统计与数据科学学院

《大数据的统计学基础》第二次作业

姓　　名：蒋贵豪

年　　级：2021 级

专　　业：应用统计学

学　　号：B+X9bo

完成日期：2021 年 11 月 21 日

目录

1 点估计

题目 1. 通过计算机模拟研究：对于非高斯分布的样本，总体方差的不同估计量的抽样分布。

解答. 我们通过编写 **R** 代码，模拟了服从 $U[0,1]$ 的**均匀分布**的样本 (图1.1) 和服从**指数分布** $exp(1)$ 的样本 (图1.2) 总体方差的不同估计量的抽样分布。其中：

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1.1}$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1.2}$$

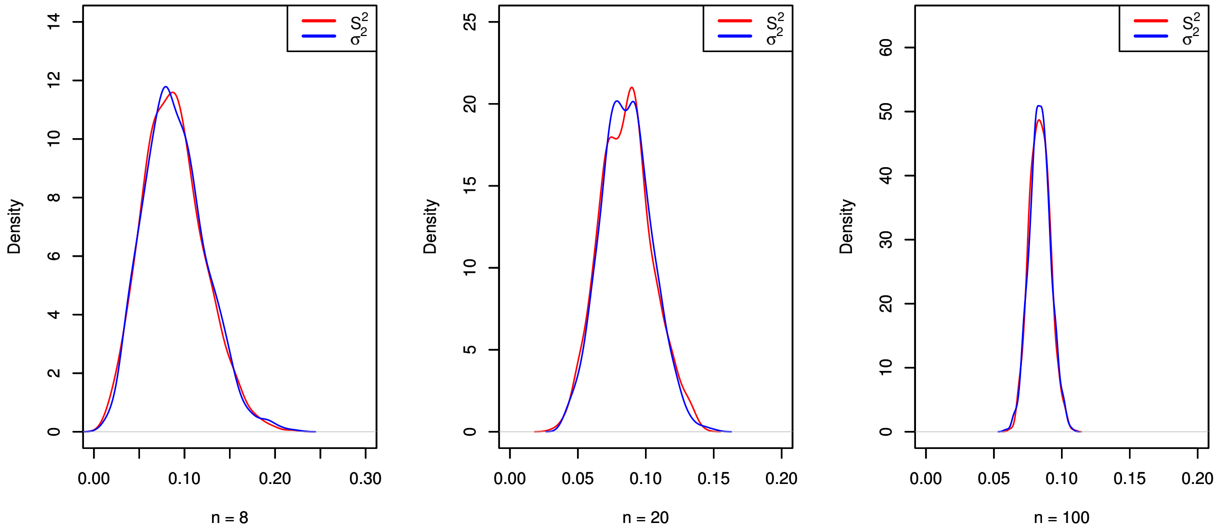

图 1.1: 抽样分布

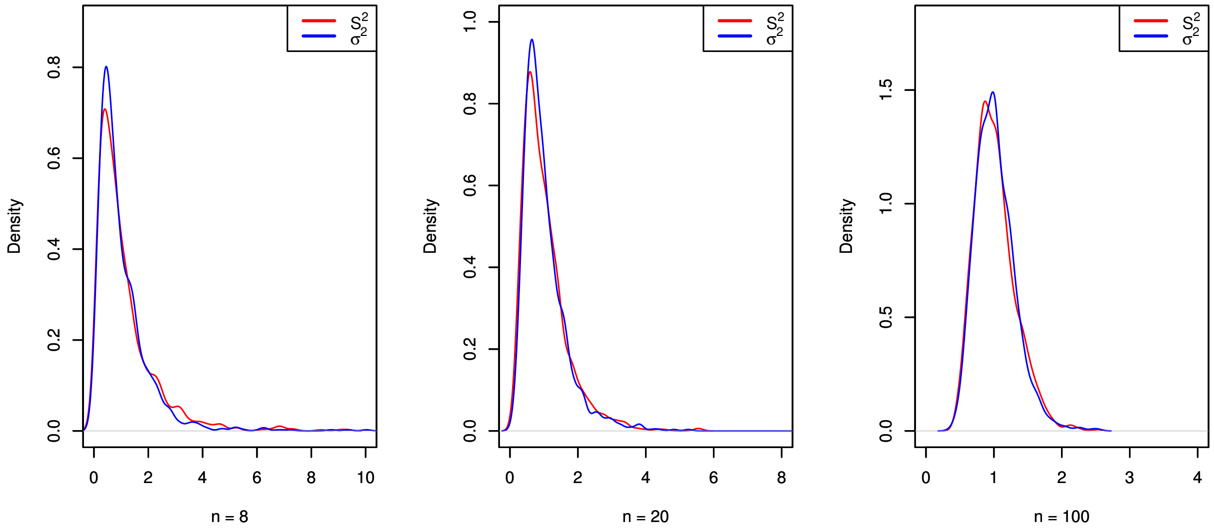

图 1.2: 抽样分布

我们知道，S^2 是总体方差 σ^2 的**无偏估计**，$\hat{\sigma}^2$ 是总体方差 σ^2 的**有偏估计**。但是，在一般情况下，**我们无法找到所谓的最好的估计量**。在渐进的意义下，S^2 和 $\hat{\sigma}^2$ 具有相同的估计效果，**并不是无偏估计就比有偏估计更好**。

```
simulate <- function(n){
  library(latex2exp)
  s_2 = vector()
  sigma_2 = vector()

  for (i in 1:1000){
    x <- rexp(n,1)
    y <- rexp(n,1)
#     x <- runif(n, min = 0, max = 1)
#     y <- runif(n, min = 0, max = 1)
    S_2 <- var(x)
    Sigma_2 <- (n-1)/n*var(y)
    s_2[i] <- S_2
    sigma_2[i] <- Sigma_2
```

```r
  }
  A = list(s_2, sigma_2)
  return(A)
}
par(mfrow = c(1,3))
A = simulate(8)
B = matrix(unlist(A),nrow=2)
plot(density(B[1,]),main = "", xlab = TeX("n = 8"), xlim = c(0,10),
     ylim = c(0,0.9), col = 'red')
#plot(density(B[1,]),main = "", xlab = TeX("n = 8"), xlim = c(0,0.3),
#     ylim = c(0,14), col = 'red')
par(new=TRUE)
plot(density(B[2,]),main = "", xlab = TeX("n = 8"), xlim = c(0,10),
     ylim = c(0,0.9), col = 'blue')
#plot(density(B[2,]),main = "", xlab = TeX("n = 8"), xlim = c(0,0.3),
#     ylim = c(0,14), col = 'blue')
legend("topright", col = c("red", "blue"),
       lty = c(1,1), lwd = c(2,2),
       legend = c(TeX("S^2"), expression(sigma^2)))

A = simulate(20)
B = matrix(unlist(A),nrow=2)
plot(density(B[1,]),main = "", xlab = TeX("n = 20"), xlim = c(0,8),
     ylim = c(0,1), col = 'red')
#plot(density(B[1,]),main = "", xlab = TeX("n = 20"), xlim = c(0,0.2),
#     ylim = c(0,25), col = 'red')
par(new=TRUE)
plot(density(B[2,]),main = "", xlab = TeX("n = 20"), xlim = c(0,8),
     ylim = c(0,1), col = 'blue')
#plot(density(B[2,]),main = "", xlab = TeX("n = 20"), xlim = c(0,0.2),
```

```r
#      ylim = c(0,25), col = 'blue')
legend("topright", col = c("red", "blue"),
          lty = c(1,1), lwd = c(2,2),
          legend = c(TeX("S^2"), expression(sigma^2)))
A = simulate(100)
B = matrix(unlist(A),nrow=2)
plot(density(B[1,]),main = "", xlab = TeX("n = 100"), xlim = c(0,4),
       ylim = c(0,1.8), col = 'red')
#plot(density(B[1,]),main = "", xlab = TeX("n = 100"), xlim = c(0,0.2),
#      ylim = c(0,64), col = 'red')
par(new=TRUE)
plot(density(B[2,]),main = "", xlab = TeX("n = 100"), xlim = c(0,4),
       ylim = c(0,1.8), col = 'blue')
#plot(density(B[2,]),main = "", xlab = TeX("n = 100"), xlim = c(0,0.2),
#      ylim = c(0,64), col = 'blue')
legend("topright", col = c("red", "blue"),
          lty = c(1,1), lwd = c(2,2),
          legend = c(TeX("S^2"), expression(sigma^2)))

```

题目 2. 设 $X_1, ..., X_n$ 是 $i.i.d$ 的，具有概率密度函数：

$$f(x|\theta) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta, \quad \theta > 0 \tag{1.3}$$

