Physical Principles in Biology
Biology 3550
Spring 2024

Lecture 4:

Introduction to Randomness and Probability

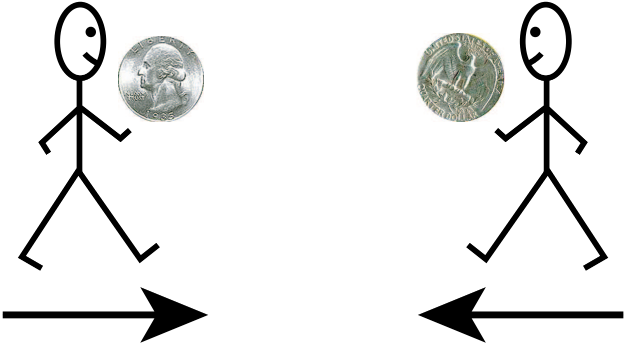
Wednesday, 17 January 2024

University of Utah
goldenberg@biology.utah.edu

Announcements

- Problem set 1:
  - Due 11:59 PM, Tuesday, 23 January.
  - Download problems from Canvas.
  - Upload work to Gradescope.
  - Work must be typed!
- Quiz 1:
  - Friday, 26 January
  - 25 min, second half of class.

A Random Walk in One Dimension



Heads - East

Tails - West

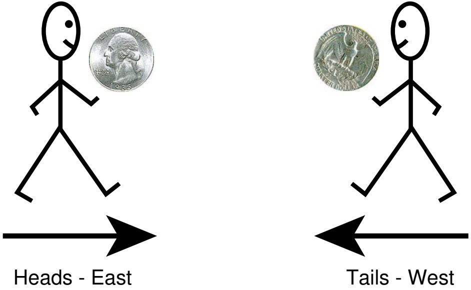
Results of the Coin Toss Experiment

T	T	H	T	T	T	T	T	H	H	H	H	T	H	H	H	H	T	H	H	11
H	T	H	H	H	H	H	T	H	T	T	T	H	H	T	T	T	H	T	T	10
T	H	T	H	T	H	T	T	T	T	H	H	H	H	T	H	T	H	T	T	9
T	T	H	T	H	T	T	H	H	H	T	T	T	H	T	H	T	H	T	H	9
H	T	T	H	H	T	H	H	T	T	T	H	T	H	T	T	T	H	H	T	9
T	T	T	T	H	H	T	H	T	T	H	T	H	H	H	T	H	T	T	H	9
H	T	T	H	H	T	T	H	H	T	T	H	T	H	T	H	T	T	H	T	9
H	H	T	H	H	H	H	T	T	H	T	H	T	T	T	T	H	T	H	H	11
T	H	T	T	H	T	H	H	H	H	H	T	H	T	T	H	H	H	H	T	12
H	T	T	H	T	T	H	T	H	H	T	T	T	H	H	T	T	T	T	H	8
H	H	H	T	T	T	H	H	T	T	H	H	T	H	H	T	T	T	H	T	10
T	T	T	H	H	T	T	H	T	T	T	H	H	T	T	H	H	T	H	T	8

Some Statistics from the Coin Toss Experiment

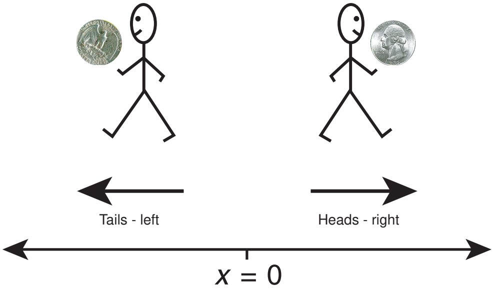
- Total sequences: 15
- Total coin tosses: 300
- Total heads: 141
- Total tails: 159
- Fraction heads: 0.47
- Fraction tails: 0.53

Two Kinds of Distances in a Random Walk



- Total distance traveled: Pedometer distance.
  The number of steps times the length of each step.
- The final displacement from the starting point.

Displacement from the Starting Point for a One-Dimensional Random Walk



- Start at position $x = 0$.
- Take n random steps to the right or left.

  $n_{\mathrm{H}} =$ no. of heads

  $n_{\mathrm{T}} =$ no. of tails

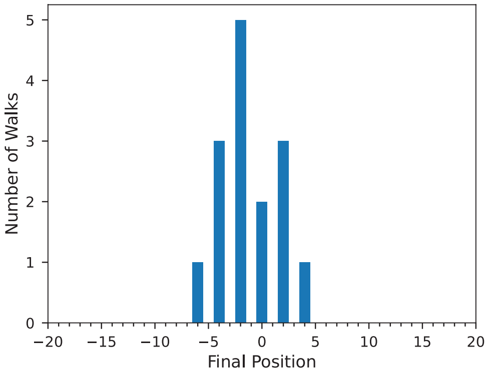
- Final position is x.

