Chapter 4

Geometric Probability and the Modern Definition

Lecture 4 | Devore 9e: 2.1–2.2

Last time: classical probability, where S is finite and all outcomes are equally likely, so $\mathbb{P}(A) = |A|/|S|$ and everything reduces to counting.

Today: what happens when S is uncountably infinite, so that $\mathbb{P}(\omega_i) = \mathbb{P}(\omega_j) =$ "$1/\infty$"? Counting is useless. We patch it, watch the patch fail, and then rebuild from scratch.

4.1 Geometric probability

Suppose S has a nice geometry — an interval, a rectangle, a disc. Then measure size by length-/area/volume instead of by counting:

Definition 4.1 (Geometric probability)**.** For $A \subset S$,

$$\mathbb{P}(A) := \frac{\text{Area}(A)}{\text{Area}(S)}$$

(with length or volume in place of area as the dimension requires).

4.1.1 A rendezvous

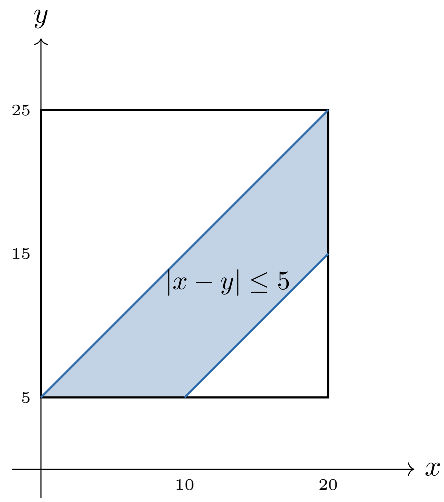
Example 4.2 (Will we meet?)**.** You arrive at the meeting point at a time uniformly distributed between 11:50 and 12:10; the person you are meeting arrives uniformly between 11:55 and 12:15, independently. Whoever arrives first waits 5 minutes and then leaves. What is the probability that you meet?

Measure time in minutes after 11:50 and let x be your arrival, y theirs. Then

$$S = \{(x, y) : 0 \le x \le 20,\ 5 \le y \le 25\},$$

a 20×20 square, so $\text{Area}(S) = 400$. You meet exactly when neither has left before the other arrives, i.e. when

$$A = \{(x, y) \in S : |x - y| \le 5\} = \{(x, y) \in S : y - 5 \le x \le y + 5\}.$$



The complement is easier to measure: it is two right triangles. Above the band, $y > x + 5$ cuts off the triangle with corners $(0,5),(0,25),(20,25)$ — the line $y = x + 5$ runs from one corner of the square to the opposite one, so both legs are 20. Below the band, $y < x - 5$ cuts off the triangle with corners $(10,5),(20,5),(20,15)$, whose legs are both 10. Hence

$$\text{Area}(A) = 400 - \tfrac{1}{2}(20)^2 - \tfrac{1}{2}(10)^2 = 400 - 200 - 50 = 150,$$

$$\mathbb{P}(A) = \frac{150}{400} = \boxed{0.375}.$$

The asymmetry is real: their window is shifted five minutes later than yours, so the two triangles are *not* the same size, and the meeting region is a trapezoid rather than the symmetric hexagon you would get from identical windows.

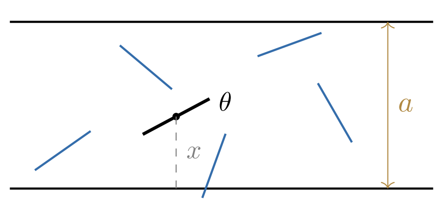
Remark 4.3. Two things to notice. First, the answer came from *areas*, not counting — there are infinitely many arrival times. Second, we computed the complement because two triangles are easier than the four-sided region itself; that move is worth remembering.

4.1.2 Buffon's needle

Example 4.4 (Buffon, 1777)**.** The plane is ruled with parallel lines a distance a apart. A needle of length $\ell < a$ is dropped at random. What is the probability it crosses a line?

