Chapter 4

Discrete Random Variables

4.1 Discrete Random Variables[1]

4.1.1 Student Learning Objectives

By the end of this chapter, the student should be able to:

- Recognize and understand discrete probability distribution functions, in general.
- Calculate and interpret expected values.
- Recognize the binomial probability distribution and apply it appropriately.
- Recognize the Poisson probability distribution and apply it appropriately (optional).
- Recognize the geometric probability distribution and apply it appropriately (optional).
- Recognize the hypergeometric probability distribution and apply it appropriately (optional).
- Classify discrete word problems by their distributions.

4.1.2 Introduction

A student takes a 10 question true-false quiz. Because the student had such a busy schedule, he or she could not study and randomly guesses at each answer. What is the probability of the student passing the test with at least a 70%?

Small companies might be interested in the number of long distance phone calls their employees make during the peak time of the day. Suppose the average is 20 calls. What is the probability that the employees make more than 20 long distance phone calls during the peak time?

These two examples illustrate two different types of probability problems involving discrete random variables. Recall that discrete data are data that you can count. A **random variable** describes the outcomes of a statistical experiment both in words. The values of a random variable can vary with each repetition of an experiment.

In this chapter, you will study probability problems involving discrete random distributions. You will also study long-term averages associated with them.

4.1.3 Random Variable Notation

Upper case letters like X or Y denote a random variable. Lower case letters like x or y denote the value of a random variable. If X **is a random variable, then** X **is defined in words.**

[1] This content is available online at <http://http://cnx.org/content/m16825/1.11/>.

For example, let X = the number of heads you get when you toss three fair coins. The sample space for the toss of three fair coins is TTT; THH; HTH; HHT; HTT; THT; TTH; HHH. Then, x = 0, 1, 2, 3. X is in words and x is a number. Notice that for this example, the x values are countable outcomes. Because you can count the possible values that X can take on and the outcomes are random (the x values 0, 1, 2, 3), X is a discrete random variable.

4.1.4 Optional Collaborative Classroom Activity

Toss a coin 10 times and record the number of heads. After all members of the class have completed the experiment (tossed a coin 10 times and counted the number of heads), fill in the chart using a heading like the one below. Let X = the number of heads in 10 tosses of the coin.

X	**Frequency of X**	**Relative Frequency of X**

Table 4.1

- Which value(s) of X occurred most frequently?
- If you tossed the coin 1,000 times, what values would X take on? Which value(s) of X do you think would occur most frequently?
- What does the relative frequency column sum to?

4.2 Probability Distribution Function (PDF) for a Discrete Random Variable[2]

A discrete **probability distribution function** has two characteristics:

- Each probability is between 0 and 1, inclusive.
- The sum of the probabilities is 1.

P(X) is the notation used to represent a discrete **probability** distribution function.

Example 4.1
A child psychologist is interested in the number of times a newborn baby's crying wakes its mother after midnight. For a random sample of 50 mothers, the following information was obtained. Let X = the number of times a newborn wakes its mother after midnight. For this example, x = 0, 1, 2, 3, 4, 5.

P(X = x) = probability that X takes on a value x.

[2] This content is available online at <http://http://cnx.org/content/m16831/1.11/>.

x	P(X = x)
0	P(X=0) $= \frac{2}{50}$
1	P(X=1) $= \frac{11}{50}$
2	P(X=2) $= \frac{23}{50}$
3	P(X=3) $= \frac{9}{50}$
4	P(X=4) $= \frac{4}{50}$
5	P(X=5) $= \frac{1}{50}$

Table 4.2

X takes on the values 0, 1, 2, 3, 4, 5. This is a discrete PDF because

1. Each P(X = x) is between 0 and 1, inclusive.
2. The sum of the probabilities is 1, that is,

$$\frac{2}{50} + \frac{11}{50} + \frac{23}{50} + \frac{9}{50} + \frac{4}{50} + \frac{1}{50} = 1 \tag{4.1}$$

Example 4.2
Suppose Nancy has classes **3 days** a week. She attends classes 3 days a week **80%** of the time, **2 days 15%** of the time, **1 day 4%** of the time, and **no days 1%** of the time.

Problem 1 ***(Solution on p. 205.)***
Let X = the number of days Nancy ____________________ .

Problem 2 ***(Solution on p. 205.)***
X takes on what values?

Problem 3 ***(Solution on p. 205.)***
Construct a probability distribution table (called a PDF table) like the one in the previous example. The table should have two columns labeled x and P(X = x). What does the P(X = x) column sum to?

4.3 Mean or Expected Value and Standard Deviation[3]

The **expected value** is often referred to as the **"long-term"average or mean** . This means that over the long term of doing an experiment over and over, you would **expect** this average.

