COMS10003 : Dealing with Uncertainty

Probability I

Andrew Calway

December 5, 2014

Introduction

So far we have dealt with Maths in which everything is certain - we know precisely the numbers and the statements that we are dealing with and the methods give us precise and reliable answers. For example, 1026 is 1026, 627 is 627, and $(1026, 627) = 57$; there is no doubt or chance involved (assuming we know the language!). This type of Maths is important in many areas of CS - not only in Cryptography, but also in designing compilers, databases, control systems, etc, in which precision, reliability and certainty is paramount.

We now turn to a part of Maths which allows us to deal with uncertainty - with entities that can take on a range of different (previously unknown) values or forms and with methods that may or may not result in the 'correct' answer. It comes under the general heading of Probability and Random Variables.

There are many areas of CS in which this type of Maths is important, especially those involving processing data in the real world. Good examples are spam filters for identifying unwanted emails (is an email containing the word 'viagra' always spam?), internet search algorithms (do you mean 'bath' as in the town or something to wash in?) and face recognition algorithms (can it recognise me when I'm smiling?). We'll take a look at some of these later but we'll start by looking at gambling - or at least how to quantify your chances of winning (or losing).

There are many many many books on probability - I encourage you to have a good look around and find what suits. For these notes I've made use of the two below, mainly because they sit on my bookcase:

Probability, Random Variables and Stochastic Processes by A.Papoulis, McGraw-Hill
Linear Algebra and Probability for Computer Science Applications by E.Davis, CRC Press

Events, Sample Spaces and Probability

We begin with a simple example and one that you are likely to find at the beginning of any textbook on probability: if we flip a fair coin many times, then we would expect that half of the time it would be heads (H) and half of the time tails (T).

In the language of Probability, we say that there is a probability of 0.5 that the coin will show H and a probability of 0.5 that it will show T and that they are both *equally likely*. To represent this we use the notation $P(H) = P(T) = 0.5$.

Flipping the coin is known as a *random experiment* - we don't know the outcome beforehand - and the set of all possible outcomes $\Omega = \{H, T\}$ is known as the *sample space*. For now, as in this case, we will be concerned with sample spaces which have a finite number of discrete elements; later we will look at spaces which are continuous and infinite.

For a given experiment, we can also talk about *events* - represented by a subset of outcomes, each of which fulfil a condition associated with the event. An event can also be associated with a single outcome or all possible outcomes, i.e. the sample space itself. For example, we can identify 3 possible events in our coin flipping example:

- $E_1 = \{H\}$, the outcome is H;
- $E_2 = \{T\}$, the outcome is T; or
- $E_3 = \Omega = \{H, T\}$, the outcome is H or T.

For completeness, we should also include the impossible event $E_0 = \emptyset$, the empty set, i.e. that the outcome is neither H nor T (we assume that we always complete a successful flipping experiment). We then get the event probabilities $P(E_0) = 0$, $P(E_1) = 0.5$, $P(E_2) = 0.5$, and $P(E_3) = 1$.

In general, $0 \leq P(E) \leq 1$ and $P(\Omega) = 1$, i.e. an outcome must belong to the sample space. Assuming all our events are equally likely, then we can use the following to compute the probability of an event:

$$P(\text{event}) = \frac{\text{number of outcomes which fulfil the event condition}}{\text{number of outcomes in the sample space}} \tag{1}$$

or: $P(E) = |E|/|\Omega|$, where $|E|$ denotes the number of elements in the set E, i.e. the probability reflects the proportion of 'event outcomes' to 'all outcomes'.

We can also estimate the probability of an event by performing our random experiment many times, e.g. by flipping our coin lots of times. If an experiment is conducted N times and the event E occurs N_E times, then the probability of the event E occurring can be estimated using $P(E) \approx N_E/N$. It's perhaps not surprising that we can show that this estimate gets better and better the more times we run the experiment, i.e. as N gets larger. In fact, we can write [1]

$$P(E) = \text{lim}_{N\rightarrow\infty} \frac{N_E}{N} \tag{2}$$

[1] As an aside, equation (2) is sometimes known as the *relative frequency* definition of probability, whilst equation (1) is known as the *classical definition*.

So far we have assumed that all our events are equally likely. That need not be the case. To deal with this we introduce the concept of weights associated with a given outcome, where the weight value represents the likelihood of the event occurring, i.e. larger weights indicating higher likelihood. The weight for an event is then given by the sum of weights of the outcomes in the event, i.e.

$$w(E) = \sum_{x \in E} w(x) \tag{3}$$

where $w(x)$ is the weight of outcome x. It follows that the probability of an event is then given by $P(E) = w(E)/w(\Omega)$. Thus, if we have a biased coin in which H is likely to occur 3 times more often than T, we have $P(H) = 3/4$ and $P(T) = 1/4$.

