

Lecture 2. Constructing Probability Spaces

This lecture describes some procedures for constructing probability spaces. We will work exclusively with discrete spaces—usually finite ones. Later, we will see that the operations introduced here actually work for more general spaces.

2.1. Trees.

Imagine the following experiment:

- Pick a library at random.
- Go there, and pick a shelf at random.
- From that shelf, pick a book at random.
- Open the book to a random page.
- On that page, put your finger on a random word.

We can visualize all the different ways in which you might complete this experiment by means of a *tree diagram*, a branching chart that illustrates the options. Draw a dot to represent a starting point. From this dot, draw several line segments—we call them *edges*—one for each of the libraries you might choose. Each edge ends at a dot labelled by a library. From each library dot, draw further edges corresponding to the shelves in that library; these terminate at dots labelled by the shelves. Each shelf dot has edges coming from it which end in dots labelled by the books on that shelf. (If there are several copies of a book on a given shelf, we will use just one edge for all the copies.) Each book dot has edges coming from it that end in page dots, and each page dot has edges from it that end in dots labelled by words. (If a word such as “and” appears several times on the page, we will use just one edge and terminal dot to represent it.) Each time you perform the experiment, you trace a path from the starting point, through a library, a shelf, a book and a page, finally ending on a word. Note that after splitting ways, paths never rejoin. The same word might occur at the end of many different paths, but the dots they are attached to are different.

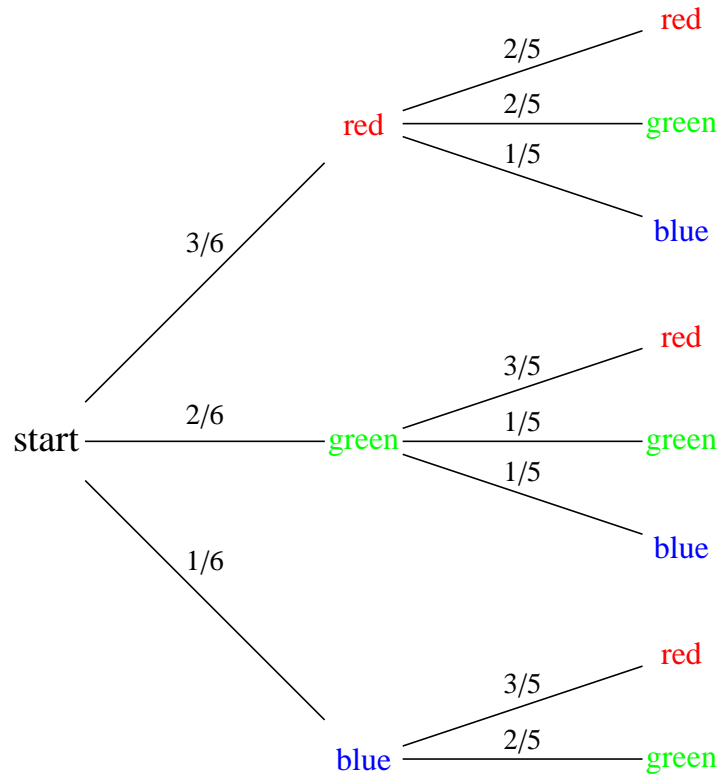
Suppose the choices at each possible stage have probabilities assigned to them. Perhaps there are three libraries: A , B and C , and the chances of choosing them are $1/2$, $1/3$ and $1/6$, respectively. The set of all libraries, in other words, is a probability space. Again, each library has many shelves and each shelf has its own probability of being chosen randomly by someone in the library. In this way, the set of all shelves at library A becomes a probability space, and the same is true for the set of shelves in B and the set of shelves in C . Similarly, the books on each shelf have a probability space to represent their chances of being chosen. The same is true of the pages in each book—they form a probability space. And finally, each page has an associated probability space. We may visualize this by labeling the edges emerging from any dot by the probabilities associated with the choices they represent.

