EngSci 213 R Notebook (Currently: Lectures 5-9)

OJM

6 April 2016

What is this? R, RStudio, Markdown and R Markdown.

Here I'm using the statistical programming language 'R' (developed at Auckland Uni but used worldwide!). See e.g. https://www.r-project.org/.

I'm interacting with R in the RStudio IDE (https://www.rstudio.com/)

Note - you can just run R from a terminal if you want!

Within this I'm using the 'notebook' environment which allows you to write in 'R Markdown' format (http://rmarkdown.rstudio.com/). This allows you to combine 'markdown' and R code and compile to e.g. HTML, PDF, Word etc. (You can write R Markdown outside of RStudio too).

We can write latex equations e.g. $\mathbb{P}(X = 1) = 0.5$

Good R documentation: https://stat.ethz.ch/R-manual/R-devel/doc/html/

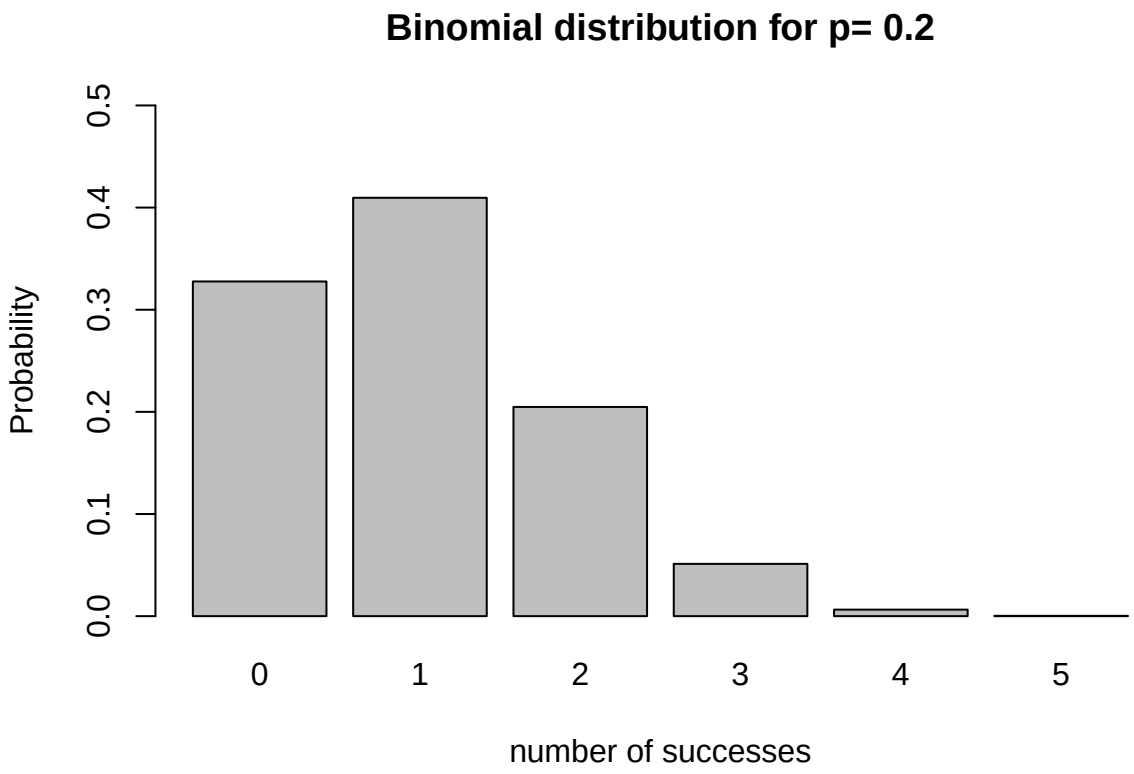
For more details on using R Markdown see http://rmarkdown.rstudio.com.

Discrete random variables

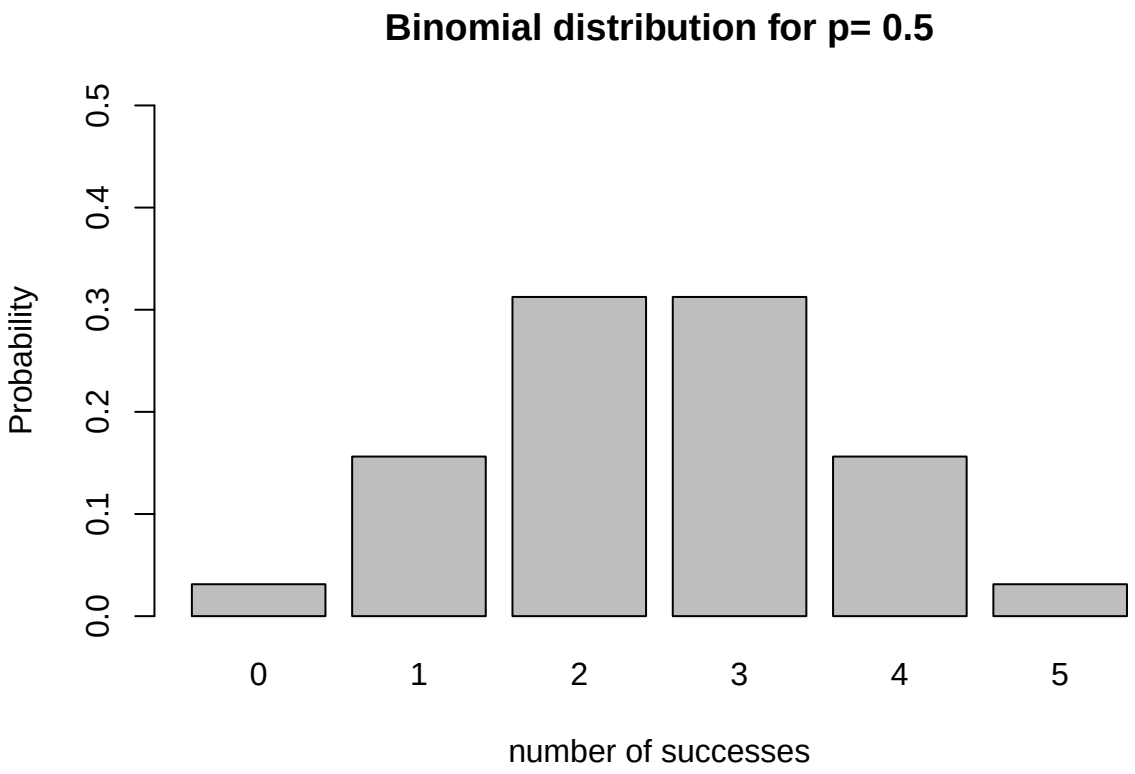
Binomial random variables

Here is an R code block to plot a binomial distribution (in terms of the probability mass function)

```
n=5
num_successes <- seq(0, n)
#try a low value...
p1 = 0.2
px1 <- dbinom(x=num_successes, size=n, p1)
# ...and a 'fair' value
p2 = 0.5
px2 <- dbinom(x=num_successes, size=n, p2)

barplot(px1, xlab="number of successes",ylab="Probability",
        main=paste("Binomial distribution for p=",
                   p1),ylim=c(0,0.5),names.arg = 0:n)
```

