第1节 随机变量及其概率分布

——贝叶斯理论基础

一、随机变量的引入

一般地，如果A为某个随机事件，则一定可以通过如下标签函数使它与数值发生联系：

$$I_A = \begin{cases} \mathbf{1}, & \text{如果A发生} \\ 0, & \text{如果A不发生} \end{cases}$$

试验的结果能用一个数 x 来表示，这个数 x 是随着试验的结果的不同而变化的，也即它是样本点的一个函数，这种量以后称为随机变量。

随机变量与普通函数不同

- 随机变量取各个值有一定的概率。普通函数则不然。
- 随机变量定义在样本空间上，即随机变量的“定义域”可以不是实数值，而普通函数的定义域是实数集或它的子集。

概率分布

一般，若R是一个实数集合，将X在R上取值写成$\{X\in \mathrm{R}\}$，它表示事件$\mathrm{B}=\{\mathrm{e}\mid X(\mathrm{e})\in \mathrm{R}\}$，即B是由$X$中使得$X(\mathrm{e})\in \mathrm{R}$的所有样本点e所组成的事件，此时有$P\{X\in \mathrm{R}\}=P(\mathrm{B})=P\{\mathrm{e}\mid X(\mathrm{e})\in \mathrm{R}\}$。$P\{X\in \mathrm{R}\}$称为随机变量$X$的概率分布。

随机分布

- 离散型随机分布
  - ◆ 二项分布、几何分布、超几何分布、泊松分布
- 连续型随机分布
  - ◆ 正态分布、均匀分布、指数分布、对数正态分布、柯西分布、Gamma分布、瑞利分布、韦伯分布、三角形分布
- 三大抽样分布
  - ◆ 卡方分布、F分布、*t*分布

二、离散型随机分布

- 定义：若随机变量 X 取$x_1, x_2, \ldots, x_n, \ldots$ 值， 且取这些值的概率依次为 $p_1, p_2, \ldots, p_n, \ldots$, 则称X为离散型随机变量，而称

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

- 为X的分布律或概率分布。可表示为：

$$X\sim\ \begin{array}{c|ccccc} X & x_1 & x_2 & \cdots & x_k & \cdots \\ \hline p_k & p_1 & p_2 & \cdots & p_k & \cdots \end{array}$$

1. 二项分布(Bernoulli distribution)

$$P(X) = \binom{n}{X} \pi^X (1-\pi)^{n-X}$$

$$E[X] = np \qquad Var[X] = np(1-p)$$

应用：*n* 次试验在相同条件下进行，各个观察单位的结果相互独立，且只能具有相互对立的一种结果，当$n \to \infty$时，二项分布近似于正态分布。

2. 几何分布

定义：在第 n 次伯努利实验，才得到第一次成功的机率。更详细的说是：n次伯努利试验，前 n-1 次皆失败，第 n 次才成功的概率。

$$P(X=k)=(1-p)^{k-1}\cdot p$$

$$E[X]=\frac{1}{p} \qquad Var[X]=\frac{1-p}{p^2}$$

应用：射击比赛等。

3. 负二项分布(Negative binomial distribution)

定义：已知一个事件在伯努利试验中每次的出现概率是 p，在一连串伯努利试验中，一件事件刚好在第$r+k$次试验出现第r次的概率。

$$P(X) = \frac{\Gamma(r+k)}{k!\,\Gamma(r)} p^r \cdot (1-p)^k$$

$$E[X] = r \cdot \frac{1-p}{p} \qquad Var[X] = r \cdot \frac{1-p}{p^2}$$

当 $r = 1$时，负二项分布等于几何分布。

4. 超几何分布

定义：在产品质量的不放回抽检中，若N件产品中有M件次品，抽检n件时所得次品数$X=k$，是一个随机变量：

$$P(X=k)=\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}} \qquad k=1, 2, \ldots$$

$$E[X]=\frac{nM}{N} \qquad Var[X]=\frac{nM}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

应用：产品质量检测等。

5. 泊松分布（Poisson distribution）

应用：泊松分布适合于描述单位时间（或空间）内随机事件发生的次数。如某一服务设施在一定时间内到达的人数，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，一块产品上的缺陷陷数，显微镜下单位分区内的细菌分布数等。

