The Binomial Distribution

Bernoulli Trials:

A Bernoulli trial is any experiment in which we observe whether or not a certain event A occurs. If A occurs we call it a success. If A does not occur we call it a failure. The probability of success is $p=P(A)$ and the probability of failure is $q=1-P(A)$. $q=1-p$.

Ex. Toss a coin. Observe whether or not we get heads. If the coin is a fair coin, $p=.5$.
If the coin is not a fair coin, p could be 1/3 and q then would be 2/3. If p is 1/4 then q is 3/4

Ex. Success can be a bad thing. Check a person for a certain disease. Since we are observing whether or not the person has the disease, success is the event the person has the disease. The value of p is the rate at which the disease occurs.

Ex. Assume electronic games are produced continuously so the supply is very large and there is some constant proportion , p, of defective games. Test a game to see if it is defective.

A Binomial Experiment consists of repeating a Bernoulli trial a fixed number of times, the trials are independent, p never changes.
Ex. Toss a coin 10 times. Each coin toss is a Bernoulli trial. The experiment of tossing the coin 10 times is the binomial experiment. The trials are independent and p is always the same.

Ex. Check 50 randomly selected people for a certain blood type
Ex. Test 15 electronic games to see if they are defective. We assume there is an unlimited supply of games and the probability of a defective game is always the same. This is different from having 100 electronic games in which 5 are defective and we withdraw 15. In that case, p would change as games were withdrawn.
Ex. A batter going up to bat 5 times. Is p constant here? If his batting average is changed each time he goes to bat then p is not constant. But we could imagine there is some fixed probability that he will hit the ball which does not change in only 5 times to bat and then p does not change.

The Binomial Distribution:
Suppose we have a binomial experiment with n trials and probability of success on any trial equal to p. Define the random variable X to be the number of successes in n trials. Then X has a binomial distribution and we write $X \sim B(n,p)$.

Ex. A coin is weighted so that the probability of heads is $p=.3$. We toss the coin six times. X is the number of heads in six tosses. Then $X \sim B(6,.3)$ We will find the distribution of X and generalize to any $B(n,p)$ random variable.

We will find $P(X=2)$ and generalize to $P(X=k)$ for any $k=0, 1, \ldots,6$

Let's consider one particular outcome in which X=2, namely HHTTTT.
P(H on 1^{st} and H on 2^{nd} and T on 3^{rd} ….and Ton 6^{th})=P(H on 1^{st})P(H on 2^{nd})P(Ton 3^{rd})…P(T on 6^{th}) because the trials are independent.
P(HHTTTT)=.3x.3x.7x.7x.7x.7x.7= $.3^2.7^4$
Consider any outcome with exactly two heads. There will be two factors of .3 and four factors of .7 so the probability is the same. It doesn't matter where the .3's occur in the string of factors.

How many outcomes have exactly two heads? Choose 2 of the 6 spaces for the H's. There are C(6,2) ways to do this. The event that X=2 has C(6,2) outcomes each with probability $.3^2.7^4$.
P(X=2)=C(6,2) $.3^2.7^4$ The same process shows
P(X=k)=C(6,k) $.3^k.7^{6-k}$
With n=6, p=.3, q=.7 we have

P(X=k)=C(n,k) $p^k q^{n-k}$ This formula works for any n and p. Recall q=1-p.

Ex. Type I diabetes occurs in about 6% of children. If 80 children are selected at random what is the probability exactly 3 of them will develop the disease?
X=the number who develop diabetes. n=80, p=.06, k=3
P(X=3)=C(80,3) $.06^3.94^{77}$

Binompdf on the calculator:
We can use a calculator function to compute this.
Locate distr=2^{nd} VARS on the calculator
Enter 0. This selects binompdf. Remember binompdf(n,p,k)=P(X=k)
Enter 80, .06, 3 and enter.
P(X=3)=.15134…

Ex. Toss a coin 75 times. The coin is weighted so P(H)=.6. X=the number of heads.
P(X=25)=C(75,25) $.6^{25}.4^{50}$ =binompdf(75,.6,25)=1.89x10^{-6}

Now we will learn to use binomcdf. The c stands for cumulative.

Ex. An urn contains 5 red and 7 non-red balls. A trial consists of drawing a ball from the urn, observing whether or not it is red and <u>then replacing the ball</u>. This trial is repeated 10 times. X=the number of red balls in 10 trials. Find the probability of getting at most 4 red balls, that is P(X$\leq$4).
X~B(10,5/12)
P(X$\leq$4)=P(X=4)+P(X=3)+P(X=2)+P(X=1)+P(X=0) We don't have to compute all these and add them up because binomcdf will do it for us.

Distr A selects binomcdf. Remember binomcdf(n,p,k)=P(X$\leq$k)

Binomcdf(10, 5/12, 4)=.591

To find $P(X \geq k)$ find $1-P(X \leq k-1)$=binomcdf(n,p,k-1)
For the example above, $P(X \geq 5)=1-P(X \leq 4)=1-.591=.409$

Mean and Standard deviation of a binomial random variable:
If X~B(n,p) then E(X)=np and Var(X)=npq. The standard deviation of X is $\sqrt{npq}$.
Ex. A test used to detect a certain disease is positive for the disease in 15% of the people who take the test. If 100 people are selected at random, what is the expected number who will test positive? What is the standard deviation of the number who test positive?

