9.1 Labelled Construction and Exponential Generating Functions

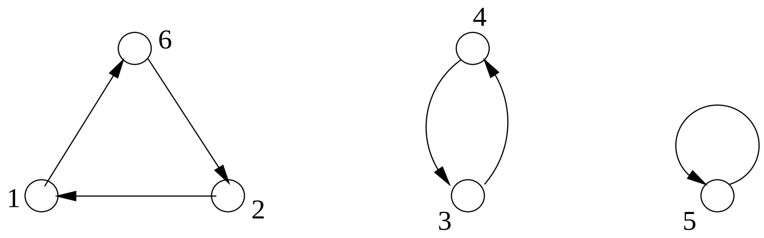
Remark. The material of this lecture is taken from Flajolet & Sedgewick [1], and in particular Chapter 2 and 3 (see [2]).

Analogously to the unlabelled constructions for OGF, we can have admissible constructions for EGF in the labelled case. We will not show these constructions in detail since they are similar to the unlabelled case. Instead, we will briefly comment on these constructions and provide similar theorems to the ones in the unlabelled case. Then, we focus on multivariate generating functions in the exponential case.

Let us assume we have nodes of a graph/tree that are labelled, so that the nodes are distinguishable. The typical example of labelled constructions are permutations seen as unordered collection of cyclic graphs whose nodes are labelled by integers. For example, the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 4 & 3 & 5 & 2 \end{pmatrix}$$

could be seen as the following graph:



Definition 1 *A* labelled construction *is formed from basic elements called* atoms *(in trees and graphs, the atoms are the nodes). The* size *of the structure is the number of atoms (that is, the number of labels).*

A structure is well-labelled *if the labels are the integers from* 1 *to a certain value* n*. When a structure is not well-labelled, we have to* re-label *the atoms. For example, for the following permutation, we may have*

$$(7 \;\; 3 \;\; 2 \;\; 8) \longmapsto (3 \;\; 2 \;\; 1 \;\; 4)$$

Notice that in this re-labelling, we have retained the order between the elements. For a formalization of the re-labelling process, see Section 2.4 in [2]. As in the unlabelled case, we have *admissible constructions for labelled structures.*

9.2 Labelled Constructions

Suppose $\mathcal{A}$ and $\mathcal{B}$ are labelled classes of combinatorial structures, with $A(z)$ and $B(z)$ as their respective exponential generating functions. Let $\mathcal{C}$ be some class constructed from $\mathcal{A}$ and $\mathcal{B}$, with $C(z)$ as its EGF. We have the following constructions.

1. **Union.** This is the same as the unlabelled case.

2. **Product.** In order to get objects of size n, we need to get k objects from $\mathcal{A}$ and $(n-k)$ objects from $\mathcal{B}$; there are $\binom{n}{k}$ ways of doing so. We do that for $k = 0, 1, \ldots, n$. Thus,

$$c_n = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}.$$

However,

$$\begin{aligned} C(z) &= \sum_{n \geq 0} \frac{1}{n!} c_n z^n = \sum_{n \geq 0} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k} z^n = \sum_{n \geq 0} \sum_{k=0}^{n} \frac{a_k}{k!} \frac{b_{n-k}}{(n-k)!} z^n \\ &= \left(\sum_{k \geq 0} \frac{a_k}{k!} z^k \right) \cdot \left(\sum_{k \geq 0} \frac{b_k}{k!} z^k \right) = A(z) \cdot B(z). \end{aligned}$$

Although the result is the same as in the unlabelled case (cartesian product translate into products of generating functions), there is something hidden here. Since the objects are labelled, relabelling of objects are needed in order to maintain a well-labelled structure.