The **mean** of a random variable X is μ. If we do an experiment many times (for instance, flip a fair coin, as Karl Pearson did, 24,000 times and let X = the number of heads) and record the value of X each time, the average gets closer and closer to μ as we keep repeating the experiment. This is known as the **Law of Large Numbers**.

NOTE: To find the expected value or long term average, μ, simply multiply each value of the random variable by its probability and add the products.

[3]This content is available online at <http://http://cnx.org/content/m16828/1.13/>.

A Step-by-Step Example

A men's soccer team plays soccer 0, 1, or 2 days a week. The probability that they play 0 days is 0.2, the probability that they play 1 day is 0.5, and the probability that they play 2 days is 0.3. Find the long-term average, μ, or expected value of the days per week the men's soccer team plays soccer.

To do the problem, first let the random variable X = the number of days the men's soccer team plays soccer per week. X takes on the values 0, 1, 2. Construct a PDF table, adding a column $xP(X = x)$. In this column, you will multiply each x value by its probability.

Expected Value Table

x	**P(X=x)**	x**P(X=x)**
0	0.2	(0)(0.2) = 0
1	0.5	(1)(0.5) = 0.5
2	0.3	(2)(0.3) = 0.6

Table 4.4: This table is called an expected value table. The table helps you calculate the expected value or long-term average.

Add the last column to find the long term average or expected value: $(0)(0.2) + (1)(0.5) + (2)(0.3) = 0 + 0.5. 0.6 = 1.1.$

The expected value is 1.1. The men's soccer team would, on the average, expect to play soccer 1.1 days per week. The number 1.1 is the long term average or expected value if the men's soccer team plays soccer week after week after week. We say $\mu = 1.1$

Example 4.3

Find the expected value for the example about the number of times a newborn baby's crying wakes its mother after midnight. The expected value is the expected number of times a newborn wakes its mother after midnight.

x	**P(X=x)**	x**P(X=x)**
0	P(X=0) $= \frac{2}{50}$	$(0)\left(\frac{2}{50}\right) = 0$
1	P(X=1) $= \frac{11}{50}$	$(1)\left(\frac{11}{50}\right) = \frac{11}{50}$
2	P(X=2) $= \frac{23}{50}$	$(2)\left(\frac{23}{50}\right) = \frac{46}{50}$
3	P(X=3) $= \frac{9}{50}$	$(3)\left(\frac{9}{50}\right) = \frac{27}{50}$
4	P(X=4) $= \frac{4}{50}$	$(4)\left(\frac{4}{50}\right) = \frac{16}{50}$
5	P(X=5) $= \frac{1}{50}$	$(5)\left(\frac{1}{50}\right) = \frac{5}{50}$

Table 4.5: You expect a newborn to wake its mother after midnight 2.1 times, on the average.

Add the last column to find the expected value. μ = Expected Value = $\frac{105}{50} = 2.1$

Problem

Go back and calculate the expected value for the number of days Nancy attends classes a week. Construct the third column to do so.

Solution

2.74 days a week.

Example 4.4
Suppose you play a game of chance in which you choose 5 numbers from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. You may choose a number more than once. You pay $2 to play and could profit $100,000 if you match all 5 numbers in order (you get your $2 back plus $100,000). Over the long term, what is your **expected** profit of playing the game?

To do this problem, set up an expected value table for the amount of money you can profit.

Let X = the amount of money you profit. The values of x are not 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Since you are interested in your profit (or loss), the values of x are 100,000 dollars and -2 dollars.

To win, you must get all 5 numbers correct, in order. The probability of choosing one correct number is $\frac{1}{10}$ because there are 10 numbers. You may choose a number more than once. The probability of choosing all 5 numbers correctly and in order is:

$$\frac{1}{10} * \frac{1}{10} * \frac{1}{10} * \frac{1}{10} * \frac{1}{10} * \ = \ 1 * 10^{-5} = 0.00001 \quad (4.2)$$

Therefore, the probability of winning is 0.00001 and the probability of losing is

$$1 - 0.00001 = 0.99999 \quad (4.3)$$

The expected value table is as follows.

	x	P(X=x)	xP(X=x)
Loss	-2	0.99999	(-2)(0.99999)=-1.99998
Profit	100,000	0.00001	(100000)(0.00001)=1

Table 4.6: Add the last column. -1.99998 + 1 = -0.99998

Since -0.99998 is about -1, you would, on the average, expect to lose approximately one dollar for each game you play. However, each time you play, you either lose $2 or profit $100,000. The $1 is the average or expected LOSS per game after playing this game over and over.

Example 4.5
Suppose you play a game with a biased coin. You play each game by tossing the coin once. P(heads) $= \frac{2}{3}$ and P(tails) $= \frac{1}{3}$. If you toss a head, you pay $6. If you toss a tail, you win $10. If you play this game many times, will you come out ahead?

Problem 1 ***(Solution on p. 205.)***
Define a random variable X.

