随机变量

司继春

[1]上海对外经贸大学

2023年9月

概览

一元随机变量

对于概率空间$(\Omega,\mathscr{F},\mathscr{P})$，样本空间$\Omega$可能是$\{\heartsuit,\spadesuit,\clubsuit,\diamondsuit\}$这样的集合，不便于研究。为了方便研究，我们通常将样本空间Ω映射到实轴$\mathbb{R}^*$上，并研究概率空间$(\mathbb{R}^*,\mathscr{B},P)$。为此我们需要引入随机变量：

随机变量

对于概率空间$(\Omega,\mathscr{F},\mathscr{P})$，映射

$$X:\Omega\to\mathbb{R}^*$$

满足：对于任意的$B\in\mathscr{B}$，有：

$$X^{-1}(B)\triangleq\{\omega:X(\omega)\in B\}\in\mathscr{F}$$

那么我们称X为随机变量（random variable, r.v.）。

一元随机变量

抛硬币

对于抛硬币的实验，$\Omega = \{H, T\}$，我们可以定义一个随机变量X如下：

$$\begin{cases} X(H) = 0 \\ X(T) = 1 \end{cases}$$

对于$\mathscr{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$，我们有

$$X^{-1}(\{0\}) = \{H\}, X^{-1}(\{1\}) = \{T\}$$

对于其他任何Borel集B，如果$1 \in B$则$T \in X^{-1}(B)$，如果$0 \in B$则$H \in X^{-1}(B)$。如此我们便定义了一个$\Omega \to \mathbb{R}^*$的随机变量X。

一元随机变量

到达次数

对于银行一天之内到达人数的问题，其$\Omega = \{0, 1, 2 \ldots\} = \mathbb{Z}$，定义$X(\omega) = \omega, \omega \in \mathbb{Z}$，同上，我们定义了从自然数集合$\mathbb{Z} \to \mathbb{R}$的随机变量$X$。

生成的$\sigma-$代数

- 实际上，可以证明，集合族$\{X^{-1}(B): B\in\mathscr{B}\}$是一个$\sigma-$代数，我们通常记为$\sigma\langle X\rangle$，成为由$X$生成的$\sigma-$代数。
- $\sigma\langle X\rangle$ 可以看做是随机变量X所携带的关于样本空间的信息，即如果我们观察到X，我们就可以知道$\sigma\langle X\rangle$ 中每一个事件是否发生。

生成的$\sigma-$代数

到达次数的信息

在到达次数的例子中，如果我们额外定义两个随机变量：
(1)退化的随机变量$X_0(\omega)=1$，即无论如何，X_0都是常数1；
(2)一个分类变量：

$$X_1(\omega)=\begin{cases}0 & if \quad \omega<10\\ 1 & if \quad \omega\geq 10\end{cases}$$

即当人数到达超过10人时$X_1=1$，否则为0。那么：

$$\sigma\langle X_0\rangle=\{\emptyset,\mathbb{Z}\}$$
$$\sigma\langle X_1\rangle=\{\emptyset,\mathbb{Z},\{\omega:\omega<10\},\{\omega:\omega\geq 10\}\}$$
$$\sigma\langle X\rangle=2^{\mathbb{Z}}$$

其中$2^{\mathbb{Z}}$代表$\mathbb{Z}$的所有子集所组成的$\sigma-$代数。

生成的$\sigma-$代数

到达次数的信息（续）

注意到：$\sigma\langle X_0\rangle \subset \sigma\langle X_1\rangle \subset \sigma\langle X\rangle$：

- X_0是一个退化的随机变量，并不能给我们带来关于样本空间的任何信息；
- 如果观察到了X_1，则只能判断实现的样本点究竟是小于10还是大于等于10；
- 而如果我们观察到X，那么我们可以对实现的样本点做出精确断言。
- 因而，三个随机变量中所携带的信息，X_0最少，X_1次之，X所携带的信息最多。

离散型随机变量

离散型随机变量

如果存在一个可数集$B \in \mathscr{B}$，满足$P(X \in B) = 1$，则随机变量X成为离散型随机变量。

随机变量的概率

以上定义了随机变量$X:\Omega\to\mathbb{R}$，然而并没有定义X在$(\mathbb{R},\mathscr{B})$上的概率。既然X定义在Ω上，而针对Ω有天然的概率空间$(\Omega,\mathscr{F},\mathscr{P})$，那么自然地，$X$在$(\mathbb{R},\mathscr{B})$上的概率函数应由$(\Omega,\mathscr{F},\mathscr{P})$相应定义。

导出的概率空间

对于一个随机变量$X:\Omega\to\mathbb{R}$，定义

$$P_X(B)=\mathscr{P}\left(X^{-1}(B)\right)=\mathscr{P}\left(\{\omega:X(\omega)\in B\}\right)$$

则P_X为概率函数，$(\mathbb{R},\mathscr{B},P_X)$为概率空间，我们称$(\mathbb{R},\mathscr{B},P_X)$为$(\Omega,\mathscr{F},\mathscr{P})$导出的概率空间。

分布函数

分布函数

对于一个随机变量X，函数

$$F_X(x) = P_X(X \le x) = \mathscr{P}\left(X^{-1}((-\infty, x])\right), \forall x \in \mathbb{R}$$

为一个分布函数（满足分布函数定义的要求），我们称其为累积分布函数（cumulative distribution function, c.d.f.）。

- 对于随机变量来说，累积分布函数包含了所有概率函数P_X的信息，因而使用P_X和使用累积分布函数F_X是等价的。
- 因而我们通常使用标记$X \sim F_X(x)$表示随机变量X服从F_X分布。
- 如果两个随机变量的累积分布函数$F_X(x) = F_Y(x)$，则我们称两个随机变量同分布（identically distributed），记为$X \sim Y$。

分布函数

泊松分布的分布函数

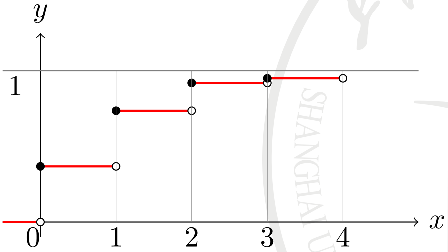
在建模一段时间内的到达次数时，我们经常使用所谓的泊松分布，即到达次数只可能取自然数值，或者$P(X \in \mathbb{Z}) = 1$，由于自然数集$\mathbb{Z}$为可数集，因而该分布是一个离散分布，且其概率函数为:

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

其中λ为这段时间的平均到达次数。其分布函数为:

$$F_X(x) = P(X \le x) = \sum_{i=0}^{x} \frac{\lambda^i}{i!} e^{-\lambda}$$

泊松分布的分布函数



分布函数

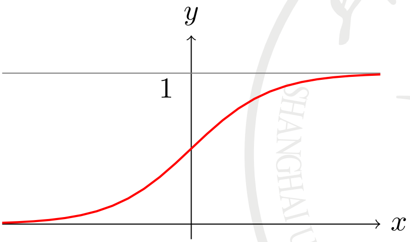
Logistic分布

如果一个随机变量X的分布函数为:

$$F_X(x) = \frac{e^x}{1+e^x}$$

那么我们称X服从Logistic分布

Logistic分布的分布函数



概率质量函数

概率质量函数

对于离散随机变量，如果$P\left(x\in B\right)=1$，B为可数集，**概率质量函数**（probability mass function, p.m.f）定义为：

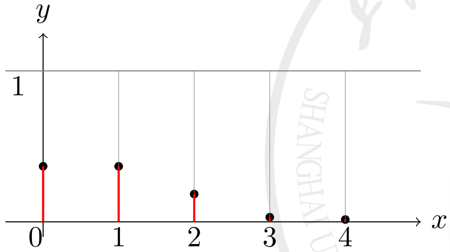
$$f_X\left(x\right)=P\left(X=x\right),\text{for all }x\in B$$

