L3: Symbolic Method for Labelled Structures & Exponential Generating Functions

Analytic Combinatorics
Yun Kuen Cheung, School of Computing, ANU

In the previous lectures, we considered unlabelled structures, where atoms carry no external labels. We now shift our focus to *labelled structures*, in which each object of size n is formed by a set of n distinct atoms, typically labelled with the integers $\{1, 2, \ldots, n\}$. The distinct labels completely remove the symmetries that complicated the unlabelled multiset, power set and cycle constructions, leading to a new set of elegant algebraic translations involving *exponential generating functions*.

Labelled Combinatorial Classes and Exponential Generating Functions

Example 1. Consider the class of connected graphs on n vertices. If the vertices carry no external labels, graphs are counted up to isomorphism, so there are exactly two graphs of size 3: a triangle (a cycle of 3 vertices) and a path of length 2. If instead the vertices are labelled $1, 2, 3$, the path of length 2 can be formed in three different ways, depending on which label is at the center of the path: $(1-2-3)$, $(2-1-3)$, and $(1-3-2)$. Thus the unlabelled count for $n = 3$ is 2, whereas the labelled count is 4. This gap grows rapidly: for $n = 4, 5, 6, 7$, the numbers of unlabelled connected graphs are $6, 21, 112, 853$, while the corresponding labelled counts are $38, 728, 26704, 1866256$.

Definition 2. A labelled combinatorial class is a pair $(\mathcal{A}, |\cdot|)$, where $\mathcal{A}$ is a set of labelled objects and $|\cdot| : \mathcal{A} \to \mathbb{N} \cup \{0\}$ is a size function. An object of size n consists of n atoms bearing distinct labels from the set $\{1, 2, \ldots, n\}$. The number of objects of size n is denoted $a_n = |\mathcal{A}_n|$, and we require each a_n to be finite for all $n \geq 0$.

Many interesting labelled classes grow as fast as or faster than $n!$. For example, when $a_n = n!$ counts the number of permutations, the OGF $\sum_{n=0}^{\infty} n! z^n$ converges only at $z = 0$, making it unusable for *analytic* purposes. Instead, we scale the terms by $n!$.

Definition 3. The exponential generating function (EGF) of a labelled combinatorial class $\mathcal{A}$ is the formal power series

$$A(z) \;=\; \sum_{n=0}^{\infty} a_n \frac{z^n}{n!}.$$

Admissible Constructions for Labelled Structures

The symbolic method applies equally well to labelled classes, provided we define our constructions to handle the distribution of labels properly. As before, we have two fundamental building blocks. The neutral class $\mathcal{E}$ consists of a single object of size 0 (the empty set of labels); its EGF is $E(z) = 1$. The atomic class $\mathcal{Z}$ consists of a single object of size 1 (bearing the label 1); its EGF is $Z(z) = z$.

Definition 4. The disjoint union of two labelled classes $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A} + \mathcal{B}$, is simply their set-theoretic disjoint union.

Theorem 5. If $\mathcal{C} = \mathcal{A} + \mathcal{B}$, then the EGF of $\mathcal{C}$ is $C(z) = A(z) + B(z)$.

Proof: The number of objects of size n in the union is the sum of the number of objects of size n in $\mathcal{A}$ and $\mathcal{B}$: $c_n = a_n + b_n$. Dividing by $n!$, multiplying by z^n, and summing over $n \geq 0$ yields $C(z) = A(z) + B(z)$. $\square$

When we combine two labelled objects, say $\alpha \in \mathcal{A}$ of size k and $\beta \in \mathcal{B}$ of size $n - k$, we must ensure the resulting composite object has a well-formed set of labels $\{1, 2, \ldots, n\}$. We cannot simply take the Cartesian product, because the labels would not be drawn from a single contiguous set $\{1, \ldots, n\}$, nor would they be disjoint. We must partition the n available labels and distribute them while preserving the relative internal ordering of labels within α and β.

Definition 6. The labelled product of $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A} \star \mathcal{B}$, is the set of pairs (α', β') obtained from some pair $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$ by relabelling the two components so that their combined labels form a partition of the set $\{1, 2, \ldots, |\alpha| + |\beta|\}$, while the relative order of the labels within each component is strictly preserved.

To form an object in $\mathcal{A} \star \mathcal{B}$ of size n, we first choose the size k of the first component. We then choose an object $\alpha \in \mathcal{A}_k$, an object $\beta \in \mathcal{B}_{n-k}$ and exactly k labels out of n to assign to α. The remaining $n - k$ labels are then assigned to β. The relative order of the labels within α and β must match their original labels. Because any set of k distinct labels has a unique strictly increasing mapping from $\{1, 2, \ldots, k\}$, the order-preserving relabeling is unique.

