

# DISCRETE RANDOM VARIABLES

Documents prepared for use in course B01.1305,  
New York University, Stern School of Business

Definitions	page 3
Discrete random variables are introduced here. The related concepts of mean, expected value, variance, and standard deviation are also discussed.	
Binomial random variable examples	page 5
Here are a number of interesting problems related to the binomial distribution.	
Hypergeometric random variable	page 9
Poisson random variable	page 15
Covariance for discrete random variables	page 19
This concept is used for general random variables, but here the arithmetic for the discrete case is illustrated.	

edit date 8 JAN 2007

© Gary Simon, 2007

Cover photo: Trout hatchery, Connetquot  
Park, Long Island, March 2006

~~~~~ DISCRETE RANDOM VARIABLES ~~~~~

## Discrete Random Variables and Related Properties



Discrete random variables are obtained by counting and have values for which there are no in-between values. These values are typically the integers 0, 1, 2, ....

Random variables are usually denoted by upper case (capital) letters. The possible values are denoted by the corresponding lower case letters, so that we talk about events of the form  $[X = x]$ . The random variables are described by their probabilities. For example, consider random variable  $X$  with probabilities

|            |      |      |      |      |      |      |
|------------|------|------|------|------|------|------|
| $x$        | 0    | 1    | 2    | 3    | 4    | 5    |
| $P[X = x]$ | 0.05 | 0.10 | 0.20 | 0.40 | 0.15 | 0.10 |

You can observe that the probabilities sum to 1.

The notation  $P(x)$  is often used for  $P[X = x]$ . The notation  $f(x)$  is also used. In this example,  $P(4) = 0.15$ . The symbol  $P$  (or  $f$ ) denotes the probability function, also called the probability mass function.

The *cumulative* probabilities are given as  $F(x) = \sum_{i \leq x} P(i)$ . The interpretation is that  $F(x)$  is the probability that  $X$  will take a value less than or equal to  $x$ . The function  $F$  is called the cumulative distribution function (CDF). This is the only notation that is commonly used. For our example,

$$\begin{aligned}
 F(3) &= P[X \leq 3] &= P[X=0] + P[X=1] + P[X=2] + P[X=3] \\
 &= 0.05 + 0.10 + 0.20 + 0.40 &= 0.75
 \end{aligned}$$

One can of course list all the values of the CDF easily by taking cumulative sums:

|            |      |      |      |      |      |      |
|------------|------|------|------|------|------|------|
| $x$        | 0    | 1    | 2    | 3    | 4    | 5    |
| $P[X = x]$ | 0.05 | 0.10 | 0.20 | 0.40 | 0.15 | 0.10 |
| $F(x)$     | 0.05 | 0.15 | 0.35 | 0.75 | 0.90 | 1.00 |

The values of  $F$  increase.

The *expected value* of  $X$  is denote either as  $E(X)$  or as  $\mu$ . It's defined as

$E(X) = \sum_x x P(x) = \sum_x x P[X = x]$ . The calculation for this example is

$$\begin{aligned}
 E(X) &= 0 \times 0.05 + 1 \times 0.10 + 2 \times 0.20 + 3 \times 0.40 + 4 \times 0.15 + 5 \times 0.10 \\
 &= 0.00 + 0.10 + 0.40 + 1.20 + 0.60 + 0.50 &= 2.80
 \end{aligned}$$

This is also said to be the mean of the probability distribution of  $X$ .



The probability distribution of  $X$  also has a standard deviation, but one usually first defines the variance. The variance of  $X$ , denoted as  $\text{Var}(X)$  or  $\sigma^2$ , or perhaps  $\sigma_X^2$ , is

$$\text{Var}(X) = \sum_x (x - \mu)^2 P(x) = \sum_x (x - \mu)^2 P(X = x)$$

This is the expected square of the difference between  $X$  and its expected value,  $\mu$ . We can calculate this for our example:

|                        |       |       |       |       |       |       |
|------------------------|-------|-------|-------|-------|-------|-------|
| $x$                    | 0     | 1     | 2     | 3     | 4     | 5     |
| $x - 2.8$              | -2.8  | -1.8  | -0.8  | 0.2   | 1.2   | 2.2   |
| $(x - 2.8)^2$          | 7.84  | 3.24  | 0.64  | 0.04  | 1.44  | 4.84  |
| $P[X = x]$             | 0.05  | 0.10  | 0.20  | 0.40  | 0.15  | 0.10  |
| $(x - 2.8)^2 P[X = x]$ | 0.392 | 0.324 | 0.128 | 0.016 | 0.216 | 0.484 |

The variance is the sum of the final row. This value is 1.560.