—–

Example : Consider the experiment of flipping a fair coin 3 times and recording the result of each flip. Each set of 3 recordings is then regarded as an outcome. The sample space is therefore $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ and the subset of outcomes for each of the following events: (a) at least two heads; (b) at least one tail; (c) exactly one head, are given by:

$E_a = \{HHT, HTH, THH, HHH\}$

$E_b = \{HHT, HTH, HTT, THH, THT, TTH, TTT\}$

$E_c = \{HTT, THT, TTH\}$

with probabilities: $P(E_a) = 4/8$, $P(E_b) = 7/8$, and $P(E_c) = 3/8$. Note the link here with logic and truth tables - if we write out all the permutations of the three recordings in a table, each event then corresponds to a subset of rows.

—–

Finally, we can also define the complement of an event - the subset of outcomes that do not fulfil the event condition. Thus the complement of E_a in the above example is $E'_a = \{HTT, THT, TTH, TTT\}$ and $P(E'_a) = 4/8$. It follows that in general $P(E') = 1 - P(E)$.

Combined Events, Conditional Probability and Independence

This or That Event

We can also consider the probability of combinations of events. For example, in the above example, consider an event E_d which occurs if E_a **or** E_c occurs, i.e. $E_d = E_a \cup E_c$. The probability of E_d is then

$$P(E_d) = \frac{|E_a \cup E_c|}{|\Omega|} = \frac{|E_a| + |E_c|}{|\Omega|} = \frac{|E_a|}{|\Omega|} + \frac{|E_c|}{|\Omega|} = P(E_a) + P(E_c) = \frac{7}{8} \tag{4}$$

i.e. the sum of the probabilities of E_a and E_c. However, **this only works because E_a and E_c don't contain any common outcomes and hence $|E_a \cup E_c| = |E_a| + |E_c|$** - they are said to be *mutually exclusive.* In fact, in general we can show for E_1, E_2, ..., E_M

$$P(E_1 \cup E_2 \cup E_3 \ldots \cup E_M) = \sum_{i=1}^{M} P(E_i) \tag{5}$$

providing that $E_i \cap E_j = \emptyset \;\; \forall i \neq j$.

What if they are not mutually exclusive? Let's consider an event E_e which occurs if E_a **or** E_b occur. In this case

$$P(E_e) \neq \frac{|E_a| + |E_b|}{|\Omega|} \tag{6}$$

Why? Because E_a and E_b have common outcomes, i.e. $E_a \cap E_b = \{HHT, HTH, THH\}$, which get 'counted twice' in the above equation. Thus we have to remove one of them to get the correct formula, i.e.

$$P(E_e) = \frac{|E_a| + |E_b| - |E_a \cap E_b|}{|\Omega|} = P(E_a) + P(E_b) - P(E_a \cap E_b) \tag{7}$$

where $P(E_a \cap E_b)$ is the probability that events E_a **and** E_b occur. This equation defines what is known as the *addition law of probability.* The formula gets more complicated if they are more than two events - see the Worksheet.

Conditional Probability

Sometimes the probability of an event occurring can be affected by the occurrence of another event. In such cases we can consider the probability of an event to be *conditional* on the other event occurring or not. An example will illustrate.

Consider event E_a from the example above and let E_f be the event that the first flip gives T. The probability of event E_a given that E_f occurs is then

$$P(E_a|E_f) = \frac{|E_a \cap E_f|}{|E_f|} = \frac{|E_a \cap E_f|}{|\Omega|}\frac{|\Omega|}{|E_f|} = \frac{P(E_a \cap E_f)}{P(E_f)} = \frac{1}{8}\frac{8}{4} = \frac{1}{4} \tag{8}$$

Thus, in general we have the following relationships for two events E_1 and E_2

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \qquad\qquad P(E_1 \cap E_2) = P(E_1|E_2)P(E_2) \tag{9}$$

where the right-hand equation gives us a formula for computing the probability of both events E_1 **and** E_2 occurring. It turns out that this an important relationship which will lead us to another one in the next lecture - Bayes' Law.

