

基本概念

随机检验

- 满足条件
 - 可以在相同的条件下重复地进行；
 - 每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果；
 - 进行一次实验之前不能确定哪一个结果会出现；

样本空间、随机事件

样本空间定义：由于每次试验的结果不能被预知，但试验的所有可能组合的集合是已知的，即随机试验 E 的所有可能结果构成的集合为 E 的样本空间 S ，每一个元素被称为样本点。

随机事件定义：称试验 E 的样本空间 S 的子集为 E 的随机事件，简称事件。每次试验，当且仅当这一子集的一个样本点出现时，称这一个事件发生。

- 由一个样本点构成的单点集称为基本事件。
- 样本空间 S 包含所有的样本点，每次试验中它总是发生， S 称为必然事件。
- 空集 \emptyset 不包含任何样本点，但也是样本空间的子集，每次试验中都不发生， \emptyset 称为不可能事件。

事件的关系：

1. 若 $A \subset B$ ，则 A 发生必然导致 B 发生；弱 $A \subset B$ ，且 $B \subset A$ ，则 $A = B$ ，称两个事件相等。
2. $A \cup B = \{x : x \in A \text{ OR } x \in B\}$ 称为 A 和 B 的和事件。
3. $A \cap B = \{x : x \in A \text{ AND } x \in B\}$ 称为 A 和 B 的积事件，也记作 AB 。
4. $A - B = \{x : x \in A \text{ AND } x \notin B\}$ 称为 A 和 B 的差事件，指当 A 发生时、 B 不发生的事件 $A - B$ 。
5. 若 $A \cap B = \emptyset$ ，称 A 与 B 互不相容，或互斥的。
6. 若 $A \cap B = \emptyset$ ，且 $A \cup B = S$ ，称 A 和 B 是互为逆事件，又称对立事件。 A 的对立事件记作 \overline{A} 。

运算定律（部分）：

- 分配律：
 - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
 - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- 德摩根律：
 - $\overline{A \cup B} = \overline{A} \cap \overline{B}$
 - $\overline{A \cap B} = \overline{A} \cup \overline{B}$

频率与概率

定义：在相同的条件下，进行 n 次试验，在这 n 次试验中，事件 A 发生的次数 n_A 称为 A 发生的频数，比例 n_A/n 称为 A 发生的频率，记作 $f_n(A)$

定义：设 E 是随机事件， S 是它的样本空间，对于 E 的每一个事件 A 赋予一个实数，记作 $P(A)$ ，称 A 的概率。

- 概率的性质
 1. $P(\emptyset) = 0$

2. $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$
3. $\text{if } A \subset B: P(B - A) = P(B) - P(A)$
 $\text{else} : P(B - A) = P(B) - P(A) + P(A - B)$
4. $P(\overline{A}) = 1 - P(A)$
5. $P(A + B) = P(A) + P(B) - P(AB)$

等可能概型（古典概型）

定义：若试验的样本空间只包含有限个元素，且每个基本事件发生的可能性相同，则称这种试验为等可能概型，即古典概型。

事件中每个基本事件的概率为 $P(\{e_i\}) = \frac{1}{n}$ ，若事件 A 包含 k 个基本事件，则

$$P(A) = \sum_{j=1}^k P(\{e_j\}) = \frac{k}{n}$$

对于每种试验，要考虑它是放回抽样还是不放回抽样。

超几何分布：设有 N 件商品，其中有 D 件次品，从中任取 n 件，其中恰好有 k 件次品的概率为：

$$p = \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}}$$

$$\text{其中 } \binom{n}{m} = C_n^m = \frac{A_n^m}{A_m^m} = \frac{n!}{(n-m)! m!}$$

条件概率

定义：若 A 已发生的条件下 B 发生的概率为条件概率，记为 $P(B|A) = \frac{P(AB)}{P(A)}$ ($P(A) > 0$)

乘法定理：通过条件概率公式可得乘法公式 $P(AB) = P(B|A)P(A)$ ($P(A) > 0$)，可推广为更多的公式，例如 $P(ABC) = P(C|AB)P(B|A)P(A)$ ($P(AB) \geq P(A) > 0$)

样本划分定义：设 S 为试验 E 的样本空间， B_1, \dots, B_n 为 E 的一组事件，若满足下列条件，则称 B_1, \dots, B_n 为 E 的一个划分。

1. $B_i B_j = \emptyset$ ($i \neq j$)
2. $\bigcup_{i=1}^n B_i = S$

全概率公式定理：设 S 为试验 E 的样本空间， B_1, \dots, B_n 为 S 的一个划分，且 $P(B_i) > 0$ ， A 为 E 的事件，则 $P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$

贝叶斯 (Bayes) 公式定理：设 S 为试验 E 的样本空间， B_1, \dots, B_n 为 S 的一个划分， A 为 E 的事件，且 $P(A) > 0$ ， $P(B_i) > 0$ ，则

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

独立性

定义：设 A, B 为两个事件，若满足 $P(AB) = P(A)P(B)$ ，则这两个事件相互独立，称 A 与 B 独立。

• 定理

1. 设 A, B 为两个事件，且 $P(A) > 0$ ，若两事件相互独立，则 $P(B|A) = P(B)$ ，反之亦然；
2. 设 A, B 相互独立，则下列各对事件也相互独立：

