Lecture notes 2
Probabilistic Models

Lecture Outline

- Basic probabilistic model: an experiment
  - Sample space
  - Events (algebra of events, set theory)
  - Probability measure (law, distribution)

Reading: Bertsekas & Tsitsiklis 1.1–1.2

Components of a Probabilistic Model

Basic model: an *experiment*, a mathematical model of a process with an outcome that is not fully predictable.

Three components:

- Sample space: A list of all "elementary" or "finest grain" outcomes of an experiment
- Algebra of events or set theory: language for manipulating collections of elementary events = sets
- Probability law: means of assigning a probability to events in a consistent and useful way.

Sample Space

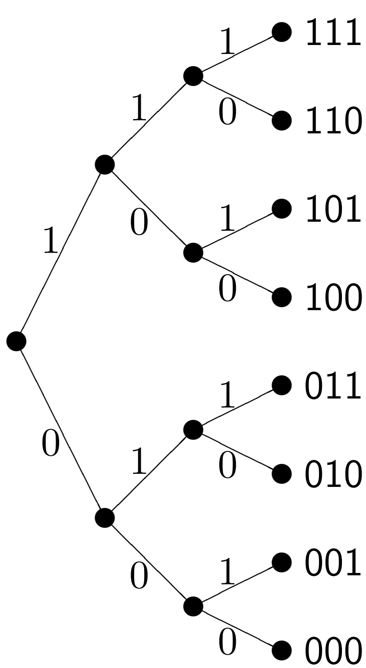
Example Three flips of a fair coin labeled 1 (or *head*) on one side and 0 (or *tail*) on the other.

Many ways to represent outcomes:

- List or table: Each column corresponds to the number of the flip.

000
001
010
011
100
101
110
111

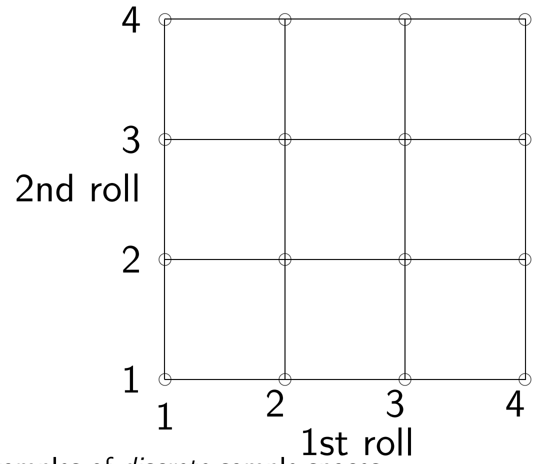
- Sequential description: a *tree*



One terminal node for each elementary outcome.

Example Roll a fair four-sided die twice.

Samplespace: table or tree as above, or a graphical representation.



Examples of *discrete* sample spaces.

An example of a *continuous* sample space:

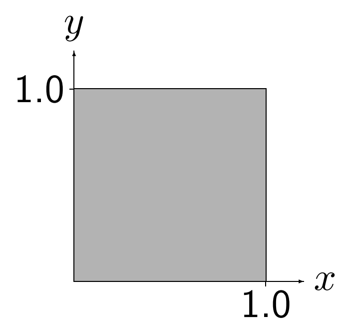
The Spinning Pointer

When the pointer stops it points to a number in $[0, 1) = \{r : 0 \leq r < 1\}$.

Suppose the pointer is spun twice, then the sample space consists of all points in the *unit square*

$$\{x, y : 0 \leq x < 1, 0 \leq y < 1\}$$

The shaded region below



A sample space is a collection of all of the possible elementary outcomes, called *sample points*. We assume that

- the sample points are all *disjoint* or *mutually exclusive*, i.e., they are separate and distinct outcomes
- the sample points are *collectively exhaustive*, i.e., together they make up the entire sample space, they consititute all possible elementary outcomes

We often use Ω to denote the sample space and $\omega \in \Omega$ to denote points in sample space.

A sample space is an example of a *set*:

A *set* is a collection of objects which are called *elements* of the set, so our sample space is just a set with sample points as elements.

So dealing with a sample space means using basic set theory.

Events

Will want to assign *probabilities* to particular events or outcomes occurring as a result of an experiment.

