Lecture 1. Probability Spaces and Models

Before beginning this lecture, write answers to the following questions. Obviously, you will struggle with some of them, but your attempts to provide answers will help you focus on the important ideas in this lecture.

1. What is a probability space? A discrete probability space?
2. What is a probability mass function (pmf)? What is the difference between a pmf and a probability measure? How is a pmf used to define a probability measure?
3. What is the difference between a probability space and a probability model? How are the words "experiment", "outcome", "event" used in probability modeling?
4. Give an example of a probability model.

Introduction. Let X be a set. A *probability measure* on X is an assignment of numerical values to some of the subsets of X. The sets to which values are assigned are called *measurable*. If $E \subseteq X$ is measurable, its assigned probability measure is denoted $P(E)$. We call $P(E)$ the *probability of* E. A set equipped with a probability measure is called a *probability space*. Probability theory is about probability spaces and how to use them to model situations that involve chance or uncertainty.

The assignment $P(\ )$ must satisfy certain technical conditions. We will describe these in detail later on, as needed. To give you a sense of what the requirements are, let me just say that the following are demanded:

- $P(\emptyset) = 0$ and $P(X) = 1$;
- $0 \le P(E) \le 1$;
- If $E_1, E_2, \ldots$ are disjoint, then $P\Big(\bigcup_{i=1}^{\infty} E_i\Big) = \sum_{i=1}^{\infty} P(E_i)$.

In this course, we often use the symbol Ω (the upper case Greek letter omega) to denote a probability space. Lower case omega—which is written ω—will be used as a variable to denote an unspecified element of Ω. Other lower case Greek letters such as α (alpha), β (beta), γ (gamma) and δ (delta) may be used to denote specific elements of Ω.

1.1. Discrete probability spaces. For the first several weeks of this course, we will be working mainly with one special kind of probability space.

Definition.[1] A *discrete probability space* is a finite or countable[2] set equipped with a *probability mass function* (or "pmf," for short). A pmf on Ω is a function $f : \Omega \to [0, 1]$ such that $\sum_{\omega \in \Omega} f(\omega) = 1$. The number $f(\omega)$ is called the *mass* assigned to ω. Given a pmf, we define a probability measure P by the rule:

$$\text{if } A \subseteq \Omega, \text{ then } P(A) := \sum_{a \in A} f(a). \tag{1}$$

[1] In mathematics, a "definition" is a statement that introduces a new terminology or notation. Definitions need to be read with great care, remembered and referred to whenever the terminology or notation in them is used.

[2] A set is called *countable* if its elements may be put in one-to-one correspondence with the counting numbers $\{1, 2, 3, \ldots\}$.

In other words, **in a discrete probability space, the probability of a set is the sum of the masses of its elements.** Note that *every* subset of a discrete probability space is measurable (i.e., has an assigned probability). In spaces that are not discrete, there are often subsets that are not measurable.

Take note! P takes *subsets of* Ω as arguments—that is, when we write $P(E)$, E must be a subset of Ω. In contrast, f takes *elements of* Ω as arguments. Note that $P\big(\{\omega\}\big) = f(\omega)$ for each $\omega \in \Omega$.

Example. We will illustrate the definition by exhibiting a finite discrete probability space. Let $\Omega = \{a, b, c\}$, and let $f : \Omega \to [0, 1]$ be the function with the following values:

$$f(a) = 1/2,\ f(b) = 1/3,\ f(c) = 1/6.$$

Since $f(a) + f(b) + f(c) = 1$, f is a pmf. The associated probability measure on Ω has the following values:

$$P\big(\{\}\big) = 0,\ P\big(\{a\}\big) = 1/2,\ P\big(\{b\}\big) = 1/3,\ P\big(\{c\}\big) = 1/6,$$

$$P\big(\{a, b\}\big) = 5/6,\ P\big(\{a, c\}\big) = 2/3,\ P\big(\{b, c\}\big) = 1/2,\ P\big(\{a, b, c\}\big) = 1.$$

Facts. Suppose P is a probability measure on a discrete probability space Ω and $E, E_i \subseteq \Omega$. Then, the following are true:

1. $0 \le P(E) \le 1$.
2. $P(\Omega) = 1$ and $P(\emptyset) = 0$.
3. $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.
4. If $E_1, E_2, \ldots$ are disjoint, $P\Big(\bigcup_{i=1}^{\infty} E_i\Big) = \sum_{i=1}^{\infty} P(E_i)$.

