# 概率论与数理统计

Somnia1337

# 1 概率论基础知识

## 概率及性质

💎 **概率**:

- 非负性：$P(A) \geqslant 0$。
- 规范性：$P(\Omega) = 1$。
- 可列可加性：对两两互斥的可列个事件，$P(\bigcup\limits_{i=1}^{\infty} A_i) = \sum\limits_{i=1}^{\infty} P(A_i)$。

🔔 概率:

- $P(A - B) = P(A\bar{B}) = P(A) - P(AB)$。
- $P(A \cup B) = P(A) + P(B) - P(AB)$。

## 等可能概型

### 古典概型

💎 **古典概型**:

- 试验的可能结果有限。
- 每个可能结果出现的可能性相等。

💎 **古典概率**：$P(A) = \frac{k}{n} = \frac{A\text{中的样本点数}}{\Omega\text{中的样本点数}}$。

💎 **超几何概率**：$N$ 个球，其中 $m$ 个为红球、其余为白球，取 $n$ 球，其中恰有 $k$ 个红球的概率为 $p_k = \frac{C_m^k C_{N-m}^{n-k}}{C_N^n}$。

### 几何概型

💎 **几何概型**：试验的所有可能结果等可能地出现在一个有界的欧式区域 $\Omega$ 内。

💎 **几何概率**：$P(A) = \frac{m(A)}{m(\Omega)}$，其中 $m(A)$ 为 $A$ 的度量（长度 / 面积 / 体积）。

# 条件概率

💎 **条件概率**：在 $B$ 发生的条件下 $A$ 发生的概率称为 $B$ 发生条件下 $A$ 发生的条件概率，简称 $A$ 对 $B$ 的条件概率，记为 $P(A|B)=\frac{P(AB)}{P(B)}$。

📍 设 $A_1$，$\cdots$，$A_n$ 为 $\Omega$ 的一个完备事件组，$B$ 为任一事件：

- **乘法公式**：$P(AB)=P(B)P(A|B)=P(A)P(B|A)$。
- **全概率公式**：$P(B)=\sum\limits_{i=1}^{n}P(A_i)P(B|A_i)$。
- **贝叶斯公式**：$P(B)>0$ 时，

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{P(B)}=\frac{P(A_i)P(B|A_i)}{\sum\limits_{j=1}^{n}P(A_j)P(B|A_j)}$$

# 事件的独立性

🔔 独立：若 $0<P(B)<1$，则 $A$ 与 $B$ 相互独立 $\iff P(A|B)=P(A|\bar{B})$。

💎 **独立试验**：将随机试验 $E$ 重复进行 $n$ 次，若各次试验的结果互不影响，称其为 $n$ 重独立试验。

💎 **伯努利试验**：对一个 $n$ 重独立试验，如果每次试验的可能结果只有"成功" / "失败"，称其为 $n$ 重伯努利试验，称相应的数学模型为**伯努利概型**。

💎 **二项概率**：在 $n$ 重伯努利试验中，设 $A$ 在各次试验中发生的概率为 $p$，则 $A$ 恰好发生 $k$ 次的概率为 $P_n(k)=C_n^k p^k(1-p)^{n-k}$。

💎 **多项概率**：在 $n$ 重独立试验中，每次试验的可能结果为 $A_1$，$\cdots$，$A_k$，且 $0<p_i=P(A_i)<1$，$\sum\limits_{i=1}^{k}p_i=1$，则 $A_1$，$\cdots$，$A_k$ 各发生 $r_1$，$\cdots$，$r_k$ 次的概率为 $\frac{n!}{r_1!\cdots r_k!}p_1^{r_1}\cdots p_k^{r_k}$。

# 2 随机变量及其分布

# 随机变量

💎 **随机变量**：对一个试验的每个样本点 $\omega\in\Omega$，规定一个实数 $X(\omega)$，由此得到一个定义域为 $\Omega$ 的实值函数 $X=X(\omega)$，称 $X$ 为随机变量.

💎 **分布函数**：$F(x)=P(X\leqslant x)$.

