Homework I: DATA130048

Biostatistics

Due Thursday, March 26th, 2020

1 Problem 1: MLE & MME

Suppose that X is a discrete random variable with the following probability mass function, where $0 \le \theta \le 1$ is a parameter. The following 10 independent observations were taken from such a distribution: $(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)$

X	0	1	2	3
$P(X)$	$\frac{2\theta}{3}$	$\frac{\theta}{3}$	$\frac{2(1-\theta)}{3}$	$\frac{1-\theta}{3}$

Derive the MLE (maximum likelihood estimator) and MME (method of moments estimator) of θ.

解: ① The probability density function of X is:

$$P(X=x) = \left[\left(\frac{2\theta}{3}\right)^{\frac{1}{6}(1-x)} \cdot \left(\frac{\theta}{3}\right)^{\frac{1}{2}x}\right]^{(2-x)(3-x)} * \left[\left(\frac{2}{3}(1-\theta)\right)^{-\frac{1}{2}(3-x)} \cdot \left(\frac{1-\theta}{3}\right)^{\frac{1}{6}(2-x)}\right]^{x(1-x)}$$

$$\Rightarrow L(\theta) = \sum_{i=1}^{n} \left[\frac{1}{6}(1-x_i)\ln(\theta) + \frac{x_i}{2}\ln(\theta)\right] * (2-x_i)(3-x_i) + \left[-\frac{1}{2}(3-x_i)\ln(1-\theta) + \frac{1}{6}(2-x_i)\ln(1-\theta)\right] * x_i(1-x_i) + \text{Const.}$$

$$\Rightarrow \frac{\partial L(\theta)}{\partial \theta} = \frac{1}{\theta}\sum_{i=1}^{n}\left(\frac{1}{6}+\frac{x_i}{3}\right)(2-x_i)(3-x_i) - \frac{1}{1-\theta}\sum_{i=1}^{n}\left(-\frac{7}{6}+\frac{x_i}{3}\right)x_i(1-x_i) = 0 \quad \Rightarrow \frac{\partial L(\theta)}{\partial \theta} = \frac{5}{\theta} - \frac{5}{1-\theta} \Rightarrow \hat{\theta} = 0.5$$

② Sample mean $\bar{X} = \frac{1}{10}\sum_{i=1}^{10} x_i = 1.5$ where the first moment $E[X] = \frac{\theta}{3} + \frac{4}{3}(1-\theta) + (1-\theta) = \frac{7}{3} - 2\theta$

$\Rightarrow$ The MME: $\hat{\theta} = \frac{5}{12} = 0.4167$.

2 Problem 2: MLE & MME

Suppose $X_1, X_2, ..., X_n$ are iid random variables with density function

$$f(x \mid \sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right)$$

Find the MLE and MME of σ.

解: ① The MLE: $L(\vec{x};\sigma) = \sum_{i=1}^{n}\left[-\frac{|x_i|}{\sigma} - \log(2\sigma)\right] \Rightarrow$ 令 $\frac{\partial L(\vec{x};\sigma)}{\partial \sigma} = \sum_{i=1}^{n}\left[|x_i| * \frac{1}{\sigma^2} - \frac{1}{\sigma}\right] \Rightarrow \hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}|x_i|$

② The MME:

$$\left.\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot \frac{1}{2\sigma}\exp\left\{-\frac{|x|}{\sigma}\right\}dx = 0 \\ E[X^2] &= \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{2\sigma}\exp\left\{-\frac{|x|}{\sigma}\right\}dx = 2\sigma^2 \end{aligned}\right\} \Rightarrow \hat{\sigma} = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} x_i^2}$$

3 Problem 3: MLE & MME

Suppose that $X_1, ..., X_n$ form a random sample from a uniform distribution on the interval $(0, \theta)$, where parameter $\theta > 0$ is unknown. Find the MME and MLE of θ.

解: ① MLE: $f(x) = \frac{1}{\theta}, \quad 0 \le x \le \theta \quad \Rightarrow \quad L(\vec{x};\theta) = \left(\frac{1}{\theta}\right)^n$

由于 $\hat{\theta} = \arg\max L(\vec{x};\theta)$，则区间 $(0,\theta)$ 应尽可能包括更多 data points $\Rightarrow \hat{\theta} = \max\{x_i, i=1,\dots,n\}$

② MME: $E[X] = \frac{2}{\theta} \quad \Rightarrow \quad \tilde{\theta} = 2\bar{X} = \frac{2}{n}\sum_{i=1}^{n} x_i$

4 Problem 4: Central Limit Theorem & Delta Method

Suppose $X_n \sim Binomial(n, p)$, with $p \neq \frac{1}{2}$. Because $\frac{X_n}{n}$ is the maximum likelihood estimator for p, the maximum likelihood estimator for $p(1-p)$ is $\delta_n = \frac{X_n(n-X_n)}{n^2}$.
Use Central Limit Theorem to show the limiting distribution for $\frac{X_n}{n}$, and use Delta Method to derive the limiting distribution for δ_n.

