Section 5: What is probability?

STOR 155.02, Spring ’21

updated 2021-02-20

What you will learn

- XXX

Resources

- Textbook: ch 3.1

Recap and motivation

We've learned ways to investigate sample datasets.

sample data comes from a larger population, roughly speaking

it matters how the sample is collected!

and to understand this we need a little bit of probability

Uses of probability and sampling extend far beyond than sample data collection

with these tools, you will have a better understanding of how to answer questions like

- how do I understand **margins of error** in economic forecasts?
- if that basketball player is on a **hot streak**, what's the chance she makes her next two shots?
- how long would it take someone randomly entering characters to guess my password?
- if I **test positive** for covid/HIV/breast cancer/... **what's the chance I actually have it?**
- if I test negative, what's the chance I don't?
- what's up with the **bell curve**, and why is it everywhere?

Warning!

This will be the toughest section of the class for most of you.

You have to work hard if you want to do well.

I am here to help you --- but you must be willing to put in the work.

image credit: deviantart.com/monchoncho

Randomness

Everything in the world is the result of some process

- metabolism converts food to energy in living organisms
- weather today driven by atmospheric forces
- my salary, determined by my employment contract
- size of your heart at age 51, in ounces
- whether or not you win the lottery *this* time
- amount of caffeine in a cup of coffee, in grams, as measured in an experiment

Definition

> > An outcome subject to randomness is one such that the same process doesn't always lead to the same outcome.

Otherwise a process is called deterministic.

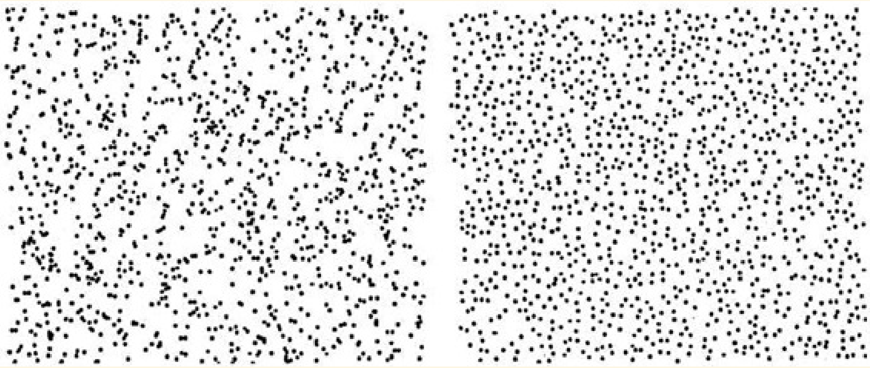
It's a philosophical question as to whether *anything* is truly random.

In this course, we will focus on things that can be best understood as random

to which the tools of probability can be applied

will build understanding by looking at examples as in data labs

e.g., which of these pictures was generated at random?



source: empiricalzeal

Other example: Throughout the probability section,

each lecture, our friend gudetama will slide off his toast at a randomly chosen slide

Definition

a random variable is the outcome of a random process

discrete random variable

takes integer values

or categorical, like the categorical data seen previously

continuous random variable

takes numeric values

in a range of possible values, finite or infinite

to sample a random variable means to observe one or more outcomes from this same random process

Probability

Probabilities associated with processes and their outcomes are

Statistical point of view

> > 'The probability of an outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.'
> >
> > textbook ch 3.1.1

Mathematical point of view

- numbers **between 0 and 1**
- one **for each possible outcome** (maybe an infinite number)
- representing the **weights used to calculate theoretical, weighted average** outcome
- and weights of all possible outcomes **must add to one**
- meaning *something* will happen for sure, though we don't yet know what.

Weighted averages and expectations

> the **expectation** of a random variable (r.v.) X is a **weighted average of all possible outcomes**, with weights given by the probabilities

$$P(X = x) \quad \text{is} \quad \text{the probability r.v. X takes outcome } x$$

For a discrete r.v. X with n possible outcomes $x_1 \dots x_n$, the expectation of X is defined mathematically as

$$E(X) = \sum_{1}^{n} x_i \, P(X = x_i) \; = x_1 \, P(X = x_1) + \dots x_n \, P(X = x_n)$$

Example

In this example, the r.v. will represent the outcome from a coin toss. Each of the two outcomes is **equally likely.**

$$X = 1 \quad \text{if coin comes up heads,} \qquad X = 0 \quad \text{if tails}$$

Using the rules of probability and expectation, tell me

$P(X = 1) =$ **?**

$E(X) =$ **?**

Rules of expectation

for any two random variables X, Y

$E(X+Y) = E(X) + E(Y)$

$E(aX) = aE(X)$ **for any number** a

$E(X) \geq \text{minimum}(x_1, x_2, \ldots x_n)$

$E(X) \leq \text{maximum}(x_1, x_2, \ldots x_n)$

example:

I flip a coin. If heads, I pay you 5 dollars. Otherwise you get nothing. X is the amount you win from this game.