泊松分布的质量函数

泊松分布为离散型随机变量，且$P\left(X\in\mathbb{Z}\right)=1$，那么其概率质量函数为：

$$P\left(X=x\right)=\frac{\lambda^x}{x!}e^{-\lambda},\forall x\in\mathbb{Z}$$

泊松分布的质量函数



概率密度函数

概率密度函数

对于连续型随机变量，概率密度函数（probability density function, p.d.f），$f_X(x)$定义为：

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt, \text{for all } x$$

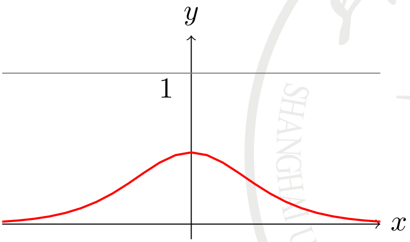
注意如果分布函数$F_X(x)$可导，那么其密度函数即其导函数。

Logistic分布的密度函数

Logistic分布的p.d.f为其分布函数的导函数：

$$f_X(x) = \frac{e^x}{(1+e^x)^2}$$

Logistic分布的密度函数



随机变量定义总结

$$
\begin{tikzcd}
\boxed{\Omega} \arrow[r, "\text{随机变量}X(\omega):"', "\Omega \to \mathbb{R}^*"'] \arrow[d, "\sigma\text{-代数}"'] & \boxed{\mathbb{R}} \arrow[r, "\text{分布函数}F_X(x):"', "\mathbb{R}^* \to {[0,1]}"'] \arrow[d, "\sigma\text{-代数}"] \arrow[dl, "X^{-1}(A): A \in \mathscr{B}"'] & \boxed{[0,1]} \\
\boxed{\mathscr{F}} & \boxed{\mathscr{B}} &
\end{tikzcd}
$$

黎曼积分

- 我们在《微积分》课程中学习的积分一般是黎曼积分（Riemann integrals），即对于任意的一个函数$f(t)$，其积分可以通过将自变量划分为网格点，计算求和并取极限的方式得到
- 我们令区间$[a,b]$上的一个划分（partition）$\Pi : a = t_0 < t_1 < \cdots < t_n = b$，并定义$\Delta_k = t_k - t_{k-1}$，那么我们可以计算：

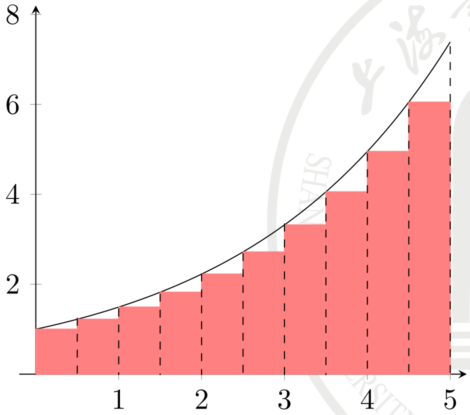
$$I_n = \sum_{k=1}^{n} f(t_{k-1})(t_k - t_{k-1}) = \sum_{k=1}^{n} f(t_{k-1}) \Delta_k$$

现在，我们令$mesh(\Pi) = \max_k \Delta_k \to 0$，此时定义：

$$\int_a^b f(t)\,dt = I \triangleq \lim_{mesh(\Pi)\to 0} I_n$$

即为函数$f(t)$在区间$[a,b]$上的黎曼积分。

黎曼积分



Riemann-Stieltjes积分

- 我们还可以仿照黎曼积分的定义，定义Riemann-Stieltjes积分，即如：$\int_a^b f(t)\,dg(t)$形式的积分。
- 我们知道，如果$g(x)$可导，那么以上积分等于$\int_a^b f(t)\,g'(t)\,dt$，然而$g(t)$并不一定是处处可导的，所以有时需要重新定义以上积分。
- 类似的，我们可以使用上述的划分，定义$\Delta_i g = g(t_i) - g(t_{i-1})$，并定义级数：

$$I_n = \sum_{k=1}^{n} f(t_{k-1})\left[g(t_k) - g(t_{k-1})\right] = \sum_{k=1}^{n} f(t_{k-1})\,\Delta_k g$$

同样，我们令$mesh(\Pi) = \max_k \Delta_k \to 0$，此时定义：

$$\int_a^b f(t)\,dg(t) = I \triangleq \lim_{mesh(\Pi)\to 0} I_n$$

如果以上极限存在，那么我们就定义了Riemann-Stieltjes积分

Riemann-Stieltjes积分

- 如果函数$g(t)$连续可导，那么根据中值定理，有：

$$I_n=\sum_{k=1}^{n} f(t_{k-1})\left[g(t_k)-g(t_{k-1})\right]=\sum_{k=1}^{n} f(t_{k-1})\, g'(t_k^*)\left[t_k-t_{k-1}\right]$$

- 从而当$mesh(\Pi)=\max_k \Delta_k \to 0$时，可以得到：

$$\int_a^b f(t)\,dg(t)=\lim_{mesh(\Pi)\to 0} I_n=\int_a^b f(t)\,g'(t)\,dt$$

- 然而如果$g(t)$不连续可导，那么以上公式就无法得到，但是使用Riemann-Stieltjes积分的定义仍然可以计算出其积分。

Riemann-Stieltjes积分

一个简单的Riemann-Stieltjes积分

假设:

$$g(x)=\begin{cases}0 & x<1\\ 1 & x\geq 1\end{cases}$$

这是一个不连续函数。如果我们要计算$\int_0^2 xdg(x)$，按照定义，我们可以使用划分$\Pi: 0=t_0<t_1<\cdots<t_n=0$，计算:

$$I_n=\sum_{k=1}^{n} t_{k-1}\left[g(t_k)-g(t_{k-1})\right]$$

注意到只有当$t_{k-1}<1\leq t_k$时，$g(t_k)-g(t_{k-1})=1$，其他情况都有$g(t_k)-g(t_{k-1})=0$，从而以上公式可以简化为:

$$I_n=\sum_{k=1}^{n} 1\left\{t_{k-1}<1\leq t_k\right\} t_{k-1}$$

Riemann-Stieltjes积分

泊松分布的Riemann-Stieltjes积分

假设$F(x) = P(X \le x) = \sum_{i=0}^{x} \frac{\lambda^i}{i!} e^{-\lambda}$，我们希望计算$\int_0^b x dF(x)$。$F(x)$是一个不连续的函数。类比上例，使用划分$\Pi : 0 = t_0 < t_1 < \cdots < t_n = b$，级数：

$$I_n = \sum_{k=1}^{n} t_{k-1} \left[F(t_k) - F(t_{k-1}) \right]$$

中当$i \le t_{k-1} < t_k < i+1$时，$F(t_k) - F(t_{k-1}) = 0$;
当$t_{k-1} < i \le t_k$时，$F(t_k) - F(t_{k-1}) = \frac{\lambda^i}{i!} e^{-\lambda}$，从而
当$mesh(\Pi) \to 0$时，可以计算：

$$\int_0^b x dF(x) = \sum_{i=0}^{\lfloor b \rfloor} i \times \frac{\lambda^i}{i!} e^{-\lambda}$$

离散型随机变量的期望

离散型随机变量期望

对于离散型随机变量$X \in B$，其数学期望（mathematical expectation）定义为：

$$\mathbb{E}(X) = \sum_{b \in B} b \cdot P(X = b)$$

离散型随机变量期望

以上定义可以使用Riemann-Stieltjes积分的形式表达：

$$\mathbb{E}(X) = \int_{\mathbb{R}} x dF_X(x)$$

离散型随机变量的期望

骰子的期望

抛一枚骰子，其样本空间为$\Omega = \{1,2,3,4,5,6\}$，假设骰子均匀，那么其概率函数可以由：

$$\mathscr{P}(\{w\}) = \frac{1}{6}, \omega \in \Omega$$

来定义。如果定义随机变量$X(\omega) = \omega$，那么其数学期望为：

$$\mathbb{E}(X) = \sum_{j \in \mathbb{Z}} j \cdot P(X = j) = \sum_{j=1}^{6} j \cdot \frac{1}{6} = \frac{7}{2}$$

