Homework 10

抹茶奶绿

May 6, 2025

Exercise 1.

（简单随机抽样）设总体大小为 N，总体均值和方差分别为 μ, σ^2，$X_i\ (i=1,\dots,n)$ 为无放回抽取的简单随机样本，$\hat{\sigma}^2 = \frac{1}{n}\sum_i (X_i - \bar{X})^2$

(1) 证明：

$$E\left[\hat{\sigma}^2\right] = \frac{N}{N-1} \cdot \frac{n-1}{n} \cdot \sigma^2$$

(2) 给出 $\mathrm{Var}(\bar{X})$ 的无偏估计

Solution.

(1) 由无放回简单随机抽样可得

$$E[X_i] = \mu, \quad \mathrm{Var}(X_i) = \sigma^2, \quad \mathrm{Cov}(X_i, X_j) = -\frac{\sigma^2}{N-1} \quad (i \neq j)$$

则

$$\begin{aligned}\mathrm{Var}(\bar{X}) &= \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}(X_i, X_j) \\ &= \frac{1}{n^2}\left(n\sigma^2 - \frac{n(n-1)\sigma^2}{N-1}\right) \\ &= \frac{\sigma^2}{n}\frac{N-n}{N-1}\end{aligned}$$

于是

$$\begin{aligned}E\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] &= E\left[\sum_i X_i^2\right] - n\,E[\bar{X}^2] \\ &= n(\sigma^2 + \mu^2) - n\big(\mathrm{Var}(\bar{X}) + \mu^2\big) \\ &= n\sigma^2 - n \cdot \frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{N(n-1)}{N-1}\sigma^2\end{aligned}$$

故

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

(2) 已知

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1}$$

令

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

由 (1) 得 $E[S^2] = \frac{N}{N-1}\sigma^2$ 故 $\tilde{\sigma}^2 = \frac{N-1}{N}S^2$ 为 σ^2 的无偏估计. 代入得

$$\widehat{\mathrm{Var}}(\bar{X}) = \frac{\tilde{\sigma}^2}{n}\frac{N-n}{N-1} = \frac{N-n}{nN(n-1)}\sum_{i=1}^n (X_i - \bar{X})^2$$

Exercise 2.
设 X 来自 Poisson 总体 $\mathrm{P}(\lambda)$
(1) 证明：对于 $g(\lambda)=e^{-2\lambda}$，其唯一无偏估计为

$$\hat{\theta}=\begin{cases}1, & X\text{ 偶数},\\ -1, & X\text{ 奇数}.\end{cases}$$

(2) 判断该估计是否合理如不合理，给出一个更合适的估计

Solution.

(1)

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

设 $\delta(X)$ 是另一个对 $g(\lambda)=e^{-2\lambda}$ 的无偏估计，则

$$E_\lambda[\delta(X)]=e^{-2\lambda},\quad \forall\lambda\geq 0$$

令

$$h(X)=\delta(X)-(-1)^X$$

则

$$E_\lambda[h(X)]=E_\lambda[\delta(X)]-E_\lambda[(-1)^X]=e^{-2\lambda}-e^{-2\lambda}=0,\quad \forall\lambda$$

但 Poisson 分布的统计量 X 是关于 λ 的完备统计量，即若 $E_\lambda[h(X)]=0$ 对所有 λ 成立
则 $h(X)=0$ a.s.，也即

$$\delta(X)=(-1)^X\quad\text{几乎处处成立}$$

(2) $\hat{\theta}$ 虽无偏，但方差

$$\mathrm{Var}(\hat{\theta})=1-e^{-4\lambda}$$

在多数 λ 值下接近 1，与 $g(\lambda)$（趋近 0）反差极大，故不够稳定
可改用 MLE 估计：

$$\tilde{\theta}=e^{-2X}$$

其期望

$$E[\tilde{\theta}]=\sum_{k=0}^{\infty}e^{-2k}\frac{e^{-\lambda}\lambda^k}{k!}=\exp\big(\lambda(e^{-2}-1)\big)$$

有偏，但均方误差

$$E\big[(\tilde{\theta}-e^{-2\lambda})^2\big]$$

通常显著小于 UMVUE 的 MSE $1-e^{-4\lambda}$，因而在 MSE 意义下更优且收敛一致

Exercise 3.

设样本 $X_i\ (i=1,\dots,n)$ 来自 $U(0,\theta)$

(1) 证明 $\hat\theta_1=\max_i X_i+\min_i X_i$ 是 θ 的无偏估计

(2) 证明存在常数 c_n 使得 $\hat\theta_2=c_n\min_i X_i$ 也是无偏估计

(3) 比较 $\hat\theta_1,\hat\theta_2,\hat\theta_3=2\bar X,\hat\theta_4=\frac{n+1}{n}\max_i X_i$ 四个无偏估计的方差

Solution.