用矩法和极大似然法估计 θ。计算两种估计量的均值和方差。哪一个应该被优先选用，为什么？

解答. 我们知道，该分布为 $[0, \theta]$ 上的均匀分布。对于**矩估计**：

$$\mu = \mathrm{E}(X) = \frac{\theta}{2} \tag{1.4}$$

于是，我们有 $\theta = 2\mu$, 从而 θ 的**矩估计**为：$\widehat{\theta} = 2\bar{X}$。

矩估计的**均值**和**方差**计算如式1.5和式1.6所示：

$$\mathrm{E}(\widehat{\theta}) = \mathrm{E}(2\bar{X}) = 2\mathrm{E}(\bar{X}) = 2 \times \frac{\theta}{2} = \theta \tag{1.5}$$

$$\mathrm{Var}(\widehat{\theta}) = \mathrm{Var}(2\bar{X}) = 4\mathrm{Var}(\bar{X}) = 4 \times \frac{(\theta-0)^2/12}{n} = \frac{\theta^2}{3n} \tag{1.6}$$

对于**极大似然估计**：

我们先计算似然函数：

$$L(\theta) = \begin{cases} (\dfrac{1}{\theta})^n & 0 < x_{(n)} \leq \theta \\ 0 & \text{其他} \end{cases} \tag{1.7}$$

为了使 $L(\theta)$ 取到最大，我们知道**极大似然估计**为 $\widehat{\theta}_{MLE} = X_{(n)}$。

我们知道，$X_{(n)}$ 的概率密度函数为：

$$p(x) = nf(x)[F(x)]^{n-1} = \begin{cases} \dfrac{nx^{n-1}}{\theta^n} & 0 \leq x \leq \theta \\ 0 & \text{其他} \end{cases} \tag{1.8}$$

于是，**极大似然估计**的**均值**如式1.9：

$$\mathrm{E}(\widehat{\theta}_{MLE}) = \int_0^\theta \frac{nx^{n-1}}{\theta^n} x \mathrm{d}x = \frac{n}{n+1}\theta \tag{1.9}$$

下面计算极大似然估计的方差，先计算 $\mathrm{E}(\widehat{\theta}^2_{MLE})$：

$$\mathrm{E}(\widehat{\theta}^2_{MLE}) = \int_0^\theta \frac{nx^{n-1}}{\theta^n} x^2 \mathrm{d}x = \frac{n}{n+2}\theta^2 \tag{1.10}$$

极大似然估计的**方差**如式1.11：

$$\mathrm{Var}(\widehat{\theta}_{MLE}) = \mathrm{E}(\widehat{\theta}^2_{MLE}) - \mathrm{E}^2(\widehat{\theta}_{MLE}) = \frac{n}{n+2}\theta^2 - (\frac{n}{n+1}\theta)^2 = \frac{n\theta^2}{(n+1)^2(n+2)} \tag{1.11}$$

从式1.5和式1.9可以看出，矩估计 $\widehat{\theta}$ 为无偏估计，而极大似然估计 $\widehat{\theta}_{MLE}$ 为有偏估计。但是，从式1.6和式1.11可以看出，矩估计的方差 $\mathrm{Var}(\widehat{\theta})$ 大于最大似然估计的方差 $\mathrm{Var}(\widehat{\theta}_{MLE})$。因此，**当样本数 n 较大时，我们选用极大似然估计，因为此时虽然存在一定的偏差，但是偏差比例随着 n 的增大会接近于 1，并且方差较小。而当样本数较小时，偏差比例较大，我们可以选用矩估计。**

题目 3. 设独立的随机变量 $X_1, ..., X_n$ 具有共同的分布：

$$P\left(X_i \leqslant x \mid \alpha, \beta\right)=\begin{cases}0 & \text{当} x<0 \\ (x/\beta)^{\alpha} & \text{当} 0 \leqslant x \leqslant \beta \\ 1 & \text{当} x>\beta\end{cases} \tag{1.12}$$

其中参数 α, β 为正。

(a) 求一个关于 (α, β) 的二维充分统计量。

(b) 求 α, β 的极大似然估计。

(c) 在篱雀的巢中找到杜鹃蛋的长度 (单位：mm) 可以用这个分布建模。根据数据 22.0, 23.9, 20.9, 23.8, 25.0, 24.0, 21.7, 23.8, 22.8, 23.1, 23.1, 23.5, 23.0, 23.0, 求 α, β 的 MLE。

解答. (a) 由题意知，该随机变量满足的概率密度函数为：

$$f(x)=\begin{cases}\dfrac{\alpha x^{\alpha-1}}{\beta^{\alpha}} & 0 \leq x \leq \beta \\ 0 & \text{其他}\end{cases} \tag{1.13}$$

于是，我们可以得到似然函数：

$$L(\alpha, \beta)=\prod_{i=1}^{n} \frac{\alpha {x_i}^{\alpha-1}}{\beta^{\alpha}} \mathbb{1}_{[0, \beta]}(x_i)=\frac{\alpha^{n}}{\beta^{n \alpha}} \mathbb{1}_{[0, \beta]}(x_{(n)}) \prod_{i=1}^{n} {x_i}^{\alpha-1} \tag{1.14}$$

由因子分解定理，我们知道**一个关于 (α, β) 的二维充分统计量为** $(\prod\limits_{i=1}^{n} X_i, X_{(n)})$。

(b) 首先，我们观察式1.14的形式，为了使 $L(\alpha, \beta)$ 取到最大，应选取 $\beta = x_{(n)}$，于是，β **的极大似然估计为** $\hat{\beta} = X_{(n)}$。

然后，我们对式1.14先取对数，然后对 α 求偏导：

$$\frac{\partial}{\partial \alpha} \log L(\alpha, \beta)=\frac{\partial}{\partial \alpha}\left[n \log \alpha-n \alpha \log \beta+(\alpha-1) \log \prod_{i=1}^{n} x_i\right]=\frac{n}{\alpha}-n \log \beta+\log \prod_{i=1}^{n} x_i \tag{1.15}$$

然后我们令式1.15为 0，可以得到 α **的极大似然估计** $\hat{\alpha}$ **如式1.16**:

$$\hat{\alpha}=\frac{n}{n \log X_{(n)}-\log \prod\limits_{i=1}^{n} X_i} \tag{1.16}$$

我们可以对式1.15再对 α 求一次偏导，并将 $\hat{\alpha}$ 代入，得到的值小于 0，因此 $\hat{\alpha}$ 为 α 的极大似然估计。

(c) 我们通过如下简单的 **Matlab** 代码实现 α,β 的 MLE 计算：

```
x = [22.0, 23.9, 20.9, 23.8, 25.0, 24.0, 21.7, 23.8, 22.8, 23.1, ...
        23.1, 23.5, 23.0, 23.0]
beta = max(x)
% beta = 25
alpha = length(x)/(length(x)*log(beta)-log(prod(x)))
% alpha = 12.5949
```

于是 α,β **的 MLE 为**：$(\hat{\alpha},\hat{\beta}) = (12.5949, 25)$。

题目 4. 设 $X_1,...,X_n$ 是来自一正态总体 $N(\theta,\sigma^2)$ 的一组随机样本，并且假定 θ 的先验分布是 $N(\mu,\tau^2)$。此处我们假定 σ^2,μ,τ^2 都已知。

(a) 求 $\bar{X}$ 和 θ 的联合概率密度函数。

(b) 证明 $\bar{X}$ 的边缘分布 $m(\bar{x}|\sigma^2,\mu,\tau^2)$ 是 $N(\mu,(\sigma^2/n)+\tau^2)$。

(c) 证明 θ 的后验分布 $\pi(\theta,|\bar{x},\sigma^2,\mu,\tau^2)$ 是正态分布，均值与方差由式1.17给出。

$$\begin{aligned} \mathrm{E}(\theta|\bar{x}) &= \frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x} + \frac{\sigma^2}{\sigma^2+n\tau^2}\mu \\ \mathrm{Var}(\theta|\bar{x}) &= \frac{\sigma^2\tau^2}{\sigma^2+n\tau^2} \end{aligned} \tag{1.17}$$

解答. (a) $\bar{X}$ 和 θ 的**联合概率密度函数如式1.18所示**:

$$f(\bar{x},\theta) = f(\bar{x}|\theta)\pi(\theta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}e^{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}} \times \frac{1}{\sqrt{2\pi}\tau}e^{-\frac{(\theta-\mu)^2}{2\tau^2}} = \frac{\sqrt{n}}{2\pi\sigma\tau}e^{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}-\frac{(\theta-\mu)^2}{2\tau^2}} \tag{1.18}$$

(b) $f(\bar{x},\theta)$ 的核函数为：

$$e^{-\dfrac{n(\bar{x}-\theta)^2}{2\sigma^2} - \dfrac{(\theta-\mu)^2}{2\tau^2}} \tag{1.19}$$