This is *not* the way that one calculates the variance, but it does illustrate the meaning of the formula. There's a simplified method, based on the result

$$\sum_x (x - \mu)^2 P[X = x] = \left\{ \sum_x x^2 P[X = x] \right\} - \mu^2. \text{ This is easier because we've already}$$

found  $\mu$ , and the sum  $\sum_x x^2 P[X = x]$  is fairly easy to calculate. For our example, this sum is  $0^2 \times 0.05 + 1^2 \times 0.10 + 2^2 \times 0.20 + 3^2 \times 0.40 + 4^2 \times 0.15 + 5^2 \times 0.10 = 9.40$ . Then

$$\sum_x (x - \mu)^2 P[X = x] = 9.40 - 2.8^2 = 9.40 - 7.84 = 1.56. \text{ This is the same number as before, although obtained with rather less effort.}$$

The standard deviation of  $X$  is determined from the variance. Specifically,  $\text{SD}(X) = \sigma = \sqrt{\text{Var}(X)}$ . In this situation, we find simply  $\sigma = \sqrt{1.56} \approx 1.2490$ .

It should be noted that random variables also obey, at least approximately, a variant on the empirical rule used with data. Specifically, for a random variable  $X$  with mean  $\mu$  and standard deviation  $\sigma$ , we have

$$P[\mu - \sigma \leq X \leq \mu + \sigma] \approx \frac{2}{3}$$

$$P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] \approx 95\%$$

1. Suppose that you are rolling a die eight times. Find the probability that the face with two spots comes up exactly twice.

SOLUTION: Let  $X$  be the number of “successes,” meaning the number of times that the face with two spots comes up. This is a binomial situation with  $n = 8$  and  $p = \frac{1}{6}$ . The probability of exactly two successes is  $P[ X = 2 ] = \binom{8}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^6 = 28 \times \frac{5^6}{6^8}$ . This can be done with a calculator. There are various strategies to organize the arithmetic, but the answer certainly comes out as about 0.260476.

Some calculators have keys like  $\boxed{x^y}$ , and these can be useful to calculate expressions of the form  $6^8$ . Of course,  $6^8$  can always be calculated by careful repeated multiplication. The calculator in Microsoft Windows will find  $6^8$  through the keystrokes 6, y, 8, =.

2. The probability of winning at a certain game is 0.10. If you play the game 10 times, what is the probability that you win at most once?

SOLUTION: Let  $X$  be the number of winners. This is a binomial situation with  $n = 10$  and  $p = 0.10$ . We interpret “win at most once” as meaning “ $X \leq 1$ .” Then

$$\begin{aligned} P[ X \leq 1 ] &= P[ X = 0 ] + P[ X = 1 ] = \binom{10}{0} 0.10^0 \times 0.90^{10} + \binom{10}{1} 0.10^1 \times 0.90^9 \\ &= 0.90^{10} + 10 \times 0.10^1 \times 0.90^9 = 0.90^{10} + 0.90^9 = 0.90^9 (0.90 + 1) \\ &= 0.90^9 \times 1.90 \approx 0.736099 \end{aligned}$$

3. If  $X$  is binomial with parameters  $n$  and  $p$ , find an expression for  $P[ X \leq 1 ]$ .

SOLUTION: This is the same as the previous problem, but it’s in a generic form.

$$\begin{aligned} P[ X \leq 1 ] &= P[ X = 0 ] + P[ X = 1 ] = \binom{n}{0} p^0 (1-p)^n + \binom{n}{1} p^1 (1-p)^{n-1} \\ &= (1-p)^n + np(1-p)^{n-1} = (1-p)^{n-1} ((1-p) + np) \\ &= (1-p)^{n-1} (1 + (n-1)p) \end{aligned}$$

4. The probability is 0.038 that a person reached on a “cold call” by a telemarketer will make a purchase. If the telemarketer calls 40 people, what is the probability that at least one sale will result?

SOLUTION: Let  $X$  be the resulting number of sales. Certainly  $X$  is binomial with  $n = 40$  and  $p = 0.038$ . This “at least one” problem can be done with this standard trick:

$$P[ X \geq 1 ] = 1 - P[ X = 0 ] = 1 - \binom{40}{0} 0.038^0 \times 0.962^{40} = 1 - 0.962^{40} \\ \approx 0.787674.$$

5. The probability is 0.316 that an audit of a retail business will turn up irregularities in the collection of state sales tax. If 16 retail businesses are audited, find the probability that

- (a) exactly 5 will have irregularities in the collection of state sales tax.
- (b) at least 5 will have irregularities in the collection of state sales tax.
- (c) fewer than 5 will have irregularities in the collection of state sales tax.
- (d) at most 5 will have irregularities in the collection of state sales tax.
- (e) more than 5 will have irregularities in the collection of state sales tax.
- (f) no more than 5 will have irregularities in the collection of state sales tax.
- (g) no fewer than 5 will have irregularities in the collection of state sales tax.

SOLUTION: Let  $X$  be the number of businesses with irregularities of this form. Note that  $X$  is binomial with  $n = 16$  and  $p = 0.316$ . The calculations requested here are far too ugly to permit hand calculation, so a program like Minitab should be used.

(a) asks for  $P[ X = 5 ] = \binom{16}{5} 0.316^5 \times 0.684^{11} = 0.2110$ . This was done by Minitab.

(b) asks for  $P[ X \geq 5 ]$ . Use Minitab to get  $1 - P[ X \leq 4 ] = 1 - 0.3951 = 0.6049$ .

