**Probability distributions for discrete random variables**

Diagram from earlier:



We discussed the basics of probability so that we can talk about probabilities more formally for the remainder of this course. This will involve examining how we can use “probability distributions”. One can think of probability distributions as population quantities since they summarize possible values that a random variable can take on.

Example: Fifty numbers from 0 to 9 (50numbers.R)

Suppose I draw 50 numbers at random from 0, 1, …, 9 with replacement and each number has an equal chance of being drawn.

One could think of this as a large population of people and each individual indicates the number of computer-like devices that they regularly use, where 10% have 0, 10% have 1, ...

Below are the results.

> y <- 0:9

> set.seed(9823)

> set1 <- sample(x = y, size = 50, replace = TRUE)

> head(set1)

[1] 1 0 5 2 9 8

> # Used specific breaks because of problems with

 hist() when 0 is part of a discrete random

 variable

> hist(x = set1, main = "50 numbers", xlab =

 "value", breaks = -1:9)

> freq.dist(data = set1)

 class Frequency Rel.Frequency

1 0 5 0.10

2 1 6 0.12

2 2 5 0.10

3 3 3 0.06

4 4 4 0.08

5 5 4 0.08

6 6 7 0.14

7 7 6 0.12

8 8 2 0.04

9 9 8 0.16



One would expect to see 10% 0’s, 10% 1’s, …, 10% 9’s because each has an equal probability of being drawn and there are 10 numbers. If this experiment were repeated over and over again, we would see the “sample” percentages become very close to what we expect.

Let Y = the number selected on a draw from 0, 1, …, 9.

Y is called a random variable because it can change from draw to draw in a random manner which is controlled by a probability structure.

We know before the experiment of drawing numbers that Y can be any number from 0, 1, …, 9, and we know the percentage of draws (probability) we *expect* Y to be any of these numbers. Thus, we can talk about a probability distribution for Y before the experiment. Note that this is for the population!

|  |  |
| --- | --- |
| **y** | **P(Y = y)** |
| 0 | P(Y = 0) = 0.1 |
| 1 | P(Y = 1) = 0.1 |
| 2 | 0.1 |
| 3 | 0.1 |
| 4 | 0.1 |
| 5 | 0.1 |
| 6 | 0.1 |
| 7 | 0.1 |
| 8 | 0.1 |
| 9 | 0.1 |

Why is there a lowercase and uppercase Y used here? In most of statistics, we formally use

* Uppercase to indicate the random variable
* Lowercase to indicate the random variable’s observed values (see the previous content on summarizing data)

Thus, a person can have Y computer-like devices. A person could be observed to own y devices.

This distinction can be difficult for students learning statistics in a formal way for the first time. What makes it more difficult is many introductory textbooks on statistics, like the commonly used Ott and Longnecker’s book, will simply just use lowercase or uppercase for both meanings in order to make it “easier” on students.

Notes:

* Why is the term “probability distribution” used? It shows how the probabilities are distributed for possible values of Y
* 
* Values of y not listed above have a probability of 0; for example, P(Y = 4.2) = 0.
* Remember that P(Y = 7) is what we expect to happen if the experiment is repeated an infinite number of times. In our sample, the percentage of time a 7 occurred was 0.04.
* The random variable used in this example is called a discrete random variable since there are a finite number of values that it can take on – 0, 1, …, 9. The more general definition of a discrete random variable is if the set of possible values for y is “countable”.

Side note: Countable corresponds to a set which is “finite” or “countably infinite”. When there are a finite number of values for Y (i.e., you can count all possible values), the random variable is discrete (as in this example). A random variable is also called “discrete” if the set of possible values of Y is “countably infinite”. There will be more on this at the later in the course.

