

HW 9

1.

概率的基本概念

- 1° 定义: 性质, 容斥
- 2° 条件概率: 乘法法则
- 3° 独立性: 乘法缩放
- 4° 全概率公式, Bayes 公式

随机变量

- 1° 定义, cdf, pdf, pmf
- 2° 离散分布: Bernoulli, Binomial, Poisson
- 3° 连续分布: Uniform, Normal, Exponential
- 4° 随机变量的函数: $g(x) \rightarrow$

联合分布: 1° 定义

2° Uniform, Normal

3° 边缘分布: $F_X(x) = \lim_{y \rightarrow \infty} F_{XY}(x, y)$, $f_X(x) = \int_{\mathbb{R}} f_{XY}(x, y) dy$

4° 条件分布: 乘法, 全概率, Bayes.

5° 独立性:

6° 随机变量的函数: $\begin{cases} P(g(X) \geq z) := \text{cdf} \\ \text{密度函数变换: } Y = g(X): \\ l(y_1, y_2) = f(h_1(\bar{y}), h_2(\bar{y})) \frac{\partial x}{\partial y} \end{cases}$

数字特征: 1° 期望;

2° 分位数: $P(X \leq a) \geq \alpha$, $P(X \geq a) \geq 1 - \alpha$.

3° 方差;

4° 协方差 $\text{Cov}(X, Y) = E((X - \mu_1)(Y - \mu_2))$

5° 相关系数: $\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_1 \sigma_2} = E\left(\frac{X - \mu_1}{\sigma_1} \cdot \frac{Y - \mu_2}{\sigma_2}\right)$

6° 矩

7° MGF: $N: e^{\frac{\sigma^2}{2}t^2 + \mu t}$, $\text{Exp}: \frac{\lambda}{\lambda - t}$

确定分布, 确定矩

8° 条件期望, 全期望

概率不等式: Markov: $P(Y \geq a) \leq \frac{E(Y)}{a}$
Chebyshev: $P(|Y - E(Y)| \geq a) \leq \frac{\text{Var}(Y)}{a^2}$
Chernoff: $P(Y \geq a) \leq \frac{E(e^{tY})}{e^{ta}}, \forall a, t \in \mathbb{R}_+$

LLN: Khinchin: iid, $\bar{X} \xrightarrow{P} \mu$
Kolmogorov: iid, $\bar{X} \xrightarrow{a.s.} \mu$

CLT: Lindeberg: $\lim_{n \rightarrow \infty} P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = \phi(x)$

2. 课程群问卷调查,

这可能导致样本均来自某个特定群体,是有偏取样.

3. 设总体 $\{x_1, \dots, x_N\}$, $\mu = \frac{1}{N} \sum_{k=1}^N x_k$, $\sigma^2 = \frac{1}{N} \sum_{k=1}^N (x_k - \mu)^2$

1) $\bar{x}_i$ 的分布是

$$P(\bar{x}_i = x_r) = \frac{1}{N} \quad (r \in [1, N])$$

$$E(\bar{x}_i) = \frac{1}{N} \sum_{r=1}^N x_r = \mu$$

$$\text{Var}(\bar{x}_i) = E((\bar{x}_i - \mu)^2) = \frac{1}{N} \sum_{r=1}^N (x_r - \mu)^2 = \sigma^2$$

$$(2) E(\bar{x}) = \frac{1}{n} \sum_{i=1}^n E(\bar{x}_i) = \mu$$

$$\begin{aligned} \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) &= \frac{1}{n^2} \text{Cov}\left(\sum_{i=1}^n x_i, \sum_{i=1}^n x_i\right) \\ &= \frac{1}{n^2} \sum_{i,j} \text{Cov}(x_i, x_j) \end{aligned}$$

当 $i \neq j$ 时:

$$\text{Cov}(x_i, x_j) = E(x_i x_j) - \mu^2$$

$$\text{而 } P(x_i x_j = x_m x_n) = \frac{1}{C_N^2} = \frac{2}{N(N-1)}$$

$$E(x_i x_j) = \frac{2}{N(N-1)} \sum_{m \neq n} x_m x_n$$

$$\text{而 } \left(\sum_k x_k\right)^2 = 2 \sum_{m \neq n} x_m x_n + \sum_{i=1}^n x_i^2 = N^2 \mu^2$$

$$\text{而 } E(X)^2 = \sigma^2 + \mu^2 = \frac{1}{N} \sum_{i=1}^n x_i^2$$

$$\Rightarrow N^2 \mu^2 = 2 \sum_{m \neq n} x_m x_n + N \sigma^2 + N \mu^2$$

$$\Rightarrow \sum_{m \neq n} x_m x_n = \frac{1}{2} (N(N-1) \mu^2 - N \sigma^2)$$

$$\Rightarrow E(x_i x_j) = \frac{1}{N-1} (N-1) \mu^2 - \sigma^2 = \frac{-\sigma^2}{N-1} + \mu^2$$

$$\text{故 } \text{Cov}(x_i, x_j) = \frac{-\sigma^2}{N-1}$$

$$\Rightarrow \text{Var}(\bar{x}) = \frac{1}{n^2} (n \text{Var}(x_i) + n(n-1) \text{Cov}(x_i, x_j))$$

$$= \frac{\sigma^2}{n} + \frac{n-1}{n} \cdot \frac{-\sigma^2}{N-1}$$

$$= \frac{\sigma^2}{n} \left(\frac{N-1}{N-1} + \frac{1-n}{N-1} \right)$$

$$= \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$$

4. ⁽¹⁾ 设 k 阶原点矩为 a_k , 中心矩为 m_k

$$\text{有 } kp = E(X) \approx a_1$$

$$kp(1-p) = \text{Var}(X) = m_2$$

$$\Rightarrow 1-p = \frac{m_2}{a_1} \Rightarrow p = 1 - \frac{m_2}{a_1}$$

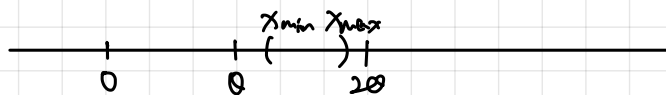
$$k = \frac{a_1}{1 - \frac{m_2}{a_1}} = \frac{a_1^2}{a_1 - m_2}$$

(2) 易受极端值影响

5. 矩估计: $E(X) = \frac{3}{2}\theta \approx a_1$

$$\Rightarrow \theta' = \frac{2}{3}a_1$$

$$\text{MLE: } L(\theta) = \begin{cases} \frac{1}{\theta^n} & (\forall i), X_i \in (\theta, 2\theta) \\ 0 & (\exists i), X_i \in (\theta, 2\theta) \end{cases}$$



$$\text{有 } \theta^* = \frac{1}{2} \max_{1 \leq i \leq n} x_i$$

6. (1) 证明: $\int_{\mathbb{R}} f dx \stackrel{x=a}{=} \frac{1}{\sqrt{2\pi} \sigma^3} \int_{\mathbb{R}} x^2 e^{-\frac{x^2}{2\sigma^2}} dx$

$$\stackrel{\frac{x}{\sigma}}{=} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^2 e^{-\frac{x^2}{2}} dx$$

$$= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x de^{-\frac{x^2}{2}}$$

$$= -\frac{xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx$$

$$= 0 + 1 = 1$$

故 $f(x; a, \sigma)$ 关于 x 是一个概率密度

$$(2) \text{ 令 } a=0, E(X) = \int_{\mathbb{R}} x f dx = 0$$

$$\text{设 } a \neq 0. E(X) = 0 + a = a.$$

故有 $a \approx a_1$.

$$\begin{aligned} \text{Var}(X) &= \int_{\mathbb{R}} (x-a)^2 f dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi} \sigma^3} (x-a)^4 e^{-\frac{(x-a)^2}{2\sigma^2}} dx \\ &\stackrel{\frac{x-a}{\sigma}}{=} \int_{\mathbb{R}} \frac{\sigma^2}{\sqrt{2\pi}} x^4 e^{-\frac{x^2}{2}} dx \\ &= - \int_{\mathbb{R}} \frac{\sigma}{\sqrt{2\pi}} x^3 d e^{-\frac{x^2}{2}} \\ &= - \frac{\sigma x^3}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} + \frac{3\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} x^2 dx \\ &= 3\sigma^2. \end{aligned}$$

$$\text{故 } \sigma \approx \sqrt{\frac{m_2}{3}}$$

$$(3) \text{ 设 } (X_1, \dots, X_n) \text{ 的 pdf 是 } f_s = \prod_{i=1}^n f(X_i)$$

$$\text{似然方程 } \begin{cases} \frac{\partial \ln f_s}{\partial a} = 0 \\ \frac{\partial \ln f_s}{\partial \sigma^2} = 0 \end{cases} \quad \begin{array}{l} \text{因为要求 } \nabla f = 0, \text{ 两式同时满足} \\ \text{故可以将第1式的结果代入第2式.} \end{array}$$

可用牛顿迭代求解

$$7. \bullet \text{矩估计: } E(X) = p \approx a_1$$

• MLE: 设 $(X_1, \dots, X_n)$ 联合分布为 f .

$$f(X_1, \dots, X_n) = p^{\sum_{i=1}^n X_i} \cdot (1-p)^{(n-\sum_{i=1}^n X_i)}$$

$$\text{设 } s = \sum_{i=1}^n X_i.$$

$$L(p) = p^s \cdot (1-p)^{n-s}$$

$$\frac{\partial \ln L}{\partial p} = \frac{\partial}{\partial p} (s \ln p + (n-s) \ln(1-p))$$

$$= \frac{s}{p} + \frac{n-s}{p-1}$$

$$\text{由 } \frac{\partial \ln L}{\partial p} = 0 \text{ 有 } p^* = \frac{s}{n} = a_1.$$

8. 设 $(x_1, \dots, x_m)$ 联合分布为 f .

$$f(x_1, \dots, x_m) = \binom{n}{x_1, \dots, x_m} \prod_{i=1}^m p_i^{x_i}$$

$$\binom{n}{x_1, \dots, x_m} = \binom{n}{x_1} \cdot \binom{n-x_1}{x_2} \cdot \dots \cdot \binom{x_m}{x_m}$$

$$= \frac{n!}{(n-x_1)! x_1!} \cdot \frac{(n-x_1)!}{(n-x_1-x_2)! x_2!} \cdot \dots \cdot 1$$

$$= \frac{n!}{\prod_{i=1}^m x_i!}$$

$$L(p_1, \dots, p_m) = f(x_1, \dots, x_m),$$

这是条件极值 $\max_{\sum_i p_i = 1} f$ 问题,

由 Lagrange 乘子法:

即 $\nabla \ln f \parallel \nabla (\sum p_i)$,

$$\text{有: } \forall i, j \quad \frac{x_i}{p_i} = \frac{x_j}{p_j},$$

$$\text{即 } p_i^* = \frac{x_i}{\sum_{j=1}^m x_j}$$

9. 矩估计:

$$\begin{aligned} E(X) &= \theta^2 + 4\theta(1-\theta) + 3(1-\theta)^2 \\ &= \theta^2 + 4\theta - 4\theta^2 + 3(\theta^2 - 2\theta + 1) \\ &= \cancel{\theta^2} + 4\theta + \cancel{3\theta^2} - 6\theta + 3 \\ &= 3 - 2\theta \end{aligned}$$

