Discrete Mathematics 114 (1993) 159–180
North-Holland

General combinatorial schemas: Gaussian limit distributions and exponential tails*

Philippe Flajolet
INRIA, Rocquencourt, F-78153 Le Chesnay, France

Michèle Soria
LRI, Université Paris-Sud, F-91405 Orsay, France

Received 12 February 1989
Revised 12 February 1991

Abstract

Flajolet, P. and M. Soria, General combinatorial schemas: Gaussian limit distributions and exponential tails, Discrete Mathematics 114 (1993) 159–180.

Under general conditions, the number of components in combinatorial structures defined as sequences, cycles or sets of components admits a Gaussian limit distribution together with an exponential tail. The results are valid, assuming simple analytic conditions on the generating functions of the components. The proofs rely on the continuity theorem for characteristic functions and Laplace transforms as well as techniques of singularity analysis applied to algebraic and logarithmic singularities. Combinatorial applications are in the fields of graphs, permutations, random mappings, ordered partitions and polynomial factorizations.

1. Introduction

Vassiliĭ Leonidovich Goncharov established in 1944 that the number of cycles in a random permutation of large size approaches a normal distribution; see Knuth's account in [19, p. 103]. Many results of a similar type are now known for a great variety of classical combinatorial structures, and extensive surveys of classical results appear in [8, 26].

Bender [1] first recognized that such limit distributions could be established for general *combinatorial schemas* under analytic conditions of a general character. This

Correspondence to: Michèle Soria, INRIA, Domaine de Voluceau-Rocqencourt, B.P. 105, 78153 Le Chesnay Cedex, France

* Based on a survey talk given at the French–Israeli Conference on Combinatorics and Algorithms, Jerusalem, 1988.

line of investigation was later pursued by Bender, Canfield, Richmond, Compton, and others [2, 4, 7, 12].

In a way, the situation parallels that of the central limit theorem in probability theory. There we know that the common scheme of taking sums of many random variables leads, under wide sets of conditions, to a general asymptotic law, a normal distribution in the limit. Here we show how common combinatorial schemes that form sequences, sets or cycles lead, under suitable conditions, to general asymptotic laws for the number of components in large random structures.

This paper adds to the already known classes a new analytic scheme that generates normal (Gaussian) distributions. Our results concern the 'weak' convergence – i.e. in the sense of distribution functions – of parameters related to the number of components in composite combinatorial structures. A corresponding statement is also often called a 'central limit theorem'. (Local limit thorems deal with density functions instead; they are discussed at length in Bender's paper [1].)

We establish companion results regarding distibution tails which are found to be of exponential decay under a very large set of conditions. The two types of results are complementary: the existence of a limit distribution provides information on distributions near the mean value, whereas exponential tail estimates entail that large deviations from the mean are extremely unlikely.

Analytically, the problem which we are confronted with here amounts to extracting information on coefficients of bivariate generating functions. These are analytic functions of two complex variables of the form[1]

$$P(z,u)=\sum_{n,k\geqslant 0} P_{n,k}u^k z^n. \tag{1}$$

We are, thus, facing a 'double' inversion problem. In some cases, real variable methods may be used; see, in particular, Compton's work [7]. The approach taken here (as well as in [1, 2, 4, 12]) relies instead on complex variable methods. It consists of a two-stage process.

- First, we consider u as a *parameter* and solve a parameterized single-variable inversion problem by estimating

$$p_n(u)\equiv\sum_{k\geqslant 0} P_{n,k}u^k=\frac{1}{2i\pi}\oint P(z,u)\frac{dz}{z^{n+1}} \tag{2}$$

  asymptotically for large n but fixed u.
- Next, once precise estimates for $p_n(u)$ have been derived for enough values of u, we can in turn 'invert' $p_n(u)$ and derive information on the coefficient $P_{n,k}=[u^k]\,p_n(u)$.

The second stage usually relies on the use of continuity theorems for Fourier transforms (Lévy's continuity theorem for characteristic functions) or for Laplace

[1] Depending upon the context, the generating functions may be ordinary or exponential. Thus, $P_{n,k}$ represents a number of structures of *size n* having k '*components*' – up to a possible factor of n!

transforms (also called moment generating functions). We refer to Billingsley's excellent treatment of these topics; see especially Sections 25, 26 and 30 of [3].

For the first stage, the asymptotic technology to be used depends on the profile of the functions under consideration, and especially on $P(z, 1)$.

- In the case of meromorphic functions, the contour can be extended to a circle of large radius, taking into account the residues of the integrand. This is the method used in the original study of Bender [1]. The related technique of 'subtracted singularities' [18, p. 442] is used by Bender and Richmond [2].
- If $P(z, u)$ has algebraic or logarithmic singularities then variations of the Darboux–Pólya method can be used [18, 22, 28]. Our treatment in this paper relies on the method of *singularity analysis* of Flajolet and Odlyzko [10], by which one can transfer on a term-by-term basis asymptotic information on a function to its coefficients. (Ultimately, the method relies on contour integration with a Hankel contour that comes close to the dominant singularity of the integrand.)
- If $P(z, u)$ is entire with exponential growth, or has essential singularities, the saddle point method normally applies, the contour is a circle crossing the saddle point and the main contribution to the integral comes from a small neighbourhood of the saddle point. This is Canfield's method [4].

We can now make our goals more precise. Our object of study is some particular analytic functions of two complex varibles that arise from combinatorial enumerations, and are taken to be of the form

$$P(z, u) = \Xi[C(z); u],$$

where $C(z)$ is generating function of the 'components' (thereby assumed to have positive coefficients), and u is a parameter. There are three major *combinatorial constructions* that form *sequences*, *cycles* and *sets*. Figure 1 describes the *analytic functionals* Ξ that correspond to the three constructions in the labelled and in the unlabelled universe. The reader unfamiliar with this symbolic approach can consult [15] as an entry point to the literature. (The generating functions are ordinary w.r.t.

Construction	Labelled	Unlabelled
Sequence($\mathscr{C}$)	$\dfrac{1}{1-uC(z)}$	$\dfrac{1}{1-uC(z)}$
Cycle($\mathscr{C}$)	$\log\dfrac{1}{1-uC(z)}$	$\displaystyle\sum_{k\geqslant 1}\frac{\phi(k)}{k}\log\frac{1}{1-u^kC(z^k)}$
Set($\mathscr{C}$)	$\exp(uC(z))$	$\exp\left(\displaystyle\sum_{k\geqslant 1}\frac{(-1)^{k+1}}{k}u^kC(z^k)\right)$

Fig. 1. The three major constructions of sequence, cycle and set together with their translation into generating functions in both the labelled and the unlabelled cases.

z for the unlabelled case, while they are exponential in the labelled case. For analytic purposes, the distinction is, however, immaterial and it suffices to take $C(z)$ as an arbitrary power series with positive coefficients.)

The analytic study of combinatorial schemas consists in finding asymptotic laws for coefficients of such bivariate generating functions, given suitable conditions on the component generating function C. For instance, a few analytic schemas giving rise to asymptotically Gaussian coefficients are described succinctly in Fig. 2. Clearly, such analytic schemas cover sequence constructions if they imply the form $1/(1-uC(z))$ cycle constructions when logarithms appear, and set constructions wherever an exponential is involved.