离散型随机变量的期望

泊松分布的期望

泊松分布随机变量的期望为:

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{j=0}^{\infty} j \cdot P(X = j) \\
&= \sum_{j=0}^{\infty} j \cdot \frac{\lambda^j}{j!} e^{-\lambda} \\
&= \lambda \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} e^{-\lambda} \\
&= \lambda
\end{aligned}
$$

期望

期望

如果概率空间$(\Omega,\mathscr{F},\mathscr{P})$导出了概率空间$(\mathbb{R},\mathscr{B},P)$，令$g$为一个可测函数，则期望：

$$\mathbb{E}\left(g\left(X\right)\right)=\int_{\mathbb{R}}g\left(x\right)P\left(dx\right)=\int_{\mathbb{R}}g\left(x\right)dF_X\left(x\right)$$

如果等式两边积分都存在。如果额外的，密度函数存在，则：

$$\mathbb{E}\left(g\left(X\right)\right)=\int_{\mathbb{R}}g\left(x\right)f_X\left(x\right)dx$$

期望的线性性

对于两个$(\Omega,\mathscr{F},\mathscr{P})$上的随机变量$X,Y$，如果期望存在，有：

$$\mathbb{E}\left(aX+bY\right)=a\mathbb{E}\left(X\right)+b\mathbb{E}\left(Y\right)$$

可积性

可积性

积分存在要求:

$$\mathbb{E}\left(|X|\right) < \infty$$

如果上式不成立，则积分不存在，从而期望也不存在。

期望的存在性

积分不存在的例子：Cauchy分布

Cauchy的密度函数为：$f(x) = \frac{1}{\pi(1+x^2)}$如果一个随机变量服从Cauchy分布，则：

$$\mathbb{E}(|X|) = \int_{\mathbb{R}} \frac{|x|}{\pi} \frac{1}{1+x^2} dx = \frac{2}{\pi} \int_{[0,\infty)} \frac{x}{1+x^2} dx$$

对于任意正数M:

$$\int_{[0,M]} \frac{x}{1+x^2} dx = \left. \frac{\log\left(1+x^2\right)}{2} \right|_0^M = \frac{\log\left(1+M^2\right)}{2}$$

因此：$\mathbb{E}(|X|) = \frac{1}{\pi} \lim_{M\to\infty} \log\left(1+M^2\right) = \infty$，因而该随机变量是不可积的，期望不存在。

正态分布的期望

正态分布的期望

给定常数μ和σ^2，如果随机变量X的分布函数为：

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx$$

那么我们称随机变量X服从正态分布，记为$X \sim N(\mu, \sigma^2)$。注意以上积分内：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

为其密度函数。

正态分布的期望

正态分布的期望（续）

$$
\begin{aligned}
F(\infty) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \sigma d\frac{x-\mu}{\sigma} \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz \\
&= 1
\end{aligned}
$$

正态分布的期望

正态分布的期望（续）

可积性:

$$
\begin{aligned}
\mathbb{E}|X| &= \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\
&= 2\int_{0}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\
&= \frac{2}{\sqrt{2\pi}\sigma} \int_{0}^{\infty} x \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\
&= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} (\mu+\sigma z) \exp\left\{-\frac{z^2}{2}\right\} dz \\
&= \frac{2\mu}{\sqrt{2\pi}} \int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz + \frac{2\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z \exp\left\{-\frac{z^2}{2}\right\} dz \\
&= \frac{2\mu}{\sqrt{2\pi}} \frac{\sqrt{2\pi}}{2} + \frac{2\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} d\frac{z^2}{2} \\
&= \mu + \frac{2\sigma}{\sqrt{2\pi}} < \infty
\end{aligned}
$$

正态分布的期望

正态分布的期望（续）

$$
\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{\sigma^2}\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma z) \exp\left\{-\frac{z^2}{2}\right\} dz \\
&= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \exp\left\{-\frac{z^2}{2}\right\} dz \\
&= \mu + \frac{\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z \exp\left\{-\frac{z^2}{2}\right\} dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{0} z \exp\left\{-\frac{z^2}{2}\right\} dz \\
&= \mu
\end{aligned}
$$

矩

矩和中心矩

- 对于任意的正数p，如果$\mathbb{E}\left(|X|^p\right) < \infty$，则记$X \in L^p = L^p\left(\Omega, \mathscr{F}, \mathscr{P}\right)$。
- 对于整数$r < p$，随机变量X的r阶矩被定义为$\mathbb{E}\left(X^r\right)$。一阶矩即为随机变量$X$的期望。
- 此外，随机变量X的r阶中心矩被定义为$\mathbb{E}\left(\left[X - E\left(X\right)\right]^r\right)$。

方差和标准差

方差和标准差

2阶中心矩：

$$\mathbb{V}(X) = \mathbb{E}[X - \mathbb{E}(X)]^2$$

即为随机变量的方差（variance），记为$\mathbb{V}(X)$或者$\sigma^2(X)$，标准差（standard deviation）定义为$\sigma(X) = \sqrt{\mathbb{V}(X)}$。

方差的性质

$$\mathbb{V}(X) = \mathbb{E}\left([X - \mathbb{E}(X)]^2\right) = \mathbb{E}\left(X^2 - 2\mathbb{E}(X)\cdot X + \mathbb{E}(X)^2\right)$$
$$= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

$$\begin{aligned}
\mathbb{V}(aX+b) &= \mathbb{E}\left[aX+b-\mathbb{E}(aX+b)\right]^2 \\
&= \mathbb{E}\left[aX+b-a\mathbb{E}(X)-b\right]^2 \\
&= a^2\mathbb{E}(X-\mathbb{E}X)^2 \\
&= a^2\mathbb{V}(X)
\end{aligned}$$

正态分布的方差

正态分布的方差

为了计算$X \sim N\left(\mu, \sigma^2\right)$的方差，我们首先必须保证$X^2$可积，即$\mathbb{E}\left|X^2\right| < \infty$。由于$X^2 > 0$且$\mathbb{V}\left(X\right) = \mathbb{E}\left(X^2\right) - \left(\mathbb{E}X\right)^2$，因而我们只要计算$\mathbb{E}X^2$即可：

$$
\begin{aligned}
\mathbb{E}\left(X^2\right) &= \int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{\sigma^2}\right\} dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^2 \exp\left\{-\frac{(x-\mu)^2}{\sigma^2}\right\} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma z)^2 \exp\left\{-\frac{z^2}{2}\right\} dz
\end{aligned}
$$

正态分布的方差

正态分布的方差

$$
\begin{aligned}
\mathbb{E}\left(X^2\right) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\mu+\sigma z)^2 \exp\left\{-\frac{z^2}{2}\right\}dz \\
&= \frac{1}{\sqrt{2\pi}}\mu^2\int_{-\infty}^{\infty}\exp\left\{-\frac{z^2}{2}\right\}dz + \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}z^2\exp\left\{-\frac{z^2}{2}\right\}dz \\
&\quad + \underbrace{\frac{2\mu\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}z\exp\left\{-\frac{z^2}{2}\right\}dz}_{=0} \\
&= \mu^2 - \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}z\,d\exp\left\{-\frac{z^2}{2}\right\}
\end{aligned}
$$

正态分布的方差

正态分布的方差

$$
\begin{aligned}
\mathbb{E}\left(X^2\right) &= \mu^2 - \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z d \exp\left\{-\frac{z^2}{2}\right\} \\
&= \mu^2 - \frac{\sigma^2}{\sqrt{2\pi}} \underbrace{\left[z \exp\left\{-\frac{z^2}{2}\right\}\right]_{-\infty}^{\infty}}_{=0} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz \\
&= \mu^2 + \sigma^2 < \infty
\end{aligned}
$$

