随机变量及其分布

随机试验

随机试验是概率论的一个基本概念。 概括地讲，在概率论中把符合下面三个特点的试验叫做随机试验：

1. 可以在相同的条件下重复的进行。
2. 每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果。
3. 进行一次试验之前不能确定哪一个结果会出现。

随机试验通常用 E（event）表示：

E： 抛1颗骰子，观察出现的点数情况。

样本空间

随机试验 E 的所有可能结果组成的集合成为 E 的**样本空间**，记为 S。

样本空间的元素，即E的每个结果，成为**样本点**。

S= {点1，点2，点3，点4，点5，点6}

点 1 到点 6 均为样本点。

随机事件

随机试验 E 的样本空间 S 的子集为 E 的**随机事件**，在每次试验中，当且仅当这一子集中的一个样本点出现时，称这一**事件发生**。由一个样本点组成的单点集，成为**基本事件**。S 本身成为**必然事件**。$\emptyset$ 不包含任何样本点，成为**不可能事件**。

A_1：抛 1 颗骰子，出现的点数大于 3。

随机变量

设随机试验的样本空间为 $S=\{e\}$，$X=X(e)$ 是定义在样本空间上的实值单值函数，则称 X 为随机变量。一般以大写字母 X，Y，Z 等表示随机变量，而以小写字母 x，y，z 等表示实数。随机变量的取值随试验的结果而定，在试验之前不能预知取值，且它的取值有一定的概率。

随机变量有**离散型**和**连续型**：

离散型随机变量有其**分布律**；

连续型随机变量可以满足一定**分布**。

X 为点数对应的数字：$X=1,2,3,4,5,6$，离散型

$A1$：$X>3$

$P(X>3)=\frac{1}{2}$

离散型随机变量分布律

0–1分布

伯努利分布（英语：Bernoulli distribution，又名**两点分布**或者**0–1分布**，是一个离散型概率分布，为纪念瑞士科学家雅各布·伯努利而命名）. 若伯努利试验成功，则伯努利随机变量取值为1. 若伯努利试验失败，则伯努利随机变量取值为0. 记其成功概率为$p(0\le p\le 1)$，失败概率为 $q=1-p$.

- 其概率密度函数为:

$$f_X(x)=p^x(1-p)^{1-x}=\begin{cases} p & \text{if } x=1, \\ q & \text{if } x=0. \end{cases}$$

- 其期望值为:

$$\mathrm{E}[X]=\sum_{i=0}^{1} x_i f_X(x)=0+p=p$$

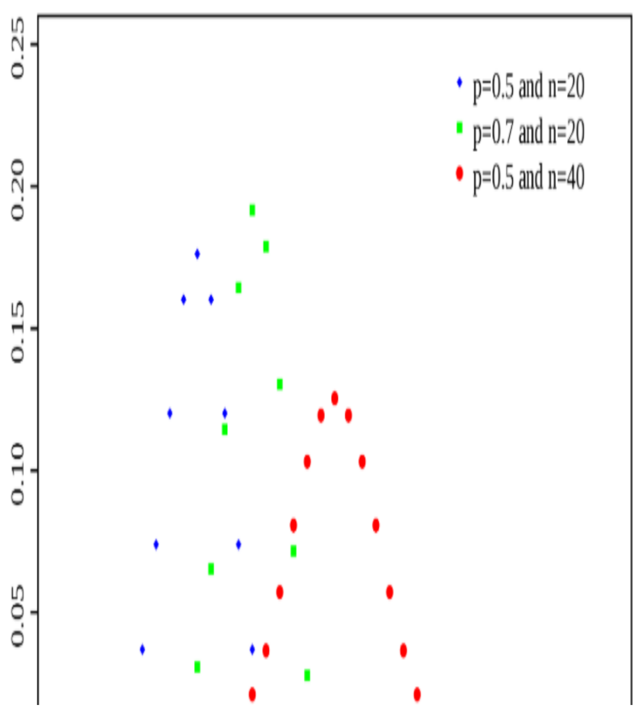
- 其方差为:

$$\mathrm{var}[X]=\sum_{i=0}^{1}(x_i-E[X])^2 f_X(x)=(0-p)^2(1-p)+(1-p)^2p=p(1-p)=pq$$