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad k = 1, 2, \ldots$$

$$E[X] = \lambda \qquad Var[X] = \lambda$$

三、连续型随机变量、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 $f(x)$ 使对于任一实数 x，有

$$F(x)=\int_{-\infty}^{x} f(t)dt$$

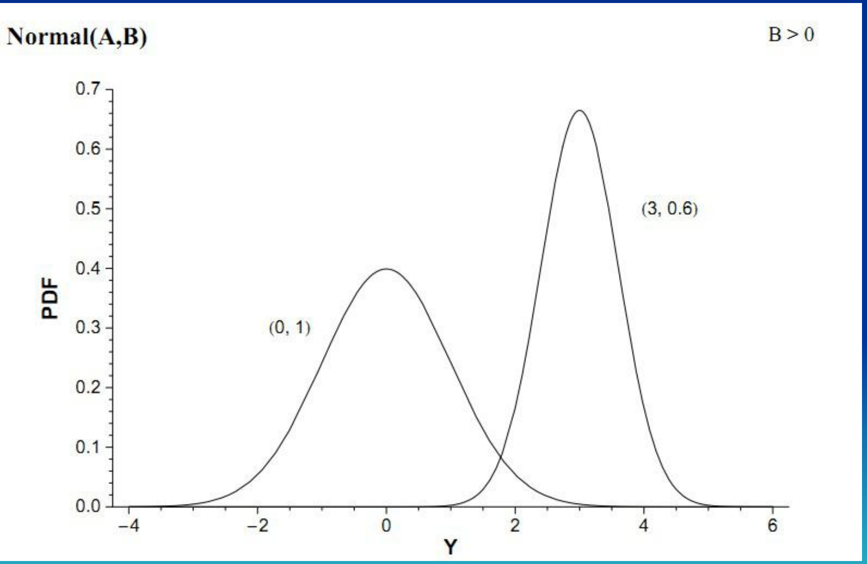
则称 X 为连续型随机变量。函数 $f(x)$ 称为随机变量 X 的**概率密度函数**。

几个常用的连续型分布

- 正态分布
- 均匀分布
- 指数分布
- 对数正态分布
- 柯西分布
- Gamma分布
- 瑞利分布
- 韦伯分布
- 三角形分布

1.正态分布 (Normal distribution)

应用：由许多微小的独立随机因素影响的变量属于正态分布。在自然现象中，大量随机变量都服从或近似服从正态分布。



设连续型随机变量X的概率密度为:

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

其中：μ ，σ（σ >0）为常数，则称X服从参数为μ, σ 的正态分布或高斯(Gauss)分布。记为：$X\sim N(\mu, \sigma^2)$。相应的分布函数为:

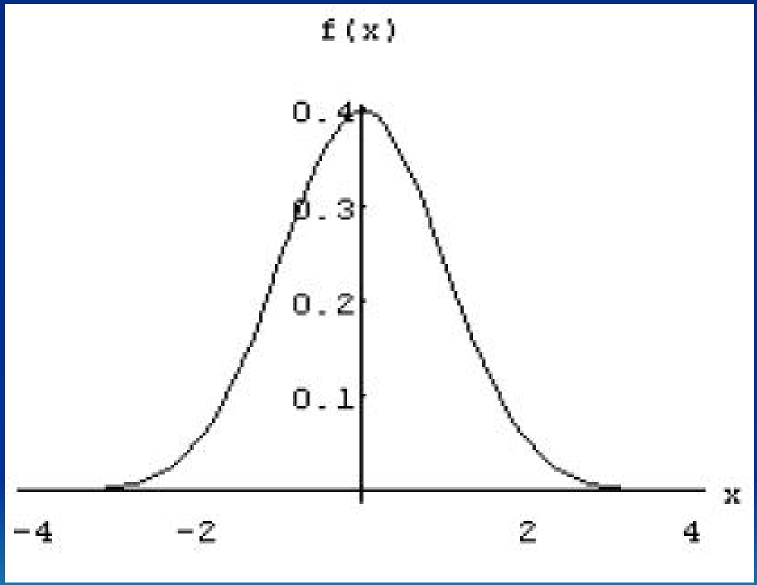
$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$

标准正态分布

当$\mu=0$，$\sigma=1$时称X服从标准正态分布，记为$X\sim N(0,1)$。其概率密度和分布函数分别为：

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$



显然 $\Phi(-x)=1-\Phi(x)$

另外，有$\Phi(x)$的函数表可查。

2. 均匀分布(Uniform distribution)

