

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

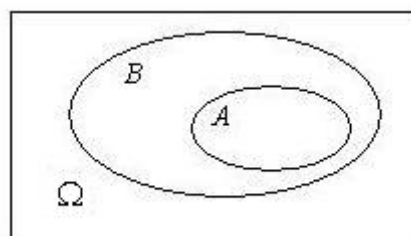
一 事件与概率

1.1 随机试验和随机事件

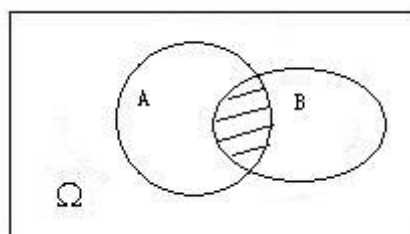
1. **随机现象**：自然界中的客观现象，当人们观测它时，所得结果不能预先确定，而仅仅是多种可能结果之一。
2. **随机试验**：随机现象的实现和对它某个特征的观测。
3. **基本事件**：随机试验中的每个单一结果，犹如分子中的原子，在化学反应中不可再分。
e.g. 硬币抛3次，有8种结果：正正正、正正反、正反正.....这8种可能结果的每一个都是基本事件。
4. **随机事件**：简称事件，在随机试验中所关心的可能出现的各种结果，它由一个或若干个基本事件组成。通常用英文大写字母表示或{一种叙述}来表示。
5. **样本空间**：随机试验中所有基本事件所构成的集合，通常用 Ω 或 S 表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

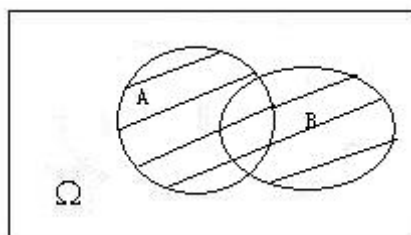
1. **子事件** $A \subset B$ ：事件 A 发生蕴含事件 B 一定发生，则事件 A 成为事件 B 的子事件。若 $A \subset B$ ，且 $B \subset A$ ，则称事件 A 与事件 B 相等，记为 $A = B$ 。



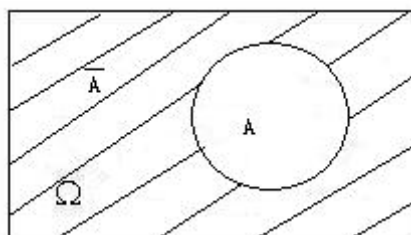
2. **事件的和** ($A \cup B$)：事件 A 和事件 B 中至少有一个发生称为事件 A 和事件 B 的和。



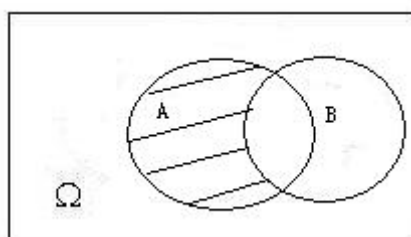
3. **事件的积** ($A \cap B$) : 事件 A 和事件 B 同时发生称为 A 和事件 B 的积。如果 $A \cap B = \phi$, 则称 A 和 B 不相容 , 即事件 A 和 B 不能同时发生。



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

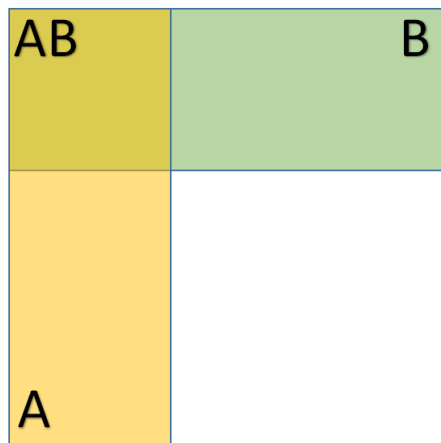
$$n!/(r_1! \dots r_k!).$$

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

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

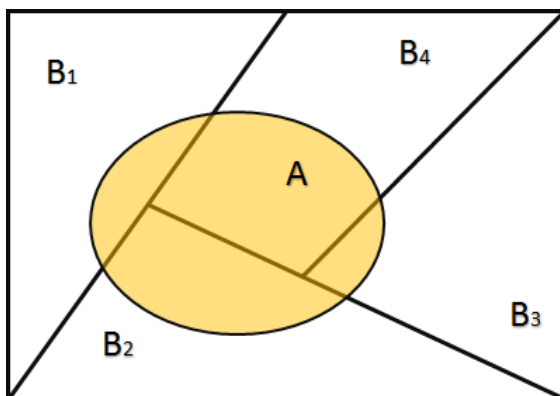
注：右边看似麻烦，其实容易算，左边看似简单，但是难算。

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

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

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

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

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

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

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

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

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 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 $\lambda = np$ ，

