

Introduction to Statistics

Nels Grevstad

Metropolitan State University of Denver
ngrevsta@msudenver.edu

September 22, 2019

Nels Grevstad

Topics

- 1 Discrete Random Variables
- 2 (Optional Section) Standard Deviation of a Discrete Probability Distribution

Nels Grevstad

Objectives

Objectives:

- Know the definition of random variables.
- Distinguish between discrete and continuous random variables.
- Use probability distributions to obtain probabilities involving discrete random variables.
- Compute and interpret the mean of a discrete probability distribution.
- (Optional) Compute and interpret the standard deviation of a discrete probability distribution.

Nels Grevstad

Discrete Random Variables (5.1)

Introduction to Random Variables

- When the outcome of a chance experiment is a numerical value, we call the outcome a **random variable**.
Thus a **random variable** is interpreted as a **numerical value** that's **determined by chance**.
- We'll use a **capital letter** such as X to denote a **random variable**.
The **actual** value of X isn't determined until the chance experiment is carried out.

Nels Grevstad

Notes

Notes

Notes

Notes

Example

Roll a six-sided die. Define the variable X to be

$X =$ The number showing when the die lands

Then X is a **random variable** whose possible values are 1, 2, 3, 4, 5, and 6.

Nels Grevstad

Example

Randomly select a person from a population. Then the **sample space** consists of the individuals in the population:

$$S = \{ \text{Stephanie Lawson,} \\ \text{Jeffrey Miller,} \\ \text{Angela DuPont,} \\ \vdots \\ \text{Karl Stevenson} \}$$

Define the variable X to be

$X =$ The selected person's age

Then X is a **random variable**.

Nels Grevstad

Example

Randomly select a person from a population. Then the **sample space** consists of the individuals in the population:

$$S = \{ \text{Stephanie Lawson, } 48, \\ \text{Jeffrey Miller } 28, \\ \text{Angela DuPont } 27, \\ \vdots \quad \quad \quad \vdots \\ \text{Karl Stevenson } 34 \}$$

Define the variable X to be

$X =$ The selected person's age

Then X is a **random variable**.

Nels Grevstad

Think of each individual as having a numerical value (their age) and then randomly selecting one individual and denoting their age by X .

Nels Grevstad

Notes

Notes

Notes

Notes

Discrete Random Variables

- A random variable is **discrete** if its possible values of are isolated numbers with gaps between them (e.g. the integers 0, 1, 2, ...).

It's **continuous** if the possible values form an entire interval, or continuum (e.g. *all* the values greater than zero, such as 3.212974541).

Exercise

Determine which of the following are **discrete** random variables and which are **continuous**.

- a) Randomly select a U.S. resident, and define the variable X to be

X = The selected person's height

- b) Roll a six-sided die and define the variable X to be

X = The number showing when the die lands

- c) Randomly select a vehicle from a highway system, and define the variable X to be

X = The number of occupants in the vehicle

- d) Randomly select a loaf of bread from a bakery, and define the variable X to be

X = The weight of the selected loaf of bread

- The **probability distribution** of a random variable X specifies two things:

1. The values that are possible for the variable.
2. The probabilities of those values.

- Probability distributions can be represented by tables, graphs, or formulas.

Example

The table below shows the **vehicle occupancy rates** on urban arterials and freeways in Miami-Dade County, Florida.

Number of Occupants	Percentage of Vehicles
1	82 %
2	12 %
3	4 %
4	2 %

(The percentage of vehicles with five or more occupants was negligibly small.)

Nels Grevstad

Consider **randomly** selecting a vehicle, and let

X = The number of occupants in the vehicle

Then X is a **discrete random variable**, with possible values 1, 2, 3, and 4.

The **probability distribution** of X is represented by the following table.

Number of occupants x	1	2	3	4
Probability $P(X = x)$	0.82	0.12	0.04	0.02

where $P(X = x)$ represents the probability that the **random variable** X (number of occupants in a randomly selected vehicle) will equal the **particular value** x , where x is either 1, 2, 3, or 4.

Nels Grevstad

For example, from the table,

$$P(X = 2) = 0.12,$$

which says there's a **12%** chance that a randomly selected vehicle will have **two** occupants.