应用：当无法区分随机变量取不同值的可能性有何不同时，可以假定随机变量服从区间上的均匀分布。

概率密度函数为：$f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0, , & others \end{cases}$

分布函数为：$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$

3. 指数分布(Exponential distribution)

应用：描述独立事件发生的时间间隔。自然界中有很多种“寿命”可以用指数分布来描述，如电子元件的寿命、动物的寿命、电话的通话时间、服务系统的服务时间等。

概率密度函数为：$f(x,\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$

分布函数为：$F(x,\lambda)=\begin{cases}1-e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$

与泊松过程的关系：

- 泊松过程中，第k次随机事件与第k+1次随机事件出现的时间间隔服从指数分布。这是因为，第k次随机事件之后长度为 t 的时间段内，第k+1次随机事件出现的概率等于1减去这个时间段内没有随机事件出现的概率。而根据泊松过程的定义，长度为t 的时间段内没有随机事件出现的概率等于：

$$\frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

所以第 k 次随机事件之后长度为 t 的时间段内，第 $k+1$ 次随机事件出现的概率等于：

$$1-e^{-\lambda t}$$

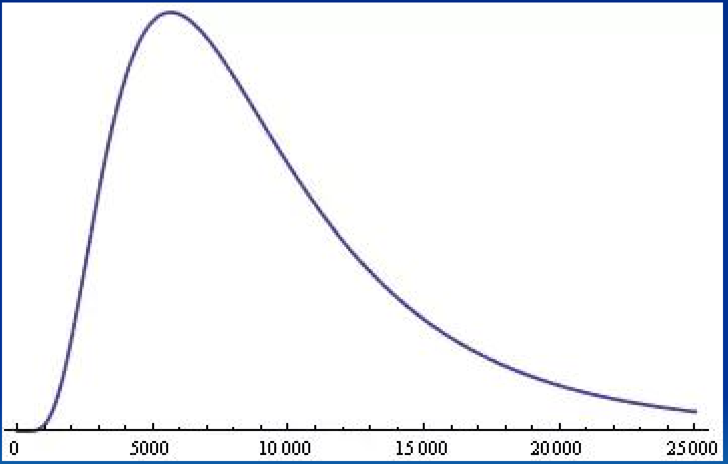
这是指数的分布函数。这还表明了泊松过程的无记忆性。

4. 对数正态分布

背景：真实世界的很多分布是不对称的。比如财富的分布，富人的有钱程度远远超出穷人的贫穷程度，即财富分布曲线有右侧的长尾。

概率密度函数为：

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



分布函数为：

$$F(x)=\frac{1}{2}+\frac{1}{2}erf\left[\frac{lnx-\mu}{\sigma\sqrt{2}}\right]$$

5. 柯西分布(Cauchy distribution)