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

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

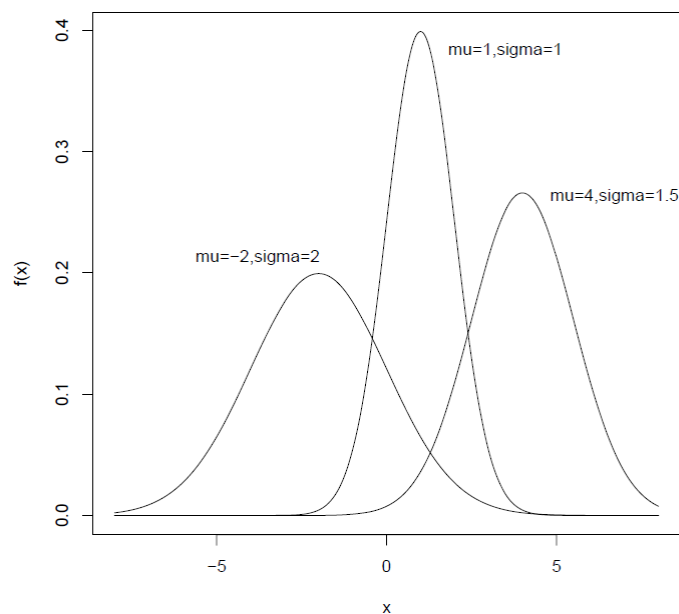
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

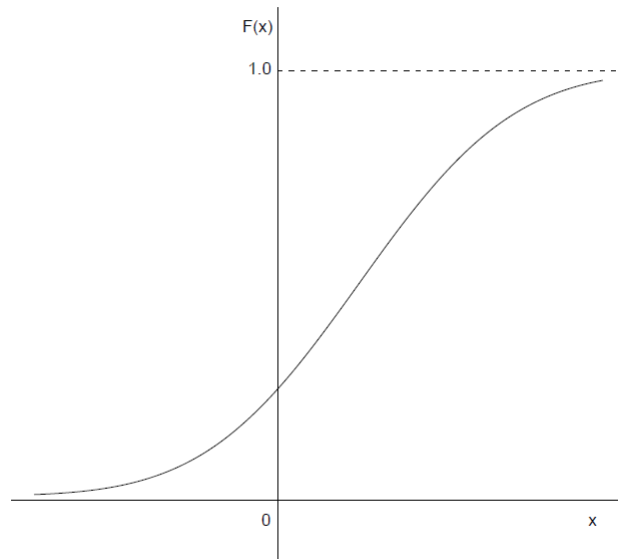
其中 $-\infty < \mu < +\infty$, $\sigma^2 > 0$ ，则称 X 为正态随机变量，并记为 $X \sim N(\mu, \sigma^2)$ 。特别地， $\mu = 0, \sigma = 1$ 的正态分布成为标准正态分布。用 $\Phi(x)$ 和 $\phi(x)$ 表示标准正态分布 $N(0, 1)$ 的分布函数和密度函数。

○ 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

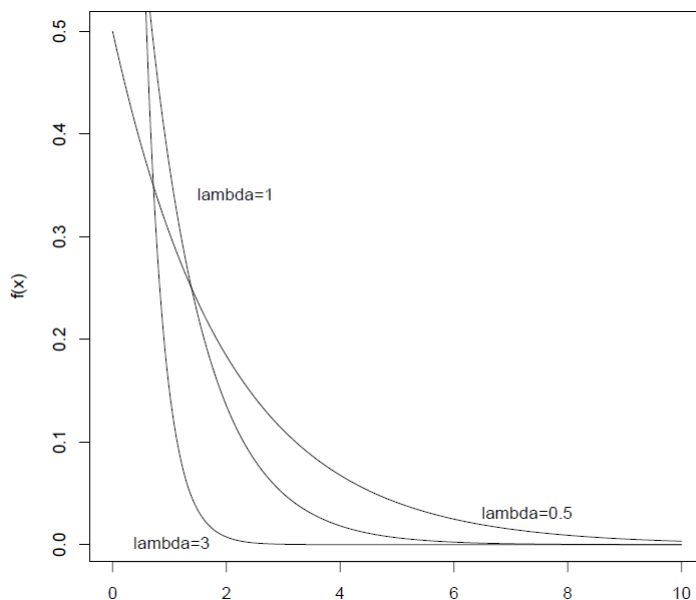
$$P(x \leq X \leq x + h | X > x) / h = \lambda \quad (h \rightarrow 0)$$

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

2.4 多维随机变量 (随机向量)

1. **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。
2. **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** : $P(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) = \frac{N!}{k_1! k_2! \dots k_n!} p_1^{k_1} p_2^{k_2} \dots p_n^{k_n}$
4. **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

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

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

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

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

○ 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 ∞ 之间做定积分，如

$$\begin{aligned} (X, Y) &\sim f(x, y) \\ f_X(u) &= \int_{-\infty}^{+\infty} f(u, v) dv \\ f_Y(u) &= \int_{-\infty}^{+\infty} f(u, v) du \end{aligned}$$

