

信息学中的概率统计

王若松

前沿计算研究中心
北京大学

参数估计

1. 统计学的基本概念
2. 点估计
3. 区间估计

1. 统计学的基本概念

- ▶ 概率论：假设概率分布已知
- ▶ 统计学：利用观察到的数据，推断出数据服从的分布的性质

- ▶ 例：给定广告的点击数据。如何估计广告的平均点击率？
- ▶ 例：给出某型号的CPU的使用寿命数据。如何估计该型号的CPU的平均使用寿命？

- ▶ **总体**：研究对象的全体
- ▶ **个体**：总体的每个成员

- ▶ 通常只关心个体的数量指标值
- ▶ 用概率分布描述总体，数量指标值为服从该分布的随机变量 X

1. 统计学的基本概念

- ▶ 从总体中随机地抽取 n 个个体 x_1, x_2, \dots, x_n
- ▶ **样本**: x_1, x_2, \dots, x_n
- ▶ **样本量**: n
- ▶ 样本是随机变量, 也写作 X_1, X_2, \dots, X_n , X_i 与总体 X 分布相同
- ▶ **简单随机样本**: X_1, X_2, \dots, X_n 相互独立, 且均服从与总体 X 相同的分布
- ▶ 若 X 的分布函数为 $F(x)$, 样本的联合分布函数为 $F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i)$
- ▶ 统计学中, 主要考虑简单随机样本
- ▶ 本部分内容除非明确说明, 否则默认考虑简单随机样本的情况
- ▶ 统计学: 利用样本 x_1, x_2, \dots, x_n , 研究总体 X 的性质

1. 统计学的基本概念

- ▶ 统计量：不依赖于任何未知参数的样本的函数
- ▶ 样本均值： $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- ▶ 样本方差： $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$
- ▶ k 阶矩： $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$
- ▶ k 阶中心矩： $B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$
- ▶ 其他统计量： $\min\{X_1, X_2, \dots, X_n\}$, $\max\{X_1, X_2, \dots, X_n\}$, 中位数

2. 点估计

- ▶ **点估计**: 估计分布中所含有的未知参数 θ , 如 $\theta = E(X)$ 和 $\theta = \text{Var}(X)$
- ▶ 例: 给定广告的点击数据, 如何估计广告的平均点击率?
 - ▶ 给定样本 x_1, x_2, \dots, x_n , 如何估计 $\theta = E(X)$?
- ▶ **估计量**: 用来估计未知参数 θ 的统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$
- ▶ **估计值**: 给定样本取值 x_1, x_2, \dots, x_n , $\hat{\theta}(x_1, x_2, \dots, x_n)$ 为估计值
- ▶ 如何评判估计量的好坏?

2. 点估计

- ▶ 已知总体 X 数学期望 $E(X)$ 和方差 $\text{Var}(X)$ 均存在
- ▶ 考虑 $\theta = E(X)$ 的下述估计量
 - ▶ $\hat{\theta}_A(X_1, X_2, \dots, X_n) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
 - ▶ $\hat{\theta}_B(X_1, X_2, \dots, X_n) = X_1$
 - ▶ $\hat{\theta}_C(X_1, X_2, \dots, X_n) = \frac{1}{n+1} \sum_{i=1}^n X_i$
- ▶ 哪个估计量最好?

2. 点估计

- ▶ 给定参数 θ 的估计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$, 定义**偏差** $\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$
- ▶ 如果 $\text{Bias}(\hat{\theta}) = 0$, 则称估计量 $\hat{\theta}$ 是**无偏的**
- ▶ 当 $n = 1$ 时
- ▶ $\hat{\theta}_A = \hat{\theta}_B = X_1, \hat{\theta}_C = \frac{1}{2}X_1$
- ▶ $E(X_1) = E(X)$
- ▶ 如果 $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$, 则称估计量 $\hat{\theta}$ 是**渐进无偏的**

2. 点估计

- ▶ 已知总体 X 数学期望 $E(X)$ 和方差 $\text{Var}(X)$ 均存在, 考虑 $\theta = E(X)$ 的下述估计量
- ▶ $\hat{\theta}_A(X_1, X_2, \dots, X_n) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- ▶ $\hat{\theta}_B(X_1, X_2, \dots, X_n) = X_1$

- ▶ 给定参数 θ 的估计量 $\hat{\theta}$, 定义**均方误差** $\text{MSE}(\hat{\theta}) = E\left((\hat{\theta} - \theta)^2\right)$
- ▶ $\text{MSE}(\hat{\theta}) = (\text{Bias}(\hat{\theta}))^2 + \text{Var}(\hat{\theta})$
 - ▶ $\text{MSE}(\hat{\theta}) = E\left((\hat{\theta} - \theta)^2\right) = E\left(\left(\hat{\theta} - E(\hat{\theta})\right)^2\right) + (E(\hat{\theta}) - \theta)^2$
- ▶ 若估计量 $\hat{\theta}$ 无偏, 则 $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta})$

2. 点估计

- ▶ 已知总体 X 数学期望 $E(X)$ 和方差 $\text{Var}(X)$ 均存在, 考虑 $\theta = E(X)$ 的下述估计量
- ▶ $\hat{\theta}_A(X_1, X_2, \dots, X_n) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- ▶ $\hat{\theta}_B(X_1, X_2, \dots, X_n) = X_1$

- ▶ 给定参数 θ 的估计量 $\hat{\theta}$, 定义**均方误差** $\text{MSE}(\hat{\theta}) = E\left((\hat{\theta} - \theta)^2\right)$
- ▶ 若估计量 $\hat{\theta}$ 无偏, 则 $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta})$
- ▶ 计算 $\hat{\theta}_A$ 和 $\hat{\theta}_B$ 的均方误差
 - ▶ $\text{MSE}(\hat{\theta}_A) = \text{Var}(\hat{\theta}_A) = \text{Var}(X)/n$
 - ▶ $\text{MSE}(\hat{\theta}_B) = \text{Var}(\hat{\theta}_B) = \text{Var}(X)$

2. 点估计

- ▶ 给定参数 θ 的估计量 $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$, 若有 $\hat{\theta}_n \xrightarrow{P} \theta$, 也即对于任意 $\epsilon > 0$,
 $\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \geq \epsilon) = 0$, 则称估计量 $\hat{\theta}_n$ 为**一致估计量**
- ▶ $\hat{\theta}_A(X_1, X_2, \dots, X_n) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- ▶ $\hat{\theta}_B(X_1, X_2, \dots, X_n) = X_1$
- ▶ $\hat{\theta}_C(X_1, X_2, \dots, X_n) = \frac{1}{n+1} \sum_{i=1}^n X_i$
- ▶ 哪个估计量是一致的?