Theorem 7. If $\mathcal{C} = \mathcal{A} \star \mathcal{B}$, then the EGF of $\mathcal{C}$ is $C(z) = A(z)B(z)$.

Proof: By the definition of the labelled product, the number of objects of size n in $\mathcal{C}$ is given by summing over all possible sizes k for the first component:

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

Dividing both sides by $n!$, we obtain

$$\frac{c_n}{n!} = \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} a_k b_{n-k} = \sum_{k=0}^{n} \frac{a_k}{k!} \frac{b_{n-k}}{(n-k)!}.$$

This discrete convolution is exactly the coefficient of z^n in the product of the two power series $A(z)$ and $B(z)$. Thus, $C(z) = A(z)B(z)$. $\square$

As in the unlabelled case, we can form sequences of labelled objects, but now using the labelled product $\star$ instead of the Cartesian product $\times$.

Definition 8. The sequence class of a labelled combinatorial class $\mathcal{A}$, denoted $\mathrm{SEQ}(\mathcal{A})$, is the class of all finite sequences (tuples) of objects from $\mathcal{A}$, where the labels are distributed across all components. Formally:

$$\mathrm{SEQ}(\mathcal{A}) = \mathcal{E} + \mathcal{A} + (\mathcal{A} \star \mathcal{A}) + (\mathcal{A} \star \mathcal{A} \star \mathcal{A}) + \cdots$$

For this construction to be admissible, we require that $a_0 = 0$.

Theorem 9. If $\mathcal{C} = \mathrm{SEQ}(\mathcal{A})$ and $a_0 = 0$, the EGF of $\mathcal{C}$ is

$$C(z) \;=\; \frac{1}{1 - A(z)}.$$

Proof: Applying Theorem 5 and Theorem 7 repeatedly to the symbolic specification, we find that the EGF of the k-fold labelled product $\mathcal{A} \star \cdots \star \mathcal{A}$ is exactly $A(z)^k$. This term describes sequences of length exactly k. Summing these geometric terms over all possible lengths k yields the infinite series $\sum_{k=0}^{\infty} A(z)^k$, which evaluates to $\frac{1}{1-A(z)}$. Since $a_0 = 0$, only finitely many terms can contribute to the coefficient of each z^n, and the formal power series is well-defined. □

We now turn to unordered collections in the labelled setting. Since the atoms of the composite object carry distinct labels from $\{1, 2, \ldots, n\}$, the collection has no hidden symmetries of the kind that appeared for unlabelled multisets, power sets and cycles. This is why the labelled constructions admit simple generating function operators.

Definition 10. The set class of a labelled combinatorial class $\mathcal{A}$, denoted $\mathrm{SET}(\mathcal{A})$, is the class of all finite unordered collections of objects from $\mathcal{A}$, with labels partitioned among the chosen objects. Formally, we define the class of sets containing exactly k components as:

$$\mathrm{SET}_k(\mathcal{A}) \;=\; \frac{\mathcal{A} \star \mathcal{A} \star \cdots \star \mathcal{A}}{k!} \quad (k \text{ times}).$$

The full set class is the disjoint union $\mathrm{SET}(\mathcal{A}) = \sum_{k=0}^{\infty} \mathrm{SET}_k(\mathcal{A})$. To avoid forming infinite collections of size zero, we require $a_0 = 0$ for the construction to be admissible.

Theorem 11. If $\mathcal{C} = \mathrm{SET}(\mathcal{A})$ and $a_0 = 0$, the EGF of $\mathcal{C}$ is

$$C(z) \;=\; \exp(A(z)).$$

Proof: A set of k objects from $\mathcal{A}$ is simply a sequence of k objects from $\mathcal{A}$, where we disregard the ordering of the k components. Since the components are built from mutually disjoint subsets of distinct labels, every unordered set of k distinct components corresponds to exactly $k!$ distinct sequences. Thus, the EGF of $\mathrm{SET}_k(\mathcal{A})$ is the EGF of the k-fold labelled product divided by $k!$: $C_k(z) = \frac{A(z)^k}{k!}$. The full set class is the disjoint union of $\mathrm{SET}_k(\mathcal{A})$ for all $k \geq 0$. Summing their EGFs gives:

$$C(z) \;=\; \sum_{k=0}^{\infty} \frac{A(z)^k}{k!} \;=\; \exp(A(z)).$$ □

Definition 12. The cycle class of $\mathcal{A}$, denoted $\mathrm{CYC}(\mathcal{A})$, is the class of directed cycles of objects from $\mathcal{A}$, where the labels are partitioned among the chosen objects. Formally, we define the class of cycles containing exactly k components as the k-fold labelled product modulo cyclic shifts:

$$\mathrm{CYC}_k(\mathcal{A}) \;=\; \frac{\mathcal{A} \star \mathcal{A} \star \cdots \star \mathcal{A}}{k} \quad (k \text{ times}).$$

