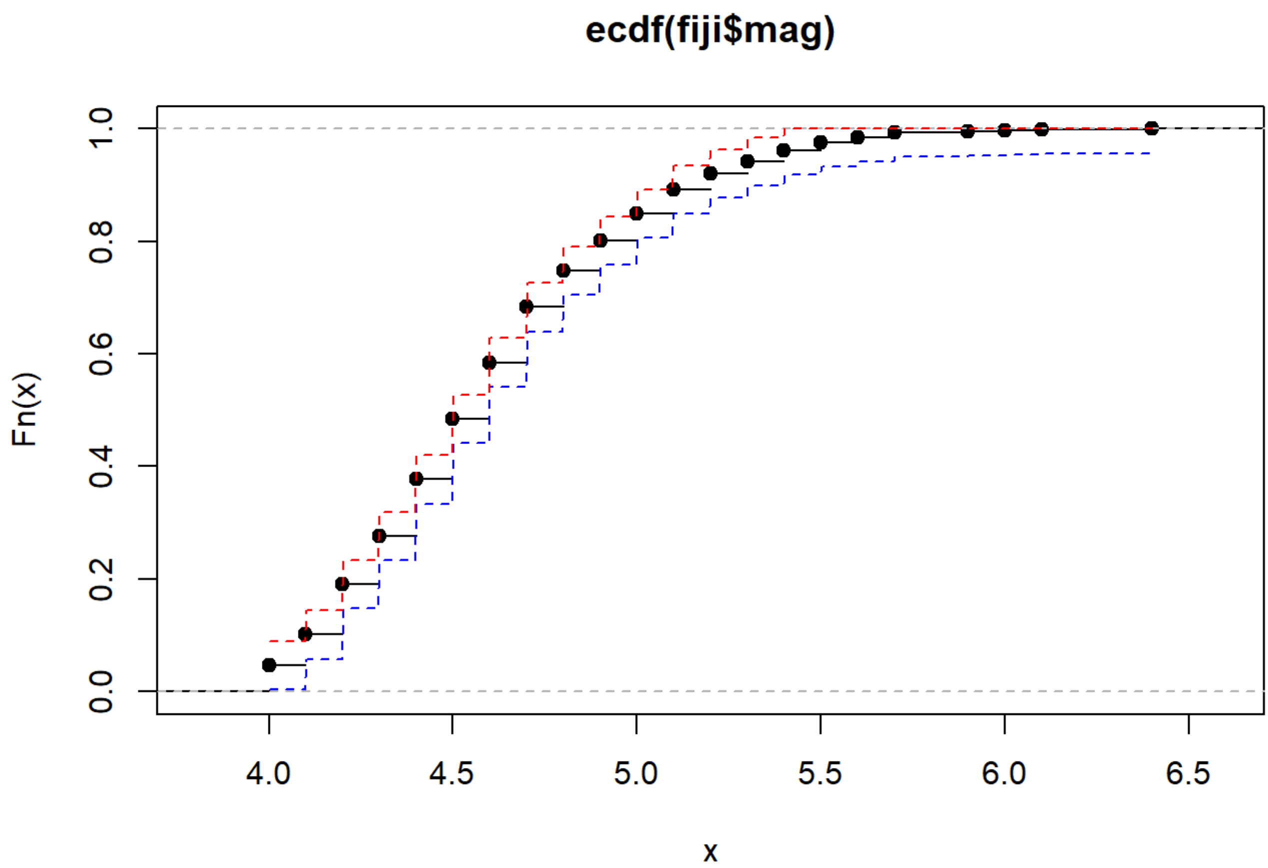
MA677 Assignment

Zhitian Liu

Fiji earthquakes

Data on the magnitudes of earthquakes near Fiji are on Blackboard under the Class-21 tab. Estimate the cdf F(x). Compute and plot a 95% confidence envelope for F. Find an approximate 95 percent confidence interval for F(4.9) − F(4.3).

```
fiji=read.csv("C:/Users/Lenovo/Desktop/fijiquakes.csv")
#The CDF of magnitude
CDF<-ecdf(fiji$mag)
plot(CDF)
#  95% confidence envelope for F
alpha <- 0.05
n <- length(fiji$mag)
epsn<-sqrt(log(2/alpha)/(2*n))
r<-max(fiji$mag)-min(fiji$mag)
grid<-seq(from=min(fiji$mag),to=max(fiji$mag),l=1000)
low.cdf<-pmax(CDF(grid)-epsn,0)
up.cdf<-pmin(CDF(grid)+epsn,1)
lines(grid,low.cdf,col="blue",lty=2)
lines(grid,up.cdf,col="red",,lty=2)
```



```
# use wilson method to calculate the confidence interval for F(4.9) - F(4.3)
a=ifelse((fiji$mag<=4.9) & (fiji$mag>4.3),1,0)
num_intherange<-sum(a)
binconf(num_intherange,length(fiji$mag),method="wilson",alpha)
```