Examples of events:

- Flip three coins and get 010. (An elementary event.)
- Flip three coins and get exactly one 1. This is a more complicated event, consisting of three elementary events 001, 010, and 100.
- Flip three coins and get an odd number of 1's: Consists of elementary events 001, 010, 100, 111.
- Flip three coins and get a result such that the sum of values is 2. 011, 101, 110.
- Spin the fair wheel and the pointer points to exactly $1/\pi$. (Elementary event, not very interesting.)

- Spin the fair wheel and the pointer points to a number between 0.0 and 0.5.
- Spin the fair wheel twice and the sum of the two numbers is less than 0.3.

Moral: Events are just *subsets* of the sample space, i.e., sets of elements which belong to Ω.

Usually use capital letters for subsets of sample space and write $F \subset \Omega$.

- If a sample point ω is in a set F which is a subset of Ω, we write $\omega \in F$. If ω is not in F, we write $\omega \notin F$.

  Note that $\in$ denotes *point inclusion*: $\omega \in F$ means a point ω is contained in a set F. $\subset$ denotes *set inclusion*, $F \subset G$ means that a set F is a subset of another set G, that is, if $\omega \in F$, then also $\omega \in G$.
- A set F might have only one point in it, e.g., $F = \{\omega\}$ for a specific $\omega \in \Omega$.

- All sets are subsets of themselves, thus the entire sample space Ω is an event, the event that "something happens."
- A set might have *no points in it*, i.e., be the *empty set* $\emptyset$. This is the event that "nothing happens," also called the *impossible event*.

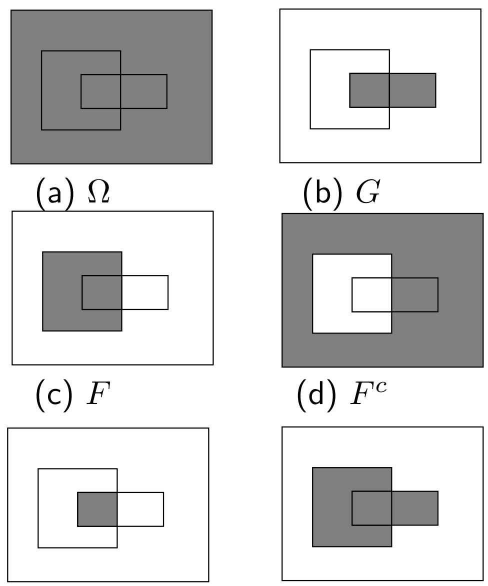
Set Theory Basics

When dealing with events will use *algebra of events* or *set theory* to combine simple events into complicated events, decompose complicated events into simple ones, and to perform other operations on sets.

The three basic operations on sets are

- complementation
- intersection
- and union

The definitions are given and illustrated by *Venn diagrams*: F and G are sets and the sample space is the full box.



(a) Ω (b) G

(c) F (d) F^c

(e) $F \cap G$ (f) $F \cup G$

Complementation

$$F^c = \{\omega : \omega \notin F\} ,$$

all of the points of Ω that are not in F.

Intersection

$$F \cap G = \{\omega : \omega \in F \text{ and } \omega \in G\} ,$$

the points which are in both sets.

If F and G have no points in common, then $F \cap G = \emptyset$, the null set, and F and G are said to be *disjoint* or *mutually exclusive*.

Union

$$F \cup G = \{\omega : \omega \in F \text{ or } \omega \in G\} ,$$

that is, the union of two sets F and G contains the points that are either in one set of the other, or both.

A few useful relations (can be rigorously proved using axioms of set theory or visualized using Venn diagrams).

First seven can be used as axioms for all of the rest.

$$\begin{aligned}
F \cup G &= G \cup F \text{ commutative law} \\
F \cup (G \cup H) &= (F \cup G) \cup H \text{ associative law} \\
F \cap (G \cup H) &= (F \cap G) \cup (F \cap H) \\
&\quad \text{distributive law} \\
(F^c)^c &= F \\
F \cap F^c &= \emptyset \\
(F \cap G)^c &= F^c \cup G^c \text{ DeMorgan's "law"} \\
F \cap \Omega &= F
\end{aligned}$$

$$\begin{aligned}
F \cap G &= G \cap F \text{ commutative law} \\
F \cap (G \cap H) &= (F \cap G) \cap H \text{ associative law} \\
(F \cup G)^c &= F^c \cap G^c \text{ DeMorgan's other "law"} \\
F \cup F^c &= \Omega \\
F \cup \emptyset &= F \\
F \cup (F \cap G) &= F = F \cap (F \cup G) \\
F \cup \Omega &= \Omega \\
F \cap \emptyset &= \emptyset \\
F \cup G &= F \cup (F^c \cap G) = F \cup (G - F) \\
F \cup (G \cap H) &= (F \cup G) \cap (F \cup H) \text{ distributive law} \\
\Omega^c &= \emptyset \\
F \cup F &= F \\
F \cap F &= F
\end{aligned}$$

notation

$\cup_{n=1}^N F_n = F_1 \cup F_2 \cup \cdots \cup F_N =$ all points in F_1 or F_2 $\cdots$ or F_N

