概率与随机变量

司继春

[1]上海对外经贸大学

2024年9月

概览

样本空间

样本空间的定义

样本空间（sample space）Ω为我们关心的随机试验的所有结果的集合，而Ω中的元素$\omega \in \Omega$称之为**样本点**（sample point）。

1. 随机从一堆扑克牌中抽取一张扑克，其花色的样本空间为：$\Omega_1 = \{\heartsuit, \spadesuit, \clubsuit, \diamondsuit\}$而样本点为$\heartsuit$、$\spadesuit$、$\clubsuit$以及$\diamondsuit$。
2. 某银行一天所接待的所有客户数，其样本空间为$\Omega_2 = \mathbb{N}$，样本点为自然数。
3. 随机从人群中抽取一个人，其身高的样本空间为$\Omega_3 = \mathbb{R}^+$，而样本点为正实数。

事件

- 我们称样本空间Ω的子集（包含Ω本身）为**事件**（event）。
- 如果A为Ω的一个子集，如果随机试验的结果为A中的一个样本点，我们称之为发生了事件A。
- 实际上，概率是定义在**事件**上的，而**非**样本点上的。

事件的运算法则

1. 交换律：$A\cup B=B\cup A$，$A\cap B=B\cap A$
2. 结合律：$A\cup(B\cup C)=(A\cup B)\cup C$，$A\cap(B\cap C)=(A\cap B)\cap C$
3. 分配率：$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$，$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$
4. 德摩根律：$(A\cup B)^c=A^c\cap B^c$，$(A\cap B)^c=A^c\cup B^c$，以及：

$$\left(\bigcup_{i=1}^{\infty}A_i\right)^c=\bigcap_{i=1}^{\infty}A_i^c$$

$$\left(\bigcap_{i=1}^{\infty}A_i\right)^c=\bigcup_{i=1}^{\infty}A_i^c$$

互斥事件

- 如果两个事件A和B满足$A \cap B = \emptyset$，我们称之为**互斥事件**（**disjoint or exclusive**）。
- 如果对于一系列事件$A_1, A_2 \ldots$，对于任意i和j，有$A_i \cap A_j = \emptyset$，则称之为**两两互斥事件**。
- 如果$A_1, A_2 \ldots$为两两互斥事件，且$\cup_{i=0}^{\infty} A_i = \Omega$，则$A_1, A_2 \ldots$为样本空间的一个**划分**（**partition**）。

$\sigma-$代数

对于一个一般的样本空间，有数量巨大的子集。我们希望挑出那些我们需要研究的子集，同时剔除那些性质不是十分良好的子集，这就诞生了$\sigma-$代数的概念。

$\sigma-$代数

如果样本空间Ω的一系列子集的集合$\mathscr{F}$满足：

1. $\emptyset \in \mathscr{F}$
2. 若$A \in \mathscr{F}$，则$A^c \in \mathscr{F}$
3. 若$A_1, A_2 \ldots \in \mathscr{F}$，则$\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$

我们称$\mathscr{F}$为一个$\sigma-$**代数**，或者$\sigma-$**域**。

根据德摩根律，以上第2和第3个条件要求可数个集合的交集也要在$\mathscr{F}$中：

$$\bigcap_{i=1}^{\infty} A_i = \left[\bigcup_{i=1}^{\infty} A_i^c\right]^c \in \mathscr{F}$$

$\sigma-$代数：直觉

为什么需要引入$\sigma-$代数这个概念?

- 限定了概率界定的范围：不是每个Ω的子集都可以被定义上概率（讲义例1.9）
- 与“信息”息息相关，比如，如果两个事件A和B是否发生是已知的，那么：
  - 空集必然不可能发生
  - A和B的补集是否发生也是已知的
  - A和B是否至少有一个发生也是已知的
  - A和B是否同时发生也是已知的

生成的$\sigma-$代数

生成的$\sigma-$代数

对于一系列的事件集合$\mathscr{C}$，我们称包含$\mathscr{C}$中所有事件的最小的$\sigma-$代数为为$\mathscr{C}$生成的$\sigma-$代数，记为$\sigma\langle\mathscr{C}\rangle$。

