Chapter 1: Experiments with random outcomes

Jess Kunke

MATH/STAT 394: Probability I (Summer 2022 A-term)

Outline

Outline

Probability models: an introduction

Random sampling

Infinitely many outcomes

Useful consequences of Kolmogorov's axioms

Random variables: an introduction

Motivation

Probability theory: building mathematical models of experiments that have random/uncertain/noisy outcomes

- Recall: probability is a tool to measure uncertainty
- How? Probabilities measure sets
- To measure finite sets, we will count the number of elements in them
  - Clever rules
  - Three basic types of problems
- We will also learn useful problem-solving strategies based on the rules of probability

We will introduce terms, concepts, and notation from set theory and counting rules **in context along the way**

- See Appendices B-C for a concise presentation with practice exercises
- Refer to App. A and D if needed for calculus, series, and more

Probability model

A **probability model**, **probability space**, or **random experiment** consists of three parts:

1. The **sample space** Ω is the set of all possible **outcomes** of the experiment
2. The **event space** $\mathcal{F}$ is the set of all possible **events** (subsets of Ω)
   - This essentially corresponds to all the questions you can ask about the experiment's result
3. The third part is a **probability measure**; more on this soon

Let's look at some examples...

- We're going to introduce a bunch of terminology and notation in context
- Don't get too hung up on remembering it all now! We will practice it together in class and you will practice it in your homework
- Do take notes and practice the correct terminology and notation
- Refer to the textbook, annotated slides, lecture recordings as needed

Probability model

Example

Consider the result of a **single coin toss**.

- What is the sample space Ω? What is the event space $\mathcal{F}$?

Probability model

Example

Consider a series of **three coin flips**.

- What is the sample space Ω? What is the event space $\mathcal{F}$?

Key terms/concepts so far

- Set (unordered collection of items without repetition)
- Tuple (ordered list of items, can repeat)
- Different notation for a set $\{1, 2, 3\}$ versus a tuple $(1, 2, 3)$
- Writing a set in words, in formulas, or as a list of elements
- Subsets, elements of a set
  - (Sometimes useful) definition of a subset: We say $A \subseteq B$ if
  $$\omega \in A \implies \omega \in B. \quad \text{(see next slide too)}$$
- Special subsets of a set A: singletons, $\emptyset$, A itself
- Cartesian products

Note:

- Authors made unfortunate choice to use lowercase and uppercase omega for elements and sample spaces (makes it awkward to read the math aloud)
- Nothing special about this choice of letters, but it is typical to use lowercase (uppercase) letters for elements (sets)

A quick but handy note on subsets

Given two sets A and B, how do you prove that A is a subset of B?

How do you prove that A is **not** a subset of B?

We'll see an example of this later, on slide 34

Probability model: some more detail

A **probability model**, **probability space**, or **random experiment** consists of three parts:

1. The **sample space** Ω is the set of all possible **outcomes** of the experiment
2. The **event space** $\mathcal{F}$ is the set of all possible **events** (subsets of Ω)
   - This essentially corresponds to all the questions you can ask about the experiment's result
   - The set of all subsets of Ω is called the **power set** of Ω
   - Sometimes we either have to or want to choose something other than the power set; more on this later
3. The **probability measure** P is a function from $\mathcal{F}$ into the real numbers $\mathbb{R}$
   - Defines how we are measuring probability/uncertainty
   - Also called probability, or probability distribution

What parts you need to define depends on the question:

- "How many ways...?" $\rightarrow$ only need sample space, event space
- "What is the probability that...?" $\rightarrow$ need full probability model

Set operations

To define probability measures, we will need some fundamental set operations:

Definition

Let A, B be two subsets of a set Ω.