式1.19中 e 的指数部分为：

$$
\begin{aligned}
&-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}-\frac{(\theta-\mu)^2}{2\tau^2}=-\frac{(\theta-\dfrac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2})^2}{\dfrac{2\sigma^2\tau^2}{n\tau^2+\sigma^2}}-\frac{(\bar{x}-\mu)}{2\times\dfrac{n\tau^2+\sigma^2}{n}}\\
&-\frac{n(\bar{x}-\theta)^2}{2\sigma^2}-\frac{(\theta-\mu)^2}{2\tau^2}=-\frac{n\tau^2(\bar{x}-\theta)^2+\sigma^2(\theta-\mu)^2}{2\sigma^2\tau^2}\\
&=\frac{-(n\tau^2+\sigma^2)\theta^2+(2n\tau^2\bar{x}+2\sigma^2\mu)\theta-\sigma^2\mu^2-n\tau^2\bar{x}^2}{2\sigma^2\tau^2}\\
&=\frac{-\theta^2+\dfrac{2n\tau^2\bar{x}+2\sigma^2\mu}{n\tau^2+\sigma^2}\theta-(\dfrac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2})^2+(\dfrac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2})^2-\dfrac{n\tau^2\bar{x}^2+\sigma^2\mu^2}{n\tau^2+\sigma^2}}{\dfrac{2\sigma^2\tau^2}{n\tau^2+\sigma^2}}\\
&=-\frac{(\theta-\dfrac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2})^2}{\dfrac{2\sigma^2\tau^2}{n\tau^2+\sigma^2}}+\frac{\dfrac{(n\tau^2\bar{x}+\sigma^2\mu)^2}{n\tau^2+\sigma^2}-(n\tau^2\bar{x}^2+\sigma^2\mu^2)}{2\sigma^2\tau^2}\\
&=-\frac{(\theta-\dfrac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2})^2}{\dfrac{2\sigma^2\tau^2}{n\tau^2+\sigma^2}}+\frac{\dfrac{(n\tau^2)^2\bar{x}^2+(\sigma^2\mu)^2+2n\tau^2\sigma^2\mu\bar{x}}{n\tau^2+\sigma^2}-(n\tau^2\bar{x}^2+\sigma^2\mu^2)}{2\sigma^2\tau^2}\\
&=-\frac{(\theta-\dfrac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2})^2}{\dfrac{2\sigma^2\tau^2}{n\tau^2+\sigma^2}}+\frac{(n\tau^2)^2\bar{x}^2+(\sigma^2\mu)^2+2n\tau^2\sigma^2\mu\bar{x}-(n\tau^2\bar{x}^2+\sigma^2\mu^2)(n\tau^2+\sigma^2)}{2\sigma^2\tau^2(n\tau^2+\sigma^2)}\\
&=-\frac{(\theta-\dfrac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2})^2}{\dfrac{2\sigma^2\tau^2}{n\tau^2+\sigma^2}}+\frac{(n\tau^2)^2\bar{x}^2+(\sigma^2\mu)^2+2n\tau^2\sigma^2\mu\bar{x}-(n\tau^2)^2\bar{x}^2-n\tau^2\sigma^2\bar{x}^2-n\tau^2\sigma^2\mu^2-(\sigma^2\mu)^2}{2\sigma^2\tau^2(n\tau^2+\sigma^2)}\\
&=-\frac{(\theta-\dfrac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2})^2}{\dfrac{2\sigma^2\tau^2}{n\tau^2+\sigma^2}}+\frac{2n\tau^2\sigma^2\mu\bar{x}-n\tau^2\sigma^2\bar{x}^2-n\tau^2\sigma^2\mu^2}{2\sigma^2\tau^2(n\tau^2+\sigma^2)}\\
&=-\frac{(\theta-\dfrac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2})^2}{\dfrac{2\sigma^2\tau^2}{n\tau^2+\sigma^2}}-\frac{(\bar{x}-\mu)^2}{2\times\dfrac{n\tau^2+\sigma^2}{n}}
\end{aligned}
\tag{1.20}
$$

于是，原来的两个相乘的正态分布函数可以重新写为两个正态分布相乘的形式：

$$
N(\theta,\frac{\sigma^2}{n})\times N(\mu,\tau^2)=N(\frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2},\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2})\times N(\mu,\frac{n\tau^2+\sigma^2}{n})\tag{1.21}
$$

从式1.21可得，$\bar{X}$ 的边缘分布 $m(\bar{x}|\sigma^2,\mu,\tau^2)$ 为 $N(\mu,\dfrac{n\tau^2+\sigma^2}{n})$。

(c) 从式1.20可得,θ 的后验分布 $\pi(\theta,|\bar{x},\sigma^2,\mu,\tau^2)$ 是正态分布,为 $N(\frac{n\tau^2\bar{x}+\sigma^2\mu}{n\tau^2+\sigma^2},\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2})$。

题目 5. 对于分布 $f(x|\theta)=\theta x^{\theta-1}$, $0<x<1$, $\theta>0$。设 $X_1,...,X_n$ 是一组随机样本。是否存在 θ 的一个函数，记为 $g(\theta)$, 对它存在一个方差达到 Cramer-Rao 下界的无偏估计量？若存在，则求出它。若不存在，说明为什么？

解答. (1) **先计算 θ 的 MLE**, θ 的似然函数为：

$$L(\theta)=\theta^n\prod_{i=1}^{n}x_i^{\theta-1} \tag{1.22}$$

对其取对数，对数似然函数为：

$$\log L(\theta)=n\log\theta+(\theta-1)\log(\prod_{i=1}^{n}x_i) \tag{1.23}$$

对 θ 求偏导：

$$\frac{\partial\log L(\theta)}{\partial\theta}=\frac{n}{\theta}+\log(\prod_{i=1}^{n}x_i) \tag{1.24}$$

令式1.24等于 0，可以得到 θ **的 MLE**:

$$\hat{\theta}_{MLE}=-\frac{n}{\log(\prod_{i=1}^{n}x_i)} \tag{1.25}$$

观察式1.25，我们取：

$$g(\theta)=\frac{1}{\theta} \tag{1.26}$$

我们知道，$g(\theta)$ **的极大似然估计为：**

$$\hat{g}(\mathbf{x})=-\frac{\log(\prod_{i=1}^{n}x_i)}{n} \tag{1.27}$$

(2) **验证 $\hat{g}(\mathbf{x})$ 为 $g(\theta)$ 的无偏估计：**

由 $f(x|\theta)=\theta x^{\theta-1}$，设 $y=-\log(x)$, 于是 $x=e^{-y}$, 那么我们有：

$$p(y|\theta)=\theta(e^{-y})^{\theta-1}e^{-y}=\theta e^{-\theta y} \tag{1.28}$$

从而，我们从式1.28中可知：$Y \sim \exp(\theta)$, 于是我们计算 $\hat{g}(\mathbf{x})$ 的期望：

$$\mathrm{E}(\hat{g}(\mathbf{x})) = \frac{1}{n}\mathrm{E}(nY) = \mathrm{E}(Y) = \frac{1}{\theta} \tag{1.29}$$

从式1.29可知，验证 $\hat{g}(\mathbf{x})$ 为 $g(\theta)$ 的无偏估计。

(3) **计算该分布的费希尔信息量** $I(\theta)$:

由 $f(x|\theta) = \theta x^{\theta-1}$，于是先对其取对数：

$$\log(f(x|\theta)) = \log(\theta) + (\theta - 1)\log(x) \tag{1.30}$$

然后对 θ 求偏导：

$$\frac{\partial \log(f(x|\theta))}{\partial \theta} = \frac{1}{\theta} + \log(x) \tag{1.31}$$

对 θ 求二阶偏导：

$$\frac{\partial^2 \log(f(x|\theta))}{\partial \theta^2} = \frac{-1}{\theta^2} \tag{1.32}$$

最后求得该分布的费希尔信息量 $I(\theta)$:

$$I(\theta) = -E(\frac{\partial^2 \log(f(x|\theta))}{\partial \theta^2}) = -E(\frac{-1}{\theta^2}) = \frac{1}{\theta^2} \tag{1.33}$$

(4) **考察** $\hat{g}(\mathbf{x})$ **的方差是否达到 C-R 下界：**

计算：

$$(g'(\theta))^2 = (\frac{-1}{\theta^2})^2 = \frac{1}{\theta^4} \tag{1.34}$$

于是 **C-R 下界为：**

$$\frac{(g'(\theta))^2}{nI(\theta)} = \frac{\frac{1}{\theta^4}}{n\frac{1}{\theta^2}} = \frac{1}{n\theta^2} \tag{1.35}$$

而我们的 $\hat{g}(\mathbf{x})$ 的方差为：

$$Var(\hat{g}) = \frac{1}{n^2} \times n \times Var(Y) = \frac{1}{n\theta^2} = \frac{(g'(\theta))^2}{nI(\theta)} \tag{1.36}$$

于是，从式1.36可知，$\hat{g}(\mathbf{x})$ **的方差达到了 C-R 下界**。综上所述，对于分布 $f(x|\theta) = \theta x^{\theta-1}$, $0 < x < 1$, $\theta > 0$, 存在 θ 的函数：$g(\theta) = 1/\theta$。它存在一个方差达到 Cramer-Rao 下界的无偏估计量为式1.27。