Plan of the paper. The basis of the method is discussed in greater detail in Section 2. Section 3 is principally concerned with an analytic schema,

$$\frac{1}{(1-uC(z))^{\alpha}}\cdot\left(\log\frac{1}{1-uC(z)}\right)^{k}, \tag{3}$$

which is applicable to sequences, cycles, as well as some other composite constructions. We obtain Gaussian limits by means of a continuity theorem; here we have taken the option of using the method of characteristic functions, although Laplace transforms could have equally well been used (see e.g. [5] for similar problems treated via Laplace transforms).

Section 4 introduces the corresponding exponential tail results that arise from a consideration of Laplace transforms. Section 5 exhibits about a dozen applications of these results to fairly classical combinatorial structures like graphs, permutations, mappings, ordered partitions or polynomial factorizations.

Since the first version of this paper was written, our results have been extended by Gao and Richmond [13]. Following the lines of [2], they show that our approach can

Analytic schema	Method	Reference
$\frac{1}{(1-uC(z))^m}$, $m\in\mathbb{N}$	Singularity of meromorphic functions	[1]
$e^{uC(z)}$, $C(z)$ polynomial	Saddle point method	[4]
$e^{uC(z)}$, $C(z)$ logarithmic	Singularity analysis	[12]
$\left.\begin{array}{l}\frac{1}{(1-uC(z))^{\alpha}}\\ \log^k\frac{1}{1-uC(z)}\end{array}\right\}$	Singularity analysis	This paper

Fig. 2. A summary of some analytic schemas leading to Gaussian distributions.

be adapted to analytic functions of $k+1$ complex variables; their results also complement our exponential tails by providing $e^{-\Theta(x^2)}$ estimates.

In another direction, Soria has pursued the investigation of probabilistic properties of general combinatorial schemas. Her work shows a wide range of distributions to occur under precise analytic conditions inside classical structures. A fairly complete typology of limit distributions in combinatorial schemas is given in [27].

2. Analytic methods

Let $P_{n,k}$ be a sequence of nonnegative numbers. By normalization, we define the probability distributions

$$\pi_{n,k}=\frac{P_{n,k}}{P_n}, \quad \text{with } P_n=\sum_j P_{n,j}; \tag{4}$$

we denote by Ω_n a random variable with probability distribution $\{\pi_{n,k}\}_{k\geq 0}$.

Our purpose here is to study the asymptotic distribution of special numbers $P_{n,k}$. In the present context, the sequence $P_{n,k}$ arises from a bivariate generating function

$$P(z,u)=\sum_{n,k\geq 0} P_{n,k}u^k z^n, \tag{5}$$

itself constructed from a function $C(z)=\sum_n C_n z^n$ by means of one of the functionals described in Fig. 1. Our only assumption at this stage is that $C(z)$ has nonnegative coefficients. The problem under consideration is, thus, of a purely analytic nature, namely, it reduces to the study of asymptotic properties of certain analytic functionals.

In combinatorial applications, we always consider two classes of structures: the class $\mathscr{C}$ of components and the class $\mathscr{P}$ of composite structures. The composite structures are of the three possible types described in Fig. 1, that is to say, sequences, cycles or sets. If $C(z)$ is a generating function of structures $\mathscr{C}$, the meaning of Ω_n is then the *random variable giving the number of $\mathscr{C}$ components in a random composite $\mathscr{P}$ structure of size n.*

We have set $P_n=\sum_j P_{n,j}$, and, by a slight abuse of terminology (in the labelled case, generating functions are exponential, so that a normalization factor of $1/n!$ is needed), we may occasionally refer to P_n as the number of composite $\mathscr{P}$ structures of size n, and to C_n as the number of component $\mathscr{C}$ structures of size n. We also define

$$P(z)=P(z,1), \quad \text{so that } P(z)=\sum_{n\geq 0} P_n z^n. \tag{6}$$

Letting $p_n(u)=\sum_k P_{n,k}u^k$, we have the following:

- The *probability generating function* of Ω_n is $p_n(u)/p_n(1)$.
- Its *characteristic function* $\phi_{\Omega_n}(\theta)$ is $p_n(e^{i\theta})/p_n(1)$.
- Its *Laplace transform* $M_{\Omega_n}(\theta)$ is $p_n(e^{\theta})/p_n(1)$.

The mean value μ_n and the variance σ_n^2 of Ω_n are readily computed by differentiation from the probability generating function:

$$\mu_n = \frac{p_n'(1)}{p_n(1)}, \qquad \sigma_n^2 = \frac{p_n''(1)}{p_n(1)} - \frac{p_n'^2(1)}{p_n^2(1)} + \frac{p_n'(1)}{p_n(1)}. \tag{7}$$

In order to establish the Gaussian limit distributions, we consider the normalized variables

$$\Omega_n^* = \frac{\Omega_n - \mu_n}{\sigma_n}.$$

We shall prove the pointwise convergence of the characteristic functions of the Ω_n^* to the characteristic function of a Gaussian variable with mean 0 and variance 1:

$$\phi_{\Omega_n^*}(\theta) \to \mathrm{e}^{-\theta^2/2}. \tag{8}$$

By the continuity theorem for characteristic functions of Paul Lévy [3, Section 26], we can then deduce from (8) the 'weak' convergence[2] of the distribution functions.

Definition 2.1. Let Ω_n be a sequence of random variables. When, for any two real constants $a<b$, we have

$$\Pr\left(a < \frac{\Omega_n - \mu_n}{\sigma_n} < b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-t^2/2}\,\mathrm{d}t, \tag{9}$$

we say that Ω_n (or its normalized form Ω_n^*) is *asymptotically Gaussian*, or that its distribution converges weakly to a Gaussian distribution, or that it satisfies a central limit theorem.

As we shall see in Section 4, a sufficient condition for the sequence of normalized random variables Ω_n^* to have (uniform) exponential tails is that the Laplace transforms be bounded by a constant K, for all θ in a fixed real neighbourhood of 0, i.e.

$$(\exists K)(\forall n), \quad M_{\Omega_n^*}(\theta) < K.$$

Therefore, the main technical problems rest with the estimation of $\phi_{\Omega_n^*}(\theta)$, and $M_{\Omega_n^*}(\theta)$. In terms of Ω_n, these are expressed as

$$\begin{aligned} \phi_{\Omega_n^*}(\theta) &= E(\mathrm{e}^{\mathrm{i}\Omega_n^*\theta}) = \mathrm{e}^{-\mathrm{i}\theta\mu_n/\sigma_n} \frac{p_n(\mathrm{e}^{\mathrm{i}\theta/\sigma_n})}{P_n}, \\ M_{\Omega_n^*}(\theta) &= E(\mathrm{e}^{\Omega_n^*\theta}) = \mathrm{e}^{-\theta\mu_n/\sigma_n} \frac{p_n(\mathrm{e}^{\theta/\sigma_n})}{P_n}. \end{aligned} \tag{10}$$

[2] In the case of weak convergence to a Gaussian distribution, we also have that the $\mathscr{L}^\infty$ distance between the distribution function of Ω_n^* and that of the standard Gaussian variate tends to 0. See, for instance, the remarks in [1, p. 91] and Section 9 of [14].

