随机变量的数字特征

讲师：Jeary

目录

目标

通过本章课程的学习，您将能够：

- 熟悉随机变量的概念
- 了解随机变量的分布
- 运用随机变量的数字特征（期望、方差、协方差、相关系数）

01

离散型随机变量

随机变量

定义

设随机试验的样本空间为$S=\{e\}$，$X=X(e)$是定义在样本空间S上的实值单值函数，称$X=X(e)$为随机变量，

即：随机变量是表示随机试验各种结果的**实值单值函数**

举例

随机试验E: 抛两枚骰子，观察可能出现的点数的和

试验的样本空间是$S=\{e\}=\{(i,j)|i,j=1,2,3,4,5,6\}$，$i$，$j$分别是两枚骰子出现的点数，以$X$记为两枚骰子出现的点数，则$X$是一个随机变量

$$X=X(e)=X(i,j)=i+j,i,j=1,2,\cdots,6$$

离散型随机变量

定义

离散型随机变量即表示在一定区间内变量取值为有限个或可数个

举例

某地区某年人口的出生数、死亡数，某药治疗某病病人的有效数、无效数等

分布律

定义

设离散型随机变量X的所有可能取值为$x_k(k=1,2,\cdots)$，X取各个可能值的概率，即事件$\{X=x_k\}$的概率，为：

$$P\{X=x_k\}=p_k,k=1,2,\cdots$$

表示形式

$\boldsymbol{X}$	x_1	x_2	…	x_n	…
$\boldsymbol{p_k}$	p_1	p_2	…	p_n	…

伯努利分布

定义

设随机变量X只可能取0与1两个值，它的分布律是：

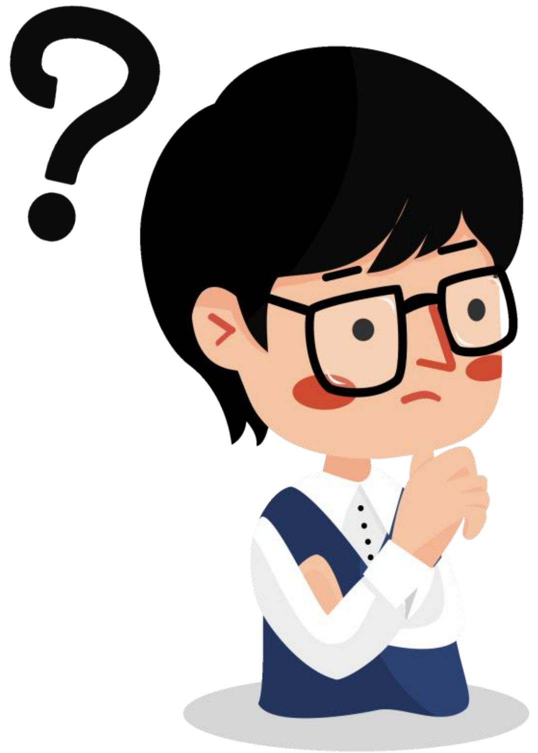
$$P\{X=k\}=p^k(1-p)^{1-k}，\ k=0，\ 1\ (0<p<1)$$

则称X服从参数为p的伯努利分布（也称为0-1分布或两点分布）

表示形式

$\boldsymbol{X}$	0	1
$\boldsymbol{p_k}$	$1-p$	p

思考题



01 硬币掉落后是正面朝上吗?

02 刚出生的小孩是个女孩吗?

03 潜在客户会买我的产品吗?

04 市民会给特定的候选人投票吗?

