

# Lab 2: Probability distributions

Sawyer Balint

2/5/2025

## Contents

<b>1</b>	<b>Continuous distributions</b>	<b>1</b>
1.1	Uniform . . . . .	1
1.2	Beta . . . . .	5
1.3	Log Normal . . . . .	10
1.4	Exponential & Laplace . . . . .	15
1.5	Gamma . . . . .	18
<b>2</b>	<b>Part 2: Discrete distributions</b>	<b>23</b>
2.1	Binomial . . . . .	23
2.2	Poisson . . . . .	27
2.3	Negative binomial . . . . .	32

## 1 Continuous distributions

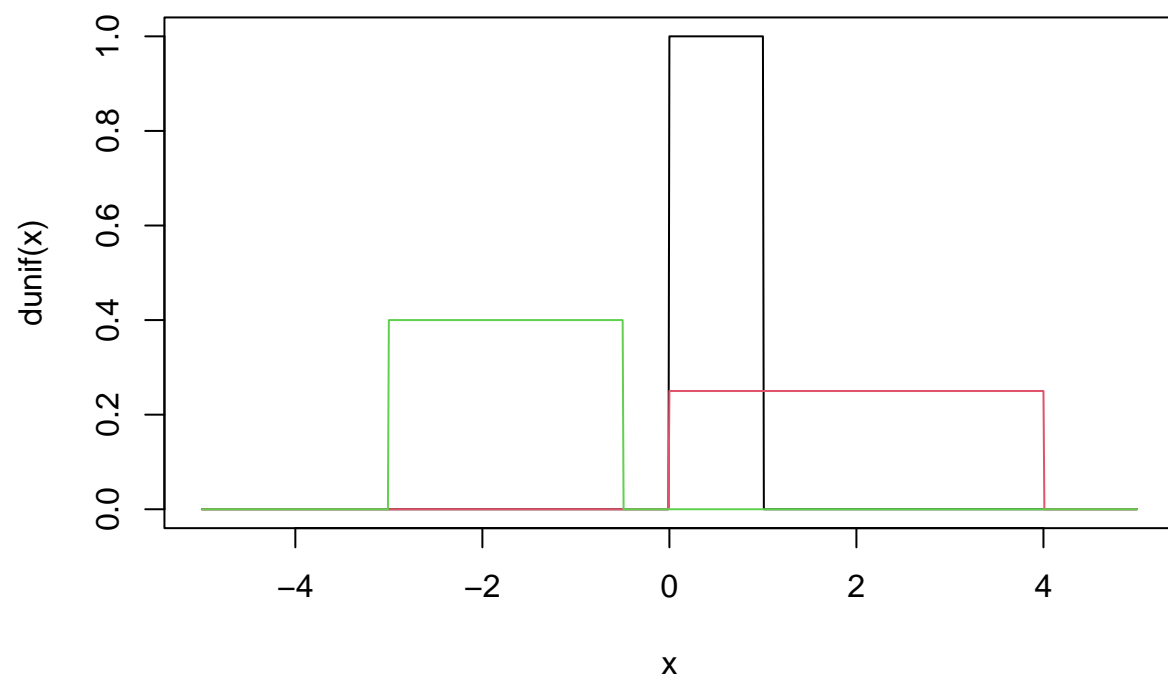
*Note, if you didn't start with the tutorial above, you'll need to define  $x$  and  $p$  variables before running the codes below.*

### 1.1 Uniform

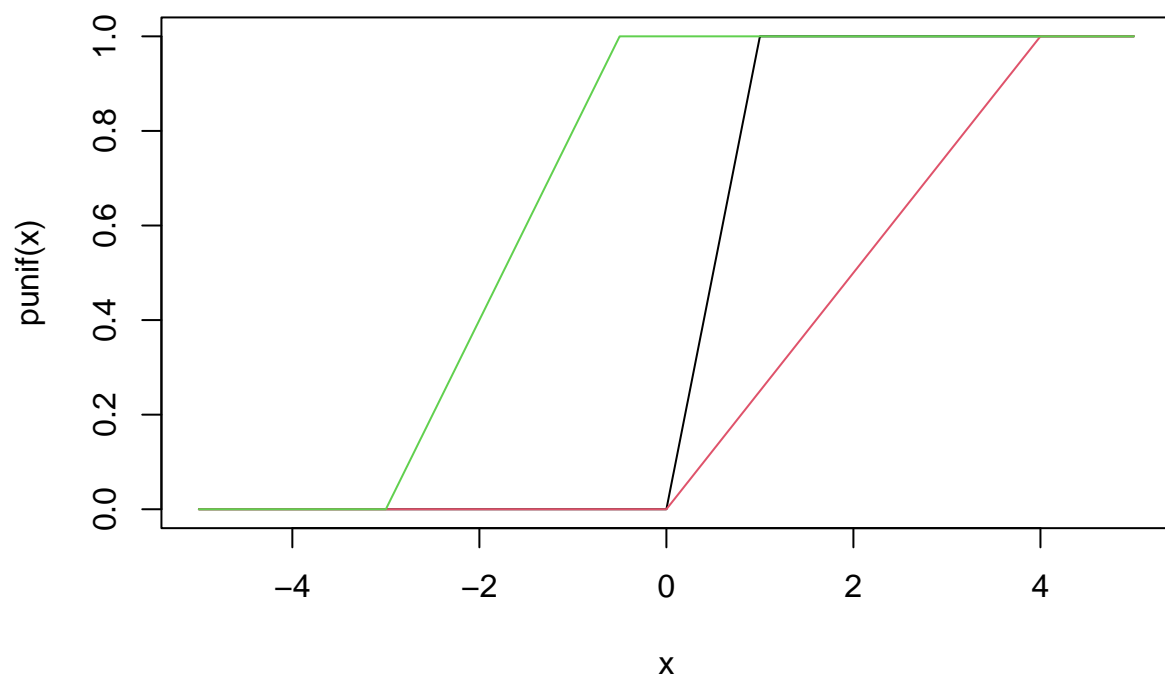
The uniform is used when there's an equal probability of an event occurring over a range of values.

```
#define variables
x = seq(-5,5,0.01)
p = seq(0,1,0.001)

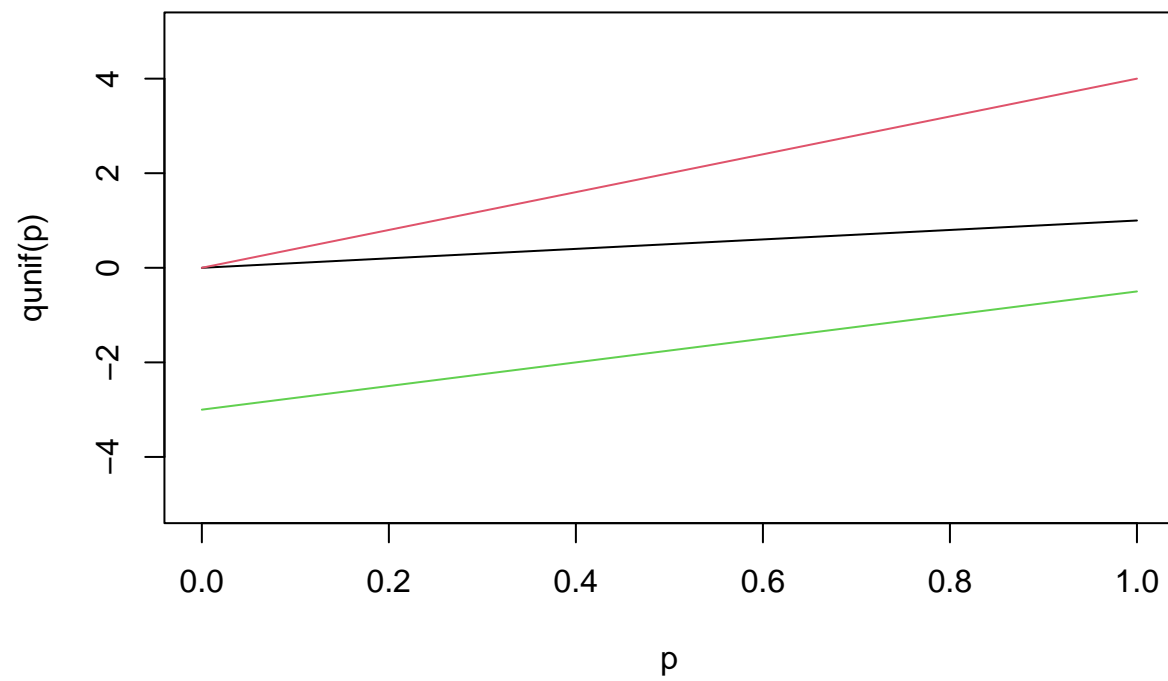
#begin original code
plot(x,dunif(x),type='l')
lines(x,dunif(x,0,4),col=2)
lines(x,dunif(x,-3,-0.5),col=3)
```



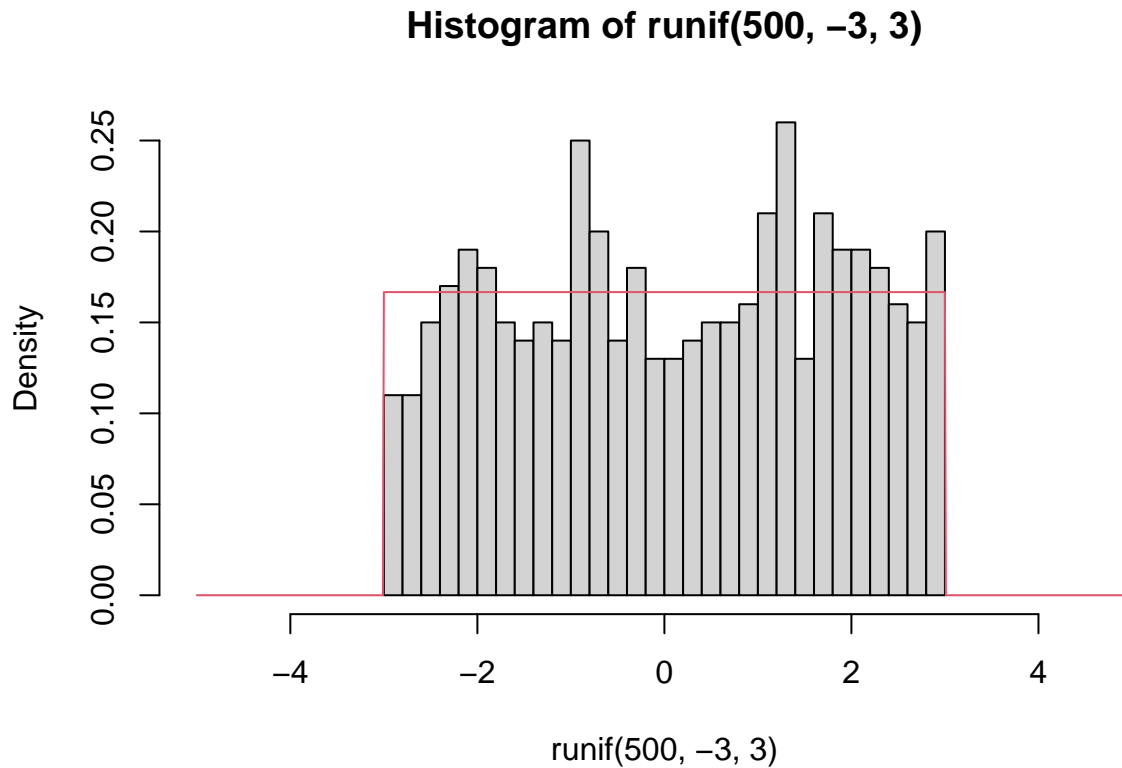
```
plot(x,punif(x),type='l')  
lines(x,punif(x,0,4),col=2)  
lines(x,punif(x,-3,-0.5),col=3)
```



```
plot(p,qunif(p),type='l',ylim=range(x))  
lines(p,qunif(p,0,4),col=2)  
lines(p,qunif(p,-3,-0.5),col=3)
```



```
hist(runif(500,-3,3),breaks=30,xlim=range(x),probability=TRUE)
lines(x,dunif(x,-3,3),col=2)
```



**Question 2:** Why does the height of the uniform PDF change as the width changes?

As the width increases, the probability of  $x$  at any given value decrease because there are more possibilities for  $x$ . Put another way, the area under the uniform PDF has to remain the same - so as width increases, the height decreases.

**Question 1:** What do the second and third arguments to the uniform specify? What are their default values?

The second argument is the minimum, and the third argument is the maximum. The default arguments are 0 and 1 for minimum and maximum, respectively.