Our analysis of limit distributions relies on the convergence of characteristic functions. The derivation of tail bounds relies on quantitative estimates of Laplace transforms.

In general, characteristic functions are a finer tool than Laplace transforms (moment generating functions) in the derivation of limit laws, since they always exist. The problems under consideration in the present paper are, however, well conditioned in the sense that both the discrete laws and the limit law have a Laplace transform; thus, as pointed out earlier, either Laplace or Fourier transforms would equally do for the purpose of establishing Gaussian limit laws. In contrast, the use of Laplace transforms for tail estimates is a necessity.

In all our problems, the standard deviation σ_n tends to infinity. Therefore, an analysis based on characteristic functions needs information on $p_n(u)$ for u in a *complex* neighbourhood of 1 along the unit circle ($u=e^{i\phi}$), while an analysis via Laplace transforms requires a knowledge of $p_n(u)$ for u in a *real* interval centred around 1. The computation of the value of $p_n(u)$, thus, appears for each case as a 'perturbation' of that of $p_n(1)$.

The rest of the paper is devoted to functions, $C(z)$ or $P(z)$, with isolated algebraic and logarithmic singularities on their circle of convergence. Thus, singularity analysis techniques [11] will be employed here. We summarize here briefly the main results of this approach.

The crucial point is the (classical) observation that there is a correspondence between function scales and coefficient scales:

$$f(z)=\frac{1}{(1-z/\rho)^{\alpha}}\left(\log\frac{1}{1-z/\rho}\right)^{k} \Rightarrow [z^n]f(z)=\rho^{-n}\frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^{k}(1+o(1)). \quad (11)$$

Under conditions of analytic continuation that are spelled out in [11], we have

$$f(z)=o\left(\frac{1}{(1-z/\rho)^{\alpha}}\left(\log\frac{1}{1-z/\rho}\right)^{k}\right) \Rightarrow [z^n]f(z)=o(\rho^{-n}n^{\alpha-1}(\log n)^{k}). \quad (12)$$

Thus, under these conditions, we are justified in translating an asymptotic expansion of a function near a singularity into a corresponding expansion for its coefficients. This fact can be systematically exploited in the case of functions given by explicit operations like in Fig. 1.

3. Gaussian limit distributions

We examine in this section two analytic schemes and obtain a Gaussian limit distribution.

The first result, Theorem 3.2, provides a limiting distribution for a scheme that generalizes the functionals arising from the sequence and cycle constructions; it is found that the mean and variance are both of order $O(n)$. Theorem 3.2 is in the line of related results of Bender and Richmond (see Theorems 3.2 and 3.4 or Corollary 1 of

[2]). The proof techniques are, however, a little different since we appeal to singularity analysis instead of the method of subtracted singularities. Since our asymptotic engine is in many ways more 'powerful', we may expect this line of attack to be of wider applicability (see [13] for recent results in this direction).

The second result is relative to the set construction which leads to an asymptotic distribution that is Gaussian; in that case, the mean and variance are of the form $O(\log n)$. The latter result was already obtained by us in [12]; we provide here a more synthetic proof that also paves the way for the exponential tail results of the next section.

Sequence and cycle constructions. The sequence construction and the cycle construction lead us to the two schemas,

$$P(z,u)=\frac{1}{1-uC(z)},\qquad P(z,u)=\log\frac{1}{1-uC(z)},$$

which we encapsulate into

$$P(z,u)=\frac{1}{(1-uC(z))^{\alpha}}\left(\log\frac{1}{1-uC(z)}\right)^{k}. \tag{13}$$

The component functions are assumed to satisfy a particular condition.

Definition 3.1. A function $C(z)=\sum_{n\geqslant 0}C_n z^n$ that is analytic at 0 is said to be *1-regular* iff

- its Taylor expansion at 0 involves only nonnegative coefficients,
- σ being the radius of convergence of $C(z)$ at 0, one has $C(\sigma)>1$.

Without loss of generality, we may freely assume that further that $C(z)$ is aperiodic, i.e. not of the form $\phi(z^d)$ for $d\geqslant 2$ and ϕ analytic at 0.

Theorem 3.2. *Consider the probability distributions defined by the bivariate generating function*

$$P(z,u)=\frac{1}{(1-uC(z))^{\alpha}}\left(\log\frac{1}{1-uC(z)}\right)^{k},$$

with $k\geqslant 0$ an integer, and $\alpha\geqslant 0$ a real number. Assume that $C(z)$ is 1-regular. Then the random variable Ω_n associated with the $P_{n,k}$ has mean μ_n and variance σ_n^2 that satisfy (*see Eqs.* (14) *and* (16))

$$\mu_n \sim c_1 n \quad \textit{and} \quad \sigma_n^2 \sim c_2 n \quad (n\to+\infty).$$

Furthermore, Ω_n converges weakly to a limiting Gaussian distribution:

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

Proof. The proof consists in evaluating in turn the number P_n of structures, the mean μ_n, the variance σ_n^2 and, finally, the probability generating function $p_n(u)$. All estimates are based on singularity analysis. For convenience, we present the proof in the case where $\alpha \neq 0$; the case where $\alpha = 0$ leads to the same results with estimates (14) and (16) for the constants c_1 and c_2 still being valid via a rather similar route, so that it will not be detailed here.

(1) Let ρ be the smallest positive real such that $C(\rho)=1$ (ρ exists by assumption of 1-regularity). Using a Taylor expansion of $C(z)$ around ρ, we get

$$\frac{1}{(1-C(z))}=\frac{1}{(1-z/\rho)}\frac{1}{\rho C'(\rho)}\cdot\left(1+\rho\frac{C''(\rho)}{2C'(\rho)}(1-z/\rho)+\mathrm{O}(1-z/\rho)^2\right).$$

We, thus, find around $z=\rho$,

$$P(z)\equiv P(z,1)=\frac{1}{\rho^\alpha C'^\alpha(\rho)}\frac{1}{(1-z/\rho)^\alpha}\cdot\left(\log\frac{1}{1-z/\rho}+\log\frac{1}{\rho C'(\rho)}\right)^k$$
$$\times\left[1+\mathrm{O}\left(\frac{1}{(1-z/\rho)^{-1}}\right)\right].$$

By the transfer principles of singularity analysis, we, thus, find an asymptotic form of the coefficients of $P(z)$, namely,

$$P_n\equiv p_n(1)\sim\frac{1}{\Gamma(\alpha)\rho^\alpha C'^\alpha(\rho)}\rho^{-n}n^{\alpha-1}(\log n)^k.$$

A more detailed expansion follows from refining the singular expansion of $P(z)$ at ρ:

$$P_n\sim\frac{1}{\Gamma(\alpha)\rho^\alpha C'^\alpha(\rho)}\rho^{-n}n^{\alpha-1}\left(Q_0+\frac{Q_1}{n}+\frac{Q_2}{n^2}+\cdots\right),$$

where the Q_i are polynomials of degree at most k in $\log n$.