- Their **intersection** is $A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$
  - Two sets A, B are called **disjoint** if $A \cap B = \emptyset$
- Their **union** is $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$
  - In logic/sets/probability, "or" means "and/or"
- Their **difference** is $A \setminus B := \{\omega \in \Omega : \omega \in A \text{ and } \omega \notin B\}$
- The **complement** of A in Ω is $A^c = \{\omega \in \Omega : \omega \notin A\}$
  - Notice that to talk about the complement of A, we have to talk about A as a subset of some larger set Ω

Let's visualize this using a Venn diagram in an example:

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad A = \{x \in \Omega : x \text{ is even}\}, \quad B = \{1, 2, 3, 4\}$$

More about probability measures

A **probability model**, **probability space**, or **random experiment** consists of three parts:

1. The **sample space** Ω is the set of all possible **outcomes** of the experiment.
2. The **event space** $\mathcal{F}$ is the set of all possible **events**, or subsets of Ω.
3. The **probability measure** P is a function from $\mathcal{F}$ into the real numbers $\mathbb{R}$. Each event A has a probability $P(A)$, and P satisfies the following three axioms (Kolmogorov's axioms):

   i $0 \leq P(A) \leq 1$ for each event A

   ii $P(\Omega) = 1$ and $P(\emptyset) = 0$

   iii If $A_1, A_2, A_3, \ldots$ is a sequence of **pairwise disjoint** events (i.e.[1] mutually exclusive, $A_i \cap A_j = \emptyset$ for $i \neq j$), then

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

Prove that the sum of the probabilities of all possible outcomes must be 1.

[1] Note: i.e. means "in other words" while e.g. means "for example"

A useful consequence of Axiom iii

Fact

If $A_1, A_2, \dots, A_n$ *are* **pairwise disjoint** *events, then*

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

Example

Roll a fair six-sided die once. Fair means that the outcomes are equally likely.

- What is the sample space Ω?
- What is the probability of each possible outcome?
- What is the probability of rolling an even number?

A useful consequence of Axiom iii

Fact

If $A_1, A_2, ..., A_n$ are **pairwise disjoint** *events, then*

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n).$$

Example

Now suppose we change the 3 on the die to another 2. Define the new probability measure.

Outline

Sampling

Sampling: choosing objects from a given set according to a probability model

- Many problems can be framed as sampling problems
- Sampling methods are the mechanisms we use to model our random experiments
- Outcomes may be equally likely or not depending on the problem

Note: Choosing "at random" (imprecise) typically implies choosing **uniformly** at random (precise), meaning that all outcomes are equally likely.

We will start by considering experiments with equally likely outcomes.

Experiments with equally likely outcomes

Probabilities must sum to 1, and if the outcomes are equally likely then the probabilities must all be equal. In this case we have the following handy result:

Fact

If the sample space Ω has finitely many elements and equally likely outcomes, then we have

$$P(\omega) = \frac{1}{|\Omega|} \text{ for any outcome } \omega \in \Omega \text{ and } P(A) = \frac{|A|}{|\Omega|} \text{ for any event } A,$$

where $|A|$ is the number of elements of A, also called the **cardinality** *of A.*

Example

Let's return to the example of flipping a fair coin three times.

- What is the sample space Ω?
- What is the probability of getting at least two heads in a row?

Types of sampling mechanisms

Three basic types of sampling with equally likely outcomes:

1. Ordered sampling with replacement
   - e.g. the last example (three coin flips)
   - Repetitions of a simple experiment
   - Cartesian products, n-tuples
2. Ordered sampling without replacement
   - Permutations, arrangements
3. Unordered sampling without replacement
   - Combinations/sets

Comments:

- What about unordered sampling with replacement? (cliffhanger! later)
- Keep straight the sample space Ω versus the underlying sets of objects
  - e.g. $C = \{H, T\}$ vs. $\Omega = \{(x_1, x_2, x_3) : x_i \in C \; \forall \, i\}$
- Remember the different notation for *unordered sets* $\{a_1, a_2, ...\}$ versus *ordered n-tuples* $(a_1, a_2, ...)$
- In general, problems can involve a combination of mechanisms
- We will start by examining one at a time

