Limit distributions and analytic methods

M. Drmota
University of Vienna
September 17, 1992

[summary by Michèle Soria]

Abstract

This paper presents a survey of analytic methods for estimating coefficients of functions
in complex variables, and their application to obtain limit distributions in combinatorial
structures. Two cases are specially investigated, with new results given: functional equations
and product schemas.

1. Analytic Methods

Let \( y(x) = \sum y_n x^n \), analytic at the origin, with \( y_n \geq 0 \); the coefficients can be evaluated by Cauchy’s
formula. Two types of well known methods are used according to the nature of \( y(x) \).

1.1. Saddle point method. For functions with at least exponential growth, for example Hayman’s
admissible functions [14], the saddle point method applies:

\[
y_n \sim \rho_n^{-n} \frac{y(\rho_n)}{\sqrt{2\pi \sigma_n^2}},
\]

where the saddle point \( \rho_n \) satisfies \( \frac{\partial}{\partial u} \log y(\rho_n e^u) \big|_{u=0} = n \), and \( \sigma_n^2 = \frac{\partial^2}{\partial u^2} \log y(\rho_n e^u) \big|_{u=0} \).

1.2. Singularity analysis. For functions with algebraic and logarithmic singularities, i.e. with local
behaviour \( y(x) = \frac{1}{(1-x)^\alpha} \log^\beta \frac{1}{1-x}, \alpha \in \mathbb{R} - \{-1, -2, \ldots\} \) and \( \beta \in \mathbb{R} \), singularity analysis on a Hankel
contour (Flajolet and Odlyzko [9]) gives

\[
y_n \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} \log^\beta(n).
\]

An interesting application of both methods is the asymptotics of coefficients of powers of generating
functions: \( [x^n] y(x)^k \), when both \( n \) and \( k \) tend to infinity.

- When \( 0 < a \leq k/n \leq b < \infty \), the saddle point of \( y(x)^k \) stays in a bounded interval \( [c, d] \), \( c > 0, d < \infty \),
and the saddle point method gives an estimation which is uniform for \( k/n \in [a, b] \). This result
was proved by Daniels [4], and extended by Good [13]. The case \( n = o(k) \) is studied by Gardy [12].
- When \( k = o(n) \), the saddle point method does not apply, but singularity analysis can be used for
some specialised functions \( y(x) \). For example, Drmota [6, 8] shows that

\[
[x^n](1 - \sqrt[4]{1-x})^k \sim \frac{k}{2n^{3/2}\sqrt{\pi}} e^{-k^2/4n}, \quad \text{uniformly for} \quad k = o(\sqrt{n \log n}).
\]

\[
[x^n]\left(\log \frac{1}{1-x}\right)^k \sim \frac{\log^k n}{n^{k/\log n}}, \quad \text{uniformly for} \quad k = o((\log n/\log \log n)^2).
\]
1.3. Multivariate functions. Let \( \mathbf{y} = [y_1, \ldots, y_m] = [x_1^n, \ldots, x_m^n] y(x_1, \ldots, x_m) \).

The saddle point method applies for admissible functions:

\[
y_1, \ldots, n_m \sim \frac{y(\rho_1, \ldots, \rho_m)}{\sqrt{(2\pi)^m} \sqrt{\det(\sigma_{ij})} \rho_1 \cdots \rho_m},
\]

with \( \frac{\partial}{\partial \sigma_{ij}} \log y(\rho_1, e^{u_1}, \ldots, \rho_m e^{u_m}) \bigg|_{u_0} = n_j \) and \( \sigma_{ij} = \frac{\partial^2}{\partial \sigma_{ij} \sigma_{kl}} \log y(\rho_1, e^{u_1}, \ldots, \rho_m e^{u_m}) \bigg|_{u_0} \).

For coefficients of powers of generating functions, the saddle point method also applies for

\[
[x_1^n, \ldots, x_m^n] y(x_1, \ldots, x_m)^k
\]

when \( k/n \in [a_j, b_j], a_j > 0 \) [4].

When \( k = o(n) \), singularity analysis can be used for some specialised functions \( y(x_1, \ldots, x_m) \). (See e.g. [7] for an illustration of a double Hankel contour.)