生成的$\sigma-$代数

如果我们投一枚骰子，那么$\Omega=\{1,2,3,4,5,6\}$，如果我们知道某次投出的结果是一个偶数（$A=\{2,4,6\}$），同时也知道这个数字大于3（$B=\{4,5,6\}$），令$\mathscr{C}=\{A,B\}$，那么：$A^c,B^c\in\sigma\langle\mathscr{C}\rangle$，其次$A,B,A^c,B^c$的任意交、并也属于$\sigma\langle\mathscr{C}\rangle$，从而

$$A\cup B=\{2,4,5,6\}\in\sigma\langle\mathscr{C}\rangle,(A\cup B)^c=\{1,3\}\in\sigma\langle\mathscr{C}\rangle$$

$$A\cap B=\{4,6\}\in\sigma\langle\mathscr{C}\rangle,(A\cap B)^c=\{1,2,3,5\}\in\sigma\langle\mathscr{C}\rangle$$

$$A^c\cup B=\{1,3,4,5,6\}\in\sigma\langle\mathscr{C}\rangle,(A^c\cup B)^c=\{2\}\in\sigma\langle\mathscr{C}\rangle$$

$$A^c\cap B=\{5\}\in\sigma\langle\mathscr{C}\rangle,(A^c\cap B)^c=\{1,2,3,4,6\}\in\sigma\langle\mathscr{C}\rangle$$

以上过程还可以继续。同时，也有一些Ω的子集是肯定不属于$\sigma\langle\mathscr{C}\rangle$的，比如$\{4\}$这个单点集。

Borel $\sigma-$代数

Borel代数

若我们关心的样本空间为$\Omega = \mathbb{R}$，我们令$\mathscr{I}$为所有开区间的集合：

$$\mathscr{I} = \{(a, b) \mid -\infty < a < b < +\infty\}$$

那么包含$\mathscr{I}$的最小$\sigma-$代数：

$$\mathscr{B} = \sigma\langle\mathscr{I}\rangle$$

称为**Borel$\sigma-$代数**或**Borel域**，而$\mathscr{B}$中的元素成为**Borel集**。

注意由于：

$$(a, b] = \bigcap_{i=1}^{\infty}\left(a, b + \frac{1}{i}\right)$$

因而所有的左开右闭区间也都是Borel集。同理可证所有的左闭右开区间$[a, b)$、闭区间$[a, b]$及其可数并、交都为Borel集。

概率的定义

经过以上准备后，可以定义概率：

概率的公理化定义（Axioms of Probability）

（Kolmogorov axioms）给定一个样本空间Ω以及相应的$\sigma-$代数$\mathscr{F}$，函数$\mathscr{P}:\mathscr{F}\to[0,1]$若满足：

1. 对于所有的事件$A\in\mathscr{F}$，$\mathscr{P}(A)\geq 0$
2. $\mathscr{P}(\Omega)=1$
3. 若$A_1,A_2,\ldots\in\mathscr{F}$为两两互斥事件，则$\mathscr{P}\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}\mathscr{P}(A_i)$（可数可加性或可列可加性）

那么我们称$\mathscr{P}$为**概率函数或概率测度**。$(\Omega,\mathscr{F},\mathscr{P})$三元组称为概率空间。

概率的定义

抛硬币

现在我们进行一项抛硬币的随机试验。记正面为H，反面为T，那么我们关心的样本空间为$\Omega = \{H, T\}$。假设硬币质地均匀，即正面和反面的概率相等，那么

$$\mathscr{P}(\{H\}) = \mathscr{P}(\{T\})$$

包含$\{H\}$,$\{T\}$ 的最小σ−代数为

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

根据概率的定义,

$$\mathscr{P}(\Omega) = \mathscr{P}(\{H\} \cup \{T\}) = \mathscr{P}(\{H\}) + \mathscr{P}(\{T\}) = 1$$

因而

$$\mathscr{P}(\{H\}) = \mathscr{P}(\{T\}) = 0.5$$

满足概率的定义

离散样本空间概率的定义

离散样本空间的概率

令$\Omega = \{s_1, s_2, \ldots, s_n\}$为有限集，令$\mathscr{F}$为$\Omega$的所有子集构成的$\sigma-$代数。令$p_1, p_2, \ldots p_n$为非负实数且$\sum_{i=1}^{n} p_i = 1$。对于任意集合$A \in \mathscr{F}$，定义

$$\mathscr{P}(A) = \sum_{\{s_i, i \in A\}} p_i$$

那么$\mathscr{P}$为$\mathscr{F}$上的概率函数。对于可数集$\Omega = \{s_1, s_2, \ldots\}$可类似构建概率函数。

