**The binomial distribution**

Corrections made in the corresponding video have been incorporated into these notes.

There are a numbered of “named” probability distributions for discrete random variables. One of the most useful is the binomial distribution. To help explain the binomial distribution and its characteristics, below is an example.

Example: Field goal kicking

Suppose a field goal kicker attempts 4 field goals during a game and each field goal has the same probability of being successful (the kick is made). Also, assume each field goal is attempted under similar conditions; i.e., distance, weather, surface, … .

Below are the characteristics that must be satisfied in order for the binomial distribution to be used.

1. There are n trials for each experiment.

n = 4 field goals attempted

1. Two possible outcomes of a trial. These are typically referred to as a success or failure.

Each field goal can be made (success) or missed (failure)

1. The trials are independent of each other.

The result of one field goal does not affect the result of another field goal.

1. The probability of success, denoted by π, remains constant for each trial. The probability of a failure is 1-π.

Suppose the probability a field goal is good is 0.6; i.e., P(success) = π = 0.6.

1. The random variable, Y, represents the number of successes.

Let Y = number of field goals that are successful. Thus, Y can be 0, 1, 2, 3, or 4.

Since these 5 items are satisfied, the binomial distribution can be used!

What is P(0 of 4 are successful) = P(Y = 0)?

Let G = Field goal is good (success) and M = Field goal is missed (failure)

P(Y = 0)

= P(1st M ∩ 2nd M ∩ 3rd M ∩ 4th M)

= P(1st M)×P(2nd M) ×P(3rd M)×P(4th M) because ind.



= P(M)×P(M)×P(M)×P(M) each trial has same prob.

= (1 – π)4

= 0.44

= 0.0256

What is P(1 good) = P(Y = 1)?

P(Y = 1)

= P(1st G ∩ 2nd M ∩ 3rd M ∩ 4th M) +

P(1st M ∩ 2nd G ∩ 3rd M ∩ 4th M) +

P(1st M ∩ 2nd M ∩ 3rd G ∩ 4th M) +

P(1st M ∩ 2nd M ∩ 3rd M ∩ 4th G)

= P(G)×P(M)×P(M)×P(M) + P(M)×P(G)×P(M)×P(M)

+ P(M)×P(M)×P(G)×P(M) + P(M)×P(M)×P(M)×P(G)

= (0.6)(0.4)(0.4)(0.4) + … + (0.4)(0.4)(0.4)(0.6)

= 4(0.6)(0.4)(0.4)(0.4)

= 4(0.6)1(0.4)3

= 0.1536

Note: P(Y = 1) = 4(0.6)1(0.4)3

* 1 success, with probability of 0.6
* 3 failures, with probability of 0.4
* 4 different ways for 1 success and 3 failures to happen.

Continuing this same process, the probability distribution can be found to be:

|  |  |
| --- | --- |
| **y** | **P(Y = y)** |
| 0 | 0.0256 |
| 1 | 0.1536 |
| 2 | 6(0.6)2(0.4)2=0.3456 |
| 3 | 4(0.6)3(0.4)1=0.3456 |
| 4 | 1(0.6)4(0.4)0=0.1296 |

In general, the equation for the binomial distribution function (binomial probability mass function) is defined as

 for y = 0, 1, 2, …, n

Notes:

* : This gives the number of unique *combinations* of ways to choose “y” items from “n” items. For this case, we are choosing y successes out of n trails which result in a success or failure. Often, it is read as “n choose y”. More information on combinations is given at the end of the binomial distribution section.
* Remember that n! = n×(n-1)×(n-2)×…×2×1.
* π is a population parameter. We will learn later how to estimate it using a sample from the population. We will also learn how to estimate it with a specific level of confidence!
* The dbinom() function can be used to find these probabilities in R.

