数理统计

大数定律

- $\overline{X} \to E\overline{X}$ 大数定律
  - 切比雪夫：$\frac{1}{n}\sum X_i = \frac{1}{n}\sum EX_i$ 条件:相互独立
  - 辛钦：$\lim\limits_{n\to\infty} P\{|\frac{1}{n}\sum X_i - \mu| \ge \epsilon\} = 0$ 条件:独立同分布,期望存在
  - 伯努利：$\frac{\mu}{n} \to p$ 伯努利实验
- 中心极限
  - 列林：$\lim\limits_{n\to\infty} P\{\frac{\sum X_i - n\mu}{\sqrt{n}\sigma} \le x\} = \Phi(x)$ 条件:独立同分布,期望方差存在
  - 拉普拉斯：条件二项分布

估计量

- 点估计
  - 矩估计
    - $\overline{X} = EX$ 一阶原点距(样本均值)=总体均值
    - $\frac{1}{n}\sum X_i^2 = E(X^2)$ 二阶原点距=总体平方的均值
  - 最大似然估计
    - 参数等于多少时,已获取的观测值的出现概率最大
    - $L(\theta) = \prod f(x_i;\theta) = \prod p(x_i;\theta)$ 似然函数取最大值时估计值
  - 评价
    - 无偏性：$E\hat{\theta} = \theta$
    - 有效性：$D\hat{\theta}_1 < D\hat{\theta}_2$ 越小越有效
    - 一致性：$\lim\limits_{n\to\infty} P\{|\hat{\theta} - \theta| \ge \epsilon\} = 0$
- 区间估计
  - 两个统计量确定的未知量区间:置信区间
    - $1-\alpha$ 置信度
    - α 显著性水平
  - 正态总体的均值和方差区间估计
    - $I = (\overline{X} - \sigma\overline{X} * z_{\frac{\alpha}{2}}, \overline{X} + \sigma\overline{X} * z_{\frac{\alpha}{2}})$
    - $I = (\overline{X} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1), \overline{X} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1))$
    - $I = (\frac{\sum(\overline{X}-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}, \frac{\sum(\overline{X}-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)})$
    - $I = (\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)})$
- 假设检验
  - $P\{否定H_0|H_0为真\}$ 第一类错误
  - $P\{肯定H_1|H_1为假\}$ 第二类错误

数字特征

- $EX = \sum x_i p_i; EX = \sum g(x_i) p_i$

统计量

- $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - $E\overline{X} = EX = \mu$
  - $D\overline{X} = \frac{1}{n}DX = \frac{\sigma^2}{n}$
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$
  - $E(S^2) = DX = \sigma^2$
- 抽样分布
  - $X \sim \chi^2(n) = \sum_{i=1}^{n} x_i^2$
    - $X_1 \sim \chi^2(n), X_2 \sim \chi^2(m) \implies X_1 + X_2 \sim \chi^2(n+m)$
    - $EX = n, DX = 2n$
  - $X \sim N(0,1), Y \sim \chi^2(n), t \sim t(n) = \frac{X}{\sqrt{Y/n}}$
    - $Et = 0$
  - $X \sim \chi^2(n_1), Y \sim \chi^2(n_2), F \sim F(n_1, n_2) = \frac{X/n_1}{Y/n_2}$
    - $\frac{1}{F} \sim F(n_2, n_1)$
    - $F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}$
    - $t \sim t(n) \to t^2 \sim F(1, n)$
  - $X \sim N(\mu, \sigma^2)$
    - $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n)$
    - $\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \sim N(0,1)$
    - $\frac{\sqrt{n}(\overline{X} - \mu)}{S} \sim t(n-1)$
      - $\frac{n(\overline{X} - \mu)^2}{S^2} \sim F(1, n-1)$
    - $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1)$
- $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
- $B_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^k$