677_hw3

Maggie Sha

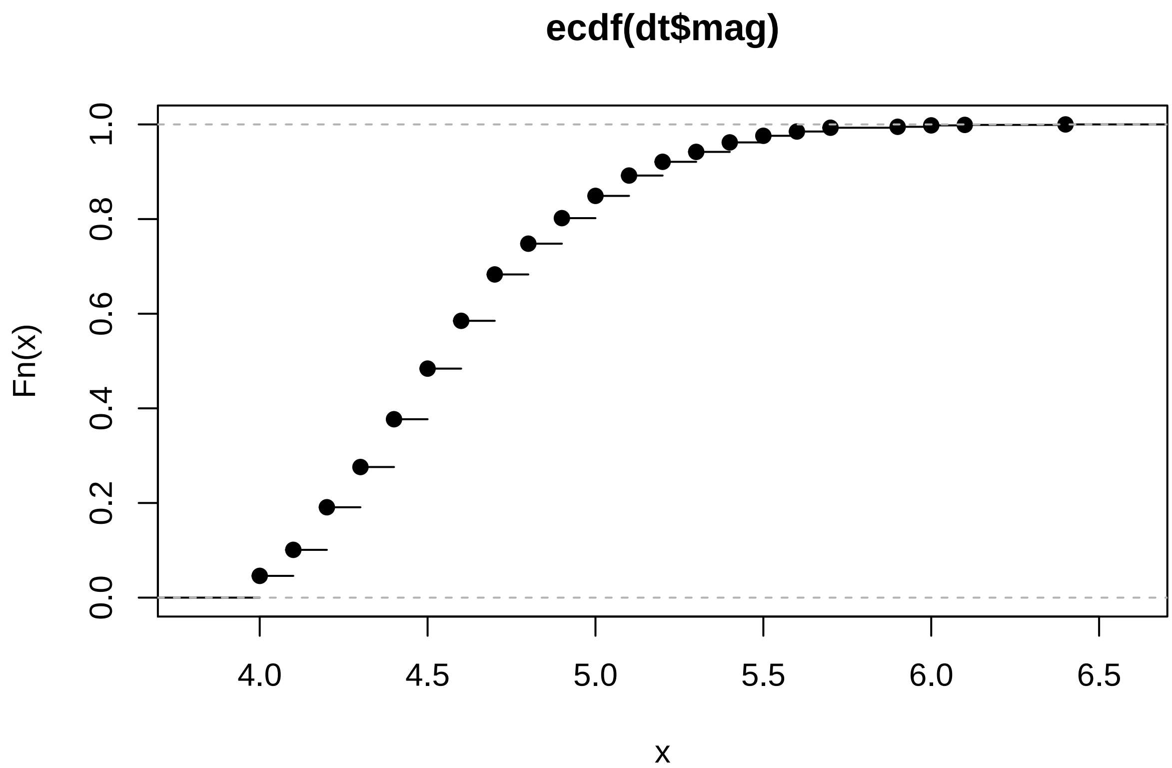
3/14/2021

Fiji earthquakes

```
options(digits=2, scipen=999)
dt <- read_table2("fijiquakes.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   obs = col_double(),
##   lat = col_double(),
##   long = col_double(),
##   depth = col_double(),
##   mag = col_double(),
##   stations = col_double()
## )
```

