Discrete Random Variables

MAT 152 – STATISTICAL METHODS I
LECTURE 1
INSTRUCTOR: DUSTIN ROTEN
FALL 2020

A Quick Review…

Suppose 50 high school students were surveyed about the number of electives they have signed up for. Given that students must take at least three core courses each year, the possible outcomes are 1-5. The data were collected and aggregated in the table to the right.

Frequency
The number of times a value occurs.

Relative Frequency
The ratio of the number of times a value occurs to the total number of outcomes.

Cumulative Relative Frequency
The sum of the relative frequencies for all values that are less than or equal to the given value.

Number of Electives	Frequency	Relative Frequency	Cumulative Relative Frequency
1	8	$\frac{8}{50} = 0.16$	0.16
2	17	$\frac{17}{50} = 0.34$	0.50
3	13	$\frac{13}{50} = 0.26$	0.76
4	10	$\frac{10}{50} = 0.20$	0.96
5	2	$\frac{2}{50} = 0.04$	1
Total	50	1	

A Quick Review…

What is the average number of electives taken by high school students?

The average can be calculated from the frequency table.

$$\bar{x} = \frac{1}{50}\sum_{i=1}^{N} x_i = \frac{1}{50}\left(1(8) + 2(17) + 3(13) + 4(10) + 5(2)\right)$$

$$\bar{x} = 2.62$$

Number of Electives	Frequency	Relative Frequency	Cumulative Relative Frequency
1	8	$\frac{8}{50} = 0.16$	0.16
2	17	$\frac{17}{50} = 0.34$	0.50
3	13	$\frac{13}{50} = 0.26$	0.76
4	10	$\frac{10}{50} = 0.20$	0.96
5	2	$\frac{2}{50} = 0.04$	1
Total	50	1	

But wait! There's more!

Remember that $P(x) = \frac{\text{outcomes of } x}{\text{all possible outcomes}}$

So, the probability of picking a student from this group that is taking 3 electives is $\frac{13}{50}$.

$$\bar{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = \frac{1}{50}(1(8) + 2(17) + 3(13) + 4(10) + 5(2))$$

What if we distribute the $\frac{1}{50}$ throughout the parentheses?

$$\bar{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = \left(1\left(\frac{8}{50}\right) + 2\left(\frac{17}{50}\right) + 3\left(\frac{13}{50}\right) + 4\left(\frac{10}{50}\right) + 5\left(\frac{2}{50}\right)\right)$$

Notice that each term in the parentheses represents the probability of picking a student with a certain number of electives.

Rewriting the Equation

$$\bar{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = \left(1\left(\frac{8}{50}\right) + 2\left(\frac{17}{50}\right) + 3\left(\frac{13}{50}\right) + 4\left(\frac{10}{50}\right) + 5\left(\frac{2}{50}\right)\right)$$

Use probability notation:

$$\bar{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = \big(1P(1) + 2P(2) + 3P(3) + 4P(4) + 5P(5)\big)$$

$$1P(1) + 2P(2) + 3P(3) + 4P(4) + 5P(5) = \sum_{k=1}^{5} x_k P(x_k)$$

$i =$ index of the individual values (All 50 of the values)

$k =$ index of the possible values (The values 1, 2, 3, 4, 5)

Formal Notation

A **random variable** describes the outcomes of a statistical experiment in words. Random variables are denoted by upper case letters and written in words.

Lower case letters denote a value of a random variable.

Example:

$X = \text{the number of heads you get when you toss three fair coins}$

$S = \{TTT, THH, HTH, HHT, HTT, THT, TTH, HHH\}$

$x = 0, 1, 2, 3$

Since the outcomes are countable, X is a discrete random variable.

Probability Distribution Function (PDF)

A discrete **probability distribution function** has two characteristics:

1. Each probability is between 0 and 1, inclusive.
2. The sum of the probabilities is 1.

Previous Example

Suppose 50 high school students were surveyed about the number of electives they have signed up for. Given that students must take at least three core courses each year, the possible outcomes are 1-5.

Example

Suppose 50 high school students were surveyed about the number of electives they have signed up for. Given that students must take at least three core courses each year, the possible outcomes are 1-5. The data were collected and aggregated in the table to the right.

$X =$ the number of electives taken by a student

$x = 1, 2, 3, 4, 5$

Both requirements of a PDF are met.

1) Each probability is between 0 and 1.
2) The sum of the probabilities is 1.

X	Frequency	$P(x)$	Cumulative Relative Frequency
1	8	$P(x=1) = \frac{8}{50}$	0.16
2	17	$P(x=2) = \frac{17}{50}$	0.50
3	13	$P(x=3) = \frac{13}{50}$	0.76
4	10	$P(x=4) = \frac{10}{50}$	0.96
5	2	$P(x=5) = \frac{2}{50}$	1
Total	50	1	

Example (cont.)

The **expected value** is referred to as the **"long-term" average or mean.**

This is the outcome you would expect after collecting LOTS of data.

The Law of Large Numbers states that, as the number of trials in an experiment increases, the difference between the expected value of an event and the actual value approaches 0.

