GAUSSIAN LIMITING DISTRIBUTIONS FOR THE NUMBER OF COMPONENTS IN COMBINATORIAL STRUCTURES

Philippe Flajolet
INRIA, Rocquencourt
78153–Le Chesnay (France)

Michèle Soria
INRIA and LRI,
Université Paris-Sud 91405–Orsay

Abstract. *Consider the number of cycles in a random permutation or a derangement, the number of components in a random mapping or a random 2-regular graph, the number of irreducible factors in a random polynomial over a finite field, the number of components in a random mapping pattern. These random variables all tend to a limiting Gaussian distribution when the sizes of the random structures tend to infinity.*

Such results, some old and some new, are derived from two general theorems that cover structures decomposed into elementary "components" in either the labelled or the unlabelled case, when the generating function of components has a singularity of a logarithmic type. The proofs are constructed by combining the continuity theorem for characteristic functions with singularity analysis techniques based on Hankel contours.

1. Introduction

Let $\mathcal{C}$ be a class of combinatorial structures, a class $\mathcal{P}$ is said to be *decomposable* over $\mathcal{C}$ if each of its elements may be uniquely decomposed into a multiset of elements of $\mathcal{C}$. We distinguish two cases, depending on whether structures under consideration are labelled or unlabelled.

In the case of *labelled* structures, an element of $\mathcal{P}$ is formed by taking a multiset of (labelled) elements of $\mathcal{C}$ and performing all consistent relabellings. That classical construction ([Goulden, Jackson 1983, Chap.3 Sec.2], [Stanley 1986, Chap.1]) originates from graphical enumerations and it was used systematically by Foata [1974] under the name of 'abelian partitional complex'. The *partitional complex* construction translates into exponential generating functions†,

$$\hat{P}(z) = \sum_{n\geq 0} P_n \frac{z^n}{n!} = \exp(\hat{C}(z))\cdot \tag{1.1}$$

† For a class of structures $\mathcal{S}$, we shall consistently denote by the same letter: the subclass $\mathcal{S}_n$ of $\mathcal{S}$ formed with elements of size n; S_n the cardinality of $\mathcal{S}_n$; $S(z) = \sum_n S_n z^n$, the ordinary generating function of $\mathcal{S}$; $\hat{S}(z) = \sum_{n\geq 0} S_n \frac{z^n}{n!}$, the corresponding exponential generating function.

For *unlabelled* structures, we have the *multiset* construction, where elements of $\mathcal{P}$ are obtained by taking arbitrary sets (with repetition allowed) of elements of $\mathcal{C}$. That construction belongs in Pólya's theory of counting. It translates into the classical relation for ordinary generating functions [Pólya 1937], [Goulden, Jackson 1983, Chap.2 Sec.2]

$$P(z) = \sum_{n\geq 0} P_n z^n = \exp\Big(\frac{C(z)}{1} + \frac{C(z^2)}{2} + \frac{C(z^3)}{3} + \cdots\Big). \tag{1.2a}$$

We shall also occasionally make use of the power–set construction, where no component appears more than once. If $\mathcal{Q}$ is formed from $\mathcal{C}$ in this way, then

$$Q(z) = \sum_{n\geq 0} Q_n z^n = \exp\Big(\frac{C(z)}{1} - \frac{C(z^2)}{2} + \frac{C(z^3)}{3} - \cdots\Big). \tag{1.2b}$$

Our objective here is to study the asymptotic distribution of the number of components in a decomposable (labelled or unlabelled) structure: we show that a limiting Gaussian distribution holds whenever the generating function for $\mathcal{C}$ has an isolated *singularity of a logarithmic type* on its circle of convergence. The Darboux–Pólya method [Henrici 1977, Chap.11 Sec.10] in asymptotic analysis provides, under suitable analytic conditions, an asymptotic expansion for coefficients of a generating function deduced from singularities of the function. Our proofs are based on a variant of that method called 'singularity analysis' and developed in [Flajolet, Odlyzko 1987], together with Lévy's continuity theorem for characteristic functions. From an analytic standpoint, the problem is thus to find asymptotic information on the coefficients of bivariate generating function corresponding to the schemas

$$\hat{P}(z,u) = \exp(u\hat{C}(z)) \qquad \text{and} \qquad P(z,u) = \exp\Big(\sum_{p\geq 1}\frac{u^p}{p}C(z^p)\Big).$$

This paper is a contribution to the study of general statistical properties of combinatorial schemata [Flajolet 1985]. What is needed in that area of growing interest, is general conditions under which Gaussian, Poisson, geometric distributions or others will appear asymptotically in combinatorial structures. Our work is especially close in scope to that of [Compton 1987] who uses real analysis to study distribution results related to the multi–set construction. It adds another class to the list of combinatorial constructions leading to Gaussian distributions for the number of components parameter: In a similar context [Canfield 1977] has shown, under natural conditions, that *entire* exponential generating functions $\hat{C}(z)$ lead to Gaussian distributions for the number–of–components in an Abelian partitional complex. In the same vein, [Bender 1973] has obtained a comparable result for the number–of–components in a sequence construction (in that case, the corresponding schema for the generating functions is $\frac{1}{1-C(z)}$). Analytically, Bender and Canfield's results are obtained using the continuity theorem for characteristic functions in conjunction with different asymptotic techniques (not usable here), namely the saddle point method or the analysis of coefficients of meromorphic functions.