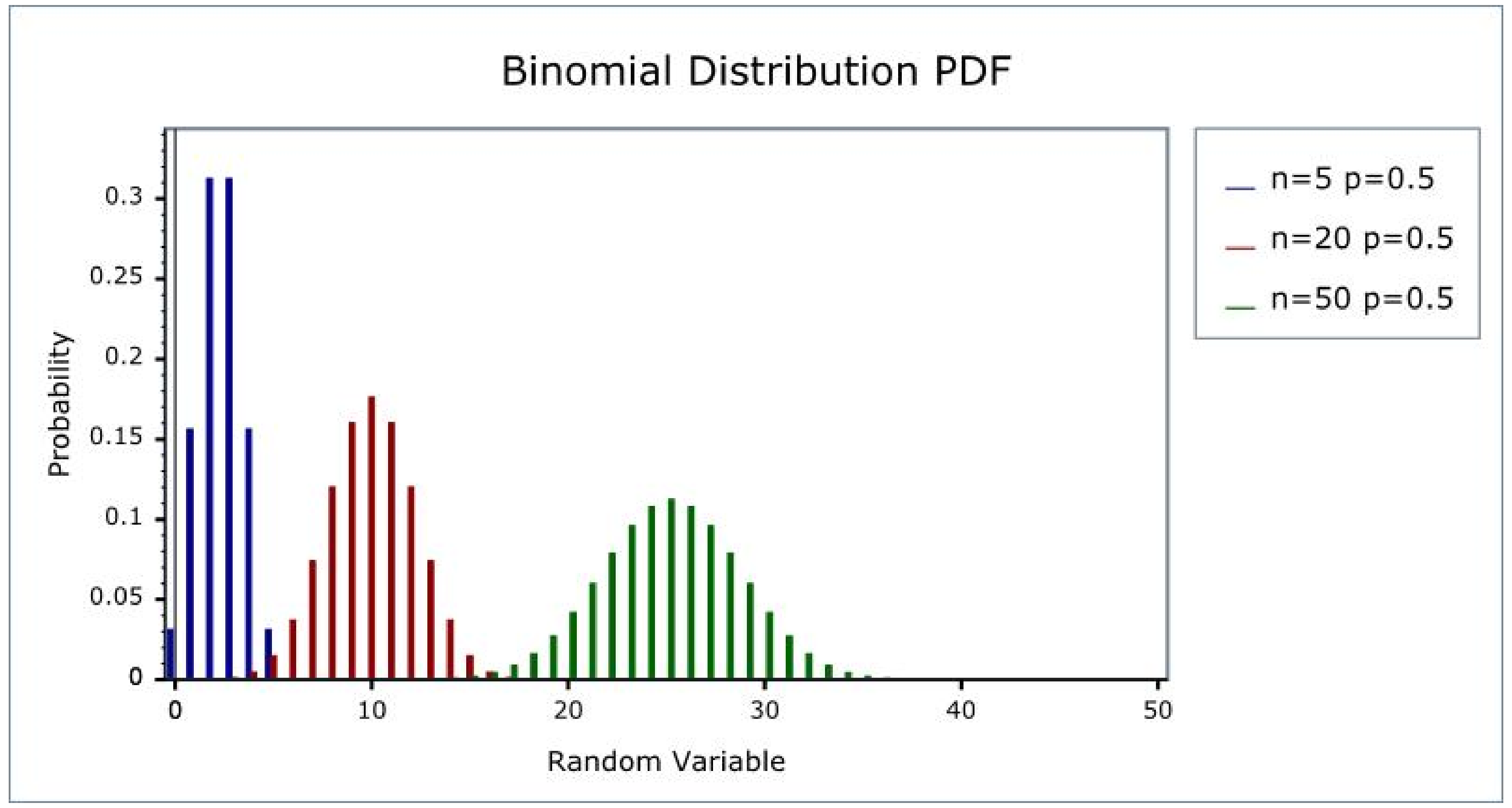
二项分布

若用X表示n重伯努利试验中事件A发生的次数，则n次试验中事件A发生k次的概率为：

$$P(X=k)=C_n^k p^k(1-p)^{n-k}, k=0,1,2,\cdots n$$

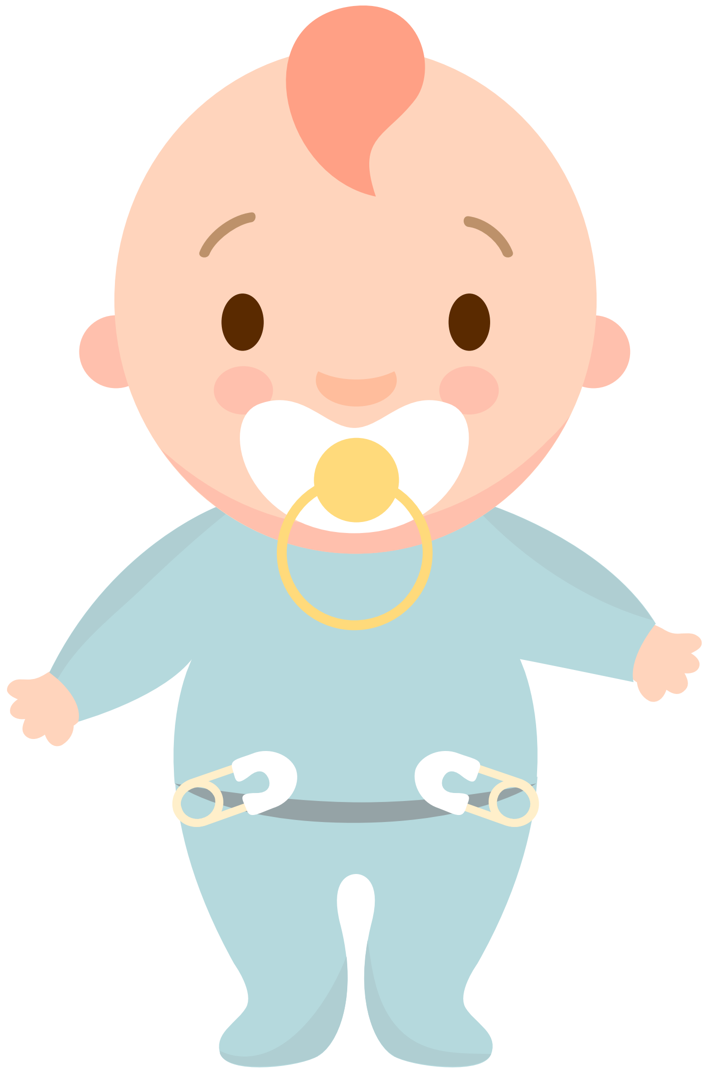
则称X服从参数为n,p的二项分布，记为：

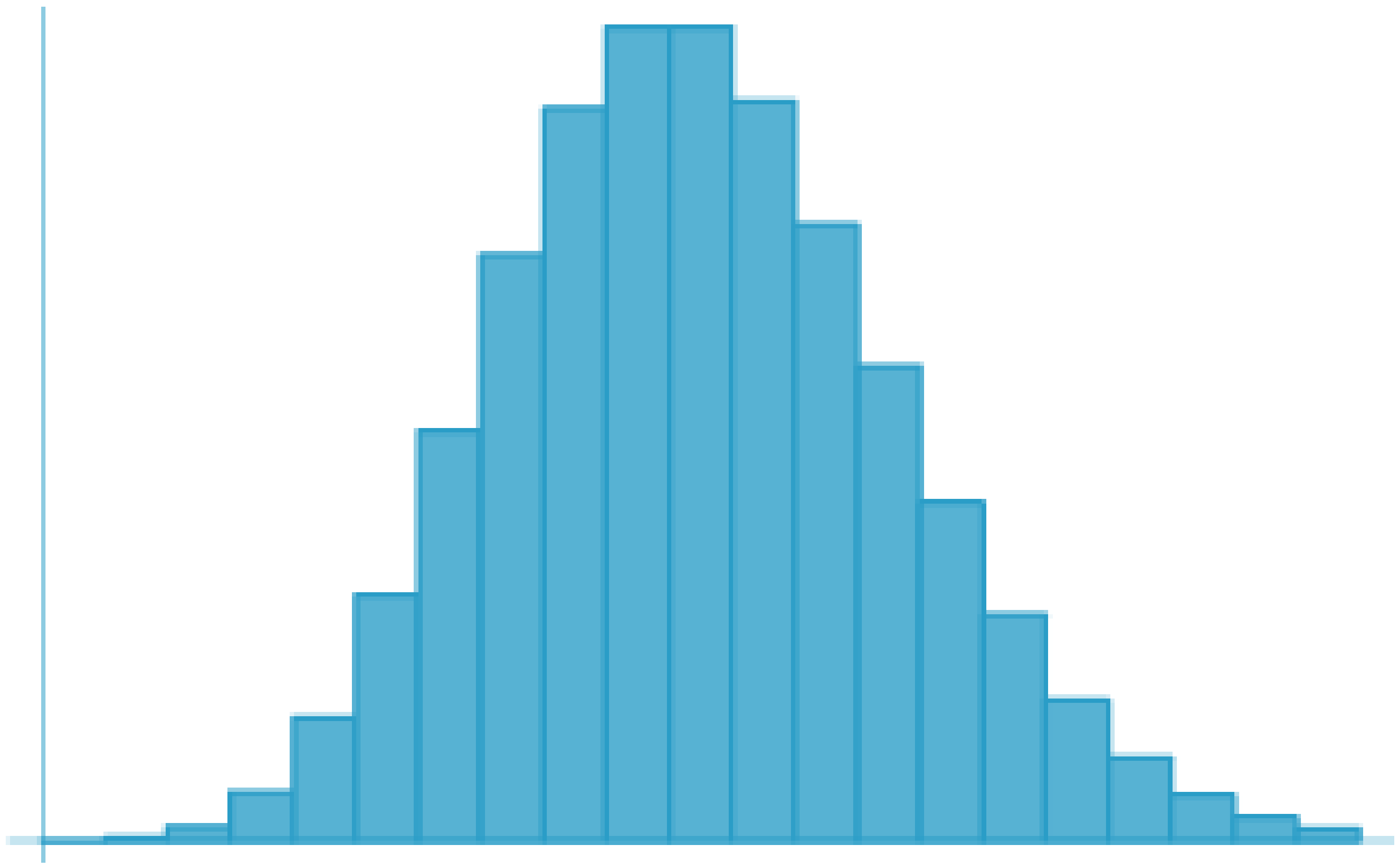
$$X\sim B(n,p)$$



案例

已知平均每小时出生3个婴儿，接下来两个小时，一个婴儿都不出生的概率。





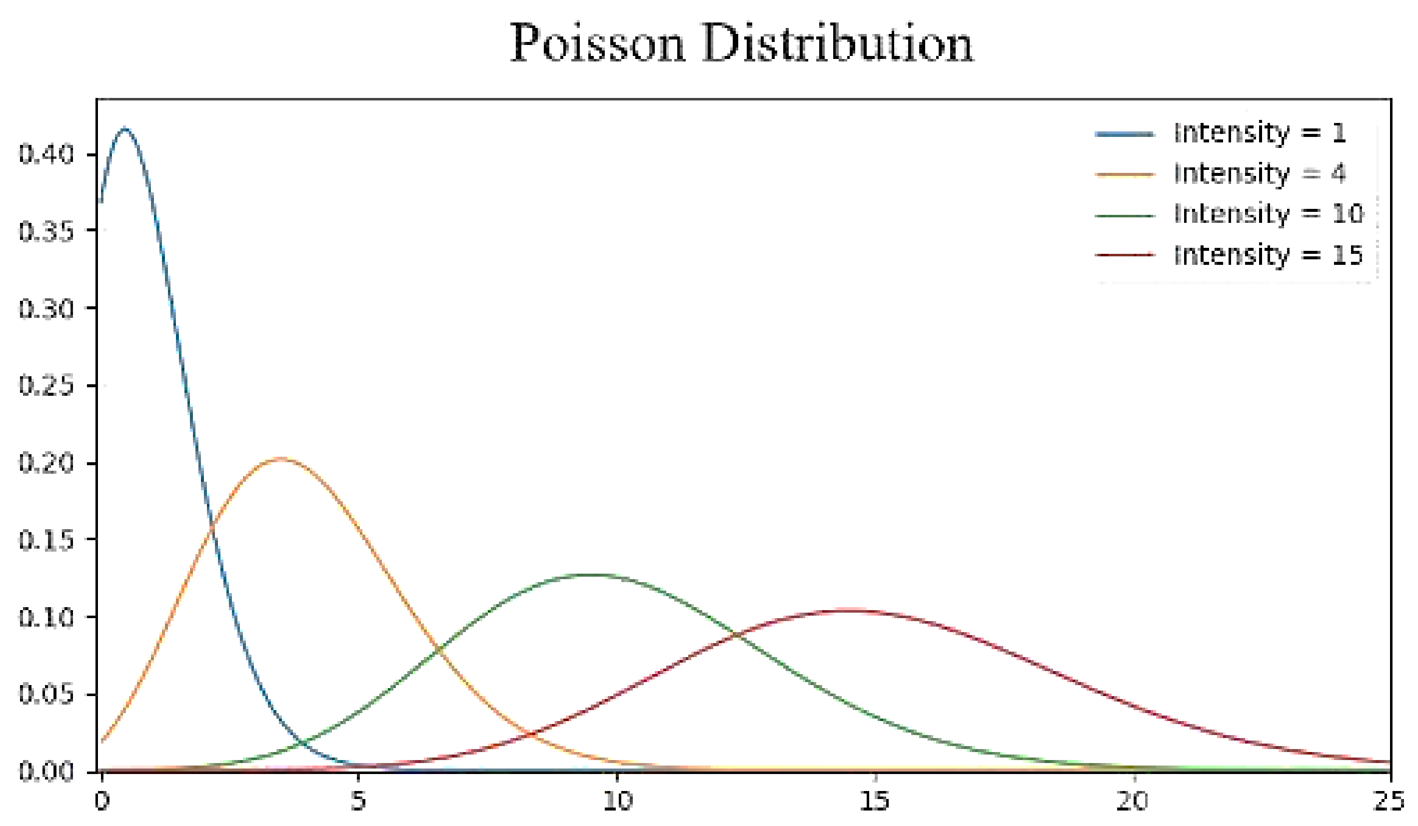