## 1.2 Beta

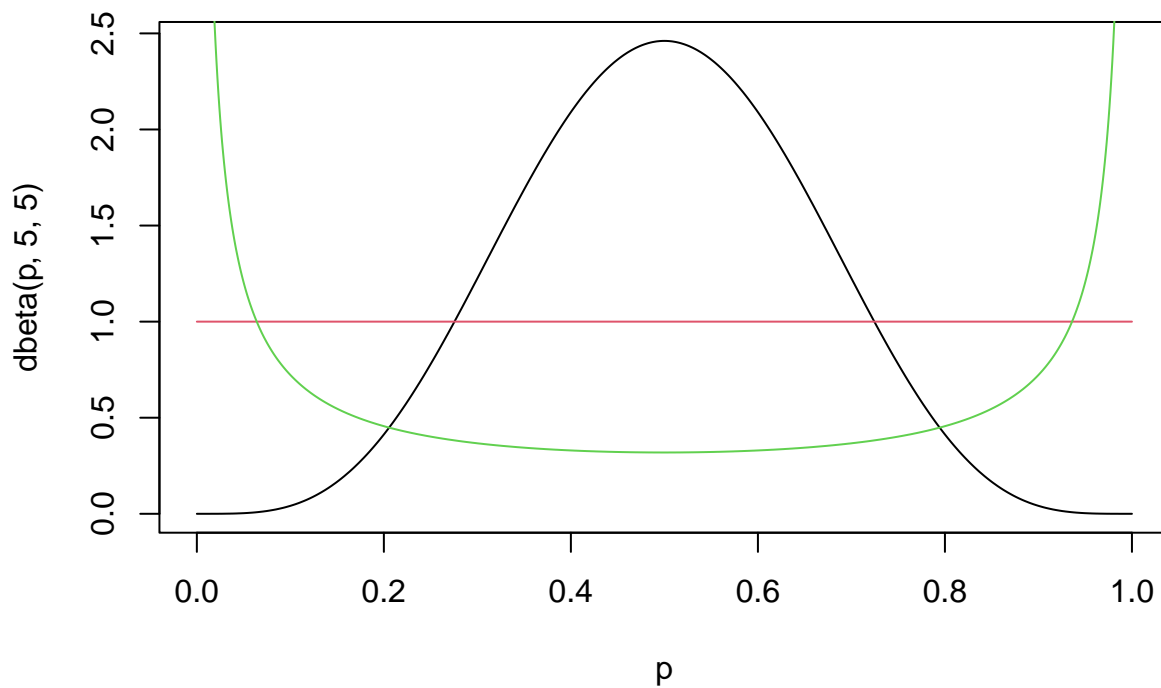
The Beta is an interesting distribution because it is bound on the range  $0 \leq X \leq 1$  and thus is very often used to describe data that is a proportion. At first glance the mathematical formula for the Beta looks a lot like the Binomial:

$$\begin{aligned} \text{Beta}(x \mid a, b) &\propto x^{a-1}(1-x)^{b-1} \\ \text{Binom}(x \mid n, p) &\propto p^x(1-p)^{n-p} \end{aligned}$$

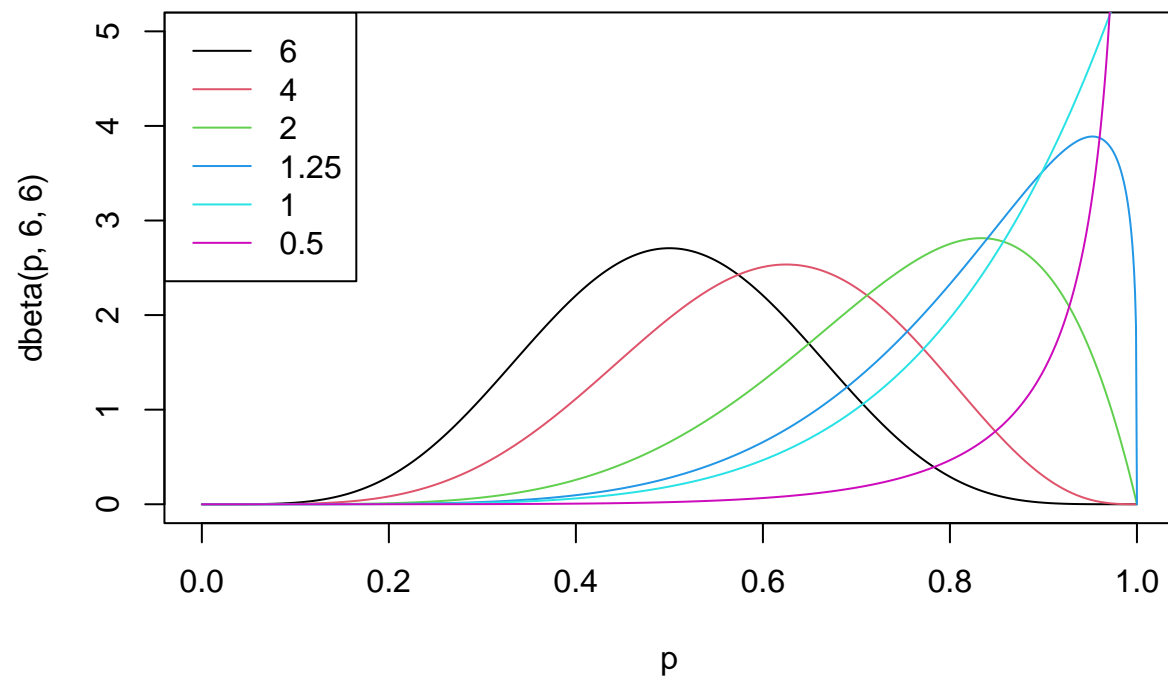
The critical difference between the two is that in the Beta the random variable  $X$  is in the base while in the Binomial it is in the exponent. Unlike many distributions you may have used in the past, the two shape parameters in the Beta do not define the mean and variance, but these can be calculated as simple functions

of  $\alpha$  and  $\beta$ . The Beta does have an interesting property of symmetry though, whereby  $\text{Beta}(\alpha, \beta)$  is the reflection of  $\text{Beta}(\beta, \alpha)$ .

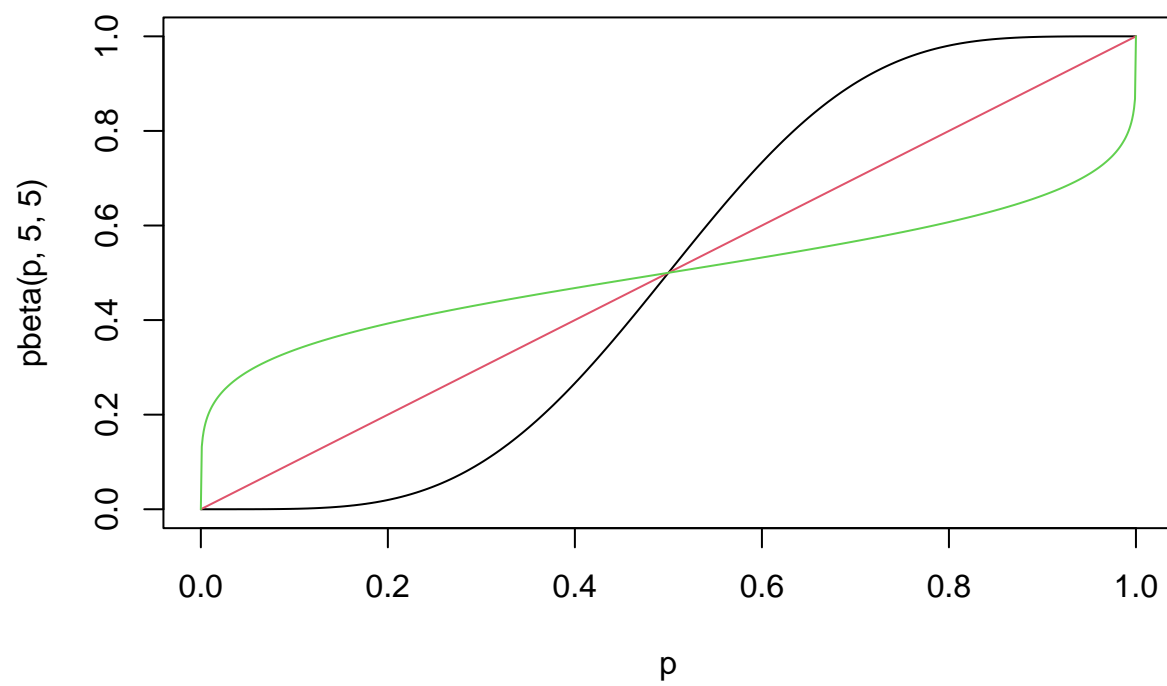
```
plot(p,dbeta(p,5,5),type='l')
lines(p,dbeta(p,1,1),col=2)
lines(p,dbeta(p,0.2,0.2),col=3)
```



```
## vary beta
plot(p,dbeta(p,6,6),type='l',ylim=c(0,5))
lines(p,dbeta(p,6,4),col=2)
lines(p,dbeta(p,6,2),col=3)
lines(p,dbeta(p,6,1.25),col=4)
lines(p,dbeta(p,6,1),col=5)
lines(p,dbeta(p,6,0.25),col=6)
legend("topleft",legend=c(6,4,2,1.25,1,0.5),lty=1,col=1:6)
```

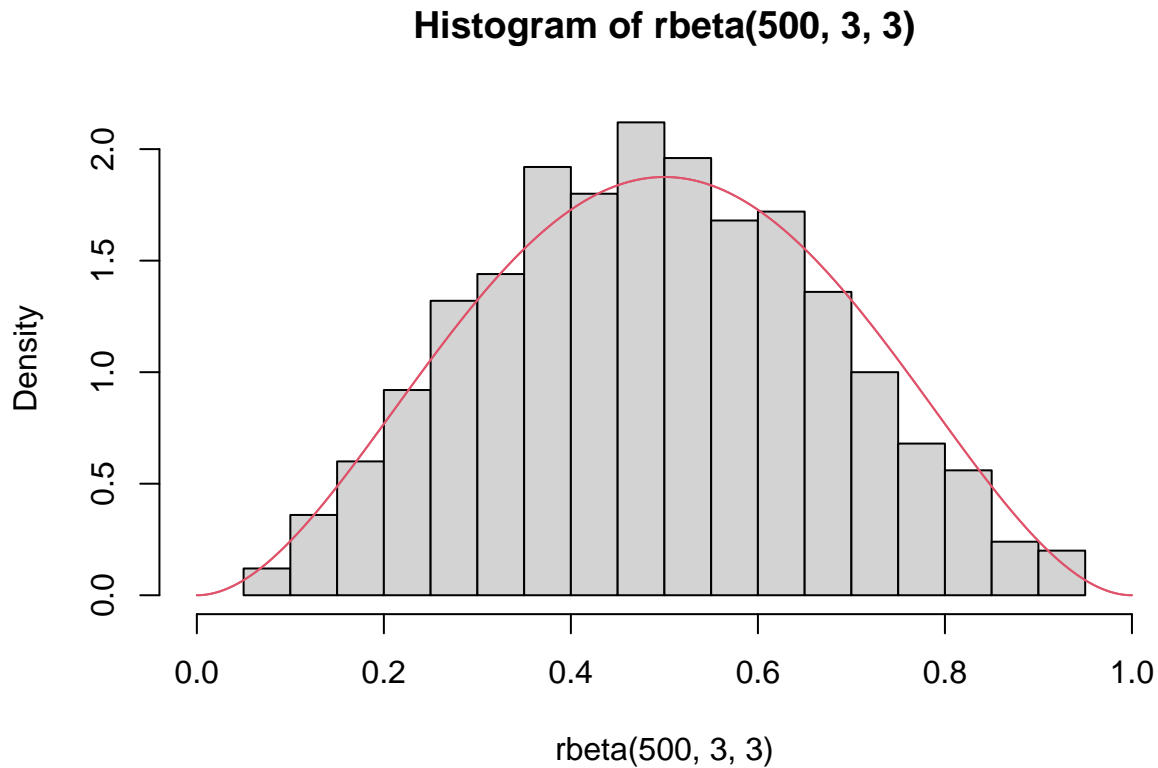


```
plot(p, pbeta(p, 5, 5), type='l')
lines(p, pbeta(p, 1, 1), col=2)
lines(p, pbeta(p, 0.2, 0.2), col=3)
```



```
hist(rbeta(500,3,3),breaks=30,xlim=range(p),probability=TRUE)
lines(p,dbeta(p,3,3),col=2)
lines(p,dbeta(p,3,3),col=2)
```





**Question 3:** The Beta has a special case, Beta(1,1) that is equivalent to what other PDF?

Beta(1,1) is equivalent to a uniform distribution.

**Question 4:** In the first panel, the mean is the same for each line (0.5). What are the variances? (Hint: Calculate this analytically. Look up the distribution in one of the recommended references.)

$$Var(X) = \frac{\alpha\beta}{(\alpha + \beta + 1) * (\alpha + \beta)^2}$$

```
#define function to calculate variance of a beta distribution
beta_var <- function(shape1, shape2){
  var = (shape1 * shape2)/((shape1 + shape2 + 1) * (shape1 + shape2)^2)
  return(var)
}

#calculate variance for each line
var1 <- round(beta_var(5,5), digits=3)
var2 <- round(beta_var(1,1), digits=3)
var3 <- round(beta_var(0.2,0.2), digits=3)
```

The variances are 0.023, 0.083, 0.179 for the black, red, and green lines, respectively.