Section 2 deals with labelled structures: we describe analytic conditions under which the parameter number–of–components has mean and variance of asymptotic order $\log n$, and an asymptotically Gaussian distribution (Theorem 1). Various examples of combinatorial objects satisfy these conditions: permutations, permutations without cycle of length one, functional digraphs...

In Section 3 we establish similar conditions and results for the case of unlabelled structures (Theorem 2). Several algebraic structures forming "arithmetical semigroups" satisfy the conditions of Theorem 2. As a consequence, we obtain an Erdös–Kac Theorem for polynomials over finite fields (*cf* [Car 1982] for closely related results): The number of irreducible factors in a random polynomial of large degree over $GF(q)$ tends to a limiting Gaussian distribution.

Section 4 presents some possible extensions of the analytic framework.

We now give a more detailed presentation of our approach.

In both the labelled and the unlabelled case, we let $P_{n,k}$ be the number of $\mathcal{P}$-structures of size n that have k components. Let Ω_n denote the random variable representing the number of components in a random $\mathcal{P}$-structure of size n†; $P_{n,k}/P_n$ is the probability that $\Omega_n = k$. Setting $p_n(u) = \sum_k P_{n,k}u^k$, we see that $p_n(u)/P_n$ is the *probability generating function* of Ω_n, and $p_n(e^{i\theta})/P_n$ is its *characteristic function* (see [Feller 1965] or [Billingsley 1986]).

The mean value μ_n and the variance σ_n^2 of Ω_n can be easily computed from the probability generating function:

$$\mu_n = \frac{p_n'(1)}{p_n(1)}; \qquad \sigma_n^2 = \frac{p_n''(1)}{p_n(1)} - \frac{p_n'^2(1)}{p_n^2(1)} + \frac{p_n'(1)}{p_n(1)}.$$

Using singularity analysis, we show that the mean value of Ω_n and its variance are both asymptotically of order $\log n$ when the generating function for $\mathcal{C}$ has an isolated singularity of logarithmic type on its circle of convergence.

We shall show under the same conditions that, as n tends to infinity, the normalized random variable obtained from Ω_n *converges weakly* to a Gaussian variable; in other words, for any two real constants $a < b$, we have

$$\Pr\left(a < \frac{\Omega_n - \mu_n}{\sigma_n} < b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt\,. \tag{1.3}$$

Levy's continuity theorem asserts that *the pointwise convergence of characteristic functions entails the convergence of distribution functions* (in the sense of (1.3)). The method of proof is thus to establish the pointwise convergence

$$\phi_{Y_n}(\theta) \to e^{-\theta^2/2},$$

where Y_n denotes the normalized variable $Y_n = (\Omega_n - \mu_n)/\sigma_n$ with characteristic function $\phi_{Y_n}(\theta)$, and $e^{-\theta^2/2}$ is the characteristic function of a normal variable of mean value 0 and variance 1.

---

† By random we mean that each $\mathcal{P}$-structure of size n is equally likely.

Therefore the main technical problem lies in the estimation of ϕ_{Y_n}.

$$\begin{aligned}\phi_{Y_n}(\theta) &= E(e^{iY_n\theta}) \\ &= e^{-i\theta\mu_n/\sigma_n}\frac{p_n(e^{i\theta/\sigma_n})}{P_n}\end{aligned} \tag{1.4}$$

The value of $p_n(e^{i\theta/\sigma_n})$ is itself computed by means of the Cauchy coefficient formula. In the case of unlabelled structures (ordinary generating functions), we have

$$p_n(u) = \frac{1}{2i\pi}\oint P(z,u)\,\frac{dz}{z^{n+1}}\ , \tag{1.5a}$$

the other case only involving an extra factor of $n!$

$$p_n(u) = \frac{n!}{2i\pi}\oint \hat{P}(z,u)\,\frac{dz}{z^{n+1}}\ \cdot \tag{1.5b}$$

To evaluate the integral, we use an integration contour of the Hankel type that comes close to the dominant singularity of the integrand. Such contours are of general use in the study of the Gamma and Zeta functions, as well as in inversion problems for integral transforms with algebraic or logarithmic singularities. Our presentation is based on the treatment of [Flajolet, Odlyzko 1982] and the more general presentation appearing in [Flajolet, Odlyzko 1987], where it is shown that Hankel contours are well-suited to analyzing large classes of generating functions with algebraic and logarithmic singularities. Related analytic methods appear in works of [Car 1984] and [Wong, Wyman 1974].

2. Labelled structures

We let $\hat{P}(z,u)$ be the *bivariate exponential generating function* of decomposable structures $\mathcal{P}$, with variable u marking the number of components in a structure, so that

$$\hat{P}(z,u) = \sum_{n,k\geq 0} P_{n,k}u^k\frac{z^n}{n!} = \sum_{n\geq 0} p_n(u)\frac{z^n}{n!}\ \cdot$$

Observe that $\hat{P}(z,1) = \hat{P}(z)$.
For labelled structures, equation (1.1) extends to the bivariate relation:

$$\hat{P}(z,u) = \exp(u\hat{C}(z))\ \cdot \tag{2.1}$$

We now study analytic conditions on $\hat{C}(z)$ under which the number of components is asymptotically normal.