2. 点估计

- ▶ 若 $\lim_{n \rightarrow \infty} \text{MSE}(\hat{\theta}_n) \rightarrow 0$, 则 $\hat{\theta}_n$ 为一致估计量
 - ▶ 对于任意 $\epsilon > 0$, $P(|\hat{\theta}_n - \theta| \geq \epsilon) = P((\hat{\theta}_n - \theta)^2 \geq \epsilon^2) \leq \text{MSE}(\hat{\theta}_n)/\epsilon^2$
 - ▶ 左右两侧取极限即可
- ▶ 有偏估计量是否为一致估计量?
- ▶ 渐进无偏估计量是否为一致估计量?
- ▶ 一致估计量是否为渐进无偏估计量?

2. 点估计

- ▶ 令总体 $X \sim U(0, \theta)$
- ▶ 判断估计量 $\hat{\theta}_A = 2\bar{X}$ 和 $\hat{\theta}_B = \max\{X_1, X_2, \dots, X_n\}$ 的无偏性
- ▶ $E(\hat{\theta}_A) = 2 \cdot E(X) = \theta$
- ▶ $F_{\hat{\theta}_B}(x) = (F_X(x))^n = \left(\frac{x}{\theta}\right)^n$ 当 $0 \leq x \leq \theta$
- ▶ $E(\hat{\theta}_B) = \int_0^{+\infty} (1 - F(x)) dx = \int_0^{\theta} \left(1 - \left(\frac{x}{\theta}\right)^n\right) dx = \frac{n}{n+1} \theta$
- ▶ $\hat{\theta}_B$ 渐进无偏
- ▶ $\hat{\theta}_C = \max\{X_1, X_2, \dots, X_n\} \cdot \frac{n+1}{n}$ 是无偏估计量

2. 点估计

- ▶ 令总体 $X \sim U(0, \theta)$, 比较 $\hat{\theta}_A = 2\bar{X}$ 和 $\hat{\theta}_C = \max\{X_1, X_2, \dots, X_n\} \cdot \frac{n+1}{n}$ 的均方误差
- ▶ $MSE(\hat{\theta}_A) = \text{Var}(\hat{\theta}_A) = \frac{\text{Var}(2X)}{n} = \frac{\theta^2}{3n}$

2. 点估计

- ▶ 令总体 $X \sim U(0, \theta)$, 比较 $\hat{\theta}_A = 2\bar{X}$ 和 $\hat{\theta}_C = \max\{X_1, X_2, \dots, X_n\} \cdot \frac{n+1}{n}$ 的均方误差
- ▶ $\hat{\theta}_B = \max\{X_1, X_2, \dots, X_n\}$
- ▶ $F_{\hat{\theta}_B}(x) = (F_X(x))^n = \left(\frac{x}{\theta}\right)^n$ 当 $0 \leq x \leq \theta$
- ▶ $F_{\hat{\theta}_B^2}(x) = F_{\hat{\theta}_B}(\sqrt{x}) = \frac{x^{\frac{n}{2}}}{\theta^n}$ 当 $0 \leq x \leq \theta^2$
- ▶ $E(\hat{\theta}_B^2) = \int_0^{\theta^2} \left(1 - \frac{x^{\frac{n}{2}}}{\theta^n}\right) dx = \frac{n}{n+2} \theta^2$
- ▶ $\text{Var}(\hat{\theta}_B^2) = \frac{n}{n+2} \theta^2 - \left(\frac{n}{n+1} \theta\right)^2$
- ▶ $\text{MSE}(\hat{\theta}_C) = \text{Var}(\hat{\theta}_C) = \left(\frac{n+1}{n}\right)^2 \cdot \text{Var}(\hat{\theta}_B) = \frac{\theta^2}{n(n+2)}$

2. 点估计

- ▶ 令总体 $X \sim U(0, \theta)$, 比较 $\hat{\theta}_A = 2\bar{X}$ 和 $\hat{\theta}_C = \max\{X_1, X_2, \dots, X_n\} \cdot \frac{n+1}{n}$ 的均方误差
- ▶ $\text{MSE}(\hat{\theta}_A) = \frac{\theta^2}{3n}$
- ▶ $\text{MSE}(\hat{\theta}_C) = \frac{\theta^2}{n(n+2)}$

- ▶ $\hat{\theta}_A$ 和 $\hat{\theta}_C$ 是否为一致估计量?

2. 点估计

- ▶ k 阶矩: $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$
- ▶ k 阶中心矩: $B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$
- ▶ A_k 是否为 $E(X^k)$ 的无偏估计?
 - ▶ $E(A_k) = \frac{1}{n} \sum_{i=1}^n E(X_i^k) = E(X^k)$
- ▶ B_2 是否为 $\text{Var}(X)$ 的无偏估计?
 - ▶ $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i^2 + \bar{X}^2 - 2X_i\bar{X}) = \sum_{i=1}^n X_i^2 - n\bar{X}^2$
 - ▶ $B_2 = A_2 - \bar{X}^2 \Rightarrow E(B_2) = E(X^2) - E(\bar{X}^2)$
 - ▶ $E(\bar{X}^2) = (E(\bar{X}))^2 + \text{Var}(\bar{X}) = (E(X))^2 + \frac{\text{Var}(X)}{n}$
 - ▶ $E(B_2) = \frac{n-1}{n} \cdot \text{Var}(X)$

2. 点估计

- $E(B_2) = E\left(\frac{1}{n}\sum_{i=1}^n(X_i - \bar{X})^2\right) = \frac{n-1}{n} \cdot \text{Var}(X)$
- 样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^n(X_i - \bar{X})^2$
- $E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^n(X_i - \bar{X})^2\right) = \text{Var}(X)$
- 样本方差 S^2 是 $\text{Var}(X)$ 的无偏估计量
- 二阶中心矩 B_2 是 $\text{Var}(X)$ 的渐进无偏估计量

