Discrete Random Variables: Binomial*

Susan Dean
Barbara Illowsky, Ph.D.

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

Abstract

This module describes the characteristics of a binomial experiment and the binomial probability distribution function.

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 help in predicting the outcome of another 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 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 $\mu = \text{np}$ and $\sigma^2 = \text{npq}$. The standard deviation, σ, is then $\sigma = \sqrt{\text{npq}}$.

Any experiment that has characteristics 2 and 3 and where n = 1 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 1

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 randomly selected elementary physics class.

*Version 1.16: May 27, 2012 3:52 am -0500
†http://creativecommons.org/licenses/by/3.0/

Example 2

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%. Each game you play is independent. 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 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 3

A fair coin is flipped 15 times. Each flip is independent. 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 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

Approximately 70% of statistics students do their homework in time for it to be collected and graded. Each student does homework independently. In a statistics class of 50 students, what is the probability that at least 40 will do their homework on time? Students are selected randomly.

Problem 1 ***(Solution on p. 4.)***

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. 4.)***

If we are interested in the number of students who do their homework, then how do we define X?

Problem 3 ***(Solution on p. 4.)***

What values does x take on?

Problem 4 ***(Solution on p. 4.)***

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. 4.)***

If $p + q = 1$, then what is q?

Problem 6 ***(Solution on p. 4.)***

The words "at least" translate as what kind of inequality for the probability question $P(x$____$40)$.

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

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 5

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.

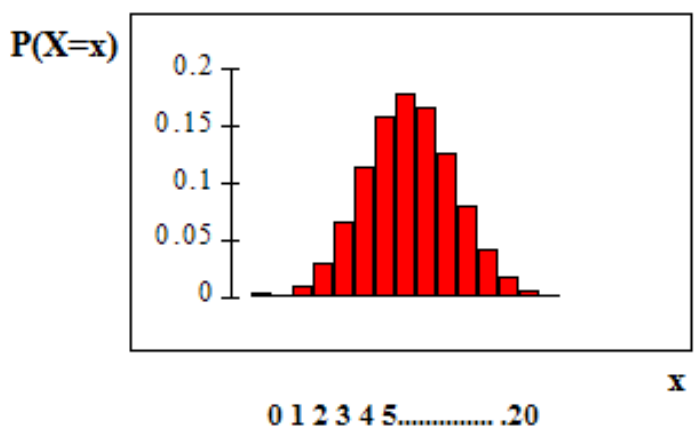
To calculate $P(x \leq$ value): binomcdf(n, p, number) If "number" is left out, the result is the cumulative binomial probability table.

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, $\mu = \text{np} = (20)(0.41) = 8.2$.

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

Example 6

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.

Solutions to Exercises in this Module

Solution to Example 4, Problem 1 (p. 2)
failure
Solution to Example 4, Problem 2 (p. 2)
X = the number of statistics students who do their homework on time
Solution to Example 4, Problem 3 (p. 2)
0, 1, 2, ..., 50
Solution to Example 4, Problem 4 (p. 2)
Failure is a student who does not do his or her homework on time.
Solution to Example 4, Problem 5 (p. 2)
$q = 0.30$
Solution to Example 4, Problem 6 (p. 2)
greater than or equal to ($\geq$)

Glossary

Definition 1: Bernoulli Trials
An experiment with the following characteristics:

- There are only 2 possible outcomes called "success" and "failure" for each trial.
- The probability p of a success is the same for any trial (so the probability $q = 1 - p$ of a failure is the same for any trial).

Definition 2: Binomial Distribution
A discrete random variable (RV) which arises from Bernoulli trials. There are a fixed number, n, of independent trials. "Independent" means that the result of any trial (for example, trial 1) does not affect the results of the following trials, and all trials are conducted under the same conditions. Under these circumstances the binomial RV X is defined as the number of successes in n trials. The notation is: $X \sim B(n, p)$. The mean is $\mu = \text{np}$ and the standard deviation is $\sigma = \sqrt{\text{npq}}$. The probability of exactly x successes in n trials is $P(X = x) = \binom{n}{x} p^x q^{n-x}$.