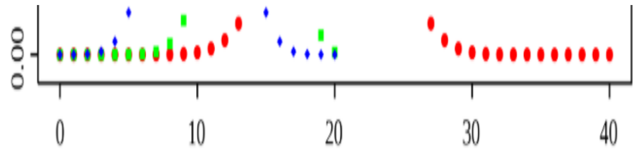
二项分布

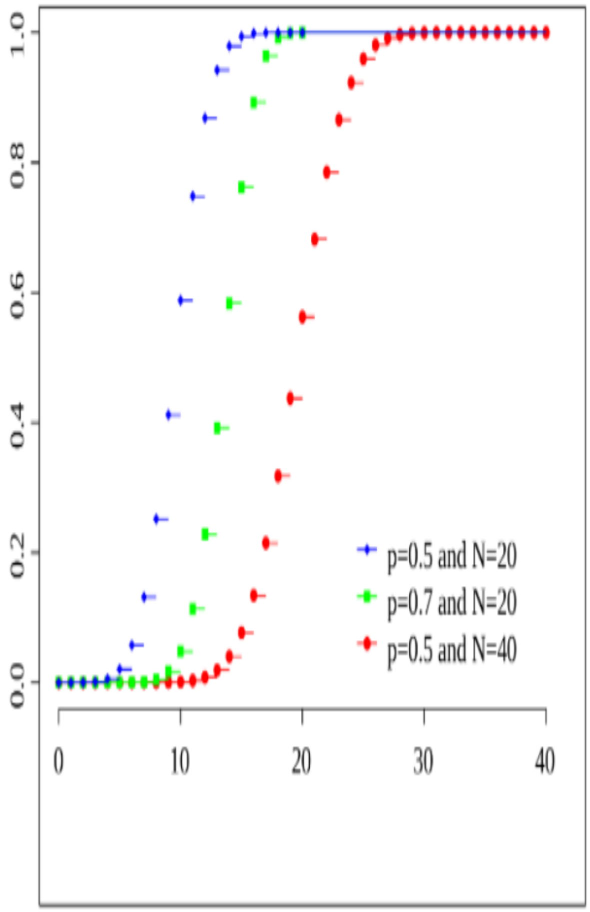
在概率论和统计学中，二项分布（英语：Binomial distribution）是n个独立的是/非试验中成功的次数的离散概率分布，其中每次试验的成功概率为p。这样的单次成功/失败试验又称为伯努利试验。实际上，当n = 1时，二项分布就是伯努利分布。二项分布是显著性差异的二项试验的基础。

下图左边两张图分别为**概率密度函数**与**累积分布函数**.



参数	$n\ge 0$ 试验次数（整数） $0\le p\le 1$ 成功概率（实数）
支撑集	$k\in\{0,\dots,n\}$
概率質量函数	$\binom{n}{k}p^k(1-p)^{n-k}$
累积分布函数	$I_{1-p}(n-\lfloor k\rfloor, 1+\lfloor k\rfloor)$
期望值	np





中位数	$\{\lfloor np\rfloor, \lceil (n+1)p\rceil\}$之一
众数	$\lfloor (n+1)p\rfloor$或$\lfloor (n+1)p\rfloor - 1$
方差	$np(1-p)$
偏度	$\frac{1-2p}{\sqrt{np(1-p)}}$
峰度	$\frac{1-6p(1-p)}{np(1-p)}$
信息熵	$\frac{1}{2}\ln(2\pi nep(1-p)) + O\left(\frac{1}{n}\right)$
动差生成函数	$(1-p+pe^t)^n$
特性函数	$(1-p+pe^{it})^n$

几何分布

在概率论和统计学中，几何分布（英语：Geometric distribution）指的是以下两种离散型概率分布中的一种：

在伯努利试验中，得到一次成功所需要的试验次数 X。X的值域是{ 1, 2, 3, ... }
在得到第一次成功之前所经历的失败次数Y = X − 1。Y的值域是{ 0, 1, 2, 3, ... }
实际使用中指的是哪一个取决于惯例和使用方便。

这两种分布不应该混淆。前一种形式（X的分布）经常被称作shifted geometric distribution；但是，为了避免歧义，最好明确地说明取值范围。

如果每次试验的成功概率是p，那么k次试验中，第k次才得到成功的概率是，

$\Pr(X=k)=(1-p)^{k-1}\,p$ $\Pr(X=k)=(1-p)^{k-1}\,p$

其中 k = 1, 2, 3, ...

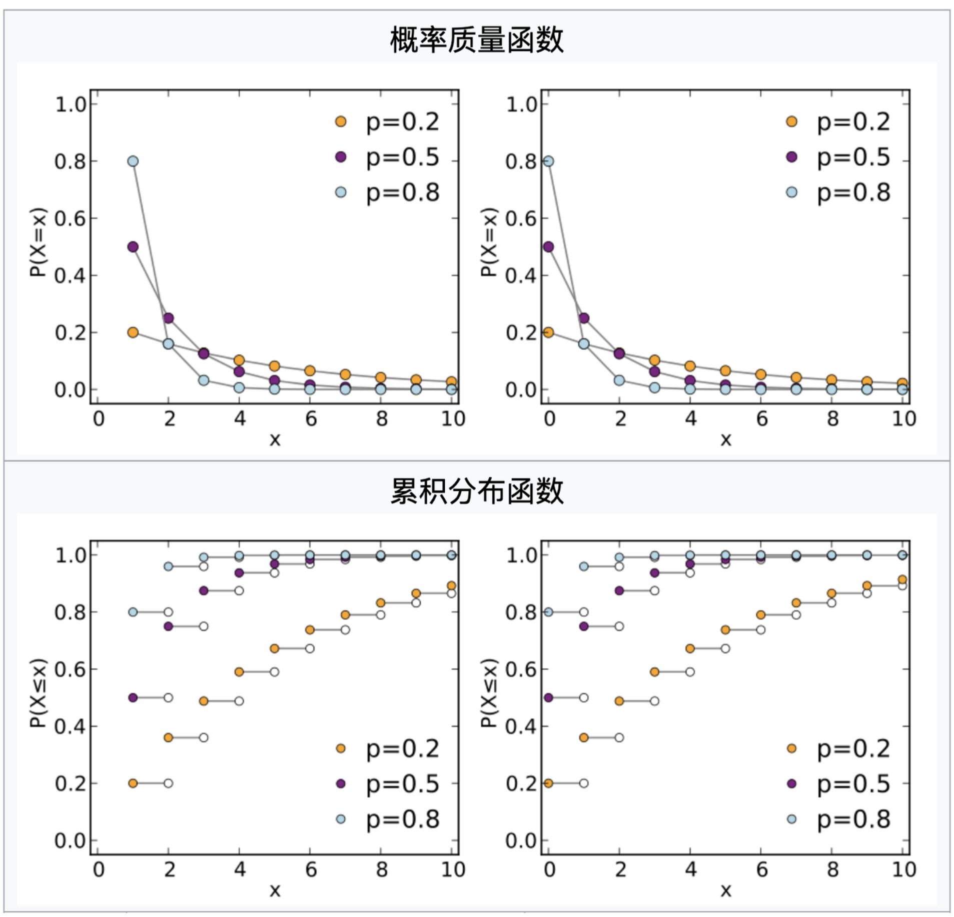
上式描述的是取得一次成功所需要的试验次数。而另一种形式，也就是第一次成功之前所失败的次数，可以写为，