3. **Sequence.** If $\mathcal{C}$ is a sequence of elements from $\mathcal{A}$, then we can write:

$$\mathcal{C} = \{\epsilon\} + \mathcal{A} + (\mathcal{A} \times \mathcal{A}) + \cdots.$$

Hence,

$$C(z) = 1 + A(z) + A^2(z) + \cdots = \frac{1}{1 - A(z)}.$$

4. **Set.** First let us consider k-sets from $\mathcal{A}$. Each k-set is associated with $k!$ permutations (the different sequences giving the same k-set). Thus, if we denote by $Set(\mathcal{A}, card = k)$ the set of k-sets from $\mathcal{A}$, then the EGF is $\frac{A^k(z)}{k!}$. And now the set constructions is

$$\begin{aligned} \mathcal{C} &= Set(\mathcal{A}) = \{\epsilon\} + \mathcal{A} + Set(\mathcal{A}, card = 2) + \cdots \\ &= \bigcup_{k \geq 0} Set(\mathcal{A}, card = k). \end{aligned}$$

Hence,

$$C(z) = 1 + A(z) + \frac{A^2(z)}{2} + \cdots = e^{A(z)}.$$

We have proven the following theorem.

Theorem 1 *The labelled constructions of union, product, sequence and set are all admissible. The associated operators are:*

$$\begin{array}{rcccccl} \textit{Union: } \mathcal{C} &=& \mathcal{A} + \mathcal{B} & \Longrightarrow & C(z) &=& A(z) + B(z) \\ \textit{Product: } \mathcal{C} &=& \mathcal{A} \times \mathcal{B} & \Longrightarrow & C(z) &=& A(z) \cdot B(z) \\ \textit{Sequence: } \mathcal{C} &=& Seq(\mathcal{A}) & \Longrightarrow & C(z) &=& \frac{1}{1-A(z)} \\ \textit{Set: } \mathcal{C} &=& Set(\mathcal{A}) & \Longrightarrow & C(z) &=& e^{A(z)}. \end{array}$$

Let us consider some examples.

1. **Permutations**
   The number of permutations with n elements in $n!$. Thus, the EGF is:
   $$\sum_{n\geq 0} n!\frac{z^n}{n!} = \sum_{n\geq 0} z^n = \frac{1}{1-z}.$$

2. **Cycles**
   Let us consider the EGF of cycles. The nth coefficient of the exponential generating function is the number of permutations of length n consisting of exactly one cycle. There are $\frac{n!}{n} = (n-1)!$ such permutations, since there are $(n-1)$ other permutations that are equivalent to it. Therefore, the EGF is:
   $$\sum_{n\geq 0}(n-1)!\frac{z^n}{n!} = \sum_{n\geq 0}\frac{z^n}{n} = \log\frac{1}{1-z}.$$

3. **Set of Cycles vs. Permutations**
   We can view a permutation as a set of cycles, and therefore the EGF of permutation is equal to $e^{A(z)}$ where $A(z)$ is the EGF of cycles. That is, it equals
   $$\exp\left(\log\frac{1}{1-z}\right) = \frac{1}{1-z},$$
   which is what we had in the first example.

4. **Involutions**
   We want to find permutations that are made up of cycles of lengths 1 and 2 only. We have the EGF of cycles of length 1 and 2 as
   $$z + \frac{z^2}{2}.$$
   Hence, the EGF for the set of cycles of lengths 1 and 2 is:
   $$\exp\left(z + \frac{z^2}{2}\right).$$

9.3 Parameters for EGFs

Let us consider the parameter "number of cycles in a permutation". We want to estimate the expectation, variance, and other moments of the number of cycles in a random permutation.

The EGF of permutations where the number of cycles is k is $\frac{A^k(z)}{k!}$, where $A(z)$ is the EGF of permutations with only 1 cycle. From an earlier example, we know that $A(z) = \log\frac{1}{1-z}$.