Let's see a slightly more general example than the three coin flips:

Example

Suppose license plates are generated randomly by choosing three letters (say from the English alphabet) followed by three digits (0 through 9), and repeats are allowed. What is the probability that you get a license plate whose first two numbers are in $\{1, 2, 3, 4, 5\}$, possibly with repeats?

$$\underset{\text{letter}}{\underline{\qquad}} \quad \underset{\text{letter}}{\underline{\qquad}} \quad \underset{\text{letter}}{\underline{\qquad}} \quad \underset{\text{number}}{\underline{\qquad}} \quad \underset{\text{number}}{\underline{\qquad}} \quad \underset{\text{number}}{\underline{\qquad}}$$

2. Ordered sampling without replacement

Example

You have 6 different dinosaur toys in a bag. You reach in, remove 4 of them one at a time, and line them up in front of you.

- Suppose you don't put them back in the bag before picking the next one (no replacement)
- Suppose that you don't look in the bag and that we can model this as uniform random sampling

How many outcomes are there? Define the sample space Ω and compute $|\Omega|$.

2. Ordered sampling without replacement

Example

If instead you draw and line up all six dinosaur toys, how many outcomes are there? If all the dinosaurs are different species, what is the probability that triceratops is either second or third in line?

2. Ordered sampling without replacement

In general, these are called permutations or arrangements:

Arrangements: If you draw k out of n objects without replacement, the number of ways to order them is

$$|\Omega| = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

Permutations: In the special case that $k = n$,

$$|\Omega| = n \cdot \cdot (n-1) \cdot (n-2) \cdots 1 = n!$$

2. Ordered sampling without replacement

Example

You have a bag containing 3 red balls and 5 orange balls. You remove 4 of them, one at a time, without putting them back in the bag. What is the probability that you draw two red balls followed by two orange balls?

3. Unordered sampling without replacement

Example

A child wants to choose three of her six dinosaur toys to bring to a friend's house. How many choices does she have? Again assuming that all the dinosaur species of the toys are different, what is the probability that triceratops and ankylosaurus are among the three toys she brings if we assume each set of three toys is equally likely?

3. Unordered sampling without replacement

In general, these are called **combinations** or sets

The number of ways to draw a subset k out of n objects is given by the **binomial coefficient**:

$$|\Omega| = \frac{n!}{k!(n-k)!}$$

This is the same as dividing (1) the number of arrangements of k out of n things by (2) the number of ways to order k things:

$$\frac{n!}{(n-k)!} \div k!$$

Practice: Identifying an appropriate sampling mechanism

Which method would you use? Then solve the problem.

1. 9 piano students are nervous about their upcoming recital, so they put their names in a bag and draw one at a time to determine who plays when. What is the probability of any particular outcome (recital program)?
2. How many different lunches can you make by choosing 1 main and 1 side from 3 different main dishes and 5 different side dishes?
3. You flip a fair coin 10 times. What is the probability of getting exactly three heads?
4. How many ways can a team of 12 basketball players choose a captain and co-captain?

More practice: Identifying an appropriate sampling mechanism

Which method would you use? Then solve the problem.

1. The four types of nucleotide bases in DNA are A, T, C, and G. Suppose you have a solution in which all four are present in equal concentrations and the bases form a sequence of three base pairs uniformly at random. What is the probability of getting the sequence ATC?
    - What if instead of molecules in solution you had a bag with scrabble tiles (2 A's, 2 T's, 2 C's, 2 G's)?

What about unordered sampling with replacement?

Let's revisit this fourth sampling method.

Example

Draw two numbers from the set $\{1, 2, 3\}$ with replacement, but then disregard order so that you define the outcomes to be the possible combinations of numbers you can draw. What are the possible outcomes, and what are their probabilities? Are they equally likely?

