1. PROBABILITY SPACE

"Random experiment" was a non-mathematical term used to describe physical situations with more than one possible outcome, for instance, "toss a fair coin and observe the outcome". In probability, although we sometimes use the same language, it is only as a quick substitute for a mathematically meaningful and precise phrasing. Consider the following examples.

(1) *"Draw a random integer from 1 to 100. What is the probability that you get a prime number?"* Mathematically, we just mean the following. Let $\Omega = \{1, 2 \ldots, 100\}$, and for each $\omega \in \Omega$, we set $p_\omega = \frac{1}{100}$. Subsets $A \subset \Omega$ are called 'events' and for each subset we define $\mathbf{P}(A) = \sum_{\omega \in A} p_\omega$. In particular, for $A = \{2, 3, 5, \ldots, 97\}$, we get $\mathbf{P}(A) = \frac{1}{4}$.

This is the setting for all of discrete probability. We have a finite or countable set Ω called sample space, and for each $\omega \in \Omega$ a number $p_\omega \geq 0$ is specified, so that $\sum_\omega p_\omega = 1$. For any $A \subset \Omega$, one defines its probability to be $\mathbf{P}(A) := \sum_{\omega \in A} p_\omega$. The whole game is to calculate probabilities of interesting events! The difficulty is of course that the set Ω and probabilities p_ω may be defined by a property which makes it hard to calculate probabilities.

Example 1. Fix $n \geq 1$ and let Ω be the set of all self-avoiding paths on length n in $\mathbb{Z}^2$ starting from $(0,0)$. That is,
$$\Omega = \{\omega = (\omega_0, \ldots, \omega_n) \; : \; \omega_i \in \mathbb{Z}^2, \omega_0 = (0,0), \omega_i - \omega_{i-1} \in \{\pm \mathbf{e}_1, \pm \mathbf{e}_2\}\}.$$
Then let $p_\omega = \frac{1}{\#\Omega}$. One interesting event is $A = \{\omega \; : \; \|\omega_n\| < n^{0.6}\}$. Far from finding $\mathbf{P}(A)$, it has not been proved to this day whether for large n, the value $P(A)$ is close to zero or one!

(2) *"Draw a number at random from the interval $[0,1]$. What is the probability that it is less than $\frac{1}{2}$? That it is rational? That its decimal expansion contains no 7?"* For the first question it seems that the answer must be $\frac{1}{2}$, but the next two questions motivate us to think more deeply about the meaning of such an assertion.

Like before we may set $\Omega = [0,1]$. What is p_ω? 'Intutively' it seems that the probability of the number falling in an interval $[a,b] \subset [0,1]$ should be $b - a$ and that forces us to set $p_\omega = 0$ for every ω. But then we cannot possibly get $\mathbf{P}([a,b])$ as $\sum_{\omega \in [a,b]} p_\omega$, even if such an uncountable sum had any meaning! So what is the basis for asserting that $\mathbf{P}[a,b] = b - a$?! Understanding this will be the first task.

A first attempt: Let us *define* the probability of any set $A \subset [0,1]$ to be the length of that set. We understand the length of an interval, but what is the length of the set of rational numbers? irrational numbers? A seemingly reasonable idea is to set
$$\mathbf{P}_*(A) = \inf \left\{ \sum_k |I_k| \; : \; \text{each } I_k \text{ is an interval and } \{I_k\} \text{ a countable cover for } A \right\}.$$

Then perhaps, $\mathbf{P}_*(A)$ should be the probability of A for every subset $A \subset [0,1]$. This is at least reasonable in that $\mathbf{P}_*[a,b] = b - a$ for any $[a,b] \subset [0,1]$. But we face an unexpected problem. One can find[1] A such that $\mathbf{P}_*(A) = 1$ and $\mathbf{P}_*(A^c) = 1$ and that violates one of the basic requirements of probability, that $\mathbf{P}_*(A \cup A^c)$ be equal to $\mathbf{P}_*(A) + \mathbf{P}_*(A^c)$! You may object that our definition of $\mathbf{P}_*$ was arbitrary, may be another definition works? Before tackling that question, we should make precise what all we properties we require probabilities to satisfy. This we do next, but let us record here that there will be two sharp differences from discrete probability.

(a) One cannot hope to define $\mathbf{P}(A)$ for all $A \subset [0,1]$, but only for a rich enough class of subsets! These will be called events.

(b) One does not start with elementary probabilities p_ω and then compute $\mathbf{P}(A)$, but probabilities of all events are part of the specification of the probability space! (If all probabilities are specified at the outset, what does a probabilist do for a living? Hold that thought till the next lecture!).

Now we define the setting of probability in abstract and then return to the second situation above.

Definition 2. A probability space is a triple $(\Omega, \mathcal{F}, \mathbf{P})$ where

(1) The *sample space* Ω is an arbitrary set.

