $2\phi(1)-1$ ۳) $2-2\phi(1)$ ۴) $2\phi(1)-0.5$

حل: گزینه ۳ درست است.

$$X \sim N(1,1) \rightarrow \mu = 1 \ , \ \sigma^2 = 1 \rightarrow \sigma = 1$$

$$P(X \geq 2 \mid X \geq 1) = \frac{P(X \geq 2 \ , X \geq 1)}{P(X \geq 1)} = \frac{P(X \geq 2)}{P(X \geq 1)} = \frac{P\left(\frac{X-\mu}{\sigma} \geq \frac{2-1}{1}\right)}{P\left(\frac{X-\mu}{\sigma} \geq \frac{1-1}{1}\right)} = \frac{P(Z \geq 1)}{P(Z \geq 0)} = \frac{1-\phi(1)}{\frac{1}{2}} = 2 - 2\phi(1)$$

مثال ۳ اگر X و Y دارای توزیع نرمال استاندارد باشند و ضریب همبستگی بین آن‌ها برابر $\frac{1}{2}$ باشد، $P(X - Y < 1)$ برابر است با: ($\varphi(0)$ تابع توزیع نرمال استاندارد است)

۱) $\varphi(0)$ ۲) $\varphi\left(\frac{1}{2}\right)$ ۳) $\varphi(1)$ ۴) $\varphi(2)$

حل: گزینه ۳ درست است.

ابتدا با توجه به خاصیت خطی بودن توزیع نرمال، توزیع $X-Y$ را به دست می‌آوریم:

$$X \sim N\left(\mu_x=0,\sigma_x^2=1\right)\ ,\ \ Y \sim N\left(\mu_y=0,\sigma_y^2=1\right)$$

$$\begin{cases} E(X-Y)=\mu_x-\mu_y=0-0=0 \\ Var(X-Y)=\sigma_x^2+\sigma_y^2+2(1)(-1)\text{cov}(x,y)=1+1-2\times\frac{1}{2}=1 \end{cases} \to\ X-Y \sim N(0,1)$$

$$\rho=\frac{Cov(x,y)}{\sigma_x\,\sigma_y}\ \to \frac{1}{2}=\frac{Cov(x,y)}{1\times1}\ \to Cov(x,y)=\frac{1}{2}$$

حال با داشتن امید و واریانس $X-Y$، احتمال خواسته‌شده را محاسبه می‌کنیم:

$$P(X-Y<1)=P\left(\frac{(X-Y)-\mu_{x-y}}{\sigma_{x-y}}<\frac{1-0}{1}\right)=P(Z<1)=\varphi(1)$$

مثال ۴ در یک توزیع نرمال با میانگین 32 و واریانس 4، تقریباً چند درصد داده‌ها بین دو عدد 38 و 26 قرار می‌گیرند؟

$$\left(S_{-\infty}^{-3}=\int_{-\infty}^{-3}f(z)\,dz=0.0013\right)$$

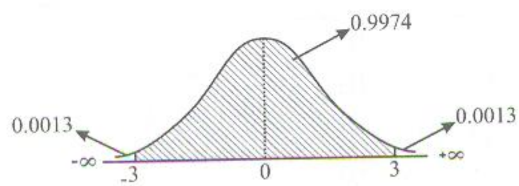
۱) 89.6 ۲) 92.3 ۳) 95.4 ۴) 99.7

حل: گزینه ۴ درست است.

$$\begin{cases} P(26<X<38)=P\left(\frac{26-32}{\sqrt{4}}<\frac{X-\mu}{\sigma}<\frac{38-32}{\sqrt{4}}\right)=P(-3<Z<3)=0.9974 \to 99.7 \text{ درصد} \\ \mu=32\ ,\ \sigma^2=4 \end{cases}$$

$$S_{-\infty}^{-3}=P(Z<-3)=P(Z>3)=0.0013$$

$$P(-3<Z<3)=1-2\times0.0013=0.9974$$



مثال ۵ متغیر تصادفی X طبق قانون نرمال با $\mu=20$ و $\sigma^2=25$ توزیع شده است. احتمال اینکه کمیت تصادفی X مقادیری بیش از 10 اختیار کند، چقدر است؟ (با فرض اینکه $F_Z(-2)=0.0228$ باشد.)

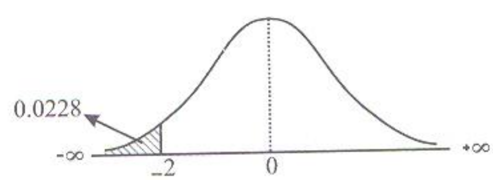
۱) 0.0228 ۲) 0.23413 ۳) 0.6587 ۴) 0.9772

حل: گزینه ۴ درست است.

$$\begin{cases} P(X>10)=P\left(\frac{X-\mu}{\sigma}>\frac{10-20}{\sqrt{25}}\right)=P(Z>-2)=0.9772 \\ \mu=20\ ,\ \sigma^2=25\ \ \to\ \ \sigma=5 \end{cases}$$

$$F_Z(-2)=P(Z\le-2)=0.0228$$

$$\to\ \ P(Z>-2)=1-0.0228=0.9772$$



مثال ۶ در یک بررسی از یک توزیع نرمال، $\mu = 100$ و $\sigma = 15$ مشاهده شده است. چه نسبتی از افراد 130 یا بالاتر از 130 هستند؟ $\left(\int_{-2}^{2} f(z)dz = 0.9544\right)$

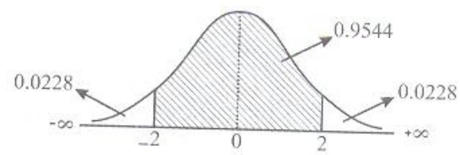
۱) 0.0228 ۲) 0.0456 ۳) 0.0912 ۴) 0.4772

حل: گزینه ۱ درست است.

$$\begin{cases} P(X \geq 130) = P\left(\dfrac{X-\mu}{\sigma} \geq \dfrac{130-100}{15}\right) = P(Z > 2) = 0.0228 \\ \mu = 100,\ \sigma = 15 \end{cases}$$

