The binomial and normal distributions are the two most important distribution in this class. If X is a binomial random variable we know these facts"

- The two parameters are n– the number of trials, and p the success probability
- The range of values for X is $0,1,2,\ldots,n$
- The *distribution* of X is
$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$$
- The mean of X is $\mu=np$ and variance $\sigma^2=np(1-p)$.

For X a normal random variable we know the following

- There are two numbers: μ, the mean, and σ^2, the variance, which determine the distribution of Z
- X can have any possible value, but the more likely values are determined by the graph giving the distribution of X.
- The probability that $a<X<b$ is given by the area under the graph between a and b.
- This graph has a special bell shape and is symmetric about μ and has roughly 68% of its area within 1 σ of μ, 95% within 2 σ and 99.7% within 3 σ.

The goal of this project is to see how to use the computer to do things with the binomial and normal distributions and to see that the normal distribution is a good approximation for the binomial distribution.

1 Simulating, finding probabilities

In R there are various functions for dealing with distributions. In particular you can find random samples of data using the "r" functions, and find probabilities using the "p" functions, and distributions with the "d" functions.

For example, to find a random sample of size 10 from the binomial with $n=10$ and $p=1/3$ we have

```
> x = rbinom(10, size = 10, p = 1/3)
> x

 [1] 2 2 3 4 5 4 2 4 1 4
```

The "r" stands for random sample and "binom" for binomial. For a normal sample, it is similar. For a random sample from normal with $\mu=10$ and $\sigma=1/3$ we would have

```
> y = rnorm(10, mean = 10, sd = 1/3)
> y
```

```
 [1]  9.817  9.858 10.377  9.428  9.902  9.525  9.496 10.374  9.791  9.721
```

Question 1: Make 10 random samples from a binomial with $n = 100$ and $p = .25$. Store the results into `x`. Use `range()` to find the smallest and largest.

Question 2: Make 10 random samples from a normal with $\mu = 25$ and $\sigma^2 = 100 \cdot .25 \cdot .75$. Store your results into `y`.

Question 3: Compare the results from the last two. Are the numbers similar in size? If not, you probabily used the wrong value for σ on the normal. Redo the exercises, only this time storing 100 answers so that you can do a side-by-side boxplot. Are the boxplots similar?

Question 4: A density plot of each data set can be done with these commands

```
> plot(density(x), type = "l")
> lines(density(y))
```

Type these in. Do the shapes look the same?

1.1 The distribution of the binomial

`R` can find probabilities of these random distributions.

The formula for the distribution of a binomial is given by `dbinom`. For example The probability of exactly 5 successes in 10 trials with success probability $p = 1/2$ can be found with

```
> dbinom(5, size = 10, p = 1/2)
```

```
[1] 0.2461
```

The "p" functions answer the questions $P(X \leq x)$. So for example, the probability that a binomial random variable with $n = 10$ and $p = 1/2$ is 5 or less is given by

```
> pbinom(5, size = 10, p = 1/2)
```

```
[1] 0.623
```

Question 5: Historically, a student has a 30 percent chance of finding a parking space near their class. What is the probability that in a week of parking (5 days), a student never gets to park close by? What is the probability the number of times is 2 or less?

Question 6: A student takes a 20 question multiple choice question by guessing for each answer. Each question has 4 parts. What is the expected number of correct answers? What is the probability he gets 70% or more on the test?

Question 7: Repeat the above for a different student. She also guesses, but for each question can eliminate 3 of the 5 possible answer for each question, so she only guesses between 2 possible answers.

Question 8: Find the probability that a binomial random variable with $n = 10$ and $p = 1/2$ actually equals its expected value.

Question 9: A survey of 1000 people finds that 62% agree with the question. Suppose the actual population percentage is really 60%. Find the probability that the sample percentage is more than 62%?

(Hint: why is this a binomial problem? Why is $P(X < x) = 1 - P(X >= x)$.)

1.2 The distribution of the normal

The same is true for the normal distribution. Most useful is the "pnorm" function. For example. The probability a standard normal is less than 2 is given by

```
> pnorm(2)

[1] 0.9772
```

The default value for μ is 0, and σ is 1. To change these, we set the arguments `mean=` and `sd=`.

