

# 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 Butler

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

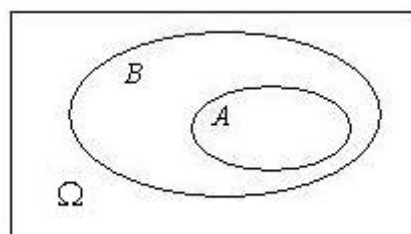
## 一 事件与概率

### 1.1 随机试验和随机事件

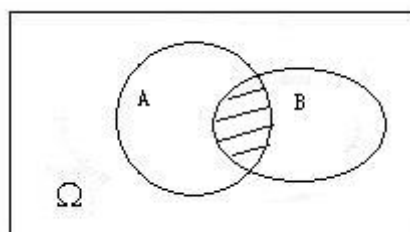
1. **随机现象**：自然界中的客观现象，当人们观测它时，所得结果不能预先确定，而仅仅是多种可能结果之一。
2. **随机试验**：随机现象的实现和对它某个特征的观测。
3. **基本事件**：随机试验中的每个单一结果，犹如分子中的原子，在化学反应中不可再分。  
e.g. 硬币抛3次，有8种结果：正正正、正正反、正反正.....这8种可能结果的每一个都是基本事件。
4. **随机事件**：简称事件，在随机试验中所关心的可能出现的各种结果，它由一个或若干个基本事件组成。通常用英文大写字母表示或{一种叙述}来表示。
5. **样本空间**：随机试验中所有基本事件所构成的集合，通常用 $\Omega$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

### 1.2 事件的运算

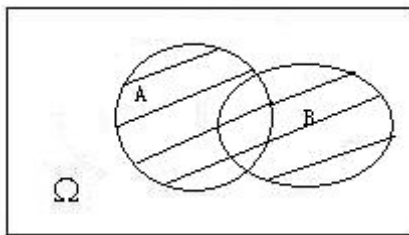
1. **子事件**  $A \subset B$ ：事件 $A$ 发生蕴含事件 $B$ 一定发生，则事件 $A$ 成为事件 $B$ 的子事件。若 $A \subset B$ ，且 $B \subset A$ ，则称事件 $A$ 与事件 $B$ 相等，记为 $A = B$ 。



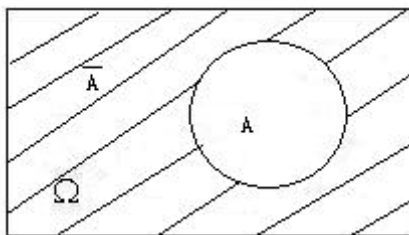
2. **事件的和** ( $A \cup B$ )：事件 $A$ 和事件 $B$ 中至少有一个发生称为事件 $A$ 和事件 $B$ 的和。



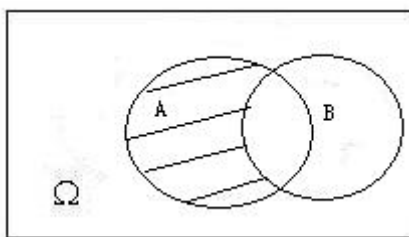
3. **事件的积** ( $A \cap B$ ) : 事件 $A$ 和事件 $B$ 同时发生称为 $A$ 和事件 $B$ 的积。如果 $A \cap B = \phi$  , 则称 $A$ 和 $B$ 不相容 , 即事件 $A$ 和 $B$ 不能同时发生。



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 $\Omega$ 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