$\Pr(Y=k)=(1-p)^{k}\,p$ $\Pr(Y=k)=(1-p)^{k}\,p$

其中k = 0, 1, 2, 3, ...

两种情况产生的序列都是几何数列。

比如，假设不停地掷骰子，直到得到1。投掷次数是随机分布的，取值范围是无穷集合{ 1, 2, 3, ... }，并且是一个p = 1/6的几何分布。



参数	$0 < p \leq 1$成功概率（实）	$0 < p \leq 1$成功概率（实）
支撑集	$k \in \{1, 2, 3, \dots\}$	$k \in \{0, 1, 2, 3, \dots\}$
概率质量函数 (pmf)	$(1-p)^{k-1}\, p$	$(1-p)^k\, p$
累积分布函数 (cdf)	$1-(1-p)^k$	$1-(1-p)^{k+1}$
期望值	$\frac{1}{p}$	$\frac{1-p}{p}$
中位数	$\left\lceil \frac{-1}{\log_2(1-p)} \right\rceil$（如果$-1/\log_2(1-p)$是整数，则中位数不唯一）	$\left\lceil \frac{-1}{\log_2(1-p)} \right\rceil - 1$（如果$-1/\log_2(1-p)$是整数，则中位数不唯一）
众数	1	0
方差	$\frac{1-p}{p^2}$	$\frac{1-p}{p^2}$
偏度	$\frac{2-p}{\sqrt{1-p}}$	$\frac{2-p}{\sqrt{1-p}}$
超值峰度	$6+\frac{p^2}{1-p}$	$6+\frac{p^2}{1-p}$
熵	$\frac{-(1-p)\log_2(1-p)-p\log_2 p}{p}$	$\frac{-(1-p)\log_2(1-p)-p\log_2 p}{p}$
动差生成函数 (mgf)	$\frac{pe^t}{1-(1-p)e^t}$, for $t < -\ln(1-p)$	$\frac{p}{1-(1-p)e^t}$
特征函数	$\frac{pe^{it}}{1-(1-p)\,e^{it}}$	$\frac{p}{1-(1-p)\,e^{it}}$

超几何分布

超几何分布是统计学上一种离散概率分布。它描述了由有限个物件中抽出 n 个物件，成功抽出指定种类的物件的个数（不归还 （without replacement））。

例如在有 N 个样本，其中 K 个是不及格的。超几何分布描述了在该 N 个样本中抽出 n 个，其中 k 个是不及格的几率：

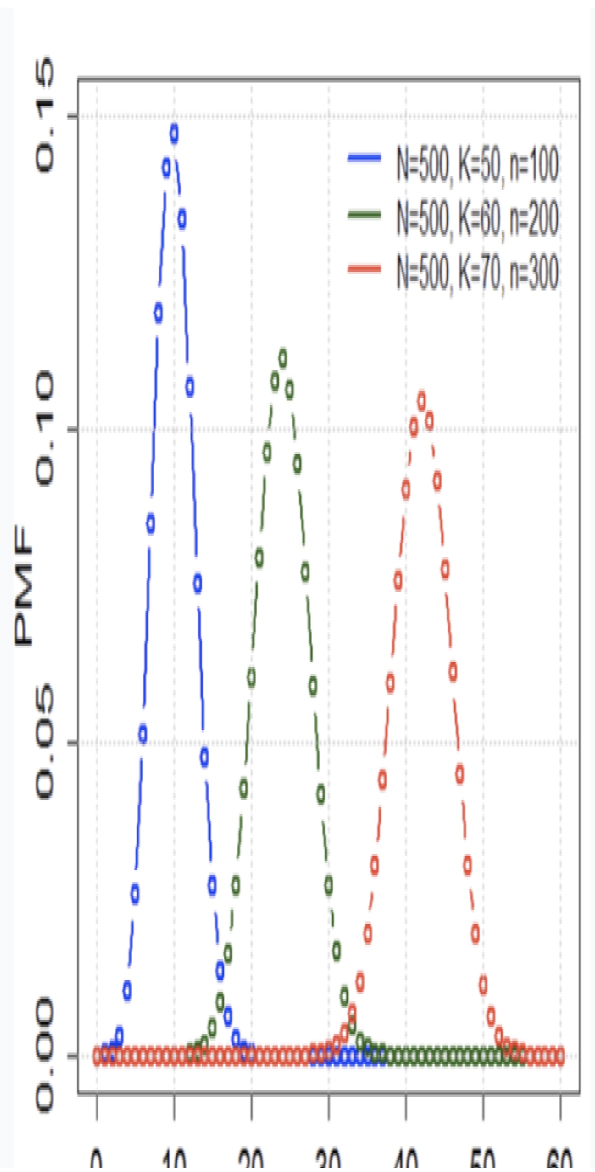
$$f(k;n,K,N)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

