In the last lecture we started with the symbolic method. This lecture presents examples of its use.

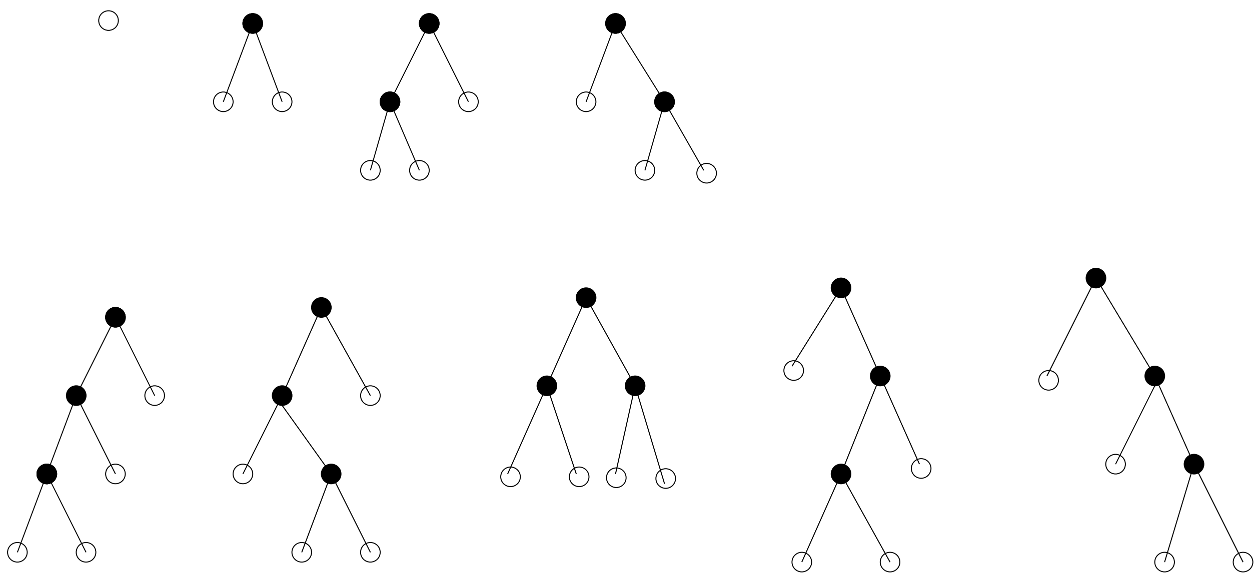
Remark. The material of this lecture is taken from Flajolet & Sedgewick [4], and in particular Chapter 1 (see [5]). Other sources are [7, 9]. The example of polynomials over finite fields is from [3].

7.1 Trees

Trees are a fundamental structure in Computer Science that underlays many applications.

Definition 1 *We define a* **binary tree** *recursively by a single external node or an internal node connected to a left subtree and a right subtree.*

Example. There is 1 tree with 1 external and 0 internal nodes, 1 tree with 2 external and 1 internal node, 2 trees with 3 external and 2 internal nodes, 5 trees with 4 external and 3 internal nodes, etc., as shown below (black nodes are internal, white are external):



In these trees the number of nodes is $2n+1$ with n internal nodes and $n+1$ external nodes. It is easy to prove by induction that always the number of external nodes is one more than the number of internal nodes.

Application: arithmetical expressions.

Other names: ordered binary trees, plane binary trees.

7.1.1 Internal Nodes

Question: How many trees are there with n internal nodes?

We use the symbolic method to give a specification for this problem.

Let $\mathcal{B}$ be the binary trees. Then, using the recursive definition of a binary tree, we have

$$\mathcal{B} = \{\varepsilon\} + (\mathcal{N} \times \mathcal{B} \times \mathcal{B}).$$

