Probability: Terminology and Examples
Class 2, 18.05
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Know the definitions of sample space, event and probability function.

2. Be able to organize a scenario with randomness into an experiment and sample space.

3. Be able to make basic computations using a probability function.

2 Terminology

2.1 Probability cast list

- Experiment: a repeatable procedure with well-defined possible outcomes.

- Sample space: the set of all possible outcomes. We usually denote the sample space by Ω, sometimes by S.

- Event: a subset of the sample space.

- Probability function: a function giving the probability for each outcome.

Later in the course we will learn about

- Probability density: a continuous distribution of probabilities.

- Random variable: a random numerical outcome.

2.2 Simple examples

Example 1. Toss a fair coin.
Experiment: toss the coin, report if it lands heads or tails.
Sample space: $\Omega = \{H, T\}$.
Probability function: $P(H) = .5, \quad P(T) = .5$.

Example 2. Toss a fair coin 3 times.
Experiment: toss the coin 3 times, list the results.
Sample space: $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.
Probability function: Each outcome is equally likely with probability 1/8.

For small sample spaces we can put the set of outcomes and probabilities into a probability table.

Outcomes	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
Probability	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

Example 3. Measure the mass of a proton
Experiment: follow some defined procedure to measure the mass and report the result.
Sample space: $\Omega = [0, \infty)$, i.e. in principle we can get any positive value.
Probability function: since there is a continuum of possible outcomes there is no probability
function. Instead we need to use a *probability density,* which we will learn about later in
the course.

Example 4. Taxis (An infinite discrete sample space)
Experiment: count the number of taxis that pass 77 Mass. Ave during an 18.05 class.
Sample space: $\Omega = \{0, 1, 2, 3, 4, \ldots\}$.
This is often modeled with the following probability function known as the Poisson distri-
bution. (Do not worry about mastering the Poisson distribution just yet):

$$P(k) = e^{-\lambda}\frac{\lambda^k}{k!},$$

where λ is the average number of taxis. We can put this in a table:

Outcome	0	1	2	3	...	k	...
Probability	$e^{-\lambda}$	$e^{-\lambda}\lambda$	$e^{-\lambda}\lambda^2/2$	$e^{-\lambda}\lambda^3/3!$	...	$e^{-\lambda}\lambda^k/k!$	...

Question: Accepting that this is a valid probability function, what is $\displaystyle\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}$?

answer: This is the total probability of all possible outcomes, so the sum equals 1. (Note,
this also follows from the Taylor series $\displaystyle e^{\lambda} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}$.)

In a given setup there can be more than one reasonable choice of sample space. Here is a
simple example.

Example 5. Two dice (Choice of sample space)
Suppose you roll one die. Then the sample space and probability function are

Outcome	1	2	3	4	5	6
Probability:	1/6	1/6	1/6	1/6	1/6	1/6

Now suppose you roll two dice. What should be the sample space? Here are two options.
1. Record the pair of numbers showing on the dice (first die, second die).
2. Record the sum of the numbers on the dice. In this case there are 11 outcomes
$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. These outcomes are not all equally likely.

As above, we can put this information in tables. For the first case, the sample space is the
product of the sample spaces for each die

$$\{(1, 1), (2, 1), (3, 1), \ldots (6, 6)\}$$

Each of the 36 outcomes is equally likely. (Why 36 outcomes?) For the probability function
we will make a two dimensional table with the rows corresponding to the number on the
first die, the columns the number on the second die and the entries the probability.

Die 2

	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

(Die 1 labels the rows)

Two dice in a two dimensional table

In the second case we can present outcomes and probabilities in our usual table.

outcome	2	3	4	5	6	7	8	9	10	11	12
probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

The sum of two dice

Think: What is the relationship between the two probability tables above?

We will see that the best choice of sample space depends on the context. For now, simply note that given the outcome as a pair of numbers it is easy to find the sum.

Note. Listing the experiment, sample space and probability function is a good way to start working systematically with probability. It can help you avoid some of the common pitfalls in the subject.

Events.

An event is a collection of outcomes, i.e. an event is a subset of the sample space Ω. This sounds odd, but it actually corresponds to the common meaning of the word.

Example 6. Using the setup in Example 2 we would describe the event that you get exactly two heads in words by $E =$ 'exactly 2 heads'. Written as a subset this becomes

$$E = \{HHT, HTH, THH\}.$$

You should get comfortable moving between describing events in words and as subsets of the sample space.

The probability of an event E is computed by adding up the probabilities of all of the outcomes in E. In this example each outcome has probability $1/8$, so we have $P(E) = 3/8$.

2.3   Definition of a discrete sample space

Definition. A discrete sample space is one that is listable, it can be either finite or infinite.

Examples. {H, T},   {1, 2, 3},   {1, 2, 3, 4, ...},   {2, 3, 5, 7, 11, 13, 17, ...} are all discrete sets. The first two are finite and the last two are infinite.

Example. The interval $0 \leq x \leq 1$ is *not* discrete, rather it is *continuous.* We will deal with continuous sample spaces in a few days.

2.4 The probability function

So far we've been using a casual definition of the probability function. Let's give a more precise one.