(c) asks for  $P[ X \leq 4 ] = 0.3951$ .

(d) asks for  $P[ X \leq 5 ] = 0.6062$ .

(e) asks for  $P[ X > 5 ]$ . Use Minitab to get  $1 - P[ X \leq 5 ] = 1 - 0.6062 = 0.3938$ .

(f) asks for  $P[ X \leq 5 ] = 0.6062$ . This is the same as (d).

(g) asks for  $P[ X \geq 5 ] = 0.6049$ . This is the same as (b).

6. A certain assembly line produces defects at the rate 0.072. If you observe 100 items from this list, what is the smallest number of defects that would cause a 1% rare-event alert? Specifically, if  $X$  is the number of defects, find the smallest value of  $k$  for which  $P[ X \geq k ] \leq 0.01$ .

SOLUTION: This will require an examination of the cumulative probabilities for  $X$ . Since Minitab computes cumulative probabilities in the  $\leq$  form, rephrase the question as

searching for the smallest  $k$  for which  $P[ X \leq k-1 ] \geq 0.99$ . Here is a set of cumulative probabilities calculated by Minitab:

| $x$ | $P[ X \leq x ]$ |
|-----|-----------------|
| 11  | 0.94417         |
| 12  | 0.97259         |
| 13  | 0.98751         |
| 14  | 0.99471         |
| 15  | 0.99791         |

The first time that the cumulative probabilities cross 0.99 occurs for  $x = 14$ . This corresponds to  $k-1$ , so we report that  $k = 15$ . The smallest number of defects which would cause a 1% rare-event alert is 15.

7. If you flip a fair coin 19 times, what is the probability that you will end up with an even number of heads?

SOLUTION: Let  $X$  be binomial with  $n = 19$  and  $p = \frac{1}{2}$ . This seems to be asking for  $P[ X = 0 ] + P[ X = 2 ] + P[ X = 4 ] + \dots + P[ X = 18 ]$ , which is an annoying calculation. However, we've got a trick. Consider the first 18 flips. The cumulative number of heads will either be even or odd. If it's even, then the 19<sup>th</sup> flip will preserve the even total with probability  $\frac{1}{2}$ . If it's odd, then the 19<sup>th</sup> flip will convert it to even with probability  $\frac{1}{2}$ . At the end, our probability of having an even number of heads must be  $\frac{1}{2}$ . This trick *only* works when  $p = \frac{1}{2}$ .

8. Suppose that you are playing roulette and betting on a single number. Your probability of winning on a single turn is  $\frac{1}{38} \approx 0.026316$ . You would like to get at least three winners. Find the minimum number of turns for which the probability of three or more winners is at least 0.80.

SOLUTION: The problem asks for the smallest  $n$  for which  $P[ X \geq 3 ] \geq 0.80$ . Since Minitab computes cumulative probabilities in the  $\leq$  form, we'll convert this question to finding the smallest  $n$  for which  $P[ X \leq 2 ] \leq 0.20$ .

This now requires a trial-and-error search. It helps to set up a diagram in which the first row is for an  $n$  that's likely to be too small and a last row for an  $n$  that is likely to be too big.

| $n$ | $P[ X \leq 2 ]$ | value of $n$ is |
|-----|-----------------|-----------------|
| 20  | 0.9851          | too small       |
|     |                 |                 |
|     |                 |                 |
| 200 | 0.1011          | too large       |

✂✂✂✂✂ BINOMIAL RANDOM VARIABLES ✂✂✂✂✂

Then intervening positions can be filled in. Let's try  $n = 100$ . This would result in  $P[ X \leq 2 ] = 0.5084$ , revealing that  $n = 100$  is too small. The table gets modified to this:

| $n$ | $P[ X \leq 2 ]$ | value of $n$ is |
|-----|-----------------|-----------------|
| 20  | 0.9851          | too small       |
| 100 | 0.5084          | too small       |
|     |                 |                 |
|     |                 |                 |
| 200 | 0.1011          | too large       |

Now try  $n = 150$ , getting  $P[ X \leq 2 ] = 0.2420$ . This says that  $n = 150$  is too small, but not by much. Here's what the table looks like:

| $n$ | $P[ X \leq 2 ]$ | value of $n$ is |
|-----|-----------------|-----------------|
| 20  | 0.9851          | too small       |
| 100 | 0.5084          | too small       |
| 150 | 0.2420          | too small       |
|     |                 |                 |
|     |                 |                 |
| 200 | 0.1011          | too large       |

After a little more effort, we get to this spot:

| $n$ | $P[ X \leq 2 ]$ | value of $n$ is |
|-----|-----------------|-----------------|
| 20  | 0.9851          | too small       |
| 100 | 0.5084          | too small       |
| 150 | 0.2420          | too small       |
| 160 | 0.2050          | too small       |
| 161 | 0.2016          | too small       |
| 162 | 0.1982          | just right!     |
| 200 | 0.1011          | too large       |

For  $n = 162$ , the cumulative probability drops below 0.20 for the first time. This is the requested number of times to play the game.