Let us define by $B(z)$ the generating function of the binary trees. Then, $[z^n]B(z)$ is the number of binary trees with n internal nodes. From the above recursive definition, it follows that the equation in generating functions is

$$B(z) = 1 + zB^2(z).$$

Solving this equation, we get the generating function for number of binary trees with internal nodes

$$B(z) = \frac{1 \pm \sqrt{1-4z}}{2z}.$$

In order to determine whether the solution is $B(z) = \frac{1+\sqrt{1-4z}}{2z}$ or $B(z) = \frac{1-\sqrt{1-4z}}{2z}$, we check $B(0) = B_0 = 1$, the number of trees with 0 internal nodes, in

$$2zB(z) = 1 \pm \sqrt{1-4z}.$$

We want the solution such that $B(0) = 1$, so we plug $z = 0$ in the above equation. We get $2 \cdot 0 \cdot B(0) = 1 \pm \sqrt{1 - 0 \cdot 4}$. That is $0 = 1 \pm 1$, and therefore the solution is

$$B(z) = \frac{1-\sqrt{1-4z}}{2z}. \tag{1}$$

We extract the nth coefficient to obtain the number of binary trees with n internal nodes. Using Newton's Theorem (Binomial Theorem) we obtain

$$\begin{aligned}
B_n &= [z^n]B(z) = [z^n]\left(\frac{1}{2z} \cdot (1 - (1-4z)^{1/2})\right) \\
&= [z^n]\left(\frac{1}{2z} \cdot (1 - \sum_{k \geq 0} \binom{1/2}{k} (-4z)^k)\right) = [z^n]\left(-\frac{1}{2z} \sum_{k \geq 1} \binom{1/2}{k} (-4z)^k)\right) \\
&= -\frac{1}{2}\binom{1/2}{n+1}(-4)^{n+1}.
\end{aligned}$$

We simplify this expression as follows

$$\begin{aligned}
B_n &= -\frac{1}{2}\binom{1/2}{n+1}(-4)^{n+1} = -\frac{1}{2}\frac{1/2 \cdot (1/2 - 1) \cdot (1/2 - 2) \cdots (1/2 - n)}{(n+1)!}(-4)^{n+1} \\
&= -\frac{1}{2}\frac{1 \cdot (-1) \cdot (-3) \cdots (-2n+1)}{(n+1)!}\frac{(-4)^{n+1}}{2^{n+1}} = \frac{1}{2}\frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{(n+1)!}2^{n+1} \\
&= \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{(n+1)!} \cdot \frac{1 \cdot 2 \cdot 3 \cdots n}{1 \cdot 2 \cdot 3 \cdots n} \cdot 2^n = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{(n+1)!} \cdot \frac{2 \cdot 4 \cdot 6 \cdots (2n)}{n!} \\
&= \frac{(2n)!}{(n+1)!n!} = \frac{1}{n+1} \cdot \frac{(2n)!}{n!n!} = \frac{1}{n+1}\binom{2n}{n}.
\end{aligned}$$

This is a Catalan Number. As an application of the Euler-Maclaurin and the Stirling formulas, one can prove that

$$B_n = \frac{4^n}{\sqrt{\pi} n^{3/2}} \left(1 - \frac{9}{8n} + \frac{145}{128n^2} - \cdots\right).$$

In other words,

$$B_n \sim \frac{4^n}{\sqrt{\pi} n^{3/2}}.$$

These numbers were introduced by Eugène Catalan in 1838 [2].

7.1.2 External Nodes

Question: How many trees are there with n external nodes?

Let's prove that the number B'_n of trees with n external nodes is equal to the number of trees with $n-1$ internal nodes, i.e. $B'_n = B_{n-1}$. Similar to the derivation for internal nodes, we have

$$\mathcal{B}' = \mathcal{N} + (\{\varepsilon\} \times \mathcal{B}' \times \mathcal{B}').$$

The equation in generating functions is

$$B'(z) = z + B'^2(z).$$

Solving this equation we obtain

$$B'(z) = \frac{1 \pm \sqrt{1-4z}}{2}.$$

Again, we determine whether $B'(z) = \frac{1+\sqrt{1-4z}}{2}$ or $B'(z) = \frac{1-\sqrt{1-4z}}{2}$ by checking $B'(0) = B'_0 = 0$, the number of trees with 0 external nodes, in

$$2B'(z) = 1 \pm \sqrt{1-4z}.$$

We want the solution such that $B'(0) = 0$, so we plug $z = 0$ in the above equation. We obtain $2 \cdot B'(0) = 1 \pm \sqrt{1 - 0 \cdot 4}$. That is, $0 = 1 \pm 1$. Therefore, the solution is

$$B'(z) = \frac{1-\sqrt{1-4z}}{2}. \tag{2}$$

We extract the nth coefficient using Newton's Theorem to get

$$\begin{aligned} B'_n &= [z^n]B'(z) = \frac{1}{2}[z^n](1-(1-4z)^{1/2}) = \frac{1}{2}[z^n]\left(1 - \sum_{k\geq 0}\binom{1/2}{k}(-4z)^k\right) \\ &= -\frac{1}{2}[z^n]\left(\sum_{k\geq 1}\binom{1/2}{k}(-4z)^k\right) = -\frac{1}{2}\binom{1/2}{n}\cdot(-4)^n = B_{n-1}. \end{aligned}$$

We could have used a more direct approach since from Equations (1) and (2)

$$B'(z) = zB(z),$$

and thus, $B'_n = B_{n-1}$.

