

# Introduction to Probability

## Experiments with random outcomes

# What is a “probability”?

**Probability** is an attempt to quantify the idea that an event may occur.

*An attempt.*

For this attempt to be mathematically rigorous, we need to define several terms precisely, and grossly restrict our playing field.

# What is a “probability”?

“[A]n event may occur” will mean that, in some **random experiment** in an observed setting (a controlled setting or not - the experiment could be “flip a coin” or “Google reports an increase in profit”), there is a “measure” out of the possible events, that this one will occur.

Certainly, the controlled setting is easier to codify and observe, and our examples of coin flips, die rolls, card deals, etc. come from this comfort level and ease of computation.

# Definition of probability

## Definition

The **sample space** of an experiment is a set.  
(Most texts call this set  $\Omega$  or  $S$ .)

The elements in a sample space are called **outcomes** or **sample points**. (We usually refer to outcomes by  $\omega$ ,  $s$ , or  $x$ .)

A subset of outcomes is called an **event**.  
We often refer to events by  $E$ ,  $F$ , etc. Note that

$$E \in \mathcal{F} \subseteq 2^\Omega \implies E \subseteq \Omega.$$

# Definition of $\sigma$ -algebra of sets

Let  $\Omega$  be the sample space of an experiment.

A  $\sigma$ -**algebra** of subsets  $\mathcal{F}$  of  $\Omega$  satisfies three properties:

- ▶  $\Omega \in \mathcal{F}$
- ▶  $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
- ▶  $A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$  (the  $\sigma$ -property).

Note that  $\mathcal{F} \subseteq 2^{\Omega}$ , and  $\emptyset \in \mathcal{F}$ .

# Definition of probability measure

For a sample space  $\Omega$  and a  $\sigma$ -algebra  $\mathcal{F}$  of subsets of  $\Omega$ , a **probability measure** (or, simply, **probability**) is a function

$$P : \mathcal{F} \rightarrow \mathbb{R}$$

that satisfies certain properties called the *axioms of probability*.

We call the triple  $(\Omega, \mathcal{F}, P)$  a **probability space**, since this triple precisely describes the outcomes, events, and probabilities of the random experiment(s) being considered.

# Kolmogorov's axioms of probability

1.  $\forall E \in \mathcal{F}, 0 \leq P(E) \leq 1$   
(probability is a percentage between 0% and 100%)
2.  $P(\Omega) = 1$  (and so  $P(\emptyset) = 0$ )  
(normalization: probability of “anything happening” is 1)
3. If  $E_1, E_2, \dots$  are an infinite sequence of **pairwise disjoint** events, that is,  $E_i \cap E_j = \emptyset$  for any  $i, j = 1, 2, 3, \dots$ , then

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

(countable additivity: probability of a bunch of different things happening is the sum of those things' individual probabilities)

# Examples in Discrete Probability

Our first two examples of sample spaces and events are examples in

**discrete probability,**

where the methods of calculating probabilities are *combinatorial*,  
i.e. deal with counting.



## Example: fair coin flips

The most basic example of a random experiment is a *fair coin flip*.

$$\Omega = \{H, T\}; P(\{H\}) = \frac{1}{2}, P(\{T\}) = \frac{1}{2}.$$

A more complex example is infinite sequences of fair coin flips:

$$\Omega = \{f_1 f_2 f_3 \dots : f_i \in \{H, T\}, i = 1, 2, 3, \dots\}$$

What is the probability that the third flip in a sequence is H?

$$P(\{**H***\dots\}) = \frac{1}{2}$$

since you're only considering one fair flip from the infinite sequence.

## Example: die rolls

Experiment: roll two six-sided dice.

What is the sample space? It consists of ordered pairs

$$\Omega = \left\{ (i, j) : i, j \in \{1, 2, 3, 4, 5, 6\} \right\}.$$

How many possible outcomes are there in this experiment?

## Example: die rolls

Each die can roll 6 different ways, and there are 2 independent dice, so the total possible outcomes, by the Multiplication Principle, count up to  $|\Omega| = 6^2 = 36$ .

What is the event that you roll a 5 on the second die?

$$\begin{aligned} E &= \text{"5 on 2nd die"} \\ &= \{(i, 5) : i \in \{1, 2, 3, 4, 5, 6\}\} \\ &= \{(1, 5), (2, 5), (3, 5), (4, 5), (5, 5), (6, 5)\}. \end{aligned}$$

## Example: die rolls

How many outcomes are in  $E$ ? We can easily count  $|E| = 6$ .