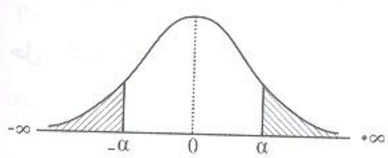
$$\int_{-2}^{2} f(z)dz = P(-2 < Z < 2) = 0.9544 \rightarrow$$

$$P(Z > 2) = P(Z < -2) = \frac{1-0.9544}{2} = 0.0228$$



۳- استفاده معکوس از جدول نرمال استاندارد

درصورتی‌که α مقدار ثابتی در بازه $-\infty < z < +\infty$ باشد، با توجه به شکل زیر داریم:



$$\rightarrow \boxed{\begin{array}{l} P(Z > \alpha) = P(Z < -\alpha) \\ P(Z < \alpha) = P(Z > -\alpha) \end{array}}$$

به عبارت دیگر اگر A و B مقادیر ثابتی در بازه $-\infty < z < +\infty$ باشند، آن‌گاه:

$$\boxed{\begin{array}{l} \left.\begin{array}{l} P(Z > A) = P(Z < B) \\ P(Z < A) = P(Z > B) \end{array}\right\} \rightarrow A = -B \\ \left.\begin{array}{l} P(Z > A) = P(Z > B) \\ P(Z < A) = P(Z < B) \end{array}\right\} \rightarrow A = B \end{array}}$$

فرض کنید:

0) X دارای توزیع نرمال به صورت $N(\mu, \sigma^2)$ است (σ^2 و μ معلوم یا مجهول).

1) احتمال $P(X < a)$ یا $P(X > a)$ (a معلوم یا مجهول) مشخص و برابر با مقدار معلوم t باشد.

2) با توجه به جدول نرمال استاندارد $P(Z < b) = t$ یا $P(Z > b) = t$ نیز مشخص است (b معلوم).

با توجه به این مفروضات، اگر یکی از مقادیر a، μ و σ مجهول باشند، برای محاسبه مقدار آن کافی است با استفاده از رابطه $Z = \dfrac{X-\mu}{\sigma}$ ابتدا مقدار استانداردشده قسمت (1) را به دست آوریم، سپس به کمک قسمت (2) مقدار مجهول را محاسبه کنیم؛ به عبارت دیگر:

$$(1)\ P(X < a) = P\left(Z < \frac{a-\mu}{\sigma}\right) = t \rightarrow \begin{cases} \xrightarrow{(2)\ P(Z<b)=t} \boxed{\dfrac{a-\mu}{\sigma} = b} \\ \xrightarrow[(2)\ P(Z>b)=t]{} \boxed{\dfrac{a-\mu}{\sigma} = -b} \end{cases}$$

یا

$$(1)\ P(X>a)=P\left(Z>\frac{a-\mu}{\sigma}\right)=t \rightarrow \begin{cases} \xrightarrow{(2)\ P(Z<b)=t} \boxed{\frac{a-\mu}{\sigma}=-b} \\ \xrightarrow{(2)\ P(Z>b)=t} \boxed{\frac{a-\mu}{\sigma}=b} \end{cases}$$

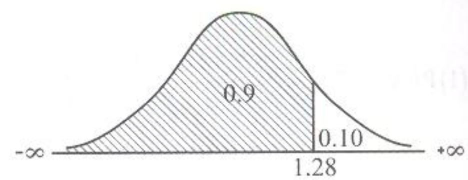
مثال ۱ در یک آزمون بزرگ، میانگین نمرات 60 و انحراف معیار نمرات 20 است. اگر 10% از شرکت‌کنندگان بتوانند نمره قبولی بگیرند، حداقل نمره قبولی چقدر است؟ $\left(S_{-\infty}^{1.28}=0.9\right)$

۱) 75.6 ۲) 75 ۳) 85 ۴) 85.6

حل: گزینه ۴ درست است.

$$(1)\begin{cases} P(X>\alpha)=0.10 \rightarrow P\left(\frac{X-\mu}{\sigma}>\frac{\alpha-60}{20}\right)=0.10 \rightarrow P\left(Z>\frac{\alpha-60}{20}\right)=0.10 \\ \mu=60\ ,\ \sigma=20 \end{cases}$$

$$(2)\ S_{-\infty}^{1.28}=P(Z<1.28)=0.90 \rightarrow P(Z>1.28)=0.1$$



حال با در نظر گرفتن روابط (۱) و (۲) داریم:

$$\begin{matrix}(1)\ P\left(Z>\frac{\alpha-60}{20}\right)=0.10 \\ (2)\ P(Z>1.28)=0.1\end{matrix} \rightarrow 1.28=\frac{\alpha-60}{20} \rightarrow \alpha=85.6$$

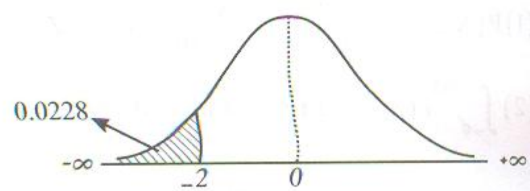
مثال ۲ اگر $P(Z\le -2)=0.0228$ و X نیز دارای توزیع نرمال با میانگین 15 و $P(X\le 35)=0.9772$ باشد، انحراف معیار X کدام است؟

۱) صفر ۲) 5 ۳) 10 ۴) 15

حل: گزینه ۳ درست است.

$$(1)\ P(X\le 35)=P\left(\frac{X-\mu}{\sigma}\le\frac{35-\mu}{\sigma}\right)=P\left(Z\le\frac{35-15}{\sigma}\right)=0.9772$$

$$(2)\ P(Z\le -2)=0.0228$$
$$\rightarrow P(Z\ge -2)=1-0.0228=0.9772$$



$$\begin{matrix}(1):P\left(Z\le\frac{20}{\sigma}\right)=0.9772 \\ (2):P(Z\ge -2)=0.9772\end{matrix} \rightarrow -2=-\left(\frac{20}{\sigma}\right) \rightarrow \sigma=10$$

مثال ۳ فرض کنید $X\sim N(20,\sigma^2)$ است. اگر $P(X\le 22+\sigma)=0.9772$ باشد، مقدار انحراف معیار (σ) کدام است؟ $\left(S_2^5=0.0228\right)$

۱) 1 ۲) 2 ۳) 3 ۴) 4

حل: گزینه ۲ درست است.