  $$x = n_{\mathrm{H}} - n_{\mathrm{T}}$$

  (in units of step lengths)

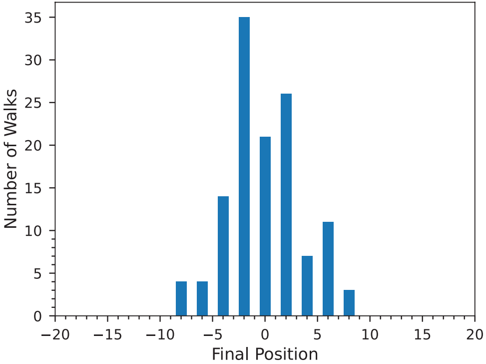
- Generally expect a distribution of x if the random walk is repeated a large number of times.

Distribution of Displacements from Our Coin Tosses



- What patterns can we find?
- We need more data!

Distribution of Displacements from 125 Experiments



- We still need more data!
- Or, a robust theory!

Clicker Question #1

Steve is very shy and withdrawn, invariably helpful but with very little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail.

Is Steve more likely to be a librarian or a farmer?

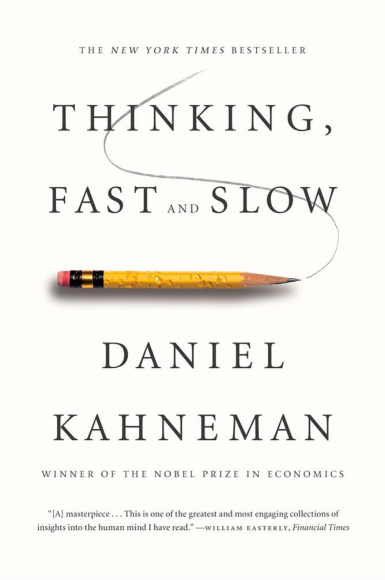
A) Librarian

B) Farmer

All answers count for now!

- Key information: There are about 15-times as many farmers than librarians in the US.
  From the Bureau of Labor Statistics Occupational Outlook Handbook: https://www.bls.gov/ooh

Thinking Fast and Slow



- Kahnemann is a psychologist who won the Nobel Prize in Economics for research on how people make decisions.
- With his collaborator, Amos Tversky, Kahnemann determined that people respond to questions or problems in two distinct ways:
  - Fast, instinctual way that is often right, but sometimes very wrong.
  - Slow, deliberate analysis.
- Highly recommended book!

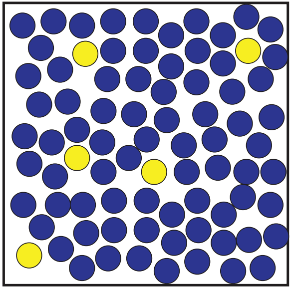
The Fast Way of Answering the Question

- Steve has the characteristics I associate with a librarian: Shy, with a need for order and structure.
- It makes more sense that Steve would be a librarian than a farmer.

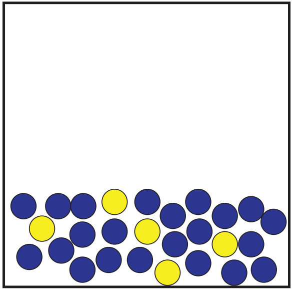
The Slow, Probabilistic Way of Answering the Question

- Steve is one of a large number of people in a population who are either librarians or farmers.
- Represent farmers and librarians as marbles in a box.

All farmers (blue) and librarians (yellow)



Shy, organized farmers and librarians



- If we pick a marble at random, are we more likely to pick a blue marble (farmer) or a yellow one (librarian)?

Probability: Some Definitions

- Outcomes – Possible results of a probabilistic process
  - For a coin toss: coin lands heads-up (H) or tails-up (T)
  - For a roll of a six-sided die: The number of spots on the side that lands up (1, 2, 3, 4, 5 and 6)
  - Distinguished from “events”, to be defined shortly
- Probability – A number with a possible value from 0 to 1, associated with a single outcome.
  - $p = 0$: Outcome will never occur.
  - $p = 1$: Outcome will always occur.
  - The sum of the probabilities of all possible outcomes of an experiment must equal 1.
  - What do we mean by this? What is implied?
How do we interpret probabilities that lie between 0 and 1?

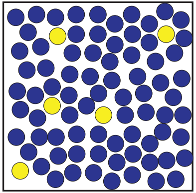
Two Interpretations of Probabilities

1. The frequentist interpretation
   - If the same experiment is repeated a large number, N, of times, an outcome with probability p will occur approximately $N \cdot p$ times.
   - "Law of large numbers"
   - Value of probabilities are defined by properties of the experiment.
2. The Bayesian interpretation
   - Quantity used to express (limited) knowledge or belief.
   - May not be able to determine probability from first principles or experiment.
   - Probability can be updated using additional information.
   - Thomas Bayes (1702-1761): Equation for calculating revised probabilities.
   - Somewhat controversial, but very important area in modern applications of probability and statistics.