Therefore, a good first guess for the question

“What is the probability of rolling a 5 on the second die?”

is the ratio (fraction)

$$P(E) = \frac{6}{36} = \frac{1}{6},$$

and we notice that this matches the probability of just rolling a 5 on one die (ignoring the first die).

# Properties of a probability

Some properties of all probability functions:

- ▶ **(something happens in an experiment)**  $P(\emptyset) = 0$
- ▶ **(complements)**  $P(E^C) = 1 - P(E)$  for any event  $E$ .
- ▶ **(finite additivity)** For events  $E_1, \dots, E_n$  mutually disjoint,

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i).$$

(Hint: countable additivity with  $E_{n+1} = E_{n+2} = \dots = \emptyset$ .)

- ▶ **(monotonicity)** If  $E \subseteq F$ , then  $P(E) \leq P(F)$ .

# Inclusion-Exclusion Principle

Recall the **inclusion-exclusion principle** of cardinality of sets: if  $E, F \subseteq \Omega$  for some finite-sized universe  $\Omega$ , then

$$|E \cup F| = |E| + |F| - |E \cap F|.$$

This applies directly to probabilities: if  $E, F$  are events in some experiment on the same sample space, then

$$P(E \cup F) = P(E) + P(F) - P(EF).$$

We can use Venn diagrams, similarly to the way we used them to count elements in sets, to figure out probabilities.

# Inclusion-Exclusion Principle

Also, just as before, inclusion-exclusion generalizes: here it is for three events,  $E$ ,  $F$ ,  $G$ :

$$\begin{aligned}P(E \cup F \cup G) = & P(E) + P(F) + P(G) \\& - P(EF) - P(EG) - P(FG) \\& + P(EFG).\end{aligned}$$

# Examples

## Example

Let's say the probability that it will rain on Wednesday is 34%, and the probability that I will have a sandwich on Wednesday is 75%.

The probability that at least one of those things happens on Wednesday is 90%.

What is the probability that both happen?



## Example: rain and sandwich?

Let  $E$  = “it will rain on Wednesday” and  
 $F$  = “I will have a sandwich on Wednesday”.

We are given  $P(E) = 0.34$ ,  $P(F) = 0.75$ , and  $P(E \cup F) = 0.90$ .  
Hence,

$$P(EF) = P(E) + P(F) - P(E \cup F) = 0.34 + 0.75 - 0.90 = 0.19.$$

The probability it will rain *and* I'll have a sandwich on Wed is 19%.

# Boole's inequality

The inclusion-exclusion principle yields an inequality called **Boole's inequality**: for any events  $E, F$ ,

$$P(E \cup F) \leq P(E) + P(F).$$

Why? Because  $0 \leq P(E \cap F) \leq 1$ .

This generalizes, just like inclusion-exclusion, and motivates the countable additivity axiom and finite additivity result: for any events  $E_1, E_2, \dots, E_n$ ,

$$P\left(\bigcup_{i=1}^n E_i\right) \leq \sum_{i=1}^n P(E_i),$$

with equality if the events are pairwise disjoint (finite additivity).

# But really... *what is* a “probability”? What does it *mean*?

There are various *interpretations* as to the *meaning* of **probability** (as opposed to the mathematical treatment we are attempting, which ascribes no *meaning*, only a reasonable methodology).

There are several ways that the notion of probability is interpreted, which can affect how they can be computed, used, or ignored.

# Classical probability = uniform sample space

The first, most common in the simple, discrete experiment setting, is often called *classical probability*.

In this setting, if the sample space  $\Omega$  is finite, and all outcomes are considered “equally likely”, then the probability of an event  $E$  on  $\Omega$  is simply

$$P(E) = \frac{|E|}{|\Omega|},$$

the ratio of the size of  $E$  to the size of  $\Omega$ .

This is the type of probability we use in all our game-based probabilities where the game is *fair* in the sense that, say, rolling a 4 on a die is just as likely as rolling a 6 or a 1.

# Classical probability = equally likely outcomes

**Classical probability** allows for immediate calculation of probabilities in many simple games of chance and other simple experiments where “equally likely outcomes” seems a very reasonable idea.