**Question 5:** In the second panel, what are the means and medians of each of these curves? (Hint: you'll need to calculate the mean analytically and use one of the variants of R's beta function to find the median.)

$$E(X) = \frac{\alpha}{\alpha + \beta}$$

```
#function to calculate mean
beta_mean <- function(shape1, shape2){
  mean <- shape1 / (shape1 + shape2)
  return(mean)
}

#calculate mean and median
#median is the 50th quantile
df <- data.frame(shape1=6,
                  shape2=c(6,4,2,1.25,1,0.5)) %>%
  mutate(mean = round(beta_mean(6, shape2), digits=3),
         median = round(qbeta(0.5, 6, shape2), digits=3))

kable(df)
```

shape1	shape2	mean	median
6	6.00	0.500	0.500
6	4.00	0.600	0.607
6	2.00	0.750	0.772
6	1.25	0.828	0.858
6	1.00	0.857	0.891
6	0.50	0.923	0.961

### 1.3 Log Normal

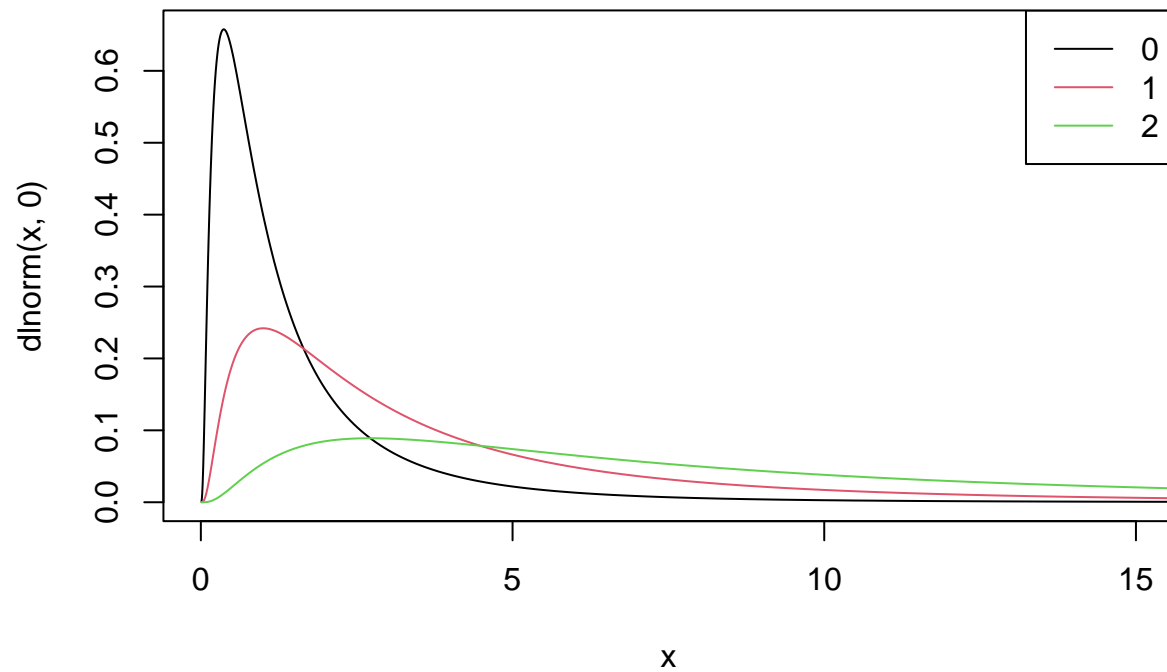
The lognormal is a log transform of the normal distribution. It is defined on the range  $X > 0$  so is commonly used for data that cannot be negative by definition. The distribution is also positively skewed so is often used for skewed data. One thing that often goes unappreciated with the log-normal is that the mean,  $E[X]$ , depends on the variance:

$$E[X] = e^{\mu + \frac{1}{2}\sigma^2}$$

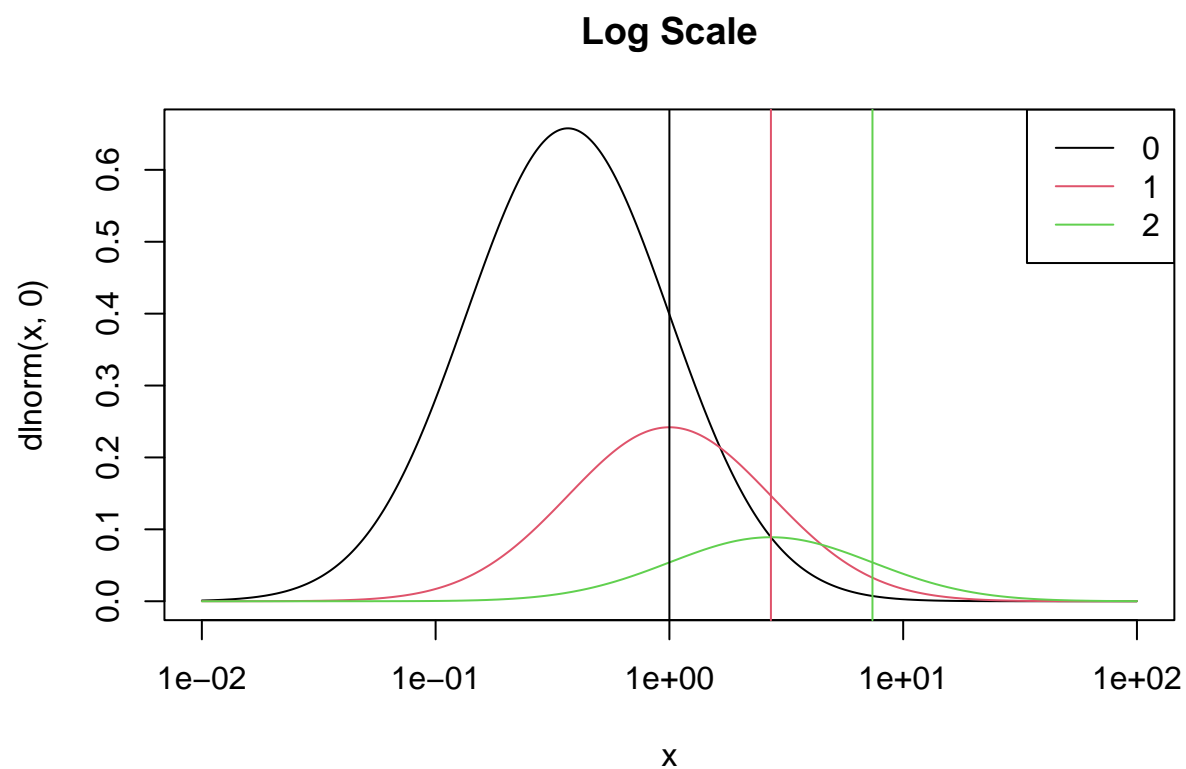
This applies not just when you explicitly use the lognormal, but also **whenever you log-transform data** and then calculate a mean or standard deviation – a fact that is vastly under-appreciated in the biological and environmental sciences and frequently missed in the published literature. In fact, ANY data transformation applied to make data “more normal” will change the mean, with the functional form of the bias depending on the transformation used. You can not simply back-transform the data without correcting for this bias. This phenomena is another illustration of Jensen’s Inequality.

```
## changing the mean
x <- 10^seq(-2,2,by=0.01)
plot(x,dlnorm(x,0),type='l',xlim=c(0,15),main="Changing the Mean")
lines(x,dlnorm(x,1),col=2)
lines(x,dlnorm(x,2),col=3)
legend("topright",legend=0:2,lty=1,col=1:3)
```

## Changing the Mean

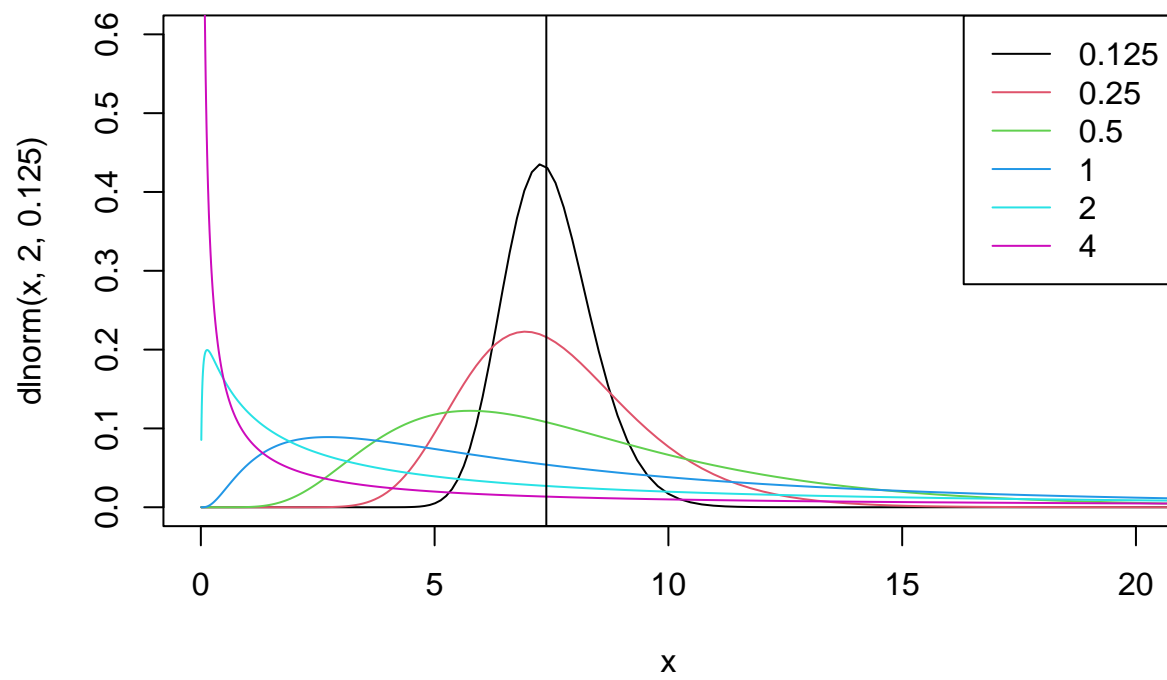


```
## on a log scale
plot(x,dlnorm(x,0),type='l',log='x',main="Log Scale")
lines(x,dlnorm(x,1),col=2)
lines(x,dlnorm(x,2),col=3)
abline(v=exp(0),col=1)
abline(v=exp(1),col=2)
abline(v=exp(2),col=3)
legend("topright",legend=0:2,lty=1,col=1:3)
```



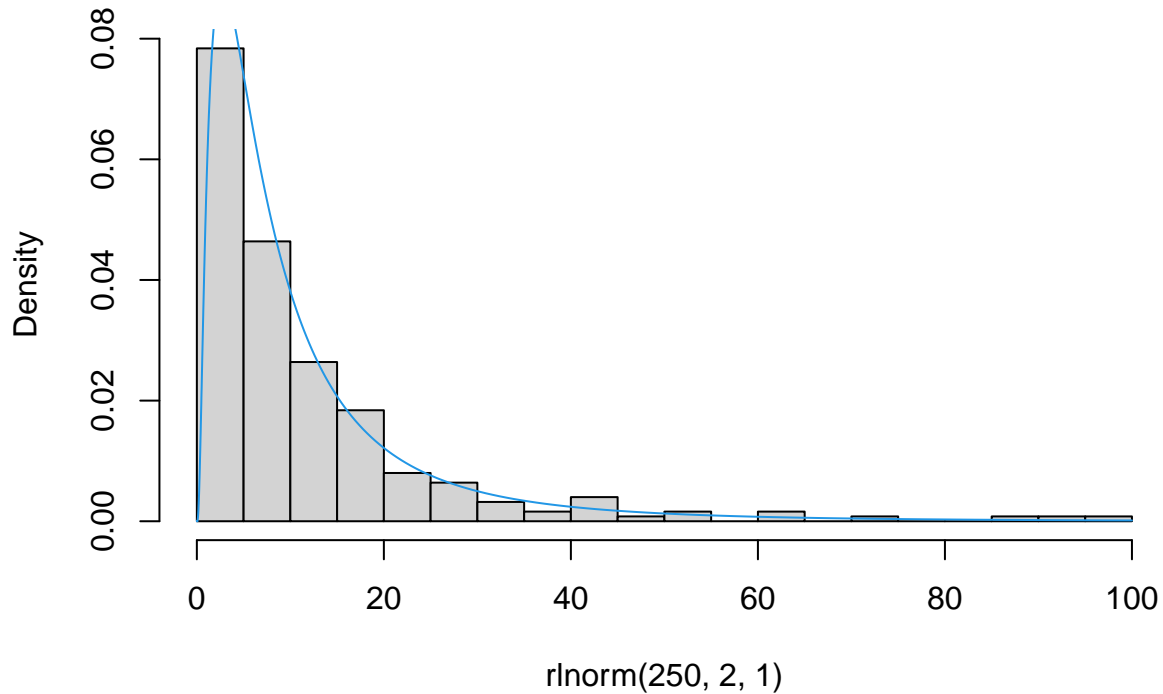
```
## changing the variance
plot(x,dlnorm(x,2,.125),type='l',xlim=c(0,20),ylim=c(0,0.6),main="Changing the Variance")
lines(x,dlnorm(x,2,0.25),col=2)
lines(x,dlnorm(x,2,0.5),col=3)
lines(x,dlnorm(x,2,1),col=4)
lines(x,dlnorm(x,2,2),col=5)
lines(x,dlnorm(x,2,4),col=6)
abline(v=exp(2),col=1)
legend("topright",legend=c(0.125,0.25,0.5,1,2,4),lty=1,col=1:6)
```

## Changing the Variance



```
## random sample  
hist(rlnorm(250,2,1),breaks=30,probability=TRUE)  
lines(x,dlnorm(x,2,1),col=4)
```

## Histogram of rlnorm(250, 2, 1)



**Question 6:** What are the arithmetic and geometric means of the three curves in the first panel? (Reminder: arithmetic means are means in the linear domain, geometric means are means in the log domain)

```
#I'm going to do this in a tidyverse way
#using rlnorm because calculating the mean of a density from dlnorm doesn't make sense to me
df <- data.frame(x=x,
                 d0=rlnorm(x,0),
                 d1=rlnorm(x,1),
                 d2=rlnorm(x,2)) %>%
  filter(x>0) %>% #for geometric mean
  pivot_longer(cols=c(d0,d1,d2), names_to="meanlog", values_to="d") %>%
  mutate(meanlog=as.numeric(substr(meanlog,2,2))) %>% #default for dlnorm
  group_by(meanlog) %>%
  summarize("Arithmetic Mean"=signif(mean(d), digits=3), #arithmetic mean
           "Arithmetic Mean (log)"=signif(log(mean(d)), digits=3), #arithmetic mean
           "Geometric Mean"=signif(exp(mean(log(d))), digits=3),
           "Geometric Mean (log)"=signif(log(exp(mean(log(d)))), digits=3)) %>% #geometric mean
  ungroup() %>%
  unique()

kable(df)
```

