

Lecture Notes, CSE 232, Fall 2014 Semester

Dr. Brett Olsen

Week 10 - Probability and Combinatorics

Combinatorics Countable discrete structures. Three major fields: permutations, combinations, and partitions.

inversions of permutations (application to, e.g., sorting)

Fibonacci Sequence The Fibonacci sequence is well known, and comes up often in a variety of contexts. There are several different ways to define and calculate the sequence, and some of these tricks can be used for related sequences.

Fibonacci numbers are defined recursively as the sum of the previous two numbers in the sequence:

$$F_0 = 0, F_1 = 1, F_n = F_{n-2} + F_{n-1} \quad (1)$$

which gives the sequence 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89...

Implementing this recursive solution directly is very slow, as it runs in $O(\phi^n)$ exponential time, where ϕ is the golden mean $\frac{1+\sqrt{5}}{2} = 1.618\dots$. This time complexity is related to the approximation of ϕ as the ratio of F_n to F_{n-1} .

We can instead calculate the sequence in linear $O(n)$ time with a simple dynamic programming solution by simply iterating from F_0 to F_n , saving the last two values of the sequence to compute the next one.

There are two other interesting, faster, methods for calculating the numbers of a Fibonacci sequence:

First, by $O(\log(n))$ solution using exponentiation of a matrix:

$$\begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^n \quad (2)$$

We can exponentiate the matrix in $\log(n)$ time by repeated squaring of the matrix.

There is an even faster $O(1)$ method: F_n is the closest integer to the real number $(\phi^n - (-\phi^{-n}))/\sqrt{5}$. For small values of n , we can do this in constant time, but beware of using it for larger values where rounding errors may cause problems.

Binomial Coefficients Binomial coefficients are defined as the k_{th} coefficients of the expansion of the n_{th} power of a binomial $(x + y)$. For example, $(x + y)^3 = 1x^3 + 3x^2y + 3xy^2 + 1y^3$, so $\{1, 3, 3, 1\}$ are the binomial coefficients for $n = 3$ and $k = \{1, 3, 3, 1\}$. These coefficients can also be interpreted as the number of ways to choose k distinct items from a set of n total items, usually denoted as $\binom{n}{k}$. A closed-form solution for binomial coefficients can be given in terms of factorials:

$$\binom{n}{k} = C(n, k) = \frac{n!}{(n-k)! \times k!} \quad (3)$$

However, this solution is slow to calculate for large n and k , and likely to cause problems with integer overflows. A much better solution is the recursive formula

$$C(n, 0) = C(n, n) = 1, C(n, k) = C(n-1, k-1) + C(n-1, k) \quad (4)$$

When large number of binomial coefficients are required, it can be worthwhile to construct the full Pascal's triangle from this recursive solution:

n=0 1 n=1 1 1 n=2 1 2 1 n=3 1 3 3 1 n=4 1 4 6 4 1 n=5 1 5 10 10 5 1 n=6 1 6 15 20 15 6 1

Catalan Numbers While I expect everyone has heard of Fibonacci number and binomial coefficients before, probably most of you are completely new to what are known as the Catalan numbers. Before I discuss what they're used for, let's define them. Just like the binomial coefficients, they can be represented as a closed-form expression in terms of factorials, which in turn can be broken down into binomial coefficients:

$$Cat(n) = \frac{2n!}{n! \times n! \times (n+1)} Cat(n) = \binom{2n}{n} \times \frac{1}{n+1} Cat(n) = \binom{2n}{n} - \binom{2n}{n+1} \quad (5)$$

We can also define them recursively:

$$Cat(0) = 1 Cat(n) = \frac{2n \times (2n-1)}{(n+1) \times n} \times Cat(n-1) \quad (6)$$

So we can use the same approaches we might take for the binomial coefficients to calculate these. The first few Catalan numbers are 1, 1, 2, 5, 14, 42, 132, 429, 1,430, 4,862, 16,796, 58,786, 208,012, 742,900, 2,674,440, 9,694,845, 35,357,670.

The Catalan numbers have an *enormous* number of possible interpretations combinatorially, most notably as the number of *Dyck words* of length $2n$, where a Dyck word is a string with n Xs and n Ys where no prefix of the string contains more Ys than Xs. Other interpretations: the number of ways to correctly match n pairs of parentheses, the number of ways to triangulate a polygon of $n+2$ sides, the number of distinct binary trees with n nodes, *etc.*, *etc.* There are far too many interpretations for me to cover them all, so in general, if you come across a combinatorial problem you're not sure about - check the first few simple values and see if they are the Catalan numbers!

OEIS When working with an unknown sequence, the [Online Encyclopedia of Integer Sequences](#) is a great resource (if the internet is available to you). Generate the first 10 or so of the sequence, which can usually be done by hand, and then run a search on the OEIS. If you're lucky and the sequence is clearly identifiable, it will tell you what the sequence is and give you the general formula on how to construct the sequence efficiently.

Probability OK, so with combinatorics we've been interested in counting the number of ways

Bayes' Theorem $P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$

Lab Section Again, we're going to work on this week's homework problem during lab section:

[Google Code Jam Qualification Round 2008 - Fly Swatter](#)