$X \sim N\left(20,\sigma^2\right)$ و $\mu = 20$ است، بنابراین:

$$(1): P(X \le 22+\sigma) = P\left(\frac{X-\mu}{\sigma} \le \frac{22+\sigma-20}{\sigma}\right) = P\left(Z < \frac{2+\sigma}{\sigma}\right) = 0.9772$$

$$(2): \int_{2}^{+\infty} = 0.0228 \quad \rightarrow \quad P(Z>2) = 0.0228 \quad \rightarrow \quad P(Z<2) = 0.9772$$

$$(1),(2): \frac{2+\sigma}{\sigma} = 2 \quad \rightarrow \quad \sigma = 2$$

مثال ۴ اگر X دارای توزیع $N(\mu\ ,\ 100)$ باشد و داشته باشیم $P(X>124) = 0.05$ و $Z_{0.05} = 1.65$، آن‌گاه مقدار μ برابر است با:

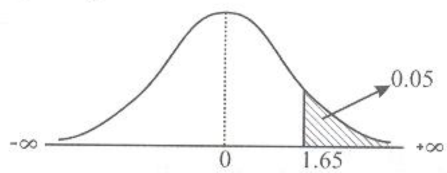
۱) 140.5 ۲) 121.9 ۳) 107.5 ۴) 104.5

حل: گزینه ۳ درست است.

با توجه به $X \sim N(\mu,100)$، مقدار $\sigma^2 = 100$ و درنتیجه $\sigma = 10$ است، بنابراین:

$$(1) P(X>124) = P\left(\frac{X-\mu}{\sigma} > \frac{124-\mu}{\sqrt{100}}\right) = P\left(Z > \frac{124-\mu}{10}\right) = 0.05$$

$$(2) Z_{0.05} = 1.65 \quad \rightarrow$$



$$\rightarrow \quad P(Z > 1.65) = 0.05$$

$$(1),(2)\ \frac{124-\mu}{10} = 1.65 \quad \rightarrow \quad \mu = 107.5$$

مثال ۵ توزیع متغیر تصادفی X نرمال با میانگین 100 و انحراف معیار 10 است. اگر $P(X \ge x) = 0.0495$ باشد، مقدار x چقدر است؟ (راهنمایی: $\int_{-4}^{-1.65} f(z)\,dz = 0.0495$)

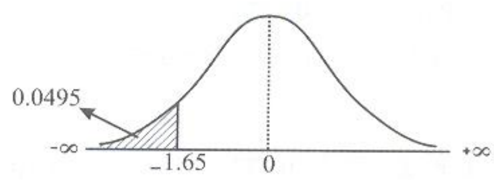
۱) 60 ۲) 83.5 ۳) 116.5 ۴) 140

حل: گزینه ۳ درست است.

$$(1) P(X \ge x) = P\left(\frac{X-\mu}{\sigma} \ge \frac{x-100}{10}\right) = P\left(Z \ge \frac{x-100}{10}\right) = 0.0495$$

$$(2) \int_{\cancel{-4}\ -\infty}^{-1.65} f(z)\,dz = P(Z < -1.65) = 0.0495$$

$$(1),(2) \rightarrow -1.65 = -\left(\frac{x-100}{10}\right) \rightarrow x = 116.5$$



مثال ۶ اگر $X \sim N(1,4)$ و $Y \sim N(2,9)$ باشد و Y, X مستقل باشند، به ازاء کدام مقدار a رابطه زیر برقرار است؟

$$P(Y-2X \le a) = P(2X+Y > a)$$

۱) $\frac{1}{2}$ ۲) 0 ۳) 2 ۴) 3

حل: گزینه ۳ درست است.

در این‌گونه مسایل به ترتیب زیر عمل می‌کنیم:

۱) با توجه به خاصیت خطی بودن توزیع نرمال، ابتدا توزیع $2X+Y$ و $Y-2X$ را به دست می آوریم:

$$\begin{cases} E(2X+Y) = 2\mu_x + \mu_y = 2\times1+2 = 4 \\ Var(2X+Y) = 2^2\sigma_x^2 + \sigma_y^2 = 2^2\times4+9 = 25 \end{cases} \rightarrow (2X+Y) \sim N(4,25)$$

$$\begin{cases} E(Y-2X) = \mu_y - 2\mu_x = 2-2\times1 = 0 \\ Var(Y-2X) = \sigma_y^2 + 2^2\sigma_x^2 = 9+2^2\times4 = 25 \end{cases} \rightarrow (Y-2X) \sim N(0,25)$$

۲) دو طرف تساوی را به نرمال استاندارد تبدیل می‌کنیم:

$$P\left(\frac{(Y-2X)-\mu_{y-2x}}{\sigma_{y-2x}} \le \frac{a-0}{5}\right) = P\left(\frac{(2X+Y)-\mu_{2x+y}}{\sigma_{2x+y}} > \frac{a-4}{5}\right) \rightarrow P\left(Z \le \frac{a}{5}\right) = P\left(Z > \frac{a-4}{5}\right)$$

۳) با استفاده معکوس از جدول نرمال استاندارد داریم:

$$\frac{a}{5} = -\frac{a-4}{5} \rightarrow 5a = -5a+20 \rightarrow 10a = 20 \rightarrow a = 2$$

تقریب توزیع‌ها به وسیله توزیع نرمال

در شرایط مشخصی، توزیع‌های گسسته پواسون و دوجمله‌ای قابل تقریب به توزیع پیوسته نرمال هستند؛ یعنی در مواردی می‌توان از توزیع نرمال به‌جای این توزیع‌ها استفاده کرد.

۱- تقریب توزیع پواسون به نرمال

اگر میانگین توزیع پواسون (λ) به حدی بزرگ شود که $\lambda > 10$ باشد، آن‌گاه توزیع نرمال تقریب مناسبی برای توزیع پواسون خواهد بود (هرچه بزرگ‌تر شود تقریب بهتر است)، به طوری که:

$$\boxed{\begin{array}{l} X \sim \text{توزیع پواسون} \\ \mu_X = \lambda \\ \sigma_X^2 = \lambda \end{array}} \xrightarrow{\lambda > 10} \boxed{\begin{array}{l} X \sim \text{توزیع نرمال} \\ \mu_X = \lambda \\ \sigma_X^2 = \lambda \to \sigma_X = \sqrt{\lambda} \end{array}}$$

Z (استاندارد) در پواسون

برای تبدیل توزیع نرمال به نرمال استاندارد داریم:

$$\boxed{Z = \frac{X-\mu}{\sigma} \xrightarrow[\sigma=\sqrt{\lambda}]{\mu=\lambda} Z = \frac{X-\lambda}{\sqrt{\lambda}}}$$

مثال ۱ در یک توزیع پواسون با $\lambda = 64$ تقریب توزیع نرمال را در نظر می‌گیریم. عدد متناظر Z برای داده $x = 70$ کدام است؟

۱) 0.5 ۲) 0.25 ۳) 0.75 ۴) 1.25

حل: گزینه ۳ درست است.

$$Z = \frac{X-\lambda}{\sqrt{\lambda}} = \frac{70-64}{\sqrt{64}} = \frac{6}{8} = \frac{3}{4} = 0.75$$

مثال ۲ در یک توزیع پواسون با میانگین 16، مقدار $P(X<12)$ تقریباً چقدر است؟

۱) 0.16 ۲) 0.84 ۳) 0.68 ۴) 0.34

حل: گزینه ۱ درست است.