- ▶ coin flip:  $\Omega = \{H, T\}$ , each outcome with probability  $\frac{1}{2}$
- ▶ 1D6 roll:  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , each with probability  $\frac{1}{6}$
- ▶ drawing a single card:

$$\Omega = \{(x, y) : x \in \{A, 2, 3, \dots, 9, 10, J, Q, K\}, y \in \{\heartsuit, \clubsuit, \diamondsuit, \spadesuit\}\},$$

each card drawn with probability  $\frac{1}{52}$ .

However, there is a severe drawback: *why* is it “reasonable” to think all these outcomes are “equally likely”?

# Frequency probability = statistical probability

An alternative way to define these is the **relative frequency** approach, appreciated by statisticians, of counting the number of times each outcome appears in *repeated*, independent trials of the experiment.

Then, we have a real, physical reason why one would believe the values used are “reasonable”.

Define  $n(E)$  as the number of times  $E$  occurs in  $n$  trials. Then

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}.$$

Problems with this approach:

- ▶ Can you run an *infinite* number of trials to calculate this limit?
- ▶ Why would we believe this limit converges, and doesn't oscillate, or fall to 0? How can we know?

# Frequency probability = statistical probability

Here is a sample run of 500 simulated die rolls in Python:

```
12133263265413165615566136521153126335632345335633
55632323651662342164564526446512243365664155364433
5362531166223525252255566422135353514323162532255
42151624543133435444566624135425265345335246414335
31352621152423146255615656121522542544146614364313
42515132113454114164651434414444422443423611326611
31211514632114162243326216256431266535434432126266
31414366462326234661612316355316135522416413334246
55641421424361155664326115163246455326531366332551
12635652262563455261636516444134116123425335552146
```

There are 88 sixes in this run. Using this run as a frequency experiment, I could say  $P(\text{roll a 6 in Python}) \approx \frac{88}{500} = 0.176$ , which is getting close to  $\frac{1}{6} = 0.1\overline{6}$ ...

# Law of Large Numbers

The **Law of Large Numbers** states that, as the number of independent, identically-run experimental trials  $n \rightarrow \infty$ , the *sample frequency ratio*  $\frac{n(E)}{n}$  converges to a “*true probability*”, which we will call  $P(E)$ .

So, there is a mathematical reason to believe that  $P(E)$  exists as a limit, but not necessarily that we can *compute* it. Also, while we can try to make sure all trials are “identical”, physical constraints make that issue difficult.

However, we can approximate with a large number of trials, and trust experimental design and tools, which is good enough for most, and really all we can do if we want results.



# Subjective probability = “measure of belief”

We call the type of probabilities coming from intuition, hunches, made up numbers and whatnot “subjective probability”.

When the mathematical tools of probability theory are used in making public policy decisions, life-affecting medical choices, large gambles on Wall St., it is important to understand where the numbers being plugged into those decisions are coming from: history, study, or whim?

Unfortunately, most people do not understand basic probability or statistics. (We may see particularly controversial examples when we reach Bayes' Law.)

## Uniform probability (finite sample space)

If the sample space of an experiment is  $\Omega = \{1, 2, \dots, N\}$ , then the **uniform (classical) probability** of any individual outcome, i.e. the event  $E_i = \{i\}$ , is

$$P(E_i) = P(\{i\}) = \frac{1}{N}.$$

In general, for any event  $E \subset \Omega$ , if  $P$  is the “equally likely” (i.e. *classical*) probability, then

$$P(E) = \frac{|E|}{|\Omega|} = \frac{\# \text{ of outcomes in } E}{\# \text{ possible outcomes in } \Omega} = \frac{|E|}{N}.$$

Obviously, then, counting is the name of the game here.

# Coin flips

## Example

For a “fair” coin flip,  $\Omega = \{H, T\}$ , and  $P(\{H\}) = P(\{T\}) = \frac{1}{2}$ .

## Example

For a sequence of three “fair” coin flips,

$$\Omega = \{H, T\}^3 = \{H, T\} \times \{H, T\} \times \{H, T\}.$$

What is the probability of getting three tails in a row: the sequence TTT?

$$P(\{TTT\}) = \frac{1}{2^3} = \frac{1}{8}.$$

# Die rolls

## Example

Two fair dice (2D6) are rolled. What is  $P(\{\text{sum is 7}\})$ ?

