

概率论与数理统计笔记

什么是统计学？

人生，是从不充分的证据开始，引出完美结论的一种艺术。——Samuel Bulter

如果我们不在同一时期，把理解了的科学知识变为我们日常生活的一部分，科学家就不可能提高他们互相拥有的知识。——J.D.Bernal

与人类有关的事实，可以由数量来表示，并且经过大量的积累重复可以导出一般规律。——英国皇家统计学会

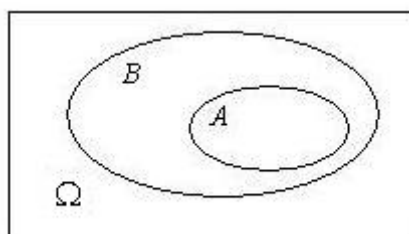
一 事件与概率

1.1 随机试验和随机事件

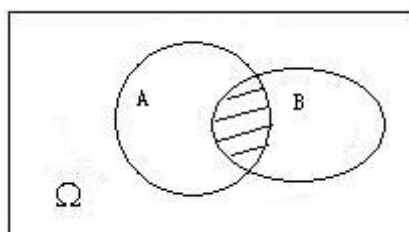
1. **随机现象**：自然界中的客观现象，当人们观测它时，所得结果不能预先确定，而仅仅是多种可能结果之一。
2. **随机试验**：随机现象的实现和对它某个特征的观测。
3. **基本事件**：随机试验中的每个单一结果，犹如分子中的原子，在化学反应中不可再分。
e.g. 硬币抛3次，有8种结果：正正正、正正反、正反正.....这8种可能结果的每一个都是基本事件。
4. **随机事件**：简称事件，在随机试验中所关心的可能出现的各种结果，它由一个或若干个基本事件组成。通常用英文大写字母表示或{一种叙述}来表示。
5. **样本空间**：随机试验中所有基本事件所构成的集合，通常用 Ω 或 S 表示。
e.g. 掷一枚骰子，观察出现的点数，则 $\Omega = \{1, 2, 3, 4, 5, 6\}$ 。
6. **必然事件** (Ω)：在试验中一定会发生的事件。
7. **不可能事件** (ϕ)：在试验中不可能发生的事件。

1.2 事件的运算

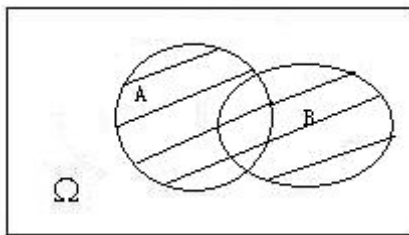
1. **子事件** $A \subset B$ ：事件 A 发生蕴含事件 B 一定发生，则事件 A 成为事件 B 的子事件。若 $A \subset B$ ，且 $B \subset A$ ，则称事件 A 与事件 B 相等，记为 $A = B$ 。



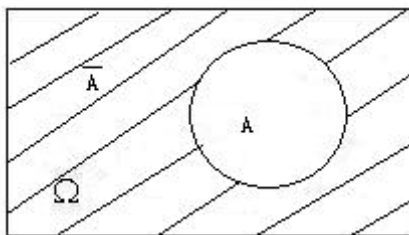
2. **事件的和** ($A \cup B$)：事件 A 和事件 B 中至少有一个发生称为事件 A 和事件 B 的和。



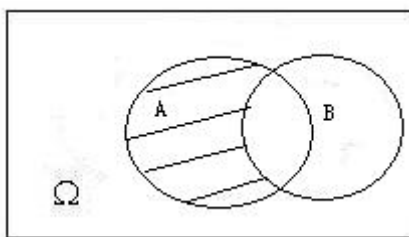
3. **事件的积** ($A \cap B$) : 事件 A 和事件 B 同时发生称为 A 和事件 B 的积。如果 $A \cap B = \phi$, 则称 A 和 B 不相容 , 即事件 A 和 B 不能同时发生。



4. **对立事件** A^c (或 \bar{A}) : A 不发生这一事件称为事件 A 的对立事件 (或余事件) 。



5. **事件 A 和事件 B 的差** ($A - B$) : 事件 A 发生而事件 B 不发生这一事件称为事件 A 和事件 B 的差 , 或等价于 AB^c .



