概率论与数理统计 第10次作业

Name: 宋昊原　　Student ID: 2022010755

November 24, 2023

1　简单随机抽样

1.1

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

考虑到对称性，有

$$E(\hat{\sigma}^2) = E(X_1^2) - 2E(X_1\bar{X}) + E(\bar{X}^2)$$

$$= E(X_1^2) - \frac{2}{n}E(X_1^2) - \frac{2(n-1)}{n}E(X_1X_2) + \frac{1}{n}E(X_1^2) + \frac{n-1}{n}E(X_1X_2)$$

$$= \frac{n-1}{n}E(X_1^2) - \frac{n-1}{n}E(X_1X_2)$$

其中

$$E(X_1^2) = \mu^2 + \sigma^2$$

$$E(X_1X_2) = \frac{1}{N(N-1)}\sum_{i\neq j}x_ix_j$$

$$= \frac{1}{N(N-1)}((\sum_i x_i)^2 - \sum_i x_i^2)$$

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

故

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

1.2

作业9-3计算过

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

故

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

当N, n都已知时

$$\frac{N-n}{N(n-1)}\hat{\sigma}^2$$

是$Var(\bar{X})$的一个无偏估计.

2 Poisson分布的估计

2.1

Poisson分布为

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

首先证明$\hat{\theta}$是无偏估计

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

$$=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k(-1)^k}{k!}e^{-\lambda}=e^{-\lambda}e^{-\lambda}=e^{-2\lambda}$$

再证明唯一性，此时样本容量为1，故估计量只能是X的函数，若此函数是无偏估计，则

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

故

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

已知

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}(-1)^k=e^{-\lambda}$$

即

$$\forall\lambda>0,\ \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\hat{\theta}(k)=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}(-1)^k$$

这两个幂级数收敛于同一函数必须要求其系数分别相等，即

$$\hat{\theta}(X)=(-1)^X$$

是唯一的无偏估计.

2.2

上述估计不合理，由于$\lambda>0$，有

$$0<e^{-2\lambda}<1$$

故上述估计给出的是完全荒谬的结果.

要给出合理的估计，可以采用极大似然估计，似然函数为

$$L(\theta)=\frac{\lambda^X}{X!}e^{-\lambda}$$

故

$$\frac{dL}{d\theta}=\frac{X\lambda^{X-1}e^{-\lambda}-\lambda^Xe^{-\lambda}}{X!}$$

令此值为0，则有

$$\lambda=X$$

这确实是最大值.
实际上，由于$E(X)=\lambda$，这和矩估计给出相同的结果.

3 均匀总体的估计

3.1

考虑最小值、最大值的CDF

$$F_1(x) = 1 - \frac{(\theta - x)^n}{\theta^n}$$

$$F_n(x) = \frac{x^n}{\theta^n}$$

则PDF

$$f_1(x) = \frac{n(\theta - x)^{n-1}}{\theta^n}$$

$$f_n(x) = \frac{nx^{n-1}}{\theta^n}$$

故

$$E(\hat{\theta_1}) = \int_0^\theta \frac{nx(\theta - x)^{n-1}}{\theta^n}dx + \int_0^\theta \frac{nx^n}{\theta^n}dx$$

$$= \int_0^\theta \frac{n(\theta - x)x^{n-1}}{\theta^n}dx + \int_0^\theta \frac{nx^n}{\theta^n}dx$$

$$= \int_0^\theta \frac{nx^{n-1}}{\theta^{n-1}}dx = \theta$$

故$\hat{\theta_1}$是θ的无偏估计.

3.2

考虑最小值X_{min}的期望

$$E(X_{min}) = \int_0^\theta \frac{nx(\theta - x)^{n-1}}{\theta^n}dx$$

$$= \int_0^\theta \frac{n(\theta - x)x^{n-1}}{\theta^n}dx = \int_0^\theta \frac{n\theta x^{n-1} - nx^n}{\theta^n}dx$$

$$= \theta - \frac{n}{n+1}\theta = \frac{\theta}{n+1}$$

于是取

$$c_n = n + 1$$

则

$$\hat{\theta_2} = (n+1)X_{min}$$

是θ的无偏估计.