You are asked to prove these facts in the problems, below. Facts 1, 2 and 3 involve only arithmetic. Fact 4 refers to limits (the infinite sum) and therefore requires some knowledge from Calculus.

Example. Let $\Omega = \{\alpha, \beta, \gamma, \delta\}$ and let $f : \Omega \to [0, 1]$ be determined by the following table:

ω	α	β	γ	δ
$f(\omega)$	8/15	4/15	2/15	1/15

Then $\sum_{\omega \in \Omega} f(\omega) = f(\alpha) + f(\beta) + f(\gamma) + f(\delta) = 1$, so f is a pmf. If P is the associated probability measure, and $A \subseteq \Omega$, then we calculate $P(A)$ by adding the values of f on the elements of A. For instance, if $A = \{\beta, \gamma, \delta\}$, then $P(A) = f(\beta) + f(\gamma) + f(\delta) = 7/15$.

Problems. In all of the following problems, $\Omega = \{\alpha, \beta, \gamma, \delta\}$, as in the example. In problem 1, we consider the pmf from the example, but in the other problems, we are concerned with different pmfs.

1. List all 16 subsets of Ω. Given the pmf f as in the example, determine the measure P of each subset.

2. A pmf is said to be *uniform* if it assigns the same mass to each element of its domain (i.e., it is a constant function). How many different uniform pmfs are there on Ω? Given a uniform pmf f on Ω, find the the associated probability measure of each subset of Ω.
3. Find a pmf on Ω that is not uniform and is different from the one in the example, and determine the measure of each subset with respect to the probability measure associated to your pmf.
4. Verify Facts 1 through 4—that is, write a sentence or two for each that explains why it follows from the definitions we have made.
5. (Harder.) Prove the Inclusion-Exclusion Principle:

$$\begin{aligned} P\big(E_1 \cup E_2 \cup \cdots \cup E_n\big) = &\sum_{i=1}^{n} P(E_i) \\ &- \sum_{i<j} P\big(E_i \cap E_j\big) \\ &+ \sum_{i<j<k} P\big(E_i \cap E_j \cap E_k\big) \\ &+ \cdots + (-1)^{n+1} P\big(E_1 \cap E_2 \cap \cdots \cap E_n\big) \end{aligned}$$

1.2. Probability models. A probability space is a clear and specific mathematical structure, as described in the previous paragraph. A *probability model* is a probability space employed to represent a situation, process or activity that involves chance or randomness and to make predictions about outcomes. Building a model means finding a probability space to represent a situation, and this typically involves insight and strategic decision-making derived from experience and practice. The way most people learn to do this is by working numerous examples.

The vocabulary of "experiments," "outcomes," and "events" is useful in making the transition from situation to model. An *experiment* is an action that can be repeated over and over to yield an *outcome*. Outcomes belong to an explicitly defined set called "sample space." Subsets of sample space are called *events*. Typically, we build a probability model by taking Ω to be the sample space associated with an experiment. This will all be clarified shortly in examples.

Note that the words "experiment," "outcome," and "event" are NOT used in probability theory with meanings that can be inferred from the common usages. These words belong to the technical jargon of probability theory. Students must train themselves to understand them and use them in the peculiar and unusual—but useful—manner of the technical field.