A 与 \overline{B} 、 \overline{A} 与 \overline{B} 、 \overline{A} 与 B

定义：设 A, B, C 为三个事件，相互独立条件为：

$$\begin{cases} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \\ P(ABC) = P(A)P(B)P(C) \end{cases}$$

随机变量及其分布

概念

随机变量定义：设随机试验的样本空间为 $S = \{e\}$ ， $X = X(e)$ 是定义在样本空间 S 上的实值单值函数。称 $X = X(e)$ 为随机变量。

离散型随机变量：若随机变量可能取到的值是有限个或可列无限多个，则称这种随机变量为离散型随机变量。

分布律定义：设离散型随机变量 X 所有可能取到的值为 $x_k (k = 1, 2, \dots)$ ， X 取各个可能值的概率，即事件 $\{X = x_k\}$ 的概率为： $P\{X = x_k\} = p_k, k = 1, 2, \dots$ 若 p_k 满足如下两个条件，则称 $P\{X = x_k\}$ 为离散型随机变量 X 的分布律。

- $p_k \geq 0, k = 1, 2, \dots$
- $\sum_{k=1}^{\infty} p_k = 1$

分布

0-1分布

0-1分布的随机变量 X 只可能取0与1，它的分布律是 $P\{X = k\} = p^k(1-p)^{1-k}, k = 0, 1$ 。

若一个随机试验的样本空间只包含两个元素 $S = \{e_1, e_2\}$ ，则能在 S 上定义一个服从0-1分布的随机变量：

$$X = X(e) = \begin{cases} 0, & e = e_1 \\ 1, & e = e_2 \end{cases}$$

伯努利试验（二项分布）

若随机试验 E 只有两种可能 A 与 \overline{A} ，则称 E 为伯努利 (Bernoulli) 试验。将 E 重复进行 n 次，则称这一串重复的独立试验为 n 重伯努利试验。以 C_i 表示第 i 次试验的结果， $C_i = A, \overline{A}$

$$P(C_1 C_2 \cdots C_n) = P(C_1)P(C_2) \cdots P(C_n)$$

当 $n = 1$ 时，伯努利分布退化为0-1分布。

泊松分布

若随机变量 X 的所有可能取的值为 $0, 2, 3, \dots$ ，而取各个值的概率为：

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots$$

其中 $\lambda > 0$ 是常数，称 X 服从参数为 λ 的泊松分布，记作 $X \sim \pi(\lambda)$ 。

泊松分布逼近二项分布的定理：

泊松定理：设 $\lambda > 0$ 是一个常数， n 为任意正整数，设 $np_n = \lambda$ ，则队以任一固定的非负整数 k ，有：

$$\lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

随机变量的分布函数

对于非离散型随机变量 X ，由于无法一一枚举可能取到的值，可通过分布函数描述它的分布律。

定义：设 X 是一个随机变量， x 为任意实数，函数 $F(x) = P\{X \leq x\}$ ， $-\infty < x < \infty$ 称为 X 的分布函数（cdf）。

连续型随机变量及其概率密度

定义：对于非离散型随机变量，若对于随机变量 X 的分布函数 $F(x)$ ，存在非负可积函数 $f(x)$ ，是对于任意实数 x ，有 $F(x) = \int_{-\infty}^x f(x)dx$ ，则称 X 为连续型随机变量， $f(x)$ 称为 X 的概率密度函数，简称为概率密度（pdf）。

均匀分布

若连续型随机变量 X 具有概率密度：

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & else \end{cases}$$

称 X 在区间 (a, b) 上服从均匀分布，记作 $X \sim U(a, b)$

指数分布

若连续型随机变量 X 具有概率密度：

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & x > 0 \\ 0, & else \end{cases}$$

其中 $\theta > 0$ 为常数，称 X 服从参数为 θ 的指数分布。

正态分布

若连续型随机变量 X 具有概率密度：

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

其中 μ, σ ($\sigma > 0$) 为常数，称 X 服从参数为 μ, σ 的正太分布或高斯（Gauss）分布，记作 $X \sim N(\mu, \sigma^2)$ 。

定理：若 $X \sim N(\mu, \sigma^2)$ ，则 $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ 。

定义： $\alpha = P\{X > z_\alpha\} = 1 - P\{X \leq z_\alpha\}$ ，称 α 为标准正态分布的上 α 分位点。

注：t分布，卡方分布等不属于连续型随机变量的概率密度，属于抽样分布的内容。

随机变量的函数的分布

就是通过已知的概率密度或概率分布转换为满足某种映射规律的新概率密度或概率分布。

定理：设随机变量 X 具有概率密度 $f_X(x)$, $-\infty < x < \infty$, 有蛇函数 $g(x)$ 处处可导且恒有 $g'(x) > 0$ (或 $g'(x) < 0$) , 则 $Y = g(X)$ 是连续型随机变量, 其概率密度为:

$$f_Y(x) = \begin{cases} f_X[h(y)] |h'(y)|, & \alpha < x < \beta \\ 0, & \text{else} \end{cases}$$