2. 点估计

- ▶ 总体 $X \sim N(\mu, \sigma^2)$ 。判断 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 和 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 的独立性。
- ▶ 考虑正交矩阵 U , 第一行每个元素均为 $\frac{1}{\sqrt{n}}$, 其余行任取
- ▶ 令随机向量 $X = (X_1, X_2, \dots, X_n)$, 令 $Y = UX$, 注意 Y 服从 n 维高斯分布, 且有
 - ▶ $E(Y) = (\sqrt{n}\mu, 0, 0, \dots, 0)$
 - ▶ $\text{Cov}(Y) = \sigma^2 \cdot I$
 - ▶ $\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n X_i^2$
- ▶ $\bar{X} = Y_1 / \sqrt{n}$
- ▶ $(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = (\sum_{i=1}^n X_i^2) - n\bar{X}^2 = (\sum_{i=1}^n Y_i^2) - Y_1^2 = \sum_{i=2}^n Y_i^2$
- ▶ $\text{Cov}(Y) = \sigma^2 \cdot I \Rightarrow Y_i$ 相互独立, 因此 \bar{X} 与 $(n-1)S^2$ 相互独立
- ▶ $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ 。 $\frac{(n-1)S^2}{\sigma^2}$ 服从何种分布?

2. 点估计

- $(n - 1)S^2 = \sum_{i=2}^n Y_i^2 \Rightarrow \frac{(n-1)S^2}{\sigma^2} = \sum_{i=2}^n \left(\frac{Y_i}{\sigma}\right)^2$
- Y_i 相互独立且对于 $i \geq 2$, $Y_i \sim N(0, \sigma^2)$
- 作业五: 若 Z_1, Z_2, \dots, Z_n 独立同分布且 $Z_i \sim N(0, 1)$, 则 $\sum_{i=1}^n Z_i^2 \sim \chi^2(n)$
- 因此有 $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n - 1)$

2. 点估计

- ▶ **点估计**: 估计分布中所含有的未知参数 θ
 - ▶ 已知总体 $X \sim U(0, \theta)$, 估计 θ
 - ▶ 已知总体 $X \sim N(\mu, \sigma^2)$, 估计 μ 和 σ^2
- ▶ **估计量**: 用来估计未知参数 θ 的统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$
- ▶ 设计估计量的常用方法
 - ▶ 矩法
 - ▶ 最大似然估计

2. 点估计

- ▶ 矩法：用样本矩去替换总体矩
- ▶ 1. 将未知参数 $\theta_1, \theta_2, \dots, \theta_k$ 表示为总体前 k 阶矩的函数
 - ▶ $\theta_i = f_i(\mu_1, \mu_2, \dots, \mu_k)$
 - ▶ μ_i 是总体 X 的*i*阶原点矩 $E(X^i)$ 或*i*阶中心矩 $E\left((X - E(X))^i\right)$
- ▶ 2. 用样本的*i*阶矩或*i*阶中心矩替换 μ_i
 - ▶ $\hat{\theta}_i = f_i(A_1, A_2, \dots, A_k)$ 或 $\hat{\theta}_i = f_i(B_1, B_2, \dots, B_k)$
 - ▶ k 阶矩： $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$
 - ▶ k 阶中心矩： $B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$
- ▶ 也可以用样本方差 S^2 替换 $\text{Var}(X)$

2. 点估计

- ▶ 矩法：用样本矩去替换总体矩
- ▶ 例1：已知总体 $X \sim U(0, \theta)$ 。用矩法设计 θ 的估计量
 - ▶ $\theta = 2E(X) \Rightarrow \hat{\theta}(X_1, X_2, \dots, X_n) = 2\bar{X}$
- ▶ 例2：已知总体 $X \sim U(a, b)$ 。用矩法设计 a 和 b 的估计量
 - ▶ $a + b = 2E(X)$
 - ▶ $\text{Var}(X) = \frac{(b-a)^2}{12} \Rightarrow b - a = \sqrt{12 \cdot \text{Var}(X)}$
 - ▶ $a = E(X) - \sqrt{3 \cdot \text{Var}(X)}, b = E(X) + \sqrt{3 \cdot \text{Var}(X)}$
 - ▶ $\hat{a} = \bar{X} - \sqrt{3 \cdot S^2}, \hat{b} = \bar{X} + \sqrt{3 \cdot S^2}$

2. 点估计

► 例3：已知总体 $X \sim \text{Exp}(\lambda)$ 。用矩法设计 λ 的估计量

- $E(X) = \frac{1}{\lambda} \Rightarrow \hat{\lambda} = \frac{1}{\bar{X}}$
- $\text{Var}(X) = \frac{1}{\lambda^2} \Rightarrow \hat{\lambda} = \frac{1}{\sqrt{S^2}}$

► 例4：已知总体 $X \sim N(\mu, \sigma^2)$ 。用矩法设计 μ 和 σ^2 的估计量

- $\mu = E(X), \sigma^2 = \text{Var}(X) \Rightarrow \hat{\mu} = \bar{X}, \hat{\sigma^2} = S^2$ 或 $\hat{\sigma^2} = B_2$
- $\mu = E(X), E(X^2) = \mu^2 + \sigma^2 \Rightarrow \hat{\mu} = \bar{X}, \hat{\sigma^2} = A_2 - \bar{X}^2$

2. 点估计

- ▶ 例4：已知总体 $X \sim N(\mu, \sigma^2)$ 。用矩法设计 μ 和 σ^2 的估计量
- ▶ σ^2 的估计量： $S^2, B_2, A_2 - \bar{X}^2$
- ▶ 比较三者的均方误差？
- ▶ $B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2 = A_2 - \bar{X}^2$
- ▶ $MSE(S^2) = \text{Var}(S^2)$, 而 $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$,
- ▶ $MSE(S^2) = \text{Var}(S^2) = \frac{2(n-1)}{(n-1)^2} \cdot \sigma^4 = \frac{2}{n-1} \cdot \sigma^4$
- ▶ $MSE(B_2) = E((B_2 - \sigma^2)^2) = E(B_2^2) + \sigma^4 - 2\sigma^2 E(B_2)$
- ▶ $E(B_2^2) = \left(\frac{n-1}{n}\right)^2 \cdot E((S^2)^2) = \left(\frac{n-1}{n}\right)^2 \cdot \frac{\sigma^4}{(n-1)^2} \cdot ((n-1)^2 + 2(n-1)) = \frac{n^2-1}{n^2} \cdot \sigma^4$
- ▶ $E(B_2) = \frac{\sigma^2(n-1)}{n} \Rightarrow MSE(B_2) = \frac{2n-1}{n^2} \sigma^4$

