STAT 703/J703
April 4th, 2005
-Lecture 23-

Instructor: Brian Habing
Department of Statistics
LeConte 203
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu

Today

Methods Based on the CDF

- The Empirical Distribution Function
- Some Statistical Properties
- Kolmogorov-Smirnov Test
- The Nonparametric Bootstrap
- Relation to Survival Functions

Recall that the definition of the cumulative distribution function (CDF) is:

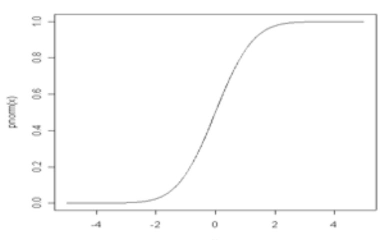
$$F_X(x)=P(X\leq x)$$

Note that:

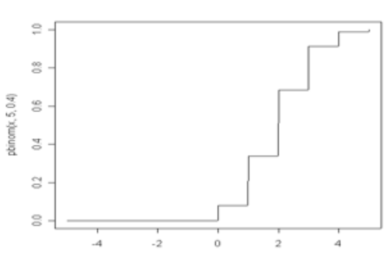
- $F_X(x)$ is non-decreasing
- $F_X(x) \to 1$ as $x\to\infty$
- $F_X(x) \to 0$ as $x\to-\infty$

The advantage of the CDF is that every random variable has one, and it has the same definition for both discrete and continuous random variables.


```
x<-(-500:500)/100
plot(x,pnorm(x),type="l")
```




```
plot(x,pbinom(x,5,.4),type="l")
```



The empirical distribution function (or empirical cumulative distribution function) is defined as:

$$F_n(x) = \frac{1}{n}\{\# x_i \le x\}$$

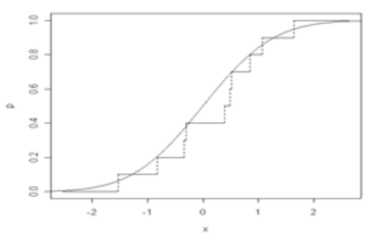
Unlike a histogram, there is only one way to plot an EDF.


```
edf<-function(y){
  x<-sort(y)
  plot(c(min(x)-1,max(x)+1),
    c(0,1),type="n",
    xlab="x",ylab="p")
 lines(c(x[1]-1,x[1]),c(0,0),
    lty=1)
 lines(c(x[1],x[1]),
    c(0,1/length(x)),lty=2)
```



```
for (i in 1:(length(x)-1)){
 lines(c(x[i],x[i+1]),
    c(i/length(x),
    i/length(x)),  lty=1)
 lines(c(x[i+1],x[i+1]),
    c(i/length(x),
    (i+1)/length(x)),lty=2)}
 lines(c(x[length(x)],
    x[length(x)]+1),c(1,1),
    lty=1) }
```

```
edf(rnorm(10))
lines(x,pnorm(x))
```



<u>Statistical properties of the EDF</u>

Note that we could write the EDF as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$$

This leads directly to the fact that

$F_n(x) \rightarrow F(x)$ as $n\rightarrow\infty$ for each x

With more theory we could prove that

$\sup_x|F_n(x) \rightarrow F(x)| \rightarrow 0$ as $n\rightarrow\infty$

The Kolmogorov-Smirnov test uses this quantity to construct a test of the null hypothesis that the data is drawn from a population with cdf F.

The test statistic is

$$\sup_x|F_n(x) \rightarrow F(x)|$$

The command in R is ks.test

```
ks.test(x,"pnorm",0,1)
```


It is interesting that the distribution of the Kolmogorov-Smirnov statistic does not depend on F !!!