Recall that the binomial coefficient  $\binom{n}{r}$  is used to count the number of possible selections of  $r$  things out of  $n$ . Using  $n = 6$  and  $r = 2$  would provide the number of possible committees that could be obtained by selecting two people out of six.

The computational formula is  $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ . For  $n = 6$  and  $r = 2$ , this would give

$$\binom{6}{2} = \frac{6!}{2! \times 4!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1) \times (4 \times 3 \times 2 \times 1)} = \frac{6 \times 5}{2 \times 1} = 15.$$

If the six people are named  $A, B, C, D, E,$  and  $F,$  these would be the 15 possible committees:

|       |       |       |
|-------|-------|-------|
| $A B$ | $B C$ | $C E$ |
| $A C$ | $B D$ | $C F$ |
| $A D$ | $B E$ | $D E$ |
| $A E$ | $B F$ | $D F$ |
| $A F$ | $C D$ | $E F$ |

There are a few useful manipulations:

$$0! = 1 \quad (\text{by agreement})$$

$$\binom{n}{r} = \binom{n}{n-r}$$

In the committee example,  $\binom{6}{2} = \binom{6}{4}$ , so that the number of selections of two people to be on the committee is exactly equal to the number of selections of four people to *leave off* the committee.

$$\binom{n}{0} = \binom{n}{n} = 1$$

$$\binom{n}{1} = n$$

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

The hypergeometric distribution applies to the situation in which a random selection is to be made from a finite set. This is most easily illustrated with drawings from a deck of cards. Suppose that you select five cards at random from a standard deck of 52 cards.

This description certainly suggests a card game in which five cards are dealt to you from a standard deck. The game of poker generally begins this way. The fact that cards will also be dealt to other players does not influence the probability calculations, as long as the identities of those cards are not known to you.

You would like to know the probability that your five cards (your “hand”) will include exactly two aces. The computational logic proceeds along these steps:

- \* There are  $\binom{52}{5}$  possible selections of five cards out of 52. These selections are equally likely.
- \* There are four aces in the deck, and there are  $\binom{4}{2}$  ways in which you can identify two out of the four.
- \* There are 48 non-aces in the deck, and there are  $\binom{48}{3}$  ways in which you can identify three of these non-aces.
- \* The number of possible ways that your hand can have exactly two aces and exactly three non-aces is  $\binom{4}{2} \times \binom{48}{3}$ . This happens because every selection of the aces can be matched with every selection of the non-aces.
- \* The probability that your hand will have exactly two aces is  $\frac{\binom{4}{2} \times \binom{48}{3}}{\binom{52}{5}}$ .

The computation is not trivial. If you are deeply concerned with playing poker, the number  $\binom{52}{5}$  will come up often. It's  $\frac{52!}{5! \times 47!} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2,598,960$ .

Now note that  $\binom{4}{2} = 6$ ,  $\binom{48}{3} = \frac{48 \times 47 \times 46}{3 \times 2 \times 1} = 17,296$ . Then you can find the desired probability  $\frac{6 \times 17,296}{2,598,960} \approx 0.039930$ . This is about 4%.

This technology can be generalized in a useful random variable notation. We will let random  $X$  be the number of special items in a sample of  $n$  taken at random from a set of  $N$ .

You will sometimes see a distinction between “sampling with replacement” and “sampling without replacement.” The issue comes down to whether or not each sampled object is returned to the set of  $N$  before the next selection is made. (Returning an item to the set of  $N$  would make it possible for that item to appear in the sample more than once.) In virtually all applications, the sampling is *without* replacement.

The sampling is usually done sequentially, but it does not have to be. In a card game, the same probabilities would apply even if you were given all your cards in a single clump from the top of a well-shuffled deck. Of course, dealing out cards in clumps violates the etiquette of the game.

The process is sometimes described as “taking a sample of  $n$  from a finite population of  $N$ .” We then use these symbols:

- $N$       population size
- $n$       sample size
- $D$       number of special items in the population
- $N - D$  number of non-special items in the population
- $X$       (random) number of special items in the sample

This table lays out the notation:

|                                              | General notation | Card example                                                                    |
|----------------------------------------------|------------------|---------------------------------------------------------------------------------|
| Population size                              | $N$              | 52 (cards in deck)                                                              |
| Sample size                                  | $n$              | 5 (cards you will be dealt)                                                     |
| Special items in the population              | $D$              | 4 (aces in the deck)                                                            |
| Non-special items in the population          | $N - D$          | 48 (non-aces in the deck)                                                       |
| Random number of special items in the sample | $X$              | Number of special items in your hand (we asked for this to be 2 in the example) |

The probability structure of  $X$  is given by the hypergeometric formula

$$P[X = x] = \frac{\binom{D}{x} \binom{N - D}{n - x}}{\binom{N}{n}}$$

For our example, this was  $\frac{\binom{4}{2} \times \binom{48}{3}}{\binom{52}{5}}$ .