故由 $3 - 2\theta \approx a_1$ 有 $\theta \approx \frac{3-a_1}{2} = \frac{5}{6}$

• MLE: 设 (X_1, X_2, X_3) 的 pmf 是 f .

$$\begin{aligned} f(1, 2, 1) &= \theta^2 \cdot (2\theta(1-\theta)) \cdot \theta^2 \\ &= 2\theta^5(1-\theta). \end{aligned}$$

$$\text{故 } L(\theta) = 2\theta^5(1-\theta)$$

$$\text{由 } \frac{\partial L}{\partial \theta} = 0 \text{ 得 } \theta^* = \frac{5}{6}.$$

10. 1D 矩估计: $E(X) = \int_0^1 x(\theta x^{\theta-1}) dx = \int_0^1 \theta x^\theta dx = \frac{\theta}{\theta+1} x^{\theta+1} \Big|_0^1 = \frac{\theta}{\theta+1}$

$$\text{有 } \hat{\theta} = \frac{a_1}{1-a_1}$$

(2) MLE: 设 $(X_1, \dots, X_n)$ 的 pdf 是 f .

$$f = \begin{cases} \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta-1} & (\forall i), x_i \in (0, 1) \\ 0 & (\exists i), x_i \notin (0, 1) \end{cases}$$

易知: 所有 $x_i \in (0, 1)$.

$$\text{故 } L(\theta) = \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta-1}$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial}{\partial \theta} \left(n \ln \theta + (\theta-1) \sum_{i=1}^n \ln x_i \right)$$

$$= \frac{n}{\theta} + \sum_{i=1}^n \ln x_i = 0$$

$$\Rightarrow \theta^* = -\frac{n}{\sum_{i=1}^n \ln x_i}$$

Binomial: $k = 10$, $p = 0.3$, $n = 100$
Moment Estimation Result

	p	k
0	0.307273	10.023669
1	0.459129	7.252862
2	0.274805	11.207940
3	0.407143	7.736842
4	0.337003	8.812968
5	0.387329	7.151552
6	0.175882	15.464883
7	0.323535	9.179831
8	0.376006	8.324327
9	0.280410	11.304871

Binomial: $k = 10$, $p = 0.3$, $n = 1000$
Moment Estimation Result

	p	k
0	0.328583	9.151418
1	0.279052	10.546435
2	0.314588	9.463166
3	0.284507	10.576197
4	0.300519	9.886241
5	0.313390	9.413196
6	0.292918	10.074492
7	0.297237	10.227534
8	0.333898	8.969801
9	0.339885	8.567613

Binomial: $k = 10$, $p = 0.5$, $n = 100$
Moment Estimation Result

	p	k
0	0.558781	9.252276
1	0.550000	8.727273
2	0.594797	8.271733
3	0.488998	10.204541
4	0.507260	10.073724
5	0.484737	9.799131
6	0.599144	8.578906
7	0.470364	10.502496
8	0.462957	11.102538
9	0.447393	11.488781

Binomial: $k = 10$, $p = 0.5$, $n = 1000$
Moment Estimation Result

	p	k
0	0.493603	10.113400
1	0.468273	10.515235
2	0.526994	9.561785
3	0.492903	10.125716
4	0.502448	10.134387
5	0.501274	9.763119
6	0.507556	9.715175
7	0.477692	10.619825
8	0.508217	9.826510
9	0.538697	9.263085

可见,当样本容量较小小时,容易出现明显不合理的估计(上图浅色线)

```

1  from scipy.stats import bernoulli
2  import numpy as np
3
4  def experiment(k, p, n):
5      data = np.array([sum(bernoulli.rvs(p=p, size=k)) for ii in range(0, n)])
6      a_1 = data.mean()
7      m_2 = data.var()
8      return (1-m_2/a_1), a_1/(1-m_2/a_1)
9
10 print("\nBinomial: k = 10, p = 0.5, n = 1000")
11 print("Moment Estimation Result\n      p      k      ")
12 for i in range(0, 10):
13     print("%d %4f %4f" % ((i,) + experiment(10, 0.5, 1000)))
14

```


HW 10

1. $f(\theta, x) = f_{\theta}(\theta) \cdot f_x(x|\theta)$

$$= \frac{1}{\theta} \quad (0 \leq x \leq \theta \leq 1)$$

$$f_x(x) = \int_x^1 f(\theta, x) d\theta = -\frac{1}{\theta^2} \Big|_x^1 = -1 + \frac{1}{x^2}$$

$$f_x(x) = \int_x^1 \frac{d\theta}{\theta} = \ln \theta \Big|_x^1 = -\ln x$$

$$f(\theta|x) = \frac{f(\theta, x)}{f_x(x)} = \frac{\frac{1}{\theta}}{\frac{1-x^2}{x^2}} = \frac{x^2}{\theta(1-x^2)}$$

$$f(\theta|x) = -\frac{1}{\theta \ln x}$$

2. 由课上例子知: Bayes估计将先验分布 $B(a, b)$ 更新为 $B(a+x, b+n-x)$

因此, 后验估计更新为 $B(1+x, 1+n-x)$

$$\text{众数是 } \frac{(1+x)-1}{(1+x) + (1+n-x) - 2} = \frac{x}{n}$$

与 MLE 给出的 $\theta^* = \frac{x}{n}$ 一致

3. 设 μ 的分布 $M \sim N(\mu_0, \sigma_0^2)$, pdf 为 $f(\mu)$ $M \sim N(\mu_0, \sigma_0^2)$ 的正态后验仍为正态.

$$f(\mu | x_1, \dots, x_n) = \frac{f(\mu, x_1, \dots, x_n)}{\int_{\mathbb{R}} f d\mu} \quad \leftarrow \text{积分值与 } \mu \text{ 无关}$$

$$f(\mu, x_1, \dots, x_n) = f(\mu) \cdot f(x_1|\mu) \cdots f(x_n|\mu) \\ = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2}(\frac{\mu-\mu_0}{\sigma_0})^2} \cdot \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2}$$

$$(1) f(\mu | x_1, \dots, x_n)_{\max} \iff f(\mu, x_1, \dots, x_n)_{\max}$$

$$\propto f = \frac{1}{(2\pi)^{\frac{n+1}{2}} \sigma_0 \sigma^n} e^{-\frac{1}{2}(\frac{\mu-\mu_0}{\sigma_0})^2 - \frac{1}{2}\sum_{i=1}^n (\frac{x_i-\mu}{\sigma})^2}$$

略去常数即可

$$\varphi(\mu) = \left(\frac{\mu-\mu_0}{\sigma_0}\right)^2 + \sum_{i=1}^n \left(\frac{x_i-\mu}{\sigma}\right)^2 \text{ 最小值.}$$

$$\varphi'(\mu) = 2 \cdot \frac{\mu-\mu_0}{\sigma_0^2} + 2 \sum_{i=1}^n \frac{\mu-x_i}{\sigma^2} = 0 \iff \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right) \mu = \frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n x_i}{\sigma^2} \\ \Rightarrow \mu = \frac{\sigma_0^{-2}\mu_0 + \sigma^{-2}\sum_{i=1}^n x_i}{\sigma_0^{-2} + n\sigma^{-2}}$$

$$(2) E(\mu | x_1, \dots, x_n) = \int_{\mathbb{R}} \mu f(\mu | x_1, \dots, x_n) d\mu$$

$$\int_{\mathbb{R}} e^{-(Ax^2+Bx+C)} dx = \sqrt{\frac{\pi}{A}} e^{-\frac{\Delta}{4A}}$$

x_i 二次项在这里.

$$\int_{\mathbb{R}} x e^{-(Ax^2+Bx+C)} dx = \frac{-B}{2A} \sqrt{\frac{\pi}{A}} e^{-\frac{\Delta}{4A}}, \quad A = \frac{1}{2} \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right), \quad B = \frac{1}{2} \left(-2 \frac{\mu_0}{\sigma_0^2} - 2 \frac{\sum x_i}{\sigma^2} \right)$$

$$\text{故 } E(\mu | x_1, \dots, x_n) = -\frac{B}{2A} = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$$

$$4. (1) E(\hat{\sigma}^2) = \frac{1}{n} E\left(\sum_{i=1}^n (x_i - \bar{x})^2\right), \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

$$= \frac{1}{n} \sum_{i=1}^n E(x_i^2 - 2x_i \bar{x} + \bar{x}^2)$$

$$= \frac{1}{n} \sum_{i=1}^n E(x_i^2) - 2 E(\bar{x}) + \frac{1}{n} E(\bar{x}^2)$$

$$= \mu^2 + \sigma^2 + \left(\frac{1}{n} - 2\right) \cdot \frac{1}{n^2} (n(\mu^2 + \sigma^2) + \sum_{i,j} E(x_i x_j))$$

$$i \neq j \text{ 时 } E(x_i x_j) = E(x_1 x_2) = \frac{\sum_{i \neq j} x_i x_j}{N(N-1)} = \frac{(\sum_{i=1}^N x_i)^2 - \sum_{i=1}^N x_i^2}{N(N-1)}$$

$$= \frac{N^2 \mu^2 - N \mu^2 - N \sigma^2}{N(N-1)} = \mu^2 - \frac{\sigma^2}{N-1}$$

$$\Rightarrow E(\hat{\sigma}^2) = \sigma^2 \cdot \frac{n-1}{n} \cdot \frac{N}{N-1}$$

$$E(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^n E(x_i - \mu)^2 + \frac{E(\bar{x} - \mu)^2}{n}$$

$\text{Var}(\bar{x})$

$-\frac{2}{n} \sum_{i=1}^n E((x_i - \mu)(\bar{x} - \mu))$

$$(2) \text{Var}(\bar{x}) = \frac{1}{n^2} E\left(\sum_{i=1}^n x_i - n\mu\right)^2 = \frac{1}{n^2} E\left(\sum_{i=1}^n x_i^2 - 2n\mu \sum_{i=1}^n x_i + n^2 \mu^2\right)$$

$$= -\mu + \frac{1}{n^2} (n(\mu^2 + \sigma^2) + n(n-1) E(x_1, x_2)) = -\mu + \frac{1}{n^2} (n\mu^2 + n\sigma^2 + n^2\mu - n\mu - n(n-1) \frac{\sigma^2}{N})$$

$$= \frac{1}{n} \sigma^2 - \frac{n-1}{n} \frac{\sigma^2}{N-1} = \frac{1}{N-1} \left(\frac{N-1}{n} + \frac{1-n}{N} \right) \sigma^2 = \frac{N-n}{(N-1)n} \sigma^2$$

故 $\frac{N-n}{N(n-1)} E(\hat{\sigma}^2)$ 是 $\text{Var}(\bar{x})$ 一个无偏估计

5. (1) 先证无偏性. $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, k \in \mathbb{N}.$

$$E(\hat{\theta}(X)) = \sum_{k=0}^{+\infty} \frac{\lambda^k}{(2k)!} e^{-\lambda} - \sum_{k=0}^{+\infty} \frac{\lambda^{2k+1}}{(2k+1)!} e^{-\lambda}$$

$$= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(-\lambda)^k}{k!}$$

$$= e^{-\lambda} \cdot e^{-\lambda} = e^{-2\lambda}.$$

$\Rightarrow \hat{\theta}(X)$ 是 $g(\lambda)$ 的无偏估计, 唯一性, 不会.....

