Motivation and Basic definitions

Probability—Motivation

- Why is Probability important?
- Anything that happens in life is uncertain.
- So whatever you try to do, you need to have some way of dealing or thinking about this uncertainty.
- Probability theory gives you the models to deal with uncertainty, in a "systematic way".

- Example: signal processing, you have to deal with the noise which is "random" and uncertain.
- Example: dealing with the stock market, it is definitely random.
- Example: playing card games, backgammon, all are uncertain.

- Probability theory is a framework for dealing with uncertainty, for dealing with situations in which we have some kind of randomness.

- Goals: understanding how to set up a probabilistic model, and what are the basic rules of the game for dealing with probabilistic models?

Probabilistic models

- There are two main components:
  - **Sample Space**: basically a description of all the things that may happen during a random experiment
  - **Probability Law**: describes our beliefs about which outcomes are more likely to occur compared to other outcomes. Probability laws have to obey certain properties that we call the **axioms of probability** (the rules of the game).

Probabilistic models—Sample Space

- Fix a particular experiment. For example: flipping a coin, rolling a dice, etc.
- List all the possible things that may happen during this experiment, which we call all the possible "outcomes".
- More formally, by saying a list, we mean a "set". That set is our sample space.
- **Sample Space**: a set whose elements are the possible outcomes of the experiment.
- Example: flipping a coin. Sample space: $\{\text{Head}, \text{Tail}\}$
- Important characteristics of sample space:
  - (a) Mutually exclusive: if A happens, then B cannot happen! Meaning that exactly one of these outcomes should happen as the result of the experiment.
  - (b) Collectively exhaustive: no matter what happens in the experiment, you're going to get one of the outcomes in the sample space (means "don't forget any possible outcome!")

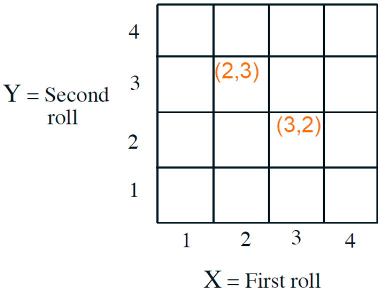
Probabilistic models—Sample Space

- How much details are you going to include in your sample space?
- Suppose your experiment is "coin flipping" and think of the following sample space:
  - head
  - tail and it's raining
  - tail and it's not raining
- It is mutually exclusive and collectively exhaustive!
- But if coin flipping inside this room is completely unrelated to the weather outside, then you're going to stick with $\{\text{Head}, \text{Tail}\}$ sample space!
- Therefore, you need judgment in order to set up an appropriate sample space.

Sample Space—Example

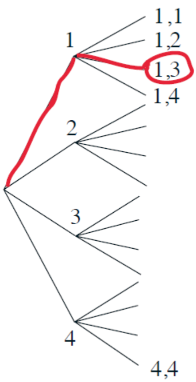
- You have a dice that only has four faces (a tetrahedron).
- When you roll it, you get a result which is 1,2,3,4 equally likely (unbiased dice).
- Experiment: rolling the dice twice (distinguish between (a, b) and (b, a) if $a \neq b$).
- Write the sample space: you get 16 elements (a finite set).

$$(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(2,4),(3,1),(3,2),(3,3),(3,4),\\(4,1),(4,2),(4,3),(4,4)$$



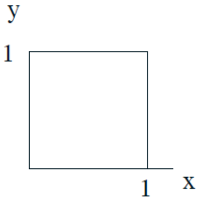
Sample Space—Sequential Way

- When you have an experiment that consists of multiple stages, you can draw a diagram that shows you how those stages evolve.
- We call it "sequential description" or a "tree-based description" by drawing a tree of the possible outcomes.
- Any path is associated to a particular outcome, any outcome is associated to a particular path.



Sample Space—Infinite set

- Experiment: you are playing darts and the target is this square: $\Omega = \{(x, y)|0 \leq x, y \leq 1\}$
- Suppose you're perfect at this game and you're sure that your darts will always fall inside the square. But where exactly your dart would fall inside that square is "random".



- All the possible points inside the square are possible outcomes of the experiment: (x, y) where x, y are real and $\in [0, 1]$
- There's infinitely many real numbers $\Rightarrow$ there's infinitely many points in the square $\Rightarrow$ our sample space is an infinite set.

Probability Law

- Which outcome is more likely to occur compared to the others?
- Should we do this by assigning probabilities to the outcomes? Not exactly!
- In the previous example, what would be the probability that you hit exactly one particular point, say $(0.5, 0.5)$ to infinite precision? Zero!
- So, if any individual outcome has zero probability, you are not giving any value information.
- What should we do? **Assign probabilities to subsets of the sample space.**
- We have our sample space, which is Ω, and we consider some subset of the sample space, say A.
- The outcome is a point and it is random. So the outcome may be
  - Inside set A, in which case we say that "event A occurred".
  - Outside the set A, in which case we say that "event A did not occur".