泊松分布

设随机变量所有可能的取值为0，1，2，…，而取每个值的概率为：

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\cdots$$

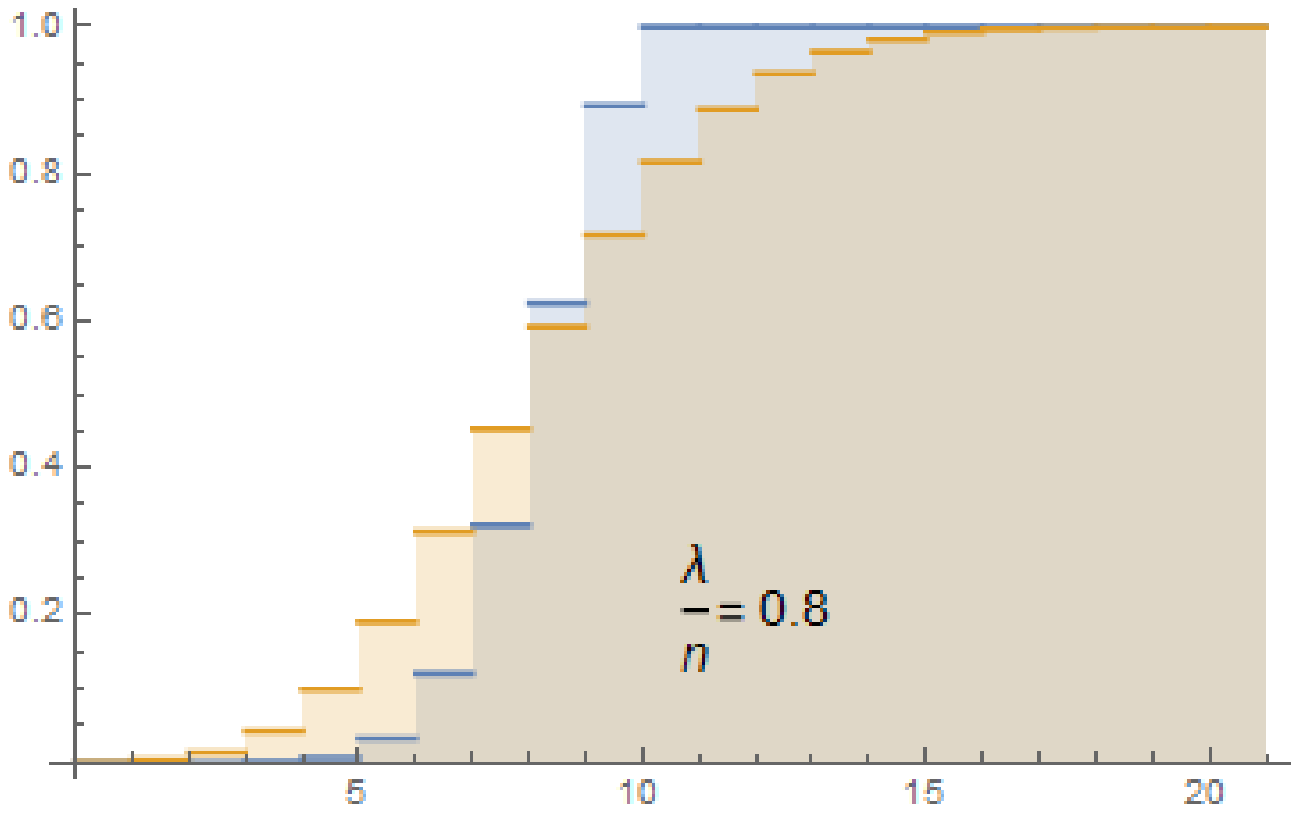
其中$\lambda>0$是常数，则称X服从参数为λ的泊松分布，记为：

$$X \sim \pi(\lambda)$$



泊松定律

当二项分布的n很大而p很小时，泊松分布可作为二项分布的近似，其中λ为np。通常当$n \geq 20, p \leq 0.05$时，就可以用泊松公式近似得计算：



02

连续型随机变量

连续型随机变量

定义

连续型随机变量在一定区间内变量取值有无限个，或数值无法一一列举出来

举例

例如某地区男性健康成人的身长值、体重值，一批新冠肺炎患者的血氧饱和度测定值等

分布函数

定义

设X是一个随机变量，x是任意实数，函数$F(x)$称为X的**分布函数**：

$$F(x) = P\{X \le x\}, -\infty < x < \infty$$

性质

1. $F(x)$是一个不减函数
2. $0 \le F(x) \le 1$，且$F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0$，$F(\infty) = \lim\limits_{x \to \infty} F(x) = 1$
3. $F(x+0) = F(x)$，即$F(x)$是右连续的