(2) 不合理, 估计可能为负值.

可给出均方误差最小的估计. 即 $\min_{\hat{\theta}} E((\hat{\theta}(X) - e^{-2\lambda})^2).$

... 不太会了...

用MLE估计 $f(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda} := L(\lambda).$

$$\ln L = -\ln x! + x \ln \lambda - \lambda$$

$$\frac{\partial \ln L}{\partial \lambda} = \frac{x}{\lambda} - 1 = 0$$

6. (1)

► 次序统计量的分布密度

以下设随机变量 $X_1, X_2, \dots, X_n$ 独立同分布, 有公共的分布函数 $F(x)$ 和密度函数 $f(x)$.

► 性质 1

$(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ 有联合密度

$$g(x_1, \dots, x_n) = \begin{cases} n! \prod_{i=1}^n f(x_i), & \text{if } x_1 < \dots < x_n \\ 0, & \text{others} \end{cases}$$

► 性质 2

$X_{(k)}$ 有密度

$$g_k(x_k) = n \binom{n-1}{k-1} [F(x_k)]^{k-1} [1 - F(x_k)]^{n-k} f(x_k)$$

最大, 最小的 pdf
求法: 先求 cdf, 再求 pdf.

不妨直接套公式: $f(x) = \frac{1}{\theta}, x \in (0, \theta); F(x) = \frac{1}{\theta}x, x \in (0, \theta)$

$$g_1(x) = n \left(1 - \frac{1}{\theta}x\right)^{n-1} \frac{1}{\theta}$$

$$g_n(x) = n \left(\frac{1}{\theta}x\right)^{n-1} \frac{1}{\theta}$$

$$\begin{aligned} E(X_{(1)}) &= \int_0^\theta x n \left(1 - \frac{x}{\theta}\right)^{n-1} \frac{1}{\theta} dx = \int_0^\theta -x n \frac{(\theta-x)^{n-1}}{\theta^n} d(\theta-x) \\ &= \int_0^\theta -\frac{x}{\theta^n} d(\theta-x)^n \\ &= -\frac{x}{\theta^n} (\theta-x)^n \Big|_0^\theta + \int_0^\theta \frac{(\theta-x)^n}{\theta^n} dx \\ &= -\frac{1}{\theta^n} \int_0^\theta (\theta-x)^n d(\theta-x) \\ &= -\frac{1}{\theta^n} \frac{1}{n+1} (\theta-x)^{n+1} \Big|_0^\theta \\ &= \frac{1}{n+1} \theta \end{aligned}$$

$$\begin{aligned} E(X_{(n)}) &= \int_0^\theta n \cdot \frac{x^{n-1}}{\theta^n} \cdot x dx \\ &= \frac{n}{n+1} \cdot \frac{x^{n+1}}{\theta^n} \Big|_0^\theta \\ &= \frac{n}{n+1} \theta \end{aligned}$$

$$\text{故 } \hat{\theta}_1 = E(X_{(1)}) + E(X_{(n)}) = \theta.$$

是无偏的.

$$(2) \text{ 令 } c_n = n+1, \text{ 则 } E(\theta_2) = (n+1) \cdot \frac{1}{n+1} \theta = \theta.$$

$$(2) E(\hat{\theta}_1^2) = E(\bar{X}_n^2 + \bar{X}_n^2 + 2\bar{X}_n \bar{X}_n) \quad \text{遇事不决, cdf 解决}$$

设 $(\bar{X}_n, \bar{X}_n)$ cdf 为

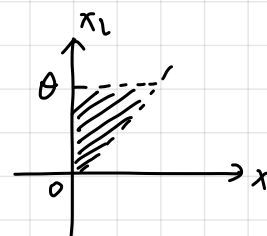
$$F(x_1, x_2) = P(\bar{X}_n \leq x_1, \bar{X}_n \leq x_2) = P(B) \cdot P(A|B)$$

$$= P(\bar{X}_n \leq x_2) - P(\bar{X}_n \geq x_1, \bar{X}_n \leq x_2)$$

$$= \left(\frac{x_2}{\theta}\right)^n - \left(\frac{x_2 - x_1}{\theta}\right)^n = P(B) (1 - P(A^c|B)) = P(B) - P(BA^c)$$

$$f(x_1, x_2) = \frac{\partial^2 F}{\partial x_1 \partial x_2} = \frac{-1}{\theta^n} \frac{\partial}{\partial x_1} (x_2 - x_1)^{n-1} = \frac{n(n-1)}{\theta^n} (x_2 - x_1)^{n-2}$$

$$E(\bar{X}_n \bar{X}_n) = \int_0^\theta \int_0^{x_2} x_1 x_2 \frac{n(n-1)}{\theta^n} (x_2 - x_1)^{n-2} dx_1 dx_2$$



$$= \int_0^\theta dx_2 \int_0^{x_2} \frac{n(n-1)}{\theta^n} x_1 (x_2 - x_1)^{n-2} dx_1$$

$$= \int_0^\theta dx_2 \int_0^{x_2} \frac{-n}{\theta^n} d(x_2 - x_1)^{n-1}$$

$$= \int_0^\theta \left(-\frac{n}{\theta^n} (x_2 - x_1)^{n-1} \Big|_0^{x_2} + \int_0^{x_2} \frac{(x_2 - x_1)^{n-1}}{\theta^n} n dx_1 \right) dx_2$$

$$= \int_0^\theta -\frac{(x_2 - x_1)^n}{\theta^n} x_2 \Big|_0^{x_2} dx_2$$

$$= \int_0^\theta \frac{x_2^{n+1}}{\theta^n} dx_2 = \frac{\theta^2}{n+2}$$

$$E(\bar{X}_n^2) = \int_0^\theta x^2 n \left(1 - \frac{x}{\theta}\right)^n \frac{1}{\theta} dx = \frac{1}{n+1} \cdot \frac{1}{n+2} \cdot 2\theta^2$$

$$E(\bar{X}_n^2) = \int_0^\theta x^2 n \cdot \frac{x^{n-1}}{\theta^n} dx = \frac{n}{n+2} \theta^2$$

$$\text{Var}(\hat{\theta}_1) = \frac{2\theta^2}{(n+1)(n+2)}$$

$$\text{Var} \hat{\theta}_2 = c_n^2 \text{Var}(\bar{X}_n) = (n+1)^2 \left(\frac{2\theta^2}{(n+1)(n+2)} - \frac{\theta^2}{(n+1)^2} \right)$$

$$= (n+1)^2 \frac{(2n+2)\theta^2}{(n+1)^2(n+2)} + \frac{(-n-2)\theta^2}{(n+1)^2(n+2)}$$

$$= \frac{n}{n+2} \theta^2$$

$$\text{Var}(\hat{\theta}_3) = 4\text{Var}(\bar{X}) = 4 \cdot \frac{\sigma^2}{n} = 4 \times \frac{1}{12} \times \frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

$$\begin{aligned} \text{Var}(\hat{\theta}_4) &= \left(\frac{n+1}{n}\right)^2 \left(E(X_{(n)}^2) - E^2(X_{(n)}) \right) \\ &= \frac{(n+1)^2}{n^2} \left(\frac{n}{n+2} \theta^2 - \frac{n^2}{(n+1)^2} \theta^2 \right) \\ &= \left(\frac{(n+1)^2}{n(n+2)} - 1 \right) \theta^2 = \frac{n^2 + 2n + 1 - n^2 - 2n}{n(n+2)} \theta^2 = \frac{\theta^2}{n(n+2)}. \end{aligned}$$

$$\text{故 } \text{Var}(\hat{\theta}_2) > \text{Var}(\hat{\theta}_3) > \text{Var}(\hat{\theta}_1) > \text{Var}(\hat{\theta}_4).$$

$$7. (1) E\left(\sum_{i=1}^n c_i \bar{x}_i\right) = \sum_{i=1}^n c_i E(\bar{x}_i) = \theta \sum_{i=1}^n c_i$$

$$\text{故 } E\left(\sum_{i=1}^n c_i \bar{x}\right) = \theta \iff \sum_{i=1}^n c_i = 1$$

$$\begin{aligned} (2) \text{Var}\left(\sum_{i=1}^n c_i \bar{x}_i\right) &= \sum_{i,j} c_i c_j \text{Cov}(\bar{x}_i, \bar{x}_j) \\ &= \sum_{i=1}^n c_i^2 \text{Var}(\bar{x}_i) \\ &= \sigma^2 \sum_{i=1}^n c_i^2 \end{aligned}$$

$$\text{由均值不等式 } \left(\frac{1}{n} \sum_{i=1}^n c_i^2\right)^{\frac{1}{2}} \geq \frac{1}{n} \sum_{i=1}^n c_i$$

当且仅当 $\forall i=j, c_i=c_j$ 时等号成立.

$$\text{Var}\left(\sum_{i=1}^n c_i \bar{x}_i\right) = \sigma^2 \sum_{i=1}^n c_i^2 \geq \frac{\sigma^2}{n}$$

当且仅当 $\forall i, c_i = \frac{1}{n}$ 时, 等号成立.

8. 经大量计算,

$$D(S^2) = \frac{2\sigma^4}{n-1}$$

$$\text{而 } m_2 = \frac{n-1}{n} S^2.$$

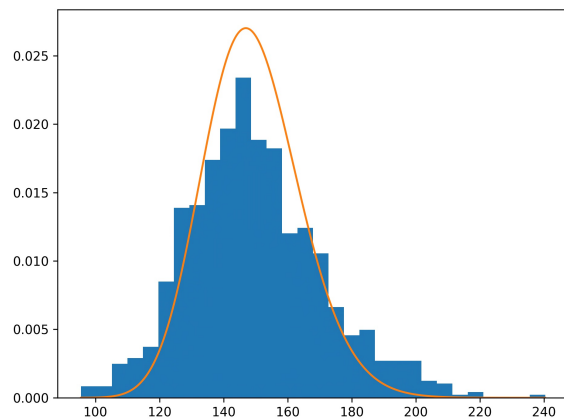
$$\text{故 } D(m_2) = \frac{n-1}{n^2} \cdot 2\sigma^4.$$

设 $n > 1$, 显然 $D(m_2) < D(S^2)$.

$$\frac{(n-1)}{\sigma^2} S^2 \sim \chi^2(n-1)$$

$$\boxed{\begin{aligned} \text{设 } Y &\sim \chi^2(n) \\ E(Y) &= n, \text{Var}(Y) = 2n \end{aligned}}$$

9. (1)-(5)



$$\log \hat{\theta} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(5, \frac{1}{100}\right)$$

$\hat{\theta}$ 的 pd 是图中橙色线，
 $\theta^{(ii)}$ 的分布是图中蓝色柱。

(6) 经计算， V_{boot} 和 $Var(\hat{\theta})$ 十分接近，
 故可以用 V_{boot} 近似 $Var(\hat{\theta})$

```

1  from random import normalvariate as norm, choice
2  from math import exp, sqrt, pi, log, e
3  from matplotlib import pyplot as plt
4  import numpy as np
5
6
7  def data_set(n: int):
8      data = []
9      observation = []
10     for i in range(0, n):
11         data += [norm(5, 1)]
12     for i in range(0, n):
13         observation += [choice(data)]
14     return exp(sum(observation) / n)
15
16
17  theta = []
18  log_norm = lambda x, mu, sigma: 1 / (x * sqrt(2 * pi) * sigma) * exp(-1 / (2 * (sigma ** 2)) * (log(x) - mu) ** 2)
19  for i in range(0, 1000):
20     theta += [data_set(100)]
21
22  dots = np.linspace(min(theta), max(theta), 1000)
23  plt.figure(dpi=300)
24  plt.hist(theta, bins=30, density=True)
25  plt.plot(dots, [log_norm(x, 5, 1/10) for x in dots])
26  plt.show()
27
28  print(np.array(theta).var(), (exp(1/100)-1)*exp(2*5+1/100))
29
    
```


HW 11

1. 设总体均值为 μ , 方差 σ^2 .

