统计学导论：R语言实验10

不同估计方法的比较和极大似然估计

主讲人：郑盼盼

Outline

1. 不同估计方法的比较
2. 极大似然估计

10.1 不同估计方法的比较

10.1.1 样本估计值和均方误差

设置参数 a 的实际值为 a_0，用计算机重复模拟样本数据。对于第 k 此模拟的样本数据计算估计量 T 的值 $T_k, 1 \leq k \leq n$，记

$$
\begin{aligned}
\bar{T} &= \frac{1}{n} \sum_{k=1}^{n} T_k \\
\widehat{\mathrm{mse}}(T, a_0) &= \frac{1}{n} \sum_{k=1}^{n} (T_k - a_0)^2
\end{aligned}
$$

依据大数定律，当 n 充分大时，$\bar{T}$ 和 $\widehat{\mathrm{mse}}(T, a_0)$ 分别接近于 $\mathbb{E}(T)$ 和 $\mathrm{mse}(T, a_0)$。

10.1.2 例

若 X_1, X_2, X_3 是来自总体变量 $X \sim N(\mu, \sigma^2)$ 的样本，就可分别以样本方差 S^2 和统计量 $T = \frac{1}{3}\sum_{k=1}^{3}(X_k - \bar{X})^2$ 估计样本方差。试用随机模拟的方法判断 S^2 和 T 中哪一个估计总体方差的效果更好。

1. 模拟真实参数为 $\mu = 0$ 和 $\sigma = 0.5, 0.8, 1, 2, 5$ 的情况，重复估计 10000 次，通过估计量的算术平均值和均方误差估计两种估计方法的效果。

10.1.2 例

```r
m <- 10000
barS2 <- c()
barT <- c()
mseS2 <- c()
mseT <- c()
for (trueSigma in c(0.5, 0.8, 1,2,5)) {
  tmpX <- matrix(rnorm(3 * m, 0, trueSigma), m, 3)
  tmpS2 <- apply(tmpX, 1, var)
  tmpT <- 2/3 * tmpS2
  barS2 <- c(barS2, mean(tmpS2))
  barT <- c(barT, mean(tmpT))
  mseS2 <- c(mseS2, mean((tmpS2 - trueSigma^2)^2))
  mseT <- c(mseT, mean((tmpT - trueSigma^2)^2))
}
myResults <- c(barS2, barT, mseS2, mseT)
myResults <- matrix(myResults, 5,4)
myResults <- data.frame(t(myResults),
                row.names = c('mean(S2)', 'mean(T)', 'mse(S2)', 'mse(T)'))
names(myResults) <- c(0.25,0.64,1,4,25)
round(myResults,3)
```

10.1.2 例

若 X_1, X_2, X_3 是来自总体变量 $X \sim N(\mu, \sigma^2)$ 的样本，就可分别以样本方差 S^2 和统计量 $T = \frac{1}{3}\sum_{k=1}^{3}(X_k - \bar{X})^2$ 估计样本方差。试用随机模拟的方法判断 S^2 和 T 中哪一个估计总体方差的效果更好。

2. 以 $\mu = 2, \sigma^2 = 4$ 为例，通过图像来对比两种估计的效果

10.1.2 例

```r
p <- 0.1
n <- 10000
myR <- data.frame("rule1" = rep(0,n), "rule2" = rep(0,n))
for (i in 1:n) {
  x <- rbinom(1, size = 1, prob = p)
  myR$rule1[i] = ifelse(x == 1, 0.9, 0.1)
  myR$rule2[i] = ifelse(x == 1, 0.1, 0.9)
}
sum(myR$rule1 == p) / n
sum(myR$rule2 == p) / n
```

10.2 极大似然估计

10.2.0 一个例子

已知一枚硬币出现正面的概率 p 或为 0.1 或为 0.9，若掷这枚硬币 1 次，如何根据实验结果判定 p？于是存在两种准则

- **准则一**：判断结论使得样本数据出现的概率大。即出现正面，判断 $p=0.9$；否则，判断 $p=0.1$
- **准则二**：判断结论使得样本数据出现的概率小。即出现正面，判断 $p=0.1$；否则，判断 $p=0.9$