2. Generating Functions and Limit Distributions

To a class \( \mathcal{A} \) of combinatorial structures is associated the generating function \( a(x) = \sum a_n x^n \) (or \( a(x) = \sum a_n x^n / n! \) in a labelled context), where \( a_n \) is the number of \( \mathcal{A} \)-structures of size \( n \). The power of generating functions in combinatorics rests on the correspondence between classical combinatorial constructions and functional operators on generating functions. This mechanism is described by Flajolet in [15].

Additional parameters of structures, \( \phi_i : \mathcal{A} \to N, i = 1, \ldots, m \) can be handled with multivariate generating functions \( a(x, z) = \sum a_{nk} x^n z^k \), where \( z = (z_1, \ldots, z_m) \) with \( z^k = z_1^{k_1} \cdots z_m^{k_m} \), and \( a_{nk} \) is the number of \( \mathcal{A} \)-structures \( \alpha \) of size \( n \) such that \( \phi_i(\alpha) = k_i, i = 1, \ldots, m \).

Consider the (discrete) random variables \( X_n = (X_{n,1}, \ldots, X_{n,m}) \) with probability distribution \( \Pr(x_n = k) = a_{nk}/a_n \). The expected value \( E X_n = (E X_{n,1}, \ldots, E X_{n,m}) \) is given by \( E X_{n,j} = \frac{1}{a_n} [x^n] a_{z_j}(x, 1) \) and the covariance matrix \( \text{Cov} X_n = (E X_{n,j} X_{n,i} - E X_{n,j} E X_{n,i}) \) can be evaluated by \( E X_{n,j}^2 = \frac{1}{a_n} [x^n] (a_{z_j z_j}(x, 1) + a_{z_j}(x, 1)) \) and \( E X_{n,j} X_{n,i} = \frac{1}{a_n} [x^n] (a_{z_j z_i}(x, 1)) \).

There are three typical cases for the limiting distribution:

- \( X_n \) tends to a discrete distribution;
- \( X_n / \sqrt{\text{Cov}(X_n)} \) tends to a one-sided continuous distribution;
- \( (X_n - E X_n) / \sqrt{\text{Cov}(X_n)} \) tends to a normal distribution.

In each of these cases it is possible to find the limit distribution by use of the characteristic function \( \Phi_{X_n}(t) = E e^{itX_n} \). This approach gives the distribution function of the limit distribution (global limit theorem). A more accurate information can be obtained by explicit or uniform asymptotic formulae for the density \( a_{nk}/a_n \) (local limit theorem). The study of limiting distributions in combinatorial schemas, initiated by Bender [1], is being pursued by several authors [2, 3, 10].

3. Functional Equations

It is a well known result that the number of leaves in plane trees satisfies a Gaussian limit theorem. The generating function of plane trees is implicitly defined by \( a(x) = \frac{x}{1 - e^{-x}} \), and the bivariate generating function, with \( z \) marking the leaves is \( a(x, z) = z x z + \frac{x a(x, z)}{1 - a(x, z)} \).

The result of this section shows that more generally, in generating functions defined by functional equations, Gaussian limit distributions are to be expected.

**Theorem 1.** [5, 6] Let \( z = (z_1, \ldots, z_m), m \geq 1 \), and \( a(x, z) = \sum a_{nk} x^n z^k \) a generating function of nonnegative numbers \( a_{nk} \) satisfying a functional equation \( F(a, x, z) = 0 \), where \( F \) is analytic in the “range of interest” (plus some minor assumptions on \( a_{nk} \) and \( F \)).

Let \( x_0, a_0 > 0 \) be solutions \((x_0, a_0) \) of

\[
F(a_0, x_0, 1) = 0
\]

\[
F_a(a_0, x_0, 1) = 0
\]
and suppose that $F_{aa}(a_0, x_0, 1)F_x(a_0, x_0, 1) > 0$.

Then the numbers $a_{nk}$ satisfy a central limit theorem with mean value $E(X_n) = \mu n + O(1)$, where

$\mu = (\mu_1, \ldots, \mu_m)$ is given by $\mu_j = F_{\gamma_j}(a_0, x_0, 1) / x_0 F_x(a_0, x_0, 1)$.