با توجه به آنکه در توزیع پواسون $\mu = \lambda = 16$ است، شرایط تقریب نرمال $(\lambda > 10)$ برقرار است و برای محاسبه احتمال از تقریب Z استفاده می‌کنیم.

$$P(X<12) = P\left(\frac{X-\mu}{\sigma} < \frac{12-\lambda}{\sqrt{\lambda}}\right) = P\left(Z < \frac{12-16}{4}\right) = P(Z<-1) = 0.16$$

۲- تقریب توزیع دوجمله‌ای به نرمال

اگر در یک توزیع دوجمله‌ای با پارامترهای n و p (n تعداد آزمایشات و p احتمال موفقیت) یکی از شرایط زیر برقرار باشد، توزیع نرمال با $\mu = np$ و $\sigma^2 = npq$ تقریب خوبی برای توزیع دوجمله‌ای است.

(a)

$$\boxed{\begin{array}{l} X \sim Bin(n,p) \\ \mu = np \\ \sigma^2 = npq \end{array}} \xrightarrow{np>5\ ,\ nq>5} \boxed{\begin{array}{l} X \sim N(\mu,\sigma^2) \\ \mu = np \\ \sigma^2 = npq \end{array}}$$

np و nq هرچه بیشتر از 5 باشند، تقریب بهتر است.

(b)

$$\boxed{\begin{array}{l} X \sim Bin(n,p) \\ \mu = np \\ \sigma^2 = npq \end{array}} \xrightarrow{\text{برای } n \text{ های کوچک } p \simeq 0.5} \boxed{\begin{array}{l} X \sim N(\mu,\sigma^2) \\ \mu = np \\ \sigma^2 = npq \end{array}}$$

در شرایط مساوی تقریب (a) از تقریب (b) قوی‌تر و بهتر است.

Z (استاندارد) در دوجمله‌ای

برای تبدیل توزیع نرمال به نرمال استاندارد داریم:

$$\boxed{Z = \frac{X-\mu}{\sigma} \xrightarrow[\sigma=\sqrt{npq}]{\mu=np} Z = \frac{X-np}{\sqrt{npq}}}$$

مثال ۱ اگر n و p دو پارامتر توزیع دوجمله‌ای باشند، کدام‌یک از این موارد را می‌توان با توزیع نرمال تقریب زد؟

۱) $n=5\ ,\ p=0.3$
۲) $n=10\ ,\ p=0.4$
۳) $n=15\ ,\ p=0.45$
۴) $n=1000\ ,\ p=0.5$

حل: گزینه ۴ درست است.

هرگاه در توزیع دوجمله‌ای $np > 5$ و $nq > 5$ باشد، آن‌گاه توزیع نرمال تقریب مناسبی برای بررسی توزیع دوجمله‌ای است:

گزینه۱) $n = 5\ ,\ p = 0.3\ ,q = 0.7 \rightarrow \begin{cases} np = 1.5 < 5 \\ nq = 3.5 < 5 \end{cases} \rightarrow$ تقریب نرمال مناسب نیست.

گزینه۲) $n = 10,\ p = 0.4\ ,\ q = 0.6 \rightarrow \begin{cases} np = 4 < 5 \\ nq = 6 > 5 \end{cases} \rightarrow$ تقریب نرمال مناسب نیست.

گزینه۳) $n = 15\ ,\ p = 0.45\ ,\ q = 0.55 \rightarrow \begin{cases} np = 6.75 > 5 \\ nq = 8.25 > 5 \end{cases} \rightarrow$ تقریب نرمال مناسب است.

گزینه۴) $n = 1000,\ p = 0.5\ ,\ q = 0.5 \rightarrow \begin{cases} np = 500 > 5 \\ nq = 500 > 5 \end{cases} \rightarrow$ تقریب نرمال مناسب است.

گزینه‌های ۳ و ۴ شرایط تقریب را دارند، اما ازآنجاکه در گزینه ۴، مقادیر np و nq به مراتب از عدد 5 بزرگ‌تر هستند، تقریب نرمال برای این گزینه بهتر از گزینه ۳ است.

مثال ۲ در یک توزیع دوجمله‌ای با پارامترهای $n = 100$ و $p = \frac{1}{2}$ تقریب توزیع نرمال را در نظر بگیرید.

الف) پارامترهای تقریب نرمال را به دست آورید.

ب) عدد متناظر Z برای داده $x = 55$ را به دست آورید.

حل:

الف) اگر در توزیع دوجمله‌ای $np > 5$ و $nq > 5$ شود، آن‌گاه از تقریب نرمال با پارامترهای $\mu = np$ و $\sigma^2 = npq$ برای آن استفاده می‌شود:

$$\begin{cases} \mu = np = 100 \times \frac{1}{2} = 50 \\ \sigma^2 = npq = 100 \times \frac{1}{2} \times \frac{1}{2} = 25 \end{cases}$$

ب) رابطه $Z=\frac{X-\mu}{\sigma}$ برای متغیر X از نرمال و Z از نرمال استاندارد همواره برقرار است، در نتیجه بنا بر قسمت (الف) داریم:

$$\begin{cases} Z=\frac{X-\mu}{\sigma}=\frac{X-np}{\sqrt{npq}}=\frac{55-50}{5}=1 \\ x=55\ ,\ \ \mu=np=50\ \ ,\ \ \sigma^2=npq=25\ \ \to\ \ \sigma=5 \end{cases}$$

مثال ۳ احتمال قبولی شرکت‌کنندگان در آزمونی 0.8 است، احتمال آنکه در یک نمونه 100 نفری حداقل 84 نفر قبول شوند، چقدر است؟

۱) 0.16 ۲) 0.84 ۳) 0.68 ۴) 0.34

حل: گزینه ۱ درست است.