In a well-formed hypergeometric calculation

the numerator upper numbers add to the denominator upper number  
(as  $4 + 48 = 52$ )

the numerator lower numbers add to the denominator lower number  
(as  $2 + 3 = 5$ )

As an interesting curiosity, it happens that we can write the hypergeometric probability in the alternate form

$$P[X = x] = \frac{\binom{n}{x} \binom{N-n}{D-x}}{\binom{N}{D}}$$

In our example, this would be  $\frac{\binom{4}{2} \times \binom{48}{3}}{\binom{52}{5}} = \frac{\binom{5}{2} \times \binom{47}{2}}{\binom{52}{4}}$ .

The program Minitab, since release 13, can compute hypergeometric probabilities. Suppose that you would like to see the probabilities associated with the number of spades that you get in a hand of 13 cards. The game of bridge starts out by dealing 13 cards to each player, so this question is sometimes of interest to bridge players.

In column 1 of Minitab, lay out the integers 0 through 13. You can enter these manually, or you can use this little trick:

**Calc** ⇒ **Make Patterned Data** ⇒ **Simple Set of Numbers** ⇒

On the resulting panel, enter the information indicated. . .

Store patterned data in: (enter C1)

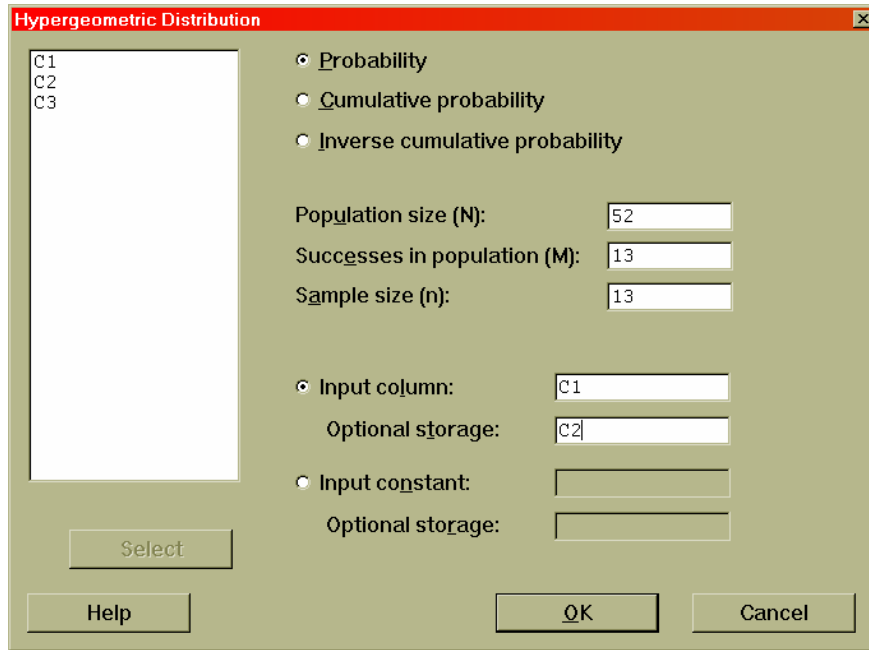
From first value: (enter 0)

To last value: (enter 13)

Minitab can then easily find all the probabilities at once:

**Calc** ⇒ **Probability Distributions** ⇒ **Hypergeometric** ⇒

Fill out the resulting panel to look like this:



The first 13, successes in population, corresponds to our symbol  $D$ . (It's unfortunate that Minitab used the letter  $X$  for this purpose.) The second 13, sample size, corresponds to our symbol  $n$  (and Minitab agrees).

When you click **OK**, the entire column of probabilities appears in column C2.

Here are some typical problems.

Example 1: What is the most likely number of spades that you will get in a hand of 13 cards?

Solution: If you examine the output that Minitab produced, you'll see

$$\begin{aligned} P[X = 2] &= 0.205873 \\ P[X = 3] &= 0.286330 \\ P[X = 4] &= 0.238608 \end{aligned}$$

All the other probabilities are much smaller. Thus, you're most likely to get three spades.