We let $\Delta(\rho,\eta)$, with $\rho > 0$ and $\eta > 0$, denote the closed domain

$$\Delta(\rho,\eta) = \{\ z \mid |z| \leq \rho+\eta\ \}$$

and Δ_0 is the open region obtained by slitting Δ along $[\rho, \rho+\eta]$:

$$\Delta_0(\rho,\eta) = \left\{ z \;\middle|\; |z| \le \rho+\eta,\ z \notin [\rho, \rho+\eta] \right\} \cdot$$

Definition. *Let $G(z)$ be a generating function which is analytic at 0 and has a unique dominant singularity ρ on its circle of convergence. We say that $G(z)$ is a logarithmic function with multiplier a and constant K if near this singularity*

$$G(z) = a \log \frac{1}{1-z/\rho} + R(z) \tag{2.2}$$

where a is a positive real number and $R(z)$ is analytic in Δ_0 and satisfies $R(z) = K + o(1)$ when z tends to ρ in Δ_0.

We shall now study the number–of–components parameter in a partitional complex of $\mathcal{C}$-structures when the exponential generating function of class $\mathcal{C}$ is a logarithmic function. smallskipWe first proceed to establish the mean and variance of the random variable Ω_n.

Proposition 1. *Let $\hat{C}(z)$ be a logarithmic function with multiplier a and constant K. The number of structures of size n in $\mathcal{P}$, with exponential generating function $\hat{P}(z) = \exp(\hat{C}(z))$, is estimated asymptotically by*

$$P_n \sim n!\, \frac{\rho^{-n} n^{a-1} e^K}{\Gamma(a)} \cdot \tag{2.3a}$$

Furthermore, the mean μ_n and variance σ_n^2 of Ω_n satisfy, as $n \to \infty$

$$\mu_n = a \log n + O(1) \qquad \text{and} \qquad \sigma_n^2 = a \log n + O(1). \tag{2.3b}$$

Proof. As stated in the introduction, all we need is to determine asymptotically $p_n'(1)$ and $p_n''(1)$, and we achieve this by finding the asymptotic expansion of their generating functions around $z = \rho$. By logarithmic derivatives from (2.1), and instantiating at $u = 1$, we get, with $f_u'(z,u)$ denoting $\partial f(z,u)/\partial u$,

$$\frac{\hat{P}_u'(z,1)}{\hat{P}(z,1)} = a \log \frac{1}{1-z/\rho} + R(z), \tag{2.4}$$

and through another differentiation,

$$\frac{\hat{P}_{u^2}''(z,1)}{\hat{P}(z,1)} = \bigl(a \log \frac{1}{1-z/\rho} + R(z)\bigr)^2. \tag{2.5}$$

Thus, taking coefficients†, and using the extension of Darboux's theorem from [Flajolet, Odlyzko 1987], we find

$$[z^n]\hat{P}(z) = \frac{\rho^{-n} n^{a-1} e^K}{\Gamma(a)} (1 + o(1)) \tag{2.6}$$

† The notation $[z^n]f(z)$ represents the coefficient of z^n in the Taylor expansion of $f(z)$, and $[u^k z^n]f(z,u) = [u^k]\bigl([z^n]f(z)\bigr)$

and

$$\mu_n = \frac{[z^n]\hat{P}'_u(z,1)}{[z^n]\hat{P}(z,1)} = a\log n + O(1). \tag{2.7}$$

The computation leading to the variance estimate is similar, though a little more complex since it is necessary to use second order terms in the asymptotic expansion of the coefficients of functions

$$\frac{1}{(1-z/\rho)^a}\log^b\big(\frac{1}{1-z/\rho}\big). \qquad \blacksquare$$

THEOREM 1. *Let $\mathcal{P}$ and $\mathcal{C}$ be two classes of combinatorial structures, such that $\mathcal{P}$ is the partitional complex of $\mathcal{C}$:*

$$\hat{P}(z,u) = \exp(u\hat{C}(z)).$$

Let Ω_n be the number of components in a random $\mathcal{P}$-structure of size n, with probability distribution

$$\Pr(\Omega_n = k) = \frac{P_{n,k}}{\sum_k P_{n,k}} \qquad \textit{with} \qquad P_{n,k} = n![u^k z^n]\exp(u\hat{C}(z)).$$

If $\hat{C}(z)$ is a logarithmic function, then Ω_n, once normalized, converges weakly to a limiting Gaussian distribution:

$$\Pr\Big(a < \frac{\Omega_n - \mu_n}{\sigma_n} < b\Big) \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}\,dt.$$

PROOF. The proof relies on the estimation of a Cauchy integral. Our starting point for asymptotic analysis is the integral formula stemming from equations (1.4) and (1.5b) in conjunction with Proposition 1

$$I = P_n\phi_n(\theta) = \frac{n!}{2i\pi}\int_\Gamma e^{-i\theta\sqrt{a\log n}}\hat{P}(z, e^{i\theta/\sqrt{a\log n}})\,\frac{dz}{z^{n+1}}. \tag{2.8}$$

The contour Γ is oriented positively and is the union of five contours $\{\Gamma_j\}_{1\le j\le 5}$ depicted on Figure 1. With $\lambda(n) = \log^2 n$, so that $1 \ll \lambda(n) \ll \sqrt{n}$, we take

$$\begin{aligned}
\Gamma_0 &= \Big\{\, z \;\;/\;\; |z-\rho| = \frac{\rho}{n},\ \Re(z) < \rho \,\Big\}\\
\Gamma_1 &= \Big\{\, z \;\;/\;\; \rho \le \Re(z) \le \rho + \frac{\lambda(n)}{n},\ \Im(z) = \frac{\rho}{n} \,\Big\}\\
\Gamma_2 &= \Big\{\, z \;\;/\;\; \rho + \frac{\lambda(n)}{n} < \Re(z),\ |z| \le \rho+\eta,\ \Im(z) = \frac{\rho}{n} \,\Big\}\\
\Gamma_3 &= \Big\{\, z \;\;/\;\; |z| = \rho+\eta,\ |\Im(z)| > \frac{\rho}{n} \,\Big\}.
\end{aligned} \tag{2.9}$$

FIGURE 1. The Hankel contour used in the proof of Theorem 1.