با توجه به آنکه آزمون برای $n=100$ نفر تکرار شده و احتمال قبولی برای هر یک $p=0.8$ است، توزیع قبولی شرکت‌کنندگان در آزمون، دوجمله‌ای است با $n=100, p=0.8$.

حال از آنجاکه $nq=20>5, np=80>5$ است، استفاده از تقریب نرمال مناسب است:

$$P(X>84)=P\left(\frac{X-\mu}{\sigma}>\frac{84-np}{\sqrt{npq}}\right)=P\left(Z>\frac{84-80}{\sqrt{16}}\right)=P(Z>1)=0.16$$

تصحیح پیوستگی (دوجمله‌ای و پواسون)

هرگاه بخواهیم توزیع‌های گسسته مانند دوجمله‌ای و پواسون را تحت شرایط مطرح‌شده به نرمال که یک توزیع پیوسته است، تقریب بزنیم، برای محاسبه احتمال‌های مختلف، از تصحیح پیوستگی به شکل زیر استفاده می‌کنیم:

$$P(X=a)=P(a-0.5<X<a+0.5)$$

$$P(X\ge a)=P(X\ge a-0.5) \quad , \quad P(X>a)=P(X\ge a+0.5)$$

$$P(X\le a)=P(X\le a+0.5) \quad , \quad P(X<a)=P(X\le a-0.5)$$

$$P(a\le X\le b)=P(a-0.5\le X\le b+0.5)$$

مثال ۱ اگر بخواهیم $P(X\ge 10)$ را در توزیع دوجمله‌ای با توزیع نرمال تقریب بزنیم (در شرایطی که تقریب مجاز باشد)، کدام‌یک از موارد زیر را باید با تصحیح پیوستگی محاسبه کنیم؟

۱) $P(X\ge 9.5)$ ۲) $P(X\ge 10.5)$ ۳) $P(X\le 10.5)$ ۴) $P(X\le 9.5)$

حل: گزینه ۱ درست است.

$$P(X\ge 10)\equiv P(X\ge 10-0.5)=P(X\ge 9.5)$$

مثال ۲ اگر بخواهیم $P(X<8)$ را در توزیع پواسون به نرمال تقریب بزنیم (در شرایطی که $\lambda>10$ باشد)، کدام‌یک از موارد زیر با تصحیح پیوستگی حل شده است؟

۱) $P(X\le 8.5)$ ۲) $P(X\ge 7.5)$ ۳) $P(X\le 7.5)$ ۴) $P(X>8.5)$

حل: گزینه ۳ درست است.

مثال ۳ متغیر X دارای توزیع پواسون با میانگین 25 است. $P(X \le 32)$ با تصحیح پیوستگی و استفاده از توزیع نرمال کدام است؟ $\left(S_0^{1.5} = 0.4332\right)$

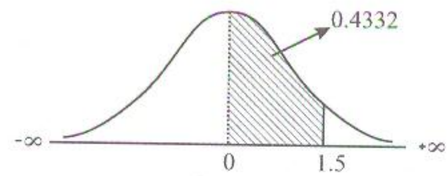
۱) 0.1336 ۲) 0.4332 ۳) 0.9332 ۴) 0.5668

حل: گزینه ۳ درست است.

$$P(X \le 32) \equiv P(X \le 32 + 0.5) = P\left(\frac{X-\mu}{\sigma} \le \frac{32.5-\lambda}{\sqrt{\lambda}}\right) = P\left(Z \le \frac{32.5-25}{5}\right) = P(Z \le 1.5)$$

$$S_0^{1.5} = P(0 < Z < 1.5) = 0.4332$$

$$\rightarrow \; P(Z < 1.5) = 0.5 + 0.4332 = 0.9332$$



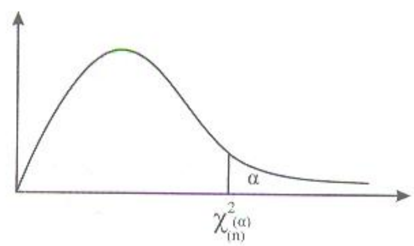
توجه: فقط هنگامی که در صورت سؤال ذکر شود با تصحیح پیوستگی، از تصحیح استفاده می‌کنیم در غیر این صورت مسئله را به شیوه معمول حل می‌کنیم.

توزیع‌های نتیجه‌شده از نرمال

۱- توزیع کای اسکور (خی دو ، کای دو، مربع کای)(Chi – Square Distribution)

متغیر تصادفی χ^2 در خانواده‌ای از توزیع‌ها قرار دارد که هر یک با یک پارامتر n مشخص می‌شوند. مقدار n همان درجه آزادی یا تعداد مشاهدات مستقل در نمونه است.

مشخصات توزیع $\chi^2_{(n)}$ (کای اسکور با n درجه آزادی)



۱- یک توزیع تک نمایی است.

۲- مقادیر $\chi^2_{(n)}$ همگی مثبت هستند $\left(\chi^2_{(n)} > 0\right)$.

۳- توزیعی نامتقارن و دارای چولگی مثبت است.

۴- امید و واریانس توزیع کای دو

اگر متغیر تصادفی X دارای توزیع $\chi^2_{(n)}$ (کای‌دو با n درجه آزادی) باشد، امید و واریانس آن برابر است با:

امید ریاضی (میانگین)	$E(X) = n$ (درجه آزادی)
واریانس	$\sigma^2(X) = 2n$ (2 برابر درجه آزادی)
انحراف معیار	$\sigma_X = \sqrt{2n}$

یادآوری: $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\ ,\ \Gamma(n) = (n-1)!$

مجموع و تفاضل دو متغیر χ^2

اگر کمیت‌های مستقل X و Y دارای توزیع χ^2 به ترتیب با m و n درجه آزادی باشند، آن‌گاه $X \pm Y$ دارای توزیع χ^2 با $m \pm n$ درجه آزادی است؛ به عبارت دیگر:

$$\begin{array}{l} X \sim \chi^2_{(m)} \\ Y \sim \chi^2_{(n)} \end{array} \rightarrow \begin{cases} \longrightarrow (X+Y) \sim \chi^2_{(m+n)} \\ \xrightarrow[m>n]{} (X-Y) \sim \chi^2_{(m-n)} \end{cases}$$

شکل توزیع χ^2 و درجه آزادی

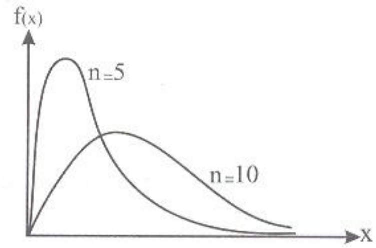
ازآنجاکه در توزیع $\chi^2_{(n)}$، درجه آزادی (n) تنها پارامتر توزیع است، شکل توزیع نیز باید توسط همین درجه آزادی مشخص شود.