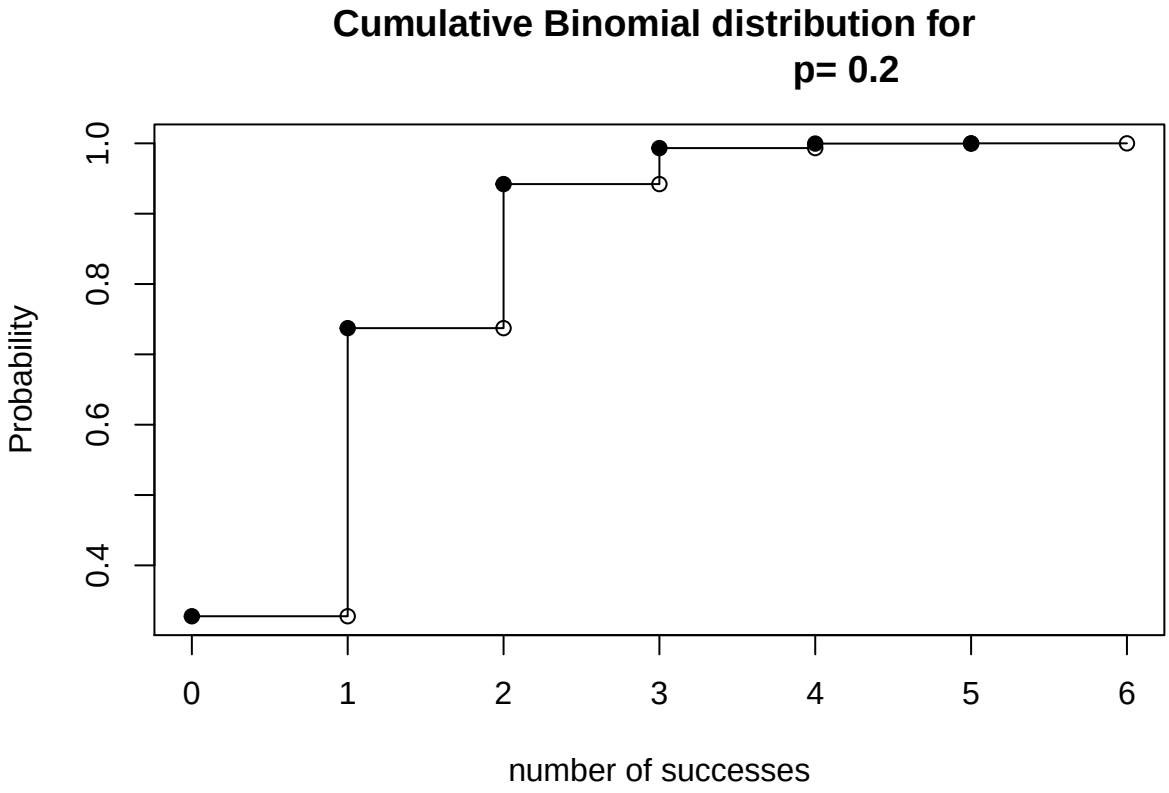



```
barplot(px2, xlab="number of successes",
  ylab="Probability",
  main=paste("Binomial distribution for p=",
             p2),ylim=c(0,0.5),names.arg = 0:n)
```

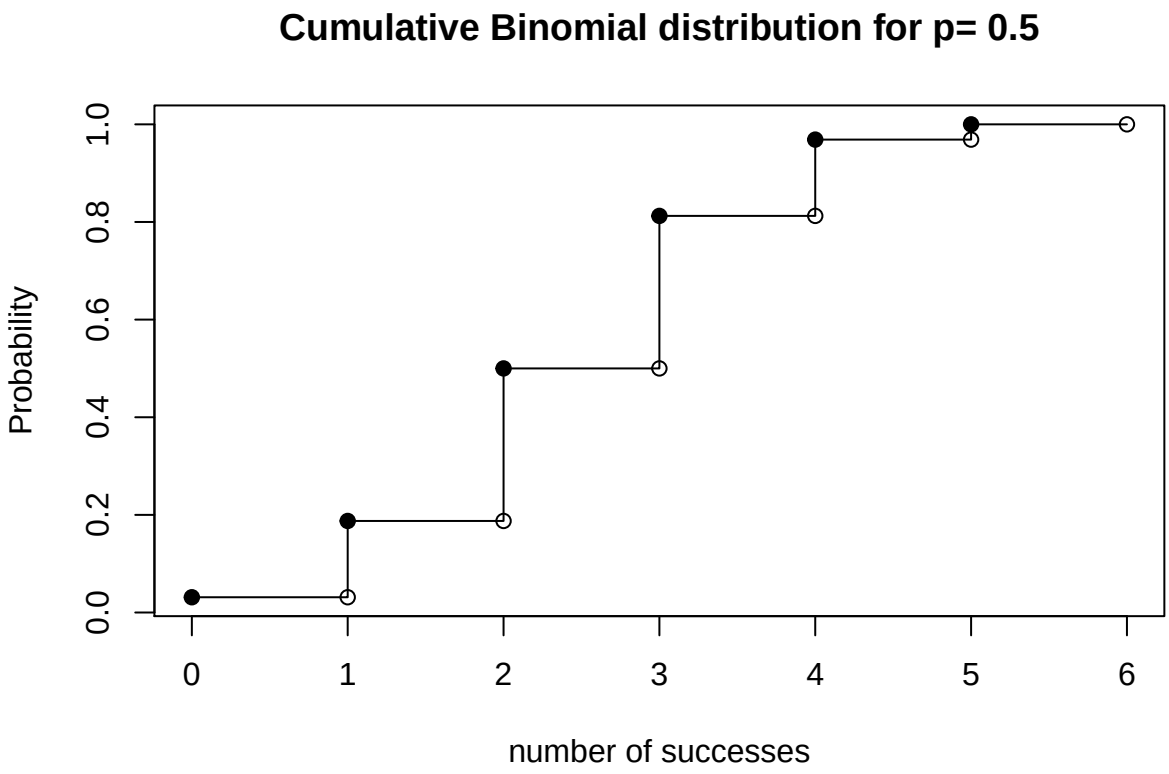


Here are the corresponding cumulative distributions

```
n=5
num_successes <- seq(0, n)
#try a low value...
p1 = 0.2
cpx1 <- pbinom(q=num_successes, size=n, p1)

# ...and a 'fair' value
p2 = 0.5
cpx2 <- pbinom(q=num_successes, size=n, p2)

plot(seq(0, n+1),c(cpx1,cpx1[n+1]), type="s",
     xlab="number of successes",
     ylab="Probability", main=paste("Cumulative Binomial distribution for
                                    p=", p1))
points(seq(0, n),cpx1,pch = 19)
points(seq(1, n+1),cpx1,pch = 21)
```




```
plot(seq(0, n+1),c(cpx2,cpx2[n+1]), type="s",
     xlab="number of successes",
     ylab="Probability",
     main=paste("Cumulative Binomial distribution for p=", p2))
points(seq(0, n),cpx2,pch = 19)
points(seq(1, n+1),cpx2,pch = 21)
```