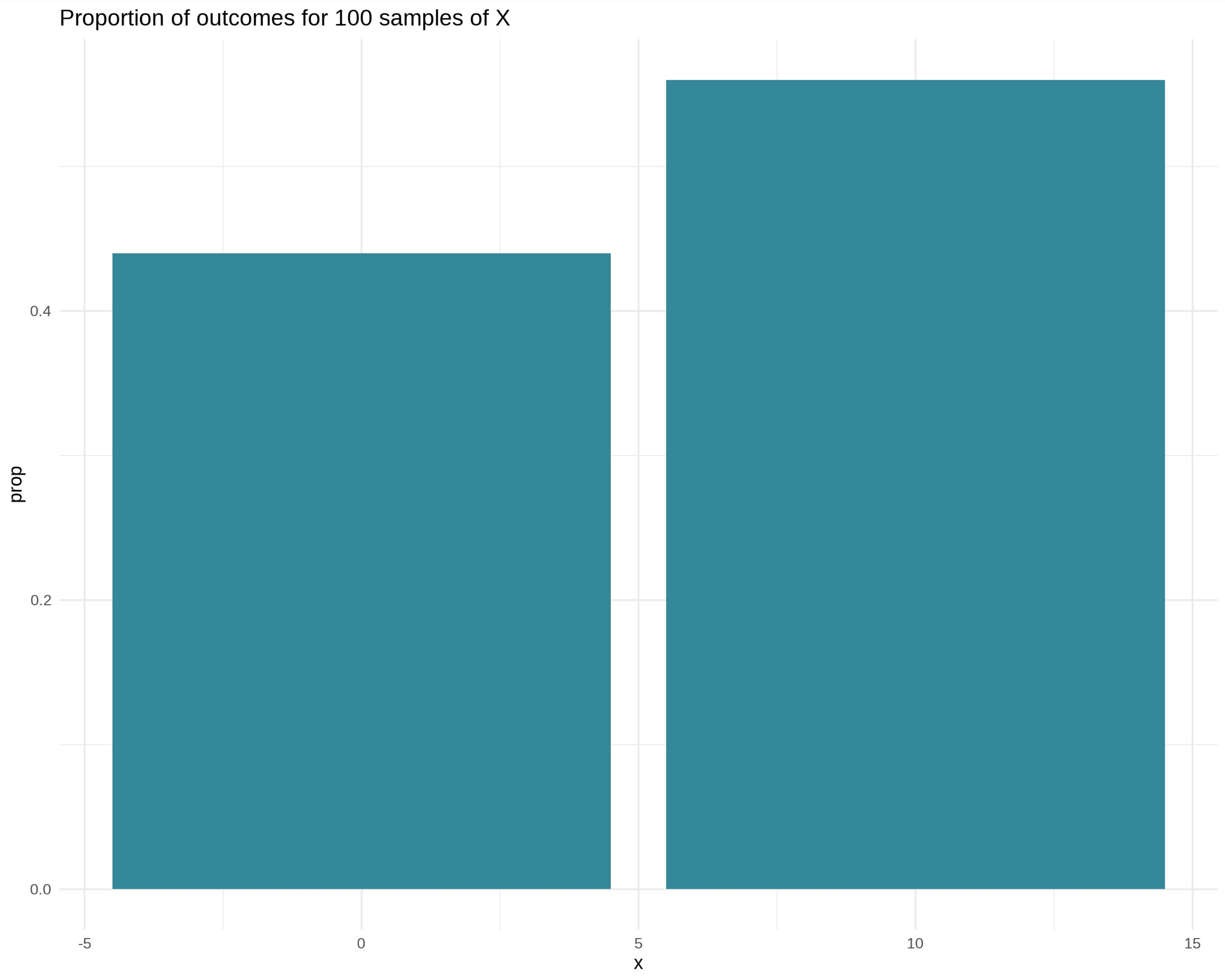
possible outcomes?

$P(X=0) =$ **?**

$E(X) =$ **?**

Basic application of probability rules (and your intuition) tell you the answer

but you could have investigated this with a simulation!


Proportion of outcomes for 100 samples of X
prop
0.0
0.2
0.4
-5
0
5
10
15
x

General method for calculating probabilities when all outcomes are equally likely

$$P(X = x_i) = \frac{\text{number of ways } x_i \text{ can happen}}{\text{number of outcomes}}$$

example: pick one random card from a briscola deck

there are **40 cards** and **four suits** with an equal number of cards: clubs, coins, cups and swords.

$$X = 1 \quad \text{if card is cups,} \qquad X = 0 \quad \text{if not}$$

answer in the polls!

PollEv.com/brendanbrown849

poll closes at __________

don't be late!


five more minutes