Problem 2 ***(Solution on p. 205.)***
Complete the following expected value table.

	x	____	____
WIN	10	$\frac{1}{3}$	____
LOSE	____	____	$\frac{-12}{3}$

Table 4.7

Problem 3 *(Solution on p. 205.)*

What is the expected value, μ? Do you come out ahead?

Like data, probability distributions have standard deviations. To calculate the standard deviation (σ) of a probability distribution, find each deviation, square it, multiply it by its probability, add the products, and take the square root . To understand how to do the calculation, look at the table for the number of days per week a men's soccer team plays soccer. To find the standard deviation, add the entries in the column labeled $(x-\mu)^2 \cdot P(X=x)$ and take the square root.

x	P(X=x)	xP(X=x)	(x -μ)^{2}P(X=x)
0	0.2	(0)(0.2) = 0	$(0-1.1)^2(.2) = 0.242$
1	0.5	(1)(0.5) = 0.5	$(1-1.1)^2(.5) = 0.005$
2	0.3	(2)(0.3) = 0.6	$(2-1.1)^2(.3) = 0.243$

Table 4.8

Add the last column in the table. $0.242 + 0.005 + 0.243 = 0.490$. The standard deviation is the square root of 0.49. $\sigma = \sqrt{0.49} = 0.7$

Generally for probability distributions, we use a calculator or a computer to calculate μ and σ to reduce roundoff error. For some probability distributions, there are short-cut formulas that calculate μ and σ.

4.4 Common Discrete Probability Distribution Functions[4]

Some of the more common discrete probability functions are binomial, geometric, hypergeometric, and Poisson. Most elementary courses do not cover the geometric, hypergeometric, and Poisson. Your instructor will let you know if he or she wishes to cover these distributions.

A probability distribution function is a pattern. You try to fit a probability problem into a **pattern** or distribution in order to perform the necessary calculations. These distributions are tools to make solving probability problems easier. Each distribution has its own special characteristics. Learning the characteristics enables you to distinguish among the different distributions.

4.5 Binomial[5]

The characteristics of a binomial experiment are:

1. There are a fixed number of trials. Think of trials as repetitions of an experiment. The letter n denotes the number of trials.
2. There are only 2 possible outcomes, called "success" and, "failure" for each trial. The letter p denotes the probability of a success on one trial and q denotes the probability of a failure on one trial. $p + q = 1$.
3. The n trials are independent and are repeated using identical conditions. Because the n trials are independent, the outcome of one trial does not affect the outcome of any other trial. Another way of saying this is that for each individual trial, the probability, p, of a success and probability, q, of a failure remain the same. For example, randomly guessing at a true - false statistics question has only two outcomes. If a success is guessing correctly, then a failure is guessing incorrectly. Suppose Joe always

[4]This content is available online at <http://http://cnx.org/content/m16821/1.5/>.
[5]This content is available online at <http://http://cnx.org/content/m16820/1.12/>.

guesses correctly on any statistics true - false question with probability $p = 0.6$. Then, $q = 0.4$ .This means that for every true - false statistics question Joe answers, his probability of success ($p = 0.6$) and his probability of failure ($q = 0.4$) remain the same.

The outcomes of a binomial experiment fit a **binomial probability distribution**. The random variable $X =$ the number of successes obtained in the n independent trials.

The mean, μ, and variance, σ^2, for the binomial probability distribution is $\mu = np$ and $\sigma^2 = npq$. The standard deviation, σ, is then $\sigma = \sqrt{npq}$.

Any experiment that has characteristics 2 and 3 is called a **Bernoulli Trial** (named after Jacob Bernoulli who, in the late 1600s, studied them extensively). A binomial experiment takes place when the number of successes is counted in one or more Bernoulli Trials.

Example 4.6
At ABC College, the withdrawal rate from an elementary physics course is 30% for any given term. This implies that, for any given term, 70% of the students stay in the class for the entire term. A "success" could be defined as an individual who withdrew. The random variable is X = the number of students who withdraw from the elementary physics course per term.

Example 4.7
Suppose you play a game that you can only either win or lose. The probability that you win any game is 55% and the probability that you lose is 45%. If you play the game 20 times, what is the probability that you win 15 of the 20 games? Here, if you define X = the number of wins, then X takes on the values X = 0, 1, 2, 3, ..., 20. The probability of a success is $p = 0.55$. The probability of a failure is $q = 0.45$. The number of trials is $n = 20$. The probability question can be stated mathematically as $P(X = 15)$.

Example 4.8
A fair coin is flipped 15 times. What is the probability of getting more than 10 heads? Let X = the number of heads in 15 flips of the fair coin. X takes on the values x = 0, 1, 2, 3, ..., 15. Since the coin is fair, p = 0.5 and q = 0.5. The number of trials is n = 15. The probability question can be stated mathematically as $P(X > 10)$.