2. 点估计

- ▶ **点估计**: 给定样本 x_1, x_2, \dots, x_n , 估计分布中所含有的未知参数 θ
- ▶ **最大似然估计**: 选择参数 θ , 使得似然函数 $L(\theta) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$ 最大
- ▶ **最大似然估计量**: 满足 $L(\hat{\theta}) = \max_{\theta} L(\theta)$ 的统计量 $\hat{\theta}$
- ▶ 例1: 已知某网站不同用户点击广告的情况服从参数为 p 的伯努利分布。给定简单随机样本 $x_1, x_2, \dots, x_n \in \{0, 1\}$, 求未知参数 p 的最大似然估计。

2. 点估计

- ▶ **最大似然估计**: 选择参数 θ , 使得似然函数 $L(\theta) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$ 最大
- ▶ 例1: 已知某网站不同用户点击广告的情况服从参数为 p 的伯努利分布。给定简单随机样本 $x_1, x_2, \dots, x_n \in \{0,1\}$, 求未知参数 p 的最大似然估计。
 - ▶ $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}$
 - ▶ 选择 p 最大化 $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; p)$
 - ▶ 对 p 求导, $\sum x_i \cdot p^{\sum x_i - 1} (1-p)^{n-\sum x_i} - (n - \sum x_i) \cdot p^{\sum x_i} (1-p)^{n-\sum x_i - 1} = 0$
 - ▶ $\sum x_i (1-p) - (n - \sum x_i) \cdot p = 0 \Rightarrow p = \sum x_i / n$
 - ▶ $\hat{p}_{MLE} = \bar{X}$

2. 点估计

- ▶ **最大似然估计 (MLE)** : 选择参数 θ , 使得似然函数 $L(\theta) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$ 最大
- ▶ 例2: 某超算中心每日收到的任务数量服从 $\pi(\lambda)$ 。给定简单随机样本 x_1, x_2, \dots, x_n , 求未知参数 λ 的最大似然估计。
 - ▶ $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod x_i!}$
 - ▶ 选择 p 最大化 $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \lambda)$
 - ▶ 求导, $e^{-n\lambda} \lambda^{\sum x_i - 1} \cdot \sum x_i - n \cdot e^{-n\lambda} \lambda^{\sum x_i} = 0 \Rightarrow \lambda = \sum x_i / n$
 - ▶ $\hat{\lambda}_{MLE} = \bar{X}$

2. 点估计

- ▶ **最大似然估计 (MLE)** : 选择参数 θ , 使得似然函数 $L(\theta) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$ 最大
- ▶ 对于简单随机样本, $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta) = \prod_{i=1}^n P(X_i = x_i; \theta)$
- ▶ 最大化 $L(\theta)$ 等价于最大化对数似然函数 $\ln L(\theta) = \sum_{i=1}^n \ln P(X_i = x_i; \theta)$
- ▶ 例2: 某超算中心每日收到的任务数量服从 $\pi(\lambda)$ 。给定简单随机样本 x_1, x_2, \dots, x_n , 求未知参数 λ 的最大似然估计。
 - ▶ $\ln L(\lambda) = \sum_{i=1}^n \ln P(X_i = x_i; \lambda) = \sum_{i=1}^n (-\lambda + x_i \ln \lambda - \ln(x_i!))$
 - ▶ 对 λ 求导, $-n + \frac{\sum x_i}{\lambda} = 0 \Rightarrow \lambda = \sum x_i / n$

2. 点估计

- ▶ 当总体 X 为连续型, 定义**似然函数** $L(\theta) = f(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$
- ▶ **最大似然估计 (MLE)** : 选择参数 θ , 使得**似然函数** $L(\theta) = f(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$ 最大
- ▶ 例3: 令总体 X 服从 $U(0, \theta)$ 。给定简单随机样本 x_1, x_2, \dots, x_n , 求未知参数 θ 的最大似然估计。
- ▶ $L(\theta) = \prod_{i=1}^n \frac{1_{x_i \leq \theta}}{\theta} = \frac{1_{x_1 \leq \theta, x_2 \leq \theta, \dots, x_n \leq \theta}}{\theta^n}$
- ▶ 最大化 $L(\theta) \Rightarrow \theta = \max\{x_1, x_2, \dots, x_n\}$

2. 点估计

- ▶ 当总体 X 为连续型, 定义**似然函数** $L(\theta) = f(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$
- ▶ **最大似然估计 (MLE)** : 选择参数 θ , 使得**似然函数** $L(\theta) = f(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$ 最大
- ▶ 例4: 令总体 X 服从 $N(\mu, \sigma^2)$ 。给定简单随机样本 x_1, x_2, \dots, x_n , 求未知参数 μ 的和 σ^2 最大似然估计。

- ▶ $\ln L(\mu, \sigma^2) = \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) = \sum_{i=1}^n \left(-\frac{(x_i - \mu)^2}{2\sigma^2} - \frac{\ln \sigma^2}{2} - \frac{\ln 2\pi}{2} \right)$
- ▶ $\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = 0 \Rightarrow \mu = \frac{\sum x_i}{n} = \bar{X}$
- ▶ $\frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = 0 \Rightarrow \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0 \Rightarrow \sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \mu)^2 = B_2$

2. 点估计

- ▶ 例4：令总体 X 服从 $N(\mu, \sigma^2)$ 。给定简单随机样本 x_1, x_2, \dots, x_n ，求未知参数 μ 的和 σ^2 最大似然估计。
- ▶ $\hat{\mu}_{MLE} = \frac{\sum x_i}{n} = \bar{X}$
- ▶ $\widehat{\sigma^2}_{MLE} = \frac{1}{n} \cdot \sum_{i=1}^n \left(x_i - \frac{\sum x_i}{n} \right)^2 = B_2$
- ▶ 如何求 σ 的最大似然估计？
 - ▶ $\hat{\sigma}_{MLE} = \sqrt{B_2}$
- ▶ 最大似然估计的不变性：给定 θ 的最大似然估计 $\hat{\theta}$ 。若 $g(\theta)$ 有单值反函数，则 $g(\theta)$ 的最大似然估计为 $g(\hat{\theta})$ 。