(2) The mean value of the distribution is $\mu_n\equiv[z^n]\ P_u'(z,1)/P_n$, where $P_u'(z,1)$ denotes the derivatives of $P(z,u)$ with respect to u, taken at $u=1$. The simplest way to carry out computations consists in reducing the study of partial derivatives to that of $P(z), P'(z)$, etc. First, we have

$$P_u'(z,1)=\frac{C(z)}{C'(z)}P'(z).$$

Thus, using a Taylor expansion of $C(z)$ around ρ, we get

$$P_u'(z,1)=\frac{1}{C'(\rho)}P'(z)(1+K(1-z/\rho))+\mathrm{O}((1-z/\rho)^2),$$

where K is expressible in terms of $\rho, C'(\rho)$, and $C''(\rho)$. Hence,

$$[z^n]\,P_u'(z,1)=(n+1)P_{n+1}\frac{1}{C'(\rho)}\left(1+\mathrm{O}\left(\frac{1}{n}\right)\right).$$

Obviously,

$$Q_0(\log(n+1)) = Q_0(\log n) + \frac{1}{n} Q_0'(\log n)\left(1 + \mathrm{O}\left(\frac{1}{n\log n}\right)\right),$$

so that

$$P_{n+1} \sim \frac{1}{\Gamma(\alpha)\rho^{\alpha}C'^{\alpha}(\rho)} \rho^{-(n+1)} n^{\alpha-1}\left(Q_0 + \frac{R_1}{n} + \frac{R_2}{n^2} + \cdots\right),$$

where the R_i are again polynomials of degree at most k in $\log n$. Returning to μ_n, we find

$$\mu_n = c_1 n\left(1 + \mathrm{O}\left(\frac{1}{n}\right)\right), \quad \text{with } c_1 = \frac{1}{\rho C'(\rho)}. \tag{14}$$

A more detailed expansion provides the constant term in μ_n (valid for the case $\alpha \neq 0$ only!):

$$\mu_n = \frac{1}{\rho C'(\rho)} n + \frac{1}{\rho C'(\rho)} + \frac{C''(\rho)}{C'^2(\rho)} - 1 + \mathrm{o}(1). \tag{15}$$

(3) The variance is $\sigma_n^2 \equiv [z^n] P''_{u^2}(z, 1)/P_n - \mu_n^2 + \mu_n$; we use the relation

$$P''_{u^2}(z, 1) = \frac{C^2(z)}{C'^2(z)} P''(z) - \frac{C^2(z)C''(z)}{C'^3(z)} P'(z).$$

Proceeding as above, it can be shown that

$$\frac{1}{P_n}[z^n]\frac{C^2(z)}{C'^2(z)} P''(z) \sim \frac{n^2}{\rho^2 C'^2(\rho)}.$$

The term of order n^2 cancels with the term coming from the square of the mean value (14); thus, the order of growth of the variance is subquadratic. More detailed computations reveal that

$$\sigma_n^2 \sim c_2 n, \quad \text{where } c_2 = \frac{1}{\rho^2 C'^2(\rho)} + \frac{C''(\rho)}{\rho C'^3(\rho)} - \frac{1}{\rho C'(\rho)}, \tag{16}$$

which turns out to be valid for the two cases $\alpha \neq 0$ and $\alpha = 0$.

(4) For the limit distribution, we have to evaluate $p_n(e^{i\theta/\sigma_n})/P_n$. The evaluation of $p_n(u)$, the coefficient of z^n in $P(z, u)$, is similar to the evaluation of $p_n(1)$.

Let $\rho(u)$ be the root of smallest modulus of the equation $C(\rho(u)) = u^{-1}$. We have $\rho = \rho(1)$ and, for u close enough to 1, by the implicit function theorem, $\rho(u)$ lies in a neighbourhood of ρ and depends analytically on u:

$$\rho(u) = \rho - \frac{1}{C'(\rho)}(u-1) - \frac{C''(\rho) - 2C'^2(\rho)}{C'^3(\rho)} \frac{(u-1)^2}{2} + \mathrm{O}((u-1)^3). \tag{17}$$

Expanding $C(z)$ around $\rho(u)$, and transferring to coefficients, we get

$$p_n(u) = \frac{1}{\Gamma(\alpha)\rho^{\alpha}(u)C'^{\alpha}(\rho(u))} \rho(u)^{-n} n^{\alpha-1} \log^k n \left(1 + \mathrm{O}\left(\frac{1}{\log n}\right)\right),$$

uniformly in u in a small neighbourhood of 1. Thus,

$$\frac{p_n(e^s)}{P_n} = \exp\left(-n\log\frac{\rho(e^s)}{\rho(1)}\right)\cdot(1+o(1)),$$

where the implied constant in the o-estimate is uniform for s sufficiently close to 0. Moreover, since the function $\rho(e^s)$ admits a full asymptotic expansion around $s=0$, we have

$$\log\frac{\rho(e^s)}{\rho(1)} = s\frac{\rho'(1)}{\rho(1)} + \frac{s^2}{2}\left(\frac{\rho''(1)}{\rho(1)} - \frac{\rho'^2(1)}{\rho^2(1)} + \frac{\rho'(1)}{\rho(1)}\right) + O(s^3).$$

Instantiating with $s = i\theta/\sigma_n$, we thus have, for the characteristic function $\phi_{\Omega_n^*}(\theta)$ as defined in Section 2,

$$\phi_{\Omega_n^*}(\theta) \sim \exp\left(-i\theta\frac{\mu_n}{\sigma_n} - n\frac{i\theta}{\sigma_n}\frac{\rho'(1)}{\rho(1)} + \frac{n\theta^2}{2\sigma_n^2}\left(\frac{\rho''(1)}{\rho(1)} - \frac{\rho'^2(1)}{\rho^2(1)} + \frac{\rho'(1)}{\rho(1)}\right) + O\left(\frac{\theta^3}{\sigma_n^3}\right)\right). \tag{18}$$

Using expansion (17), as well as estimates (14) and (16) of μ_n and σ_n, we find

$$\phi_{\Omega_n^*}(\theta) \to e^{-\theta^2/2}.$$

Thus, the sequence $\{\Omega_n^*\}$ converges weakly to a Gaussian limit distribution. □

Set constructions. For the second theorem, we need a precise statement of the conditions for the generating function of components to be of logarithmic type.

First, we let $\Delta_0(\rho, \eta, \phi)$, with $\rho > 0$, $\eta > \rho$, and $0 < \phi < \pi/2$, denote the indented disk

$$\Delta_0(\rho, \eta, \phi) = \{z \mid |z| \leqslant \eta \text{ and } |\mathrm{Arg}(z-\rho)| \geqslant \phi\}.$$

Definition 3.3. 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 dominant singularity ρ, multiplier a and constant K) if it is analytic inside a domain Δ_0, and there we have

$$G(z) = a\log\frac{1}{1-z/\rho} + K + o\left(\left(\log\frac{1}{1-z/\rho}\right)^{-1}\right) \tag{19}$$

as $z \to \rho$.

Note. An oversight in our earlier work [12] led us to pose a definition of a logarithmic function that is a little too loose. The definition of a logarithmic function in [12, p. 169, Eq. (2.2)] (with a requirement that the error term in (19) be only $K + o(1)$) should be changed to Eq. (19) above. With this correction, the statements of theorems and the examples of [12] remain unaffected.