题目 6. 设 $X_1,...,X_n$ 是来自一个均值为 μ, 方差为 σ^2 的总体的一组随机样本。

(a) 证明：如果 $\sum_{i=1}^{n} a_i = 1$, 则估计量 $\sum_{i=1}^{n} a_i X_i$ 是 μ 的一个无偏估计量。

(b) 在所有的这种形式的估计量 (称为线性无偏估计量) 中求一个具有最小方差者，并计算其方差。

解答. (a) 我们计算 $\sum_{i=1}^{n} a_i X_i$ 的期望：

$$\mathrm{E}(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i \mathrm{E}(X_i) = \sum_{i=1}^{n} a_i \mu = \mu \sum_{i=1}^{n} a_i = \mu \tag{1.37}$$

由式1.37可得 **估计量 $\sum_{i=1}^{n} a_i X_i$ 是 μ 的一个无偏估计量**。

(b) 我们计算 $\sum_{i=1}^{n} a_i X_i$ 的方差：

$$\mathrm{Var}(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) = \sum_{i=1}^{n} a_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^{n} a_i^2 \tag{1.38}$$

对于1.38中的 $\sum_{i=1}^{n} a_i^2$ 由柯西不等式，我们有：

$$\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} 1 \geq (\sum_{i=1}^{n} a_i)^2 = 1 \tag{1.39}$$

由柯西不等式的取等条件，我们知道，**取到最小方差时**：

$$a_i = \frac{1}{n}, \quad i = 1, 2, ..., n \tag{1.40}$$

最小方差为：

$$\min Var(\sum_{i=1}^{n} a_i X_i) = \frac{\sigma^2}{n} \tag{1.41}$$

题目 7. 刀切法是减少估计量偏倚的一种普遍技术。一个一步刀切法估计量是如下定义的。设 $X_1,...,X_n$ 是一组随机样本，又设 $T_n = T_n(X_1,...,X_n)$ 是参数 θ 的某个估计量。为了刀切 T_n, 我们计算 n 个统计量 $T_n^{(i)}, i = 1,...,n$, 其中 $T_n^{(i)}$ 是如同 T_n 那样计算

不过却是利用样本中移去 X_i 之后的 $n-1$ 个观测值计算出来的。θ 的刀切估计量记作 $\mathrm{JK}(T_n)$, 由式1.42给出：

$$\mathrm{JK}(T_n)=nT_n-\frac{n-1}{n}\sum_{i=1}^{n}T_n^{(i)} \tag{1.42}$$

现在特别地，设 $X_1,...,X_n$ 为 $i.i.d$ 的 Bernoulli(θ)。目的是估计 θ^2。

(a) 证明：θ^2 的 MLE, 即 $(\frac{\sum\limits_{i=1}^{n}X_i}{n})^2$ 是 θ^2 的一个有偏估计量。

(b) 由 MLE 导出一步刀切估计量。

(c) 证明：这个一步刀切估计量是 θ^2 的一个无偏估计量。

解答. (a) 我们知道，θ 的 MLE 为 $\hat{\theta}_{MLE}=\bar{X}$, 于是 θ^2 的 MLE 为 $\bar{X}^2$, 下面证明其是有偏的：

$$\mathrm{E}(\bar{X}^2)=\mathrm{E}^2(\bar{X})+\mathrm{Var}(\bar{X})=\theta^2+\frac{\theta(1-\theta)}{n}=\frac{n-1}{n}\theta^2+\frac{\theta}{n}\neq\theta^2 \tag{1.43}$$

因此，$(\frac{\sum\limits_{i=1}^{n}X_i}{n})^2$ **是 θ^2 的一个有偏估计量**。

(b) 由 (a)，我们知道，$T_n^{(i)}$ 的 MLE 为：$(\frac{\sum\limits_{i\neq j}X_j}{n-1})^2$, 于是：

$$\begin{aligned}
&n(\frac{\sum\limits_{i=1}^{n}X_i}{n-1})^2-\frac{n-1}{n}\sum_{i=1}^{n}(\frac{\sum\limits_{i\neq j}X_j}{n-1})^2=\frac{(\sum\limits_{i=1}^{n}X_i)^2}{n}-\frac{1}{n(n-1)}\sum_{i=1}^{n}\Big(\sum_{i\neq j}X_j\Big)^2\\
&=\frac{(\sum\limits_{i=1}^{n}X_i)^2}{n}-\frac{1}{n(n-1)}(\sum_{i=1}^{n}(n-1)X_i^2+2\sum_{i\neq j}(n-2)X_iX_j)\\
&=\frac{(\sum\limits_{i=1}^{n}X_i)^2}{n}-\frac{(n-2)(\sum\limits_{i=1}^{n}X_i)^2}{n(n-1)}-\frac{\sum\limits_{i=1}^{n}X_i^2}{n(n-1)}\\
&=\frac{(\sum\limits_{i=1}^{n}X_i)^2-\sum\limits_{i=1}^{n}X_i}{n(n-1)}
\end{aligned} \tag{1.44}$$

由式1.44可知，**一步刀切估计量为：**$\frac{(\sum\limits_{i=1}^{n}X_i)^2-\sum\limits_{i=1}^{n}X_i}{n(n-1)}$。

(c) 我们计算 (b) 中估计量的期望：

$$
\begin{aligned}
&\mathrm{E}(\frac{(\sum_{i=1}^{n} X_i)^2-\sum_{i=1}^{n} X_i}{n(n-1)})=\frac{\mathrm{Var}(\sum_{i=1}^{n} X_i)+\mathrm{E}^2(\sum_{i=1}^{n} X_i)-\mathrm{E}(\sum_{i=1}^{n} X_i)}{n(n-1)} \\
&=\frac{n\theta(1-\theta)+n^2\theta^2-n\theta}{n(n-1)}=\frac{n^2\theta^2-n\theta^2}{n(n-1)}=\theta^2
\end{aligned}
\tag{1.45}
$$

由式1.45可知，**这个一步刀切估计量是 θ^2 的一个无偏估计量。**

2 区间估计

题目 8. 独立随机变量 $X_1,...,X_n$ 具有共同的分布

$$
P\left(X_i\leqslant x\right)=\begin{cases}0 & \text{当} x<0 \\ (x/\beta)^{\alpha} & \text{当} 0\leqslant x\leqslant\beta \\ 1 & \text{当} x>\beta\end{cases}
\tag{2.1}
$$

(a) 在**题目 3** 里求出了 α,β 的 MLE。如果 α 是一个已知常数 α_0, 求 β 的一个置信系数为 0.95 的置信上界。

(b) 用**题目 3** 里的数据构造一个关于 β 的区间估计，假定 α 已知并且等于它的 MLE。

解答. (a) 我们由**题目 3** 知道 β 的 MLE 为 $\hat{\beta}=X_{(n)}$, 于是我们来构造 β 的置信区间。首先，$X_{(n)}$ 的累计分布函数为：

$$
F(x)=\begin{cases}0 & \text{当} x<0 \\ ((x/\beta)^{\alpha_0})^n & \text{当} 0\leqslant x\leqslant\beta \\ 1 & \text{当} x>\beta\end{cases}
\tag{2.2}
$$

于是，为了构造 β 的 $1-\alpha$ 置信区间, 我们令：

$$
\alpha<(\frac{x}{\beta})^{n\alpha_0}
\tag{2.3}
$$

可以得到：

$$
0<\beta<\frac{x}{\alpha^{\frac{1}{n\alpha_0}}}
\tag{2.4}
$$

从式2.4可知，β 的 $1-\alpha$ 置信区间为 $(0, \frac{X_{(n)}}{\alpha^{\frac{1}{n\alpha_0}}})$, 于是 β **的一个置信系数为 0.95 的置信上界为：** $\frac{X_{(n)}}{0.05^{\frac{1}{n\alpha_0}}}$。

(b) 在 (a) 中，我们已经得到了 β 的 $1-\alpha$ 置信区间, 我们代入**题目 3** 的数据，即 $(\hat{\alpha}, \hat{\beta}) = (12.5949, 25)$, $n = 14$, $X_{(n)} = 25$, 可知 $\frac{X_{(n)}}{0.05^{\frac{1}{n\alpha_0}}} = 25.4285$。

因此，**题目 3** 里的数据关于 β 的 95% 区间估计为 $(0, 25.4285)$。

题目 9. 证明表2.1中的三个量都是枢轴量。

表 2.1: 位置-尺度枢轴

pdf 形式	枢轴量
$f(x-\mu)$	$\bar{X}-\mu$
$\frac{1}{\sigma}f(\frac{x}{\sigma})$	$\frac{\bar{X}}{\sigma}$
$\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$	$\frac{\bar{X}-\mu}{S}$

解答. (a) 若:

$$X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} f(x-u) \tag{2.5}$$

我们令：

$$Y_i = X_i - \mu \tag{2.6}$$

则我们有：

$$Y_1, Y_2, \ldots, Y_n \overset{i.i.d}{\sim} f(y) \tag{2.7}$$

于是，从式2.7中我们得到：$\bar{Y} = \bar{X} - \mu$ **的分布于** μ **无关。故** $\bar{X} - \mu$ **为枢轴量**。

(b) 若：

$$X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} \frac{1}{\sigma}f(\frac{x}{\sigma}) \tag{2.8}$$