3.3

考虑X_{min}, X_{max}的联合累积分布函数

$$F(x_m, x_M) = P(X_{min} \le x_m \wedge X_{max} \le x_M)$$

$$= P(X_{max} \le x_M) - P(X_{min} > x_m \wedge X_{max} \le x_M)$$

$$= (\frac{x_M}{\theta})^n - (\frac{x_M - x_m}{\theta})^n$$

于是其联合PDF为

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

$0 \le x \le \theta$时

$$\hat{f}_1(x) = \int_0^{\frac{x}{2}} f(\xi, x-\xi) d\xi = \int_0^{\frac{x}{2}} \frac{n(n-1)(x-2\xi)^{n-2}}{\theta^n} d\xi$$
$$= -\frac{1}{2} \int_0^{\frac{x}{2}} \frac{n(n-1)(x-2\xi)^{n-2}}{\theta^n} d(x-2\xi)$$
$$= \frac{nx^{n-1}}{2\theta^n}$$

$\theta < x \le 2\theta$时

$$\hat{f}_1(x) = \int_{x-\theta}^{\frac{x}{2}} f(\xi, x-\xi) d\xi = \int_{x-\theta}^{\frac{x}{2}} \frac{n(n-1)(x-2\xi)^{n-2}}{\theta^n} d\xi$$
$$= \frac{n(2\theta - x)^{n-1}}{2\theta^n}$$

故

$$Var(\hat{\theta}_1) = E(\hat{\theta_1}^2) - \theta^2$$
$$= \int_0^{\theta} \frac{nx^{n+1}}{2\theta^n} dx + \int_{\theta}^{2\theta} \frac{n(2\theta - x)^{n-1} x^2}{2\theta^n} dx - \theta^2$$
$$= \frac{n\theta^2}{2(n+2)} + \int_0^{\theta} \frac{nx^{n-1}(2\theta - x)^2}{2\theta^n} dx - \theta^2$$
$$= \frac{n\theta^2}{n+2} - \frac{2n\theta^2}{n+1} + 2\theta^2 - \theta^2$$
$$= \frac{2\theta^2}{(n+1)(n+2)}$$

而

$$Var(\hat{\theta}_2) = (n+1)^2 Var(X_{min})$$
$$= (n+1)^2 E(X_{min}^2) - \theta^2$$
$$= (n+1)^2 \int_0^{\theta} \frac{n(\theta - x)^{n-1} x^2}{\theta^n} dx - \theta^2$$
$$= \frac{n\theta^2}{n+2}$$
$$Var(\hat{\theta}_3) = 4Var(\bar{X}) = \frac{\theta^2}{3n}$$
$$Var(\hat{\theta}_4) = \frac{(n+1)^2}{n^2} E(X_{max}^2) - \theta^2$$
$$= \frac{(n+1)^2}{n^2} \int_0^{\theta} \frac{nx^{n+1}}{\theta^n} dx - \theta^2$$
$$= \frac{\theta^2}{n(n+2)}$$

在$n \to \infty$的渐近条件下有

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

这说明，在估计中引入较大的次序统计量的权重越大，估计的方差就越小.

4 加权平均估计

4.1

$$E(\sum_{i=1}^{n} c_i X_i) = \sum_{i=1}^{n} c_i E(X_i) = \theta \sum_{i=1}^{n} c_i$$

故

$$E(\sum_{i=1}^{n} c_i X_i) = \theta \Leftrightarrow \sum_{i=1}^{n} c_i = 1$$

4.2

设总体方差为σ^2

$$E((\sum_{i=1}^{n} c_i X_i)^2) = (\theta^2 + \sigma^2) \sum_{i=1}^{n} c_i^2 + \theta^2 \sum_{i \neq j} c_i c_j$$

$$= \theta^2 (\sum_{i=1}^{n} c_i)^2 + \sigma^2 \sum_{i=1}^{n} c_i^2$$

$$= \theta^2 + \sigma^2 \sum_{i=1}^{n} c_i^2$$

$$\geq \theta^2 + \sigma^2 \frac{\sum_{i=1}^{n} c_i}{n}$$

$$= \theta^2 + \frac{1}{n} \sigma^2$$

上式取等当且仅当

$$c_1 = ... = c_n = \frac{1}{n}$$

5 正态估计的均方误差

由于

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

可知

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

而

$$Var(m_2) = \frac{(n-1)^2}{n^2} Var(S^2) = \frac{2(n-1)\sigma^4}{n^2}$$

$$E((m_2 - \sigma^2)^2) = Var(m_2 - \sigma^2) + E^2(m_2 - \sigma^2) = \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2}$$