🔔 分布函数：非减，极限，右连续.

- 对任意 $x_1 < x_2$，$F(x_1) \leqslant F(x_2)$，即 $F(x)$ 非递减.
- $F(-\infty) = 0$，$F(+\infty) = 1$.
- 对任意 $x_0$，$F(x_0) = F(x_0 + 0)$，即 $F(x)$ 右连续.
- 对任意 $x_0$，$P(X = x_0) = F(x_0) - F(x_0 - 0)$.

# 离散型随机变量

💎 **离散型随机变量**：可能的取值数量为有限或可数无穷的随机变量.

💎 **概率分布**：离散型随机变量 $X$ 可能的取值为 $x_1$，$\cdots$，$x_n$，称 $X$ 取各值的概率 $p_k = P(X = x_k)$ 为 $X$ 的概率分布（概率函数 / 分布律）.概率分布可用表格 / 矩阵 / 图表表示.

- 表格：

| $X$ | $x_1 \quad x_2 \quad \cdots \quad x_n$ |
|---|---|
| $p_k$ | $p_1 \quad p_2 \quad \cdots \quad p_n$ |

- 矩阵：

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

🔔 概率分布：这 2 条性质为概率分布的特征.

- 对任意 $k$，$p_k \geqslant 0$.
- $\sum\limits_k p_k = 1$.

💎 **可列重伯努利试验**：可一直重复下去的 $n$ 重伯努利试验.在成功概率为 $p$ 的可列重伯努利试验中，首次“成功”出现在第 $k$ 次试验中的概率 $p_k = p(1-p)^{k-1}$.

💎 常见离散型分布：

- **几何分布**：$X \sim G(p)$，$p_k = p(X = k) = p(1-p)^{k-1}$.
- **超几何分布**：$X \sim H(n, m, N)$，$p_k = P(X = k) = \dfrac{C_m^k C_{N-m}^{n-k}}{C_N^n}$.
- **二项分布**：$X \sim B(n, p)$，$P_n(k) = P(X = k) = C_n^k p^k (1-p)^{n-k}$.
  - **0-1 分布**：$B(1,\ p)$，$P(X = k) = p^k (1-p)^{1-k}$.
  - **最可能次数**：
    - 若 $(n+1)p$ 为整数，$k_0 = (n+1)p$ 及 $(n+1)p - 1$.

    - 若 $(n+1)p$ 不为整数，$k_0=[(n+1)p]$.
- **泊松分布**：$X\sim P(\lambda)$，$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$.
    - $p\leqslant 0.1$，$n\geqslant 50$ 的二项分布可近似为 $\lambda=np$ 的泊松分布.

# 连续型随机变量

💎 **连续型随机变量**：$F(x)$ 为随机变量 $X$ 的分布函数，如果存在一个非负函数 $f(x)$，使得对任意实数 $x$，有

$$F(x)=\int_{-\infty}^{x}f(t)\mathrm{d}t$$

称 $X$ 为连续型随机变量，称 $f(x)$ 为 $X$ 的**概率密度函数**（**密度函数** / **密度**）.

🔔 密度函数：这 2 条性质为密度函数的特征.

- $f(x)\geqslant 0,\ x\in(-\infty,\ +\infty)$.
- $\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=1$.

🛡 $X$ 为连续型随机变量，$F(x)$ 为分布函数，$f(x)$ 为密度函数：

- 对任意常数 $a<b$，有 $P(a<X\leqslant b)=\int_a^b f(x)\mathrm{d}x$.
- 对任意常数 $C$，有 $P(X=C)=0$.
- $F(x)$ 连续.
- 在 $f(x)$ 的连续点，有 $F'(x)=f(x)$.

💎 常见连续型分布：

**均匀分布**：$X\sim U(a,b)$，

- 密度函数：$f(x)=\frac{1}{b-a},\ a\leqslant x\leqslant b$.
- 分布函数：$F(x)=\frac{x-a}{b-a},a\leqslant x\leqslant b$.