The full cycle class is $\mathrm{CYC}(\mathcal{A}) = \sum_{k=1}^{\infty} \mathrm{CYC}_k(\mathcal{A})$. We require $a_0 = 0$ for admissibility.

Theorem 13. If $\mathcal{C} = \text{CYC}(\mathcal{A})$ and $a_0 = 0$, the EGF of $\mathcal{C}$ is

$$C(z) \;=\; \log \frac{1}{1 - A(z)}.$$

Proof: A cycle of length k is formed by taking a sequence of k objects and factoring out the k cyclic shifts. The EGF of $\text{CYC}_k(\mathcal{A})$ is therefore the EGF of the sequence divided by k: $C_k(z) = \frac{A(z)^k}{k}$. The full cycle class is the disjoint union of $\text{CYC}_k(\mathcal{A})$ for all $k \geq 1$. Summing their EGFs gives:

$$C(z) \;=\; \sum_{k=1}^{\infty} \frac{A(z)^k}{k} \;=\; \log \frac{1}{1 - A(z)}. \qquad \square$$

Alignments, Set Partitions, Surjections, Bell Numbers and Stirling Numbers

The set construction is perfectly tailored to combinatorial problems in which a population is partitioned into disjoint non-empty blocks. Some classical examples are *alignments*, *set partitions* and *surjective functions*. In the next few examples, we will use the class $\mathcal{U}$ of a single non-empty set of labelled atoms. Its specification is $\mathcal{U} = \text{SET}_{\geq 1}(\mathcal{Z})$, and its EGF is $U(z) = e^z - 1$.

Example 14. Consider the problem of arranging n items into an ordered sequence of groups, where the order of the items within each group does not matter, but the order of the groups does. For instance, this models classifying competitors in a tournament into ranked tiers (1st place, 2nd place, etc.), where multiple competitors can tie for the same rank. Such an arrangement is called an *alignment*. With three items $\{a, b, c\}$, there are 13 alignments: 1 sequence of 1 group ($\{a, b, c\}$), 6 sequences of 2 groups (like $\{a\}, \{b, c\}$ and $\{b, c\}, \{a\}$), and 6 sequences of 3 groups (the 6 permutations of the items).

The specification for the class of alignments is $\mathcal{A} = \text{SEQ}(\mathcal{U})$. By Theorem 9, its EGF is

$$A(z) \;=\; \frac{1}{1 - U(z)} \;=\; \frac{1}{1 - (e^z - 1)} \;=\; \frac{1}{2 - e^z}.$$

We expand $A(z)$ as a geometric series:

$$A(z) \;=\; \frac{1}{2} \cdot \frac{1}{1 - \frac{e^z}{2}} \;=\; \frac{1}{2} \sum_{k=0}^{\infty} \frac{e^{kz}}{2^k} \;=\; \sum_{k=0}^{\infty} \frac{1}{2^{k+1}} \sum_{n=0}^{\infty} \frac{(kz)^n}{n!}.$$

The coefficients of this generating function, $a_n = n![z^n]A(z)$, are known as the *ordered Bell numbers* or *Fubini numbers*. This gives the exact infinite sum formula:

$$a_n \;=\; \sum_{k=0}^{\infty} \frac{k^n}{2^{k+1}}.$$

Of course, a_n is an integer since it counts the number of assignments. However, if we were given the infinite sum without the above combinatorial context, it is highly non-trivial to show that the sum converges to an integer for all $n \geq 0$.