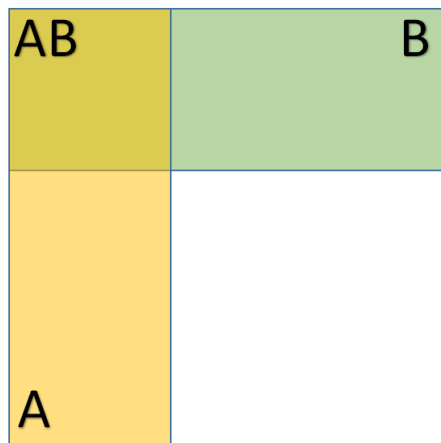
$$n!/(r_1! \dots r_k!).$$

## 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 $\Omega$ 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

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

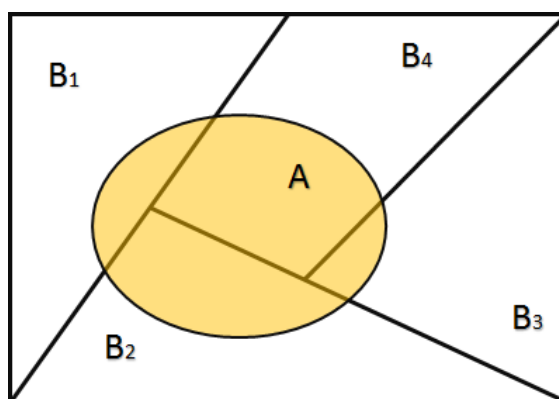
注：右边看似麻烦，其实容易算，左边看似简单，但是难算。

## 1.6 全概率

设 $B_1, B_2, \dots, B_n$ 是样本空间 $\Omega$ 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 $\Omega$ 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 $\Omega$ 的一个分割,  $A$ 为 $\Omega$ 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 $\Omega$ 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

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

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

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

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

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

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

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

## 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  $\lambda$  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

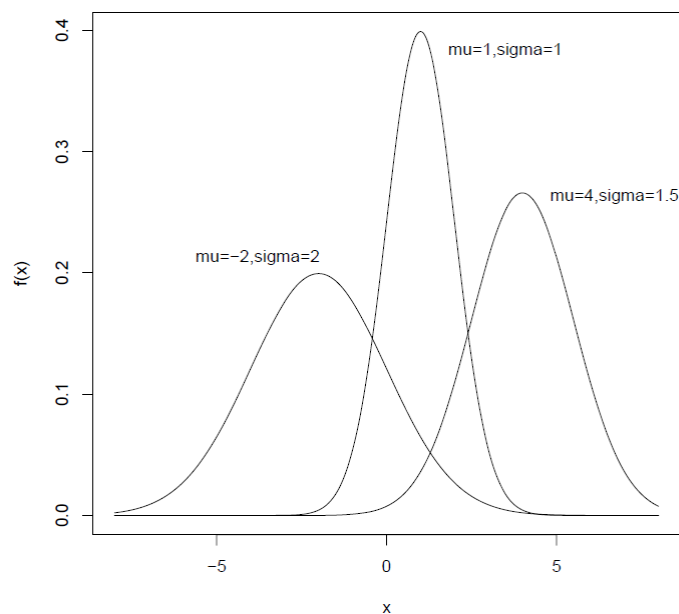
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

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。

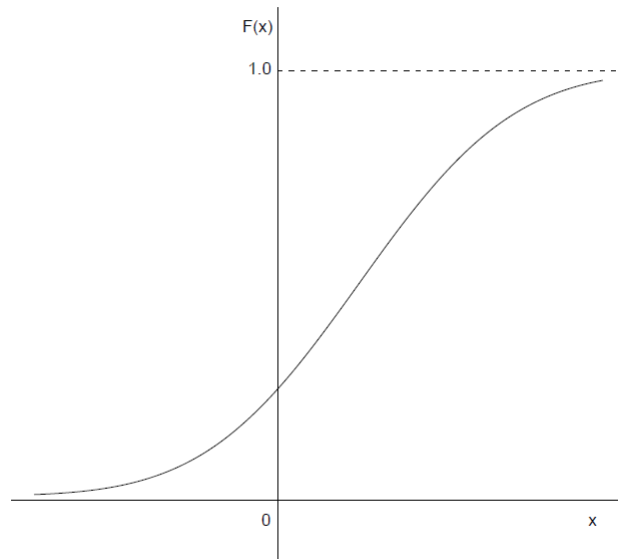
○ 性质：

- 正态分布的密度函数是以  $x = \mu$  为对称轴的对称函数， $\mu$  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。
- $\sigma$  的大小决定了密度函数的陡峭程度，通常称  $\sigma$  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  $\lambda$  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