次序统计量

$$X_{(1)}=\min_{1\le i\le n}X_i \qquad X_{(n)}=\max_{1\le i\le n}X_i$$

- $X_{(1)}$:

$$F_{X_{(1)}}(x)=1-P(X_1>x,\dots,X_n>x)=1\Big(1-\frac{x}{\theta}\Big)^n,\quad 0\le x\le\theta,$$
$$f_{X_{(1)}}(x)=\frac{d}{dx}F_{X_{(1)}}(x)=\frac{n}{\theta}\Big(1-\frac{x}{\theta}\Big)^{n-1},\quad 0\le x\le\theta.$$

- $X_{(n)}$:

$$F_{X_{(n)}}(x)=P(X_1\le x,\dots,X_n\le x)=\Big(\frac{x}{\theta}\Big)^n,\quad 0\le x\le\theta,$$
$$f_{X_{(n)}}(x)=\frac{d}{dx}F_{X_{(n)}}(x)=\frac{n}{\theta^n}x^{n-1},\quad 0\le x\le\theta.$$

(1)

$$E[X_{(n)}]=\int_0^\theta x\,f_{X_{(n)}}(x)\,dx=\int_0^\theta x\,\frac{n}{\theta^n}x^{n-1}\,dx=\frac{n}{n+1}\theta,$$
$$E[X_{(1)}]=\int_0^\theta x\,f_{X_{(1)}}(x)\,dx=\int_0^\theta x\,\frac{n}{\theta}\Big(1-\frac{x}{\theta}\Big)^{n-1}dx=\frac{1}{n+1}\theta.$$
$$E[\hat\theta_1]=E[X_{(n)}]+E[X_{(1)}]=\frac{n}{n+1}\theta+\frac{1}{n+1}\theta=\theta.$$

(2)

$$\hat\theta_2=(n+1)X_{(1)},\quad E[\hat\theta_2]=(n+1)\frac{\theta}{n+1}=\theta$$

(3)

$$\mathrm{Var}(X_{(1)})=\frac{n}{(n+1)^2(n+2)}\theta^2 \qquad \mathrm{Var}(X_{(n)})=\frac{n}{(n+1)^2(n+2)}\theta^2$$
$$\mathrm{Cov}(X_{(1)},X_{(n)})=\frac{1}{(n+1)^2(n+2)}\theta^2 \qquad \mathrm{Var}(\bar X)=\frac{\theta^2}{12n}$$

于是

$$\mathrm{Var}(\hat\theta_1)=2\frac{n}{(n+1)^2(n+2)}\theta^2+2\frac{1}{(n+1)^2(n+2)}\theta^2=\frac{2}{(n+1)(n+2)}\theta^2$$
$$\mathrm{Var}(\hat\theta_2)=(n+1)^2\frac{n}{(n+1)^2(n+2)}\theta^2=\frac{n}{n+2}\theta^2$$
$$\mathrm{Var}(\hat\theta_3)=4\frac{\theta^2}{12n}=\frac{\theta^2}{3n}$$
$$\mathrm{Var}(\hat\theta_4)=\frac{(n+1)^2}{n^2}\frac{n}{(n+1)^2(n+2)}\theta^2=\frac{1}{n(n+2)}\theta^2$$

故

$$\mathrm{Var}(\hat\theta_4)\ <\ \mathrm{Var}(\hat\theta_1)\ <\ \mathrm{Var}(\hat\theta_3)\ <\ \mathrm{Var}(\hat\theta_2)$$

Exercise 4.

设 X_i $(i=1,\dots,n)$ 来自均值为 θ、方差有限的总体

(1) 设估计 $\sum_i c_i X_i$，证明无偏当且仅当 $\sum_i c_i = 1$

(2) 在此类估计中，方差最小当且仅当 $c_i = 1/n\ \forall i$

Solution.

(1) $E[\sum c_i X_i] = \theta \sum c_i = \theta \iff \sum c_i = 1$

(2) $\mathrm{Var}(\sum c_i X_i) = \sigma^2 \sum c_i^2$，在 $\sum c_i = 1$ 的意义下，由柯西不等式取等条件即知 $c_i = 1/n$

Exercise 5.

设 X_i $(i=1,\dots,n)$ 来自 $N(\mu,\sigma^2)$，两种估计 m_2 和 S^2 均可估计 σ^2，比较它们的均方误差

Solution.