应用：物理学中描述共振微分方程的解，描述被共振或者其他机制加宽的谱线形状。

概率密度函数为：

$$f(x)=\frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

分布函数为：

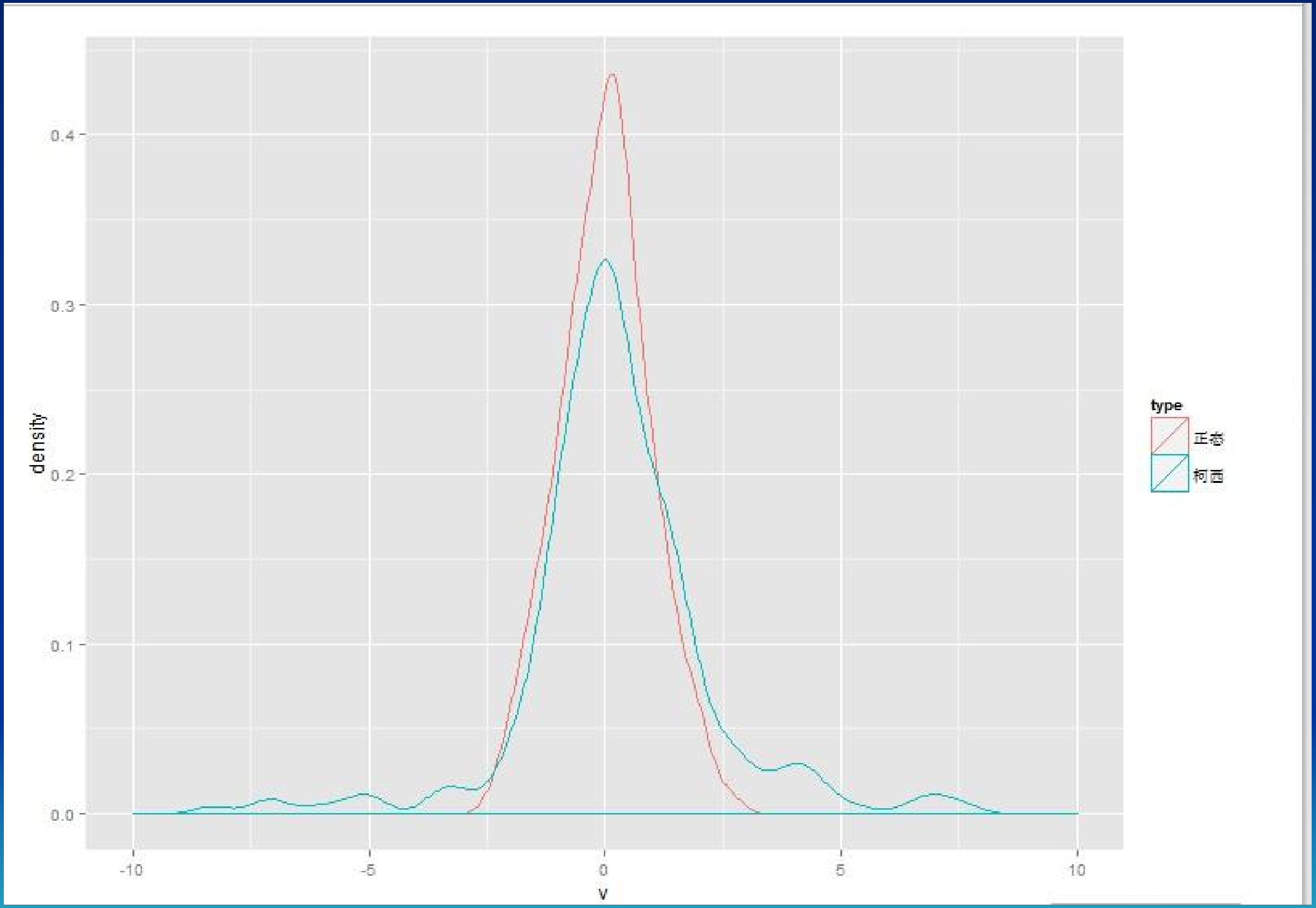
$$F(x)=\frac{1}{\pi}arctan\left(\frac{x-x_0}{\gamma}\right)+\frac{1}{2}$$

柯西分布的特性：

1) 期望和方差均不存在。

柯西分布概率密度函数的广义积分不是绝对收敛的，所以说柯西分布的数学期望不存在。至于它的方差不存在，也是基于同样的道理。

2) 图形与正态分布相似，亦为钟形。



3) 与正态分布的关系。

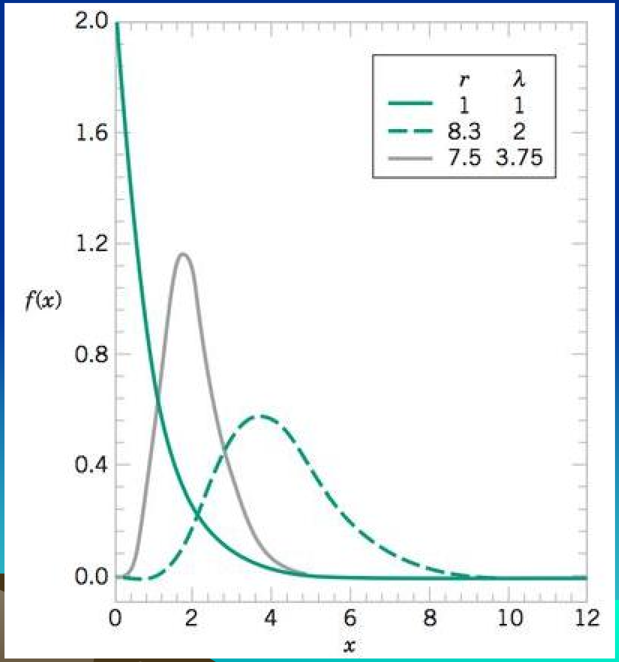
如果 随机变量U与V是两个服从期望值为 0、方差为 1 的独立正态分布，那么比值 U / V 为柯西分布。

