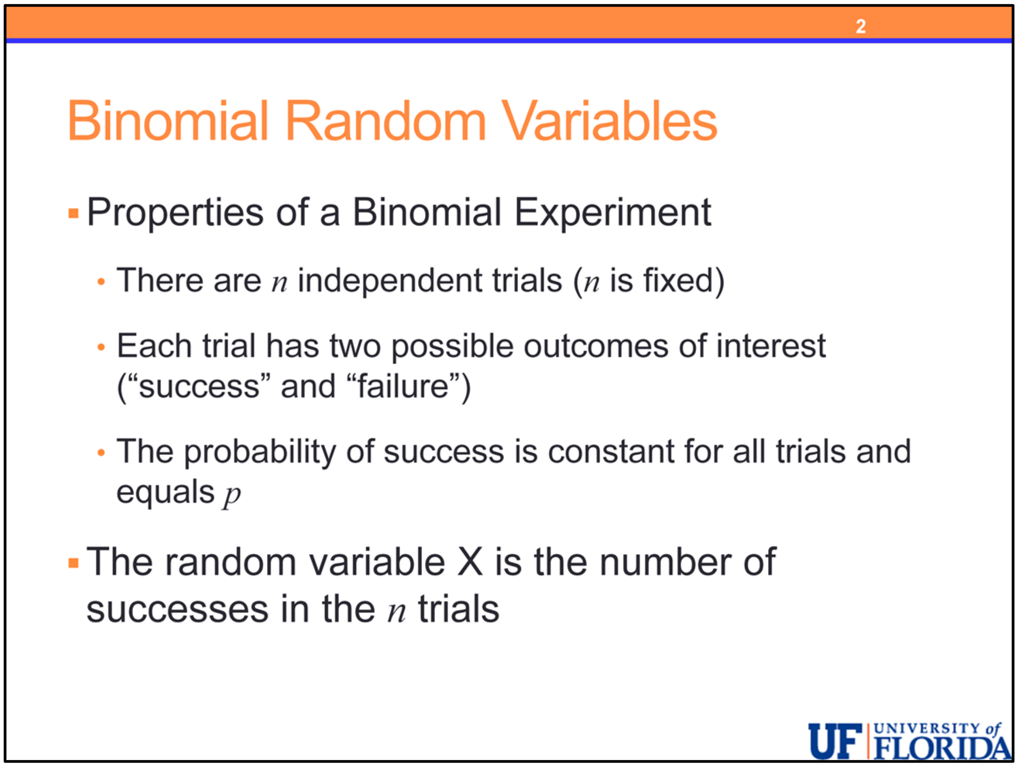
1

BINOMIAL RANDOM VARIABLES

Unit 3B: Random Variables

UF UNIVERSITY of FLORIDA

Now we discuss the details about binomial random variables which are a special type of discrete random variable.



There are four properties which much be verified in order to confirm that a binomial random variable is appropriate.