(2) The σ-*field* or σ-*algebra* $\mathcal{F}$ is a set of subsets of Ω such that (i) $\emptyset, \Omega \in \mathcal{F}$, (ii) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, (iii) if $A_n \in \mathcal{F}$ for $n = 1, 2 \ldots$, then $\cup A_n \in \mathcal{F}$. In words, $\mathcal{F}$ is closed under complementation and under countable unions, and contains the empty set. Elements of $\mathcal{F}$ are called *measurable sets*.

---

[1]Not obvious!

(3) The *probability measure* is any function $\mathbf{P} : \mathcal{F} \to [0,1]$ is such that if $A_n \in \mathcal{F}$ and are pairwise disjoint, then $\mathbf{P}(\cup A_n) = \sum \mathbf{P}(A_n)$ (countable additivity) and such that $\mathbf{P}(\Omega) = 1$. $\mathbf{P}(A)$ is called *the probability of A*.

Measurable sets are what we call *events* in probability theory. It is meaningless to ask for the probability of a subset of Ω that is not measurable. The σ-field is closed under many set operations and the usual rules of probability also hold. If one allows $\mathbf{P}$ to take values in $[0,\infty]$ and drops the condition $\mathbf{P}(\Omega) = 1$, then it is just called a *measure*. Measures have the same basic properties as probability measures, but probabilistically crucial concepts of *independence* and *conditional probabilities* (to come later) don't carry over to general measures and that is mainly what makes probability theory much richer than general measure theory.

Exercise 3. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space.
  (1) $\mathcal{F}$ is closed under finite and countable unions, intersections, differences, symmetric differences. Also $\Omega \in \mathcal{F}$.
  (2) If $A_n \in \mathcal{F}$, then $\limsup A_n := \{\omega : \omega$ belongs to infinitely many $A_n\}$ and $\liminf A_n := \{\omega : \omega$ belongs to all but finitely many $A_n\}$ are also in $\mathcal{F}$. In particular, if A_n increases or decreases to A, then $A \in \mathcal{F}$.
  (3) $\mathbf{P}(\phi) = 0$, $\mathbf{P}(\Omega) = 1$. For any $A, B \in \mathcal{F}$ we have $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$. If $A_n \in \mathcal{F}$, then $\mathbf{P}(\cup A_n) \leq \sum \mathbf{P}(A_n)$.
  (4) If $A_n \in \mathcal{F}$ and A_n increases (decreases) to A, the $\mathbf{P}(A_n)$ increases (decreases) to $\mathbf{P}(A)$.

Some examples of probability spaces.

Example 4. Let Ω be a finite or countable set. Let $\mathcal{F}$ be the collection of all subsets of Ω. Then $\mathcal{F}$ is a σ-field. Given any numbers p_ω, $\omega \in \Omega$ that add to 1, we set $\mathbf{P}(A) = \sum_{\omega \in A} p_\omega$. Then $\mathbf{P}$ is a probability measure. More generally, let Ω be any set and let $R \subset \Omega$ be a countable set. Let $\mathcal{F}$ be the powerset of Ω. Fix nonnegative numbers p_x, $x \in R$ that add to 1. Then define $\mathbf{P}(A) = \sum_{x \in R \cap A} p_x$. This is a probability measure on $\mathcal{F}$.

This means that a discrete measure, say Binomial distribution, can be considered as a p.m. on $\{1, 2, \ldots, n\}$ or on $\mathbb{R}$. The problem of not being able to define probability for all subsets does not arise when the p.m. is so simple.

Example 5. Let Ω be an arbitrary set. Let $\mathcal{F} = \{A \subset \Omega : \text{either } A \text{ or } A^c \text{ is countable}\}$ (where 'countable' includes finite and empty sets). Define $\mathbf{P}(A) = 0$ if A is countable and $\mathbf{P}(A) = 1$ if A^c is countable. This is just a frivolous example of no particular importance.

Exercise 6. Check that $\mathcal{F}$ is a σ-field and that $\mathbf{P}$ is a probability measure on $\mathcal{F}$.

In the most interesting cases, one cannot explicitly say what the elements of $\mathcal{F}$ are, but only require that it is rich enough. Here is an exercise to introduce the important idea of a σ-field generated by a collection of sets.

Exercise 7. Let Ω be a set and let S be a collection of subsets of Ω. Show that there is a smallest sigma filed $\mathcal{F}$ containing all elements of S. That is, if $\mathcal{G}$ is any σ-field of subsets of Ω and $\mathcal{G} \supset S$, then $\mathcal{G} \supset \mathcal{F}$. $\mathcal{F}$ is called *the σ-field generated by S* and often denoted $\sigma(S)$.

Now we come to the most interesting probability spaces for Probability theory.