Example 4.9
Approximately 70% of statistics students do their homework in time for it to be collected and graded. In a statistics class of 50 students, what is the probability that at least 40 will do their homework on time?

Problem 1 ***(Solution on p. 205.)***
This is a binomial problem because there is only a success or a __________, there are a definite number of trials, and the probability of a success is 0.70 for each trial.

Problem 2 ***(Solution on p. 205.)***
If we are interested in the number of students who do their homework, then how do we define X?

Problem 3 ***(Solution on p. 205.)***
What values does X take on?

Problem 4 ***(Solution on p. 205.)***
What is a "failure", in words?

The probability of a success is p = 0.70. The number of trial is n = 50.

Problem 5 ***(Solution on p. 205.)***
If $p + q = 1$, then what is q?

Problem 6 *(Solution on p. 205.)*

The words "at least" translate as what kind of inequality?

The probability question is $P(X \geq 40)$.

4.5.1 Notation for the Binomial: B = Binomial Probability Distribution Function

$X \sim B(n, p)$

Read this as "X is a random variable with a binomial distribution." The parameters are n and p. n = number of trials p = probability of a success on each trial

Example 4.10

It has been stated that about 41% of adult workers have a high school diploma but do not pursue any further education. If 20 adult workers are randomly selected, find the probability that at most 12 of them have a high school diploma but do not pursue any further education. How many adult workers do you expect to have a high school diploma but do not pursue any further education?

Let X = the number of workers who have a high school diploma but do not pursue any further education.

X takes on the values 0, 1, 2, ..., 20 where n = 20 and p = 0.41. q = 1 - 0.41 = 0.59. $X \sim B(20, 0.41)$

Find $P(X \leq 12)$. $P(X \leq 12) = 0.9738$. (calculator or computer)

Using the TI-83+ or the TI-84 calculators, the calculations are as follows. Go into 2nd DISTR. The syntax for the instructions are

To calculate (X **= value): binompdf(**n**,** p**, number)** If "number" is left out, the result is the binomial probability table.

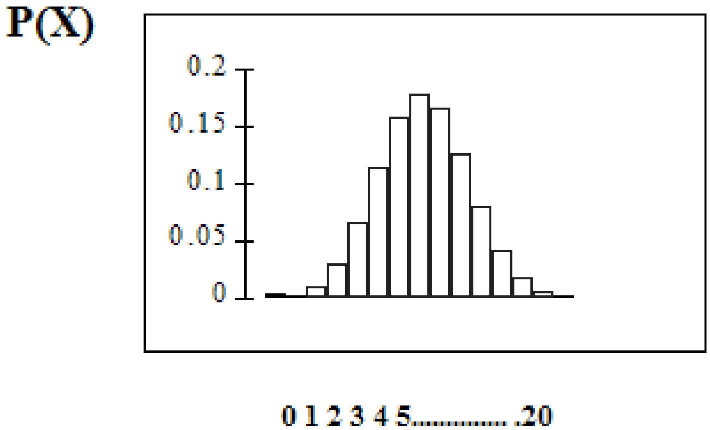
To calculate $P(X \leq$ **value**$)$**: binomcdf(**n**,** p**, number)** If "number" is left out, the result is the cumulative binomial probability table.

For this problem: After you are in 2nd DISTR, arrow down to A:binomcdf. Press ENTER. Enter 20,.41,12). The result is $P(X \leq 12) = 0.9738$**.**

NOTE: If you want to find $P(X = 12)$, use the pdf (0:binompdf). If you want to find $P(X > 12)$, use 1 - binomcdf(20,.41,12).

The probability at most 12 workers have a high school diploma but do not pursue any further education is 0.9738

The graph of $X \sim B(20, 0.41)$ is:



The y-axis contains the probability of X, where X = the number of workers who have only a high school diploma.

The number of adult workers that you expect to have a high school diploma but not pursue any further education is the mean, $\mu = np = (20)\,(0.41) = 8.2$.

The formula for the variance is $\sigma^2 = npq$. The standard deviation is $\sigma = \sqrt{npq}$. $\sigma = \sqrt{(20)\,(0.41)\,(0.59)} = 2.20$.

Example 4.11
The following example illustrates a problem that is **not** binomial. It violates the condition of independence. ABC College has a student advisory committee made up of 10 staff members and 6 students. The committee wishes to choose a chairperson and a recorder. What is the probability that the chairperson and recorder are both students? All names of the committee are put into a box and two names are drawn **without replacement**. The first name drawn determines the chairperson and the second name the recorder. There are two trials. However, the trials are not independent because the outcome of the first trial affects the outcome of the second trial. The probability of a student on the first draw is $\frac{6}{16}$. The probability of a student on the second draw is $\frac{5}{15}$, when the first draw produces a student. The probability is $\frac{6}{15}$ when the first draw produces a staff member. The probability of drawing a student's name changes for each of the trials and, therefore, violates the condition of independence.