其中, $\alpha = \min\{g(-\infty), g(\infty)\}$, $\beta = \max\{g(-\infty), g(\infty)\}$, $g(x)$ 的反函数为 $h(x) = g^{-1}(x)$ 。

多维随机变量及其分布

二维随机变量

定义：设 E 是一个随机试验, 它的样本空间为 $S = \{e\}$ 。设 $X = X(e), Y = Y(e)$ 是定义在 S 上的随机变量, 由它们构成的一个向量 (X, Y) 叫做二维随机变量或二位随机向量。对于任意实数 x, y , 二元函数:

$$F(x, y) = P\{(X \leq x) \cap (Y \leq y)\} \stackrel{\text{记作}}{=} P\{X \leq x, Y \leq y\}$$

称该函数为二维随机变量 (X, Y) 的分布函数, 或称为随机变量 X 和 Y 的联合分布函数。

$$P\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\} = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

定义：对于二维离散型随机变量, 称 $P\{X = x_i, Y = y_j\} = p_{ij}$, $i, j = 1, 2, \dots$ 为二位离散型随机变量 (X, Y) 的分布律, 或称为随机变量 X 和 Y 的联合分布律。

定义：对于二维随机变量 (X, Y) 的分布函数 $F(x, y)$, 若存在非负可积函数 $f(x, y)$, 使对于任意 x, y , 满足下列公式, 则称 (X, Y) 是连续型随机变量, 函数 $f(x, y)$ 为它的概率密度, 或称为随机变量 X 和 Y 的联合概率密度。

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(u, v) du dv$$

边缘分布

定义：二维随机变量 (X, Y) 作为一个整体具有分布函数 $F(x, y)$, 而 X 和 Y 都是随机变量, 各自也有分布函数, 将它们分别记做 $F_X(x), F_Y(y)$, 依次称为随机变量 (X, Y) 关于 X 和关于 Y 的边缘分布函数。

$$F_X(x) = P\{X \leq x\} = P\{X \leq x, Y < \infty\} = F(x, \infty)$$

随机变量 (X, Y) 关于 X 的边缘概率密度为:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

条件分布

定义：二维随机变量 (X, Y) 的 X, Y 的边缘分布率存在的条件下, 若事件 $\{Y = y_j\}$ 已发生的条件下事件 $\{X = x_i\}$ 发生的概率为 $P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{P_{ij}}{P_{.j}}$, $i = 1, 2, \dots$ 称其为

$Y = y_j$ 条件下随机变量 X 的条件分布律。 $f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$ 为 $Y = y_j$ 条件下随机变量 X 的条件概率密度。

相互独立的随机变量

定义： 设 $F(x, y)$ 及 $F_X(x), F_Y(y)$, 若对所有 x, y 有 $P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}$, 即满足 $F(x, y) = F_X(x)F_Y(y)$, 则称随机变量 X 和 Y 是相互独立的。

同理 $f(x, y) = f_X(x)f_Y(y)$

定理： 设 (X_1, X_2, \dots, X_m) 和 (Y_1, Y_2, \dots, Y_n) 相互独立, 且 X_i 和 Y_j 相互独立。若 h, g 是连续函数, 则 $h(X_1, X_2, \dots, X_m)$ 与 $g(Y_1, Y_2, \dots, Y_n)$ 相互独立。

两个随机变量的函数的分布

$Z = X + Y$ 的分布

设二维随机变量 (X, Y) 具有概率密度函数 $f(x, y)$, 则 $Z = X + Y$ 仍为连续型随机变量, 其概率密度为:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(z-y, y) dy = \int_{-\infty}^{\infty} f(x, z-x) dx$$

若 X 与 Y 相互独立, 则其概率密度可再被划分为:

$$f_{X+Y}(z) = f_X * f_Y = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y) dy = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x) dx$$

有限个相互独立的正态分布的随机变量的线性组合仍然服从正态分布。

$Z = \frac{Y}{X}$, $Z = XY$ 的分布

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f(x, xz) dx = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz) dx$$
$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, \frac{z}{x}) dx = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(\frac{z}{x}) dx$$

$M = \max\{X, Y\}$, $N = \min\{X, Y\}$ 的分布

$$P\{M \leq z\} = P\{X \leq z, Y \leq z\}$$
$$P\{N \leq z\} = 1 - P\{X > z, Y > z\}$$

随机变量的数字特征

数学期望

定义： 设离散型随机变量 X 的分布律为: $P\{X = x_k\} = p_k, k = 1, 2, \dots$, 若级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛, 则称级数 $\sum_{k=1}^{\infty} x_k p_k$ 的和为随机变量 X 的数学期望, 记作 $E(X)$ 。

对于连续型随机变量 X , 其概率密度为 $f(x)$, 若积分 $\int_{-\infty}^{\infty} x f(x) dx$ 绝对收敛, 则称该积分的值为随机变量 X 的数学期望。

数学期望简称期望, 又称为均值。

定理： 设 Y 是随机变量 X 的函数: $Y = g(X)$ (g 是连续函数)

1. 若 X 是离型随机变量, 且 $\sum_{k=1}^{\infty} g(x_k) p_k$ 绝对收敛, 则 $E(Y) = E[g(x)] = \sum_{k=1}^{\infty} g(x_k) p_k$
2. 若 X 是连续型随机变量, 且 $\int_{-\infty}^{\infty} g(x)f(x) dx$ 绝对收敛, 则 $E(Y) = E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x) dx$

