Smallest Components in Combinatorial Structures

Daniel Panario
University of Toronto

February 16, 1998

[summary by Philippe Flajolet]

Abstract

The smallest size of components in random decomposable combinatorial structures is studied in a general framework. The results apply to several combinatorial structures in both the labelled and the unlabelled case. Typical examples are the cycle decomposition of permutations and the factorization of polynomials over finite fields into irreducible factors.

1. Introduction

Many types of combinatorial objects decompose as *sets* of simpler basic objects diversely known as "prime", "irreducible", or "connected" components. For instance, a permutation decomposes as a set of cyclic permutations, a polynomial as a (multi)set of irreducible factors, and a graph as a set of connected components. Such situations are combinatorial analogues of the fact that natural numbers uniquely decompose as products of primes.

Let $\mathcal{I}$ be a class of basic objects, $\mathcal{F}$ the class of all sets of objects from $\mathcal{I}$, that is

$$\mathcal{F} = \mathtt{Set}(\mathcal{I}).$$

As usual, this schema covers both the labelled case (L) where sets are built upon labelled products, and the unlabelled case (U) where multisets are intended. Enumeration is treated by generating functions [5]. The generating functions (gf's) $F(z), I(z)$ corresponding to $\mathcal{F}, \mathcal{I}$, are taken to be either the exponential generating function (egf) in the labelled case or the ordinary generating function in the unlabelled case,

$$\begin{aligned} (L): &\qquad F(z) = \sum_n F_n \frac{z^n}{n!} &&\qquad I(z) = \sum_n I_n \frac{z^n}{n!} \\ (U): &\qquad F(z) = \sum_n F_n z^n &&\qquad I(z) = \sum_n I_n z^n, \end{aligned}$$

with F_n, I_n the number of objects of size n in $\mathcal{F}, \mathcal{I}$. Then, the fundamental relations between generating functions are given by the exponential formulæ:

$$\begin{aligned} &(L): \quad F(z) = e^{I(z)} \\ &(U): \quad F(z) = \prod_{k=1}^{\infty} (1 - z^k)^{-I_k} = \exp\left(I(z) + \frac{1}{2} I(z^2) + \frac{1}{3} I(z^3) + \cdots \right). \end{aligned} \tag{1}$$

The construction covers a number of classical combinatorial structures like permutations (cyclic, general), monic polynomials over a finite field of cardinality q (irreducible, general), functional

graphs (connected, general) in either the labelled or the unlabelled case. In fact, the examples just cited all belong to an interesting class called the "exp-log" class that was introduced in [4].

Definition 1. A pair $(\mathcal{I}, \mathcal{F})$ is said to have the exp-log property if $I(z)$ has a unique dominant singularity ρ of the logarithmic type,

$$I(z) \underset{z\to\rho}{\sim} a \log \frac{1}{1 - z/\rho} + c_0 + O((1 - z/\rho)^{\epsilon}), \tag{2}$$

for some $\epsilon > 0$, where a is called the multiplier. Accordingly, one has

$$F(z) \sim e^{I(z)} \sim c_1 (1 - z)^{-a}, \qquad c_1 = e^{c_0}. \tag{3}$$

It is understood that these expansions should hold in an indented disk of the type required by singularity analysis.

Based on the known facts for integers [12] and on specific combinatorial examples, the following properties are expected to hold true:

1. **Prime Number Theorem:** The *asymptotic density of irreducible objects* satisfies
$$\frac{I_n}{F_n} \sim (a e^{-c_0} \Gamma(a)) \frac{1}{n^a}.$$
2. **Gaussian law:** The *number of irreducible components* in a random $\mathcal{F}$-object of size n is asymptotically Gaussian with mean and variance each asymptotic to $a \log n$.
3. **Dickman's law:** The density of $\mathcal{F}$-object of size n whose *largest* $\mathcal{I}$-component is of size $m = n/u$ involves a function of which a prototype is the Dickman function $\rho(u)$ classically defined by the difference-differential equation
$$\rho(u) = 1 \ (0 \le u \le 1), \qquad u\rho'(u) + \rho(u - 1) = 0 \ (u > 1).$$
4. **Buchstab's law:** The density of $\mathcal{F}$-object of size n whose *smallest* $\mathcal{I}$-component is of size $m = n/u$ involves a function of which a prototype is the Buchstab function $\omega(u)$ classically defined by the difference-differential equation
$$u\omega(u) = 1 \ (1 \le u \le 2), \qquad (u\omega(u))' = \omega(u - 1) \ (u > 2).$$