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

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

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

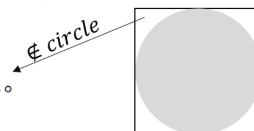
例 2.6.4. 如果随机变量 X_1, \dots, X_n 相互独立，则容易证明其中任何一部分随机变量也相互独立。然而一般来说，仅由某一部分独立却无法推出 X_1, \dots, X_n 相互独立。如见下例：

例 2.6.5. 若 ξ, η 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 ζ, ξ, η 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 χ_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

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

若 $\sigma_1^2 = \sigma_2^2$ ，则

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

三 随机变量的数字特征

3.1 数学期望（均值）与中位数

1. 数学期望

○ 定义：设随机变量 X 只取有限个可能值 a_1, \dots, a_m ，其概率分布为

$P(X = a_i) = p_i \quad (i = 1, \dots, m)$. 则 X 的数学期望记作 $E(X)^*$ 或 $E(X)$ ，定义为

$E(X) = a_1p_1 + a_2p_2 + \dots + a_mp_m$. 数学期望也常称为“均值”，即指以概率为权的加权平均。

○ 离散型变量的数学期望： $E(X) = \sum_{i=1}^{\infty} a_i p_i$. (当级数绝对收敛，即 $\sum_{i=1}^{\infty} |a_i| p_i < \infty$)

○ 连续型变量的数学期望： $E(X) = \int_{-\infty}^{\infty} x f(x) dx$. (当 $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$)

○ 常见分布的数学期望：

- 泊松分布： $E(X) = \lambda$.
- 二项分布： $E(X) = np$.
- 均匀分布： $E(X) = \frac{1}{2}(a+b)$.
- 指数分布： $E(X) = \lambda^{-1}$.
- 正态分布： $E(X) = \mu$.
- 卡方分布： $E(X) = n$.

- **t分布** : $E(X) = 0 \quad (n > 1).$
- **F分布** : $E(X) = n/(n-2) \quad (n > 2).$

○ **性质** :

- 若干个随机变量之和的期望等于各变量的期望值和, 即

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n).$$

- 若干个**独立**随机变量之积的期望等于各变量的期望之积, 即

$$E(X_1 X_2 \dots X_n) = E(X_1) E(X_2) \dots E(X_n).$$

- 设随机变量 X 为离散型, 有分布 $P(X = a_i) = p_i (i = 1, 2, \dots)$; 或者为连续型, 有概率密度函数 $f(x)$. 则

$$E(g(x)) = \sum_i g(a_i) p_i \quad (\text{当 } \sum_i |g(a_i)| p_i < \infty \text{ 时})$$

或

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (\text{当 } \int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty \text{ 时})$$

- 若 c 为常数, 则 $E(cX) = cE(X).$

2. 条件数学期望

- **定义** : 随机变量 Y 的条件期望就是它在给定的某种附加条件下的数学期望。

$E(Y|x) = \int_{-\infty}^{\infty} y f(y|x) dy$. 它反映了随着 X 取值 x 的变化 Y 的平均变化的情况如何。在统计上, 常把条件期望 $E(Y|x)$ 作为 x 的函数, 称为 Y 对 X 的回归函数。

○ **性质** :

- $E(Y) = \int_{-\infty}^{\infty} E(Y|x) f_X(x) dx.$
- $E(Y) = E[E(Y|X)].$

3. 中位数

- **定义** : 设连续型随机变量 X 的分布函数为 $F(x)$, 则满足条件 $P(X \leq m) = F(m) = 1/2$ 的数 m 称为 X 或分布 F 的中位数。即 m 这个点把 X 的分布从概率上一切两半。

○ **性质** :

- 与期望值相比, 中位数受特大值或特小值影响很小, 而期望不然。
- 中位数可能不唯一, 且在某些离散型情况下, 中位数不能达到一分两半的效果。

3.2 方差与矩

1. 方差与标准差

- **定义** : 设 X 为随机变量, 分布为 F , 则 $Var(X) = E(X - EX)^2$ 称为 X (或分布 F) 的方差, 其平方根 $\sqrt{Var(X)}$ (取正值) 称为 X (或分布 F) 的标准差。

○ **常见分布的方差** :

- **泊松分布** : $Var(X) = \lambda.$
- **二项分布** : $Var(X) = np(1-p).$
- **正态分布** : $Var(X) = \sigma^2.$
- **指数分布** : $Var(X) = 1/\lambda^2.$
- **均匀分布** : $Var(X) = (b-a)^2/12.$
- **卡方分布** : $Var(X) = 2n.$
- **t分布** : $Var(X) = n/(n-2).$
- **F分布** : $Var(X) = 2n^2(m+n-2)/[m(n-2)^2(n-4)] \quad (n > 4).$

○ **性质** :

- $Var(X) = E(X^2) - (EX)^2.$
- 常数的方差为 0, 即 $Var(c) = 0.$
- 若 c 为常数, 则 $Var(X+c) = Var(X).$

- 若 c 为常数, 则 $Var(cX) = c^2 Var(X)$.
- **独立**随机变量和的方差等于各变量方差和, 即 $Var(X_1 + \dots + X_n) = Var(X_1) + \dots + Var(X_n)$.