Furthermore let $\rho = (\rho_1, \ldots, \rho_m)$, and $a_{\rho}, x_{\rho}, z_{\rho}$ be the solution of

$$
F(a_0, x_0, 1) = 0,
\quad F_{a}(a_0, x_0, 1) = 0,
\quad z_j F_{z_j}(a, x, z) = \rho_j x F_x(a, x, z),
$$

then the covariance matrix is given by $\text{Cov}\; X_n = \sigma^2_j(\rho) n + O(1)$, where $\sigma^2_j(\rho) = G_{ji}(a_{\rho}, x_{\rho}, z_{\rho})$ and $G_{ji}$ is expressed in terms of first and second derivatives of $F$ w.r.t. $a, x, z_j, z_l$.

Moreover let $\rho = \frac{b}{n} = \left(\frac{b_1}{n}, \ldots, \frac{b_m}{n}\right)$, then the numbers $a_{nk}$ satisfy a local limit theorem:

$$
a_{nk} = x_{\rho}^{-n} z_{\rho}^{-k} \cdot \frac{n - \frac{nk}{n+1}}{(2\pi)^{m/2} \sqrt{\det(\sigma^2_j(\rho))}} \sqrt{\frac{x_{\rho} F_x(a_{\rho}, z_{\rho}, z_{\rho})}{F_{aa}(a_{\rho}, x_{\rho}, z_{\rho})}} (1 + O(1/n)),
$$

uniformly for $k/n$ in a compact set containing only positive components.

**Proof.** The proof is in three steps. First the implicit function theorem gives a local representation of $a(x, z)$. Second extract the coefficient of $x^k$ in $a(x, z)$ by singularity analysis. And finally use saddle point method to obtain the coefficient of $x^n z^k$ in $a(x, z)$. □

An application of the theorem to independent subsets of simply generated trees is given by Drmota [5]: the number of independent subsets (a subset of a tree is independent if two different nodes are not adjacent) has a normal limit distribution with asymptotic mean value $n/3$; and the number of maximal independent subsets (an independent subset is maximal if any node not contained in it is adjacent to a node contained in it) has a normal limit distribution with asymptotic mean value $n/2$.

## 4. Product Schemas

The analysis of functional composition $F(uw(x))$, that translates into generating functions the combinatorial operation of substitution, has been largely investigated [1, 3, 10, 11]. It leads to discrete, or normal, or special distributions, according to analytic properties of $F$ and $w$. In the case of product schemas

$$
y(x, u) = g(x) F(uw(x))
$$

studied by Drmota and Soria [8], the limit distribution may be dictated either by $g(x)$, or by $F(uw(x))$, or it should involve both $g$ and $F$.

The analytic criterion to distinguish between these cases is singular order. Let $f_1(x)$ and $f_2(x)$ be analytic at the origin, with non negative coefficients, we say that $f_1$ is of higher singular order than $f_2$ if either the radius of convergence of $f_1$ is smaller than the one of $f_2$, or $f_1$ and $f_2$ have the same radius of convergence and the saddle point of $f_1$ is asymptotically smaller than the one of $f_2$.

We always assume that the coefficients of the Taylor expansions of $g(x)$, $w(x)$ and $F(w(x))$ can be evaluated, by saddle point method or singularity analysis.

### 4.1. $g$ is dominated

If $F(uw(x))$ is of higher singular order than $g(x)$, and $w(x)$ is of higher or equal singular order than $g(x)$, then $g(x)$ is (usually) dominated in $y(x, u)$.

When $g$ is dominated, the factor $g(x)$ has actually no influence over the limit distribution, i.e. the limit distribution of $y(x, u)$ is the same as the limit distribution of $F(uw(x))$. 

83
III Asymptotic Analysis

4.2. $g$ dominates. If $g(x)$ is of higher singular order than $F(w(x))$ and $F'(w(x))$, then $g(x)$ (usually) dominates in $g(x, u)$.

When $g$ dominates there are only a few kinds of limit distributions, which can be classified in the following way.

- If \( \lim_{x \to u} w(x) = w(r) \) exists and $F(z)$ is regular at $w(r)$, then $X_n$ (related to $g(x, u)$) has a discrete limit distribution.
- If \( \lim_{x \to u} w(x) = \infty \) and $F(z)$ is admissible, then $X_n$ is asymptotically normally distributed.
- If \( \lim_{x \to u} w(x) = w(r) \) exists and $F(z)$ is singular at $z = w(r)$
  * if $F(z)$ is admissible then $X_n$ is asymptotically normally distributed.
  * if $F(z)$ has an algebraic-logarithmic singularity, then $X_n$ is asymptotically Gamma distributed.

Bibliography