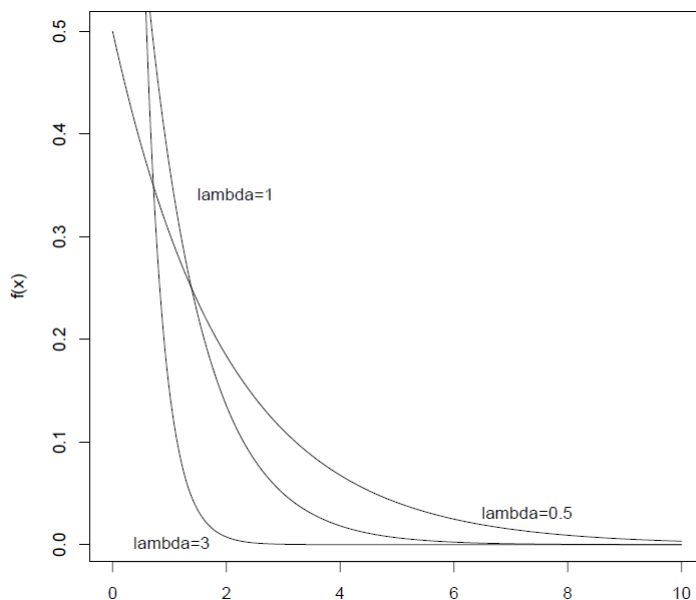
$$P(x \leq X \leq x + h | X > x) / h = \lambda \quad (h \rightarrow 0)$$

证 :

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

- $\lambda$  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： $\forall R(c, d) \subset R(a, b), P(c < X < d) = \frac{d-c}{b-a}$

## 2.4 多维随机变量 ( 随机向量 )

- 随机向量**：设  $X = \{X_1, \dots, X_n\}$ 。如果每个  $X_i$  都是一个随机变量， $i = 1, \dots, n$ ，则称  $X$  为  $n$  维随机变量或者随机向量。
- 离散型随机向量的分布**：如果每一个  $X_i$  都是一个离散型随机变量， $i = 1, \dots, n$ ，则称  $X = \{X_1, \dots, X_n\}$  为一  $n$  维离散型随机变量。设  $X_i$  的所有可能取值为  $\{a_{i1}, a_{i2}, \dots\}$ ， $i = 1, \dots, n$ ，则称

$$p(j_1, \dots, j_n) = P(X_1 = a_{1j_1}, \dots, X_n = a_{nj_n}), \quad j_1, \dots, j_n = 1, 2, \dots$$

为  $n$  维随机变量  $X$  的概率函数，这也是其联合分布。

其具有下列性质：

- $p(j_1, \dots, j_n) \geq 0, \quad j_i = 1, 2, \dots, \quad i = 1, 2, \dots, n;$
- $\sum_{j_1, \dots, j_n} p(j_1, \dots, j_n) = 1.$