泊松分布

泊松分布的定义

我们关心在一个小时之内到达次数问题，样本空间为$\Omega = \{0, 1, 2 \ldots\}$。取一个非常大的自然数$n$，我们可以把这段时间（标准化为1）分解为等长的$n$段，即$(0, \frac{1}{n}], (\frac{1}{n}, \frac{2}{n}] \ldots (\frac{n-1}{n}, 1]$，当$n$很大时，一个区间段内有两个到达的概率几乎可以忽略不计。假设每段时间到达的概率相等，且反比于n，不妨假设为$\frac{\lambda}{n}$，那么一小时内总的次数：

$$\mathscr{P}^*(\{k\}) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

泊松分布

泊松分布的定义

令$n \to \infty$，则

$$\frac{\binom{n}{k}}{n^k} = \frac{n!}{k! \cdot (n-k)! \cdot n^k} \to \frac{1}{k!}$$

$$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$$

因而

$$\mathscr{P}\left(\{k\}\right) = \frac{\lambda^k}{k!} e^{-\lambda} > 0$$

且$\sum_{k=0}^{\infty} \mathscr{P}\left(\{k\}\right) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1$（$e^{\lambda x}$在$x = 0$处泰勒展开可得），因而根据以上定理，上式定义的概率函数$\mathscr{P}$即定义了样本空间Ω上任意$\sigma-$代数的概率函数。

$\mathbb{R}$上的概率定义

为了在$\mathbb{R}$上定义概率函数，我们首先引入**分布函数**（**distribution function**, d.f.）的概念：

分布函数

如果函数$F:\mathbb{R}\to\mathbb{R}$满足：

1. 单调性：$F(a)\le F(b), a\le b$
2. 右连续：$\lim_{x\to x_0^+}F(x)=F(x_0)$
3. $F(-\infty)=0, F(\infty)=1$

则称F为分布函数。

分布函数

- 令

$$\mathbb{1}\{x \geq t\} = \begin{cases} 0 & x < t \\ 1 & x \geq t \end{cases}$$

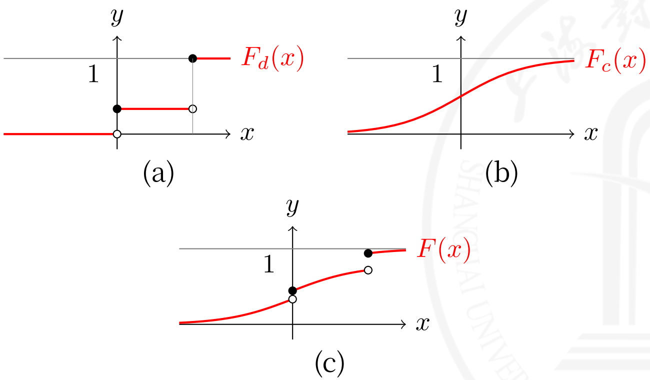
其中$\mathbb{1}\{x \geq t\}$被成为**指示函数**（indicator function），即大括号中的条件成立则为1，否则为0。

- 如果将多个指示函数进行线性组合，即令$g(x) = \sum_j b_j \cdot \mathbb{1}\{x \in A_j\}$，其中$\{A_j\}$为不重不漏的集合，我们就得到了**阶梯函数**（step function）。
- 若$\{a_j\}$为可数集，$b_j > 0, \sum_j b_j = 1$，则$F(x) = \sum_j b_j \cdot \mathbb{1}\{x \geq a_j\}$为分布函数，我们称之为**离散型分布函数**（discrete d.f.）
- 处处连续的分布函数称为**连续型分布函数**（continuous d.f.）

分布函数的分解

每个分布函数都可以写为一个离散型分布函数和一个连续型分布函数的凸组合，且该分解唯一。

分布函数



$\mathbb{R}$上的概率定义

可以使用分布函数定义$\mathbb{R}$上的概率函数。定义:

$$P\left((-\infty,x]\right)=F\left(x\right)$$

则对于任意的$-\infty<a<b<+\infty$，有:

$$\begin{cases} P\left((a,b]\right)=F\left(b\right)-F\left(a\right) \\ P\left((a,b)\right)=F\left(b-\right)-F\left(a\right) \\ P\left([a,b)\right)=F\left(b-\right)-F\left(a-\right) \\ P\left([a,b]\right)=F\left(b\right)-F\left(a-\right) \end{cases}$$