2. 点估计

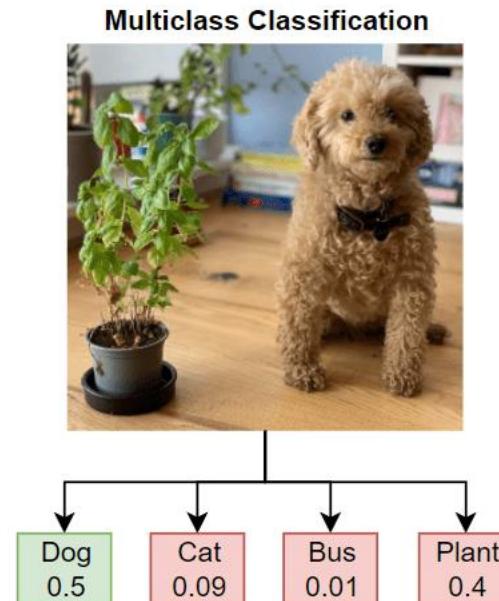
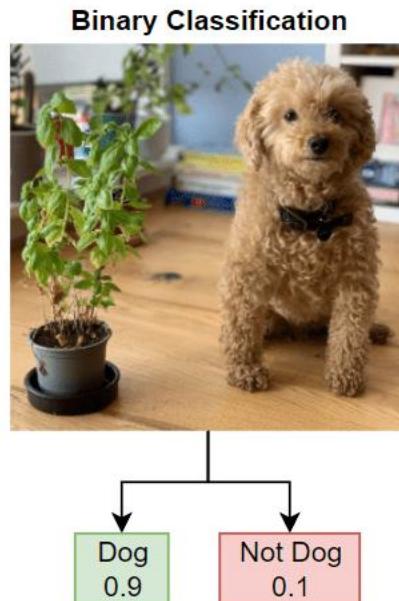
- ▶ 当总体 X 为连续型, 定义**似然函数** $L(\theta) = f(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$
- ▶ **最大似然估计 (MLE)** : 选择参数 θ , 使得**似然函数** $L(\theta) = f(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$ 最大
- ▶ 例5: 总体 X 的概率密度函数为
 - ▶ $f(x) = \frac{1}{\theta} \cdot e^{-(x-\mu)/\theta}$ 若 $x \geq \mu$
 - ▶ 否则 $f(x) = 0$
- ▶ 给定简单随机样本 x_1, x_2, \dots, x_n , 求未知参数 μ 和 θ 的最大似然估计和矩法估计。
- ▶ 矩法: $X - \mu \sim \text{Exp}\left(\frac{1}{\theta}\right) \Rightarrow E(X) = \mu + \theta, \text{Var}(X) = \theta^2$
- ▶ $\hat{\theta} = \sqrt{B_2}, \mu = \bar{X} - \sqrt{B_2}$

2. 点估计

- ▶ 例5：总体 X 的概率密度函数为
 - ▶ $f(x) = \frac{1}{\theta} \cdot e^{-(x-\mu)/\theta}$ 若 $x \geq \mu$
 - ▶ 否则 $f(x) = 0$
- ▶ 给定简单随机样本 x_1, x_2, \dots, x_n , 求未知参数 μ 和 θ 的最大似然估计和矩法估计。
 - ▶ $L(\mu, \theta) = \frac{1}{\theta^n} \cdot e^{-\frac{1}{\theta} \sum (x_i - \mu)}$ $\cdot 1_{x_1 \geq \mu, x_2 \geq \mu, \dots, x_n \geq \mu}$
 - ▶ 当 $x_1 \geq \mu, x_2 \geq \mu, \dots, x_n \geq \mu$, $L(\mu, \theta)$ 关于 μ 为增函数
 - ▶ $\hat{\mu}_{MLE} = \min\{x_1, x_2, \dots, x_n\}$
 - ▶ $\ln L(\mu, \theta) = -n \ln \theta - \frac{1}{\theta} \sum (x_i - \mu)$
 - ▶ 求导, $-\frac{n}{\theta} + \frac{1}{\theta^2} \sum (x_i - \mu) = 0 \Rightarrow \hat{\theta}_{MLE} = \bar{X} - \mu = \bar{X} - \min\{x_1, x_2, \dots, x_n\}$

2. 点估计

- ▶ **最大似然估计**: 选择参数 θ , 使得似然函数 $L(\theta) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$ 最大
- ▶ **分类问题**: 给定数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 。其中 $x_i \in \mathcal{X}$ 为数据, $y_i \in \{1, 2, \dots, C\}$ 为分类标签。利用数据学习分类函数 $f: \mathcal{X} \rightarrow \{1, 2, \dots, C\}$ 。



2. 点估计

- ▶ **分类问题**: 给定数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 。利用数据学习分类函数 $f: \mathcal{X} \rightarrow \{1, 2, \dots, C\}$ 。
- ▶ **机器学习模型**: 给定参数 θ , 对于给定的数据 $x \in \mathcal{X}$, 模型 f_θ 输出 $\{1, 2, \dots, C\}$ 上的一个分布
- ▶ $f_\theta(y|x)$ 表示给定输入数据 $x \in \mathcal{X}$, 分类为 y 的概率
- ▶ **分类函数**: $\operatorname{argmax}_y f_\theta(y|x)$
- ▶ 给定数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, 如何选取参数 θ ?

