Successes and Failures of Analytic Methods in Asymptotic Enumeration

A. M. Odlyzko
AT&T Bell Laboratories
Murray Hill, New Jersey 07974

Extended Abstract

Analytic methods give a powerful method for obtaining asymptotic estimates in combinatorial enumeration. When they can be used, they usually provide extremely precise results. However, there are also many situations where they do not apply, and one has to use elementary or probabilistic reasoning. Problems in various classes will be presented, and general principles of what kinds of methods are best for various situations will be discussed.

1 Introduction

Analytic methods provide extremely powerful tools for asymptotic enumeration. Their need is steadily being extended by new research. However, there are also many cases where analytic methods have failed to yield useful information, even when it seemed that they ought to apply. The goal of this paper is to illustrate both successes and failures of analytic methods, indicate where additional research is needed, and draw some conclusions about the applicability of such methods to various problems.

The most frequent and obvious reason for failure of analytic methods to yield enumeration information is the lack of a useful analytic generating function. Of course, for any sequence $a_0, a_1, \ldots$, we can find another explicit sequence $b_0, b_1, \ldots$, such that the generating function

$$f(z) = \sum_{n=0}^{\infty} a_n b_n^{-1} z^n \tag{1.1}$$

is analytic near $z = 0$. However, what is needed for a function to be useful for asymptotic enumeration is for it to be in a form that can be used to deduce the analytic behavior of that function. Usually this requires a generating function (typically ordinary or exponential one) that reflects the combinatorial structure of the sequence that is being enumerated. In many situations no such generating function is known, and so it is no surprise that analytic methods do not apply.

This paper will concentrate on situations where an explicit analytic generating function is known, but there are difficulties in exploiting this feature. We first briefly review the standard analytic methods. We then illustrate with a series of examples of recent successes and failures of such techniques.

The purpose of this note is not to present a general introduction to the use of analytic methods in combinatorial enumeration and analysis of algorithms. That is done in the author's recent survey and tutorial [36]. This note was inspired and is to some extent based on that work. There are many other sources for basic asymptotic methods, such as [1, 2, 3, 7, 8, 9, 10, 22, 23, 24, 27, 28, 30, 31, 32, 33, 42, 43].

2 Standard analytic methods

When a sequence has an explicit single-variable generating function, there is a wealth of techniques that usually suffice to determine the asymptotics of that sequence with little effort. The coefficients of generating functions $f(z)$ arising in combinatorics or analysis of algorithms are usually nonnegative. Hence there is typically a single singularity at $z = x_0 \in R$ on the circle of convergence of

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \tag{2.2}$$

whose influence dominates the behavior of a_n. It is referred to as the *dominant singularity*. If it is the only singularity on $z = x_0$, everything simplifies. When there are other singularities, the situation is more complicated but often still tractable.

There are two basic classes of methods that apply to single-variables analytic generating functions. If the dominant singularity is small, so that $f(z)$ in an appropriate neighborhood of $z = x_0$ behaves like $(z - x_0)^\alpha$ or $(\log(z - \alpha))^\beta$, then transfer methods are very effective. For example, the generating function of 2-regular graphs is known [4] to be

$$f(z) = (1 - z)^{-1/2} \exp(-z/2 - z^2/4) , \tag{2.3}$$

so that the dominant singularity is at $z = 1$ and there are no other singularities on $z = 1$. Transfer theorems, such as those of [16] immediately show that

$$a_n \sim \pi^{-1/2} \exp(-3/4) n^{-1/2} \quad \text{as} \quad n \to \infty . \tag{2.4}$$

If the dominant singularity is large, so that $f(z)$ blows up rapidly as z approaches x_0 along the real axis from the left, then another class of methods, based largely on the saddle point method, are most productive. A surprisingly large number of problems can be solved by using theorems of Hayman [26], which present conditions that guarantee that the saddle point estimates do apply. For example, the Bell numbers B_n, which have the exponential generating function

$$B(z) = \sum_{n=0}^{\infty} B_n \frac{z^n}{n!} = \exp(\exp(z) - 1) , \tag{2.5}$$

satisfy

$$B_n \sim n!(2\pi x_0^2 \exp(x_0))^{-1/2} \exp(\exp(x_0) - 1) x_0^{-n} \quad \text{as} \quad n \to \infty , \tag{2.6}$$

where $x_0 \exp(x_0) = n$. The estimate (2.6) can be derived immediately from the results of [26].

These two basic methods for determining the asymptotics of sequences from univariate generating functions suffice for most examples in books such as [4]. However, there are many problems where more sophisticated approaches are needed.

