STA 2023

Module 3B
Discrete Random Variables

Learning Objectives

Upon completing this module, you should be able to

1. Determine the probability distribution of a discrete random variable.
2. Construct a probability histogram.
3. Describe events using random-variable notation, when appropriate.
4. Use the frequentist interpretation of probability to understand the meaning of probability distribution of a random variable.
5. Find and interpret the mean and standard deviation of a discrete random variable.

What is a Random Variable?

Random Variable

A **random variable** is a quantitative variable whose value depends on chance.

Discrete Random Variable

A **discrete random variable** is a random variable whose possible values can be listed.

What does it mean?

A discrete random variable usually involves a count of something.

What is a Probability Distribution?

Siblings x	Probability $P(X = x)$
0	0.200
1	0.425
2	0.275
3	0.075
4	0.025
	1.000

This is an example of a probability distribution of a discrete random variable X, where x represents the possible number of siblings of a randomly selected student.

Note the difference between X and x.

Probability Distribution and Probability Histogram

Probability Distribution and Probability Histogram

Probability distribution: A listing of the possible values and corresponding probabilities of a discrete random variable, or a formula for the probabilities.

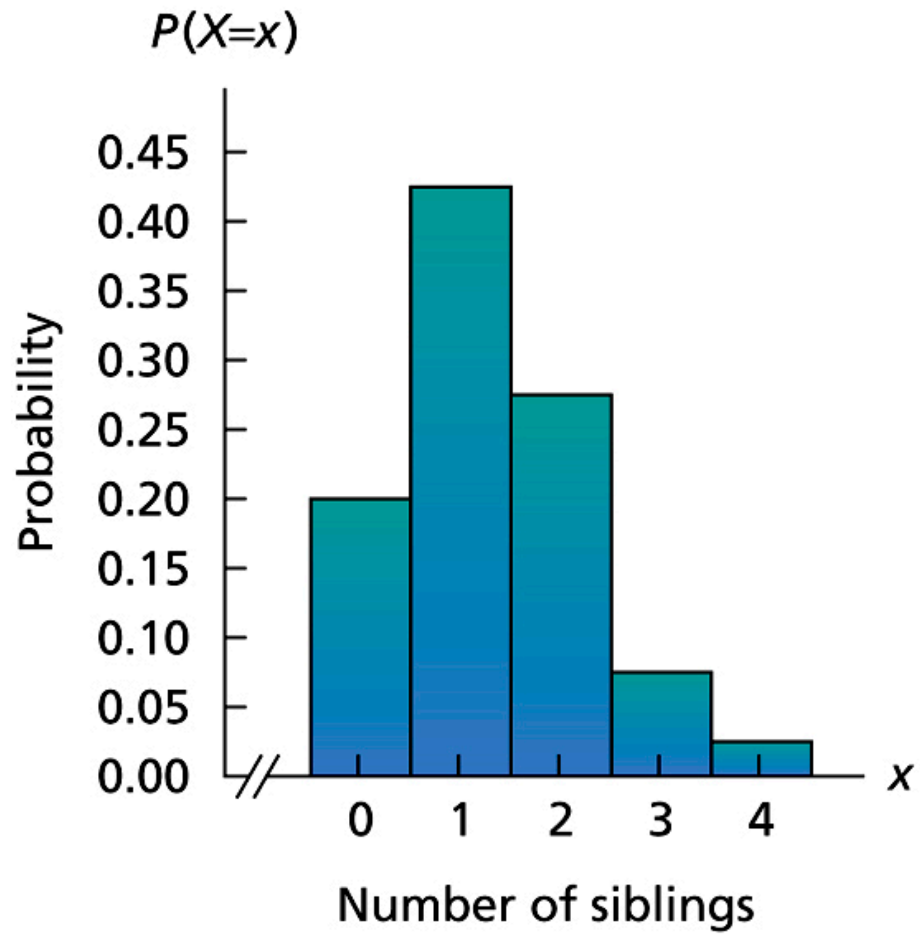
Probability histogram: A graph of the probability distribution that displays the possible values of a discrete random variable on the horizontal axis and the probabilities of those values on the vertical axis. The probability of each value is represented by a vertical bar whose height equals the probability.

The probability distribution and probability histogram of a discrete random variable show its possible values and their likelihood.

Probability Distribution and Probability Histogram (Cont.)