而

$$E((S^2 - \sigma^2)^2) = Var(S^2) = \frac{2\sigma^4}{n-1} = \frac{(2n-1)\sigma^4}{n^2 - \frac{3}{2}n + \frac{1}{2}}$$

故

$$E((S^2 - \sigma^2)^2) > E((m_2 - \sigma^2)^2)$$

6 正态分布与t分布

分母

$$\frac{X_3^2+X_4^2}{4}\sim\chi^2(2)$$

分子

$$\frac{X_1+X_2}{2\sqrt{2}}\sim N(0,1)$$

故

$$\frac{\frac{X_1+X_2}{2\sqrt{2}}}{\sqrt{\frac{X_3^2+X_4^2}{8}}}=\frac{X_1+X_2}{\sqrt{X_3^2+X_4^2}}\sim t(2)$$

于是$a=1$，此时t分布的自由度为2.

7 正态分布估计实操1

由于

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

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

有

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

故μ的$1-\alpha$置信区间为

$$(\bar{X}-\frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1),\bar{X}+\frac{S}{\sqrt{n}}t_{\frac{\alpha}{2}}(n-1))$$

现有参数如下

$$\bar{X}=503.75$$

$$S=6.20215$$

$$n=16$$

$$\alpha=0.05$$

计算得区间为

$$(500.45,507.05)$$

8 正态分布估计实操2

$$\bar{X}=1160$$

$$S=99.74969$$

$$n=5$$

$$\alpha=0.05$$

需求出

$$\bar{X}-\frac{S}{\sqrt{n}}t_{\alpha}(n-1)$$

解得下限为1064.9.

9 正态分布估计实操3

9.1

由于$X \sim N(\mu_1, \sigma^2)$和$Y \sim N(\mu_2, \sigma^2)$独立，$X, Y$样本容量分别为$n_1, n_2$，则有

$$(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2) \sim N(0, \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2})$$

又

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \sim \chi^2(n_1 - 1)$$
$$\frac{(n_2 - 1)S_2^2}{\sigma^2} \sim \chi^2(n_2 - 1)$$

故

$$\frac{\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}}{\sqrt{\frac{\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2}}{n_1+n_2-2}}} \sim t(n_1 + n_2 - 2)$$

化简得

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}} \sim t(n_1 + n_2 - 2)$$

故$1 - \alpha$置信区间为

$$(\bar{X} - \bar{Y} - \sqrt{(\frac{1}{n_1} + \frac{1}{n_2})\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}t_{\frac{\alpha}{2}}(n_1 + n_2 - 2),$$

$$\bar{X} - \bar{Y} + \sqrt{(\frac{1}{n_1} + \frac{1}{n_2})\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}t_{\frac{\alpha}{2}}(n_1 + n_2 - 2))$$

已知参数

$$\bar{X} = 91.73$$
$$\bar{Y} = 93.75$$
$$S_1^2 = 3.89$$
$$S_2^2 = 4.02$$
$$n_1 = 20$$
$$n_2 = 30$$
$$\alpha = 0.05$$

解得

$$(-3.18, -0.86)$$

9.2

两个催化剂有显著差别. 上述计算得知，在95%的置信水平下都有$\mu_1 < \mu_2$，可见有较高的置信水平保证新型催化剂的期望产率高于原催化剂.