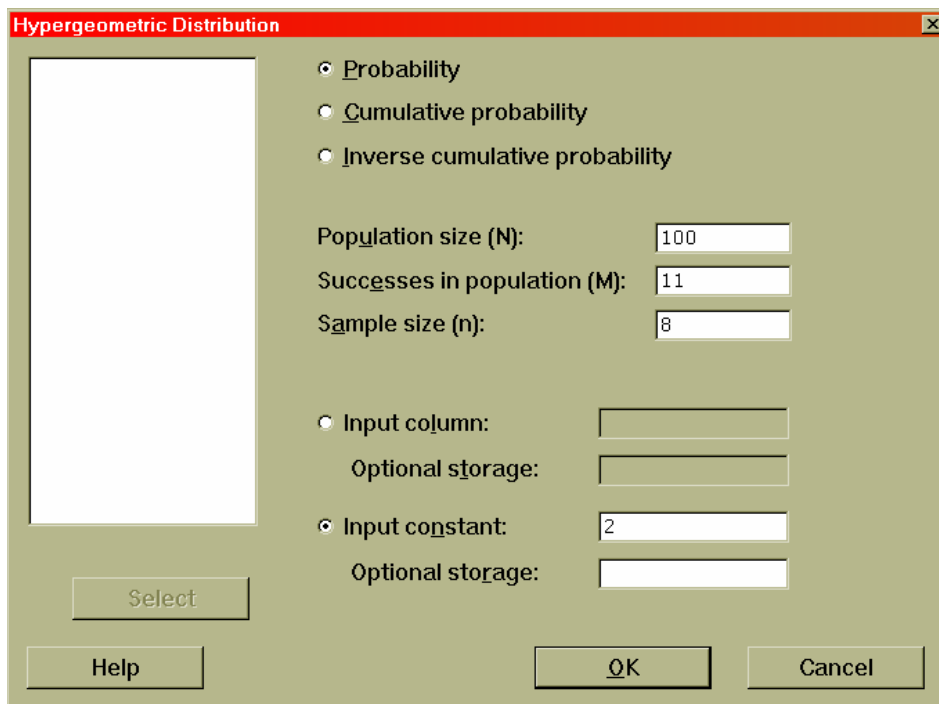
☺☺☺☺☺ THE HYPERGEOMETRIC RANDOM VARIABLE ☺☺☺☺☺

Example 2: Suppose that a shipment of 100 fruit crates has 11 crates in which the fruit shows signs of spoilage. A quality control inspection selects 8 crates at random, opens these selected crates, and then counts the number (out of 8) in which the fruit shows signs of spoilage. What is the probability that exactly two crates in the sample show signs of spoilage?

Solution: Let  $X$  be the number of bad crates in the sample. This is a hypergeometric random variable with  $N = 100$ ,  $D = 11$ ,  $n = 8$ , and we ask  $P[ X = 2 ]$ . This probability is

$$\frac{\binom{11}{2} \binom{89}{6}}{\binom{100}{8}}$$

The arithmetic is possible, but it's annoying. Let's use Minitab for this. The detail panel should be this:



Minitab will produce this information in its session window:

**Probability Density Function**

Hypergeometric with  $N = 100$ ,  $M = 11$ , and  $n = 8$

```
x  P( X = x )
2   0.171752
```

The requested probability is 0.171752.

The Poisson random variable is obtained by counting outcomes. The situation is not governed by a pre-set sample size, but rather we observe over a specified length of time or a specified spatial area. There is no conceptual upper limit to the number of counts that we might get. The Poisson would be used for

- The number of industrial accidents in a month
- The number of earthquakes to strike Turkey in a year
- The number of maple seedlings to sprout in a 10 m × 10 m patch of meadow
- The number of phone calls arriving at your help desk in a two-hour period

The Poisson has some similarities to the binomial and hypergeometric, so we'll lay out the essential differences in this table:

|                   | Binomial                                   | Hypergeometric           | Poisson                         |
|-------------------|--------------------------------------------|--------------------------|---------------------------------|
| Number of trials  | $n$                                        | $n$                      | no concept of sample size       |
| Population size   | Infinite (trials could go on indefinitely) | $N$                      | no concept of population size   |
| Event probability | $p$                                        | $\frac{D}{N}$            | no concept of event probability |
| Event rate        | no concept of event rate                   | no concept of event rate | $\lambda$                       |

The Poisson probability law is governed by a rate parameter  $\lambda$ . For example, if we are dealing with the number of industrial accidents in a month,  $\lambda$  will represent the expected rate. If  $X$  is this random variable, then the probability law is

$$P[ X = x ] = e^{-\lambda} \frac{\lambda^x}{x!}$$

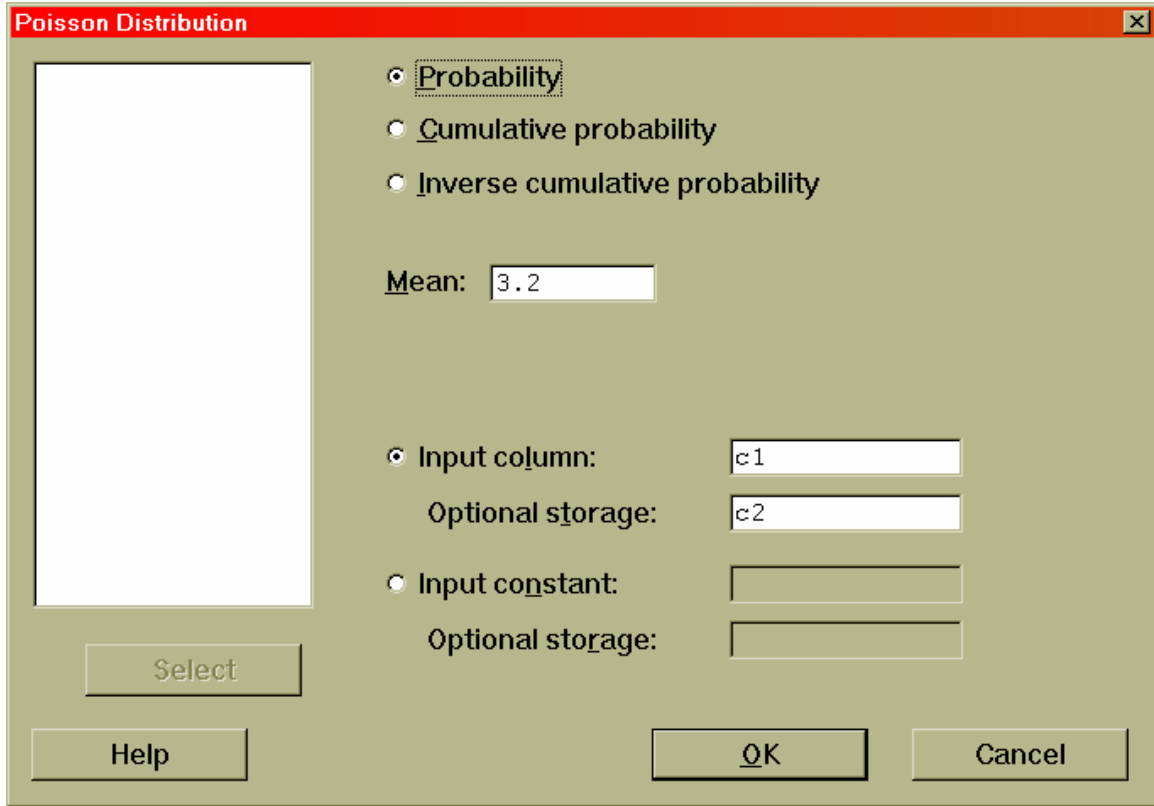
This calculations can be done for  $x = 0, 1, 2, 3, 4, \dots$  There is no upper limit.