从而

$$
\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - (\mathbb{E}X)^2 = \sigma^2
$$

偏度

偏度（**skewness**）为随机变量的三阶中心矩，即定义：

$$\gamma = \mathbb{E}\left(\left[\frac{X-\mathbb{E}(X)}{\sigma(X)}\right]^3\right) = \frac{\mathbb{E}\left[X-\mathbb{E}(X)\right]^3}{\sigma^3(X)}$$

- 如果X为对称分布，那么必然有$\gamma = 0$。
- 当$\gamma > 0$时，分布函数右边的尾巴比较厚，我们称其分布为右偏（**positive skew**）分布
- 反之为左偏（**negative skew**）分布。

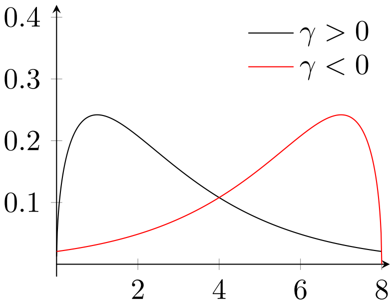
峰度

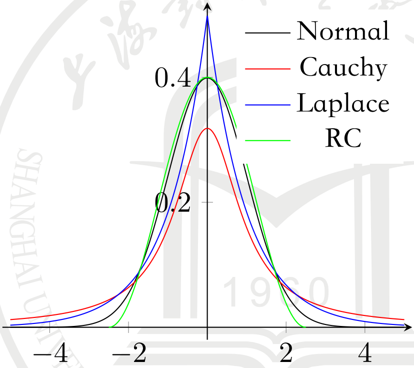
峰度是随机变量的四阶中心矩，即：

$$\mathsf{Kurt}\left(X\right)=\mathbb{E}\left(\left[\frac{X-\mathbb{E}\left(X\right)}{\sigma\left(X\right)}\right]^4\right)=\frac{\mathbb{E}\left(\left[X-\mathbb{E}\left(X\right)\right]^4\right)}{\left[\mathbb{V}\left(X\right)\right]^2}$$

- 尽管我们一般将其称为「峰度」，然而这一称呼并不准确，更准确的称呼应该为「尾厚度」。
- 正态分布的峰度为3。
- 我们一般会把峰度与正态分布的峰度相比，定义超额峰度（**excess kurtosis**）为峰度减3
- 因而如果超额峰度>0，那么其尾巴比正态分布的尾巴要厚，而如果超额峰度<0，那么其尾巴要比正态分布的尾巴要薄。

偏度和峰度


0.4
0.3
0.2
0.1
2
4
6
8
$\gamma > 0$
$\gamma < 0$


0.4
0.2
−4
−2
2
4
Normal
Cauchy
Laplace
RC

Chebyshev不等式

Chebyshev不等式

如果函数ψ满足：$\psi(u) = \psi(-u) \geq 0$，且在$(0, \infty)$上单调递增，$X$为随机变量，且$\psi(X) < \infty$，则对于$u > 0$，有：

$$P(|X| \geq u) \leq \frac{\mathbb{E}[(\psi(X))]}{\psi(u)}$$

证明

$$\begin{aligned}\mathbb{E}[\psi(X)] &= \int_{\mathbb{R}} \psi(X) P(dX) \\ &\geq \int_{\{|X| \geq u\}} \psi(X) P(d\omega) \\ &\geq \psi(u) P(|X| \geq u)\end{aligned}$$

Chebyshev不等式

Chebyshev不等式应用

令

$$Y = \frac{[X - \mathbb{E}(X)]}{\sigma(X)}$$

则$\mathbb{E}(Y) = 0$，$\mathbb{E}\left(Y^2\right) = 1$。令$\psi(x) = x^2$，有:

$$P(|Y| \geq u) \leq \frac{1}{u^2}$$

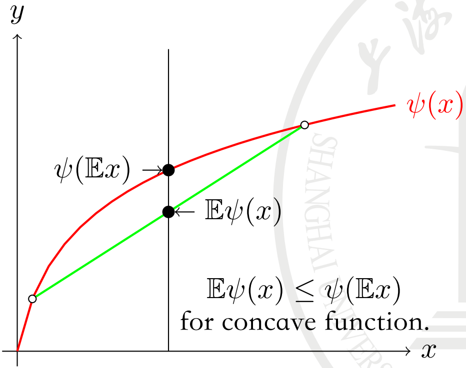
Jensen不等式

Jensen不等式

如果ψ为凹函数，且随机变量X和$\psi(X)$可积，则：

$$\psi(\mathbb{E}(X)) \geq \mathbb{E}[\psi(X)]$$

Jensen不等式



Hölder不等式

Hölder不等式

对于两个随机变量X, Y，若$\mathbb{E}|X|^p < \infty, \mathbb{E}|Y|^q < \infty$，且

$$1 < p < \infty, 1 < q < \infty, \frac{1}{p} + \frac{1}{q} = 1$$

有:

$$|\mathbb{E}(XY)| \le \mathbb{E}(|XY|) \le [\mathbb{E}(|X|^p)]^{\frac{1}{p}} [\mathbb{E}(|Y|^q)]^{\frac{1}{q}}$$

当且仅当存在常数$0 \le c_1, c_2 < \infty$使得$P(c_1|X|^p = c_2|Y|^q) = 1$时，等号成立。

Cauchy-Schwarz不等式

Hölder不等式中，当$p = q = 2$时，有

Cauchy-Schwarz不等式

对于两个随机变量X, Y，若满足可积性，有：

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq \left[\mathbb{E}\left(|X|^2\right)\right]^{\frac{1}{2}} \left[\mathbb{E}\left(|Y|^2\right)\right]^{\frac{1}{2}}$$

当且仅当存在常数a, b使得$P(Y = aX + b) = 1$成立时，等号成立。

伯努利分布

伯努利分布即伯努利试验（Bernoulli trial）的结果。伯努利试验即试验结果只有两种可能性（成功，失败），定义为$X = 1\ if$ 成功, $= 0\ if$ 失败。如果成功的概率为p，那么伯努利分布（**Bernoulli Distribution**）$X \sim Ber\left(p\right)$即：

$$X = \begin{cases} 1 & with\ probability\ p \\ 0 & with\ probability\ 1-p \end{cases} \quad 0 \le p \le 1$$

- 期望：$\mathbb{E}\left(X\right) = p$
- 方差：$\mathbb{V}\left(X\right) = p\left(1-p\right)$

二项分布

二项分布即独立重复N次伯努利实验，成功次数$Y=\sum_{i=1}^{N}X_i, X_i\sim Ber\left(p\right)$的分布。简单计算可以得到，随机变量$Y$的概率质量函数：

$$P\left(Y=y|N,p\right)=\left(\begin{array}{c}N\\y\end{array}\right)p^y\left(1-p\right)^{N-y},y=1,2,...,N$$

我们成Y服从二项分布（**Binomial Distribution**），记为$Y\sim Bi\left(N,p\right)$。

- 期望：$\mathbb{E}\left(Y\right)=Np$
- 方差：$\mathbb{V}\left(Y\right)=Np\left(1-p\right)$

负二项分布

负二项分布（Negative Binomial Distribution）用于建模为了达到r次成功所需要的试验次数，如果记为Z，那么:

$$
\begin{aligned}
P(Z=z|r,p) &= \binom{z-1}{r-1} p^{r-1}(1-p)^{z-r}\cdot p \\
&= \binom{z-1}{r-1} p^{r}(1-p)^{z-r}, z=r,r+1,\ldots
\end{aligned}
$$