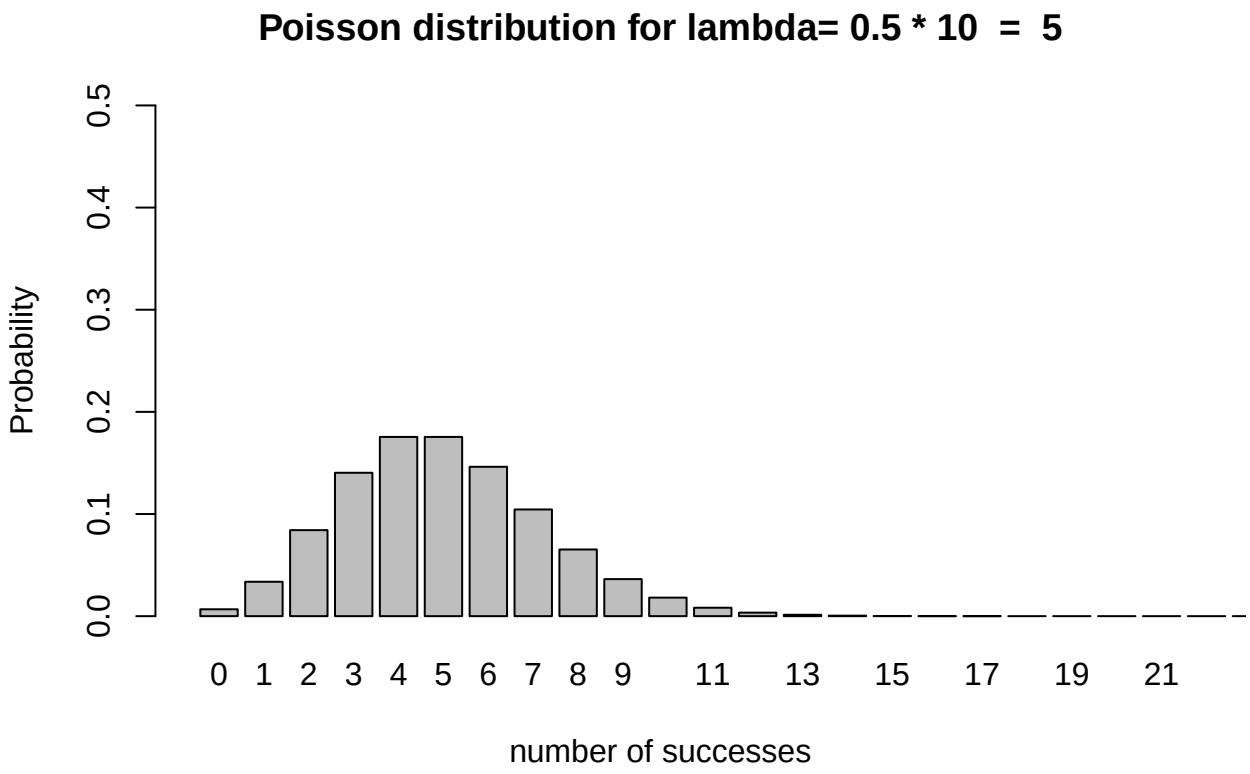


Poisson random variables

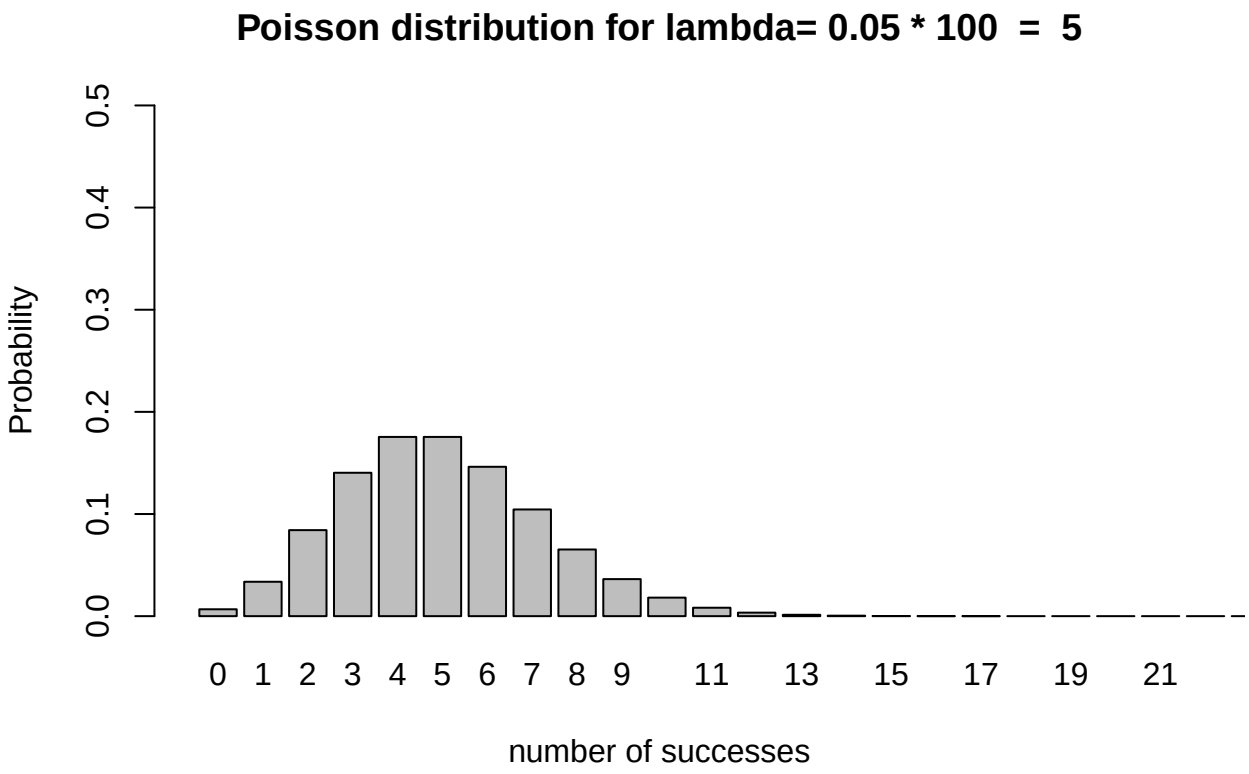
Let's plot some Poisson random variables. Note - we only need *one* parameter λ for the Poisson but we will give it in terms of $\lambda = np$ so we can compare to the corresponding Binomial distributions.

We'll choose our (n, p) combinations such that np is constant but n gets large while p gets small.

```
#Choose some (n,p) combinations as above.
n1=10
p1 = 0.5
n2 = 100
p2 = 0.05

x1 = seq(from=0,to=5*p1*n1,by=1)
x2 = seq(from=0,to=5*p2*n2,by=1)

ppx1 <- dpois(x=seq(0, 5*p1*n1),lambda = p1*n1)
ppx2 <- dpois(x=seq(0, 5*p2*n2),lambda = p2*n2)

barplot(ppx1, xlab="number of successes",ylab="Probability",
        main=paste("Poisson distribution for lambda=",
                   p1,"*",n1," = ",p1*n1),
        ylim=c(0,0.5),xlim=c(0, 5*p1*n1),names.arg = x1)
```