上式可如此理解：$\binom{N}{n}$ 表示所有在 N 个样本中抽出 n 个的方法数目。$\binom{K}{k}$ 表示在 K 个样本中，抽出 k 个的方法数目，即组合数，又称二项式系数。剩下来的样本都是及格的，而及格的样本有 $N-K$ 个，剩下的抽法便有 $\binom{N-K}{n-k}$ 种。

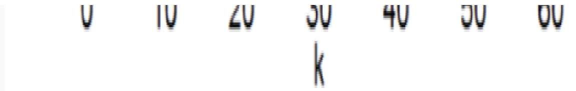
若 n=1，超几何分布还原为伯努利分布。

记号

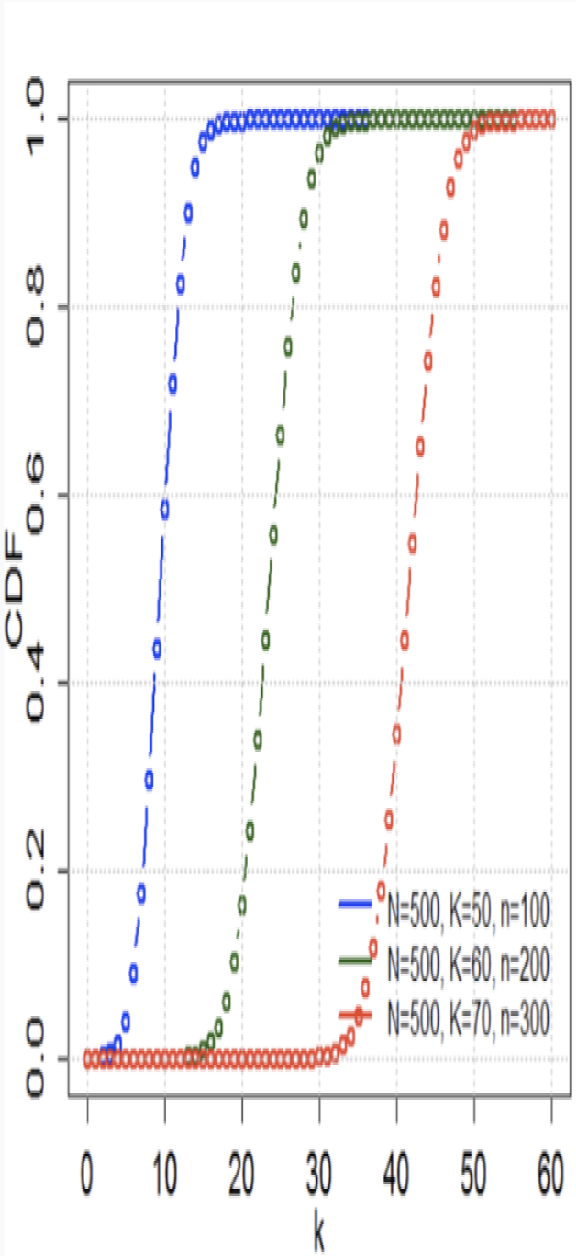
若随机变量 X 服从参数为 n,K,N 的超几何分布，则记为 $X\sim H(n,K,N)$。



参数	$N\in\{0,1,2,\dots\}$ $K\in\{0,1,2,\dots,N\}$ $n\in\{0,1,2,\dots,N\}$
支撑集	$k\in\{\max(0,n+K-N),\dots,\min(n,K)\}$
概率質量函数	$\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$
累积分布函数	$1-\frac{\binom{n}{k+1}\binom{N-n}{K-k-1}}{\binom{N}{K}}\,{}_3F_2\left[\begin{matrix}1,\,k+1-K,\,k+1-n\\k+2,\,N+k+2-K-n\end{matrix};1\right]$ 其中 ${}_pF_q$ 为广义超几何函数
期望值	$n\frac{K}{N}$



概率质量函数



累积分布函数

众数	$\left\lceil \frac{(n+1)(K+1)}{N+2} \right\rceil - 1, \left\lfloor \frac{(n+1)(K+1)}{N+2} \right\rfloor$
方差	$n\frac{K}{N}\frac{(N-K)}{N}\frac{N-n}{N-1}$
偏度	$\frac{(N-2K)(N-1)^{\frac{1}{2}}(N-2n)}{[nK(N-K)(N-n)]^{\frac{1}{2}}(N-2)}$
峰度	$\frac{1}{nK(N-K)(N-n)(N-2)(N-3)} \cdot \left[(N-1)N^2\Big(N(N+1)-6K(N-K)-6n(N-n)\Big)+ +6nK(N-K)(N-n)(5N-6)\right]$
动差生成函数	$\frac{\binom{N-K}{n}{}_2F_1(-n,-K;N-K-n+1;e^t)}{\binom{N}{n}}$
特性函数	$\frac{\binom{N-K}{n}{}_2F_1(-n,-K;N-K-n+1;e^{it})}{\binom{N}{n}}$