6. De Morgan对偶法则及其推广

$$\overline{A \cup B} = \bar{A} \cap \bar{B},$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

上式可推广到 n 个事件 :

$$\overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \bar{A}_i,$$

$$\overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \bar{A}_i,$$

1.3 概率的定义

概率是随机事件发生可能性大小的数字表征 , 其值在0和1之间 , 即概率是事件的函数。概率有以下定义 :

1.3.1 古典概率

设一个试验有 N 个等可能的结果 , 而事件 E 恰包含其中的 M 个结果 , 则事件 E 的概率 , 记为 $P(E)$, 定义为

$$P(E) = M/N$$

或

$$P(E) = \#(M) / \#(N),$$

其中， $\#(M)$ 为事件 M 中基本事件的个数。

古典概型有**两个条件**：

- 有限性，试验结果只有有限个（记为 n ），
- 等可能性，每个基本事件发生的可能性相同。

注：古典概率可引申出“几何概率”。

1.3.2 概率的统计定义

古典概率的两个条件往往不能满足，但可以将事件的随机试验独立反复做 n 次（Bernoulli试验），设事件 A 发生了 n_A 次，称比值 $\frac{n_A}{n}$ 为事件 A 发生的频率，当 n 越来越大时，频率会在某个值 p 附近波动，且波动越来越小，这个值 p 就定义为事件 A 的概率。该学派为频率派。

注：不能写为 $\lim_{n \rightarrow \infty} \frac{n_A}{n} = p$ ，因为 $\frac{n_A}{n}$ 不是 n 的函数。

1.3.3 主观概率

主观概率可以理解作为一种心态或倾向性。究其根由，大抵有二：一是根据其经验和知识，二是根据其利害关系。该学派在金融和管理有大量的应用，这一学派成为Bayes学派。

1.3.4 概率的公理化定义

对概率运算规定一些简单的基本法则：

1. 设 A 是随机事件，则 $0 \leq P(A) \leq 1$,
2. 设 Ω 为必然事件，则 $P(\Omega) = 1$,
3. 若事件 A 和 B 不相容，则 $P(A \cup B) = P(A) + P(B)$,

可推广至无穷： $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

注：

1. 一般情况下， $P(A \cup B) = P(A) + P(B) - P(AB)$ ，
 $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$
2. $P(\overline{A}) = 1 - P(A)$
3. $P(A - B) = P(A) - P(AB)$

1.4 古典概率计算

1.4.1 排列组合

- **选排列**：从 n 个不同元素中取 r 个不同取法（ $1 \leq r \leq n$ ）， $P_r^n = n(n-1)\dots(n-r+1)$.
- **重复排列**：从 n 个不同元素中可重复地取 r 个不同取法（ $1 \leq r \leq n$ ）， $P_r^n = n^r$.
- **组合**：同选排列，但不考虑次序， $\binom{n}{r} = \frac{P_r^n}{r!}$.

注：

1. 排列英文为Permutation，组合英文为Combination.
2. $0!$ 为1。当 r 不是非负整数时，记号 $r!$ 没有意义.
3. 一些书中将组合写成 C_n^r 或 C_r^n ，更通用的是 $\binom{n}{r}$.

1.4.2 其他公式

- 组合系数 $\binom{n}{r}$ 又常称为二项式系数

$$(a+b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}$$

- n 个相异物件分成 k 堆，各堆物件数分为 r_1, \dots, r_k 的方法是

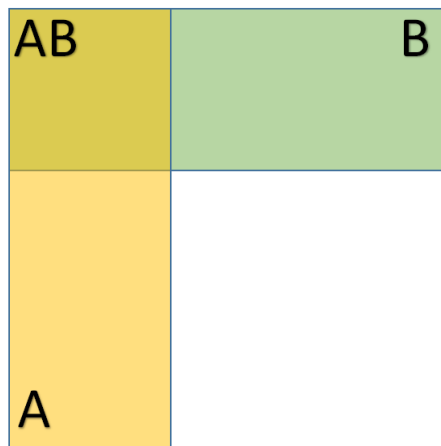
$$n!/(r_1! \dots r_k!).$$

1.5 条件概率

条件概率就是知道了一定信息下得到的随机事件的概率。设事件 A 和 B 是随机试验 Ω 中的两个事件， $P(B) > 0$ ，称

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

为事件 B 发生条件下事件 A 发生的条件概率，可用图形表示：



注：事实上，我们所考虑的概率都是在一定条件下计算的，因为随机试验就是在一定条件下进行的。

1.5.1 条件概率性质

给定 A 发生， $P(A) > 0$ ：

- $0 \leq P(B|A) \leq 1$
- $0 \leq P(\Omega|A) = 1$
- 若 $B_1 \cap B_2 = \phi$ ，则 $P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A)$ ，可推广至无穷。