**指数分布**：$X\sim e(\lambda)$，

- 密度函数：$f(x)=\lambda e^{-\lambda x}(\lambda>0),x>0$.
- 分布函数：$F(x)=1-e^{-\lambda x},x>0$.

**$\Gamma$ 分布**：$X\sim\Gamma(\alpha,\beta)$，

- $\Gamma$ 函数：$\Gamma(\alpha)=\int_0^{+}x^{\alpha-1}e^{-x}\mathrm{d}x$.
    - $\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$.

- $\Gamma(1)=1$，$\Gamma(\frac{1}{2})=\sqrt{\pi}$.
- 对正整数 $n$，$\Gamma(n)=(n-1)!$.

- 密度函数：$f(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$.

# 3 多维随机变量及其分布

## 二维随机变量

💎 **二维随机变量**：设 $X$，$Y$ 为定义在同一样本空间 $\Omega$ 上的两个随机变量，称 $(X,\ Y)$ 为二维随机变量。

💎 **二维分布函数**：称 $F(x,\ y)=P(X\leqslant x,\ Y\leqslant y)$ 为 $(X,\ Y)$ 的二维分布函数 / 联合分布函数。

🔔 二维分布函数：

- $F(x,\ y)$ 分别关于 $x$，$y$ 单调不减。
- $F(x,\ y)$ 分别关于 $x$，$y$ 右连续。
- $F(-\infty,\ -\infty)=F(-\infty,\ y)=F(x,\ -\infty)=0$，$F(+\infty,\ +\infty)=1$。
- 对任意 $x_1<x_2$，$y_1<y_2$，有 $F(x_2,\ y_2)-F(x_2,\ y_1)-F(x_1,\ y_2)+F(x_1,\ y_1)\geqslant 0$。

💎 **二维离散型随机变量**：只取有限个或可数无穷个点对 $(x_i,\ y_i)$ 的 $(X,\ Y)$。

💎 **二维概率分布**：记 $p_{ij}=P(X=x_i,\ Y=y_j)$ 为 $(X,\ Y)$ 的二维概率分布 / 联合概率分布。

🔔 二维概率分布：

- $p_{ij}\geqslant 0$。
- $\sum p_{ij}=1$。

💎 **二维连续型随机变量**：

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

则称 $(X,\ Y)$ 为二维连续型随机变量，称 $f(x,\ y)$ 为 **二维概率密度函数** / **联合密度函数**。

🔔 二维密度函数：

- $f(x,\ y) \geqslant 0$，$(x,\ y) \in \mathbf{R}^2$。
- $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,\ y)\mathrm{d}x\mathrm{d}y = 1$。

🛡️ 二维连续型随机变量：$(X,\ Y)$ 的密度函数为 $f(x,\ y)$，

- $F(x,\ y)$ 连续，且在 $f(x,\ y)$ 的连续点 $(x,\ y)$ 有 $f(x,\ y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$。
- 对平面上任意区域 $G \subset \mathrm{R}^2$，若 $f(x,\ y)$ 在 $G$ 上可积，有 $P((X,\ Y) \in G) = \iint\limits_G f(x,\ y)\mathrm{d}x\mathrm{d}y$。
- 对平面上任意一条曲线 $L$，有 $P((X,\ Y) \in L) = 0$。

## 边缘分布

## 条件分布

## 二维随机变量函数的分布

📍 卷积公式：$Z = X + Y$，

- $f_Z(z) = \int_{-}^{+} f(x, z-x)\mathrm{d}x = \int_{-}^{+} f(z-y, y)\mathrm{d}y$。
- 如果 $X$ 与 $Y$ 独立，$f_Z(z) = \int_{-}^{+} f_X(x)f_Y(z-x)\mathrm{d}x = \int_{-}^{+} f_X(z-y)f_Y(y)\mathrm{d}y$。

# 4 随机变量的数字特征

| $X$ | $E(X)$ | $D(X)$ |
|---|---|---|
| $B(n,p)$ | $np$ | $np(1-p)$ |
| $P(\lambda)$ | $\lambda$ | $\lambda$ |
| $\Gamma(\alpha,\beta)$ | $\frac{\alpha}{\beta}$ | $\frac{\alpha}{\beta^2}$ |
| $e(\lambda)$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| $U(a,b)$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |

## 数学期望

💎 **数学期望**：

- 对离散型随机变量 $X$，概率分布 $P(X = x_k) = p_k$，如果级数 $\sum x_k p_k$ 绝对收敛，称其为 $X$ 的数学期望。

- 对连续型随机变量 $X$，密度 $f(x)$，如果反常积分 $\int_{-}^{+} xf(x)\mathrm{d}x$ 绝对收敛，称其为 $X$ 的数学期望。

🔔 数学期望：

- $E(X \pm Y) = E(X) \pm E(Y)$（线性性，不依赖于独立性）。
- $E(XY) = E(X)E(Y)$（$X$ 和 $Y$ 独立）。

# 方差

💎 **方差**：如果期望 $E[X - E(X)]^2$ 存在，称其为 $X$ 的方差，计算式为 $D(X) = E(X^2) - [E(X)]^2$。

🔔 方差：

- $D(aX + b) = a^2 D(X)$。
  - eg. $D(-2X + 1) = 4D(X)$。
- 如果 $X$ 和 $Y$ 独立，
  - $D(X \pm Y) = D(X) + D(Y)$。
  - $D(XY) = D(X)D(Y) + [E(X)]^2 D(Y) + [E(Y)]^2 D(X)$。

💎 **变异系数**：称 $C_v = \frac{\sqrt{D(X)}}{|E(X)|}$ 为 $X$ 的变异系数。

💎 矩：

- **原点矩**：称 $m_k = E(X^k)$ 为 $X$ 的 $k$ 阶原点矩。
- **中心矩**：称 $\mu_k = E[X - E(X)]^k$ 为 $X$ 的 $k$ 阶中心矩。

# 协方差

💎 **协方差**：称 $\mathrm{Cov}(X, Y) = E[[X - E(X)][Y - E(Y)]]$ 为 $(X, Y)$ 的协方差，计算式 $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$。

🔔 协方差：

- $\mathrm{Cov}(X, X) = D(X)$。
- $\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X, Y)$。
- $\mathrm{Cov}(X \pm Y, Z) = \mathrm{Cov}(X, Z) \pm \mathrm{Cov}(Y, Z)$。
- $D(X \pm Y) = D(X) + D(Y) \pm 2\mathrm{Cov}(X, Y)$。
- $\mathrm{Cov} > 0$ 正线性相关，$\mathrm{Cov} < 0$ 负线性相关，$\mathrm{Cov} = 0$ 无线性相关。
  - $X$ 与 $Y$ 独立 $\implies \mathrm{Cov}(X, Y) = 0$。

💎 **均值向量**：称向量 $(E(X), E(Y))$ 为 $(X, Y)$ 的均值向量。

💎 **协方差阵**：称矩阵 $\mathbf{V}=\begin{pmatrix} D(X) & \operatorname{Cov}(X,Y) \\ \operatorname{Cov}(X,Y) & D(Y)\end{pmatrix}$ 为 $(X, Y)$ 的协方差阵。

## 相关系数

💎 **标准化随机变量**：称 $X^*=\frac{X-E(X)}{\sqrt{D(X)}}$ 为 $X$ 的标准化随机变量。

🔔 标准化随机变量：$E(X^*)=0$，$D(X^*)=1$。

💎 **相关系数**：称 $R(X,Y)=\operatorname{Cov}(X^*,Y^*)$ 为 $X$ 和 $Y$ 的相关系数，计算式 $R(X,Y)=\frac{\operatorname{Cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$。

🔔 相关系数：

- $|R(X,Y)| \leqslant 1$。
- $|R(X,Y)|=1 \iff$ 存在常数 $a \neq 0$，$b$，使 $P(Y=aX+b)=1$。
- $R(X,Y)=1$ 完全正线性相关，$R(X,Y)=-1$ 完全负线性相关，$R(X,Y)=0$ 不相关。
    - 独立 $\implies$ 不相关，但不相关是指没有线性相关关系，可能有其他相关关系。