泊松分布

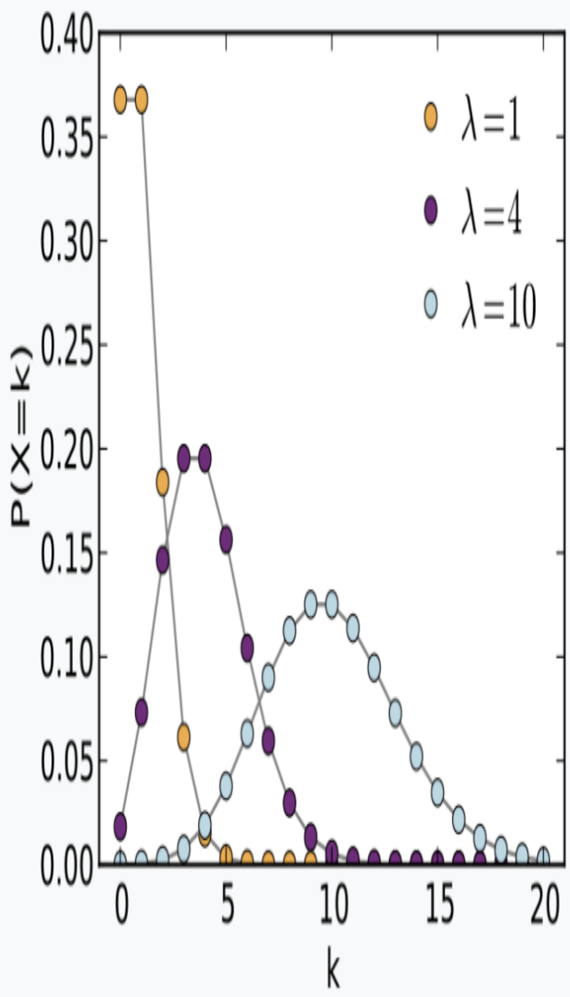
泊松分布（法语：loi de Poisson，英语：Poisson distribution）又称帕松分布、普阿松分布、布瓦松分布、布阿松分布、波以松分布、卜氏分配、泊松小数法则（Poisson law of small numbers），是一种统计与概率学里常见到的离散概率分布，由法国数学家西莫恩·德尼·泊松在1838年时发表。

泊松分布适合于描述单位时间内随机事件发生的次数的概率分布。如某一服务设施在一定时间内受到的服务请求的次数，电话交换机接到呼叫的次数、汽车站台的候客人数、机器出现的故障数、自然灾害发生的次数、DNA序列的变异数、放射性原子核的衰变数、激光的光子数分布等等。

泊松分布的概率质量函数为：

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

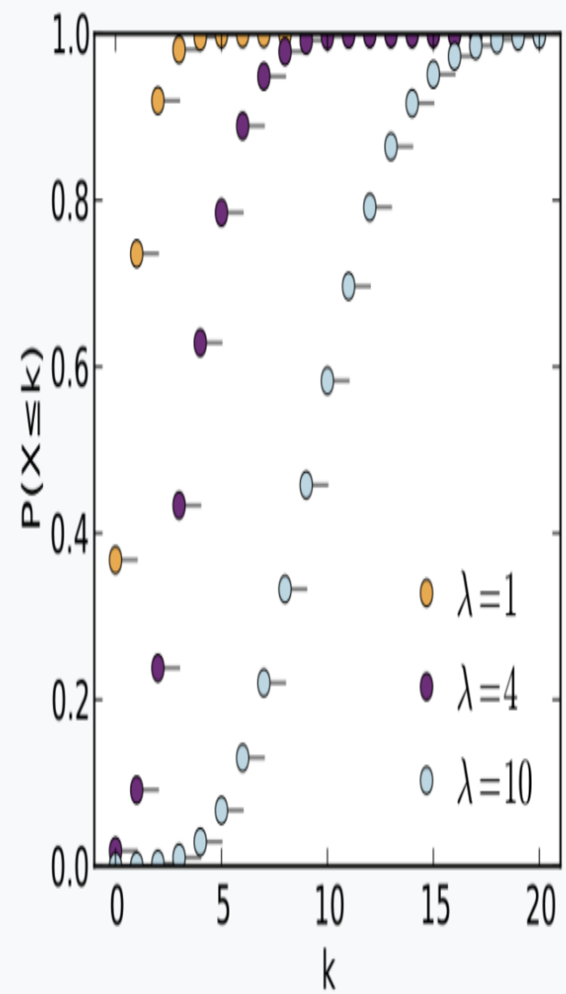
泊松分布的参数λ是单位时间（或单位面积）内随机事件的平均发生率。



横轴是索引k，发生次数。该函数只定义在k为整数的时候。连接线是只为了指导视觉。

概率质量函数

参数	$\lambda>0$ (实数)
支撑集	$k\in\{0,1,2,3,\dots\}$
概率质量函数	$\frac{\lambda^k}{k!}e^{-\lambda}$
累积分布函数	$\frac{\Gamma(\lfloor k+1\rfloor,\lambda)}{\lfloor k\rfloor!}$，或$e^{-\lambda}\sum_{i=0}^{\lfloor k\rfloor}\frac{\lambda^i}{i!}$，或 $Q(\lfloor k+1\rfloor,\lambda)$ (对于$k\geq 0$，其中$\Gamma(x,y)$是不完全Γ函数，$\lfloor k\rfloor$是高斯符号，Q是规则化Γ函数)
期望值	λ
中位数	$\approx\lfloor\lambda+1/3-0.02/\lambda\rfloor$
众数	$\lceil\lambda\rceil-1, \lfloor\lambda\rfloor$



