**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.

> x <- 0:9

> set.seed(9823)

> set1 <- sample(x = x, 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).