# 5 正态分布与自然指数分布族

## 正态分布

💎 **标准正态分布**：记为 $N(0,1)$，

- 密度函数

$$\phi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$$

- 分布函数

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t$$

💎 **正态分布**：记为 $N(\mu,\sigma^2)$，满足 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$，分布函数

$$F(x)=\Phi(\frac{x-\mu}{\sigma})$$

$3\sigma$ 原则：$X$ 落在 $(\mu-3\sigma,\mu+3\sigma)$ 的概率为 $0.9974$，因此常将该区间作为 $X$ 的实际取值区间。

🔔 正态分布：$X\sim(\mu,\sigma^2)$，

- $\Phi(x)$
  - $\Phi(0)=\frac{1}{2}$
  - $\Phi(x)+\Phi(-x)=1$
  - 无法表示为初等函数
- $\phi(x)$
  - 关于 $x=\mu$ 对称
  - 顶点为 $\max\{f(x)\}=\frac{1}{\sqrt{2\pi}}\sigma$
  - $\sigma$ 越小，分布越集中，曲线越"瘦高"
- $E(X)=\mu$，$D(x)=\sigma^2$
- $Y=aX+b\sim N(a\mu+b,a^2\sigma^2)$

🛡 **正态分布的可加性**：$n$ 个正态分布 $X_i\sim N(\mu_i,\sigma_i^2)$ 独立，有

$$Z=\sum C_iX_i\sim N(\sum C_i\mu_i,\sum C_i^2\sigma_i^2)$$

推论：$n$ 个变量 $X_i$ 独立同分布于 $N(\mu,\sigma^2)$，有

$$\overline{X}=\frac{1}{n}\sum X_i\sim N(\mu,\frac{\sigma^2}{n})$$

# 二维正态分布

💎 **二维正态分布**：设二维随机变量 $(X,Y)$ 有密度函数

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2}+\frac{(y-\mu_2)^2}{\sigma_2^2}-2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}]}$$

，称 $(X,Y)$ 服从二维正态分布，记为 $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;r)$。

🛡 二维正态分布：

- 如果 $(X,Y)\sim N(\mu_1,\mu_2;\sigma_1^2,\sigma_2^2;r)$，
  - $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$
  - $X$ 与 $Y$ 独立 $\iff r=0$
- $(X,Y)$ 服从二维正态分布 $\iff X$ 与 $Y$ 的任意非零线性组合 $Z=aX+bY$ 服从一维正态分布

# 6 极限定理

## 大数律

💎 **切比雪夫不等式**：

$$P(|X-E(X)|<\epsilon) \geqslant 1-\frac{D(X)}{\epsilon^2}$$

**大数律**：随机变量序列 $\{X_n\}$ 中，$E(X_n)=\mu_n$，$D(X_n)=\sigma^2$，如果 $n\to\infty$ 时有 $\sigma^2\to 0$，则 $X_n-\mu_n \xrightarrow{P} 0$。

**切比雪夫大数律**：对独立随机变量序列 $\{X_k\}$，如果 $E(X_k)$，$D(X_k)$ 存在且存在常数 $C$ 使 $D(X_k)<C$，则

$$\frac{1}{n}\sum_{k=1}^{n} X_k-\frac{1}{n}\sum_{k=1}^{n} E(X_k) \xrightarrow{P} 0$$

