MA 361: Probability Theory

Naman Mishra

August 2024

Contents

The course

Grading

Lecture 1.
Thursday
August 1

- Homework: 20%
- Two midterms: 15% each
- Final: 50%

Chapter I

Review of discrete probability

Definition I.1 (Discrete probability space)**.** A discrete probability space is a pair (Ω, p) where Ω is a finite or countable set called *sample space* and $p : \Omega \to [0, 1]$ is a function giving the *elementary probabilities* of each $\omega \in \Omega$ such that

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

Examples.

- "Toss a fair n times" is modeled as

  $$\Omega = \{0, 1\}^n$$

  with

  $$p(\omega) \equiv \frac{1}{2^n}.$$

- "Throw r balls randomly into m bins" is modeled as

  $$\Omega = [m]^r$$

  with p given by the multinomial distribution (assuming uniformity).

- "A box has N coupons, draw one of them."

  $$\Omega = [N]$$
  $$p = \omega \mapsto \frac{1}{N}.$$

- "Toss a fair coin countably many times." The set of outcomes is clear: $\Omega = \{0, 1\}^{\mathbb{N}}$. What about the elementary probabilities?

  Probabilities of some events are also fairly intuitive. For example, the event

  $$A = \{\underline{\omega} \in \Omega \mid \omega_1 = 1, \omega_2 = 1, \omega_3 = 0\}$$

  has probability 1/8. Similarly $B = \{\underline{\omega} \in \Omega \mid \omega_1 = 1, \omega_2 = 0\}$ has probability 1/4. Where does this come from?

What about this event:

$$C = \{\underline{\omega} \in \Omega \;\Big|\; \frac{1}{n}\sum_{i=1}^{n} \omega_i \to 0.6\}$$

What about:

$$D = \{\underline{\omega} \in \Omega \;\Big|\; \sum_{i=1}^{n} \omega_i = \frac{n}{2} \text{ for infinitely many } n\}$$[1]

- "Draw a number uniformly at random from $[0,1]$." Ω is obviously $[0,1]$. Again some events have obvious probabilities.

  $$A = [0.1, 0.3] \implies \mathbf{P}(A) = 0.2$$

  Similarly

  $$B = [0.1, 0.2] \cup (0.7, 1) \implies \mathbf{P}(B) = 0.4$$

  What about $C = \mathbb{Q} \cap [0,1]$? What about D, the $\frac{1}{3}$-Cantor set?

  The $\frac{1}{3}$-Cantor set is given by the limit of the following sequence of sets.

  $$\begin{aligned} K_0 &= [0,1] \\ K_1 &= [0,1/3] \cup [2/3,1] \\ K_2 &= [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1] \\ &\vdots \end{aligned}$$

  where each K_{n+1} is obtained by removing the middle third of each interval in K_n.[2]

The resolution for the above examples is achieved by taking the 'obvious' cases as definitions.

What we agree on:

(∗) $\mathbf{P}([a,b]) = b - a$ for all $0 \le a \le b \le 1$.

What we wish for:

(#1) If $A \cap B = \varnothing$, then $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.

(#2) If $A_n \downarrow A$, then $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$.

Question: Does there exist a $\mathbf{P}\colon 2^{[0,1]} \to [0,1]$ that satisfies (∗), (#1) and (#2)? **No.**

Question: Does there exist a $\mathbf{P}\colon 2^{[0,1]} \to [0,1]$ that satisfies (∗), (#1) and even *translational invariance*? **Yes!**
However, it is not unique.

[1] $\mathbf{P}(C) = 0$ and $\mathbf{P}(D) = 1$.
[2] $\mathbf{P}(C) = \mathbf{P}(D) = 0$.

What about the same for a probability measure on $[0, 1]^2$ that is translation and rotation invariant?
What about $[0, 1]^3$?[3]

Lack of uniqueness is a disturbing issue. The way out is the following: restrict the class of sets on which $\mathbf{P}$ is defined to a σ-algebra.

[3]The Banach-Tarski paradox gives a "no" for the 3D case.

Chapter II

Measure-theoretic probability

II.1 σ-algebras

Definition II.1 (σ-algebra)**.** Given a set Ω, a collection $\mathcal{F} \subseteq 2^{\Omega}$ is called a σ-algebra if

(ς1) $\varnothing \in \mathcal{F}$.

(ς2) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$.

(ς3) If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

This gives us a modified question.

Question: Does there exist *any* σ-algebra $\mathcal{F}$ on $[0,1]$ and a function $\mathbf{P}\colon \mathcal{F} \to [0,1]$ that satisfies (#), ($*$) and ($**$)?