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\},$$

ordered pairs of individual die rolls: (3,5) and (5,3) are two ways to roll a 3 and a 5. Thus,  $|\Omega| = 6^2 = 36$ .

Next, note that our event is

$$E = \{(i, j) \in \Omega : i+j = 7\} = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

There are  $|E| = 6$  possible outcomes for this event. Thus,

$$P(E) = \frac{|E|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}.$$

# Venn diagrams can help in uniform probability

## Example

In an old survey of 75 college students on their reading of three magazines, it was found that:

- ▶ 23 read Time (and possibly more)
- ▶ 18 read Newsweek (and possibly more)
- ▶ 14 read US News (and possibly more)
- ▶ 10 read Time and Newsweek (and possibly more)
- ▶ 9 read Time and US News (and possibly more)
- ▶ 8 read Newsweek and US News (and possibly more)
- ▶ 5 read all three.

Question: If a student is selected at random from the survey to win a subscription, what is the probability they read none of the above?

# Balls in urns: without replacement

“Balls in urns” is a typical type of counting / probability problem when asking about indistinguishable items (balls).

## Example

Say I have an urn with 6 blue and 5 yellow balls in it. I draw 3 balls at the same time, *without replacement* (I keep all 3 out). What is the probability that I draw 1 blue and 2 yellow?

$\Omega$  is the space of all possible ball draws *without replacement*. First, we'll consider this without ordering. There are 11 items total, and I am drawing 3 of them. How many ways are there to do that?

$$|\Omega| = \binom{11}{3} = 165.$$

# Balls in urns: without replacement: without order

The event in question is  $E = \{ 1 \text{ blue, } 2 \text{ yellow drawn} \}$ .

How many ways are there to do this?

There are 6 blue and 5 yellow, so there are

$$|E| = \binom{6}{1} \binom{5}{2} = 60$$

ways. Thus, the probability of this event is

$$P(E) = \frac{\binom{6}{1} \binom{5}{2}}{\binom{11}{3}} = \frac{60}{165} = \frac{4}{11}.$$

## Balls in urns: without replacement: with order

Now, we'll consider the problem with ordering, i.e. draw the first ball, then second, then third, instead of all three at once.

With ordering, the number of possible draws is

$$|\Omega| = P(11, 3) = \frac{11!}{8!} = 11 \cdot 10 \cdot 9 = 990.$$



## Balls in urns: without replacement: with order

There are 3 orders in  $E = \{ 1 \text{ blue, 2 yellow drawn} \}$  to consider:

$$BYY, YBY, YYB.$$

There are

- ▶  $6 \cdot 5 \cdot 4 = 120$  ways to draw  $BYY$ ,
- ▶  $5 \cdot 6 \cdot 4 = 120$  ways to draw  $YBY$ , and
- ▶  $5 \cdot 4 \cdot 6 = 120$  ways to draw  $YYB$ .

Therefore,  $|E| = 360$ , and so

$$P(E) = \frac{360}{990} = \frac{4}{11}.$$

# Balls in urns: with replacement

## Example

Same urn, different experiment: draw 3 balls, *with replacement* (draw 1, note its color, put it back in the urn, repeat).

Now, what is the probability of drawing 1 blue and 2 yellow?

$\Omega$  is now the space of all possible draws *with replacement*.

With replacement, each draw comes from 11 balls, so  $|\Omega| = 11^3$ .

With replacement, there are  $\binom{3}{1} = 3$  positions the blue ball could show up (in ordering): BYY, YBY, YYB. Thus,

$$E = \{BYY, YBY, YYB\}, \text{ so } |E| = 3(6)(5)(5).$$

Therefore,

$$P(E) = \frac{3(6)(5)(5)}{11^3} = \frac{450}{1331}.$$

## Preview: Binomial Random Variables

Say  $X$  = the number of blue balls drawn in the example.

Then the event  $E = \{ \text{exactly 1 blue drawn} \}$  can be written

$$E = \{X = 1\}.$$

$X$  is an example of a **binomial random variable**.

What are the probabilities of each different value of  $X$  showing up?

Think about the binomial theorem.