由于

$$m_2 = \frac{n-1}{n} S^2, \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

那么

$$E[S^2] = \sigma^2, \quad \mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

$$E[m_2] = \frac{n-1}{n}\sigma^2, \quad \mathrm{Var}(m_2) = \left(\frac{n-1}{n}\right)^2 \mathrm{Var}(S^2) = \frac{2(n-1)\,\sigma^4}{n^2}$$

所以

$$\mathrm{MSE}(S^2) = \mathrm{Var}(S^2) + \left(E[S^2] - \sigma^2\right)^2 = \frac{2\sigma^4}{n-1}$$

$$\mathrm{MSE}(m_2) = \mathrm{Var}(m_2) + \left(E[m_2] - \sigma^2\right)^2 = \frac{2(n-1)\sigma^4}{n^2} + \left(-\frac{\sigma^2}{n}\right)^2 = \frac{2n-1}{n^2}\sigma^4$$

我们有

$$\frac{2n-1}{n^2} < \frac{2}{n-1} \implies \mathrm{MSE}(m_2) < \mathrm{MSE}(S^2)$$

即使 m_2 有偏，其均方误差仍小于无偏估计 S^2

Exercise 6.

设 $X_1,\dots,X_4 \sim N(0,4)$，令

$$Y = \frac{X_1 + X_2}{\sqrt{X_3^2 + X_4^2}} \cdot a$$

已知 Y 服从 t 分布，求 a 及自由度

Solution.

令

$$X_i = 2Z_i, \quad Z_i \sim N(0,1)$$

则

$$X_1 + X_2 = 2(Z_1 + Z_2), \quad X_3^2 + X_4^2 = 4(Z_3^2 + Z_4^2)$$

令

$$U = \frac{Z_1 + Z_2}{\sqrt{2}} \sim N(0,1), \qquad V = Z_3^2 + Z_4^2 \sim \chi^2_2$$

于是

$$Y = \frac{2(Z_1 + Z_2)}{2\sqrt{V}}\, a = \frac{\sqrt{2}\,U}{\sqrt{V}}\, a = a\,\frac{U}{\sqrt{V/2}}$$

因 $U \sim N(0,1)$, $V \sim \chi_2^2$ 独立，故

$$\frac{U}{\sqrt{V/2}} \sim t_2$$

要使 $Y \sim t$ 分布且无额外因子，则

$$a = 1, \quad \text{自由度} = 2$$

Exercise 7.
16 袋糖果重量（克）如下：

$$506, 508, 499, 503, 504, 510, 497, 512, 514, 505, 493, 496, 506, 502, 509, 496.$$

假设正态，求总体均值的 95% 置信区间；若以样本均值为估计，误差范围及其意义

Solution.

$$\bar{x} = 503.75, \quad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = 6.2022.$$

又

$$t_{0.975,15} = 2.13145.$$

那么

$$\bar{x} \pm t_{0.975,15}\frac{s}{\sqrt{n}} = 503.75 \pm 2.13145\frac{6.2022}{\sqrt{16}} = 503.75 \pm 3.3049$$

即

$$[500.445, \ 507.055]$$

误差范围（置信半宽）为

$$E = t_{0.975,15}\frac{s}{\sqrt{n}} = 3.3049$$

其意义是：若多次重复抽样并构造相同形式的置信区间，则约有 95% 的区间会包含真实的总体均值 μ

Exercise 8.
5 只灯泡寿命（小时）：

$$1050, 1100, 1120, 1250, 1280.$$

假设正态，求均值的 95% 单侧下限

Solution.

$$\bar{x} = 1160 \qquad s = \sqrt{9950} \approx 99.75$$

又

$$t_{0.05,4} = 2.132$$

故总体均值 μ 的单侧 95% 置信下限为

$$\mu \geq \bar{x} - t_{0.05,4} \cdot \frac{s}{\sqrt{n}} = 1160 - 2.132 \cdot \frac{99.75}{\sqrt{5}} \approx 1160 - 95.14 = 1064.86$$

Exercise 9.
原催化剂 20 次：$\bar{x}_1 = 91.73,\ s_1^2 = 3.89$；新催化剂 30 次：$\bar{x}_2 = 93.75,\ s_2^2 = 4.02$ 假设正态且方差相等，
(1) 求两均值差的 95% 置信区间；
(2) 判断是否显著差异

Solution.