10.2.0 一个例子

已知一枚硬币出现正面的概率 p 或为 0.1 或为 0.9，若掷这枚硬币 1 次，如何根据实验结果判定 p? 于是存在两种准则

- **准则一**：判断结论使得样本数据出现的概率大。即出现正面，判断 $p=0.9$；否则，判断 $p=0.1$
- **准则二**：判断结论使得样本数据出现的概率小。即出现正面，判断 $p=0.1$；否则，判断 $p=0.9$

```r
p <- 0.1
n <- 10000
myR <- data.frame("rule1" = rep(0,n), "rule2" = rep(0,n))
for (i in 1:n) {
  x <- rbinom(1, size = 1, prob = p)
  myR$rule1[i] = ifelse(x == 1, 0.9, 0.1)
  myR$rule2[i] = ifelse(x == 1, 0.1, 0.9)
}
sum(myR$rule1 == p) / n
sum(myR$rule2 == p) / n
```

10.2.1 极大似然估计 (maximum likelihood estimation)

一般地，总体变量 X 的重复观测样本 $X_1, \ldots, X_n$

$$L(a) \triangleq p(x_1) \times p(x_2) \times \cdots \times p(x_n)$$

其中

$$p(x_i) = \begin{cases} \mathbb{P}(X = x_i), & \text{若}X\text{为离散型随机变量} \\ f(x_i), & \text{若}X\text{的密度函数为}f(x) \end{cases}$$

因此使得 $L(a)$ 达到最大的 $\hat{a}$ 是 a 的极大似然估计，称 $L(a)$ 为**似然函数**。

由于对数函数是严格增函数，因此**对数似然函数**

$$l(a) \triangleq \log L(a) = \sum_{i=1}^n \log(p(x_i))$$

的极大值点也是参数 a 的极大似然估计。

10.2.2 极大似然估计的R语言实现

1. **根据题目定义似然函数 `myL(a， x)`**
    - a 为我们估计的参数
    - x 为我们所得到的一系列观测值
2. 通过优化函数 `optimize(f,x,interval,maximum)` 获得 `myL(a,x)` 的最大值点，即极大似然估计结果
    - f 为我们优化的函数，即似然函数 `myL`
    - x 为观测值的具体数值，对应于 `myL(a,x)` 中的 x
    - `interval` 为我们估计参数 a 的取值范围
    - `maximum` 由于我们希望 `myL()` 能取到最大值，此处为 TRUE

10.2.2 极大似然估计的R语言实现

例1 假设总体 $\xi \sim B(10000, p)$，其中总体参数 $p \in [0,1]$，若 X 为总体的容量为 1 的观测样本。使用极大似然估计参数 p

1. 二项分布密度函数

$$p(k) = \mathbb{P}(\xi = k) = \binom{10000}{k} p^k (1-p)^{10000-k}$$

因此，观测值 X 发生的概率为：

$$p(X) = \mathbb{P}(\xi = X) = \binom{10000}{X} p^X (1-p)^{10000-X}$$

10.2.2 极大似然估计的R语言实现

例1 假设总体 $\xi \sim B(10000, p)$，其中总体参数 $p \in [0,1]$，若 X 为总体的容量为 1 的观测样本。使用极大似然估计参数 p

2. 根据似然函数的定义（此处只有一个观测值）：

$$L(p, X) = \binom{10000}{X} p^X (1-p)^{10000-X}$$

10.2.2 极大似然估计的R语言实现

例1 假设总体 $\xi \sim B(10000, p)$，其中总体参数 $p \in [0,1]$，若 X 为总体的容量为 1 的观测样本。使用极大似然估计参数 p

2. 根据似然函数的定义（此处只有一个观测值）：

$$L(p, X) = \binom{10000}{X} p^X (1-p)^{10000-X}$$

3. 因此，对数似然函数：

$$\begin{aligned} l(p, X) &= \binom{10000}{X} + X \log p + (10000 - X) \log(1-p) \\ &\propto X \log p + (10000 - X) \log(1-p) \end{aligned}$$

10.2.2 极大似然估计的R语言实现

例1 假设总体 $\xi \sim B(10000, p)$，其中总体参数 $p \in [0, 1]$，若 X 为总体的容量为 1 的观测样本。使用极大似然估计参数 p

3. 因此，对数似然函数：

$$\begin{aligned} l(p, X) &= \binom{10000}{X} + X \log p + (10000 - X) \log(1 - p) \\ &\propto X \log p + (10000 - X) \log(1 - p) \end{aligned}$$

```r
myL <- function(p, x){
  return (x * log(p) + (10000-x) * log(1-p))
}
y <- rbinom(1, 10000, 0.2)
hatP <- optimize(f = myL, x = y, interval = c(0,1),
maximum = T)
hatP[[1]]
```

