Estimation Methods Application Solutions

Lt Col Ken Horton   Professor Bradley Warner

30 June, 2020

Exercises

1. In the Notes, we found that if we take a sample from the uniform distribution $\mathsf{Unif}(0, \theta)$, the method of moments estimate of θ is $\hat{\theta}_{MoM} = 2\bar{x}$. Suppose our sample consists of the following values:

$$0.2 \quad 0.9 \quad 1.9 \quad 2.2 \quad 4.7 \quad 5.1$$

a) What is $\hat{\theta}_{MoM}$ for this sample?

```
x<-c(0.2,0.9,1.9,2.2,4.7,5.1)
thetamom<-2*mean(x)
thetamom
```

```
## [1] 5
```

b) What is wrong with this estimate?

The estimate for θ is 5, which is an impossible value for θ, since one of our observations (5.1) is beyond this value.

c) Show that this estimator is unbiased.

We need to show that
$$E\left(\hat{\theta}_{MoM}\right) = \theta$$
We proceed as follows
$$E\left(2\bar{X}\right) = 2E\left(\sum \frac{X_i}{n}\right) = \frac{2}{n}E\left(\sum X_i\right)$$
$$= \frac{2}{n}\sum E\left(X_i\right) = \frac{2}{n}\sum \frac{\theta}{2} = \frac{n\theta}{n} = \theta$$
Notice that in performing this derivation, we treated X as a random variable and not as x a data value.

d) ADVANCED: Use simulation in R to find out how often this happens ($\hat{\theta}_{MoM} < \max x$). Report an answer for various sizes of samples.

Let's start by writing code for one sample size and then generalize.

1

```
check <- function(n=10){
  temp<-runif(n,max=5)
  2*mean(temp)<max(temp)
}
```

```
set.seed(3030)
check(10)
```

```
## [1] FALSE
```

Now let's repeat

```
(do(10000)*check(10)) %>%
  summarize(mean(check)) %>%
  pull()
```
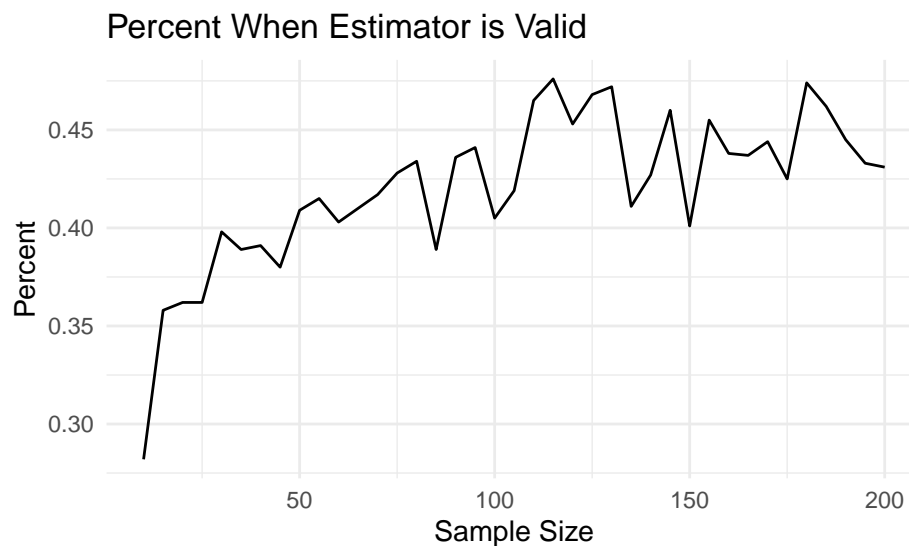
```
## [1] 0.285
```

Let's make `check` a vectorized function

```
check <- Vectorize(check)
```

```
my_data<-(do(1000)*check(seq(10,200,5))) %>%
  summarise_all(mean) %>%
  pivot_longer(everything(),names_to = "Sample",values_to = "Percent") %>%
  mutate(Sample=seq(10,200,5))
```

```
my_data %>%
  gf_line(Percent~Sample,xlab="Sample Size",title="Percent When Estimator is Valid") %>%
  gf_theme(theme = theme_minimal())
```
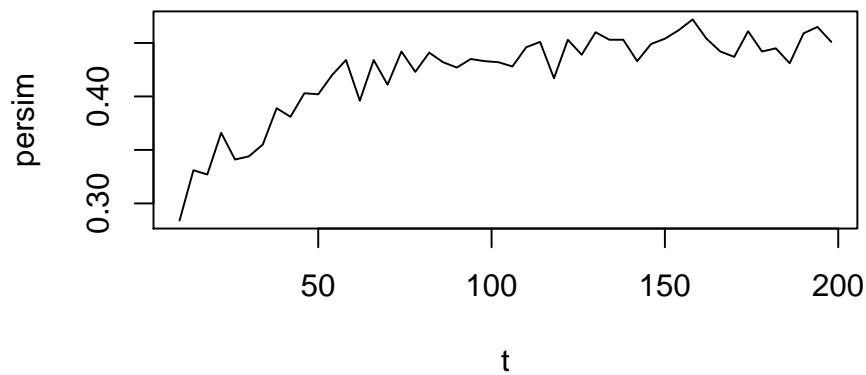

Here is more traditional, old school `R` code.

```
simn<-function(n){
  y<-replicate(1000,{
    x<-runif(n)
    (2*mean(x))<max(x)
  })
  mean(y)
}

t<-seq(10,200,4)
persim<-sapply(t,simn)
plot(t,persim,type="l")
```


2. Let $x_1, x_2, ..., x_n$ be a simple random sample from an exponentially distributed population with parameter λ. Find $\hat{\lambda}_{MoM}$.

