On some enumeration problems related to bicolored objects

Danièle GARDY*

Abstract

We consider the problem of enumerating combinatorial objects built on two types of basic blocks, according to their total size and to the excess of one type of blocks. We show that the generating functions involved can be obtained in a systematic way, and that they may be amenable to a systematic treatment, then consider the existence of a limiting distribution for the number of components of specified balance.

Résumé Cet article traite de l'énumération d'objets combinatoires bicolores, selon leur taille totale, et selon le déséquilibre des couleurs. Nous indiquons comment la série génératrice bivariée peut être obtenue systématiquement, puis étudions la loi limite du nombre de composants de déséquilibre donné.

A well-known combinatorial problem is to determine the number of components of type $\mathcal{A}$ of an object of type $\mathcal{C} = \mathcal{B}(\mathcal{A})$, when the total size of the $\mathcal{C}$ object is known. There are many results on this subject for *admissible constructs*. Classical examples are the number of cycles in a random permutation, or the number of non-empty subsets in the partition of a set of n objects, enumerated by Stirling numbers of the first and second kind; see for example [13] for a recent survey.

Now assume that the combinatorial construct $\mathcal{A}$ is no longer built as usual on a single type of basic objects, or atoms, but on two types : red and blue atoms. We can classify the objects of type $\mathcal{A}$ according to the relative excess of red elements on blue ones : the *balance,* or according to the dominant color. It is easy to determine the probability that an object of $\mathcal{A}$ belongs to one of these classes. A more involved question is to study the number of $\mathcal{A}$-components in a given class, for an object of $\mathcal{C}$ with total size n. Our interest in this problem comes from the paper [4], where the problem we studied there (evaluation of the generalization error in a learning framework) was modelled by introducing a sequence of sets, i.e. an occupancy urn model, and computing the number of urns with positive balance.

We present our framework and an overview of our results in Section 1, and consider an example in Section 2. We then show in Section 3 how generating functions for bicolored objects can be obtained in a systematic way, before turning to the asymptotic study of the number of components with specified balance in Section 4.

1 Overview of the model

Let us make precise the type of results we are seeking. We consider bicolored combinatorial objects of type $\mathcal{C} = \mathcal{B}(\mathcal{A})$; let c_n be the number of objects of $\mathcal{C}$ with size n, and $c_{n,q,k}$ be the number of such objects that have k $\mathcal{A}$-components of balance q. For example, we consider permutations as sets of cycles ($\mathcal{A}$ is a cycle and $\mathcal{B}$ a set); $c_n = 2^n n!$ is the number of bicolored permutations on n elements, where each element may be red or blue; the balance of a cycle is the difference between the number of red elements and the number of blue elements of the cycle; and $c_{30,2,5}$ is the number of permutations on 30 elements that have exactly five cycles with an excess of two red atoms each, and an unspecified (possibly null) number of cycles with different balances.

*Laboratoire PRISM, Université de Versailles St-Quentin and CNRS UMR 8636, Versailles, France.

Define a random variable $X_{n,q}$ on objects of $\mathcal{C}$ to be equal to the number of $\mathcal{A}$-components that have a specified balance q, conditioned on the total size n of the $\mathcal{C}$ structure. We have that

$$Proba\ (X_{n,q} = k) = \frac{c_{n,q,k}}{c_n}.$$

We would like to compute the expectation and variance of $X_{n,q}$, and to determine if a limiting distribution exists when the global size n goes to infinity and q is fixed.

The enumeration of the objects of $\mathcal{A}$ with a specified balance q and a given size n presents no difficulty. Let a_n be the number of (non colored) objects of $\mathcal{A}$ that have size n; the number of such objects with balance q is $\binom{n}{(n-q)/2} a_n$ for $-n \leq q \leq n$, 0 otherwise. However, we want to be able to combine the condition on the balance (or a "majority" condition, which requires that the balance of a component be non-negative) with other, so-called "admissible" constructs (see for example [12] for an introduction to admissible constructs and associated generating functions). To this effect, we need to be able to write the generating functions associated with $\mathcal{A}$-components of given balance, in a way that will allow us to inject them into other generating functions, and to do asymptotics on the resulting function.

More specifically, we begin by considering the combinatorial objects of $\mathcal{A}$ obtained by the following rule, which we call the *specified balance rule* :

> *Start from a given type $\mathcal{A}$, built by some combinatorial construct from basic objects, which we call "atoms". Assume that the atoms can take two colors : blue and red. We are interested in the objects of $\mathcal{A}$, such that the number of atoms of one color (say, red) differs from the number of atoms of the other color (blue) by a specified number q ($q \in Z$). Define $\tilde{\mathcal{A}}_q$ as the set of such objects.*

We next consider the *majority rule,* i.e. the rule that builds the set $\tilde{\mathcal{A}}$ of combinatorial objects with null or positive balance, starting from a basic type $\mathcal{A}$:

$$\tilde{\mathcal{A}} = \bigoplus_{q \geq 0} \tilde{\mathcal{A}}_q.$$

For example, if $\mathcal{A}$ is a set, the elements of $\tilde{\mathcal{A}}_q$ are sets of p blue atoms and $p+q$ red atoms, for any $p \geq 0$, and the elements of $\tilde{\mathcal{A}}$ are the sets with a number of red atoms equal to or greater than the number of blue atoms. If $\mathcal{A}$ is a cycle, the elements of $\tilde{\mathcal{A}}_q$ are cycles with p blue atoms and $p+q$ red atoms (the relative order of appearance counts), and the elements of $\tilde{\mathcal{A}}$ are the cycles with more red atoms than blue atoms.

We shall see in Section 3 that the generating functions enumerating the objects of $\mathcal{A}$ with a specified excess of red elements, or the objects of $\mathcal{A}$ with a majority of red elements, can be written as Hadamard products of the generating function enumerating $\mathcal{A}$ and of some functions related to Dyck paths, and give a combinatorial explanation of this result.