1.5.2 乘法定理

由 $P(A|B) = \frac{P(AB)}{P(B)} \Rightarrow P(AB) = P(A|B)P(B)$ ，可推广至

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 \dots A_{n-1})$$

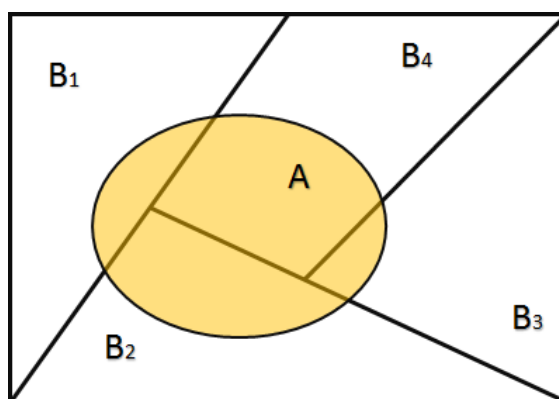
注：右边看似麻烦，其实容易算，左边看似简单，但是难算。

1.6 全概率

设 B_1, B_2, \dots, B_n 是样本空间 Ω 中的两两不相容的一组事件，即 $B_i B_j = \phi$ ， $i \neq j$ ，且满足 $\bigcup_{i=1}^n B_i = \Omega$ ，则称 B_1, B_2, \dots, B_n 是样本空间 Ω 的一个分割（又称为**完备事件群**，英文为 *partition*）。

设 $\{B_1, B_2, \dots, B_n\}$ 是样本空间 Ω 的一个分割, A 为 Ω 的一个事件, 则

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$



推导：

$$\begin{aligned} P(A) &= P(A \cap \Omega) \\ &= P\left(A \cap \sum_{i=1}^n B_i\right) \\ &= P\left(\sum_{i=1}^n AB_i\right) \\ &= \sum_{i=1}^n P(AB_i) \\ &= \sum_{i=1}^n P(A|B_i)P(B_i) \\ &= \sum_{i=1}^n P(A|B_i)P(B_i) \end{aligned}$$

注：有时不易直接计算事件 A 的概率，但是在每个 B_i 上 A 的条件概率容易求出

1.7 Bayes公式

设 $\{B_1, B_2, \dots, B_n\}$ 是样本空间的一个分割, A 为 Ω 中的一个事件, $P(B_i) > 0, i = 1, 2, \dots, n, P(A) > 0$, 则

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

注：当有因果关系互换时必须用Bayes公式。

1.8 事件的独立性

设 A, B 是随机试验中的两个事件, 若满足 $P(AB) = P(A)P(B)$, 则称事件 A 和 B 相互独立。判断事件的独立, 应该是从实际出发, 如果能够判断事件 B 的发生与否对事件 A 的发生与否不产生影响, 则事件 A, B 即为独立。

设 \tilde{A} 表示事件 A 发生和不发生之一, \tilde{B} 表示事件 B 发生和不发生之一。有独立性的定义可推至 $P(\tilde{A}\tilde{B}) = P(\tilde{A})P(\tilde{B})$ (一共有四个等式)。可推广至：

$$P(\tilde{A}_1 \tilde{A}_2 \dots \tilde{A}_n) = P(\tilde{A}_1) \dots P(\tilde{A}_n)$$

上面有 2^n 个等式。

注：独立 (independent) 和不相容 (exclusive) 是不同的两个概念，前者有公共部分，后者没有公共部分，独立一定相容。

1.9 重要公式与结论

$$(1) P(\bar{A}) = 1 - P(A)$$

$$(2) P(A \cup B) = P(A) + P(B) - P(AB)$$

$$(3) P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

$$(4) P(A - B) = P(A) - P(AB)$$

$$(5) P(\overline{AB}) = P(A) - P(AB), P(A) = P(AB) + P(\overline{AB}), \\ P(A \cup B) = P(A) + P(\overline{AB}) = P(AB) + P(\overline{AB}) + P(\overline{AB})$$

$$(6) P(\overline{A_1}|B) = 1 - P(A_1|B), P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 A_2|B) \\ P(A_1 A_2|B) = P(A_1|B)P(A_2|A_1 B)$$

$$(7) \text{ 若 } A_1, A_2, \dots, A_n \text{ 相独立, 则 } P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i), P\left(\bigcup_{i=1}^n A_i\right) = \prod_{i=1}^n (1 - P(A_i))$$