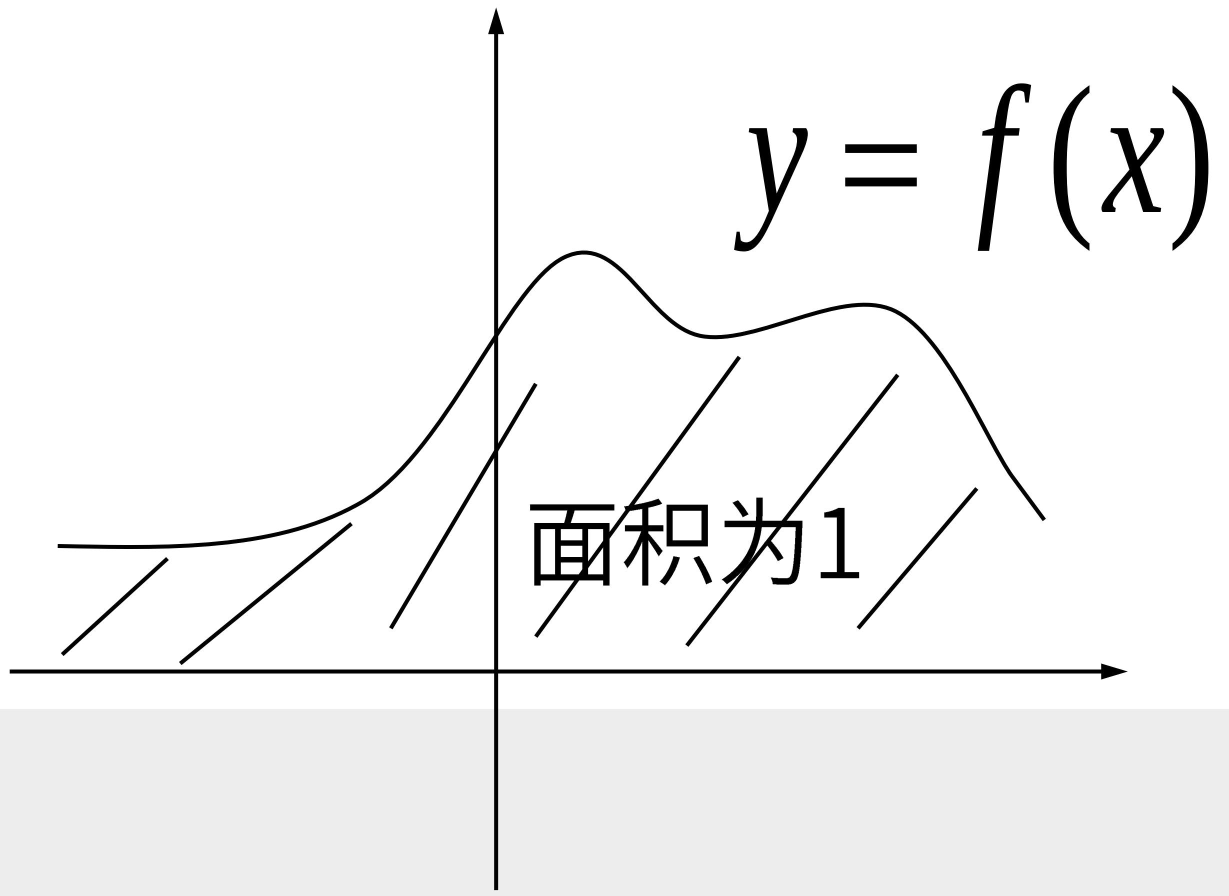
概率密度函数

如果对于随机变量X的分布函数$F(x)$，存在非负函数$f(x)$，使对于任意实数x有

$$F_X(a) = \int_{-\infty}^{a} f_X(x)dx$$

则称X为连续型随机变量，$f_X(x)$为X的概率密度函数，简称**概率密度**

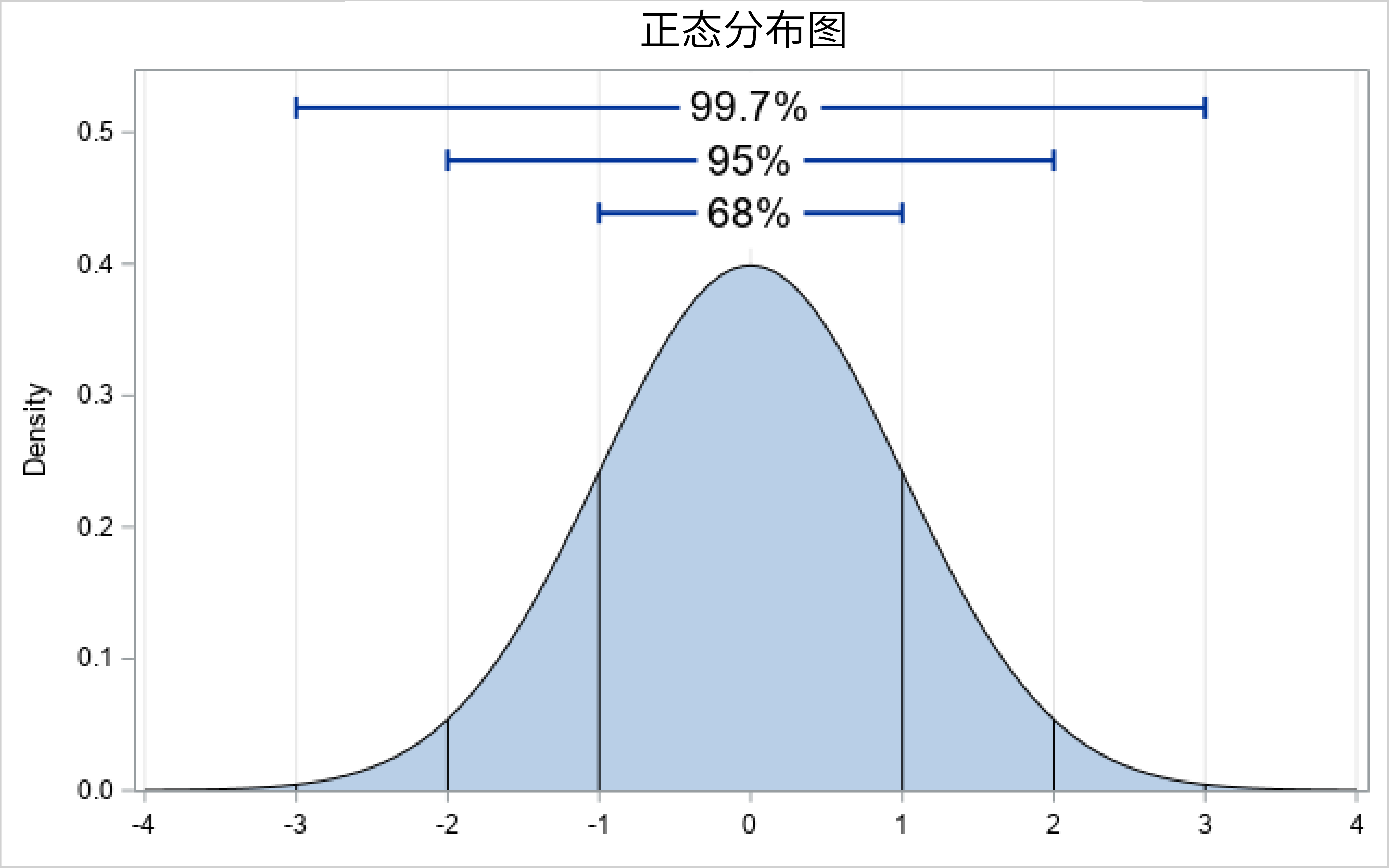
概率密度函数



■ 性质

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. 对于任意实数x_1，$x_2(x_1 < x_2)$，$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx$
4. 若$f(x)$在点x处连续，则有 $F'_X(x) = f_X(x)$
5. 连续型随机变量X取任一实数的概率值为0，即$P(X = a) = 0$

