第五次作业

1907402030 熊雄

2021 年 11 月 16 日

Question 1. (课件思考题)

Consider a random sample $X_1, X_2, \cdots, X_n \sim Unif(0, \theta)$.

1. Find the estimator for θ through MoM, denoted by $\hat{\theta}_{MM}$.
2. Find the MLE $\hat{\theta}_{MLE}$.
3. What are the expectation and variance of $\hat{\theta}_{MM}$ and $\hat{\theta}_{MLE}$? Which estimator is better?

Answer.

1. **Solve.**

   设母体 $X \sim Unif(0, \theta)$, 故

   $$EX = \int_0^\theta \frac{x}{\theta} dx = \frac{\theta}{2}.$$

   因此用矩法估计得方程:

   $$\overline{X} = \frac{\theta}{2}.$$

   从而得到 θ 的矩估计为

   $$\hat{\theta}_{MM} = 2\overline{X}.$$

2. **Solve.**

设子样的观测值分别为 $x_1, x_2, \cdots, x_n$. 似然函数

$$L(\theta; x_1, x_2, \cdots, x_n) = \frac{1}{\theta^n},\ 0 < x_i \le \theta,\ i = 1, \cdots, n$$

是 θ 的一个单调递减函数. 由于每一个 $x_i \le \theta$, 最大次序统计量的观测值 $x_{(n)} = \max\limits_{1\le i\le n} x_i \le \theta$. 在 $0 < x_i \le \theta, i = 1, \cdots, n$ 中要使 $L(\theta; x_1, x_2, \cdots, x_n) = \frac{1}{\theta^n}$ 达到极大, 就要使 θ 达到最小. 但 θ 不能小于 $x_{(x_n)}$, 否则子样的观测值为 $x_1, x_2, \cdots, x_n$ 就不是来自这一母体, 所以 $\hat{\theta}_L = x_{(n)}$ 是 θ 的极大似然估计值. 于是 $\hat{\theta}_L(X_1, X_2, \cdots, X_n) = X_{(n)}$ 即最大次序统计量是参数 θ 的极大似然估计量, 即

$$\hat{\theta}_{MLE} = X_{(n)}.$$

3. **Solve.**

显然, 由矩法估计的性质我们有:

$$E(\hat{\theta}_{MM}) = \theta,$$

$$Var(\hat{\theta}_{MM}) = Var(2\overline{X}) = \frac{n}{4} Var(X_1) = \frac{\theta^2}{3n}.$$

$\hat{\theta}_{MLE}$ 的分布函数为:

$$F(x) = P(\hat{\theta}_{MLE} < x) = P(X_{(n)} < x) = \begin{cases} 0, & x \le 0 \\ (\frac{x}{\theta})^n, & 0 < x < \theta \\ 1, & x \ge \theta \end{cases}$$

从而 $\hat{\theta}_{MLE}$ 的密度函数为:

$$f_{\hat{\theta}_{MLE}}(x) = \begin{cases} \frac{nx^{n-1}}{\theta^n}, & 0 < x < \theta \\ 0, & else \end{cases}$$

则很容易可求得

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

$$E(\hat{\theta}_{MLE}^2) = \int_0^\theta x \cdot \left(\frac{nx^{n-1}}{\theta^n}\right)^2 dx = \frac{n}{n+2}\theta^2,$$

$$Var(\hat{\theta}_{MLE}) = E(\hat{\theta}^2_{MLE}) - \left(E(\hat{\theta}_{MLE})\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2.$$

由以上的计算可以看出 $\hat{\theta}_{MM}$ 是 θ 的无偏估计, $\hat{\theta}_{MLE}$ 是 θ 的有偏估计, 但是二者的阶 (order) 偏差较小, 且其方差的阶 (order) 比 $\hat{\theta}_{MM}$ 小, 并且可以比较得到在 $n > 0$ 时 $Var(\hat{\theta}_{MLE}) < Var(\hat{\theta}_{MM})$. 综上所述, 选择 $\hat{\theta}_{MLE}$ 作为 θ 的估计更好.