其中$F\left(b-\right)=\lim_{x\to b^-}F\left(x\right)$。据此可以定义所有Borel集的概率，从而产生概率空间$(\mathbb{R},\mathscr{B},P)$。

概率的性质

1. $\mathscr{P}(A) \leq 1$
2. $\mathscr{P}(A^c) = 1 - \mathscr{P}(A)$
3. $\mathscr{P}(\emptyset) = 0$
4. $\mathscr{P}(A \cup B) + \mathscr{P}(A \cap B) = \mathscr{P}(A) + \mathscr{P}(B)$
5. $A \subset B \Rightarrow \mathscr{P}(A) = \mathscr{P}(B) - \mathscr{P}(B \backslash A) \leq \mathscr{P}(B)$
6. $\mathscr{P}(\bigcup_i A_i) \leq \sum_i \mathscr{P}(A_i)$
7. 如果$C_1, C_2, \ldots$为样本空间Ω的一个划分，那么

$$\mathscr{P}(A) = \sum_{i=1}^{\infty} \mathscr{P}(A \cap C_i)$$

条件概率

条件概率

如果A和B为Ω中的两个事件，且$\mathscr{P}(B) > 0$，那么给定B，事件A发生的条件概率为：

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

全概率公式

如果$C_1, C_2, \ldots$为样本空间Ω的一个划分，那么：

$$\mathscr{P}(A) = \sum_{i=1}^{\infty} \mathscr{P}(A|C_i) \cdot P(C_i)$$

特别的，对于任意事件B，有

$$\mathscr{P}(A) = \mathscr{P}(A|B) \cdot \mathscr{P}(B) + \mathscr{P}(A|B^c) \cdot \mathscr{P}(B^c)$$

贝叶斯法则

如果事件A和B都有正的概率，那么我们可以同时定义$\mathscr{P}(A|B)$以及$\mathscr{P}(B|A)$，根据定义：

$$\mathscr{P}(A\cap B)=\mathscr{P}(A|B)\,\mathscr{P}(B)=\mathscr{P}(B|A)\,\mathscr{P}(A)$$

从而：

$$\mathscr{P}(A|B)=\frac{\mathscr{P}(A\cap B)}{\mathscr{P}(B)}=\frac{\mathscr{P}(B|A)\cdot\mathscr{P}(A)}{\mathscr{P}(B)}$$

以上关系式我们称之为**贝叶斯法则（Bayes’ rule）**。

贝叶斯公式

贝叶斯公式

根据全概率公式：

$$\mathscr{P}(A|B)=\frac{\mathscr{P}(B|A)\cdot P(A)}{\mathscr{P}(B)}$$
$$=\frac{\mathscr{P}(B|A)\cdot\mathscr{P}(A)}{\mathscr{P}(B|A)\cdot\mathscr{P}(A)+\mathscr{P}(B|A^c)\cdot\mathscr{P}(A^c)}$$

贝叶斯公式

如果$B_1,B_2,\ldots$为样本空间Ω的一个划分，令$A\in\mathscr{B}$，那么：

$$\mathscr{P}(B_i|A)=\frac{\mathscr{P}(A|B_i)\,\mathscr{P}(B_i)}{\sum_{j=1}^{\infty}\mathscr{P}(A|B_j)\,\mathscr{P}(B_j)}$$

贝叶斯法则

双胞胎

如果在所有双胞胎中，同卵双胞胎的比率是1/3，异卵双胞胎的比率是2/3。生物学理论告诉我们，同卵双胞胎性别一定相同，而异卵双胞胎性别相同的概率为1/2。如果观察到双胞胎的性别相同，那么其为同卵双胞胎的概率为：

$$\mathscr{P}(\text{同卵双胞胎}|\text{性别相同}) = \frac{\mathscr{P}(\text{性别相同}|\text{同卵双胞胎})\,\mathscr{P}(\text{同卵双胞胎})}{\mathscr{P}(\text{性别相同})}$$