Contours Γ_4, Γ_5 are symmetrical to Γ_2, Γ_1 with respect to the real axis. The various contributions of the contour to integral (2.8) will be evaluated starting from the outside.

Γ_3: When $n \to \infty$, $u = \exp(i\theta/\sqrt{a \log n})$ stays in a fixed neighbourhood of 1. On Γ_3, $|1 - z/\rho|$ is lower–bounded by an absolute constant, so that, since $R(z)$ is analytic in Δ_0, $|\hat{P}(z, u)| = O(1)$. We thus get the bound

$$\int_{\Gamma_3} |\hat{P}(z, u)| \frac{|dz|}{|z|^{n+1}} = O((\rho + \eta)^{-n}), \tag{2.10}$$

a quantity exponentially smaller than ρ^{-n}.

$\Gamma_2 \cup \Gamma_4$: We still have on that part of the contour $|uR(z)| = O(1)$. On $\Gamma_2 \cup \Gamma_4$, z is still far from ρ and $|\log(1 - z/\rho)^{-1}| = O(\log n)$ so that $|\hat{P}(z, u)| = O(n^{c_1})$ for some constant c_1. On the other hand, z is far enough from ρ, so that $\rho^n |z^{-n}| < \exp(-c_2 \lambda(n))$. The contribution arising from $\Gamma_2 \cup \Gamma_4$ is again exponentially small.

$\Gamma_0 \cup \Gamma_1 \cup \Gamma_5$: On that part of the contour, z is close to the singularity ρ and we use local expansions. We set

$$\rho^{-1} z = 1 + \frac{t}{n}$$

so that $\Re(t) \leq \lambda(n)$. All the implied constants in the $O(.)$ error terms that follow are uniform in t and u and thus represent absolute constants. We have

$$\begin{cases} R(z) = K + o(1) \\ u = 1 + \dfrac{i\theta}{\sqrt{a \log n}} - \dfrac{\theta^2}{2a \log n} + O(\dfrac{1}{\log^{3/2} n}) \\ \log \dfrac{1}{1 - z/\rho} = \log \dfrac{n}{-t} \\ z^{-n} = \rho^{-n}(1 + \dfrac{t}{n})^{-n} = \rho^{-n} e^{-t + O(t^2/n)} \end{cases} \tag{2.11}$$

Plugging these expansions into integral I, we get

$$I = e^{-\theta^2/2} e^K \rho^{-n} n^{a-1} \cdot \frac{n!}{2i\pi} \int_\gamma e^{-t} \big(1 + o(1)\big) \frac{dt}{(-t)^a}, \tag{2.12}$$

where the integration contour γ is the image of $\Gamma_0 \cup \Gamma_1 \cup \Gamma_5$ under the mapping $z \mapsto \rho(1 + t/n)$. The integral in (2.12) expands as

$$\int_\gamma e^{-t} \frac{dt}{(-t)^a} \; + o(1)\cdot \tag{2.13}$$

This last integral can be extended to a contour γ^* going all the way to $+\infty$ introducing only exponentially small error terms. In this way, we get a classical Hankel integral [Whittaker, Watson 1935, p.244]

$$\frac{1}{2i\pi} \int_{\gamma^*} e^{-t} \frac{dt}{(-t)^a} = \frac{1}{\Gamma(a)}\cdot \tag{2.14}$$

Thus, since $\phi_n(\theta) = I/P_n$, we have by (2.3a), (2.8) and (2.12)

$$\phi_n(\theta) \to e^{-\theta^2/2}. \tag{2.15}$$

The proof of Theorem 1 is now complete. ∎

We now indicate several classes of combinatorial structures that satisfy conditions of Theorem 1.

Example 1. *Cycles in permutations.* The bivariate generating function for *permutations* with u marking the number of cycles is well known to be

$$\hat{P}(z,u) = \sum S_{n,k} u^k \frac{z^n}{n!} = \exp(u \log \frac{1}{1-z})\cdot$$

The function inside the exponential is the simplest case of a logarithmic function; thus the *number of cycles in a random permutation of size n converges to a Gaussian limiting distribution.* This classical result stating the asymptotically normal distribution of the Stirling numbers (of the first kind) constitutes Goncharov's Theorem. It can also be obtained via a direct application of the Central Limit Theorem. However, Theorem 1 can be applied to various families of restricted permutations. For instance, permutations without cycles of length one, called *derangements* [Comtet 1974, p.182], have the following generating function :

$$\hat{D}(z,u) = \frac{e^{-zu}}{(1-z)^u} = \exp\big(u(\log \frac{1}{1-z} - z)\big),$$

another clear case of application of Theorem 1. More generally, *the number of cycles is asymptotically normal in generalized derangements where a finite set of cycle lengths are forbidden.*