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

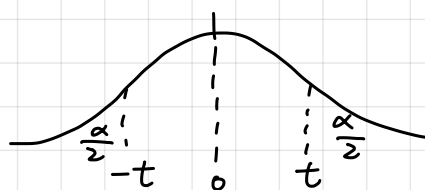
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}} \sqrt{\frac{(n-1)S^2}{n-1}}} \sim t(n-1)$$

$$\bar{X} = 503.75, S = 6.202$$

$$\frac{1}{2}\alpha = 0.05$$

$$t_{\frac{\alpha}{2}}(15) = 2.131 = t$$



$$-t \leq \frac{\mu - \bar{X}}{\frac{S}{\sqrt{n}}} \leq t$$

$$\bar{X} - \frac{S}{\sqrt{n}}t \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}}t$$

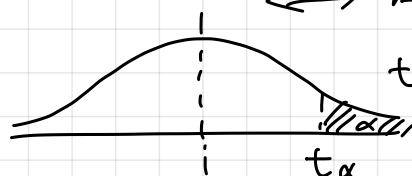
故 95% 置信区间是 (500.449, 507.054)

2. $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

$$P(\mu > \hat{\mu}) = P\left(\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < \frac{\bar{X} - \hat{\mu}}{\frac{S}{\sqrt{n}}}\right)$$



$\Leftrightarrow$ 让 $\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ 这个 t 分布 $< t_\alpha$ 的概率 $\geq 95\%$

$$\frac{1}{2}\alpha = 0.05$$

$$\text{有 } \frac{\bar{X} - \hat{\mu}}{\frac{S}{\sqrt{n}}} = t_\alpha \Rightarrow \hat{\mu} = \bar{X} - \frac{S}{\sqrt{n}}t_\alpha = 1064.893$$

// 注意 S^2 开根号.

3. ⁽¹⁾ 设 $X_i \sim N(\mu_1, \sigma^2)$, $n=20$ $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, (\frac{1}{n} + \frac{1}{m})\sigma^2)$
 $Y_j \sim N(\mu_2, \sigma^2)$, $m=30$.

$$S^2 = \frac{n-1}{n+m-2} S_1^2 + \frac{m-1}{n+m-2} S_2^2 = 3.9685.$$

$$\frac{(n+m-2)S^2}{\sigma^2} \sim \chi^2(n+m-2)$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \cdot \sqrt{\frac{(n+m-2)S^2}{\sigma^2}} \sim t(n+m-2)$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2), \quad \wedge \alpha = 0.05$$

$$-t_{\frac{\alpha}{2}} \leq \frac{\Delta\mu - \Delta\bar{X}}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \leq t_{\frac{\alpha}{2}}$$

$$\Rightarrow \bar{X} - \bar{Y} - S t_{\frac{\alpha}{2}} \sqrt{\frac{1}{n} + \frac{1}{m}} \leq \Delta\mu \leq \bar{X} - \bar{Y} + S t_{\frac{\alpha}{2}} \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$t_{\frac{\alpha}{2}} = 2.0106$$

故 95% 置信区间估计是 $(-4.323, 0.283)$
 $-3.176, -0.8637$

(2) 用 S_1, S_2 分别近似 σ_1, σ_2

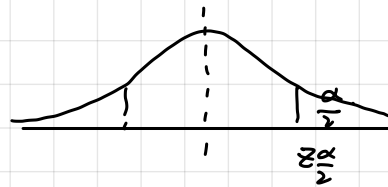
$$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n}), \quad \bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{m}).$$

$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}).$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$$

$$\wedge \alpha = 0.05.$$

$$-Z_{\frac{\alpha}{2}} \leq \frac{(\mu_1 - \mu_2) - (\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \leq Z_{\frac{\alpha}{2}}$$



$$-Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} + (\bar{X} - \bar{Y}) \leq \mu_1 - \mu_2 \leq Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} + (\bar{X} - \bar{Y}).$$

$$\begin{cases} Z_{\frac{\alpha}{2}} = 1.96 \\ \sigma_1^2 = s_1^2 \\ \sigma_2^2 = s_2^2 \end{cases}$$

得: $(-3.1434, -0.8966)$.

不算很显著.

4. 设 $M = \max\{Z_1, \dots, Z_n\}$.

令 $P(M \leq \theta \leq c_n M)$

$$= P\left(\frac{\theta}{c_n} \leq M \leq \theta\right) = 1 - \alpha$$

对 $\forall$ 给定 θ ,

$$P\left(\frac{\theta}{c_n} \leq M \leq \theta\right) = P\left(M \geq \frac{\theta}{c_n}\right)$$

$$= 1 - P\left(M < \frac{\theta}{c_n}\right)$$

$$= 1 - \left(\int_0^{\frac{\theta}{c_n}} \frac{1}{\theta} dx\right)^n$$

$$\Rightarrow \int_0^{\frac{\theta}{c_n}} \frac{1}{\theta} dx = \alpha^{\frac{1}{n}}$$

$$\Rightarrow c_n = \alpha^{-\frac{1}{n}}$$

所以 $\exists c_n$, s.t. $(M, c_n M)$ 是 θ 的一个 $(1-\alpha)$ 置信区间.

将求 θ 落在 M

转化成 M 落在 θ .

这 $\Leftrightarrow$ 求 M 的分布

⑤. 思路: 由 CLT, 认为 $\bar{X} \sim N(\lambda, \frac{\lambda}{n})$

$$|\frac{\lambda - \bar{X}}{\sqrt{\frac{\lambda}{n}}}| \leq Z_{\frac{\alpha}{2}}$$

$$\lambda^2 - 2\lambda\bar{X} + \bar{X}^2 \leq Z_{\frac{\alpha}{2}}^2 \cdot \frac{\lambda}{n}$$

$$\lambda^2 - (2\bar{X} + \frac{\lambda}{n} Z_{\frac{\alpha}{2}}^2) \lambda + \bar{X}^2 \leq 0$$

$$\Delta = (2\bar{X} + \frac{\lambda}{n} Z_{\frac{\alpha}{2}}^2)^2 - 4\bar{X}^2$$

则相应的置信区间是 $(\bar{X} + \frac{\lambda}{2n} Z_{\frac{\alpha}{2}}^2 - \frac{\sqrt{\Delta}}{2}, \bar{X} + \frac{\lambda}{2n} Z_{\frac{\alpha}{2}}^2 + \frac{\sqrt{\Delta}}{2})$

6. 设 $X_i \sim B(p)$ 由 CLT: $\bar{X} \sim N(p, \frac{p(1-p)}{n})$

$$-Z_{\frac{\alpha}{2}} \leq \frac{p - \bar{X}}{\sqrt{\frac{p(1-p)}{n}}} \leq Z_{\frac{\alpha}{2}} \quad \text{令 } \mu = \bar{X}$$

$$\text{即 } \frac{p^2 - 2\mu p + \mu^2}{\frac{p - p^2}{n}} \leq Z_{\frac{\alpha}{2}}^2$$

$$p^2 - 2\mu p + \mu^2 \leq -\frac{Z_{\frac{\alpha}{2}}^2}{n} p^2 + \frac{Z_{\frac{\alpha}{2}}^2}{n} p$$

$$(1 + \frac{Z_{\frac{\alpha}{2}}^2}{n}) p^2 + (-2\mu - \frac{Z_{\frac{\alpha}{2}}^2}{n}) p + \mu^2 \leq 0$$

则 p 的一个 $(1-\alpha)$ 置信区间是.

$$(\mu + \frac{Z_{\frac{\alpha}{2}}}{n} - \sqrt{\Delta}, \mu + \frac{Z_{\frac{\alpha}{2}}}{n} + \sqrt{\Delta})$$

带入数据, 得置信区间是 $(0.309, 0.498)$

$$7. (1) f(x_1, \dots, x_n; \sigma) = \prod_{i=1}^n \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} \right).$$

$$L(\sigma) = f(x_1, \dots, x_n; \sigma)$$

$$\frac{\partial \ln L(\sigma)}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2} \left(\sum_{i=1}^n \left(-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right) \left(-\frac{1}{2} \ln \sigma^2 - \frac{1}{2} \ln 2\pi \right) \right)$$

$$= \frac{\partial}{\partial \sigma^2} \left(-\frac{n}{2} \ln \sigma^2 + -\frac{1}{2} \cdot \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right)$$

$$= -\frac{1}{2} \left(\frac{n}{\sigma^2} + -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \right).$$

$$\text{令 } \frac{\partial \ln L}{\partial \sigma^2} = 0, \text{ 得 } (\sigma^2)^* = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.$$

经验证, $(\sigma^2)^*$ 是 $L(\sigma)$ 的极大值点.

$$(2) I(\sigma) = -E\left(\frac{\partial^2 \ln L}{\partial \sigma^2}\right); L = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

$$\ln L = -\frac{1}{2} \cdot \frac{1}{\sigma^2} (x - \mu)^2 - \ln \sigma - \frac{1}{2} \ln 2\pi.$$

$$\frac{\partial \ln L}{\partial \sigma} = (x - \mu)^2 \sigma^{-3} - \sigma^{-1}$$

$$\frac{\partial^2 \ln L}{\partial \sigma^2} = -3(x - \mu)^2 \sigma^{-4} + \sigma^{-2}$$

$$I(\sigma) = E(3(x - \mu)^2 \sigma^{-4}) - \sigma^{-2}$$

$$= \frac{3\sigma^2}{\sigma^4} - \frac{1}{\sigma^2} = \frac{2}{\sigma^2}$$

$$I_n(\sigma) = \frac{2n}{\sigma^2}$$

$$\hat{se} = \sqrt{\frac{1}{I_n(\sigma)}} = \frac{\sqrt{2n}}{\sigma}$$

(3) 可以利用 $N(\sigma^*, \frac{2n}{\sigma^{*2}})$ 近似得出 σ^* 的置信区间. 再平方, 取对数即可. (不确定)

$$8. (1) f_M(x) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu_0}{\sigma_0} \right)^2}$$

$$f_{Z_i}(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}$$

$$f_M(\mu | Z_1, \dots, Z_n) = \frac{f_M(\mu) \cdot f(Z_1, \dots, Z_n | \mu)}{\int_{\mathbb{R}} f_M(\mu) f(Z_1, \dots, Z_n | \mu) d\mu}$$

$$= \frac{f_M(\mu) f(Z_1, \dots, Z_n | \mu)}{\sigma_0 \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\mu - \mu_0}{\sigma_0} \right)^2} \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2}$$

$$= \frac{1}{\sigma_0 \sigma^n (2\pi)^{\frac{n+1}{2}}} \exp \left\{ -\frac{1}{2} \left[\left(\frac{\mu - \mu_0}{\sigma_0} \right)^2 + \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \right] \right\}$$