4.6 Geometric (optional)[6]

The characteristics of a geometric experiment are:

1. There are one or more Bernoulli trials with all failures except the last one, which is a success. In other words, you keep repeating what you are doing until the first success. Then you stop. For example, you throw a dart at a bull's eye until you hit the bull's eye. The first time you hit the bull's eye is a "success" so you stop throwing the dart. It might take you 6 tries until you hit the bull's eye. You can think of the trials as failure, failure, failure, failure, failure, success. STOP.
2. In theory, the number of trials could go on forever. There must be at least one trial.
3. The probability,p, of a success and the probability, q, of a failure is the same for each trial. $p + q = 1$ and $q = 1 - p$. For example, the probability of rolling a 3 when you throw one fair die is $\frac{1}{6}$. This is

[6]This content is available online at <http://http://cnx.org/content/m16822/1.13/>.

true no matter how many times you roll the die. Suppose you want to know the probability of getting the first 3 on the fifth roll. On rolls 1, 2, 3, and 4, you do not get a face with a 3. The probability for each of rolls 1, 2, 3, and 4 is $q = \frac{5}{6}$, the probability of a failure. The probability of getting a 3 on the fifth roll is $\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \cdot = 0.0804$

The outcomes of a geometric experiment fit a geometric probability distribution. The random variable $X =$ the number of independent trials until the first success. The mean and variance are in the summary in this chapter.

Example 4.12
You play a game of chance that you can either win or lose (there are no other possibilities) **until** you lose. Your probability of losing is $p = 0.57$. What is the probability that it takes 5 games until you lose? Let X = the number of games you play until you lose (includes the losing game). Then X takes on the values 1, 2, 3, ... (could go on indefinitely). The probability question is $P(X = 5)$.

Example 4.13
A safety engineer feels that 35% of all industrial accidents in her plant are caused by failure of employees to follow instructions. She decides to look at the accident reports **until** she finds one that shows an accident caused by failure of employees to follow instructions. On the average, how many reports would the safety engineer **expect** to look at until she finds a report showing an accident caused by employee failure to follow instructions? What is the probability that the safety engineer will have to examine at least 3 reports until she finds a report showing an accident caused by employee failure to follow instructions?

Let X = the number of accidents the safety engineer must examine **until** she finds a report showing an accident caused by employee failure to follow instructions. X takes on the values 1, 2, 3, ... The first question asks you to find the **expected value** or the mean. The second question asks you to find $P(X \geq 3)$. ("At least" translates as a "greater than or equal to" symbol).

Example 4.14
Suppose that you are looking for a chemistry lab partner. The probability that someone agrees to be your lab partner is 0.55. Since you need a lab partner very soon, you ask every chemistry student you are acquainted with **until** one says that he/she will be your lab partner. What is the probability that the fourth person says yes?

This is a geometric problem because you may have a number of failures before you have the one success you desire. Also, the probability of a success stays the same each time you ask a chemistry student to be your lab partner. There is no definite number of trials (number of times you ask a chemistry student to be your partner).

Problem 1
Let X = the number of ____________ you must ask ____________ one says yes.

Solution
Let X = the number of **chemistry students** you must ask **until** one says yes.

Problem 2 *(Solution on p. 205.)*
What values does X take on?

Problem 3 *(Solution on p. 205.)*
What are p and q?

Problem 4 *(Solution on p. 206.)*
The probability question is P(_______).

4.6.1 Notation for the Geometric: G = Geometric Probability Distribution Function

$X \sim G(p)$

Read this as "X is a random variable with a geometric distribution." The parameter is p. p = the probability of a success for each trial.

Example 4.15

Assume that the probability of a defective computer component is 0.02. Find the probability that the first defect is caused by the 7th component tested. How many components do you expect to test until one is found to be defective?

Let X = the number of computer components tested until the first defect is found.

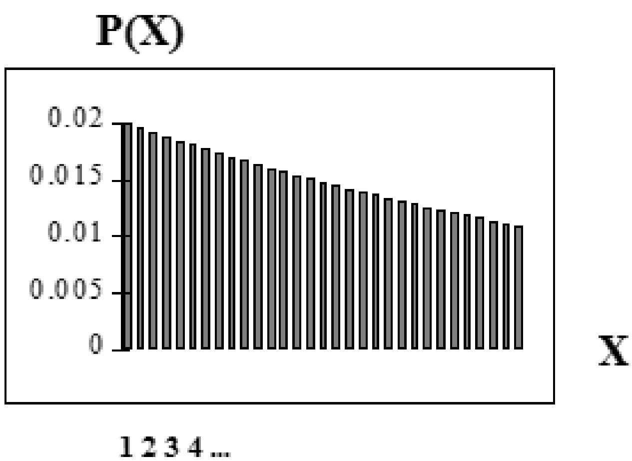
X takes on the values 1, 2, 3, ... where $p = 0.02$. $X \sim$ G(0.02)

Find $P(X = 7)$. $P(X = 7) = 0.0177$. (calculator or computer)

TI-83+ and TI-84: **For a general discussion, see this example (binomial)**. The syntax is similar. The geometric parameter list is (p, number) If "number" is left out, the result is the geometric probability table. For this problem: **After you are in 2nd DISTR, arrow down to D:geometpdf. Press ENTER. Enter .02,7). The result is $P(X = 7) = 0.0177$.**

The probability that the 7th component is the first defect is 0.0177.