When evaluating the long-term results of statistical experiments, the long-term average is known as the **expected value** or mean.

$$\mu = \text{expected value} = \text{mean}$$

X	Frequency	$P(x)$
1	8	$P(x = 1) = \frac{8}{50}$
2	17	$P(x = 2) = \frac{17}{50}$
3	13	$P(x = 3) = \frac{13}{50}$
4	10	$P(x = 4) = \frac{10}{50}$
5	2	$P(x = 5) = \frac{2}{50}$
Total	50	1

Calculating the Standard Deviation

Like data, probability distributions have standard deviations. To calculate the standard deviation (σ) of a probability distribution, find each deviation from its expected value, square it, multiply it by its probability, add the products, and take the square root.

X	Frequency	$P(x)$	$(x-\mu)^2 \cdot P(x)$	$(x-\mu)^2 \cdot P(x)$
1	8	$P(x=1)=\frac{8}{50}$	$(1-2.62)^2 \cdot \frac{8}{50}$	$=0.419904$
2	17	$P(x=2)=\frac{17}{50}$	$(2-2.62)^2 \cdot \frac{17}{50}$	$=0.130696$
3	13	$P(x=3)=\frac{13}{50}$	$(3-2.62)^2 \cdot \frac{13}{50}$	$=0.037544$
4	10	$P(x=4)=\frac{10}{50}$	$(4-2.62)^2 \cdot \frac{10}{50}$	$=0.38088$
5	2	$P(x=5)=\frac{2}{50}$	$(5-2.62)^2 \cdot \frac{2}{50}$	$=0.226576$
Total	50	1		$\sigma^2=1.1956$

$$\sigma=\sqrt{\sigma^2}=\sqrt{1.1956}=1.093435$$

A calculator or computer can be used to calculate μ, σ^2, σ.

(Expected value, variance, and standard deviation

Example

Suppose two fair, six-sided die were tossed.

1) What are all the possible outcomes?

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Table 4.13

2) Suppose $X = \text{the number of faces that show an even number};\ x = 0, 1, 2$

Example (cont.)

$$P(x = 0) = \frac{9}{36}$$

9 outcomes where **both** faces show odd numbers

18 outcomes where **one** face shows even numbers

9 outcomes where **two** faces show even numbers

$$P(x = 1) = \frac{18}{36}$$

$$P(x = 2) = \frac{9}{36}$$

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Table 4.13

Example (cont.)

$\boldsymbol{x}$	$\boldsymbol{P(x)}$	$\boldsymbol{xP(x)}$	$\boldsymbol{(x-\mu)^2 \cdot P(x)}$
0	$P(x=0)=\frac{9}{36}$	$0\cdot\left(\frac{9}{36}\right)=0$	$(0-1)^2\cdot\left(\frac{9}{36}\right)=\frac{9}{36}$
1	$P(x=1)=\frac{18}{36}$	$1\cdot\left(\frac{18}{36}\right)=\frac{18}{36}$	$(1-1)^2\cdot\left(\frac{18}{36}\right)=0$
2	$P(x=2)=\frac{9}{36}$	$2\cdot\left(\frac{9}{36}\right)=\frac{18}{36}$	$(2-1)^2\cdot\left(\frac{9}{36}\right)=\frac{9}{36}$
TOTAL	1	$\mu=\sum xP(x)=1$	$\sigma^2=\sum(x-\mu)^2P(x)=\frac{18}{36}$

(Add the columns)

Probability of an event (adds to 1)

Expected value (μ)

Variance (σ^2)

$$\sigma=\sqrt{\sigma^2}=\sqrt{\frac{18}{36}}=0.7071$$

(Standard Deviation)

Example

Suppose you owned a small delivery business in which you earn $50 for every successful delivery (no damage to the product). In cases where the product is damaged, customers do not have to pay the delivery fee AND you provide $20 for damages. Fortunately, damage only occurs 5% of the time. How much money will you make after a year of deliveries?

1) Determine *X*.

$X = \text{the money made from a delivery; } x = \{\$50, -\$20\}$

2) Determine the probabilities of the events

$P(\text{an undamaged delivery}) = 0.95 \text{ (or 95\%)}$

$P(\text{a damaged delivery}) = 0.05 \text{ (or 5\%)}$

3) Calculate!

$\sigma = \$15.26$

Delivery	x	$P(x)$	$xP(x)$	$(x-\mu)^2P(x)$
Undamaged	$50	0.95	$\$50 \cdot 0.95 = \$47.5$	$(\$50 - \$46.50)^2 \cdot (0.95) = 11.6375$
damaged	-$20	0.05	$-\$20 \cdot 0.05 = -\$1$	$((-\$20) - \$46.50)^2 \cdot (0.05) = 221.1125$
Total:		1	$\mu = \$46.50$ per delivery	$\sigma^2 = 232.75$ Variance ($\$^2$)

A Quick Review

Expected value (long-term mean) after many experiments: $\mu = \sum_i x_i P(x_i)$

Variance after many experiments: $\sigma^2 = \sum_i (x_i - \mu)^2 P(x_i)$

Standard deviation: $\sigma = \sqrt{\sigma^2} = \sqrt{\sum_i (x_i - \mu)^2 P(x_i)}$