6. Gamma分布

定义：若干个独立指数分布变量的和的分布。参数 r 称为形参，λ 称为逆尺度参数。

概率密度函数为：$f(x)=\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}$

分布函数为：$F(x)=\dfrac{\Gamma(r,\lambda x)}{\Gamma(r)}$



几种特殊的Gamma分布：

1) 指数分布。

当$r=1$时，$\Gamma(r)=\Gamma(1)=1, x^{r-1}=x^{0}=1$。

此时$\Gamma-$分布的概率密度函数为

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

这是参数为λ的指数分布的概率密度函数.

2) 爱尔朗分布。

假设有r 个串联的服务台相互独立，则一个顾客接受完r个服务台服务 所需的总时间$f(x)$就服从 r 阶的爱尔朗分布。

$$f(x)=\begin{cases}\dfrac{\lambda^r}{(r-1)!}x^{r-1}e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$$

3) χ^2-分布。

当$r=\frac{n}{2}$(n为正整数)，$\lambda=\frac{1}{2}$时，此时$Gamma-$分布的概率密度函数为：

$$f(x)=\begin{cases}\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}} & x>0\\ 0 & x\le 0\end{cases}$$

这是自由度为n的χ^2-分布的概率密度函数。

记为　$\xi\sim\chi^2(n)$

Gamma－分布的重要性质：

- **定 理:**

如果随机变量$\xi_1, \xi_2, \cdots, \xi_n$相互独立，且$\xi_i$服从参数为$\lambda$，$r_i$ 的Gamma－分布 $(i = 1,2, \cdots n)$，则$\eta = \xi_1 + \xi_2 + \cdots + \xi_n$ 服从λ，r的Gamma－分布。

其中：$r = r_1 + r_2 + \cdots + r_n$

推 论:

如果随机变量 $\xi_1, \xi_2, \cdots, \xi_n$ 相互独立，且

$\xi_i \sim \chi^2(n_i)$，$(i = 1,2,\cdots n)$，则

$$\eta = \xi_1 + \xi_2 + \cdots + \xi_n$$

- 服从自由度 $n = n_1 + n_2 + \cdots + n_n$ 的 $\chi^2 -$ 分布。

即：$$\eta = \xi_1 + \xi_2 + \cdots + \xi_n \sim \chi^2(n_1 + n_2 + \cdots + n_n)$$

7. 瑞利分布(Rayleigh distribution)

定义：当一个随机二维向量的两个分量呈独立的、有着相同的方差的正态分布时，这个向量的"模"呈瑞利分布。

概率密度函数为：

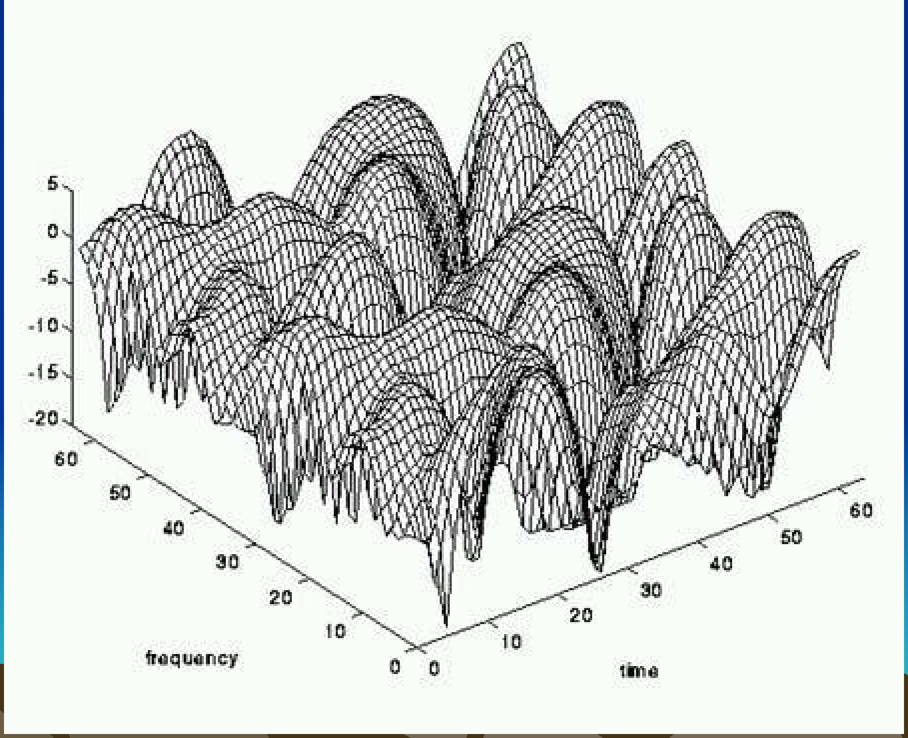
$$f(x)=\frac{x\cdot exp\left(\frac{-x^2}{2\sigma^2}\right)}{\sigma^2}$$