The graph of $X \sim$ G(0.02) is:



The y-axis contains the probability of X, where X = the number of computer components tested.

The number of components that you would expect to test until you find the first defective one is the mean, μ = 50.

The formula for the mean is $\mu = \frac{1}{p} = \frac{1}{0.02} = 50$

The formula for the variance is $\sigma^2 = \frac{1}{p} \cdot \left(\frac{1}{p} - 1\right) = \frac{1}{0.02} \cdot \left(\frac{1}{0.02} - 1\right) = 2450$

The standard deviation is $\sigma = \sqrt{\frac{1}{p} \cdot \left(\frac{1}{p} - 1\right)} = \sqrt{\frac{1}{0.02} \cdot \left(\frac{1}{0.02} - 1\right)} = 49.5$

4.7 Hypergeometric (optional)[7]

The characteristics of a hypergeometric experiment are:

1. You take samples from **2** groups.
2. You are concerned with a group of interest, called the first group.
3. You sample **without replacement** from the combined groups. For example, you want to choose a softball team from a combined group of 11 men and 13 women. The team consists of 10 players.
4. Each pick is **not** independent, since sampling is without replacement. In the softball example, the probability of picking a women first is $\frac{13}{24}$. The probability of picking a man second is $\frac{11}{23}$ if a woman was picked first. It is $\frac{10}{23}$ if a man was picked first. The probability of the second pick depends on what happened in the first pick.
5. You are **not** dealing with Bernoulli Trials.

The outcomes of a hypergeometric experiment fit a **hypergeometric probability** distribution. The random variable $X =$ the number of items from the group of interest. The mean and variance are given in the summary.

Example 4.16
A candy dish contains 100 jelly beans and 80 gumdrops. Fifty candies are picked at random. What is the probability that 35 of the 50 are gumdrops? The two groups are jelly beans and gumdrops. Since the probability question asks for the probability of picking gumdrops, the group of interest (first group) is gumdrops. The size of the group of interest (first group) is 80. The size of the second group is 100. The size of the sample is 50 (jelly beans or gumdrops). Let X = the number of gumdrops in the sample of 50. X takes on the values x = 0, 1, 2, ..., 50. The probability question is $P(X = 35)$.

Example 4.17
Suppose a shipment of 100 VCRs is known to have 10 defective VCRs. An inspector chooses 12 for inspection. He is interested in determining the probability that, among the 12, at most 2 are defective. The two groups are the 90 non-defective VCRs and the 10 defective VCRs. The group of interest (first group) is the defective group because the probability question asks for the probability of at most 2 defective VCRs. The size of the sample is 12 VCRs. (They may be non-defective or defective.) Let X = the number of defective VCRs in the sample of 12. X takes on the values 0, 1, 2, ..., 10. X may not take on the values 11 or 12. The sample size is 12, but there are only 10 defective VCRs. The inspector wants to know $P(X \leq 2)$ ("At most" means "less than or equal to").

Example 4.18
You are president of an on-campus special events organization. You need a committee of 7 to plan a special birthday party for the president of the college. Your organization consists of 18 women and 15 men. You are interested in the number of men on your committee. What is the probability that your committee has more than 4 men?

This is a hypergeometric problem because you are choosing your committee from two groups (men and women).

Problem 1 *(Solution on p. 206.)*
Are you choosing with or without replacement?

Problem 2 *(Solution on p. 206.)*
What is the group of interest?

[7] This content is available online at <http://http://cnx.org/content/m16824/1.12/>.

Problem 3 *(Solution on p. 206.)*
How many are in the group of interest?

Problem 4 *(Solution on p. 206.)*
How many are in the other group?

Problem 5 *(Solution on p. 206.)*
Let X = _________ on the committee. What values does X take on?

Problem 6 *(Solution on p. 206.)*
The probability question is P(_______).

4.7.1 Notation for the Hypergeometric: H = Hypergeometric Probability Distribution Function

$X \sim H\,(r,\ b,\ n)$

Read this as "X is a random variable with a hypergeometric distribution." The parameters are r, b, and n. r = the size of the group of interest (first group), b = the size of the second group, n = the size of the chosen sample

Example 4.19
A school site committee is to be chosen from 6 men and 5 women. If the committee consists of 4 members, what is the probability that 2 of them are men? How many men do you expect to be on the committee?

