Tables and Probability (and You)

Grinnell College

September 11, 2026

Probability

We will be concerning ourselves with *outcomes* associated with *random processes*

Probability of an *outcome* is the proportion of times the outcome would occur if we observed the *random process* an infinite number of times

Simple examples include:

- Flipping a coin
- Rolling a dice
- Sampling marbles from a jar
- Drawing a card from a deck

A set of all possible outcomes, denoted $\mathcal{S}$, is called a **sample space**.

Consider rolling a dice, where the set of possible outcomes is

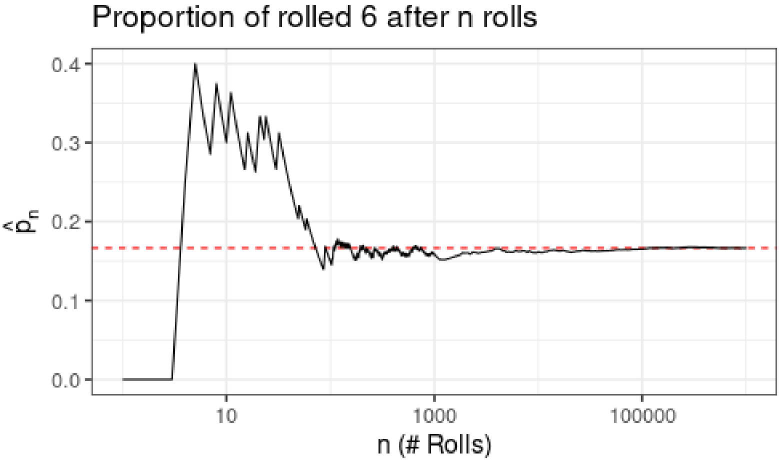
$$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$$

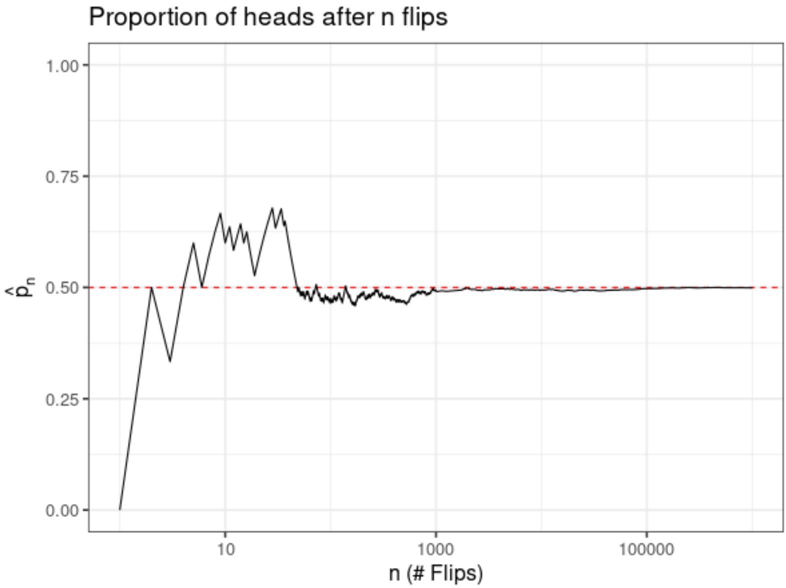
We express probability of an outcome as such

$$P(\text{rolling a 4}) = \frac{1}{6}$$

It's worth asking ourselves: *how can we know the probability of rolling a four is equal to* $\frac{1}{6}$*?*

As more observations are collected (n increases), the size of fluctuations of p_n around p will begin to shrink. This tendency to stabilize is known as the **Law of Large Numbers**



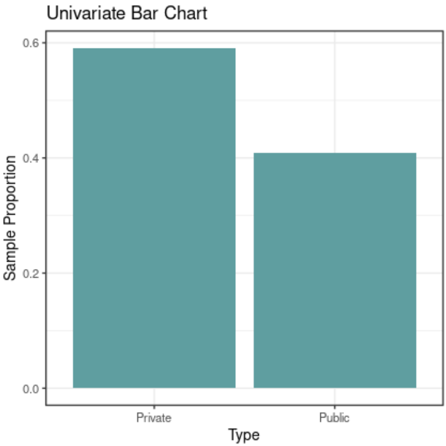

Proportion of heads after n flips
1.00
0.75
0.50
0.25
0.00
$\hat{p}_n$
10
1000
100000
n (# Flips)

Sample Proportions

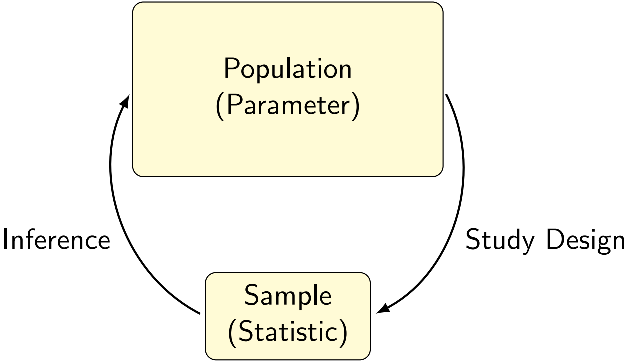
We can consider the true proportion of outcomes for each category to be a **parameter**.