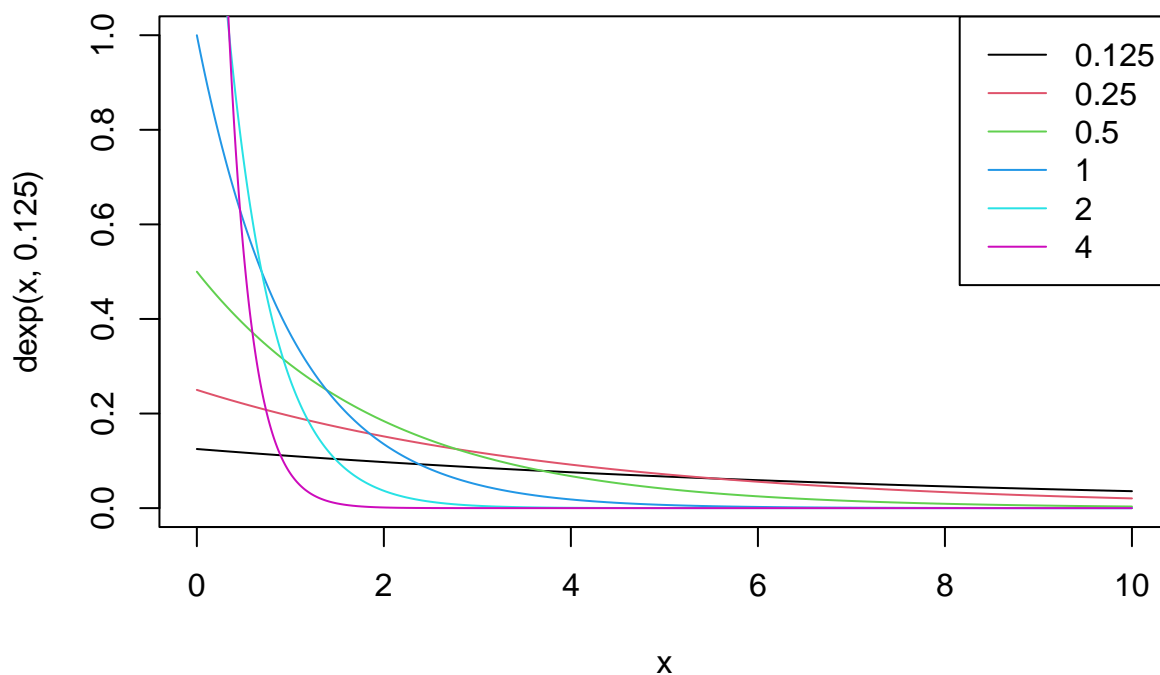
meanlog	Arithmetic Mean	Arithmetic Mean (log)	Geometric Mean	Geometric Mean (log)
0	1.72	0.542	1.02	0.0198
1	4.82	1.570	2.75	1.0100
2	12.50	2.520	7.46	2.0100

Above, I calculated the arithmetic and geometric mean as well as their log values. We see that the log of the geometric mean accurately reflects the meanlog parameter passed to `rlnorm()`

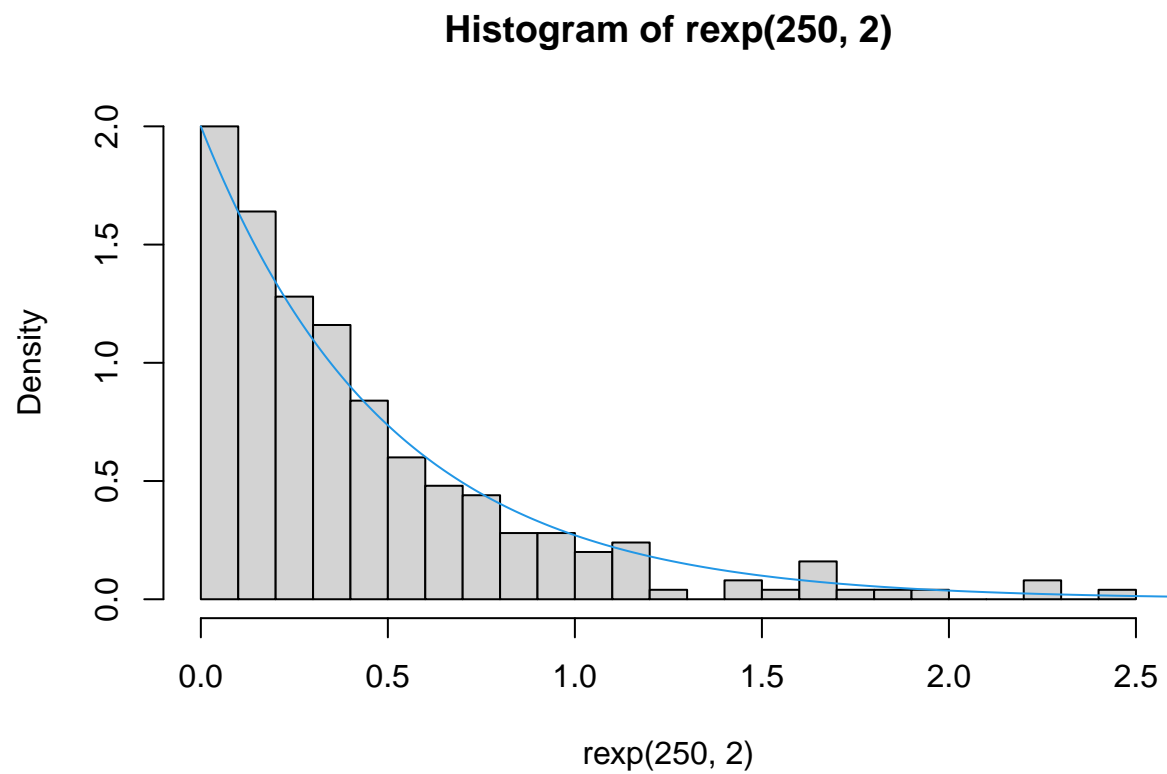
## 1.4 Exponential & Laplace

The exponential distribution arises naturally as the time it takes for an event to occur when the average rate of occurrence,  $r$ , is constant. The exponential is a special case of the Gamma (discussed next) where  $Exp(X | r) = Gamma(X | 1, r)$ . The exponential is also a special case of the Weibull,  $Exp(X | r) = Weibull(X | r, 1)$ , where the Weibull is a generalization of the exponential that allows the rate parameter  $r$  to increase or decrease with time. The Laplace is basically a two-sided exponential and arises naturally if one is dealing with absolute deviation,  $|x - m|$ , rather than squared deviation,  $(x - m)^2$ , as is done with the normal.

```
## changing the mean
x <- seq(0,10,by=0.01)
plot(x,dexp(x,0.125),type='l',ylim=c(0,1))
lines(x,dexp(x,0.25),col=2)
lines(x,dexp(x,0.5),col=3)
lines(x,dexp(x,1),col=4)
lines(x,dexp(x,2),col=5)
lines(x,dexp(x,4),col=6)
legend("topright",legend=c(0.125,0.25,0.5,1,2,4),lty=1,col=1:6)
```

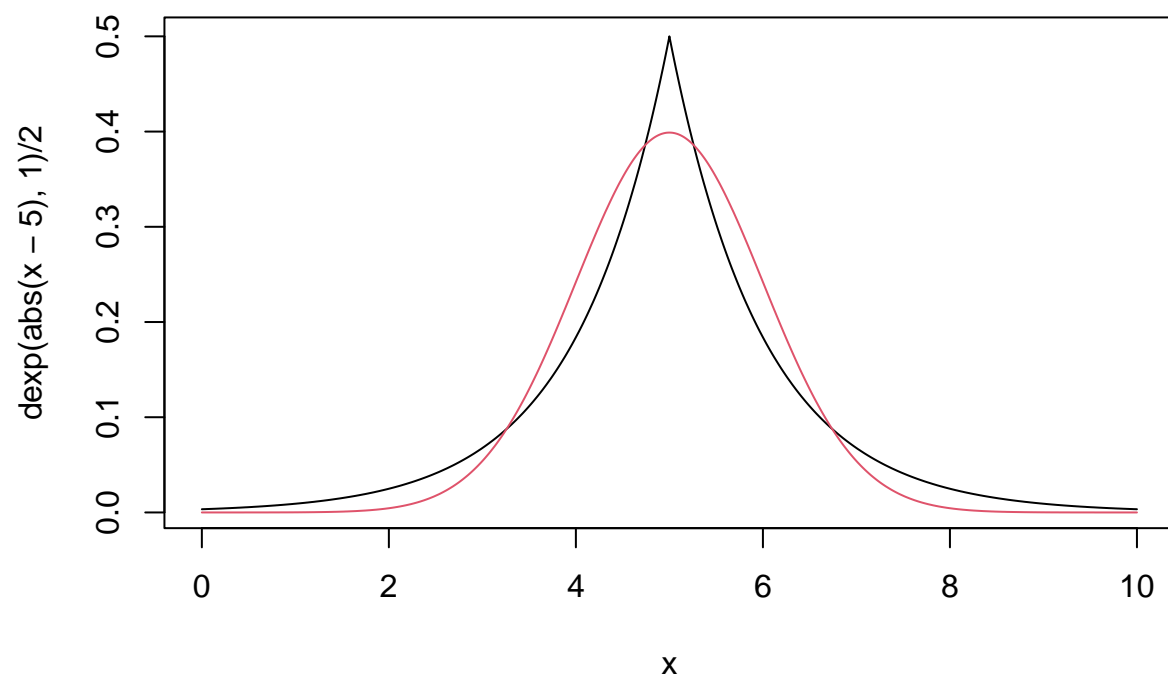


```
## random sample
hist(rexp(250,2),breaks=30,probability=TRUE)
lines(x,dexp(x,2),col=4)
```

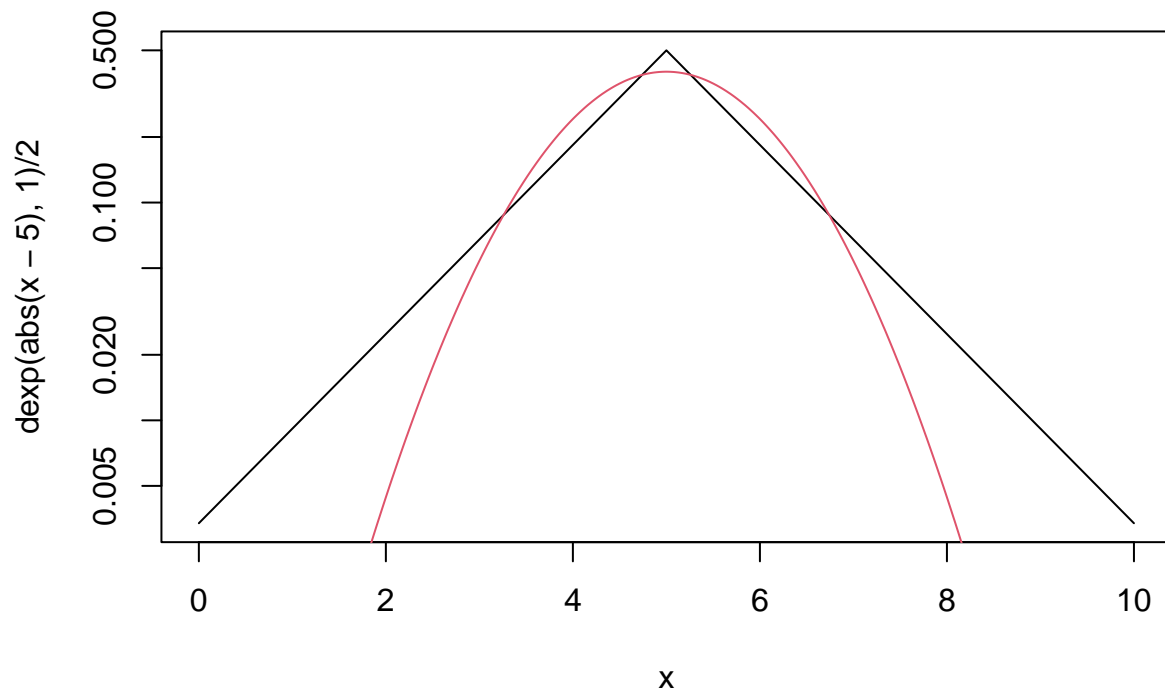


```
## laplace vs Gaussian
plot(x,dexp(abs(x-5),1)/2,type='l')
lines(x,dnorm(x,5),col=2)
```





```
plot(x,dexp(abs(x-5),1)/2,type='l',log='y')    ## same plot as last but on a log scale  
lines(x,dnorm(x,5),col=2)
```



**Question 7)** The last two panels compare a normal and a Laplace distribution with the same mean and variance. How do the two distributions compare? In particular, compare the difference in the probabilities of extreme events in each distribution.

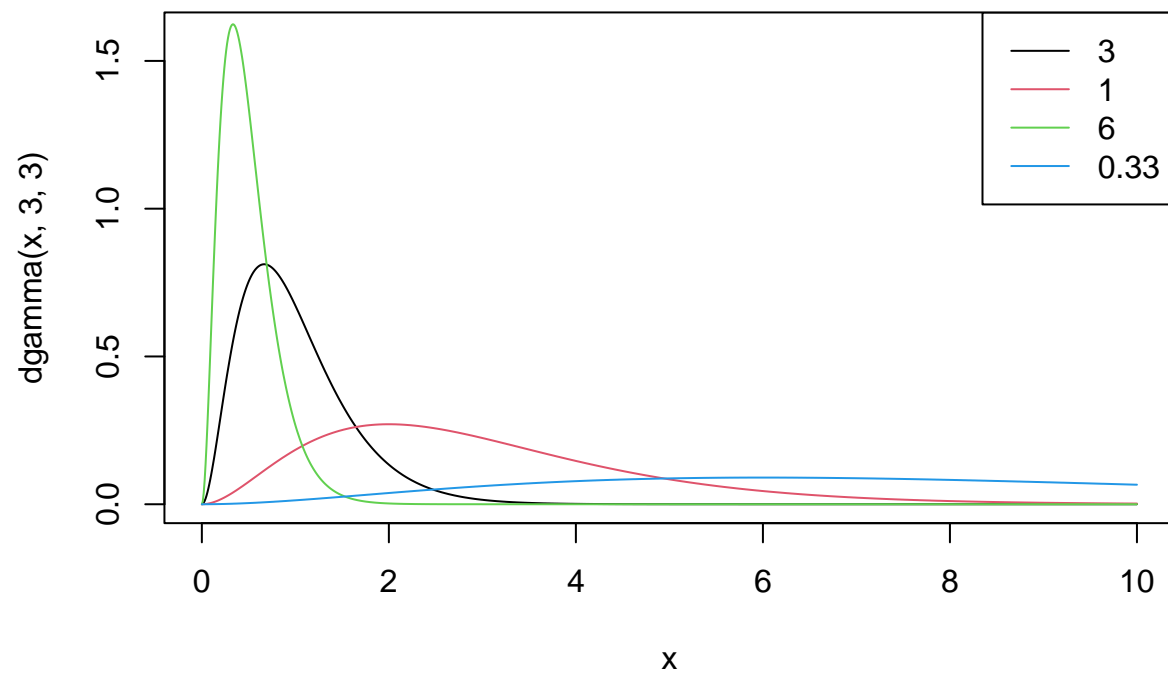
The Laplace distribution has a higher probability of values at the mean (i.e., the sharp point) and also has a higher probability of extreme events (i.e., the fatter tails). The normal distribution has a higher likelihood of values that have a relatively small deviation from the mean.