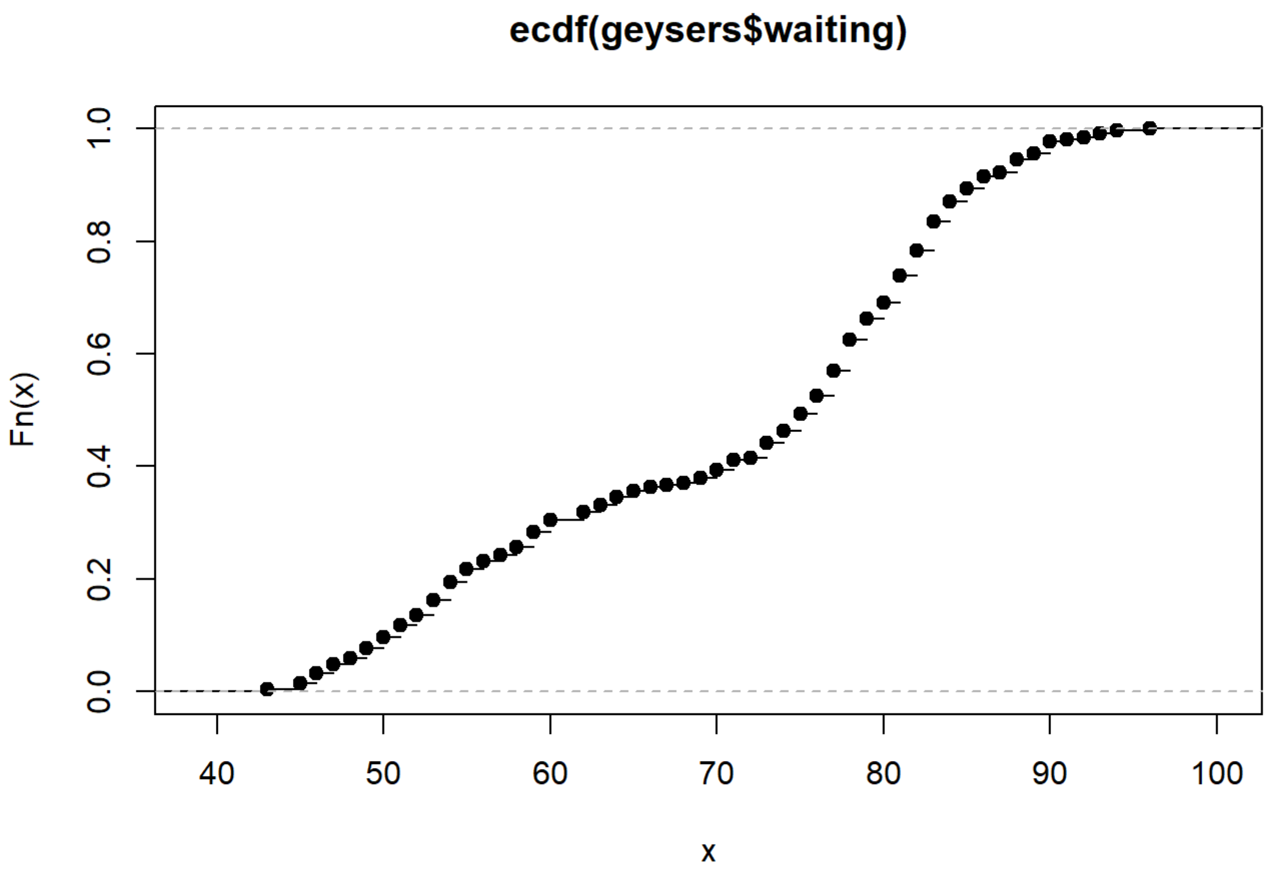
```
##  PointEst     Lower     Upper
##     0.526 0.4950118 0.5567892
```

```
#Hence the 95% confidence interval for F(4.9) - F(4.3) would be (0.495,0.557)
```

Old Faithful

Data on eruption times and waiting times between eruptions of the old faithful geyser (located in Yellowstone National Park) are posted on Blackboard under the Class-21 tab. Estimate the mean waiting time and compute a standard error for the estimate. Also, calculate a 90 percent confidence interval for the mean waiting time. Finally, estimate the median waiting time. We will use this calculation of the median in our discussions next week.

```
geysers=read.csv("C:/Users/Lenovo/Desktop/geysers.csv")
CDF=ecdf(geysers$waiting)
plot(CDF)
```



```
wt=geysers$waiting
#mean waiting time
mean_wt=mean(wt)
#standard error of the estimate
SEM=sd(wt)/sqrt(length(wt))
#90 percent confidence interval for the mean waiting time
low=mean_wt-qnorm(0.95)*SEM
up=mean_wt+qnorm(0.95)*SEM
#the median waiting time
mid=median(wt)
#print
paste("the mean of the waiting time is ",mean_wt )
```

```
## [1] "the mean of the waiting time is  70.8970588235294"
```

```
paste("the standard error of the estimate waiting time is ",SEM )
```

```
## [1] "the standard error of the estimate waiting time is  0.824316366377517"
```

```
paste("the median of the waiting time is ",mid )
```

```
## [1] "the median of the waiting time is  76"
```

```
paste("the 90 percent confidence interval for the mean waiting time is from",low,"to",up)
```

```
## [1] "the 90 percent confidence interval for the mean waiting time is from 69.5411790585379 t
o 72.2529385885209"
```