Example 8. Let $\Omega = [0,1]$. Let S be the collection of all intervals, to be precise let us take all right-closed, left-open intervals $(a,b]$, with $0 \leq a < b \leq 1$ as well as intervals $[0,b]$, $b \leq 1$. If we are trying to make precise the notion of '*drawing a number at random from $[0,1]$*', then we would want $\mathbf{P}(a,b] = b - a$ and $\mathbf{P}[0,b] = b$. The precise mathematical questions can now be formulated as follows. (i) Let $\mathcal{G}$ be the σ-field of all subsets of $[0,1]$. Is there a p.m. $\mathbf{P}$ on $\mathcal{G}$ such that $\mathbf{P}(a,b] = b - a$ and $\mathbf{P}[0,b] = b$ for all $0 \leq a < b \leq 1$? If the answer is 'No', we ask for the less ambitious (ii) Is there a smaller σ-field large enough to contain all interval $(a,b]$, say $\mathcal{F} = \sigma(S)$ such that $\mathbf{P}(a,b] = b - a$?

The answer to the first question is 'No', which is why we need the notion of σ-fields, and the answer to the second question is 'Yes', which is why probabilists still have their jobs. Neither answer is obvious, but we shall answer them in coming lectures.

Example 9. Let $\Omega = \{0,1\}^{\mathbb{N}} = \{\omega = (\omega_1, \omega_2, \ldots) : \omega_i \in \{0,1\}\}$. Let S be the collection of all subsets of Ω that depend on only finitely many co-ordinates (such sets are called cylinders). More precisely, a cylinder set is of the form $A = \{\omega : \omega_{k_1} = \varepsilon_1, \ldots \omega_{k_n} = \varepsilon_n\}$ for some given $n \geq 1$, $k_1 < k_2 < \ldots < k_n$ and $\varepsilon_i \in \{0,1\}$ for $i \leq n$.

What are we talking about? If we want to make precise the notion of '*toss a coin infinitely many times*', then clearly Ω is the sample space to look at. It is also desirable that elements of S be in the σ-field as we should be

able to ask questions such as 'what is the chance that the fifth, seventh and thirtieth tosses are head, tail and head respectively' which is precisely asking for the probability of a cylinder set.

If we are '*tossing a coin with probability p of turning up Head*', then for a cylinder set $A = \{\omega : \omega_{k_1} = \varepsilon_1, \ldots \omega_{k_n} = \varepsilon_n\}$, it is clear that we would like to assign $\mathbf{P}(A) = \prod_{i=1}^{n} p_i^{\varepsilon_i} q^{1-\varepsilon_i}$ where $q = 1 - p$. So the mathematical questions are: (i) If we take $\mathcal{F}$ to be the σ-field of all subsets of Ω, does there exist a p.m. $\mathbf{P}$ on $\mathcal{F}$ such that for cylinder sets $\mathbf{P}(A)$ is as previously specified. (ii) If the answer to (i) is 'No', is there a smaller σ-field, say the one generated by all cylinder sets and a p.m. $\mathbf{P}$ on it with probabilities as previously specified for cylinders?

Again, the answers are 'No' and 'Yes', respectively.

The σ-fields in these two examples can be captured under a common definition.

Definition 10. Let (X, d) be a metric space. The σ-field $\mathcal{B}$ generated by all open balls in X is called the Borel sigma-field of X.

First consider $[0, 1]$ or $\mathbb{R}$. Let $S = \{(a, b]\} \cup \{[0, b]\}$ and let $T = \{(a, b)\} \cup \{[0, b)\} \cup \{(a, 1]\}$. We could also simply write $S = \{(a, b] \cap [0, 1] : a < b \in \mathbb{R}\}$ and $T = \{(a, b) \cap [0, 1] : a < b \in \mathbb{R}\}$. Let the sigma-fields generated by S and T be denoted $\mathcal{F}$ (see example above) and $\mathcal{B}$ (Borel σ-field), respectively. Since

$$(a, b) = \cup_n (a, b - 1/n], \qquad (a, b] = \cap_n (a, b + 1/n), \qquad [0, b] = \cap_n$$

it is clear that $S \subset \mathcal{B}$ and $T \subset \mathcal{F}$. Hence $\mathcal{F} = \mathcal{B}$.

In the countable product space $\Omega = \{0, 1\}^{\mathbb{N}}$ or more generally $\Omega = X^{\mathbb{N}}$, the topology is the one generated by all sets of the form $U_1 \times \ldots \times U_n \times X \times X \times \ldots$ where U_i are open sets in X. Clearly each of these sets is a cylinder set. Conversely, each cylinder set is an open set. Hence $\mathcal{G} = \mathcal{B}$. More generally, if $\Omega = X^{\mathbb{N}}$, then cylinders are sets of the form $A = \{\omega \in \Omega : \omega_{k_i} \in B_i, i \le n\}$ for some $n \ge 1$ and $k_i \in \mathbb{N}$ and some Borel subsets B_i of X. It is easy to see that the σ-field generated by cylinder sets is exactly the Borel σ-field.