NOTES ON PROBABILITY

Suppose we plan to perform an experiment, in which certain activities are undertaken and we record a list of results. For instance, the activities might be flipping a coin two times, and the results might be the pair of locations where the coin comes to rest each time (these locations would presumably be x and y coordinates, hence we'd end up with two pairs of real numbers, one pair for each flip). Or, we might open the curtains and see whether the sun is shining and record yes (shining) or no (not shining).

Whatever the "experiment", the set of all possible outcomes (records) is called the sample space, S, and its elements e are called elementary events.

Examples of sample spaces:

S1 = { <x1,y1> , <x2,y2> > | for all reals x1,x2,y1,y2 } in the case of the two coin flips above. Here the sample space is infinite.

S2 = {yes,no} in the case of the sun shining or not.

S3 = {HH,HT,TH,TT} where the activity is flipping a nickel and a dime and recording heads or tails for each.

S4 = {p1,p2,...,pn} where each pi is person in the class, and the activity is picking one volunteer to work a problem at the blackboard.

An "event", E, is a subset of S. Thus for S = S3 above, all sixteen subsets of S are events, including the empty set { }, S itself, and for instance E1 = {HH,HT,TH} and E2 = {TT}. The intuition is this: we think of E1 as the event of "getting at least one head", and E2 as getting no heads (ie getting 2 Ts).

E1 does *not* mean we do the experiment three times and get HH, HT, and TH. It is to be thought of as something that might be true of *one* run of the experiment, in that the result might be one of the elementary events in it.

We say an event E holds or is true for (one run of) an experiment if the outcome is of that type, ie if it is (one of the elementary events e) in E. So E1 here is the "at least one head" event, and E2 is the "no heads" event. And E3 = {HT,TH} is the "exactly one head" event, which is also the "exactly one tail" event. If we flip the coins (do the experiment) and get HT, then events E1 and E3 hold, and E2 does not.

Now we can introduce probabilities. For a given sample space S, a probability distribution is a function Pr on the domain of all events, ie the domain is the powerset of S, with codomain R, the real numbers, such that the following axioms are true:

(i) P(E) >= 0 for all events E

(ii) P(S) = 1

(iii) if (E int F) = { }, then P(E u F) = P(E) + P(F)

From these axioms one can prove

(iv) P({ }) = 0

(v) P(complement of E) = 1 - P(E)

and many other results.

Example 1.

Consider the sample space S = S3 = {HH,HT,TH,TT} above. Let P be the probability distribution determined by supposing all the elementary events e to have the same probability: P(e) = constant. Such a distribution is said to be "uniform". It follows that P(e) = 1/4 since they must (by repeated use of axioms (ii) and (iii)) add up to 1. And now we can compute the probability of any event, for instance P(at least one head) = 1 - P(no heads) = 1 - 1/4 = 3/4.

Example 2.

We flip a coin n times and record the sequence of heads and tails. So S now is all strings of H's and T's, of length n; there are 2^n such strings. If again we choose the uniform distribution, so all these strongs are equally likely, then each string (ie each elementary event) has probability 1/2^n.

Now consider the event Ek of getting exactly k heads (and thus n-k tails). This is not an elementary event, rather it is made up of all those elementary events that happen to have exactly k H's. So the probability of Ek is just 1/2^n times the number of elementary events in E (ie, times the cardinality of Ek). What is this cardinality?

This is the same as asking how many ways we can build a string of length n with exactly k H's. But this is just the number of ways we can select k positions out of n: the k positions will be the ones that get H's. And there are n-take-k ways to do this. So P(Ek) = n-take-k / 2^n = n! / [k! (n-k! 2^n]. For instance when n = 5 and k = 2, P(E2) = 5! / [2! 3! 32] = 10/32 = 5/16, roughly 1/3. If we now think of flipping a fair coin (that's what the uniform distribution means) 5 times, we'd expect to get about half heads, ie two or three; we might also get less than two or more than three but this seems unlikely, so the chance of getting two ought to be almost 50% (and the same for getting three). So 1/3 seems about right: it is close to 1/2 but 1/3 for two and another 1/3 for three (it is easy to check that P(E3) is also 1/3) leaves a little extra for the rarer cases of less than two or more than three.

This suggests a useful intuition about probabilities. We can think of them (the probability values) as being weight (or candy, whatever metaphor you prefer) that comes in a total of one kilogram that is spread around among the different elementary events, and therefore also among all the events. So in our example, 2 heads and 3 heads each get a little less than half the weight (or candy) and the rest (also 1/3) is spread over the remaining cases (0,1,4, or 5 heads).

Example 3.

Now we consider a case of non-uniform distribution.

... to be continued ...