شکل توزیع χ^2 بر اساس درجه آزادی (n) دو حالت دارد:

الف) اگر درجه آزادی کم باشد $(n \le 10)$، توزیع دارای چولگی مثبت است.

ب) اگر درجه آزادی بزرگ باشد $(n > 10)$، چولگی توزیع کم می‌شود و به سمت توزیع نرمال (قرینه) میل می‌کند.

$$\chi^2_{(n)} \xrightarrow{n>10} N\left(\mu = n , \sigma^2 = 2n\right)$$



رابطه χ^2 و نرمال استاندارد

بین توزیع χ^2 و توزیع نرمال استاندارد رابطه نزدیکی وجود دارد، به طوری که مجذور هر متغیر تصادفی نرمال استاندارد یک متغیر تصادفی χ^2 است. قضایای زیر، این رابطه را بررسی می‌کنند.

قضیه ۱: اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد، آن‌گاه:

$$\chi^2_{(1)} = \left(\frac{X-\mu}{\sigma}\right)^2$$

نتیجه: اگر متغیرهای تصادفی مستقل $X_1, X_2, \ldots, X_n$ دارای توزیع نرمال با میانگین μ_i و واریانس σ_i^2 باشند، آن‌گاه:

$$\chi^2_{(n)} = \sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$

قضیه ۲: اگر متغیر تصادفی Z دارای توزیع نرمال استاندارد با میانگین 0 و واریانس 1 باشد، آن‌گاه:

$$\chi^2_{(1)} = Z^2$$

نتیجه: اگر متغیرهای تصادفی مستقل $Z_1, Z_2, \ldots, Z_n$ دارای توزیع نرمال استاندارد با میانگین 0 و واریانس 1 باشند، آن‌گاه:

$$\chi^2_{(n)} = \sum_{i=1}^{n} z_i^2$$

مثال ۱ اگر Z_1, Z_2, Z_3 مستقل از هم دارای توزیع نرمال استاندارد باشند، توزیع $\sum_{i=1}^{3} z_i^2$ کدام است؟

۱) Z

۲) $t^2_{(3)}$

۳) $\chi^2_{(3)}$ (کای دو با 3 درجه آزادی)

۴) هیچ توزیعی را نشان نمی‌دهد.

حل: گزینه ۳ درست است.

با توجه به قضیه ۲ $\rightarrow$ $\left(Z_1^2 + Z_2^2 + Z_3^2\right) \sim \chi^2_{(3)}$

مثال ۲ درصورتی‌که کمیت‌های مستقل X و Y بر طبق توزیع خی‌دو (کای دو) با واریانس‌های 16 و 10 باشند، کمیت X + Y دارای چه توزیعی خواهد بود؟

۱) نرمال با واریانس $4+\sqrt{5}$

۲) χ^2 با 26 درجه آزادی

۳) نرمال با واریانس 16

۴) χ^2 با 13 درجه آزادی

حل: گزینه ۴ درست است.

$$\left.\begin{array}{l}\sigma_X^2 = 2m = 16 \rightarrow m = 8 \text{ درجه آزادی} \\ \sigma_Y^2 = 2n = 10 \rightarrow n = 5 \text{ درجه آزادی}\end{array}\right\} \rightarrow (X+Y) \sim \chi^2_{m+n} = \chi^2_{13}$$

دقت کنید اگر مقادیر 16 و 10 به عنوان امید ریاضی مطرح می‌شدند، همان درجه آزادی بودند و گزینه ۲ پاسخ درست بود، زیرا:

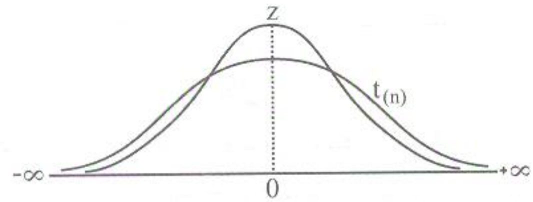
$$\begin{cases}\mu_X = m = 16 \rightarrow X \sim \chi^2_{(16)} \\ \mu_Y = n = 10 \rightarrow Y \sim \chi^2_{(10)}\end{cases} \rightarrow (X+Y) \sim \chi^2_{(16+10)} = \chi^2_{26}$$

نکته:

۱- هرگاه X دارای توزیع $\chi^2_{(n)}$ با n درجه آزادی باشد، توزیع X معادل توزیع گاما با پارامترهای $\alpha = \frac{n}{2}$ و $\lambda = \frac{1}{2}$ است.

۲- هرگاه X دارای توزیع $\chi^2_{(2)}$ با 2 درجه آزادی باشد، توزیع X معادل توزیع نمایی با پارامتر $\lambda = \frac{1}{2}$ است. (دقت کنید که این رابطه فقط برای $n = 2$ برقرار است.)

۲- توزیع t– استیودنت (Student's t – Distribution)



توزیع t مانند توزیع نرمال متقارن است، اما نسبت به آن پراکندگی بیشتری دارد. هر متغیر t با یک درجه آزادی مشخص می‌شود که تعیین‌کننده ارتفاع و میزان پراکندگی توزیع است (درجه آزادی تعداد مشاهده مستقل در هر نمونه دلخواه از جامعه است).