我们令：

$$Y_i = \frac{X_i}{\sigma} \quad 即 \quad X_i = \sigma Y_i \tag{2.9}$$

则我们有：

$$Y_1, Y_2, \ldots, Y_n \overset{i.i.d}{\sim} f(y) \times \frac{1}{\sigma} \times \sigma = f(y) \tag{2.10}$$

于是，从式2.10中我们得到：$\bar{Y}=\dfrac{\bar{X}}{\sigma}$ **的分布于** σ **无关。故** $\dfrac{\bar{X}}{\sigma}$ **为枢轴量。**

(c) 若：

$$X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} \frac{1}{\sigma} f(\frac{x-\mu}{\sigma}) \tag{2.11}$$

我们令：

$$Y_i=\frac{X_i-\mu}{\sigma} \quad \text{即} \quad X_i=\sigma Y_i+\mu \tag{2.12}$$

则我们有：

$$Y_1, Y_2, \ldots, Y_n \overset{i.i.d}{\sim} f(y) \times \frac{1}{\sigma} \times \sigma=f(y) \tag{2.13}$$

于是，从式2.13中我们得到：$\dfrac{\bar{X}-\mu}{S_X}=\dfrac{\sigma \bar{Y}}{\sqrt{\sigma^2} S_Y}=\dfrac{\bar{Y}}{S_Y}$ **的分布于** μ, σ **无关。故** $\dfrac{\bar{X}-\mu}{S}$ **为枢轴量。**

3 渐进评价

题目 10. 建立式3.1中的两个收敛结果：

$$\begin{aligned} -\frac{1}{\sqrt{n}} l^{\prime}\left(\theta_0 \mid \boldsymbol{X}\right) & \xrightarrow{L} N\left(0, I\left(\theta_0\right)\right) \\ \frac{1}{n} l^{\prime \prime}\left(\theta_0 \mid \boldsymbol{X}\right) & \xrightarrow{P} I\left(\theta_0\right) \end{aligned} \tag{3.1}$$

(a) 证明：

$$\frac{1}{\sqrt{n}} l^{\prime}\left(\theta_0 \mid \boldsymbol{X}\right)=\sqrt{n}\left[\frac{1}{n} \sum_i W_i\right] \tag{3.2}$$

其中 $W_i=\dfrac{\frac{\partial}{\partial \theta} f\left(X_i \mid \theta\right)}{f\left(X_i \mid \theta\right)}$ 有均值 0 和方差 $I(\theta_0)$。再利用中心极限定理证明收敛于 $N(0, I(\theta_0))$。

(b) 证明：

$$-\frac{1}{n} l^{\prime \prime}\left(\theta_0 \mid \boldsymbol{X}\right)=\frac{1}{n} \sum_i W_i^2-\frac{1}{n} \sum_i \frac{\frac{\partial^2}{\partial \theta^2} f\left(X_i \mid \theta\right)}{f\left(X_i \mid \theta\right)} \tag{3.3}$$

并且第一部分的均值是 $I(\theta_0)$, 而第二部分的均值是 0，应用弱大数定律。

解答. (a) 先证明式3.2:

$$l(\theta \mid \mathbf{X})=\sum_{i=1}^n \log f(X_i \mid \theta) \tag{3.4}$$

对式1.42中的 θ 求偏导：

$$l'(\theta|\mathbf{X})=\sum_{i=1}^{n}\frac{\frac{\partial}{\partial\theta}f(X_i|\theta)}{f(X_i|\theta)} \tag{3.5}$$

结合，$W_i=\frac{\frac{\partial}{\partial\theta}f(X_i\mid\theta)}{f(X_i\mid\theta)}$，我们可以得到：

$$\frac{1}{\sqrt{n}}l'(\theta_0\mid\boldsymbol{X})=\sqrt{n}\left[\frac{1}{n}\sum_i W_i\right] \tag{3.6}$$

因此，我们证明了式3.2。

下面计算 W_i 的均值和方差：

$$\mathrm{E}_\theta[W_i]=\mathrm{E}[\frac{\frac{\partial}{\partial\theta}f(X_i|\theta)}{f(X_i|\theta)}]\Bigg|_{\theta=\theta_0}=\frac{\partial}{\partial\theta}\mathrm{E}[\frac{f(X_i|\theta)}{f(X_i|\theta)}]\Bigg|_{\theta=\theta_0}=\frac{\partial}{\partial\theta}\mathrm{E}[1]|_{\theta=\theta_0}=0 \tag{3.7}$$

$$\mathrm{Var}_\theta[W_i]=\mathrm{E}[\frac{\frac{\partial}{\partial\theta}f(X_i|\theta)}{f(X_i|\theta)}-\mathrm{E}_\theta[W_i]]^2\Bigg|_{\theta=\theta_0}=\mathrm{E}[\frac{\frac{\partial}{\partial\theta}f(X_i|\theta)}{f(X_i|\theta)}]^2\Bigg|_{\theta=\theta_0}=I(\theta_0) \tag{3.8}$$

因此，**我们从式3.7和式3.8中可知，W_i 的均值为 0，方差为** $I(\theta_0)$。对式3.6应用中心极限定理：

$$\sqrt{n}\bar{W}=\sqrt{n}[\frac{1}{n}\sum_i W_i]\xrightarrow{L}N(0,I(\theta_0)) \tag{3.9}$$

于是，我们得到了：

$$-\frac{1}{\sqrt{n}}l'(\theta_0\mid\boldsymbol{X})\xrightarrow{L}N(0,I(\theta_0)) \tag{3.10}$$

这也证明了式3.1的第一个式子。

(b) 计算：

$$\begin{aligned}
&l''(\theta|\mathbf{X})=\frac{\partial^2}{\partial\theta^2}l(\theta|\mathbf{X})=\frac{\partial}{\partial\theta}l'(\theta|\mathbf{X})\\
&=\frac{\partial}{\partial\theta}[\sum_{i=1}^{n}\frac{\frac{\partial}{\partial\theta}f(X_i|\theta)}{f(X_i|\theta)}]=\sum_{i=1}^{n}\frac{(\frac{\partial^2}{\partial\theta^2}f(X_i|\theta))f(X_i|\theta)-(\frac{\partial}{\partial\theta}f(X_i|\theta))^2}{(f(X_i|\theta))^2}\\
&=\sum_{i=1}^{n}\frac{\frac{\partial^2}{\partial\theta^2}f(X_i|\theta)}{f(X_i|\theta)}-\sum_{i=1}^{n}W_i^2
\end{aligned} \tag{3.11}$$

于是，从式3.11中可知：

$$-\frac{1}{n} l^{\prime \prime}\left(\theta_0 \mid \boldsymbol{X}\right)=\frac{1}{n} \sum_i W_i^2-\left.\frac{1}{n} \sum_i \frac{\frac{\partial^2}{\partial \theta^2} f\left(X_i \mid \theta\right)}{f\left(X_i \mid \theta\right)}\right|_{\theta=\theta_0} \tag{3.12}$$

因此我们证明了式3.3。

计算式3.12第一部分的均值：

$$\mathrm{E} W_i^2=\mathrm{E}(W_i-0)^2=\mathrm{E}(W_i-\mathrm{E} W_i)^2=\operatorname{Var}(W_i)=I(\theta_0) \tag{3.13}$$

$$\mathrm{E}(\frac{1}{n} \sum_i W_i^2)=I(\theta_0) \tag{3.14}$$

于是，由式3.14, **式3.3第一部分的均值是** $I(\theta_0)$。再计算第二部分的均值：

$$\mathrm{E}(\frac{\frac{\partial^2}{\partial \theta^2} f(X_i|\theta)}{f(X_i|\theta)})=\frac{\partial^2}{\partial \theta^2} \mathrm{E}(\frac{f(X_i|\theta)}{f(X_i|\theta)})=\frac{\partial^2}{\partial \theta^2} \mathrm{E}(1)=0 \tag{3.15}$$

对式3.12应用弱大数定律，我们有：

$$\frac{1}{n} l^{\prime \prime}\left(\theta_0 \mid \boldsymbol{X}\right) \stackrel{P}{\rightarrow} I\left(\theta_0\right) \tag{3.16}$$

这也便证明了式3.1的第二个式子。

题目 11. 假定 $X_1, ..., X_n$ 是 $i.i.d$Poisson(λ) 的。求下列量的最佳无偏估计量：

(a) $e^{-\lambda}$, 这是 $X=0$ 的概率。

(b) $\lambda e^{-\lambda}$, 这是 $X=1$ 的概率。

(c) 对 (a) 和 (b) 中的这些最佳无偏估计量，计算它们相对于 MLE 的渐进相对效率，你更喜欢哪个估计？为什么？

(d) 对于可能致癌的化合物，可以通过测量暴露于这种化合物下的微生物的突变率来进行基本检测。试验人员把这种化合物放在 15 个皮氏培养皿中，记录到下列数目的突变群体：

$$10,7,8,13,8,9,5,7,6,8,3,6,6,3,5$$

估计 $e^{-\lambda}$, 即没有突变群体出现的概率，以及 $\lambda e^{-\lambda}$, 也就是只有一个突变群体出现的概率。计算最佳无偏估计和 MLE。