3 Recent successes and failures

Single variable asymptotics are well understood. That is not the case when multivariate generating functions are required. Both transfer theorems and saddle point methods can be generalized (see [36]), but their applicability is more limited than in the univariate case. However, there have been remarkable successes in extending some of these methods even to cases where the number of variables in the problem grows rapidly. A good example comes from the work of McKay and Wormald [34, 35].

Example 3.1 *Simple labeled graphs of high degree.* Let $G(n; d_1, \ldots, d_n)$ be the number of labeled simple graphs on n vertices with degree sequence $d_1, d_2, \ldots, d_n$. Then

$G(n; d_1, \ldots, d_n)$ is the coefficient of $z_1^{d_1} z_2^{d_2} \cdots z_n^{d_n}$ in

$$F = \prod_{\substack{j,k=1 \\ j<k}}^{n} (1 + z_j z_k) , \tag{3.7}$$

and so by Cauchy's theorem

$$G(n; d_1, \ldots, d_N) = (2\pi i)^{-n} \int \cdots \int F z_1^{-d_1-1} \cdots z_n^{-d_n-1} dz_1 \cdots dz_n , \tag{3.8}$$

where each integral is on a circle centered at the origin. The difficulty of this problem arises from the large number of variables that are used. The general principle of the saddle point method is to choose an appropriate contour of integration so that a small region of the space of integration gives the dominant contribution to the integral. This method, which is well understood for a single variable, can also be extended easily to a fixed number of variables. When the number of variables increases, though, formidable difficulties arise, and it was a great achievement for McKay and Wormald to overcome all the technical problems. Let all the radii be equal to some $r > 0$. The integrand takes on its maximum absolute value on the product of these circles at precisely the two points $z_1 = z_2 = \cdots = z_n = r$ and $z_1 = z_2 = \cdots = z_n = -r$. If $d_1 = d_2 = \cdots = d_n$, so that we consider only regular graphs, McKay and Wormald [35] show that for an appropriate choice of the radius r, these two points are actually saddle points of the integrand, and succeed through careful analysis in proving that if dn is even, and $\min(d, n-d-1) > cn(\log n)^{-1}$ for some $c > 2/3$, then

$$G(n, d, d, \ldots, d) = 2^{1/2}(2\pi n \lambda^{d+1}(1-\lambda)^{n-d})^{-n/2} \exp\left(\frac{-1 + 10\lambda - 10\lambda^2}{12\lambda(1-\lambda)} + O(n^{-\zeta})\right) \tag{3.9}$$

as $n \to \infty$ for any $\zeta < \min(1/4, 1/2 - 1/(3c))$, where $\lambda = d/(n-1)$. □

Even univariate problems still present challenges. Some of the problems arise when the relevant generating functions do not have explicit forms, but are defined by recursions.

Example 3.2 *Heights of binary trees.* We let B_n denote the number of binary trees of size n, so that $B_0 = 1$ (by convention), $B_1 = 1$, $B_2 = 2$, $B_3 = 5, \ldots$. Let

$$B(z) = \sum_{n=0}^{\infty} B_n z^n . \tag{3.10}$$

Since each nonempty binary tree consists of the root and two binary trees (the left and right subtrees), we obtain the functional equation

$$B(z) = 1 + zB(z)^2 . \tag{3.11}$$

This implies that

$$B(z) = \frac{1 - (1-4z)^{1/2}}{2z} , \tag{3.12}$$

so that

$$B_n = \frac{1}{n+1}\binom{2n}{n} , \tag{3.13}$$

and the B_n are the Catalan numbers. Stirling's formula then shows that

$$B_n \sim \pi^{-1/2} n^{-3/2} 4^n \quad \text{as} \quad n \to \infty . \tag{3.14}$$

Let $B_{h,n}$ be the number of binary trees of size n and height $\leq h$, and let

$$b_h(z) = \sum_{n=0}^{\infty} B_{h,n} z^n . \tag{3.15}$$

Then

$$b_0(z) = 0, \quad b_1(z) = 1 , \tag{3.16}$$

and an extension of the argument that led to the relation (3.11) yields

$$b_{h+1}(z) = 1 + z b_h(z)^2 , \quad h \geq 0 . \tag{3.17}$$

The $b_h(z)$ are polynomials in z of degree $2^{h-1} - 1$ for $h \geq 1$. Unfortunately there is no simple formula for them like Eq. (3.12) for $B(z)$, and one has to work with the recurrence (3.17) to obtain most of the results about heights of binary trees. Different problems involve study of the recurrence in different ranges of values of z, and the behavior of the recurrence varies drastically.