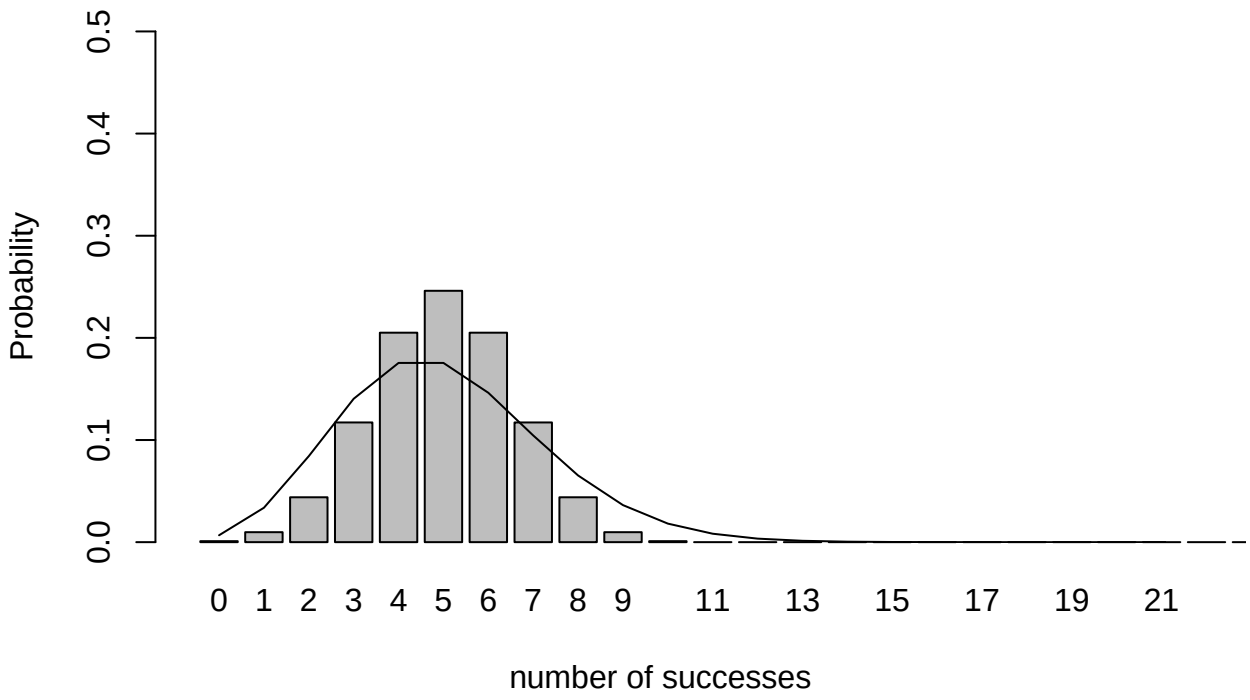



```
barplot(ppx2, xlab="number of successes",
  ylab="Probability",
  main=paste("Poisson distribution for lambda=",
                  p2,"*",n2," = ",p2*n2),ylim=c(0,0.5),xlim=c(0, 5*p2*n2),names.arg = x2)
```

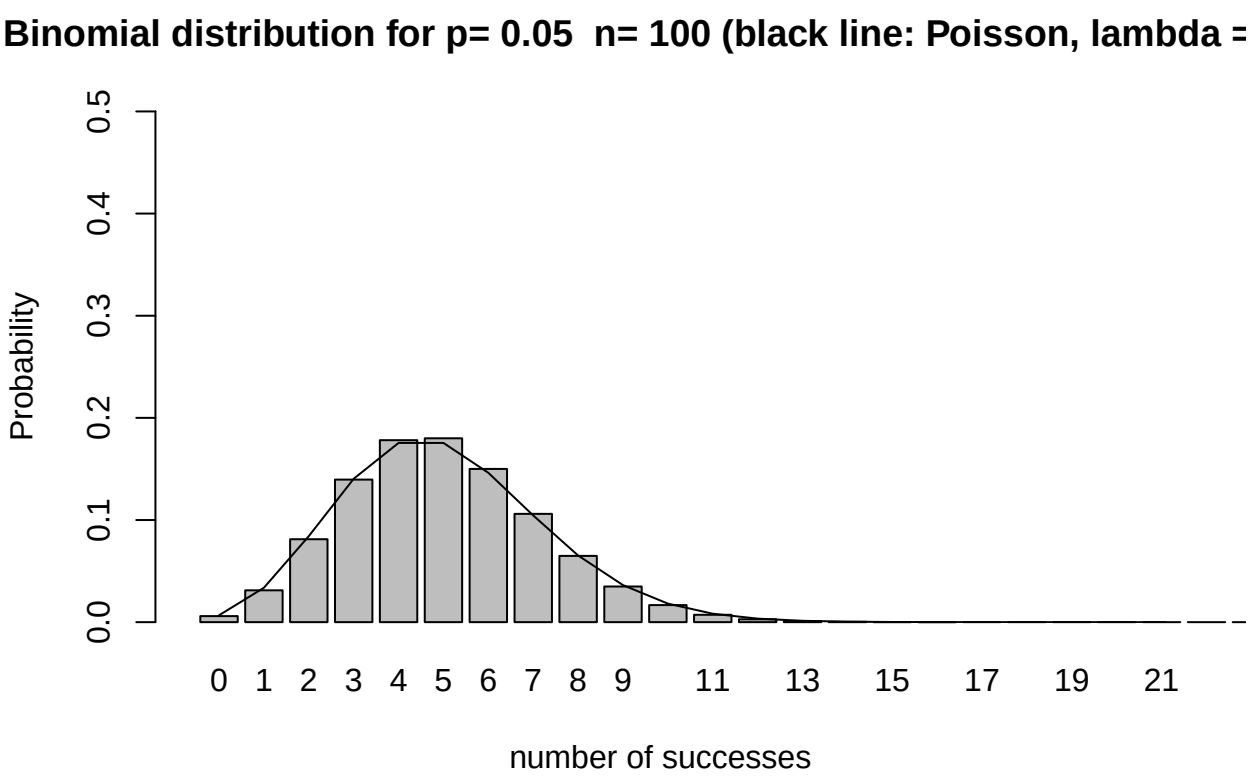


Now we compare to the Binomial for the given n and p values above.

```
px1 <- dbinom(x=x1, size=n1, p1)
px2 <- dbinom(x=x2, size=n2, p2)

df.bar<-barplot(px1, xlab="number of successes",ylab="Probability",
        main=paste("Binomial distribution for p=",
                   p1," n=", n1,
                   "(black line: Poisson, lambda =",p1*n1,")"),
        ylim=c(0,0.5),xlim=c(0,5*p1*n1),names.arg = x1)
lines(x=df.bar,y=ppx1,xlim=c(0,5*p1*n1))# note bar plots on different
```