Throughout the paper, we shall use consistently the variables y, z and u as follows : y denotes the (total) size of the combinatorial object currently under study; z marks the balance of the global object, i.e. the relative difference between the numbers of red and blue atoms; for a composed object, u is the number of its components that have a specified balance.

2 Bicolored Permutations

A permutation is a set of cycles, i.e. a construct $\mathcal{P} = \mathcal{B}(\mathcal{A})$, with $\mathcal{A}$ a cycle and $\mathcal{B}$ a set. The exponential generating functions are respectively $\hat{A}(y) = \log(1/(1-y))$, $\hat{B}(y) = e^y$, and $\hat{P}(y) = 1/(1-y)$.

2.1 Enumeration of bicolored permutations by size and balance on the number of balls

Let $f_{n,q}$ be the number of permutations on n elements with balance q, i.e. built on $(n+q)/2$ red balls and $(n-q)/2$ blue balls; we have that $f_{n,q} = n!\binom{n}{(n-q)/2}$, and the bivariate generating function enumerating the permutations by their size and balance is

$$\hat{F}(y,z) := \sum_{n,q} f_{n,q} \frac{y^n}{n!} z^q = exp\left(\log \frac{1}{1-y(z+1/z)}\right) = \frac{1}{1-y(z+1/z)}.$$

The sum $\sum_q f_{n,q} = n![y^n]\hat{F}(y,1)$ is equal to the number $n!\,2^n$ of bicolored permutations. The exponential generating function for the cycles of balance q is $\hat{\varphi}_q(y) := [z^q]\hat{F}(y,z) = \sum_n f_{n,q} y^n/n!$. For positive q,

$$\hat{\varphi_q}(y) = \sum_r \frac{1}{2r+q}\binom{2r+q}{r} y^{q+2r}.$$

From [15, p. 201-203], we have that

$$\sum_{r:2r+q>0} \frac{1}{2r+q}\binom{2r+q}{r} t^r = \frac{1}{q} B_2(t)^q, \tag{1}$$

with $B_2(t) = (1-\sqrt{1-4t})/(2t)$ the generating function for Catalan numbers; hence the expression of $\hat{\varphi_q}$ $(q>0)$ as $(1/q)$ $(yB_2(y^2))^q = (1/q)\left(\frac{1-\sqrt{1-4y^2}}{2y}\right)^q$. For balanced cycles $(q=0)$, we get $\hat{\varphi_0}(y) = \sum_{r>0}\binom{2r}{r}\frac{y^{2r}}{2r} = \log\frac{1-\sqrt{1-4y^2}}{2y^2} = \log B_2(y^2)$. The function $\hat{\Phi}(y)$ describing the predominantly red cycles is $\hat{\Phi}(y) = \sum_{q\geq 0}\hat{\varphi_0}(y) = \log\frac{B_2(y^2)}{1-yB_2(y^2)} = \log\frac{2}{1-2y+\sqrt{1-4y^2}}$. We have simple closed-form expressions for the numbers $\Phi_n := [y^n/n!]\hat{\Phi}(y)$, according to their parity : $\Phi_{2p} = (2p-1)!\left(2^{2p-1}+\frac{1}{2}\binom{2p}{p}\right)$ and $\Phi_{2p+1} = (2p)!\,2^{2p}$.

2.2 Number of permutations on cycles with specified balance

Define $\mathcal{P}_q$ as the set of permutations whose cycles all have the same balance q : $\mathcal{P}_q = \mathcal{S}et(\mathcal{C}_q)$, with $\mathcal{C}_q$ the cycles of balance q, i.e. with any number p of blue balls and $p+q$ red balls ($p \in N$, $q \in Z$ and $p+q \geq 0$). The generating function for $\mathcal{P}_q$ is

$$\hat{P_q}(y) := \sum_n p_{q,n}\frac{y^n}{n!} = exp(\hat{\varphi_q}(y)),$$

with $p_{q,n}$ the number of bicolored permutations on n elements, with an unspecified number of cycles, and such that each cycle has balance q.

The generating function enumerating the permutations on balanced cycles is $\hat{P_0}(y) = exp(\hat{\varphi_0}(y) = B_2(y^2)$, and $p_{0,2n+1} = 0$, $p_{0,2n} = (2n)!\ C_n$. Such a permutation can be seen as the product of a standard permutation on $2n$ elements and of a Dyck path of length $2n$, where each return to 0 corresponds to the end of a cycle.

Assume now that q is strictly positive. Then $\hat{\varphi_q}(y) = (1/q)(yB_2(y^2))^q$, which gives $\hat{P_q}(y) = \sum_n p_{q,n}y^n/n! = exp[(1/q)(yB_2(y^2))^q]$. What is the asymptotic number of these bicolored permutations? We shall evaluate it by singularity analysis applied to the function $\hat{\varphi_q}(y)$ or to related functions [11]. To proceed further, we have to take into account the parity of q :

- For q even, $q = 2p$, we have that $\hat{P}_{2p}(y) = \tilde{P}_p(4y^2)$, with

$$\tilde{P}_p(z) := exp\left(\frac{(1-\sqrt{1-z})^{2p}}{2p\,z^p}\right).$$

  Hence $p_{2p,2n+1} = 0$ and $p_{2p,2n} = (2n)!\,4^n[z^n]\tilde{P}_p(z)$. The function $\tilde{P}_p$ has an algebraic singularity at $z=1$; near this singularity $\tilde{P}_p(z) \sim e^{1/2p}\left(1-\sqrt{1-z}+O(1-z)\right)$, which gives, by a transfer lemma

$$[z^n]\tilde{P}_p(z) \sim \frac{e^{1/2p}}{2n\sqrt{\pi n}}.$$

- For q odd, $q = 2p+1$, the function $\hat{P}_{2p+1}(y)$ is equal to $\tilde{P}_{2p+1}(2y)$, with

$$\tilde{P}_{2p+1}(z) := exp\left(\frac{\left(1-\sqrt{1-z^2}\right)^{2p+1}}{(2p+1)z^{2p+1}}\right),$$

and $p_{2p+1,n} = n!\,2^n[z^n]\tilde{P}_{2p+1}(z)$. Computing the asymptotic value of the coefficient $[z^n]\tilde{P}_{2p+1}(z)$ requires us to take into account the two algebraic singularities of $\tilde{P}_{2p+1}$ at ± 1. Again a transfer lemma gives us the result :

$$[z^n]\tilde{P}_{2p+1}(z) \sim \frac{e^{1/(2p+1)} + (-1)^{n+1}e^{-1/(2p+1)}}{n\sqrt{2\pi n}}.$$