Thus, the bivariate EGF P such that $[u^k z^n]P(u,z)$ is the number of permutations of length n with k cycles is

$$\begin{aligned} P(u,z) &= \sum_{k\geq 0} u^k \frac{1}{k!}\left(\log\frac{1}{1-z}\right)^k = \sum_{k\geq 0}\frac{(u\log\frac{1}{1-z})^k}{k!} \\ &= \exp\left(u\log\frac{1}{1-z}\right) = \frac{1}{(1-z)^u}. \end{aligned}$$

Then, the expected number of cycles in a random permutation is given by:

$$\frac{\partial P(u,z)}{\partial u} = \frac{\partial}{\partial u}\left[\frac{1}{(1-z)^u}\right] = \frac{1}{(1-z)^u}\log\frac{1}{1-z}.$$

We set $u=1$ in $\frac{\partial P(u,z)}{\partial u}$, and obtain $\frac{1}{1-z}\log\frac{1}{1-z}$. We extract the nth coefficient to obtain the expected number of cycles in a random permutation of size n: $H_n = \log n + \gamma + o(\frac{1}{n})$. Thus, we expect to have $\log n$ cycles in a permutation of length n.

For the variance we differentiate again with respect to u and get

$$\frac{\partial^2 P(u,z)}{\partial u^2} = \frac{1}{(1-z)^u}\left(\log\frac{1}{1-z}\right)^2.$$

Setting $u=1$ and extracting the nth coefficient, we obtain

$$\left(\sum_{k=1}^{n}\frac{1}{k}\right)^2 - \sum_{k=1}^{n}\frac{1}{k^2}.$$

The standard deviation is thus approximately $\sqrt{\log n}$. The standard deviation is of smaller order than the expectation, and thus, we have concentration around the mean. Goncharov [4] proved that there is a Gaussian distribution on the limit. Flajolet and Soria [3] proved that there are many other structures that behave like this.

What we just did for permutations holds for other combinatorial classes.

Theorem 2 *Let $\mathcal{A}$ be an unlabelled class and consider admissible constructions of the type* Seq, Set, Multiset. *Then the OGF for the construction with u marking the number of component is*

$$\begin{aligned} \textit{Seq} &: \frac{1}{1-uA(z)}, \\ \textit{Set} &: \exp\left(\sum_{k=1}^{\infty}(-1)^{k-1}\frac{u^k}{k!}A(z^k)\right), \\ \textit{Multiset} &: \exp\left(\sum_{k=1}^{\infty}\frac{u^k}{k}A(z^k)\right). \end{aligned}$$

For a labelled class $\mathcal{A}$, the EGFs are

$$\begin{aligned} Seq &: \frac{1}{1-uA(z)}, \\ Set &: e^{uA(z)}. \end{aligned}$$

Proof. First, let us show the unlabelled cases. We know that the OGF of k components is $A^k(z)$. Thus the OGF for *Sequence* is:

$$B(u,z) = \sum_{k\geq 0} u^k A^k(z) = \sum_{k\geq 0}[uA(z)]^k = \frac{1}{1-uA(z)}.$$

The OGF for *Set* is:

$$\begin{aligned} B(u,z) &= \prod_{\alpha\in\mathcal{A}}(1+uz^{|\alpha|}) = \exp\left(\sum_{\alpha\in\mathcal{A}}\log(1+uz^{|\alpha|}\right) \\ &= \exp\left(\sum_{\alpha\in\mathcal{A}}\sum_{k=1}^{\infty}(-1)^{k-1}u^k\frac{z^{k|\alpha|}}{k}\right) \\ &= \exp\left(\sum_{k=1}^{\infty}(-1)^{k-1}\frac{u^k}{k}A(z^k)\right). \end{aligned}$$

For the labelled case, the *Sequence* is the same as in the unlabelled case. As for the *Set*, it is the same as in the permutation example, i.e.

$$B(u,z) = \sum_{k\geq 0}\frac{u^k(A(z))^k}{k!} = \sum_{k\geq 0}\frac{(uA(z))^k}{k!} = e^{uA(z)}.$$

References

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

[2] 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.

[3] P. Flajolet and M. Soria. Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory,* Series A, 53:165–182, 1990.

[4] V. Goncharov. Sur la distribution des cycles dans une permutation. *Acad. Sci. URSS (N.S.)*, 35:267–269, 1942.