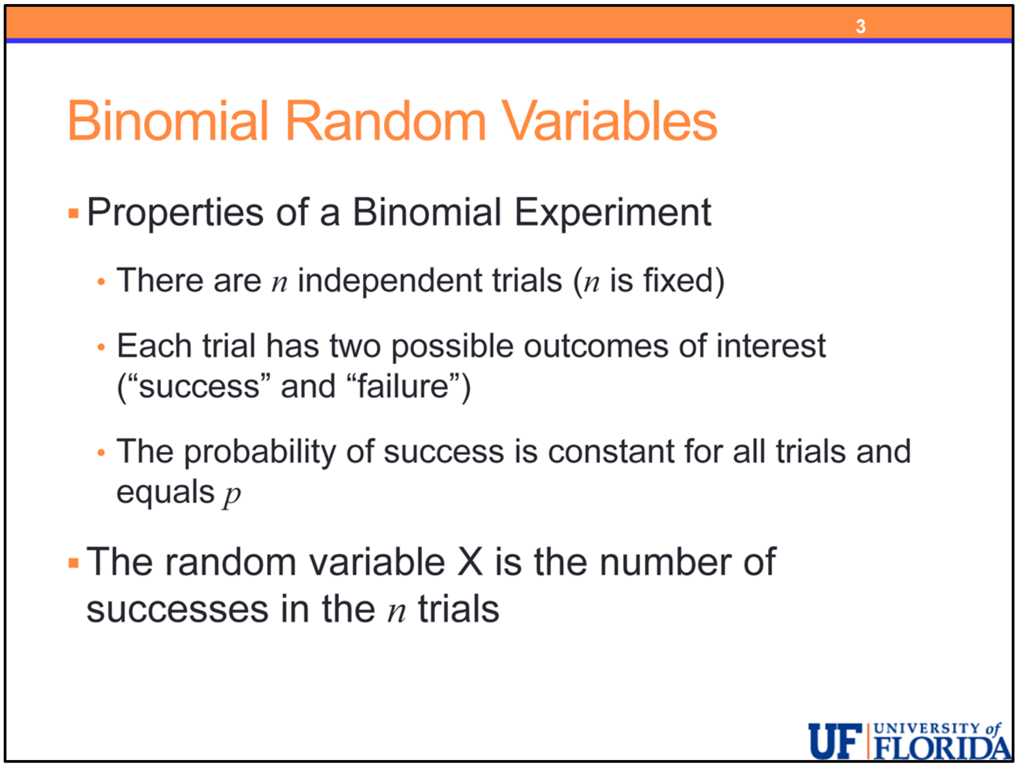
- We must have a fixed number of trials, represented by little n
- These trials must be independent
- Each trial has only two possible outcomes: success and failure.
- The probability of success must be constant for all trials and we call it little p

A few additional comments:

First, what we call a success or a failure can be anything as long as the two events are opposites of each other.

Many situations can be put in this framework where the success is everything of interest to us (in one group) and failure is everything else.

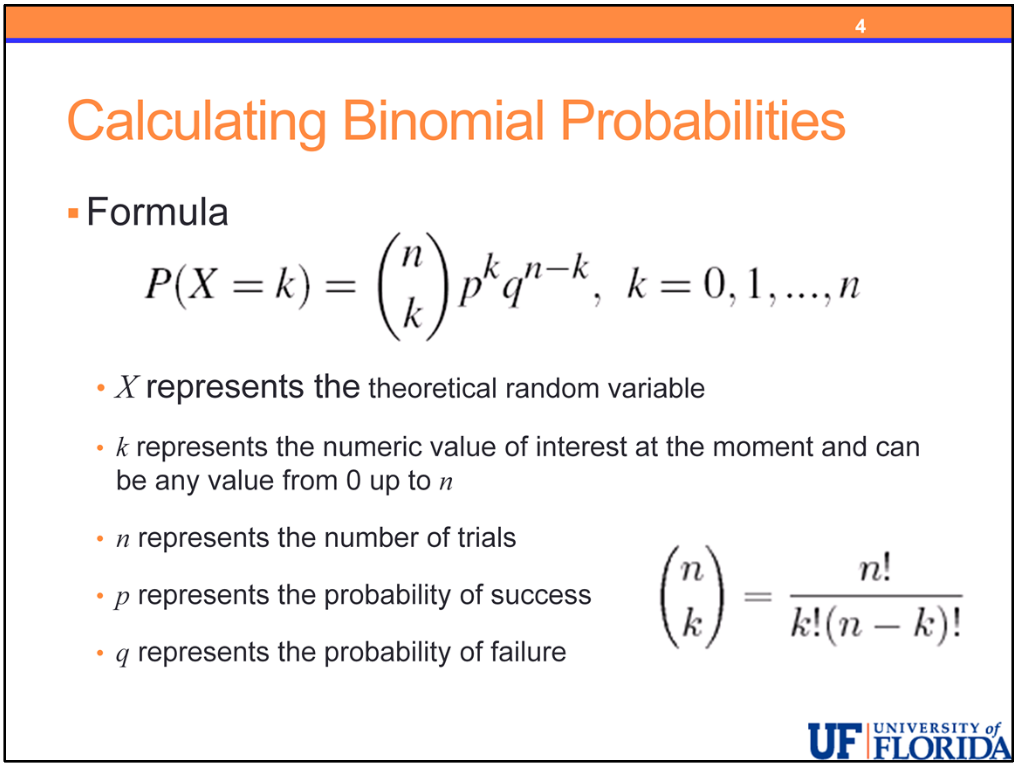
The words success and failure, are not indicative of the practical result in each case. A success might be a bad thing. It might be a diagnosis of breast cancer. That is the success in my study. It's not a success in the sense that is a good outcome but it is what I am interested in looking at so we call it here a success.