Theorem 3.4. *Consider the probability distributions defined by the bivariate generating function*

$$P(z,u)=\exp(uC(z)). \tag{20}$$

Assume that $C(z)$ is a logarithmic function with multiplier a. Then the random variable Ω_n associated with the $P_{n,k}$ converges weakly to a limiting Gaussian distribution. The mean μ_n and variance σ_n^2 of Ω_n satisfy, as $n\to\infty$,

$$\mu_n \sim a\log n \quad \text{and} \quad \sigma_n^2 \sim a\log n.$$

Proof. Let ρ be the dominant singularity of $C(z)$, and set

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

Then $P(z,u)$ is of the form

$$P(z,u)=\exp(uR(z))\frac{1}{(1-z/\rho)^{au}}.$$

By virtue of singularity analysis, this gives $P_n=\rho^{-n}n^{a-1}e^K/\Gamma(a)(1+o(1/\log n))$. The asymptotic forms of μ_n and σ_n follow through an identical argument.

Considering u as a parameter, we derive in the same vein

$$p_n(u)=\frac{\rho^{-n}n^{au-1}e^{uK}}{\Gamma(au)}\left(1+o\left(\frac{1}{\log n}\right)\right).$$

This estimation is uniform, for u in a sufficiently small complex neighbourhood of 1. Thus, we have

$$\frac{p_n(e^s)}{P_n}=n^{a(e^s-1)}\left(1+o\left(\frac{1}{\log n}\right)\right).$$

Now $i\theta/\sigma_n$ tends to 0 when n tends to infinity; so,

$$\frac{p_n(e^{i\theta/\sigma_n})}{p_n(1)}=\exp\left(a\log n\left(\frac{i\theta}{\sigma_n}-\frac{\theta^2}{2\sigma_n^2}+O(\theta^3)\right)\right)(1+o(1)). \tag{21}$$

Substituting the values of μ_n and σ_n^2, we get

$$\phi_{\Omega_n^*}(\theta)\equiv e^{-i\theta\mu_n/\sigma_n}\frac{p_n(e^{-i\theta/\sigma_n})}{P_n}\to e^{-\theta^2/2},$$

which implies the weak convergence of $\{\Omega_n^*\}$ to a Gaussian limit distribution. □

The proof technique of [12] consisted in going back to the original Hankel contour that is at the basis of singularity analysis methods. The proof outlined here takes advantage of the uniformity of estimates provided by singularity analysis.

4. Exponential tails

Weak convergence of probability distributions to a limit provides information on distributions near their centre (whence the denomination of 'central limit theorems'). Such results are, thus, useful for characterizing relatively frequent cases. However, for applications to statistics, combinatorics or analysis of algorithms, it is often useful to characterize the rarity of extreme cases or, in other words, find information on possible 'large deviations' from the average. An important concept in this area is that of distributions with an exponential tail. It turns out that the distributions considered in this paper all have exponential tails, so that large deviations are extremely unlikely, and have a lower probability of occurrence than would be predicted from a Chebyshev moment inequality of arbitrary order.

Definition 4.1. Let Y be a normalized random variable with mean 0 and standard deviation 1. We say that Y has an *exponential tail with parameter* $\alpha<1$ if

$$\exists C>0,\ \forall k>0,\quad \Pr(|Y|>k)<C\alpha^k.$$

Similarly, if $\{Y_n\}_{n\geqslant 0}$ is a sequence of normalized random variables, we say that $\{Y_n\}_{n\geqslant o}$ has an *exponential tail with parameter* $\alpha<1$ if

$$\exists C>0,\ \forall k>0,\ \forall n,\quad \Pr(|Y_n|>k)<C\alpha^k.$$

The last part of the definition is, therefore, a *uniform* version of the first one. We also extend the definition to unnormalized variables: a sequence Ω_n of random variables is said to have an exponential tail whenever the normalized sequence Ω_n^* itself has an exponential tail. Variables with an exponential tail have an exponentially vanishing probability of large deviations from the expected values.

Observe first that the weak convergence of a sequence $\{Y_n\}$ to a limit Y with an exponential tail (e.g. a Gaussian distribution) does not necessarily entail that the Y_n themselves have an exponential tail according to the definition above: It suffices to consider a probability distribution with mass $1/n$ concentrated at point $x=\sqrt{n}$, and everywhere else with a Gaussian density normalized by a factor of $1-1/n$. Exponential tail estimates are, therefore, a useful complement to weak convergence results.

For a single random variable Y, it is well known (see e.g. Sections 9 and 22 of [3]) that there are relations between tail distributions and inequalities satisfied by the moment generating function. For a completeness of exposition, we state the following proposition.

Proposition 4.2. (*i*) *Let Y be a random variable whose Laplace transform $M(\theta)\equiv E(e^{\theta Y})$ is defined for θ in an interval $I=[\theta_0,\theta_1]$ enclosing* 0. *Then Y has an exponential tail, with*

$$C=\sup_{\theta\in I} M(\theta) \quad \textit{and} \quad \alpha=e^{-\min(-\theta_0,\theta_1)}.$$

(ii) Similarly, for a sequence $\{Y_n\}$ with Laplace transforms $M_n(\theta)$, if we have

$$\exists C>0,\ \forall n,\quad M_n(\theta)<C,$$

for all θ in some finite interval $[\theta_0, \theta_1]$ around 0, then the sequence $\{Y_n\}$ admits an exponential tail in the sense of our definition.

Proof. For a single variable, we consider the upper tail estimate $\Pr(Y>k)$ for $k>0$. We have, for any $\theta>0$,

$$\begin{aligned}\Pr(Y>k) &= \Pr(e^{\theta Y}>e^{\theta k})\\ &\leqslant e^{-\theta k}E(e^{\theta Y})\\ &\leqslant Ce^{-\theta k}.\\ &\leqslant Ce^{-\theta_1 k}.\end{aligned} \tag{22}$$

The first upper bound follows by Markov's inequality [3, p. 283] applied to the moment generating function $E(e^{\theta Y})$. The other two result from the definition of C and the 'best' choice of $\theta=\theta_1$.

The lower tail estimate and the extension to sequences of random variables follow from identical arguments. □

A nice consequence of analytic estimates derived in Section 3 is that we get, with a little additional work, exponential tail results for combinatorial distributions that admit a Gaussian limiting law.

Theorem 4.3. *Let $p_n(u)$ be defined by*

$$\sum_n p_n(u)z^n = \frac{1}{(1-uC(z))^\alpha}\left(\log\frac{1}{1-uC(z)}\right)^k,$$

where k is an integer and α is a real number $\geqslant 0$. Assume that $C(z)$ is 1-regular. Let Ω_n be the random variable with probability generating function $p_n(u)/p_n(1)$. Then the sequence of random variables Ω_n admits an exponential tail.