2. 点估计

- ▶ **最大似然估计**: 选择参数 θ , 使得似然函数 $L(\theta) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n; \theta)$ 最大
- ▶ **分类问题**: 给定数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 。利用数据学习分类函数 $f: \mathcal{X} \rightarrow \{1, 2, \dots, C\}$ 。
- ▶ **机器学习模型**: 给定参数 θ , 对于给定的数据 $x \in \mathcal{X}$, 模型 f_θ 输出 $\{1, 2, \dots, C\}$ 上的一个分布
- ▶ 对于分类问题: 最大化似然函数 $L(\theta) = \prod_{i=1}^n f_\theta(y_i | x_i)$
- ▶ 实践中, 通常改为最小化 $-\ln L(\theta) = \sum_{i=1}^n -\ln f_\theta(y_i | x_i)$

2. 点估计

- ▶ 对于两个分布 p 和 q , 定义**交叉熵** $H(p, q) = -E_p(\ln q)$
 - ▶ 对于离散分布, $H(p, q) = -\sum_x p(x) \ln q(x)$
 - ▶ 对于连续分布, $H(p, q) = -\int P(x) \ln Q(x) dx$, $P(x)$ 和 $Q(x)$ 为概率密度函数
- ▶ p 为数据真实标签对应的分布, 也即 $p(y|x_i) = 1_{y=y_i}$
- ▶ $-\ln L(\theta) = \sum_{i=1}^n -\ln f_\theta(y_i|x_i) = \sum_{i=1}^n H_i(p, f_\theta)$
 - ▶ H_i 为数据 x_i 真实标签分布和模型输出分布的交叉熵
 - ▶ $H_i(p, f_\theta) = -\sum_y p(y|x_i) \ln f_\theta(y|x_i) = -\ln f_\theta(y_i|x_i)$
- ▶ 结论: 在分类问题中, 最大似然估计等价于最小化交叉熵损失函数

2. 点估计

- ▶ **机器学习模型**: 给定参数 θ , 对于给定的数据 $x \in \mathcal{X}$, 模型 f_θ 输出 $\{1, 2, \dots, C\}$ 上的一个分布
 - ▶ 满足 $f_\theta(y|x) \geq 0$ 和 $\sum_y f_\theta(y|x) = 1$
- ▶ 如何放松上述要求?
- ▶ $\text{softmax}_i(a_1, a_2, \dots, a_C) = \frac{e^{a_i}}{\sum_j e^{a_j}}$
- ▶ 给定 $f_\theta: \mathcal{X} \rightarrow \mathbb{R}^C$, 将softmax应用于 f_θ 的输出即可得到 $\{1, 2, \dots, C\}$ 上的分布
- ▶ 实践中, 交叉熵损失函数往往与softmax配合使用

3. 区间估计

- ▶ **点估计**: 给定样本 x_1, x_2, \dots, x_n , 估计分布中所含有的未知参数 θ
- ▶ **区间估计**: 设计统计量 $\hat{\theta}_L(X_1, X_2, \dots, X_n)$ 和 $\hat{\theta}_U(X_1, X_2, \dots, X_n)$, 使得 $\theta \in [\hat{\theta}_L, \hat{\theta}_U]$ 的概率尽量大
- ▶ $P(\theta \in [\hat{\theta}_L, \hat{\theta}_U])$ 越大, 则导致区间长度 $\hat{\theta}_U - \hat{\theta}_L$ 更长
- ▶ **置信区间**: 若统计量 $\hat{\theta}_L(X_1, X_2, \dots, X_n)$ 和 $\hat{\theta}_U(X_1, X_2, \dots, X_n)$, 对于任意 θ , 满足 $P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) \geq 1 - \alpha$, 则 $[\hat{\theta}_L, \hat{\theta}_U]$ 为 θ 的**置信水平**为 $1 - \alpha$ 的**置信区间**。
 - ▶ $\hat{\theta}_L$: 置信下限
 - ▶ $\hat{\theta}_U$: 置信上限

3. 区间估计

- ▶ **置信区间**: 若统计量 $\hat{\theta}_L(X_1, X_2, \dots, X_n)$ 和 $\hat{\theta}_U(X_1, X_2, \dots, X_n)$, 对于任意 θ , 满足 $P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) \geq 1 - \alpha$, 则 $[\hat{\theta}_L, \hat{\theta}_U]$ 为 θ 的置信水平为 $1 - \alpha$ 的置信区间。
- ▶ **单侧置信下限**: 若统计量 $\hat{\theta}_L(X_1, X_2, \dots, X_n)$, 对于任意 θ , 满足 $P(\hat{\theta}_L \leq \theta) \geq 1 - \alpha$, 则 $\hat{\theta}_L$ 为 θ 的置信水平为 $1 - \alpha$ 的单侧置信下限。
- ▶ **单侧置信上限**: 若统计量 $\hat{\theta}_U(X_1, X_2, \dots, X_n)$, 对于任意 θ , 满足 $P(\theta \leq \hat{\theta}_U) \geq 1 - \alpha$, 则 $\hat{\theta}_U$ 为 θ 的置信水平为 $1 - \alpha$ 的单侧置信上限。
- ▶ 若 $\hat{\theta}_L$ 为 θ 的置信水平为 $1 - \alpha_1$ 的单侧置信下限, $\hat{\theta}_U$ 为 θ 的置信水平为 $1 - \alpha_2$ 的单侧置信上限, 则 $[\hat{\theta}_L, \hat{\theta}_U]$ 为 θ 的置信水平为 $1 - \alpha_1 - \alpha_2$ 的置信区间。

3. 区间估计

- ▶ 例1：总体 $X \sim N(\mu, \sigma^2)$, σ^2 已知。设计 μ 的置信水平为 $1 - \alpha$ 的置信区间
- ▶ 利用 \bar{X} 构造置信区间？
- ▶ 注意到 $\bar{X} \sim N(\mu, \sigma^2/n)$, 也即 $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$
- ▶ $P\left(c \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq d\right) = 1 - \alpha \Rightarrow P\left(\bar{X} - \frac{d\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - \frac{c\sigma}{\sqrt{n}}\right) = 1 - \alpha$
- ▶ 如何选择 c 与 d 使得区间最短？
 - ▶ 令 $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ 为标准正态分布的分布函数
 - ▶ 取 $c = \Phi^{-1}(\alpha/2)$, $d = \Phi^{-1}(1 - \alpha/2)$, 注意到 $c + d = 0$
- ▶ 结论： $P\left(\bar{X} - \frac{\Phi^{-1}(1-\alpha/2)\cdot\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{\Phi^{-1}(1-\alpha/2)\cdot\sigma}{\sqrt{n}}\right) = 1 - \alpha$

3. 区间估计

- ▶ 例1：总体 $X \sim N(\mu, \sigma^2)$, σ^2 已知。设计 μ 的置信水平为 $1 - \alpha$ 的置信区间
- ▶ \bar{X} 为无偏估计量。基于集中不等式构造置信区间？
- ▶ 回顾：若 $X \sim N(\mu, \sigma^2)$

