Tutorial 3

Rui Qiu

2018-03-06

Q1

2.6

In the case of a random sample $X_1, X_2, \ldots, X_n$ from the Bernoulli distribution with probability function

$$f(x;\theta) = \theta^x (1-\theta)^{1-x}, x = 0, 1, 0 \leq \theta \leq 1,$$

find the Cramér-Rao lower bound.

Solution:

$$\begin{aligned}
l(x|\theta) &= \log f(x|\theta) = x \log \theta + (1-x)\log(1-\theta) \\
l'(x|\theta) &= \frac{x}{\theta} + \frac{1-x}{1-\theta} \\
l''(x|\theta) &= -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}
\end{aligned}$$

We know that $E(X) = \theta$, then

$$\begin{aligned}
I(x|\theta) &= -E[l''(x|\theta)] = \frac{E(x)}{\theta^2} + \frac{1-E(X)}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)} \\
I_n(\theta) &= n \cdot I(x|\theta) = \frac{n}{\theta(1-\theta)}
\end{aligned}$$

Therefore, the Cramér-Rao lower bound is $I_n^{-1}(\theta) = \frac{\theta(1-\theta)}{n}$.

Similarly, for estimator of θ^2.

$$\begin{aligned}
l''(\theta^2, x) &= \frac{\theta(3x+1) - 2\theta^2 - 2x}{4(\theta-1)^2\theta^4} \\
I(\theta) &= -E\left[l''(\theta, x)\right] = -\left[\frac{\theta(3\theta+1) - 2\theta^2 - 2\theta}{4(\theta-1)^2\theta^4}\right] \\
&= \frac{1}{4(\theta-1)\theta^3} \\
I_n(\theta) &= \frac{n}{4(\theta-1)\theta^3}
\end{aligned}$$

So the C-R LB is $\frac{4(\theta-1)\theta^3}{n}$.

2.8

Suppose $X_1, X_2, \ldots, X_n$ form a random sample from the normal distribution with unknown variance σ^2. Show that the sample variance

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

does not attain the Cramér-Rao lower bound for finite n, but does so as n tends to infinity. For what value of c does the estimator

$$c\sum_{i=1}^{n} (X_i - \bar{X})^2$$

of σ^2 have the smallest MSE?

Solution:

$$\begin{aligned}
f(x|\theta) &= \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(x-\mu)^2}{2\theta}\right) \\
l(x|\theta) &= -\frac{(x-\mu)^2}{2\theta} - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\theta \\
l'(x|\theta) &= \frac{(x-\mu)^2}{2\theta^2} - \frac{1}{2\theta} \\
l''(x|\theta) &= -\frac{(x-\mu)^2}{\theta^3} + \frac{1}{2\theta^2} \\
I(\theta) &= -E[l''(x|\theta)] = -E\left[-\frac{(X-\mu)^2}{\theta^3} + \frac{1}{2\theta}\right] = \frac{1}{2\theta^2} \\
I_n(\theta) &= nI(\theta) = \frac{n}{2\theta^2}
\end{aligned}$$

The Cramér-Rao lower bound is $\frac{2\theta^2}{n}$. The sample variance is $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$$\begin{aligned}
\frac{(n-1)S^2}{\theta} &\sim \chi^2_{n-1} \\
V\left(\frac{n-1}{\theta}S^2\right) &= \frac{(n-1)^2}{\theta^2}V(S^2) = 2(n-1) \\
V(S^2) &= \frac{2\theta^2}{n-1} > \frac{2\theta^2}{n}
\end{aligned}$$

But when $n \to \infty$, the sample variance attains the lower bound.

Then for the estimator $c\sum_{i=1}^{n}(X_i - \bar{X})^2$, we have

$$\begin{aligned}
MSE &= E\left[c\sum_{i=1}^{n}(X_i - \bar{X})^2 - \sigma^2\right]^2 + V\left[c\sum_{i=1}^{n}(X_i - \bar{X})^2\right] \\
&= (c(n-1)-1)^2(\sigma^2)^2 + 2c^2\sigma^4(n-1) \\
&= \sigma^4((n^2-1)c^2 - 2(n-1)c + 1)
\end{aligned}$$

Hence $c = \frac{1}{n+1}$ is a minimizer of MSE.

2.10

Suppose that $X_1, X_2, \ldots, X_n$ form a random sample from $N(\theta, \sigma^2)$ where σ^2 is known. Use Lemma 2.1 to show that $I_\theta = n/\sigma^2$.

Lemma 2.1: Under the same regularity conditions as for the Cramér-Rao inequaility,

$$I_\theta = -E\left[\frac{d^2 l}{d\theta^2}\right]$$

Solution:

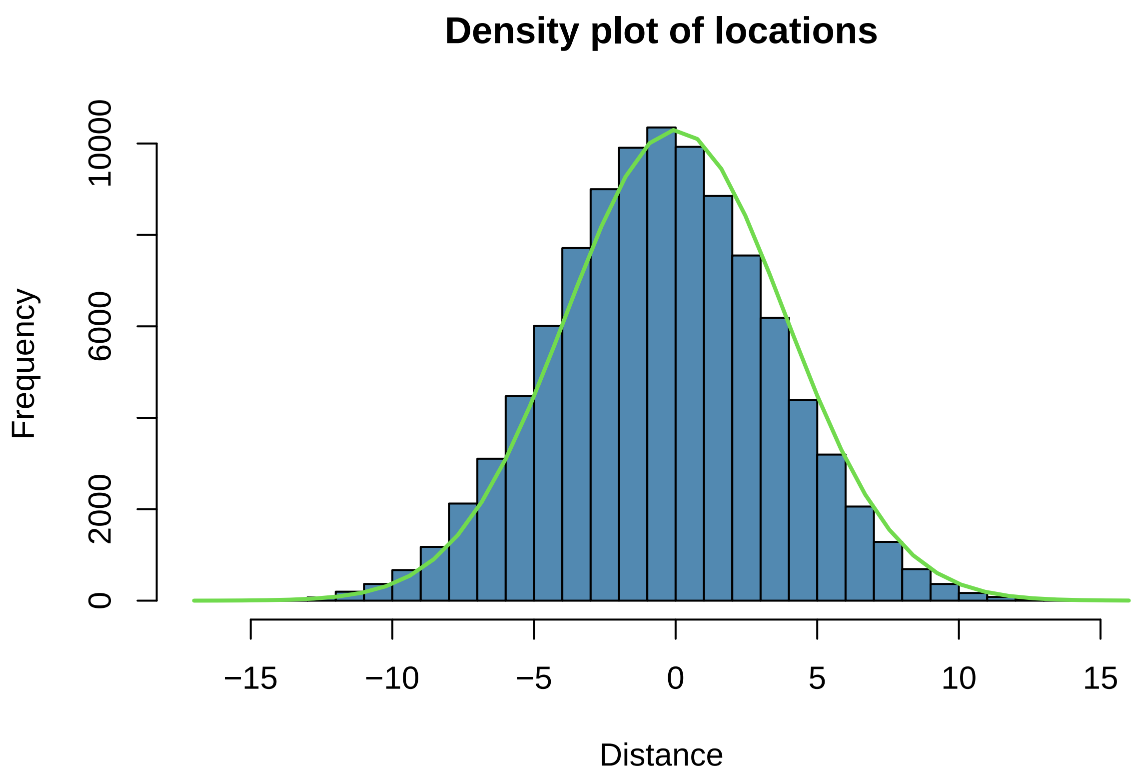
$$\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\theta)^2}{2\sigma}\right) \\
l(x) &= -\frac{(x-\theta)^2}{2\sigma^2} - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma \\
\frac{dl}{d\theta} &= -\frac{1}{2\sigma^2}(\theta^2 - 2x) \\
&= -\frac{1}{\sigma^2}(\theta - x) \\
\frac{d^2 l}{d\theta^2} &= -\frac{1}{\sigma^2} \\
I_\theta &= \frac{n}{\sigma^2}
\end{aligned}$$

Q2

A drunkard executes a "random walk" in the following way: Each minute he takes a step north or south, with probability 1/2 each, and his successive step directions are independent. His step length is 50 cm. Use the central limit theorem to approximate the probability distribution of his location after 1 hour. Where is he most likely to be? Can you also code this in R?

Solution:

```r
set.seed(8027)
sim = 100000
loc = rep(NA, sim)
for (i in 1:sim) {
    loc[i] <- sum(sample(c(-0.5,0.5),60,replace = T))
}
h<-hist(loc, breaks=25, col="#5289B1", xlab="Distance",
        main="Density plot of locations")
xfit<-seq(min(loc),max(loc),length=40)
yfit<-dnorm(xfit,mean=mean(loc),sd=sd(loc))
yfit <- yfit*diff(h$mids[1:2])*length(loc)
lines(xfit, yfit, col="#71DA4E", lwd=2)
```




```
summary(loc)
```

```
##       Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -17.00000  -3.00000   0.00000   0.00443   3.00000  16.00000
```

Most likely still at the starting point.

Q3

Use the Monte Carlo method with $n = 100$ and $n = 1000$ to estimate:

$$\int_0^1 \cos(2\pi x)dx$$

Compare it with the exact answer.

```
for (n in c(100, 1000, 10000, 100000)) {
    x <- runif(n, 0, 1)
    cat("MC estimate with n=", n, "is", 1/n*sum(cos(2*pi*x)), "\n")
}
```

```
## MC estimate with n= 100 is -0.09874592
## MC estimate with n= 1000 is 0.01007239
## MC estimate with n= 10000 is 0.001567561
## MC estimate with n= 1e+05 is -0.0005153421
```

Analytic solution:

$$\int_0^1 \cos(2\pi x)dx = \frac{\sin(2\pi x)}{2\pi}\bigg|_0^1 = \frac{1}{2\pi} \cdot \cdot 0 = 0$$

Q4

Use the Monte Carlo method with $n = 100$ and $n = 1000$ to estimate:

$$\int_0^1 \cos(2\pi x^2)dx$$

No exact answer (i.e. closed form analytical solution) exists.

```
for (n in c(100, 1000, 10000, 100000)) {
    x <- runif(n, 0, 1)
    cat("MC estimate with n=", n, "is", 1/n*sum(cos(2*pi*x^2)), "\n")
}
```

```
## MC estimate with n= 100 is 0.2975664
## MC estimate with n= 1000 is 0.2498523
## MC estimate with n= 10000 is 0.2436607
## MC estimate with n= 1e+05 is 0.2402221
```