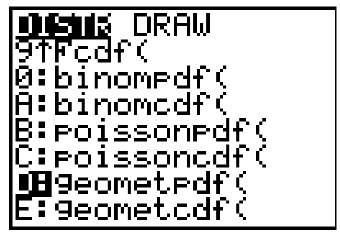
Chapter 8 – Probability Models

We've already used the calculator to find probabilities based on normal models. There are many more models which are useful. This chapter explores three such models.

Many types of random variables are based on Bernoulli trials experiments. These involve independent trials with only two outcomes possible, and a constant probability of success called *p*. Two of the more common of these variables have either a Geometric or Binomial model. The Poisson random variable is related as well – it can be viewed as a Binomial model with small probability of success and large n, or as a process where only "successes" are seen: phone call arrivals at a switchboard, for example.

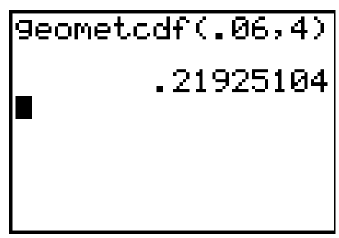
GEOMETRIC MODELS

The Geometric probability model is used to find the chance the first success occurs on the n^{th} trial. If the first success is on the n^{th} trial, it was preceded by n - 1 failures. Because trials are independent we can multiply the probabilities of failure and success on each trial so $P(X = n) = (1-p)^{n-1}p$ which is sometimes written as $P(X = n) = (q)^{n-1}p$. This is generally easy enough to find explicitly, but the calculator has a built-in function to find this quantity, as well as the probability of the first success occurring somewhere on or before the n^{th} trial.

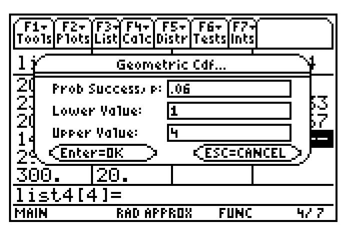


Suppose we are interested in finding blood donors with O-negative blood; these are called "universal donors." Only about 6% of people have O-negative blood. In testing a group of people, what is the probability the first O-negative person is found on the 4^{th} person tested? We want P(*X* = 4). Press 2nd VARS (`Distr`). If you are using a TI-89, the `Distr` menu is F5 in the Stats/List Editor application. We want menu choice `D:geometpdf (F` on an 89). To find this menu option (or any of those discussed in this chapter), pressing the up arrow to find them is probably the easiest. Press ENTER to select the option.

Enter the two parameters for the command: *p* and *x* (*n*). Here, *p* = .06 and *n* = 4. Press ENTER to find the result. We see there is about a 5% chance to find the first O-negative person on the 4^{th} person tested (assuming of course that the individuals being tested are independent of each other).

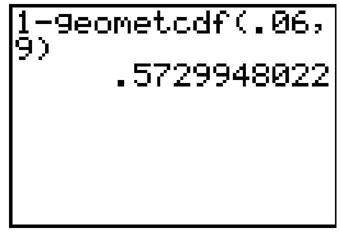


There are some other related questions that can be asked. What is the probability the first O-negative person will be found in the first four persons tested? We want to know $P(X \le 4)$. We could find all the individual probabilities for 1, 2, 3, and 4, and add them together but there is an easier way. We really want to "accumulate" all those probabilities into one, or find the *cumulative* probability.

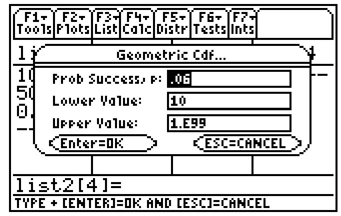


This uses menu choice `E:Geometcdf (G` on an 89). The cdf part stands for cumulative distribution function. On a TI-83/84 we input the probability of a success (still .06); and the high end of interest (4). Using a TI-89, you will be prompted for the low end of interest (`Lower Value`) and the upper end of interest (`Upper Value`).