$\cap_{n=1}^N F_n = F_1 \cap F_2 \cap \cdots \cap F_N =$ all points in F_1 and F_2 $\cdots$ and F_N

Can also have infinite unions and intersections

$\cup_{n=1}^\infty F_n =$ all points which are in some F_n (i.e., for at least one n)

$\cap_{n=1}^\infty F_n =$ all points which are in all of the F_n

Probablity

Probability measure: an assignment of a real number $P(F)$ to every event F (subset of sample space Ω) such that the following three simple axioms are satisfied:

1. $P(F) \geq 0$ for all F (nonnegativity)

2. $P(\Omega) = 1$ (normalization)

3. If F and G are disjoint ($F \cap G = \emptyset$), then

$$P(F \cup G) = P(F) + P(G) \text{ (additivity)}$$

The third axiom needs to be strengthened when dealing with infinite sample spaces and limits.

Rhetorical question: Do weird sets have probabilities?

Intuition for probability axioms: Mimic relative frequencies, e.g., perform a sequence of N experiments (e.g., roll a die N times): relative frequency of an event $F =$

$$\frac{\text{number of times } F \text{ occurs}}{N}$$

- Probabilities are nonnegative (like relative frequencies).
- Probability something happens is 1 (again like relative frequencies).
- Probabilities of *disjoint* events add. (again like relative frequencies)

Regardless of intuition (which eventually will be made precise by results), key point is this:

A function P defined for all subsets of Ω is a probability measure if and only if it satisfies the three axioms of probability.

Except for normalization, behavior is like

- weights
- areas
- volumes
- sums
- integrals

Elementary Properties

Axioms immediately imply several simple, but important, properties of probabilities

Assume that P is a probability measure defined on a sample space Ω. Then

(a) $P(F^c) = 1 - P(F)$.

(b) $P(F) \leq 1$.

(c) $P(\emptyset) = 0$.

(d) If an event F is the union $\{F_i; i = 1, \ldots, n\}$ of a finite collection of disjoint events, i.e., if $F_i \cap F_k = \emptyset$ when $i \neq k$ and $F = \cup_{i=1}^n F_i$, then

$$P(F) = \sum_{i=1}^{n} P(F_i). \tag{1}$$

(e) "Total Probability Theorem" If $\{F_i;\ i = 1, 2, \ldots, K\}$ is a finite partition of Ω, i.e., if $F_i \cap F_k = \emptyset$ when $i \neq k$ and $\bigcup_{i=1} F_i = \Omega$, then

$$P(G) = \sum_{i=1}^{K} P(G \cap F_i) \tag{2}$$

(f) If $F \subset G$, then $P(F) \leq P(G)$.

(g) $P(F \cup G) = P(F) + P(G) - P(F \cap G)$.

(h) $P(F \cup G) \leq P(F) + P(G)$. (Union bound or Bonferroni inequality.)

Proof:

(a) $F \cup F^c = \Omega$ implies $P(F \cup F^c) = 1$ (Axiom 2). $F \cap F^c = \emptyset$ implies $1 = P(F \cup F^c) = P(F) + P(F^c)$ (Axiom 3), which implies (a).

(b) $P(F) = 1 - P(F^c) \leq 1$ (Axiom 1 and (a) above).

(c) Axiom 2 and (a) above.

(d) Iteratively applying Axiom 3

$$\begin{aligned} P(F) &= P(\cup_{i=1}^n F_i) \\ &= P(F_1) + P(\cup_{i=2}^n F_i) \\ &\vdots \\ &= \sum_{i=1}^n P(F_i). \end{aligned}$$

(e)

$$\begin{aligned} P(G) &= P(G \cap \Omega) \text{ set theory} \\ &= P(G \cap (\bigcup_i F_i)) \text{ set theory} \\ &= P(\bigcup_i (G \cap F_i)) \text{ set theory} \\ &= \sum_i P(G \cap F_i). \text{ additivity} \end{aligned}$$

(f) From set theory, if $F \subset G$, then $G = F \cup (G \cap F^c)$. Since this is a union of disjoint events, $P(G) = P(F) + P(G \cap F^c) \geq P(F)$ from Axiom 1.