By collecting a sample, we can find a **statistic** that estimates this parameter

These statistics allow me to estimate the true **distribution** of a variable



Statistics


Population
(Parameter)
Inference
Study Design
Sample
(Statistic)

Probabilities and distributions

This brings us to a trio of definitions:

The **marginal probability** describes the probability of a *single* variable without regard to others, e.g., the probability of event A is $P(A)$

The **joint probability** describes the probabilities for two or more variables together, e.g., the probability of events A and B both is $P(A \text{ and } B)$

Conditional probability describes the probability of one event based on the assumed outcome of another, e.g., the conditional probability of event A *given* B is denoted $P(A|B)$

Example

Suppose we took a collection of 9 individuals, giving 5 of them a placebo and 4 of them an active drug. We then recorded how many from each group were sick once in the next 100 days

	Sick	Not Sick	Total
Placebo	3	2	5
Drug	1	3	4
Total	4	5	9

Marginal Probabilities

	Sick	Not Sick	Total
Placebo	3	2	5
Drug	1	3	4
Total	4	5	9

From this, we may ask ourselves: what is the probability that a randomly selected person is sick?

$$P(\text{Sick}) = \frac{\#\text{ of sick people}}{\text{Total }\#\text{ of people}} = \frac{4}{9} = 0.44$$

Meanwhile, the probability that a randomly selected person is *not* sick would be

$$P(\text{Not Sick}) = \frac{\#\text{ of not sick people}}{\text{Total }\#\text{ of people}} = \frac{5}{9} = 0.56$$

Each of these represents as **marginal probability**. All together, these marginal probabilities make up the **marginal distribution** for the variable `Sick`

Joint Probabilities

	Sick	Not Sick	Total
Placebo	3	2	5
Drug	1	3	4
Total	4	5	9

Or we may ask: what is the probability that a randomly selected person is sick and has been given a placebo?

$$\begin{aligned} P(\text{Is sick and got placebo}) &= \frac{\#\text{ of sick people with placebo}}{\text{Total }\#\text{ of people}} \\ &= \frac{3}{9} = 0.33 \end{aligned}$$

Just as with marginal probabilities, the **joint probabilities**, all taken together, make up the **joint distribution** of both variables

Marginal and Joint Together

We can derive the marginal and joint probabilities of variables by dividing both the cells of the table, as well as the margins, by the total number of observations. In this case, $n = 9$

	Sickness		
	Sick	Not Sick	Marginal Prob
Placebo	0.33	0.22	0.55
Drug	0.11	0.33	0.45
Marginal Prob	0.44	0.56	1

Conditional Probability

When looking at **conditional probabilities**, we limit ourselves to the row that satisfy the conditional

	Sick	Not Sick	Total
Placebo	3	2	5
Drug	1	3	4
Total	4	5	9

Does knowing that a person received a placebo change our estimate of the probability that they were sick

$$
\begin{aligned}
P(\text{Is sick } \textit{given} \text{ placebo}) &= P(\text{Sick} \mid \text{Placebo}) \\
&= \frac{\#\text{ of sick people with placebo}}{\text{Total } \#\text{ given placebo}} \\
&= \frac{3}{5} \\
&= 0.6
\end{aligned}
$$

Conditional Probability

However, we can also compute this same thing from *marginal* and *joint* probabilities

	Actual Photo		
	Sick	Not Sick	Marginal Prob
Placebo	0.33	0.22	0.56
Drug	0.11	0.33	0.44
Marginal Prob	0.44	0.56	1

$$
\begin{aligned}
P(\text{Is sick } \textit{given} \text{ placebo}) &= P(\text{Sick} \mid \text{Placebo}) \\
&= \frac{P(\text{Sick and Placebo})}{P(\text{Placebo})} \\
&= \frac{0.33}{0.55} \\
&= 0.6
\end{aligned}
$$

Association and Independence

One way we can assess whether two categorical variables are associated is by comparing the marginal and conditional distributions

For example, suppose that some event has occurred (say, receiving a drug), and we want to assess the probability that a second event has also occurred (say, not being sick)

If two events A and B are **independent**, then the fact B has occurred will have no impact on the probability of A, and we will find that

$$P(A|B) = P(A)$$

Two events that are not independent are said to be associated

Change note

Should very clearly have an example here showing independence using tables

Summary

Key Ideas:

- Probability and proportion
- Law of Large Numbers
- Marginal, joint, and conditional probabilities
- Association and independence