记为$Z\sim NB(r,p)$。

- 期望：$\mathbb{E}(Z)=\frac{r}{p}$
- 方差：$\mathbb{V}(Y)=\frac{r(1-p)}{p^2}$

几何分布

- **几何分布（Geometric Distribution）是最简单形式的负二项分布，如果一个随机变量$V \sim NB(1,p)$，则随机变量V服从几何分布：**

$$P(V=v|p)=p(1-p)^{v-1}$$

  我们记$V \sim G(p)$。

- 几何分布具有无记忆性，即：

$$P(V>s|V>t)=P(V>s-t), s>t$$

泊松分布

如果随机变量X的概率质量函数为：

$$P\left(X=x\right)=\frac{\lambda^{x}}{x!}e^{-\lambda},x=0,1,2,\ldots$$

那么我们称随机变量X服从泊松分布（Poisson Distribution），记为$X\sim P\left(\lambda\right)$。泊松分布经常被用来建模次数、个数等相关问题。

- 期望：$\mathbb{E}\left(X\right)=\lambda$
- 方差：$\mathbb{V}\left(X\right)=\lambda$

均匀分布

如果随机变量X在区间$[a,b]$内的概率密度函数为常数，即：

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & otherwise \end{cases}$$

则称随机变量X服从区间$[a,b]$上的均匀分布（**Uniform Distribution**），记为$X \sim U(a,b)$。

- 期望：$\mathbb{E}(X) = \frac{a+b}{2}$
- 方差：$\mathbb{V}(X) = \frac{(b-a)^2}{12}$

指数分布

若随机变量X的概率密度函数为:

$$f\left(x|a,\beta\right)=\frac{1}{\beta}e^{-\frac{x-a}{\beta}},x>a$$

那么我们称X服从**指数分布**（Exponential Distribution），记为$X\sim E\left(a,\beta\right)$。

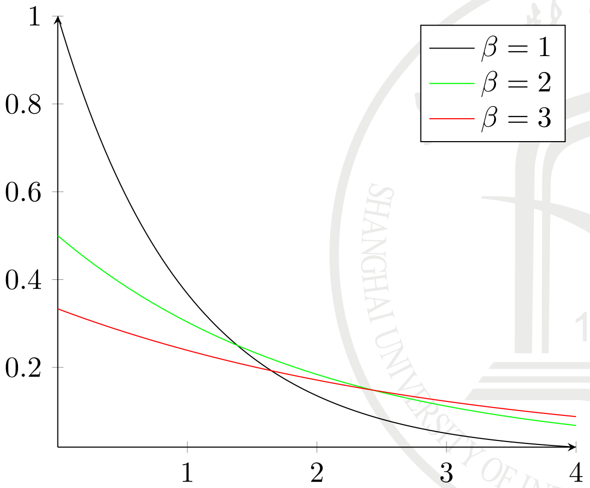
- 分布函数：$F\left(x\right)=1-e^{-\frac{x-a}{\beta}},x>a$
- 期望：$\mathbb{E}\left(X\right)=\beta+a$
- 方差：$\mathbb{V}\left(X\right)=\beta^2$

指数分布

- 如果一个随机变量$Y\sim P\left(\lambda\right)$，其中参数$\lambda$代表一段时间$T$内到达的平均次数，所以对于时间$t$内，平均到达的人数即$\lambda\frac{t}{T}$，因而在t时间内只有0次到达的概率为$e^{-\frac{\lambda t}{T}}$。
- 如果在时间t内有0次到达，意味着等待时间大于t，因而等待时间X大于t的概率$P\left(X>t\right)=e^{-\frac{\lambda t}{T}}$，所以$P\left(X\leq t\right)=1-e^{-\frac{\lambda t}{T}}$，因而等待时间X服从指数分布，即$X\sim E\left(0,\frac{T}{\lambda}\right)$。
- 同样，指数分布也具有无记忆性，即如果令$\beta=\frac{T}{\lambda}$，那么$X\sim E\left(0,\beta\right)$，那么

$$P\left(X>s|X>t\right)=P\left(X>s-t\right),s>t$$

指数分布



Beta分布

如果一个随机变量$X \in (0,1)$，且其分布函数为：

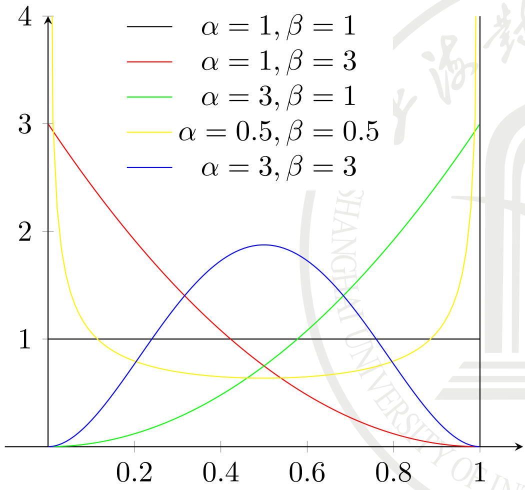
$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1, \alpha > 0, \beta > 0$$

那么我们称X服从**Beta分布**（**Beta Distribution**）。其中

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- 期望：$\mathbb{E}(X) = \frac{\alpha}{\beta+a}$
- 方差：$\mathbb{V}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Beta分布


4
3
2
1
$\alpha = 1, \beta = 1$
$\alpha = 1, \beta = 3$
$\alpha = 3, \beta = 1$
$\alpha = 0.5, \beta = 0.5$
$\alpha = 3, \beta = 3$
0.2
0.4
0.6
0.8
1

正态分布

如果一个随机变量X潜在受到非常多的独立因素的影响，即$X = f_1 + f_2 + \cdots$，而每个f_i又不能单独对X有非常大的影响，那么一般来说X将会服从正态分布（**normal distribution**）或者高斯分布（**Gaussian distribution**）。正态分布的密度函数为:

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

记为$X \sim N(\mu, \sigma^2)$。

- 期望：$\mathbb{E}(X) = \mu$
- 方差：$\mathbb{V}(X) = \sigma^2$

正态分布

- 如果$X \sim N\left(\mu, \sigma^2\right)$，如果令$Z = \frac{X-\mu}{\sigma}$，则$E\left(Z\right) = \frac{E(X)-\mu}{\sigma} = 0$，$\mathbb{V}\left(Z\right) = \frac{1}{\sigma^2} \cdot \mathbb{V}\left(X\right) = 1$，随机变量$Z$服从$\mu = 0, \sigma = 1$的正态分布，即$Z \sim N\left(0, 1\right)$，我们称之为标准正态分布（standard normal distribution）。
- 标准正态分布的密度函数为：

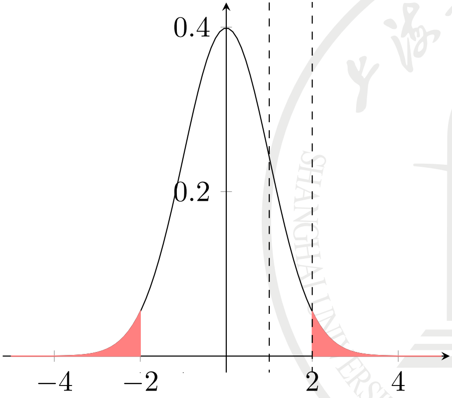
$$\phi\left(x\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- 而其分布函数为：