For example, we get $p_{1,n} \sim n!2^n(e + (-1)^{n+1}e^{-1})/n\sqrt{2\pi n}$, which depends on the parity of n.

To sum up, we have that $p_{2p,2n+1} = 0$ and

$$\begin{aligned} p_{2p,2n} &\sim 2^{2n}(2n)!\frac{e^{1/2p}}{2n\sqrt{\pi n}}; \\ p_{2p+1,n} &\sim 2^n n!\frac{e^{1/(2p+1)} + (-1)^{n+1}e^{-1/(2p+1)}}{n\sqrt{2\pi n}}. \end{aligned}$$

As a consequence, the proportion of permutations whose cycles all have a specified balance q is either 0 (if q is even and n is odd) or of order $n^{-3/2}$.

2.3 Number of cycles with specified balance in a random permutation

Assume now that we are interested in the number X of cycles with specified balance in a random permutation (for example cycles of balance 0, or of positive balance). Let $g_{n,l}$ be the number of bicolored permutations on n elements with exactly l cycles of the desired balance, and an unspecified number of cycles of different balance; and define $\hat{f}(y)$ as the function enumerating the bicolored cycles with the desired balance. Then the enumerating function, exponential in the total size and ordinary in the number of cycles, is :

$$G(u,y) := \sum_{n,l} g_{n,l}u^l\frac{y^n}{n!} = \frac{e^{(u-1)\hat{f}(y)}}{1-2y}.$$

The average number of cycles with specified balance is

$$E[X] = \frac{[y^n]\partial G/\partial u(1,y)}{[y^n]G(1,y)} = \frac{1}{2^n}[y^n]\left\{\frac{\hat{f}(y)}{1-2y}\right\} = [y^n]\lambda(y),$$

with $\lambda(y) = \hat{f}(y/2)/(1-y)$.

- For cycles with balance 0, $\hat{f}(y) = \hat{\varphi}_0(y) = \log B_2(y^2)$. The function $\lambda(y) := \frac{1}{1-y}\log\frac{2(1-\sqrt{1-y^2})}{y^2}$ has algebraic singularities at ± 1; the main contribution comes from the singularity at $+1$, and we get

$$E[X] = [y^n]\lambda(y) = \log 2 - \sqrt{2/\pi n} + O(n^{-3/2});$$

  the average number of balanced cycles has a limit equal to $\log 2 = 0.6931\ldots$.

- For cycles with a strictly positive balance $q > 0$, the enumerating function is $\hat{f}(y) = \hat{\varphi}_q(y)$, and the average number of cycles with balance q is

$$E[X] = \frac{1}{q}\left\{[y^n]\frac{1}{1-y}\left(\frac{1-\sqrt{1-y^2}}{y}\right)^q\right\}.$$

  Singularity analysis gives $E[X] = 1/q + O(1/\sqrt{n})$.

- For cycles with positive balance, $\hat{f}(y) = \hat{\Phi}(y)$ and $\lambda(y) = 1/(1-y)\ \log 2/(1-y+\sqrt{1-y^2})$. The singularities of λ are $y = \pm 1$. The contribution from the singularity $+1$ is $(1/2)\log n(1+o(1))$, and the singularity -1 gives a smaller order term : As could be expected, the average number of cycles with a positive balance is equal to half the average total number of cycles.

The variance is obtained from the generating function by

$$\sigma^2(X) = \frac{[y^n]\partial^2 G/\partial u^2(1,y)}{[y^n]G(1,y)} + E[X] - E[X]^2 = \frac{1}{2^n}[y^n]\left\{\frac{\hat{f}(y)^2}{1-2y}\right\} + E[X] - E[X]^2.$$

We do not discuss its asymptotics, which could be obtained in a similiar vein, as it is more interesting to turn at once to the limiting distribution for cycles of specified balance : The existence and form of this distribution are direct applications of a former study by Drmota and Soria [8, 22]; see the discussion in Section 4.1. We sum up the asymptotic results in

Theorem 2.1 *The number of balanced cycles in a random bicolored permutation has average value and variance* $\log 2$. *It has a discrete limiting distribution : for each* k, $Pr(k) \sim (1/2)(\log 2)^k/k!$. *The number of cycles of balance* $q > 0$ *in a random bicolored permutation has average value and variance equal to* $1/q$. *Its limiting distribution is given by* $Pr(k) \sim e^{-1/q}(1/q)^k/k!$. *The number of cycles with positive balance follows asymptotically a Gaussian limiting distribution, of mean* $(1/2)\log n$ *and variance* $\sqrt{\log n}$.