Outline

Infinitely many outcomes

All our examples until now dealt with finitely many outcomes. This example

- involves infinitely many outcomes
- demonstrates we can sometimes turn a problem with unequally likely outcomes into a series of problems with equally likely ones

Example

Define a probability model (all three parts) for this problem: you flip a coin until it comes up tails. The outcome is the number of coin flips up to and including the first tails. What is the probability that you get tails on exactly the third coin toss?

Different kinds of infinite sets

Consider two other examples with infinite outcomes:

- What is the probability that you get a bullseye if you throw a dart uniformly at random at a target?
- If you choose a real number x uniformly at random from the interval [0,1], what is the probability that $x \in (0.25, 0.5)$?

The previous example (coin flips) is inherently different from the two examples above in an important way:

The sets involved in the coin flips example are **countable** or **discrete**

- Elements can be labeled in one-to-one correspondence with the positive integers (natural numbers) $\mathbb{N}$ (infinite) or a finite subset of them (finite)

Sets in the two examples above are **uncountable** or **continuous**

- These sets cannot be covered by or mapped to $\mathbb{N}$

Measuring uncountable sets

For an uncountable set A such as the real numbers,

$$P(x) = 0 \quad \forall\, x \in A.$$

Therefore a probability measure much be based not on points or counting the number of elements, but rather on lengths, areas, etc.

We can generalize our previous result on equally likely outcomes:

Fact

If the sample space Ω has uncountably many elements and equally likely outcomes, then for any event A we have

$$P(\omega) = 0 \text{ for any outcome } \omega \in \Omega \text{ and } P(A) = \frac{\nu(A)}{\nu(\Omega)} \text{ for any event } A,$$

where $\nu(A)$ (pronounced "nu of A") is an appropriately defined **measure** *of A for the problem of interest (e.g. length or area).*

This is somewhat imprecise but suffices for this course; a measure theoretic probability course will make this more precise

A note about Axiom iii

Note that Axiom iii in our statement of the probability model is only required to hold for countable sets: If $A_1, A_2, A_3, \ldots$ is a sequence of **pairwise disjoint** events (i.e. mutually exclusive, $A_i \cap A_j = \emptyset$ for $i \neq j$), then

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

This is called countable additivity.

An aside: σ-algebras

For uncountably infinite sets, we actually cannot use the power set as the event space

- Instead, replace this with a **$\boldsymbol{\sigma}$-algebra**, any collection of sets satisfying three properties (see textbook §1.6)

Even for countable sets (finite or infinite), it can be useful to define the event space as a σ-algebra that is not the power set

- Allows you to model information available to an observer/experimenter
- Arises in advanced probability concepts such as filtrations, martingales, optional stopping theorem
- Applications include selective inference, mathematical finance

Back to an uncountably infinite example

Example

What is the probability that you hit the bullseye if you throw a dart uniformly at random at a target?

- What additional information do you need? *This is an important part of research and consulting in which the problems are not handed to you fully defined*
- Define the sample space and the event of interest (not the whole event space)
- Answer the original question above

Outline

Useful consequences of the axioms

The following useful strategies are consequences of the rules of probability:

1. Partitioning an event
   - Decomposing an event into simpler pieces
2. Relating an event to its complement
   - Sometimes the probability of the complement is easier to compute
3. Leveraging the monotonicity of probability
   - If event B contains event A, then B must be as or more likely than A
4. Applying inclusion-exclusion rules
   - Relating the probability of a union to the probability of an intersection
5. Using de Morgan's law
   - Relating the complement of a union/intersection to the intersection/union of the complements

1. Partitioning an event

Definition

A **partition** of a set A is any collection of sets $\{A_1, ..., A_n\}$ such that

1. the sets are pairwise disjoint:

$$A_i \cap A_j = \emptyset \; \forall \; i \neq j$$

2. and their union is A:

$$\bigcup_{i=1}^{n} A_i = A.$$

Using the additivity of probabilities,

$$P(A) = P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) + ... + P(A_n).$$

1. Partitioning an event

Example

Arlo and Emery take turns rolling a fair six-sided die.