For any fixed z with $z \leq 1/4$, $b_h(z) \to B(z)$ as $h \to \infty$. For $z > 1/4$ the behavior of $b_h(z)$ is more complicated, and is a subject of of nonlinear dynamics. For any real z with $z > 1/4$, $b_h(z) \to \infty$ as $h \to \infty$. To study the distribution of the $B_{h,n}$ as n varies for h fixed, but large, it is necessary to investigate this range of rapid growth. It can be shown [14] that for any λ_1 and λ_2 with $0 < \lambda_1 < \lambda_2 < 1/2$,

$$B_{h,n} = \frac{\exp(2^{h-1}(\beta(r) - r\beta'(r)\log r))}{2^{(h-1)/2}(2\pi(r^2\beta''(r) + r\beta'(r)))^{1/2}}(1 + O(2^{-h/2})) \tag{3.18}$$

uniformly as $h, n \to \infty$ with

$$\lambda_1 < n/2^h < \lambda_2 . \tag{3.19}$$

The function $\beta(x)$ is defined for $1/4 < x < \infty$ by

$$\beta(x) = \log x + \sum_{j=1}^{\infty} 2^{-j} \log \left(1 + \frac{1}{b_j(x) - 1}\right) , \tag{3.20}$$

and r is the unique solution in $(1/4, \infty)$ to

$$r\beta'(r) = n2^{-h+1} . \tag{3.21}$$

The proof of the estimate (3.18) is derived from the estimate

$$b_h(z) = \exp(2^{h-1}\beta(z) - \log z)(1 + O(\exp(-\epsilon 2^h))) , \tag{3.22}$$

valid in a region along the half-axis $x > 1/4$. The estimates for the coefficients $B_{h,n}$ are obtained by applying the saddle point method.

The methods that are used to study the average height are different from those used for trees of a fixed height. The basic approach of [13] is to let

$$H_n = \sum_{\substack{T \\ T=n}} ht(T) ,$$

where the sum is over the binary trees T of size n, and $ht(T)$ is the height of T. Then the average height is just H_n/B_n. The generating function for the H_n is

$$H(z) = \sum_{n=0}^{\infty} H_n z^n = \sum_{h \geq 0} (B(z) - b_h(z)) , \tag{3.23}$$

and the analysis of [13] proceeds by investigating the behavior of $H(z)$ near $z = 1/4$. If we let

$$\epsilon(z) = (1 - 4z)^{1/2} , \tag{3.24}$$

$$e_h(z) = (B(z) - b_h(z))/(2B(z)) , \tag{3.25}$$

then the recurrence (3.17) yields

$$e_{h+1}(z) = (1 - \epsilon(z))e_h(z)(1 - e_h(z)) , \quad e_0(z) = 1/2 . \tag{3.26}$$

Extensive analysis of this relation yields an approximation to $e_h(z)$ of the form

$$e_h(z) \approx \frac{\epsilon(z)(1 - \epsilon(z))^h}{1 - (1 - \epsilon(z))^h} , \tag{3.27}$$

valid for $\epsilon(z)$ sufficiently small, Arg $\epsilon(z) < \pi/4 + \delta$ for a fixed $\delta > 0$. (The precise error terms in this approximation are complicated, and are given in [13].) This then leads to an expansion for $H(z)$ in a sector $z - 1/4 < \alpha$, $\pi/2 - \beta <$ Arg$(z - 1/4) < \pi/2 + \beta$ of the form

$$H(z) = -2\log(1 - 4z) + K + O(1 - 4z^v) , \tag{3.28}$$

where v is any constant, $v < 1/4$, and K is a fixed constant. Transfer theorems then yield the asymptotic estimate

$$H_n \sim 2n^{-1}4^n \quad \text{as} \quad n \to \infty . \tag{3.29}$$

When we combine (3.29) with (3.14), we obtain the desired result that the average height of a binary tree of size n is $\sim 2(\pi n)^{1/2}$ as $n \to \infty$.

For extremely small and large heights, different methods are used. It follows from [11] that

$$\frac{B_{h,n} - B_{h-1,n}}{B_n} \leq \exp(-c(h^2/n + n/h^2)) \tag{3.30}$$

for a constant $c > 0$, which shows that extreme heights are infrequent. (The estimates in [11] are more precise than (3.30).) Bounds of the above form for small heights are obtained in [11] by studying the behavior of the $b_h(z)$ almost on the boundary between convergence and divergence. Let x_h be the unique positive root of $b_h(z) = 2$. Note that $B(1/4) = 2$, and each coefficient of the $b_h(z)$ is nondecreasing as $h \to \infty$. Therefore $x_2 > x_3 > \cdots > 1/4$. More effort shows [11] that x_h is approximately $1/4 + \alpha h^{-2}$ for a certain $\alpha > 0$. This leads to an upper bound for $B_{h,n}$. Bounds for trees of large heights are even easier to obtain, since they only involve upper bounds for the $b_h(z) - b_{h-1}(z)$ inside the disk of convergence $z < 1/4$. □

There are some univariate generating functions which so far have not yielded to analytic approaches.