## 1.5 Gamma

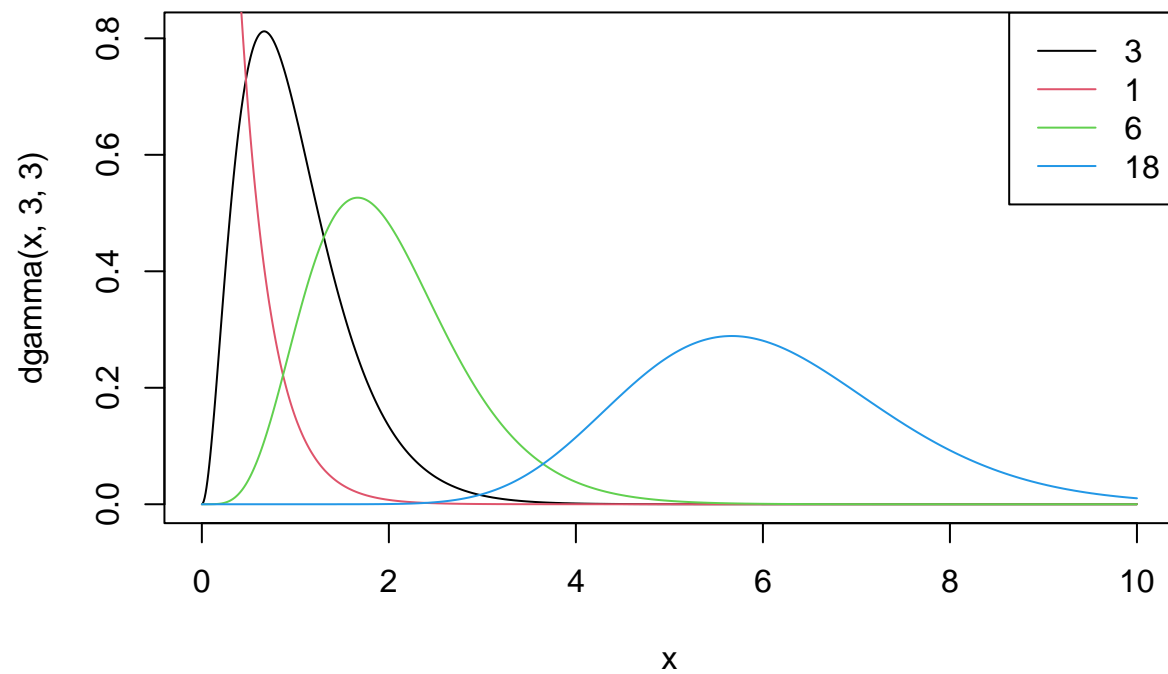
The gamma and inverse-gamma distribution are flexible distributions defined for positive real numbers. These are frequently used to model the distribution of variances or precisions (precision = 1/variance), in which case the shape and rate parameters are related to the sample size and sum of squares, respectively. The gamma is frequently used in Bayesian statistics as a prior distribution, and also in mixture distributions for inflating the variance of another distribution

```
x <- seq(0,10,by=0.01)

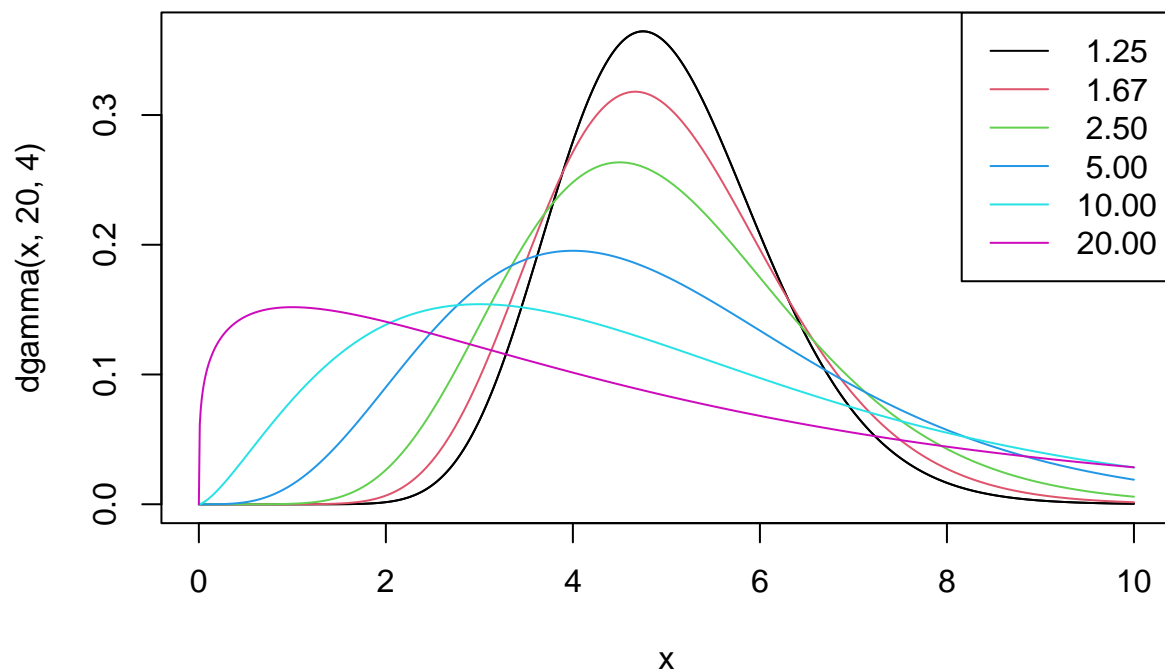
## change rate
plot(x,dgamma(x,3,3),type='l' ,ylim=c(0,1.6))
lines(x,dgamma(x,3,1),col=2)
lines(x,dgamma(x,3,6),col=3)
lines(x,dgamma(x,3,1/3),col=4)
legend("topright",legend=c(3,1,6,0.33),lty=1,col=1:4)
```



```
## change shape
plot(x,dgamma(x,3,3),type='l')
lines(x,dgamma(x,1,3),col=2)
lines(x,dgamma(x,6,3),col=3)
lines(x,dgamma(x,18,3),col=4)
legend("topright",legend=c(3,1,6,18),lty=1,col=1:4)
```



```
## change variance
a <- c(20,15,10,5,2.5,1.25)
r <- c(4,3,2,1,0.5,0.25)
plot(x,dgamma(x,20,4),type='l')
for(i in 1:6){
  lines(x,dgamma(x,a[i],r[i]),col=i)}
var = a/r^2
mean = a/r
legend("topright",legend=format(var,digits=3),lty=1,col=1:6)
```



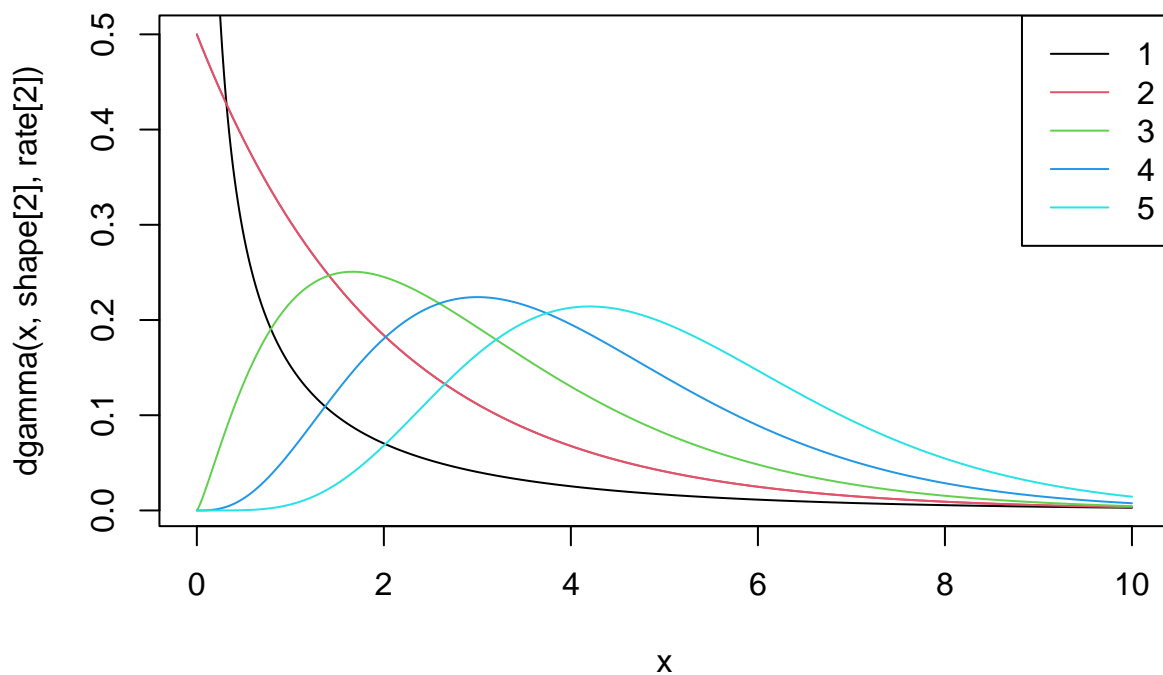
```
var
```

```
## [1] 1.250000 1.666667 2.500000 5.000000 10.000000 20.000000
```

```
mean
```

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

```
## change mean
var = 4
mean = c(1,2,3,4,5)
rate = mean/var
shape = mean^2/var
plot(x,dgamma(x,shape[2],rate[2]),type='l',col=2)
for(i in 1:5){
  lines(x,dgamma(x,shape[i],rate[i]),col=i)
}
legend("topright",legend=1:5,lty=1,col=1:5)
```



```
rate
```

```
## [1] 0.25 0.50 0.75 1.00 1.25
```

```
shape
```

```
## [1] 0.25 1.00 2.25 4.00 6.25
```

Question 8) Looking at the ‘change variance’ figure, how does the variance change as a and r increase? Qualitatively, how does this affect the mode and skew? Quantitatively, how does this affect the median (relative to the mean)?

```
df <- data.frame(a = c(20,15,10,5,2.5,1.25),
                 r = c(4,3,2,1,0.5,0.25)) %>%
  mutate(var = signif(a/r^2, digits=3),
         mean = a/r,
         median = signif(qgamma(0.5,a,r), digits=3))
```

```
kable(df)
```

a	r	var	mean	median
20.00	4.00	1.25	5	4.92
15.00	3.00	1.67	5	4.89
10.00	2.00	2.50	5	4.83
5.00	1.00	5.00	5	4.67
2.50	0.50	10.00	5	4.35
1.25	0.25	20.00	5	3.75

As  $a$  and  $r$  increase, the variance decreases. additionally, the mode increases, the skew decreases, and the median approximates the mean.

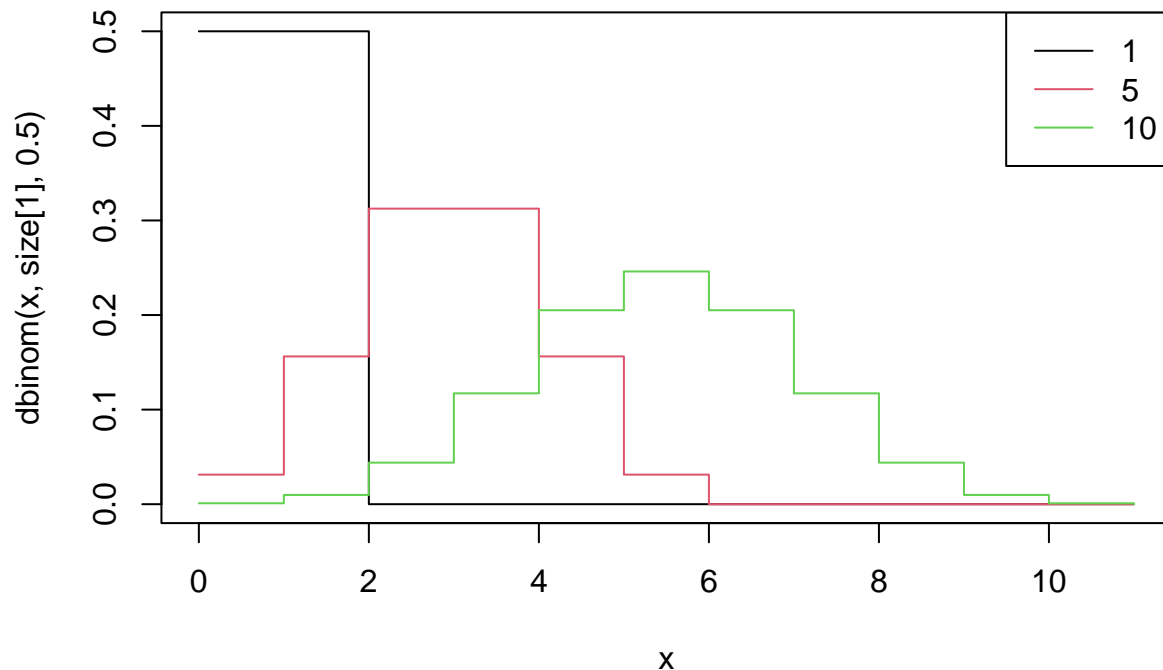
## 2 Part 2: Discrete distributions

### 2.1 Binomial

The binomial arises naturally from counts of the number of successes given a probability of success,  $p$ , and a sample size,  $n$ . You are probably already familiar with the binomial in the context of coin toss examples.

```
x <- 0:11

## vary size of sample (number of draws)
size = c(1,5,10)
plot(x,dbinom(x,size[1],0.5),type='s')
for(i in 2:3){
  lines(x,dbinom(x,size[i],0.5),type='s',col=i)
}
legend("topright",legend=size,lty=1,col=1:3)
```