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

没有初等函数表示，因而一般我们用$\Phi\left(z\right)$来代表标准正态的分布函数。

正态分布



正态分布

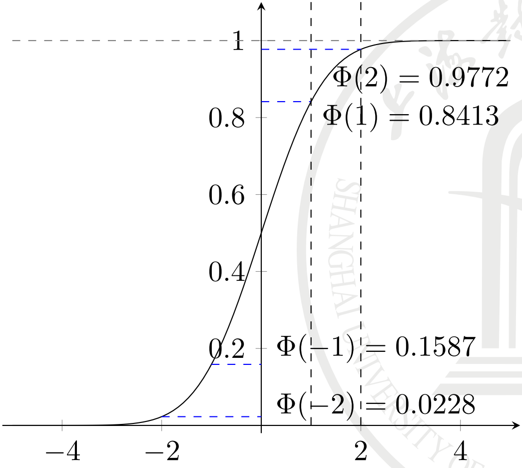
- 由于正态分布为对称分布，即$\phi(x)=\phi(-x)$，因而分布函数$\Phi(x)=1-\Phi(-x)$。
- 如果$X\sim N\left(\mu,\sigma^2\right)$，那么$Z=\frac{X-\mu}{\sigma}\sim N(0,1)$，根据标准正态的分布函数$\Phi(x)$可以计算$X$在区间内取值的概率，比如：

$$P(|X-\mu|\le\sigma)=P(|Z|<1)=\Phi(1)-\Phi(-1)=2\Phi(1)-1=0.6827$$

同理，有：

$$\begin{aligned}
P(|X-\mu|\le 1.65\sigma)&=P(|Z|<1.65)\approx 0.90\\
P(|X-\mu|\le 1.96\sigma)&=P(|Z|<1.96)\approx 0.95\\
P(|X-\mu|\le 2\sigma)&=P(|Z|<2)=0.9545\\
P(|X-\mu|\le 2.58\sigma)&=P(|Z|<2.58)\approx 0.99\\
P(|X-\mu|\le 3\sigma)&=P(|Z|<3)=0.9973\\
P(|X-\mu|\le 5\sigma)&=P(|Z|<5)\ge 1-10^{-6}\\
P(|X-\mu|\le 6\sigma)&=P(|Z|<6)\ge 1-10^{-8}
\end{aligned}$$

正态分布



对数正态分布

若随机变量$Y = e^X, X \sim N\left(\mu, \sigma^2\right)$，则我们称随机变量$Y$服从对数正态分布（lognormal distribution），记为$Y \sim LN\left(\mu, \sigma^2\right)$。

- 期望：$\mathbb{E}\left(Y\right) = e^{\mu + \frac{\sigma^2}{2}}$
- 方差：$\mathbb{V}\left(Y\right) = e^{2\left(\mu+\sigma^2\right)} - e^{2\mu+\sigma^2}$

χ^2分布

K个独立的标准正态分布的平方和的分布被称为卡方分布（Chi-square distribution）或者χ^2分布。即，如果$X_1, X_2, ..., X_K$为K个独立的标准正态分布，那么

$$X = \sum_{i=1}^{K} X_i^2 \sim \chi^2(K)$$

其中参数K为卡方分布的自由度（degrees of freedom）。

- 期望：$\mathbb{E}(X) = K$
- 方差：$\mathbb{V}(X) = 2K$

F分布

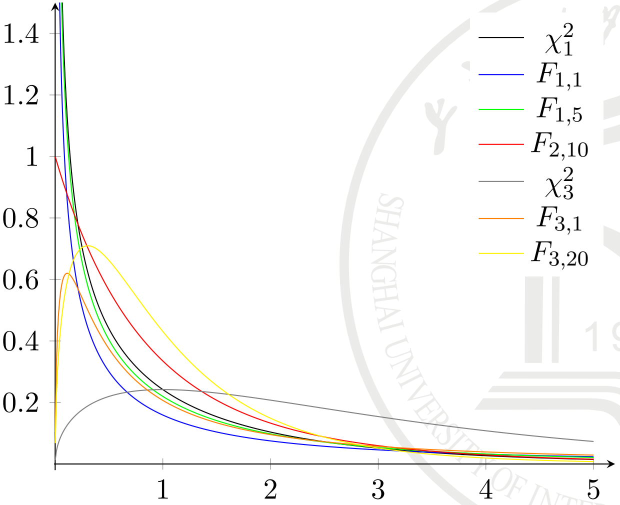
如果存在两个独立的χ^2分布$X_1 \sim \chi^2(n), X_2 \sim \chi^2(m)$，那么随机变量

$$F = \frac{X_1/n}{X_2/m}$$

即服从F分布（F-**distribution**），记为$F \sim F(n, m)$，参数n, m为自由度。

- 期望：$\mathbb{E}(X) = \frac{m}{m-2}, m > 2$
- 方差：$\mathbb{V}(X) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}, m > 4$

χ^2分布与F分布



t分布

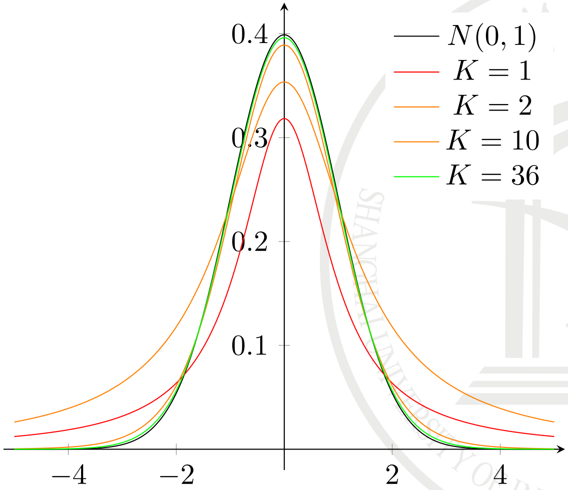
如果存在一个标准正态分布$Z \sim N(0,1)$，以及一个χ^2分布$X \sim \chi^2(K)$，且Z和X独立，那么随机变量

$$T = \frac{Z}{\sqrt{X/K}}$$

即服从学生氏t分布（Students' t-distribution）或者简称t分布（t-distribution），记为$T \sim t(K)$，参数K为自由度。

- 期望：$\mathbb{E}(X) = 0, K > 1$
- 方差：$\mathbb{V}(X) = \frac{K}{K-2}, K > 2$

标准正态分布与t分布



柯西分布

若随机变量X的概率密度函数为:

$$f(x|\mu,\sigma)=\frac{1}{\pi\sigma}\left[1+\left(\frac{x-\mu}{\sigma}\right)^2\right]^{-1}$$

那么我们称X服从**柯西分布（Cauchy distribution）**，记为$X\sim C(\mu,\sigma)$。

- 期望：不存在
- 方差：不存在

由于柯西分布的一阶矩不存在，因而通常作为不可积的分布的例子。此外，可以得到，两个独立的标准正态分布X,Y，其比例$U=\frac{Y}{X}$刚好服从柯西分布。此外，$C(0,1)\sim t(1)$。

逻辑斯蒂分布

若随机变量X的分布函数为:

$$F(x|\mu,\sigma)=\frac{e^{\frac{x-\mu}{\sigma}}}{1+e^{\frac{x-\mu}{\sigma}}}$$

那么我们称X服从逻辑斯蒂分布（logistic distribution），其密度函数为:

$$f(x|\mu,\sigma)=\frac{1}{\sigma}\frac{e^{-\frac{x-\mu}{\sigma}}}{\left[1+e^{-\frac{x-\mu}{\sigma}}\right]^2}$$

记为：$X\sim LG(\mu,\sigma)$。

- 期望：$\mathbb{E}(X)=\mu$
- 方差：$\mathbb{V}(X)=\sigma^2\frac{\pi^2}{3}$

分布族

- 给定任意的一个$(\Omega, \mathscr{F})$，在这个空间里面我们可以定义各种不同的概率函数，统计学需要解决的问题就是使用样本数据去推断总体的概率函数$\mathscr{P}$。
- 然而一般情况下，概率函数$\mathscr{P}$的可能性太多，因而我们经常将概率函数限制在一个可能集合内，为此我们可以定义参数族的概念。

参数族

如果对于每一个已知的$\theta \in \Theta, \Theta \subset \mathbb{R}^d$，$P_\theta$为在$(\Omega, \mathcal{F})$上的一个已知的概率函数，那么$\{P_\theta, \theta \in \Theta\}$即被称为参数族（**parametric family**），其中Θ为参数空间（**parametric space**），正整数d为参数空间的维数。

位置尺度族

对于一个随机变量X，我们令$Y=\sigma X+\mu,\sigma>0$，那么其分布函数：

$$\begin{aligned}F_Y\left(y\right)=P\left(Y\leq y\right)&=P\left(\sigma X+\mu\leq y\right)\\&=P\left(X\leq\frac{y-\mu}{\sigma}\right)\\&=F_X\left(\frac{y-\mu}{\sigma}\right)\end{aligned}$$