解答. 我们现在直接求泊松概率：

$$P_\lambda(k)=\frac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,... \tag{3.17}$$

的 UMVUE。

我们知道，$T_n=\sum_{i=1}^{n}X_i$ 是泊松分布族的完备充分统计量，因为泊松分布是指数分布族的成员。由泊松分布具有可加性，T_n 服从参数为 $n\lambda$ 的泊松分布，即：

$$P(T_n=t)=\frac{(n\lambda)^t}{t!}e^{-n\lambda},\quad t=0,1,2,... \tag{3.18}$$

容易看出，统计量：

$$\varphi_k(\mathbf{X})=\begin{cases}1 & 当X_1=k\\ 0 & 当X_1\neq k\end{cases} \tag{3.19}$$

是 $P_\lambda(k)$ 的无偏估计。所以 $P_\lambda(k)$ 的 UMVUE 为：

$$\begin{aligned}\hat{P}_\lambda(k)&=E_\lambda\left\{\varphi_k(\mathbf{X})\mid T_n=t\right\}\\&=P\left(X_1=k\mid T_n=t\right)=\frac{P\left(X_1=k,T_n=t\right)}{P\left(T_n=t\right)}\\&=\frac{P\left(X_1=k,X_2+\cdots+X_n=t-k\right)}{P\left(T_n=t\right)}\end{aligned} \tag{3.20}$$

考虑到 $X_1,...,X_n$ 是 $i.i.d$ 的, 且 $X_2+X_3+...+X_n$ 服从参数为 $(n-1)\lambda$ 的泊松分布，所以：

$$\begin{aligned}\hat{P}_\lambda(k)&=\frac{\dfrac{\lambda^k\mathrm{e}^{-\lambda}}{k!}\cdot\dfrac{[(n-1)\lambda]^{t-k}}{(t-k)!}\mathrm{e}^{-(n-1)\lambda}}{\dfrac{(n\lambda)^t}{t!}\mathrm{e}^{-n\lambda}}\\&=\mathrm{C}_t^k\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{t-k},\quad k=0,1,2,\cdots\end{aligned} \tag{3.21}$$

于是，$P_\lambda(k)$ 的 UMVUE 为：

$$\hat{P}_\lambda(k)=\mathrm{C}_{T_n}^k\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{T_n-k},\quad k=0,1,2,\cdots \tag{3.22}$$

(a) 在式3.22中，取 $k=0$, 知：$e^{-\lambda}$ **的 UMVUE 为：**

$$\widehat{P}_\lambda(0)=\left(1-\frac{1}{n}\right)^{T_n}=\left(1-\frac{1}{n}\right)^{\sum_{i=1}^n X_i} \tag{3.23}$$

(b) 在式3.22中，取 $k=1$, 知：$\lambda e^{-\lambda}$ **的 UMVUE 为：**

$$\hat{P}_\lambda(1)=C_{T_n}^1\left(\frac{1}{n}\right)^1\left(1-\frac{1}{n}\right)^{T_n-1}=\frac{\sum\limits_{i=1}^n X_i}{n}\left(1-\frac{1}{n}\right)^{\sum\limits_{i=1}^n X_i-1} \tag{3.24}$$

(c) 我们先计算 $e^{-\lambda}$ 和 $\lambda e^{-\lambda}$ 的 MLE。首先，我们先写出 λ 的似然函数：

$$L(\lambda)=\frac{\lambda^{\sum\limits_{i=1}^n X_i}e^{-n\lambda}}{X_1!X_2!...X_n!} \tag{3.25}$$

对数似然函数为：

$$\log(L(\lambda))=\sum_{i=1}^n X_i\log\lambda-n\lambda-\log(X_1!X_2!...X_n!) \tag{3.26}$$

对 λ 求偏导：

$$\frac{\partial}{\partial\lambda}\log(L(\lambda))=\frac{\sum\limits_{i=1}^n X_i}{\lambda}-n \tag{3.27}$$

我们令式3.27为 0，得到 λ 的 MLE 为：

$$\hat{\lambda}=\frac{\sum\limits_{i=1}^n X_i}{n}=\bar{X} \tag{3.28}$$

结合式3.28, 我们知道 $e^{-\lambda}$ 和 $\lambda e^{-\lambda}$ 的 MLE 分别为：$e^{-\bar{X}}$ 和 $\bar{X}e^{-\bar{X}}$。然后我们分别计算它们的方差，对于 $e^{-\bar{X}}$：

$$\mathrm{Var}(e^{-\bar{X}})\approx((e^{-\lambda})')^2\mathrm{Var}(\bar{X})=(-e^{-\lambda})^2\times\frac{\lambda}{n}=(e^{-\lambda})^2\times\frac{\lambda}{n} \tag{3.29}$$

对于 $\lambda e^{-\lambda}$：

$$\mathrm{Var}(\bar{X}e^{-\bar{X}})\approx((\lambda e^{-\lambda})')^2\mathrm{Var}(\bar{X})=(e^{-\lambda}-\lambda e^{-\lambda})^2\times\frac{\lambda}{n}=(e^{-\lambda}(1-\lambda))^2\times\frac{\lambda}{n} \tag{3.30}$$

然后我们计算 $e^{-\lambda}$ 的 UMVUE 式3.23的方差：

$$\mathrm{Var}\left(\left(1-\frac{1}{n}\right)^{\sum\limits_{i=1}^n X_i}\right)=\left(\frac{\partial}{\partial\lambda}\left(1-\frac{1}{n}\right)^{n\lambda}\right)^2\times\mathrm{Var}(\bar{X})=\left(n\left(1-\frac{1}{n}\right)^{n\lambda}\log\left(1-\frac{1}{n}\right)\right)^2\times\frac{\lambda}{n} \tag{3.31}$$

再计算 $\lambda e^{-\lambda}$ 的 UMVUE 式3.24的方差：

$$\begin{aligned}\mathrm{Var}\left(\frac{\sum\limits_{i=1}^n X_i}{n}\left(1-\frac{1}{n}\right)^{\sum\limits_{i=1}^n X_i-1}\right)&=\left(\frac{\partial}{\partial\lambda}\lambda\left(1-\frac{1}{n}\right)^{n\lambda-1}\right)^2\mathrm{Var}(\bar{X})\\&=\left(\frac{n}{n-1}\left(1-\frac{1}{n}\right)^{n\lambda}+\frac{n^2\lambda}{n-1}\left(1-\frac{1}{n}\right)^{n\lambda}\log\left(1-\frac{1}{n}\right)\right)^2\times\frac{\lambda}{n}\end{aligned} \tag{3.32}$$

结合式3.29和式3.31, 我们可以得到 $e^{-\lambda}$ **的 UMVUE 相对于 MLE 的渐进相对效率为:**

$$\text{ARE}\left(\left(1-\frac{1}{n}\right)^{n\lambda}, e^{-\lambda}\right) = \left[\frac{e^{-\lambda}}{n\left(1-\frac{1}{n}\right)^{n\lambda}\log\left(1-\frac{1}{n}\right)}\right]^2 \tag{3.33}$$

结合式3.30和式3.32, $\lambda e^{-\lambda}$ **的 UMVUE 相对于 MLE 的渐进相对效率为**:

$$\text{ARE}\left(\lambda\left(1-\frac{1}{n}\right)^{n\lambda-1}, \lambda e^{-\lambda}\right) = \left[\frac{(1-\lambda)e^{-\lambda}}{\frac{n}{n-1}\left(1-\frac{1}{n}\right)^{n\lambda}+\frac{n^2\lambda}{n-1}\left(1-\frac{1}{n}\right)^{n\lambda}\log\left(1-\frac{1}{n}\right)}\right]^2 \tag{3.34}$$

由于:

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = \frac{1}{e} \tag{3.35}$$

所以，在式3.33和式3.34中渐进效率是收敛于 1 的，在这种情况下，**我们更倾向选择 MLE, 因为其与 UMVUE 的渐进效率是相同的，并且 MLE 具有更简单的形式。**

(d) 我们用如下简单的 **Matlab** 代码实现 $e^{-\lambda}$ 和 $\lambda e^{-\lambda}$ 的 MLE 和 UMVUE 的计算，**得到的结果如表3.1所示：**

表 3.1: $e^{-\lambda}$ 和 $\lambda e^{-\lambda}$ 的 MLE 和 UMVUE

	MLE	UMVUE
$e^{-\lambda}$	0.000975	0.000765
$\lambda e^{-\lambda}$	0.006758	0.005684

```
X = [10,7,8,13,8,9,5,7,6,8,3,6,6,3,5];
X_bar = mean(X);
n = length(X);
sum_X = sum(X);
MLE_0 = exp(-X_bar) % 9.7475e-04
MLE_1 = X_bar*exp(-X_bar) % 0.0068
UMVUE_0 = (1-1/n)^sum_X % 7.6529e-04
UMVUE_1 = X_bar*(1-1/n)^(sum_X-1) % 0.0057
```