Also we said the probability must be constant for each trial. In Example A, we are talking about the prevalence being 8 percent. We need to be convinced that while we are studying this scenario, that prevalence isn't changing. In practice, we try to keep our time frame fairly short. If I study diabetes over very long term I expect that prevalence to be changing, that would not satisfy the properties of a binomial distribution.

The random variable X will count the number of successes.

If you can verify the properties and have a random variable which counts the number of successes then you have a binomial experiment and can use the formulas associated with this special case instead of those for general discrete random variables.



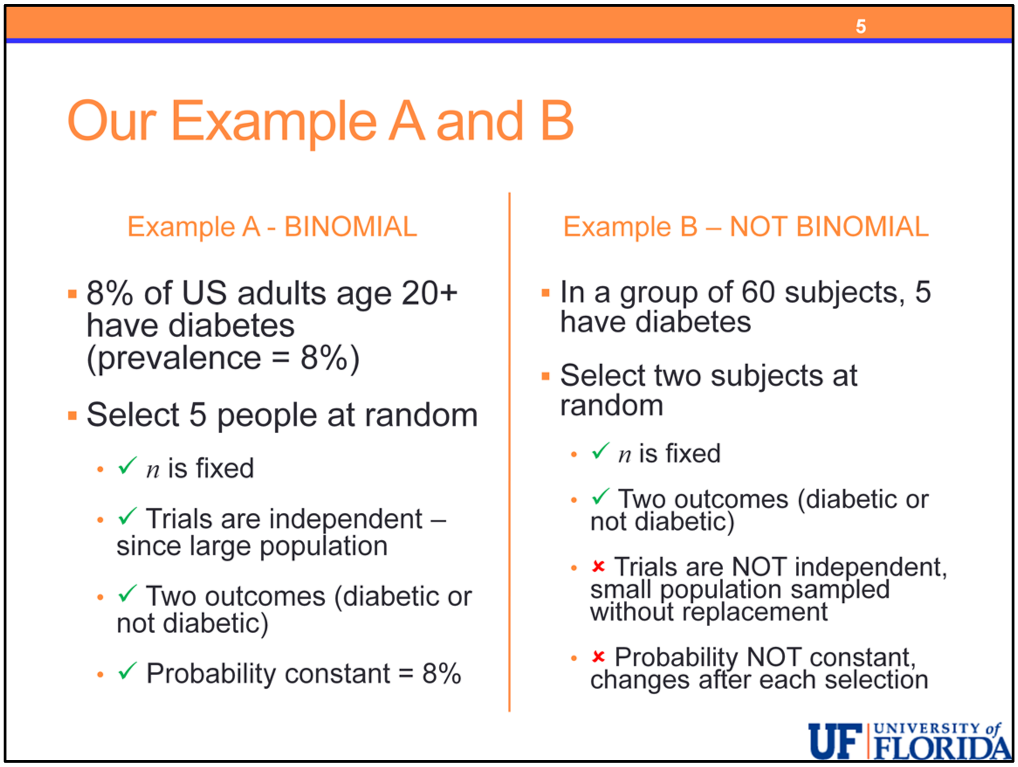
To calculate probabilities, we have the following formula.

We will see that this is simply going through the process we discussed earlier and putting it into a formula which can be applied for any n and p.

The first part of the equation with the n choose k is the biggest issue if you don't remember what this is. I have given the translation in the lower right here. On the left of this equation we have the notation for n choose k that we tend to give in statistics which is just two big parentheses.

Notice there's no fraction line here. It's not n divided by k. It is n choose k and it means n factorial in the numerator divided by k factorial times n minus k factorial in the denominator.

So we have the probability that the random variable X is equal to k is n choose k times p to the k times q to the (n-k) where k = 0, 1, 2, 3, up to n.
We need to substitute values for k, p, q, and n and perform the calculation to get P(X = k).



Let's look at our Examples A and B and determine if they represent binomial experiments. For Example A, we have 8% of all US adults age 20+ have diabetes. We will select 5 people from this population at random.