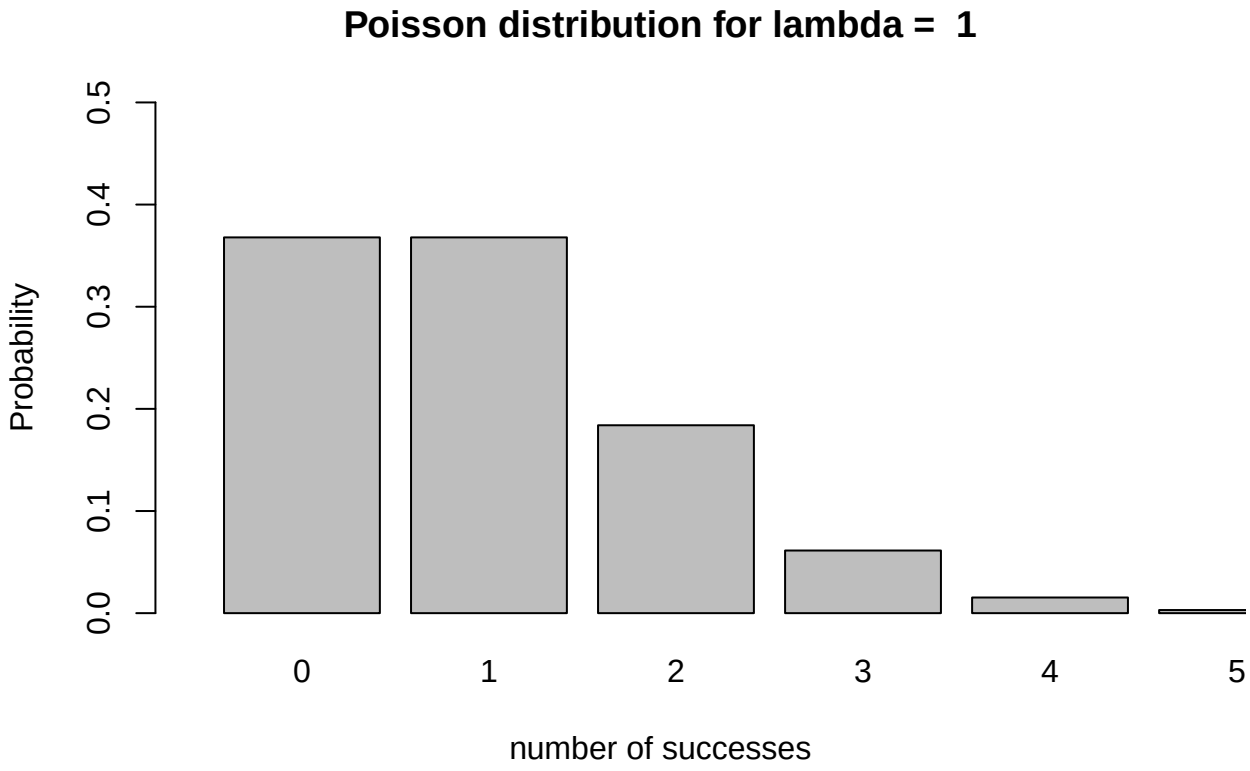



```
df.bar <- barplot(px2, xlab="number of successes",ylab="Probability",
        main=paste("Binomial distribution for p=",
                   p2," n=", n2,
                   "(black line: Poisson, lambda = ", p2*n2,")"),
        ylim=c(0,0.5),xlim=c(0,5*p2*n2),names.arg = x2)
lines(x=df.bar,y=ppx2,xlim=c(0,5*p2*n2))# note bar plots on different scale.
```

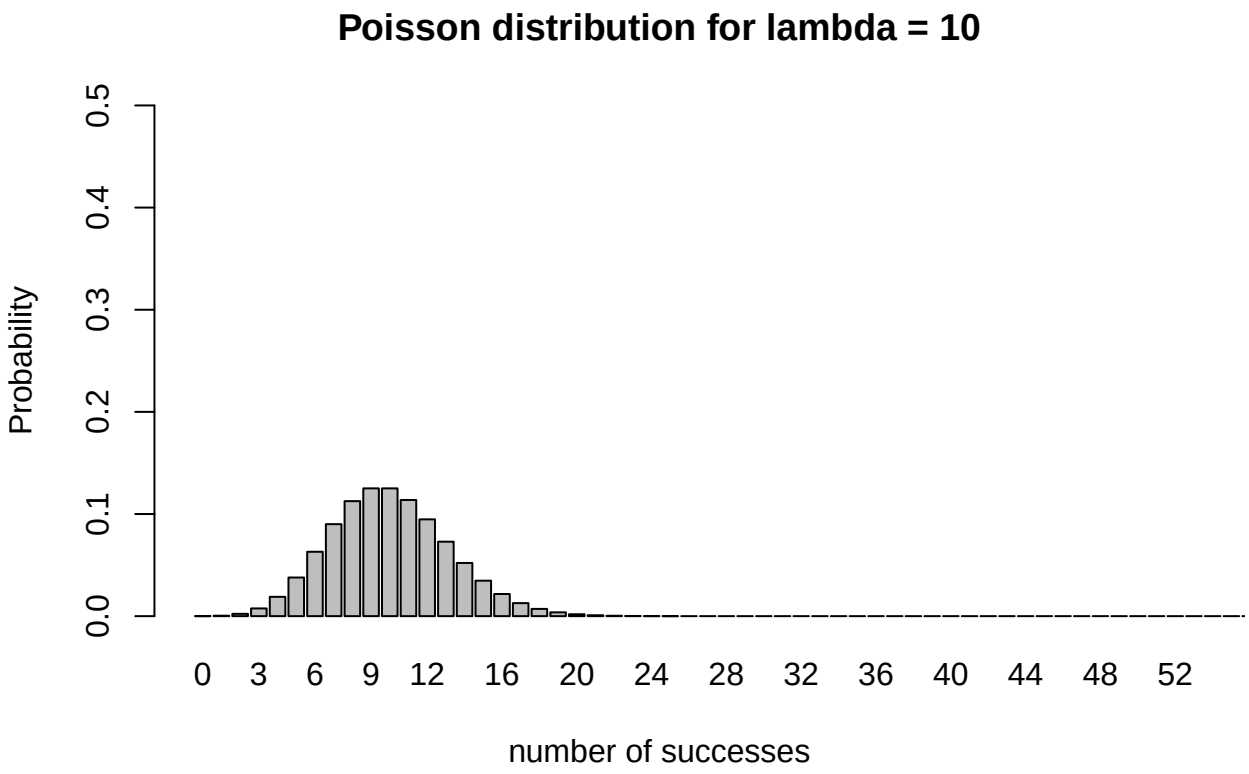


Let's just plot some other straight Poisson distributions, not worrying about any 'underlying' n or p.

```
#Choose some (n,p) combinations as above.
lambda1 = 1
lambda2 = 10

x1 = seq(from=0,to=6*lambda1,by=1)
x2 = seq(from=0,to=6*lambda2,by=1)

ppx1 <- dpois(x=seq(0, 6*lambda1),lambda = lambda1)
ppx2 <- dpois(x=seq(0, 6*lambda2),lambda = lambda2)

barplot(ppx1, xlab="number of successes",ylab="Probability",
        main=paste("Poisson distribution for lambda = ",lambda1),
        ylim=c(0,0.5),xlim=c(0, 6*lambda1),names.arg = x1)
```



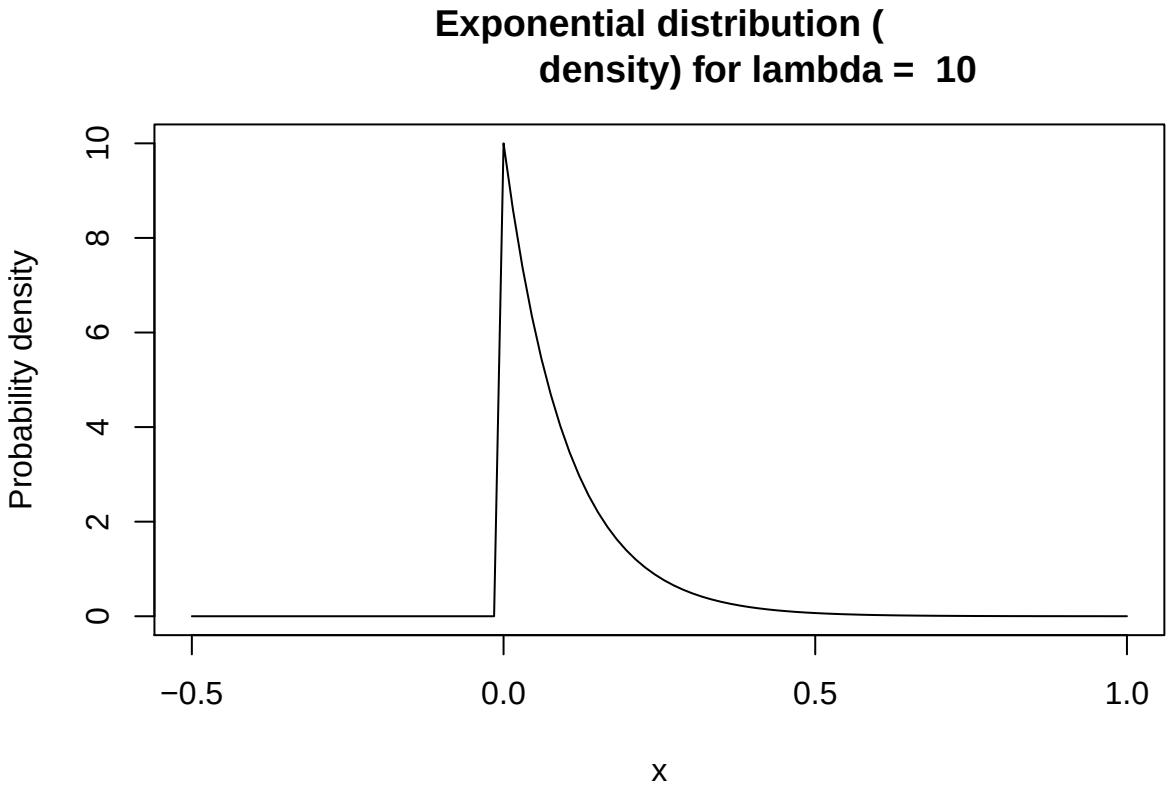

```
barplot(ppx2, xlab="number of successes",ylab="Probability",
        main=paste("Poisson distribution for lambda =",
                   lambda2),ylim=c(0,0.5),
        xlim=c(0, 6*lambda2),names.arg = x2)
```