题目 12. Efron 分析了法学院入学数据，目的在于考察 LAST 分数与一年级 GPA 之间的相关性。对于 15 个法学院，我们有数据对 (平均 LAST, 平均 GPA)：

$$(576, 3.39)(635, 3.30)(558, 2.81)(578, 3.03)(666, 3.44)$$
$$(580, 3.07)(555, 3.00)(661, 3.43)(651, 3.36)(605, 3.13)$$
$$(653, 3.12)(575, 2.74)(545, 2.76)(572, 2.88)(594, 2.96)$$

(a) 计算 LAST 分数和 GPA 之间的相关系数。

(b) 用非参数自助法估计相关系数的标准差。用 $B = 1000$ 个重抽样本，并画出这些样本相关系数的直方图。

(c) 用参数自助法估计相关系数的标准差。假定 (LAST, GPA) 有二元正态分布，估计其中的 5 个参数。然后从这个二元正态分布产生容量为 15 的 1000 个样本。

(d) 如果 (X, Y) 是二元正态的，相关系数为 ρ, r 为样本相关系数，则用 Δ 方法可以证明：

$$\sqrt{n}(r-\rho) \to N\left(0, \left(1-\rho^2\right)^2\right) \tag{3.36}$$

用这个事实估计 r 的标准差，它与自助法估计相比如何？给出 r 的近似概率密度函数。

解答. (a) 计算相关系数的 **R** 语言代码如下，**LAST 分数和 GPA 之间的相关系数为：** 0.7763745。

```r
F = c(576,635,558,578,666,580,555,661,651,605,653,575,545,572,594)
G = c(3.39,3.3,2.81,3.03,3.44,3.07,3,3.43,3.36,3.13,3.12,2.74,
      2.76,2.88,2.96)
rho = cor(F,G) # 0.7763745
```

(b) 计算非参数自助法相关系数的标准差的 **R** 语言代码如下，**代码得到的标准差为：** 0.1378396。这些样本相关系数的直方图如图3.1所示。

```r
F = c(576,635,558,578,666,580,555,661,651,605,653,575,545,572,594)
G = c(3.39,3.3,2.81,3.03,3.44,3.07,3,3.43,3.36,3.13,3.12,2.74,
      2.76,2.88,2.96)
X = cbind(F,G)
```

```r
library(bootstrap)
n = length(F)
N = 1000
Rho = function(x, X){
  cor(X[x,1],X[x,2])
}
Sample = bootstrap(1:n,N,Rho,X,func=sd)
sd_bootstrap_nonparametric = Sample[2]
# sd_bootstrap_nonparametric = 0.1378396
hist(Sample[[1]],xlab = expression(rho), main = '' )
```

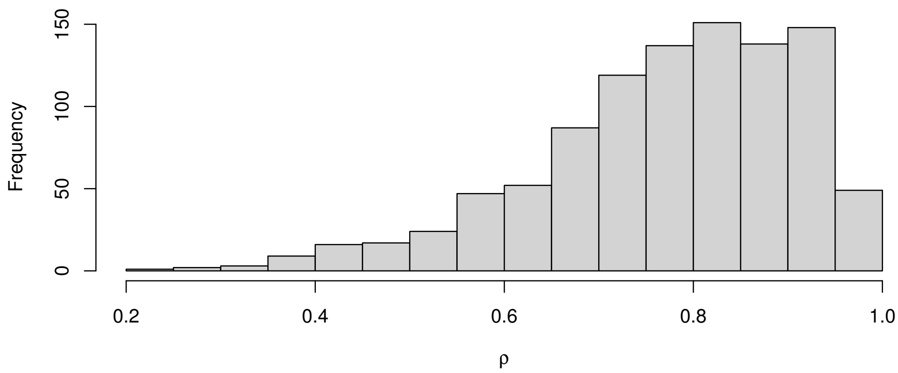

图 3.1: 非参数自助法抽取的样本相关系数 ρ 的直方图

(b) 计算参数自助法相关系数的标准差的 **R** 语言代码如下。**我们首先得到二元正态分布（LAST，GPA）的 5 个参数的估计为：**$\hat{\mu}_{LAST} = 600.2667$，$\hat{\mu}_{GPA} = 3.094667$，$\hat{\sigma}_{LAST} = 41.79451$，$\hat{\sigma}_{GPA} = 0.243512$，$\hat{\rho} = 0.7763745$。**代码得到的参数自助法相关系数的标准差为：**0.1148616。这些样本相关系数的直方图如图3.2所示。

```r
F = c(576,635,558,578,666,580,555,661,651,605,653,575,545,572,594)
G = c(3.39,3.3,2.81,3.03,3.44,3.07,3,3.43,3.36,3.13,3.12,2.74,
      2.76,2.88,2.96)
n = length(F)
N = 1000
```

```r
F_mean = mean(F)
G_mean = mean(G)
F_sd = sd(F)
G_sd = sd(G)
rho = cor(F,G)
DATA = rho
for (i in 1:N){
  y = rnorm(n,mean = G_mean,sd = G_sd)
  x = rnorm(n,mean = F_mean+rho*F_sd/G_sd*(y-G_sd),
            sd = F_sd*sqrt(1-rho^2))
  DATA = c(DATA,cor(x,y))
}
sd_DATA = sd(DATA)
# d_DATA = 0.1148616
hist(DATA,xlab = expression(rho), main = '' )
```

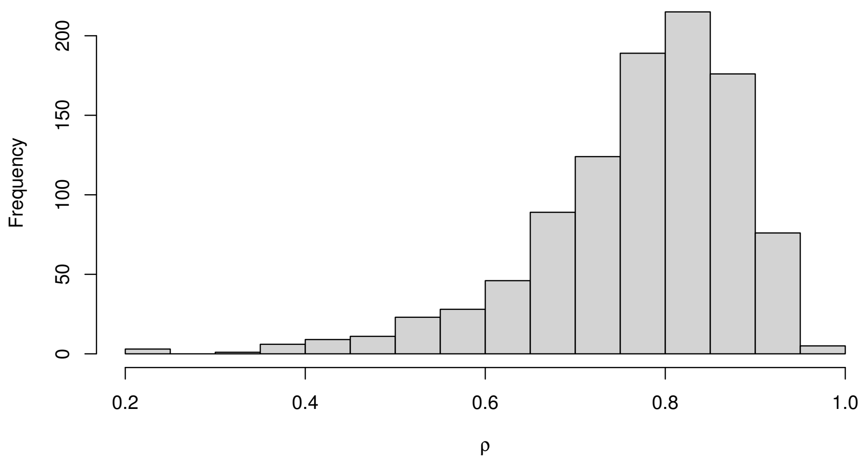

图 3.2: 参数自助法抽取的样本相关系数 ρ 的直方图

(d) 由式3.36, 我们可以得到：

$$r \sim AN\left(\rho, \frac{(1-\rho^2)^2}{n}\right) \tag{3.37}$$

于是 r 的标准差估计为：

$$\sqrt{\frac{(1-\rho^2)^2}{n}}=\sqrt{\frac{(1-0.7763745)^2}{15}}=0.1025676 \tag{3.38}$$

从式3.38中可以看出，该方法的标准差估计比自助法更小。

由式3.37, 我们可以得到 r 的近似概率密度函数；

$$p(r)=\frac{\sqrt{n}}{\sqrt{2\pi}\left(1-\rho^2\right)}e^{-\frac{n(r-\rho)^2}{2(1-\rho^2)^2}} \tag{3.39}$$

4 附加题

题目 13. 现在我们要估计 Bernoulli 分布的方差 $p(1-p)$, 这个方差的 MLE 为 $\hat{p}(1-\hat{p})$。

(a) 计算估计量 $\hat{p}(1-\hat{p})$ 的真实方差。

(b) 推导 Δ 方法得到的方差近似结果。

(c) 利用参数/非参数自助法近似估计量的方差。

解答. (a) 我们先来计算估计量 $\hat{p}(1-\hat{p})$ 的真实方差：

首先，Bernoulli 分布为：

$$p_k=p^k(1-p)^{1-k}, k=0,1 \tag{4.1}$$

然后，由样本均值抽样分布的定理，若 $X_1,...,X_n$ 为来自某个总体的样本，$\bar{X}$ 为其均值。若总体分布未知或不是正态分布，但是 $\mathrm{E}(X)=\mu$, $\mathrm{Var}(X)=\sigma^2$ 存在，则 n 较大时 $\bar{X}\sim AN(\mu,\sigma^2/n)$。于是我们知道，从 Bernoulli 分布4.1中抽取的样本 $X_1,...,X_n$, 其均值满足：

$$\bar{X}\sim AN(p,\frac{p(1-p)}{n}) \tag{4.2}$$

而 Bernoulli 分布的参数 p 的估计量为 $\hat{p}=\bar{X}$, 于是:

$$\hat{p}\sim AN(\hat{p},\frac{\hat{p}(1-\hat{p})}{n}) \tag{4.3}$$

下面我们计算 $\hat{p}(1-\hat{p})$ 的真实方差，由：

$$\begin{aligned}\mathrm{Var}\left(\hat{p}(1-\hat{p})\right)&=\mathrm{E}[\hat{p}(1-\hat{p})]^2-\left(\mathrm{E}\left[\hat{p}(1-\hat{p})\right]\right)^2\\&=\mathrm{E}\left[\hat{p}^2\right]+\mathrm{E}\left[\hat{p}^4\right]-2\mathrm{E}\left[\hat{p}^3\right]-\left(\mathrm{E}\left[\hat{p}\right]-\mathrm{E}\left[\hat{p}^2\right]\right)^2\end{aligned} \tag{4.4}$$