**注**：对于高维离散型随机变量，一般不使用分布函数

## 3. 多项式分布

- 定义**：设  $A_1, A_2, \dots, A_n$  是某一试验之下的完备事件群，分别以  $p_1, p_2, \dots, p_n$  记事件  $A_1, A_2, \dots, A_n$  的概率，则  $p_i \geq 0, \quad p_1 + \dots + p_n = 1$ 。将试验独立地重复  $N$  次，以  $X_i$  记在这  $N$  次试验中事件  $A_i$  出现的次数 ( $i = 1, \dots, n$ )，则  $X = (X_1, \dots, X_n)$  为一个  $n$  维随机向量。该分布记作  $M(N; p_1, \dots, p_n)$ 。
  - 概率分布函数**： $P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) = \frac{N!}{k_1! k_2! \dots k_n!} p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}$
- 连续型随机向量的分布**： $X = \{X_1, \dots, X_n\}$  为  $n$  维连续型随机变量，如果存在  $\mathbb{R}^n$  上的非负函数  $f(x_1, \dots, x_n)$ ，使得对任意的  $-\infty < a_1 \leq b_1 < +\infty, \dots, -\infty < a_n \leq b_n < +\infty$ ，有

$$P(a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, \dots, x_n) dx_1 \dots dx_n$$

则称为 $f$ 为 $X$ 的概率密度函数。有

$$P(a_1 \leq X_1 \leq b_1, \dots, a_n \leq X_n \leq b_n) = F(x_1, \dots, x_n)$$

则称为 $F$ 为 $X$ 的（联合）分布函数。其中分布函数 $F(X_1, \dots, X_n)$ 具有下述性质：

- $F(x_1, \dots, x_n)$  单调非降；
  - 对任意的  $1 \leq j \leq n$ ，有  $\lim_{x_j \rightarrow -\infty} F(x_1, \dots, x_n) = 0$ ；
  - $\lim_{x_1 \rightarrow \infty, \dots, x_n \rightarrow \infty} F(x_1, \dots, x_n) = 1$
5. **边缘分布**：因为 $X$ 的每个分量 $X_i$ 都是一维随机变量，故它们都有各自的分布 $F_i$  ( $i = 1, \dots, n$ )，这些都是一维分布，称为随机向量 $X$ 或其分布 $F$ 的边缘分布。

#### ○ 离散型随机向量

$\begin{matrix} X \\ Y \end{matrix}$	$x_1$	$x_2$	$\dots$	$x_n$	行和
$y_1$	$p_{11}$	$p_{21}$	$\dots$	$p_{n1}$	$p_{\cdot 1}$
$y_2$	$p_{12}$	$p_{22}$	$\dots$	$p_{n2}$	$p_{\cdot 2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$y_m$	$p_{1m}$	$p_{2m}$	$\vdots$	$p_{nm}$	$p_{\cdot m}$
列和	$p_{1\cdot}$	$p_{2\cdot}$	$\dots$	$p_{n\cdot}$	1

行和与列和就是边缘分布。即固定某个 $x_i$ ，即可计算边缘分布，故有

$$p_X(x_i) = P(X = x_i) = \sum_j^m P(X = x_i, Y = y_j) = \sum_j^m p_{ij} = p_{i\cdot}, \quad i = 1, 2, \dots, n$$

$$p_Y(y_j) = P(Y = y_j) = \sum_i^m P(X = x_i, Y = y_j) = \sum_i^m p_{ij} = p_{\cdot j}, \quad j = 1, 2, \dots, n$$

#### ○ 连续型随机向量

为求某分量 $X_i$ 的概率密度函数，只需把 $f(x_1, \dots, x_n)$ 中的 $x_i$ 固定，然后对 $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ 在 $-\infty$ 到 $\infty$ 之间做定积分，如

$$(X, Y) \sim f(x, y)$$

$$f_X(u) = \int_{-\infty}^{+\infty} f(u, v) dv$$

$$f_Y(u) = \int_{-\infty}^{+\infty} f(u, v) du$$

**注**：二维正态分布 $N(a, b, \sigma_1^2, \sigma_2^2, \rho)$ 的边缘分布密度分别是一维正态分布 $N(a, \sigma_1^2)$ 和 $N(b, \sigma_2^2)$ 。因此联合分布可推边缘分布，而边缘分布不可推联合分布。

## 2.5 条件分布和随机变量的独立性

1. **离散型随机变量的条件分布**：设 $(X, Y)$ 为二维离散型随机变量，对于给定的事件 $\{Y = y_j\}$ ，其概率 $P(Y = y_j) > 0$ ，则称