7.1.3 Total Number of Nodes

Question: How many trees are there with n total nodes?

Let us denote by B''_n the number of trees with n total nodes. The process of extracting B''_n is similar to above. First, we derive

$$\mathcal{B}'' = \mathcal{N} + (\mathcal{N} \times \mathcal{B}'' \times \mathcal{B}'').$$

From this equation, we have

$$B''(z) = z + zB''^2(z),$$

that is,

$$B''(z) = \frac{1 \pm \sqrt{1-4z^2}}{2z}.$$

We can write

$$B''(z) = z\left(\frac{1 \pm \sqrt{1-4z^2}}{2z^2}\right) = zB(z^2).$$

Extracting the nth coefficient we obtain

$$B''_n = [z^n]B''(z) = [z^n]zB(z^2) = [z^n]z\sum_{k\geq 0} B_k z^{2k} = [z^n]\sum_{k\geq 0} B_k z^{2k+1}.$$

Finally,

$$B''_{2n+1} = B_n.$$

7.1.4 General Trees

Definition 2 *A* **tree** *is defined recursively by a node connected to a sequence of disjoint trees. We call a* **forest** *this sequence of disjoint trees.*

Symbolically,

$$\mathcal{G} = \mathcal{N} \times Seq(\mathcal{G}),$$

where $Seq(\mathcal{G})$ is a sequence of disjoint trees, and we require at least one node in the tree.

Therefore, the generating function for general trees is

$$G(z) = z\frac{1}{1-G(z)}.$$

Solving this equation as in the previous examples implies

$$G(z) = \frac{1-\sqrt{1-4z}}{2}.$$

This is the generating function for external nodes, thus

$$[z^n]G(z) = B'_n = B_{n-1},$$

that is, the number of general trees with n nodes is the same as the number of trees with n external nodes or with $n-1$ internal nodes.

For the forest, we have a specification by cutting the root

$$\mathcal{F} = Seq(\mathcal{G}).$$

We have

$$F(z) = \frac{1}{1-G(z)},$$

and thus the equation

$$G(z) = zF(z).$$

Finally, we obtain

$$F_{n-1} = [z^{n-1}]F(z) = [z^n]G(z) = B'_n = B_{n-1}.$$

In other words, the number of forests with $n-1$ nodes is the same as the number of binary trees with $n-1$ nodes. We note that a bijection can be constructed between F_n and B_n (the bijection is "first child, next sibling").

Remark 1. As we saw, Catalan numbers model ordered trees. Sometimes we say that these are Catalan trees. We also say Catalan statistics or Catalan model to refer to average and other moment studies of trees when each tree is equally likely to occur. The probability that a particular tree occurs is the inverse of a Catalan number. There are other important trees not included in this equally-likely model. For instance, BST and other balanced trees. No analysis is known for these balanced trees under random insertions; just an enumeration study of them is a difficult problem.

Remark 2. The questions we have been commenting on are related with counting the number of trees. We would like to answer more interesting questions like what is the expected height (maximum level among all nodes), expected path length (sum of the level of each of the nodes), internal and external path length (sum of level of internal/external nodes), etc. Fortunately, the symbolic method allows these studies. For that we will consider *bivariate generating functions* $A(u,z)$ such that $[u^k z^n]A(u,z)$ is the number of objects of size n with the parameter of interest equal to k. In addition, it will be easy to obtain variances and other moments by successive differentiation of the bivariate generating function with respect to the parameter u, and substitution of $u=1$. That is the material of next lecture.

7.2 Polynomials Over Finite Fields

We briefly consider another important combinatorial class: polynomials over a finite field. We plan to comment by the end of the course on how to apply these techniques to study the average-case of algorithms that deal with polynomials over finite fields.

Let $\mathbb{F}_q$ be a finite field, where $q = p^k$ and p is prime.

Fact: There are irreducible polynomials of any degree and over any finite field.