Example 15. A *set partition* of $\{1, 2, \ldots, n\}$ is an unordered collection of non-empty, disjoint subsets (called blocks) whose union is the original set. For example, there are 5 set partitions of $\{1, 2, 3\}$: $\{\{1, 2, 3\}\}$, $\{\{1, 2\}, \{3\}\}$, $\{\{1, 3\}, \{2\}\}$, $\{\{2, 3\}, \{1\}\}$ and $\{\{1\}, \{2\}, \{3\}\}$. The order of the blocks does not matter, and the blocks themselves are non-empty sets of labelled atoms. Thus, a set partition is simply a "set of sets", and the class of set partitions is $\mathcal{B} = \mathrm{SET}(\mathcal{U})$. By Theorem 11, its EGF is

$$B(z) \;=\; \exp(U(z)) \;=\; \exp(e^z - 1).$$

The coefficients of this generating function, $b_n = n![z^n] \exp(e^z - 1)$, are known as the *Bell numbers*. Expanding gives

$$\exp(e^z - 1) \;=\; e^{-1} \sum_{k=0}^{\infty} \frac{e^{kz}}{k!} \;=\; e^{-1} \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=0}^{\infty} \frac{(kz)^n}{n!} \;=\; \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{k=0}^{\infty} \frac{k^n}{e \cdot k!},$$

and hence

$$b_n \;=\; \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

As in the previous example, it seems hard to show the above infinite sum converges to an integer if the combinatorial context is not provided.

A second extraction, leading to a recurrence, follows by differentiating the EGF:

$$B'(z) \;=\; e^z \exp(e^z - 1) \;=\; e^z B(z).$$

Since

$$B'(z) \;=\; \sum_{n=0}^{\infty} b_{n+1} \frac{z^n}{n!} \qquad \text{and} \qquad e^z B(z) \;=\; \left(\sum_{j=0}^{\infty} \frac{z^j}{j!} \right) \left(\sum_{m=0}^{\infty} b_m \frac{z^m}{m!} \right) \;=\; \sum_{n=0}^{\infty} \left(\sum_{m=0}^{n} \binom{n}{m} b_m \right) \frac{z^n}{n!},$$

coefficient comparison gives the following recurrence relation for $n \geq 1$:

$$b_n \;=\; \sum_{m=0}^{n-1} \binom{n-1}{m} b_m, \quad \text{with } b_0 = 1.$$

This recurrence can also be derived combinatorially, where $\binom{n-1}{m} = \binom{n-1}{n-1-m}$ chooses $(n - 1 - m)$ integers from $\{2, 3, \ldots, n\}$ to form a block together with 1, and b_m counts the number of set partitions of the remaining m integers.

Exercise 16. Consider the classes of set partitions in which the size of each block is

(a) exactly 1 or 2;

(b) an odd integer;

(c) a positive even integer.

Give the EGF for each class. Extract the coefficients in each case to establish a summation formula and a recurrence relation. Give a combinatorial interpretation for each recurrence relation.

Example 17. Example 15 considers set partitions with an unconstrained number of blocks. In contrast, the *Stirling numbers of the second kind*, denoted $\left\{ {n \atop k} \right\}$, count the number of ways to partition a set of n elements into exactly k blocks. The specification is $\mathrm{SET}_k(\mathcal{U})$, and the EGF is

$$\sum_{n=k}^{\infty} \left\{ {n \atop k} \right\} \frac{z^n}{n!} \;=\; \frac{1}{k!}(e^z - 1)^k.$$

We can expand the right-hand side using the binomial theorem to obtain an explicit formula for $\left\{ {n \atop k} \right\}$:

$$\frac{1}{k!}(e^z-1)^k \;=\; \frac{1}{k!}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}e^{jz} \;=\; \frac{1}{k!}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}\sum_{n=0}^{\infty}\frac{(jz)^n}{n!}.$$

Extracting the coefficient of $z^n/n!$ gives:

$$\left\{ {n \atop k} \right\} \;=\; \frac{1}{k!}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}j^n.$$

As we shall see in the next example, this summation formula can be derived using the principle of inclusion-exclusion for counting surjective functions.

Example 18. Fix $k \geq 1$. A *surjective function* $f : \{1, \ldots, n\} \to \{1, \ldots, k\}$ requires that every element in the codomain has at least one preimage. The specification of the class of all such surjections is $\mathcal{S}^{(k)} = \mathrm{SEQ}_k(\mathcal{U})$. Its EGF is

$$S^{(k)}(z) \;=\; (e^z - 1)^k.$$

To find the number of surjective functions, we extract the coefficient using the binomial theorem just as we did for the Stirling numbers:

$$(e^z-1)^k \;=\; \sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}e^{jz} \;=\; \sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}\sum_{n=0}^{\infty}\frac{(jz)^n}{n!}.$$

Taking the coefficient of $z^n/n!$, we find that the number of surjections is exactly

$$n![z^n]S^{(k)}(z) \;=\; \sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}j^n.$$

This formula can also be obtained by the principle of inclusion-exclusion: start from all k^n functions, subtract the functions missing at least one codomain element, add back those missing at least two, and so on. The factor $\binom{k}{j} = \binom{k}{k-j}$ counts the subsets of size $(k-j)$ that are excluded from the image; once these elements are excluded, there are j^n maps into the remaining j codomain elements.