Parametrize a landing by two numbers: x, the distance from the needle's center to the nearest line, and θ, the angle the needle makes with the lines. Then

$$S = \left\{(x,\theta)\ :\ 0 \le \theta \le \pi,\ \ 0 \le x \le \tfrac{a}{2}\right\}, \qquad \text{Area}(S) = \frac{a}{2}\cdot\pi.$$



The needle crosses a line exactly when its half-projection onto the perpendicular reaches the line, i.e.

$$A = \Big\{(x, \theta)\ :\ x \le \tfrac{\ell}{2}\sin\theta\Big\}.$$

Therefore

$$\mathbb{P}(A) = \frac{\int_0^\pi \frac{\ell}{2}\sin\theta\,\mathrm{d}\theta}{\frac{a}{2}\cdot\pi} = \frac{\frac{\ell}{2}\cdot 2}{\frac{a}{2}\pi} = \boxed{\frac{2\ell}{\pi a}}.$$

Remark 4.5. Solving for π turns this into an experiment: drop a needle many times, and $\pi \approx 2\ell N/(a\,N_{\mathrm{cross}})$. This is arguably the first Monte Carlo method.

4.1.3 Bertrand's paradox — where "at random" breaks

Example 4.6 (Bertrand, 1889)**.** Choose a chord of the unit circle *at random*. What is the probability that its length exceeds the side of the inscribed equilateral triangle (which is $\sqrt{3}$)?

Three perfectly reasonable readings of "at random":

How the chord is chosen	**Answer**
Fix one endpoint, pick the other endpoint uniformly on the circle. The chord is long when the second endpoint is in the far arc, which is 1/3 of the circumference.	$\frac{1}{3}$
Pick the chord's *midpoint* distance from the center uniformly on $[0, 1]$. The chord is long when that distance is $< \frac{1}{2}$.	$\frac{1}{2}$
Pick the chord's *midpoint* uniformly in the disc. The chord is long when the midpoint lands in the disc of radius $\frac{1}{2}$, of area $\pi/4$ out of π.	$\frac{1}{4}$

All three are correct. The phrase "at random" does not specify a probability model, and geometric probability silently assumed one.

4.1.4 So where are we

	requires	
Classical $\mathbb{P}$	S finite,	$\mathbb{P}(\omega_i) = \mathbb{P}(\omega_j)$
Geometric $\mathbb{P}$	S a nice shape,	$\mathbb{P}(\omega_i) = \mathbb{P}(\omega_j)$

Limitations. Neither handles

- uneven probabilities across S (a loaded die),
- countably infinite S,
- uncountable S with a complicated shape.

We need a definition that never asks where the numbers came from.

4.2 The modern definition

The change of viewpoint: stop assigning probability to *outcomes* and start assigning it to *events*. So $\mathbb{P}$ becomes a function whose input is a set.

$$\text{experiment} \Rightarrow \text{outcomes } \omega_i \Rightarrow S \Rightarrow \text{events (subsets of } S) \Rightarrow \mathbb{P} \text{ a function on events.}$$

But which subsets are allowed to be events? The natural first answer is "all of them", i.e. the power set $\mathcal{P}(S)$ (also written 2^S). For a finite or countable S that works. For an uncountable S it does not: one can construct subsets of $[0, 1]$ so pathological that *no* sensible assignment of length to them exists.

So we retreat, deliberately: **do not try to define $\mathbb{P}$ on every subset — choose a modest family of subsets and define $\mathbb{P}$ only there.** The family must be closed under the operations we actually perform on events (complement, and countable unions), or the theory would be useless. That requirement, and nothing more, is the next definition.

Definition 4.7 (σ-algebra)**.** Let $\mathcal{F}$ be a collection of subsets of S such that

(i) $S \in \mathcal{F}$;

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$;

(iii) $A_i \in \mathcal{F}$ for $i = 1, 2, \cdots \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Then $\mathcal{F}$ is a **σ-algebra** (or *event field*), and its elements are called **events**.

Remark 4.8. Note (iii) uses *countable* unions — this is exactly why Lecture 2 spent time on countable versus uncountable. Also note that $\mathcal{F}$ is a set *of sets*; $\mathcal{F} = \emptyset \cup S$ is a type error, the correct smallest example is $\mathcal{F} = \{\emptyset, S\}$.

Example 4.9.

- $\mathcal{F} = \{\emptyset, S\}$ — the smallest σ-algebra.
- For a fixed $A \subset S$: $\mathcal{F} = \{\emptyset, A, A^c, S\}$.
- $\mathcal{F} = \mathcal{P}(S)$, all subsets — the largest σ-algebra. This is the right choice whenever S is finite or countable.

The pair $(S, \mathcal{F})$ is now fixed, and we put a function on it.