正态分布



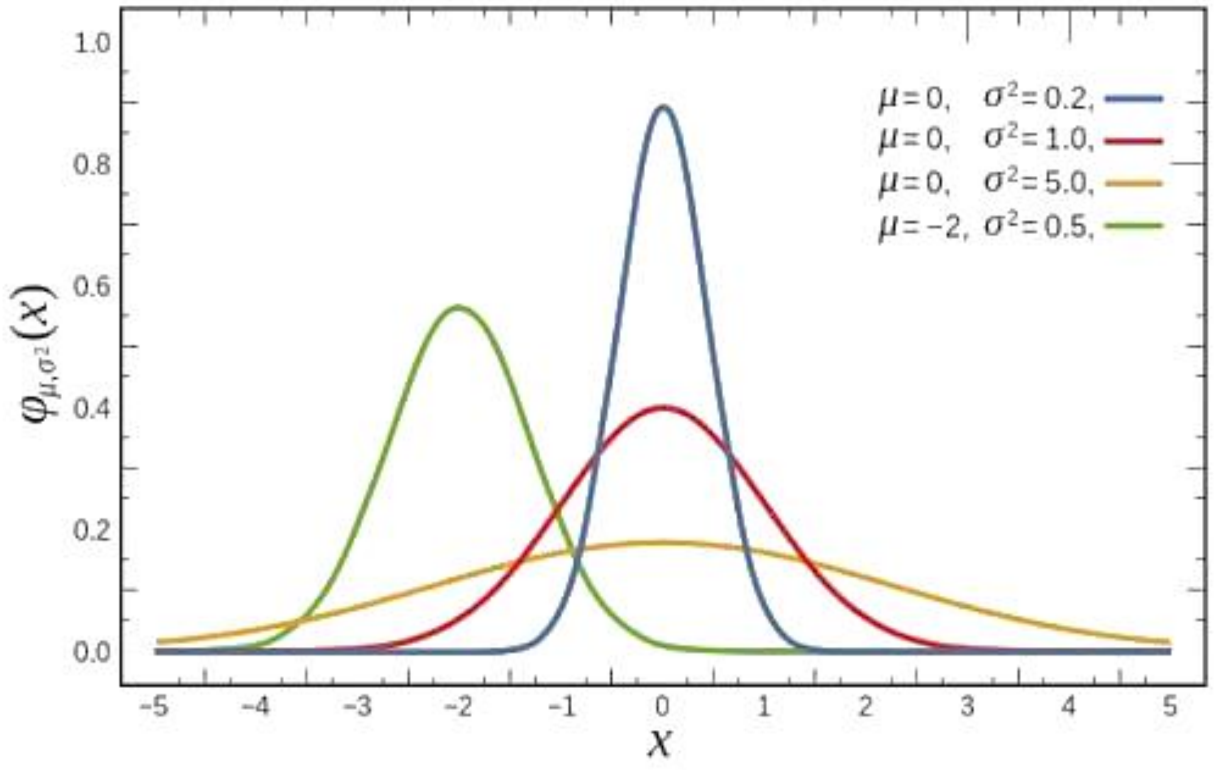
正态分布

若连续型随机变量X的概率密度函数为：

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

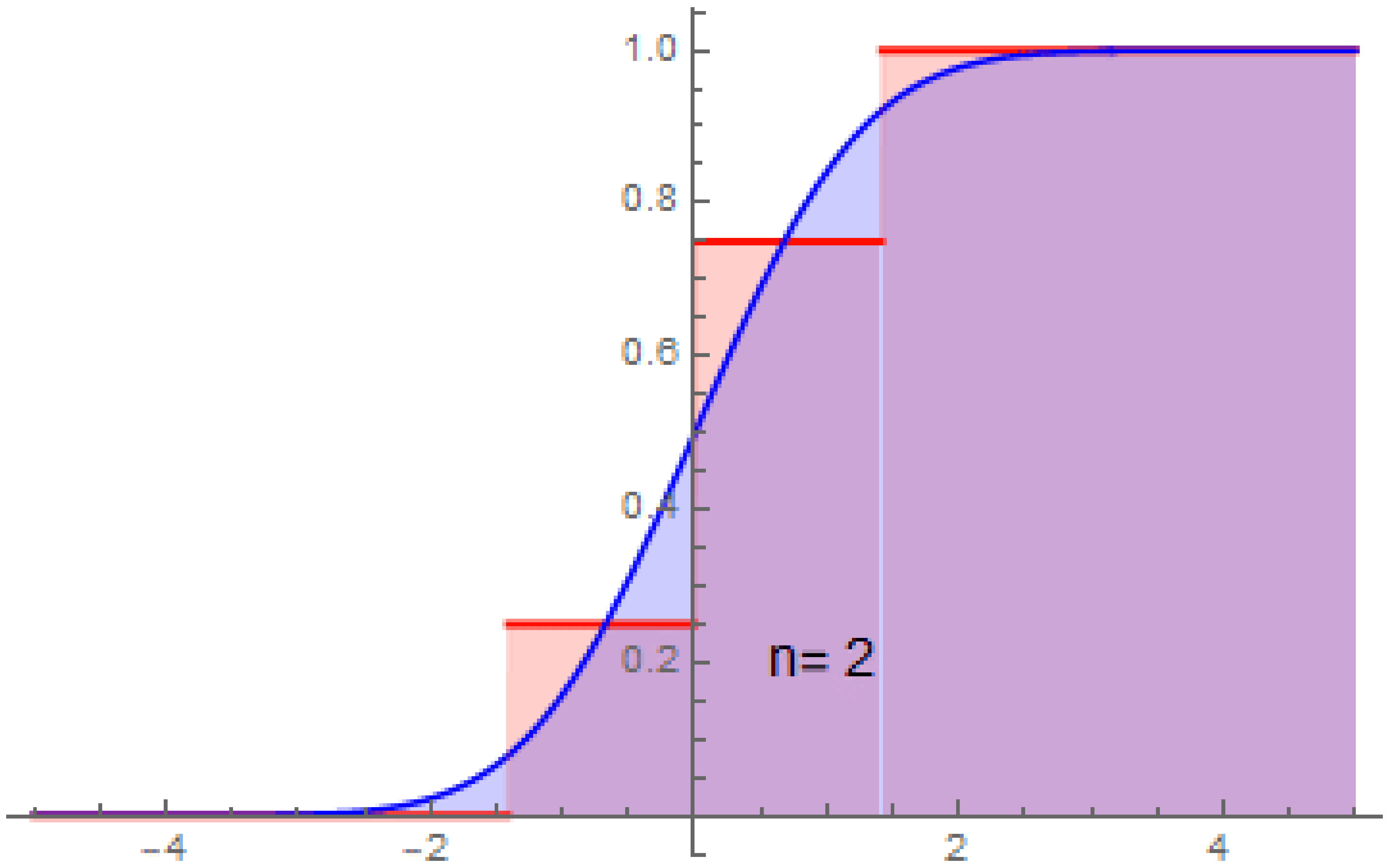
其中$\mu,\sigma(\sigma>0)$为常数，则称X服从参数为μ,σ的正态分布或高斯分布，记为X~$N(\mu,\sigma$^2).特别当 μ=0, σ=1 时称随机变量X服从标准正态分布，记为：

$$X\sim N(0,1)$$



正态分布

■ 二项分布逼近正态分布的过程



案例

考生的外语成绩近似成正态分布，平均成绩为72分，96分以上的占比2.3%，考生的外语成绩在60至84之间的概率是多少?