We can also consider conditional probabilities involving more than two events. For example, consider mutually exclusive events E_1 and E_2, and another event E_3, which is not mutually exclusive to either E_1 or E_2. The probability of event E_1 **or** event E_2 occurring **given** that event E_3 occurs is then

$$P(E_1 \cup E_2 | E_3) = \frac{|(E_1 \cup E_2) \cap E_3|}{|E_3|} = \frac{|E_1 \cap E_3| + |E_2 \cap E_3|}{|E_3|} = P(E_1|E_3) + P(E_2|E_3) \quad (10)$$

Similarly, if E_1 and E_2 are not mutually exclusive, then the probability of both E_1 **and** E_2 occurring **given** that E_3 occurs is

$$P(E_1 \cap E_2 | E_3) = \frac{|E_1 \cap E_2 \cap E_3|}{|E_3|} = \frac{|E_1 \cap E_2 \cap E_3|}{|E_2 \cap E_3|} \frac{|E_2 \cap E_3|}{|E_3|} = P(E_1|E_2 \cap E_3) P(E_2|E_3) \quad (11)$$

and hence that

$$P(E_1 \cap E_2 \cap E_3) = P(E_1|E_2 \cap E_3) P(E_2|E_3) P(E_3) \quad (12)$$

Independent Events

If the occurrence of one event has no impact on the occurrence of another event, then we say that they are *independent*. For two independent events E_1 and E_2, we therefore have

$$P(E_1|E_2) = P(E_1) \Rightarrow P(E_1 \cap E_2) = P(E_1|E_2)P(E_2) = P(E_1)P(E_2) \quad (13)$$

i.e. the probability of both E_1 **and** E_2 occurring is given by the product of the individual probabilities, *providing* that they are independent events.

—–

Example : A classic example of using conditional probabilities is when drawing coloured balls from a bag. Consider a bag containing 3 red balls and 2 blue balls. Let E_R be the event of drawing a red ball first and E_B the event of drawing a blue ball. Then if we always replace a ball after drawing it, we have $P(E_R) = 3/5$ and $P(E_B) = 2/5$. But if we first draw a red ball and then draw another *without replacement*, then $P(E_R|E_R) = 1/2$ and $P(E_B|E_R) = 1/2$.

It is important to consider this example in terms of sets. Denote the 3 red balls by R_1, R_2 and R_3 and the blue balls by B_1 and B_2. A given outcome of experiment has the form R_iB_j and the relevant subsets of outcomes and hence probabilities are

$$E_R = \{R_1R_2, R_1R_3, R_1B_1, R_1B_2, R_2R_1, \ldots, R_3B_2\} \quad (14)$$

$$E_B \cap E_R = \{R_1B_1, R_1B_2, R_2B_1, \ldots, R_3B_2\} \quad (15)$$

$$P(E_R) = \frac{|E_R|}{|\Omega|} = \frac{12}{20} = \frac{3}{5} \quad \text{and} \quad P(E_B \cap E_R) = \frac{|E_B \cap E_R|}{|\Omega|} = \frac{6}{20} = \frac{3}{10} \tag{16}$$

$$P(E_B|E_R) = \frac{P(E_B \cap E_R)}{P(E_R)} = \frac{3}{10}\frac{5}{3} = \frac{1}{2} \tag{17}$$

—

Permutations and Combinations

Many problems in probability involve the use of permutations - ways of ordering items of a set - and combinations - ways of selecting a subset of items from a set. It is therefore useful to know some formulae that enable us to compute the number of permutations and combinations for a given case. Note that in permutations, the order matters, whilst it doesn't in combinations. Examples are given below.

Permutations of n items

Given n items, how many different orderings of those items are there? There are n possibilities for the first item, then $n-1$ possibilities for the second, and so on, giving $n(n-1)(n-2)\ldots 2.1 = n!$ possible permutations.

Permutations of k out of n items

Given n items, how many permutations of k items can we select from the n items? There are n possibilities for the first item, $n-1$ for the second, down to $n-k+1$ for the kth item, giving the number of permutations as $n(n-1)(n-2)\ldots(n-k+1) = n!/(n-k)!$. This is often denoted $P(n,k)$.

Combinations of k items out of n

Related to the above is the number of subsets of k items that can be selected from n items. In this case, the order doesn't matter. We can compute the number by noting the number of permutations of k items from n is $n!/(n-k)!$ and that there must be $k!$ permutations of each subset of k items. Hence the number of combinations of k items must be

$$C(n,k) = \left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!(n-k)!} \tag{18}$$

—

Example : Playing cards provide intriguing probability problems. Consider the probability of obtaining a flush (all cards the same suit) in 5 card poker. There are 52 cards in the deck and so there are $C(52,5)$ combinations of 5 cards. How many of these are flushes? There are four different suits and 13 cards in each suit. Hence there are $4C(13,5)$ possible flushes and so the probability of getting a flush is $4C(13,5)/C(52,5) = 0.002$.