Example 3.3 *Heights of random binary search trees.* Example 3.2 is set in the standard combinatorial counting model, in which all trees of a given size are counted equally. In many applications it is desirable to have different weights for different trees. For example, if random permutations are used to construct binary search trees, then the two trees of maximal heights will have probability of occurring of $1/n!$ each, whereas more balanced trees will have exponentially larger probabilities. The average height in this model turns out to be $\sim c \log n$ as $n \to \infty$, where $c = 4.311\ldots$ is a certain constant. This was proved by Devroye [5, 6] using branching processes methods. (See also [33].) Generating function approaches lead to functional equations of the type

$$f(z) = 1 + \int_0^z f(u)^2 du \; , \tag{3.31}$$

which have not been solved so far. It would be nice to develop the analytic generating function approach to this problem, since it might than be used to obtain new information about the distribution of heights, for example. □

There are some problems with rather simple generating functions where analytic methods fail, but not because of the complexity of the function, but rather because of the basic limitations of known analytic methods. This occurs in a set partition problem described below.

Example 3.4 *Set partitions with distinct block sizes.* Let a_n be the number of partitions of a set of n elements into blocks of distinct sizes. Then $a_n = b_n \cdot n!$, where $b_n = [z^n]f(z)$, with

$$f(z) = \prod_{k=1}^{\infty} \left(1 + \frac{z^k}{k!}\right) \; . \tag{3.32}$$

The function $f(z)$ is entire and has nonnegative coefficients, so it might appear as an ideal candidate for an application of the methods for dealing with large singularities, such as the saddle point technique. However, on circles $z = (n + 1/2)/e$, $n \in Z^+$, $f(z)$ does not vary much, so there are technical problems in applying these analytic methods. On the other hand, combinatorial estimates can be used to show [29] that the b_n behave in a "regularly irregular" way, so that, for example,

$$b_{m(m+1)/2-1} \sim b_{m(m+1)/2} \quad \text{as} \quad m \to \infty \; , \tag{3.33}$$

$$b_{m(m+1)/2} \sim m b_{m(m+1)/2+1} \quad \text{as} \quad m \to \infty \; . \tag{3.34}$$

The term $b_n z^n$ for $n = m(m+1)/2$ for example, comes almost entirely from the product of $z^k/k!$, $1 \le k \le m$, all other products contributing an asymptotically negligible amount. There do not seem to be any analytic methods that obtain these results without essentially redoing the combinatorial estimates. □

References

[1] E. A. Bender, Asymptotic methods in enumeration, *SIAM Review, 16* (1974), pp. 485–515.

[2] E. A. Bender and S. G. Williamson, *Foundations of Applied Combinatorics*, Addison-Wesley, 1991.

[3] N. G. de Bruijn, *Asymptotic Methods in Analysis*, North-Holland, Amsterdam, 1958.

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

[5] L. Devroye, A note on the expected height of binary search trees, *J. ACM, 33* (1986), pp. 489–498.

[6] L. Devroye, Branching processes in the analysis of the heights of trees, *Acta Informatica, 24* (1987), pp. 277–298.

[7] G. P. Egorychev, *Integral Representation and the Computation of Combinatorial Sums*, Amer. Math. Soc. 1984.

[8] A. Erdélyi, *Asymptotic Expansions*, Dover reprint, 1956.

[9] M. V. Fedoryuk, Asymptotic methods in analysis, pp. 83–191 in *Analysis I*, R. V. Gamkrelidze, ed., Springer 1989.

[10] P. Flajolet, Analytic analysis of algorithms, *Proc. ICALP '92*, Springer Lecture Notes in Computer Science, 1992, to be published.

[11] P. Flajolet, Z. Gao, A. M. Odlyzko, and B. Richmond, The height of binary trees and other simple trees, *Combinatorics, Probability, and Computing* (1993), to appear.

[12] P. Flajolet and T. Lafforgue, Search costs in quadtrees and singularity perturbation asymptotics, to be published.

[13] P. Flajolet and A. M. Odlyzko, The average height of binary trees and other simple trees, *J. Comput. System Sci., 25* (1982), pp. 171–213.