方差

样本 X 与其预期 $E(x)$ 之间存在偏离程度，偏离程度的均值就是方差。

定义：设 X 是一个随机变量，若 $E\{[X - E(X)]^2\}$ 存在，则称 $E\{[X - E(X)]^2\}$ 为 X 的方差，记作 $D(X)$ 或 $Var(X)$ 。定义 $\sigma(X) = \sqrt{D(X)}$ ，称为标准差或均方差。

$X^* = \frac{X - \mu}{\sigma}$ ，数学期望为0，方差为1， X^* 称为 X 的标准化变量。

性质：

1. 设 C 为常数， $D(C) = 0$
2. 设 X 为随机变量， C 为常数，则有 $D(CX) = C^2 D(X)$ $D(X + C) = D(X)$
3. 设 X, Y 为两个随机变量，则有 $D(X + Y) = D(X) + D(Y) + 2E\{(X - E(X))(Y - E(Y))\}$ ，若二者相互独立，则 $D(X + Y) = D(X) + D(Y)$
4. $D(X) = 0$ 的充要条件为 $P\{X = E(X)\} = 1$

切比雪夫不等式：设随机变量 X 具有数学期望 $E(X) = \mu$ ，方差 $D(X) = \sigma^2$ ，则对于任意正数 ε ，有：

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

协方差及相关系数

协方差描述的是两个或多个随机变量的相关程度，协方差为零则相互独立。

定义： $E\{[X - E(X)][Y - E(Y)]\}$ 称为随机变量 X, Y 的协方差，记作 $Cov(X, Y)$ 。

$\rho_{XY} = \frac{Cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ 称为随机变量 X 与 Y 的相关系数。

有之前式子展开，可得 $Cov(X, Y) = E(XY) - E(X)E(Y)$

性质：

1. $Cov(aX, bY) = abCov(X, Y)$
2. $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$

定理：

1. $|\rho_{XY}| \leq 1$
2. $|\rho_{XY}| = 1$ 的充要条件为 $P\{Y = a + bX\} = 1$ ，即 X, Y 线性相关。

矩、协方差矩阵

定义：

- $E(X^k), k = 1, 2, \dots$ 存在，则称它为 X 的 k 阶原点矩，简称 k 阶矩。
- $E\{[X - E(X)]^k\}, k = 1, 2, \dots$ 存在，则称它为 X 的 k 阶中心矩。
- $E(X^k Y^l), k, l = 1, 2, \dots$ 存在，则称它为 X, Y 的 $k + l$ 阶混合矩。
- $E\{[X - E(X)]^k [Y - E(Y)]^l\}, k, l = 1, 2, \dots$ 存在，则称它为 X, Y 的 $k + l$ 阶混合中心矩。
- 设 n 维随机变量 (X_1, X_2, \dots, X_n) ， $c_{ij} = Cov(X_i, X_j)$ 存在，则称这个矩阵为维随机变量 (X_1, X_2, \dots, X_n) 的协方差矩阵。

大数定律及中心极限定理

大数定理

弱大数定理 (辛钦大数定理) : 设 X_1, X_2, \dots 相互独立, 服从同一个分布的随机变量序列, 且具有数学期望 $E(X_k) = \mu$, 做前 n 个变量的算术平均 $\frac{1}{n} \sum_{k=1}^n X_k$, 则对任意 $\varepsilon > 0$ 有:

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \varepsilon \right\} = 1$$

即序列 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ 依概率收敛于 μ , 即 $\bar{X} \xrightarrow{P} \mu$

伯努利大数定理: 设 f_A 是 n 次独立重复试验中事件 A 发生的次数, p 是事件 A 在每次试验中发生的概率, 则对于任意整数 $\varepsilon > 0$ 有:

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{f_A}{n} - p \right| < \varepsilon \right\} = 1$$

其他还有如切比雪夫大数定理、马尔可夫大数定理、泊松大数定理。

强大数定理: 设 X_1, X_2, \dots 相互独立, 服从同一个分布的随机变量序列, 满足下列公式。

$$P \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n [X_k - E(X_k)] = 0 \right\} = 1$$

还有如博雷尔强大数定理、柯尔莫哥洛夫强大数定理。

中心极限定理

独立同分布的中心极限定理: 设 X_1, X_2, \dots 相互独立, 服从同一个分布, 且具有数学期望和方差 $E(X_k) = \mu, D(X_k) = \sigma^2 > 0$, 则随机变量之和 $\sum_{k=1}^n X_k$ 的标准化变量

$$Y_n = \frac{\sum_{k=1}^n X_k - E(\sum_{k=1}^n X_k)}{\sqrt{D(\sum_{k=1}^n X_k)}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\begin{aligned} \lim_{n \rightarrow \infty} F_n(x) &= \lim_{n \rightarrow \infty} \left\{ \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq x \right\} \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x) \end{aligned}$$

即, 可用正态分布近似表示。

李雅普诺夫 (Lyapunov) 定理: 设 X_1, X_2, \dots 相互独立, 服从同一个分布, 且具有数学期望和方差 $E(X_k) = \mu_k, D(X_k) = \sigma_k^2 > 0$, 记 $B_n^k = \sum_{k=1}^n \sigma_k^2$ 。若存在正数 δ , 使得当 $n \rightarrow \infty$ 时:

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^n \{|X_k - \mu_k|^{2+\delta}\} \rightarrow 0$$

则随机变量之和 $\sum_{k=1}^n X_k$ 的标准化变量

$$Z_n = \frac{\sum_{k=1}^n X_k - E(\sum_{k=1}^n X_k)}{\sqrt{D(\sum_{k=1}^n X_k)}} = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n}$$

的分布函数 $F_n(x)$ 对任意 x , 满足

$$\begin{aligned}\lim_{n \rightarrow \infty} F_n(x) &= \lim_{n \rightarrow \infty} \left\{ \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n} \leq x \right\} \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)\end{aligned}$$

棣莫弗—拉普拉斯 (De Moivre-Laplace) 定理: 设随机变量 η_n 服从参数为 n, p 的二项分布, 则对于任意 x , 有

$$\begin{aligned}\lim_{n \rightarrow \infty} F_n(x) &= \lim_{n \rightarrow \infty} \left\{ \frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x \right\} \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)\end{aligned}$$

样本及抽样分布

随机样本

- **定义:**
 - 总体: 实验中所有可能的观察值;
 - 个体: 每一个可能的观察值;
 - 容量: 总体中所包含的个体的个数;
 - 有限总体: 容量为有限的总体;
 - 无限总体: 容量为无限的总体;
 - 样本: 设 X 是具有分布函数 F 的随机变量, 若 X_1, X_2, \dots, X_n 是具有同一分布函数 F 的, 相互独立的随机变量, 则称它们为从分布函数 F (或总体 F , 或总体 X) 得到的容量为 n 的简单随机样本, 简称样本, 它们的观察值 x_1, x_2, \dots, x_n 称为**样本值**, 又称 X 的 **n 个独立的观察值**。

抽样分布

样本平均值

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

样本方差

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$$

样本标准差

$$S = \sqrt{S^2}$$

样本 k 阶 (原点) 矩

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots$$

样本 k 阶中心矩

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad k = 1, 2, \dots$$

表示上，将大写转为小写即为它们的观察值。

对于样本 k 阶（原点）矩，根据辛钦大数定理，

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{P} \mu_k$$
$$\Rightarrow g(A_1, \dots, A_n) \xrightarrow{P} g(\mu_1, \dots, \mu_n)$$

其中 g 为连续函数，这也是矩估计法的理论依据。

经验分布函数：按观察值得到的分布函数。

χ^2 分布

设 X_1, X_2, \dots, X_n 是来自总体 $N(0, 1)$ 的样本，则统计量 $\chi^2 = X_1^2 + \dots + X_n^2$ 服从自由度为 n 的 χ^2 分布，记作 $\chi^2 \sim \chi^2(n)$

$$f(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}, & y > 0 \\ 0, & else \end{cases}$$

特性：

- 可加性： $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$
- 期望与方差： $E(\chi^2) = n, D(\chi^2) = 2n$
- 上分位点： $P\{\chi^2 > \chi_\alpha^2(n)\} = \int_{\chi_\alpha^2(n)}^\infty f(x)dx = \alpha$
 - 当 $n > 40$ ，可用近似 $\chi_\alpha^2 \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$
 - z_α 为正态分布的上分位点。

t 分布（学生氏分布）

设 $X \sim N(0, 1), Y \sim \chi^2(n)$ ，且相互独立，则 $t = \frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布，记作 $t \sim t(n)$

。

上分位点： $P\{t > t_\alpha(n)\} = \int_{t_\alpha(n)}^\infty h(t)dt = \alpha$

- $t_{1-\alpha}(n) = -t_\alpha(n)$
- $t_\alpha(n) \approx z_\alpha$, when $n > 45$

F 分布

设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$ ，且二者相互独立，则 $F = \frac{U/n_1}{V/n_2}$ 服从自由度为 (n_1, n_2) 的 F 分布，

记作 $F \sim F(n_1, n_2)$

上分位点： $P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1, n_2)}^\infty \psi(x)dx = \alpha$

- $F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}$

正态分布的样本均值与样本方差的分布

设总体 X 的均值为 μ ，方差为 σ^2 ， X_1, X_2, \dots, X_n 是来自 X 的一组样本， \bar{X} , S^2 分别是样本均值和样本方差，则有：

$$E(\bar{X}) = \mu, D(\bar{X}) = \sigma^2/n$$
$$E(S^2) = \sigma^2$$

定理： 设 X_1, X_2, \dots, X_n 是来自于正态分布 $N \sim (\mu, \sigma^2)$ 的样本， \bar{X} 是样本均值，则 $\bar{X} \sim N(\mu, \sigma^2/n)$

定理： 设 X_1, X_2, \dots, X_n 是来自于正态分布 $N \sim (\mu, \sigma^2)$ 的样本， \bar{X} , S^2 是样本均值和样本方差，则有：

1. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
2. \bar{X} , S^2 相互独立

定理： 设 X_1, X_2, \dots, X_n 是来自于正态分布 $N \sim (\mu, \sigma^2)$ 的样本， \bar{X} , S^2 是样本均值和样本方差，则有 $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$