而根据全概率公式，在双胞胎中，性别相同的概率：

$$\begin{aligned}\mathscr{P}(\text{性别相同}) &= \mathscr{P}(\text{性别相同}|\text{同卵双胞胎})\,\mathscr{P}(\text{同卵双胞胎}) \\ &+ \mathscr{P}(\text{性别相同}|\text{异卵双胞胎})\,\mathscr{P}(\text{异卵双胞胎}) \\ &= 1\times\frac{1}{3}+\frac{1}{2}\times\frac{2}{3}=\frac{2}{3}\end{aligned}$$

因而如果观察到性别相同，那么该双胞胎为同卵双胞胎的概率为：

$$\mathscr{P}(\text{同卵双胞胎}|\text{性别相同}) = \frac{1\times\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

独立性

统计独立性

如果两个事件A和B满足：

$$\mathscr{P}(A\cap B)=\mathscr{P}(A)\cdot\mathscr{P}(B)$$

那么我们称事件A和B为独立事件。

如果假设$\mathscr{P}(B)\neq 0$，且事件A和事件B独立，那么根据条件概率的定义：

$$\mathscr{P}(A|B)=\frac{\mathscr{P}(A\cap B)}{\mathscr{P}(B)}=\frac{\mathscr{P}(A)\cdot\mathscr{P}(B)}{\mathscr{P}(B)}=\mathscr{P}(A)$$

因而如果条件概率存在，那么事件A和事件B独立意味着两个事件之间不能相互预测。

独立性

三个事件的独立性

现在抛掷两枚骰子，我们关心如下三个事件：

$$A = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),\}$$
$$B = \{\text{两枚骰子之和介于7和10之间}\}$$
$$C = \{\text{两枚骰子之和为2,7或者8}\}$$

可以计算得到：$\mathscr{P}(A) = \frac{1}{6}$，$\mathscr{P}(B) = \frac{1}{2}$，$\mathscr{P}(C) = \frac{1}{3}$，而：

$$\begin{aligned}\mathscr{P}(A \cap B \cap C) &= \mathscr{P}(\{(4,4)\}) \\ &= \frac{1}{36} = \mathscr{P}(A) \cdot \mathscr{P}(B) \cdot \mathscr{P}(C)\end{aligned}$$

然而：$\mathscr{P}(B \cap C) = \frac{11}{36} \neq \mathscr{P}(B) \cdot \mathscr{P}(C)$，因而事件B和C并不独立。

统计独立性

独立性

我们称一系列事件$A_1, A_2, ...A_n$为相互独立的（mutually independent or jointly independent），如果对于任意的子列$A_{i_1}, A_{i_2}, ...A_{i_k}$，有：

$$\mathscr{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathscr{P}\left(A_{i_j}\right)$$

其中$\{i_1, ..., i_k\} \subset \{1, 2, ..., n\}, 1 \leq k \leq n$。

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.）**。

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。

随机变量的函数

- 随机变量X作为一个从Ω到$\mathbb{R}^*$的映射，如果有一个函数$f:\mathbb{R}^*\to\mathbb{R}^*$，那么$f(X)$也是一个从$\Omega$到$\mathbb{R}^*$的映射，从而同样是一个随机变量。
- 有时$f(X)$的取值可能会退化为一个常数，比如在上例中，如果定义$f(x)=|x-0.5|$，那么不管硬币是证明还是反面，$P[f(X)=0.5]=1$，此时$f(X)$就退化成了一个常数。

离散型随机变量

离散型随机变量

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

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$X$。

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

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

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

$$F_X(x) = P_X(X \le x), \forall x \in \mathbb{R}$$

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

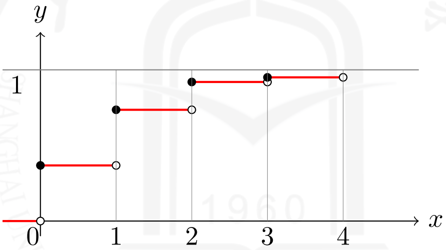
泊松分布的分布函数

在建模一段时间内的到达次数时，经常使用泊松分布。由于自然数集$\mathbb{Z}$为可数集且$P(X \in \mathbb{Z}) = 1$，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：

