Discrete Random Variables

MATH 130, *Elements of Statistics I*

J Robert Buchanan

Department of Mathematics

Fall 2023

Objectives

During this lesson we will learn to:

- distinguish between discrete and continuous random variables,
- identify discrete probability distributions,
- construct probability histograms,
- compute and interpret the mean of a random variable,
- interpret the mean of a discrete random variable as an expected value,
- compute the standard deviation of a discrete random variable.

Random Variables

Definition

A **random variable** is a numerical measure of the outcome of an experiment. Random variables are denoted using letters such as X.

Remark: the numerical value of a random variable is unknown until the experiment is completed.

Example

If X is the number of students absent from class August 1, 2023, then the value of X is unknown until after class on August 1, 2023.

Discrete *vs.* Continuous Random Variables

Definition

A **discrete random variable** has either a finite or countable number of values. The values of a discrete random variable can be plotted on a number line with gaps between each point.

A **continuous random variable** has infinitely many values from a range without gaps.

Examples

Which of the following describe discrete or continuous random variables?

1. The sum of a roll of a pair of fair dice.
2. The batting average of a professional baseball player in the 2023 season.
3. The number of A's earned by students in a section of statistics with 39 students enrolled.
4. The number of hours lost to injuries in the workplace for a company next week.
5. The price of the stock of company XYZ tomorrow.
6. The number of chocolate chips in a cookie.

Probability Distributions

Definition

The **probability distribution** of a discrete random variable X provides the possible values of the random variable and their corresponding probabilities. A probability distribution can be in the form of a table, graph, or mathematical formula.

Example

Suppose a single die has the following probabilities associated with its outcomes (this is not a "fair" die since the probabilities of the values of the random variable are not equal).

X	$P(X)$
1	0.170
2	0.173
3	0.170
4	0.177
5	0.151
6	0.159

This table provides a **probability distribution** for the die.

Properties of a Discrete Probability Distribution

Theorem (Rules for a Discrete Probability Distribution)

Let $P(x)$ denote the probability that the random variable X equals x; then

1. $\sum P(x) = 1$
2. $0 \leq P(x) \leq 1$

Note: capital X denotes the random variable whose value is unknown and lowercase x denotes a value (a number typically) that X can assume.

Examples

Which of the following describe a discrete probability distribution?

x	$P(x)$
1	0.170
2	0.173
3	0.170
4	0.177
5	0.151
6	0.159

x	$P(x)$
1	0.27
2	0.31
3	0.07
4	0.17
5	0.15
6	0.19

x	$P(x)$
1	0.87
2	0.58
3	-0.65
4	0.86
5	-0.98
6	0.32

Examples

Which of the following describe a discrete probability distribution?

x	P(x)
1	0.170
2	0.173
3	0.170
4	0.177
5	0.151
6	0.159

x	P(x)
1	0.27
2	0.31
3	0.07
4	0.17
5	0.15
6	0.19

x	P(x)
1	0.87
2	0.58
3	-0.65
4	0.86
5	-0.98
6	0.32

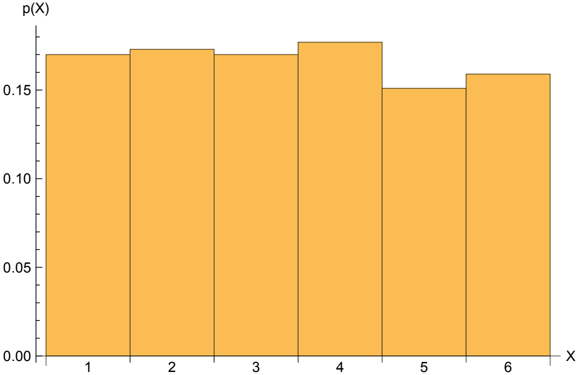
Only the first table represents a probability distribution.

Probability Histogram

Definition

A **probability histogram** is a histogram in which the horizontal axis corresponds to the value of the random variable and the vertical axis represents the probability of each value of the random variable.

x	$P(x)$
1	0.170
2	0.173
3	0.170
4	0.177
5	0.151
6	0.159



Mean of a Discrete Random Variable

Theorem (Mean of a Discrete Random Variable)

The mean of a discrete random variable is given by the formula

$$\mu_X = \sum [x \cdot P(x)]$$

where x is the value of the random variable and $P(x)$ is the probability of observing the random variable x.