We can also deal with the cycle decomposition of permutations *with non isomorphic cycles* (no two cycles are order–isomorphic) which have exponential generating function

$$\hat{H}(z,u) = \exp\big(u(\log\frac{1}{1-z} - R(z,u))\big),$$

where

$$R(z,u) = \sum_{k,n} \frac{(-1)^k u^k}{kn} \frac{z^{kn}}{(n!)^{k-1}}$$

can be treated by similar methods as explained in the next section. ∎

EXAMPLE 2. *Children's yards* of [Stanley 1978] are combinatorial configurations representing sets made of rounds (directed cycles) with one child (node) in the center of each round. The exponential generating function for this class is $(1-z)^{-z}$ with corresponding bivariate generating function

$$\exp\big(u(\log\frac{1}{1-z} + (z-1)\log\frac{1}{1-z})\big).$$

Since $(z-1)\log\frac{1}{1-z} \to 0$ when z tends to 1, we obtain: *The number of rounds in a children's yard has asymptotic mean and variance* $\log n$, *and a Gaussian limit distribution.* ∎

EXAMPLE 3. *Clouds and 2-regular graphs.* "Clouds" are defined in [Comtet 1974 p.274]: let n straight lines in the plane be given, with $\binom{n}{2}$ intersecting points; a cloud of size n is a (maximal) set of n intersecting points no three of which are collinear. There is a one–to–one correspondence between clouds and a class of undirected graphs called *2–regular graphs*: A 2–regular graph of size n is a subgraph with n lines of the complete graph on n points, such that each vertex has degree 2. Any 2–regular graph may be decomposed into a product of connected components that are (undirected) cycles of length at least 3. Hence the bivariate exponential generating function for 2–regular graphs, or clouds is:

$$\frac{e^{-uz/2-uz^2/4}}{(1-z)^{u/2}} = \exp\big(u(\frac{1}{2}\log\frac{1}{1-z} - \frac{z}{2} - \frac{z^2}{4})\big).$$

The function $R(z) = z/2 + z^2/4$ is entire, so that conditions of Theorem 1 are satisfied, and *the number of connected components in a 2–regular graph, or equivalently the number of polygones in a cloud, has a Gaussian limiting distribution.* ∎

EXAMPLE 4. *Random mappings.* Let f denote a function that maps the set $N = \{1, 2, \cdots, n\}$ into itself. Function f may be represented by a directed graph G_f with point–set N and arc–set $\{(i, f(i)); i \in N\}$. Such graphs, in which every point has out–degree one, are called *functional digraphs* [Harary and Palmer 1973 p.68]. A functional digraph may be viewed as a partitional complex of components that are themselves cycles of rooted labelled trees. The bivariate generating function for functional digraphs is

$$\hat{F}(z,u) = \exp\big(u(\log\frac{1}{1-a(z)})\big),$$

where the generating function (of rooted labelled trees) $a(z)$ is defined implicitly by the relation $a(z) = z\exp(a(z))$. By the inversion theorem for implicit functions (see [Meir, Moon 1978] for similar applications), we get

$$a(z) = 1 - \sqrt{2(1-ez)} + \sum_{k\geq 2} c_k (1-ez)^{k/2}.$$

Thus,

$$\hat{F}(z,u) = \exp\left\{u\left(\frac{1}{2}\log\frac{1}{1-ez} + H((1-ez)^{1/2})\right)\right\},$$

where $H(v)$ is analytic at $v = 0$, with $H(0) = 0$. From this form and Theorem 1, we obtain Stepanov's theorem [1969]: *the number of components in functional digraphs has a limiting Gaussian distribution.*

That approach could be extended to functional digraphs satisfying various degree constraints as considered by Bender et al. Such analyses are of (some) relevance to integer factorization, using Pollard's "rho" method [Knuth 1981]. ∎

3. Unlabelled structures

In the case of unlabelled structures, we have the classical relation

$$\sum_{n\geq 0} P_n z^n = \prod_{n\geq 1} (1-z^n)^{-C_n}. \tag{3.1a}$$

By taking logarithms in (3.1a) and reorganising the corresponding series, we get the alternative form

$$P(z) = \exp\left(\frac{C(z)}{1} + \frac{C(z^2)}{2} + \frac{C(z^3)}{3} + \cdots\right). \tag{3.1b}$$

An easy combinatorial argument extending the preceding equation shows that

$$P(z,u) = \sum_{n,k\geq 0} P_{n,k} u^k z^n = \exp\left(\frac{u}{1}C(z) + \frac{u^2}{2}C(z^2) + \frac{u^3}{3}C(z^3) + \cdots\right) \tag{3.2}$$

As in the preceding section, we are interested in structures such that $C(z)$ is a generating function satisfying the logarithmic condition of Definition 1:

$$C(z) = a\log\frac{1}{1-z/\rho} + R(z), \tag{3.3}$$

where a is a positive real number and $R(z)$ is an analytic function in Δ_0 such that $R(z) = K + o(1)$ when z tends to ρ in Δ_0.

Furthermore, in many cases of application, ρ is strictly less than 1. In that case, each $C(z^k)$, for $k \geq 2$, is analytic at $z = \rho$; Moreover, as $|C(z^k)| < c^{st}|z|^k$ for $k \geq 2$ and $|z| < \rho^{1/2}$, the sum $\sum_{k\geq 2} C(z^k)$ is uniformly convergent for $|z| = \rho$. Hence

$$P(z,u) = \exp\big(uC(z) + S(z,u)\big), \tag{3.4}$$

where $S(z,u) = \sum_{k\geq 2} \frac{u^k}{k} C(z^k)$ is analytic for $|z| < \rho + \epsilon$ and $|u| < 1 + \delta$, for some $\epsilon > 0$ and $\delta > 0$. Under these conditions we can establish results similar to those of Section 2.