$$= \frac{1}{\sigma_0 \sigma^n (2\pi)^{\frac{n+1}{2}}} \exp \left\{ -\frac{1}{2} \left[\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right) \mu^2 - 2 \frac{\mu_0 \mu}{\sigma_0^2} + \frac{\mu_0^2}{\sigma_0^2} + \sum_{i=1}^n \frac{x_i^2}{\sigma^2} + \sum_{i=1}^n -\frac{2x_i \mu}{\sigma^2} \right] \right\}$$

$$= \frac{1}{\sigma_0 \sigma^n (2\pi)^{\frac{n+1}{2}}} \exp \left\{ -\frac{1}{2} \left[\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right) \mu^2 - 2 \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2} \right) \mu + \frac{\mu_0^2}{\sigma_0^2} + \sum_{i=1}^n \frac{x_i^2}{\sigma^2} \right] \right\}$$

$$\int_{\mathbb{R}} f(\mu, x_1, \dots, x_n) d\mu = \sqrt{\frac{2\pi}{\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)}} e^{\frac{B^2 - 4AC}{2\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)}}$$

$$A = \frac{1}{2} \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right), B = \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2} \right)$$

$$C = \frac{\mu_0^2}{\sigma_0^2} + \sum_{i=1}^n \frac{x_i^2}{\sigma^2}$$

$$f_M(\mu | Z_1, \dots, Z_n) = \frac{1}{\sigma_0 \sigma^n (2\pi)^{\frac{n}{2}} \sqrt{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}} e^{-\left(A\mu^2 + B\mu + C \right) - \frac{B^2 - 4AC}{4A}}$$

$$= (\text{---}) e^{-A\mu^2 - B\mu - \frac{B^2}{4A}}$$

故 μ 的后验分布是 $N\left(\frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}\right)$

$$\text{令 } \mu' = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \sigma'^2 = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$$

$$\Rightarrow \frac{\mu - \mu'}{\sigma'} \sim N(0, 1)$$

设 $N(0, 1)$ 的上侧 β 分位数是 z_β .

$$-z_{\frac{\alpha}{2}} \leq \frac{\mu - \mu'}{\sigma'} \leq z_{\frac{\alpha}{2}}$$

$$\text{故 } a = \mu' - \sigma' z_{\frac{\alpha}{2}}, b = \mu' + \sigma' z_{\frac{\alpha}{2}}.$$

这时区间长度最小.

$$(2) \text{ 令 } \sigma_0 \rightarrow +\infty, \mu' \rightarrow \bar{x}, \sigma'^2 \rightarrow \frac{\sigma^2}{n}$$

$$\text{这时所求置信区间为 } \left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}\right)$$

与经典方法一致.

这是因为 $\sigma_0 \rightarrow +\infty$ 时, 相当于没有任何先验信息.

9. $\log \hat{\theta} \sim N(\mu, \frac{\sigma^2}{n_0})$, $n_0 = 100$.

$$E(\hat{\theta}) = e^{\mu} \cdot e^{\frac{\sigma^2}{2n_0}} = \theta e^{\frac{\sigma^2}{2n_0}}$$

$$\text{Var}(\hat{\theta}) = (e^{\sigma^2} - 1) e^{2\mu + \sigma^2} = \theta^2 (e^{\frac{\sigma^2}{n_0}} - 1) e^{\frac{\sigma^2}{n_0}}, \quad \text{令 } \alpha = e^{\frac{\sigma^2}{2n_0}} = e^{\frac{1}{200}}$$

$$\bar{\theta} \stackrel{\text{Jrqr}}{\underset{\text{CLT}}{\sim}} N(\alpha\theta, \frac{\theta^2(\alpha-1)\alpha}{n}) \quad n=1000$$

$$-Z_{\frac{\alpha}{2}} \leq \frac{\alpha\theta - \bar{\theta}}{\theta \sqrt{\frac{(\alpha-1)\alpha}{n}}} \leq Z_{\frac{\alpha}{2}}$$

$$\frac{\bar{\theta}}{\alpha + Z_{\frac{\alpha}{2}} \sqrt{\frac{\alpha(\alpha-1)}{n}}} \leq \theta \leq \frac{\bar{\theta}}{\alpha - Z_{\frac{\alpha}{2}} \sqrt{\frac{\alpha(\alpha-1)}{n}}}$$

(2) 另一思路, 先估计 $\bar{x}$.

$$\bar{x} \sim N(\mu, \frac{1}{100}) \quad \bar{\bar{x}} \sim N(\mu, 10^5)$$

(如何利用1000次这一信息?)

$$-Z_{\frac{\alpha}{2}} \leq \frac{\bar{\bar{x}} - \mu}{\frac{1}{10}} \leq Z_{\frac{\alpha}{2}}$$

$$\bar{\bar{x}} - 10^5 Z_{\frac{\alpha}{2}} \leq \mu \leq \bar{\bar{x}} + 10^5 Z_{\frac{\alpha}{2}}$$

再取指数.

```
Python 3.8.11 (default, Aug 6 2021, 09:57:55) [MSC v.1916 64 bit (AMD64)] on win32
>>> runfile('C:\\Users\\joshu\\PycharmProjects\\scientificProject\\hw10_11.py', wdir
(1): [149.4269175093138, 150.74081789569183]
(2): [146.44812310117808, 148.2748113831372]
```

本题代码

```
1 from random import normalvariate as norm, choice
2 from math import exp, sqrt, log
3
4 def data_set(n: int):
5     data = []
6     observation = []
7     for i in range(0, n):
8         data += [norm(5, 1)]
9     for i in range(0, n):
10        observation += [choice(data)]
11    return exp(sum(observation) / n)
12
13
14 theta = []
15 for i in range(0, 1000):
16     theta += [data_set(100)]
17 mean = sum(theta)/1000
18 alpha = exp(1/200)
19 print("(1):", [mean/(alpha+1.96*sqrt(alpha*(alpha-1)/1000)), mean/(alpha-1.96*sqrt(alpha*(alpha-1)/1000))])
20 x = sum([log(data_set(100)) for i in range(0, 1000)]) / 1000
21 print("(2):", [exp(x - 1.96 * (10**(-2.5))), exp(x + 1.96 * (10**(-2.5)))]])
```

HW 12

1. 设 $X \sim \text{Exp}(\lambda)$. $E(X) = \frac{1}{\lambda}$, $\text{Var}(X) = \frac{1}{\lambda^2}$

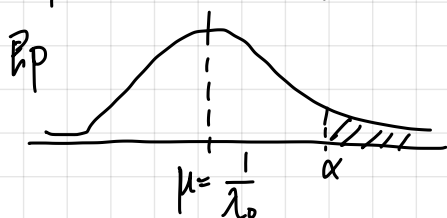
采用大样本方法, 由 CLT:

$$\bar{X} \xrightarrow{d} N\left(\frac{1}{\lambda}, \frac{1}{n\lambda^2}\right)$$

① 单边. $H_0: \lambda > \lambda_0$. $H_1: \lambda \leq \lambda_0$.

设 $\bar{X} \geq c$ 时拒绝 H_0 .

则当 H_0 成立时: $P(\bar{X} \geq c) \leq \alpha$.



$$P\left(\frac{\bar{X} - \frac{1}{\lambda_0}}{\frac{1}{\lambda_0\sqrt{n}}} \geq \frac{c - \frac{1}{\lambda_0}}{\frac{1}{\lambda_0\sqrt{n}}}\right) \leq \alpha.$$

or max??

$$\text{令 } \sqrt{n}\lambda_0(c - \frac{1}{\lambda_0}) = z_\alpha$$

$$\text{有: } \sqrt{n}\lambda_0 c - \sqrt{n} = z_\alpha$$

$$c = \frac{z_\alpha}{\lambda_0\sqrt{n}} + \frac{1}{\lambda_0}.$$

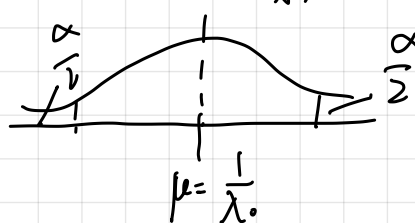
即 $\bar{X} \geq \frac{1}{\lambda_0} + \frac{z_\alpha}{\lambda_0\sqrt{n}}$ 时拒绝 H_0 .

② 双边 $H_0: \lambda = \lambda_0$. $H_1: \lambda \neq \lambda_0$.

设 $|\bar{X} - \frac{1}{\lambda_0}| \geq c$ 时, 拒绝 H_0 .

当 H_0 成立时, 要求 $P(|\bar{X} - \frac{1}{\lambda_0}| \geq c) \leq \alpha$.

$$\text{即 } P\left(\frac{|\bar{X} - \frac{1}{\lambda_0}|}{\frac{1}{\lambda_0\sqrt{n}}} \geq \lambda_0 c \sqrt{n}\right) \leq \alpha$$



$$\text{令 } \lambda_0 c \sqrt{n} = z_{\frac{\alpha}{2}} \text{ 即 } c = \frac{z_{\frac{\alpha}{2}}}{\lambda_0\sqrt{n}}$$

得到检验: $|\bar{X} - \frac{1}{\lambda_0}| \geq \frac{z_{\frac{\alpha}{2}}}{\lambda_0\sqrt{n}}$ 时拒绝 H_0 .

2. 设 $X \sim U(0, \theta)$. $E(X) = \frac{\theta}{2}$, $\text{Var}(X) = \frac{\theta^2}{12}$

(1) 用大样本方法. 认为 $\bar{X} \sim N(\frac{\theta}{2}, \frac{\theta^2}{12n})$

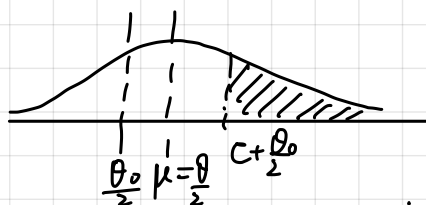
在 $\bar{X} - \frac{\theta_0}{2} \geq c$ 时拒绝 H_0 .

当 H_0 成立时, 要求 $P(\bar{X} - \frac{\theta_0}{2} \geq c) \leq \alpha$.

$$P\left(P\left(\frac{\bar{X} - \frac{\theta_0}{2}}{\frac{\theta}{\sqrt{12n}}} \geq \frac{c\sqrt{12n}}{\theta}\right) \leq \alpha\right.$$

$$\text{令 } \frac{c\sqrt{12n}}{\theta} = z_\alpha$$

$$\text{即 } c = \frac{z_\alpha \theta}{\sqrt{12n}} = \frac{\theta z_\alpha \sqrt{3}}{6n}, \quad \bar{X} \geq \frac{\theta_0}{2} + \frac{\sqrt{3} \theta z_\alpha}{6n}$$



$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\alpha(R) = \alpha.$$

$$1 - \beta(R) = \int_{c + \frac{\theta_0}{2}}^{+\infty} f(x; \frac{\theta}{2}, \frac{\theta^2}{12n}) dx.$$

$$\text{功效函数 } g(\theta) = \int_{c + \frac{\theta_0}{2}}^{+\infty} f(x; \frac{\theta}{2}, \frac{\theta^2}{12n}) dx$$

$$(\theta \geq \theta_0)$$

(2) 在 MLE 中: $\theta^* = \max\{x_1, \dots, x_n\}$

设 $Y = \max\{x_1, \dots, x_n\}$.

$$P(Y \leq x) = P(\forall i \in \{1, \dots, n\}, x_i \leq x), \quad x \in (0, \theta).$$

$$= F^n(x)$$

$$f(x) = \frac{1}{\theta}, \quad x \in (0, \theta).$$

$$F(x) = \int_0^x f(t) dt = \frac{x}{\theta}$$

$$P(Y \leq x) = \frac{x^n}{\theta^n}$$

认为 $Y > c$ 时拒绝 H_0 .