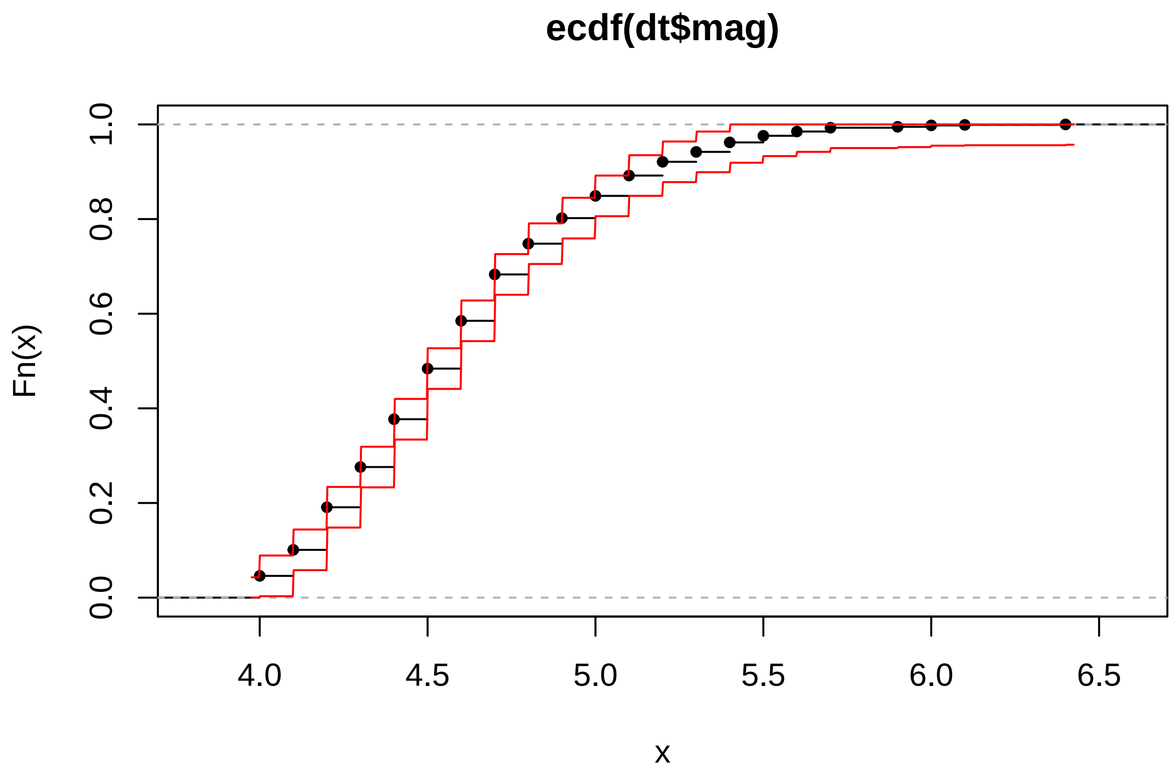
```
cdf <- ecdf(dt$mag)
plot(cdf)
```




```r
n <- length(dt$mag)
alpha =  .05

epsilon <- sqrt((1/(2*n))*log(2/alpha))
r <- max(dt$mag) - min(dt$mag)
x <- seq(min(dt$mag)-0.01*r, max(dt$mag)+0.01*r, length = n)
fx <- cdf(x)
L <- pmax(fx - epsilon, 0)
U <- pmin(fx + epsilon, 1)

plot.ecdf(cdf, pch=20)
lines(x, L, type="l", lty=1, col="red" )
lines(x, U, type="l", lty=1, col="red" )
```




```
## confidence interval
s = sum((dt$mag <= 4.9)&(dt$mag >= 4.3))
epsilon <- sqrt((1/(2*s))*log(2/alpha))
L4.9 <- pmax(cdf(4.9) - epsilon, 0)
L4.3 <- pmax(cdf(4.3) - epsilon, 0)
U4.9 <- pmin(cdf(4.9) + epsilon, 1)
U4.3 <- pmin(cdf(4.3) + epsilon, 1)
max <- U4.9 - L4.3
min <- L4.9 - U4.3
c(min,max)
```

```
## [1] 0.42 0.64
```

Old Faithful

```
dt2 <- read.csv("geysers.csv")
avg_time <- mean(dt2$waiting)
n = length(dt2$waiting)

sd <- sqrt((sum((dt2$waiting - avg_time)^2) / n))
sd #14
```

```
## [1] 14
```

```r
# z-score for 90% confidence interval is 1.645
n = length(dt2$waiting)
CI <- c((avg_time - 1.645*sd/sqrt(n)), (avg_time + 1.645*sd/sqrt(n)))
CI # 70 72
```

```
## [1] 70 72
```

```r
median(dt2$waiting) # 76
```

```
## [1] 76
```

KS problem

```r
table <- 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(table,"punif", 0, 1)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  table
## D = 0.2, p-value = 0.4
## alternative hypothesis: two-sided
```

```r
# One-sample Kolmogorov-Smirnov test
#
# data:  table
# D = 0.2, p-value = 0.4
# alternative hypothesis: two-sided

# P value is not significant at 5% significance-level,
# this table is not from a uniform distribution on the interval [0,1].


f = function(x){
  if(x <= 0.5 & x > 0){
    x = 1.5
  }
  else if (0.5 < x & x < 1){
    x = 0.5
  }
  else{
    x = 0
  }
  return (x)
```

```
}
ks.test(table, f)
```

```
## Warning in if (x <= 0.5 & x > 0) {: the condition has length > 1 and only the
## first element will be used
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  table
## D = 2, p-value = 0.0000000000000003
## alternative hypothesis: two-sided
```

```
#   One-sample Kolmogorov-Smirnov test
#
# data:  table
# D = 2, p-value = 0.0000000000000003
# alternative hypothesis: two-sided

# P value is significant at 5% significance-level,
# this table is from this continuous distribution
```

KS problem

$$E(\bar{X} - \bar{Y}) = \frac{1}{n} * \sum success_1 - \frac{1}{m} * \sum success_2 = p - q$$
$$The\ 90 percent\ C.I.\ is\ given\ by$$
$$((\bar{X} - \bar{Y}) - z * \sqrt{(}\frac{p * (1-p)}{n} + \frac{q * (1-q)}{m}, (\bar{X} - \bar{Y}) + z * \sqrt{(}\frac{p * (1-p)}{n} + \frac{q * (1-q)}{m})$$

Problem 4

$$We\ have\ \hat{F}_n(x)\ is\ a\ empirical\ distribution$$
$$So\ I(X_i) <= x) = 1\ if\ X_i <= x\ and\ 0\ if\ X_i > x$$

From the central limit theorem, for a large n, the limiting distribution has variance that goes to 0. So

$$I(X_i <= x) \simeq F(x)$$

for every x.

$$F_n(x) \simeq n^{-1} \sum_{i=1}^{n} F(x) = F(x)$$

So F is the limiting distribution of F_n.