3 Generating functions for bicolored objects

3.1 How do we obtain them?

The balance of an element α of $\mathcal{A}$ is defined as the difference between the number of red balls and the number of blue balls in α; it is a relative integer. Let us assume that the ordinary function enumerating the objects of type $\mathcal{A}$ is $A(y) = \sum_{p\geq 0} a_p y^p$ (later on, we shall see that exponential generating functions behave in the same way). We introduce the variable z to keep track of the balance. The number of objects of size n and balance q is $a_n\binom{n}{(n-q)/2}$, and we obtain the bivariate generating function describing the objects of $\mathcal{A}$, enumerated w.r.t. the size (y) and their balance (z) as

$$A\left(y\left(z+\frac{1}{z}\right)\right) = \sum_{n\geq 0, q\in Z} a_n \binom{n}{\frac{n-q}{2}} z^q y^n =: \sum_{q\in Z} \varphi_q(y) z^q,$$

with the functions $\varphi_q(y)$ enumerating the objects of $\mathcal{A}$ with balance q. We shall also wish to use the generating function associated with the set $\tilde{\mathcal{A}}$ of objects that have a majority of red atoms; this function is $\Phi(y) := \sum_{q\geq 0} \varphi_q(y)$. Computing the functions $\varphi_q(y)$ gives :

$$\varphi_q(y) := \sum_{\substack{0\leq r\leq p\\ p-2r=q}} a_p \binom{p}{r} y^p = \sum_{0\leq r\leq q+2r} a_{q+2r}\binom{q+2r}{r} y^{q+2r}. \tag{2}$$

The summation is on r such that $0 \leq r \leq q+2r$, i.e. $r \geq 0$ for $q \geq 0$, and $r \geq -q = |q|$ for $q < 0$. For $q = 0$, $\varphi_0(y) = \sum_{0\leq r} a_{2r}\binom{2r}{r} y^{2r}$. For negative $q = -p$, we have the obvious relation $\varphi_{-p} = \varphi_p$, and we can express the function Φ in a simple form :

$$2\Phi(y) = \varphi_0(y) + A(2y).$$

3.2 Hadamard products

It is interesting to see the functions φ_q and Φ as Hadamard products of the enumeration function $A(y)$ and of another function. The Hadamard product of two power series $f(t) = \sum_n f_n t^n$ and $g(t) = \sum_n g_n t^n$ is defined as

$$(f \odot g)(t) := \sum_n f_n g_n t^n.$$

For positive q, and according to Equation (2), the function φ_q can be written as a variation of the Hadamard product of $A(y)$ and of the function $y \mapsto \sum_r \binom{q+2r}{r} y^r = B_2(t)^q/\sqrt{1-4t}$. Define

$$f_q(y) := \sum_r \binom{q+2r}{r} y^{q+2r} = \frac{(y\, B_2(y^2))^q}{1-2y^2 B_2(y^2)};$$

then $\varphi_q(y) = (A \odot f_q)(y)$. The expression (1) holds for all real r, and we get for $q=0$

$$f_0(y) = \frac{1}{\sqrt{1-4y^2}} = \frac{1}{1-2y^2 B_2(y^2)}.$$

Now, by linearity of the Hadamard product,

$$\Phi(y) = \sum_{q\geq 0} \varphi_q(y) = \sum_{q\geq 0} (A \odot f_q)(y) = (A \odot \sum_{q\geq 0} f_q)(y) = (A \odot F)(y),$$

with $F(y) := \sum_{q\geq 0} f_q(y)$. Simplifying, we obtain :

$$F(y) = \frac{1}{(1-2y^2B_2(y^2))\,(1-yB_2(y^2))} = \frac{1}{2}\left(\frac{1}{\sqrt{1-4y^2}} + \frac{1}{1-2y}\right).$$

If the objects of $\mathcal{A}$ are best enumerated by the exponential generating function $\hat{A}(z) = \sum_p a_p z^p/p!$ (this happens for example with labelled objects), the bivariate function enumerating them w.r.t. their size and balance is $\sum_{q\in Z} \hat{\varphi}_q(y)\, z^q$, with

$$\hat{\varphi}_q(y) = \sum_r a_{q+2r} \binom{q+2r}{r} \frac{y^{q+2r}}{(q+2r)!}.$$

Hence $\hat{\varphi}_q = \hat{A} \odot f_q$, where the function f_q is the same as for ordinary generating functions. We sum up our results so far in

Theorem 3.1 *The ordinary generating functions φ_q and Φ and the exponential generating functions $\hat{\varphi}_q$ and $\hat{\Phi}$ enumerating the objects of $\tilde{A}_q$ and $\tilde{A}$, i.e. the objects with a specified balance q or with a positive or null balance, can be obtained from the ordinary or exponential generating functions $A(y)$ or $\hat{A}(y)$ by taking their Hadamard product with suitable functions, as follows :*

$$\varphi_q = A \odot f_q; \quad \hat{\varphi}_q = \hat{A} \odot f_q; \qquad \Phi = A \odot F; \quad \hat{\Phi} = \hat{A} \odot F,$$

with $f_{-q} = f_q$ and, for $q \geq 0$:

$$f_q(y) := \frac{1}{\sqrt{1-4y^2}} \left(\frac{1-\sqrt{1-4y^2}}{2y}\right)^q;$$

$$F(y) := \frac{1}{2}\left(\frac{1}{\sqrt{1-4y^2}} + \frac{1}{1-2y}\right).$$

The functions Φ and $\hat{\Phi}$ can also be expressed as

$$\Phi(y) = \frac{1}{2}(A(2y) + \varphi_0(y)); \qquad \hat{\Phi}(y) = \frac{1}{2}(\hat{A}(2y) + \hat{\varphi}_0(y)).$$

More generally, the function associated with the set $\cup_{q\in\mathcal{E}} \tilde{A}_q$ of objects such that their balance q belongs to a set $\mathcal{E}$ is obtained as the Hadamard product of $A(y)$ or $\hat{A}(y)$, and of $\sum_{q\in\mathcal{E}} f_q$. An important consequence of recognizing Hadamard products is that, as the functions f_q and F are algebraic, with singularities at $\pm 1/2$, we have some a-priori information on the location of the singularities of their Hadamard product, and can sometimes decide their types [7, 21] : The singularities of φ_q and Φ are among the points $\pm\alpha/2$, where α is a singularity of $A(y)$, and their radius of convergence is half the radius of convergence of $A(y)$. If $A(y)$ is entire, then so is φ_q; if $A(y)$ is a rational function, then φ_q is algebraic; however, if $A(y)$ is itself an algebraic function, then the Hadamard product of A and f_q may be a transcendental function. We shall see examples of this when considering some common combinatorial constructs $\mathcal{A}$.