Observe that $n![z^n]S^{(k)}(z) = k!\left\{ {n \atop k} \right\}$. This has a direct combinatorial explanation: a surjection first partitions the domain into k non-empty sets, which can be done in $\left\{ {n \atop k} \right\}$ ways, and then assigns the k distinct target values to these blocks, which can be done in $k!$ ways.

Exercise 19. In the setting of Example 17, prove that for $1 \leq k \leq n$,

$$\left\{ {n \atop k} \right\} \;=\; \sum_{m=k-1}^{n-1}\binom{n-1}{m}\left\{ {m \atop k-1} \right\}$$

by algebraic manipulation of generating function. Then give a combinatorial interpretation of the equation.

Example 20. Consider the number of functions $f : \{1,\dots,n\} \to \{1,\dots,k\}$ such that every element in the codomain has at least 4 pre-images, i.e., $|f^{-1}(j)| \geq 4$ for all $j \in \{1,\dots,k\}$. In the ball-and-bin model, this means we distribute n balls into k bins such that every bin contains at least 4 balls.

The class of a single such restricted bin is $\mathcal{U}_{\geq 4} = \mathrm{SET}_{\geq 4}(\mathcal{Z})$. Its EGF is the exponential series with the first four terms removed:

$$U_{\geq 4}(z) \;=\; e^z - 1 - z - \frac{z^2}{2} - \frac{z^3}{6}.$$

The class of all such valid functions is $\mathcal{S}_{\geq 4}^{(k)} = \mathrm{SEQ}_k(\mathcal{U}_{\geq 4})$. Its EGF is:

$$S_{\geq 4}^{(k)}(z) \;=\; \left(e^z - 1 - z - \frac{z^2}{2} - \frac{z^3}{6}\right)^k.$$

The explicit formula for $n![z^n]S_{\geq 4}^{(k)}(z)$ is more complex than the standard surjection case, but the symbolic method allows us to write down the EGF directly.

Example 21. For an integer $n \geq 0$, choose a set partition of $\{1,2,\dots,n\}$ uniformly at random, and let K_n denote its number of blocks. To mark the number of blocks, introduce the bivariate EGF

$$B(z,u) \;=\; \exp\left(u(e^z - 1)\right).$$

Differentiating with respect to u and then setting $u = 1$ weights each set partition by its number of blocks. Hence

$$\left.\frac{\partial}{\partial u}B(z,u)\right|_{u=1} \;=\; (e^z - 1)\exp(e^z - 1) \;=\; \frac{\mathsf{d}}{\mathsf{d}z}\exp(e^z - 1) - \exp(e^z - 1) \;=\; \sum_{n=0}^{\infty}(b_{n+1} - b_n)\frac{z^n}{n!}.$$

It follows that the total number of blocks over all set partitions of an n-element set is $b_{n+1} - b_n$. Since there are b_n such partitions, we obtain the following neat formula:

$$\mathbb{E}\left[K_n\right] \;=\; \frac{b_{n+1} - b_n}{b_n} \;=\; \frac{b_{n+1}}{b_n} - 1.$$

Laplace-Borel Summation for Discrete-Time Random Processes

In many discrete-time random processes, an expectation can be written as an infinite sum over possible running times. For instance, a non-negative integer-valued stopping time satisfies $\mathbb{E}\left[T\right] = \sum_{n\geq 0}\mathbb{P}\left[T > n\right]$. When the events at time n are counted by an EGF, this produces sums of the form $\sum_{n\geq 0} f_n/r^n$, where the factorial denominator in the EGF must be cancelled. The *Laplace-Borel transform* provides exactly this cancellation through the identity $\int_0^\infty z^n e^{-z}\,\mathsf{d}z = n!$.