分布函数为：

$$F(x)=1-exp\left(\frac{-x^2}{2\sigma^2}\right)$$

应用:

瑞利分布常用于描述平坦衰落信号接收包络或独立多径分量接受包络统计时变特性。如两个正交高斯噪声信号之和的包络线服从瑞利分布。

- 

8. 韦伯分布（Weibull distribution）

定义：韦氏分布或威布尔分布，是可靠性分析和寿命检验的理论基础。

概率密度函数为：$f(x)=\begin{cases}\dfrac{k}{\lambda}\left(\dfrac{x}{\lambda}\right)^{k-1}\cdot e^{-\frac{x^k}{\lambda}} & x\geq 0\\ 0 & x<0\end{cases}$

分布函数为：$F(x)=1-e^{-(x/\lambda)^k}$

韦伯分布的参数及性质：

$\lambda > 0$是比例参数，$k > 0$是形参。它的累积分布函数是扩展的指数分布函数。

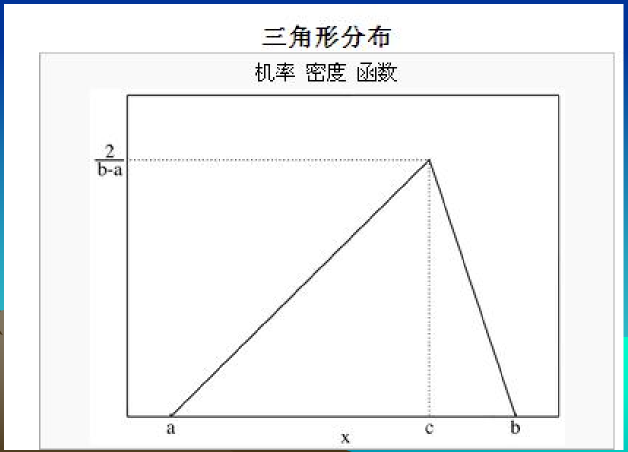
当$k = 1$时，它是 指数分布；$k = 2$时，它

- 是瑞利分布。

9. 三角分布(Triangle distribution)

定义：三角形分布是低限为 a、最高点为 c、上限为 b 的连续概率分布。

$$f(x|a,b,c)=\begin{cases}\dfrac{2(x-a)}{(b-a)(c-a)} & a\le x\le c\\[2ex] \dfrac{2(b-x)}{(b-a)(b-c)} & c\le x\le b\end{cases}$$



三角形分布的应用：

- 三角形分布经常用于商务决策，尤其是计算机模拟领域。通常，如果对结果的概率分布所知信息很少，例如仅仅知道最大值与最小值，那么可以使用平均分布模型。但是，如果已经知道了最可能出现的结果，那么就可以用三角形分布进行模拟。
- 三角形分布以及Beta分布在项目管理中大量地用作项目评估与审核技术以及关键途径的输入信息，以建立在最大值与最小值之间事件发生的概率模型。