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \dots$$

为在给定 $Y = y_j$ 的条件下 $X$ 的条件分布律。类似的，称

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i\cdot}}, \quad j = 1, 2, \dots$$

为在给定  $X = x_j$  的条件下  $Y$  的条件分布律。

2. **连续型随机变量的条件分布**：设  $(X, Y)$  为二维连续型随机变量，对于给定条件  $Y = y$  下的条件概率密度为

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, \quad f_Y(y) > 0.$$

类似的，在  $X = x$  下的条件概率密度为

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \quad f_X(x) > 0.$$

二维正态分布  $\rho = 0$  时，其联合密度分布等于条件密度分布的乘积。

### 3. 随机变量的独立性

称随机变量  $X_1, \dots, X_n$  相互独立，

#### ○ 离散型随机变量

则联合分布律等于各自的边缘分布律的乘积，即

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n)$$

其中  $(x_1, \dots, x_n)$  为  $(X_1, \dots, X_n)$  的值域中的任意一点。

#### ○ 连续型随机变量

则联合密度等于各自的边缘密度的乘积，即

$$f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

#### ○ 更具一般地

设  $X_1, \dots, X_n$  为  $n$  个随机变量，如果它们的联合分布函数等于各自边缘分布函数的乘积，即

$$F(X_1, \dots, x_n) = F_1(x_1) \dots F_n(x_n), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

则称随机变量  $X_1, \dots, X_n$  相互独立。

#### 一些重要的结论

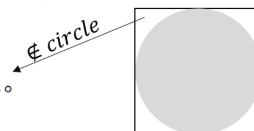
例 2.6.4. 如果随机变量  $X_1, \dots, X_n$  相互独立，则容易证明其中任何一部分随机变量也相互独立。然而一般来说，仅由某一部分独立却无法推出  $X_1, \dots, X_n$  相互独立。如见下例：

例 2.6.5. 若  $\xi, \eta$  相互独立，都服从  $-1$  和  $1$  这两点上的等可能分布，而  $\zeta = \xi\eta$ 。则  $\zeta, \xi, \eta$  两两独立但不相互独立。

例 2.6.6. 设  $(X, Y) \sim N(a, b, \sigma_1^2, \sigma_2^2, \rho)$ ，则  $X$  与  $Y$  相互独立的充要条件是  $\rho = 0$ 。

例 2.6.7. 设  $(X, Y)$  服从矩形  $D = [a, b] \times [c, d]$  上的均匀分布，则  $X$  与  $Y$  相互独立。

例 2.6.8. 设  $(X, Y)$  服从单位圆上的均匀分布，则  $X$  与  $Y$  不独立。



例 2.6.9. 设有  $n$  个事件：  $A_1, A_2, \dots, A_n$ ，对于每个事件  $A_i$ ，定义：  $X_i = I_{A_i}$  ( $A_i$  的示性函数)，  $i = 1, 2, \dots, n$ ，则可证明：  $A_1, A_2, \dots, A_n$  独立  $\iff X_1, X_2, \dots, X_n$  独立。

## 2.6 随机变量的函数的概率分布

最简单的情形，是由一维随机变量 $X$ 的概率分布去求其一给定函数 $Y = g(X)$ 的分布。较为常见的，是由 $(X_1, X_2, \dots, X_n)$ 的分布去求 $Y = g(X_1, X_2, \dots, X_n)$ 的分布。更一般地，由 $(X_1, X_2, \dots, X_n)$ 的分布去求 $(Y_1, Y_2, \dots, Y_m)$ 的分布，其中 $Y_i = g_i(X_1, X_2, \dots, X_n)$ ,  $i = 1, 2, \dots, m$ .