Lemma 22. Laplace-Borel Transform. Let $F(z) = \sum_{n=0}^{\infty} f_n \frac{z^n}{n!}$ be an EGF with non-negative coefficients. Fix $r > 0$, and assume that the power series for $F(t)$ converges for every $t \geq 0$ and that the integral below is finite. Then

$$\sum_{n=0}^{\infty}\frac{f_n}{r^n} \;=\; \int_0^\infty F(z/r)e^{-z}\,\mathsf{d}z.$$

Example 23. We can use the EGF for surjections to derive the expected number of trials required to collect all k coupons in the *coupon collector problem*. Let T be the random variable representing the number of trials needed. The event $T > n$ means that after n trials, the sequence collected so far is *not* surjective onto the set of k coupon types. Let

$$F(z) \;=\; e^{kz} - (e^z - 1)^k.$$

Here $n![z^n]e^{kz} = k^n$ counts all length-n coupon sequences, while $n![z^n](e^z - 1)^k$ counts the surjective sequences (see Example 18). Thus, if $f_n = n![z^n]F(z)$, then

$$\mathbb{P}\left[T > n\right] \;=\; \frac{f_n}{k^n}.$$

A fundamental identity for non-negative integer-valued random variables gives the expectation as the sum of survival probabilities: $\mathbb{E}\left[T\right] = \sum_{n=0}^{\infty} \mathbb{P}\left[T > n\right]$. Applying Lemma 22 with $r = k$ gives

$$\mathbb{E}\left[T\right] \;=\; \sum_{n=0}^{\infty} \frac{f_n}{k^n} \;=\; \int_0^\infty F(z/k)e^{-z}\,\mathsf{d}z \;=\; \int_0^\infty \left(1 - (1 - e^{-z/k})^k\right)\mathsf{d}z.$$

We evaluate this integral using the substitution $y = 1 - e^{-z/k}$, which implies $e^{-z/k} = 1 - y$ and $\mathsf{d}y = \frac{1}{k}e^{-z/k}\,\mathsf{d}z = \frac{1-y}{k}\,\mathsf{d}z$. The limits of integration change from $[0, \infty)$ to $[0, 1]$. The integral becomes:

$$\mathbb{E}\left[T\right] \;=\; k\int_0^1 \frac{1 - y^k}{1 - y}\,\mathsf{d}y \;=\; k\int_0^1 (1 + y + y^2 + \cdots + y^{k-1})\,\mathsf{d}y \;=\; k\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{k}\right) \;=\; kH_k,$$

where H_k is the k-th harmonic number. The expected time kH_k can also be derived by elementary means, by noting that the expected number of trials to obtain a new coupon when j uncollected coupons remain is k/j.

Exercise 24. Consider the variant of the coupon collector problem in which the process stops only after every one of the k coupon types has appeared at least twice. Let T_2 be the corresponding stopping time. Use the Laplace-Borel transform to show that

$$\mathbb{E}\left[T_2\right] \;=\; k\int_0^\infty \left(1 - \left(1 - (1 + y)e^{-y}\right)^k\right)\mathsf{d}y.$$

The asymptotic analysis is not easy; we only mention here that $\mathbb{E}\left[T_2\right] = k\log k + k\log\log k + O(k)$.

Example 25. Run the coupon collector process until all k coupons have appeared, and let X be the number of coupon types that have appeared exactly once at the stopping time. The last coupon collected always contributes one such type. We ask how many of the other $k - 1$ coupon types are also singletons at that moment.

Fix the type of the last coupon. The preceding trials form a word over the other $k - 1$ coupon types, and each of these types must occur at least once. We mark a coupon type that appears exactly once with u. For one fixed coupon type, the EGF contribution is uz if it appears exactly once and $e^z - 1 - z$ if it appears at least twice. Hence the bivariate EGF for the prefix is

$$G(z, u) \;=\; \left(uz + e^z - 1 - z\right)^{k-1} \;=\; \left(e^z - 1 + (u - 1)z\right)^{k-1}.$$

Thus $g_{n,m} = n![z^n u^m]G(z,u)$ is the number of length-n prefixes in which exactly m of the $(k-1)$ old coupon types appear exactly once, while all remaining old coupon types appear at least twice.

For a fixed prefix counted by $g_{n,m}$, there are k possible types for the final newly collected coupon, and the corresponding stopped sequence has probability $k^{-(n+1)}$. Such a stopped sequence contributes $m+1$ to X. Therefore

$$\mathbb{E}[X] \;=\; \sum_{n,m\geq 0}(m+1)\frac{k g_{n,m}}{k^{n+1}} \;=\; \sum_{n,m\geq 0}(m+1)\frac{g_{n,m}}{k^n}.$$

Differentiating $G(z,u)$ with respect to u weights each prefix by the number of old coupon types that are singletons, and the extra $G(z,1)$ term accounts for the final coupon. By Lemma 22,

$$\mathbb{E}[X] \;=\; \int_0^\infty e^{-z}\left(\frac{\partial}{\partial u}G(z/k,u)\bigg|_{u=1} + G(z/k,1)\right)\mathsf{d}z.$$