The **probability distribution table** on the last slide is just the **relative frequency distribution table** of the number of occupants in the **population** of vehicles in Miami-Dade County.

Nels Grevstad

The **probability distribution** of a **discrete** random variable can be graphed in a so-called **probability histogram** like the one on the next slide.

The height of each bar is the **probability** of the number of vehicle occupants shown at the base of the bar.

We can think of this as the **relative frequency histogram** for the **population** of vehicles.

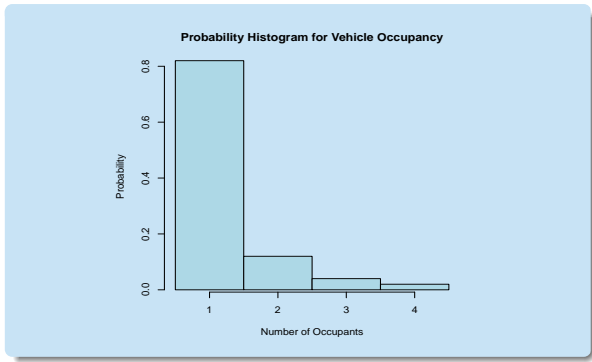
Nels Grevstad

Notes

Notes

Notes

Notes



Nels Grevstad

- As in the last example, a **probability distribution** often represents a **population** from which an individual is **randomly selected**.

Nels Grevstad

Exercise

Using the **probability distribution** table from the last example, find the following **probabilities**.

- $P(X = 3)$, the probability that the vehicle will have three occupants.
- $P(X \leq 3)$, the probability that the vehicle will have three or fewer occupants.
- $P(X < 3)$, the probability that the vehicle will have fewer than three occupants.
- $P(2 \leq X \leq 4)$, the probability that the vehicle will have between two and four occupants, inclusive.

Nels Grevstad

The Mean of a Discrete Probability Distribution (5.2)

- The **mean** (denoted μ) of the probability distribution of a **discrete** random variable X is defined as:

Mean of a Discrete Probability Distribution:

$$\mu = \sum xP(X = x)$$

where the summation is over all possible values x of the random variable X .

μ is also sometimes also called the **expected value** of X .

Nels Grevstad

Notes

Notes

Notes

Notes

- **Interpretation:** The mean μ of a **probability distribution** represents the **center of the probability distribution** of X (i.e. its **"balancing point"**).

Depending on the context, μ can also be thought of as:

- The **long-run average** value of the random variable X upon repeating the chance experiment a **very large** number of times.
- The **population mean** when the probability distribution represents a population.
- A **weighted average** of the possible values x of the random variable X , where the "weights" are the probabilities of the x values.

Nels Grevstad

Exercise

When a roulette wheel is spun, the ball is equally likely to land in any of **38** slots, **18** of which are red, **18** black, and **2** green.



Nels Grevstad

A bet of \$1.00 on red **pays a dollar** if the ball lands in a red slot. Otherwise, you **lose your dollar**. Let

X = Your (net) winnings after a \$1.00 bet on red

- a) Determine the **probability distribution** of X , and fill in its values in the table below.

Net winnings x	Probability $P(X = x)$

Nels Grevstad

- b) Find the **mean** μ of the probability distribution of X .
 c) Give an interpretation of μ as a **long-run average**.

Nels Grevstad

Notes

Notes

Notes

Notes

Notes

Example

Consider rented housing units in the U.S. Let

X = The number of rooms in randomly selected unit

Relative frequencies provided in the Current Housing Report by the U.S. Census Bureau give the **probability distribution** of X :

Number of rooms x	1	2	3	4	5	6	7	8
Probability $P(X = x)$	0.01	0.03	0.25	0.35	0.20	0.10	0.04	0.02

Nels Grevstad

Discrete Random Variables

(Optional Section) Standard Deviation of a Discrete Probability Dis

The mean μ of this probability distribution is

$$\begin{aligned}\mu &= \sum xP(X = x) \\ &= 1(0.01) + 2(0.03) + 3(0.25) + 4(0.35) + 5(0.20) + 6(0.10) \\ &\quad + 7(0.04) + 8(0.02) \\ &= 4.26\end{aligned}$$