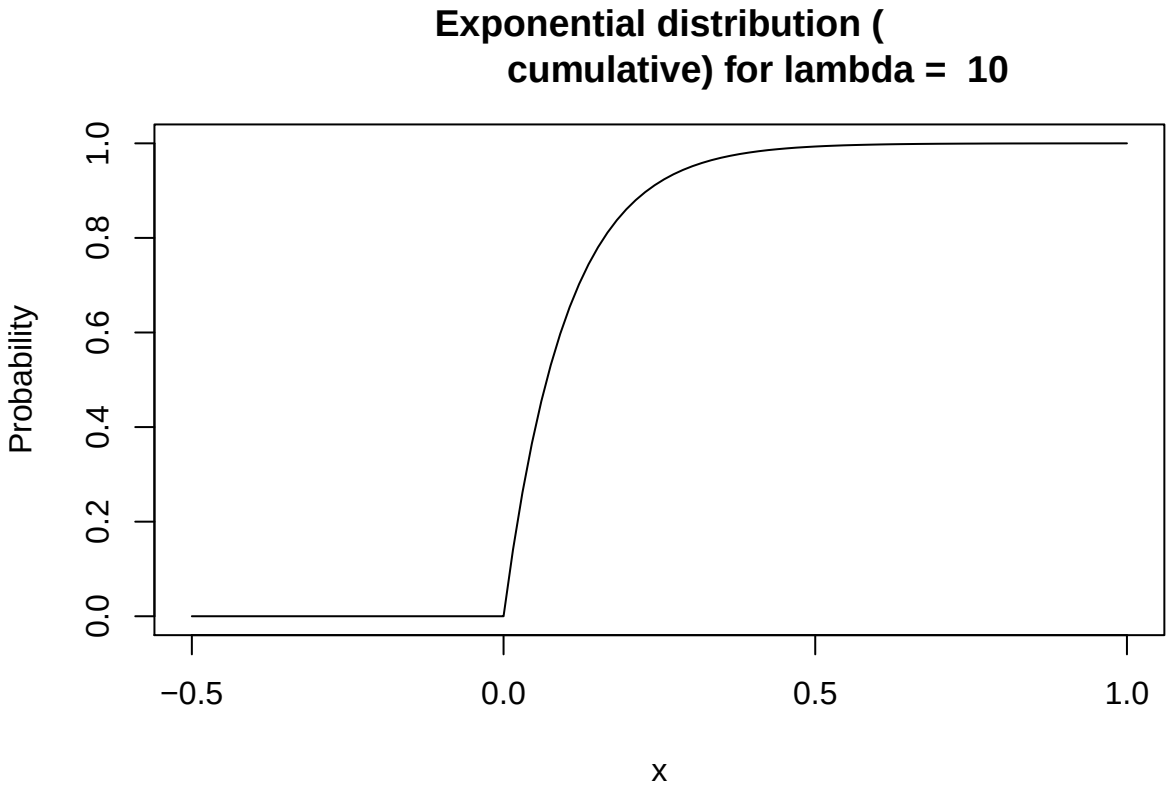
Continuous random variables

Exponential

```
rate=10

x = seq(from=-0.5,to=1,length=100)
hx <- dexp(x,rate = rate)
cx <- pexp(x,rate = rate)

plot(x,hx, type = 'l', xlab="x",ylab="Probability density",
        main=paste("Exponential distribution (
                    density) for lambda = ",
                    rate))
```

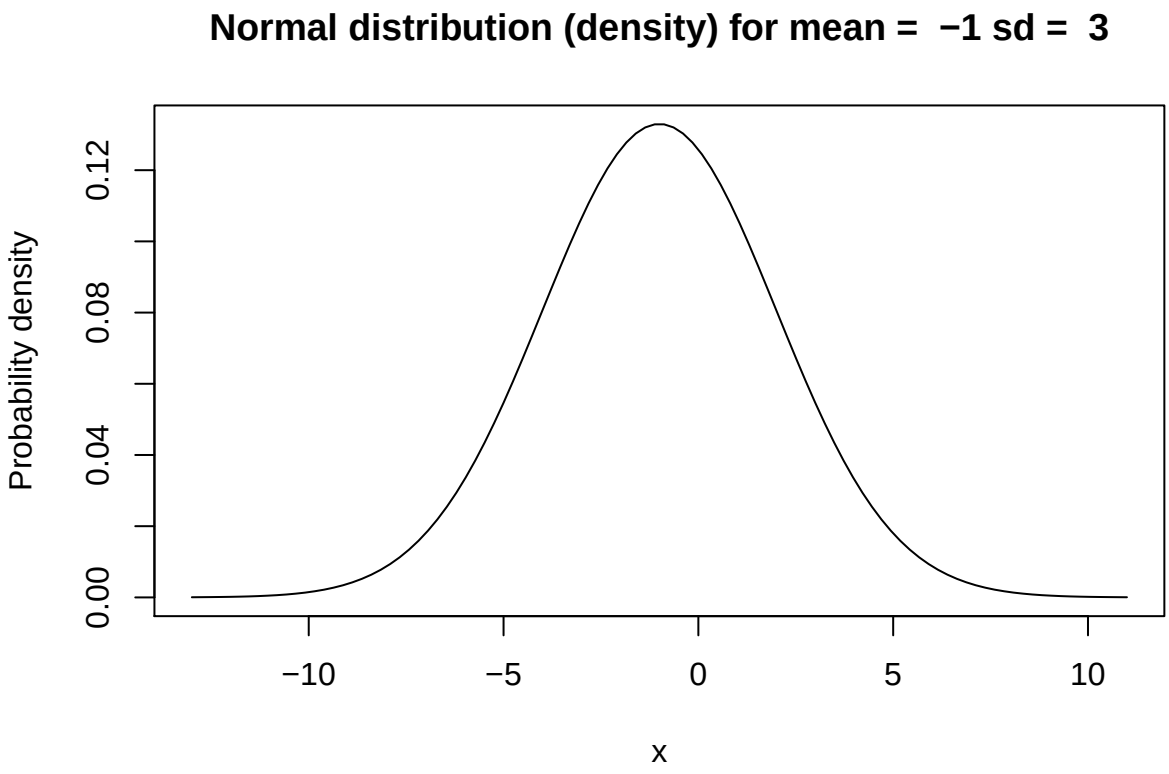



```
plot(x,cx, type = 'l', xlab="x",ylab="Probability",
        main=paste("Exponential distribution (
                    cumulative) for lambda = ",
                    rate))
```