For $k=1$, this expectation is plainly $1 = H_1$. Assume now that $k \geq 2$. We have

$$G(z/k,1) \;=\; (e^{z/k}-1)^{k-1}, \qquad \frac{\partial}{\partial u}G(z/k,u)\bigg|_{u=1} \;=\; \frac{(k-1)z}{k}(e^{z/k}-1)^{k-2}.$$

Substituting $y = 1-e^{-z/k}$ gives

$$\mathbb{E}[X] \;=\; k\int_0^1 y^{k-1}\,\mathsf{d}y + k(k-1)\int_0^1 -y^{k-2}(1-y)\log(1-y)\,\mathsf{d}y.$$

Using the integral identity $\int_0^1 -x^r\log(1-x)\,\mathsf{d}x = \frac{H_{r+1}}{r+1}$, which holds for any non-negative integer r, we obtain

$$\mathbb{E}[X] \;=\; 1 + k(k-1)\left(\frac{H_{k-1}}{k-1} - \frac{H_k}{k}\right) \;=\; 1 + kH_{k-1} - (k-1)H_k \;=\; H_k.$$

Exercise 26. Consider the random variable X from Example 25. Show that its variance is $(H_k)^2 - 3H_k + 2\sum_{i=1}^k \frac{1}{i^2}$.

Exercise 27. Let $k \geq 1$ be an integer and let $0 < p < 1$. On each independent draw, a special ball occurs with probability p, while each of k ordinary ball types occurs with probability $(1-p)/k$. Stop when the special ball first occurs, and let X be the number of distinct ordinary types observed before stopping. By using the Laplace-Borel transform, prove that

$$\mathbb{E}[X] \;=\; \frac{k(1-p)}{1+(k-1)p}.$$

Permutations and Restricted Cycles

Every permutation decomposes uniquely into disjoint cycles on the underlying labels. This makes permutations a natural testing ground for the labelled cycle construction: once a permutation is viewed as a set of cycles, restrictions on cycle lengths or statistics such as the number of cycles translate directly into symbolic specifications and hence into EGFs.

Example 28. Let $\mathcal{P}$ be the class of permutations. Recall that every permutation of $\{1, 2, \ldots, n\}$ can be uniquely factored into a set of disjoint cycles. The components of these cycles are single elements, so the class of such cycles is $\mathcal{C} = \mathrm{CYC}(\mathcal{Z})$. By Theorem 13, its EGF is $C(z) = \log \frac{1}{1-z}$. The class of permutations is $\mathcal{P} = \mathrm{SET}(\mathcal{C})$. Applying Theorem 11, the EGF of permutations is:

$$P(z) \;=\; \exp(C(z)) \;=\; \exp\left(\log \frac{1}{1-z}\right) \;=\; \frac{1}{1-z} \;=\; \sum_{n=0}^{\infty} z^n \;=\; \sum_{n=0}^{\infty} n! \frac{z^n}{n!}.$$

Thus, we recover the familiar fact that the number of permutations of n elements is $p_n = n!$.

Example 29. A *derangement* is a permutation with no fixed points. In other words, a derangement is a set of cycles in which all cycles have length at least 2. The class of cycles of length at least 2 is $\mathcal{C}_{\geq 2} = \mathrm{CYC}_{\geq 2}(\mathcal{Z})$. Its EGF is obtained by subtracting the length-1 term from the full cycle EGF:

$$C_{\geq 2}(z) \;=\; \log \frac{1}{1-z} - z.$$

The class of derangements is $\mathcal{D} = \mathrm{SET}(\mathcal{C}_{\geq 2})$. Its EGF is

$$D(z) \;=\; \exp\left(\log \frac{1}{1-z} - z\right) \;=\; \frac{e^{-z}}{1-z}.$$

To find the number of derangements d_n, we extract the coefficients:

$$d_n \;=\; n![z^n]\left(\frac{1}{1-z} e^{-z}\right) \;=\; n! \sum_{k=0}^{n} \left([z^{n-k}] \frac{1}{1-z}\right)\left([z^k] e^{-z}\right) \;=\; n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

This reproduces the classic inclusion-exclusion formula for derangements. The limiting probability that a random permutation is a derangement is

$$\lim_{n \to \infty} \frac{d_n}{n!} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = \frac{1}{e}.$$

Exercise 30. An *involution* is a permutation that is its own inverse. Equivalently, its cycle decomposition consists exclusively of 1-cycles (fixed points) and 2-cycles (transpositions).