**辛钦大数律**（**独立同分布大数律**）：对独立同分布随机变量序列 $\{X_k\}$，$E(X_k)=\mu$，$D(X_k)=\sigma^2$，则

$$\overline{X} \xrightarrow{P} \mu$$

**伯努利大数律**：在 $n$ 次伯努利试验中，事件 $A$ 发生的频率为 $f_n(A)=\frac{n_A}{n}$，发生的概率为 $p=P(A)$，则

$$f_n(A) \xrightarrow{P} p=P(A)$$

## 中心极限定理

**独立同分布的中心极限定理**：$n$ 充分大时，$n$ 个独立同分布的随机变量之和近似服从正态分布，即

$$\sum_{k=1}^{n} X_k \dot{\sim} N(n\mu, n\sigma^2)$$

或

$$\frac{1}{n}\sum_{k=1}^{n} X_k \dot{\sim} N(\mu, \frac{\sigma^2}{n})$$

**二项分布的中心极限定理**：$n$ 充分大时，$n$ 个独立同分布的二项分布之和近似于正态分布，即

$$X_n \dot{\sim} N(np, np(1-p))$$

# 7 数理统计的基础知识

## $\chi^2$ 分布，$t$ 分布，$F$ 分布

💎 **$\chi^2$ 分布**：$X_1, \cdots, X_n$ 独立同服从 $N(0,1)$，称

$$\chi^2 = \sum_{i=1}^{n} X_i^2$$

服从自由度为 $n$ 的 $\chi^2$ 分布，记为 $\chi^2 \sim \chi^2(n)$.

🔔 $\chi^2$ 分布：

- $\chi^2(n)$ 等同于 $\Gamma(\frac{n}{2}, \frac{1}{2})$.
  - $\chi^2 \sim \chi^2(n) \Rightarrow E(\chi^2) = n$，$D(\chi^2) = 2n$.
- 可加性：$\chi^2(n_1) + \chi^2(n_2) \sim \chi^2(n_1 + n_2)$（独立）.

💎 **$t$ 分布**：$X \sim N(0,1)$，$Y \sim \chi^2(n)$，$X$ 与 $Y$ 独立，称

$$t = \frac{X}{\sqrt{Y/N}}$$

服从自由度为 $n$ 的 $t$ 分布，记为 $t \sim t(n)$.

🔔 $t$ 分布：

- $n = 1$ 时 $E(t)$ 不存在，$n \geqslant 2$ 时 $E(t) = 0$.
- $n$ 充分大（$\geqslant 45$）时，$t(n)$ 近似于 $N(0,1)$.
- $X \sim N(0,1) \to -X \sim N(0,1) \Rightarrow t \sim t(n) \to -t \sim t(n)$.

💎 **$F$ 分布**：$X \sim \chi^2(n)$，$Y \sim \chi^2(m)$，$X$ 与 $Y$ 独立，称

$$F = \frac{X/n}{Y/m}$$

服从自由度为 $(n, m)$ 的 $F$ 分布，记为 $F \sim F(n, m)$.

🔔 $F$ 分布：$F \sim F(n,m) \to \frac{1}{F} \sim F(m,n)$.

💎 **$p$ 分位点**：$0 < p < 1$，称满足 $F(a_p) = P(X \leqslant a_p) = p$ 的 $a_p$ 为 $X$ 的 $p$ 分位点，称 $a_{\frac{1}{2}}$ 为中位数.

📍 $p$ 分位点：

- $N(0,1)$：$u_p = -u_{1-p}$.
- $\chi^2(n)$：
  - $n \leqslant 45$ 时，查表.
  - $n > 45$ 时，$\chi_p^2(n) \approx \frac{1}{2}(u_p + \sqrt{2n-1})^2$.
- $t(n)$：
  - $t_p = -t_{1-p}$.
  - $n > 45$ 时，$t_p \approx u_p$.
- $F(n,m)$：$F_p(n,m) = \frac{1}{F_{1-p}(m,n)}$.

# 统计量

💎 **样本函数**：样本 $X_1, \cdots, X_n$ 的连续函数 $g(X_1, \cdots, X_n)$.

💎 **统计量**：不含未知参数的样本函数.

常用统计量：

- 均值：

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

- 方差：

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

- 标准差：

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

- $k$ 阶原点矩：

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

- $k$ 阶中心矩：

$$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$$

均值 $\overline{X}$ 为 1 阶原点矩 $A_1$，方差 $S^2$ 不是 2 阶中心矩 $B_2$.