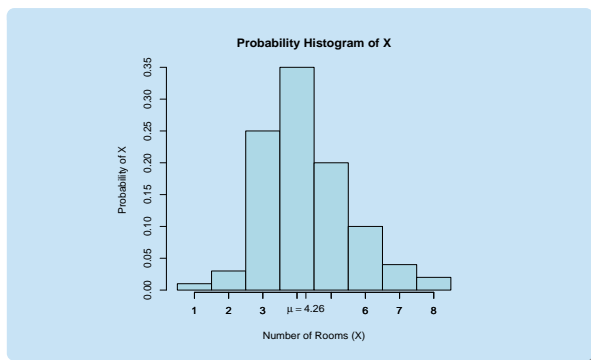
The value of μ is marked on the horizontal axis of the **probability histogram** for this distribution on the next slide.

Notes

Nels Grevstad

Discrete Random Variables

(Optional Section) Standard Deviation of a Discrete Probability Dis



Nels Grevstad

Discrete Random Variables

(Optional Section) Standard Deviation of a Discrete Probability Dis

Because the probability distribution represents the **population** of housing units, we can interpret μ as the **population mean** number of rooms.

Note also that μ is the **center** ("balancing point") of the distribution.

Notes

Notes

Nels Grevstad

(Optional Section) Standard Deviation of a Discrete Probability Distribution (5.2)

- The **variance** (denoted σ^2) of the probability distribution of a discrete random variable X is:

Variance of a Discrete Probability Distribution:

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

Nels Grevstad

- The **standard deviation** (denoted σ) is the square root of the variance:

Standard Deviation of a Discrete Probability Distribution:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum (x - \mu)^2 P(X = x)}$$

Nels Grevstad

- Interpretation:** The standard deviation σ of a **probability distribution** measures **the amount of spread** in the **probability distribution**.

Depending on the context, σ can also be thought of as:

- A measure of **how much X varies** from one run of the chance experiment to the next.

In particular, it represents a **typical deviation of X away from the mean μ** .

- The **standard deviation of the population** when the probability distribution represents a population.

Nels Grevstad

Example

Let X again be the **number of rooms** in randomly selected U.S. rental housing unit.

Recall that the **mean** of the probability distribution of X is $\mu = 4.26$.

The intermediate calculations required for computing σ are shown on the next slide.

Nels Grevstad

Notes

Notes

Notes

Notes

x	$P(X = x)$	$(x - \mu) = (x - 4.26)$	$(x - \mu)^2$
1	0.01	-3.26	10.63
2	0.03	-2.26	5.12
3	0.25	-1.26	1.59
4	0.35	-0.26	0.07
5	0.20	0.74	0.55
6	0.10	1.74	3.03
7	0.04	2.74	7.51
8	0.02	3.74	13.99

Multiplying each squared deviation in the last column by $P(X = x)$ in the second column, then summing the results, gives the **variance** of the **probability distribution** of X :

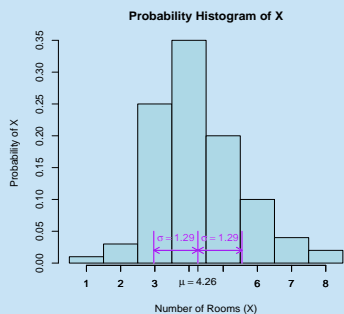
$$\begin{aligned} \sigma^2 &= \sum (x - \mu)^2 P(X = x) \\ &= 10.63(0.01) + 5.12(0.03) + 1.59(0.25) + 0.07(0.35) \\ &\quad + 0.55(0.20) + 3.03(0.10) + 7.51(0.04) + 13.99(0.02) \\ &= \mathbf{1.67}. \end{aligned}$$

Taking the square root gives the **standard deviation**:

$$\begin{aligned} \sigma &= \sqrt{\sigma^2} \\ &= \sqrt{1.67} \\ &= \mathbf{1.29}. \end{aligned}$$

In the **probability histogram** on the next slide, σ is marked as a "typical" deviation of X away from μ , and serves as a measure of the **spread** of the distribution.

Because the probability distribution represents the *population* of housing units, we can interpret σ as the **standard deviation of the population**.



Notes

Notes

Notes

Notes