(g) Can rewrite a union as a disjoint union: $F \cup G = F \cup (G \cap F^c)$ so that $P(F \cup G) = P(F) + P(G \cap F^c)$. Also know that

$$\begin{aligned} P(G) &= P(G \cap \Omega) \text{ set theory} \\ &= P(G \cap (F \cup F^c) \text{ set theory} \\ &= P((G \cap F) \cup (G \cap F^c)) \text{ set theory} \\ &= P(G \cap F) + P(G \cap F^c) \text{ Axiom 3} \end{aligned}$$

which the previous formula proves the result.

(h) Follows from (e) and Axiom 1.

The fourth property provides the key for classical probability using counting arguments:

Suppose that a sample space has a finite number of points, say $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$.

Since the sample points are by definition disjoint, the last property implies that for any event F

$$P(F) = \sum_{k: \omega_k \in F} P(\{\omega_k\})$$

I.e., can find the probability of any event by adding up the probabilities of all the sample points in the event.

Great for computing probabilities for discrete experiments (coins, dice, etc.), *not* for continous experiments! *(Why not??)*

Important Special Case: Uniform Probability Law — All sample points are equally probable, i.e., $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ and

$$P(\{\omega_k\}) = \frac{1}{N};\ k = 1, 2, \ldots, N$$

This implies

$$P(F) = \frac{\text{number of sample points in F}}{\text{total number of sample points}}$$

Can compute probabilities by *counting* for experiments with finite sample spaces and a uniform probability law.

e.g.,

Probability get exactly one 1 in three coin flips: $\frac{3}{8}$

Probability get an odd number of 1's in three coin flips: $\frac{4}{8} = \frac{1}{2}$

Probability the sum of the outcomes of three coin flips is 2: $\frac{3}{8}$

What if discrete, but not finite?

E.g., *countably infinite* like the integers.

Uniform law does not make sense (but will find useful nonuniform laws using the additivity property).

Need stronger form of Axiom 3: If $F_1, F_2, F_3, \ldots$ are disjoint, then

$$P(\bigcup_{k=1}^{\infty} F_k) = \sum_{k=1}^{\infty} P(F_k)$$

countable additivity, limiting form of Axiom 3.

Example Suppose the sample space is $\{1, 2, 3, \ldots\}$ and $P(n) = 2^{-n}$.

$$\begin{aligned}\Pr(\text{ outcome is even }) &= P(\{2, 4, 6, 8, \ldots\}) \\ &= P(2) + P(4) + P(6) + \ldots\end{aligned}$$

$$\begin{aligned}&= \sum_{k=1}^{\infty} P(2k) \\ &= \sum_{k=1}^{\infty} 2^{-2k} = \frac{1}{3}\end{aligned}$$

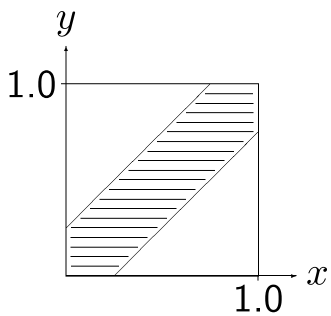
What if continuous sample space, e.g., uncountably infinite?

Idea of adding up probabilities of elementary points does *NOT* work!

Use continuous analogy to summing: integrate.

Example Romeo and Juliet have a date. Each arrives late with a random delay of up to 1 hour, where the pair of delays is equivalent to that achievable by spinning two identical fair wheels. Each will wait only 1/4 of an hour before leaving.

- What is the probability that Romeo and Juliet will meet?



Crosshatched region $= \{(x,y) : |x-y| \le \frac{1}{4}\}$

Answer: $\frac{\text{Area of crosshatched region}}{\text{Area of sample space}} = 1 - 2 \times \frac{1}{2}(.75)^2 = .4375$

So continuous uniform probability law:

$$P(F) = \frac{\text{area of } F}{\text{total area of sample space}}$$

Only make sense if $P(\Omega)$ is finite.

In higher dimensions, use volume.