PROPOSITION 2. *If $C(z)$ is a logarithmic function with radius of convergence $\rho < 1$, then the mean μ_n and variance σ_n^2 of Ω_n satisfy, as $n \to \infty$*

$$\mu_n = a \log n + O(1) \qquad \text{and} \qquad \sigma_n^2 = a \log n + O(1).$$

PROOF. Proceeding as in Proposition 1 we find:

$$\frac{P'_u(z,1)}{P(z,1)} = a \log \frac{1}{1 - z/\rho} + R(z) + \sum_{k\geq 2} C(z^k) \tag{3.5}$$

and

$$\frac{P''_{u^2}(z,1)}{P(z,1)} = \big(a \log \frac{1}{1 - z/\rho} + R(z)\big)^2 + \sum_{k\geq 2} (k-1) C(z^k), \tag{3.6}$$

and we can conclude by taking coefficients (the two sums in the preceding equations are analytic at ρ). ∎

THEOREM 2. *Let $\mathcal{P}$ and $\mathcal{C}$ be two classes of combinatorial structures, such that $\mathcal{P}$ is the multiset construction over $\mathcal{C}$:*

$$P(z) = \exp\big(\sum_{p\geq 1} \frac{u^p}{p} C(z^p)\big).$$

Let Ω_n be the number of irreducible components in a random $\mathcal{P}$-structure of size n. If $C(z)$ is a logarithmic function with radius of convergence $\rho < 1$, then Ω_n, once normalized, converges weakly to a limiting Gaussian distribution:

$$\Pr\big(a < \frac{\Omega_n - \mu_n}{\sigma_n} < b\big) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\, dt.$$

PROOF. By equations (3.3) and (3.4), we have

$$P(z,u) = \exp\big(u \log \frac{1}{1 - z/\rho} + uR(z) + S(z,u)\big), \tag{3.7}$$

where $S(z,u)$ is analytic for $|z| < \rho + \epsilon$ and $|u| < 1 + \delta$, for some $\epsilon > 0$ and $\delta > 0$. Thus we can use a proof entirely similar to the one of Theorem 1: the integral stemming from (1.4) and (1.5) is evaluated on the same contour as integral (2.8). The contribution arising from Γ_2, Γ_3, Γ_4 is still exponentially small; On the other parts of the contour, where z is close to the singularity ρ, we use the local expansions of (2.11) together with the expansion of the analytic function $S(z,u)$:

$$S(z,u) = \sum_{k\geq 2} \frac{1}{k} C(\rho^k) + O(|z-\rho|) + O(|u-1|) = L + O(\frac{\lambda(n)}{n} + \frac{1}{\sqrt{\log n}}). \quad \blacksquare$$

Remark. There is an obvious analogue of Theorem 2 for the set–of construction (see equation (1.2b)), as well as several other constructions belonging in Pólya's theory of counting [Pólya 1937].

Example 1. *Random mapping patterns.* Let f and g be two functions mapping the set $\{1,2,\cdots,n\}$ into itself. Mappings f and g are said to be equivalent if there exists a permutation π of $\{1,2,\cdots,n\}$ such that $f(i)=j$ iff $g(\pi(i)) = \pi(j)$. Random mapping patterns are equivalence classes of mapping functions. From example 4 of Section 2, it is clear that random mapping patterns are functional digraphs on unlabelled points; they correspond to multisets of cycles of rooted unlabelled trees. The ordinary generating function for rooted unlabelled trees satisfies the implicit relation $A(z) = z\exp(\sum \frac{1}{k}A(z^k))$, and [Otter 1947] proved that

$$A(z) = 1 - c_1\sqrt{(1-z/\eta)} + \sum_{k\geq 2} c_k(1-z/\eta)^k, \tag{3.8}$$

for some $\eta < 1$.
On the other hand, it is known [Read 1961, De Bruijn and Klarner 1982] that a class of combinatorial structures $\mathcal{C}$ made of *cycles* of unlabelled structures in class $\mathcal{A}$ has generating function

$$C(z) = \sum_{k\geq 1} \frac{\phi(k)}{k} \log \frac{1}{1-A(z^k)}, \tag{3.9}$$

where $\phi(k)$ is the Euler totient function. In the present context, since $A(z)$ has radius of convergence η strictly less than 1,

$$C(z) = \log \frac{1}{1-A(z)} + S(z), \tag{3.10}$$

where $S(z)$ is analytic at η. Finally the bivariate generating function for random mapping patterns verifies

$$\begin{aligned} M(z,u) &= \exp\big(\sum_{k\geq 1} u^k \frac{C(z^k)}{k}\big) \\ &= \exp\big(u \log \frac{1}{1-A(z)} + uS(z) + T(z,u)\big) \\ &= \exp\Big(\frac{u}{2}\log\frac{1}{1-z/\eta} + uH\big((1-z/\eta)^{1/2}\big) + uS(z) + T(z,u)\Big). \end{aligned} \tag{3.11}$$

where $S(z)$ is analytic at η, $T(z,u)$ is analytic for $z=\eta$ and $u=1$, and H is analytic around 0, with $H(0)=0$. Thus conditions for applying Theorem 2 are satisfied: *the number of components in random mapping patterns has a Gaussian limiting distribution.* The mean value is equal to $\frac{1}{2}\log n$ (this result appears in [Meir and Moon 1984]), and the variance is also $\frac{1}{2}\log n$.