(a) Formulate the symbolic specification for the class of involutions, $\mathcal{I}$, and derive its EGF $I(z)$.

(b) Show that $I'(z) = (1+z)I(z)$, and use this to prove the recurrence relation $i_n = i_{n-1} + (n-1)i_{n-2}$ for $n \geq 2$, with $i_0 = i_1 = 1$. Give a combinatorial interpretation of the recurrence relation.

(c) If σ is drawn uniformly at random from all permutations of $\{1, 2, \ldots, n\}$, show that the limiting probability that σ is an involution is 0.

Exercise 31. Let $\mathcal{F}$ be the class of permutations σ for which σ^4 has no fixed points.

(a) Derive the EGF $F(z)$ of $\mathcal{F}$.

(b) If σ is drawn uniformly at random from all permutations of $\{1, 2, \ldots, n\}$, show that the limiting probability that σ^4 has no fixed points is $e^{-7/4}$.

Example 32. The unsigned *Stirling numbers of the first kind*, denoted $\left[{n \atop k}\right]$, count the number of permutations of length n consisting of exactly k disjoint cycles. When k is fixed, the class is $\mathrm{SET}_k(\mathrm{CYC}(\mathcal{Z}))$, and its EGF is

$$\sum_{n \geq k} \begin{bmatrix} n \\ k \end{bmatrix} \frac{z^n}{n!} \;=\; \frac{1}{k!}\left(\log \frac{1}{1-z}\right)^k.$$

We can track the number of cycles using a secondary variable u: $\mathcal{P} = \mathrm{SET}(u \cdot \mathrm{CYC}(\mathcal{Z}))$. This yields the bivariate EGF:

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

By definition, $P(z,u) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \left[{n \atop k}\right] u^k \frac{z^n}{n!}$. Using the generalized binomial theorem, we can extract the coefficients in z:

$$[z^n](1-z)^{-u} \;=\; (-1)^n \binom{-u}{n} \;=\; \frac{u(u+1)\cdots(u+n-1)}{n!}.$$

Therefore, the polynomial $\sum_{k=0}^{n} \left[{n \atop k}\right] u^k$ is exactly the *rising factorial* $u^{\overline{n}} = u(u+1)\cdots(u+n-1)$.

We can use $P(z,u)$ to find the expected number of cycles in a random permutation of length n. To find the total number of cycles across all permutations of length n, we differentiate with respect to u and set $u = 1$:

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

To extract the coefficient of z^n, we write this expression as the product of two power series: $\frac{1}{1-z} = \sum_{k=0}^{\infty} z^k$ and $\log \frac{1}{1-z} = \sum_{m=1}^{\infty} \frac{z^m}{m}$. Multiplying a power series by $\frac{1}{1-z}$ corresponds to taking the partial sums of its coefficients. Therefore:

$$[z^n]\left(\frac{1}{1-z} \log \frac{1}{1-z}\right) \;=\; \sum_{m=1}^{n} [z^m] \log \frac{1}{1-z} \;=\; \sum_{m=1}^{n} \frac{1}{m} \;=\; H_n.$$

The total number of cycles across all permutations of length n is $n![z^n]\frac{\partial P}{\partial u} = n!H_n$. Dividing by the $n!$ possible permutations, the expected number of cycles is exactly H_n.

An interesting combinatorial bijection known as *Foata's fundamental correspondence* links cycles to records in a permutation. In a permutation sequence, a *record* (or left-to-right maximum) is an element that is strictly greater than all elements to its left. Foata's correspondence maps each permutation to another permutation by writing each cycle with its maximum element first, sorting the cycles in increasing order of their maximum elements, and then erasing the cycle parentheses. The left-to-right maxima of the resulting sequence are exactly the cycle leaders. Thus, $\left[{n \atop k}\right]$ also counts the number of permutations of length n with exactly k records.

The correspondence gives a simple way to compute the expected number of cycles. For any element ℓ, it is a record if and only if it is the first element among $\{\ell, \ell+1, \ldots, n\}$ to appear in the permutation. This occurs with probability $\frac{1}{n-\ell+1}$, and hence the expected number of cycles is $\sum_{\ell=1}^{n} \frac{1}{n-\ell+1} = H_n$.

Exercise 33. In Example 32, the expected number of cycles is shown to be H_n. Show that the variance of the number of cycles is $H_n - \sum_{i=1}^{n} \frac{1}{i^2}$.