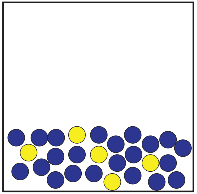
A Bayesian Considers Steve

- Steve is one of a large number of people in a population who are either librarians or farmers, but more are farmers.

All farmers (blue) and librarians (yellow)



Shy, organized farmers and librarians



- Initial information: Relative number of farmers and librarians in the population.
- New information: Steve's personality traits and the average traits of librarians and farmers.
- New information allows us to increase our estimate of the probability that Steve is a librarian.

The Sample Space, S

- Set of all possible outcomes
  - For a single coin toss:
    $S = \{T, H\}$
    Curly braces are used to indicate sets.
  - For two independent coin tosses:
    $S = \{(H, H), (H, T), (T, H), (T, T)\}$
    Ordered pairs representing the results of the two tosses.
- The sample set must be complete: It must include all possible outcomes.
- The elements in the sample set must not overlap.
- The sum of the probabilities for all of the elements in the sample set must equal 1.
- Are there other possible sample sets that could be defined for two coin tosses?

Clicker Question #2

How many outcomes are there in the (simplest) sample set for three coin tosses?

A) 3

B) 4

C) 6

D) 8

E) 10

Three Coin Tosses

- For each coin there are two possible results, the total number of possible results is:

$$2 \times 2 \times 2 = 8$$

- The sample set:

$$\begin{aligned} S = & \{(H, H, H), (H, T, H), (H, H, T), (H, T, T) \\ & (T, H, H), (T, T, H), (T, H, T), (T, T, T)\} \end{aligned}$$

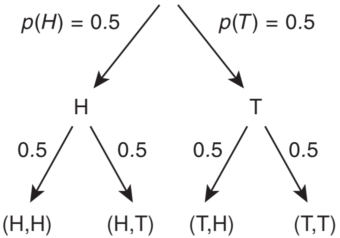
- The order matters!

Events

- An event is a subset of the sample space.
- Some possible events defined for two coin tosses:
  - Two heads: $2H = \{(H, H)\}$
  - Two tails: $2T = \{(T, T)\}$
  - One heads and one tails: $1H1T = \{(H, T), (T, H)\}$
- Each outcome defined in the sample space is an event, but additional events can usually be defined by grouping outcomes together.
- Some other events that can be defined for two coin tosses:
  - One or more heads: $1^{+}H = \{(H, H), (H, T), (T, H)\}$
  - One or more tails: $1^{+}T = \{(H, T), (T, H), (T, T)\}$

Calculating Probabilities: Sequential Trials

- Two coin tosses:



- Probabilities are multiplied

$$p\big((H,H)\big) = p(H)p(H) = 0.5 \times 0.5 = 0.25$$
$$p\big((H,T)\big) = p(H)p(T) = 0.5 \times 0.5 = 0.25$$
$$p\big((T,H)\big) = p(T)p(H) = 0.5 \times 0.5 = 0.25$$
$$p\big((T,T)\big) = p(T)p(T) = 0.5 \times 0.5 = 0.25$$

- Multiplication of probabilities is usually associated with “and”.

Clicker Question #3

What is the probability of three heads in three coin tosses?

A) 0

B) 1/8

C) 1/3

D) 3/8

E) 3/4

The Probability of Three Heads in Three Coin Tosses

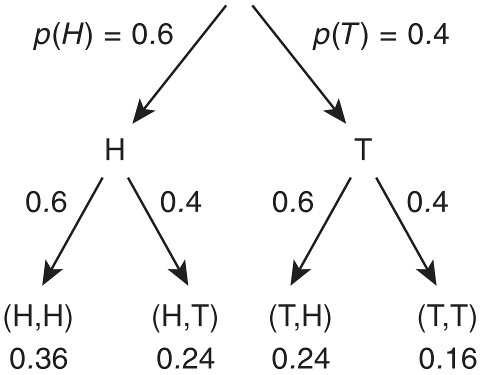
- Three sequential events

$$\xrightarrow{p=0.5} \text{H} \xrightarrow{p=0.5} \text{H} \xrightarrow{p=0.5} \text{H}$$

$$p = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$$

- Consider the sample set:
  - There are 8 elements in the sample set.
  - All outcomes in the sample set have equal probabilities (if the coin is fair).
  - For each outcome, $p = 1/8$.
  - Only one outcome has 3 heads, (H,H,H).

Two Tosses of a Bad Coin



- Sum of probabilities:

$$p\big((H,H)\big) + p\big((H,T)\big) + p\big((T,H)\big) + p\big((T,T)\big) = 1$$