YData: Introduction to Data Science

Lecture 14: Chance

Overview

For loops continued

The Monty Hall problem

If there is time: some elementary probability

Announcements

To give home time for project 1, homework 5 just a practice homework

- i.e., it will not be graded

Keep working on project 1!

For loops

For loops

For loops repeat a process many times, iterating over a sequence of items

- Often we are iterating over an array of sequential numbers

```
animals = make_array("cat", "dog", "bat")
for creature in animals:
    print(creature)

for i in np.arange(4):
    print(i**2)
```

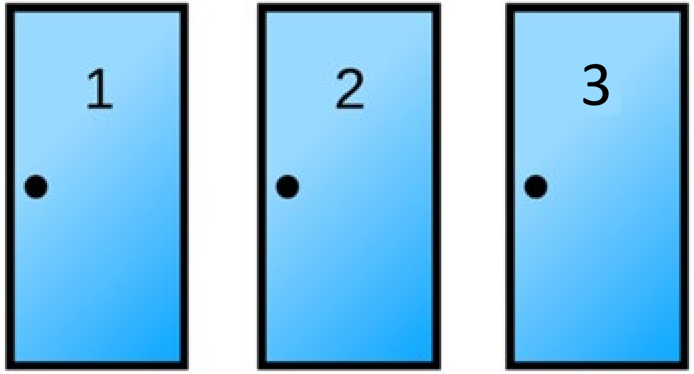
Let's explore this in Jupyter!

The Monty Hall Problem

The Monty Hall problem

"The Monty Hall Problem" comes from TV game show from the 1960s called "Let's Make a Deal"

Contestant is presented three closed doors



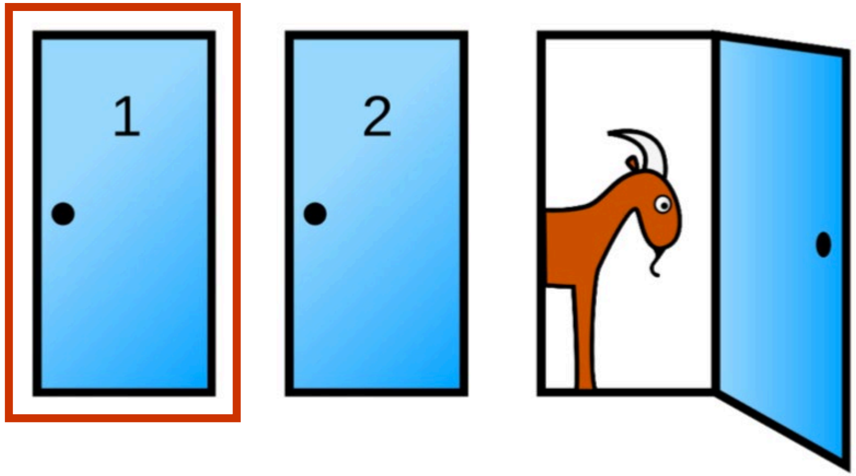
Behind one door is a fancy car, and goats are behind the other two doors

- The contestant does not know which door has the car

The Monty Hall problem

Steps of the game:

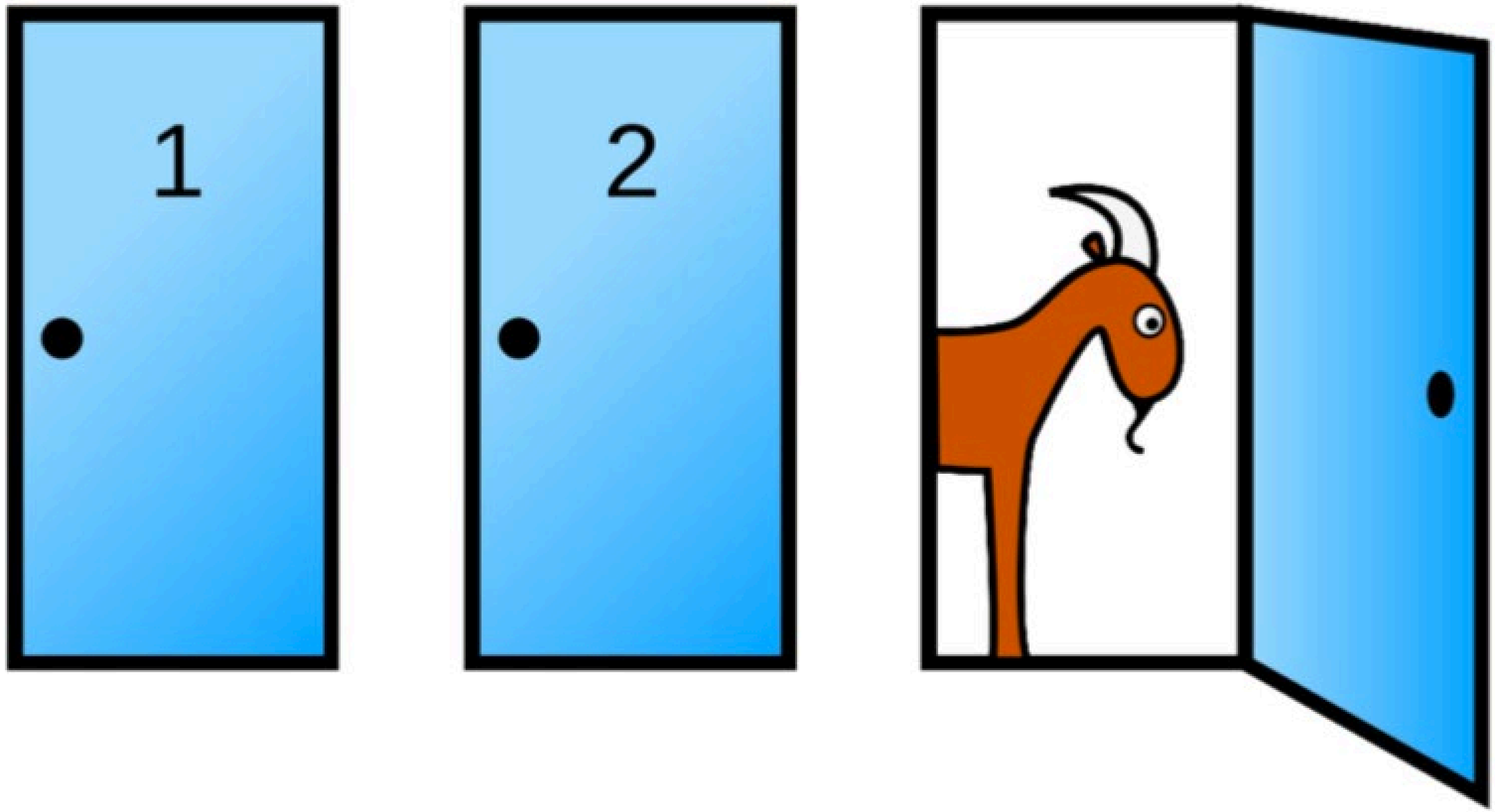
1. Contestant makes an initial choice of door, but the door stays closed
2. One of the other doors with a goat behind it is opened
3. There are two closed doors remaining (one being the contestant's initial). The contestant now gets to choose which of the two doors to open.



What should the contestant do?

- Stick with her initial choice? or
- Switch to the other door?

The Monty Hall problem



Let's explore this in Jupyter!

Probability

The basics

A probability model assigns values to random events

Lowest value: 0

- Chance of event that is impossible

Highest value: 1   (100% chance)

- Chance of event that is certain

The probability an event doesn't occur is 1 minus the probability an even does occur:

- E.g., if there is a 0.7 change an event occurs, then the probability it doesn't occur is:
- 1 - 0.7 = 0.3

Equally likely outcomes

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

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

Example

Suppose there are three tickets: Red, Green, and Blue

When sampling **without replacement**, what's the chance of getting GR? i.e.,

- a green ticket on first draw
- and then a red ticket on second draw?

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

Multiplication rule

The chance that two events A and B both happen is:

= P(A happens) x P(B happens given that A has happened)

When sampling **without replacement**, what's the chance of getting GR?

- RB RG BR BG GR GB = P(GR) = 1/6

P(G) = 1/3

P(R given G) = 1/2

P(GR) = 1/3 x 1/2 = 1/6

Stage 1: 1/3

Stage 2: 1/2

Addition rule

If event A can happen in exactly one of two (mutually exclusive) ways, then:

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

What is the chance of getting a red or a green on a single draw?

$$P(R \text{ or } G) \quad = \quad R \quad G \quad B \quad = \; 2/3$$

$$P(R \text{ or } G) \quad = \quad P(R) \quad + \quad P(G) \quad = \quad 1/3 \quad + \quad 1/3 \quad = \quad 2/3$$

Example

What is the probability of getting at least one head out of *k* coin flips?

In 3 tosses:

- Any outcome except TTT
- P(TTT) = (1/2) x (1/2) x (1/2) = 1/8
- P(at least one head) = 1 - P(TTT) = 7/8 = 87.5%

In 10 tosses:

- $1 - (1/2)^{10}$
- 99.9%

Let's explore this in Jupyter!