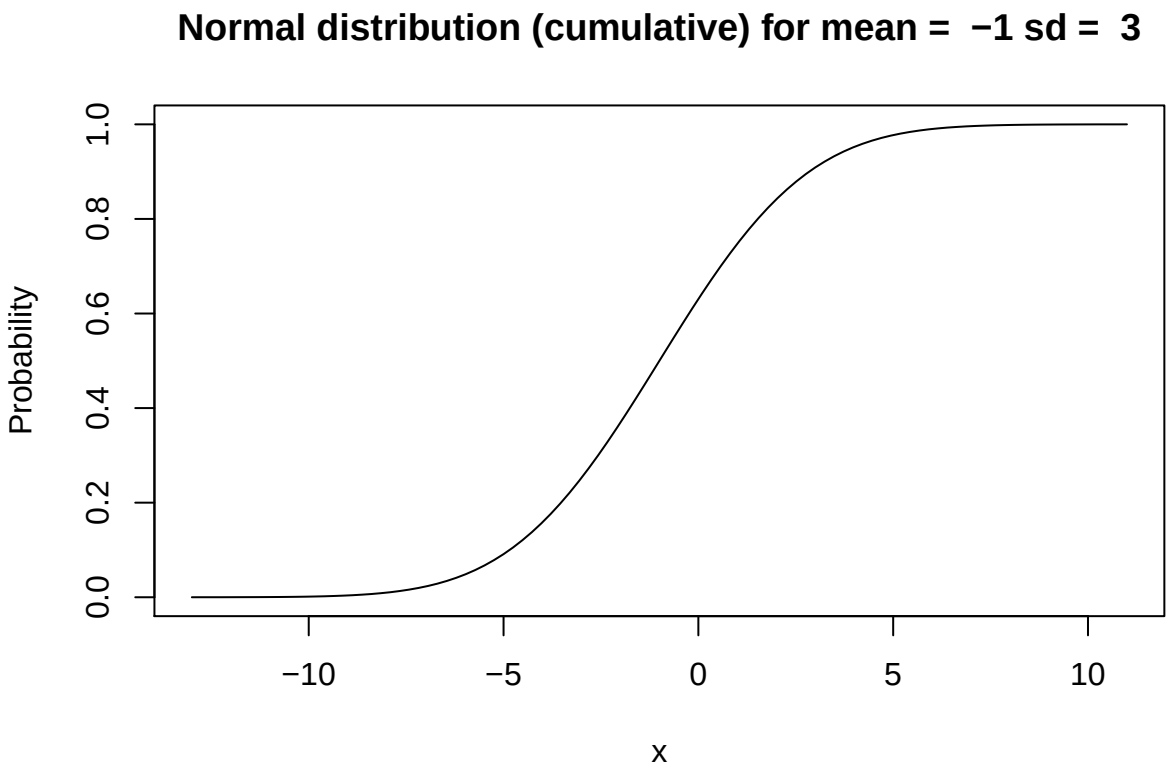


Normal

```
mean=-1.0; sd=3.0

x <- seq(-4,4,length=100)*sd + mean
hx <- dnorm(x,mean,sd)
cx <- pnorm(x,mean,sd)

plot(x,hx, type = 'l', xlab="x",ylab="Probability density",
        main=paste("Normal distribution (density) for mean = ",
                   mean, 'sd = ', sd))
```



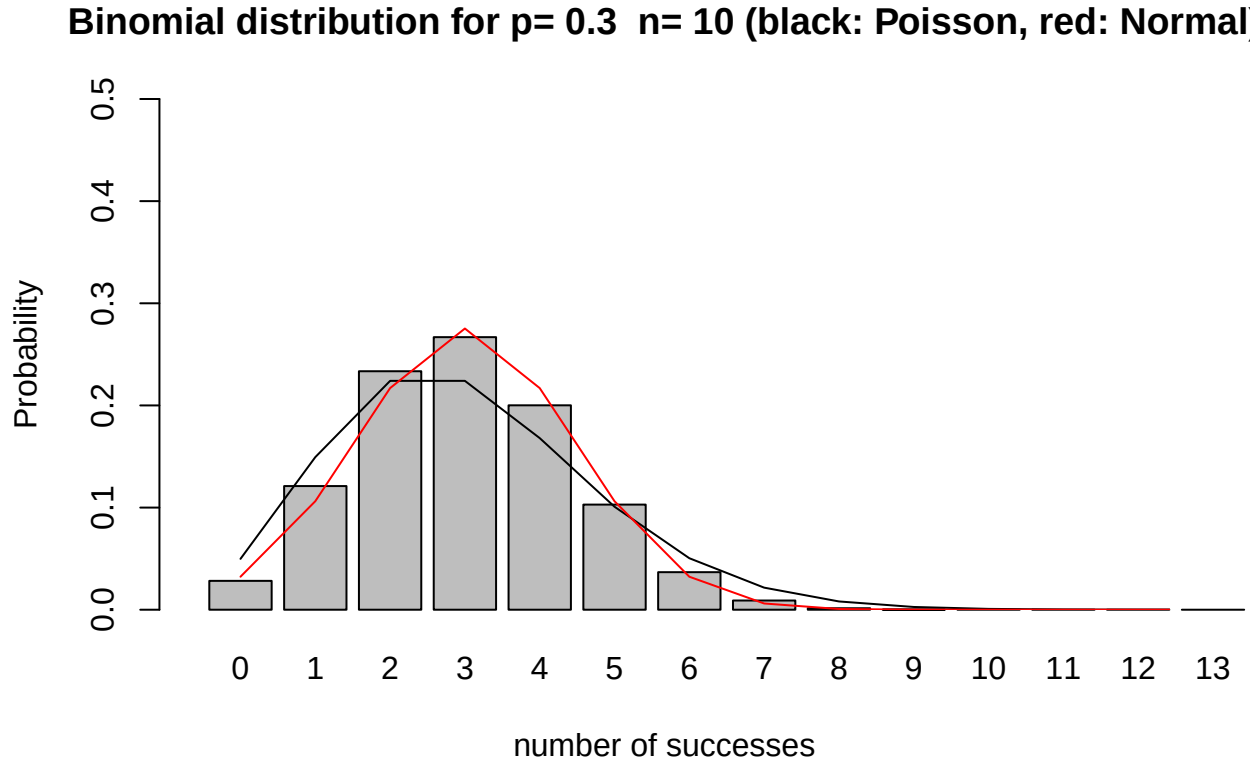

```
plot(x,cx, type = 'l', xlab="x",ylab="Probability",
       main=paste("Normal distribution (cumulative) for mean = ",
                  mean, 'sd = ', sd))
```