- ▶ $P(X - E(X) \geq k\sigma) \leq e^{-\frac{k^2}{2}}$
- ▶ $P(X - E(X) \leq -k\sigma) \leq e^{-\frac{k^2}{2}}$
- ▶ $\bar{X} - \mu \sim N(0, \sigma^2/n) \Rightarrow P\left(|\bar{X} - \mu| \geq \frac{k\sigma}{\sqrt{n}}\right) \leq 2e^{-\frac{k^2}{2}}$
- ▶ $2e^{-\frac{k^2}{2}} = \alpha \Rightarrow k = \sqrt{2\ln(2/\alpha)}$
- ▶ 结论： $P\left(\bar{X} - \frac{\sigma\sqrt{2\ln(2/\alpha)}}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{\sigma\sqrt{2\ln(2/\alpha)}}{\sqrt{n}}\right) \leq 1 - \alpha$
- ▶ 与 $P\left(\bar{X} - \frac{\Phi^{-1}(1-\alpha/2)\cdot\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{\Phi^{-1}(1-\alpha/2)\cdot\sigma}{\sqrt{n}}\right) = 1 - \alpha$ 的联系？

3. 区间估计

- ▶ **置信区间**: 若统计量 $\hat{\theta}_L(X_1, X_2, \dots, X_n)$ 和 $\hat{\theta}_U(X_1, X_2, \dots, X_n)$, 对于任意 θ 满足 $P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) \geq 1 - \alpha$, 则 $[\hat{\theta}_L, \hat{\theta}_U]$ 为 θ 的置信水平为 $1 - \alpha$ 的置信区间。
- ▶ **枢轴量法**
- ▶ 1. 设计枢轴量 $G = G(x_1, x_2, \dots, x_n)$, 使得 G 的分布与未知参数 θ 无关。
- ▶ 2. 选择 c 和 d , $P(c \leq G \leq d) = 1 - \alpha$, $c \leq G \leq d$ 可变形为 $\hat{\theta}_L \leq \theta \leq \hat{\theta}_U$
- ▶ 3. 得到结论 $P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha$
- ▶ 设计枢轴量通常从 θ 的点估计出发
- ▶ 通常选择 c 和 d 使得 $P(G < c) = P(G > d) = \alpha/2$

3. 区间估计

- ▶ 例2：总体 $X \sim U(0, \theta)$ 。设计 θ 的置信水平为 $1 - \alpha$ 的置信区间。
- ▶ 令 $\hat{\theta} = \max\{X_1, X_2, \dots, X_n\}$ 为 θ 的最大似然估计
- ▶ $F_{\hat{\theta}}(x) = (F_X(x))^n = \left(\frac{x}{\theta}\right)^n$ 当 $0 \leq x \leq \theta$
- ▶ 枢轴量 $G = \hat{\theta}/\theta$
- ▶ $F_G(x) = x^n$ 当 $0 \leq x \leq 1$
- ▶ $P(c \leq G \leq d) = 1 - \alpha \Rightarrow P\left(\frac{\hat{\theta}}{d} \leq \theta \leq \frac{\hat{\theta}}{c}\right) = 1 - \alpha$
- ▶ $d^n - c^n = 1 - \alpha \Rightarrow d = 1, c = \alpha^{1/n}$
- ▶ 结论： $P(\hat{\theta} \leq \theta \leq \hat{\theta} \cdot \alpha^{-1/n}) = 1 - \alpha$

3. 区间估计

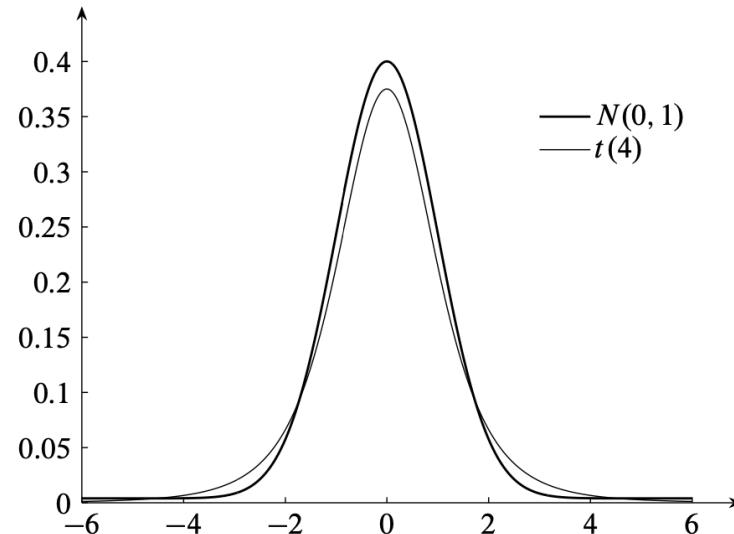
- 例3：总体 $X \sim N(\mu, \sigma^2)$ 。设计 σ^2 的置信水平为 $1 - \alpha$ 的置信区间。
- 回顾： S^2 是 σ^2 的无偏估计， $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
- 令 $F(x)$ 为 $\chi^2(n-1)$ 的分布函数，取 $c = F^{-1}(\alpha/2), d = F^{-1}(1-\alpha/2)$
- 则有 $P\left(c \leq \frac{(n-1)S^2}{\sigma^2} \leq d\right) = 1 - \alpha \Rightarrow P\left(\frac{(n-1)S^2}{d} \leq \sigma^2 \leq \frac{(n-1)S^2}{c}\right) = 1 - \alpha$

3. 区间估计

- ▶ 例4：总体 $X \sim N(\mu, \sigma^2)$, μ, σ^2 均为未知参数。设计 μ 的置信水平为 $1 - \alpha$ 的置信区间。
- ▶ $\frac{\bar{X} - \mu}{\sqrt{S^2}}$ 服从何种分布？
- ▶ 回顾： $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, 且 \bar{X} 和 S^2 相互独立
- ▶ **t 分布**：令 $X_1 \sim N(0, 1)$, $X_2 \sim \chi^2(n)$, X_1 和 X_2 相互独立。称 $T = \frac{X_1}{\sqrt{X_2/n}}$ 的分布为自由度为 n 的 **t 分布**, 记为 $T \sim t(n)$