Thus we do have a fixed number of trials, 5. The trials are independent since our population is very large. We only have two outcomes – diabetic (success) or not diabetic (failure). And, as long as we study this situation over a reasonable time period, it is reasonable that the probability of success is constant at 8%.

Therefore, we can consider Example A to be a binomial experiment with n = 5 and p = 0.08.

In Example B, we have a group of 60 subjects, of which 5 have diabetes. We will select two subjects at random.

Here there is a fixed number of trials, 2 and we only have two outcomes – diabetic (success) or not diabetic (failure). However, the trials are not independent since we have a small population sampled without replacement and the probability is not constant, it will change for each trial as we have seen when calculating probabilities in this scenario.

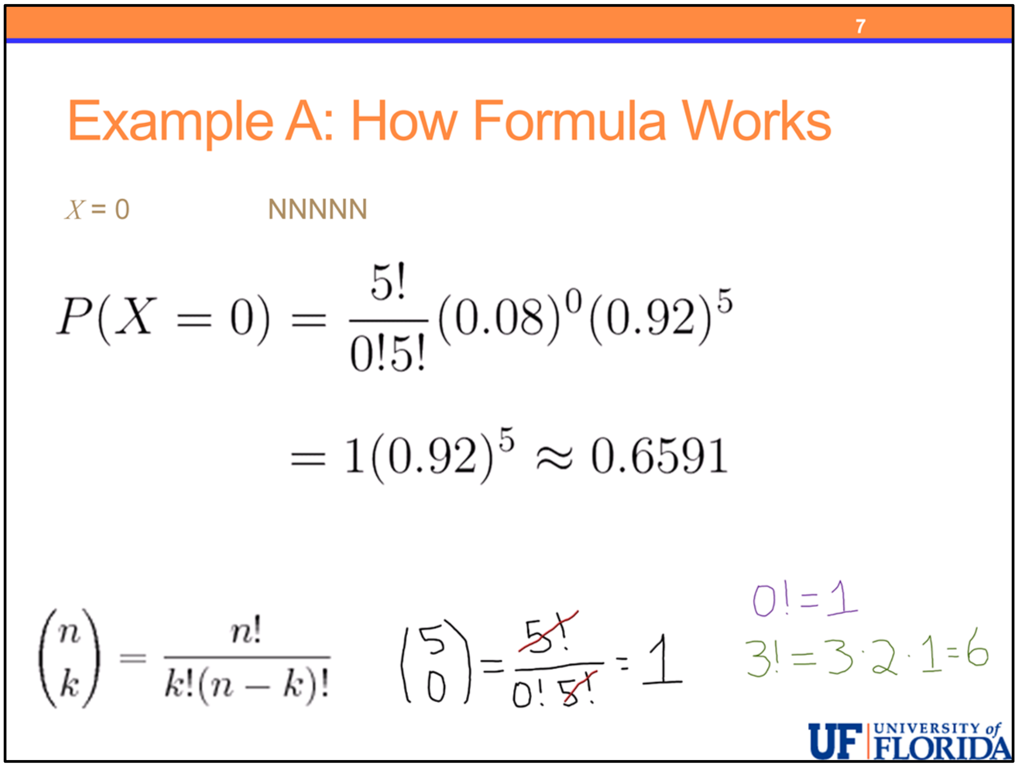
Thus in Example B, we do not have a binomial experiment.

Example A: How Formula Works

$X = 0$	NNNNN
$X = 1$	NNNND NNNDN NNDNN NDNNN DNNNN
$X = 2$	NNNDD NNDND NDNND DNNND NNDDN NDNDN DNNDN NDDNN DNDNN DDNNN
$X = 3$	DDDNN DDNDN DNDDN NDDDN DDNND DNDND NDDND DNNDD NDNDD NNDDD
$X = 4$	NDDDD DNDDD DDNDD DDDND DDDDN
$X = 5$	DDDDD

Here we have the list of all possible outcomes for Example A organized by the value of X = number of diabetics in our sample of 5 individuals. So we have X = 0, X = 1, through X = 5 and we have listed the appropriate outcomes for each value of X. We can see that there is one way to get X = 0, five ways to get X = 1, 10 ways to get X = 2, 10 ways to get X = 3, five ways to get X = 4 and one way to get X = 5.