The Prime Number Theorem for exp-log classes derives immediately from basic singularity analysis theorems. The Gaussian law was established in [4] by means of characteristic functions, thanks to the uniformity afforded by singularity analysis; it is an analogue of the classical Erdős-Kac theorem for the number of prime divisors of integers. The Dickman law is known originally from number theory [12] and it holds as well for the cycle decomposition of permutations [10], its extension to the general framework of exp-log classes being due to Gourdon [7]. The purpose of the talk is precisely to establish for exp-log structures the Buchstab law of smallest components by building upon Gourdon's analysis of largest components.

2. Cycles in Permutations

In its simplest terms the problems are well exemplified by the analysis of smallest and largest cycles in permutations. In an important paper, Shepp and Lloyd [10] established the Dickman law and the Buchstab law for permutations. Their approach is however based on an asymptotic-probabilistic model of permutations as sums of Poisson random variables of rates $1, \frac{1}{2}, \frac{1}{3}, \ldots$ relayed by nonconstructive Tauberian arguments. Gourdon [7] was instead able to push the analytic approach to its ultimate limits, thereby solving the long-standing Golomb-Knuth conjecture; see [6].

From standard methods of enumerative combinatorics the egf's of permutations with all their cycles of size at most m ($P^{[\le m]}(z)$) or at least $m+1$ ($P^{[>m]}(z)$) are given by

$$P^{[\le m]}(z) = \exp\left(\frac{z}{1} + \frac{z^2}{2} + \cdots + \frac{z^m}{m}\right) \tag{4}$$
$$= \frac{1}{1-z}\exp\left(-\frac{z^{m+1}}{m+1} - \frac{z^{m+2}}{m+2} - \cdots\right)$$
$$P^{[>m]}(z) = \exp\left(\frac{z^{m+1}}{m+1} + \frac{z^{m+2}}{m+2} + \cdots\right) \tag{5}$$
$$= \frac{1}{1-z}\exp\left(-\frac{z}{1} - \frac{z^2}{2} - \cdots - \frac{z^m}{m}\right).$$

Let L_n and S_n be the random variables that represent the largest cycle and the smallest cycle in a random permutation of size n. Equations (4) and (5) give access to probabilities, as

$$\Pr\{L_n \le m\} = [z^n]P^{[\le m]}(z), \qquad \Pr\{S_n > m\} = [z^n]P^{[>m]}(z).$$

In the analytic perspective, an important rôle is thus played by the decomposition of the logarithm into its partial sum and remainder,

$$\log\frac{1}{1-z} = s_m(z) + r_m(z), \qquad s_m(z) := \sum_{k=1}^{m}\frac{z^k}{k}, \qquad r_m(z) := \sum_{k>m}\frac{z^k}{k}.$$

Consider now smallest cycles. For any *fixed* m, singularity analysis at $z=1$ immediately implies a formula for generalized derangements,

$$P_n^{[>m]} \equiv [z^n]P^{[>m]}(z) = e^{-H_m} + o(1), \tag{6}$$

where $H_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m}$ is the harmonic number and the error term is exponentially small. There is no claim to uniformity, but this argument suggests for m tending to ∞ (at least sufficiently slowly) the approximate formula

$$P_n^{[>m]} \approx \frac{e^{-\gamma}}{m}. \tag{7}$$

Let S_n be length of the smallest cycle in a random permutation of size n. The estimate above suggests that the expectation of S_n satisfies