3.3 Basic constructs

The case where $\mathcal{A}$ *is a set,* or equivalently an infinite urn, was treated in the paper [4]. It was proved there that the function $\hat{\varphi}_q(y)$ is (almost) a Bessel coefficient. We can derive easily this result by taking the exponential generating function $\hat{A}(y) = e^y$: $a_p = 1$, and we get

$$\hat{\varphi_q}(y) = \sum_r \frac{1}{(q+r)!r!} y^{q+2r} = I_q(2y) \qquad with \qquad I_q(t) = \sum_r \frac{(t/2)^{q+2r}}{r!(q+r)!} \qquad (q \in Z).$$

The sequence of balances is a random walk, and the apparition of Bessel functions was already noted in [10, p. 59-60]. Now the function associated with predominantly red sets is

$$\hat{\Phi}(y) = \sum_{q \geq 0} I_q(2y) =: \sum_{n \geq 0} \Phi_n \frac{y^n}{n!},$$

and we have simple expressions for the Φ_n, according to the parity of n : $2\Phi(y) = e^{2y} + I_0(2y)$.

When $\mathcal{A}$ *is a sequence,* we start from the ordinary generating function for $\mathcal{A}$: $A(y) = 1/(1-y) = \sum_{p \geq 0} y^p$. This is the neutral element for the Hadamard product : $(A \odot f)(y) = f(y)$. Hence φ_q is equal to f_q, and Φ is equal to F (see Theorem 3.1 for the definition of these functions).

When $\mathcal{A}$ *is a cycle,* we use exponential generating functions : $\hat{A}(y) = \log 1/(1-y) = \sum_{p \geq 1} y^p/p$. We have already seen in Section 2 that

$$\begin{aligned}
\hat{\varphi_q}(y) &= \frac{1}{q}\ (yB_2(y^2))^q = \frac{1}{q}\left(\frac{1-\sqrt{1-4y^2}}{2y}\right)^q \qquad (q > 0);\\
\hat{\varphi_0}(y) &= \log B_2(y^2);\\
\hat{\Phi}(y) &= \frac{1}{2}\log\frac{1}{1-2y} + \frac{1}{2}\log B_2(y^2) = \log\frac{2}{1-2y+\sqrt{1-4y^2}}.
\end{aligned}$$

Finally, let's consider the case of *Catalan trees.* The generating function for $\mathcal{A}$ is an ordinary one : $A(y) = B_2(y) = (1-\sqrt{1-4y})/(2y)$. We get

$$\varphi_q(y) = \sum_r C_{q+2r}\binom{q+2r}{r} y^{q+2r} = \sum_r \binom{2q+4r}{q+2r}\binom{q+2r}{r}\frac{y^{q+2r}}{q+2r+1}.$$

For example,

$$\varphi_0(y) = \sum_r C_{2r}\binom{2r}{r} y^{2r} = \sum_r \binom{4r}{2r}\binom{2r}{r}\frac{y^{2r}}{2r+1}.$$

We also have an expression of the function enumerating predominantly red trees :

$$\Phi(y) = \frac{1}{2}(\varphi_0(y) + B_2(2y)) = \frac{1}{2}\left(\sum_r C_{2r}\binom{2r}{r} y^{2r} + \sum_r C_r\, 2^r\, y^r\right).$$

The functions φ_q are transcendental series, as well as $\Phi(y) = (1/2)(A(2y) + \varphi_0(y))$ [1]. We can write these functions as hypergeometric functions, following the method outlined in [15, p. 207-208]. Let us see what happens for $\varphi_0(y)$: Define $t_k := C_{2k}\binom{2k}{k}y^{2k}$; $t_0 = 1$ and the ratio of two consecutive terms is a rational fraction in k :

$$\frac{t_k}{t_{k+1}} = \frac{4y^2\,(4k+1)\,(2k+1)\,(4k+3)}{(2k+3)(k+1)^2} = (64\,y^2)\,\frac{(k+1/4)\,(k+1/2)\,(k+3/4)}{(k+3/4)(k+1)^2};$$

hence $\varphi_0(y)$ can be expressed in terms of an hypergeometric function :

$$\varphi_0(y) = \sum_{k \geq 0} C_{2k}\binom{2k}{k} y^{2k} = F\left(\begin{matrix}\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\\ 1, \frac{3}{2}\end{matrix}\ \middle|\ 64y^2\right).$$

In the same vein, we have that

$$\varphi_q(y) = y^q\ F\left(\begin{matrix}\frac{1+2q}{4}, \frac{1+q}{2}, \frac{3+2q}{4}\\ q+1, \frac{3+q}{2}\end{matrix}\ \middle|\ 64y^2\right).$$

3.4 Hypergeometric functions

We have seen that the functions enumerating Catalan trees with specified balance are expressible using hypergeometric functions. This is a general property of most, if not all, of these functions. Let us define $h_{q,r} := a_{q+2r}\binom{q+2r}{r}$ and $h_q(z) := \sum_r h_{q,r} z^r$; then $\varphi_q(y) = y^q h_q(y^2)$. Can we characterize the constructs leading to hypergeometric functions?