Proof. Let $M_{\Omega_n^*}(\theta)$ denote the Laplace transform of Ω_n^*:

$$M_{\Omega_n^*}(\theta) = e^{-\theta\mu_n/\sigma_n}E(e^{\Omega_n\theta/\sigma_n}) = e^{-\theta\mu_n/\sigma_n}\frac{p_n(e^{\theta/\sigma_n})}{P_n}.$$

Using the same estimate as in the proof of Theorem 3.2, we find

$$M_{\Omega_n^*}(\theta) = e^{-\theta\mu_n/\sigma_n}\left(\frac{\rho(e^{\theta/\sigma_n})}{\rho(1)}\right)^{-n}(1+o(1)),$$

the estimation being uniform for θ lying in a fixed (that may be *arbitrarily* chosen!) real neighbourhood I of 0. Expanding function ρ around 1, we get a formula analogous to

Eq. (18):

$$M_{\Omega_n^*}(\theta) \sim \exp\left(-\theta\frac{\mu_n}{\sigma_n} - \frac{n\theta}{\sigma_n}\frac{\rho'(1)}{\rho(1)} + \mathrm{O}\left(\frac{n\theta^2}{2\sigma_n^2}\right)\right).$$

Since σ_n^2 is of order n, and $\mu_n = -n\rho'(1)/\rho(1) + \mathrm{O}(1)$, we conclude that $M_{\Omega_n^*}(\theta)$ is uniformly $\exp(\mathrm{O}(1))$, which means uniformly bounded for θ staying in the fixed interval I. □

Along the same principle of proof, we can add an exponential tail result to Theorem 3.4.

Theorem 4.4. *Let $p_n(u)$ be defined by*

$$\sum_n p_n(u) z^n = \exp(uC(z)),$$

where $C(z)$ is a logarithmic function. Then the sequence of random variables Ω_n, with generating function $p_n(u)/p_n(1)$, admits an exponential tail.

Proof. We have the counterpart of Eq. (21),

$$\mathrm{e}^{-\theta\mu_n/\sigma_n}\frac{p_n(\mathrm{e}^{\theta/\sigma_n})}{p_n(1)} = \exp\left((a\log n - \mu_n)\left(\frac{\theta}{\sigma_n}\right) + \mathrm{O}\left(\frac{\theta^2}{2\sigma_n^2}\right)\right)(1+\mathrm{o}(1)). \tag{23}$$

The proof now relies on the fact that the error terms of $\mu_n - a\log n$ are much smaller than σ_n and on the fact that σ_n^2 is of order $\log n$. Thus, for θ in a fixed interval I, $M_{\Omega_n^*}(\theta)$ remains uniformly bounded. □

As a conclusion, note that it is also possible to derive superexponential bounds[3] with the same methods. An alternative approach to the problem could be to consider asymptotic estimates for densities ('local limit theorems'), in the style of Bender [1]. This may involve, however, rather delicate estimates away from the centre.

Exponential tail results should prove sufficient for many practical applications. For instance, the first nontrivial upper bound on the height of binary search trees was obtained by Robson [24] using exponential tail properties of Stirling numbers of the first kind (in that case, explicit generating functions are available and the analysis is, therefore, easier).

5. Examples and extensions

There are many cases of applications of the techniques reviewed here, owing to the generality of the combinatorial schemas under consideration. A small sample is given below and we also indicate a few directions into which our results could be extended.

[3] From the proofs of Theorems 4.3 and 4.4 it would be possible to optimize on the bounds that one derives by adequately selecting the interval I.

Example 5.1 (*Ordered partitions and cyclic partitions*). The ordered partitions of an n-set are described by the bivariate generating function

$$\frac{1}{1-u(e^z-1)},$$

where u marks the number of blocks. The corresponding distribution is $P_{n,k}=k!\,S_{n,k}$, with $S_{n,k}$ a Stirling number of the second kind, and P_n is sometimes referred to as 'preferential arrangement' or 'surjection' number. From Theorem 3.2, the $P_{n,k}$ are asymptotically normal, with exponential tails (Theorem 4.3). This example is well known and asymptotic normality already follows from Bender's results [1].

If we consider, instead, cyclic partitions of an n-set, we are led to a generating function

$$\log\frac{1}{1-u(e^z-1)},$$

which does not fall into Bender's class. The enumeration sequence becomes $P_{n,k}=(k-1)!\,S_{n,k}$. From the logarithmic case of Theorem 3.2 ($\alpha=0$, $k=1$), the distribution of the number of blocks is again asymptotically Gaussian and, from Theorem 4.3, it has exponential tails.

The mean and the variance of the number of blocks in an n-partition satisfy

$$\mu_n \sim \frac{1}{2\log 2}n, \qquad \sigma_n^2 \sim \left(\frac{1}{4\log^2 2}-\frac{1}{4\log 2}\right)n$$

in both the sequence and the cycle cases.

Example 5.2 (*Permutations and 2-regular graphs*). Several examples of the application of Theorem 3.4 have been given in [12], and will not be duplicated here. Let us just say that prototypes of application are the functions

$$\exp\left(u\log\frac{1}{1-z}\right) \quad \text{and} \quad \exp\left(\frac{u}{2}\left(\log\frac{1}{1-z}-z-\frac{z^2}{2}\right)\right),$$

corresponding to the distribution of cycles in *permutations* and of connected components in 2-*regular graphs*. Another interesting example, which goes back to early work on random mappings, is the distribution of connected components in random mappings. The bivariate generating function is

$$\exp\left(u\log\frac{1}{1-a(z)}\right),$$

where $a(z)=ze^{a(z)}$ is the generating function of Cayley trees.

Example 5.3 (*Trees of cycles and cycles of trees*). More generally, Theorem 3.2 and its companion Theorem 4.3 express asymptotic properties for objects obtained by 'composing' a class of structures having a generating function with an algebraico-logarithmic singularity (e.g. cycles, trees) and a suitably regular generating function for the class of components. As typical instances, Gaussian distributions and exponential tails will be present in the two bivariate schemas

$$\lambda(u\beta(z)) \quad \text{and} \quad \beta(u\lambda(z)),$$

where

$$\lambda(z)=\log\frac{1}{1-z} \quad \text{and} \quad \beta(z)=\frac{1-\sqrt{1-2z^2}}{z}, \tag{24}$$

corresponding to cycles of trees and trees of cycles. Here, trees are binary, labelled and nonplane:

$$\beta(z)=z+\frac{z}{2}(\beta(z))^2.$$

The case of $\lambda(\beta(z))$ is an application of Theorems 3.2 and 4.3 with $\beta(z)$ being 1-regular ($\beta(1/\sqrt{2})=\sqrt{2}>1$). The case of $\beta(\lambda(z))$ illustrates an extension to negative exponents ($\alpha=-1/2$) of Theorems 3.2 and 4.3.

Variations on analytic conditions. The methods developed in the previous sections are applicable to a variety of analytic schemes. We may allow various types of modifications in the basic schemes considered so far – a typical example being the functionals attached to *unlabelled constructions* in the next subsection – as well as allowing for '*error terms*' of various sorts.

An easy qualitative analysis of generating functions provides, in a large number of cases, direct proofs of Gaussian approximations and exponential tails estimates for combinatorial enumerations. The general methodology appears to be quite robust and we proceed to indicate a few direct extensions whose proofs follow the same path.