横轴是索引k，发生次数。CDF在整数k处不连续，且在其他任何地方都是水平的，因为服从泊松分布的变量只针对整数值。

累积分布函数

方差	λ
偏度	$\lambda^{-1/2}$
峰度	λ^{-1}
信息熵	$\lambda[1-\log(\lambda)]+e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k\log(k!)}{k!}$ (for large λ) $\frac{1}{2}\log(2\pi e\lambda)-\frac{1}{12\lambda}-\frac{1}{24\lambda^2}-\frac{19}{360\lambda^3}+O\left(\frac{1}{\lambda^4}\right)$
动差生成函数	$\exp(\lambda(e^t-1))$
特性函数	$\exp(\lambda(e^{it}-1))$

连续型随机变量的分布

分布函数、概率密度函数、随机变量的函数的分布

均匀分布

连续型均匀分布，如果连续型随机变量 X 具有如下的概率密度函数，则称 X 服从 $[a,b]$ 上的均匀分布（uniform distribution）,记作 $X\sim U[a,b]$

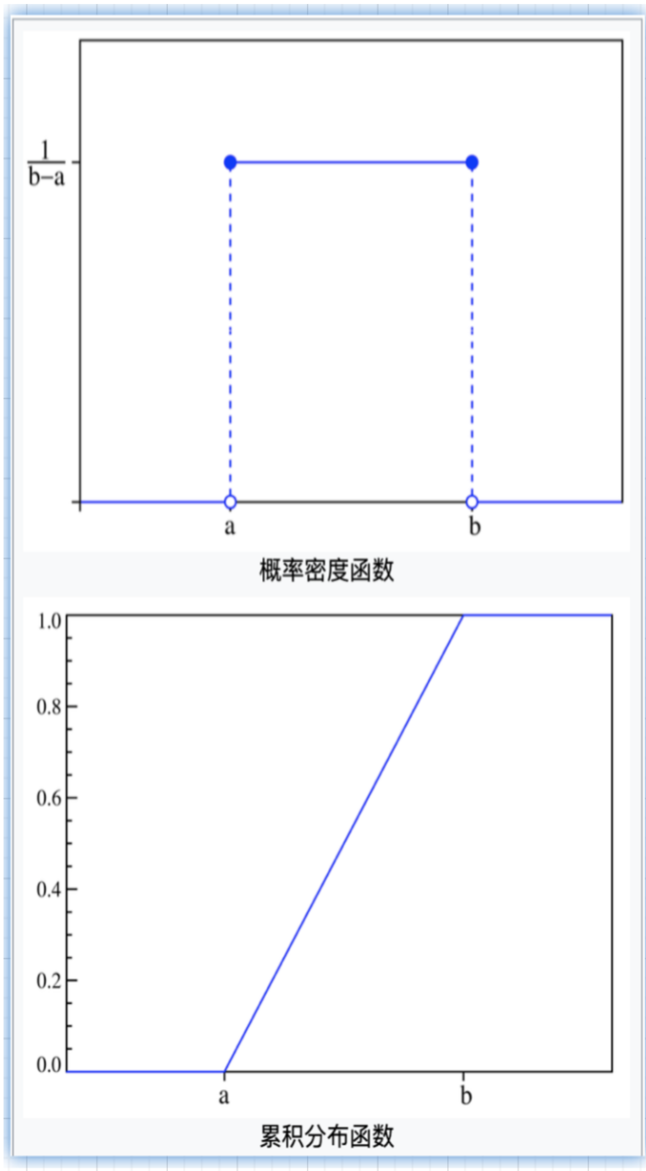
一个均匀分布在区间 [a,b] 上的连续型随机变量 X 可给出如下函数：

概率密度函数：

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

累积分布函数：

$$F(x)=\begin{cases}0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \le x < b \\ 1 & \text{for } x \ge b\end{cases}$$



参数	$a, b \in (-\infty, \infty)$
支撑集	$a \le x \le b$
概率密度函数	$\frac{1}{b-a}$ for $a \le x \le b$ 0 for $x < a$ or $x > b$
累积分布函数	0 for $x < a$ $\frac{x-a}{b-a}$ for $a \le x < b$ 1 for $x \ge b$
期望值	$\frac{a+b}{2}$
中位数	$\frac{a+b}{2}$
众数	任何$[a,b]$内的值
方差	$\frac{(b-a)^2}{12}$
偏度	0
峰度	$-\frac{6}{5}$
信息熵	$\ln(b-a)$
动差生成函数	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
特性函数	$\frac{e^{itb}-e^{ita}}{it(b-a)}$

指数分布

在概率论和统计学中，指数分布（英语：Exponential distribution）是一种连续概率分布。指数分布可以用来表示独立随机事件发生的时间间隔，比如旅客进入机场的时间间隔、打进客服中心电话的时间间隔、中文维基百科新条目出现的时间间隔等等。