EXAMPLE 2. *Polynomial factorization.* In many applications related to Pólya's construction, the function $R(z)$ appearing in expression (3.3) of $C(z)$ satisfies a stronger condition, namely that $R(z)$ is analytic in a disk with radius strictly greater than ρ. In this case, it can be shown using Moebius Inversion that a logarithmic singularity for $C(z)$ is equivalent to a polar singularity at ρ for $P(z)$. As an illustration, we shall study the factorization of polynomials over a finite field.

Fix a finite field $K=GF(q)$ and consider the class $\mathcal{P}$ of monic polynomials (having leading coefficient 1) in $K[z]$, with $\mathcal{I}$ the subclass of irreducible polynomials. Let $\mathcal{P}_n$ and $\mathcal{I}_n$ represent the sets formed with polynomials of degree n. Obviously, $P_n=q^n$, so that

$$P(z)=(1-qz)^{-1}. \tag{3.12}$$

Because of the unique factorization property, a polynomial is a multiset of irreducible polynomial, whence the relation

$$P(z)=\exp\left(\frac{I(z)}{1}+\frac{I(z^2)}{2}+\frac{I(z^3)}{3}+\cdots\right) \tag{3.13}$$

The preceding relation can be inverted using Moebius inversion. If we set $L(z)=\log P(z)$, then we have

$$I(z)=\sum_{k\geq 1}\mu(k)\frac{L(z^k)}{k}=\log\frac{1}{1-qz}+\sum_{k\geq 2}\mu(k)\frac{L(z^k)}{k}, \tag{3.14}$$

where μ is the Moebius function.

Since $L(z^k)$ is analytic for $|z|<q^{-1/2}$ whenever $k\geq 2$, and $|L(z^k)|<c^{st}|z|^k$, the sum $\sum_{k\geq 2}\mu(k)L(z^k)/k$ is analytic for $|z|\leq\tau$, with $q^{-1}<\tau<q^{-1/2}$. Hence $I(z)$ has an isolated singularity of logarithmic type at $z=q^{-1}<1$.

Thus the average number of irreducible factors in a polynomial, and its variance, are equal to $\log n+O(1)$ (this result appears in [Knuth 1981, Ex.4.6.2.5]). Moreover we can state:

COROLLARY 1. *Let Ω_n be the random variable representing the number of irreducible factors of a random polynomial of degree n over $GF(q)$, each factor being counted with its order of multiplicity. Then as n tends to infinity, for any two real constants $\lambda<\mu$, we have*

$$\Pr\{\log n+\lambda\sqrt{\log n}<\Omega_n<\log n+\mu\sqrt{\log n}\}\to\frac{1}{\sqrt{2\pi}}\int_\lambda^\mu e^{-t^2/2}\,dt. \tag{3.15}$$

This statement is an analogue of the Erdös–Kac Theorem [1940] for the number of prime divisors of natural numbers (with $\log n$ here replacing $\log\log n$).
This result, as well as several other results concerning limiting distribution of factors in polynomials can be deduced from Car's work on factorization of polynomials [1982,1984]: She shows a Poisson limiting distribution, which entails Corollary 1.

As stated in the proof of Theorem 2, the result of a limiting Gaussian distribution holds true whenever $P(z,u)$ can be put in the form of equation(3.4). Consider the bivariate generating function $P(z,u)$ with u "marking" the number of *different* irreducible factors in a polynomial. We have (compare with (3.1a))

$$\begin{aligned} P(z,u) &= \sum_{n,k\geq 0} P_{n,k} u^k z^n = \prod_{n\geq 1} (1 + uz^n + uz^{2n} + uz^{3n} + \cdots)^{I_n} \\ &= \exp\Big(\sum_{n\geq 1} I_n \log\big(1 + \frac{uz^n}{1-z^n}\big)\Big) \\ &= \exp\Big(uI(z) + \sum_{k\geq 2} \frac{1+(-1)^{k-1}(u-1)^k}{k} I(z^k)\Big). \end{aligned} \tag{3.16}$$

So that we can state the following result: *The number of* distinct *irreducible factors of a random polynomial of degree n over $GF(q)$ tends to a Gaussian limiting distribution, in the sense of Corollary 1.*

Finally, let us mention that, according to (1.2b), the same result holds true for *square–free polynomials* (products of *different* irreducible factors). ∎

EXAMPLE 3. *Arithmetical semigroups.* Knopfmächer [1979] defines an arithmetical semigroup as a semigroup with unique factorization, and a notion of size (or degree) so that

$$|xy| = |x| + |y|$$

and the number of elements of a fixed size is finite. If $\mathcal{P}$ is an arithmetical semigroup and $\mathcal{I}$ its set of 'primes' (irreducibles elements), axiom $A^{\#}$ of Knopfmächer asserts the condition

$$\mathrm{card}\{x \in \mathcal{P} \ / \ |x| = n\} = cq^n + O(q^{\alpha n}) \quad (\alpha < 1).$$

It is shown by Knopfmächer that several algebraic structures forming arithmetical semigroups satisfy axiom $A^{\#}$, and thus the conditions of Theorem 2. Therefore, that theorem fits into the framework of Knopfmacher's "abstract analytic number theory", since it reveals general conditions under which theorems of the Erdös–Kac type will hold true. Let us mention as examples of applications: Galois polynomial rings of the preceding example, finite modules or semisimple finite algebras over a finite field $K = GF(q)$, integral divisors in algebraic function fields, ideals in the principal order of a algebraic function field, finite modules, or semisimple finite algebras over a ring of integral functions.