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



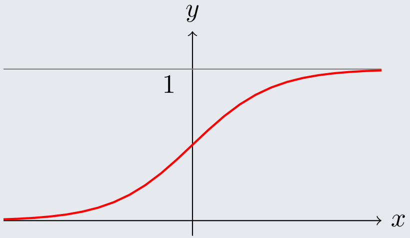
分布函数

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



概率质量函数

概率质量函数

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

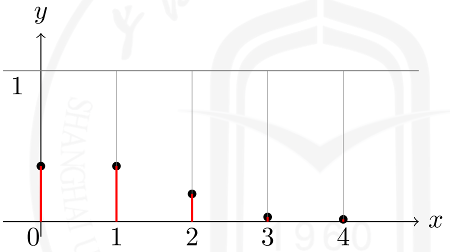
$$f_X(x) = P(X = x), \text{for all } x \in DB$$

泊松分布的质量函数

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:

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



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：

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

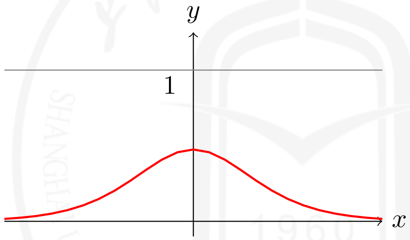
注意如果分布函数$F_X(x)$可导，那么其密度函数即其导函数。

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:

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



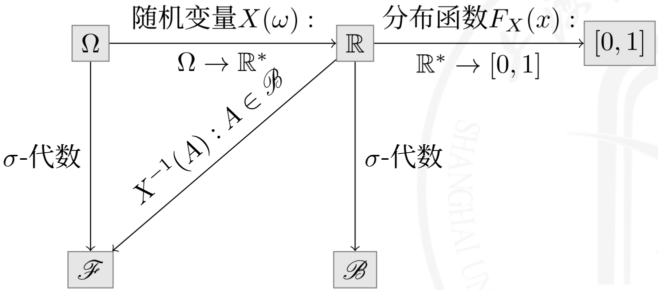
支撑集

- 由于概率质量函数和概率密度函数都可以认为是通过分布函数进行微分（差分）得到的，所以我们接下来在不引起混淆的前提下，会统一称为密度函数。
- 此外，我们还可以定义密度函数的**支撑集**（support）为

$$\mathrm{supp}\,(f) = \{x \in \mathbb{R} : f\,(x) \neq 0\}$$

由于密度函数总是大于等于0的，从而支撑集也就是密度函数严格大于0的部分，从而我们可以将支撑集近似解释为随机变量的取值范围。

随机变量定义总结


Ω
随机变量$X(\omega)$:
$\Omega \to \mathbb{R}^*$
ℝ
分布函数$F_X(x)$:
$\mathbb{R}^* \to [0,1]$
[0, 1]
σ-代数
$X^{-1}(A) : A \in \mathscr{B}$
σ-代数
$\mathscr{F}$
$\mathscr{B}$

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$g(t)$，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：

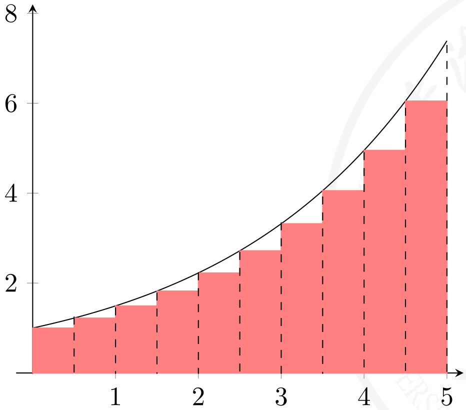
$$I_n = \sum_{k=1}^{n} g(t_{k-1})(t_k - t_{k-1}) = \sum_{k=1}^{n} g(t_{k-1})\Delta_k$$

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

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

即为函数$g(t)$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：$\int_a^b g(t)\,dF(t)$形式的积分，其中$F(t)$单调递增。
- 我们知道，如果$F(x)$可导，那么以上积分等于$\int_a^b g(t)\,F'(t)\,dt$，然而$F(t)$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$\Delta_i F = F(t_i) - F(t_{i-1})$，并定义级数：

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

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

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

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：

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

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

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

- 然而如果$F(t)$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

实际上对于单调递增的阶梯函数$F(x)=\sum_j b_j\cdot\mathbb{1}\{x\geq a_j\}$，其Riemann-Stieltjes积分就等价于求和

$$\int_I g(t)\,dF(t)=\sum_j b_j g(a_j)$$

一个简单的Riemann-Stieltjes积分

假设$F(x)=\mathbb{1}\{x\geq a\}$这是一个不连续函数。如果我们要计算$\int_0^2 x dF(x)$，按照定义，我们可以使用划分$\Pi: 0=t_0<t_1<\cdots<t_n=0$，计算：