Answer: Yes, and it is sort-of unique.

Exercise II.2. *Prove that ($**$) is equivalent to the following: if $(B_n)_{\mathbb{N}}$ are pairwise disjoint, then*

$$\mathbf{P}\Big(\bigcup B_n\Big) = \sum \mathbf{P}(B_n).$$

Solution. ■

A σ-algebra that works for our case is the *smallest* one that contains all intervals.

Exercise II.3. *If $\{\mathcal{F}_i\}_{i\in I}$ are σ-algebras on Ω, then $\bigcap_{i\in I} \mathcal{F}_i$ is also a σ-algebra.*

Proof. $\varnothing$ is in each $\mathcal{F}_i$ and hence in the intersection. If A is in each $\mathcal{F}_i$, then so is A^c. If $A_1, A_2, \ldots$ are in each $\mathcal{F}_i$, then so is $\bigcup_{n=1}^{\infty} A_n$. ■

This allows us to make sense of the word 'smallest' above.

Definition II.4. Let $\mathcal{S} \subseteq 2^\Omega$. The *smallest* σ-algebra containing $\mathcal{S}$ is given by the intersection of all σ-algebras on Ω that contain $\mathcal{S}$. We denote this by $\sigma(\mathcal{S})$.

This will contain $\mathcal{S}$ since 2^Ω itself is a σ-algebra.

Example (Borel σ-algebra). The *Borel σ-algebra* on $[0,1]$ is the smallest σ-algebra containing all intervals in $[0,1]$. It is denoted by $\mathcal{B}_{[0,1]}$.

II.2 Probability spaces

Lecture 2. Tuesday August 6

Definition II.5 (probability space)**.** A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbf{P})$, where Ω is a non-empty set called the *sample space*, $\mathcal{F}$ is a σ-algebra on Ω, and $\mathbf{P}$ is a *probability measure* on $\mathcal{F}$.

A *probability measure* on a σ-algebra $\mathcal{F}$ is a function $\mathbf{P}\colon \mathcal{F} \to [0,1]$ such that $\mathbf{P}(\Omega) = 1$ and

$$\mathbf{P}\left(\bigsqcup_n A_n\right) = \sum_n \mathbf{P}(A_n)$$

for any sequence of pairwise disjoint sets $A_n \in \mathcal{F}$ (countable additivity).

Countable additivity is a stronger condition than finite additivity.

Exercise II.6. *Prove that countable additivity is equivalent to the following two conditions taken together:*

(i) ***finite additivity:*** *if $A \cap B = \varnothing$, then $\mathbf{P}(A \sqcup B) = \mathbf{P}(A) + \mathbf{P}(B)$*

(ii) If $A_n \uparrow A$, then $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$.

Solution. We first show that these follow from countable additivity.

If one takes $A_1 = A$ and $A_n = \varnothing$ for $n \geq 2$, then countable additivity implies $\mathbf{P}(\varnothing) = 0$. Finite additivity is immediate by taking only $A_3, A_4, \ldots$ to be empty.

Let $A_n \uparrow A$. Define $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$. Then $(B_n)_{\mathbb{N}}$ are pairwise disjoint and $\bigcup B_n = A$. By countable additivity,

$$\mathbf{P}(A) = \sum_n \mathbf{P}(B_n).$$

But the partial sums of this series are exactly $\mathbf{P}(A_n)$ by finite additivity. Thus $\mathbf{P}(A) = \lim \mathbf{P}(A_n)$. ■

Where do Ω, $\mathcal{F}$, and $\mathbf{P}$ come from?

Ω is simply the set of all possible outcomes.

II.2.1 The σ-algebra

$\mathcal{F} = 2^\Omega$ and $\mathcal{F} = \{\varnothing, \Omega\}$ are bullshit choices. In reality, $\mathcal{F}$ is always chosen to be the smallest σ-algebra containing some specified sets of interest. That is, for some $\mathcal{S} \subseteq 2^\Omega$, $\mathcal{F} = \sigma(\mathcal{S})$.

This is sometimes called the σ-algebra "generated by" $\mathcal{S}$. However, this can create a misconception. Recall the similar notion of the *span* of a set of vectors. We can define the span of a set $S \subseteq V$ of vectors in two ways:

- (external) the smallest subspace containing S.
- (internal) the set of all linear combinations of vectors in S.

For $\sigma(\mathcal{S})$, there is no "internal" definition. $\sigma(\mathcal{S})$ cannot be generated by unions, intersections, etc. of sets in $\mathcal{S}$.

A frequent choice for $\mathcal{S}$ is the following.