二 随机变量及其分布

2.1 随机变量的概念

1. **随机变量 (Random variable)**：值随机会而定的变量，研究随机试验的一串事件。可按维数分为一维、二维至多维随机变量。按性质可分为**离散型随机变量**以及**连续型随机变量**。
2. **分布 (Distribution)**：事件之间的联系，用来计算概率。
3. **示性函数 (Indication function)**： $I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{反之} \end{cases}$ ，事件A有随机变量 I_A 表示出来， I_A 称为事件A的示性函数。

2.2 离散型随机变量及其分布

1. **离散型随机变量**：设X为一随机变量，如果X只取有限个或可数个值，则称X为一个（一维）离散型随机变量。
2. **概率函数**：设X为一随机变量，其全部可能值为 $\{a_1, a_2, \dots\}$ ，则 $p_i = P(X = a_i), i = 1, 2, \dots$ 称为X的概率函数。
3. **概率分布**：离散型随机变量的概率分布可以用分布表来表示：

可能值	a_1	a_2	...	a_i	...
概率	p_1	p_2	...	p_i	...

4. **概率分布函数**：

◦ **定义**：设X为一随机变量，则函数

$$F(X) = P(X \leq x) \quad (-\infty < x < \infty)$$

称为X的分布函数。（注：这里并未限定X为离散型的，它对任何随机变量都有定义。）

- **性质**：

- $F(x)$ 是单调非降的：当 $x_1 < x_2$ 时，有 $F(x_1) \leq F(x_2)$ 。
- 当 $x \rightarrow \infty$ 时， $F(x) \rightarrow 1$ ；当 $x \rightarrow -\infty$ 时， $F(x) \rightarrow 0$ 。

• **离散型随机变量分布函数：**

对于离散型随机变量，

$$F(x) = P(X \leq x) = \sum_{\{i|a_i \leq x\}} p_i, \quad p_i = P(X = i) = F(i) - F(i-1).$$

5. **二项分布 (Binomial distribution)：**

- **定义：**设某事件 A 在一次试验中发生的概率为 p ，先把试验独立地重复 n 次，以 X 记 A 在这 n 次试验中发生的次数，则 X 取值 $0, 1, \dots, n$ ，且有

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

称 X 服从二项分布，记为 $X \sim B(n, p)$ 。

- **服从二项分布的条件：**1. 各次试验的条件是稳定的，即事件 A 的概率 p 在各次试验中保持不变；2. 各次试验的独立性

6. **泊松分布 (Poisson distribution)：**

- **定义：**设随机变量 X 的概率分布为

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, \quad i = 0, 1, 2, \dots, \quad \lambda > 0$$

则称 X 服从参数为 λ 的 Poisson 分布，并记 $X \sim P(\lambda)$ 。

- **特点：**

- 描述稀有事件发生概率
- 作为二项分布的近似。若 $X \sim B(n, p)$ ，其中 n 很大， p 很小，而 $np = \lambda$ 不太大时（一般 $n > 30, np \leq 5$ ），则 X 的分布接近泊松分布 $P(\lambda)$ 。

推导：

若事件 $A \sim B(n, p)$ ，且 n 很大， p 很小，而 $np = \lambda$ 不太大时，设 $\lambda = np$ ，

$$\begin{aligned} P(X = i) &= \lim_{n \rightarrow \infty} \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \lambda^i \lim_{n \rightarrow \infty} \frac{\binom{n}{i}}{n^i} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \lambda^i e^{-\lambda} \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2)\dots(n-i+1)}{i! n^i} \\ &= \lambda^i e^{-\lambda} \lim_{n \rightarrow \infty} \frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\dots\left(1 - \frac{i-1}{n}\right)}{i!} \\ &= \frac{\lambda^i}{i!} e^{-\lambda} \end{aligned}$$