指数分布

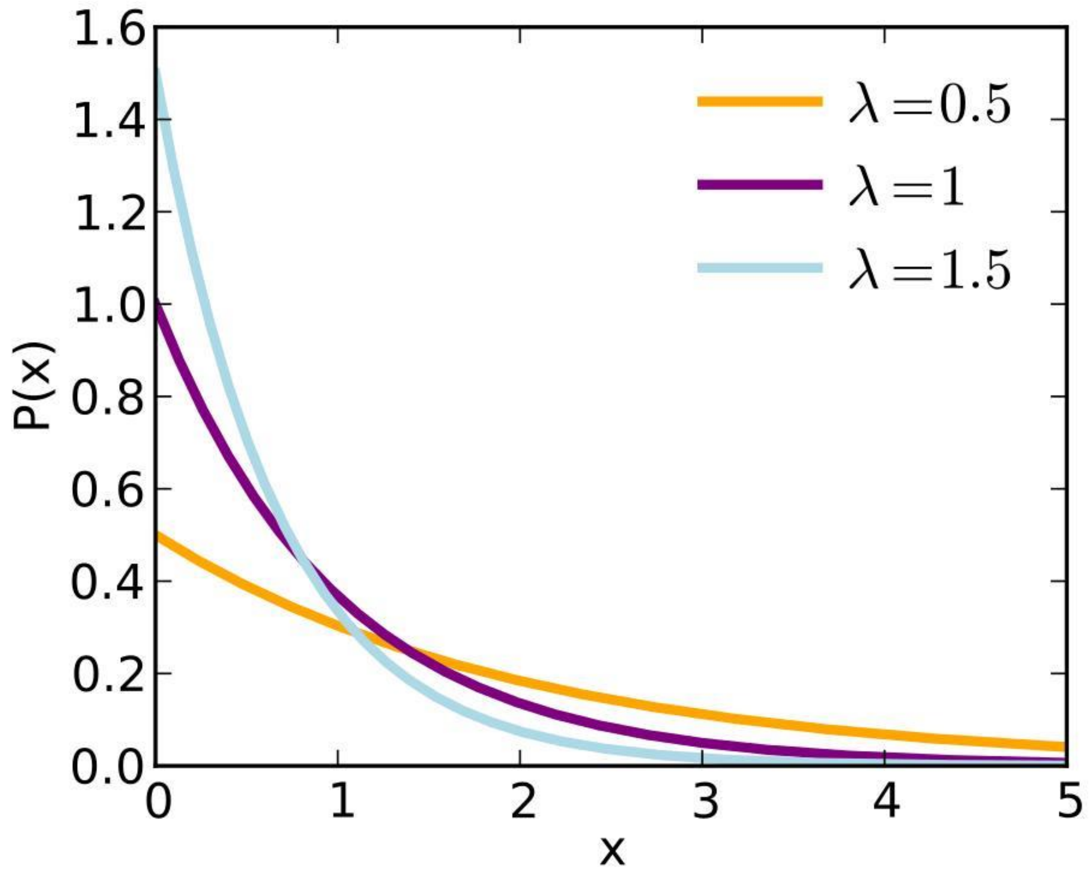
若连续型随机变量X的概率密度为：

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, x>0\\ 0 \qquad\ , x\le 0\end{cases}$$