We recall (see for example [15, p. 205]) that an hypergeometric function is a series whose coefficients can be expressed in terms of rising factorial powers $a^{\bar{n}} = a(a+1)\dots(a+n-1)$, as follows :

$$F\left(\begin{array}{c} a_1,\dots,a_m \\ b_1,\dots,b_p \end{array} \middle| z\right) = \sum_{n\geq 0} \frac{a_1^{\bar{n}}\dots a_m^{\bar{n}}}{b_1^{\bar{n}}\dots b_p^{\bar{n}}}\frac{z^n}{n!} =: \sum_n f_n z^n.$$

In such a case, the first coefficient is equal to 1, and the ratio of two consecutive coefficients is a rational function of the index n :

$$\frac{f_{n+1}}{f_n} = \frac{(n+a_1)\dots(n+a_m)}{(n+b_1)\dots(n+b_p)(n+1)}. \tag{3}$$

The reverse of (3) is also true, and we shall use it as a characterization of hypergeometric functions (see again [15, p. 207-208]) : *Let $f(z) = \sum_n f_n z^n$; if the ratio f_{n+1}/f_n can be written as*

$$\frac{f_{n+1}}{f_n} = \lambda\frac{(n+a_1)\dots(n+a_m)}{(n+b_1)\dots(n+b_p)(n+1)}, \tag{4}$$

and if $f_0 = 1$, then

$$f(z) = F\left(\begin{array}{c} a_1,\dots,a_m \\ b_1,\dots,b_p \end{array} \middle| \lambda z\right).$$

For fixed q, the ratio of two consecutive coefficients of h_q is :

$$\frac{h_{q,r+1}}{h_{q,r}} = \frac{a_{q+2r+2}}{a_{q+2r}}\frac{\binom{q+2r+2}{r+1}}{\binom{q+2r}{r}} = \frac{4}{r+1}\cdot\frac{a_{q+2r+2}}{a_{q+2r}}\cdot\frac{(r+1+q/2)(r+(1+q)/2)}{r+q+1}.$$

If the ratio a_{q+2r+2}/a_{q+2r} for fixed q is a rational function of r, then so is the ratio $h_{q,r+1}/h_{q,r}$ and, by (4), h_q is an hypergeometric function. More precisely, we can prove the

Theorem 3.2 *The functions $\varphi_q(y)$ can be written as $\varphi_q(y) = y^q h_q(y^2)$, where h_q is an hypergeometric function in the following cases :*

1. *If the function $A(y) = \sum_n a_n y^n$ is such that the ratio of two consecutive coefficients of the same parity is a rational function of n :*

$$\frac{a_{n+2}}{a_n} = \lambda\frac{(n+\alpha_1)\dots(n+\alpha_m)}{(n+\beta_1)\dots(n+\beta_p)},$$

then

$$h_q(t) = a_q F\left(\begin{array}{c} \frac{\alpha_1+q}{2},\dots,\frac{\alpha_m+q}{2}, 1+\frac{q}{2}, \frac{1+q}{2} \\ \frac{\beta_1+q}{2},\dots,\frac{\beta_p+q}{2}, 1+q \end{array} \middle| 2^{m-p+2}\lambda t\right).$$

2. *If the function $A(y)$ is itself an hypergeometric function :*

$$A(y) = F\left(\begin{array}{c} \alpha_1,\dots,\alpha_m \\ \beta_1,\dots,\beta_p \end{array} \middle| y\right),$$

then

$$h_q(t) = a_q F\left(\begin{array}{c} \frac{\alpha_1+q}{2},\dots,\frac{\alpha_m+q}{2}, \frac{\alpha_1+q+1}{2},\dots,\frac{\alpha_m+q}{2}, \\ q+1, \frac{\beta_1+q}{2},\dots,\frac{\beta_p+q}{2}, \frac{\beta_1+q+1}{2},\dots,\frac{\beta_p+q+1}{2} \end{array} \middle| 4^{m-p+2}t\right).$$

3.5 Bessel generating functions

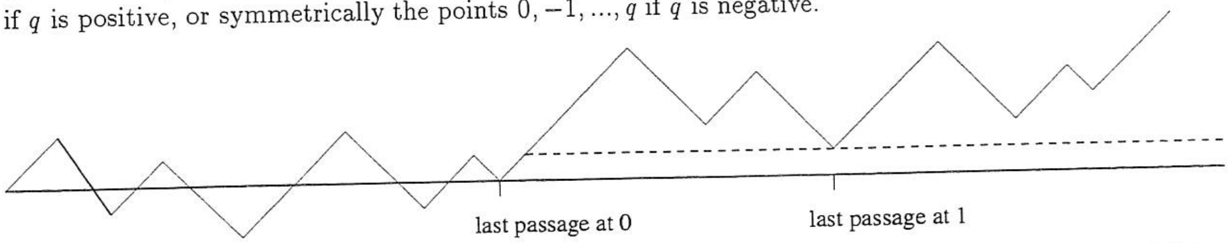
The Bessel generating function of a sequence $\{a_n\}$ is $\sum_n a_n z^n/(n!)^2$; it was used by Camion and Sole [5], and later by Fedou and Rawlings [9]; it is related to the enumeration of *pairs* of objects. The function $\hat{\varphi_0}(t) = \sum_r a_{2r} \binom{2r}{r} t^{2r}/(2r)! = \sum_r a_{2r} t^{2r}/(r!)^2$ can be seen as the Bessel generating function of the sequence a_{2r}. Anticipating on the combinatorial explanation given below, we can say that we are enumerating simultaneously two objects, both of size n, one of them being built on red atoms and the other being built on blue atoms, from which we obtain the final object of size $2n$ and balance 0.

When working with unbalanced objects, we are considering a generalization of Bessel generating functions, where the pairs of objects (the one on blue atoms and the other on red atoms) can have different sizes.

3.6 Combinatorial interpretation

It is interesting to notice that the functions $f_q(y)$ and $F(y)$ are related only to the balance or to the dominant color of a set, not to the construct $\mathcal{A}$ that operates on the elements of this set, nor to the type (ordinary or exponential) of the generating function for $\mathcal{A}$. The functions f_q and F involve the function $B_2(t)$, enumerating classical objects : binary trees, Dyck paths,..., and it is only natural to try and seek a combinatorial explanation.

