# Definition of Probability

What does it mean when someone makes a claim like "the probability that you find a pearl in an oyster is 1 in 5,000?" or "the probability that it will rain tomorrow is 52%"?

## Events and Experiments

When we speak about probabilities, there is always an implied context, which we formally call the "experiment". For example: flipping two coins is something that probability folks would call an experiment. In order to precisely speak about probability, we must first define two sets: the set of all possible outcomes of an experiment, and the subset that we consider to be our event ([what is a set?](#)).

---

**Definition: Sample Space, $S$**

A Sample Space is the set of all possible outcomes of an experiment. For example:

- Coin flip: $S$ = {Heads, Tails}
- Flipping two coins: $S$ = {(H, H), (H, T), (T, H), (T, T)}
- Roll of 6-sided die: $S$ = {1, 2, 3, 4, 5, 6}
- The number of emails you receive in a day: $S = \{x | x \in \mathbb{Z}, x \geq 0\}$ (non-neg. ints)
- YouTube hours in a day: $S = \{x | x \in \mathbb{R}, 0 \leq x \leq 24\}$

---

**Definition: Event, $E$**

An Event is some subset of $S$ that we ascribe meaning to. In set notation ($E \subseteq S$).For example:

- Coin flip is heads: $E$ = {Heads}
- At least 1 head on 2 coin flips = {(H, H), (H, T), (T, H)}
- Roll of die is 3 or less: E = {1, 2, 3}
- You receive less than 20 emails in a day: $E = \{x | x \in Z, 0 \leq x < 20\}$ (non-neg. ints)
- Wasted day ($\geq 5$ YouTube hours): $E = \{x | x \in R, 5 \leq x \leq 24\}$

Events can be represented as capital letters such as $E$ or $F$.

---

[todo] In the world of probability, events are binary: they either happen or they don't.

## Definition of Probability

It wasn't until the 20th century that humans figured out a way to precisely define what the word probability means:

$$P(\text{Event}) = \lim_{n \to \infty} \frac{\text{count}(\text{Event})}{n}$$

In English this reads: lets say you perform $n$ trials of an "experiment" which could result in a particular "[Event](#)" occurring. The probability of the event occurring, $P(\text{Event})$, is the ratio of trials that result in the event, written as $\text{count}(\text{Event})$, to the number of trials performed, $n$. In the limit, as your number of trials approaches infinity, the ratio will converge to the true probability. People also apply other semantics to the concept of a probability. One common meaning ascribed is that $P(E)$ is a measure of the chance of event E occurring.

---

***Example***: Probability in the limit

Here we use the definition of probability to calculate the probability of event $E$, rolling a "5" or a "6" on a fair [six-sided dice](#). Hit the "Run trials" button to start running trials of the experiment "roll dice". Notice how $P(E)$, converges to 2/6 or 0.33 repeating.

Event $E$: Rolling a 5 or 6 on a six-sided dice.

Dice outcome:

$n = 0$

$count(E) = 0$

$P(E) \approx \frac{count(E)}{n} =$

Observed Fraction

1.0

0.5

0

E          not E

True Probability

0.8
0.6
0.4
0.2
0

E          not E

**Measure of uncertainty**: It is tempting to think of probability as representing some natural randomness in the world. That might be the case. But perhaps the world isn't random. I propose a deeper way of thinking about probability. There is so much that we as humans don't know, and probability is our robust language for expressing our belief that an event will happen given our limited knowledge. This interpretation acknowledges your own uncertainty of an event. Perhaps if you knew the position of every water molecule, you could perfectly predict tomorrow's weather. But we don't have such knowledge and as such we use probability to talk about the chance of rain tomorrow given the information that we have access to.

**Origins of probabilities**: The different interpretations of probability are reflected in the many origins of probabilities that you will encounter in the wild (and not so wild) world. Some probabilities are calculated analytically using mathematical proofs. Some probabilities are calculated from data, experiments or simulations. Some probabilities are just made up to represent a belief. Most probabilities are generated from a combination of the above. For example, someone will make up a prior belief, that belief will be mathematically updated using data and evidence. Here is an example of calculating a probability from data:

**Probabilities and simulations**: Another way to compute probabilities is via simulation. For some complex problems where the probabilities are too hard to compute analytically you can run simulations using your computer. If your simulations generate believable trials from the sample space, then the probability of an event E is approximately equal to the fraction of simulations that produced an outcome from E. Again, by the definition of probability, as your number of simulations approaches infinity, the estimate becomes more accurate.