Example

x	$\text{P}(x)$	$x \cdot \text{P}(x)$
1	0.170	
2	0.173	
3	0.170	
4	0.177	
5	0.151	
6	0.159	

Example

x	$P(x)$	$x \cdot P(x)$
1	0.170	0.170
2	0.173	0.346
3	0.170	0.510
4	0.177	0.708
5	0.151	0.755
6	0.159	0.954
		3.443

Thus $\mu_X = 3.4$.

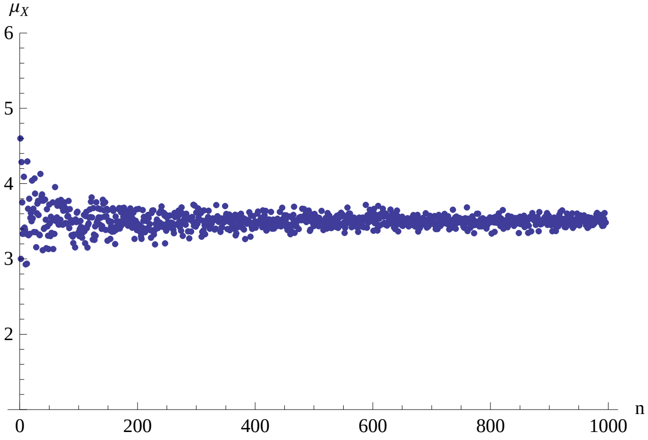
Interpretation

Suppose an experiment is conducted n independent times and the value of the random variable X is recorded. As the number of repetitions n of the experiment is increased, the mean value of X recorded will approach the value μ_X.

Example

Adopting the classical approach, the probability of the outcome X obtained from rolling a fair die is $P(X) = 1/6$. Thus

$$\mu_X = \sum [x \cdot P(x)] = \frac{1+2+3+4+5+6}{6} = 3.5$$



Expected Value

Definition

The mean of a random variable is also called its **expected value**.

Example

Suppose a lottery ticket costs \$5 and has a probability $P(\text{win}) = 0.001$ of winning a \$1000 prize. What is the expected value of the ticket holder's net gain/loss?

Solution

x	$P(x)$
-5	0.999
995	0.001

$$\mu_X = (-5)(0.999) + (995)(0.001) = -4$$

Solution

x	$P(x)$
-5	0.999
995	0.001

$$\mu_X = (-5)(0.999) + (995)(0.001) = -4$$

Interpretation: the mean (average) value of the lottery ticket is \$1. If the ticket costs \$5, then the expected gain/loss made by purchasing a lottery ticket is $-\$4$.

Variance and Standard Deviation

The **variance** of a discrete random variable is given by

$$\sigma_X^2 = \sum \left[x^2 \cdot P(x) \right] - \mu_X^2,$$

where x is the value of the random variable, μ_X is the mean of the random variable, and $P(x)$ is the probability of observing the random variable x. The **standard deviation** of the discrete random variable is the square root of the variance.

$$\sigma_X = \sqrt{\sigma_X^2}.$$

Example

x	$P(x)$	$x^2 \cdot P(x)$
1	0.170	
2	0.173	
3	0.170	
4	0.177	
5	0.151	
6	0.159	

Example

x	$P(x)$	$x^2 \cdot P(x)$
1	0.170	0.170
2	0.173	0.692
3	0.170	1.530
4	0.177	2.832
5	0.151	3.775
6	0.159	5.724

Example

x	$\mathrm{P}(x)$	$x^2 \cdot \mathrm{P}(x)$
1	0.170	0.170
2	0.173	0.692
3	0.170	1.530
4	0.177	2.832
5	0.151	3.775
6	0.159	5.724
		14.723

Example

x	$\mathrm{P}(x)$	$x^2 \cdot \mathrm{P}(x)$
1	0.170	0.170
2	0.173	0.692
3	0.170	1.530
4	0.177	2.832
5	0.151	3.775
6	0.159	5.724
		14.723

$$\sigma_X^2 = \sum \left[x^2 \cdot P(x)\right] - \mu_X^2 = 14.723 - (3.443)^2 = 2.869 \approx 2.9$$

Example

x	P(x)	$x^2 \cdot \mathrm{P}(x)$
1	0.170	0.170
2	0.173	0.692
3	0.170	1.530
4	0.177	2.832
5	0.151	3.775
6	0.159	5.724
		14.723

$$\sigma_X^2 = \sum \left[x^2 \cdot P(x)\right] - \mu_X^2 = 14.723 - (3.443)^2 = 2.869 \approx 2.9$$

$$\sigma_X = \sqrt{2.869} = 1.694 \approx 1.7$$