2.3 连续型随机变量及其分布

1. **连续型随机变量：**设 X 为一随机变量，如果 X 不仅有无限个而且有不可数个数值，则称 X 为一个连续型随机变量。

2. **概率密度函数：**

- **定义：**设连续型随机变量 X 有概率分布函数 $F(x)$ ，则 $F(x)$ 的导数 $f(x) = F'(x)$ 称为 X 的概率密度函数。

- **性质：**

- 对于所有的 $-\infty < x < +\infty$ ，有 $f(x) \geq 0$ ；
- $\int_{-\infty}^{+\infty} f(x) dx = 1$ ；
- 对于任意的 $-\infty < a \leq b < +\infty$ ，有 $P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x) dx$ 。

○ 注：

- 对于任意的 $-\infty < x < +\infty$, 有 $P(X = x) = \int_x^x f(u)du = 0$.
- 假设有总共一个单位的质量连续地分布在 $a \leq x \leq b$ 上 , 那么 $f(x)$ 表示在点 x 的质量密度且 $\int_c^d f(x)dx$ 表示在区间 $[c, d]$ 上的全部质量。

3. 概率分布函数：设 X 为一连续型随机变量，则

$$F(x) = \int_{-\infty}^x f(u)du, \quad -\infty < x < +\infty$$

4. 正态分布 (Normal distribution)：

○ 定义：如果一个随机变量具有概率密度函数

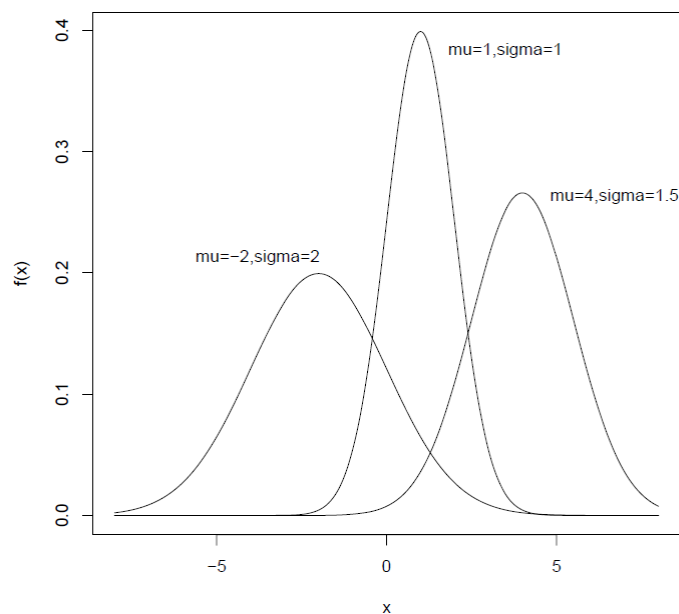
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

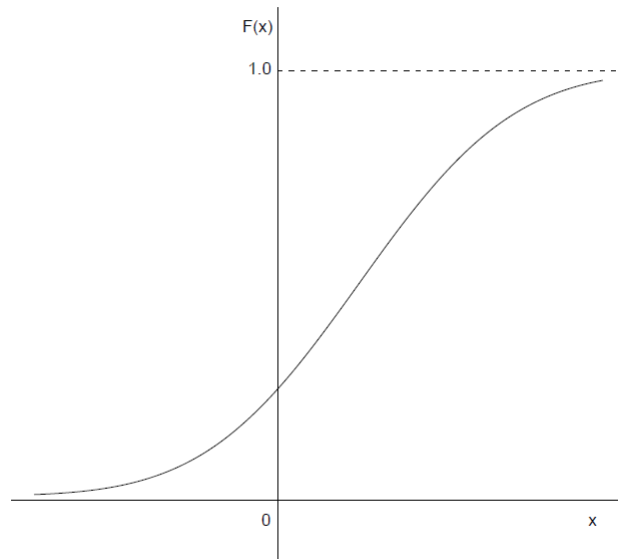
其中 $-\infty < \mu < +\infty$, $\sigma^2 > 0$, 则称 X 为正态随机变量，并记为 $X \sim N(\mu, \sigma^2)$. 特别地， $\mu = 0, \sigma = 1$ 的正态分布成为标准正态分布。用 $\Phi(x)$ 和 $\phi(x)$ 表示标准正态分布 $N(0, 1)$ 的分布函数和密度函数。