2. 矩

- **定义**: 设 X 为随机变量, c 为常数, k 为正整数。则量 $E[(X - c)^k]$ 称为 X 关于 c 点的 k 阶矩。特别地, 有两种重要的情况:

(1) $c = 0$. 这时 $a_k = E(X^k)$ 称为 X 的 k 阶原点矩。

(2) $c = E(X)$. 这时 $\mu_k = E[(X - EX)^k]$ 称为 X 的 k 阶中心矩。

一阶原点矩就是期望, 一阶中心距 $\mu_1 = 0$, 二阶中心距 μ_2 就是 X 的方差 $Var(X)$ 。

- **两种重要应用**:

- **偏度系数**: $\beta_1 = \mu_3 / \mu_2^{3/2}$. 衡量概率分布函数 $f(x)$ 是否关于均值对称。如果 $\beta > 0$, 则称分布为正偏或右偏; 如果 $\beta < 0$, 则称分布为负偏或左偏; 如果 $\beta = 0$, 则对称。
(注: $\mu_2^{3/2}$ 为标准差的三次方, 可将 μ_3 缩放到一次因次)
- **峰度系数**: $\beta_2 = \mu_4 / \mu_2^2$. 衡量概率分布函数 $f(x)$ 在均值附近的陡峭程度。若 X 有正态分布 $N(\mu, \sigma^2)$, 则 $\beta_2 = 3$ 。(注: μ_2^2 为标准差的四次方, 将 μ_4 缩放到一次因次。为了迁就正态分布, 也常定义 $\mu_4 / \mu_2^2 - 3$ 为峰度系数, 以使正态分布的峰度系数为0)

3.3协方差与相关系数

两者都反映了随机变量之间的关系。

1. 协方差 (Covariance)

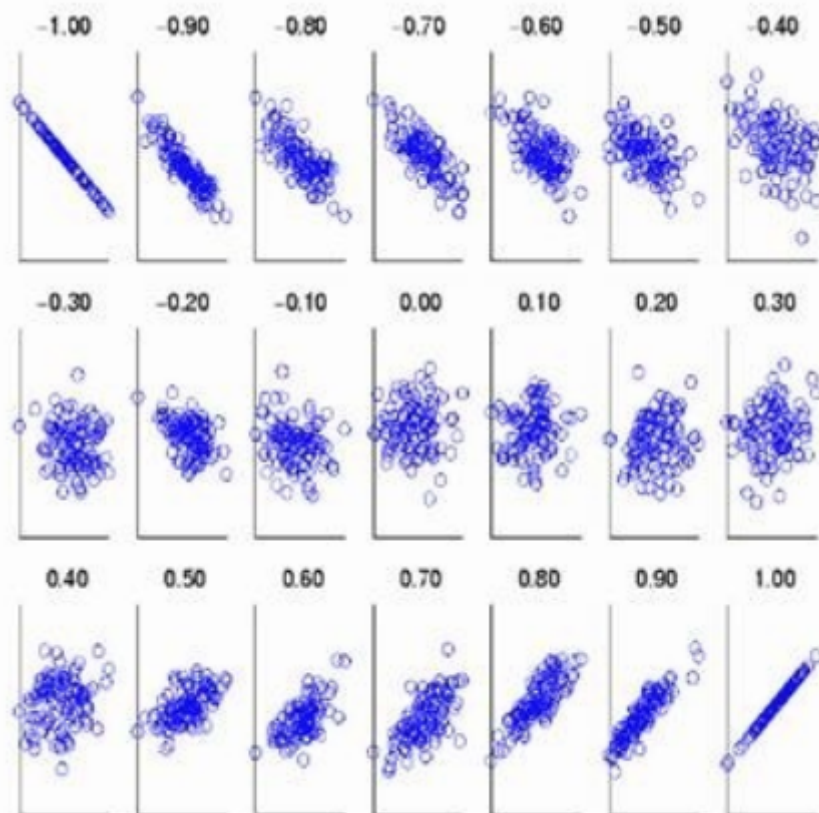
- **定义**: 称 $E[(X - m_1)(Y - m_2)]$ 为 X, Y 的协方差, 并记为 $Cov(X, Y)$.
- **性质**:
 - $Cov(X, Y)$ 与 X, Y 的次序无关, 即 $Cov(X, Y) = Cov(Y, X)$.
 - $Cov(c_1 X + c_2, c_3 Y + c_4) = c_1 c_3 Cov(X, Y)$.
 - $Cov(X, Y) = E(XY) - E(X)E(Y)$.
 - 若 X, Y 独立, 则 $Cov(X, Y) = 0$.
 - $[Cov(X, Y)]^2 \leq \sigma_1^2 \sigma_2^2$. 等号当且仅当 X, Y 之间有严格线性关系 ($Y = a + bX$) 时成立。