Now consider the collection of all paths from the starting point all the way to a word. A path contains a complete history of all the random choices up to and including the word finally chosen at the end. How can we compute the probability of following a given path? If we have labelled the edges as suggested at the end of the previous paragraph, then we

simply multiply the edge labels as we walk along the path. The justification is that the labels on the edges show the proportion of times each of the options will be taken. For example, if Library A has a $1/2$ chance of being chosen and if a given shelf in library A has a $1/100$ chance of being chosen, then at the start there is a $1/200$ chance of arriving at that shelf.

Tree diagrams are often used with so-called *urn models*. Imagine a large urn into which The Dealer has placed red, green and blue beads.² You take a bead at random from the urn, and then depending on the color chosen, The Dealer adds or removes some additional red, green or blue beads. The contents are mixed, and then you draw again. Again, depending on the color you draw, The Dealer adds or removes some beads. And so it goes. If the number of beads at the outset is known and if the rules for adding or removing beads are known, then you can use a tree diagram to calculate the probability of any particular sequence of draws.

Example 1. Suppose there are initially 3 red, 2 green and 1 blue bead. When you remove a bead, it is not replaced. No other beads are added or removed. The figure below shows the tree diagram that describes the possibilities if two draws are made. Note that the denominator is 6 at the first draw, but 5 at the second. This is because the bead taken on the first draw is not replaced.



This gives us a probability space with 8 elements, one for each path. The elements in the

² Many examples involve a situation that is set up and controlled by a agent who follows certain rules but then stands back from the chance events that you subsequently encounter. We will call him "The Dealer."

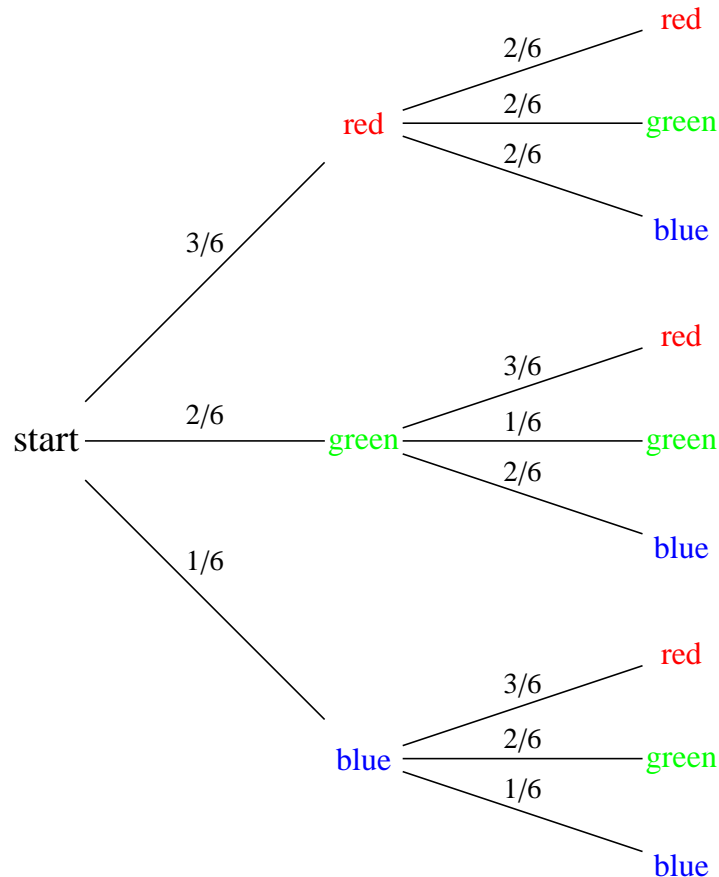
space, together with the values of the pmf for the space are given in the following table:

ω		rr	rg	rb	gr	gg	gb	br	bg	
$f(\omega)$		$6/30$	$6/30$	$3/30$	$6/30$	$2/30$	$2/30$	$3/30$	$2/30$,

where xy refers to the outcome consisting of a bead of type x followed by a bead of type y . The event of getting exactly one red is $\{rg, rb, gr, br\}$, and the probability of this event is

$$P(\{rg, rb, gr, br\}) = f(rg) + f(rb) + f(gr) + f(br) = \frac{6}{30} + \frac{3}{30} + \frac{6}{30} + \frac{3}{30} = \frac{18}{30} = \frac{3}{5}.$$

Example 2. Now, suppose we start with the same initial conditions, but when you remove a bead The Dealer replaces it with a blue bead. The figure below shows the tree diagram that describes the possibilities if two draws are made. In this case, the denominator is always 6, since the first drawn bead is replaced with a blue bead.