## Preview: Binomial Random Variables

As before, the total number of possible draws with replacement is  $|\Omega| = 11^3$ . Then, the possible cases for  $X$  are

$$\begin{aligned}P(X = 0) &= \frac{\binom{3}{0} 6^0 5^3}{11^3} & P(X = 1) &= \frac{\binom{3}{1} 6^1 5^2}{11^3} \\P(X = 2) &= \frac{\binom{3}{2} 6^2 5^1}{11^3} & P(X = 3) &= \frac{\binom{3}{3} 6^3 5^0}{11^3}\end{aligned}$$

which total up to, by the binomial theorem,

$$\sum_{i=0}^3 \frac{\binom{3}{i} 6^i 5^{3-i}}{11^3} = \frac{(6+5)^3}{11^3} = 1.$$

# Random variables

In general, a **random variable**  $X$  is a function from a sample space  $\Omega$  into the real numbers  $\mathbb{R}$ :

$$X : \Omega \rightarrow \mathbb{R}.$$

Thus, if  $E$  is an event describing certain values of  $X$ , we can write the event in terms of a **pre-image** of  $X$ .

For example, the event

$$\{X \in \{2, 5\}\} = \{\omega \in \Omega : X(\omega) \in \{2, 5\}\} = X^{-1}(\{2, 5\}).$$

# Infinitely many outcomes? They cannot be “equally likely”.

Notice that, for a probability to have equally likely outcomes, there must only be a finite number of outcomes possible.

## Example

Flip a fair coin until tails appears. What is the probability distribution of the number of flips it takes until this happens?

The sample space  $\Omega = \{\infty, 1, 2, 3, \dots\}$  are the possible flip count outcomes. We put  $\infty$  in the sample space because there is an infinitely-long sequence:

*HHHHHH...*

where T *never* shows up. (We will see that the probability of this happening is 0.)

# How many flips until T appears?

Let  $Y$  = the number of flips counted (another example of a **random variable**).

What is the probability  $P(Y = 4)$ , that it takes 4 flips for T to show up? That is, what is the probability of getting the event

$$\{Y = 4\} = \{\omega \in \Omega : Y(\omega) = 4\} = \{HHHT\}?$$

$$P(HHHT) = \frac{1}{2^4} = \frac{1}{16}.$$

## How many flips until T appears? Geometric distribution.

In general,  $P(Y = k) = \left(\frac{1}{2}\right)^k$  for  $k = 1, 2, 3, \dots$ .  $Y$  is called a **geometric random variable**, and as the geometric series yields

$$\sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = 1,$$

this accounts for all probability in this experiment; thus,

$$P(Y = \infty) = 0.$$



# Discrete random variable, probability mass function (pmf)

A random variable is called **discrete** if there is a countable or finite set of values it may take with positive probability.

The function  $p_X$  defined by

$$p_X(k) = P(X = k)$$

is called the **probability mass function (pmf)** of  $X$ .

If  $\{k_1, k_2, k_3, \dots\}$  is the set of all possible values of  $X$ , then

$$\sum_{k=1}^{\infty} p_X(k_i) = 1.$$

We will see more details of pmfs later in the semester.

# Uncountably many outcomes? We need intervals.

If the sample space of an experiment has *uncountably* many outcomes, our distribution will not be summable as a series.

In this case, as we will see with **continuous random variables** later on, you will need to check the probability that a random variable lands in an interval of possible values.

# Uncountably many outcomes? We need intervals.

## Example

Pick a number  $X$  “uniformly” from the unit interval  $(a,b)$ .

( $X$  here is called a **uniform random variable**.)

Then the sample space  $\Omega = (a, b)$ , and the probability that  $X \in (c, d)$ , where  $a \leq c \leq d \leq b$ , is

$$P(X \in (c, d)) = \frac{d - c}{b - a}.$$

Note that, for any point  $x \in (a, b)$ ,  $P(X = x) = \frac{x-x}{b-a} = 0$ .

## Example: dart throwing

If the sample space of a uniform rv is multidimensional, then the probability of landing in a particular region is the relative area or volume of that region.

If  $X$  is uniformly distributed on the set  $B$ , then

$$P(X \in C) = \frac{\text{volume}(C \cap B)}{\text{volume}(B)}.$$

### Example

Throw a dart uniformly at random at a circular dart board of radius 9 in. If the bullseye is a disc of radius 1 in in the center of the board, what is the probability of hitting the bullseye?

$$P(\text{bullseye}) = \frac{\text{area}(\text{bullseye})}{\text{area}(\text{board})} = \frac{\pi(1)^2}{\pi(9)^2} = \frac{1}{81}.$$