Let X = the number of men on the committee of 4. The men are the group of interest (first group).

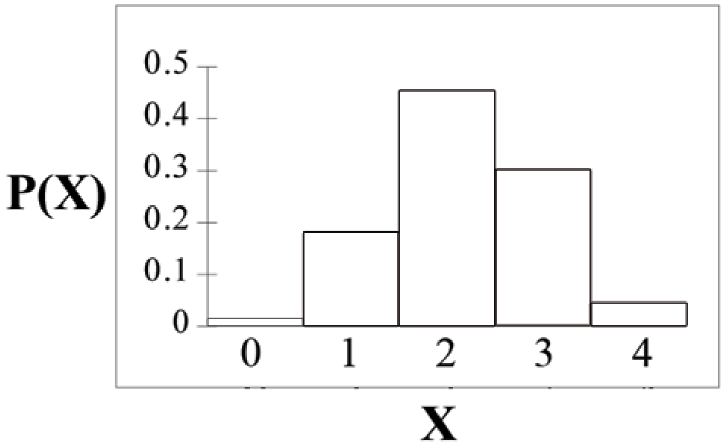
X takes on the values 0, 1, 2, 3, 4, where $r = 6$, $b = 5$, and $n = 4$. $X \sim H(6, 5, 4)$

Find $P\,(X = 2)$. $P\,(X = 2) = 0.4545$(calculator or computer)

NOTE: Currently, the TI-83+ and TI-84 do not have hypergeometric probability functions. There are a number of computer packages, including Microsoft Excel, that do.

The probability that there are 2 men on the committee is about 0.45.

The graph of $X \sim H(6, 5, 4)$ is:



The y-axis contains the probability of X, where X = the number of men on the committee.

You would expect $m = 2.18$(about 2) men on the committee.

The formula for the mean is $\mu = \frac{n \cdot r}{r+b} = \frac{4 \cdot 6}{6+5} = 2.18$

The formula for the variance is fairly complex. You will find it in the Summary of the Discrete Probability Functions Chapter (Section 4.9).

4.8 Poisson[8]

Characteristics of a Poisson experiment are:

1. You are interested in the number of times something happens in a certain **interval**. For example, a book editor might be interested in the number of words spelled incorrectly in a particular book. It might be that, on the average, there are 5 words spelled incorrectly in 100 pages. The interval is the 100 pages.
2. The Poisson may be derived from the binomial if the probability of success is "small" (such as 0.01) and the number of trials is "large" (such as 1000). You will verify the relationship in the homework exercises. n is the number of trials and p is the probability of a "success."

The outcomes of a Poisson experiment fit a **Poisson probability distribution**. The random variable $X =$ the number of occurrences in the interval of interest. The mean and variance are given in the summary.

Example 4.20
The average number of loaves of bread put on a shelf in a bakery in a half-hour period is 12. What is the probability that the number of loaves put on the shelf in 5 minutes is 3? Of interest is the number of loaves of bread put on the shelf in 5 minutes. The time interval of interest is 5 minutes.

Let X = the number of loaves of bread put on the shelf in 5 minutes. If the average number of loaves put on the shelf in 30 minutes (half-hour) is 12, **then the average number of loaves put on the shelf in 5 minutes is**

$\left(\frac{5}{30}\right) \cdot 12 = 2$ loaves of bread

[8]This content is available online at <http://http://cnx.org/content/m16829/1.12/>.

The probability question asks you to find $P(X = 3)$.

Example 4.21
A certain bank expects to receive 6 bad checks per day. What is the probability of the bank getting fewer than 5 bad checks on any given day? Of interest is the number of checks the bank receives in 1 day, so the time interval of interest is 1 day. Let X = the number of bad checks the bank receives in one day. If the bank expects to receive 6 bad checks per day then the average is 6 checks per day. The probability question asks for $P(X < 5)$.

Example 4.22
Your math instructor expects you to complete 2 pages of written math homework every day. What is the probability that you complete more than 2 pages a day?

This is a Poisson problem because your instructor is interested in knowing the number of pages of written math homework you complete in a day.

Problem 1 ***(Solution on p. 206.)***
What is the interval of interest?

Problem 2 ***(Solution on p. 206.)***
What is the average number of pages you should do in one day?

Problem 3 ***(Solution on p. 206.)***
Let X = _____________. What values does X take on?

Problem 4 ***(Solution on p. 206.)***
The probability question is P(______).

4.8.1 Notation for the Poisson: P = Poisson Probability Distribution Function

$X \sim$ P(μ)

Read this as "X is a random variable with a Poisson distribution." The parameter is μ (or λ). μ (or λ) = the mean for the interval of interest.