10 均匀分布区间估计

$$F_n(x) = \frac{x^n}{\theta^n}$$

要使

$$P(X_{max} < \theta < c_n X_{max}) = 1 - \alpha$$

只需

$$P(X_{max} > \frac{\theta}{c_n}) = 1 - \alpha$$

即

$$1 - F_n(\frac{\theta}{c_n}) = 1 - \alpha$$

故

$$c_n^n = \alpha$$

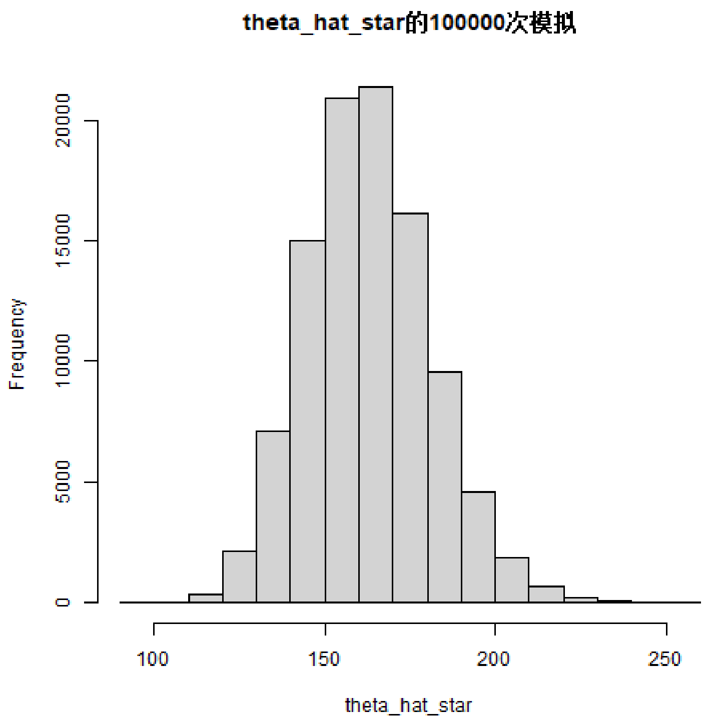
只需

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

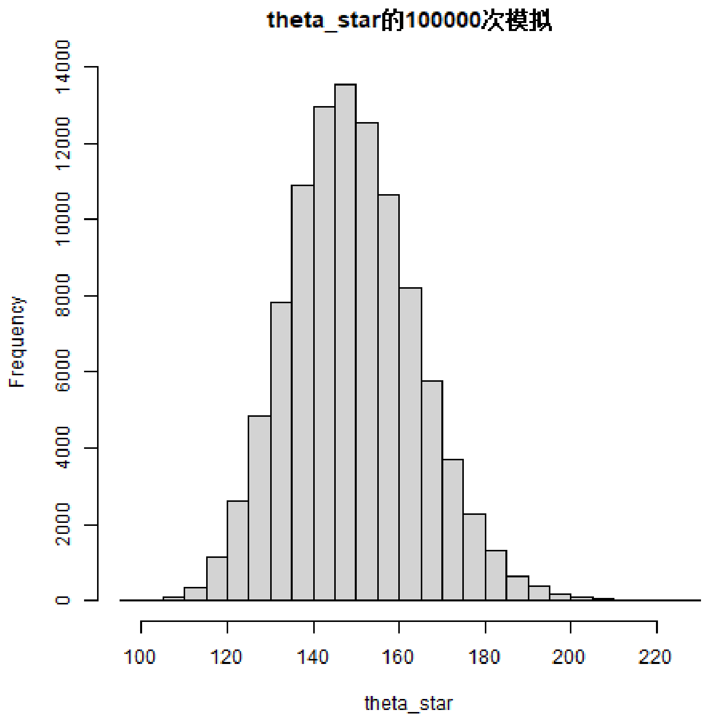
即可.

11 计算机实验：自助法Bootstrap

我进行了100000次模拟，这是$\hat{\theta}^*$的分布：



这是$\hat{\theta}$的分布：



可以看到二者的分布形状上比较接近，但方差有一定差异.
接下来，我做了10次模拟并比较二者的方差（实际上是样本方差），如下：

```
[1] "第1次模拟，Vboot=240.521705，Var=222.132836"
[1] "第2次模拟，Vboot=222.984021，Var=223.007122"
[1] "第3次模拟，Vboot=232.957451，Var=223.450371"
[1] "第4次模拟，Vboot=303.018521，Var=222.906810"
[1] "第5次模拟，Vboot=290.774069，Var=222.934047"
[1] "第6次模拟，Vboot=180.703187，Var=223.359109"
[1] "第7次模拟，Vboot=206.476294，Var=222.653401"
[1] "第8次模拟，Vboot=224.740107，Var=225.476957"
[1] "第9次模拟，Vboot=166.406862，Var=224.667907"
[1] "第10次模拟，Vboot=279.831986，Var=223.626337"
```

可见V_{boot}具有对$Var(\hat{\theta})$的估计作用，但由于V_{boot}实际上只基于一组给定的样本，它的估计会产生较大的偏差.