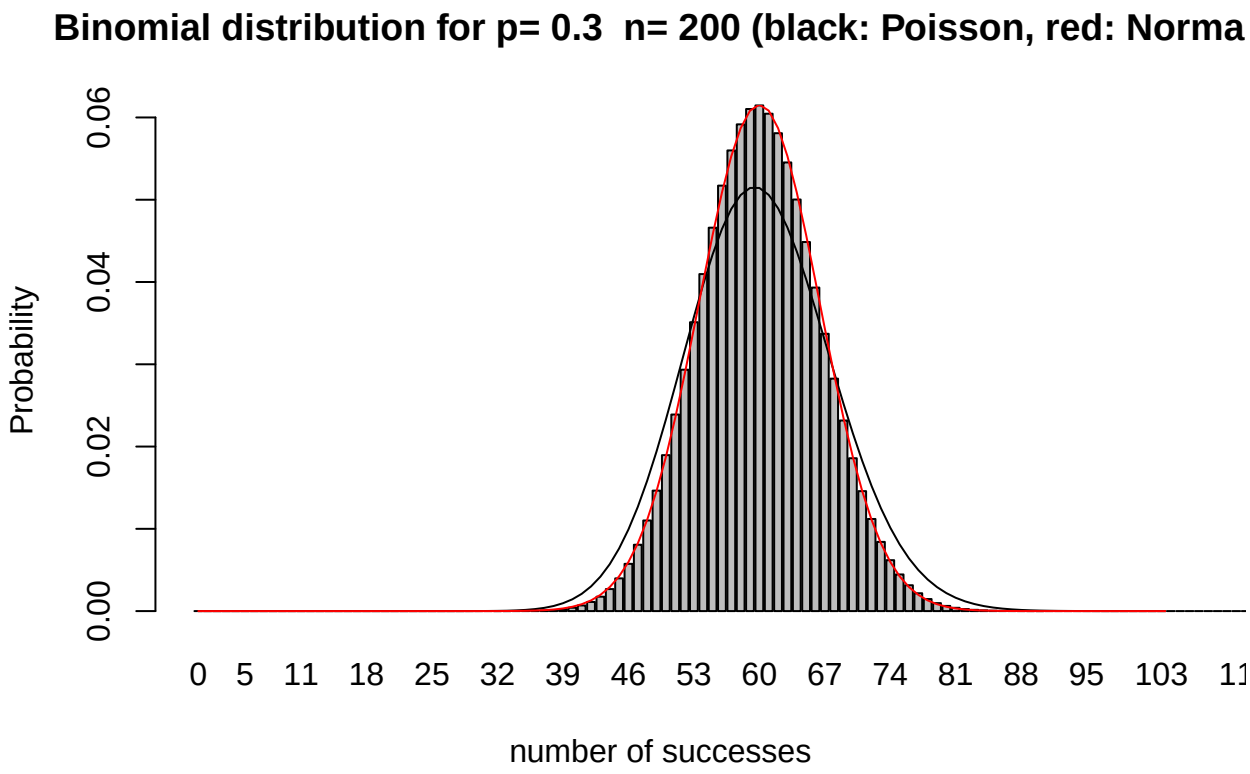
CLT illustrations

```
n1=10
p1 = 0.3

n2 = 200
p2 = 0.3

x1 = seq(from=0,to=5*p1*n1,by=1)
x2 = seq(from=0,to=2*p2*n2,by=1)


px1 <- dbinom(x=x1, size=n1, p1)
ppx1 <- dpois(x=seq(0, 5*p1*n1),lambda = p1*n1)
nx1 <- dnorm(x=x1,n1*p1,sqrt(n1*p1*(1-p1)))#note: R takes SD not var!

px2 <- dbinom(x=x2, size=n2, p2)
ppx2 <- dpois(x=seq(0, 2*p2*n2),lambda = p2*n2)
nx2 <- dnorm(x=x2,n2*p2,sqrt(n2*p2*(1-p2)))#note: R takes SD not var!


df.bar<-barplot(px1, xlab="number of successes",ylab="Probability",
        main=paste("Binomial distribution for p=",
                   p1," n=", n1,
```

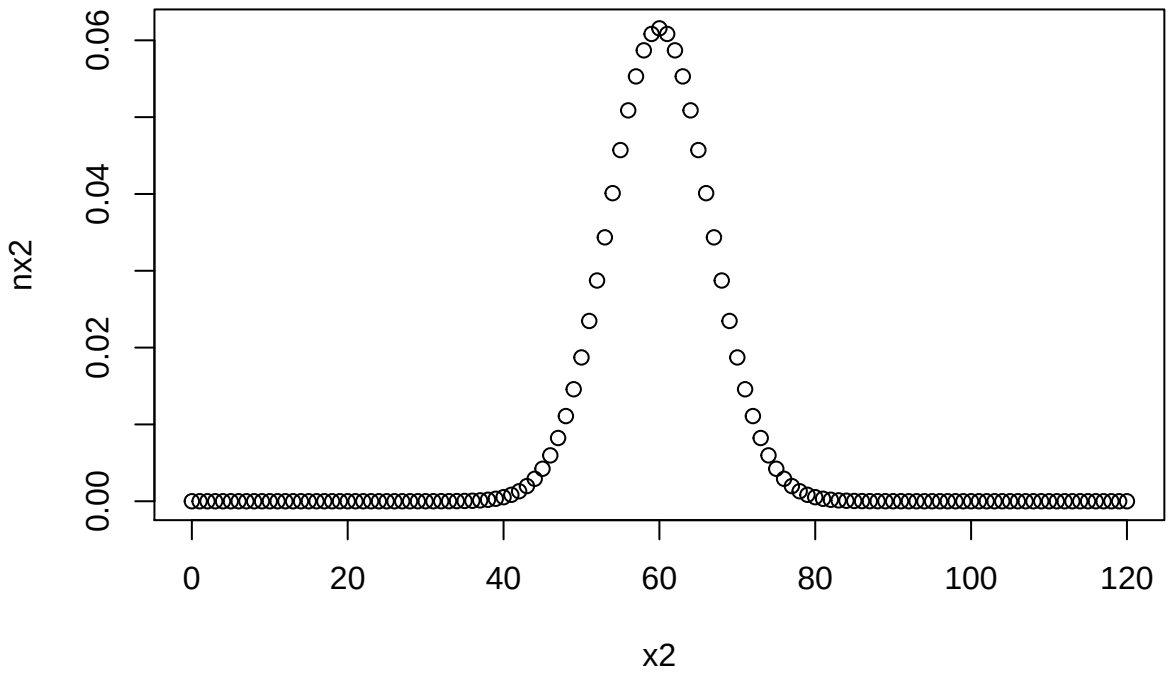
```
                  "(black: Poisson, red: Normal)"),
        ylim=c(0,0.5),xlim=c(0,5*p1*n1),names.arg = x1)
lines(x=df.bar,y=ppx1,xlim=c(0,5*p1*n1))# note bar plots on different scale.
lines(x=df.bar,y=nx1,xlim=c(0,5*p1*n1),col="red")
```




```
df.bar <- barplot(px2, xlab="number of successes",ylab="Probability",
        main=paste("Binomial distribution for p=",
                   p2," n=", n2,
                   "(black: Poisson, red: Normal)"),
        xlim=c(0,2*p2*n2),
        names.arg = x2)
lines(x=df.bar,y=ppx2,xlim=c(0,2*p2*n2))# note bar plots on different scale.
lines(x=df.bar,y=nx2,xlim=c(0,2*p2*n2),col="red")
```


Binomial distribution for p= 0.3 n= 200 (black: Poisson, red: Norma
Probability
0.00
0.02
0.04
0.06
0 5 11 18 25 32 39 46 53 60 67 74 81 88 95 103 11
number of successes

```
plot(x2,nx2)
```


nx2
0.00
0.02
0.04
0.06
0
20
40
60
80
100
120
x2