所以其密度函数满足：

$$f_Y\left(y\right)=\frac{1}{\sigma}f_X\left(\frac{x-\mu}{\sigma}\right)$$

同时，其期望和方差满足：

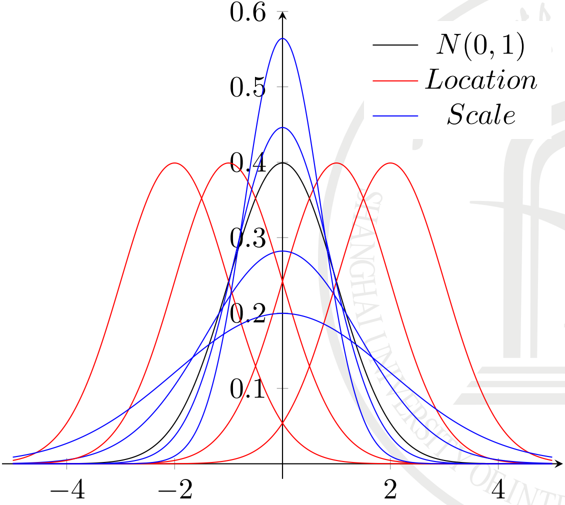
$$\mathbb{E}\left(Y\right)=\sigma\mathbb{E}\left(X\right)+\mu$$
$$\mathbb{V}\left(Y\right)=\sigma^2\mathbb{V}\left(X\right)$$

位置尺度族

位置尺度族

令$f(x)$为任意的密度函数，那么形如$\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$的密度函数就形成了以$\theta=(\mu,\sigma)$为参数的参数族，我们称之为位置尺度族（location-scale family）。位置尺度族即对于任意的随机变量做位移和数乘所得到的随机变量的分布组成的分布族。其中参数μ一般称为位置参数（location parameter），而σ一般称为尺度参数（scale parameter）。

位置尺度族



位置尺度族

正态分布

标准正态分布的密度函数为

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

对于任意的$\mu\in\mathbb{R},\sigma\in\mathbb{R}^+$，其位置尺度族的密度函数可以写为:

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

即正态分布的密度函数。因而正态分布族$\{P_{\mu,\sigma},\mu\in\mathbb{R},\sigma\in\mathbb{R}^+\}$为一个位置尺度族。

标准化

- 正如正态分布一样，很多时候我们会将一个随机变量变换为期望为0、方差为1的随机变量，即令随机变量

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

易得

$$\mathbb{E}(Z) = 0$$
$$\mathbb{V}(Z) = 1$$

我们称之为标准化（**standardized**）。

- 由于$P(X \leq x) = P(Z \leq \sigma x + \mu)$，因而应用中对于位置尺度族，经常首先研究标准化之后的随机变量，进而推广到其位置尺度族。

单参数指数分布族

指数分布族

对于一个参数族$\{P_\theta, \theta \in \Theta\}$，如果其概率密度（质量）函数可以写成如下形式：

$$f(x|\theta) = h(x) \cdot \exp\{\eta(\theta) \cdot T(x) - B(\theta)\}$$

那么我们称$\{P_\theta, \theta \in \Theta\}$为单参数指数分布族（one-parameter exponential family）。

单参数指数分布族

由于很多随机变量的取值范围不是$\mathbb{R}$，然而随机变量的值域为$\mathbb{R}$，因而我们一般使用指示函数（indicator function）来表示随机变量的取值范围，即定义

$$1_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

例如$1_{(-\infty,0]}(x)$即当$x \le 0$时，$1_{(-\infty,0]}(x) = 1$；当$x > 0$时，$1_{(-\infty,0]}(x) = 0$。

单参数指数分布族

二项分布

令N为已知的正整数，如果随机变量$X \sim Bi(N, p)$，参数$\theta = p \in \Theta = (0, 1)$，其质量函数：

$$
\begin{aligned}
P(x|p) &= \binom{N}{x} p^x (1-p)^{N-x} \\
&= \binom{N}{x} \exp\{x \ln(p) + (N-x)\ln(1-p)\} \\
&= \binom{N}{x} \exp\{x[\ln(p) - \ln(1-p)] + N\ln(1-p)\} \\
&= \binom{N}{x} \exp\left\{x\left[\ln\left(\frac{p}{1-p}\right)\right] + N\ln(1-p)\right\}
\end{aligned}
$$

故令$h(x) = \binom{N}{x}$，$\eta(\theta) = \ln\left(\frac{p}{1-p}\right)$，$T(x) = x$，$B(\theta) = -N\ln(1-p)$，所以二项分布属于指数分布族。

单参数指数分布族

泊松分布

泊松分布参数为$\theta = \lambda \in \Theta = (0, \infty)$，其质量函数为:

$$P(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda} = \frac{1}{x!} \exp\left\{x \ln(\lambda) - \lambda\right\}$$

故令$h(x) = \frac{1}{x!}$，$\eta(\theta) = \ln(\lambda)$，$T(x) = x$，$B(\theta) = \lambda$，所以泊松分布属于指数分布族。

单参数指数分布族

指数分布

指数分布的密度函数为:

$$
\begin{aligned}
f(x|\beta) &= \frac{1}{\beta} e^{-\frac{x}{\beta}} \cdot 1_{(0,\infty)}(x) \\
&= 1_{(0,\infty)}(x) \exp\left\{-x \cdot \frac{1}{\beta} - \ln \beta\right\}
\end{aligned}
$$

因而
令$h(x) = 1_{(0,\infty)}(x)$，$\eta(\theta) = -\frac{1}{\beta}$，$T(x) = x$，$B(\theta) = \ln\beta$，故指数分布属于指数分布族。

单参数指数分布族

非指数分布族

若某一密度函数为:

$$f\left(x|\beta\right)=\frac{1}{\beta}\exp\left\{1-\frac{x}{\beta}\right\},x>\beta>0$$

可知$f\left(x|\beta\right)$为密度函数。然而:

$$\begin{aligned} f\left(x|\beta\right)&=\frac{1}{\beta}\exp\left\{1-\frac{x}{\beta}\right\}\cdot 1_{(\beta,\infty)}\left(x\right)\\ &=1_{(\beta,\infty)}\left(x\right)\cdot\exp\left\{-\frac{x}{\beta}-\ln\beta+1\right\}\end{aligned}$$

由于$1_{(\beta,\infty)}\left(x\right)$不仅仅依赖于$x$，而且依赖于$\beta$，因而这一分布不属于指数分布族。

指数分布族的规范形式

指数分布的规范形式

为了方便起见，我们会把密度函数重新参数化，即对于指数分布族

$$f(x|\theta) = h(x) \cdot \exp\{\eta(\theta) \cdot T(x) - B(\theta)\}$$

我们令$\lambda = \eta(\theta)$，那么指数分布族可以写为：

$$f(x|\lambda) = h(x) \cdot \exp\{\lambda \cdot T(x) - C(\lambda)\}$$

并将这种形式成为规范形式（canonical form）。

指数分布族的矩

如果随机变量X的概率密度（质量）函数为规范形式，那么

$$\mathbb{E}[T(X)] = C'(\lambda)$$

$$\mathbb{V}[T(X)] = C''(\lambda)$$

指数分布族矩的证明

对于规范形式$f\left(x|\lambda\right)=h\left(x\right)\cdot\exp\left\{\lambda\cdot T\left(x\right)-C\left(\lambda\right)\right\}$，由于其为密度函数，因而：

$$1=\int_{\mathbb{R}}f\left(x|\lambda\right)dx=\int_{\mathbb{R}}h\left(x\right)\cdot\exp\left\{\lambda\cdot T\left(x\right)-C\left(\lambda\right)\right\}dx$$