Recall that $E(X) = \frac{1}{\lambda}$. Thus,

$$\hat{\lambda}_{MoM} = \frac{1}{\bar{x}}$$

3. Let $x_1, x_2, ..., x_n$ be an iid random sample from an exponentially distributed population with parameter λ. Find $\hat{\lambda}_{MLE}$.

Recall that

$$f_X(x; \lambda) = \lambda e^{-\lambda x}$$

So the likelihood function is:

$$L(\lambda; \boldsymbol{x}) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}$$

And the log-likelihood function is:

$$l(\lambda; \boldsymbol{x}) = n \log \lambda - \lambda \sum x_i$$

Taking the derivative with respect to λ and setting to 0:

$$\frac{\mathrm{d}l(\lambda; \boldsymbol{x})}{\mathrm{d}\lambda} = \frac{n}{\lambda} - \sum x_i = 0$$

Note that $\frac{\mathrm{d}^2 l(\lambda; \boldsymbol{x})}{\mathrm{d}\lambda^2} = -\frac{n}{\lambda^2} < 0$, so l is always concave down. Thus, any optimum found is a maximum.

So,

$$\hat{\lambda}_{MLE} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

4. It is mathematically difficult to determine if the estimators found in questions 2 and 3 are unbiased. Since the sum is in the denominator, mathematically we may have to work with the joint pdf. So instead, use simulation to get an sense of whether the method of moments estimator for the exponential distribution is unbaised.
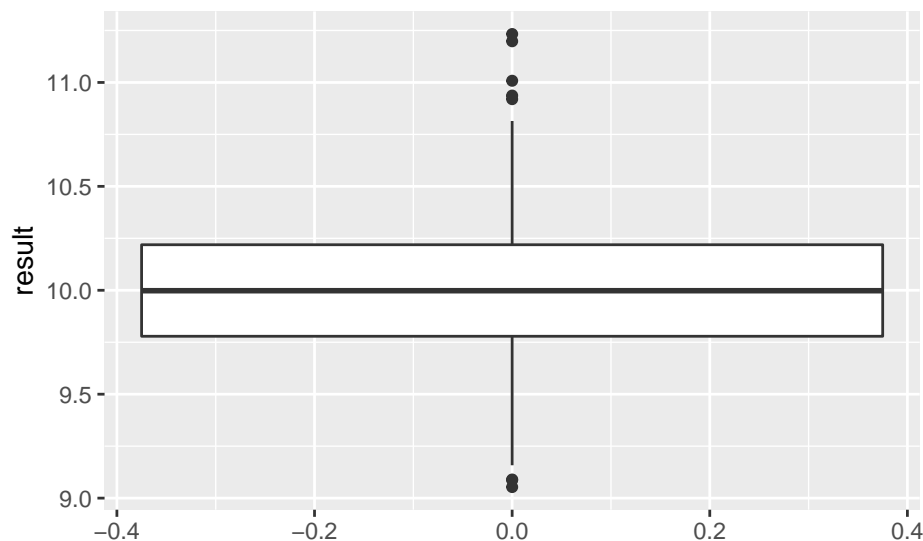
We need to sample data from an exponential and then compare the the reciprocal of the mean to the parameter.

```
set.seed(630)
1/mean(rexp(1000,rate=10))
```

```
## [1] 10.09422
```

This is close, maybe we just got lucky. Let's repeat many times.

```
(do(1000)*(1/mean(rexp(1000,rate=10)))) %>%
  gf_boxplot(~result)
```


Looks like it has the potential to be unbiased. We would need to investigate other values for λ.

5. Find a maximum likelihood estimator for θ when $X \sim \mathsf{Unif}(0, \theta)$. Compare this to the method of moments estimator we found. Hint: Do not take the derivative of the likelihood function.

$$L(\theta; \boldsymbol{x}) = \frac{1}{\theta^n}, \quad \text{only if all } x_i \leq \theta$$

Recall that $f(x_i; \theta) = \frac{1}{\theta}$ if $x_i \in [0, \theta]$ and 0 elsewhere. And since the likelihood function is simply the product of these pdfs, if any x_i is beyond θ, then the entire likelihood function is 0.

You can picture L as a decreasing function of θ, but remembering that L takes the value 0 if θ is smaller than at least one x_i. Thus, L achieves its maximum at $\theta = \max x_i$.

This estimate is more intuitive than the method of moments estimate $(2\bar{x})$. The method of moments estimate is sometimes not feasible. Meanwhile, the MLE $(\max x_i)$ is always feasible.

File Creation Information

- File creation date: 2020-06-30
- Windows version: Windows 10 x64 (build 17763)
- R version 3.6.3 (2020-02-29)
- `mosaic` package version: 1.6.0
- `tidyverse` package version: 1.3.0
- `fastR2` package version: 1.2.1