Stat 135
07/01/2021

Lab 4 Solution

1. Data Cleaning in R

Do the following tasks:
1) Load beeswax into R as a dataframe called bees.
2) Examine the structure of the bees using command str(bees). Notice that the data is factor data with the numbers as strings. This is terrible to work with. Best to convert to a numeric vector.
3) Give the columns names: meltingpoint and hydrocarbon
4) Switch the order of the columns
5) Keep just meltingpoints > 63 and hydrocarbons > 14 using filter()
6) Use mutate() to multiply every value of hydrocarbon by 2.

```
bees <- read.csv("Your_local_directory/beeswax.txt",header = TRUE)
head(bees)
bees <- cbind(bees[,1],bees[,2])
bees <- as.data.frame(bees)
colnames(bees)=c("meltingpoint","hydrocarbon")
bees <- bees[,c(2,1)]
head(bees)
bees <- bees %>% filter(meltingpoint>63,hydrocarbon>14) %>% mutate(hydrocarbon=hydrocarbon*2)
head(bees)
```

2. Delta Method

We need to first find the asymptotic distribution $g(\bar{X}_n) = \sqrt{\bar{X}_n}$.

Since $\sqrt{n}(\bar{X}_n - \mu) \to N(0, \sigma^2)$ as $n \to \infty$,

it follows that $\sqrt{n}\left(g(\bar{X}_n) - g(\mu)\right) \to g'(\mu) N(0, \sigma^2)$.

Note $g'(\mu) = \frac{1}{2\sqrt{\mu}}$. Thus,

$$E[g(\bar{X}_n)] \approx g(\mu)$$

and $\mathrm{var}[g(\bar{X}_n)] \approx \frac{[g'(\mu)]^2 \sigma^2}{n} = \frac{\sigma^2}{4n\mu}$.

3. Rice 8.10.19

Solution:

1. The log of the normal density is

$$\log f(x|\mu,\sigma) = \log\frac{1}{\sqrt{2\pi}} - \log\sigma - \frac{(x-\mu)^2}{2\sigma^2}$$

So the log likelihood function is

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

a. If μ is a known constant then all you have to do is differentiate the log likelihood with respect to σ, set equal to 0, and solve:

$$-\frac{n}{\hat\sigma} + \frac{\sum_{i=1}^n (X_i-\mu)^2}{\hat\sigma^3} = 0$$

so $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$, no big surprise. Take the square root to get the MLE of σ.

b. This time treat σ as the constant and differentiate the log likelihood with respect to μ:

$$\frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\hat\mu) = 0$$

so $\hat\mu = \bar X$, again no big surprise.

c. We know that $\hat\mu$ is unbiased and has variance σ^2/n. Now

$$\frac{d}{d\mu}[\log\frac{1}{\sqrt{2\pi}} - \log\sigma - \frac{(x-\mu)^2}{2\sigma^2}] = \frac{x-\mu}{\sigma^2}$$

So the Fisher information is

$$I(\mu) = E[(\frac{X-\mu}{\sigma^2})^2] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

The Cramer-Rao bound says that no unbiased estimate has variance less than $1/nI(\mu) = \sigma^2/n = Var(\hat\mu)$. So $\hat\mu$ has the smallest variance among all unbiased estimates.

4. Rice 8.10.21

2a. Notice that the density is that of $T+\theta$ where T has the exponential density with parameter 1. Therefore the first moment of the density is $\mu_1 = E(T)+\theta = 1+\theta$, and therefore $\theta = \mu_1 - 1$. Therefore the MOM estimate is $\hat\theta_{MOM} = \bar X - 1$.

b. It is important to notice that with probability 1, $\theta \le \min(X_1, X_2, \ldots, X_n)$. So the likelihood function is

$$e^{-\sum_{i=1}^n (X_i-\theta)} = e^{-n\bar X}\cdot e^{n\theta}$$

for $\theta \le \min(X_1, X_2, \ldots, X_n)$. This is an increasing function of θ so there is no need to differentiate it to find its maximum. The function is maximized by the maximum possible value of θ, which is $\min(X_1, X_2, \ldots, X_n)$ by our earlier observation. So $\hat\theta_{MLE} = \min(X_1, X_2, \ldots, X_n)$.

c. The sample mean is a sufficient statistic by the factorization theorem (h(x)=1).