If the rate is 3.2 accidents/month, then the probability that there will be exactly two accidents in any month is

$$P[ X = 2 ] = e^{-3.2} \frac{3.2^2}{2!} \approx 0.040762 \frac{10.24}{2} \approx 0.2087$$

Minitab can organize these calculations easily. In a column of the data sheet, say C1, enter the integers 0, 1, 2, 3, 4, ..., 10. It's easy to enter these directly, but you could also use **Calc** ⇒ **Make Patterned Data**. Then do **Calc** ⇒ **Probability Distributions** ⇒ **Poisson**.

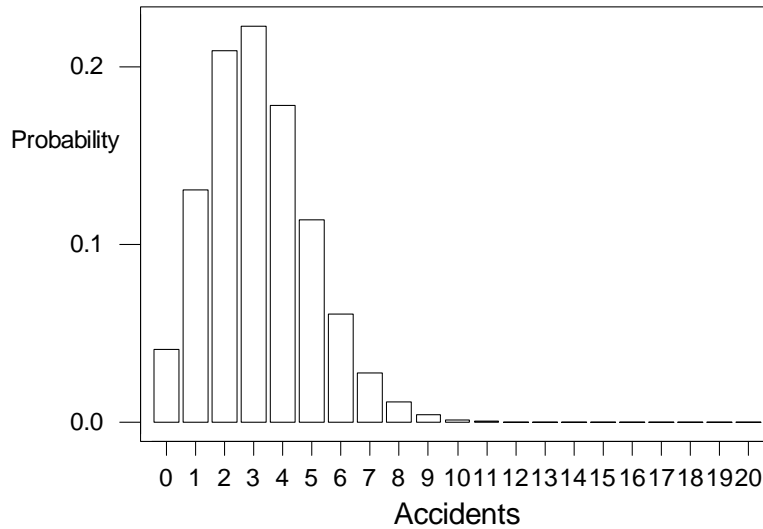
The information panel should then be filled as indicated:



The probabilities that result from this operation are these:

| Accidents | Probability |
|-----------|-------------|
| 0         | 0.040762    |
| 1         | 0.130439    |
| 2         | 0.208702    |
| 3         | 0.222616    |
| 4         | 0.178093    |
| 5         | 0.113979    |
| 6         | 0.060789    |
| 7         | 0.027789    |
| 8         | 0.011116    |
| 9         | 0.003952    |
| 10        | 0.001265    |

Here is a graph of the probability function for this Poisson random variable:



This is drawn all the way out to 20 accidents, but it's clear that nearly all the probability action is below 12.

**EXAMPLE:** The number of calls arriving at the Swampside Police Station follows a Poisson distribution with rate 4.6/hour. What is the probability that exactly six calls will come between 8:00 p.m. and 9:00 p.m.?

**SOLUTION:** Let  $X$  be the random number arriving in this one-hour time period. We'll use  $\lambda = 4.6$  and then find  $P[X = 6] = e^{-4.6} \frac{4.6^6}{6!} \approx 0.1323$ .

**EXAMPLE:** In the situation above, find the probability that exactly 7 calls will come between 9:00 p.m. and 10:30 p.m.

**SOLUTION:** Let  $Y$  be the random number arriving during this 90-minute period. The Poisson rate parameter expands and contracts appropriately, so the relevant value of  $\lambda$  is  $1.5 \times 4.6 = 6.9$ . We find  $P[Y = 7] = e^{-6.9} \frac{6.9^7}{7!} \approx 0.1489$ .

The Poisson random variable has an expected value that is exactly  $\lambda$ . The standard deviation is  $\sqrt{\lambda}$ .