In our formula, the n choose k is counting how many ways that particular value of k can happen and we will see that we will get 1, 5, 10, 10, 5, and 1 in this case. This list is just to convince you of where the values are coming from!



For P(X = 0), we substitute zero for k, 5 for n, 0.08 for p and 0.92 for q (for 1 minus p). To calculate the 5 choose zero we have 5 factorial divided by the product of zero factorial and 5 factorial. We need to be careful, zero factorial is defined to be equal to one. All other values are the product of the value and all integers smaller. So 5 factorial Is 5 times 4 times 3 times 2 times 1.

Here we don’t need to calculate 5 factorial since there is one in the numerator and one in the denominator which cancel to leave the result of 5 choose zero equals one. In fact n choose zero is always equal to one. There is only one way to get no successes in n trials, regardless of the value of n.

Then we combine this with the rest of our formula – we have 0.08 to the zero power – this is equal to 1 and 0.92 to the 5th power. Notice this is what we found using the multiplication rule and is about 0.6591.

Example A: How Formula Works

X = 1 NNNND NNNDN NNDNN NDNNN DNNNN

$$P(X = 1) = \frac{5!}{1!4!}(0.08)^1(0.92)^4$$

$$= 5(0.08)^1(0.92)^4 \approx 0.2866$$

$$\binom{5}{1} = \frac{5!}{1!4!} = \frac{5\cdot 4\cdot 3\cdot 2\cdot 1}{1\cdot 4\cdot 3\cdot 2\cdot 1} = 5$$

UF UNIVERSITY of FLORIDA

For P(X = 1), we substitute 1 for k, 5 for n, 0.08 for p and 0.92 for q (for 1 minus p). To calculate 5 choose 1, I get five factorial in the numerator divided by k, which is one, factorial times n minus k, which is 4, factorial and when I do these by hand this is how I would do it.

I would write out the numerator as five times four times three times two times one. Then write out my denominator as 1 times 4 times 3 times 2 times 1, cancel what I can, ... and we get five fairly easily. So it is not hard to do by hand. You can also use your calculator if it has the ability to do combinations which might use the notation nCr.

Notice we did see that there were 5 ways to have one diabetic in our 5 individuals. After multiplying the entire probability calculation we get approximately 0.2886.

Example A: How Formula Works

$X = 2$ NNNDD NNDND NDNND DNNND NNDDN

NDNDN DNNDN NDDNN DNDNN DDNNN

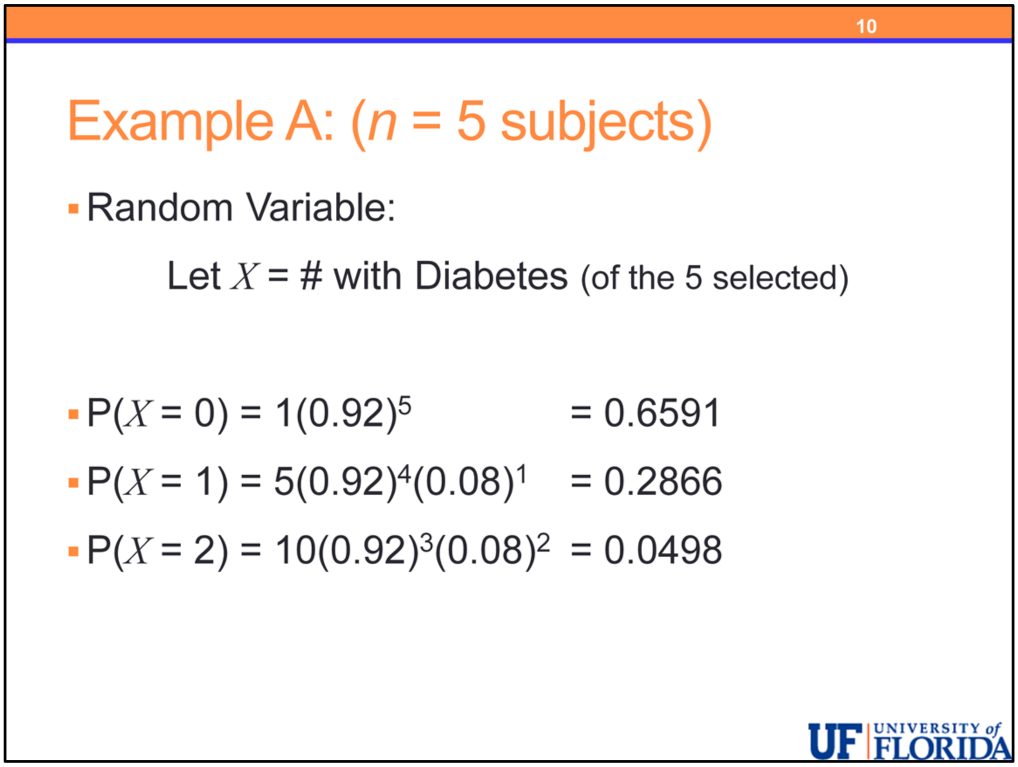
$$P(X=2) = \frac{5!}{2!3!}(0.08)^2(0.92)^3$$