When the number of trials is 1 (i.e., n = 1), the binomial distribution simplifies to

 for y = 0, 1

because  is 1 for y = 0 or 1. This special case of the binomial called a Bernoulli distribution. Also, suppose Y1, Y2, …, Yn are independent random variables with a Bernoulli probability distribution. Then  has a binomial distribution of



Because of this relationship, one often refers to each “trial” for a binomial setting as a “Bernoulli trial”.

Example: Field goal kicking (FG.R)

> #y = 2, n = 4, pi = 0.6

> dbinom(x = 2, size = 4, prob = 0.6)

[1] 0.3456

> #n = 4, pi = 0.6, y = 0, 1, 2, 3, 4

> dbinom(x = 0:4, size = 4, prob = 0.6)

[1] 0.0256 0.1536 0.3456 0.3456 0.1296

> #Nice display

> data.frame(y = 0:4, Prob.y = dbinom(x = 0:4, size

= 4, prob = 0.6))

y Prob.y

1 0 0.0256

2 1 0.1536

3 2 0.3456

4 3 0.3456

5 4 0.1296

> #Plots

> n <- 4

> y <- 0:n

> pi <- 0.6

> plot(x = y, y = dbinom(x = y, size = n, prob =

pi), type = "h", xlab = "y", ylab = "P(y)",

main = paste("Plot of a binomial distribution

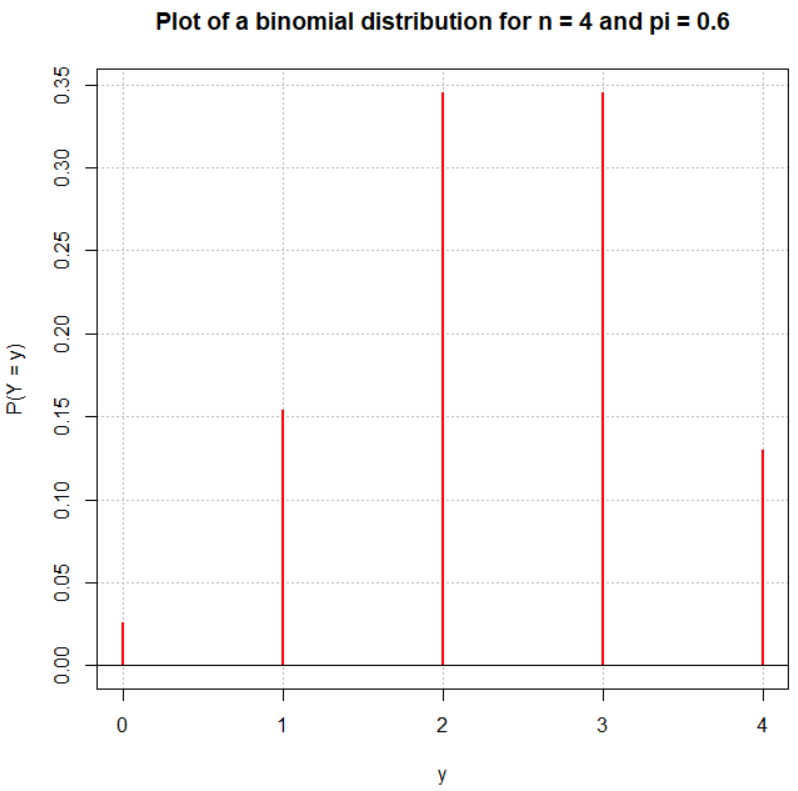
for n =", n, "and pi =", pi), panel.first =

grid(col="gray", lty="dotted"), lwd = 2, col =

"red", ylim = c(0, max(dbinom(x = y, size = n,

prob = pi))))

> abline(h = 0)



What would we expect the number of successes to be on average?

This can be found by using an “expected value”. An expected value of a quantity simply denotes what we would expect on average the quantity to be. In general for any discrete random variable, the expected value of Y is



This is equivalent to the population mean that we first saw earlier in the course! Compare this to the formula we used there. Note that books often use μ to be E(Y).

Without going into a mathematical proof, the expected value of Y for the binomial setting here is μ = nπ. Why does this value make some intuitive sense?