$$E[S_n] \equiv \sum_{m\ge 1} P_n^{[>m]} = e^{-\gamma}\log n(1+o(1)),$$

where the asymptotic estimate matches what is otherwise known about the distribution of S_n. However, an approximation of the form (7) cannot hold all the way up to $m = n-1$ since

$$P_n^{[>(n-1)]} = \frac{1}{n}, \tag{8}$$

corresponding to cyclic permutations. A natural way to reconcile (7) and (8) is to look for a version that is of the form

$$P_n^{[>m]} \approx \frac{\omega(n/m)}{m}, \tag{9}$$

where one should have $\omega(1) = 1$ and $\omega(+\infty) = e^{-\gamma}$. It turns out that an amended form of (9) does hold true with $\omega(u)$ in (9) being precisely the Buchstab function.

3. The exp-log Class

The main theorem of the talk deals with the general exp-log case. We state it here in the case of a multiplier $a = 1$ where the standard Buchstab function appears. Also, we develop the main ideas in the representative case of the cycle structure of permutations.

Theorem 1. *For a random element of size n in an exp-log class $\mathcal{F}$ of multiplier $a = 1$, the probability that the smallest component S_n is of size greater than m satisfies*

$$\Pr\{S_n > m\} = \frac{1}{m}\omega\left(\frac{n}{m}\right) + O\left(\frac{1}{m^2} + \frac{\log n}{nm}\right),$$

uniformly over the range $\{0, \ldots, , n-1\}$.

The proof starts from Cauchy's coefficient formula

$$P_n^{[>m]} = \frac{1}{2i\pi}\int_C P^{[>m]}(z)\frac{dz}{z^{n+1}}. \tag{10}$$

With the purpose of "capturing the singularity", the integration contour is taken to be a circle of radius close to 1, namely $e^{-1/n}$. Set

$$z = e^{-t/n},$$

where t ranges from $1 - ni\pi$ to $1 + ni\pi$. Then z^{-n} normalizes to an exponential e^t. The form (5) of the gf $P^{[>m]}(z)$ involves $r_m(z)$ that is none other than a Riemann sum relative to the exponential integral,

$$E(v) := \int_v^{+\infty} e^{-w}\frac{dw}{w}.$$

Thus, everything rests on a uniform approximation of the Riemann sum $r_m(z)$ by the exponential integral. This is provided by the following key lemma of [6].

Lemma 1 (Gourdon). *One has uniformly for $\Re(h) > 0$ and $|\Im(h)| \le \pi$,*

$$r_m(e^{-h}) = E(mh) + O\left(\frac{e^{-mh}}{m}\right).$$

(The proof of the lemma is based on the integral formula

$$r_m(e^{-h}) = \frac{1}{m}\int_{mh}^{+\infty} e^{-s}\frac{1}{1 - e^{-s/m}}\,ds,$$

and the decomposition

$$\frac{1}{1-e^{-z}} = \left(\frac{1}{1-e^{-z}} - \frac{1}{z}\right) + \frac{1}{z},$$

where the first term is analytic near $z = 0$.)

Using Lemma 1, one can justify replacing the remainder logarithm in the expression of

$$[z^n](P^{[>m]}(z) - 1)$$

by an exponential integral. In this way, one establishes rigourously the chain of approximations

$$
\begin{aligned}
P_n^{[>m]} &= \frac{1}{2i\pi n}\int_{1-in\pi}^{1+in\pi}(e^{r_m(e^{-t/n})}-1)e^t dt \\
&\sim \frac{1}{2i\pi n}\int_{1-i\infty}^{1+i\infty}(e^{E(\mu t))}-1)e^t dt \\
&\sim \frac{1}{2i\pi m}\int_{1-i\infty}^{1+i\infty}(e^{E(t)}-1)e^{t/\mu} dt,
\end{aligned}
\tag{11}
$$

where $\mu = m/n$. (This is easier said than done!)