概率密度函数

一个指数分布的概率密度函数是：

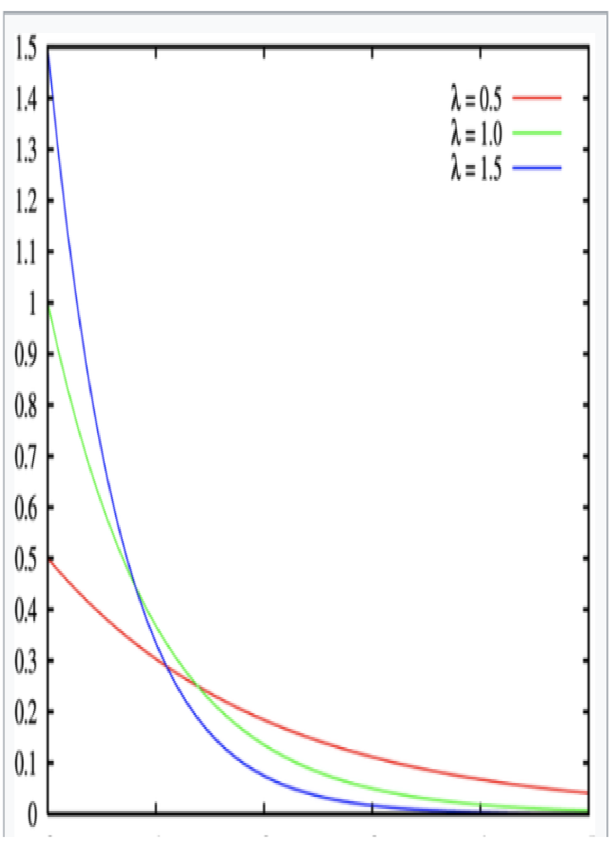
$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & ,\ x \geq 0, \\ 0 & ,\ x < 0. \end{cases}$$

其中λ > 0是分布的一个参数，常被称为率参数（rate parameter）。即每单位时间发生该事件的次数。指数分布的区间是[0,∞)。 如果一个随机变量X 呈指数分布，则可以写作：X ~ Exponential（λ）。

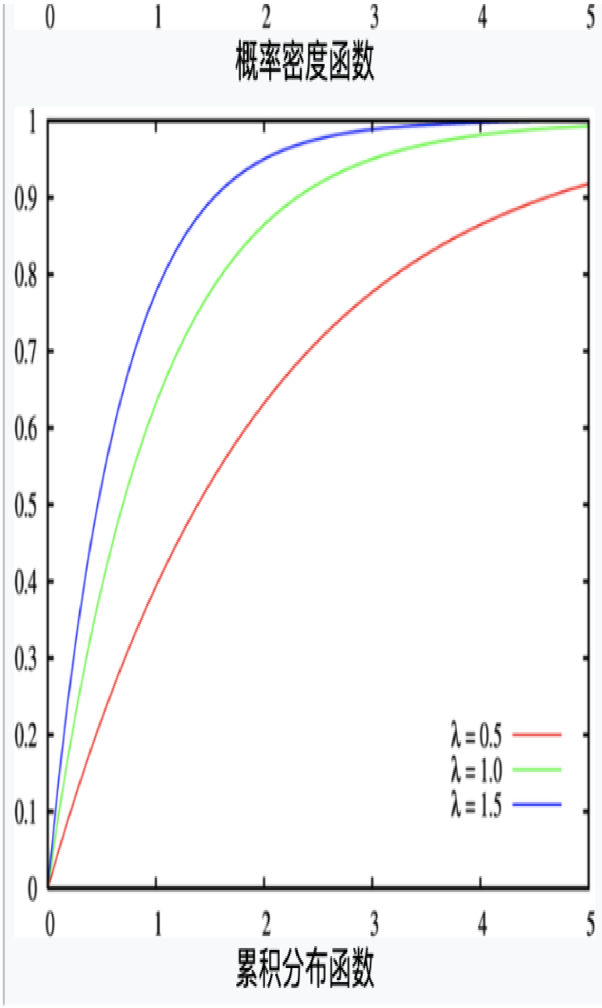
累积分布函数

累积分布函数可以写成：

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & ,\ x \geq 0, \\ 0 & ,\ x < 0. \end{cases}$$



参数	$\lambda > 0$ 率
支撑集	$x \in [0; \infty)$
概率密度函数	$\lambda e^{-\lambda x}$
累积分布函数	$1 - e^{-\lambda x}$
期望值	λ^{-1}
中位数	$\ln(2)/\lambda$
众数	0



方差	λ^{-2}
偏度	2
峰度	6
信息熵	$1 - \ln(\lambda)$
动差生成函数	$\left(1 - \frac{t}{\lambda}\right)^{-1}$
特性函数	$\left(1 - \frac{it}{\lambda}\right)^{-1}$

正态分布

正态分布（台湾作常态分布，英语：normal distribution）又名高斯分布（英语：Gaussian distribution），是一个非常常见的连续概率分布。正态分布在统计学上十分重要，经常用在自然和社会科学来代表一个不明的随机变量。[1][2]

