Chapter 1

Introduction

1.1 Arrays of numbers

The main subject of these lecture notes is an array of numbers

$$\{a_{r_1,\ldots,r_d} : r_1, \ldots, r_d \in \mathbb{Z}^+\}\,.$$

This will usually be written as $\{a_{\mathbf{r}} : \mathbf{r} \in (\mathbb{Z}^+)^d\}$, where $\mathbb{Z}^+$ is understood to start with zero: $\mathbb{Z}^+ := \{0, 1, 2, \ldots\}$. The numbers $a_{\mathbf{r}}$ may be integers, real numbers or even complex numbers. As a notational aside, we will always use d to denote the dimension of the array. The variables r, s and t are reserved as synonyms for r_1, r_2 and r_3, respectively, so as to avoid subscripts in examples of dimensions up to three.

The numbers $a_{\mathbf{r}}$ usually come with a story — a reason they are interesting. Often they count a class of objects parametrized by $\mathbf{r}$. For example, it could be that $a_{\mathbf{r}}$ is the multinomial coefficient $a_{\mathbf{r}} := \binom{\sum_{j=1}^d r_j}{r_1 \cdots r_d}$, in which case $a_{\mathbf{r}}$ counts sequences with r_1 ones, r_2 twos, and so forth up to r_d occurrences of the symbol d. Another frequent source of these arrays is in probability theory. Here, the numbers $a_{\mathbf{r}} \in [0, 1]$ are probabilities of events parametrized by $\mathbf{r}$. For example, a_{rs} might be the probability that a simple random walk of r steps ends at the integer point s.

How might one understand an array of numbers? There might be a simple, explicit formula. The multinomial coefficients, for example, are given by ratios of factorials. As Richard Stanley[1] points out in the introduction to [Sta97], a formula of this brevity seldom exists; when it does, we don't need fancy techniques to describe the array. Often, if a formula exists at all, it will not be in closed form but will have a summation in it. As Stanley says, "There are actually formulas in the literature (nameless here forevermore) for certain counting functions whose evaluation requires listing all of the objects being counted! Such a 'formula' is completely worthless." Less egregious are the formulæ

[1]The next few paragraphs are adapted from Stanley – see the notes to this chapter.

containing functions that are rare or complicated and whose properties are not immediately familiar to us. It is not clear how much good it does to have this kind of formula.

Another way of describing arrays of numbers is via recursions. The simplest recursions are finite linear recursions, such as the recursion

$$a_{r,s} = a_{r-1,s} + a_{r,s-1}$$

for the binomial coefficients. A recursion for $a_{\mathbf{r}}$ in terms of values $\{a_{\mathbf{s}} : \mathbf{s} < \mathbf{r}\}$ whose indices precede $\mathbf{r}$ in the coordinatewise partial order may be pretty unwieldy, perhaps requiring evaluation of a complicated function of all $a_{\mathbf{s}}$ with $\mathbf{s} < \mathbf{r}$. But if the recursion is of bounded complexity, such as a linear recursion $a_{\mathbf{r}} = \sum_{\mathbf{j} \in F} c_{\mathbf{j}} a_{\mathbf{r}-\mathbf{j}}$ for some finite set $\{c_{\mathbf{j}} : \mathbf{j} \in F\}$ of constants, then the recursion gives a polynomial time algorithm for computing $a_{\mathbf{r}}$. Still, we will see that even in this case, the estimation of $a_{\mathbf{r}}$ is not at all straightforward. Thus, while we look for recursions to help us understand number arrays, recursions rarely provide definitive descriptions.

A third way of understanding an array of numbers is via an estimate. If one uses Stirling's formula

$$n! \sim \frac{n^n}{e^n} \sqrt{2\pi n}\,,$$

one obtains an estimate for binomial coefficients

$$a_{r,s} \sim \left(\frac{r+s}{r}\right)^r \left(\frac{r+s}{s}\right)^s \sqrt{\frac{r+s}{2\pi r s}} \tag{1.1}$$

and a similar estimate for multinomial coefficients. If number theoretic properties of $a_{\mathbf{r}}$ are what you want, then you are better off sticking with the formula $(r+s)!/(r!\,s!)$, but when the approximate size of $a_{\mathbf{r}}$ is paramount, then the estimate (1.1) is better.