当 H_0 成立时: $P(Y > c) \leq \alpha$

$$1 - \frac{c^n}{\theta_0^n} \leq \alpha$$

$$\frac{c^n}{\theta_0^n} \geq 1 - \alpha$$

$$c \geq (1 - \alpha)^{\frac{1}{n}} \theta_0$$

$$\text{取 } c = \theta_0 (1 - \alpha)^{\frac{1}{n}}$$

$$\alpha(R) = \alpha.$$

$$1 - \beta(R) = P_\theta(Y > c) = 1 - P_\theta(Y \leq c) = 1 - \frac{\theta_0^n}{\theta^n} (1 - \alpha)$$

$$\text{故功效函数是 } g(\theta) = 1 - \frac{\theta_0^n}{\theta^n} (1 - \alpha)$$

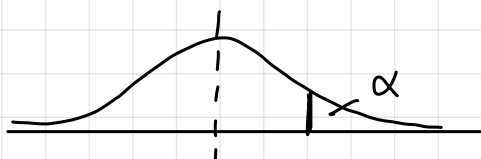
3. 设该公司病假天数为 $\bar{X} \sim N(\mu, \sigma^2)$, $\mu_0 = 5$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t(n-1) \quad ; \quad n-1 = 48$$

设 H_0 : 这些员工不比正常人容易生病 即 $\mu \leq \mu_0$

设显著水平为 α . 设 $\bar{X} - \mu_0 \geq c$ 时拒绝 H_0 . (H_0 成立时, V 参数, $\alpha(R) \leq \alpha$)



$$\text{即 } \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \geq \frac{c\sqrt{n}}{s}$$

$$\text{令 } \frac{c\sqrt{n}}{s} = t_\alpha$$

$$c = \frac{s t_\alpha}{\sqrt{n}}$$

令 $\alpha = 0.05$, 有 $t_\alpha = 1.6772$

$\Rightarrow \bar{X} \geq 5.699$ 时, 拒绝 H_0 .

$\Rightarrow$ 该公司员工确实更易生病.

4. ⁽¹⁾ 设 $X \sim N(\mu, \sigma^2)$, $\bar{X} \sim (\mu, \frac{\sigma^2}{n})$

$H_1: \mu \geq 1180$.

$H_0: \mu < 1180$. 令 $\mu_0 = 1180$, $S = 99.7497$, $\bar{X} = 1160$.

设 $\bar{X} - \mu_0 \geq c$ 就拒绝 H_0 . 当 H_0 成立时,

要求 $P(\bar{X} - \mu_0 \geq c) \leq \alpha$

当 $\mu = \mu_0$ 时:

$$\frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

$$\text{令 } \frac{c\sqrt{n}}{S} = t_\alpha$$

$$\text{有 } c = \frac{t_\alpha S}{\sqrt{n}}, \text{ 令 } \alpha = 0.05, t_\alpha = 2.1318$$

$$\bar{X} \geq \mu_0 + \frac{t_\alpha S}{\sqrt{n}} = 1275.1 \text{ 拒绝 } H_0.$$

$\bar{X}$ 不满足拒绝 H_0 的条件.

故认为不合格.

(2) 互换: $H_0: \mu \geq 1180 = \mu_0$, $H_1: \mu < 1180$.

易知: $\bar{X} < \mu_0 - \frac{t_\alpha S}{\sqrt{n}}$ 时拒绝 H_0 .

即 $\bar{X} \geq \mu_0 - \frac{t_\alpha S}{\sqrt{n}} = 1084.9$ 时接受 $\mu \geq 1180$.

故认为合格.

原假设是受保护的, 而我们仅限制 $\alpha(R) \leq \alpha$.

这样达到的效果是: 没有充分的证据就不拒绝 H_0 .

(3) 令 $\alpha \rightarrow 1$. 则 (1) 中总拒绝 H_0 . 即总认为产品合格.

(2) 中总拒绝 H_0 . 即总认为不合格.

5. 设 $\bar{X} \sim N(\mu, \sigma^2)$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), S = 98.726.$$

H_1 : 寿命 $\mu > \mu_0 = 225$.

H_0 : $\mu \leq \mu_0$.

设 $\bar{X} - \mu_0 \geq c$ 拒绝 H_0 .

当 H_0 成立时,

$$P(\bar{X} - \mu_0 \geq c) \leq \alpha.$$

即 $\mu = \mu_0$ 时:

$$P(\bar{X} - \mu_0 \geq c) \leq \alpha$$

$$\text{而此时: } \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t(n-1)$$

$$\therefore \text{令 } \frac{c\sqrt{n}}{S} = t_\alpha \Rightarrow c = \frac{S t_\alpha}{\sqrt{n}}$$

$$\text{令 } \alpha = 0.05, t_\alpha = 1.753.$$

$$\Rightarrow \bar{X} \geq \mu_0 + \frac{S t_\alpha}{\sqrt{n}} = 268.3 \text{ 拒绝 } H_0.$$

而 $\bar{X} = 241.5$ 不足以拒绝 H_0 .

故 $\alpha = 0.05$ 时, 不应认为寿命大于 225 小时.

6. 认为 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ 其中 $\mu = \sigma^2 = 1$

用临界值检验, 认为 $|\bar{X} - \mu| \geq c$ 就拒绝 H_0 .

当 H_0 成立时, 需 $P(|\bar{X} - \mu| \geq c) \leq \alpha$.

$$\text{即 } P\left(\frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} \geq \frac{c\sqrt{n}}{\sigma}\right) \leq \alpha$$



$$\text{故取 } \frac{c\sqrt{n}}{\sigma} = z_{\frac{\alpha}{2}} \Rightarrow c = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

故一个可取的检验是 $|\bar{X} - \lambda_0| \geq z_{\frac{\alpha}{2}} \sqrt{\frac{\lambda}{n}}$, 就拒绝 H_0 .

7. (1) $\frac{1}{20}$

(2) $\frac{2}{7}$...

(3) $\frac{1}{2}$

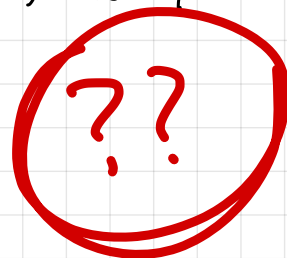
(4) 功效 $(1 - \beta(R)) = \frac{1}{2}$.

8. 不科学.

设 H_0 : 化疗不比外科方法有效. 设外科治愈率 p_0 , 化疗治愈率 p .

设每个患者治愈与否服从 $B(p)$, p 是治愈率.

用 CLT, 认为 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$. $\mu = p$, $\sigma^2 = p(1-p)$.

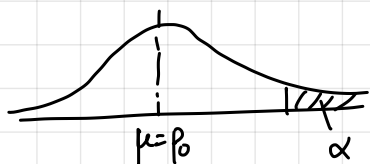


$$H_0 \Leftrightarrow p \leq p_0$$

给出检验: $\bar{X} - p_0 \geq c$ 时拒绝 H_0 .

当 H_0 成立时, 需 $P(\bar{X} - p_0 \geq c) \leq \alpha$

$$\text{即 } P\left(\frac{\bar{X} - p_0}{\frac{\sigma}{\sqrt{n}}} \geq \frac{c\sqrt{n}}{\sigma}\right) \leq \alpha$$



$$\text{取 } \frac{c\sqrt{n}}{\sigma} = z_{\alpha}$$

$$\text{即 } C = z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} = z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}$$

本题 $\bar{x} - p_0 = 0.01$.

令 $\alpha = 0.01$, $z_{\alpha} = 2.236$.

$C = 0.0221$, 不能拒绝 H_0 .

令 $\alpha = 0.1$, $z_{\alpha} = 1.282$

$C = 0.0127$, 不能拒绝 H_0 .

可以令 $z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}} = 0.01$

有 $z_{\alpha} = 1.0102$.

$\phi(z_{\alpha}) = 0.844$.

故在显著水平 $\alpha = 0.156$ 的程度上, 才能认为化疗更有效. 而这 α 太大.

9. 重复这实验 10 次, 可见实际上犯第 I 类错误的比例和 α 很接近

```
Python 3.8.11 (default, Aug 6 2021, 09:57:55) [MSC v.1916 64 bit (AMD64)] on win32
>>> runfile('C:/Users/joshu/PycharmProjects/scientificProject/hw12.py', wdir='C:/User
Error of type I 0.042
Error of type I 0.056
Error of type I 0.063
Error of type I 0.067
Error of type I 0.048
Error of type I 0.051
Error of type I 0.061
Error of type I 0.053
Error of type I 0.058
Error of type I 0.05
```

```
1 from scipy.stats import poisson, norm
2 import numpy as np
3 import math
4
5 def experiment(count: int, mu: float, alpha: float):
6     rej = 0
7     n = 20
8     for i in range(0, count):
9         sample = np.array(poisson.rvs(mu=mu, size=n))
10        if abs(sample.mean()-mu) >= norm.ppf(1-alpha/2)*math.sqrt(mu/n):
11            rej += 1
12    print('Error of type I', rej/count)
13
14 for i in range(10):
15     experiment(1000, 1, 0.05)
```


HW13

1. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t(n-1)$$

① 假设检验: (α)

$$H_0: \mu \leq \mu_0$$

$$H_1: \mu > \mu_0$$

设当 $\bar{X} - \mu_0 \geq c$ 拒 H_0 .

$$H_0 \text{ 成立时: } P(\bar{X} - \mu_0 \geq c) \leq \alpha$$

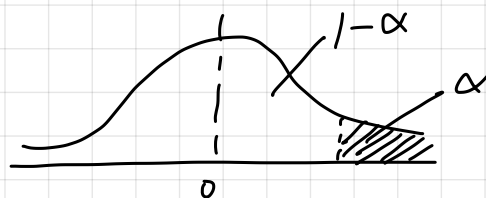
$$P\left(\frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \geq \frac{c\sqrt{n}}{s}\right) \leq \alpha$$

$$\text{由 } \frac{c\sqrt{n}}{s} = t_\alpha$$

$$\therefore \text{不拒绝} \iff \bar{X} < \mu_0 + \frac{t_\alpha s}{\sqrt{n}}$$

② 置信区间

$$\frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t(n-1)$$



$$\frac{\mu_0 - \bar{X}}{\frac{s}{\sqrt{n}}} > -t_\alpha$$

$$\text{即 } \mu_0 > \bar{X} - \frac{St_\alpha}{\sqrt{n}} \iff \bar{X} < \mu_0 + \frac{St_\alpha}{\sqrt{n}}$$

可得对偶关系: 在置信区间中 $\iff$ 不拒绝 H_0 .