四、三大抽样分布

1. 卡方分布

定义：若 k 个随机变量$z_1, z_2, \ldots, z_k$ 相互独立，且服从标准正态分布，则随机变量：$X^2 = \sum_{n=1}^{k} Z_n^2$ 称为自由度为 k 的卡方分布，记作：$X \sim \chi^2(k)$。

$$f_k(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-x/2}$$

应用：常用于假设检验和置信区间的计算。

2. F 分布

定义：设 X_1 服从自由度为n_1的χ^2分布，设X_2服从自由度为n_2的χ^2分布，且X_1, X_2相互独立，则称变量：

$$F = \frac{X_1/n_1}{X_2/n_2}$$

服从F 分布，自由度为(n_1, n_2)。

应用：假设检验。

F分布与正态分布的关系:

自一个正态分布中随机抽取数量为n_1及n_2的两个样本，其方差的比值服从 F 分布，分子的自由度为 n_1-1，分母的自由度为 n_2-1。

- $$F=\frac{S_{n_1-1}^2}{S_{n_2-1}^2}$$

这样，可通过分析任意两样本方差，判断它们是否服从同一正态分布。

3. t 分布

定义：设X_1服从标准正态分布$N(0,1)$，X_2服从自由度为n的χ^2分布，且X_1，X_2相互独立，则称变量：

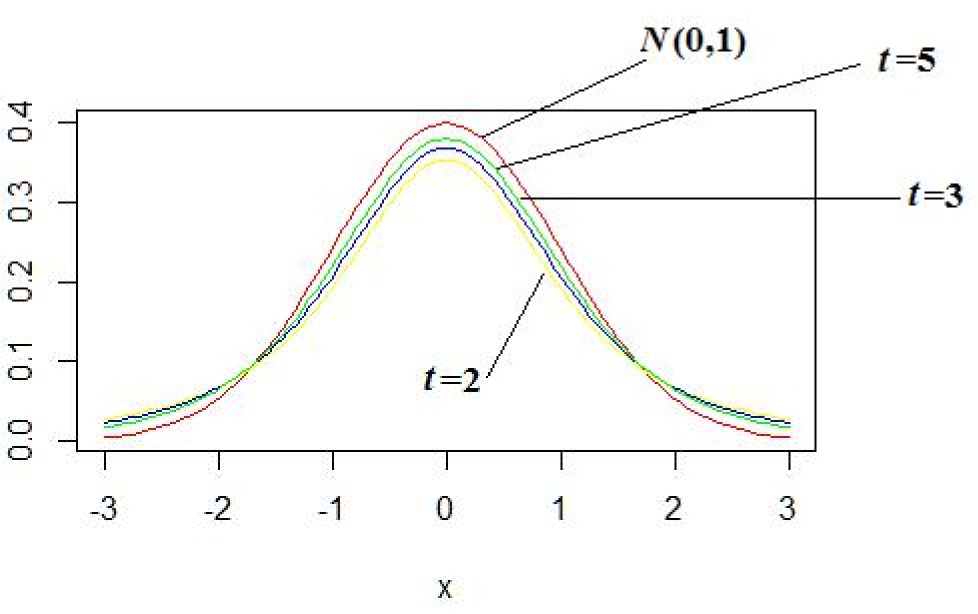
$$t=\frac{X_1}{\sqrt{X_2/n}}$$

服从自由度为 n 的 t 分布。

应用：假设检验。

t分布与正态分布的关系:

t 分布曲线形态与自由度n大小有关。与标准正态分布曲线相比，自由度n越小，t 分布曲线愈平坦，曲线中间愈低，曲线双侧尾部翘得愈高；自由度n愈大，t 分布曲线愈接近正态分布曲线，当自由度$n \to \infty$ 时，t 分布曲线变为标准正态分布曲线。

- 

五、条件概率与独立性原理

定理：设A，B是两个事件，且$\mathrm{P}(B) > 0$，称：

$$P(A|B) = \frac{P(AB)}{P(B)}$$

为事件 B 发生的条件下，A 发生的概率。

例如：投掷骰子，观察点数，A表示“出现 3 点”，B表示“出现奇数点”，求$P(A)$及已知B发生的条件下A发生的概率$P(A|B)$.

$P(AB)$ ——联合概率，表示在样本空间Ω中，AB两个事件同时发生的概率，计算方法如下：

$$P(AB)=\frac{AB\ \text{中样本点数}}{\Omega\ \text{中样本点数}}$$

$P(B)$ —全概率或者边缘概率，表示在样本空间Ω中，B事件发生的概率，计算方法如下：

$$P(B)=\frac{B\ \text{中样本点数}}{\Omega\ \text{中样本点数}}$$

n 个事件的独立性

定义：若事件 A_1，A_2，…，A_n 中任意两个事件相互独立，即对于一切 $1 \le i < j \le n$, 有

$$P(A_iA_j) = P(A_i)P(A_j)$$

则称A_1，A_2，$\cdots A_n$两两相互独立.