Example 1. Suppose darts are thrown blindly at a wall that is painted in red, white and blue, with 2/5 of the wall red, 2/5 white and 1/5 blue. If numerous darts are thrown, the we would expect 2/5 to stick in the red region, 2/5 to stick in the white region and 1/5 to stick in the blue region, though on any given throw, we cannot predict with certainty which color it will hit. We can represent this situation by using a probability space that consists of the set $\{r, w, b\}$ (with the letters obviously chosen to stand for the colors) and the pmf $f(r) = 2/5$, $f(w) = 2/5$ and $f(b) = 1/5$. The event of the dart sticking in red is represented by the set $\{r\}$. The event of the dart sticking in some color is represented by the set $\{r, w, b\}$. The event of the dart sticking in some color *other than* blue is represented by the set $\{r, w\}$. We can use the "sample space" $\{r, w, b\}$ to keep track of the probabilities of all these events.

$$\begin{array}{llccccccccc} \text{If} & E & = & \{\} & \{r\} & \{w\} & \{b\} & \{r,w\} & \{r,b\} & \{w,b\} & \{r,w,b\} \\ \text{then} & P(E) & = & 0 & 2/5 & 2/5 & 1/5 & 4/5 & 3/5 & 3/5 & 1 \end{array} .$$

Example 2. Suppose a card is drawn at random from a deck of 52 different cards. The probability space associated to this situation is based on the set $\{A\spadesuit, 2\spadesuit, \ldots, Q\diamondsuit, K\diamondsuit\}$, with one symbol for each card. Each card is equally likely to be chosen, so the mass assigned to each card must be the same. The total of the masses must be 1, so each mass will be 1/52. The event of drawing a king is represented by the set $\{K\spadesuit, K\heartsuit, K\clubsuit, K\diamondsuit\}$, and the associated probability is $4/52 = 1/13$.

Example 3. Suppose you flip a fair coin until a head appears and then stop. The probability of obtaining a head on any flip is 1/2. What is the probability that you will flip the first head on the N^{th} try? Imagine a vast number of people performing this experiment. Half of them will get a head on the first flip. Of the remainder, half will get a head on the second flip, so 1/4 of the people will stop after the second flips. Similarly, 1/8 will stop after the third flips, 1/16 after the 4^{th} and in general $1/2^n$ will stop after the n^{th} flip. A probability model to represent this situation is based on the set $W = \{1, 2, 3, \ldots\}$. To each $n \in W$, we assign the mass $f(n) = 1/2^n$. This meets the requirement for a pmf because the values of f are in $[0, 1]$ and $f(1) + f(2) + f(3) + \cdots = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1$. It is a reasonable model of the coin flipping experiment for the reasons described. Let us compute the probabilities of some events, based on the model. Let N be the number of flips required to get the first head.

- The probability of getting a head on or before the k^{th} flip is denoted $P(N \leq k)$. This is the probability measure of the set $\{\, n \mid n \leq k \,\}$. By rule (1), this is $f(1) + f(2) + \cdots + f(k) = \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^k} = \frac{2^k - 1}{2^k}$.
- The probability that N is even is denoted $P(N \text{ is even})$. By rule (1), this is

$$f(2) + f(4) + f(6) + \cdots = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \cdots = \frac{1}{4}\sum_{i=0}^{\infty}(1/4)^i = \frac{1/4}{1 - (1/4)} = \frac{1}{3}.$$

  (Here we have used the formula for summing a geometric series: If $x \in (-1, 1)$, $1 + x + x^2 + \cdots = 1/(1 - x)$.)

Problems

1. Suppose the wall is painted red, green, purple and gold in proportions 2/15, 1/5, 1/10 and 17/30, respectively. What are the chances of throwing a dart into an LSU color? Into a Christmas color? A non-metallic color?
2. When drawing a single card from a deck, what is the probability of each of the following events:
   a. Drawing a face card?
   b. Drawing a diamond?
   c. Drawing a red card?
   d. Drawing a card with a number between 2 and 8 (inclusive)?
3. Suppose that the coin in Example 3 is not fair, and the probability of getting a head is 1/3, not 1/2. Then, what pmf should we use to model this? What is $P(N \leq k)$ in this case? What is $P(N \text{ is even})$?

Mathematica Project. Please refer to the accompanying Mathematica work book, Notebook1. If you are new to Mathematica, select “Help” on the menu bar.