1. Arlo wins if he rolls a 1 or 2
2. Emery wins if he rolls a 4, 5, or 6
3. Arlo rolls first

What is the probability that Arlo wins and rolls no more than 4 times?

2. Relating an event to its complement

Either an event A happens or it doesn't happen (A^c), and this partitions the sample space:

$$A \cup A^c = \Omega$$

Therefore,

$$P(A) + P(A^c) = P(\Omega) = 1.$$

If the quantity of interest is $P(A)$, but computing $P(A^c)$ is easier, we can compute $P(A) = 1 - P(A^c)$.

Example

In five rolls of a fair six-sided die, what is the probability that any number appears more than once?

2. Relating an event to its complement

Another use of set complements

Suppose you are interested in the probability of an event A. Sometimes it is helpful to partition A by intersecting it with another event B and its complement:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

3. Leveraging the monotonicity of probability

If event B contains event A, then B must be as or more likely than A:

$$A \subseteq B \implies P(A) \leq P(B)$$

Example

Prove that repeated flips of a fair coin will eventually come up tails with probability 1.

Proof.

Define

$$A = \text{never see tails},$$
$$A_n = \text{first } n \text{ flips come up heads.}$$

A is the complement of the event we want to show has probability 1, so our goal is to show that $P(A) = 0$. *Which useful consequence did we just use?*

$A \implies A_n$ for any $n \in \mathbb{N}$, so which event is a subset of the other? (Look back at subset definition.) Does it go both ways, and why/why not?

Therefore, $P(\_\_\_) \leq P(\_\_\_)$.

$P(A_n) =$ ________ $\implies$ ________________

Therefore, $P(A) = 0$. $\square$

A quick but handy note on subsets

Given two sets A and B, how do you prove that A is a subset of B?

How do you prove that A is **not** a subset of B?

4. Applying inclusion-exclusion rules

Relating the probability of a union to the probability of an intersection:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof.

We'll use a Venn diagram.

$\square$

Example

A bag contains 10 red, 4 green, and 6 yellow balls. Draw two without replacement. What is the probability that your sample contains exactly one red ball or exactly one yellow ball?

- Recall: "or" means "and/or", not exclusive "or"

5. Using de Morgan's law

Relating the complement of a union/intersection to the intersection/union of the complements

Lemma

2D version:
For two subsets A, B of a set Ω,

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c.$$

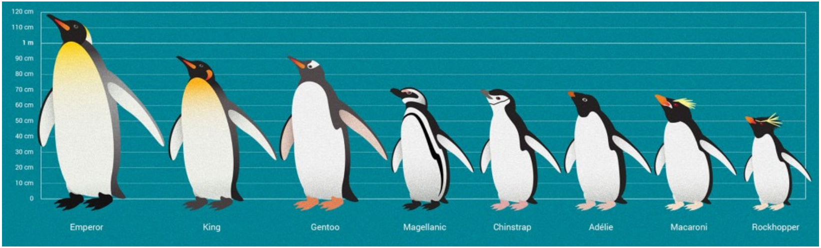
nD version:
Given subsets $A_1, A_2, \ldots$ of a set Ω,

$$\left(\bigcup_i A_i\right)^c = \bigcap_i A_i^c, \quad \left(\bigcap_i A_i\right)^c = \bigcup_i A_i^c.$$

Proof.

□

Putting these strategies together



Example

In a particular community of penguins, 20% of the penguins are rockhopper penguins and 15% of the penguins have interacted with humans. Rockhopper penguins who have interacted with humans make up 3% of the community. If you select one penguin uniformly at random, what is the probability of choosing a penguin that has never interacted with humans and is not a rockhopper?