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

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

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

如此当$mesh(\Pi)\to 0$时，$t_{k-1}\to a$，从而$\int_0^2 x dF(x)=a$。

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:

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

其中$\lfloor b \rfloor$为小于等于b的最大整数。如此，我们就可以使用以上求和方法计算出$\int_0^b x dF(x)$的值。

Riemann-Stieltjes积分的性质

- 线性性，对于任意实数c_1, c_2，有

$$\int_a^b \left[c_1 g_1(x) + c_2 g_2(x)\right] dF(x) = c_1 \int_a^b g_1(x)\, dF(x) + c_2 \int_a^b g_2(x)\, dF(x)$$

- 如果在$[a,b]$上$g_1(x) \le g_2(x)$，那么

$$\int_a^b g_1(x)\, dF(x) \le \int_a^b g_2(x)\, dF(x)$$

- 可加性，如果$a < c < b$，那么

$$\int_a^c g(x)\, dF(x) + \int_c^b g(x)\, dF(x) = \int_a^b g(x)\, dF(x)$$

- 对于任意非负实数c_1, c_2，有

$$\int_a^b g(x)\, d\left[c_1 F(x) + c_2 G(x)\right] = c_1 \int_a^b g(x)\, dF(x) + c_2 \int_a^b g(x)\, dG(x)$$

- 在一定条件下，积分和微分可以互换：

$$\frac{d}{d\theta} \int_a^b g(x,\theta)\, dF(x) = \int_a^b \frac{dg(x,\theta)}{d\theta} dF(x)$$

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$X \in D$，其数学期望（mathematical expectation）定义为:

$$\mathbb{E}\left(X\right) = \sum_{a \in D} a \cdot P\left(X = a\right)$$

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:

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

离散型随机变量的期望

指示函数的期望

如果令$Y = \mathbb{1}\{X \in A\}$，那么其期望:

$$\begin{aligned}\mathbb{E}\left(\mathbb{1}\{X \in A\}\right) &= 1 \cdot P\left(X \in A\right) + 0 \cdot P\left(X \in A^c\right) \\ &= P\left(X \in A\right)\end{aligned}$$

即指示函数的期望等于该令指示函数等于1的集合的概率。

泊松分布的期望

泊松分布随机变量的期望为:

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

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$g$为一个函数，则期望：

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

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：

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

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

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

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

$$m_r = \mathbb{E}\left(X^r\right)$$

一阶矩m_1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

$$\mathbb{E}\left(\left[X - \mathbb{E}\left(X\right)\right]^r\right)$$

方差和标准差

方差和标准差

二阶中心矩:

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

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

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

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

正态分布的方差

正态分布的方差

为了计算$X \sim \mathrm{N}\left(\mu, \sigma^2\right)$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$\mathbb{V}\left(X\right) = \mathbb{E}\left(X^2\right) - \left[\mathbb{E}\left(X\right)\right]^2$，因而我们只要计算$\mathbb{E}\left(X^2\right)$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。

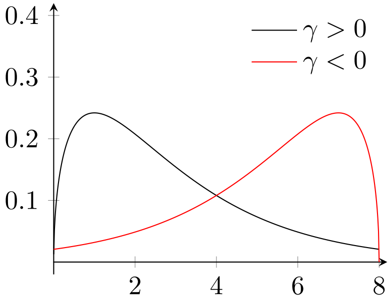
峰度

峰度是随机变量的四阶中心矩，即:

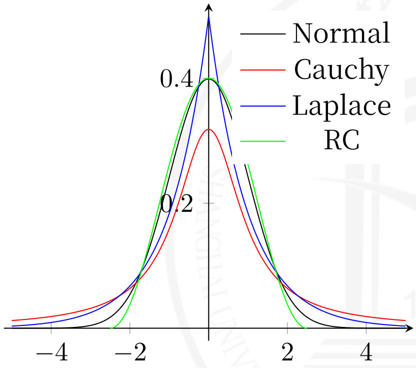
$$\mathrm{Kurt}\left(X\right)=\mathbb{E}\left(\left[\frac{X-\mathbb{E}\left(X\right)}{\sigma\left(X\right)}\right]^4\right)=\frac{\mathbb{E}\left(\left[X-\mathbb{E}\left(X\right)\right]^4\right)}{\left[\mathbb{V}\left(X\right)\right]^2}$$