After pressing ENTER we see that finding the first O-negative person within the first 4 people tested should happen about 21.9% of the time.



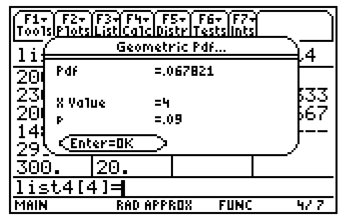
What's the chance we'll have to test at least ten before we find one with type O-negative blood? We want P($X \geq 10$). Since there could (possibly) be an infinite number of people tested to find the first O-negative person. Using a TI-83/84 we will use complements to answer this question. The complement (opposite) of needing at least ten people to find the first universal donor is finding the first one somewhere in the first nine tested. Using the complements rule, $P(X \geq 10) = 1 - P(X \leq 9)$. It makes no difference whether we find $P(X \leq 9)$ and then subtract from 1 or do both operations in a single step.



Using a TI-89, since the lower end of interest and the upper end of interest are explicitly asked for, here we will say the lower end is 10, and the upper end is [1][EE][9][9] for infinity.

The probability we will find the first O-negative person somewhere on the tenth person or later is 57.3%.

Spam and Geometric Models

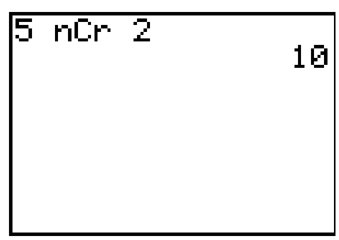


According to *Postini*, about 91% of all email traffic is spam. That means only about 9% is legitimate email. Since spam comes from so many sources, assuming messages are independent is reasonable. What's the chance you have to check four messages in your inbox to find the first legitimate message? According to the calculation at right, abut 6.8%.

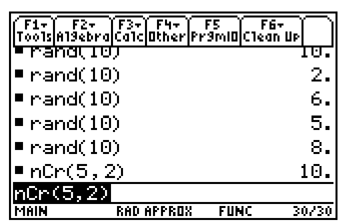
BINOMIAL MODELS

Binomial models are interested in the chance of k successes occurring when there are a fixed number (n) of Bernoulli trials.

Suppose you plan on buying five boxes of cereal that each have a 20% chance of having a picture of Tiger Woods in them. Assuming the boxes are chosen at random, they should be independent of one another. We are interested in the probability that there are exactly two pictures of Tiger in the five boxes. How many arrangements are possible so that there are two Tiger pictures? You could get the two on the first two boxes and not on the last three, or on the second and fifth, or on the third and fourth, etc. How many arrangements are possible to get exactly two pictures of Tiger and three of someone else? The Binomial coefficient provides the answer to this question. The coefficient itself is variously written as $\binom{n}{r}$ or nCr and is read as "n choose r."



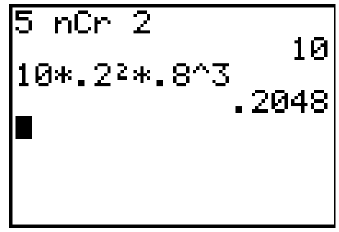
On the TI-83/84 home screen, type in the desired n (5, here), then press [MATH] then arrow to PRB. Either arrow down to 3:nCr(and press [ENTER] or just press [3]. The command is transferred to the home screen. Finally, enter k (3 here). Execute the command by pressing [ENTER].



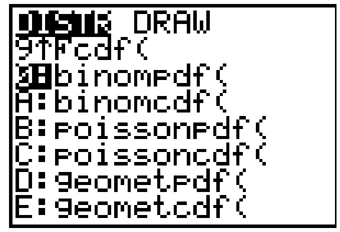
If you are using a TI-89, there are two ways to find the number of combinations. On the home screen, press [2nd][5] for the Math menu, then press [4] to expand the probability menu. Then choose nCr, which is menu option 3. Press [3] to transfer the shell to the input area. Type in n and r separated by a comma and close the parentheses. Press [ENTER] to complete the computation. You can also perform the calculation in the in the same manner while in the Statistics/List

Editor using the `Probability` menu under [F4] (`Calc`).