* We will discuss continuous random variables shortly where there are an infinite number of values a random variable can take on within a particular region.
* The value “observed” from the first draw was 1. The value “observed” from the second draw was 0. All of these numbers constitute a sample of size 50 from a population which has the specified probability distribution.
* In the sample, the percentage of times 0, 1, …, 9 were observed are somewhat similar to the probabilities in the probability distribution. If the sample size was larger, say 5,000, we would expect these percentages to be much closer to the probability distribution. Below are the results when this is actually done:

> set.seed(9823)

> set2 <- sample(x = x, size = 5000, replace =

 TRUE)

> head(set2)

[1] 1 0 5 2 9 8

> hist(x = set2, main = "5,000 numbers", xlab =

 "value", breaks = -1:9)

> freq.dist(data = set2, numb.breaks = -1:9)

 class Frequency Rel.Frequency

1 0 494 0.10

2 1 501 0.10

3 2 500 0.10

4 3 475 0.10

5 4 477 0.10

6 5 485 0.10

7 6 540 0.11

8 7 494 0.10

9 8 521 0.10

10 9 513 0.10



Questions:

* Suppose another sample of size 50 is taken. Would you expect the same observed values for the sample to be found?
* Suppose the probability distribution of

| **y** | **P(Y = y)** |
| --- | --- |
| 0 | 0.01 |
| 1 | 0.03 |
| 2 | 0.05 |
| 3 | 0.07 |
| 4 | 0.09 |
| 5 | 0.11 |
| 6 | 0.13 |
| 7 | 0.15 |
| 8 | 0.17 |
| 9 | 0.19 |

is used instead. What do you expect would happen with the sample?

Characteristics of a probability distribution for a discrete random variable:

* P(Y = y) ≥ 0 – all probabilities are ≥ to zero
* P(Y = y) ≤ 1 – all probabilities are ≤ to one
*  – sum up all of the probabilities and get 1

A probability distribution for a discrete random variable is often referred to as a probability mass function (PMF).

There is another notation that is used with these probabilities! Rather than use P( ), function notation is used instead:

P(Y = y) = f(y)

Thus, f(y) is the probability for a specific value of Y; i.e., the probability is a *function* of y. Providing f(y) for all possible values Y gives the probability distribution function (PDF).

Relative to our original example in this section, the probability distribution is

|  |  |
| --- | --- |
| **y** | **f(y)** |
| 0 | f(0) = 0.1 |
| 1 | f(1) = 0.1 |
| 2 | 0.1 |
| 3 | 0.1 |
| 4 | 0.1 |
| 5 | 0.1 |
| 6 | 0.1 |
| 7 | 0.1 |
| 8 | 0.1 |
| 9 | 0.1 |

Note that it is **INCORRECT** to write f(Y), f(Y = 0), or f(y = 0).

A useful function that we will use with probability distributions is a cumulative distribution function (CDF). This function for a discrete random variable Y, denoted by F(y), is

F(y) = P(Y ≤ y) = 

Notice how F(y) “cumulates” probabilities as y increases and how the function ranges from 0 to 1.

Example: Fifty numbers from 0 to 9 (50numbers.R)

| **y** |  **P(Y = y) = f(y)** | **F(y)** |
| --- | --- | --- |
| 0 | 0.1 | F(0) = 0.1 |
| 1 | 0.1 |  F(1) = 0.1+0.1 = 0.2  |
| 2 | 0.1 | 0.3 |
| 3 | 0.1 | 0.4 |
| 4 | 0.1 | 0.5 |
| 5 | 0.1 | 0.6 |
| 6 | 0.1 | 0.7 |
| 7 | 0.1 | 0.8 |
| 8 | 0.1 | 0.9 |
| 9 | 0.1 | 1 |

Questions:

* What is F(0)?
* What is F(1.5)?
* What is F(1.99999)?
* What is F(2)?
* What is F(-1)?

Plot of the CDF

> y <- 0:9

> plot.ecdf(x = y, verticals = TRUE, do.p = FALSE,

 panel.first = grid(), col = "red", lwd = 2, main =

 "CDF plot", ylab = "F(y)", xlab = "y")



Notes:

* CDFs for discrete random variables are often referred to as “step functions” because their plots look like stairs.
* Once y = 9 is reached, F(y) should be equal to 1 across the remainder of the plot. This is not shown as well as I would like on the plot.
* To create this particular plot, I am using a function that plots the “estimated” CDF. More on this later in the course.

Example: Let’s play Plinko! (plinko.xlsx)





Video at <https://youtu.be/naUppHrHJpI?t=262>.

From Haws (1995): The contestant drops a circular disk down a board with nails arranged in the pattern indicated in Figure 1. Assume also that the disk is equally likely to go to the left or to the right at each nail it encounters. The contestant wins the amount of cash indicated on the reservoir in which the disk lands.