This gives us a probability space with 9 elements, since now it is possible to draw two blues in a row. In this space,

$$P(\{rg, rb, gr, br\}) = f(rg) + f(rb) + f(gr) + f(br) = \frac{6}{36} + \frac{6}{36} + \frac{6}{36} + \frac{3}{36} = \frac{21}{36} = \frac{7}{12}.$$

Problems.

1. In Example 1, what is the probability of getting exactly one green? Exactly one blue? Two reds? Two whites? No reds? Answer the same questions for example 2.
2. Suppose we extend Example 1 by making a third draw. Draw the associated tree diagram. List the elements in the associated probability space, and make a table to show the values of the pdf. Do the same for Example 2.

2.2. Products, quotients and subspaces.

There are three very general procedures for building complex probability spaces from simpler ones (products) and for modifying probability spaces to produce new ones (quotients and subspaces). Products, quotients and subspaces occur throughout mathematics. Their occurrence in probability theory is well understood, but I am not aware of any expositions that emphasize them. This lecture, therefore, is something of an experiment. I hope that even if the operations seem abstract, the examples will show right away that they are indeed useful organizing principles.

Products. Recall that the product of two sets A and B —denoted $A \times B$ —is the set of all ordered pairs (a, b) with $a \in A$ and $b \in B$.

Example. Let \mathbb{Z} denote the integers: $0, \pm 1, \pm 2, \dots$. Then $\mathbb{Z} \times \mathbb{Z}$ may be identified with the set of “lattice points” in the plane.

Example. Let $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K\}$ and let $B = \{\spadesuit, \heartsuit, \clubsuit, \diamondsuit\}$. Then

$$A \times B = \{(1, \spadesuit), (2, \spadesuit), \dots, (K, \spadesuit), (1, \heartsuit), \dots, (K, \diamondsuit)\}.$$

It is easier to visualize $A \times B$ as a grid. Below, I have labeled the columns by elements of A and the rows by elements of B . To save space, I have written $1\spadesuit$ as an abbreviation for $(1, \spadesuit)$ and have abbreviated the other entries similarly.

	1	2	3	4	5	6	7	8	9	10	J	Q	K
\spadesuit	$1\spadesuit$	$2\spadesuit$	$3\spadesuit$	$4\spadesuit$	$5\spadesuit$	$6\spadesuit$	$7\spadesuit$	$8\spadesuit$	$9\spadesuit$	$10\spadesuit$	$J\spadesuit$	$Q\spadesuit$	$K\spadesuit$
\heartsuit	$1\heartsuit$	$2\heartsuit$	$3\heartsuit$	$4\heartsuit$	$5\heartsuit$	$6\heartsuit$	$7\heartsuit$	$8\heartsuit$	$9\heartsuit$	$10\heartsuit$	$J\heartsuit$	$Q\heartsuit$	$K\heartsuit$
\clubsuit	$1\clubsuit$	$2\clubsuit$	$3\clubsuit$	$4\clubsuit$	$5\clubsuit$	$6\clubsuit$	$7\clubsuit$	$8\clubsuit$	$9\clubsuit$	$10\clubsuit$	$J\clubsuit$	$Q\clubsuit$	$K\clubsuit$
\diamondsuit	$1\diamondsuit$	$2\diamondsuit$	$3\diamondsuit$	$4\diamondsuit$	$5\diamondsuit$	$6\diamondsuit$	$7\diamondsuit$	$8\diamondsuit$	$9\diamondsuit$	$10\diamondsuit$	$J\diamondsuit$	$Q\diamondsuit$	$K\diamondsuit$