- 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
$\gamma > 0$
$\gamma < 0$



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

矩母函数

包括期望、方差、偏度和峰度在内的矩是描述随机变量的重要指标，然而其计算往往比较繁琐。而**矩母函数**（**moment-generating function**）则提供了一个方便的计算高阶矩的工具。

- 考虑一个特定的函数$m_X(t) = e^{tX}$在$t = 0$处的泰勒展开：

$$e^{tX} = 1 + Xt + \frac{X^2}{2!}t^2 + \frac{X^3}{3!}t^3 + \frac{X^4}{4!}t^4 + \cdots$$

  由于X是一个随机变量，考虑将其取期望，得到

$$\mathbb{E}\left(e^{tX}\right) = 1 + m_1 t + \frac{m_2}{2!}t^2 + \frac{m_3}{3!}t^3 + \frac{m_4}{4!}t^4 + \cdots$$

  现在，这个期望就是一个关于t的非随机的函数了，我们定义该函数为矩母函数，记为

$$M_X(t) = \mathbb{E}\left(e^{tX}\right)$$

- 矩母函数（如果存在）与分布函数一一对应。

矩母函数

- 如果我们对矩母函数求导数，并计算在$t=0$处的导数值，有

$$\begin{cases} \left.\dfrac{\partial M_X(t)}{\partial t}\right|_{t=0} = m_1 \\ \left.\dfrac{\partial^2 M_X(t)}{\partial t^2}\right|_{t=0} = m_2 \\ \vdots \end{cases}$$

  这也就是矩母函数名字的来历：矩母函数的n阶导数在$t=0$处的取值也就是X的n阶矩。

- 如果我们对随机变量X做一个线性变换，即令$Y=aX+b$，那么

$$M_Y(t) = \mathbb{E}\left(e^{tY}\right) = \mathbb{E}\left(e^{t(aX+b)}\right) = e^{bt}\mathbb{E}\left(e^{atX}\right) = e^{bt}M_X(at)$$

矩母函数

正态分布的矩母函数

对于标准正态分布$Z \sim \mathrm{N}(0,1)$，其矩母函数为

$$
\begin{aligned}
M_Z(t) = \mathbb{E}\left(e^{tZ}\right) &= \int_{\mathbb{R}} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\left(\frac{z^2}{2} - tz\right)} dz = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\left(\frac{z^2-2tz+t^2}{2} - \frac{t^2}{2}\right)} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(z-t)^2}{2}} e^{\frac{t^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}} \int_{\mathbb{R}} e^{-\frac{(z-t)^2}{2}} d(z-t) \\
&= e^{\frac{t^2}{2}}
\end{aligned}
$$

从而对于任意的正态分布$X = \sigma Z + \mu$，矩母函数为$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$。

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：

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

证明

注意到F单调递增，从而

$$\begin{aligned}\mathbb{E}\left[\psi(X)\right] &= \int_{\mathbb{R}} \psi(X)\,dF \\ &\geq \int_{\{|X|\geq u\}} \psi(X)\,dF \\ &\geq \psi(u)\,P(|X|\geq u)\end{aligned}$$

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：

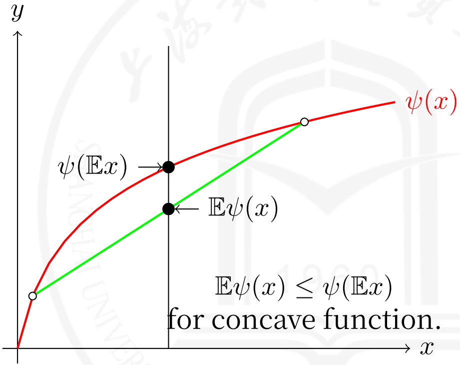
$$P(|Y| \geq u) \leq \frac{1}{u^2}$$

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：

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



Hölder不等式

Hölder不等式

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

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

有：

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

当且仅当存在常数$0 \leq c_1, c_2 < \infty$使得$P\left(c_1\left|X\right|^p = c_2\left|Y\right|^q\right) = 1$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。

作业

- 1.6, 1.8, 1.11, 1.12, 1.13, 1.14