4. Extensions

The analytic framework under which our results still hold can be extended in various ways.

a) It is possible to weaken the conditions on the logarithmic function:

- It can have several singularities on its circle of convergence: for example, there is an asymptotic Gaussian distribution for the number of cycles in permutations containing only cycles of even length.
- $R(z)$ need be analytic only in an indented disk Δ_ϕ, with $\phi < \pi/2$:

$$\Delta_\phi = \left\{ z \mid |z| < \rho, \operatorname{Arg}(z) > \phi \right\}.$$

Though we do not have a natural example, similar situations do arise in combinatorial enumerations, (see [Odlyzko 1982] for an example of a generating function with a fractal boundary).

b) Let $f(z)$ be an analytic function at 0, with either a dominant singularity greater than ρ, or with an algebraico–logarithmic singularity at ρ. Then our theorems can be extended to the functions

$$f(z)\exp(uC(z)) \qquad \text{and} \qquad f(z)\exp(\sum u^k C(z^k)/k).$$

For example, the number of cycles with odd (resp. even) length, and the number of cycles with length greater than a fixed value s, in a random permutation, all have Gaussian limiting distributions.

References.

E. Bender [1973]. "Central and Local Limit Theorems Applied to Asymptotic Enumeration", *J. Combinatorial Theory Series A* **15**, 1973, 91-111.

P. Billingsley [1986]. *Probability and Measure*, Academic Press, 1986.

E. R. Canfield [1977]. "Central and Local Limit Theorems for Coefficients of Binomial Type", *J. Combinatorial Theory Series A* **23**, 1977, 275-290.

M. Car [1982]. "Factorization dans $\mathbf{F}_q[X]$", *C. R. Acad. Sciences Paris Série I* **294**, 1982, 147-150.

M. Car [1984]. "Ensembles de polynômes irréductibles et théorèmes de densité", *Acta Arithmetica* **44**, 1984, 323-342.

K. J. Compton [1987]. "Some methods for computing component distribution probabilities in relational structures", *Discrete Mathematics* **66**, 1987, 59-77.

L. Comtet [1974]. *Advanced Combinatorics.* Reidel, Dordrecht, 1974.

N. G. De Bruijn and D. A. Klarner [1982]. "Multisets of aperiodic cycles", *Siam J. Alg. Disc. Meth.* **3**, 1982, 359-368.

P. Erdös and M. Kac [1940]. "The gaussian law of errors in the theory of additive number–theoretic functions", Amer. J. Math. **62**, 1940, 738–742.

W. Feller [1965]. *An Introduction to Probability Theory and Its Applications*, 2 Volumes, Wiley, New York, 1965.

P. Flajolet [1985]. “Elements of a general theory of combinatorial structures”, in *Proc. FCT Conf., Lecture Notes in Comp. Sc*, Springer Verlag, 1985, 112-127.

P. Flajolet and A. M. Odlyzko [1982]. “The Average Height of Binary Trees and Other Simple Trees”, *J. Computer and System Sciences* **25**, 1982, 171-213.

P. Flajolet and A. M. Odlyzko [1987]. “Singularity Analysis of Generating Functions”, preprint, 1987.

D. Foata [1974]. *La série génératrice exponentielle dans les problèmes d'énumérations.* Les presses de l'Université de Montréal, 1974.

I. Goulden and D. Jackson [1983]. *Combinatorial Enumerations.* Wiley, New York, 1983.

F. Harary and E. Palmer [1973]. *Graphical Enumerations*, Academic Press, New–York, 1973.

P. Henrici [1977]. *Applied and Computational Complex Analysis.* Three Volumes. Wiley, New York, 1977.

J. Knopfmacher [1979]. *Analytic Arithmetic of Algebraic Function Fields*, Lect. Notes in Pure and Applied Maths.,M. Dekker, Inc., New–York, 1979.

D. E. Knuth [1981]. *The Art of Computer Programming.* Volume 2: *Semi-Numerical Algorithms.* Addison-Wesley, Reading, MA, second edition 1981.

A. Meir and J.W. Moon [1978]. “On the altitude of nodes in random trees”, *Canadian Journal of Mathematics* **30**, 1978, 997-1015.

A. Meir and J.W. Moon [1984]. “On random mapping patterns”, *Combinatorica* **4**, 1984, 61-70.

A. M. Odlyzko [1982]. “Periodic Oscillations of Coefficients of Power Series that Satisfy Functional Equations”, *Advances in Math.* **44**, 1982, 180-205.

R. Otter [1948]. “The number of trees”, *Annals of Mathematics* **49**, 1948, 583-599.

G. Pólya [1937]. “Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen”, *Acta Mathematica* **68**, 1937, 145–254. Translated in: G. Pólya and R. C. Read, *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds*, Springer, New-York, 1987.

R. C. Read [1961]. “A note on the number of functional digraphs”, *Math. Ann.* **143**, 1961, 109-110.

R. P. Stanley [1978]. “Generating Functions,” in *Studies in Combinatorics*, edited by G-C. Rota, M. A. A. Monographs, 1978.

R. P. Stanley [1986]. *Enumerative Combinatorics*, Wadsworth and Brooks/Cole, Monterey, 1986.

V. E. Stepanov [1969]. "Limit Distributions of Certain Characteristics of Random Mappings," *Theory of Prob. and Appl.* **14**, 1969, 612-626.

E. T. Whittaker, G. N. Watson [1935]. *A Course of Modern Analysis.* Cambridge University Press, 1935.

R. Wong, M. Wyman [1974]. "The Method of Darboux," *Journal of Approximation Theory* **10**, 1974, 159-171.