$$= 10(0.08)^2(0.92)^3 \approx 0.0498$$

$$\binom{5}{2} = \frac{5!}{2!\,3!} = \frac{5\cdot 4\cdot 3\cdot 2\cdot 1}{2\cdot 1\cdot 3\cdot 2\cdot 1} = 10$$

UF UNIVERSITY of FLORIDA

For X = 2, we get 5 choose 2 equals 10 and multiplying the probability calculation gives around 0.0498. You can go through the values for X = 3, 4, and 5 and check them with the results on the next slide.



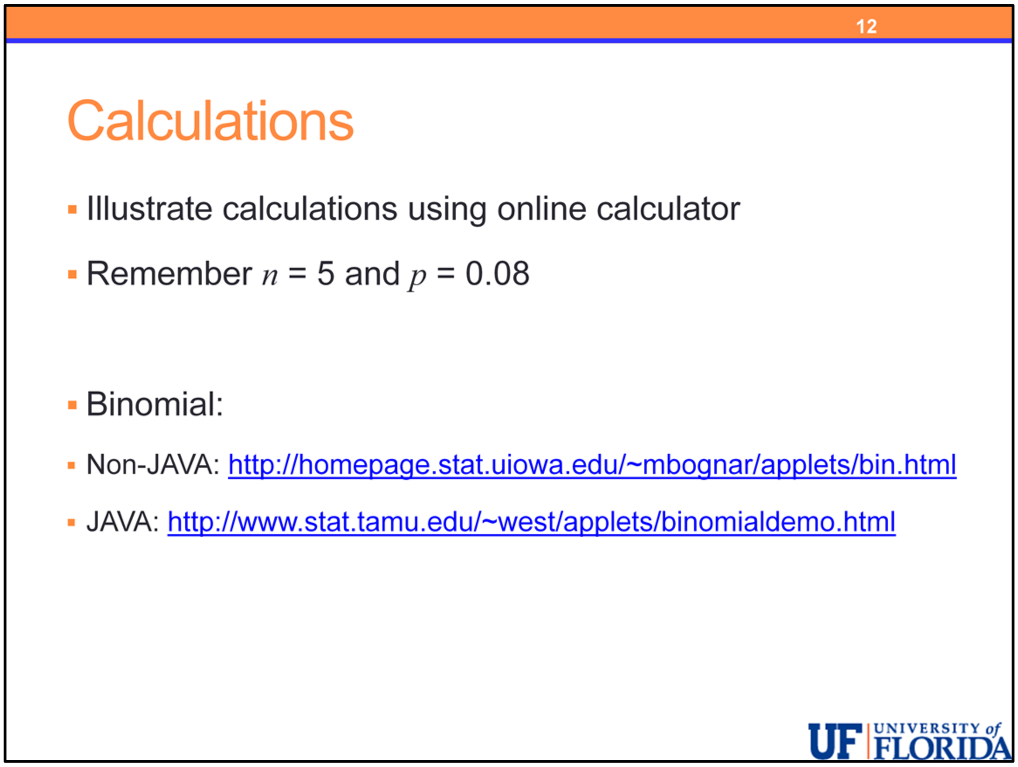
Here is a summary of the calculations and results.