3. 区间估计

- ▶ **t 分布**: 令 $X_1 \sim N(0,1)$, $X_2 \sim \chi^2(n)$, X_1 和 X_2 相互独立。称 $T = \frac{X_1}{\sqrt{X_2/n}}$ 的分布为自由度为 n 的 **t 分布**, 记为 $T \sim t(n)$
- ▶ 自由度为 n 的 t 分布概率密度函数: $f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$
- ▶ 当 $n = 1$, $f(t) = \frac{1}{1+t^2} \cdot \frac{1}{\pi}$
- ▶ 当自由度 n 较大, 近似为标准正态分布



3. 区间估计

- ▶ 例4：总体 $X \sim N(\mu, \sigma^2)$ 且 μ, σ^2 均为未知参数。设计 μ 的置信水平为 $1 - \alpha$ 的置信区间。
- ▶ **t 分布：**令 $X_1 \sim N(0,1)$, $X_2 \sim \chi^2(n)$, X_1 和 X_2 相互独立。称 $T = \frac{X_1}{\sqrt{X_2/n}}$ 的分布为自由度为 n 的 **t 分布**, 记为 $T \sim t(n)$
- ▶ 回顾: $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, 且 \bar{X} 和 S^2 相互独立
- ▶ $\frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{S^2}} \sim t(n-1)$
- ▶ $P\left(c \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{S^2}} \leq d\right) = 1 - \alpha \Rightarrow P\left(\bar{X} - d \cdot \frac{\sqrt{S^2}}{\sqrt{n}} \leq \mu \leq \bar{X} - c \cdot \frac{\sqrt{S^2}}{\sqrt{n}}\right) = 1 - \alpha$
- ▶ 令 $F(x)$ 为 $t(n-1)$ 的分布函数, 取 $c = F^{-1}(\alpha/2)$, $d = F^{-1}(1 - \alpha/2)$
- ▶ 注意到 $c + d = 0$

3. 区间估计

- ▶ 例5：总体 $X \sim B(1, p)$ 。设计 p 的置信水平为 $1 - \alpha$ 的置信区间。
- ▶ 由中心极限定理， $\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}}$ 近似服从 $N(0,1)$
- ▶ $P\left(-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \leq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) \approx 1 - \alpha$
 - ▶ $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ 为标准正态分布的分布函数
- ▶ 解方程可得到置信水平为 $1 - \alpha$ 的近似置信区间

3. 区间估计

- ▶ 例5：总体 $X \sim B(1, p)$ 。设计 p 的置信水平为 $1 - \alpha$ 的置信区间。
- ▶ \bar{X} 为无偏估计
- ▶ Chernoff bound: $P(|\bar{X} - p| > \epsilon) \leq 2e^{-2n\epsilon^2}$
- ▶ $2 \cdot e^{-2n\epsilon^2} = \alpha \Rightarrow \epsilon = \sqrt{\ln(2/\alpha)/2n}$
- ▶ $P(\bar{X} - \sqrt{\ln(2/\alpha)/2n} \leq p \leq \bar{X} + \sqrt{\ln(2/\alpha)/2n}) \geq 1 - \alpha$

3. 区间估计

- ▶ 有 n 台游戏机，第 i 台游戏机中奖概率为未知参数 p_i 。
- ▶ 每一轮可从 n 台游戏机选择一台并进行游戏
- ▶ 用最少的轮数，找到中奖概率最高的游戏机
- ▶ 应用：推荐系统，医疗，...

- ▶ 第 t 轮中：
 - ▶ 选择游戏机 i
 - ▶ 观测到结果 $X_t \sim B(1, p_i)$
- ▶ 输出：中奖概率最高的游戏机 $o \in \{1, 2, \dots, n\}$

- ▶ 均匀采样：对每台游戏机采集相同数量的样本 N ，返回样本均值最大的游戏机
- ▶ 要求：返回游戏机 $o \in \{1, 2, \dots, n\}$ ，使得 $P(p_o \geq \max_i p_i - \epsilon) \geq 2/3$ 。 N 应当取多大？



3. 区间估计

- ▶ 若对第 i 台游戏机采集 N 个样本
- ▶ $P\left(\bar{X}_i - \sqrt{\ln(2/\alpha)/2N} \leq p_i \leq \bar{X}_i + \sqrt{\ln(2/\alpha)/2N}\right) \geq 1 - \alpha$
- ▶ 若取 $N = O(\ln(1/\alpha)/\epsilon^2)$, 则有 $P(|p_i - \bar{X}_i| \leq \epsilon) \geq 1 - \alpha$
- ▶ α 应当如何选择?
- ▶ 取 $\alpha = \frac{1}{3n}$, $N = O(\ln n / \epsilon^2)$, 则 $P(|p_i - \bar{X}_i| \leq \epsilon) \geq 1 - \frac{1}{3n}$
- ▶ Union bound: $P(\forall i, |p_i - \bar{X}_i| \leq \epsilon) \geq 2/3$
- ▶ 此时, 样本均值最大的游戏机 o 满足 $p_o \geq \max_i p_i - 2\epsilon$
- ▶ $p_o \geq \bar{X}_o - \epsilon \geq \bar{X}_i - \epsilon \geq p_i - 2\epsilon$

3. 区间估计

- ▶ 均匀采样：对每台游戏机采集相同数量的样本 N ，返回样本均值最大的游戏机
- ▶ 如何使 $P(p_o \geq \max_i p_i - \epsilon) \geq 1 - \delta$ ？

- ▶ $P\left(\bar{X}_i - \sqrt{\ln(2/\alpha)/2N} \leq p_i \leq \bar{X}_i + \sqrt{\ln(2/\alpha)/2N}\right) \geq 1 - \alpha$
- ▶ 取 $\alpha = \delta/n$, $N = O(\ln(n/\delta)/\epsilon^2)$, 则 $P(|p_i - \bar{X}_i| \leq \epsilon) \geq 1 - \delta/n$
- ▶ Union bound: $P(\forall i, |p_i - \bar{X}_i| \leq \epsilon) \geq 1 - \delta$

- ▶ 对每台游戏机采集样本数量: $O(\ln(n/\delta)/\epsilon^2)$