定理： 设 X_1, X_2, \dots, X_{n_1} 是来自于正态分布 $N \sim (\mu_1, \sigma_1^2)$ 的样本， \bar{X} , S_1^2 是样本均值和样本方差；设 Y_1, Y_2, \dots, Y_{n_2} 是来自于正态分布 $N \sim (\mu_2, \sigma_2^2)$ 的样本， \bar{Y} , S_2^2 是样本均值和样本方差。则有：

1. $\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$
2. 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$
$$\text{where } S_w^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

参数估计

点估计

根据频率估计分布的参数。

流程：

1. 设总体 X 的分布函数 $F(x; \theta)$ 的形式为已知， θ 是待估参数。 X_1, X_2, \dots, X_n 是 X 的一个样本， x_1, x_2, \dots, x_n 是相应的一个样本值。
2. 构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$ ，用它的观察值 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 作为未知参数 θ 的近似值。其中， $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为**估计量**， $\hat{\theta}(x_1, x_2, \dots, x_n)$ 为**估计值**。统称二者为**估计**，简记为 $\hat{\theta}$ 。
3. 将样本值作为输入，求出对参数 θ 的估计值，并将该估计值作为未知参数 θ 的近似值。

构造估计量可理解为构造一个能求出近似值的样本到目标参数的映射关系，常用：矩估计法和最大似然估计法。

矩估计法

设 $\theta_1, \theta_2, \dots, \theta_k$ 为待估参数, X_1, X_2, \dots, X_n 是来自 X 的样本, 设总体 X 的前 k 阶矩:

$$\begin{aligned}\mu_l &= E(X^l) = \int_{-\infty}^{\infty} x^l f(x; \theta_1, \theta_2, \dots, \theta_k) dx \quad (X \text{ 为连续型}) \\ \mu_l &= E(X^l) = \sum_{x \in R_X} x^l p(x; \theta_1, \theta_2, \dots, \theta_k) \quad (X \text{ 为离散型})\end{aligned}$$

R_X 为 X 可能取值的范围。

将 k 个总体矩联立解方程组, 求出 $\theta_1, \theta_2, \dots, \theta_k$ 的方程组:

$$\begin{cases} \mu_1 = \mu_1(\theta_1, \theta_2, \dots, \theta_k) \\ \mu_2 = \mu_2(\theta_1, \theta_2, \dots, \theta_k) \\ \vdots \\ \mu_k = \mu_k(\theta_1, \theta_2, \dots, \theta_k) \end{cases} \Rightarrow \begin{cases} \theta_1 = \theta_1(\mu_1, \mu_2, \dots, \mu_k) \\ \theta_2 = \theta_2(\mu_1, \mu_2, \dots, \mu_k) \\ \vdots \\ \theta_k = \theta_k(\mu_1, \mu_2, \dots, \mu_k) \end{cases}$$

用样本矩 $A_l = \frac{1}{n} \sum_{i=1}^n X_i^l$ 替换解出的总体矩 μ_l :

$$\begin{cases} \theta_1 = \theta_1(\mu_1, \mu_2, \dots, \mu_k) \\ \theta_2 = \theta_2(\mu_1, \mu_2, \dots, \mu_k) \\ \vdots \\ \theta_k = \theta_k(\mu_1, \mu_2, \dots, \mu_k) \end{cases} \Rightarrow \begin{cases} \theta_1 = \theta_1(A_1, A_2, \dots, A_k) \\ \theta_2 = \theta_2(A_1, A_2, \dots, A_k) \\ \vdots \\ \theta_k = \theta_k(A_1, A_2, \dots, A_k) \end{cases}$$

这种估计量称为矩估计量, 估计值称为矩估计值。

最大似然估计法

设 $\theta \in \Theta$ 为待估参数, Θ 为 θ 可能的取值范围, X_1, X_2, \dots, X_n 是来自 X 的样本, 则事件 $\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$ 发生的概率 (离散, 连续则定义为落在那个坐标的邻域内) 为:

$$\begin{aligned}L(\theta) &= L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p(x_i; \theta) \quad (X \text{ 为离散型}) \\ L(\theta) &= L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (X \text{ 为连续型}) \\ L(x_1, x_2, \dots, x_n; \hat{\theta}) &= \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)\end{aligned}$$

由于试验中能取到这组样本, 则认为这组样本出现的概率是最大的, 最终将方程转化为求一个 θ 使该样本出现的概率最大, 即最大值问题。

其中, $L(\theta)$ 为**似然函数**, $\theta(x_1, x_2, \dots, x_n)$ 为 θ 的**最大似然估计值**, 称 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为**最大似然估计量**。

求解则求解方程 $\frac{d}{d\theta} L(\theta) = 0$, 由于 $L(\theta), \ln L(\theta)$ 在同一个 θ 取极值, 因此为了方便, 使用对数似然方程 $\frac{d}{d\theta} \ln L(\theta) = 0$ 进行求解。

估计量的评估

无偏性

设 X_1, X_2, \dots, X_n 是总体 X 的一个样本, $\theta \in \Theta$ 是包含在总体 X 的分布中的待估参数。若估计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 的数学期望 $E(\hat{\theta})$ 存在, 且 $E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 是 θ 的无偏估计量。