# 抽样分布定理

## 一个正态总体

$X_i$ 来自正态总体 $N(\mu, \sigma^2)$:

- $$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$
- $$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
- $$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$
- $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$
- $\overline{X}$ 与 $S^2$ 独立.

$X_i$ 来自任意总体 $X$，$E(X) = \mu$，$D(X) = \sigma^2$，且 $n$ 充分大:

- $$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$
- $$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0,1)$$

## 两个正态总体

两个总体 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$:

- $$U = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

- $$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

  其中

  $$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

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

# 8 参数估计

一个总体 $X$ 的分布函数往往含有未知参数或未知参数向量 $\theta$，从而可记总体分布函数为 $F(x,\theta)$。

解决实际问题时需要了解未知参数或未知参数向量 $\theta$，因此可以利用样本提供的信息，对 $\theta$ 有一个基本的估计，这就是参数的估计问题。

参数的估计问题分为点估计和区间估计两大类。

## 点估计

💎 **点估计**：$X$ 的分布函数为 $F(x,\theta)$，抽取样本 $X_1,\cdots,X_n$，其观测值为 $x_1,\cdots,x_n$，构造某个统计量，用其观测值 $\hat{\theta}(x_1,\cdots,x_n)$ 估计 $\theta$，称 $\hat{\theta}(x_1,\cdots,x_n)$ 为 $\theta$ 的点估计值，称 $\hat{\theta}(X_1,\cdots,X_n)$ 为 $\theta$ 的点估计量，点估计值或点估计量称为 $\theta$ 的点估计，均可简记为 $\hat{\theta}$。

点估计分为矩估计、极大似然估计。

💎 **矩估计**：$X$ 的分布函数为 $F(x,\theta)$，$m=E(X)$ 一般为 $\theta$ 的函数，反解出 $\theta=g(m)$，用样本均值 $\overline{X}$ 代替，得到 $\hat{\theta}=g(\overline{x})$ 称为 $\theta$ 的矩估计值。

🛡 矩估计：$X$ 服从任何分布，记 $\mu=E(X)$，$\sigma^2=D(X)$，有：

- $\hat{\mu}=\overline{X}$
- $\hat{\sigma}^2=B_2$

💎 **似然函数**：

- 离散型：$L(\theta)=\Pi p(x_i,\theta)$

- 连续型：$L(\theta)=\Pi f(x_i,\theta)$

💎 **极大似然估计**：如果存在 $\theta$ 的一个取值 $\hat{\theta}(x_1,\cdots,x_n)=\theta$，使得 $L(\hat{\theta})=\max\{L(\theta)\}$，称 $\hat{\theta}$ 为 $\theta$ 的极大似然估计值，由 $\frac{\mathrm{d}(\ln L(\theta))}{\mathrm{d}\theta}=0$ 求解。

# 估计量的评选标准

💎 **无偏估计量**：如果 $\theta$ 的估计量 $\hat{\theta}$ 满足 $E(\hat{\theta})=\theta$，称 $\hat{\theta}$ 为无偏估计量。

💎 **渐进无偏估计量**：称 $b(\hat{\theta})=E(\hat{\theta})-\theta$ 为 $\hat{\theta}$ 的 **偏差**，如果样本容量 $n\to+\infty$ 时，$b(\hat{\theta})\to 0$，称 $\hat{\theta}$ 为渐进无偏估计量。

🛡️ 无偏估计量：$X$ 服从任何分布，

- $\mu=E(X)$，样本均值 $\overline{X}$ 为 $\mu$ 的无偏估计量。
- $m_k=E(X^k)$，样本 $k$ 阶矩 $A_k$ 为 $m_k$ 的无偏估计量。
- $\sigma^2=D(X)$，样本方差 $S^2$ 为 $\sigma^2$ 的无偏估计量。

💎 **有效性**：$\hat{\theta}_1$，$\hat{\theta}_2$ 为 $\theta$ 的无偏估计量，如果 $D(\hat{\theta}_1)\leqslant D(\hat{\theta}_2)$，称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效。