Let X be a random variable denoting the amount won for one chip.[[1]](#footnote-1) Below are the probability distributions that Haws (1995) finds for winning a specific dollar amount for one Plinko chip.

|  |  |
| --- | --- |
|  | **Drop Plinko chip in column above:** |
|  | **$100** | **$500** | **$1,000** | **$0** | **$10,000** |
| **x** | **f(x)** | **f(x)** | **f(x)** | **f(x)** | **f(x)** |
| **$100** | 0.1667 | 0.1339 | 0.0822 | 0.0359 | 0.0176 |
| **$500** | 0.3571 | 0.3080 | 0.2204 | 0.1287 | 0.0882 |
| **$1,000** | 0.2976 | 0.2991 | 0.2862 | 0.2545 | 0.2353 |
| **$0** | 0.1429 | 0.1920 | 0.2796 | 0.3713 | 0.4118 |
| **$10,000** | 0.0357 | 0.0670 | 0.1316 | 0.2096 | 0.2471 |

For example, the probability of winning $10,000 when the chip is dropped above one of the $100 slots is f(10000) =0.0357.

Questions:

* If you wanted to maximize your chance to win $10,000, where should you drop the chip?
* If you wanted to maximize your chance to win >$0, where should you drop the chip?
* For Price is Right fans: In computing these probability distributions, Haws assumes that when a chip hits a peg (nail) on the Plinko board, it could only go to the opening to the left or right of that peg. Is this a correct assumption?

Below is the observed data for all Price is Right shows that appeared during one year.

|  |  |
| --- | --- |
|  | **Drop Plinko chip in column above:** |
|  | **$100** | **$500** | **$1,000** | **$0** | **$10,000** |
| **x** | **Count** | **Count** | **Count** | **Count** | **Count** |
| **$100** | 0 | 3 | 2 | 5 | 2 |
| **$500** | 0 | 3 | 12 | 5 | 6 |
| **$1,000** | 1 | 3 | 12 | 19 | 11 |
| **$0** | 0 | 2 | 9 | 21 | 8 |
| **$10,000** | 0 | 1 | 4 | 9 | 4 |

|  |  |
| --- | --- |
|  | **Drop Plinko chip in column above:** |
|  | **$100**  | **$500**  | **$1,000**  | **$0**  | **$10,000**  |
| **x** | **Proportion** | **Proportion** | **Proportion** | **Proportion** | **Proportion** |
| **$100**  | 0.00 | 0.25 | 0.05 | 0.08 | 0.06 |
| **$500**  | 0.00 | 0.25 | 0.31 | 0.08 | 0.19 |
| **$1,000**  | 1.00 | 0.25 | 0.31 | 0.32 | 0.35 |
| **$0**  | 0.00 | 0.17 | 0.23 | 0.36 | 0.26 |
| **$10,000**  | 0.00 | 0.08 | 0.10 | 0.15 | 0.13 |

Why are the observed probabilities different from the true probabilities?

Example: Field goal kicking (FieldGoal-Intro.R)

Let Y be a random variable denoting the number of successful field goals out of 5 attempts where each attempt had the exact same probability of success. The probability distribution for Y is



for y = 0, 1, 2, 3, 4, 5 and f(y) = 0 otherwise

This is actually a special type of probability distribution called a binomial probability distribution. Characteristics about the binomial will be discussed later in the course.

Below are the PDF and CDF written out:

| **y** | **f(y)** | **F(y)** |
| --- | --- | --- |
| 0 | f(0) = (0.6)0(0.4)5 = 0.0102 | F(0) = 0.0102 |
| 1 | f(1) = 5(0.6)1(0.4)4 = 0.0768 | 0.0102 + 0.0768 = 0.0870 |
| 2 | 0.2304 | 0.3174 |
| 3 | 0.3456 | 0.6630 |
| 4 | 0.2592 | 0.9222 |
| 5 | 0.0776 | 1 |

Example calculations from R

> factorial(5)

[1] 120

> factorial(5)/(factorial(2)\*factorial(3))

[1] 10

> #f(2)

> factorial(5)/(factorial(2)\*factorial(3)) \* 0.6^2 \*

 0.4^3

[1] 0.2304

Below is a plot of the PDF.