We find there are 10 possible ways to get exactly two pictures of Tiger in five boxes of cereal. We can then find the probability of exactly two pictures of Tiger as $10(.2)^2(.8)^3 = 0.2048$. This means about 20% of the time, if you bought five boxes of cereal (independently of each other) you would get two pictures of Tiger.

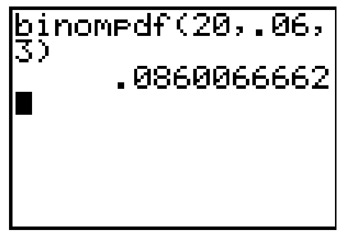


There is an easier way to find these binomial probabilities. Looking back at the portion of the `Distr` menu shown on the first page of this chapter, there are two menu options that will help us here: `0:Binompdf` and `A:Binomcdf`. On a TI-89, these are options `B` and `C` in the `Distr` menu. Just as in the case of the geometric model discussed above, the pdf menu choice gives P($X = x$) and the cdf gives $P(X \leq x)$. The parameters for both commands are n, p, x. Three examples follow.

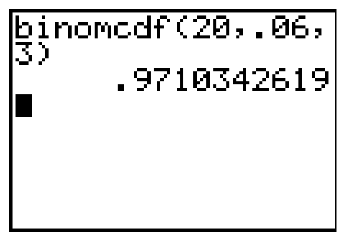


These commands are similar on a TI-89, with the exception that for the cdf one explicitly specifies the lower value of interest and the upper value of interest, just as we did with the geometric model.

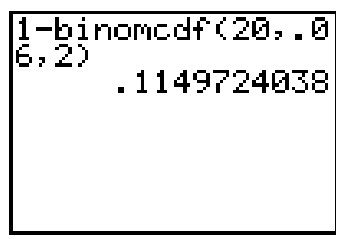
Returning to the prior example about blood donors, what is the probability that if 20 donors come to the blood drive, there will be exactly three O-negative donors? From the screen at right, we see this is 8.6%.



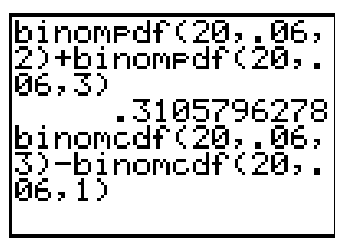
As with geometric models, we can find the use the cdf command to find cumulative probabilities. Remember, the calculator adds individual terms for the values of interest specified. For instance, what is the chance of at most three O-negatives in a blood drive with 20 donors? We see this is very likely to happen. We'll expect three 3 or fewer O-negative donors in 20 people about 97% of the time.



What's the chance there would be more than two O-negative donors in a group of 20? Again, since we are looking for P($X > 2$) we use complements. The complement of more than 2 is 2 or less. So we find $P(X > 2)$ as $1 - P(X \leq 2)$.



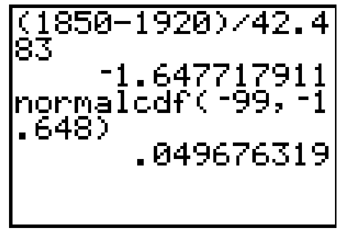
An example in the text asks for the probability of two or three universal donors in a group of 20. Using a TI-89, this could be done specifying a lower value of 2 and an upper value of 3. With the TI83/84 series, there are two possible ways to accomplish the calculation as shown at right. The first explicitly adds the two individual probabilities. The second uses the binomial cdf function and finds the probability of three or fewer which includes the possibility of 0, 1, 2, or 3. We then subtract the probability of 0 or 1 from the result because we are interested in only 2 or 3. This method of calculation is very efficient in case you wanted to find, for example, the probability of between 12 and 19 universal donors – adding individual terms would become tedious. Find the cumulative probability for the upper end of interest, then subtract what was included that is not of interest.