Careful definition of the probability function.
For a discrete sample space S a probability function P assigns to each outcome ω a number $P(\omega)$ called the probability of ω. P must satisfy two rules:

- Rule 1. $0 \le P(\omega) \le 1$ (probabilities are between 0 and 1).

- Rule 2. The sum of the probabilities of all possible outcomes is 1 (something must occur)

In symbols Rule 2 says: if $S = \{\omega_1, \omega_2, \ldots, \omega_n\}$ then $P(\omega_1) + P(\omega_2) + \ldots + P(\omega_n) = 1$. Or, using summation notation: $\sum_{j=1}^{n} P(\omega_j) = 1$.

The probability of an event E is the sum of the probabilities of all the outcomes in E. That is,

$$P(E) = \sum_{\omega \in E} P(\omega).$$

Think: Check Rules 1 and 2 on Examples 1 and 2 above.

Example 7. Flip until heads (A classic example)
Suppose we have a coin with probability p of heads and we have the following scenario.

Experiment: Toss the coin until the first heads. Report the number of tosses.
Sample space: $\Omega = \{1, 2, 3, \ldots \}$.
Probability function: $P(n) = (1 - p)^{n-1} p$.

Challenge 1: show the sum of all the probabilities equals 1 (hint: geometric series).

Challenge 2: justify the formula for $P(n)$ (we will do this soon).

Stopping problems. The previous toy example is an uncluttered version of a general class of problems called stopping rule problems. A stopping rule is a rule that tells you when to end a certain process. In the toy example above the process was flipping a coin and we stopped after the first heads. A more practical example is a rule for ending a series of medical treatments. Such a rule could depend on how well the treatments are working, how the patient is tolerating them and the probability that the treatments would continue to be effective. One could ask about the probability of stopping within a certain number of treatments or the average number of treatments you should expect before stopping.

3 Some rules of probability

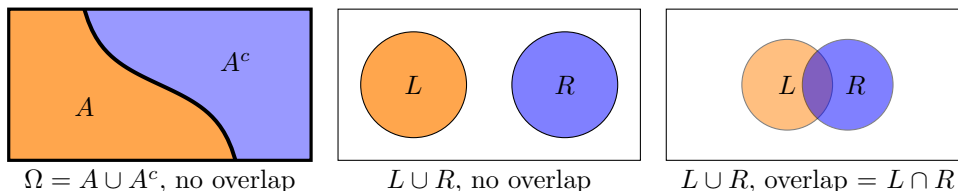For events A, L and R contained in a sample space Ω.

Rule 1. $P(A^c) = 1 - P(A)$.

Rule 2. If L and R are disjoint then $P(L \cup R) = P(L) + P(R)$.

Rule 3. If L and R are not disjoint, we have the **inclusion-exclusion principle**:

$$P(L \cup R) = P(L) + P(R) - P(L \cap R)$$

We visualize these rules using Venn diagrams.

$\Omega = A \cup A^c$, no overlap $L \cup R$, no overlap $L \cup R$, overlap $= L \cap R$

We can also justify them logically.

Rule 1: A and A^c split Ω into two non-overlapping regions. Since the total probability $P(\Omega) = 1$ this rule says that the probabiity of A and the probability of 'not A' are complementary, i.e. sum to 1.

Rule 2: L and R split $L \cup R$ into two non-overlapping regions. So the probability of $L \cup R$ is is split between $P(L)$ and $P(R)$

Rule 3: In the sum $P(L) + P(R)$ the overlap $P(L \cap R)$ gets counted twice. So $P(L) + P(R) - P(L \cap R)$ counts everything in the union exactly once.

Think: Rule 2 is a special case of Rule 3.

For the following examples suppose we have an experiment that produces a random integer between 1 and 20. The probabilities are not necessarily uniform, i.e., not necessarily the same for each outcome.

Example 8. If the probability of an even number is .6 what is the probability of an odd number?

answer: Since being odd is complementary to being even, the probability of being odd is 1-.6 = .4.

Let's redo this example a bit more formally, so you see how it's done. First, so we can refer to it, let's name the random integer X. Let's also name the event 'X is even' as A. Then the event 'X is odd' is A^c. We are given that $P(A) = .6$. Therefore $P(A^c) = 1 - .6 = \boxed{.4}$.

Example 9. Consider the 2 events, A: 'X is a multiple of 2'; B: 'X is odd and less than 10'. Suppose $P(A) = .6$ and $P(B) = .25$.
(i) What is $A \cap B$?
(ii) What is the probability of $A \cup B$?

answer: (i) Since all numbers in A are even and all numbers in B are odd, these events are disjoint. That is, $\boxed{A \cap B = \emptyset.}$

(ii) Since A and B are disjoint $\boxed{P(A \cup B) = P(A) + P(B) = .85.}$

Example 10. Let A, B and C be the events X is a multiple of 2, 3 and 6 respectively. If $P(A) = .6$, $P(B) = .3$ and $P(C) = .2$ what is $P(A \text{ or } B)$?

answer: Note two things. First we used the word 'or' which means union: 'A or B' $= A \cup B$. Second, an integer is divisible by 6 if and only if it is divisible by both 2 and 3.

This translates into $C = A \cap B$. So the inclusion-exclusion principle says

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = .6 + .3 - .2 = \boxed{.7}.$$

18.05 Introduction to Probability and Statistics
Spring 2014