امید و واریانس توزیع t

امید ریاضی (میانگین)	$E(X)=0 \quad , \quad n>1$
واریانس	$\sigma^2(X)=\dfrac{n}{n-2} \quad , \quad n>2$
انحراف معیار	$\sigma_X=\sqrt{\dfrac{n}{n-2}} \quad , \quad n>2$

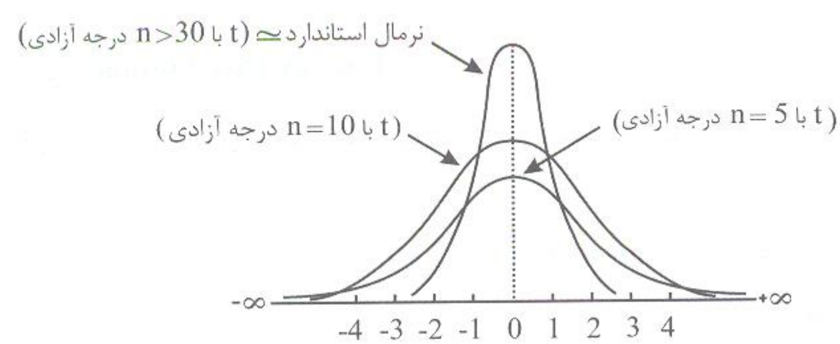
رابطه توزیع t با توزیع نرمال استاندارد و کای‌دو

اگر متغیر تصادفی Z دارای توزیع **نرمال استاندارد** باشد $\left(Z \sim N(0,1)\right)$ و متغیر تصادفی $\chi^2_{(n)}$ (**کای اسکور با n درجه آزادی**) مستقل از آن مفروض باشد، آن‌گاه کسر زیر دارای توزیع t **استیودنت با n درجه آزادی** است:

$$t_{(n)}=\frac{Z}{\sqrt{\dfrac{\chi^2_{(n)}}{n}}}$$

درواقع حاصل تقسیم یک متغیر نرمال استاندارد به جذر متغیر χ^2 به درجه آزادی‌اش، برابر توزیع t با درجه آزادی χ^2 است.

هر چه حجم نمونه مستقل یا همان n (درجه آزادی) بیشتر باشد، توزیع t متمرکزتر و پراکندگی آن کمتر می‌شود به طوری که برای نمونه‌های بزرگ $(n>30)$، توزیع t تقریباً نرمال است.



رابطه مقادیر استاندارد و سطح زیر منحنی با کشیدگی

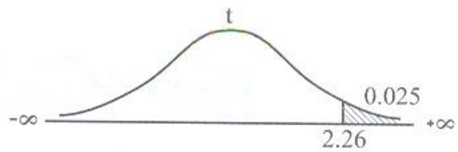
هرگاه یک توزیع متقارن، کوتاه‌تر یا بلندتر از توزیع نرمال استاندارد باشد، مقادیر استاندارد و سطح زیر منحنی برای آن توزیع نسبت به توزیع نرمال، به صورت زیر بررسی می‌شود:

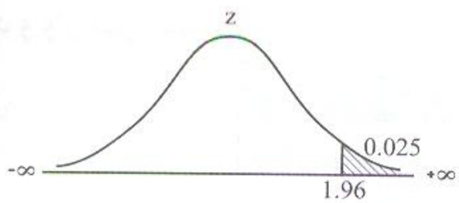
توزیع کوتاه‌تر از نرمال

هرگاه کشیدگی یک توزیع متقارن مانند t کمتر از کشیدگی نرمال استاندارد باشد (کوچک‌تر از 3) آنگاه:

۱- برای سطوح مشابه، مقدار استاندارد در توزیع مورد نظر **بزرگ‌تر** از مقدار استاندارد نرمال است.

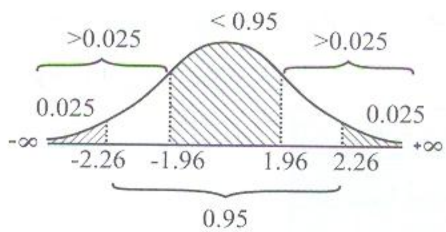
برای مثال، اگر سطح 0.025 را برای توزیع‌های نرمال و t درنظر بگیریم، داریم:

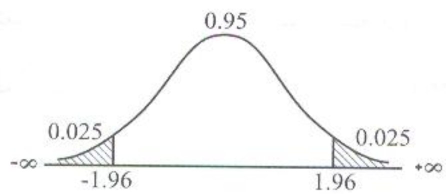




۲- برای هر بازه دلخواه از مقادیر استاندارد (بحرانی یا معنی‌دار) مانند $(-k,k)$، سطح بین دو مقدار، در توزیع نرمال و سطح خارج از دو مقدار، در توزیع مورد نظر بزرگ‌تر است.

برای مثال، اگر بازه $(-1.96,1.96)$ را برای توزیع‌های نرمال و t در نظر بگیریم، داریم:





توزیع بلندتر از نرمال

هرگاه کشیدگی یک توزیع متقارن بیشتر از کشیدگی نرمال استاندارد باشد (بزرگ‌تر از 3)، شرایط دقیقاً عکس حالتی است که توزیع مورد نظر کوتاه‌تر از نرمال باشد.

۳- توزیع کوشی (Cauchy Distribution)

هر گاه متغیر تصادفی پیوسته X دارای توزیع کوشی با پارامترهای (μ,σ) باشد، آن را به صورت $C(\mu,\sigma)$ نمایش می‌دهند و تابع چگالی آن برابر است با:

$$f(x)=\frac{1}{\pi\sigma}.\frac{1}{1+\left(\frac{x-\mu}{\sigma}\right)^2} \quad ; \quad -\infty<x<+\infty$$

کوشی استاندارد: اگر X دارای توزیع کوشی با پارامتر های $\mu = 0$ و $\sigma = 1$ باشد، $\left(C(0,1)\right)$ به کوشی استاندارد معروف است. که در این صورت با توزیع $t -$ استیودنت با درجه آزادی یک برابر است.

$$f(x) = \frac{1}{\pi\left(1+x^2\right)} \sim t_{(1)} \quad ; \quad -\infty < x < +\infty$$

نکته: توزیع کوشی تابع مولد گشتاور ندارد؛ بنابراین امید ریاضی هیچ توانی از X را نیز نخواهد داشت؛ به عبارت دیگر توزیع کوشی فاقد میانگین و واریانس است.

قضیه: توزیع $t_{(1)}$ با یک درجه آزادی حاصل تقسیم دو متغیر تصادفی نرمال استاندارد است که در این وضعیت برابر با توزیعی به نام کوشی استاندارد با تابع چگالی زیر است که امید و واریانس ندارد.

$$t_{(1)} = \frac{Z}{Z_i} = \frac{1}{\pi\left(1+x^2\right)}$$

اثبات:

$$t_{(1)} = \frac{Z}{\sqrt{\frac{\chi^2_{(1)}}{1}}} = \frac{Z}{\sqrt{\frac{Z_i^2}{1}}} = \frac{Z}{Z_i}$$

۴- توزیع فیشر (F – Distribution)

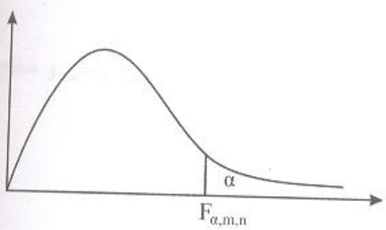
توزیع F (فیشر) در خانواده‌ای از توزیع‌ها قرار دارد که با دو پارامتر m و n مشخص می‌شوند و از تقسیم دو متغیر تصادفی مستقل کای‌دو بر درجه آزادی آن‌ها به دست می‌آید.

$$F_{m,n}=\frac{\frac{\chi^2_{(m)}}{m}}{\frac{\chi^2_{(n)}}{n}}=\frac{n\chi^2_{(m)}}{m\chi^2_{(n)}}$$

مقدار n، درجه آزادی توزیع $\chi^2_{(n)}$ است که در صورت کسر قرار گرفته است.

