Probability

The probability or likelihood of an event is defined as total number of successes (or frequency of an event occurring) divided by the sample space (or total number of possible times for the event to occur.) For example, the probability of rolling an even number on a standard six-sided die is $3/6 = \frac{1}{2}$ because there are six possible outcomes (1,2,3,4,5,6) of which three are even. It is IMPORTANT to note that each of the outcomes in the sample space MUST be equally probable for this definition to be valid. For example, if rolling a 1 was 5 times more likely than each of a 2, 3, 4, 5, or 6, then $\frac{1}{2}$ would not be the answer to the question above. We can denote the probability of an event A occurring as p(A).

Some probability rules:
1) The sum of the probabilities of all events/outcomes occurring is always 1. (Each event/outcome must be disjoint.)

2) The probability of any event is in between 0 and 1, inclusive.

3) If two events A and B are disjoint, then the probability of either event occurring is the sum of the probability of A occurring and of B occurring. Symbolically, we have, if $p(A \cap B) = 0$, then $p(A \cup B) = p(A) + p(B)$.

4) If two events are independent, meaning that one event does not affect the probability of another occurring, such as two consecutive flips of a fair coin, then the probability of both occurring is the product of the probability of each occurring. Symbolically, $p(A \cap B) = p(A)p(B)$, for independent events A and B.

5) A conditional probability, written as $p(A|B)$ is defined as the probability of an event A occurring given that event B has occurred. $p(A|B) = p(A \cap B)/p(B)$. In essence, the numerator accounts for all events where both A and B occur, but the sample space has to be confined to the events where B occurs, or $p(B)$.

6) The inclusion-exclusion principle holds: $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.

7) Bayes Law for conditional probabilities is $p(A|B) = p(A)p(B|A)/p(B)$. This can be derived from the equations for Bayes Law. ($p(A|B) = p(A \cap B)/p(B)$ and $p(B|A) = p(B \cap A)/p(A)$.)

Random Variables

We can model events through random variables. A random variable is one that takes on various values with various probabilities. For example, a random variable X that models a die roll is one that is equal to 1, $1/6^{th}$ of the time, 2 $1/6^{th}$ of the time, etc. Given a random variable X, we define its expected value as follows:

$E(X) = \Sigma \, xP(X=x)$, where the sum is over all possible values of the random X, and x is the index variable. For example, for a unbiased die roll, the expectation is $1/6(1+2+3+4+5+6) = 7/2$. You can think of this as the probable average value of a die roll over a long string of die rolls. (The expected value of a random variable is it's average value.)

There is a linearity of expectation, therefore, with random variables X and Y, we have $E(X+Y) = E(X)+E(Y)$. Furthermore, if two events X and Y are independent, we have $E(XY) = E(X)E(Y)$. For now, we will accept these results without proof.

A good example here is the expected sum of rolling two dice. The expected sum of rolling one die is 7/2 as stated above. Thus, using the linearity of expectation and letting event X be the value of the first die roll and event Y be the value of the second die roll, we find:

$E(X+Y) = E(X) + E(Y) = 7/2+7/2 = 7.$
Also, we have $E(XY) = E(X)E(Y) = 49/4.$

Consider this example from algorithm analysis that involves probability:

We have discussed the best and worst case analysis of a binary search on n sorted items. Now, let's do the average case analysis (assuming that the item is always in the array and to simplify things, I'll assume that $n=2^k - 1$ for some integer k):

Given an array of n items, the probability that the value is found on the first comparison is 1/n, assuming that each array element is equally likely to be searched for.

If that first search is incorrect, then the probability of finding the element in the second comparison is 1/(floor(n/2)), since there are only floor(n/2), or $2^{k-1}-1$ elements left to search.

Continuing with this logic the probability the element is found on the third comparison is $1/(n/2^2)$, etc. We know that the element MUST BE found on the $\log_2 n$ comparison, approximately. (This is the maximum number of steps in the binary search.)

Now, let X be a random variable equal to the number of comparisons done in a binary search of an array with n sorted items. We have to solve for the expectation of X in terms of n to approximate the AVERAGE case running time of the binary search. We have:

$E(X) = 1[1/n] + 2[2/n] + 3[2^2/n]+... + k[2^{k-1}/n]$

$= [1(2^0) + 2(2^1) + 3(2^2) + ... + k(2^{k-1})]/n$

Now, this sum isn't anything we recognize, but consider the following technique:

Let $S = 1(2^0) + 2(2^1) + 3(2^2) + \ldots + k(2^{k-1})$

Now, multiply this equation by 2:

$$2S = \quad 1(2^1) + 2(2^2) + \ldots + (k-1)(2^{k-1}) + k(2^k)$$

Now, subtract the second equation from the first:

$$-S = (2^0) + (2^1) + (2^2) + \ldots + (2^{k-1}) - k(2^k)$$

Solve for S:

$$S = k(2^k) - [2^0 + 2^1 + 2^2 + \ldots + 2^{k-1}]$$

The sum to be subtracted from $k(2^k)$ is a geometric sequence, with the first term 1, and a common ratio of 2. Using the formula given in the book, we find:

$$S = k(2^k) - (2^k - 1)/(2-1)$$

$$S = k(2^k) - 2^k + 1$$

$$S = (k-1)2^k + 1$$

$$S = (n+1)[\log_2(n+1) - 1] + 1$$

Plugging back into our equation for E(X), we have

$$E(X) = S/n = ((n+1)[\log_2(n+1) - 1] + 1)/n = \theta(\log_2 n),$$

showing that the average case running time is proportional to $\log_2 n$.