> y <- c(0,1,2,3,4,5)

> fy <- c(0.0102, 0.0768, 0.2304, 0.3456, 0.2592, 0.0776)

> sum(fy) # Not equal to 1 due to rounding

[1] 0.9998

> plot(x = y, y = fy, xlab = "y", ylab = "f(y)", col =

 "red", type = "h")

> abline(h = 0)



Another way to draw the PDF is as follows:



f(y)

y

This plot was made outside of R, and you do not need to reproduce it. Each bar is specially drawn so that it has a width of 1. This particular type of plot is sometimes called a probability histogram. Because of the way it is drawn, we can get probabilities from the areas of the bars. Thus,

f(2) = P(Y=2) = area of the rectangle for Y = 2

= width × height

= 1 × 0.2304

= 0.2304

Also, P(2 ≤ Y ≤ 3) = 1 × 0.2304 + 1 × 0.3456 = 0.5760.

Therefore, probabilities correspond to areas of the bars on the plot. This type of thinking will be very important soon in our course!

Question: Is it likely that not all of the 5 field goals will be made? Explain.

Example: Insurance example (Insurance.R, insurance.ipynb)

Suppose probabilities for the number of injury claims per month for a company can be modeled by a random variable X with

f(x) =  for x = 0, 1, 2, … .

as the PDF. There is not a set number of finite values that X can take on here! However, X is still a discrete random variable. Why?

The set of X values is countably infinite. In this case, there is a “bijective mapping” between the set of possible X values and the set of natural numbers.

This is a PDF because

1. f(x) ≥ 0;  > 0 for all possible x’s here (x’s are all ≥ 0 meaning the denominator is > 0).
2. f(x) ≤ 1;  < 1 for all possible x’s here. Notice that the maximum value occurs when x = 0 (f(x) = ½).
3. ;

 by the use of partial fractions. Then





 where a = x – 1

= 1

Sage (to be discussed later):



R (not an exact answer):

> x <- 0:100

> fx <- 1/((x+1)\*(x+2))

> Fx <- cumsum(fx)

> set1 <- data.frame(x, fx, Fx)

> head(set1)

 x fx Fx

1 0 0.50000000 0.5000000

2 1 0.16666667 0.6666667

3 2 0.08333333 0.7500000

4 3 0.05000000 0.8000000

5 4 0.03333333 0.8333333

6 5 0.02380952 0.8571429

> tail(set1)

 x fx Fx

96 95 1.073883e-04 0.9896907

97 96 1.051967e-04 0.9897959

98 97 1.030715e-04 0.9898990

99 98 1.010101e-04 0.9900000

100 99 9.900990e-05 0.9900990

101 100 9.706853e-05 0.9901961

What is the probability of more than 1 claim during the month?

Using the PDF,

P(X > 1) = P(X ≥ 2)

= 1 – P(X < 2)

= 1 – P(X ≤ 1)

= 1 – P(X = 0) – P(X = 1)

= 1 – f(0) – f(1)
= 1- 1/2 - 1/6 = 1/3

Using the CDF,

P(X > 1) = 1 – P(X ≤ 1) = 1 – F(1) = 1 – 2/3 = 1/3

> # P(X > 1) = P(X >= 2)

> 1 - (fx[1] + fx[2])

[1] 0.3333333

> 1 - Fx[2]

[1] 0.3333333

Would having 20 or more claims during a month seem unusual? Why?

P(X ≥ 20) = 1- P(X ≤ 19) = 1 – F(19) = 1 – 0.9524 = 0.0476.

> # P(X >= 20)

> set1[19:22,]

 x fx Fx

19 18 0.002631579 0.9500000

20 19 0.002380952 0.9523810

21 20 0.002164502 0.9545455

22 21 0.001976285 0.9565217

> 1 - Fx[20]

[1] 0.04761905

Yes, this is somewhat unusual since we would expect this to happen only about 5% of the time. Looking at items that are “unusual” relative to probabilities will be very useful later in the course.

This example was based upon Society of Actuaries exam question.

Final note: There are many special probability distributions for discrete random variables that have been shown to work well in practice. These will be discussed in soon in our course!

1. Because changes in the amount won have occurred over time and will likely occur in the future, I decided to use these particular amounts. [↑](#footnote-ref-1)