

YData: An Introduction to Data Science

Lecture 14: Chance

Jessi Cisewski-Kehe and John Lafferty
Statistics & Data Science, Yale University
Spring 2019

Credit: data8.org



Announcements

For today

- Control statements
- Some elementary probability
- The Monty Hall puzzle

Control Statements

Control Statements

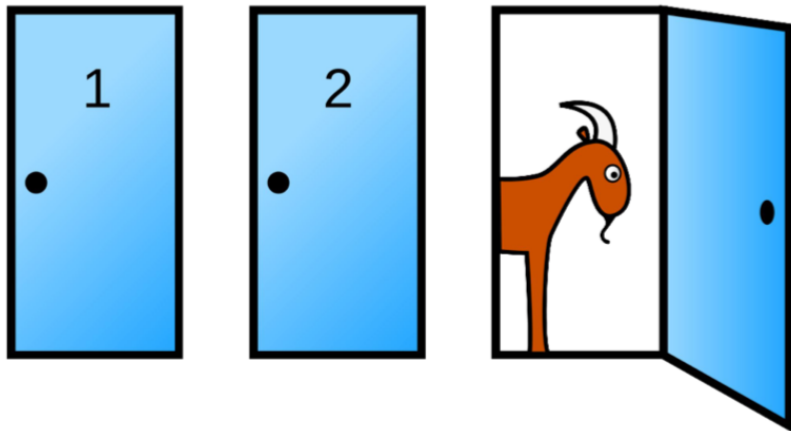
These statements control the sequence of computations that are performed in a program

- The keywords `if` and `for` begin control statements
- The purpose of `if` is to define functions that choose different behavior based on their arguments
- The purpose of `for` is to perform a computation for every element in a list or array

(DEMO)

The Monty Hall Problem

Monty Hall Problem



Probability

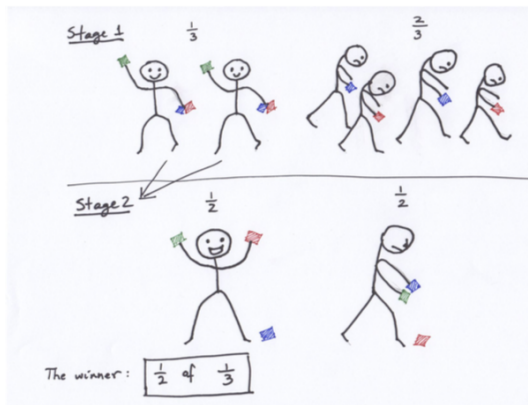
- Lowest value: 0
 - Chance of event that is impossible
- Highest value: 1 (or 100%)
 - Chance of event that is certain
- If an event has chance 70%, then the chance that it doesn't happen is
 - $100\% - 70\% = 30\%$
 - $1 - 0.7 = 0.3$

Equally Likely Outcomes

Assuming all outcomes are equally likely, the chance of an event A is:

$$P(A) = \frac{\text{number of outcomes that make } A \text{ happen}}{\text{total number of outcomes}}$$

Fraction of a Fraction



- There are three tickets: Red Green Blue
- What's the chance of GR when sampling two tickets *without* replacement?
RB RG BR BG GR GB $\implies P(\text{GR}) = 1/6$

Multiplication Rule (aka the Chain Rule)

Chance that two events A and B both happen

$$= P(A \text{ happens}) \times P(B \text{ happens given that A has happened})$$

- The answer is *less than or equal to* each of the two chances being multiplied
- The more conditions you have to satisfy, the less likely you are to satisfy them all
- Tickets: Red Green Blue: What's the chance of GR when sampling two tickets without replacement?

$$RB \ RG \ BR \ BG \ GR \ GB \implies P(GR) = 1/6$$

$$P(G) = 1/3, P(R \text{ given } G) = 1/2 \implies P(GR) = 1/3 \times 1/2 = 1/6$$

Addition Rule

If event A can happen in exactly one of two ways, then

$$P(A) = P(\text{first way}) + P(\text{second way})$$

- The answer is greater than or equal to the chance of each individual way
- Tickets: Red Green Blue: What is the chance of one R and one G?
RB RG BR BG GR GB \implies
 $P(\text{one R and one G}) = P(\text{GR}) + P(\text{RG}) = 1/6 + 1/6$

Example: At Least One Head

- In 3 tosses:
 - Any outcome except TTT
 - $P(\text{TTT}) = (\frac{1}{2}) \times (\frac{1}{2}) \times (\frac{1}{2}) = \frac{1}{8}$
 - $P(\text{at least one head}) = 1 - P(\text{TTT}) = \frac{7}{8} = 87.5\%$
- In 10 tosses:
 - $1 - (\frac{1}{2})^{10}$
 - 99.9%

(DEMO)



<https://www.youtube.com/watch?v=Asm7w6q7DsA>

Rethinking an Experiment

Some experiments that purport to show cognitive-dissonance effects might be explainable by statistics alone.

Experiment Psychologists first observe that a monkey seeks out red, blue and green M&Ms about equally.



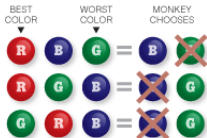
The monkey is given a choice of a red or blue candy. It chooses red.



If the monkey is then given a choice of blue or green, it is more likely to choose green.

Psychological explanation The monkey rationalizes its initial rejection of blue by telling itself it doesn't really like blue.

Statistical explanation If the monkey slightly prefers red over blue, there are only three possible ways it can rank green. In two of the three possibilities, the monkey will prefer green over blue.



THE NEW YORK TIMES