有效性

在参数附近越密集, 则有效性越好, 即方差越小越好。

设 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \dots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \dots, X_n)$ 都是 θ 的无偏估计量, 若 $D(\hat{\theta}_1) \leq D(\hat{\theta}_2)$, 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效。

相合性

样本越大, 估计值越稳定于某个值, 则代表具有相合性。

若 $\lim_{n \rightarrow \infty} P\{|\hat{\theta} - \theta| < \varepsilon\} = 1, \forall \varepsilon > 0$, 则称 $\hat{\theta}$ 是 θ 的相合估计量。

区间估计

置信区间: 目标参数 θ 有 $1 - \alpha$ 的概率落在 $\hat{\theta}$ 的置信区间内。

正态总体均值与方差的区间估计

(0-1) 分布参数的区间估计

单侧置信区间

假设检验

• 定义

- H_0 : 原假设;
- H_1 : 备择假设, 与 H_0 相互对立的假设

假设检验中作为假设的大多为均值, 对于非均值的要进行转换。

重申: $1 - \alpha$ 是落入置信区间的概率, α 是显著性水平。如 $\alpha = 0.05$, 则有95%的概率使假设落入置信区间, 称有95%的把握认为原假设显著。

由样本参数可得一个置信区间, 而由显著性水平 α 可得一个分位点。定义统计量

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \in [\underline{\mu}, \bar{\mu}], \text{ 由显著性水平 } \alpha \text{ 和原假设可得一个分位点 } k。$$

$$P\{\text{拒绝原假设 } H_0\} = P_{\mu_0}\{|Z| \geq k\} = \alpha$$

本质是弃真和纳伪, 分别为第一类错误和第二类错误。通过约定显著性水平 α , 可减少犯第一类错误的可能, 而不考虑犯第二类错误, 这种检验叫显著性检验。

拒绝原假设可得到拒绝域。由于置信区间内的值包含一部分落入显著性水平规定的区间外, 即存在一部分值大于显著性水平对应的分位点, 即该置信区间包含错误部分, 因此要拒绝原假设。

原假设是双边备择假设, 则该假设检验是双边假设检验, $z = |Z| \in [0, \max\{\underline{\mu}, \bar{\mu}\}]$, 邻接点 $z = k = z_{\alpha/2}$ 。

原假设是单边备择假设, $z = Z \in [\underline{\mu}, \bar{\mu}]$, 则该假设检验是单边假设检验, $H_0: \mu \leq \mu_0$ 时, 拒绝域 $z \geq k = z_\alpha$; $H_0: \mu \geq \mu_0$ 时, 拒绝域 $z \leq k = -z_\alpha$ 。

正态总体均值的假设检验

单个总体 $N(\mu, \sigma^2)$ 均值 μ 的检验

1. σ^2 已知

- 使用 Z 检验:

$$\begin{aligned} H_0: \mu &= \mu_0 \\ Z &= \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \in [\underline{\mu}, \bar{\mu}] \\ k &= z_{\alpha/2} \end{aligned}$$

2. σ^2 未知

- 使用 t 检验:

$$\begin{aligned} H_0: \mu &= \mu_0 \\ t &= \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \in [\underline{\mu}, \bar{\mu}] \\ S &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \\ k &= t_{\alpha/2}(n-1) \end{aligned}$$

详细略

原假设 H_0	检验统计量	拒绝域
$\mu \leq \mu_0$ $\mu \geq \mu_0$ $\mu = \mu_0$ $(\sigma^2 \text{已知})$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$z \geq z_\alpha$ $z \leq -z_\alpha$ $ z \geq z_{\alpha/2}$
$\mu \leq \mu_0$ $\mu \geq \mu_0$ $\mu = \mu_0$ $(\sigma^2 \text{未知})$	$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	$t \geq t_\alpha(n-1)$ $t \leq -t_\alpha(n-1)$ $ t \geq t_{\alpha/2}(n-1)$
$\mu_1 - \mu_2 \leq \delta$ $\mu_1 - \mu_2 \geq \delta$ $\mu_1 - \mu_2 = \delta$ $(\sigma_1^2, \sigma_2^2 \text{已知})$	$Z = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$z \geq z_\alpha$ $z \leq -z_\alpha$ $ z \geq z_{\alpha/2}$
$\mu_1 - \mu_2 \leq \delta$ $\mu_1 - \mu_2 \geq \delta$ $\mu_1 - \mu_2 = \delta$ $(\sigma_1^2 = \sigma_2^2 = \sigma^2 \text{未知})$	$Z = \frac{\bar{X} - \bar{Y} - \delta}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ $S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$	$t \geq t_\alpha(n_1 + n_2 - 1)$ $t \leq -t_\alpha(n_1 + n_2 - 1)$ $ t \geq t_{\alpha/2}(n_1 + n_2 - 1)$
$\sigma^2 \leq \sigma_0^2$ $\sigma^2 \geq \sigma_0^2$ $\sigma^2 = \sigma_0^2$ $(\mu \text{未知})$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2 \geq \chi_\alpha^2(n-1)$ $\chi^2 \leq \chi_{1-\alpha}^2(n-1)$ $\chi^2 \geq \chi_{\alpha/2}^2(n-1) \text{ OR}$ $\chi^2 \leq \chi_{1-\alpha/2}^2(n-1)$
$\sigma_1^2 \leq \sigma_2^2$ $\sigma_1^2 \geq \sigma_2^2$ $\sigma_1^2 = \sigma_2^2$ $(\mu_1, \mu_2 \text{未知})$	$F = \frac{S_1^2}{S_2^2}$	$F \geq F_\alpha(n_1 - 1, n_2 - 1)$ $F \leq F_{1-\alpha}(n_1 - 1, n_2 - 1)$ $F \geq F_{\alpha/2}(n_1 - 1, n_2 - 1) \text{ OR}$ $F \leq F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$
$\mu_D \leq \mu_0$ $\mu_D \geq \mu_0$ $\mu_D = \mu_0$ (成对数据)	$t = \frac{\bar{D} - 0}{S_D/\sqrt{n}}$	$t \geq t_\alpha(n-1)$ $t \leq -t_\alpha(n-1)$ $ t \geq t_{\alpha/2}(n-1)$