Definition II.7 (Borel σ-algebra)**.** Let (X, d) be a metric space. The *Borel σ-algebra* on X is the smallest σ-algebra containing all open sets in X, and is denoted $\mathcal{B}(X)$.

II.2.2 The probability measure

There is some collection $\mathcal{S} \subseteq \Omega$ for which we know what the probabilities "should" be, $\mathbf{P} \colon \mathcal{S} \to [0, 1]$.

Question II.8. *Does $\mathbf{P}$ extend to a probability measure on $\sigma(\mathcal{S})$? If so, is it unique?*

Uniqueness does not hold.

Example. Let $\Omega = \{1, 2, 3, 4\}$ and $\mathcal{S} = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$. $\mathcal{F} = \sigma(\mathcal{S}) = 2^\Omega$.

Then the probability measures given by

$$\underline{p} = (.25, .25, .25, .25)$$
$$\underline{q} = (.5, 0, .5, 0)$$

agree on $\mathcal{S}$ but differ on $\mathcal{F}$.

When does uniqueness hold?

Uniqueness

Definition II.9 (π-system)**.** A collection $\mathcal{S} \subseteq 2^\Omega$ is a *π-system* if it is closed under finite intersections. That is, for any $A, B \in \mathcal{S}$, $A \cap B \in \mathcal{S}$.

Definition II.10 (λ-system)**.** A collection $\mathcal{C} \subseteq 2^\Omega$ is a λ-*system* if it contains Ω and is closed under

- proper differences: if $A, B \in \mathcal{C}$ and $B \subseteq A$, then $A \setminus B \in \mathcal{C}$.
- increasing limits: if $A_n \in \mathcal{C}$ and $A_n \uparrow A$, then $A \in \mathcal{C}$.

Theorem II.11. *If $\mathcal{F} = \sigma(\mathcal{S})$ where $\mathcal{S}$ is a π-system and P, Q are probability measures on $\mathcal{F}$ that agree on $\mathcal{S}$, then $P = Q$.*

Proof sketch. Consider $\mathcal{G} = \{A \in \mathcal{F} \mid P(A) = Q(A)\}$. Then $\mathcal{G} \supseteq \mathcal{S}$. Further, if $A \in \mathcal{G}$, then $A^c \in \mathcal{G}$ since $P(A^c) = 1 - P(A) = 1 - Q(A) = Q(A^c)$.

If $A, B \in \mathcal{G}$ are disjoint, then

$$P(A \sqcup B) = P(A) + P(B) = Q(A) + Q(B) = Q(A \sqcup B).$$

But how do we deal with A, B not disjoint? We need to show that $A, B \in \mathcal{G} \implies A \cap B \in \mathcal{G}$.

Resolution: Show that $\mathcal{G}$ is a λ-system, and then apply the π-λ theorem. Suppose $A, B \in \mathcal{G}$ with $B \subseteq A$. Then $P(A \setminus B) = P(A) - P(B) = Q(A) - Q(B) = Q(A \setminus B)$. Thus $\mathcal{G}$ is closed under proper differences.

If $A_n \uparrow A$ are in $\mathcal{G}$, then $P(A_n) \uparrow P(A)$ and $Q(A_n) \uparrow Q(A)$. But $P(A_n) = Q(A_n)$ for all n, so $P(A) = Q(A)$. Thus $\mathcal{G}$ is closed under increasing limits.

$\mathcal{G}$ contains Ω since $P(\Omega) = Q(\Omega) = 1$.

Thus by the π-λ theorem, $\mathcal{G}$ is a σ-algebra and thus $\mathcal{G} \supseteq \mathcal{F}$. ■

Theorem II.12 (π-λ theorem)**.** *Let $\mathcal{S}$ be a π-system and $\mathcal{C}$ be a λ-system. If $\mathcal{C} \supseteq \mathcal{S}$, then $\mathcal{C} \supseteq \sigma(\mathcal{S})$.*

This is due to Sierpiński and Dynkin.

What about existence?

Existence

In the general case, obviously not. Consider $\Omega = [0, 1]$ with

$$\mathcal{S} = \{(0, \frac{1}{2}), (0, \frac{1}{4}), (\frac{1}{4}, \frac{1}{2})\}$$
$$\mathbf{P}(a, b) = (b - a)^2.$$

Then the sum of $\mathbf{P}(0, \frac{1}{4})$ and $\mathbf{P}(\frac{1}{4}, \frac{1}{2})$ is less that $\mathbf{P}(0, \frac{1}{2})$.

Let us impose some necessary conditions.