Binomial Distributions with large n.

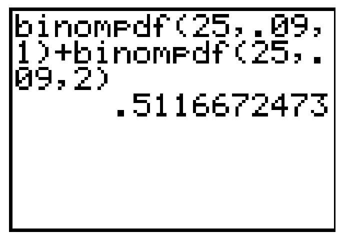
Older TI-83 calculators (and most computer applications as well) cannot deal with large values of n. This is because the binomial coefficient becomes too large very quickly. However, when n and p are sufficiently large (generally, *both* of $np \geq 10$ and $n(1-p) \geq 10$ must be true to move the distribution away from the ends so it can become symmetric) binomials can be approximated with a normal model. One uses the `normalcdf` command described in Chapter 4 specifying the mean as the mean of the binomial ($\mu = np$) and the standard deviation as that of the binomial ($\sigma = \sqrt{np(1-p)}$). Newer models of the TI-83 (and 84) actually use an approximation to find the probabilities in these situations.

Suppose the Red Cross anticipates the need for at least 1850 units of O-negative blood this year. They anticipate having about 32,000 donors. What is the chance they will not get enough? We desire $P(X < 1850)$. We have calculated the mean to be 1920 and the standard deviation to be 42.483. On this scale, 1850 has a z-score of -1.648. It appears there is about a 5% chance there will not be enough O-negative in the scenario discussed.



Spam and Binomial Models

Suppose there were 25 new email messages in your inbox. As we've seen before, there's only about a 9% chance that any one of them is real (assuming you're not using a good filter). What's the chance you'll find only one or two good messages? This calculation is similar to the one done above where we were interested in 2 or 3 universal blood donors. We need the sum of two binomial results. From the screen at right, we see the probability of only 1 or 2 real messages in 25 is about 51.2%.

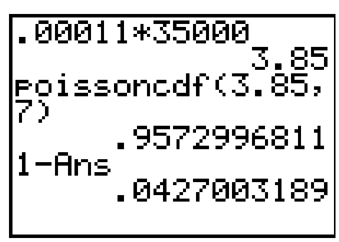


POISSON MODELS

Poisson models can be viewed in terms of several scenarios: an approximation to the binomial when the number of trials is large and the probability of a success is small (like the Woburn leukemia example in the text) or as realizations of a process in which we can really only see the successes. Classic examples are phone calls to a switchboard (we know about those who actually got through but not people who got busy signals, thought about calling but didn't, etc) and catching fish in a lake (we don't know about the fish who maybe ignored the bait). Lucky for us that the name of the individual who developed the model was Simeon Poisson (his name is French for fish).

Poisson models are generally characterized by a rate called λ (lambda). In the binomial setting, this is the mean of the binomial variable, $\lambda = \mu = np$. In the other setting, we think in terms of the rate of occurrence: we normally get 10 phone calls per hour, or I can typically catch 5 fish per day.

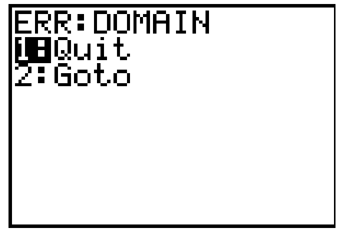
Revisiting the Woburn example, the national incidence rate for new leukemia cases each year is about 0.00011. That's p. The town of Woburn had about 35,000 people. Multiplying the two tells us that we'd expect about 3.85 new leukemia cases per year in a town that size. We want to find the probability of at least 8 cases in a single year. Since 8 or more is the complement of 7 or less, we use the calculator and first find the probability of 7 or less cases (using `poissoncdf` from the `DISTR` menu) , then subtract that result from 1. The chance of at least 8 new cases is about 4.3%. That result is pretty small (less than 1 in 20 chance), but not totally unreasonable.



WHAT CAN GO WRONG?