10.2.2 极大似然估计的R语言实现

例2 下表是来自泊松分布 $P(\lambda)$ 的独立同分布样本数据，求 λ 的极大似然估计：

表 5-4　泊松分布样本数据

4	5	1	4	4	2	1	5	1	2	3	0	4	2	3	5	5	5	4	2
1	2	3	1	3	1	2	0	0	1	6	4	1	0	2	1	3	5	4	7

10.2.2 极大似然估计的R语言实现

例2 下表是来自泊松分布 $P(\lambda)$ 的独立同分布样本数据，求 λ 的极大似然估计：

1. 泊松分布密度函数

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

因此，观测值 X 发生的概率为：

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

10.2.2 极大似然估计的R语言实现

例2 下表是来自泊松分布 $P(\lambda)$ 的独立同分布样本数据，求 λ 的极大似然估计：

2. 对于 n 个观测值，$X_1, \ldots, X_n$，根据似然函数的定义，其似然函数为：

$$L(\lambda, X) = \frac{\lambda^{X_1}}{X_1!}e^{-\lambda}\frac{\lambda^{X_2}}{X_2!}e^{-\lambda}\cdots\frac{\lambda^{X_n}}{X_n!}e^{-\lambda} = (e^{-\lambda})^n\prod_{i=1}^{n}\frac{\lambda^{X_i}}{X_i!}$$

3. 对数似然函数为：

$$\begin{aligned} l(\lambda, X) &= -n\lambda + \sum_{i=1}^{n}[X_i \log\lambda - \log(X_i!)] \\ &\propto -n\lambda + \sum_{i=1}^{n} X_i \log\lambda \end{aligned}$$

10.2.2 极大似然估计的R语言实现

例2 下表是来自泊松分布 $P(\lambda)$ 的独立同分布样本数据，求 λ 的极大似然估计：

3. 对数似然函数为：

$$
\begin{aligned}
l(\lambda, X) &= -n\lambda + \sum_{i=1}^{n}[X_i \log\lambda - \log(X_i!)] \\
&\propto -n\lambda + \sum_{i=1}^{n} X_i \log\lambda
\end{aligned}
$$

```r
y = c(4,5,1,4,4,2,1,5,1,2,3,0,4,2,3,5,5,5,4,2,
      1,2,3,1,3,1,2,0,0,1,6,4,1,0,2,1,3,5,4,7)
tmpF <- function(lambda, x){
  n <- length(x)
  return (-n * lambda + log(lambda) * sum(x))
}
hatLambda <- optimize(f = tmpF, x = y,interval = c(0,100), maximum=T)
hatLambda[[1]]
```

10.2.2 极大似然估计的R语言实现

例3 下表来自于正态总体 $N(0,\sigma^2)$ 的独立同分布样本数据，求参数 σ 的极大似然估计

表 5-5　来自正态总体 $N(0,\sigma^2)$ 的独立同分布样本数据

3.14	0.72	−1.30	−0.10	1.35	1.36	−0.42	−4.93	1.17	0.47
2.23	−2.28	0.76	−0.74	4.49	−3.62	−1.58	−1.75	0.46	1.41

10.2.2 极大似然估计的R语言实现

例3 下表来自于正态总体 $N(0,\sigma^2)$ 的独立同分布样本数据，求参数 σ 的极大似然估计

1. $N(0,\sigma^2)$ 的密度函数

$$\varphi_{0,\sigma}(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)$$

因此，观测值 X 的概率密度为：

$$\varphi_{0,\sigma}(X)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{X^2}{2\sigma^2}\right)$$

10.2.2 极大似然估计的R语言实现

例3 下表来自于正态总体 $N(0,\sigma^2)$ 的独立同分布样本数据，求参数 σ 的极大似然估计

2. 对于 n 个观测值，$X_1,\ldots,X_n$，根据似然函数的定义，其似然函数为：

$$
\begin{aligned}
L(\sigma,X) &= \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{X_1^2}{2\sigma^2}\right)\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{X_2^2}{2\sigma^2}\right)\cdots\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{X_n^2}{2\sigma^2}\right)\\
&= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\prod_{i=1}^{n}\exp\left(-\frac{X_i^2}{2\sigma^2}\right) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\exp\left(-\sum_{i=1}^{n}\frac{X_i^2}{2\sigma^2}\right)
\end{aligned}
$$