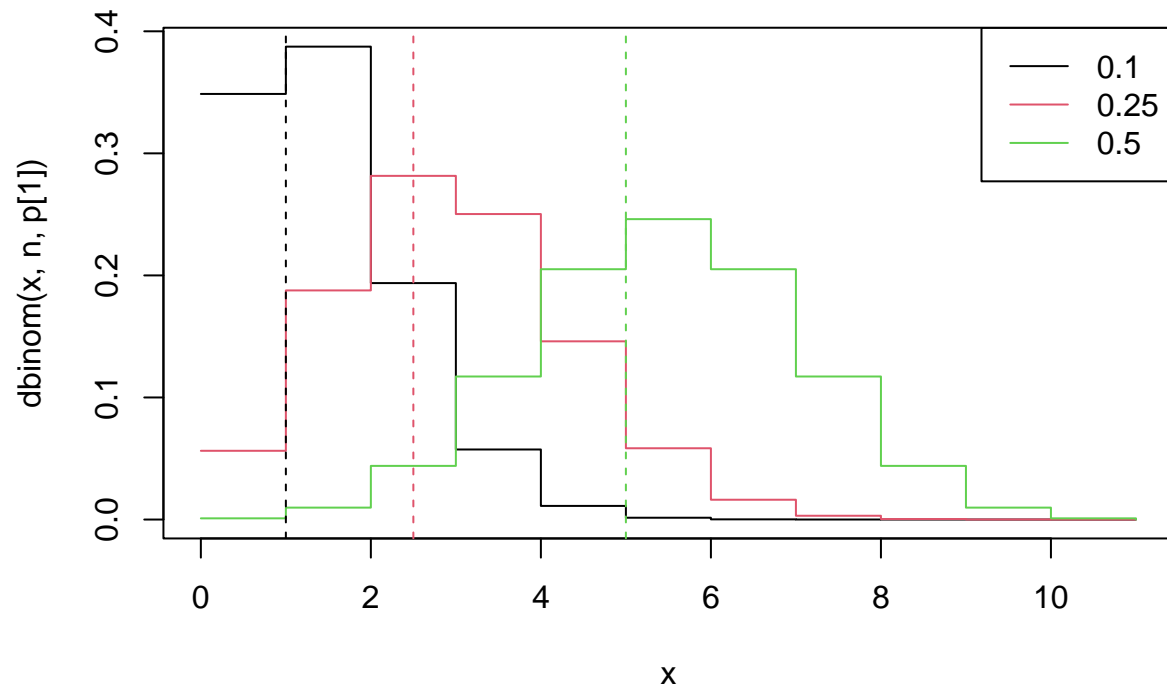


```
## vary probability
n = 10
p = c(0.1,0.25,0.5)
plot(x,dbinom(x,n,p[1]),type='s')
```

```

for(i in 2:3){
  lines(x,dbinom(x,n,p[i]),col=i,type='s')
}
abline(v = n*p,col=1:3,lty=2)
legend("topright",legend=p,lty=1,col=1:3)

```

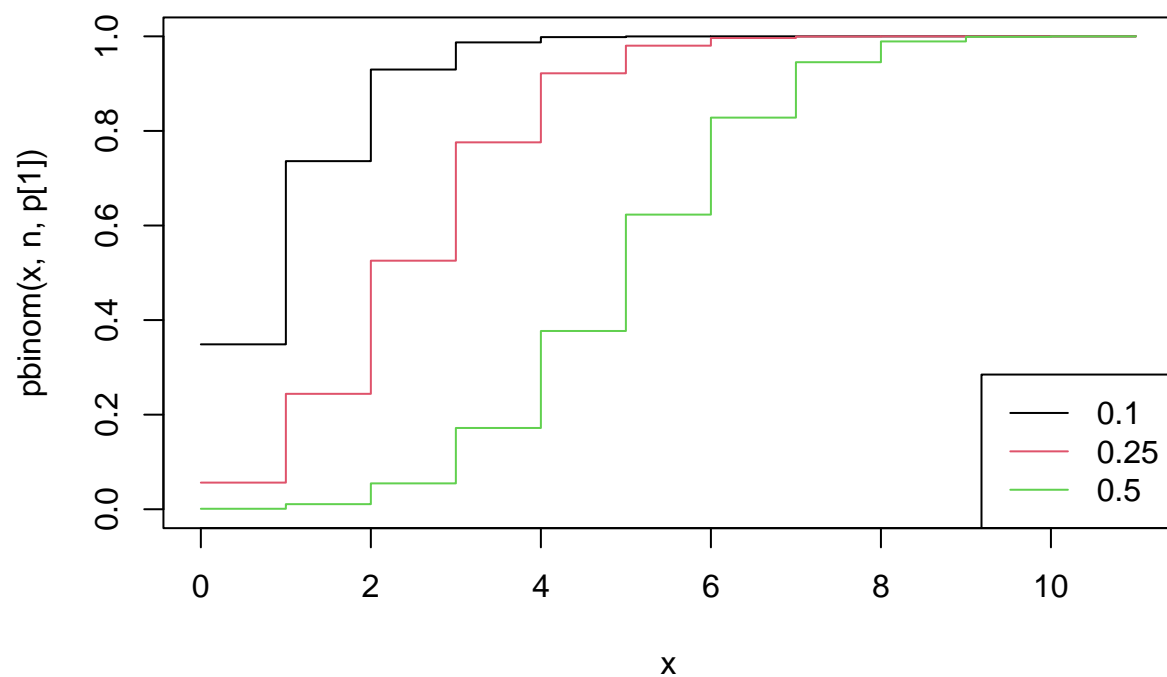


```

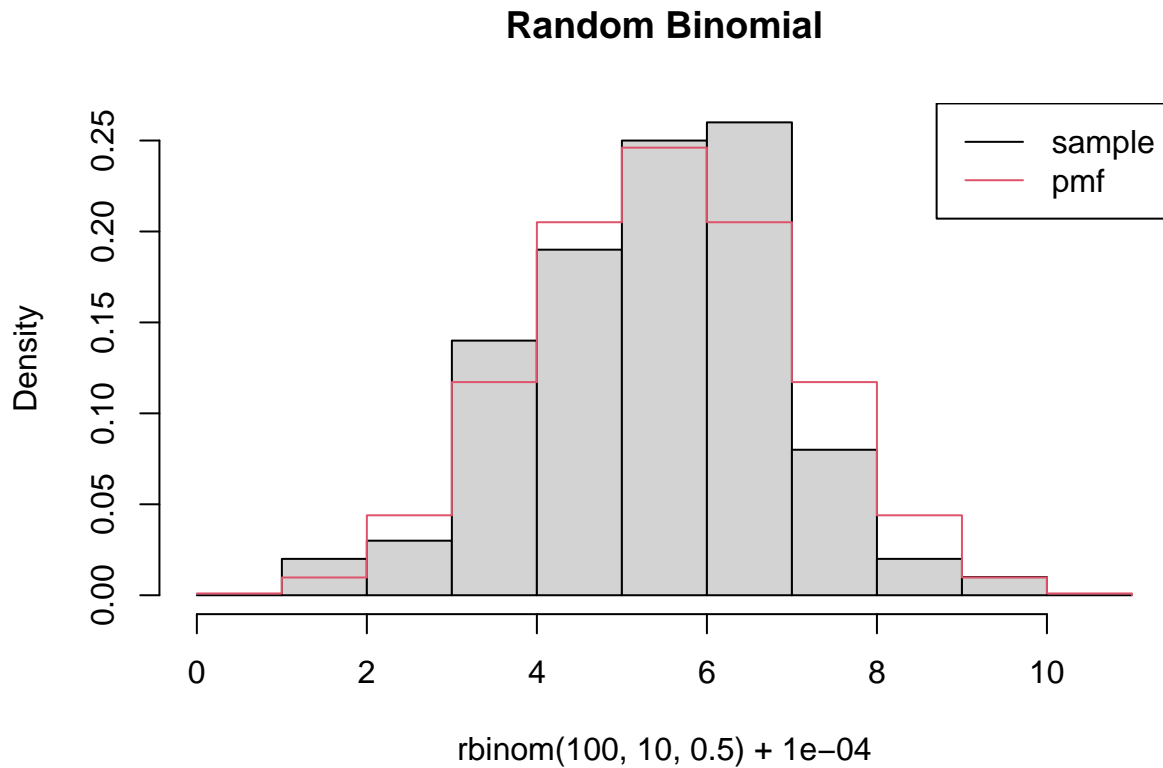
## CDF
plot(x,pbinom(x,n,p[1]),type='s',ylim=c(0,1))
for(i in 2:3){
  lines(x,pbinom(x,n,p[i]),col=i,type='s')
}
legend("bottomright",legend=p,lty=1,col=1:3)

```





```
## Random samples
hist(rbinom(100,10,0.5)+0.0001, ## small amount added because
     probability=TRUE,         ## of way R calculates breaks
     breaks=0:11,main="Random Binomial")
lines(x,dbinom(x,10,0.5),type='s',col=2) #Analytical solution
legend("topright",legend=c("sample","pmf"),lty=1,col=1:2)
```



Question 9) Consider a binomial distribution that has a constant mean,  $np$ . What are the differences in the shape of this distribution if it has a high  $n$  and low  $p$  vs. a low  $n$  and high  $p$ ?

```
x <- 0:11
np <- 5 #set a constant mean
n <- c(6,10,100)
p <- np/n

result.list <- list()

for (i in 1:length(p)){

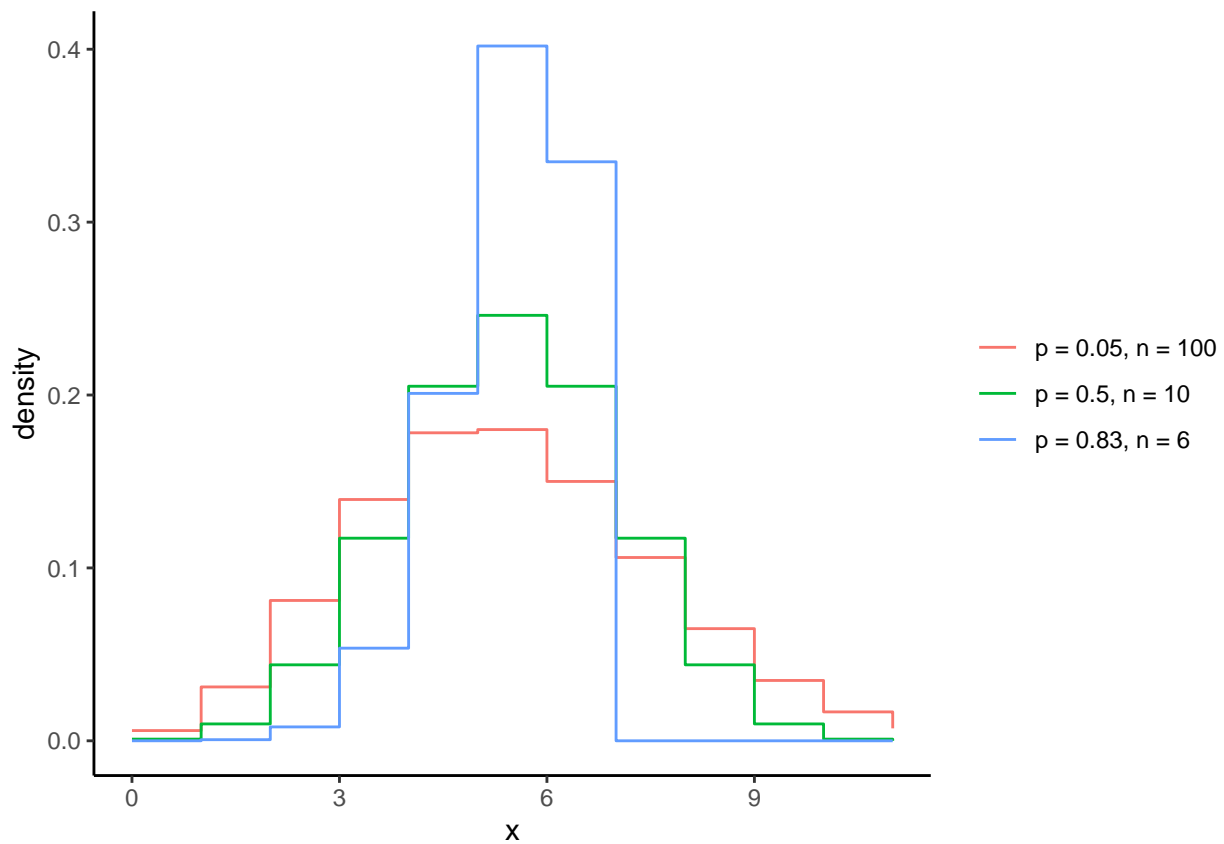
  result.df <- data.frame(p=p[i],
                        n=n[i],
                        iteration=paste0("p = ",round(p[i], digits=2),
                                           ", n = ",n[i]),
                        x=x,
                        y=dbinom(x,n[i],p[i]))

  result.list <- c(result.list, list(result.df))
}

df <- bind_rows(result.list)

ggplot(df, aes(x,y,color=iteration))+
  theme_classic()+
```

```
geom_step()+
labs(x="x", y="density", color=NULL)
```