注: 协方差的结果受随机变量量纲影响。

2. 相关系数 (Correlation coefficient)

- **定义**: 称 $Cov(X, Y) / (\sigma_1 \sigma_2)$ 为 X, Y 的相关系数, 并记为 $Corr(X, Y)$.
- **性质**:
 - 若 X, Y 独立, 则 $Corr(X, Y) = 0$.
 - $-1 \leq Corr(X, Y) \leq 1$, 或 $|Corr(X, Y)| \leq 1$, 等号当且仅当 X 和 Y 有严格线性关系时达到。当 $Corr(X, Y) = 0$ 时, 推出 X, Y 非线性相关。

注: 相关系数常称为“**线性相关系数**”, 实际上相关系数并不是刻画了 X, Y 之间消除量纲后“一般”关系的程度, 而只是“线性关系的程度”。即使 X 与 Y 有某种严格的函数关系但非线性关系, $|Corr(X, Y)|$ 不仅不必为1, 还可以为0。



3.4大数定理和中心极限定理

1. 大数定理

“大数”的意思，就是指涉及大量数目的观察值 X_i ，它表明这种定理中指出现象只有在大量次数的试验和观察之下才能成立。

- **定义：**设 $X_1, X_2, \dots, X_n, \dots$ 是独立同分布的随机变量，记它们的公共均值为 a 。又设它们的方差存在并记为 σ^2 。则对任意给定的 $\varepsilon > 0$ ，有 $\lim_{n \rightarrow \infty} P(|\bar{X}_n - a| \geq \varepsilon) = 0$ 。（该式表明，当 n 很大时， \bar{X}_n 接近 a ）

2. 中心极限定理

即和的分布收敛于正态分布。

- **定义：**设 X_1, X_2, \dots, X_n 为独立同分布的随机变量， $E(X_i) = a, Var(X_i) = \sigma^2 (0 < \sigma^2 < \infty)$ 。则对任何实数 x ，有 $\lim_{n \rightarrow \infty} P\left(\frac{1}{\sqrt{ns}}(X_1 + \dots + X_n - na) \leq x\right) = \Phi(x)$ 。（ $\Phi(x)$ 为标准正态分布 $N(0, 1)$ 的分布函数）
- **特例：**设 X_1, X_2, \dots, X_n 独立同分布， X_i 分布是 $P(X_i = 1) = p, P(X_i = 0) = 1 - p (0 < p < 1)$ 。则对任何实数 x ，有 $\lim_{n \rightarrow \infty} P\left(\frac{1}{\sqrt{np(1-p)}}(X_1 + \dots + X_n - np) \leq x\right) = \Phi(x)$ 。

注：如果 t_1, t_2 是两个正整数， $t_1 < t_2$ 。则当 n 相当大时，近似地有

$$P(t_1 \leq X_1 + \dots + X_n \leq t_2) \approx \Phi(y_2) - \Phi(y_1),$$

其中

$$y_i = (t_i - np) / \sqrt{np(1-p)} \quad (i = 1, 2).$$

若把 y_1, y_2 修正为

$$y_1 = (t_1 - \frac{1}{2} - np) / \sqrt{np(1-p)},$$

$$y_2 = (t_2 - \frac{1}{2} - np) / \sqrt{np(1-p)}$$

在应用上式，则一般可提高精度。

四 统计量及其分布

4.1 总体与样本

1. 总体

在一个统计问题里，研究对象的全体叫做总体，构成总体的每个成员称为个体。根据个体的数量指标数量，定义总体的维度，如每个个体只有一个数量指标，总体就是一维的，同理，个体有两个数量指标，总体就是二维的。**总体就是一个分布，数量指标就是服从这个分布的随机变量。**

总体根据个体数分为**有限总体**和**无限总体**，当有限总体的个体数充分大时，其可以视为无限总体。

2. 样本

◦ 定义：

从总体中随机抽取的部分个体组成的集合称为样本，**样本个数称为样本容量。**

◦ 性质：

■ **二重性**：抽取前随机，是随机变量；抽取后确定，是一组数值。

■ **随机性**：每个个体都有同等的机会被选入样本。

◦ **独立性**：每个样本的取值不影响其他样本取值，即分部独立。

满足后面两个性质称为**简单随机样本**，则

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i),$$

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i),$$

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i)$$

3. 分组样本

只知样本观测值所在区间，而不知具体值的样本称为分组样本。**缺点**：与完全样本相比损失部分信息。**优点**：在样本量较大时，用分组样本既简明扼要，又能帮助人们更好地认识总体。