样本容量的选取

由于显著性检验只能保证减少第二类错误，若想用样本容量控制犯第二类错误的概率，可构建**OC函数**进行计算。

分布拟合检验

单个分布

上述的检验法都是基于知道总体服从什么类型的分布的基础上进行的，若分布类型未知，则需进行关于分布的检验。

$$\chi^2 = \sum_{i=1}^k \frac{f_i^2}{np_i} - n \geq \chi_\alpha^2(k-1)$$

其中， f_i 为样本 i 出现的次数，需合并 $np_i < 5$ 的样本， k 为最后的组数。

详见书P199及其案例。

分布族

偏度、峰度检验

偏度、峰度检验和夏皮罗-威尔克法式最有效的正态性检验方法。

秩和检验

见书案例P210

p值法

以上讨论的都是**临界值法**，p值法是个更直观的方法，它可得到最小显著水平。

本质上是反对原假设 H_0 的依据的强度，p数值越小即依据越强，只要这个概率大于显著水平，那么就能接受原假设，

理解：在临界值法中，求的是由 α 构成的置信区间，若样本计算后的观察值落在置信区间外，则说明原假设大概率不成立，换句话说，就是观察值落在总体平均的位置，但不能偏离某个范围；在p值法中，设原假设 $H_0: \mu \leq \mu_0$ 根据观察值 $z_0 = Z(\mu_0)$ 可得总体落在观察值外的概率 $P\{Z \geq z_0\} = p$ ，则说明有p的概率看到备择假设发生，若p值越小，反而能证明原假设不符合分布中表现出来的规律（因为原假设成立那么p应该表现得更像是被随机噪声影响后的样子，即越大），当p低于显著性水平时，就拒绝原假设。

假设有一家披萨店送餐时间的样本均值和方差分别为10和1，符合正态分布，样本数为100，
 $H_0: time < 20$ ， $p = P\{Z \geq z_0\} = P\{Z \geq \frac{10-20}{1/10} = -100\} = 1 - \Phi(-100)$ ，很大，显然成立。

假设有一家披萨店送餐时间的样本均值和方差分别为30和1，符合正态分布，样本数为100，
 $H_0: time < 20$ ， $p = P\{Z \geq z_0\} = P\{Z \geq \frac{30-20}{1/10} = 100\} = 1 - \Phi(100)$ ，很小，显然不成立。

方差分析及回归分析

单因素试验的方差分析

- 定义
 - 指标：在试验中，受各种因素影响的事物；
 - 试验指标：要考察的指标；
 - 因素：影响试验指标的条件；方差
 - 因素的水平：因素所处的状态；
 - 单因素试验：只有一个因素在改变的试验；
 - 多因素试验：有多于一个因素在改变的试验；

在单因素试验中，考虑的是该因素对结果的影响是否显著。为了实现这个目的，使用**误差平方和**和**效益平方和**进行方差分析。

单因素试验方差分析表

方差来源	平方和	自由度	均方	F 比
因素 A	S_A	$s - 1$	$\bar{S}_A = \frac{S_A}{s - 1}$	$F = \frac{\bar{S}_A}{\bar{S}_E}$
误差	S_E	$n - s$	$\bar{S}_E = \frac{S_E}{n - s}$	
总和	S_T	$n - 1$		

平方和公式见书P228

双因素试验的方差分析

略，应用到时再套公式吧。

主要添加了考虑两因素之间相互影响的交互项。

一元线性回归

见书。

bootstrap方法

设总体的分布 F 未知，但已有一个容量为 n 的来自分布 F 的数据样本，自这一样本按放回抽样的方法抽取一个容量为 n 的样本，这种样本称为bootstrap 样本或自助样本。

非参数bootstrap方法

不需要假设所研究的总体的分布函数的形式，通过bootstrap样本计算待求参数。

总已有样本中抽多次抽样本进行分组估计。

参数bootstrap方法

需要假设所研究的总体的分布函数的形式，通过bootstrap样本计算待求参数。

先假设分布，然后用估计法估计出分布参数，最后多次用随机数总分布函数中随机出多个样本进行分组估计。