KS problem

Use the Kolmogorov-Smirnov test to test the hypothesis that the 25 values in the table below form a random sample from the uniform distribution on the interval [0, 1]. 0.42 0.06 0.88 0.40 0.90 0.38 0.78 0.71 0.57 0.66 0.48 0.35 0.16 0.22 0.08 0.11 0.29 0.79 0.75 0.82 0.30 0.23 0.01 0.41 0.09 Using the table above, test the hypothesis that the 25 values are a random sample from a continuous distribution with pdf:……

```
x=c(0.42,0.06,0.88,0.40,0.90,
0.38,0.78,0.71,0.57,0.66,
0.48,0.35,0.16,0.22,0.08,
0.11,0.29,0.79,0.75,0.82,
0.30,0.23,0.01,0.41,0.09)
ks.test(x,punif)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.18, p-value = 0.3501
## alternative hypothesis: two-sided
```

Since the p-value of the One-sample Kolmogorov-Smirnov test is 0.35, we accept the null hypothesis, The sample is drawn from uniform distribution.

```
x=c(0.42,0.06,0.88,0.40,0.90,
0.38,0.78,0.71,0.57,0.66,
0.48,0.35,0.16,0.22,0.08,
0.11,0.29,0.79,0.75,0.82,
0.30,0.23,0.01,0.41,0.09)
#CDF of the given distribution
p=function(x){
  if(x<=0.5&x>=0) {
    1.5
  } else if (0.5<x&x<1 ) {
   0.5+0.5*x
  } else {
   0
  }
}
ks.test(x,p)
```

```
## Warning in if (x <= 0.5 & x >= 0) {: 条件的长度大于一，因此只能用其第一元素
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 1.5, p-value = 3.331e-16
## alternative hypothesis: two-sided
```

Since the p-value of the One-sample Kolmogorov-Smirnov test is small enough, we reject the null hypothesis, The sample isn't drawn from the given distribution.

Exercises 8.5.2

Let $X_1, X_2, X3....X_n\ Bernoulli(p)$ and let $Y_1, Y_2, Y3....Y_n\ Bernoulli(q)$. Find the plug-in estimator and estimated standard error for p. Find an approximate 90 percent confidence interval for p. Find the plug-in estimator and estimated standard error for p-q. Find an approximate 90 per cent confidence interval for p-q.

Answer: The plug in estimator For p is $\hat{p} = \overline{X}_n$. The plug in estimator for standard error of Bernoulli(p) is $\sqrt{\hat{p}(1-\hat{p})} = \sqrt{\overline{X}_n(1-\overline{X}_n)}$.

approximate 90 percent confidence interval for p is

$$\hat{p} + -z\hat{se} = \overline{X}_n + -z\sqrt{\overline{X}_n(1-\overline{X}_n)}$$

The plug in estimator for p-q is $\hat{\theta} = \hat{p} - \hat{q} = \overline{X}_n - \overline{Y}_n$. The plug in estimator for standard error of p-q is $se = \sqrt{V(\hat{p}-\hat{q})} = \sqrt{V(\hat{p}) + V(\hat{q})} = \sqrt{(\overline{X}_n(1-\overline{X}_n))^2 + \bar{(}Y_n(1-\overline{Y}_n))^2}$

approximate 90 per cent confidence interval for p-q is

$$\overline{X}_n - \overline{Y}_n + -z\sqrt{(\overline{X}_n(1-\overline{X}_n))^2 + \bar{(}Y_n(1-\overline{Y}_n))^2}$$

Exercises 8.5.4

Let x and y be two distinct points. Find $Cov(\hat{F_n}(x), \hat{F_n}(y))$

Answer:

$$
\begin{aligned}
Cov(\hat{F_n}(x), \hat{F_n}(y)) &= E(\hat{F_n}(x)\hat{F_n}(y)) - E(\hat{F_n}(x))E(\hat{F_n}(y)) \\
&= E(\hat{F_n}(x)\hat{F_n}(y)) - F(x)F(y)
\end{aligned}
$$

Then we have

$$
E(\hat{F_n}(x)\hat{F_n}(y)) = \frac{1}{n}F(\min\{x,y\}) + \left(1 - \frac{1}{n}\right)F(x)F(y)
$$

Assume $x < y$ $$

$$
\begin{aligned}
Cov(\hat{F_n}(x), \hat{F_n}(y)) &= E(\hat{F_n}(x)\hat{F_n}(y)) - E(\hat{F_n}(x))E(\hat{F_n}(y)) \\
&= E(\hat{F_n}(x)\hat{F_n}(y)) - F(x)F(y) \\
&= \frac{1}{n}F(\min\{x,y\}) + \left(1 - \frac{1}{n}\right)F(x)F(y) - F(x)F(y) \\
&= \frac{F(x)(1-F(y))}{n}
\end{aligned}
$$

$$