Complete Distribution

X	P(X = x)
0	0.659082
1	0.286557
2	0.049836
3	0.004334
4	0.000188
5	0.000003

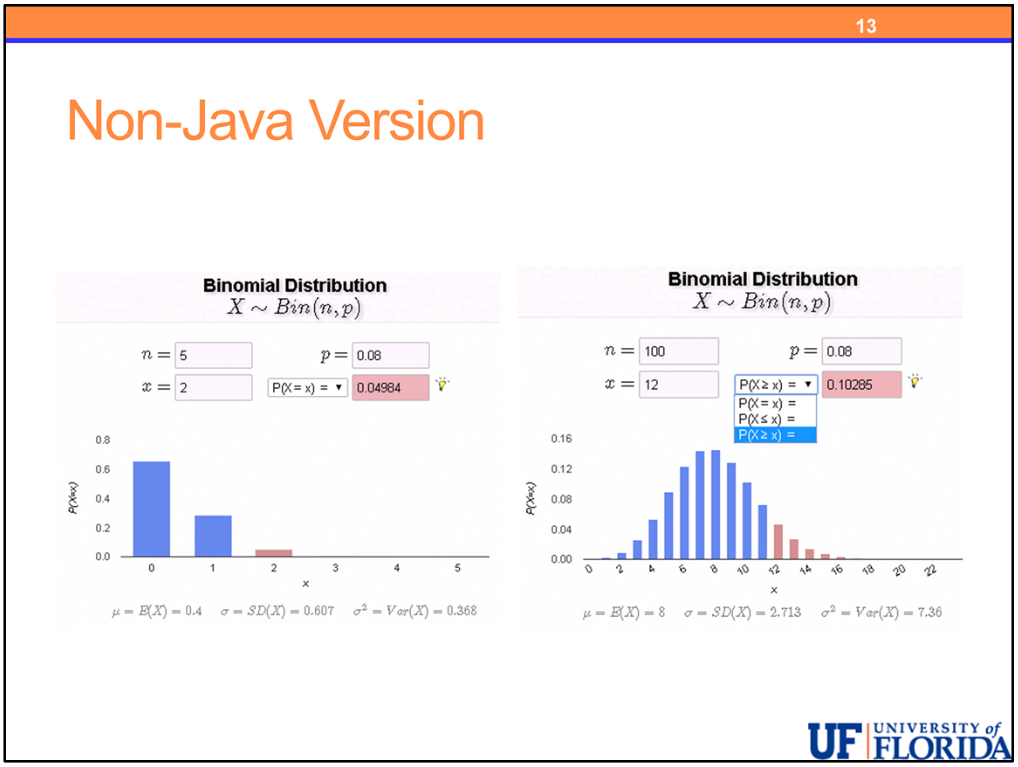
UF UNIVERSITY of FLORIDA

And here we have the entire distribution.



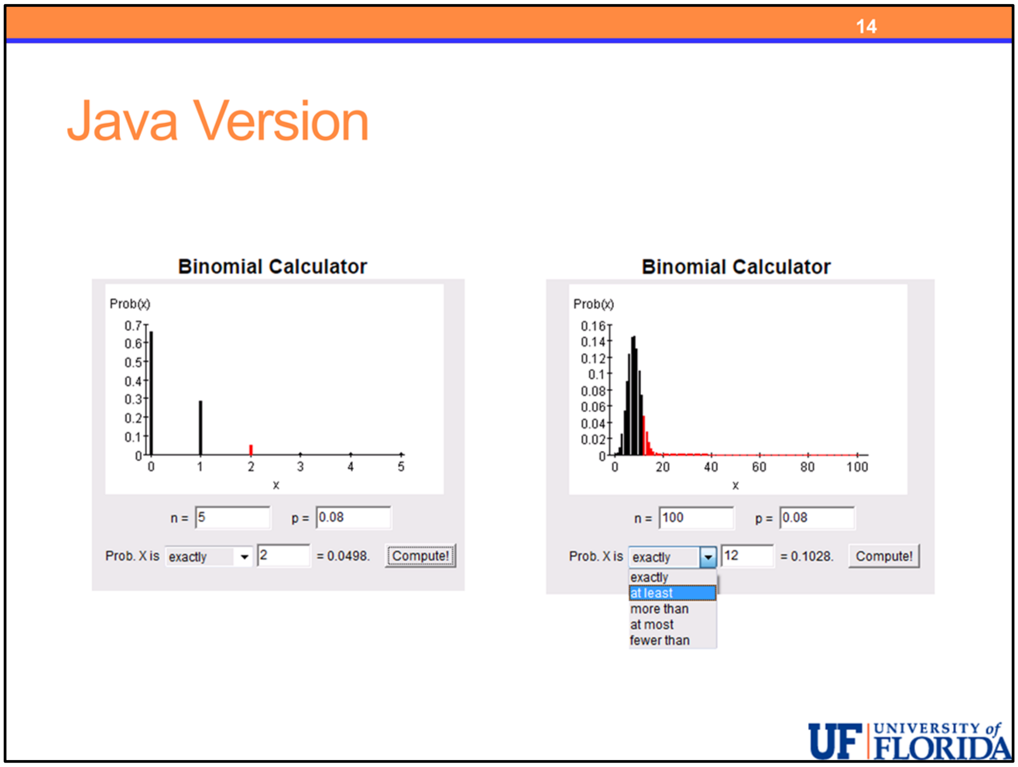
Now we want to illustrate the online calculators. There are two that we have linked from the materials but there are likely many others available and you are welcome to use any that provides correct answers.