The Hadamard product corresponds to building a combinatorial object of $\mathcal{A}_q$ of size n from a combinatorial object of $\mathcal{A}$ and one of the objects enumerated by f_q, both of the same size n, so a natural question is : *What are the objects enumerated by f_q?* The answer is simple : They are the *sequence of the n successive balances* obtained when we build an object of $\mathcal{A}$ by adding red or blue atoms one at a time. This sequence, which can have integer positive or negative values, and ends at the value q, can be enumerated by considering the points at which it takes for the last time the values $0, 1, ..., q$ in this order, if q is positive, or symmetrically the points $0, -1, ..., q$ if q is negative.



We give below a bijection for the sequence of balances when q is positive. We shall use a for the addition of a red ball, which increases the balance and which we can represent by an *up* step, and b for a blue ball, or similarly a *down* step.

- The addition of red and blue atoms (one at a time) gives a balance that can be positive or negative, and that at some point takes for the last time the value 0. This corresponds to a generalized Dyck path, on *up* and *down* steps, that can pass under 0, and that ends at 0.
- Then we take an *up* step, which gives a balance (height of the path) 1. Afterwards, the balance stays at least equal to 1. Its behaviour between this point and the point of last passage to 1 is described by a Dyck path. This point is followed by an *up* step, then again by a Dyck path until the point of last passage to 2, and so on until the path ends at height q. This behaviour can be summarized as $(a.Dyck)^q$.

To sum up, the sequence of balances is given as a generalized Dyck path, followed by the prefix of a Dyck path, ending at height q.

The case of negative q can be treated in the same way, by a symmetry after the end of the generalized Dyck path : The combinatorial explanation for the function $F(y)$ related to components with non-negative balance follows the same lines : A sequence of balances either ends at level 0, i.e. is a generalized Dyck path, or is a generalized Dyck path, followed by the prefix of a (standard) Dyck path : The path ends at an unspecified, but (strictly) positive level.

We can sum up this section very simply : *To build a combinatorial object with balance q, we just build it while keeping track of the balance, and mix the two phenomena "term by term", as implied by the Hadamard product.*

4 Number of components of specified balance

Many results are known on the number of components of a decomposable combinatorial structure, and we can expect similar results for the number of components with a specified balance : See [2, 6, 3, 14, 16, 8, 18, 19] for various results, and [13] for a global presentation and summary.

For bicolored objects, the (ordinary or exponential) generating function enumerating the objects of $\mathcal{B}(\mathcal{A})$ according to their size and to the number of components whose balance is in a given set $\mathcal{E}$ is

$$C(y,u) = B(A(2y) + (u-1)f(y)),$$

with $f(y)$ enumerating the objects of $\mathcal{A}$ with balance in $\mathcal{E}$. The asymptotic study of the number of such $\mathcal{A}$-components in a random $\mathcal{C}$ object requires knowledge of the singularities of the function C, which are in part determined by those of the functions $A(y)$ and $f(y)$, and in part by the type of the function B.

4.1 $\mathcal{B}$ is a set

In this case,

$$C(y,u) = e^{A(2y)-f(y)}\, e^{uf(y)}.$$

This is a product scheme according to the definition of Drmota and Soria [22]. Define r as the radius of convergence of $f(y)$; then r is also the radius of convergence of $y \mapsto e^{A(2y)-f(y)}$ and is half the radius of convergence of $A(y)$. The number X of components of specified balance can behave in only one of the following two ways (see the discussion in [22]) :

- If the function f has a finite limit $f(r)$ for $y \to r^-$, then the r.v. X has a discrete limiting distribution, with $Pr(X=k) \sim f(r)^k e^{-f(r)}/k!$.
- If $f(y) \to \infty$ for $y \to r^-$, then the limiting distribution is Gaussian, with mean $f(\rho)$, where ρ is the saddle point, defined by the equation $2yA'(2y) - yf'(y) = n$.

Now the behaviour of f depends on the exact balance condition, i.e. on the set $\mathcal{E}$. For example, for permutations, $A(y) = \log 1/(1-y)$; the radius of convergence is $r = 1/2$, and

- for components with balance equal to q, $f(1/2) = \varphi_q(1/2)$ is finite and we get a discrete limiting distribution;
- for components with positive balance, $f(y) = \Phi(y) \to \infty$ for $y \to 1/2$ and we get a Gaussian limiting distribution with mean $(1/2)\log n$.

4.2 $\mathcal{B}$ is a sequence

Now $B(y) = 1/(1-y)$ and

$$G(y,u) = \frac{1}{1 - A(2y) - (u-1)f(y)}.$$

This is a meromorphic scheme of the type studied by Bender [2]. Following the presentation given in [13, p. 54-55], we see that the limiting distribution is Gaussian, provided that the equation $A(2y) = 1$ has a single solution of modulus smaller than half the radius of convergence of A and that some "variability condition" is satisfied. Assume from now on that there exists a solution ρ to the equation $A(2y) = 1$; the variability condition can be written as

$$Var > 0, \qquad with \qquad Var := \frac{f(\rho)}{2\rho A'(2\rho)} \left(\frac{f(\rho)}{2\rho A'(2\rho)} + 1 - \frac{f(\rho)}{A'(2\rho)} - \frac{f(\rho)A''(2\rho)}{A'(2\rho)^2} \right). \tag{5}$$

If this holds, the limiting distribution has for average value $nf(\rho)/2\rho A'(2\rho)$, and for variance $nVar$.

For example, let us take non-empty sets as components : $A(y) = e^y - 1$ and $G(y,1)$ has a single pole at $\rho = (1/2)\log 2$. For a fixed balance equal to q, the condition (5) becomes

$$\left(\frac{1}{\log 2} - 1\right) I_q(\log 2) + 2 - I_q'(\log 2) > 0$$

and it is easy to check that it is satisfied for all $q > 0$: The number of sets with balance q follows asymptotically a Gaussian limiting distribution with an average value $nI_q(\log 2)/2\log 2$, and a variance also of order n.