1. 离散型分布的情形：设 $X$ 的分布律为 $P(X = x_i) = p_i$ ,  $i = 1, 2, \dots$

$g: R \rightarrow R$ , 令 $Y = g(X)$ , 则 $Y$ 的分布律为

$$P(Y = y_j) = P(g(X) = y_j) = \sum_{x_i: g(x_i) = y_j} P(X = x_i) = \sum_{i: g(x_i) = y_j} p_i$$

即把 $Y = g(X_1, \dots, X_n)$ 可以取的不同值找出来，把与某个值相应的全部 $(X_1, \dots, X_n)$ 值的概率加起来，即得 $Y$ 取这个值的概率。

2. 连续型分布的情形

### ○ 一个变量的情况

设 $X$ 有密度函数 $f(x)$ . 设 $Y = g(x)$ ,  $g$ 是一个严格单调的函数，即当 $x_1 < x_2$ 时，必有 $g(x_1) < g(x_2)$ 或当 $x_1 > x_2$ 时，必有 $g(x_1) > g(x_2)$ . 又设 $g$ 的导数 $g'$ 存在。由于 $g$ 的严格单调性，其反函数 $X = h(Y)$ 存在，且 $h$ 的导数 $h'$ 也存在。有 $g(X)$ 的密度函数 $l(y)$ 为

$$l(y) = f(h(y))|h'(y)|.$$

### ○ 多个变量的情形

以两个为例，设 $(X_1, X_2)$ 的密度函数 $f(x_1, x_2)$ ,  $Y_1, Y_2$ 都是 $(X_1, X_2)$ 的函数：

$$Y_1 = g_1(X_1, X_2), \quad Y_2 = g_2(X_1, X_2),$$

要求 $(Y_1, Y_2)$ 的概率密度函数 $l(y_1, y_2)$ . 假定 $(X_1, X_2)$ 到 $(Y_1, Y_2)$ 的一一对应变换有逆变换：

$$X_1 = h_1(Y_1, Y_2), \quad X_2 = h_2(Y_1, Y_2)$$

即雅可比行列式

$$J(y_1, y_2) = \begin{vmatrix} \partial h_1 / \partial y_1 & \partial h_1 / \partial y_2 \\ \partial h_2 / \partial y_1 & \partial h_2 / \partial y_2 \end{vmatrix}$$

不为0. 在 $(Y_1, Y_2)$ 的平面上任取一个区域 $A$ ，变换后到 $(X_1, X_2)$ 平面的区域 $B$ ，则有

$$P((Y_1, Y_2) \in A) = P((X_1, X_2) \in B) = \iint_B f(x_1, x_2) dx_1 dx_2$$

$$P((Y_1, Y_2) \in A) = \iint_A f(h_1(y_1, y_2), h_2(y_1, y_2)) |J(y_1, y_2)| dy_1 dy_2$$

### ○ 随机变量和的密度函数

设 $(X_1, X_2)$ 的联合密度函数为 $f(x_1, x_2)$ ,  $Y = X_1 + X_2$ 的密度函数：

- 一般的， $l(y) = \int_{-\infty}^{\infty} f(x_1, y - x_1) dx_1 = \int_{-\infty}^{\infty} f(x, y - x) dx$ .
- 若 $X_1, X_2$ 独立，则 $l(y) = \int_{-\infty}^{\infty} f_1(x) f_2(y - x) dx = \int_{-\infty}^{\infty} f_1(y - x) f_2(x) dx$ .

两个独立的正态变量的和仍服从正态分布，且有关的参数相加，其逆命题也成立。

### ○ 随机变量商的密度函数 设 $(X_1, X_2)$ 的联合密度函数为 $f(x_1, x_2)$ , $Y = X_1/X_2$ 的密度函数：

- 一般的， $l(y) = \int_0^{\infty} x_1 f(x_1, x_1 y) dx_1$ .
- 若 $X_1, X_2$ 独立，则 $l(y) = \int_0^{\infty} x_1 f_1(x_1) f_2(x_1 y) dx_1$ .