Definitions:

- a) Suppose Ω_0 and Ω_1 are discrete probability spaces with pmfs f_0 and f_1 . Then we make $\Omega_0 \times \Omega_1$ into a probability space by assigning it the pmf obtained by multiplying f_0 and f_1 :

$$f_{01}(\omega_0, \omega_1) := f_0(\omega_0) \cdot f_1(\omega_1).$$

- b) Suppose $\Omega_i, i = 1, \dots, k$ are discrete probability spaces with pmfs f_i . Let $X := \Omega_1 \times \Omega_2 \times \dots \times \Omega_k$ and let $g : X \rightarrow [0, 1]$ be defined by

$$g(\omega_1, \omega_2, \dots, \omega_k) := f_1(\omega_1) \cdot f_2(\omega_2) \cdot \dots \cdot f_k(\omega_k).$$

Then X with the probability measure induced by g is called the *product* of the spaces (Ω_i, f_i) .

- c) If in Part b) all the (Ω_i, f_i) are equivalent to (Ω, f) , we denote the product $(\Omega, f)^k$, or if the f is understood, Ω^k .

Problems.

1. True or false? Picking a card at random when all cards are equally likely is equivalent to picking a suit at random, with all suits equally likely and then picking a rank at random, with all ranks equally likely.
2. If we remove all the face cards from the deck, do the remaining cards have a natural product structure?
3. If we remove a single card from the deck, do the remaining cards have a natural product structure?
4. Suppose $\Omega = \{H, T\}$, with the pmf $f(H) = p, f(T) = 1 - p$. How many elements in Ω^2 ? In Ω^3 ? In Ω^k ? Show that Ω^k with its associated pmf can be used to model the process of flipping a biased coin k times. List the elements of Ω^4 and compute the value of the product pmf at each one.

Quotients. To partition a set means to select a family of subsets so that no two have any points in common and so that all together they contain every element of the set. For example, here is a partition of $\{0, 1, 2, 3, 4, 5, 6, 7\}$:

$$\left\{ \{0\}, \{1, 2, 4\}, \{3, 5, 6\}, \{7\} \right\}.$$

Definition. Suppose $\widehat{\Omega} := \{Y_1, \dots, Y_k\}$ is a partition of Ω and f is a pmf on Ω . Define a pmf \widehat{f} on $\widehat{\Omega}$ by $\widehat{f}(Y_i) = \sum \{ f(y) \mid y \in Y_i \}$.

Problem. Show that \widehat{f} is indeed a pmf.

Example. Let $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. Partition Ω into the sets $Y_0 = \{HHH\}, Y_1 = \{HHT, HTH, THH\}, Y_2 = \{HTT, THT, TTH\}, Y_3 = \{TTT\}$. Let f be the uniform pmf on Ω , so $f(\omega) = 1/8$ for all $\omega \in \Omega$. Then $\widehat{f}(Y_0) = 1/8, \widehat{f}(Y_1) = 3/8, \widehat{f}(Y_2) = 3/8, \widehat{f}(Y_3) = 1/8$.

Example. Let $\Omega = \{0, 1, 2, 3, 4, 5, 6\}$, with pmf $f(\omega) = 1/6$. Then Ω^2 can be used to represent the result of throwing a pair of dice:

	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Partition Ω^2 so that $Y_n = \{(i, j) \mid i + j = n\}$. Then $\hat{f}(Y_7) = 6/36$, while $\hat{f}(Y_2) = 1/36$.

Subspaces. Suppose that Ω is a set with pmf f . Let A be a subset of Ω . The restriction of f to A is a pmf only if $P(A) = 1$. If $P(A) \neq 0$, we can correct this problem by dividing by $P(A)$.

Definition. If $A \subseteq \Omega$, define $f_A(\omega) := f(\omega)/P(A)$. A together with the probability measure on A induced by f_A is said to be *conditioned on* A . The pmf $f_A(\omega)$ is sometimes denoted $f(\omega|A)$. The probability measure induced by f_A denoted $P_A(_)$ or $P(_|A)$.