A fourth way to understand an array of numbers is to give its generating function. The **generating function** for the array $\{a_{\mathbf{r}}\}$ is the series $F(\mathbf{z}) := \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$. Here $\mathbf{z}$ is a d-dimensional vector of indeterminates $(z_1, \ldots, z_d)$, and $\mathbf{z}^{\mathbf{r}}$ denotes the monomial $z_1^{r_1} \cdots z_d^{r_d}$. In our running example of multinomial coefficients, the generating function

$$F(\mathbf{z}) = \sum_{\mathbf{r}} \binom{\sum_{j=1}^d r_j}{r_1 \,\cdots\, r_d} z_1^{r_1} \cdots z_d^{r_d}$$

is written more compactly as $1/(1 - r_1 - \cdots - r_d)$. Stanley calls the generating function "the most useful but the most difficult to understand" method for describing an array.

One reason a generating function is useful is that the algebraic form of the function is intimately related to recursions for $a_{\mathbf{r}}$ and combinatorial decompositions for the objects enumerated by $a_{\mathbf{r}}$. Another reason is that estimates (and exact formulæ if they exist) may be extracted from a generating function. In other words, formulæ recursions and estimates all ensue once a generating function is known.

1.2 Generating functions and asymptotics

I will employ the usual asymptotic notation, as follows. If f, g are real valued functions then the statement "$f = O(g)$" is shorthand for the statement "$\limsup_{x \to x_0} |f(x)|/|g(x)| < \infty$". It must be made clear at which value, x_0, the limit is taken; if f and g depend on parameters other than x, it must also be made clear which is the variable being taken to the limit. Most commonly, $x_0 = +\infty$; in the statement $a_n = O(g(n))$, the limit is always taken at infinity. The statement "$f = o(g)$" is shorthand for $f(x)/g(x) \to 0$, again with the limiting value of x specified. Lastly, the statement "$f \sim g$" means $f/g \to 1$ and is equivalent to "$f = (1 + o(1)) \cdot g$" or "$f - g = o(g)$"; again the variable and its limiting value must be specified. Two more useful notations are $f = \Omega(g) \Leftrightarrow g = O(f)$, and $f = \Theta(g) \Leftrightarrow$ both $f = O(g)$ and $g = O(f)$. An **asymptotic expansion**

$$f \sim \sum_{j=0}^{\infty} g_j$$

for a function f in terms of a sequence $\{g_j : j \in \mathbb{Z}^+\}$ satisfying $g_{j+1} = o(g_j)$ is said to hold if for every M, $f - \sum_{j=0}^{M-1} g_j = O(g_M)$. This is equivalent to $f - \sum_{j=0}^{M-1} g_j = o(g_{M-1})$. A function is said to be **rapidly decreasing** if it is $O(x^{-n})$ at ∞ for every n, **exponentially decaying** if it is $O(e^{-\gamma n})$ for some $\gamma > 0$ and **super-exponentially decaying** if it is $O(e^{-\gamma n})$ for every $\gamma > 0$.

All these notations hold in the multivariate case as well, except that if the limit value of $\mathbf{z}$ is infinity, then a statement such as $f(\mathbf{z}) = O(g(\mathbf{z}))$, must also specify how $\mathbf{z}$ approaches the limit. Our chief concern will be with the asymptotics of $a_{\mathbf{r}}$ as $\mathbf{r} \to \infty$ in a given direction. More specifically, by a **direction**, I mean an element $\overline{\mathbf{r}}$ of $(d-1)$-dimensional projective space whose class contains a d-tuple of positive real numbers. It turns out that a typical asymptotic formula for $a_{\mathbf{r}}$ is $C|\mathbf{r}|^\alpha \mathbf{z}^{-\mathbf{r}}$ where $|\mathbf{r}|$ is the sum of the coordinates of $\mathbf{r}$, and the d-tuple $\mathbf{z}$ and the multiplicative constant C depend on $\mathbf{r}$ only through its projection $\overline{\mathbf{r}}$. In hindsight, formulæ such as these make it natural to consider $\mathbf{r}$ projectively and take $\mathbf{r}$ to infinity in prescribed directions.

When I quoted Stanley, I should have mentioned that he was talking chiefly about univariate arrays, that is, the case $d = 1$. As will be seen in Chapter 3, it is indeed true that the generating function $f(z)$ for a univariate sequence $\{a_n : n \in \mathbb{Z}^+\}$ leads, almost automatically, to asymptotic estimates for a_n as $n \to \infty$. [Another notational aside: I will use $f(z)$ and a_n instead of $F(z)$ and a_r in one variable, so as to coincide with notation in the univariate literature.]

