

# Basic Probability

## Notes.

Main concepts

- Arrangements when some objects are alike
- Counting integer solutions
- Binomial Theorem

## What is an event

An experiment is a procedure that results in one out of a number of possible outcomes. The set of all possible outcomes is called the sample space of the experiment. A subset of the sample space is called an event.

Examples are

- Consider a simple example of an experiment in which a red and a blue die are thrown. We will denote a single outcome by an ordered pair where the first number denotes the outcome on the blue one and second number denotes the outcome on the red one. Then the sample space is  $S = \{(x, y) | 1 \leq x \leq 6, 1 \leq y \leq 6\}$ . Any subset of the sample space is called an event. So for instance the event  $E$  or having doubles show up on the dice can be written explicitly as

$$E = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}.$$

- Another common experiment is playing cards. Let us say we are playing poker and we get a 5 card hand. Obviously there are several possible outcomes. Every single outcome is one of the 5 element subsets of the set of 52 cards.

The sample space therefore is of size  $\binom{52}{5}$ .

A particular event in this case for instance is a royal flush, which means A, K, Q, J and 10, all of the same suit. Let's call that event

$$R = \{\{A\clubsuit, K\clubsuit, Q\clubsuit, J\clubsuit, 10\clubsuit\}, \{A\spadesuit, K\spadesuit, Q\spadesuit, J\spadesuit, 10\spadesuit\}, \\ \{A\heartsuit, K\heartsuit, Q\heartsuit, J\heartsuit, 10\heartsuit\}, \{A\diamondsuit, K\diamondsuit, Q\diamondsuit, J\diamondsuit, 10\diamondsuit\}\}$$

There are 4 such royal flushes.

## Basic probability definition

A natural question with regards to an experiment is to say, what is the likelihood (the probability) that a particular event occurs.

A probability distribution over an experiment with a sample space of  $S$  is a function  $p$  from  $S$  to the set  $[0, 1]$ .

Certain properties need to be satisfied by this probability distribution.

$$\sum_{s \in S} p(s) = 1$$

The probability of the single outcome  $s$  is  $p(s)$ . If  $E \subseteq S$  is an event, then the probability of the event  $E$  is

$$p(E) = \sum_{s \in E} p(s)$$

This definition is especially useful when dealing with situations where not every outcome is equally likely.

Loaded dice for instance. (More detail in the zybook)

Let's say a die is twice as likely to show up 4 as anything else. To make this work, we basically want to assign a value of  $2/7$  to  $p(4)$  and a value of  $1/7$  to the probability of every other outcome.

Given these individual outcome probabilities, you can now work out the probabilities of any event

For example the event that the die shows a number greater than 3 becomes  $p(4) + p(5) + p(6)$  which is  $4/7$ .

## Uniform distribution

In many scenarios, the probability of every outcome in the sample space is the same. The probability distribution in which every outcome has the same probability is called the uniform distribution.

Since there are  $|S|$  outcomes and their probabilities sum to 1, under the uniform distribution, for each  $s \in S$ ,  $p(s) = 1/|S|$ . The uniform distribution reduces questions about probabilities to questions about counting because for every event  $E$ ,

$$p(E) = \frac{|E|}{|S|}$$

Example (from zybook)

Consider a situation in which files are stored on a distributed network. Multiple copies of each files are stored around the network so that if one or more computers crash, the data is more likely to be available from at least one source. Suppose that three copies of a files are stored at different locations in a network of 30 computers and that at a particular moment, five random computers fail. Each subset of 5 computers are equally likely to be the five that have failed. What is the probability that there are no copies left of the file?

The experiment over here is choosing 5 computers that fail. So total number of distinct outcomes is  $\binom{30}{5}$ . Every computer is equally likely to fail. That means every subset of 5 is equally likely to fail.

How many of these subsets have all 3 of the computers that contain this file? The file is in 3 of the computers, so we have to have all 3 of these computers being picked. But we still have a choice for the remaining 2 computers. 27 other computers exist. 2 of them have to be picked.

therefore  $\frac{\binom{27}{2}}{\binom{30}{5}}$