为了计算 $\mathrm{E}\left[\hat{p}^k\right]$, 我们首先计算正态分布 $N(\mu,\sigma^2)$ 的 1-4 阶矩。
我们知道，正态分布 $N(\mu,\sigma^2)$ 的特征函数为：

$$f(t) = e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2} \tag{4.5}$$

对式4.5分别求 1-4 阶导，我们有：

$$\begin{aligned}
&f'(t) = (\mathrm{i}\mu - \sigma^2 t)e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2}\\
&f''(t) = -\sigma^2 e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2} + (\mathrm{i}\mu - \sigma^2 t)^2 e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2}\\
&f^{(3)}(t) = -3\sigma^2(\mathrm{i}\mu - \sigma^2 t)e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2} + (\mathrm{i}\mu - \sigma^2 t)^3 e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2}\\
&f^{(4)}(t) = 3\sigma^4 e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2} - 6\sigma^2(\mathrm{i}\mu - \sigma^2 t)^2 e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2} + (\mathrm{i}\mu - \sigma^2 t)^4 e^{\mathrm{i}\mu t - \frac{1}{2}\sigma^2 t^2}
\end{aligned} \tag{4.6}$$

在式4.6中，我们令 $t = 0$, 得到：

$$\begin{aligned}
&f'(0) = \mathrm{i}\mu\\
&f''(0) = -\sigma^2 - \mu^2\\
&f^{(3)}(0) = -3(\sigma^2\mu + \mu^3)\mathrm{i}\\
&f^{(4)}(0) = 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4
\end{aligned} \tag{4.7}$$

由：

$$f^{(k)}(0) = \mathrm{i}^k \mathrm{E}\left[\hat{p}^k\right] \tag{4.8}$$

结合式4.7和4.8，代入式4.4，可以得到：

$$\mathrm{Var}\left(\hat{p}(1-\hat{p})\right) = 2\sigma^4 + 4\sigma^2\mu^2 - 4\sigma^2\mu + \sigma^2 \tag{4.9}$$

其中 $\mu = \hat{p}$, $\sigma^2 = \dfrac{p(1-p)}{n}$。最后我们得到了**估计量 $\hat{p}(1-\hat{p})$ 的真实方差**:

$$\mathrm{Var}\left(\hat{p}(1-\hat{p})\right) = \frac{(2\hat{p}(1-\hat{p}) + n(2\hat{p}-1)^2)(1-\hat{p})\hat{p}}{n^2} \tag{4.10}$$

编写如下 **R** 代码，计算在 $\hat{p}$ 不同取值时候估计量的真实方差：

```
n = 24
p = 0.25  # p = 0.5, p = 2/3
var_real = (2*p*(1-p)+n*(2*p-1)^2)*p*(1-p)/n/n
#var_real_0.25 = 0.002075195
#var_real_0.5 = 0.0002170139
#var_real_2/3 = 0.001200274
```

(b) 根据 Δ 方法与 MLE 的渐进有效性，$h(\hat{p})$ 的方差近似为：

$$
\begin{aligned}
\operatorname{Var}(h(\hat{p}) \mid p) &\approx \frac{[h'(p)]^2}{I_n(p)} \\
&= \frac{[h'(p)]^2}{\mathrm{E}_p\left(-\dfrac{\partial^2}{\partial p^2}\log L(p \mid \boldsymbol{X})\right)} \\
&\approx \frac{[h'(p)]^2\big|_{p=\hat{p}}}{-\dfrac{\partial^2}{\partial p^2}\log L(p \mid \boldsymbol{X})\Big|_{p=\hat{p}}}
\end{aligned} \tag{4.11}
$$

利用式4.11，可知**用 Δ 方法的 $\hat{p}(1-\hat{p})$ 的方差近似为**：

$$
\begin{aligned}
\widehat{\operatorname{Var}}(\hat{p}(1-\hat{p})) &= \frac{\left[\dfrac{\partial}{\partial p}(p(1-p))\right]^2\Bigg|_{p=\hat{p}}}{-\dfrac{\partial^2}{\partial p^2}\log L(p \mid \boldsymbol{x})\Big|_{p=\hat{p}}} \\
&= \frac{(1-2p)^2\big|_{p=\hat{p}}}{\dfrac{n}{p(1-p)}\Big|_{p=\hat{p}}} \\
&= \frac{\hat{p}(1-\hat{p})(1-2\hat{p})^2}{n}
\end{aligned} \tag{4.12}
$$

编写如下 **R** 代码，计算在 $\hat{p}$ 不同取值时使用 Δ 方法的 $\hat{p}(1-\hat{p})$ 的方差近似：

```
n = 24
p = 0.25 # # p = 0.5, p = 2/3
var_△1 = (2*p-1)^2*p*(1-p)/n
#var_△1_0.25 = 0.001953125
#var_△1_0.5 = 0
#var_△1_2/3 = 0.001028807
```

而在式4.12中，若 $\hat{p} = 0.5$ 此时计算得到的方差为 0，这显然不是一个很好的估计。此

时我们需要用到二阶 Δ 方法来计算 $\hat{p}(1-\hat{p})$ 的方差近似：

$$\begin{aligned}\widehat{\mathrm{Var}}(\hat{p}(1-\hat{p})) &= \frac{\left[\frac{\partial^2}{\partial p^2}(p(1-p))\right]^2\Big|_{p=\hat{p}}}{2\left(-\frac{\partial^2}{\partial p^2}\log L(p\mid \mathbf{x})\right)^2\Big|_{p=\hat{p}}} \\ &= \frac{4|_{p=\hat{p}}}{2\times\left(\frac{n}{p(1-p)}\right)^2\Big|_{p=\hat{p}}} = \frac{2\hat{p}^2(1-\hat{p})^2}{n^2}\end{aligned} \tag{4.13}$$

编写如下 **R** 代码，计算在 $\hat{p}$ 不同取值时使用 Δ 二阶方法的 $\hat{p}(1-\hat{p})$ 的方差近似：

```
n = 24
p = 0.5   # p = 0.25, p = 2/3
var_△2 = (2*p*p*(1-p)^2)/n/n
#var_△2_0.25 = 0.0001220703
#var_△2_0.5 = 0.0002170139
#var_△2_2/3 = 0.0001714678
```

(c) 我们编写的非参数自助法近似估计量的 **R** 代码如下：

```
n = 24
p = 0.5 # p = 0.25, p = 2/3
X = rbinom(n,1,p)
library(bootstrap)
N = 1000
Var = function(x,X){
  mean(X[x])*(1-mean(X[x]))
}
Sample = bootstrap(1:n,N,Var,X,func = sd)
var_bootstrap_nonparametric = Sample[[2]]*Sample[[2]]
#var_bootstrap_nonparametric_0.25 = 0.001814157
#var_bootstrap_nonparametric_0.5 = 0.0002064469
#var_bootstrap_nonparametric_2/3 = 0.001198679
```

我们编写的参数自助法近似估计量的 **R** 代码如下：

```
n = 24
p = 0.5 # p = 0.25, p = 2/3
N = 1000
X = rbinom(n,1,p)
var = mean(X)*(1-mean(X))
DATA = var
for (i in 1:N){
  x = rbinom(n,1,p)
  DATA = c(DATA,mean(x)*(1-mean(x)))
}
var_DATA = var(DATA)
#var_DATA_0.25 = 0.001917107 , var_DATA_0.5 = 0.0002101721 , ...
        var_DATA_2/3 = 0.001042534
```

综上，我们得到了不同方法计算 $\hat{p}(1-\hat{p})$ 的结果，如表4.1所示：

表 4.1: $\hat{p}(1-\hat{p})$ 不同方法的估计

	$\hat{p}=1/4$	$\hat{p}=1/2$	$\hat{p}=2/3$
Δ 一阶方法	0.001953125	0	0.001028807
Δ 二阶方法	0.000122070	0.000217014	0.000171468
非参数自助法	0.001814157	0.000206447	0.001198679
参数自助法	0.001917107	0.000210172	0.001042534
真值	0.002075195	0.000217014	0.001200274

从表4.1中可以看出，Δ 一阶方法是一个一阶的近似，它是基于泰勒展开的第一项。Δ 二阶方法是一个二阶的近似，它是基于泰勒展开的第二项。只有一阶的近似等于 **0** 的时候，才需要使用二阶 Δ 方法。而自助法拥有类似于“二阶”方法的精确度，往往比 Δ 一阶方法具有更为精确的结果。本题可参考：

https://www.journalajpas.com/index.php/AJPAS/article/view/24547/45892