Think of success as testing positive for the disease. Let X be the number of the 100 who test positive. Then X~B(100,0.15) and E(X)=100x0.15=15.
$\sigma(X) = \sqrt{100\ x 0.15\ x 0.85}$ =3.57

Ex. For the example above, what is the probability that at least 15 people will test positive?
$P(X \geq 15)=1-P(X \leq 14)$=binomcdf(100, 0.15, 14)=.457

Ex. For the same example, what is the probability that between 11 and 19 people will test positive.
$P(11 \leq X \leq 19)=P(X \leq 19)-P(X \leq 10)$=binomcdf(100, 0.15, 19)-binomcdf(100, 0.15, 10)=.794.

The Normal Distribution:

A random variable with a normal distribution is one example of a continuous random variable. A continuous random variable takes on values in an interval, that is its domain is an interval. Some examples are the heights of people in a population or the time it takes to finish a task.

A continuous random variable has a probability density function, f(x), such that $f(x) \geq 0$ for all x and the total area under the graph of f is 1.

$P(a<X<b)$=area under the graph of f over the interval (a,b). P(X=a)=0 because the area under f over a single point is 0.

The probability density function for a normal random variable is called a normal curve and is bell shaped and symmetric about the mean. It approaches 0 as x approaches $+\infty$ and as x approaches $-\infty$.

If X has a normal distribution with mean μ and standard deviation σ we write X~N(μ, σ) . There is no setup to recognize. You have to study the data to see if a distribution is approximately normal. Your problems will state that the random variable is normally distributed or approximately normally distributed. Population heights and weights are often normally distributed. Exam scores are often approximately normal.

If X~N(μ, σ) then $\frac{X-\mu}{\sigma}$~N(0,1) and is called standard normal, denoted by Z.

For any random variable P($(\mu - 3\sigma < X < \mu + 3\sigma)$ =.9973 and
P($\mu - \sigma < X < \mu + \sigma$)=.6827
Symmetry Property: $P(Z<-a)=P(Z>a)$ $P(Z<0)=P(Z>0)=.5$

$P(X< \mu - a) = P(X> \mu + a)$ $P(X< \mu) = P(X > \mu) =$

On the calculator use 2^{nd} Distr normalcdf(a, b, μ, σ) for P(a<X<b) when X is normal. We do not use normalpdf. If the mean and standard deviation are not entered they are assumed to be 0 and 1. That is normalcdf(a,b)=P(a<Z<b).

Ex. Exam scores are found to be approximately normally distributed with a mean of 70 and a standard deviation of 12. Find the probability that a randomly selected score

a) is between 65 and 75. b) is more than 80. c) is less than 75.

a) P(65<X<75)=normalcdf(65, 75, 70, 12)=.323

b)P(X>80)=normalcdf(80, 1E9, 70,12)=.2023

c) P(X<75)=normalcdf(-1E9, 75,70,12)=.6615

invNorm(p, μ, σ) :

Find a so that P(X<a)=p when X~N(μ, σ). a=invNorm(p, μ, σ)

Ex. A professor finds his exam scores are normally distributed with mean 75 and standard deviation 8. He wants to give A's to the top 7% of the class. What should be the A cutoff score?

If X is a randomly selected exam score then X~N(75,8). Let a be the A cutoff score.
The professor wants $P(X \geq a)$ to be .07. This means P(X<a)=.93
a=invNorm(.93, 75, 8)=86.8
Suppose he wants to give B's to 15% of the class. What should be the B cutoff score?
Let b be the B cutoff score. If 15% get B's and 7% get A's then 22% are above the B cutoff score. So 78% are below the B cutoff score. Then P(X<b)=.78 so b=invNorm(.78, 75, 8)=81.18

Approximation of a binomial with a normal:

The normal distribution can be used to approximate a binomial distribution. This is especially useful when n is large and np and nq are both greater than 5.

If X~B(n,p) then $\mu = np, \sigma = \sqrt{npq}$. Suppose Y~N(np, $\sqrt{npq}$) . Then
$P(X \leq k)$ is approximately P(Y<k+.5) and $P(X \geq k)$ is approximately P(Y>k-.5).
$P(j \leq X \leq k)$ is approximately P(j-.5<Y<k+.5).
For these problems, you have to recognize a binomial setup and then find the mean and standard deviation and approximate with a normal.

Ex. At a bank it is estimated that 5% of 400 loan accounts will be delinquent. Estimate using a normal distribution the probability that 25 or more accounts will be delinquent.
X=the number of delinquent accounts out of 400
n=400 p=.05 $\mu = 400\ x .05 = 20$, $\sigma = \sqrt{400\ x .05\ x .95}$ =4.359
$P(X \geq 25)$ is approximately normalcdf(24.5, 1E9, 20, 4.359)=.150955.
Compare to 1-binomcdf(400, .05, 24)=.15102

Ex. Suppose X~B(100, .2) $P(10 \leq X \leq 30)$ is approximately
normalcdf(9.5, 30.5, 20, 4)=.991335