Definition II.13 (Algebra)**.** A collection $\mathcal{A} \subseteq 2^\Omega$ is an *algebra* if it is closed under complements and finite unions.

Theorem II.14 (Carathéodory's extension theorem)**.** *Let $\mathcal{S}$ be an algebra. Assume that $P\colon \mathcal{S} \to [0,1]$ is countably additive. Then there exists an extension of P to a probability measure $\mathbf{P}$ on $\mathcal{F} = \sigma(\mathcal{S})$.*

Corollary II.15. *The above extension is unique.*

Proof. An algebra is a π-system. Theorem II.11 applies. ■

II.3 Existence of Lebesgue measure

Theorem II.16. *There is a unique probability measure λ on $([0,1], \mathcal{B}_{[0,1]})$ such that*

$$\lambda([a,b]) = b - a \quad \textit{for all } 0 \le a \le b \le 1.$$

Proof. Let $\Omega = [0,1)$.

Let $\mathcal{S}_0 = \{[a,b) \mid 0 \le a \le b \le 1\}$. Half-open intervals are nice because they are closed under complements: $[a,b)^c = [0,a) \sqcup [b,1)$.

Let

$$\mathcal{S} = \{I_1 \sqcup \dots \sqcup I_k \mid k \ge 1, I_j \in \mathcal{S}_0\}$$

be the collection of all finite disjoint unions of half-open intervals. This is an algebra. ■

Lecture 3. Thursday August 8

Definition II.17. Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two sets with σ-algebras. A function $T\colon \Omega \to \Omega'$ is *measurable* if

$$T^{-1}(B) \in \mathcal{F} \quad \text{for all } B \in \mathcal{F}'.$$

II.3.1 Push forward

Lemma II.18. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\Omega', \mathcal{F}')$ be a measurable space. Let $T\colon \Omega \to \Omega'$ be measurable. Then $Q := P \circ T^{-1}$ is a probability measure on $\mathcal{F}'$.*

Proof. Notice that $T^{-1}(B)^c = T^{-1}(B^c)$ and that if B_1 and B_2 are disjoint, so are $T^{-1}(B_1)$ and $T^{-1}(B_2)$. ■

Definition II.19 (cumulative distributive function)**.** A *cumulative distributive function* (CDF) is a function $F\colon \mathbb{R} \to [0,1]$ such that

(i) (increasing) $x \le y \implies F(x) \le F(y)$

(ii) (right-continuous) $\lim_{h \searrow 0} F(x+h) = F(x)$

(iii) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

Lecture 4

Lecture 4. Tuesday August 13

Let $P(\mathbb{R})$ be the set of all probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. If $\mu \in P(\mathbb{R})$, then $F_\mu(x) := \mu(-\infty, x]$ is a CDF (increasing, right-continuous with $F(-\infty) = 0$, $F(\infty) = 1$)).

Theorem II.20. *Given a CDF $F\colon \mathbb{R} \to [0,1]$, there exists a unique probability measure $\mu \in P(\mathbb{R})$ such that $\mu(-\infty, x] = F(x)$ for all $x \in \mathbb{R}$.*

Proof. Consider $((0,1), \mathcal{B}, \lambda)$ and define

$$\begin{aligned} T\colon (0,1) &\to \mathbb{R} \\ u &\mapsto \inf\{x \in \mathbb{R} : F(x) \geq u\} \end{aligned}$$

The set is non-empty since $F(x) \to 1$ as $x \to \infty$. Moreover, T is increasing since

$$\{x \in \mathbb{R} : F(x) \geq u\} \subseteq \{x \in \mathbb{R} : F(x) \geq v\}$$

whenever $u \leq v$. T is left-continuous.

Finally, $T(u) \leq x \iff F(x) \geq u$. (This is reminiscent of the inverse property: $T(u) = x \iff F(x) = u$.) If $F(x) \geq u$, then $x \in F^{-1}[u, 1)$, so $T(u) \leq x$. If $T(u) \leq x$, then $x + \frac{1}{n} \in F^{-1}[u, 1)$ for all $n \in \mathbb{N}$. By right-continuity, $F(x) \geq u$.

Now T is Borel-measurable, so

$$\mu := \lambda \circ T^{-1}$$

is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Further, $\mu(-\infty, x] = \lambda(T^{-1}(-\infty, x]) = \lambda(0, F(x)] = F(x)$.

Uniqueness if by the π-system thingy. ■

Examples.

- Take $f\colon \mathbb{R} \to [0, \infty)$ measurable whose total integral is 1. Then $F = x \mapsto \int_{-\infty}^{x} f(u)\,\mathrm{d}u$ is a CDF.