- Tip: Define events and draw a Venn diagram.
- Try to solve it with and without de Morgan's law, and state which strategies you are using.

Outline

Motivation

Consider the example of rolling a die three times.

- So far we've asked questions like these:
  - What is the probability of rolling the same number more than once?
  - What is the probability that the sum of the rolls is at least 10?
- What about questions like these:
  - What is a "typical" die roll?
  - What die roll would we expect on average if we roll many times?
  - Suppose this is a carnival game and the prize you win is some function of the sequence of die rolls. What prize would you expect to win on average?

Random variables can make it easier to handle the first type of question, and they will enable us to address the second type of question

Random variables

Definition

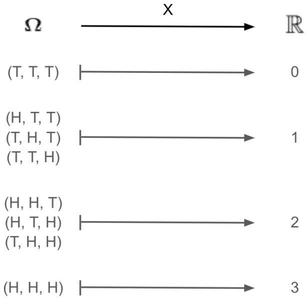
A **random variable** X on a sample space Ω is a function from Ω into the real numbers, or a real-valued function on Ω:

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

- Usually denoted by a capital letter

Example

Flip a coin three times. Define X = number of heads.



Events and random variables

By definition, the preimage in Ω of any subset B of the codomain (set of possible outputs/values) of a random variable X on Ω is an event:

$$\{X \in B\} = \{\omega \in \Omega : X(\omega) \in B\}$$

This gives us two ways to define an event:

- In our previous example, consider the event "the coin comes up heads once". We've seen that we can define this by which outcomes in Ω meet this criterion:
$$\{(H, T, T), (T, H, T), (T, T, H)\}$$
- We now see that we can also view this as the subset of Ω that is mapped by X to the real number 1:
$$\{\omega \in \Omega : X(\omega) = 1\}$$
- These are two ways of describing the same subset of Ω (the same event)

Degenerate random variables

Definition

A random variable is **degenerate** if $\exists\ b \in \mathbb{R}$ such that $P(X = b) = 1$.

Example

Consider either (1) flipping a coin once (e.g. whether it is raining today or not) or (2) drawing a number uniformly at random from the unit interval $[0, 1]$.

- Define a probability measure P on Ω and a random variable X such that X is degenerate.
- Define a different random variable Y (and possibly a different probability measure $\tilde{P}$) such that Y is degenerate.

A few points on notation

Both:

$$P(X < 3, Y = 4) = P(X < 3 \text{ and } Y = 4) = P(X < 3) \cup P(Y = 4)$$

Either/or:

$$P(X < 3 \text{ or } Y = 4) = P(X < 3) \cap P(Y = 4)$$

Probability distributions and mass functions

Now that we have random variables, we can define the following as well:

Definition

Let X be a random variable.

- The **probability distribution** of X is the collection of probabilities $P\{X \in B\}$ for sets $B \subseteq \mathbb{R}$.
- X is a **discrete random variable** if $\exists$ a finite or countably infinite set $\{k_1, k_2, ...\}$ of real numbers such that

$$\sum_i P(X = k_i) = 1.$$

- The **probability mass function** or pmf of a discrete random variable X is the function p (or p_X) defined by

$$p(k) = P(X = k)$$

  for all possible values k of X.

Example 1: Finding the pmf of a discrete RV

Let's model daily stock price as a coin flip. Every time the coin comes up heads (tails), the price increases (decreases) and you earn \$1 (\$1). Define X as your net profit or loss after five flips.

- What is the set of possible values of X?
- Define the probability mass function for X.

Example 2: Finding the pmf of a discrete RV

Suppose we have the target drawn below; the black numbers along the horizontal segment are the difference in radii of the circles, and the blue vertical numbers are the number of points you get if your arrow lands in that region. Suppose your arrow lands uniformly at random on the target, and define X to be the number of points you get.

- What are the possible values of X (the codomain of X)?
- Define the pmf of X. Verify that the probabilities sum to 1.