One class of applications concerns composite structures with structural definitions of the type

$$\mathscr{P}=\mathscr{F}\times \mathit{Sequence}(\mathscr{C})$$

as well as their set or cycle counterparts. The generating function form becomes

$$P(z,u)=f(z)\cdot\frac{1}{1-uC(z)}.$$

If the generating function $f(z)$ of $\mathscr{F}$ is regular at the dominant singularity of $P(z,1)$ or if it has there only a dominant algebraico-logarithmic singularity, it plays the role of a small perturbation, and distributions remain Gaussian in the limit, with exponential tails.

Situations where multiple dominant singularities (of the proper type) appear can also be treated by our methods, just using composite Hankel contours. The net result, valid for Theorems 3.2, 3.4, 4.3 and 4.4 is still the occurrence of Gaussian limit distributions and exponential tails.

Example 5.4 (*Semipermutations*). We define a semipermutation as a set of undirected cycles. The class of semipermutations of a component class has generating function

$$P(z,u)=\exp\left(\frac{1}{2}\log\frac{1}{1-uC(z)}+\frac{uC(z)}{2}+\frac{uC^2(z)}{4}\right)\equiv\frac{e^{uC(z)/2+u^2C^2(z)/4}}{\sqrt{1-uC(z)}},$$

where $C(z)$ is the components generating function, and u marks the number of components. The asymptotic distribution of the number of components remains Gaussian provided that $C(z)$ is a 1-regular function. For example, we can take for $C(z)$ the generating function $\beta(z)$ of (24). We find that the number of components in a semipermutation has a distribution which is asymptotically Gaussian, with mean $\mu_n \sim n/3$ and variance $\sigma_n^2 \sim 0.88n$.

Example 5.5 (*Cycles in restricted permutations*). The decomposition of permutations into cycles corresponds to the generating function equation

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

Goncharov's well-known result states that the associated Ω_n (the distribution of Stirling numbers of the first kind) is asymptotically normal, with mean and variance asymptotic to $\log n$.

Consider the distribution of the number of cycles in permutation where all cycles are restricted to have *odd length*. The analytic form is

$$\exp\left(\frac{u}{2}\left(\log\frac{1+z}{1-z}\right)\right).$$

We now have two dominant singularities at $z=\pm 1$, but, combining local expansions at ± 1, one still derives the Gaussian property, with mean and variance asymptotic to $\frac{1}{2}\log n$.

Similarly, these asymptotic properties of the distribution are preserved for the number of cycles of odd length in general permutations, which corresponds to the generating function

$$\frac{1}{\sqrt{1-z^2}}\exp\left(\frac{u}{2}\left(\log\frac{1+z}{1-z}\right)\right).$$

Example 5.6 (*Unary nodes in* 1–2 *trees*). The bivariate generating function of 1–2 trees, with u marking the number of unary nodes, satisfies

$$T(u,z)=z(1+uT(u,z)+T^2(u,z)),$$

whose solution is

$$T(u,z)=\frac{1-zu}{2z}-\frac{\sqrt{1-z(u-2)}}{2z}\sqrt{1-z(u+2)}.$$

For u close to 1, $T(u, z)$ has dominant singularity at $\rho(u)=1/(u+2)$. The conjugate root at $z=1/(u-2)$ introduces only a small perturbation. Using a natural extension of the proof of Theorem 3.2 adapted to $\alpha=-1/2$, we can derive a Gaussian limit distribution for this parameter.[4]

Unlabelled structures. Unlabelled constructions like set, multiset or cycle lead to schemes that involve the component generating function taken at points of the form z^l (see Fig. 1). Under frequently satisfied conditions, the terms $C(z^l)$, with $l \geqslant 2$, only tend to modify (additive or multiplicative) constants in singular expansions of generating functions. This situation is well known in graphical enumerations [16].

Theorem 5.7. *Consider the probability distributions defined by the bivariate generating function*[5]

$$P(z, u)=\log\frac{1}{1-uC(z)}+\sum_{l\geqslant 2}\frac{\phi(l)}{l}\log\frac{1}{1-u^l C(z^l)}, \tag{25}$$

which corresponds to an unlabelled cycle construction. Assume that $C(z)$ is 1-regular, and also that the smallest positive root ρ of the equation $C(x)=1$ satisfies $\rho<1$. Then the random variable Ω_n associated with the $P_{n,k}$ has mean μ_n and variance σ_n^2 that satisfy

$$\mu_n \sim c_1 n \quad \textit{and} \quad \sigma_n^2 \sim c_2 n \quad (n \to +\infty).$$

Furthermore, Ω_n converges weakly to a limiting Gaussian distribution, and it admits exponential tails.

Proof. The condition $\rho<1$ implies that $P(z, u)$ is driven by its first term, namely $\log(1-uC(z))^{-1}$. From this same condition, we see that each of the remaining terms in expansion (25) is analytic in a polydisc $|u| \leqslant 1+\varepsilon$ and $|z| \leqslant \rho+\varepsilon$ for some fixed $\varepsilon>0$, where we can also impose the conditions $(\rho+\varepsilon)(1+\varepsilon)<1-\varepsilon$. Moreover, for $l \geqslant 2$, $|C(z^l)|<K\cdot|z|^l$ when $|z| \leqslant \rho+\varepsilon$ for some constant K. Then we have

$$\begin{aligned}\left|\sum_{l\geqslant 2}\frac{\phi(l)}{l}\log\frac{1}{1-u^l C(z^l)}\right| &\leqslant K\cdot\sum_{l\geqslant 2}\frac{\phi(l)}{l}\sum_{k\geqslant 1}\frac{|u|^{lk}|z|^{lk}}{k}\\ &\leqslant K\cdot\sum_{n}\left(\sum_{d\mid n}\phi(d)\right)\frac{|u|^n|z|^n}{n}\\ &\leqslant K\cdot\sum_{n}|uz|^n=\frac{K}{1-|zu|}.\end{aligned}$$

[4] Actually, stronger multivariate normal distribution results are known for simple families of trees, as can be seen through Lagrange inversion and reduction to multinomial distributions. The present example is only meant to illustrate a simple application to certain generating functions with algebraic singularities when the variables u and z need not be 'separated'.

[5] In this formula, $\phi(n)$ represents Euler's totient function, i.e. the number of integers in the interval $[1, n-1]$ that are coprime to n.

Thus, analytically, $P(z, u)$ behaves in the vicinity of $u=1$ like its first term, to which Theorems 3.2 and 4.3 can be applied. □

Example 5.8 (*Cyclic compositions of integers*). Positive integers have generating function $A(z)=z/(1-z)$, which is 1-regular, and reaches 1 for $\rho=1/2$, so that the conditions of the theorem are satisfied. In accordance with (25), the bivariate generating function for cyclic compositions, with variable u marking the number of summands, is

$$P(z, u)=\log\frac{1}{1-uz/(1-z)}+\sum_{l\geqslant 2}\frac{\phi(l)}{l}\log\frac{1}{1-u^l z^l/(1-z^l)}.$$

Thus, the distribution of summands is asymptotically Gaussian, and it admits exponential tails. The mean and variance of Ω_n are

$$\mu_n \sim \tfrac{1}{2}n, \qquad \sigma_n^2 \sim \tfrac{1}{4}n. \tag{26}$$

Note the similarity with the distribution of summands in *linear* compositions of integers, with generating function $1/(1-uA(z))$, which leads to mean and variance of the same asymptotic form (26).