- (Cantor measure) Consider the $\frac{1}{3}$-Cantor set $K = K_1 \cap K_2 \cap \dots$ where

$$\begin{aligned} K_1 &= \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right] \\ K_2 &= \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right] \\ &\vdots \end{aligned}$$

Notice that

$$K = \{x \in [0,1] : x = \sum_{n=1}^{\infty} \frac{x_n}{3^n}, x_n = 0 \text{ or } 2\}.$$

We can construct the measurable function

$$\begin{aligned} T\colon [0,1] &\to \mathbb{R} \\ \sum_{n=1}^{\infty} \frac{x_n}{2^n} &\mapsto \sum_{n=1}^{\infty} \frac{2x_n}{3^n} \end{aligned}$$

where we are considering the non-terminating binary expansion of x on the left. It is obvious that T maps only to K. Since $T^{-1}(K) = [0,1]$, we have that $\mu(K) = 1$. However, $\lambda(K) = 0$. Thus the CDF cannot arise from a density. However, the CDF is continuous!

- (just for fun) Fix a $\theta > 2$ and define

$$\begin{aligned} T_\theta\colon [0,1] &\to [0,1] \\ \sum_{n=1}^{\infty} \frac{x_n}{2^n} &\mapsto \sum_{n=1}^{\infty} \frac{x_n}{\theta^n} \end{aligned}$$

define $\mu_\theta = \lambda \circ T_\theta^{-1}$. $\mu_2 = \lambda$. It is known that for $\theta > 2$, μ_θ has no density. What about $1 < \theta < 2$? This is an open problem. "Bernoulli convolution problem".

II.3.2 Structure of $P(\Omega, \mathcal{F})$

What is the structure of $P(\Omega, \mathcal{F})$? Is it a vector space? A group?

One thing to note is that $P(\Omega, \mathcal{F})$ is convex. That is, given any $\mu, \nu \in P(\Omega, \mathcal{F})$ and $0 \le t \le 1$, $(1-t)\mu + t\nu \in P(\Omega, \mathcal{F})$. This is called a *mixture* of μ and ν.

We would like to study *closeness* of probability measures. Consider a computer generating a random number between 0 and 1, by generating a sequence of 8 random bits. The computer is actually sampling from the uniform distribution

$$\mu_{2^8} = \mathrm{Unif}\left\{ \frac{0}{2^8}, \frac{1}{2^8}, \ldots, \frac{2^8 - 1}{2^8} \right\}.$$

However, we do accept μ as an approximation of λ. We will thus attempt to define a *metric* on $P(\mathbb{R})$.

Attempt 1. (total variation distance) Define

$$d(\mu, \nu) = \sup_{A \in \mathcal{B}_\mathbb{R}} |\mu(A) - \nu(A)|.$$

This does not work for out for our use case, as

$$d(\mu_{2^8}, \lambda) = 1.$$

Attempt 2. (Kolmogorov-Smirnov metric) Choose a suitable $\mathcal{C} \in \mathcal{B}_\mathbb{R}$ and

define
$$d(\mu,\nu) = \sup_{A\in\mathcal{C}} |\mu(A) - \nu(A)|.$$
$\mathcal{C}$ should be "measure-determining".

Attempt 3. (Lévy metric)

$$\begin{aligned} d(\mu,\nu) = \inf\{\varepsilon > 0 : F_\mu(x+\varepsilon) + \varepsilon \geq F_\nu(x) \text{ and} \\ F_\nu(x+\varepsilon) + \varepsilon \geq F_\mu(x) \text{ for all } x \in \mathbb{R}\}. \end{aligned}$$

This is symmetric by sheer obviousness. For $\triangle$, consider three measures μ, ν, ρ.

$$\begin{aligned} t &> d(\mu,\nu) && \implies F_\mu(x+t) + t \geq F_\nu(x) \\ s &> d(\nu,\rho) && \implies F_\nu(x+s) + s \geq F_\rho(x) \end{aligned}$$

Thus

$$F_\mu(x+t+s) + t + s \geq F_\nu(x+s) + t \geq F_\rho(x)$$

Thus $t+s \geq d(\mu,\rho)$. $\triangle$ holds.

Finally, suppose $d(\mu,\nu) = 0$. Let $\varepsilon_n \downarrow 0$ be a sequence such that $F_\mu(x+\varepsilon_n) + \varepsilon_n \geq F_\nu(x)$ for all x for all n. Taking limits, we have $F_\mu(x) \geq F_\nu(x)$ by right-continuity. By symmetry, $F_\mu(x) = F_\nu(x)$.