定义：设 A_1，A_2，…，A_n为n 个事件，

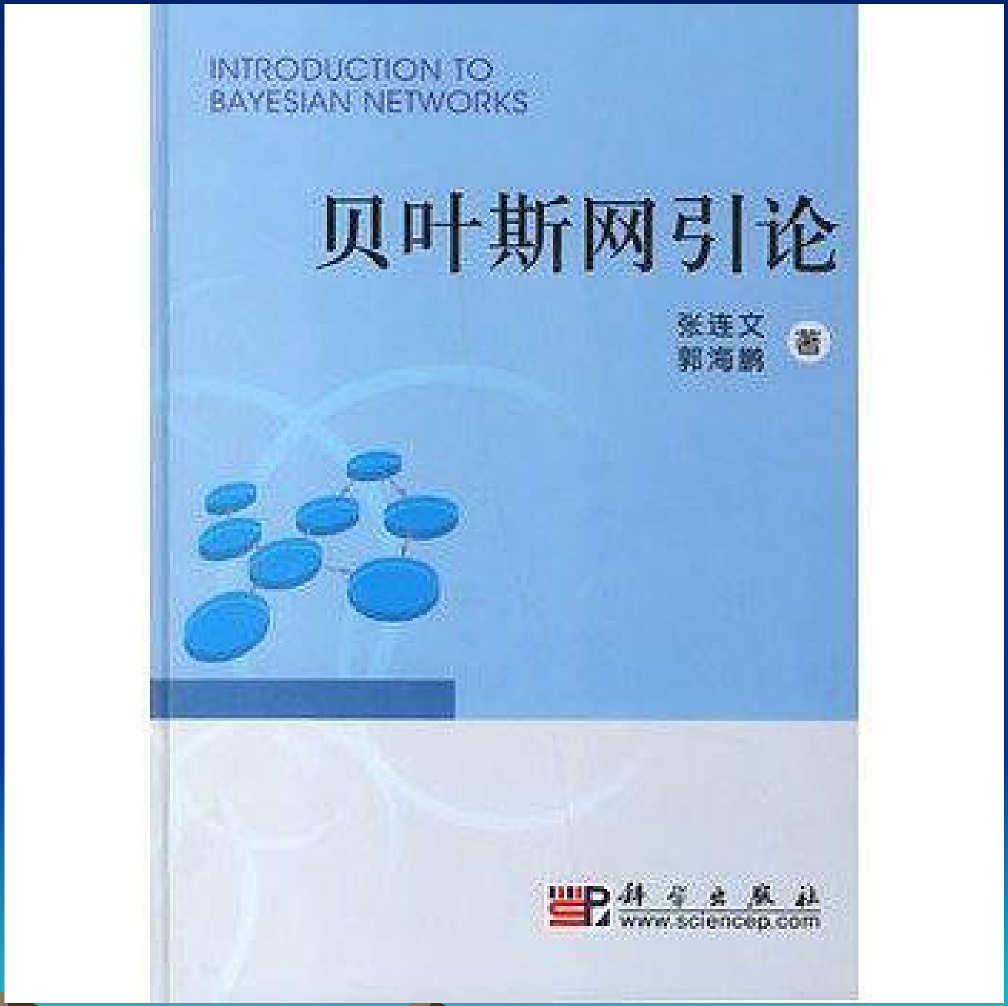
若对于任意$k(1 \le k \le n)$, 及 $1 \le i_1 < i_2 < \cdots < i_k \le n$

有 $P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$

则称A_1，A_2，$\cdots A_n$相互独立.

参考文献：

1. 贝叶斯网引论。张连文，郭海鹏著，科学出版社出版，2006年11月第一版。



参考文献：

2. 贝叶斯统计。茆诗松，汤银才著，中国统计出版社出版，2012年9月第二版。

参考文献：

3. 概率图模型原理与技术。[美]Daphne Koller，[以色列]Nir Friedman著，王飞跃，韩素青译，清华大学出版社，2015年3月第一版。