مقدار m، درجه آزادی توزیع $\chi^2_{(m)}$ است که در مخرج کسر قرار گرفته است.

مشخصات توزیع $F_{m,n}$ (فیشر با m , n درجه آزادی)



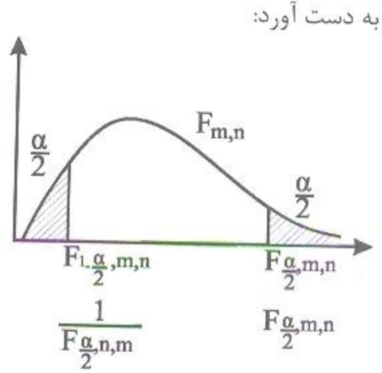
۱- یک توزیع تک نمایی است.

۲- مقادیر $F_{m,n}$ همگی غیر منفی (مثبت) هستند ($F_{m,n}>0$).

۳- توزیعی نامتقارن است که چولگی آن مثبت است.

۴- اگرچه توزیع F نامتقارن است اما نوعی تقارن معکوس در آن وجود دارد که بر اساس آن می‌توان حدود پایین را به صورت زیر به دست آورد:

$$F_{1-\frac{\alpha}{2},m,n}=\frac{1}{F_{\frac{\alpha}{2},n,m}}$$



۵- امید و واریانس توزیع F

اگر متغیر تصادفی X دارای توزیع $F_{m,n}$ با m و n درجه آزادی باشد، آن‌گاه امید و واریانس آن عبارت است از:

$$E(X)=\frac{n}{n-2}\qquad n>2$$

$$\sigma_X^2=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}\ \overset{\text{در بعضی منابع}}{=\!=}\ 2\left(\frac{1}{m}+\frac{1}{n}\right)\qquad n>4$$

۶- توزیع F با زیاد شدن درجات آزادی ($n,m>30$) به سمت توزیع **نرمال** میل می‌کند.

$$F_{n,m}\sim N$$
$$n,m\to\infty$$

مثال ۱ اگر متغیرهای تصادفی U و V مستقل از یکدیگر بوده و هر یک به ترتیب دارای توزیع χ^2 با $m-1$ و $n-1$ درجه آزادی باشند، آن‌گاه متغیر تصادفی $\frac{(n-1)U}{(m-1)V}$ دارای چه توزیعی خواهد بود؟

۱) $F_{(n-1,m-1)}$ ۲) $F_{(n,m)}$ ۳) $F_{(m-1,n-1)}$ ۴) $F_{(m,n)}$

حل: گزینه ۳ درست است.

$$\frac{(n-1)u}{(m-1)v}=\frac{\frac{u}{m-1}}{\frac{v}{n-1}}=\frac{\frac{\chi^2_{(m-1)}}{m-1}}{\frac{\chi^2_{(n-1)}}{n-1}}=F_{m-1,n-1}$$

مثال ۲ توزیع $F_{1,n}$ (فیشر با $1,n$ درجه آزادی) معادل کدام توزیع است؟

۱) $\chi^2_{(n)}$ ۲) Z^2 ۳) $t^2_{(n)}$ ۴) Z

حل: گزینه ۳ درست است.

$$F_{1,n}=\frac{\frac{\chi^2_{(1)}}{1}}{\frac{\chi^2_{(n)}}{n}}= \xrightarrow{Z^2=\chi^2_{(1)}} =\frac{\frac{Z^2}{1}}{\frac{\chi^2_{(n)}}{n}}=\left(\frac{Z}{\sqrt{\frac{\chi^2_{(n)}}{n}}}\right)^2=t^2_{(n)}$$

نتیجه: توزیع $F_{1,n}$ همان مجذور توزیع $t_{(n)}$ است.

$$\boxed{\left(t_{(n)}\right)^2=F_{1,n}}$$

مثال ۳ اگر $F_{0.05,2,10}=4.1$ باشد، $F_{0.95,10,2}$ کدام است؟

۱) 0.24 ۲) 4.1

۳) 3.1 ۴) باید جدول F در دسترس باشد.

حل: گزینه ۱ درست است.

$$F_{0.95,10,2}=\frac{1}{F_{0.05,2,10}}=\frac{1}{4.1}=0.24$$

مثال ۴ اگر X_1 و X_2 دارای توزیع نرمال استاندارد مستقل باشند، توزیع متغیر تصادفی $Y=\frac{(X_1+X_2)^2}{(X_1-X_2)^2}$ عبارت است از:

۱) مربع کای با یک درجه آزادی ۲) مربع کای با دو درجه آزادی

۳) توزیع F با یک و یک درجه آزادی ۴) توزیع t با یک درجه آزادی

حل: گزینه ۳ درست است.

$$X_1+X_2\sim N(0,2)$$

$$X_1-X_2\sim N(0,2)$$

$$Y=\frac{\left(\frac{X_1+X_2-0}{\sqrt{2}}\right)^2}{\left(\frac{X_1-X_2-0}{\sqrt{2}}\right)^2}=\frac{(X_1+X_2)^2}{(X_1-X_2)^2}=\frac{Z_1^2}{Z_2^2}=\frac{\chi^2_{(1)}}{\chi^2_{(1)}}=F_{1,1}$$

مثال ۵ با افزایش حجم نمونه و میل کردن آن به سمت بی‌نهایت $(n \to \infty)$، توزیع‌های F, t, χ^2 به سمت ....

۱) توزیع پواسون میل می‌کنند.
۲) توزیع دوجمله‌ای میل می‌کنند.
۳) توزیع یکنواخت میل می‌کنند.
۴) توزیع نرمال میل می‌کنند.

حل: گزینه ۴ درست است.

در تمام توزیع‌های پیوسته $P(X = a) = 0$ است.