Probability Basics
Part 1: What is Probability?

INFO-1301, Quantitative Reasoning 1

University of Colorado Boulder

September 26, 2016

Prof. Michael Paul

Prof. William Aspray

Variables

We can describe events like coin flips as variables

Domain: {Heads, Tails}

X = Heads

Random variables

What if we haven't flipped the coin yet?

Domain: {Heads, Tails}

A **random variable** is a variable whose value is unknown
(but we know the **probability** of the possible values)

$X = ?$

$P(X = \text{Heads}) = 0.5$
$P(X = \text{Tails}) = 0.5$

Also called a **random process**

Random variables

What if we haven't flipped the coin yet?

Domain: {Heads, Tails}

The domain of a random variable is called the **sample space**

$X = ?$

$P(X = \text{Heads}) = 0.5$
$P(X = \text{Tails}) = 0.5$

The values of random variables are called **outcomes**

Random variables

Random variables can be any variables with unknown values

Confusingly, the outcomes of random variables aren't necessarily random

- the winner of the upcoming election
- the weather tomorrow

  ➢ These examples are unknown and can be treated as random variables with probabilities

What is probability?

"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."

$X = ?$

If we kept flipping a coin forever, half of the outcomes would be heads and half would be tails

This property is known as the **Law of Large Numbers**

What is probability?

Since probabilities correspond to proportions, probabilities are between 0 and 1 (inclusive)

- Or written as a percentage: (0%, 100%)
- Can also be written as a fraction, like ½

A **distribution** is a table of the probabilities of all possible outcomes of a random variable (that is, all values in the sample space)

- The sum of all probabilities in a distribution must equal 1 (or 100%)

What is probability?

What about outcomes that can't happen more than once?

- the probability that Clinton wins the 2016 election
- the probability that I am telling the truth

An alternative way to define probability is as a *degree of belief*

Why probability?

Probability allows us to reason about data
even when it is uncertain

We can predict what will happen in the future and make decisions accordingly

- If you are 95% certain it will rain tomorrow, go ahead cancel your plans
- If you are 55% certain it will rain, you should wait and see what happens

Why probability?

Probability allows us to reason about data
even when it is uncertain

We can estimate long-term tendencies to determine risk

- If you invest in a stock that has a 0.53 probability of increasing in value on any day, then you have a near-equal chance of gaining or losing money on a given day
- But long term, you can expect to gain more than you lose

Central Tendency

You learned about mean, median, and mode as measures of *central tendency* of variables

For random variables, the standard measure of central tendency is the **expected value**

- What do we "expect" the outcome to be?

"the expected value of X" ➡ $\mathbf{E}[X] = \Sigma_x \, P(X = x)\, x$

The probability that *X* has value x times the value x

Sum over all values (denoted 'x') in the sample space

The expected value is equivalent to the *mean* of the outcomes if you repeat a process forever

Central Tendency

Example: Let X be the number of times a coin comes up Heads after 3 flips

$P(X = 0) = 0.125$

$P(X = 1) = 0.375$

$P(X = 2) = 0.375$

$P(X = 3) = 0.125$

This is the **distribution** of X

$$\mathbf{E}[X] = 0 \times 0.125 + 1 \times 0.375 + 2 \times 0.375 + 3 \times 0.125$$
$$= 1.5$$

Central Tendency

If you take the average of multiple outcomes of a random variable, the average will most often be close to the expected value

This is proven by the **Central Limit Theorem**

- More formally, the theorem states that if you take the average of multiple random outcomes multiple times, the averages will form a bell curve where the mean is the expected value of that random variable