بسمه تعالی

موضوع : مبانی آمار و احتمال برای ورود به علم داده و تحلیل داده

بخش پانزدهم: انواع تابع توزیع پیوسته

- یکنواخت
- نرمال (Normal Distribution)
- نمایی

در این پارت از اموزش به بررسی **انواع تابع توزیع پیوسته** می پردازیم .

در اموزش های قبلی راجب **انواع توابع توزیع گسسته** صحبت کردیم.

در بخش قبل درباره توزیع‌های گسسته صحبت کردیم؛ یعنی متغیرهایی که مقادیرشان قابل شمارش بود، مثل:

- تعداد مشتریان
- تعداد خرابی‌ها
- تعداد تماس‌ها
- تعداد موفقیت‌ها

اما همه متغیرهای تصادفی قابل شمارش نیستند.

مثلاً فرض کنید می‌خواهیم سن، قد، وزن، زمان انتظار مشتری یا زمان خرابی یک دستگاه را بررسی کنیم.این متغیرها می‌توانند در یک بازه، مقادیر بسیار زیادی داشته باشند و به صورت پیوسته تغییر کنند.به این نوع متغیرها، متغیر تصادفی پیوسته (Continuous Random Variable) می‌گوییم.

در این حوزه توزیع های مهم عبارتند از :

- یکنواخت
- نرمال
- نمایی

اما توزیع های دیگری نیز وجود دارد:

- توزیع گاما(Gamma)
- توزیع بتا(Beta)
- توزیع کای‌دو(Chi-Square)
- توزیع t استیودنت(Student's t)
- توزیع F

در توزیع یکنواخت، احتمال وقوع مقادیر در یک بازه مشخص به صورت یکنواخت توزیع شده است.

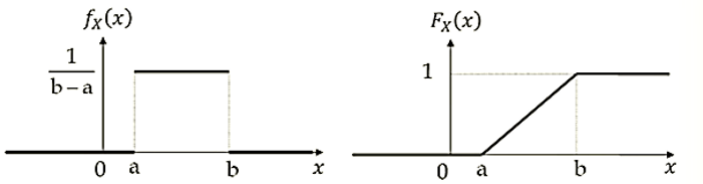
فرض کنید زمان ورود یک مشتری به یک بانک در فاصله ۱۰ تا ۲۰ دقیقه، بدون ترجیح خاصی نسبت به لحظات مختلف این بازه مدل شود.

در این حالت می‌توان از توزیع یکنواخت استفاده کرد.

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

تابع توزیع تجمعی این متغیر تصادفی به صورت زیر محاسبه می شود:

$$\begin{cases}0 & x\le a\\ \frac{x-a}{b-a} & a<x<b\\ 1 & b\le x\end{cases}$$



امید ریاضی و واریانس متغیر تصادفی یکنواخت پیوسته به صورت زیر محاسبه می شود:

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

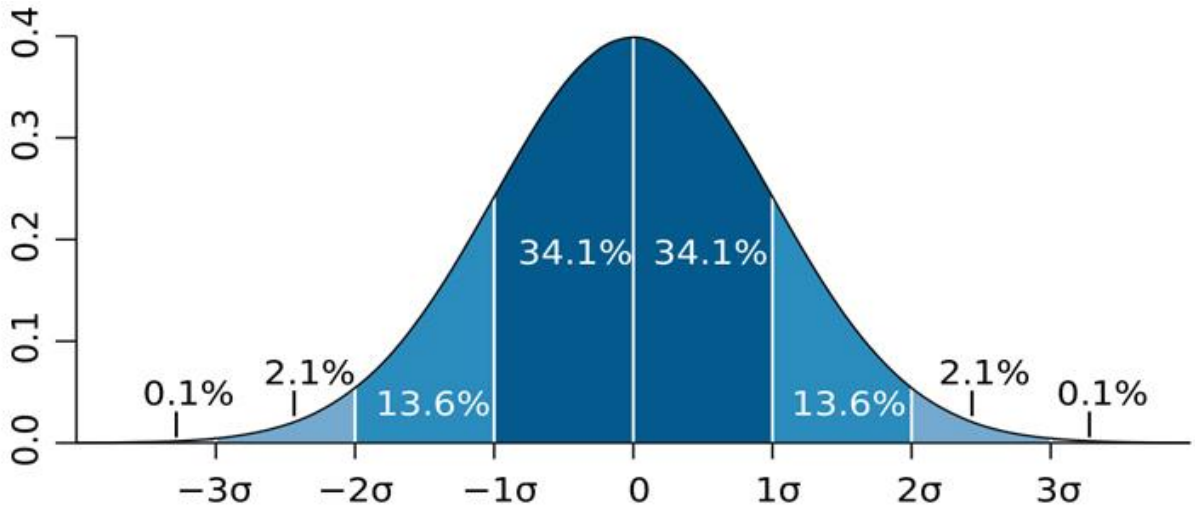
$$Var(X)=\frac{(b-a)^2}{12}$$

حالا به یکی از مهم‌ترین توزیع‌های آماری می‌رسیم:

توزیع نرمال:

این توزیع در بسیاری از مسائل آماری کاربرد دارد.