Probability Axioms

- To assign probabilities to events, there are some ground rules which we call **probability axioms**.
- (1) **Nonnegativity**: $P(A) \geq 0$.
  Probabilities should be non-negative (that's our convention).
- (2) **Normalization**: $P(\Omega) = 1$.
  The probability of the entire sample space Ω is equal to one. Because the outcome is certain to be an element of the sample space (collectively exhaustive).
- (3) **Additivity**: If $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$.
  If two events A and B have no common elements, then the total probability of A together with B has to be equal to the sum of the individual probabilities.
- $A \cap B$: intersection of A and B
- $A \cup B$: union of A and B

Probability Axioms

- Using probability axioms 1,2,3, we prove that probabilities are always less than 1: $P(A) \leq 1$

$$\begin{aligned} 1 = P(\Omega) = P(A \cup A^c) &\leftarrow \text{ Axiom 2} \\ 1 = P(A) + P(A^c) &\leftarrow \text{ Axiom 3} \\ P(A) = 1 - P(A^c) \leq 1 &\leftarrow \text{ Axiom 1} \end{aligned}$$

- Additivity for three disjoint sets. Suppose $A \cap B = A \cap C = B \cap C = \emptyset$, then:

$$\begin{aligned} P(A \cup B \cup C) = P((A \cup B) \cup C) &= P(A \cup B) + P(C) \\ &= P(A) + P(B) + P(C) \end{aligned}$$

- Additivity for m disjoint sets. Suppose $A_1, A_2, \cdots, A_m$ are pairwise disjoint. Then using "induction":

$$P(A_1 \cup A_2 \cup \cdots \cup A_m) = P(A_1) + P(A_2) + \cdots + P(A_m)$$

Probability Axioms

- Suppose we have a finite set of outcomes $< a_1, a_2, \cdots, a_n >$. Put them together in a set $A = \{a_1, a_2, \cdots, a_n\}$. Then:

$$P(A) = P(\{a_1, a_2, \cdots, a_n\}) = P(a_1) + P(a_2) + \cdots + P(a_n)$$

- Example. Roll an unbiased dice with 6 faces. What is the probability of seeing an odd number?
$\Omega = \{1, 2, 3, 4, 5, 6\}, A = \{1, 3, 5\}, P(A) = P(1) + P(3) + P(5) = \frac{3}{6} = \frac{1}{2}$
- Example. Roll an unbiased dice with 4 faces twice and let every possible outcome have probability $\frac{1}{16}$. Let X be the outcome of the first roll and Y be the outcome of the second roll.
    - $P\big((X, Y) \text{ is } (1,1) \text{ or } (1,2)\big) = \frac{2}{16}$
    - $P(\{X = 1\}) = \frac{4}{16}$
    - $P(X + Y \text{ is odd }) = \frac{8}{16}$
    - $P(\min(X, Y) = 2) = \frac{5}{16}$

Discrete uniform law

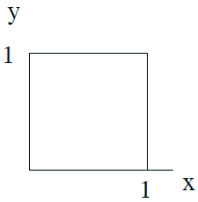
- Let all outcomes be equally likely, then:

$$P(A) = \frac{\text{number of elements of } A}{\text{total number of sample points}}$$

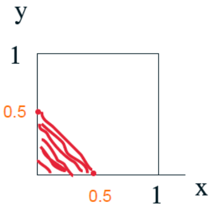
- Computing probabilities $\equiv$ counting
- Defines fair coins, fair dice, well-shuffled decks

Continuous uniform law

- Two "random" numbers in $[0,1]$.



- **Uniform law**: "Probability = Area"
- $P(X+Y \leq \frac{1}{2})$ = Area of the triangle$= 0.5 * 0.5 * 0.5 = \frac{1}{8}$



Countable additivity axiom

- If $A_1, A_2, \cdots$ are pairwise disjoint events, then:

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

- Example. Suppose you have a fair coin, and you keep flipping it until you see a head, then you will stop. Outcome i happens if you see a head (for the first time) in the i-th step. Suppose you are given $P(n) = \frac{1}{2^n}$ for $n = 1, 2 \cdots$
- Find $P(\text{outcome is even})$
- $P(\text{outcome is even}) = P(\{2, 4, 6, \cdots\}) = P(2) + P(4) + P(6) + \cdots = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \cdots = \frac{1}{3}$
- If $a < 1$ then $1 - a^n = (1 - a)(1 + a + a^2 + \cdots + a^{n-1})$
- When $n \to \infty$: $(1 - a)^{-1} = 1 + a + a^2 + \cdots$
- $a = \frac{1}{4}$ then $1 + \frac{1}{4} + \frac{1}{4^2} + \cdots = \frac{1}{\frac{3}{4}} = \frac{4}{3} \Rightarrow \frac{1}{4} + \frac{1}{4^2} + \cdots = \frac{1}{3}$
- Note that $\frac{1}{4} + \frac{1}{4^2} + \cdots = \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \cdots$
- $\Rightarrow \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \cdots = \frac{1}{3}$