### ○ 统计学三大分布

引入两个重要的特殊函数：

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt \quad (x > 0) \text{ 和 } B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt \quad (x > 0, y > 0)$$

其中， $\Gamma(1) = 1$ ,  $\Gamma(1/2) = \sqrt{\pi}$ ,  $\Gamma(n) = (n-1)!$

$$B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$$

■ 卡方分布，记作 $\chi_n^2$

密度函数： $k_n(x) = \frac{1}{\Gamma(\frac{n}{2}2^{n/2})} e^{-x/2} x^{(n-2)/2} I_{(0,\infty)}(x)$

性质：1. 设 $X_1, X_2$ 独立， $X_1 \sim \chi_m^2, X_2 \sim \chi_n^2$ ，则 $X_1 + X_2 \sim \chi_{m+n}^2$

2. 若 $X_1, \dots, X_n$ 独立，且都服从指数分布，则 $X = 2\lambda(X_1 + \dots + X_n) \sim \chi_{2n}^2$

■  $t$ 分布，记作 $t_n$

设 $X_1, X_2$ 独立， $X_1 \sim \chi_n^2, X_2 \sim N(0, 1)$ ，而 $Y = X_2 / \sqrt{X_1/n}$ ，则 $Y \sim t_n$ 。

密度函数： $t_n(y) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} (1 + \frac{y^2}{n})^{-(\frac{n+1}{2})}$

性质：密度函数关于原点对称，其图形与正态分布 $N(0, 1)$ 的密度函数的图形相似。

■  $F$ 分布，记作 $F_{mn}$

设 $X_1, X_2$ 独立， $X_1 \sim \chi_n^2, X_2 \sim \chi_m^2$ ，而 $Y = m^{-1}X_2 / (n^{-1}X_1)$ ，则 $Y \sim F_{mn}$

密度函数： $f_{mn}(y) = m^{m/2} n^{n/2} \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} y^{m/2-1} (my + n)^{-(m+n)/2} \quad (y > 0)$

### 三大分布的几个重要性质

1. 设 $X_1, \dots, X_n$ 独立同分布，有公共的正态分布 $N(\mu, \sigma^2)$ .记

$\bar{X} = (X_1 + \dots + X_n) / n, S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$ . 则

$(n-1)S^2 / \sigma^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / \sigma^2 \sim \chi_{n-1}^2$ .

2. 设 $X_1, \dots, X_n$ 的假定同1，则 $\sqrt{n}(\bar{X} - \mu) / S \sim t_{n-1}$

3. 设 $X_1, \dots, X_n, Y_1, \dots, Y_m$ 独立， $X_i$ 各有分布 $N(\mu_1, \sigma_1^2)$ ， $Y_j$ 各有分布 $N(\mu_2, \sigma_2^2)$ ，则

\$\$

$$\frac{[\sum_{j=1}^m (Y_j - \bar{Y})^2 / (\sigma_2^2 (m-1))]}{[\sum_{i=1}^n (X_i - \bar{X})^2 / (\sigma_1^2 (n-1))]} \sim F_{m-1, n-1}$$

\$\$

$$\frac{[\sum_{j=1}^m (Y_j - \bar{Y})^2 / (\sigma_2^2 (m-1))]}{[\sum_{i=1}^n (X_i - \bar{X})^2 / (\sigma_1^2 (n-1))]} \sim F_{m-1, n-1}$$

若 $\sigma_1^2 = \sigma_2^2$ ，则

$$\sqrt{\frac{nm(n+m-2)}{n+m}} [(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)] / \left[ \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2 \right]^{1/2} \sim t_{n+m-2}$$

Author：钱小z

Email：[qz\\_gis@163.com](mailto:qz_gis@163.com)

Bio：GISer，Spatiotemporal data mining

GitHub：QianXzhen