How much variability would we expect to see for the number of successes from one set of n trials to another?

There is another expected value that is often of interest. This expected value measures the squared deviation of possible values of Y from its expected value. The expected value of (Y – μ)2 is





This is equivalent to the population variance that we first saw in earlier in the course! Compare this to the formula we used there. Note that books often use σ2 as the symbol for . Also, note that most books also use  for  as well.

Without going into a mathematical proof, the expected value of y for the binomial setting here is σ2 = nπ(1 – π).

Question: How could we use the “rule of thumb” from Chapter 2 with μ and σ2 here?

Example: FG kicking (FG.R)

Remember that n = 4 and π = 0.6

Find the mean and variance.

μ = 4×0.6 = 2.4

Var(Y) = σ2 = 4×0.6×(1-0.6) = 0.96

If one used the long way to find μ,

| **y** | **P(Y=y)** | **y×P(Y=y)** |
| --- | --- | --- |
| 0 | 0.0256 | 0 |
| 1 | 0.1536 | 0.1536 |
| 2 | 0.3456 | 0.6912 |
| 3 | 0.3456 | 1.0368 |
| 4 | 0.1296 | 0.5184 |

μ =  = 0 + 0.1536 + 0.6912 + 1.0368 + 0.5184   
 = 2.4

μ ± 2σ = 2.4 ± 2× = (0.44, 4.36)

Suppose a field goal kicker made 0 out of 4 field goals during a game. Based on the information above, what could we conclude about the kicker?

He had a very unusual game. Alternatively, perhaps π = 0.6 is too high. Maybe we could use this information to test a hypothesis about the true value of π.

Example: Binomial distribution plots (binomial\_plots.R)

See the program for code. Note that P(Y = y) is displayed on the y-axis despite the label being omitted.









Examine the following with these plots:

* When are the plots “symmetric” and when are the plots “skewed”?
* Where is the largest probability?
* Notice the μ ± 2σ lines.
* We will learn later in this course how to estimate π using a sample from a population. Given the results of these plots, why do you think it is important to estimate it with a sample instead of just setting it to a particular value of choice?

Final notes:

* The multinomial probability distribution is an extension of the binomial distribution to the case of more than two possible categories of outcomes. This distribution is discussed later in this course.
* One limitation of the binomial distribution is that π remains constant for each trial. How could we remove this limitation? For example, the probability of success for a field goal is highly dependent of its distance. How could we incorporate distance into a binomial distribution setting?

More about “combinations”

The number of combinations of n distinct objects take r at a time is



Note that the order in selecting the objects (items) is not important. Often  is read as “n choose r”.

Example: How many ways can TWO of the letters a, b, and c be chosen from the three?

First, it is instructive to answer the question, “How many ways can two of the letters a, b, and c be arranged?”

|  | **Letter 1** | **Letter 2** |
| --- | --- | --- |
| 1 | a | b |
| 2 | a | c |
| 3 | b | a |
| 4 | b | c |
| 5 | c | a |
| 6 | c | b |

To answer the original question of “How many ways can two of the letters a, b, and c be chosen from the three?” there is no longer a distinction between cases like (a,b) and (b,a). Thus, order is no longer important. Then,

|  |  |  |
| --- | --- | --- |
|  | **Letter 1** | **Letter 2** |
| 1 | a | b |
| 2 | a | c |
| 3 | b | a |
| 4 | b | c |
| 5 | c | a |
| 6 | c | b |

only (a,b), (a,c), and (b,c) remain. We could also calculate .

Example: How many different number combinations are there in the Pick 5 game of a lottery (5 numbers 1 through 38 are picked)?

|  | **#1** | **#2** | **#3** | **#4** | **#5** |
| --- | --- | --- | --- | --- | --- |
| 1 | 1 | 2 | 3 | 4 | 5 |
| 2 | 1 | 2 | 3 | 4 | 6 |
|  |  |  |  |  |  |
| 501,942 | 34 | 35 | 36 | 37 | 38 |

 = 501,942