Mathematical Toolkit **Autumn 2018**

Lecture 11: November 8, 2018

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 $\mu : \Omega \to [0, 1]$ be a function such that

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

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

$$\mathbb{P}[E] = \sum_{\omega \in E} \mu(\omega).$$

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:

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

It is important to observe that $\mathbb{E}$ is a linear transformation from the space of random variables to $\mathbb{R}$ (check that if X and Y are random variables, then $\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$).

1.3 Conditioning

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

$$\mu_E(\omega) = \begin{cases} \frac{\mu(\omega)}{\mathbb{P}[E]} & \text{if } \omega \in E \\ 0 & \text{otherwise} \end{cases} .$$

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

$$\mathbb{P}[F \mid E] = \sum_{\omega \in F} \mu_E(\omega) = \sum_{\omega \in E \cap F} \frac{\mu(\omega)}{\mathbb{P}[E]} = \frac{\mathbb{P}[E \wedge F]}{\mathbb{P}[E]} .$$

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

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

with μ_E as above. Verify the following identities.

Proposition 1.1 (Total Probability and Total Expectation) *Let Ω be a finite "outcome space" with probabiliy measure μ. Let $E, F \subseteq \Omega$ be events, and $X : \Omega \to \mathbb{R}$ be a random variable. Then*

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

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 $\{Y = y\} := \{\omega : Y(\omega) = y\}$.

Note that the notion of independence defined above is pairwise. It is imprecise to say, for $n > 2$, that events $A_1, \cdots, A_n$ or random variables $X_1, \cdots, X_n$ are independent. Of course one could require any two pairs of events or random variables to be independent. But what is usually meant with random variables is that *any* (measurable) function of *any* two disjoint subsets of the random variables, bringing us to a collection of random variable pairs, are independent. For the finite Ω case, this means that if we take *any* disjoint subsets $I, J \subseteq [n]$ and consider the random variables

$$X_I = \{X_i\}_{i \in I} \quad \text{and} \quad X_J = \{X_i\}_{i \in J},$$

then X_I and X_J are independent. Note that this is a stronger definition than requiring that every pair of variables X_i, X_j are independent. Also, we cheated above and defined random variables X_I and X_J which take values not in $\mathbb{R}$ but in $\mathbb{R}^I$ and $\mathbb{R}^J$, but the independence for random variables (on a finite Ω) taking values in any space is defined similarly.

1.5 The countable case

Everything defined above can also be extended to countable spaces but we need to be careful about the convergence of the above summations.