4.3 $\mathcal{B}$ is a finite sequence, or urn models

When $\mathcal{B}$ is a finite sequence of m elements, and $\mathcal{A}$ is a set, we are in fact considering urn models, where a combinatorial object is simply a random allocation of n balls into a sequence of m urns. This is a classical problem of discrete probability, for which many results are known in the non-colored case; see for example the book of Johnson and Kotz [20]. We have $B(y) = y^m$, $\hat{A}(y) = e^y$, and the generating function for the allocation of bicolored balls is

$$G(u,y) = \left(e^{2y} + (u-1)\hat{f}(y)\right)^m,$$

where $\hat{f}(y) = I_q(2y)$ for urns with fixed balance q and $\hat{f}(y) = \sum_{q\geq 0} I_q(2y)$ for urns with non-negative balance. What is the probability that there are k urns with a specified (equal to some q or positive) balance? The relevant generating function is either $f(y) = \hat{\Phi}(y)$ or $f(y) = I_q(2y)$; then

$$Proba(k/n) = \frac{\binom{m}{k}}{n!\,(2m)^n}[y^n]\{f^k(y)(e^{2y} - f(y))^{m-k}\}.$$

When the number m of urns and the number n of balls have the same growth rate, the number $X_{n,m}$ of urns with a specified balance, either positive of equal to some q, can be analyzed in the quasi-powers framework of Hwang [17], and follows a limiting Gaussian distribution. Its average value is obtained by a saddle point approximation; here the saddle point is $\rho = n/2m$, and $E[X_{n,m}] \sim mf(\rho)e^{-\rho}$. The variance is also of order m.

We may extend this approach to consider more general "urns" with different $\mathcal{A}$-components. If the saddle-point equation $2yA'(2y)/A(2y) = n/m$ has a solution smaller than half the radius of convergence of A, then we expect a limiting Gaussian distribution; otherwise the singularity of smallest modulus of A will determine the asymptotic behaviour.

5 Possible extensions

We have presented in this paper results on the enumeration of bicolored composed objects of the type $\mathcal{B}(\mathcal{A})$, and on the number of their components when the construct $\mathcal{B}$ is a finite or infinite sequence, or a set. We should be able to extend this approach to take into account other classical combinatorial constructs. We might also consider iterating the majority rule, i.e. define colors on the $\mathcal{B}$ components of objects $\mathcal{C}(\mathcal{B}(\mathcal{A}))$, according to the majority of their $\mathcal{A}$ components; what kind of results can be expected? Interesting extensions might also take into account unequal probabilities for the two colors, and three or more colors.

Acknowledgements are due to J.P. Allouche, P. Flajolet and D. Gouyou-Beauchamps for helpful discussions about various aspects of this work.

References

[1] J.P. ALLOUCHE. Transcendence of formal power series with rational coefficients. *Theoretical Computer Science*, to appear.

[2] E. BENDER. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory (A)*, 15:91–111, 1973.

[3] E.A. BENDER and L.B. RICHMOND. Central and local limit theorems applied to asymptotic enumeration II: Multivariate generating functions. *Journal of Combinatorial Theory, Series A*, 34:255–265, 1983.

[4] S. BOUCHERON and D. GARDY. An urn model from learning theory. *Random Structures and Algorithms*, 10:43–67, January 1997. Special issue on mathematical analysis of algorithms.

[5] P. CAMION and P. SOLÉ. The Bessel generating function. Unpublished manuscript, 1989.

[6] E. R. CANFIELD. Central and local limit theorems for the coefficients of polynomials of binomial type. *Journal of Combinatorial Theory (A)*, 23:275–290, 1977.

[7] P. DIENES. *The Taylor Series.* Dover, New York, 1958. A reprint of the first Oxford University Press edition, 1931.

[8] M. DRMOTA and M. SORIA. Marking in combinatorial constructions : generating functions and limiting distributions. *Theoretical Computer Science*, 144:67–99, 1995.

[9] J-M. FEDOU and D. RAWLINGS. Statistics on pairs of permutations. *Discrete Mathematics*, 143:31–45, 1995.

[10] W. FELLER. *An introduction to probability theory and its applications (Vol. 2).* Wiley, 1971.

[11] P. FLAJOLET and A. ODLYZKO. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3(2):216–240, 1990.

[12] P. FLAJOLET and R. SEDGEWICK. *An introduction to the analysis of algorithms.* Addison-Wesley, 1996.

[13] P. FLAJOLET and R. SEDGEWICK. The average-case analysis of algorithms : multivariate asymptotics and limit distributions. Technical Report 3162, INRIA, 1997.

[14] P. FLAJOLET and M. SORIA. Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory*, 153:165–182, 1990.

[15] R.L. GRAHAM, D.E. KNUTH, and O. PATASHNIK. *Concrete Mathematics.* Addison-Wesley, 1989.

[16] J.C. HANSEN. Order statistics for decomposable combinatorial structures. *Random Structures and Algorithms*, 5(4):517–533, 1994.

[17] H.K.HWANG. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19:329–343, 1998.

[18] H.-K. HWANG. A Poisson * geometric convolution law for the number of components in combinatorial structures. Technical report, Institute of Statistical Science, Academia Sinica, TAIWAN, 1995.

[19] H.-K. HWANG. A Poisson $*$ geometric convolution law for the number of components in unlabelled combinatorial structures. *Combinatorics, Probability and Computing*, 7:89–110, 1998.

[20] N.L. JOHNSON and S. KOTZ. *Urn models and their application.* Wiley & Sons, 1977.

[21] R. JUNGEN. Sur les séries de Taylor n'ayant que des singularités algébrico-logarithmiques sur leur cercle de convergence. *Commentarii Mathematici Helvetici*, 3:266–306, 1931.

[22] M. SORIA. Limiting distributions in product schemes. In *Séminaire Algorithmes, Inria-Rocquencourt, 1993-94*, pages 83–86, 1994.