The analytic schemes corresponding to the unlabelled set and multiset constructions are, respectively,

$$\exp\left(\sum_{l\geqslant 1}(-1)^l\frac{u^l}{l}C(z^l)\right) \quad \text{and} \quad \exp\left(\sum_{l\geqslant 1}\frac{u^l}{l}C(z^l)\right). \tag{27}$$

Both formulae combine the exponential $\exp(u(C(z)))$ that we find in the labelled case and factors involving the $\{C(z^l)\}_{l\geqslant 2}$. If $C(z)$ is of logarithmic type, the Gaussian limit still holds true, as shown in [12]. A modification of the proof of Theorem 4.4 also permits us to extend the exponential tail result to this schema.

Theorem 5.9. *Consider the probability distributions corresponding to the set and the multiset schemas:*

$$\sum_{n=0}^{\infty} p_n(u)z^n \equiv \exp\left(\sum_{l\geqslant 1}\pm\frac{u^l}{l}C(z^l)\right).$$

If $C(z)$ is logarithmic and has a radius of convergence strictly less than 1, *then the random variable Ω_n with generating function $p_n(u)$ is asymptotically normal and it admits exponential tails.*

Example 5.10 (*Polynomial factorization*). It is well known that the distribution of the number of prime factors in a random integer from the interval $[1, n]$ is asymptotically normal. This is classically known as the Erdős–Kac theorem. As a consequence of Theorem 5.9, the authors derived in [12] an 'Erdős–Kac theorem' for polynomial over finite fields: The number of irreducible factors in a random monic polynomial of large degree over GF(q) tends to a limiting Gaussian distribution. An exponential tail property also holds in such a case.

6. Conclusion

Many combinatorial schemes are now known to be at the origin of limit distributions, with the simplest cases leading to Gaussian, Poisson geometric or other classical distributions. The nature of laws arising in nonrecursive structures generated by sequence, cycle and set construction is, at present, better understood and we can foresee a typology emerging from the discussion of [9, 27].

At the same time, results about counting and mean values are even decidable for suitable classes of combinatorial problems, as shown in [11]. In a related context, that of the so-called zero–one laws and asymptotic laws, large classes of enumerative problems in logic are known to have asymptotic distributions in the limit (the limits are often from the set $\{0, 1\}$, whence the name). We refer the reader to the works of Lynch [21], regarding random mappings, or Compton [6], regarding general logical frameworks.

The classification of distributions that arise in recursive structures represents an appreciably more difficult problem. For instance, path lengths in planar trees and in binary search trees are described by the two *functional equations*

$$P(z,u)=\frac{z}{1-P(zu,u)} \quad \text{and} \quad \frac{\partial P(z,u)}{\partial z}=(P(zu,u))^2.$$

It is only for the first equation that we have an expression for the limit law since Louchard [20] derived a representation involving the Airy function. The second problem – which is identical to that of the distribution of costs for Quicksort – still represents an intriguing and (partly) open problem [17, 23, 25].

Acknowledgment

This research was supported in part by the Esprit II Basic Research Actions Program of the EC under contract No. 3075 (Project ALCOM).

References

[1] E.A. Bender, Central and local limit theorems applied to asymptotic enumeration, J. Combin. Theory 15 (1973) 91–111.

[2] E.A. Bender and L.B. Richmond, Central and local limit theorems applied to asymptotic enumeration II: multivariate generating functions, J. Combin. Theory Ser. A 34 (1983) 255–265.

[3] P. Billingsley, Probability and Measure, 2nd edn. (Wiley, New York, 1986).

[4] E.R. Canfield, Central and local limit theorems for the coefficients of polynomials of binomial type, J. Combin. Theory. Ser. A 23 (1977) 275–290.

[5] Ch.A. Charalambides and A. Kyriakoussis, An asymptotic formula for the exponential polynomials and a central limit theorem for their coefficients, Discrete Math. 54 (1985) 259–270.

[6] K.J. Compton, A logical approach to asymptotic combinatorics. I. First order properties, Adv. Math. 65 (1987) 65–96.

[7] K.J. Compton, Some methods for computing component distribution probabilities in relational structures, Discrete Math. 66 (1987) 59–77.

[8] F.N. David and D.E. Barton, Combinatorial Chance (Charles Griffin, London, 1962).

[9] P. Flajolet, Elements of a general theory of combinatorial structures, in: L. Budach, ed., Fundamentals of Computation Theory, Lecture Notes in Computer Science, Vol. 199 (Springer, Berlin, 1985) 112–127; Proc. FCT '85, Cottbus, 1985.

[10] P. Flajolet and A.M. Odlyzko, Singularity analysis of generating functions, SIAM J. Discrete Math. 3(2) (1990) 216–240.

[11] P. Flajolet, B. Salvy and P. Zimmermann, Automatic average-case analysis of algorithms, Theoret. Comput. Sci. 79(1) (1991) 37–109.

[12] P. Flajolet and M. Soria, Gaussian limiting distributions for the number of components in combinatorial structures, J. Combin. Theory. Ser. A 53 (1990) 165–182.

[13] Z. Gao and L.B. Richmond, Central and local limit theorems applied to asymptotic enumerations IV: multivariate generating functions, J. Comput. Appl. Math. 41 (1992) 177–186.

[14] B.V. Gnedenko and A.N. Kolmogorov, Limit Distributions for Sums of Independent Random Variables (Addison–Wesley, Reading, MA, 1968).

[15] I.P. Goulden and D.M. Jackson, Combinatorial Enumeration (Wiley, New York, 1983).

[16] F. Harary and E.M. Palmer, Graphical Enumeration (Academic Press, New York, 1973).

[17] P. Hennequin, Combinatorial analysis of quicksort algorithm, RAIRO Theoret. Inform. Appl. 23(3) (1989) 317–333.

[18] P. Henrici, Applied and Computational Complex Analysis, Vol. 2 (Wiley, New York, 1977).

[19] D.E. Knuth, The Art of Computer Programming, Vol. 1: Fundamental Algorithms (Addison-Wesley, Reading, MA, 1968).

[20] G. Louchard, The Brownian excursion: a numerical analysis, Comput. Math. Appl. 10(6) (1984) 413–417.

[21] J.F. Lynch, Probabilities of first-order sentences about unary functions, Trans. Amer. Math. Soc. 287(2) (1985) 543–568.

[22] F.W.J. Olver, Asymptotics and Special Functions (Academic Press, New York 1974).

[23] M. Régnier, A limiting distribution for quicksort, RAIRO Theoret. Inform. Appl. 23(3) (1989) 335–343.

[24] J.M. Robson, The height of binary search trees, Austral. Comput. J. 11 (1979) 151–153.

[25] U. Rösler, A limit theorem for quicksort, RAIRO Theoret. Inform. Appl. 25 (1991) 85–100.

[26] V.N. Sachkov, Verojatnostnye Metody v Kombinatornom Analize (Nauka, Moscow, 1978).

[27] M. Soria, Méthodes d'analyse pour les constructions combinatoires et les algorithmes, Doctorat d'etat, Université de Paris-Sud, Orsay, 1990.

[28] R. Wong and M. Wyman, The method of Darboux, J. Approx. Theory 10 (1974) 159–171.