Example 4.23
Leah's answering machine receives about 6 telephone calls between 8 a.m. and 10 a.m. What is the probability that Leah receives more than 1 call **in the next 15 minutes?**

Let X = the number of calls Leah receives in 15 minutes. (The **interval of interest** is 15 minutes or $\frac{1}{4}$ hour.)

X takes on the values 0, 1, 2, 3, ...

If Leah receives, on the average, 6 telephone calls in 2 hours, and there are eight 15 minutes intervals in 2 hours, then Leah receives

$\frac{1}{8} \cdot 6 = 0.75$

calls in 15 minutes, on the average. So, μ = 0.75 for this problem.

$X \sim$ P(0.75)

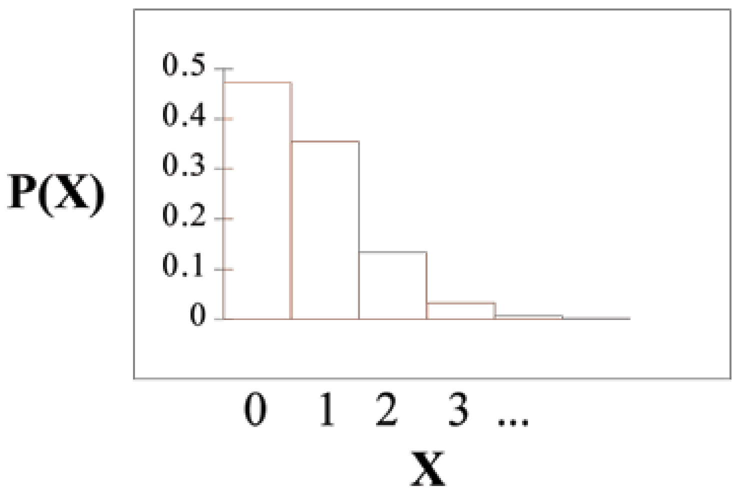
Find $P(X > 1)$. $P(X > 1) = 0.1734$ (calculator or computer)

TI-83+ and TI-84: For a general discussion, see **this example (Binomial)**. The syntax is similar. The Poisson parameter list is (μ for the interval of interest, number). **For this problem:**

Press 1- and then press 2nd DISTR. Arrow down to C:poissoncdf. Press ENTER. Enter .75,1). The result is $P(X > 1) = 0.1734$**. NOTE: The TI calculators use** λ **(lambda) for the mean.**

The probability that Leah receives more than 1 telephone call in the next fifteen minutes is about 0.1734.

The graph of $X \sim$ P(0.75) is:



The y-axis contains the probability of X where X = the number of calls in 15 minutes.

4.9 Summary of Functions[9]

Formula 4.1: Binomial
$X \sim B(n, p)$

X = the number of successes in n independent trials

n = the number of independent trials

X takes on the values $x = 0,1, 2, 3, ...,n$

p = the probability of a success for any trial

q = the probability of a failure for any trial

$p + q = 1 \quad q = 1 - p$

The mean is $\mu = np$. The standard deviation is $\sigma = \sqrt{npq}$.
Formula 4.2: Geometric
$X \sim G(p)$

X = the number of independent trials until the first success (count the failures and the first success)

X takes on the values x= 1, 2, 3, ...

p = the probability of a success for any trial

q = the probability of a failure for any trial

$p + q = 1$

$q = 1 - p$

The mean is $\mu = \frac{1}{p}$

The standard deviation is $\sigma = \sqrt{\frac{1}{p}\left(\left(\frac{1}{p}\right) - 1\right)}$
Formula 4.3: Hypergeometric
$X \sim H(r, b, n)$

X = the number of items from the group of interest that are in the chosen sample.

X may take on the values x= 0, 1, ..., up to the size of the group of interest. (The minimum value for X may be larger than 0 in some instances.)

r = the size of the group of interest (first group)

b= the size of the second group

n= the size of the chosen sample.

$n \leq r + b$

The mean is: $\mu = \frac{nr}{r+b}$

[9]This content is available online at <http://http://cnx.org/content/m16833/1.8/>.

The standard deviation is: $\sigma = \sqrt{\frac{rbn(r+b+n)}{(r+b)^2(r+b-1)}}$

Formula 4.4: Poisson
$X \sim \mathrm{P}(\mu)$

X = the number of occurrences in the interval of interest

X takes on the values x = 0, 1, 2, 3, ...

The mean μ is typically given. (λ is often used as the mean instead of μ.) When the Poisson is used to approximate the binomial, we use the binomial mean $\mu = np$. n is the binomial number of trials. p = the probability of a success for each trial. This formula is valid when n is "large" and p "small" (a general rule is that n should be greater than or equal to 20 and p should be less than or equal to 0.05). If n is large enough and p is small enough then the Poisson approximates the binomial very well. The standard deviation is $\sigma = \mu$.