若随机变量 X 服从一个位置参数为 μ 、尺度参数为 σ 的正态分布，记为：

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

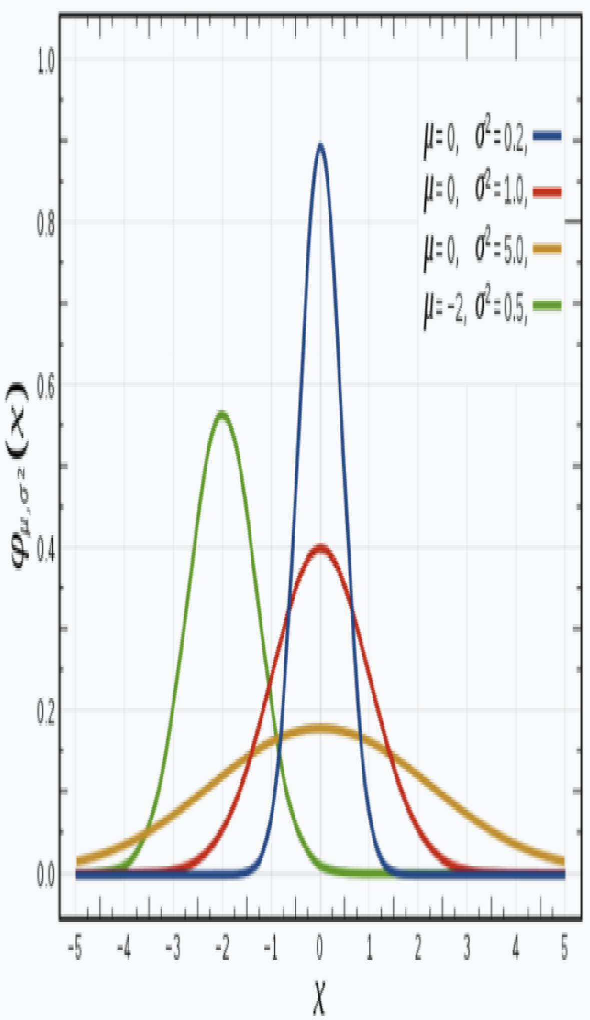
则其概率密度函数为:

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

正态分布的数学期望值或期望值 μ 等于位置参数，决定了分布的位置；其方差 σ^2 的开平方或标准差 σ

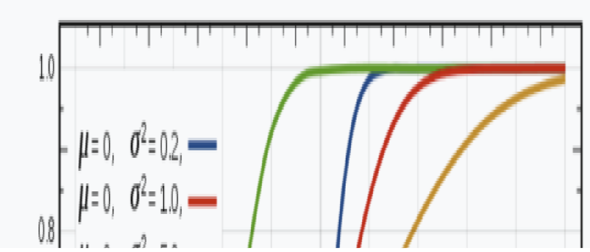
正态分布的概率密度函数曲线呈钟形，因此人们又经常称之为钟形曲线（类似于寺庙里的大钟，因此得名）。我们通常所说的标准正态分布是位置参数 $\mu = 0$，尺度参数 $\sigma^2 = 1$ 的正态分布（见下图中红色曲线）.

正态分布

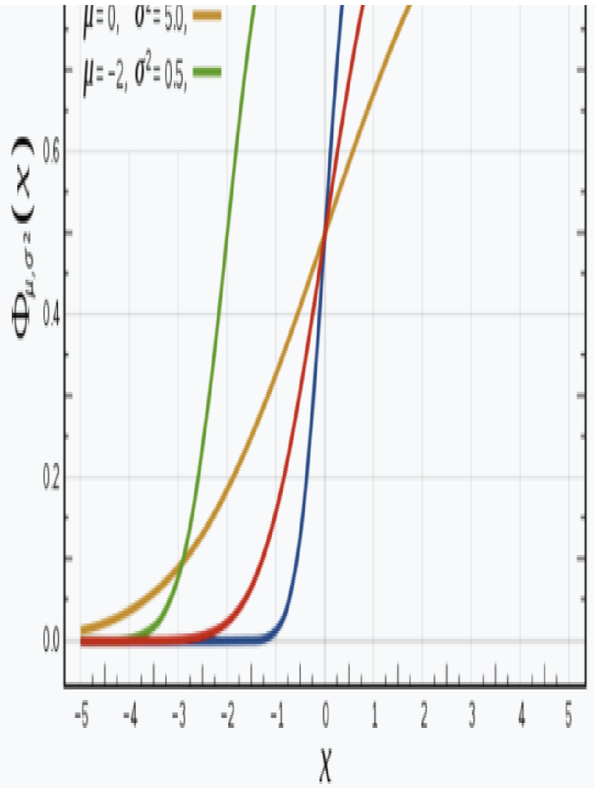


红线代表标准正态分布

概率密度函数



参数	μ 数学期望（实数） $\sigma^2 > 0$ 方差（实数）
支撑集	$x \in (-\infty; +\infty)$
概率密度函数	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
累积分布函数	$\frac{1}{2}\left(1 + \operatorname{erf}\frac{x-\mu}{\sigma\sqrt{2}}\right)$
期望值	μ
中位数	μ
众数	μ
方差	σ^2



颜色与概率密度函数同

累积分布函数

偏度	0
峰度	0
信息熵	$\ln\left(\sigma\sqrt{2\pi e}\right)$
动差生成函数	$M_X(t) = \exp\left(\mu t + \sigma^2 \frac{t^2}{2}\right)$
特性函数	$\phi_X(t) = \exp\left(\mu i t - \frac{\sigma^2 t^2}{2}\right)$