EXAMPLE: If  $X$  is a Poisson random variable with  $\lambda = 225$ , would it be unusual to get a value of  $X$  which is less than 190?

SOLUTION: If we were asked for the exact number, we'd use Minitab to find  $P[X \leq 189] \approx 0.0077$ . This suggests that indeed it would be unusual to get an  $X$  value below 190. However, we can get a quick approximate answer by noting that  $E X = \text{mean of } X = \lambda = 225$ , and  $SD(X) = \sqrt{\lambda} = \sqrt{225} = 15$ . The value 190 is  $\frac{225 - 190}{15} \approx 2.33$  standard deviations below the mean; yes, it would be unusual to get a value that small.

This chart summarizes some relevant facts for the three useful discrete random variables.

| Random variable | Description                                                                                               | Probability function<br>$P[X = x]$                          | Expected value<br>(Mean) | Standard deviation                                                  |
|-----------------|-----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|--------------------------|---------------------------------------------------------------------|
| Binomial        | Number of successes in $n$ independent trials, each having success probability $p$                        | $\binom{n}{x} p^x (1-p)^{n-x}$                              | $np$                     | $\sqrt{np(1-p)}$                                                    |
| Hyper-geometric | Number of special items obtained in a sample of $n$ from a population of $N$ containing $D$ special items | $\frac{\binom{D}{x} \times \binom{N-D}{n-x}}{\binom{N}{n}}$ | $n \frac{D}{N}$          | $\sqrt{n \frac{D}{N} \left(1 - \frac{D}{N}\right) \frac{N-n}{N-1}}$ |
| Poisson         | Number of events observed over a specified period of time (or space) at event rate $\lambda$              | $e^{-\lambda} \frac{\lambda^x}{x!}$                         | $\lambda$                | $\sqrt{\lambda}$                                                    |

Suppose that the two random variables  $X$  and  $Y$  have this probability structure:

|          |         |         |
|----------|---------|---------|
|          | $Y = 1$ | $Y = 2$ |
| $X = 8$  | 0.12    | 0.18    |
| $X = 10$ | 0.20    | 0.40    |
| $X = 12$ | 0.02    | 0.08    |

We can check that the probability sums to 1. The easiest way to do this comes in appending one row and one column to hold totals:

|          |         |         |       |
|----------|---------|---------|-------|
|          | $Y = 1$ | $Y = 2$ | Total |
| $X = 8$  | 0.12    | 0.18    | 0.30  |
| $X = 10$ | 0.20    | 0.40    | 0.60  |
| $X = 12$ | 0.02    | 0.08    | 0.10  |
| Total    | 0.34    | 0.66    | 1.00  |

Thus  $P(Y = 1) = 0.34$  and  $P(Y = 2) = 0.66$ . Then

$$E Y = 0.34 \times 1 + 0.66 \times 2 = 1.66 = \mu_Y$$

$$E Y^2 = 0.34 \times 1^2 + 0.66 \times 2^2 = 2.98$$

$$\sigma_Y^2 = \text{Var}(Y) = E Y^2 - (E Y)^2 = 2.98 - 1.66^2 = 2.98 - 2.7556 = 0.2244$$

$$\sigma_Y = \text{SD}(Y) = \sqrt{0.2244} \approx 0.4737$$

Using  $P(X = 8) = 0.30$ ,  $P(X = 10) = 0.60$ , and  $P(X = 12) = 0.10$ . Then

$$E X = 0.30 \times 8 + 0.60 \times 10 + 0.10 \times 12 = 9.6 = \mu_X$$

$$E X^2 = 0.30 \times 8^2 + 0.60 \times 10^2 + 0.10 \times 12^2 = 93.6$$

$$\sigma_X^2 = \text{Var}(X) = E X^2 - (E X)^2 = 93.6 - 9.6^2 = 93.6 - 92.16 = 1.44$$

$$\sigma_X = \text{SD}(X) = \sqrt{1.44} = 1.2$$

👉 👉 👉 COVARIANCE FOR DISCRETE RANDOM VARIABLES 👉 👉 👉

Let's introduce the calculation of Covariance( $X, Y$ ) =  $\text{Cov}(X, Y)$ . This is defined as

$$\text{Cov}(X, Y) = E[ (X - \mu_X)(Y - \mu_Y) ]$$

but it is more easily calculated as

$$\text{Cov}(X, Y) = E[ X Y ] - \mu_X \mu_Y$$

Here  $E[ X Y ] =$

$$\begin{aligned} & 0.12 \times 8 \times 1 + 0.18 \times 8 \times 2 \\ & + 0.20 \times 10 \times 1 + 0.40 \times 10 \times 2 \\ & + 0.02 \times 12 \times 1 + 0.08 \times 12 \times 2 = 16.00 \end{aligned}$$

Then  $\text{Cov}(X, Y) = 16 - 9.6 \times 1.66 = 16 - 15.936 = 0.064$ .

We can then find the *correlation* of  $X$  and  $Y$  as

$$\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{0.064}{1.2 \times 0.4737} \approx 0.1126$$