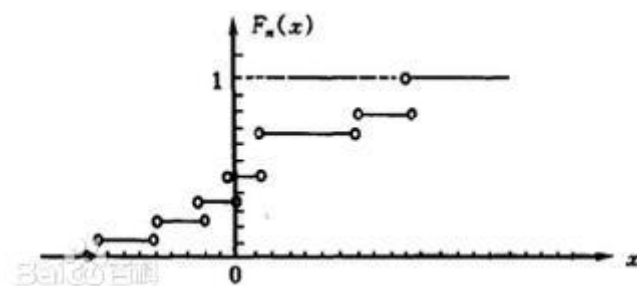
4.2 样本数据的整理与显示

1. 经验分布函数

若将样本观测值 x_1, x_2, \dots, x_n 由小到大进行排列，得到有序样本 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ ，用有序样本定义如下函数

$$F_n(x) = \begin{cases} 0 & \text{当 } x < x_{(1)} \\ k/n & \text{当 } x_{(k)} \leq x < x_{(k+1)}, k = 1, 2, \dots, n-1 \\ 1 & \text{当 } x \geq x_{(n)} \end{cases}$$

则称为 $F_n(x)$ 为该样本的经验分布函数。



2. 格里文科定理

设 x_1, x_2, \dots, x_n 是取自总体分布函数为 $F(x)$ 的样本, $F_n(x)$ 是该样本的经验分布函数, 则当 $n \rightarrow +\infty$ 时, 有

$$P(\sup_{-\infty < x < +\infty} |F_n(x) - F(x)| \rightarrow 0) = 1$$

表明当 n 相当大时, 经验分布函数 $F_n(x)$ 是总体分布函数 $F(x)$ 的一个良好的近似。它是经典统计学的一块基石。

3. 频数频率分布表

有样本 x_1, x_2, \dots, x_n 制作频数频率分布表的操作步骤如下:

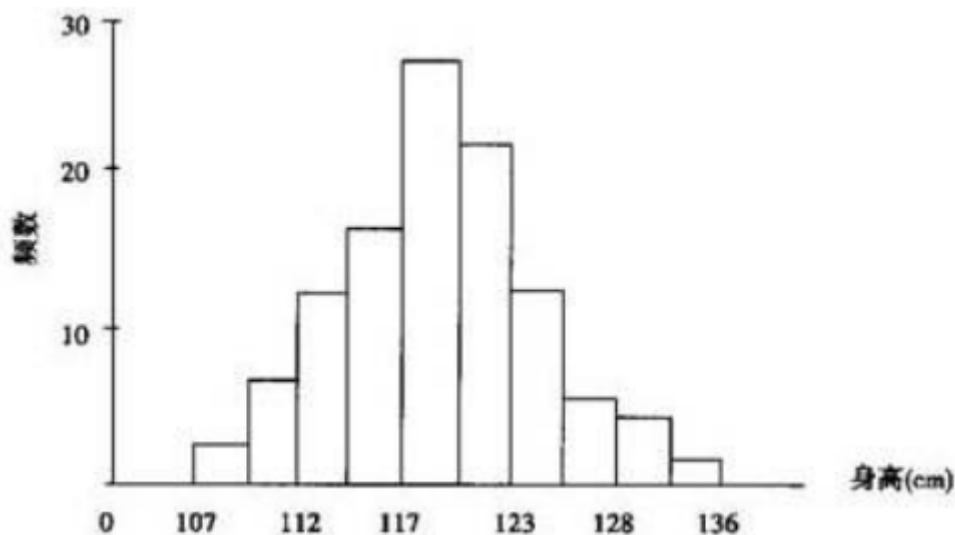
- 确定组数 k ;
- 确定每组组距, 通常取每组组距相等为 d (方便起见, 可选为整数);
- 确定组限 (下限 a_0 略小于最小观测值, 上限 a_k 略大于最大观测值);
- 统计样本数据落入每个区间的频数, 并计算频率。

该表能够简明扼要地把样本特点表示出来。不足之处是该表依赖于分组, 不同的分组方式有不同的频数频率分布表。

组段 (1)	频数, f (2)	组中值, X (3)	频率/% (4)	累计频率/% (5)
95~	1	96.5	0.83	0.83
98~	7	99.5	5.83	6.67
101~	10	102.5	8.33	15.00
104~	18	105.5	15.00	30.00
107~	25	108.5	20.83	50.83
110~	21	111.5	17.50	68.33
113~	15	114.5	12.50	80.83
116~	15	117.5	12.50	93.33
119~	7	120.5	5.83	99.17
122~125	1	123.5	0.83	100.00
合计	120	100.0		

4. 直方图

- 利用频数频率分布表上的区间 (横坐标) 和频数 (纵坐标) 可作为频数直方图;
- 若把纵坐标改为频率就得频率直方图;
- 若把纵坐标改为频率/组距, 就得到单位频率直方图。这时长条矩形的面积之和为1.



