Mathematical Toolkit **Autumn 2023**

Lecture 11: October 31, 2023

Lecturer: Madhur Tulsiani

Probability theory is a mathematical framework used to model uncertainty and variability in nature. It is by no means the only contender for this role, but has weathered many trials through time. A good deal of probability theory was developed long before being formalized in the way that we're familiar with now, which is due to Kolmogorov. One could cite the works of Laplace, Poisson, Gauss, to name a few. So in some sense the formalization we present here is not strictly necessary, at least for most simple problems. But it does place the whole field on a very stable foundation, which is also helpful whenever something challenges our grasp of this otherwise intuitive discipline.

1 Basics of probability: the finite case

We recall very briefly the basics of probability and random variables. For a much better and detailed introduction, please see the lecture notes by Terry Tao, linked from the course homepage.

1.1 Probability spaces

Let Ω be a finite set. Let $\nu : \Omega \to [0,1]$ be a function such that

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

We often refer to Ω as a sample space or outcome space and the function ν as a probability distribution on this space. An event can be thought of as a subset of outcomes i.e., any $A \subseteq \Omega$ defines an event, and we define its probability as

$$\mathbb{P}[A] \;=\; \sum_{\omega \in A} \nu(\omega)\,.$$

1.2 Random Variables and Expectation

In a finite probability space, a real-valued random variable over Ω is any function $X : \Omega \to \mathbb{R}$. So a random variable is technically neither random (it's quite deterministic) nor a variable (it's a function), but it's a terminology that has stuck.

In a finite probability space, we define the expectation of a random variable X as:

$$\mathbb{E}[X] := \sum_{\omega \in \Omega} \nu(\omega) \cdot X(\omega).$$

An extremely useful fact about expectation is that is a linear transformation from the space of random variables to $\mathbb{R}$. In particular, if X and Y are random variables, then $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, and $\mathbb{E}[c \cdot X] = c \cdot \mathbb{E}[X]$.

Proposition 1.1 (Linearity of Expectation) *For any two random variables X and Y, $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, and $\mathbb{E}[c \cdot X] = c \cdot \mathbb{E}[X]$.*

Proof: This follows directly from the definition.

$$\mathbb{E}[X+Y] = \sum_{\omega \in \Omega} \nu(\omega) \cdot (X(\omega) + Y(\omega)) = \sum_{\omega \in \Omega} \nu(\omega) \cdot X(\omega) + \sum_{\omega \in \Omega} \nu(\omega) \cdot Y(\omega) = \mathbb{E}[X] + \mathbb{E}[Y].$$

The proof for $\mathbb{E}[c \cdot X] = c \cdot \mathbb{E}[X]$ is similar. ■

Example: Card shuffling Suppose we unwrap a fresh deck of cards and shuffle it until the cards are completely random. How many cards do we expect to be in the same position as they were at the start? To solve this, let's think formally about what we are asking. We are looking for the expected value of a random variable X denoting the number of cards that end in the same position as they started. We can write X as a sum of indicator random variables X_i, one for each card, where $X_i = 1$ if the ith card ends in position i and $X_i = 0$ otherwise. These X_i are easy to analyze: $\mathbb{P}(X_i = 1) = 1/n$ where n is the number of cards. $\mathbb{P}(X_i = 1)$ is also $\mathbb{E}[X_i]$. Now we use linearity of expectation:

$$\mathbb{E}[X] = \mathbb{E}[X_1 + \ldots + X_n] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n] = 1.$$

So, this is interesting: no matter how large a deck we are considering, the expected number of cards that end in the same position as they started is 1.

1.3 Conditioning

Conditioning on an event A is equivalent to restricting the probability space to the set A. We then consider the conditional probability measure ν_A defined as

$$\nu_A(\omega) = \begin{cases} \frac{\nu(\omega)}{\mathbb{P}[A]} & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}.$$

Thus, one can define the conditional probability of an event B as

$$\mathbb{P}[B \mid A] \;=\; \sum_{\omega \in B} \nu_A(\omega) \;=\; \sum_{\omega \in A \cap B} \frac{\nu(\omega)}{\mathbb{P}[A]} \;=\; \frac{\mathbb{P}[A \wedge B]}{\mathbb{P}[A]}\,.$$

For a random variable X and an event A, we similarly define the *conditional expectation* of X given A as

$$\mathbb{E}[X \mid A] \;=\; \sum_{\omega} \nu_A(\omega) \cdot X(\omega)\,,$$

with ν_A as above. Verify the following identities.

Proposition 1.2 (Total Probability and Total Expectation) *Let Ω be a finite "outcome space" with probabiliy measure ν. Let $A, B \subseteq \Omega$ be events, and $X : \Omega \to \mathbb{R}$ be a random variable. Then*

1. $\mathbb{P}[B] \;=\; \mathbb{P}[A] \cdot \mathbb{P}[B \mid A] + \mathbb{P}[A^{\mathsf{c}}] \cdot \mathbb{P}[B \mid A^{\mathsf{c}}]$,
2. $\mathbb{E}[X] \;=\; \mathbb{P}[A] \cdot \mathbb{E}[X \mid A] + \mathbb{P}[A^{\mathsf{c}}] \cdot \mathbb{E}[X \mid A^{\mathsf{c}}]$.

1.4 Independence

Now that we have the notion of conditioning, we can define independence. Two non-zero probability events A and B are independent if $\mathbb{P}[A \mid B] = \mathbb{P}[A]$. One can verify that this is equivalent to $\mathbb{P}[B \mid A] = \mathbb{P}[B]$. In other words, restricting to one event does not change the probability of the other event. Independence is a joint property of events and the probability measure: one cannot make judgment about independence without knowing the probability measure.

Two random variables X and Y defined on the same finite probability space are defined to be independent if

$$\mathbb{P}[X = x \mid Y = y] = \mathbb{P}[X = x]$$

for all non-zero probability events $\{X = x\} := \{\omega : X(\omega) = x\}$ and $\{Y = y\} := \{\omega : Y(\omega) = y\}$.

In general (not necessarily finite) probability spaces, random varaibles X and Y are said to be independent if

$$\mathbb{P}\{X \in X(A) \mid Y \in Y(B)\} = \mathbb{P}\{X \in X(A)\}\,,$$

for all non-zero probability events A and B, where $X(A)$ denotes the images of $A \subseteq \Omega$ under the function X, and $\{X \in X(A)\} \;:=\; \{\omega : \exists \omega' \in A,\; X(\omega) = X(\omega')\}$.