شکل آن معمولاً به صورت یک منحنی زنگوله‌ای و متقارن است.



ویژگی‌های مهم آن:

- متقارن است .
- میانگین، میانه و نما در حالت نرمال دقیقاً برابرند .
- دو طرف منحنی مشابه یکدیگر هستند .
- بخش عمده داده‌ها در اطراف میانگین قرار می‌گیرند.

پارامترهای توزیع نرمال:

توزیع نرمال با دو پارامتر مشخص می‌شود:

μ = میانگین

σ = انحراف معیار

یا به صورت:

$X \sim N(\mu, \sigma^2)$

نوشته می‌شود.

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

برای تابع تجمعی آن به دلیل اینکه محاسبه آن از روی تابع چگالی سخت و زمان بر است راه حلی ارائه کردند که در آن ابتدا متغیر تصادفی را به آماره استاندارد تبدیل کرده و بعد احتمال وقوع آن را از جدول تجمعی توزیع نرمال محاسبه می کنیم.

آماره استاندار عبارت است از :

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

مثلاً اگر:

$\mu = 50$

$\sigma = 10$

و یک فرد مقدار:

$X = 70$

داشته باشد:

Z = (70−50)/10 = 2

یعنی مقدار ۷۰، دو انحراف معیار بالاتر از میانگین قرار دارد.

و در اخر **توزیع نمایی** یکی از توزیع‌های بسیار مهم در کاربردهای عملی است.

این توزیع معمولاً برای مدل‌سازی زمان انتظار تا وقوع یک رویداد استفاده می‌شود.

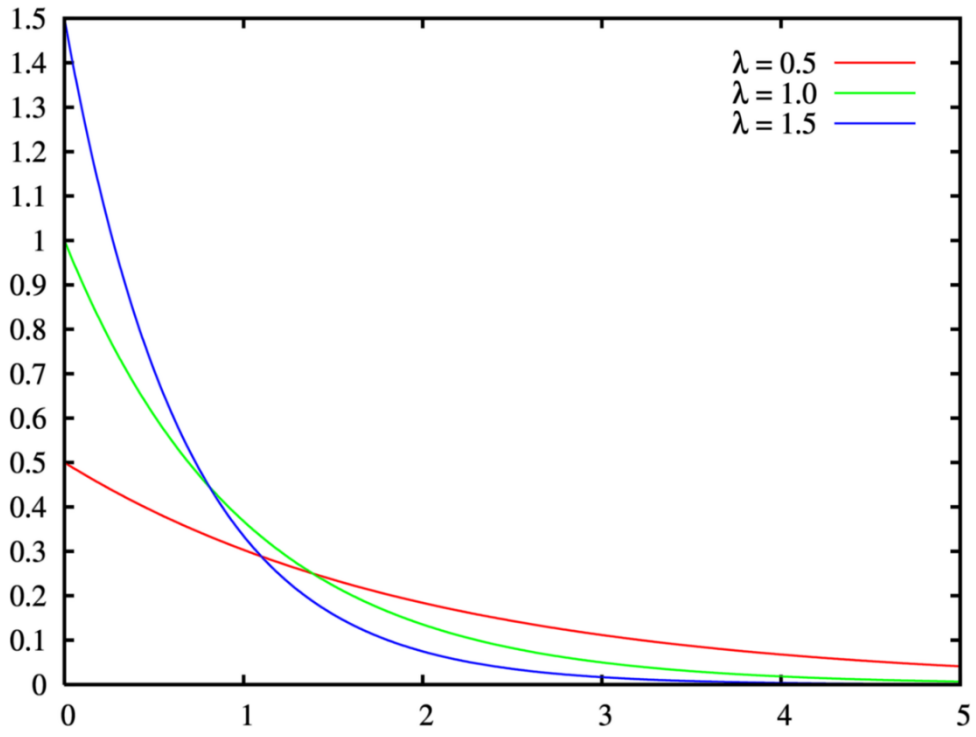
مثلاً:

- زمان تا خرابی یک دستگاه
- زمان انتظار تا تماس بعدی
- زمان بین ورود دو مشتری
- زمان تا وقوع یک خرابی

اگر نرخ وقوع رویداد را با λ نشان دهیم. تابع توزیع آن به صورت زیر خواهد بود:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

که نموداری مشابه تصویر زیر دارد:



تابع توزیع تجمعی متغیر تصادفی نمایی برابر است با :

$$\boldsymbol{F(a) = 1 - e^{-\lambda a}}$$

همچنین امید ریاضی و واریانس متغیر تصادفی نمایی برابر است با :

$$E(X) = \frac{1}{\lambda} \qquad Var(X) = \frac{1}{\lambda^2}$$

مثلاً اگر نرخ وقوع خرابی: $\lambda = 0.2$ خرابی در ساعت باشد، میانگین زمان انتظار تا خرابی: $E(X) = 1/0.2 = 5$ ساعت است.

این توزیع در مباحث **قابلیت اطمینان و نگهداری و تعمیرات** هم کاربرد بسیار زیادی دارد.