5. 茎叶图

把样本中的每个数据分为茎与叶，把茎放于一侧，叶放于另一侧，就得到一张该样本的茎叶图。比较两个样本时，可画出背靠背的茎叶图。茎叶图保留数据中全部信息，当样本量较大，数据很分散，**横跨二、三个数量级时，茎叶图并不适用。**

甲				乙		
	8		0			
4	6	3	1	2	5	
3	6	8	2	5	4	
3	8	9	3	1	6	1 6 7 9
	2		4	4	9	
	1		5	0		

4.3 统计量及其分布

1. 统计量

不含未知参数的样本函数称为统计量。统计量的分布称为抽样分布。

2. 样本均值

○ 定义：

样本 x_1, x_2, \dots, x_n 的算数平均值称为样本均值，记为 \bar{x} 。分组样本均值 $\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i f_i$ ，其中 n 为样本量， k 为组数， x_i 与 f_i 为第 i 组的组中值和频率，**分组样本均值是完全样本均值的一种较好的近似。**

样本均值是样本的位置特征，样本中大多数值位于 \bar{x} 左右。平均可消除一些随机干扰，等价交换也是在平均数中实现的。

○ 性质：

- $\sum_{i=1}^n (x_i - \bar{x}) = 0$ ，样本数据 x_i 对样本均值 \bar{x} 的偏差之和为零；
- 样本数据 x_i 与样本均值 \bar{x} 的偏差平方和最小，即对任意的实数 c 有 $\sum_{i=1}^n (x_i - \bar{x})^2 \leq \sum_{i=1}^n (x_i - c)^2$ ；
- 若总体分布为 $N(\mu, \sigma^2)$ ，则 \bar{x} 的精确分布为 $N(\mu, \sigma^2/n)$ ；
- 若总体分布未知，但其期望 μ 与方差 σ^2 存在，则当 n 较大时， \bar{x} 的渐进分布为 $N(\mu, \sigma^2/n)$ ，这里渐进分布是指 n 较大时的近似分布。

3. 样本方差与样本标准差

样本方差有两种， $s_*^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ 与 $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ ，后者为无偏方差，也是最常用的。（这是因为当 σ^2 为总体方差时，总有 $E(s_*^2) = \frac{n-1}{n} \sigma^2$, $E(s^2) = \sigma^2$ ，表明 s_*^2 有系统偏小的误差， s^2 无此系统偏差。）称 $\sqrt{s^2}$ 为样本标准差。

样本方差是样本的散布特征， **s^2 越大样本越分散， s^2 越小分布越集中**，样本标准差比样本方差使用更频繁，因为前者和样本均值有着相同的单位。

s^2 的计算有如下三个公式可供选用：

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = \frac{1}{n-1} \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right] = \frac{1}{n-1} (\sum x_i^2 - n\bar{x}^2)$$

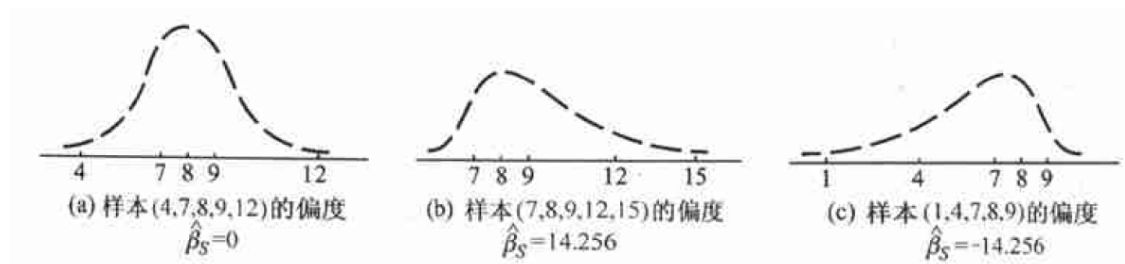
在分组样本场合，样本方差的近似计算公式为

$$s^2 = \frac{1}{n-1} \sum_{i=1}^k f_i (x_i - \bar{x})^2 = \frac{1}{n-1} (\sum_{i=1}^k f_i x_i^2 - n\bar{x}^2)$$

其中 k 为组数， x_i, f_i 分别为第 i 个区间的组中值与频数， \bar{x} 为分组样本的均值。

4. 样本矩及其函数

- 样本的 k 阶原点矩 $a_k = \frac{1}{n} \sum_{i=1}^n x_i^k$ ，样本均值 \bar{x} 为样本的一阶原点矩；
- 样本的 k 阶中心矩 $b_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$ ，样本方差 s^2 和 s_*^2 都为样本的二阶中心矩；
- 样本变异系数 $C_r = s/\bar{x}$ ；
- 样本的偏度 $\hat{\beta}_3 = b_3/b_2^{3/2}$ ，反映样本数据与对称性偏离程度和偏离方向；
- 样本的峰度 $\hat{\beta}_4 = \frac{b_4}{b_2^2} - 3$ ，反映总体分布密度曲线在其峰值附近的陡峭程度和尾部粗细。