The probability a normal with mean 100 and standard deviation 20 is more than 110 is the difference

```
> 1 - pnorm(110, mean = 100, sd = 20)

[1] 0.3085
```

(Why is there a 1−?

The probability a normal with mean 400 and standard deviation 50 is in the interval $[325, 425]$ is given by

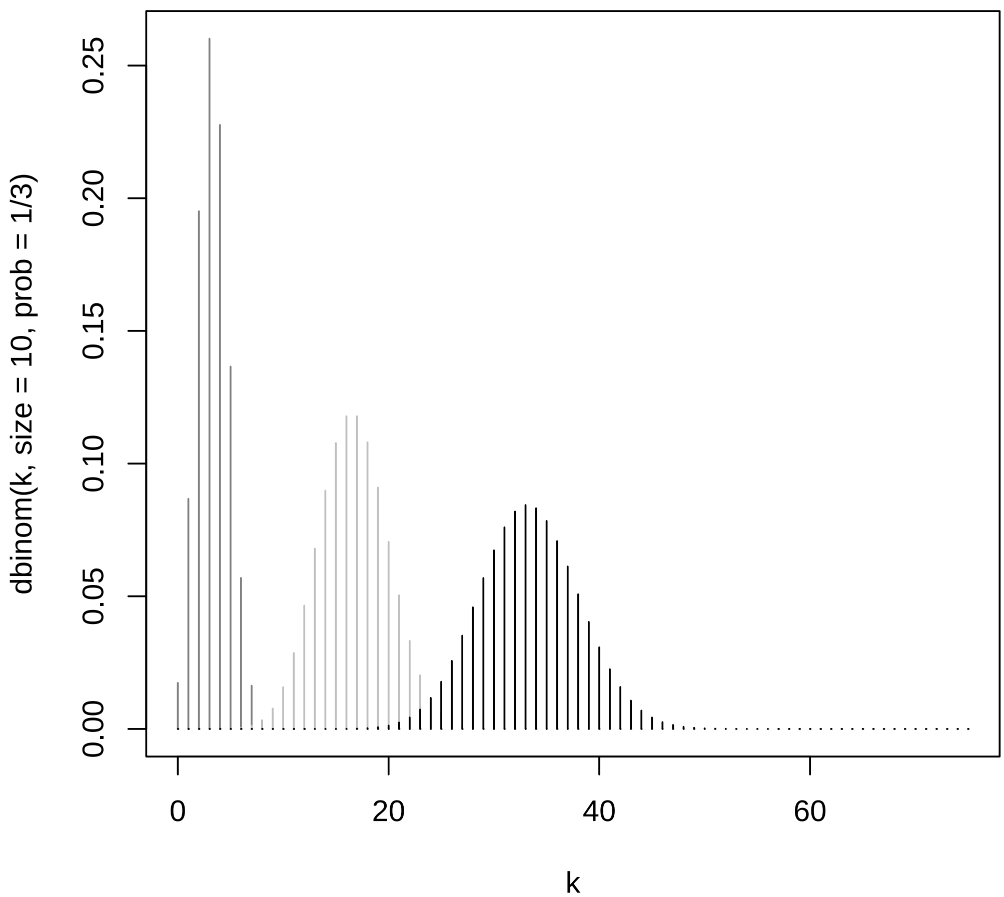
```
> pnorm(425, mean = 400, sd = 50) - pnorm(325, mean = 400, sd = 50)

[1] 0.6247
```

Question 10: Find probability that Z, a standard normal, is between −2 and 2.

Question 11: The table in the book finds the probabilities a standard normal is between 0 and z for $z > 0$. Use "pnorm" to find $P(0 < Z < 1.5)$.