其中$\lambda>0$为常数，表示随机事件发生一次的时间，则称X服从参数为λ的指数分布，

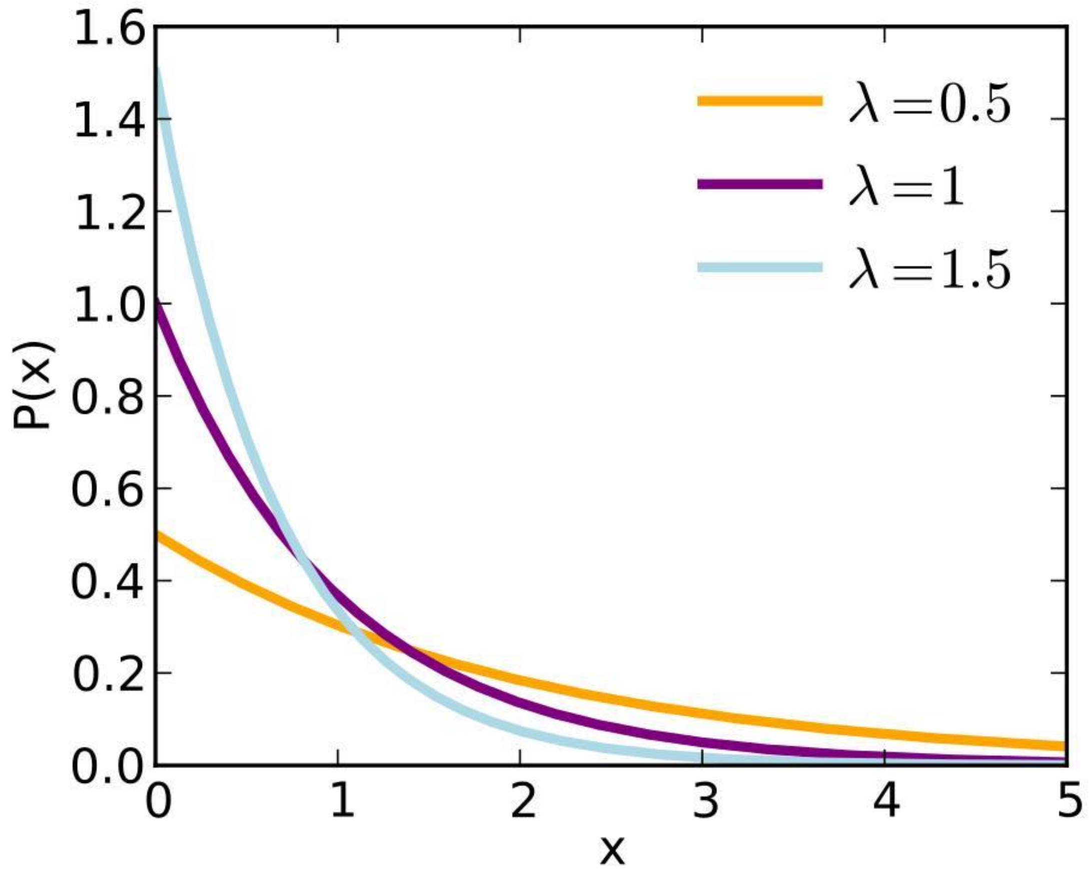
记为：

$$X\sim E(\lambda)$$

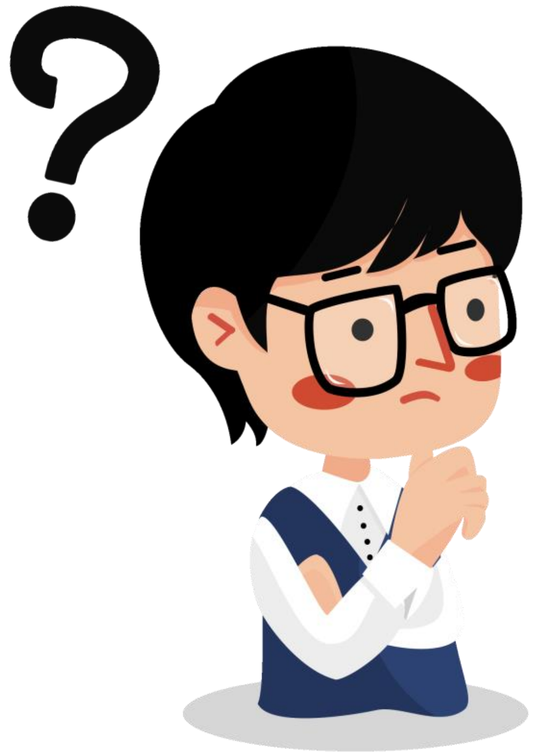


案例

如果下一个婴儿要间隔时间 t，就等同于 t 之内没有任何婴儿出生，接下来15分钟婴儿出生的概率是多少



思考题



01 生活中还有哪些正态分布的例子?

03

数学期望

定义

数学期望是试验中每次可能结果的概率乘以其结果的总和，是最基本的数学特征之一

它反映随机变量平均取值的大小，简称期望，也称均值，记为$E(X)$， X为随机变量

离散型随机变量的数学期望：

$$E(X)=\sum_{k=1}^{\infty}x_k p_k$$

连续型随机变量的数学期望：

$$E(X)=\int_{-\infty}^{\infty}xf(x)dx$$

期望和平均数的区别

平均数：

实验后根据实际结果统计得到的样本的平均值

期望：

实验前根据概率分布‘预测’的样本的平均值

案例

某医院当新生儿诞生时，医生要根据婴儿的皮肤颜色、肌肉弹性、反应的敏感性、心脏的搏动等方面的情况进行评分，新生儿的得分X是一个随机变量。据以往的资料表明X的分布律为：

X	0	1	2	3	4	5	6	7	8	9	10
p_k	0.002	0.001	0.002	0.005	0.02	0.04	0.18	0.37	0.25	0.12	0.01

试求X的数学期望$E(X)$

04

方差

概念

方差是随机变量和其数学期望之间偏离程度的度量

设X为一个随机变量，若$E\{[X-E(X)]^2\}$存在，则称$E\{[X-E(X)]^2\}$为X的方差，记为$D(X)$或$Var(X)$，即：

$$D(X)=Var(X)=E\{[X-E(X)]^2\}$$

而$\sqrt{D(X)}$记为$\sigma(X)$，称为标准差或均方差

离散型随机变量的方差：

$$D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2 p_k$$

连续型随机变量的方差：

$$D(X)=\int_{-\infty}^{\infty}[x_k-E(X)]^2 f(x)dx$$

常见数据分布的期望和方差

分布	期望	方差
伯努利分布	$E(X) = p$	$D(X) = p(1-p)$
二项分布	$E(X) = np$	$D(X) = np(1-p)$
泊松分布	$E(X) = \lambda$	$D(X) = \lambda$
指数分布	$E(X) = \frac{1}{\lambda}$	$D(X) = \frac{1}{\lambda^2}$
正态分布	$E(X) = \mu$	$D(X) = \sigma^2$
标准正态分布	$E(X) = 0$	$D(X) = 1$

05

协方差和相关系数

定义

协方差

在某种意义上给出了两个变量线性相关性的强度以及这些变量的尺度

$$Cov(X, Y) = E\big(X - E(X)\big)E\big(Y - E(Y)\big)$$

相关系数

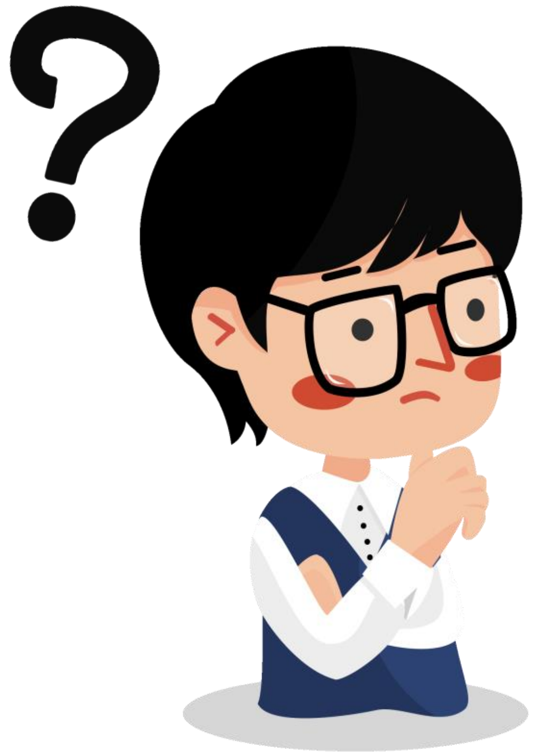
又叫线性相关系数，用来度量两个变量间的线性关系

$$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

06

总结

思考题



01

在证券投资组合选择模型中，随机变量的数字特征都有哪些作用?

总结

本章包含五小节内容：

第一、二节

· 离散型和连续型随机变量分布

第三节

· 数学期望的概念以及和平均数的差别

第四节

· 常见分布的方差和期望

第五节

· 协方差的概念和计算方法

谢谢观看

参考书目：概率论与数理统计·第四版（浙江大学） 高等教育出版社