5. 次序统计量及其分布

设 x_1, \dots, x_n 是取自某总体的一个样本， $x_{(i)}$ 称为该样本的第 i 个次序统计量（升序排序后，第 i 个样本）。

- $x_{(1)} = \min\{x_1, \dots, x_n\}$ 称为该样本的**最小次序统计量**；
- $x_{(n)} = \max\{x_1, \dots, x_n\}$ 称为该样本的**最大次序统计量**；
- $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ 称为该样本的次序统计量，即**不独立也非同分布**；
- $R = x_{(n)} - x_{(1)}$ 称为样本极差。设总体 X 的密度函数为 $f(x)$ ，分布函数为 $F(x)$ ， x_1, \dots, x_n 为样本，则有
- 样本第 k 个次序统计量 $x_{(k)}$ 的密度函数为

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} (F(x))^{k-1} (1-F(x))^{n-k} f(x);$$

- 样本第 i 个与第 j 个次序统计量的联合密度函数为

$$f_{ij}(y, z) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(y))^{i-1} (F(z) - F(y))^{j-i-1} (1-F(z))^{n-j} f(y)f(z), \quad y \leq z, 1 \leq i < j \leq n$$

6. 样本中位数与样本分位数

设 x_1, \dots, x_n 是取自某总体的样本， $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ 为该样本的次序统计量，则样本中位数 $m_{0.5}$ 定义为

$$m_{0.5} = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ 为奇数} \\ \frac{1}{2} (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & n \text{ 为偶数} \end{cases}$$

样本的 p 分位数 m_p 定义为

$$m_p = \begin{cases} x_{[np+1]} & np \text{不是整数} \\ \frac{1}{2}(x_{(np)} + x_{(np+1)}) & np \text{是整数} \end{cases}$$

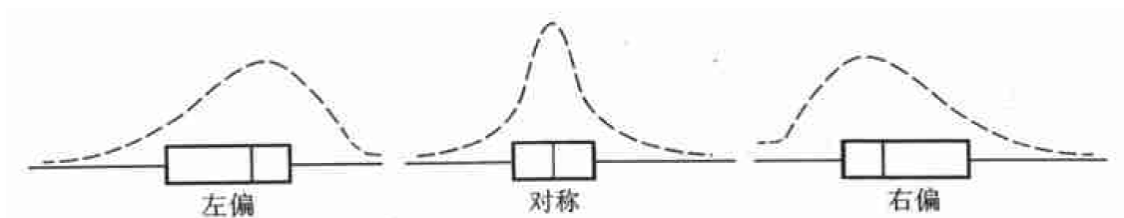
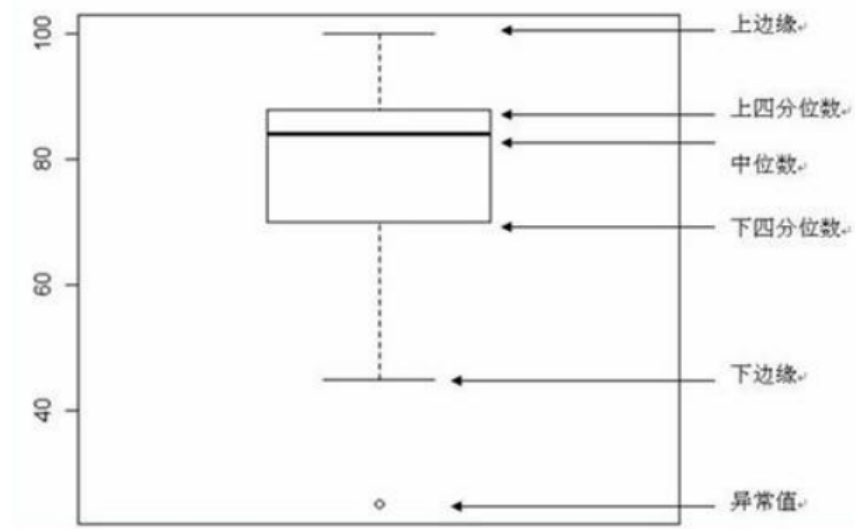
其中 $[x]$ 表示向下取整。**中位数对样本的极端值有抗干扰性，或称有稳健性。** **样本分位数的渐近分布：**设总体的密度函数为 $f(x)$ ， x_p 为总体的 p 分位数。若 $p(x)$ 在 x_p 处连续且 $p(x_p) > 0$ ，则当 n 充分大时，有

$$m_p \sim N\left(x_p, \frac{p(1-p)}{n \cdot p^2(x_p)}\right),$$

$$m_{0.5} \sim N\left(x_{0.5}, \frac{1}{4n \cdot p^2(x_{0.5})}\right)$$

7. 五数概括与箱线图

五数指用样本的五个次序统计量，即最小观测值，最大观测值，中位数，第一四分位数和第三四分位数。其图形为箱线图，可描述样本分布形状。



Author : 钱小z

Email : qz_gis@163.com

Bio : GISer , Spatiotemporal data mining

GitHub : QianXzhen