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

```
> rbinom(10,size=10,p=1/3)
 [1] 1 3 7 2 5 4 4 3 4 2
```

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

```
> rnorm(10, mean=10,sd=1/3)
 [1]  9.661  9.588 10.599  9.964  9.983 10.252 10.202 10.054
 [9]  9.912  9.963
```

1.1 Make 10 random samples from a binomial with $n = 100$ and $p = .25$. write down the answers and circle the largest and smallest.

1.2 Make 10 random samples from a normal with $\mu = 25$ and $\sigma^2 = 300/16$. Write down your answers using 1 decimal place (eg. 25.1 not 25.10324) and circle the largest and smallest.

1.3 Compare the results from the last two. Are the numbers similar in size? Redo the exercises, only this time storing the answers so that you can do a side-by-side boxplot. Sketch the boxplot.

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 $P(X < x)$. 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
```

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

```
> pnorm(2)                    # note mu=0, sd=1 is deafult!
[1] 0.9772
```

And the probability a standard normal is between 2 and 1 is given by the difference

```
> pnorm(2) - pnorm(1)
[1] 0.1359
```

1.4 Find probability that Z, a standard normal, is between -2 and 2.

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

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

1.7 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 $P(X < x) = 1 - P(X >= x)$.)

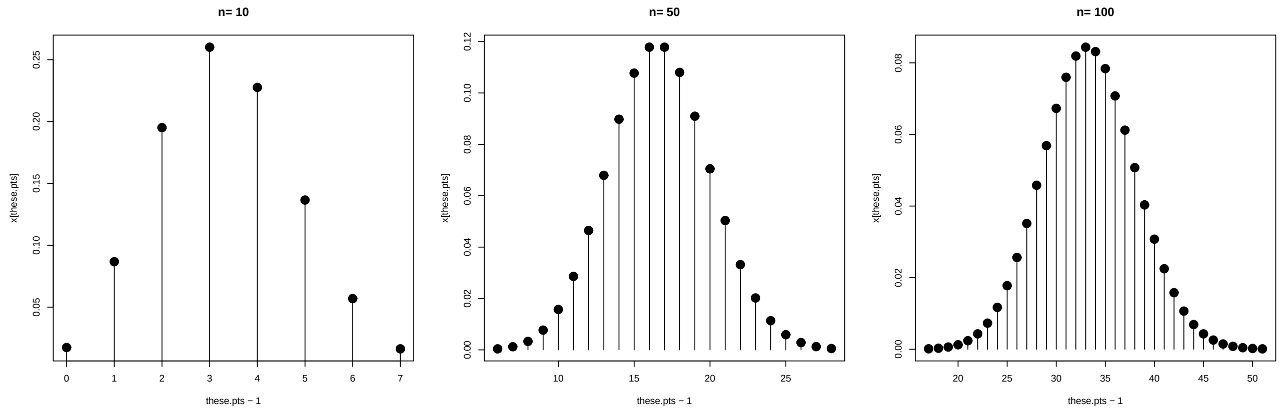

Figure 1: Binomial with n=10,50,100, p = 1/2

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 $p = 1/3$ these graphs show the binomial for $n = 10, 50$ and 100.

Notice the shape is 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*.

2.1 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.

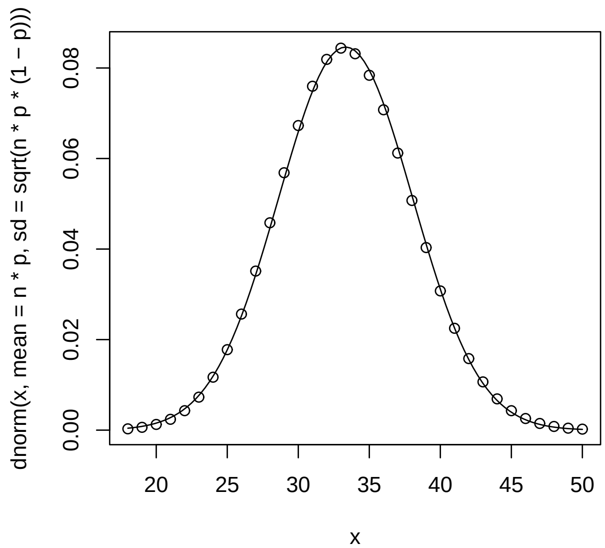

Figure 2: Normal with same mean and sd as binomial. Close fit.

2.2 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.)