Siblings x	Probability $P(X = x)$
0	0.200
1	0.425
2	0.275
3	0.075
4	0.025
	1.000

Can you tell what is represented on the *y*-axis of the probability histogram?



Probability of an Event

Siblings x	Probability $P(X = x)$
0	0.200
1	0.425
2	0.275
3	0.075
4	0.025
	1.000

- The probability of the event that the students have exactly two siblings (when x is 2) can be represented by $P(X = 2)$. In this case, $P(X = 2) = 0.275$.

- The probability of the event that the students have exactly three siblings (when x is 3) can be represented by $P(X = 3)$. In this case, $P(X = 3) = 0.075$.

Probability of an Event (Cont.)

Siblings x	Probability $P(X = x)$
0	0.200
1	0.425
2	0.275
3	0.075
4	0.025
	1.000

The probability of the event that the students have less than three siblings (when x is 3) can be represented by $P(X < 3)$.

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= 0.200 + 0.425 + 0.275 = 0.900$$

The probability of the event that the students have less than two siblings (when x is 2) can be represented by $P(X < 2)$.

$$P(X < 2) = P(X = 0) + P(X = 1)$$
$$= 0.200 + 0.425 = 0.625$$

Probability of an Event (Cont.)

Siblings x	Probability $P(X = x)$
0	0.200
1	0.425
2	0.275
3	0.075
4	0.025
	1.000

The probability of the event that the students have less than or equal to two siblings (when x is 2) can be represented by $P(X \le 2)$.

$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 0.200 + 0.425 + 0.075 = 0.9$$

Note that $P(X \le 2) = P(X < 3)$

$$= P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 0.9.$$

Probability of an Event (Cont.)

Siblings x	Probability $P(X = x)$
0	0.200
1	0.425
2	0.275
3	0.075
4	0.025
	1.000

The probability of the event that the students have at most two siblings (when x is 2) can be represented, again, by $P(X \leq 2)$.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 0.200 + 0.425 + 0.075 = 0.9$$

Similarly, the probability of the event that the students have at most one siblings (when x is 1) can be represented by $P(X \leq 1)$.

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$

$$= 0.200 + 0.425 = 0.645$$

Probability of an Event (Cont.)

Siblings x	Probability $P(X = x)$
0	0.200
1	0.425
2	0.275
3	0.075
4	0.025
	1.000

The probability of the event that the students have more than two siblings (when x is 2) can be represented, by $P(X > 2)$.

$$P(X > 2) = P(X = 3) + P(X = 4)$$

$$= 0.075 + 0.025 = 0.1$$

Alternatively, the probability of the event that the students have more than two siblings (when x is 2) can be solved by using the complementation rule.

$P(X > 2) = 1 - P(X \le 2) = 1 - 0.9 = 0.1$

Probability of an Event (Cont.)

Siblings x	Probability $P(X = x)$
0	0.200
1	0.425
2	0.275
3	0.075
4	0.025
	1.000

Can you tell what is the probability of the event that the students have more than one but less than four siblings?

This can be represented by

$P(1 < X)$ and $P(X < 4) = P(1 < X < 4)$.

$P(1 < X < 4) = P(X = 2) + P(X = 3)$

$= 0.275 + 0.075 = 0.35$

What is the Mean of a Discrete Random Variable?

Mean of a Discrete Random Variable

The **mean of a discrete random variable** X is denoted $\boldsymbol{\mu}_X$ or, when no confusion will arise, simply $\boldsymbol{\mu}$. It is defined by

$$\mu = \Sigma x P(X = x).$$

The terms **expected value** and **expectation** are commonly used in place of the term *mean*.†

Hint: To obtain the mean of a discrete random variable, multiply each possible value by its probability and then add those products.

The Mean of a Random Variable (Cont.)

Interpretation of the Mean of a Random Variable

In a large number of independent observations of a random variable X, the average value of those observations will approximately equal the mean, μ, of X. The larger the number of observations, the closer the average tends to be to μ.

What does is mean?

The mean of a random variable can be considered the long-run-average value of the random variable in repeated independent observations.

Expected Value: Center

A random variable assumes a value based on the outcome of a random event.

- We use a capital letter, like X, to denote a random variable.
- A particular value of a random variable will be denoted with a lower case letter, in this case x.

Expected Value: Center (cont.)

There are two types of random variables:

- Discrete random variables can take one of a finite number of distinct outcomes.
  - Example: Number of credit hours
- Continuous random variables can take any numeric value within a range of values.
  - Example: Cost of books this term

Expected Value: Center (cont.)