Now, the form (11) is an inverse Laplace integral evaluated at $1/\mu$. It can be matched against the Laplace transform of $\omega(u)$, itself directly derived from the defining difference-differential equation. Thus eventually, *the Buchstab law arises from Cauchy's coefficient integral upon using a contour close to the singularity $z = 1$ with a "renormalization" that leads to the appearance of a Laplace transform—the transform of Buchstab's function.*

The technique adapts gracefully to all exp-log structures with multiplier $a = 1$ since these behave analytically very nearly like permutations. For other multipliers $a \neq 1$, a function $\omega_a(u)$ closely related to the Buchstab function must be introduced (work in progress). Finally, like in Gourdon's treatment of largest components, other problems can be dealt with including: (i) local and central limit laws; (ii) distribution estimates for the rth largest component for small fixed r.

4. **Applications**

The analysis sketched here follows closely a preprint by Panario and Richmond [9] and the related works on largest components [6, 7]. It applies to all exp-log classes. In particular, it specializes to polynomials over finite fields and hence has consequences on the analysis of corresponding algorithms. We may cite here:

1. The comparative analysis of several halting rules for the Distinct Degree Factorization phase of univariate polynomial factorization in [3], which requires knowledge of the degrees of the two largest irreducible factors.
2. The analysis of the trial-and-error construction of irreducible polynomials by Ben-Or's algorithms [9], where only partial factorisations are attempted and a candidate polynomial is discarded as soon as its factor of smallest degree has been found.

More generally, the analogy between the prime decomposition of integers and exp-log structures is a striking fact that constitutes a valuable addition to the abstract theory of combinatorial schemas initiated by Soria [11]. (Other general approaches have been recently developed by a variety of authors in a stochastic perspective; see [1, 2, 8].)

Bibliography

[1] Arratia (Richard), Barbour (A. D.), and Tavaré (Simon). – Random combinatorial structures and prime factorizations. *Notices of the American Mathematical Society*, vol. 44, n° 8, 1997, pp. 903–910.

[2] Cameron (Peter J.). – On the probability of connectedness. *Discrete Mathematics*, vol. 167/168, 1997, pp. 175–187.

[3] Flajolet (Philippe), Gourdon (Xavier), and Panario (Daniel). – Random polynomials and polynomial factorization. In Meyer auf der Heide (F.) and Monien (B.) (editors), *Automata, Languages, and Programming*, *Lecture Notes in Computer Science*, pp. 232–243. – 1996. Proceedings of the 23rd ICALP Conference, Paderborn, July 1996. Journal version submitted to*SIAM J. Computing* and available as INRIA Res. Rep. 3370, 1998, 28 pages.

[4] Flajolet (Philippe) and Soria (Michèle). – Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory,* Series A, vol. 53, 1990, pp. 165–182.

[5] Goulden (Ian P.) and Jackson (David M.). – *Combinatorial Enumeration.* – John Wiley, New York, 1983.

[6] Gourdon (Xavier). – *Combinatoire, Algorithmique et Géométrie des Polynômes.* – PhD thesis, École polytechnique, June 1996.

[7] Gourdon (Xavier). – Largest component in random combinatorial structures. *Discrete Mathematics*, vol. 180, n° 1-3, 1998, pp. 185–209.

[8] Hansen (Jennie C.). – A functional central limit theorem for random mappings. *Annals of Probability*, vol. 17, n° 1, 1989.

[9] Panario (Daniel) and Richmond (Bruce). – Analysis of Ben-Or's polynomial irreducibility test. – Preprint, 1997. 16 pages. Submitted to *Random Structures and Algorithms*.

[10] Shepp (L. A.) and Lloyd (S. P.). – Ordered cycle lengths in a random permutation. *Transactions of the American Mathematical Society*, vol. 121, 1966, pp. 340–357.

[11] Soria-Cousineau (Michèle). – *Méthodes d'analyse pour les constructions combinatoires et les algorithmes.* – Doctorat ès Sciences, Université de Paris-Sud, Orsay, July 1990.

[12] Tenenbaum (Gérald). – *Introduction à la théorie analytique des nombres.* – Institut Élie Cartan, Nancy, France, 1990, vol. 13.