**Probabilities and percentages**: You might hear people refer to a probability as a percent. That the probability of rain tomorrow is 32%. The proper way to state this would be to say that 0.32 is the probability of rain. Percentages are simply probabilities multiplied by 100. "Percent" is latin for "out of one hundred".

*Problem:* Use the definition of probability to approximate the answer to the question: "What is the probability a new-born elephant child is male?" Contrary to what you might think the gender outcomes of a newborn elephant are not equally likely between male and female. You have data from a report in Animal Reproductive Science which states that 3,070 elephants were born in Myanmar of which 2,180 were male [1]. Humans also don't have a 50/50 sex ratio at birth [2].

*Answer:* The Experiment is: A single elephant birth in Myanmar.

The sample space is the set of possible sexes assigned at birth, {Male, Female, Intersex}.

$E$ is the event that a new-born elephan child is male, which in set notation is the subset {Male} of the sample space. The outcomes are not equally likely.

By the definition of probability, the ratio — of trials that result in the event, to the total number of trials — will tend to our desired probability:

$$P(\text{Born Male}) = P(E)$$
$$= \lim_{n \to \infty} \frac{\text{count}(E)}{n}$$
$$\approx \frac{2,180}{3,070}$$
$$\approx 0.710$$

Since 3,000 is quite a bit less than infinity, this is an approximation. It turns out, however, to be a rather good one. A few important notes: there is no guarantee that our estimate applies to elephants outside Myanmar. Later in the class we will develop language for "how confident we can be in a number like 0.71 after 3,000 trials?" Using tools from later in class we can say that we have 98% confidence that the true probability is within 0.02 of 0.710.

## Axioms of Probability

Here are some basic truths about probabilities that we accept as axioms:

| | |
|---|---|
| **Axiom 1**: $0 \leq P(E) \leq 1$ | All probabilities are numbers between 0 and 1. |
| **Axiom 2**: $P(S) = 1$ | All outcomes must be from the Sample Space. |
| **Axiom 3**: If $E$ and $F$ are mutually exclusive, then $P(E \text{ or } F) = P(E) + P(F)$ | The probability of "or" for mutually exclusive events |

These three axioms are formally called the Kolmogorov axioms and they are considered to be the foundation of probability theory. They are also useful identities!

You can convince yourself of the first axiom by thinking about the math definition of probability. As you perform trials of an experiment it is not possible to get more events than trials (thus probabilities are less than 1) and its not possible to get less than 0 occurrences of the event (thus probabilities are greater than 0). The second axiom makes sense too. If your event is the sample space, then each trial must produce the event. This is sort of like saying; the probability of you eating cake (event) if you eat cake (sample space that is the same as the event) is 1. The third axiom is more complex and in this textbook we dedicate an entire chapter to understanding it: Probability of or. It applies to events that have a special property called "mutual exclusion": the events do not share any outcomes.

These axioms have great historical significance. In the early 1900s it was not clear if probability was somehow different than other fields of math -- perhaps the set of techniques and systems of proofs from other fields of mathematics couldn't apply. Kolmogorov's great success was to show to the world that the tools of mathematics did in fact apply to probability. From the foundation provided by this set of axioms mathematicians built the edifice of probability theory.

## Provable Identities

We often refer to these as corollaries that are directly provable from the three axioms given above.

| | |
|---|---|
| **Identity 1**: $P(E^C) = 1 - P(E)$ | The probability of event E not happening |
| **Identity 2**: If $E \subseteq F$, then $P(E) \leq P(F)$ | Events which are subsets |

This first identity is especially useful. For any event, you can calculate the probability of the event *not* occurring which we write in probability notation as $E^C$, if you know the probability of it occurring -- and vice versa. We can also use this identity to show you what it looks like to prove a theorem in probability.

**Proof:** $P(E^C) = 1 - P(E)$

$$P(S) = P(E \text{ or } E^C) \qquad E \text{ or } E^C \text{ covers every outcome in the sample space}$$
$$P(S) = P(E) + P(E^C) \qquad \text{Events } E \text{ and } E^C \text{ are mututally exclusive}$$
$$1 = P(E) + P(E^C) \qquad \text{Axiom 2 of probability}$$
$$P(E^C) = 1 - P(E) \qquad \text{By re-arranging}$$