- A probability model for a random variable consists of:
  - The collection of all possible values of a random variable, and
  - the probabilities that the values occur.
- Of particular interest is the value we expect a random variable to take on, notated μ (for population mean) or $E(X)$ for expected value.

Expected Value: Center (cont.)

The expected value of a (discrete) random variable can be found by summing the products of each possible value and the probability that it occurs:

$$\mu = E(X) = \sum x \cdot P(x)$$

- Note: Be sure that every possible outcome is included in the sum and verify that you have a valid probability model to start with.

First Center, Now Spread…

For data, we calculated the standard deviation by first computing the deviation from the mean and squaring it. We do that with discrete random variables as well.

The variance for a random variable is:

$$\sigma^2 = Var(X) = \sum (x-\mu)^2 \cdot P(x)$$

The standard deviation for a random variable is:

$$\sigma = SD(X) = \sqrt{Var(X)}$$

More About Means and Variances

Adding or subtracting a constant from data shifts the mean but doesn’t change the variance or standard deviation:

$$E(X \pm c) = E(X) \pm c \qquad Var(X \pm c) = Var(X)$$

- Example: Consider everyone in a company receiving a $5000 increase in salary.

More About Means and Variances (cont.)

In general, multiplying each value of a random variable by a constant multiplies the mean by that constant and the variance by the *square* of the constant:

$$E(aX) = aE(X) \qquad Var(aX) = a^2 Var(X)$$

– Example: Consider everyone in a company receiving a 10% increase in salary.

More About Means and Variances (cont.)

In general,

- The mean of the sum of two random variables is the sum of the means.
- The mean of the difference of two random variables is the difference of the means.

$$E(X \pm Y) = E(X) \pm E(Y)$$

- If the random variables are *independent*, the variance of their sum *or* difference is always the sum of the variances.

$$Var(X \pm Y) = Var(X) + Var(Y)$$

Combining Random Variables
(The Bad News)

It would be nice if we could go directly from models of each random variable to a model for their sum.

But, the probability model for the sum of two random variables is *not* necessarily the same as the model we started with *even when the variables are independent*.

Thus, even though expected values may add, the probability model itself is different.

Continuous Random Variables

Random variables that can take on any value in a range of values are called continuous random variables.

Continuous random variables have means (expected values) and variances.

We won't worry about how to calculate these means and variances in this course, but we can still work with models for continuous random variables when we're given the parameters.

Combining Random Variables (The Good News)

Nearly everything we’ve said about how discrete random variables behave is true of continuous random variables, as well.

When two independent continuous random variables have Normal models, so does their sum or difference.

This fact will let us apply our knowledge of Normal probabilities to questions about the sum or difference of independent random variables.

What Can Go Wrong?

Probability models are still just models.

- Models can be useful, but they are not reality.
- Question probabilities as you would data, and think about the assumptions behind your models.

If the model is wrong, so is everything else.

What Can Go Wrong? (cont.)

- Don't assume everything's Normal.
  - You must *Think* about whether the Normality Assumption is justified.
- Watch out for variables that aren't independent:
  - You can add expected values for *any* two random variables, but
  - you can only add variances of *independent* random variables.

What Can Go Wrong? (cont.)

- Don't forget: Variances of independent random variables add. Standard deviations don't.
- Don't forget: Variances of independent random variables add, even when you're looking at the difference between them.
- Don't write independent instances of a random variable with notation that looks like they are the same variables.

What have we learned so far?

We know how to work with random variables.

– We can use a probability model for a discrete random variable to find its expected value and standard deviation.

The mean of the sum or difference of two random variables, discrete or continuous, is just the sum or difference of their means.

And, *for independent random variables*, the variance of their sum or difference is always the *sum* of their variances.

What have we learned so far? (cont.)

Normal models are once again special.

- Sums or differences of Normally distributed random variables also follow Normal models.

What have we learned?

We have learned to:

1. Determine the probability distribution of a discrete random variable.
2. Construct a probability histogram.
3. Describe events using random-variable notation, when appropriate.
4. Use the frequentist interpretation of probability to understand the meaning of probability distribution of a random variable.
5. Find and interpret the mean and standard deviation of a discrete random variable.

Credit

Some of these slides have been adapted/modified in part/whole from the slides of the following textbooks.

- Weiss, Neil A., Introductory Statistics, 8th Edition
- Bock, David E., Stats: Data and Models, 3rd Edition