○ 性质：

- 正态分布的密度函数是以 $x = \mu$ 为对称轴的对称函数， μ 称为位置参数，密度函数在 $x = \mu$ 处达到最大值，在 $(-\infty, \mu)$ 和 $(\mu, +\infty)$ 内严格单调。
- σ 的大小决定了密度函数的陡峭程度，通常称 σ 为正态分布的形状参数。
- 若 $X \sim N(\mu, \sigma^2)$, 则 $Y = (X - \mu)/\sigma \sim N(0, 1)$.
- $\Phi(-k) = 1 - \Phi(k)$

○ 图像 (密度和分布函数图)：





4. 指数分布 (Exponential distribution) :

- **定义** : 若随机变量 X 具有概率密度函数

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases} = \lambda e^{-\lambda x} I_{(0, \infty)}(x)$$

其中 $\lambda > 0$ 为常数, 则称 X 服从参数为 λ 的指数分布。

- **概率分布函数** : $F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases} = (1 - e^{-\lambda x}) I_{(0, \infty)}(x)$

- **性质** :

- 无后效性, 即无老化, 用来描述寿命 (如元件等) 的分布。

证明 :

“无老化”就是说在时刻 x 正常工作的条件下, 其失效率总保持为某个常数 $\lambda > 0$, 与 x 无关, 可表示

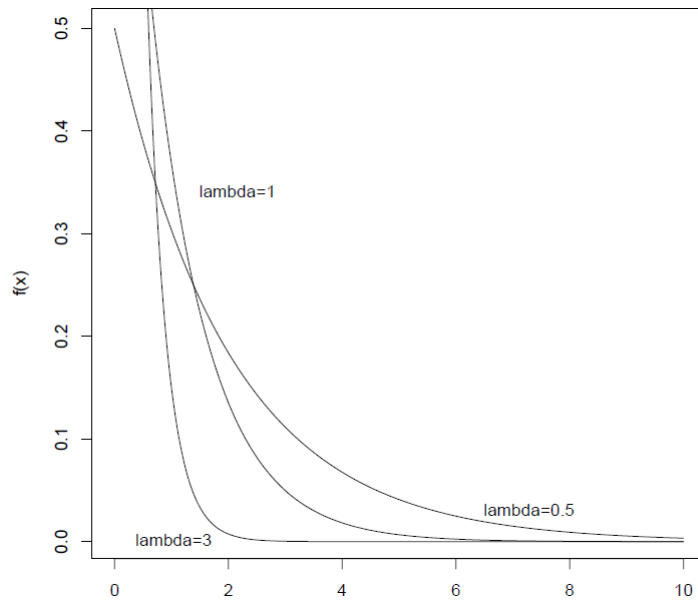
$$P(x \leq X \leq x + h | X > x) / h = \lambda \quad (h \rightarrow 0)$$

证 :

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{P(x \leq X \leq x + h | X > x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{P(x \leq X \leq x + h, X > x)}{P(X > x)h} \\ &= \lim_{h \rightarrow 0} \frac{P(x < X \leq x + h)}{P(X > x)h} \\ &= \lim_{h \rightarrow 0} \frac{-e^{-\lambda t} \Big|_x^{x+h}}{-e^{-\lambda t} \Big|_x^{\infty} h} \\ &= \lim_{h \rightarrow 0} \frac{e^{-\lambda x} - e^{-\lambda x - \lambda h}}{e^{-\lambda x} h} \\ &= \lim_{h \rightarrow 0} \frac{1 - \frac{1}{e^{\lambda h}}}{h} \\ &= \lim_{h \rightarrow 0} \lambda e^{-\lambda h} \\ &= \lambda \end{aligned}$$

- λ 为失效率, 失效率越高, 平均寿命就越小。

- **图像 (密度函数)** :



5. 均匀分布 (Uniform distribution) :

- **定义** : 设 $a < b$, 如果分布 $F(x)$ 具有密度函数

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{其它} \end{cases} = \frac{1}{b-a} I_{(a,b)}(x)$$

则该分布为区间 $[a, b]$ 上的均匀分布。

- **概率分布函数** : $F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ 1 & x > b \end{cases}$
- **性质** : $\forall R(c, d) \subset R(a, b), P(c < X < d) = \frac{d-c}{b-a}$

Author : 钱小z

Email : qz_gis@163.com

Bio : GISer , Spatiotemporal data mining

GitHub : QianXzhen