[14] P. Flajolet and A. M. Odlyzko, Limit distributions for coefficients of iterates of polynomials with application to combinatorial enumeration, *Math. Proc. Cambridge Phil. Soc., 96* (1984), pp. 237–253.

[15] P. Flajolet and A. M. Odlyzko, Random mapping statistics, in *Advances in Cryptology: Proceedings of Eurocrypt '89*, J-J. Quisquater, ed., Springer Lecture Notes in Computer Science, *434* (1990), pp. 329–354.

[16] P. Flajolet and A. M. Odlyzko, Singularity analysis of generating function, *SIAM J. Discrete Math., 3* (1990), pp. 216–240.

[17] P. Flajolet, B. Salvy, and P. Zimmermann, Automatic average–case analysis of algorithms, *Theoretical Computer Science, 79* (1991), 37–109.

[18] P. Flajolet and M. Soria, Gaussian limiting distributions for the number of components in combinatorial structures, *J. Combinatorial Theory,* Series A, *53* (1990), pp. 165–182.

[19] P. Flajolet and M. Soria, The cycle construction, *SIAM J. Discrete Math., 3* (1990).

[20] G. H. Gonnet and R. Baeza-Yates, *Handbook of Algorithms and Data Structures*, 2nd ed., Addison-Wesley, 1991.

[21] I. Goulden and D. Jackson, *Combinatorial Enumeration*, John Wiley, New York, 1983.

[22] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison Wesley, 1989.

[23] D. H. Greene and D. E. Knuth, *Mathematics for the Analysis of Algorithms*, 2nd ed., Birkhäuser, Boston, 1982.

[24] F. Harary and E. M. Palmer, *Graphical Enumeration*, Academic Press, 1973.

[25] F. Harary, R. W. Robinson, and A. J. Schwenk, Twenty–step algorithm for determining the asymptotic number of trees of various species, *J. Austral. Math. Soc.* (Series A), *20* (1975), pp. 483–503.

[26] W. K. Hayman, A generalization of Stirling's formula, *J. reine angew. Math., 196* (1956), pp. 67–95.

[27] M. Hofri, *Probabilistic Analysis of Algorithms*, Springer, 1987.

[28] R. Kemp, *Fundamentals of the Average Case Analysis of Particular Algorithms*, Wiley, 1984.

[29] A. Knopfmacher, A. Odlyzko, B. Richmond, G. Szekeres, and N. Wormald, manuscript in preparation.

[30] D. E. Knuth, *The Art of Computer Programming Vol. 1: Fundamental Algorithms*, 2nd ed., Addison–Wesley, Reading, 1973.

[31] D. E. Knuth, *The Art of Computer Programming Vol. 2: Semi-Numerical Algorithms*, 2nd ed., Addison–Wesley, Reading, 1981.

[32] D. E. Knuth, *The Art of Computer Programming Vol. 3: Sorting and Searching*, Addison–Wesley, Reading, 1973.

[33] H. S. Mahmoud, *Evolution of Random Search Trees*, Wiley, 1992.

[34] B. D. McKay, The asymptotic numbers of regular tournaments, eulerian digraphs, and eulerian and oriented graphs, *Combinatorica, 10* (1990), 367–377.

[35] B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs of high degree, *European J. Combinatorics, 11* (1990), 565–580.

[36] A. M. Odlyzko, Asymptotic enumeration methods, *Handbook of Combinatorics*, in R. L. Graham, M. Grötschel, and L. Lovász, eds., North-Holland, to appear.

[37] A. M. Odlyzko, B. Poonen, H. Widom, and H. S. Wilf, manuscript in preparation.

[38] A. Regev, Asymptotic values for degrees associated with strips of Young diagrams, *Advances Math., 41* (1981), 115–136.

[39] J. Riordan, *Introduction to Combinatorial Analysis*, John Wiley, New York, 1958.

[40] J. Riordan, *Combinatorial Identities*, Wiley, 1968.

[41] R. P. Stanley, *Enumerative Combinatorics*, vol. 1, Wadsworth and Brooks/Cole, Monterey, 1986.

[42] J. Vitter and P. Flajolet, Analysis of algorithms and data structures, *Handbook of Theoretical Computer Science*, Vol. A: Algorithms and Complexity, Ch. 9, pp. 432–524, J. Van Leeuwen, ed., North Holland, 1990.

[43] H. S. Wilf, *Generatingfunctionology*, Academic Press, 1990.

[44] J. Wimp, Current trends in asymptotics: some problems and some solutions, *J. Computational Appl. Math., 35* (1991), 53–79.