2. 设 $\bar{X} \sim N(\mu, \sigma^2)$. $\mu_0 = 225$

$$H_0: \mu \leq 225$$

$$H_1: \mu > 225, \alpha = 0.05, \bar{X} \text{ 的观测值 } \bar{x}.$$

$$p\text{值}: P_{H_0}(\bar{X} > \bar{x})$$

$$= P_{H_0}\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P_{H_0}(T_n > 0.668)$$

$$= 0.257$$

$$p > \alpha$$

没有理由认为 $\mu > 225$.

3. 设 H_0 : 标注准确. $\mu = \mu_0$

$$H_1: \text{标注不准确 } \mu \neq \mu_0$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

H_0 成立时:

$$p = P(|\bar{X} - \mu_0| > |\bar{x} - \mu_0|)$$

$$= P\left(\frac{|\bar{X} - \mu_0|}{\frac{\sigma}{\sqrt{n}}} > \frac{|\bar{x} - \mu_0|}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P(|Z| > 2.828)$$

$$= 0.00468$$

$$\text{对 } \alpha = 0.05, 0.01$$

$$\text{均有 } p < \alpha$$

故拒绝 H_0 . 认为标注不准确

$$4. \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$H_0: \sigma \geq 0.1$$

$$H_1: \sigma < 0.1, \text{ 设 } S \text{ 的观测值 } x, \Sigma \sim \chi^2(n-1)$$

$$P_{H_0}(S < x)$$

$$= P_{H_0}\left(\frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)x^2}{\sigma^2}\right)$$

$$= P_{H_0}\left(\Sigma < \frac{(n-1)x^2}{\sigma^2}\right)$$

$$\leq P_{H_0}\left(\Sigma < \frac{(n-1)x^2}{\sigma_0^2}\right)$$

$$= P_{H_0}(\Sigma < 1.669)$$

$$= 0.00431 < 0.05$$

故拒绝 H_0 , 认为标准差小于 0.1

$$5. \text{ 设无心脏病的人 } \bar{X} \sim N(\mu_1, \sigma_1^2), \bar{Y} \sim N(\mu_2, \sigma_2^2)$$

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2 \quad \bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right), \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right).$$

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

用 s_1^2, s_2^2 估计 σ_1^2, σ_2^2 .

$$H_0 \text{ 成立时: } \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1)$$

$$p = P_{H_0}\left(\left|\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right| > \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}\right) = P(|Z| > 3.7719) = 0.000163.$$

$p < \alpha$, 拒绝 H_0 . 认为有显著不同.

6. (1) 錯

(2) 錯

(3) 錯

(4) 錯

(5) 錯. $1 - \beta(R)$ 未知

(6) 錯

$$7. (1) \frac{P(H_0|x)}{P(H_1|x)} = \frac{\frac{1}{2} \times \binom{10}{x} (\frac{1}{2})^{10}}{\frac{1}{2} \times \binom{10}{x} 0.7^x 0.3^{10-x}} = \frac{(\frac{1}{2})^{10}}{0.7^x 0.3^{10-x}}$$

检验: $\frac{(\frac{1}{2})^{10}}{0.7^x 0.3^{10-x}} > 1$, 不拒绝 H_0 .

$$\frac{(\frac{1}{2})^{10}}{0.7^x 0.3^{10-x}} < 1, \text{ 拒绝 } H_0.$$

$$\text{拒绝} \Leftrightarrow (0.5)^{10} < \left(\frac{0.7}{0.3}\right)^x (0.3)^{10}$$

$$\Leftrightarrow \left(\frac{5}{3}\right)^{10} < \left(\frac{7}{3}\right)^x$$

$$10 \ln \frac{5}{3} < x \ln \frac{7}{3}$$

$$x > 10 \frac{\ln 5 - \ln 3}{\ln 7 - \ln 3} = 6.02$$

故 $x \geq 7$ 拒绝 H_0

$x \leq 6$ 接受 H_0

$$\alpha(R) = P_{H_0}(X \geq 7) = 1 - P_{H_0}(X \leq 6) = 0.172$$

$$\beta(R) = P_{H_1}(X \leq 6) = 0.35$$

$$(2) \frac{P(H_0|x)}{P(H_1|x)} = \frac{\frac{10}{11} \binom{10}{x} 0.5^{10}}{\frac{1}{11} \binom{10}{x} 0.7^x 0.3^{10-x}}$$

不拒绝: $10 \cdot 0.5^{10} > \left(\frac{7}{3}\right)^x 0.3^{10}$

$$10 \cdot \left(\frac{5}{3}\right)^{10} > \left(\frac{7}{3}\right)^x$$

$$\ln 10 + 10 \ln \frac{5}{3} > x \ln \frac{7}{3}$$

$$x < \frac{10 \ln \frac{5}{3} + \ln 10}{\ln 7 - \ln 3} = 8.75$$

故 $x \leq 8$: 接受 H_0 .

$x \geq 9$: 拒绝 H_0 .

$$\alpha(R) = P_{H_0}(X \geq 9) = 1 - P_{H_0}(X \leq 8) = 0.011$$

$$\beta(R) = P_{H_1}(X \leq 8) = 0.85$$

8. ⁽¹⁾ H_0 : 均匀; H_1 : 不均匀

$$\chi^2 = \sum_{i=1}^6 \frac{(O_i - E_i)^2}{E_i} = \frac{1}{10} (1 + 1 + 0 + 4 + 4 + 0) = 1$$

$$\chi^2 \overset{\text{近似}}{\sim} \chi^2(5)$$

$$P(\chi^2 > 1) = 1 - P(\chi^2 \leq 1) = 0.96$$

P 很大, 认为 H_0 成立, 骰子均匀.

$$^{(2)} \chi^2 = \sum_{i=1}^6 \frac{(O_i - E_i)^2}{E_i} = \frac{1}{100} (100 + 100 + 400 + 400 + 0) = 10$$

$$\chi^2 \overset{\text{近似}}{\sim} \chi^2(5)$$

$P(\chi^2 > 10)$ 很小, 认为不均匀

9.

```
hw13
Python 3.8.11 (default, Aug 6 2021, 09:57:55) [MSC v.1916 64 bit (AMD64)] on win32
>>> runfile('C:/Users/joshu/PycharmProjects/scientificProject/hw13.py', wdir='C:/Users/joshu/PycharmProjects/scientificProject')
0.0
0.0
0.0
0.0
0.001
0.0
0.0
0.0
0.0
0.0
```

将这1000次实验再重复10次.

可见几乎没有犯第一类错误.

```
1 from scipy.stats import norm, t
2 import numpy as np
3
4 for j in range(10):
5     cnt = 0
6     for i in range(0, 1000):
7         arr = np.array(norm.rvs(loc=5.0, scale=1, size=100))
8         p = 1 - t.cdf((arr.mean()-5.2)/arr.std(ddof=True)*10, 99)
9         if p < 0.05:
10             cnt += 1
11
12 print(cnt/1000)
```

HW14

1. H_0 : 符合 Poisson 分布

$$\text{设 } X \sim \text{Poisson}(\lambda), P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$L(\lambda) = \prod_{k=0}^5 \left(\frac{\lambda^k e^{-\lambda}}{k!} \right)^{x_k}$$

$$\ln L = \sum_{k=0}^5 x_k (k \ln \lambda - \ln k! - \lambda)$$

$$\frac{d \ln L}{d \lambda} = \sum_{k=0}^5 x_k \left(\frac{k}{\lambda} - 1 \right)$$

$$= \sum_{k=0}^5 \left(\frac{x_k k}{\lambda} - x_k \right) = 0 \quad \text{得:}$$

$$\lambda n = \sum_{k=0}^5 k x_k$$

$$\Rightarrow \lambda = \sum_{k=0}^5 \frac{x_k k}{n} = \sum_{k=0}^5 f_k \cdot k = \bar{x} \Rightarrow \lambda^* = \frac{161}{200}$$

$$E_i = P(X=i) \times 200.$$

	0	E
0	92	89.42
1	68	71.98
2	28	28.97
3, 4, 5	12	9.63

$$\chi_0^2 = \sum_{r=1}^4 \frac{(O_r - E_r)^2}{E_r} = 0.9103$$

χ^2 : 合并组 $\Rightarrow E_i \geq 5$
自由度 $k-1-s$

$$\text{自由度为 } k-1-s = 4-1-1 = 2.$$

$$P\text{值} = P(\chi^2(2) \geq \chi_0^2) = 0.63 > \alpha = 0.05$$

不拒绝 H_0 , 认为服从 Poisson 分布

2. 假设是真实的,

$$E_1 = 25, E_2 = 25$$

$$O_1 = 22, O_2 = 28$$

$$\chi_0^2 = \sum_{i=1}^2 \frac{(O_i - E_i)^2}{E_i} = 0.72, \text{ 取 } \alpha = 0.1$$

临界值为 $\chi^2(1) = 0.15$.

$\chi_0^2 > 0.15$ 不拒绝 H_0 .

认为数据没有造假

3. H_0 : 是真实的

$$\chi^2 = \sum_{k=1}^4 \frac{(O_k - E_k)^2}{E_k}$$

$$= 0.5103$$

$$P_{\text{值}} = P(\chi^2(3) < \chi^2) = 0.08$$

$$P > \alpha$$

不拒绝 H_0 , 不认为是伪造的

4. H_0 : 设两者独立

类型 \ 人数 \ 烟量	x 支/日	y 支/日	z 支/日	总计
患病者人数	26	147	37	210
健康者人数	30	123	22	175
总计	56	270	59	385

$$p_{1+}^* = \frac{210}{385}, \quad p_{2+}^* = \frac{175}{385}$$

$$p_{+1}^* = \frac{56}{385}, \quad p_{+2}^* = \frac{270}{385}, \quad p_{+3}^* = \frac{59}{385}$$

期望值

30.55	147.27	20.18
25.45	122.73	26.82

$$\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 16.37$$

$$\chi^2(3)_\alpha = 6.25$$

$\chi^2 > 6.25$, 拒绝 H_0 , 认为有关联

5. 进行齐性检验.

居民	支持	反对	总计
城市	143	57	200
郊区	98	102	200
农村	13	87	100
总计	254 p_1^*	246 p_2^*	500

H_0 : 各类居民对政策意见一致

$$p_1^* = \frac{254}{500} = 0.508$$

$$p_2^* = \frac{246}{500} = 0.492$$

期望值

101.6	98.4
101.6	98.4
50.8	49.2

$$\begin{aligned}\chi_0^2 &= \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \\ &= 91.72\end{aligned}$$

$$\chi^2(3)_\alpha = 3.84.$$

$$\chi_0^2 > 3.84, \text{ 拒绝 } H_0.$$

认为各群体的想法不一致

6. Fisher 精确检验

Fisher 指出对 2×2 列联表. 在齐性的情况下.

	Men	Women	Row Total
Studying	a	b	$a + b$
Non-studying	c	d	$c + d$
Column Total	$a + c$	$b + d$	$a + b + c + d (=n)$

这表出现的概率为

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}} = \frac{\binom{a+b}{b} \binom{c+d}{d}}{\binom{n}{b+d}} = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{a! b! c! d! n!}$$

这样就可以枚举, 以计算不依赖大样本的精确 P 值.