Question 2. (课本 P151 d5.9)

在研究国家财政收入时, 我们把财政收入按收入形式分为: 各项税收收入、企业收入、债务收入、国家能源交通重点建设基金收入、基本建设贷款归还收入、国家预算调节基金收入、其他收入等。为了建立国家财政收入回归模型, 我们以财政收入 y（亿元）为因变量, 自变量如下: x_1 为农业增加值（亿元）;x_2 为工业增加值（亿元）;x_3 为建筑业增加值（亿元）;x_4 为人口数（万人）;x_5 为社会消费总额（亿元）;x_6 为受灾面积（万公顷）。据《中国统计年鉴》获得 1978 ~ 1998 年共 21 个年份的统计数据, 见**课本 P151 表 5-5**。由定性分析知，所选自变量与变量 y 有较强的相关性, 分别用后退法和逐步回归法做自变量选元。

Answer.

```r
# —- 《中国统计图鉴》1978-1998 年的统计数据 —-
y <- as.matrix(c(1132.3,1146.4,1159.9,1175.8,1212.3,1367.0,1642.9,2004.8,
                 2122.0,2199.4,2357.2,2664.9,2937.1,3149.5,3483.4,4349.0,
                 5218.1,6242.2,7408.0,8651.1,9876.0) )

x <- matrix(c(1018.4,1258.9,1359.4,1545.6,1761.6,1960.8,2295.5,2541.6,
              2763.9,3204.3,3831.0,4228.0,5017.0,5288.6,5800.0,6882.1,
              9457.2,11993.0,13844.2,14211.2,14599.6,
              1607.0,1769.7,1996.5,2048.4,2162.3,2375.6,2789.0,3448.7,
              3967.0,4585.8,5777.2,6484.0,6858.0,8087.1,10284.5,14143.8,
              19359.6,24718.3,29082.6,32412.1,33429.8,
              138.2,143.8,195.5,207.1,220.7,270.6,316.7,417.9,
              525.7,665.8,810.0,794.0,859.4,1015.1,1415.0,2284.7,
              3012.6,3819.6,4530.5,4810.6,5262.0,
              96259,97542,98705,100072,101654,103008,104357,105851,
              107507,109300,111026,112704,114333,115823,117171,118517,
```

```
                 119850,121121,122389,123626,124810,
                  2239.1,2619.4,2976.1,3309.1,3637.9,4020.5,4694.5,5773.0,
                  6542.0,7451.2,9360.1,10556.5,11365.2,13145.9,15952.1,20182.1,
                 26796.0,33635.0,40003.9,43579.4,46405.9,
                 50760,39370,44530,39790,33130,34710,31890,44370,
                 47140,42090,50870,46990,38470,55470,51330,48830,
                 55040,45821,46989,53429,50145
),nrow=21,ncol=6)

# （1）建立 y 对 x1-x6 的线性回归方程 —-
lm5.9 <- lm(y ~ x[,1] + x[,2] +x[,3] +x[,4] +x[,5] +x[,6])
summary(lm5.9)
# 回归方程为 y=1348 - 0.641x1 - 0.317x2 - 0.4125x3 - 0.002111x4 - 0.6711x5 - 0.00752x6


# （2）用后退法选择自变量 —-
lm5.9.back <- step(lm5.9,direction='backward')
summary(lm5.9.back)

# 后退法剔除 P 值最大的 x3,x4,x6，保留 x1,x2,x5 作为最终模型，模型的参数均通过显著性检验。
# 得回归方程为到 y=874.60021 - 0.61119x1 - 0.35305x2 + 0.63671x5
# 模型表明国家财政收入 y 与农业增加值 x1、工业增加值 x2、社会消费总额 x5 有显著线性关系。


# （3）用逐步回归法选择自变量 —-
lm5.9_step <- step(lm5.9,direction='both')
summary(lm5.9_step)
# 得到 y=874.60021 - 0.61119x1 - 0.35305x2 + 0.63671x5
# 使用 step() 函数进行逐步回归，其结果与后退法一致。
```