Let $\mathcal{I}$ be the collection of *monic* irreducible polynomials over $\mathbb{F}_q$. Symbolically, the multiset and set constructions give

$$\mathcal{P} = \prod_{\omega \in \mathcal{I}} \frac{1}{1-\omega}, \qquad \text{and} \qquad \mathcal{Q} = \prod_{\omega \in \mathcal{I}} (1+\omega),$$

where $\mathcal{P}$ is the class of monic polynomials and $\mathcal{Q}$ is the class of squarefree polynomials in $\mathbb{F}_q$. Their generating functions are

$$P(z) = \prod_{\omega \in \mathcal{I}} \frac{1}{1-z^{|\omega|}} = \prod_{n \geq 1} \left(\frac{1}{1-z^n} \right)^{I_n},$$

and

$$Q(z) = \prod_{\omega \in \mathcal{I}} (1+z^{|\omega|}) = \prod_{n \geq 1} (1+z^n)^{I_n},$$

where I_n is the number of irreducible polynomials of degree n. Since the number of monic polynomials in $\mathbb{F}_q$ is q^n, we have $[z^n]P(z) = q^n$, and therefore

$$P(z) = \sum_{n \geq 0} q^n z^n = \frac{1}{1-qz}.$$

Thus, we have the equality

$$\frac{1}{1-qz} = \prod_{n \geq 1} \left(\frac{1}{1-z^n} \right)^{I_n}.$$

We have an expression for I_n in an implicit form. One can obtain the approximation

$$I_n = \frac{q^n}{n} + O\left(\frac{q^{\frac{n}{2}}}{n} \right).$$

Therefore, the probability of a polynomial of degree n being irreducible is very close to $\frac{1}{n}$

$$\frac{I_n}{q^n} = \frac{1}{n} + O\left(\frac{1}{n q^{\frac{n}{2}}} \right).$$

The first algorithm for finding an irreducible polynomial of degree n was given by Rabin [8], although it was already hinted by Galois [6].

How to obtain square free polynomials? Carlitz [1] proved the following

$$[z^n]Q(z) = Q_n = \begin{cases} q^n & \text{if } n = 0, 1, \\ q^{n-1}(q-1) & \text{if } n \geq 2. \end{cases}$$

Writing a polynomial as its squarefree part times an arbitrary polynomial we obtain

$$P(z) = Q(z) \cdot P(z^2).$$

This implies

$$Q(z) = \frac{P(z)}{P(z^2)} = \frac{1}{1-qz} \cdot \frac{1-qz^2}{1} = \frac{1-qz^2}{1-qz} = (1-qz^2) \cdot \left(\sum_{k \geq 0} q^k z^k \right),$$

and extracting coefficients we obtain Carlitz's result.

Remark. As it was commented for trees, we can study interesting parameters by considering bivariate generating functions.

References

[1] L. Carlitz. The arithmetic of polynomials in a Galois field. *Amer. J. Math.*, 54:39–50, 1932.

[2] E. Catalan. Note sur une équation aux différences finies. *J. Math. Pures Appl.*, 3:508–516, 1838.

[3] P. Flajolet, X. Gourdon, and D. Panario. The complete analysis of a polynomial factorization algorithm over finite fields. Submitted March 1998. [Extended abstract in *Proc. 23rd ICALP Symp.*, Lecture Notes in Computer Science, vol. 1099, p. 232–243, 1996.] Full version in technical report 3370, INRIA, March 1998.

[4] P. Flajolet and R. Sedgewick. Analytic combinatorics. In preparation.

[5] P. Flajolet and R. Sedgewick. The average case analysis of algorithms: counting and generating functions. Technical Report 1888, INRIA, Rocquencourt, Le Chesnay, France, April 1993.

[6] É. Galois. Sur la théorie des nombres. In R. Bourgne and J.P. Arza, editors, *Écrits et mémoires d'Évariste Galois*, pages 112–128. Gauthier-Villars, 1830.

[7] I.P. Goulden and D.M. Jackson. *Combinatorial Enumeration.* John Wiley, New York, 1983.

[8] M.O. Rabin. Probabilistic algorithms in finite fields. *SIAM J. Comp.*, 9:273–280, 1980.

[9] R. Sedgewick and P. Flajolet. *An Introduction to the Analysis of Algorithms.* Addison-Wesley, Reading MA, 1996.