Definition 4.10 (Probability)**.** A **probability** is a non-negative function $\mathbb{P}$ defined on $\mathcal{F}$ such that

(i) $\mathbb{P}(S) = 1$;

(ii) (*countable additivity*) if $A_1, A_2, \cdots \in \mathcal{F}$ are pairwise disjoint, then

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

The triple $(S, \mathcal{F}, \mathbb{P})$ is a **probability space**.

That is the entire foundation. Everything else in the course is a consequence.

4.3 Properties of probability

Theorem 4.11 (Basic properties)**.** *For events $A, B, A_1, \ldots, A_n$:*

(i) $\mathbb{P}(\emptyset) = 0$*;*

(ii) (finite additivity) if $A_1, \ldots, A_n$ are pairwise disjoint then $\mathbb{P}\big(\bigcup_{i=1}^n A_i\big) = \sum_{i=1}^n \mathbb{P}(A_i)$*;*

(iii) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$*;*

(iv) $\mathbb{P}(A - B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$*;*

(v) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$*.*

Proof.

(i) Write $S = S \cup \emptyset \cup \emptyset \cup \cdots$, a disjoint union. Countable additivity gives $\mathbb{P}(S) = \mathbb{P}(S) + \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) + \cdots$, so $\sum_{i \geq 1} \mathbb{P}(\emptyset) = 0$, forcing $\mathbb{P}(\emptyset) = 0$.

(ii) Pad the finite union with empty sets: $\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1 \cup \cdots \cup A_n \cup \emptyset \cup \cdots) = \sum_{i=1}^n \mathbb{P}(A_i) + \sum \mathbb{P}(\emptyset) = \sum_{i=1}^n \mathbb{P}(A_i)$.

(iii) $S = A \cup A^c$ disjointly, so $1 = \mathbb{P}(A) + \mathbb{P}(A^c)$.

(iv) $A = (A - B) \cup (A \cap B)$ disjointly.

(v) $A \cup B = (A - B) \cup B$ disjointly, so by (iv) $\mathbb{P}(A \cup B) = \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B)$. □

Theorem 4.12 (Inclusion–exclusion)**.**

$$\mathbb{P}\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} \mathbb{P}(A_1 \cap \cdots \cap A_n).$$

Proof. Induction on n, with (v) as the base case. (See Homework 1.) □

4.4 Finite and countable sample spaces

The axioms cover both cases classical probability could not touch.

4.4.1 A loaded die

Example 4.13. A die is weighted so that

$$\mathbb{P}(1) = \tfrac{1}{2}, \quad \mathbb{P}(2) = \tfrac{1}{6}, \quad \mathbb{P}(3) = \mathbb{P}(4) = \mathbb{P}(5) = \mathbb{P}(6) = \tfrac{1}{12}.$$

This is legitimate: the six numbers are non-negative and $\frac{1}{2} + \frac{1}{6} + 4 \cdot \frac{1}{12} = \frac{1}{2} + \frac{1}{6} + \frac{1}{3} = 1$. Take $\mathcal{F} = \mathcal{P}(S)$; then for any event, additivity gives the answer, e.g.

$$\mathbb{P}(\{1, 3, 4\}) = \tfrac{1}{2} + \tfrac{1}{12} + \tfrac{1}{12} = \tfrac{8}{12} = \tfrac{2}{3}.$$

Classical probability cannot express this at all — it assumes all outcomes are equally likely, and here they are not. The axioms never ask *where* the numbers came from, only that they are non-negative and add to 1.

4.4.2 Countably infinite sample spaces

Let $S = \{\omega_1, \omega_2, \dots\}$ be countably infinite and take $\mathcal{F} = \mathcal{P}(S)$. Assign each outcome a number $\mathbb{P}(\omega_i) \geq 0$ with $\sum_{i=1}^{\infty} \mathbb{P}(\omega_i) = 1$; then for any event A,

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\omega_i).$$

No equal-likelihood assumption is needed — and indeed *no* uniform assignment on a countably infinite set is possible.

Example 4.14 (Toss until the first head)**.** The outcome is the number of tosses, so $S = \{1, 2, 3, \dots\}$ and $\mathbb{P}(n) = \dfrac{1}{2^n}$. This is legitimate because $\sum_{n=1}^{\infty} 2^{-n} = 1$. For instance $\mathbb{P}(\text{even number of tosses}) = \sum_{k\geq 1} 2^{-2k} = \frac{1}{3}$.