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{19\cdot 3.89 + 29\cdot 4.02}{48} = \frac{190.49}{48} \approx 3.9685, \quad s_p = \sqrt{3.9685} \approx 1.9921.$$

自由度：$\nu = n_1 + n_2 - 2 = 48$，查表得 $t_{0.975,48} \approx 2.010$
样本均值差：$\bar{x}_1 - \bar{x}_2 = -2.02$
标准误差：

$$SE = s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1.9921 \cdot \sqrt{\frac{1}{20} + \frac{1}{30}} \approx 1.9921 \cdot \sqrt{0.0833} \approx 0.5756$$

置信区间为：

$$\bar{x}_1 - \bar{x}_2 \pm t \cdot SE = -2.02 \pm 2.010 \cdot 0.5756 = -2.02 \pm 1.157 = [-3.177,\ -0.863]$$

因置信区间不包含 0，故两个催化剂效果有显著差异

Exercise 10.
设 $X_i \sim U(0,\theta)$，证明：对任意 $\alpha \in (0,1)$，存在 c_n 使

$$\left[\max_i X_i,\ \max_i X_i \cdot c_n\right]$$

是 θ 的 $1-\alpha$ 置信区间

Solution.
记最大值 $M = \max_i X_i$，有

$$F_M(x) = \left(\frac{x}{\theta}\right)^n, \quad 0 < x < \theta,$$

$$f_M(x) = \frac{d}{dx}F_M(x) = \frac{n}{\theta^n}x^{n-1}, \quad 0 < x < \theta.$$

我们要求：

$$P\left(\theta \in [M,\ c_nM]\right) = 1-\alpha \quad \Longleftrightarrow \quad P\left(\theta \le c_nM\right) = 1-\alpha \quad \Longleftrightarrow \quad P\left(M \ge \frac{\theta}{c_n}\right) = 1-\alpha$$

利用 M 的 CDF，得

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

故

$$c_n = \alpha^{-1/n}$$

Exercise 11.
（Bootstrap 自助法）balabala

Solution.

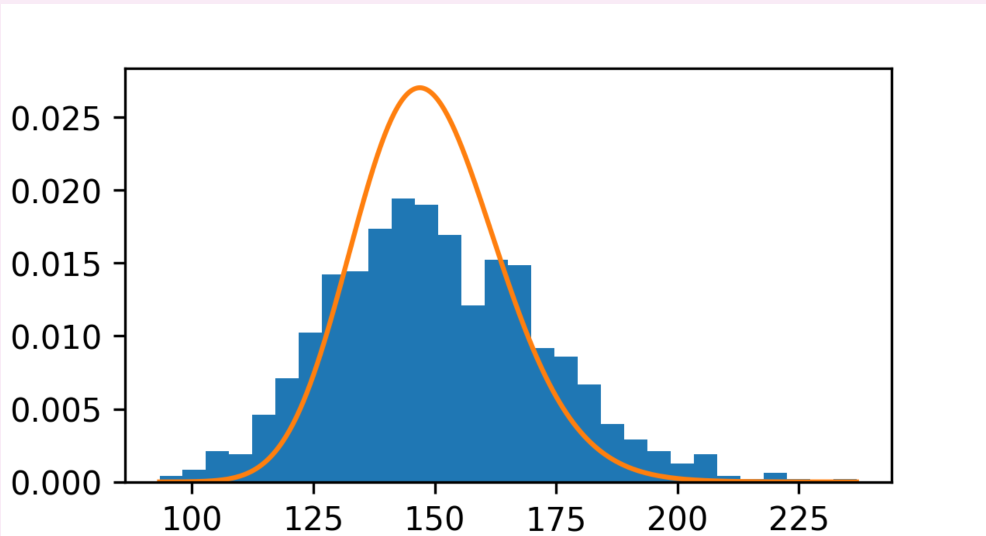
```
from random import normalvariate as norm, choice
from math import exp, sqrt, pi, log, e
from matplotlib import pyplot as plt
import numpy as np

def data_set(n: int):
    data = []
    observation = []
    for i in range(0, n):
        data += [norm(5, 1)]
    for i in range(0, n):
        observation += [choice(data)]
    return exp(sum(observation) / n)

theta = []
log_norm = lambda x, mu, sigma: 1 / (x * sqrt(2 * pi) * sigma) * exp(-1 / (2 * (sigma **
for i in range(0, 1000):
    theta += [data_set(100)]

dots = np.linspace(min(theta), max(theta), 1000)
plt.figure(dpi=300)
plt.hist(theta, bins=30, density=True)
plt.plot(dots, [log_norm(x, 5, 1/10) for x in dots])
plt.show()

print(np.array(theta).var(), (exp(1/100)-1)*exp(2*5+1/100))
```



$\log\hat{\theta} \sim N(5, 1/100)$，分布图如上所示，
而 V_{boot} 与 $Var(\hat{\theta})$ 十分接近，故可以.