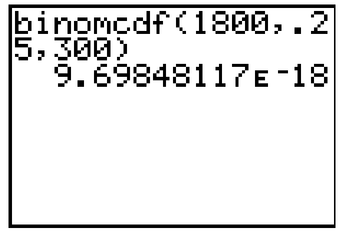
What does Err: Domain mean?

This error is normally caused in these types of problems by specifying a probability as a number greater than 1 (in percent possibly instead of a decimal) or a value for n or x which is not an integer. Reenter the command giving p in decimal form. Pressing [ESC] will return you to the input screen to correct the error. This will also occur in older TI-83 calculators if n is too large in a binomial calculation; in that case, you need to use the normal approximation.



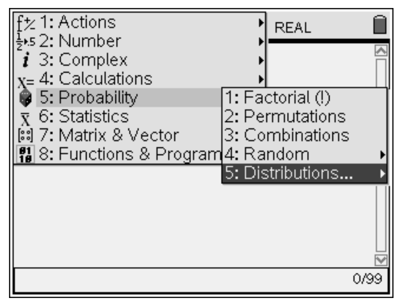
How can the probability be more than 1?

It can't. As we've said before, if it looks more than 1 on the first glance, check the right hand side. This value is 9.7×10^{-18} or seventeen zeros followed by the leading 9.

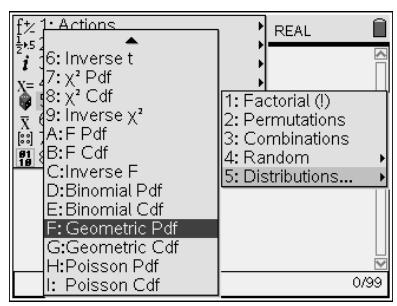


Commands for the TI-Nspire™ Handheld Calculator

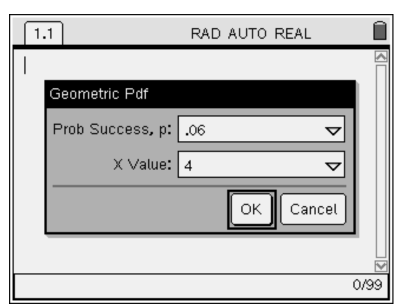
For probability distribution, start on a Calculator page, press (menu) , and then select Probability and Distributions.

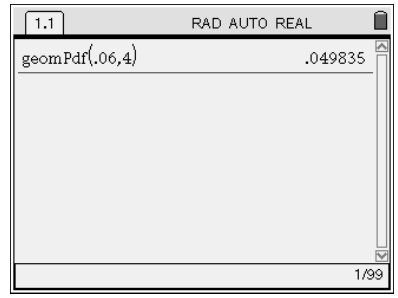


Press the number or letter for the desired distribution. Scroll if necessary to see additional choices. We selected a geometric distribution.

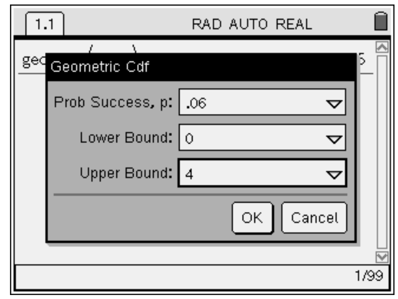


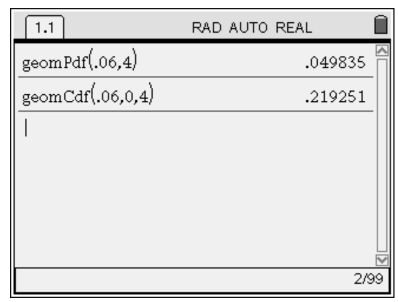
In the input box, type the probability of success and the value for the variable. For the blood donor example, use .06 and 4. (tab) to OK and (enter) .





For cumulative distributions, enter both a lower and upper bound.





Other distributions, such as Binomial and Poisson, are found on the same Distribution menu.