从而：

$$e^{C(\lambda)}=\int_{\mathbb{R}}h\left(x\right)\cdot\exp\left\{\lambda\cdot T\left(x\right)\right\}dx$$

两边对λ求导可得：

$$\begin{aligned}e^{C(\lambda)}C'\left(\lambda\right)&=\frac{\partial}{\partial\lambda}\int_{\mathbb{R}}h\left(x\right)\cdot\exp\left\{\lambda\cdot T\left(x\right)\right\}dx\\&=\int_{\mathbb{R}}h\left(x\right)\cdot\frac{\partial}{\partial\lambda}\exp\left\{\lambda\cdot T\left(x\right)\right\}dx\\&=\int_{\mathbb{R}}h\left(x\right)\cdot T\left(x\right)\exp\left\{\lambda\cdot T\left(x\right)\right\}dx\end{aligned}$$

整理可得:

$$C'\left(\lambda\right)=\int_{\mathbb{R}}T\left(x\right)\cdot h\left(x\right)\cdot\exp\left\{\lambda\cdot T\left(x\right)-C\left(\lambda\right)\right\}dx=\mathbb{E}\left[T\left(X\right)\right]$$

指数分布族的规范形式

二项分布的规范形式

对于二项分布：

$$P(x|p) = \begin{pmatrix} N \\ x \end{pmatrix} \exp\left\{ x \left[\ln\left(\frac{p}{1-p} \right) \right] + N \ln(1-p) \right\}$$

如果令$\lambda = \ln\left(\frac{p}{1-p}\right)$，那么

$$P(x|\lambda) = \begin{pmatrix} N \\ x \end{pmatrix} \exp\left\{ \lambda \cdot x - N \ln\left(1 + e^{\lambda}\right) \right\}$$

即为二项分布的质量函数的规范形式。其中，$T(x) = x$，因而

$$\mathbb{E}[T(X)] = \mathbb{E}(X) = C'(\lambda) = \frac{Ne^{\lambda}}{1+e^{\lambda}} = Np$$

$$\mathbb{V}[T(X)] = \mathbb{V}(X) = C''(\lambda) = \frac{Ne^{\lambda}}{(1+e^{\lambda})^2} = Np(1-p)$$

随机变量的变换

- 对于一个可测函数$g(\cdot)\,\mathbb{R}\to\mathbb{R}$，$Y=g(X)$也是一个随机变量。如果我们已知$X$的分布，如何获得新的随机变量$Y$的分布呢?
- 首先仿照随机变量的定义，我们定义函数$g(\cdot)$对于一个集合的逆为:

$$g^{-1}(A)=\{x\in\mathbb{R}:g(x)\in A\}$$

因而对于单点集

$$g^{-1}(\{y\})=\{x\in\mathbb{R}:g(x)=y\}$$

离散随机变量的变换

对于离散型的随机变量X，$g(X)$也是离散型的，因而随机变量Y的p.m.f.为：

$$
\begin{aligned}
f_Y(y) &= P(Y=y) \\
&= \sum_{x \in g^{-1}(\{y\})} P(X=x) \\
&= \sum_{x \in g^{-1}(\{y\})} f_X(x)
\end{aligned}
$$

由此可以确定随机变量$Y=g(X)$的概率质量函数。

离散随机变量的变换

抛骰子

若假设X为一次抛骰子的随机试验的结果，其概率质量函数为:

X	1	2	3	4	5	6
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

如果我们希望得到$Y=g(X)=(X-3.5)^2$的概率质量函数，可以通过以上步骤，比如:

$$g^{-1}(\{6.25\})=\left\{x\in\mathbb{R}:(x-3.5)^2=6.25\right\}=\{1,6\}$$

因而$P(Y=6.25)=P(X\in\{1,6\})=\frac{1}{3}$。以此类推，我们得到$Y$的概率质量函数

Y	6.25	2.25	0.25
$P(X=x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

连续随机变量的变换

对于一个一般的随机变量X，随机变量$Y=g(X)$的累积分布函数可以使用如下定义计算：

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(g(X) \le y) \\
&= P(\{x : g(x) \le y\}) \\
&= \int_{\{x:g(x)\le y\}} dF_X
\end{aligned}
$$

特别的，如果$g(\cdot)$严格单调递增，则：$F_Y(y) = F_X\left(g^{-1}(y)\right)$；
如果$g(\cdot)$严格单调递减，则：$F_Y(y) = 1 - F_X\left(g^{-1}(y)\right)$

连续随机变量的变换

均匀分布的变换

如果$U \sim U(-1,1)$，欲求$Y = U^2$的密度函数，可以首先求其累计分布函数，对于$y \geq 0$，有：

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) = P\left(U^2 \leq y\right) \\
&= P\left(\{-\sqrt{y} \leq U \leq \sqrt{y}\}\right) \\
&= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2} du = \sqrt{y}
\end{aligned}
$$

因而其密度函数为：

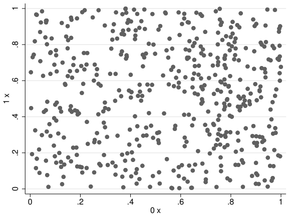
$$
f_Y(y) = \frac{1}{2\sqrt{y}}
$$

均匀分布随机数生成

- 随机数生成是基于模拟（simulation）方法，比如蒙特卡洛法（Monte Carlo）、马尔可夫链蒙特卡洛方法（Markov Chain Monte Carlo, MCMC）等等的基础
- 计算机无法生成真正的随机数，而是根据某些算法（如KISS算法）计算生成的伪随机数（Pseudo random number）
- 几乎所有的编程语言都有生成$(0,1)$上均匀分布的函数：
  - C语言：<stdlib.h>中的rand()函数产生一个0到RAND_MAX的整数，除以RAND_MAX就得到了$(0,1)$上的均匀分布随机数
  - Python：random包中的random()函数产生$(0,1)$均匀分布的随机数
  - Stata用runiform()函数产生$(0,1)$均匀分布的随机数
- seed：产生一系列随机数的一个初始参数，seed确定则随机数序列确定

均匀分布随机数的随机性

以下程序检验了Stata中均匀分布随机数的随机性：
check_random.do



逆变换方法

- 一个自然的问题是，如果给定一个分布函数$F(\cdot)$，如何产生一个分布函数为$F(\cdot)$的随机数？
- 最简单的办法——逆变换方法，即：
    - 先产生一个均匀分布随机变量$U \sim U(0,1)$
    - 令$X = F^{-1}(U) \sim F$
- 细节：如果分布函数$F(\cdot)$存在「平台」，即$F(x) = c,\ for\ a \leq x < b$，那么定义

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}$$

逆变换方法

原理：随机变量$X = F^{-1}(U)$的分布函数为：

$$
\begin{aligned}
F_X(x) &= \int_{F^{-1}(u)\le x} dG(u) \\
&= \int_{F^{-1}(u)\le x} du \\
&= \int_{u\le F(x)} du \\
&= F(x)
\end{aligned}
$$

其中$G(u) = u, 0 \le u \le 1$为均匀分布的分布函数。

逆变换方法

生成指数分布随机数

若$U\sim U\left(0,1\right)$，令$F\left(x\right)=1-e^{-x},x>0$，即指数分布的分布函数。令$X=F^{-1}\left(U\right)$，则随机变量$X$的分布函数为：

$$F_X\left(x\right)=\int_{F^{-1}(u)\leq x}dG\left(u\right)=\int_{F^{-1}(u)\leq x}du=\int_{u\leq(1-e^{-x})}du=1-e^{-x}$$

因而为了生成指数分布的随机变量，只要生成均匀分布U，并令$U=F\left(X\right)$，由于：

$$1-e^{-X}=U\Leftrightarrow X=-\ln\left(1-U\right)$$

从而，为了生成$X\sim F\left(x\right)$，只要令$X=-\ln\left(1-U\right)$即可。
代码：exponential.do

作业

- 2.2, 2.3, 2.4, 2.5, 2.7, 2.8