解: $E\left[\frac{X_n}{n}\right] = p \ ; \ Var\left[\frac{X_n}{n}\right] = \frac{p}{n}(1-p)$

由中心极限定理，有: $\dfrac{\frac{X_n}{n} - p}{\sqrt{p(1-p)/n}} \xrightarrow{D} N(0,1)$ as $n \to \infty$ 即 $\frac{X_n}{n} \xrightarrow{D} N\left(p, \frac{p(1-p)}{n}\right)$ as $n \to \infty$

由于 $\delta_n = \frac{X_n}{n} - \left(\frac{X_n}{n}\right)^2$. 由 Delta Method 可知:

$$\left[g\left(\frac{X_n}{n}\right) - g(p)\right] \xrightarrow{D} N\left(0, \frac{p(1-p)}{n} * [g'(p)]^2\right) \quad \text{as } n \to \infty$$

即 $\delta_n = g\left(\frac{X_n}{n}\right) \xrightarrow{D} N\left(p(1-p), \frac{1}{n} p(1-p)(1-2p)^2\right)$ as $n \to \infty$

5 Problem 5: Central Limit Theorem & Delta Method

Suppose $X_1, X_2, ...$ are iid with mean μ and finite variance σ^2, and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. What is the limiting distribution of $\bar{X}_n^2$?

解: Already known $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

Therefore $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ has $E[\bar{X}] = \mu$, $Var(\bar{X}) = \frac{\sigma^2}{n}$

By the CLT we have $\dfrac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} \xrightarrow{D} N(0,1)$ as $n \to \infty$

- $g(x) = x^2$, $g'(x) = 2x$, $g''(x) = 2$.

① If $g'(\mu) = 2\mu \neq 0$, then by the first order delta method:

$$\sqrt{n}\left[g(\bar{X}_n) - g(\mu)\right] \longrightarrow N\left(0, \sigma^2 g'(\mu)^2\right) \quad \Rightarrow \quad \bar{X}_n^2 \xrightarrow{D} N\left(\mu^2, \frac{4\sigma^2\mu^2}{n}\right) \quad \text{as } n \to \infty$$

② If $g'(\mu) = 2\mu = 0$, then by the second order delta method:

$$n\left[g(\bar{X}_n) - g(\mu)\right] \longrightarrow \frac{\sigma^2}{2} g''(\mu) \cdot \chi_1^2 \quad \Rightarrow \quad \bar{X}_n^2 - \mu^2 \xrightarrow{D} \frac{\sigma^2}{n}\chi_1^2 \quad \text{as } n \to \infty$$

6 Problem 6: Confidence intervals

Suppose that a clinical trial was conducted to estimate the response rate p of an experimental drug. We observed 13 responses among 65 subjects treated by this drug. Based on the binomial distribution, the estimated response rate is $\hat{p} = 0.2$. Construct the 95% confidence interval for p using the asymptotic normal approximation.

解: Suppose $X_1, \dots, X_n \overset{iid}{\sim}$ Bernoulli (p)

Then $f(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}$ Thus $L(\tilde{x}; p) = \sum_{i=1}^{n} x_i \log p + (1-x_i)\log(1-p)$

$\Rightarrow \hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow$ In this case, $\hat{p} = \frac{13}{65} = 0.2$

By the asymptotic normal approximation, 95% confidence interval is

$$p \in \left[\hat{p} \pm z_{\alpha/2}\sqrt{\frac{1}{n}\hat{p}(1-\hat{p})}\right] \Rightarrow p \in [0.1028, 0.2972]$$

》 Rcodes

```
# Read in Data
setwd("./Desktop/MyRcodes")
data <- read.table("./gamma-arrivals.txt")
#data <- as.numeric(data)

# Make a histogram of arrival time
hist(data[,], main = "histogram of number of gamma-arrivals", prob=TRUE, right=FALSE, xlim=c(0, 800), ylim = c(0, 0.012))

# Compute the MME of data
sample_mean = mean(data[,])
sample_variance = var(data[,])
mme_a = sample_mean**2 / sample_variance
mme_b = sample_mean / sample_variance

# compute the MLE of data
X <- data[,]
n <- length(X)
negloglike <- function(param){
  param[2]*sum(X) + n*log(gamma(param[1])) - n*param[1]*log(param[2]) - (param[1]-1)*sum(log(X))
}
out <- nlm(negloglike, p = c(0.5, 0.5), hessian = TRUE)
mle_a <- out$estimate[1]
mle_b <- out$estimate[2]

# plot gamma(alpha=shape, beta=rate) distribution
x <- seq(0, 800, 0.1)
lines(x, dgamma(x, shape=mme_a, rate = mme_b), lty = 2, col = "blue") # MME.dist
lines(x, dgamma(x, shape=mle_a, rate=mle_b), lty = 4, col = "red") # MLE.dist
```