💎 **线性无偏估计**：$\forall(a_i)$，如果 $\sum a_i=1$，称 $\hat{\mu}=\sum a_iX_i$ 为 $\mu$ 的线性无偏估计。

样本均值 $\overline{X}$ 是最有效的线性无偏估计。

# 区间估计

💎 **置信区间**：$X$ 的分布函数为 $F(x,\theta)$，$\hat{\theta}_1$，$\hat{\theta}_2$ 为统计量，如果对给定概率 $1-\alpha$，有 $P(\hat{\theta}_1<\theta<\hat{\theta}_2)=1-\alpha$，称区间 $(\hat{\theta}_1,\hat{\theta}_2)$ 为 $\theta$ 的置信度为 $1-\alpha$ 的置信区间。

$\alpha$ 越小，置信度越高，区间越宽（误差越大）。

📍 一个正态总体下参数的置信区间：$X\sim N(\mu,\sigma^2)$，

| 估计 | 已知 | 置信区间($1-\alpha$) |
|---|---|---|
| $\mu$ | $\sigma_0$ | $(\overline{X}-u_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}},\overline{X}+u_{1-\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}})$ |
| | $s$ | $(\overline{X}-t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}},\overline{X}+t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}})$ |
| $\sigma^2$ | $s$ | $(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)})$ |

单侧置信限：

- $\mu$ 的置信度为 $1-\alpha$ 的单侧置信下限为 $\overline{X}-t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}$。

- $\sigma^2$ 的置信度为 $1-\alpha$ 的单侧置信上限为 $\frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)}$。

# 9 假设检验

## 假设检验

💎 假设检验（显著性检验）：

- **原假设**：$H_0$。
- **备择假设**：$H_1$。
- **显著性水平**：很小的正数 $\alpha$。
- **检验统计量**：服从已知分布的量，$U$ / $t$ / $\chi^2$。
- **临界值**：小概率事件发生的临界分位点。
- **拒绝域**：小概率事件发生时检验统计量的值域。

## 正态总体下的假设检验

- $\sigma^2$ 已知，检验 $\mu$：$U=\frac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}}\sim N(0,1)$

| $H_0$ | $H_1$ | 拒绝域 |
| --- | --- | --- |
| $\mu=\mu_0$ | $\mu\neq\mu_0$ | $\lvert U\rvert>u_{1-\frac{\alpha}{2}}$ |
| $\mu\leqslant\mu_0$ | $\mu>\mu_0$ | $U>u_{1-\alpha}$ |
| $\mu\geqslant\mu_0$ | $\mu<\mu_0$ | $U<-u_{1-\alpha}$ |

- $\sigma^2$ 未知，检验 $\mu$：$t=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$

| $H_0$ | $H_1$ | 拒绝域 |
| --- | --- | --- |
| $\mu=\mu_0$ | $\mu\neq\mu_0$ | $\lvert t\rvert>t_{1-\frac{\alpha}{2}}(n-1)$ |
| $\mu\leqslant\mu_0$ | $\mu>\mu_0$ | $t>t_{1-\alpha}(n-1)$ |
| $\mu\geqslant\mu_0$ | $\mu<\mu_0$ | $t<-t_{1-\alpha}(n-1)$ |

- $\mu$ 未知，检验 $\sigma^2$：$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$

| $H_0$ | $H_1$ | 拒绝域 |
| --- | --- | --- |
| $\sigma^2=\sigma_0^2$ | $\sigma^2\neq\sigma_0^2$ | $\chi^2<\chi^2_{\frac{\alpha}{2}}(n-1),\chi^2>\chi^2_{1-\frac{\alpha}{2}}(n-1)$ |
| $\sigma^2\leqslant\sigma_0^2$ | $\sigma^2>\sigma_0^2$ | $\chi^2>\chi^2_{1-\alpha}(n-1)$ |

| $H_0$ | $H_1$ | 拒绝域 |
| --- | --- | --- |
| $\sigma^2 \geqslant \sigma_0^2$ | $\sigma^2 < \sigma_0^2$ | $\chi^2 < \chi_\alpha^2(n-1)$ |