如表

	Men	Women	Row total
Studying	1	9	10
Not-studying	11	3	14
Column total	12	12	24

出现的概率就是

$$p = \frac{\binom{10}{1} \binom{14}{11}}{\binom{24}{12}} = \frac{10! 14! 12! 12!}{1! 9! 11! 3! 24!} \approx 0.001346076$$

7.

O_i	E_i
315	313
101	104
108	104
32	35

$$\chi_0^2 = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} = 0.5103$$

H_0 : 理论有效, $\alpha = 0.1$

$$P\text{值} = P(\chi^2(3) \geq \chi_0^2) = 0.9167 > \alpha.$$

故不拒绝 H_0 , 认为理论有效

8. 先用以下代码预实验

```
1 from scipy.stats import cauchy, norm, chi2
2 import numpy as np
3
4 def experiment(n):
5     print(n)
6     data = np.array(cauchy.rvs(size=n))
7     cnt = [0, 0, 0, 0, 0, 0]
8     mu = data.mean()
9     sig = data.std()
10    expe = np.array([
11        norm.cdf(-20, loc=mu, scale=sig) - 0,
12        norm.cdf(-10, loc=mu, scale=sig) - norm.cdf(-20, loc=mu, scale=sig),
13        norm.cdf(0, loc=mu, scale=sig) - norm.cdf(-10, loc=mu, scale=sig),
14        norm.cdf(10, loc=mu, scale=sig) - norm.cdf(0, loc=mu, scale=sig),
15        norm.cdf(20, loc=mu, scale=sig) - norm.cdf(10, loc=mu, scale=sig),
16        1 - norm.cdf(20, loc=mu, scale=sig),
17    ]) * n
18    for d in data:
19        if d < -20:
20            cnt[0] += 1
21        elif d < -10:
22            cnt[1] += 1
23        elif d < 0:
24            cnt[2] += 1
25        elif d < 10:
26            cnt[3] += 1
27        elif d < 20:
28            cnt[4] += 1
29        else:
30            cnt[5] += 1
31
32    print(cnt)
33    print(expe)
```

得到

```
critical value 16.918977684628448
50
[0, 0, 13, 16, 0, 1]
[ 0.08456393 -1.87636852 -10.02778456 -13.27042117 -4.38308287
 0.35777895]
100
[1, 0, 41, 56, 1, 1]
[ 3.73516852e+04 -1.08313295e+08 -4.22375266e+01 -5.43992522e+01
 -2.35797329e+00 1.74148897e-03]
1000
[15, 20, 400, 460, 14, 11]
[ 118.2054188 -164.51984726 -231.0261366 -226.36928988 -154.76891926
 105.11038901]
```

可见对 $n=30$, $n=100$, 将分组合并后
 $k-1-s \leq 1$. 无法使用 Pearson 卡方检验.

对 $n=1000$, 使用以下代码计算 χ^2

```
sm = 0
for i in range(0, 6):
    sm += (expe[i] - cnt[i])**2 / expe[i]
print(sm)
```

得到

```
1000
[21, 15, 470, 459, 19, 16]
[280.56121689 106.50421009 115.81631983 115.57221956 105.83220641
 275.71382723]
2738.292587566665
```

可见 χ^2 很大, 拒绝 H_0 . 不符合正态分布.

HW 15

1. 设 $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$, $\mu^* = \bar{x}$

$$L(\mu) = \prod_{i=1}^n f(x_i)$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

$$\Lambda = \frac{L(\mu)}{L(\mu^*)} = e^{-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (x_i - \mu)^2 - \sum_{i=1}^n (x_i - \mu^*)^2 \right)}$$

$$= e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (2x_i - \mu - \mu^*) (\mu^* - \mu)}$$

$$\begin{aligned} -2 \ln \Lambda &= \frac{1}{\sigma^2} (\mu^* - \mu) (\sum_{i=1}^n 2x_i - n\mu - n\mu^*) \\ &= \frac{n}{\sigma^2} (\mu^* - \mu)^2 \end{aligned}$$

$$-2 \ln \Lambda \sim \chi^2(1)$$

而我们又通 $\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$

故其与Z检验结果一致

2. H_0 : 符合理论. $p_1 = \frac{9}{16}$, $p_2 = p_3 = \frac{3}{16}$, $p_4 = \frac{1}{16}$

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{L(\theta^*)}, \quad L(p_1, \dots, p_k) = \binom{n}{n_1, n_2, n_3, n_4} \prod_{i=1}^4 p_i^{n_i}$$

$$\frac{L_{H_0}(p_1, \dots, p_4)}{L(p_1^*, \dots, p_4^*)} = \prod_{i=1}^4 \left(\frac{p_i}{p_i^*} \right)^{n_i}$$

$$-2 \ln \Lambda = 2 \ln \prod_{i=1}^4 \left(\frac{p_i^*}{p_i} \right)^{n_i} = 2 \left(\sum_{i=1}^4 n_i \left(\ln \frac{n_i}{p_i n} \right) \right)$$

$$x \ln \frac{x}{x_0} = 0 + (x - x_0) + \frac{(x - x_0)^2}{2x_0} + o((x - x_0)^2)$$

$$\text{故 } -2\ln \Lambda \approx 2 \left(\sum_{i=1}^4 (O_i - E_i) + \sum_{i=1}^4 \frac{(O_i - E_i)^2}{2E_i} \right) \\ = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} = 0.9103$$

$$P\text{值} = P(\chi^2(3) \geq 0.9103) = 0.63 > 0.05$$

不拒绝 H_0

$$3. L_p(p) = \binom{n}{x} p^x (1-p)^{n-x}, L = L_1(p_1) L_2(p_2)$$

$$H_0: p_1 = p_2$$

$$H_1: p_1 > p_2$$

$$\Lambda = \frac{\sup_{p_1=p_2} L_1(p_1) L_2(p_2)}{L_1(p_1^*) L_2(p_2^*)}$$

$$\frac{\partial \ln L}{\partial p_i} = \frac{\partial \ln L_i}{\partial p_i}, \text{ 故 } p_1^* = 0.0217$$

$$p_2^* = 0.0126$$

$$\max_{p_1=p_2} \ln L \Rightarrow \nabla \ln L \parallel (1, -1)$$

$$\frac{\partial \ln L_1}{\partial p} + \frac{\partial \ln L_2}{\partial p} = 0$$

$$-2\ln \Lambda = -2\ln \frac{p_1^{x_1+x_2} (1-p_1)^{n_1+n_2-(x_1+x_2)}}{p_1^{x_1} p_2^{x_2} (1-p_1)^{n_1-x_1} (1-p_2)^{n_2-x_2}}$$

$$= -2 \left((x_1+x_2) \ln p^* + (n_1+n_2-x_1-x_2) \ln (1-p^*) \right) \Rightarrow \frac{x_1+x_2}{p^*} + \frac{n_1+n_2-(x_1+x_2)}{1-p^*} = 0$$

$$-x_1 \ln p_1^* - x_2 \ln p_2^* - (n_1-x_1) \ln (1-p_1^*)$$

$$- (n_2-x_2) \ln (1-p_2^*)$$

$$\text{即 } (1-p^*)(x_1+x_2) = p^*(n_1+n_2) - p^*(x_1+x_2)$$

$$\Rightarrow p^* = \frac{x_1+x_2}{n_1+n_2} \text{ (多个多项分布)}$$

若情况相同, 数据合并)

$$= 27.2612$$

$$\chi_{\alpha}^2(1) = 3.84 \text{ 拒绝域 } -2\ln \Lambda > \chi_{\alpha}^2(1)$$

$-2\ln \Lambda > 3.84$, 拒绝 H_0 , 认为对降低发病率有效

4. $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 \neq \sigma_2^2$

$$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1)$$

$$\frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$$

$$\text{令 } F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}}$$

H_0 成立时 $F = \frac{S_1^2}{S_2^2} \sim F(4, 6)$

取 $\alpha = 0.05$.

$$F_{\frac{\alpha}{2}}(4, 6) = 6.227$$

$$F_{1-\frac{\alpha}{2}}(4, 6) = \frac{1}{F_{\frac{\alpha}{2}}(6, 4)} = 0.108$$

拒绝域: $F < 0.108$ 或 $F > 6.227$.

而 $F = 0.086 < 0.108$

故拒绝 H_0 : 认为方差⁴⁴不相等.

5. ¹¹⁾ $H_0: \mu_1 = \mu_2$
 设 $\bar{X} \sim N(\mu_1, \sigma^2)$, $\bar{Y} \sim N(\mu_2, \sigma^2)$

$$\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, (\frac{1}{n} + \frac{1}{m}) \sigma^2)$$

$$\bar{X} = 2.063, S_1,$$

$$\bar{Y} = 2.059, S_2,$$

$$S^2 = \frac{1}{n+m-2} ((n-1)S_1^2 + (m-1)S_2^2) = 7.22 \times 10^{-6}$$

$$S = 0.0269.$$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2)$$

当 H_0 成立时: $t_0 = 3.32$

$$t_{\frac{\alpha}{2}}(18) = 2.101$$

$$t_0 > 2.101 \Rightarrow \text{拒绝 } H_0.$$

认为 $\mu_1 \neq \mu_2$, 生产不稳定 ~~稳定~~

12) $H_0: \text{方差 } \sigma_1^2 = \sigma_2^2, \alpha = 0.05$

$$F = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(9, 9)$$

$$F_{\frac{\alpha}{2}}(9, 9) = 4.026$$

$$F_{1-\frac{\alpha}{2}}(9, 9) = \frac{1}{F_{\frac{\alpha}{2}}(9, 9)} = 0.248$$

当 H_0 成立时: $F = \frac{S_1^2}{S_2^2} = 1.95$, 不拒绝 H_0 , 认为方差一致

6. 构造差异列 $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \sigma^2)$

[17, -2, 11, 13, 18, 15, 7, -1, 13, -2]

设 $D \sim N(\mu, \sigma^2)$, σ^2 未知, $S = 7.906$

$H_0: \mu = 0, H_1: \mu \neq 0$

当 H_0 成立时 $\frac{\bar{D}}{\frac{S}{\sqrt{n}}} \sim t(n-1)$

$t_0 = 3.55$

$t_{\frac{\alpha}{2}}^{(n-1)} = 2.26$

$|t_0| > t_{\frac{\alpha}{2}}^{(n-1)}$ 拒绝 H_0

认为两组成绩有差异

化成正态总体的
双样本检验

7. (1) 先看 A: 当 H_0 成立时: $\frac{\bar{X} - 100}{\frac{S}{\sqrt{10}}} \sim t(9)$

而 $\frac{\bar{X} - 100}{\frac{S_A}{\sqrt{10}}} = -2.45$



$t_{1-\alpha}(9) = -1.83 \Rightarrow -2.45 < -1.83$ $S_B = 50\sqrt{2}$

拒绝 H_0

$\bar{Y} = -150, S_B^2 = \frac{1}{1} \times 2 \times 2500$

再看 B: H_0 成立时: $\frac{\bar{Y} - 100}{\frac{S_B}{\sqrt{2}}} = \frac{-250\sqrt{2}}{S_B} = -5$

~~$\Rightarrow (n-1)$~~

~~$t_{1-\alpha}(2) = -2.92 \Rightarrow -5 < -2.92$~~ 拒绝 H_0

~~6.31~~

(2) 相样本数量较大时, 不明显的差异也会被放大造成不符合直观的结果