3. 对数似然函数为：

$$
\begin{aligned}
l(\sigma,X) &= -n\log\sigma-\frac{n}{2}\log 2\pi-\sum_{i=1}^{n}\frac{X_i^2}{2\sigma^2}\\
&\propto -n\log\sigma-\frac{1}{2\sigma^2}\sum_{i=1}^{n}X_i^2
\end{aligned}
$$

10.2.2 极大似然估计的R语言实现

例3 下表来自于正态总体 $N(0,\sigma^2)$ 的独立同分布样本数据，求参数 σ 的极大似然估计

3. 对数似然函数为：

$$l(\sigma,X) = -n\log\sigma - \frac{n}{2}\log 2\pi - \sum_{i=1}^{n}\frac{X_i^2}{2\sigma^2}$$

$$\propto -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}X_i^2$$

```r
y <- c(3.14,0.72,-1.30,-0.10,1.35,1.36,-0.42,-4.93,1.17,0.47,
       2.23,-2.28,0.76,-0.74,4.49,-3.62,-1.58,-1.75,0.46,1.41)
myF <- function(sigma, x){
  n <- length(x)
  return(- n * log(sigma) - 1/(2*sigma^2)*sum(x^2))
}
hatSigma <- optimize(f = myF, x = y, interval = c(0,1000),maximum = T)
hatSigma[[1]]
```

10.2.2 极大似然估计的R语言实现

3. 多参数的估计，这时需要使用函数 `optim(par, fn, x)`
    - `par` 多参数的初始值
    - `fn` 优化函数
    - `x` 观测值对应于 `myL` 中的 `x`

10.2.2 极大似然估计的R语言实现

例4 下表来自于正态总体 $N(\mu, \sigma^2)$ 的独立同分布样本数据，求参数 $a = (\mu, \sigma)$ 的极大似然估计

表 5-5 来自正态总体 $N(0, \sigma^2)$ 的独立同分布样本数据

3.14	0.72	−1.30	−0.10	1.35	1.36	−0.42	−4.93	1.17	0.47
2.23	−2.28	0.76	−0.74	4.49	−3.62	−1.58	−1.75	0.46	1.41

10.2.2 极大似然估计的R语言实现

例4 下表来自于正态总体 $N(\mu, \sigma^2)$ 的独立同分布样本数据，求参数 $a = (\mu, \sigma)$ 的极大似然估计

1. $N(\mu, \sigma^2)$ 的密度函数

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

因此，观测值 X 的概率密度为：

$$\varphi_{\mu,\sigma}(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(X-\mu)^2}{2\sigma^2}\right]$$

10.2.2 极大似然估计的R语言实现

例4 下表来自于正态总体 $N(\mu, \sigma^2)$ 的独立同分布样本数据，求参数 $a = (\mu, \sigma)$ 的极大似然估计

2. 对于 n 个观测值，$X_1, \ldots, X_n$，根据似然函数的定义，其似然函数为：

$$L(a, X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(X_1-\mu)^2}{2\sigma^2}\right] \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(X_2-\mu)^2}{2\sigma^2}\right] \cdots \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(X_n-\mu)^2}{2\sigma^2}\right]$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} \exp\left[-\frac{(X_i-\mu)^2}{2\sigma^2}\right] = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{2\sigma^2}\right]$$

3. 对数似然函数为：

$$l(a, X) = -n\log\sigma - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$\propto -n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

10.2.2 极大似然估计的R语言实现

例4 下表来自于正态总体 $N(\mu,\sigma^2)$ 的独立同分布样本数据，求参数 $a=(\mu,\sigma)$ 的极大似然估计

3. 对数似然函数为：

$$l(a,X)=-n\log\sigma-\frac{n}{2}\log 2\pi-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$\propto -n\log\sigma-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

```
y <- c(3.14,0.72,-1.30,-0.10,1.35,1.36,-0.42,-4.93,1.17,0.47,
       2.23,-2.28,0.76,-0.74,4.49,-3.62,-1.58,-1.75,0.46,1.41)
tmpF <- function(a, x){
  n <- length(x)
  mu <- a[1]
  sigma <- a[2]
  f <- -n*log(sigma) - sum((x - mu)^2)/(2*sigma^2)
  return(-f)
}
hatA <- optim(par=c(0,2), fn=tmpF, x=y)
hatA[[1]]
```