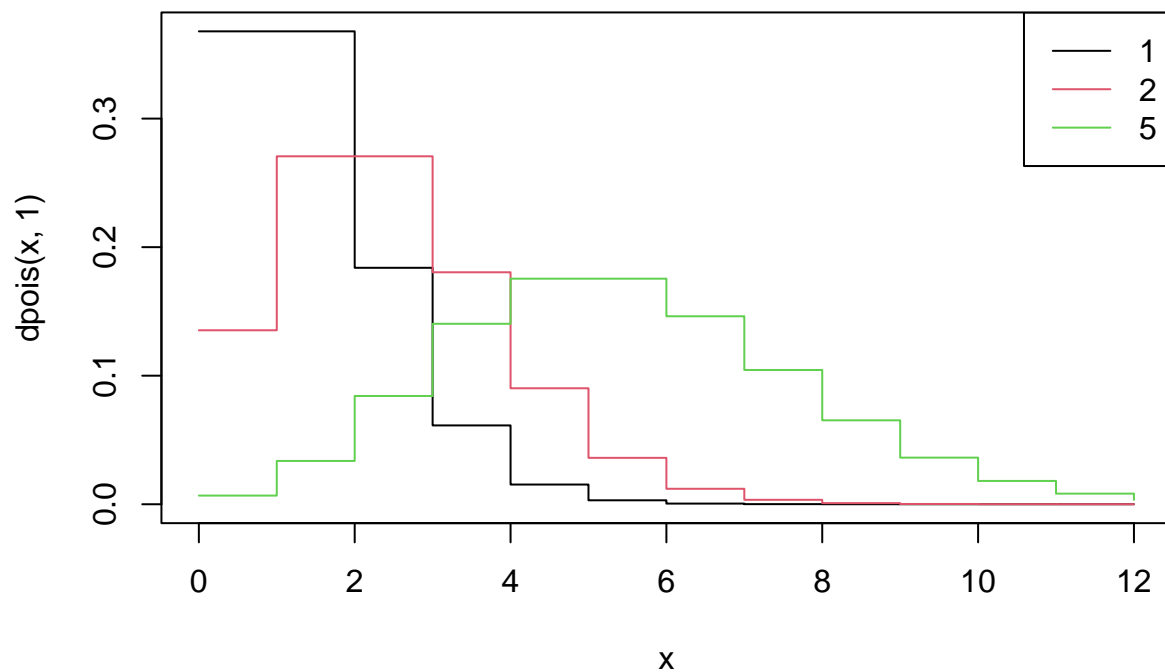


In the figure above, I created three binomial distributions that each have a mean of 5. A distribution with a high  $p$  and a low  $n$  will be more narrow (e.g., blue) while distribution with a low  $p$  and a high  $n$  will be more wide (e.g., red).

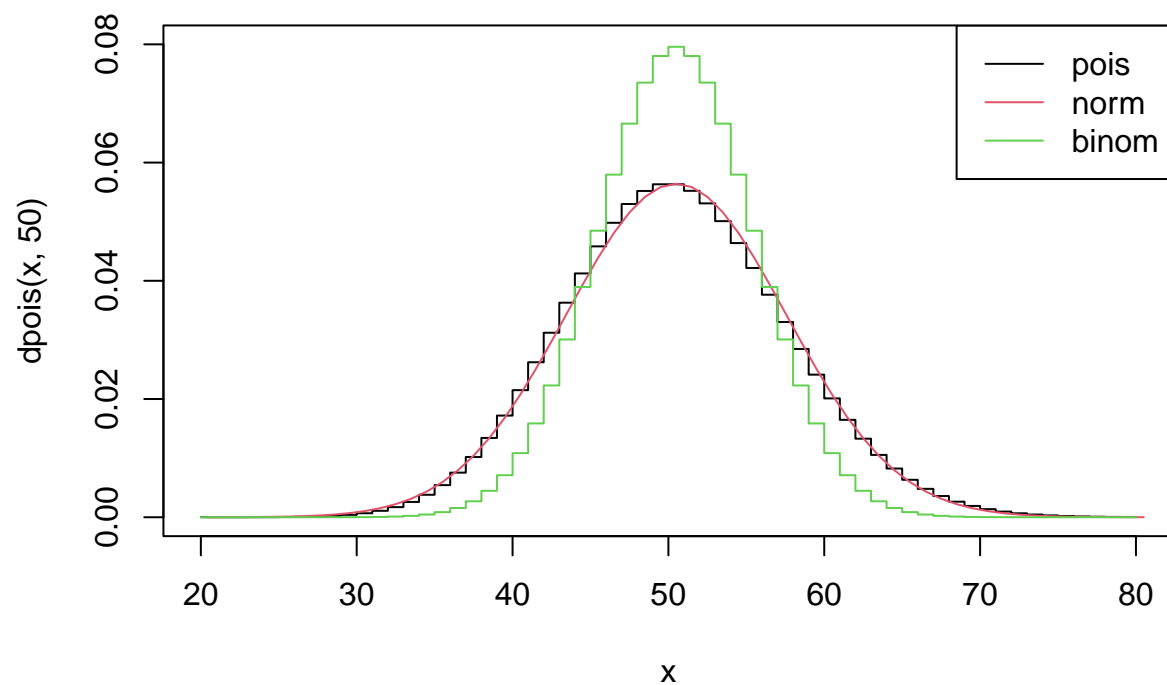
## 2.2 Poisson

The Poisson is also very common for count data and arises as the number of events that occur in a fixed amount of time (e.g. number of bird sightings per hour), or the number of items found in a fixed amount of space (e.g. the number of trees in a plot). Unlike the Binomial distribution the Poisson doesn't have a fixed upper bound for the number of events that can occur.

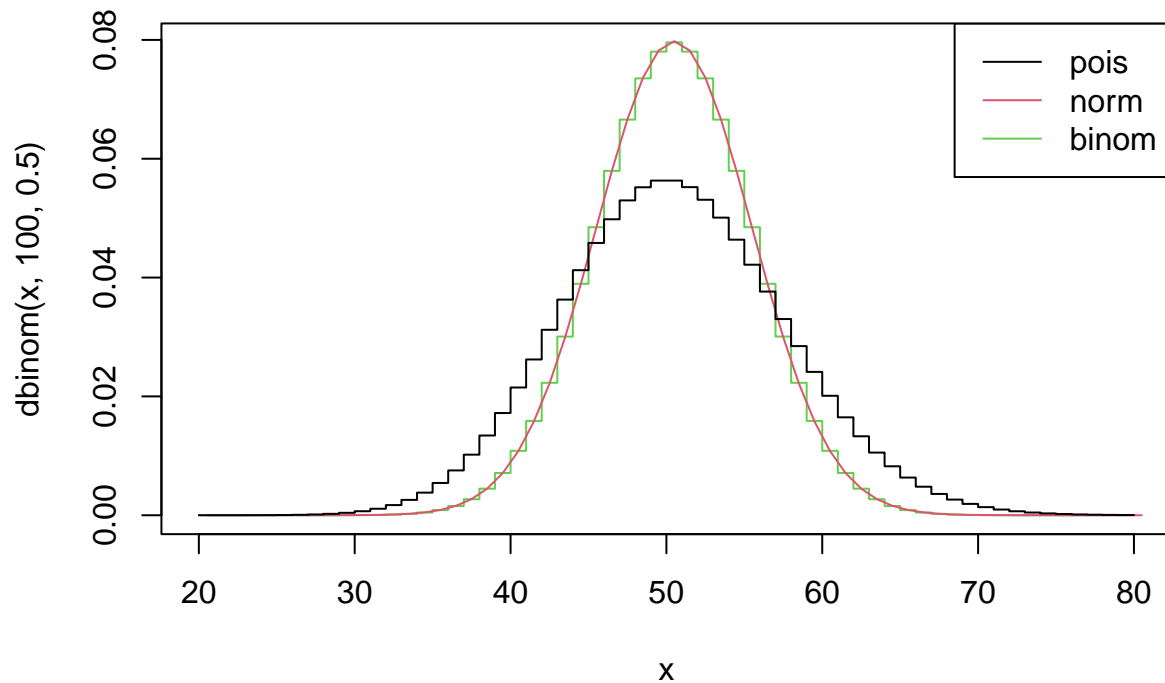
```
x <- 0:12
plot(x,dpois(x,1),type='s')
lines(x,dpois(x,2),type='s',col=2)
lines(x,dpois(x,5),type='s',col=3)
legend("topright",legend=c(1,2,5),lty=1,col=1:3)
```



```
x <- 20:80
plot(x,dpois(x,50),type='s',ylim=c(0,0.08))      #Poisson with mean 50 (variance = 50)
lines(x+0.5,dnorm(x,50,sqrt(50)),col=2)         #Normal with mean and variance of 50
lines(x,dbinom(x,100,0.5),col=3,type='s')       #Binomial with mean 50 (variance = 25)
legend("topright",legend=c("pois","norm","binom"),col=1:3,lty=1)
```



```
plot(x,dbinom(x,100,0.5),type='s',col=3)           #Binomial with mean 50 (variance = 25)
lines(x+0.5,dnorm(x,50,sqrt(25)),col=2)           #Normal with mean 50 and variance of 25
lines(x,dpois(x,50),col=1,type='s')               #Poisson with mean 50 (variance = 50)
legend("topright",legend=c("pois","norm","binom"),col=1:3,lty=1)
```



The last two panels depict a comparison of the Poisson, Normal, and Binomial with the same mean and a large sample size. The Poisson and Binomial are identical in the two figures but in the first the normal has the same variance as the Poisson and the second it is the same as the binomial.

**Question 10) Normal distributions are often applied to count data even though count data can only take on positive integer values. Is this fair is this to do in these two examples? (i.e. how good is the normal approximation)**

In these two examples, applying the normal distribution to the count data is probably a good approximation because the count data is far from the bounds of the Poisson and Binomial distributions (i.e., zero).

**Question 11) Would the normal be a fair approximation to the Poisson curves for small numbers (the first panel)? How about for the Binomial for small numbers (earlier panel of figures on the Binomial)?**

No, the normal would be a poor approximation of either the Poisson or Binomial distributions at low values because the normal would suggest a relatively high likelihood of negative values (which are impossible for count data.)

**Question 12) Is the Poisson a good approximation of the Binomial?**

The Poisson can be a good approximation of the Binomial if the variance of the Binomial is close to it's mean. Here, I derive values of  $n$  and  $p$  such that the mean is 50 and the variance is 49:

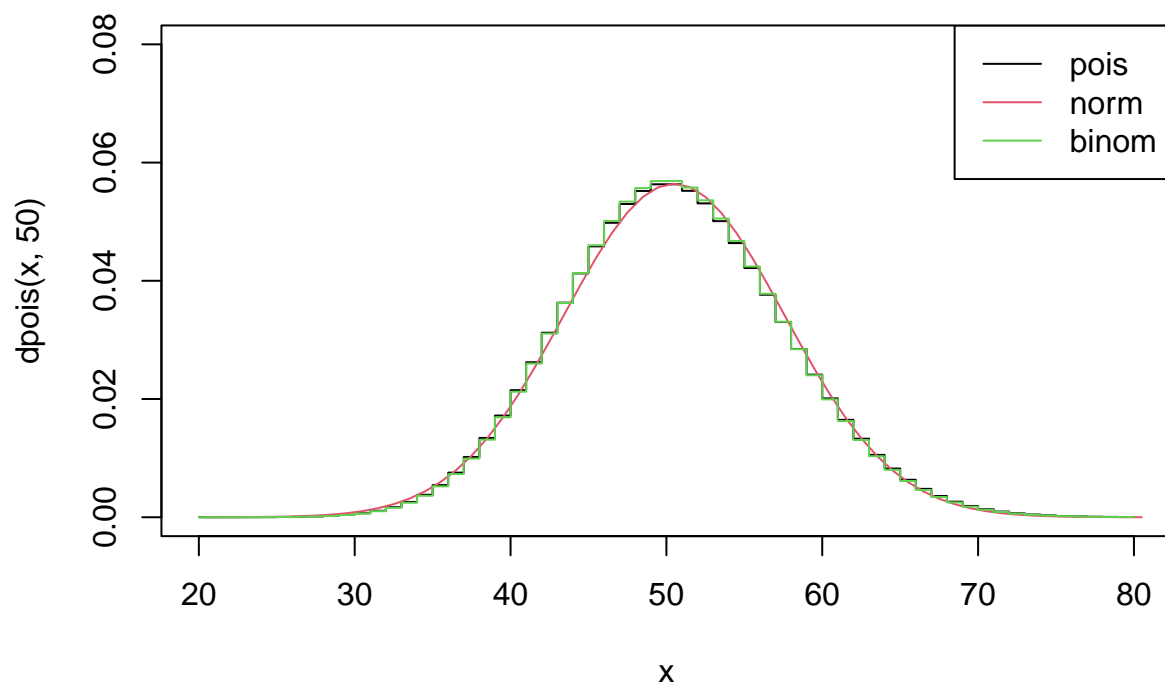
$$\begin{aligned}
 E(x) &= np \\
 50 &= np \\
 n &= \frac{50}{p} \\
 Var(x) &= np(1-p) \\
 49 &= np(1-p) \\
 49 &= \frac{50}{p}p(1-p) \\
 49 &= 50(1-p) \\
 1 &= 50p \\
 p &= \frac{1}{50} \\
 n &= \frac{50}{0.02}
 \end{aligned}$$

```

#derive n and p such that mean = 50 and variance = 49
#n <- 50/p

x <- 20:80
plot(x,dpois(x,50),type='s',ylim=c(0,0.08))      #Poisson with mean 50 (variance = 50)
lines(x+0.5,dnorm(x,50,sqrt(50)),col=2)          #Normal with mean and variance of 25
lines(x,dbinom(x,2500,0.02),col=3,type='s')      #Binomial with mean 50 (variance = 49)
legend("topright",legend=c("pois","norm","binom"),col=1:3,lty=1)

```



13) Is it possible to choose the parameters so that the Poisson and Binomial to both have the same mean and variance? If so what is this parameterization?

No. The mean and variance of the Poisson are equal to each other, but setting the variance of a binomial equal to its mean will result in a  $p$  of 0. I tried it anyways:

$$\begin{aligned}E(x) &= np \\Var(x) &= np(1 - p) \\np &= np(1 - p) \\1 &= (1 - p) \\p &= 0\end{aligned}$$

> If make  $p$  extremely small, it will be possible to have the Binomial approximate the Poisson. But, we cannot make them equal.