There is a JAVA version and a non-JAVA version.



Let's start with the non-Java Version. Here we have a few images of this applet in use. On the left, we entered n = 5, p = 0.08, and x = 2 (this would k in the earlier formula but x is also a very common notation which we have also used in the written materials). We left the probability at the default of P(X = x) to get 0.04984 which is the same as we obtained previously.

The nice thing about these calculators is we are no longer restricted to small values of n or simple probabilities. On the right, we have used n = 100, p = 0.08, x = 12 and have chosen to calculate P(X ≥ x). The calculator adds all probabilities for 12, 13, … all the way to 100 to get 0.10285.



Here we have illustrated the same calculations using the JAVA version. The only major differences between these two calculators is that this one requires JAVA which can be painful these days and this version uses verbal descriptions of the probability where the other one uses probability notation.

Regardless of which calculator you use (if any) be careful to correctly interpret any language in the question. Use the table on the page on discrete random variables as needed.

15

Binomial Random Variables

Expected Value = $E(X) = \mu = np$

Variance = $Var(X) = \sigma^2 = np(1-p) = npq$

Standard Deviation: $\sigma_X = \sqrt{np(1-p)}$

Finally, we need to discuss the mean and standard deviation for a binomial random variable.

The mean (expected value) of a binomial is very easy to calculate and we already did this using logic earlier. If we select 100 individuals from all US adults age 20+ where the prevalence of diabetes is 8% in the population, how many would we expect to be diabetic. We said 8% of 100 or 100 times 0.08 which is n times p, the equation we have here. This does NOT work for any discrete random variable, only for binomial random variables derived from a binomial experiment.

Then we have the formula for the variance which is n times p times q but of course we are most interested in the standard deviation which is the square root of n times p times q. Where q is equal to (1 – p).

In Example A

$$\mu = E(X) = np = 5(0.08) = 0.4$$

$$\sigma = \sqrt{npq} = \sqrt{5(0.08)(0.92)}$$

$$= \sqrt{0.368} = 0.60663$$

UF UNIVERSITY of FLORIDA

In Example A where n = 5 and p = 0.08, the mean is 5 times 0.08 which is 0.4.

And the standard deviation is the square root of 5 times 0.08 times 0.92 which is 0.60663.

We illustrated both of these using the formulas for general discrete random variables earlier and here we are obtaining basically the same values using the special case formulas for binomial random variables. The difference is that previously we did have some rounding error so these values are more accurate.

BINOMIAL RANDOM VARIABLES

Unit 3B: Random Variables

Binomial distributions have many applications!!

UF UNIVERSITY of FLORIDA

That is everything for binomial random variables. Binomial random variables have many applications in real-world situations.

We need to be able to determine if the binomial distribution is appropriate by verifying the properties. For binomial random variables, we can then find probabilities using the equation for P(X = k) by hand or using an online calculator (or other calculator) and finally we can calculate the mean and standard deviation for a binomial random variable using the special case formulas presented here.