2 The binomial distribution and normal distribution

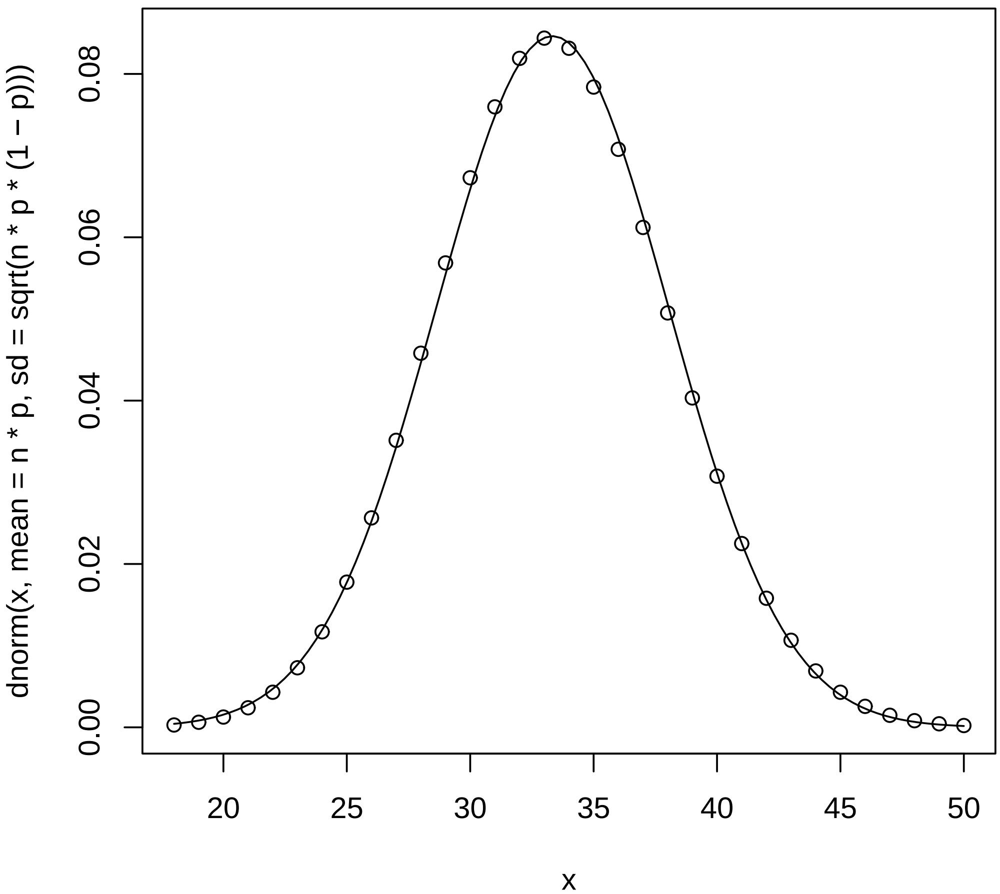
We can graph the binomial distributions and get a sense of what they look like. For example. For $p = 1/3$ this graph shows the binomial for $n = 10, 50$ and 100.



Notice the shapes are fairly bell shaped. If you didn't, don't worry Bernoulli and others did. They saw that they could use the normal curve to *approximate* the binomial. As binomials were hard to compute before computers, this was very advantageous.

Let's compare the graph of the binomial with $n = 100$, $p = 1/3$ with a normal with the same mean and standard deviation ($\mu = 100/3$, $\sigma^2 = 100(1/3)(2/3)$.

```
> n = 100
> p = 1/3
> curve(dnorm(x, mean = n * p, sd = sqrt(n * p * (1 - p))), 18, 50)
> x = 18:50
> points(x, dbinom(x, n, p))
```



To take advantage of this relationship, the **normal approximation** says that the area to the left of $b+1/2$ for the normal should be the probability a binomial is b or less. And the area to the right of $a-1/2$ should the probability a binomial is a or more.

For example

```
> n = 100
> p = 1/3
> b = 40
> pbinom(b, n, p)

[1] 0.9341

> pnorm(b + 1/2, n * p, sqrt(n * p * (1 - p)))

[1] 0.9358
```

These are close, but not exact as it is only *approximately correct*.

Question 12: Use the normal approximation to find the probability that in 1000 samples with $p = .6$ the number of successes is between 620 and 1000.

(Hint, think binomial $n = 1000$, $p = .6$ and use the normal approximation.)

Write down your commands.

Question 13: There are 50,000 fans at a rock concert and historically 1% try to sneak in bottles. Find the probability that security nabs fewer than 200 at a given concert. Write the answer and your commands to find it.

(Hint, again, this is meant to be a binomial problem.)