## 2.3 Negative binomial

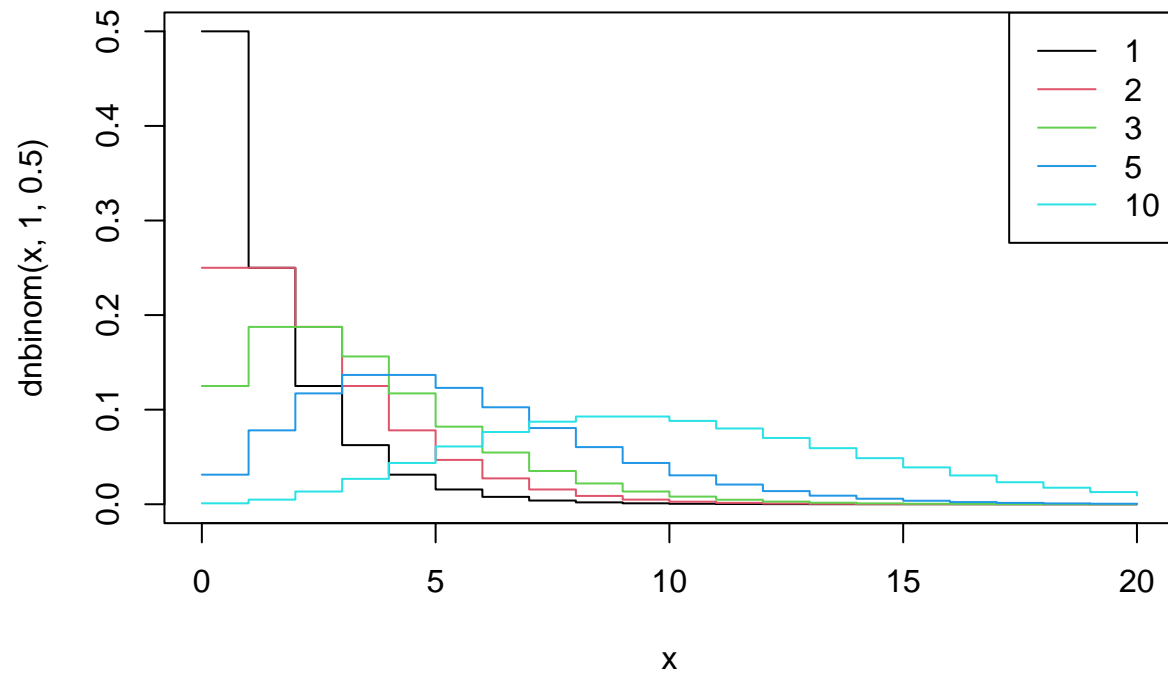
The negative binomial has two interesting interpretations that are subtly different from either the Poisson or Binomial. In the first case, it is the number of trials needed in order to observe a fixed number of occurrences, which is the opposite from the Binomial's number of occurrences in a fixed trial size and thus where it gets its name. The Negative Binomial also arises as the distribution of number of events that occur in a fixed space or time when the rate is not constant (as in the Poisson) but varies according to a Gamma distribution. Hence the Negative Binomial is also used to describe data that logically seems to come from a Poisson process but has greater variability that is expected from the Poisson (which by definition has a variance equal to its mean). The Geometric distribution arises as a special case of the negative binomial where the number of occurrences is fixed at 1.

```
x <- 0:20
## negative binomial

## vary size
plot(x,dnbinom(x,1,0.5),type="s",main="vary size")
lines(x,dnbinom(x,2,0.5),type="s",col=2)
lines(x,dnbinom(x,3,0.5),type="s",col=3)
lines(x,dnbinom(x,5,0.5),type="s",col=4)
lines(x,dnbinom(x,10,0.5),type="s",col=5)
legend("topright",legend=c(1,2,3,5,10),col=1:5,lty=1)
```

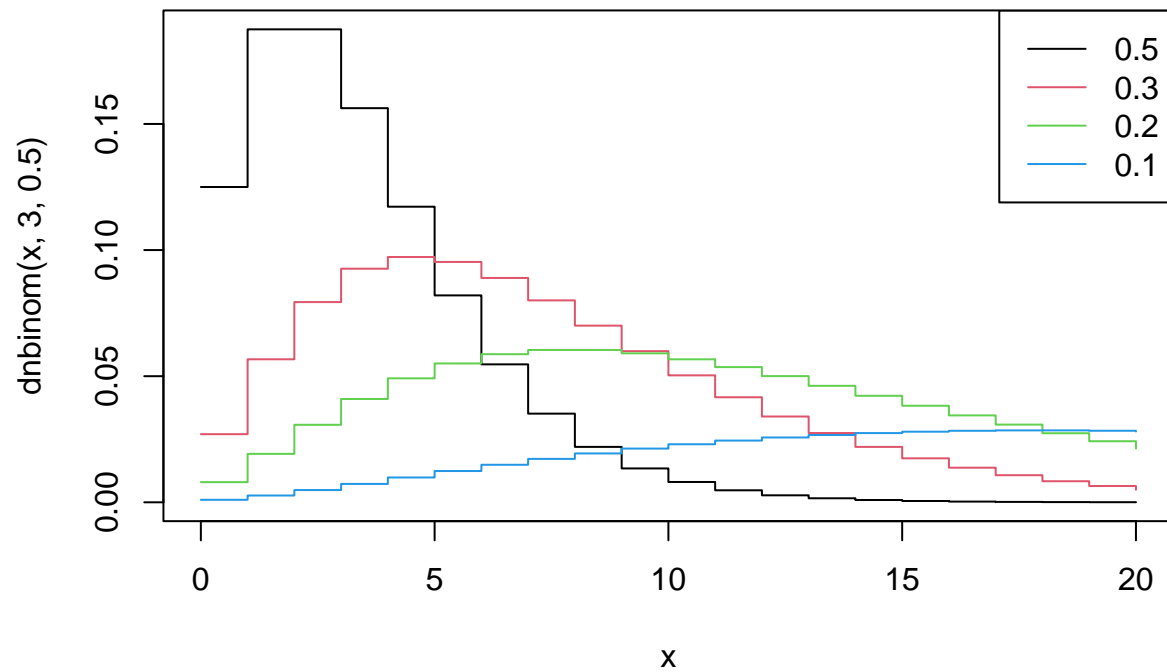


## vary size

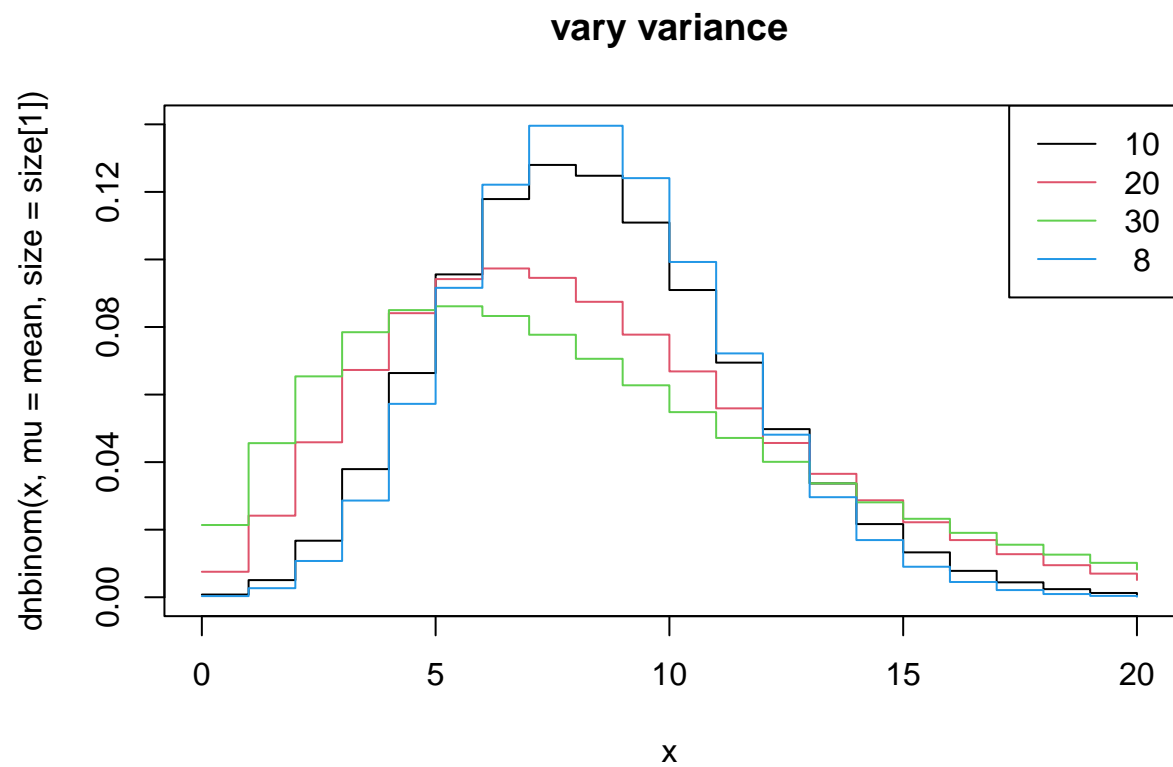


```
## vary probability
plot(x,dnbinom(x,3,0.5),type="s",main="vary probability")
lines(x,dnbinom(x,3,0.3),type="s",col=2)
lines(x,dnbinom(x,3,0.2),type="s",col=3)
lines(x,dnbinom(x,3,0.1),type="s",col=4)
legend("topright",legend=c(0.5,0.3,0.2,0.1),col=1:5,lty=1)
```

## vary probability



```
## vary variance , alternate parameterization
mean = 8
var = c(10,20,30)
size = mean^2/(var-mean)
plot(x,dnbinom(x,mu=mean,size=size[1]),type="s",ylim=c(0,0.14),main="vary variance")
lines(x,dnbinom(x,mu=mean,size=size[2]),type="s",col=2)
lines(x,dnbinom(x,mu=mean,size=size[3]),type="s",col=3)
legend('topright',legend=format(c(var,mean),digits=2),col=1:4,lty=1)
lines(x,dpois(x,mean),col=4,type="s")
```

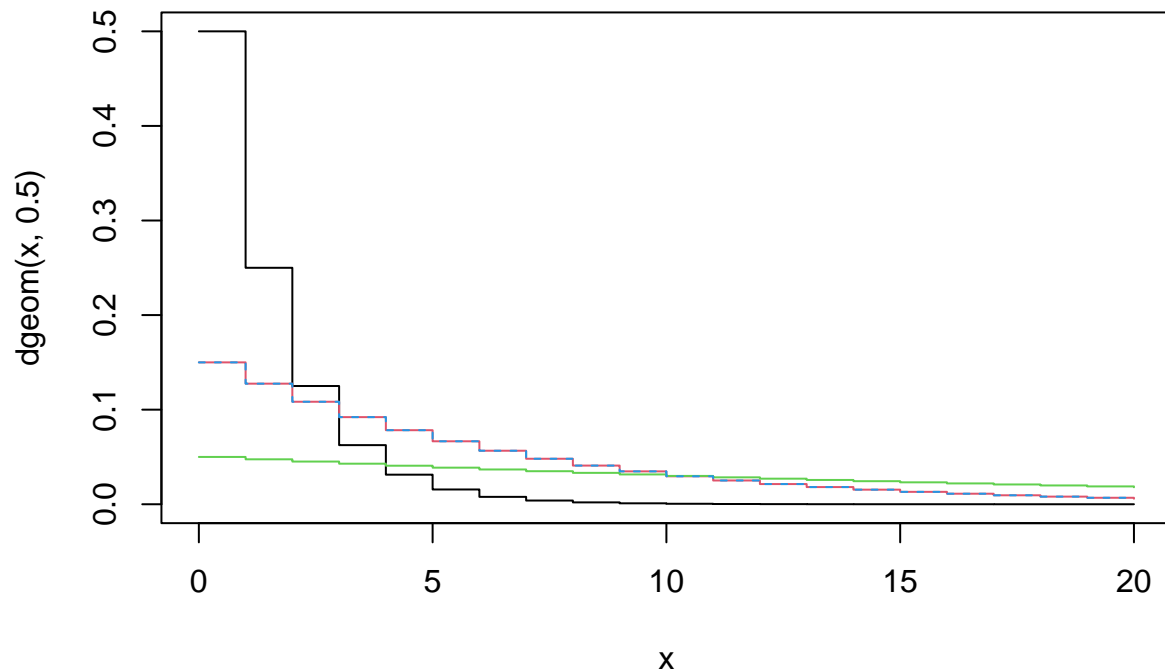


```
## NB as generalization of pois with inflated variance
```

```
## geometric
```

```
plot(x,dgeom(x,0.5),type="s",main="Geometric")
lines(x,dgeom(x,0.15),type="s",col=2)
lines(x,dgeom(x,0.05),type="s",col=3)
lines(x,dnbinom(x,1,0.15),type="s",col=4,lty=2)
```

## Geometric



```
## geometric as special case of NB where size = 1
```

Question 14) In the ‘vary size’ panel, what are the means of the curves?

$$E(x) = \frac{n(1-p)}{p}$$

$$E(x) = \frac{n(1-0.5)}{0.5}$$

$$E(x) = n$$

```
df <- data.frame(size = c(1,2,3,5,10),
                 p=0.5) %>%
  mutate(mean=(size*(1-p))/p)
```

```
kable(df)
```

size	p	mean
1	0.5	1
2	0.5	2
3	0.5	3
5	0.5	5
10	0.5	10

Question 15) In the “vary variance” panel, how does the shape of the Negative Binomial compare to a Poisson with the same mean?

The Poisson (blue) looks similar to a Negative Binomial with a variance of 10 (black), although the higher variance of the negative binomial is seen as slightly broader tails.