Class 2
Probability: Terminology and Examples
18.05, Spring 2013

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 defined 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.
- Probability density: (We'll get to this later in the course.)
- Random variable: a random numerical outcome (We'll get to this later in the course.)

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 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 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: sinces 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. (An infinite sample space)
Experiment: Watch a bacterium until it divides. Then count the number of mutations.
Sample space: $\Omega = \{0, 1, 2, 3, 4, \dots\}$.

There is no obvious probability function. Here is a possible probability function: $P(k) = \mathrm{e}^{-\lambda}\dfrac{\lambda^k}{k!}$, where λ is the average number of mutations.

We can put this in a table:

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

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

In a given setup there can be more than one possible sample space.

Example 5. (Choice of sample space)
Suppose you roll two dice. What should be the sample space? That is, how should we record the outcomes. In this case there are two obvious choices.
1. Record the pair of numbers showing on the dice (first die, second die). In this case there are 36 (why?) equally likely outcomes.
2. Record the sum of the numbers on two dice. In this case there are 11 outcomes $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. These outcomes are not equally likely.

As above, we can put this information in tables. For the first case will make a two dimensional table with the rows representing 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	1/6	1/6	1/6	1/6	1/6	1/6
	2	1/6	1/6	1/6	1/6	1/6	1/6
Die 1	3	1/6	1/6	1/6	1/6	1/6	1/6
	4	1/6	1/6	1/6	1/6	1/6	1/6
	5	1/6	1/6	1/6	1/6	1/6	1/6
	6	1/6	1/6	1/6	1/6	1/6	1/6

Two dice in a two dimensional table

outcomes	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

We will see that there are good reasons for both choices. For now simply note that given the outcome as a pair of numbers it is easy to convert it to 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.

We defined an *event* as a collection of outcomes. In other words and 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 an event in words by saying something like, $E =$ 'the event you get 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.

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

2.3 Definition of a discrete set

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

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

Example. The interval $0 \le x \le 1$ 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. Since in all the examples it was clear that every outcome was equally likely The probability function tells you how to compute the probability of an single outcome or an event (collection of outcomes).

Careful definition of the probability function.
For a discrete sample space a *probability function* assigns to each outcome ω a number $P(\omega)$ called the probability. It must satisfy:
1. $0 \le P(\omega) \le 1$ (the probability is between 0 and 1).
2. The sum of the probabilities of all possible outcomes is 1 (something must happen).
3. The probability of an event E is the sum of the probabilities of all the outcomes in E.

In symbols rule 2 says: if $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ then $P(\omega_1) + P(\omega_2) + \ldots + P(\omega_n) = 1$. Using summation notation: $\displaystyle\sum_{j=1}^{n} P(\omega_j) = 1$.

Problem. Check rules 1, 2 and 3 on examples 1 and 2 above.

Example 7. (A classic example)
Suppose we have a coin with probability p of heads.
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.
Challenge 2: show $P(n)$ has the formula given (we will do this soon).

Stopping problems. As usual with our toy examples, the previous 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 example the process was flipping a coin and we stopped after the first heads. A possibly more practical sounding example might be a rule for ending a series of medical treatments. One could ask about the probability of stopping within a certain number of treatments or the average number of treatments you should expect.

3 Some rules of probability

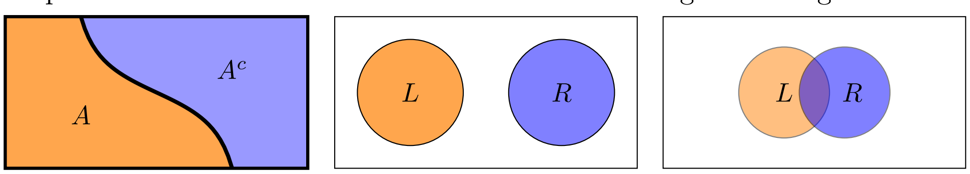
For events A, L, R

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. In general, $P(L \cup R) = P(L) + P(R) - P(L \cap R)$.

We can express these rules in words and visualize them using Venn diagrams.



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

Rule 1: A and A^c split into 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 into $L \cup R$ 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.

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

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

<u>**answer:**</u> 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.
a. 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 $A \cup B$?
(iii) Describe the set $A \cup B$ in words.

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

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

(iii) $A \cup B$ is the set of all numbers that are either even or odd and less than 20. That is $\boxed{A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20\}.}$

Example 10. Let A, B and C be the events X is a multiple of 2, 3, 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 English word 'or' this means the same thing as union. A or $B = A \cup B$. Second, $6 = 2 \times 3$, which translates into $C = A \cap B$. Therefore, y rule (3) for probability

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

End of class 2 notes