To estimate a_n when f is known, begin with Cauchy's integral formula:

$$a_n = \frac{1}{2\pi i} \int z^{-n-1} f(z)\, dz\,. \tag{1.2}$$

The integral is a complex contour integral on a contour encircling the origin, and one may apply complex analytic methods to estimate the integral. The necessary knowledge of residues and contour shifting may be found in an introductory complex variables text such as [Con78, BG91], although one obtains a better idea of univariate saddle point integration from [Hen88, Hen91].

The situation for multivariate arrays is nothing like the situation for univariate arrays. In 1974, when Bender published the review article [Ben74] on asymptotic enumeration, the asymptotics of multivariate generating functions was largely a gap in the literature. Bender's concluding section urges research in this area:

> Practically nothing is known about asymptotics for recursions in two variables even when a generating function is available. Techniques for obtaining asymptotics from bivariate generating functions would be quite useful.

In the 1980's and 1990's, a small body of results was developed by Bender, Richmond, Gao and others, giving the first partial answers to questions of asymptotics of generating functions in the multivariate setting. The first paper to concentrate on extracting asymptotics from multivariate generating functions was [Ben73], already published at the time of Bender's survey, but the seminal paper is [BR83]. The hypothesis is that F has a singularity of the form $A/(z_d - g(\mathbf{x}))^q$ on the graph of a smooth function g, for some real exponent q, where $\mathbf{x}$ denotes $(z_1, \ldots, z_{d-1})$. They show, under appropriate further hypotheses on F, that the probability measure μ_n one obtains by renormalizing $\{a_\mathbf{r} : r_d = n\}$ to sum to 1 converges to a multivariate normal when appropriately rescaled. Their method, which we call the **GF-sequence method**, is to break the d-dimensional array $\{a_\mathbf{r}\}$ into a sequence of $(d-1)$-dimensional slices and consider the sequence of $(d-1)$-variate generating functions

$$f_n(\mathbf{x}) = \sum_{\mathbf{r}: r_d = n} a_\mathbf{r} \mathbf{z}^\mathbf{r} \, .$$

They show that, asymptotically as $n \to \infty$,

$$f_n(\mathbf{x}) \sim C_n g(\mathbf{x}) h(\mathbf{x})^n \tag{1.3}$$

and that sequences of generating functions obeying (1.3) satisfy a central limit theorem and a local central limit theorem.

These results always produce Gaussian (central limit) behavior. The applicability of the entire GF-sequence method is limited to the single, though important, case where the coefficients $a_\mathbf{r}$ are nonnegative and possess a Gaussian limit. The work of [BR83] has been greatly expanded upon, but always in a similar framework. For example, it has been extended to matrix recursions [BRW83] and the applicability has been extended from algebraic to algebraico-logarithmic singularities of the form $F \sim (z_d - g(\mathbf{x}))^q \log^\alpha(1/(z_d - g(\mathbf{x})))$ [GR92]. The difficult step is always deducing asymptotics from the hypotheses $f_n \sim C_n \, g \cdot h^n$. Thus some papers in this stream refer to such an assumption in their titles [BR99], and the term "quasi-power" has been coined for such a sequence $\{f_n\}$.

1.3 New multivariate methods

On a personal note, my interest in this problem began when I had, with some effort, found a generating function for a particular bivariate array of probabilities that were of interest to me. I then consulted the literature, sure that I would find results, well known and neatly packaged, that gave asymptotic estimates for the probabilities. At that time, the most recent and complete reference on asymptotic enumeration was Odlyzko's 1995 survey [Odl95]. Only six of its over one hundred pages are devoted to multivariate asymptotics, mainly to the GF-sequence results of Bender *et al*, and this section closes with a call for further work in this area. After several iterations of replacing my generating function by simpler toy problems, I became convinced that a general asymptotic formula or method was not known, even for the simplest imaginable class, namely rational functions. This stands in contrast to the univariate theory of rational functions, which is trivial (see Chapter 3). The relative difficulty of the problem in higher dimensions surprised me, but the connections to other areas of mathematics such as Morse theory were intriguing and caused me to pursue this line of research long after my interest in the original array of probabilities had faded.

Odlyzko describes why he believes multivariate coefficient estimation to be difficult. First, the singularities are no longer isolated, but form $(d-1)$-dimensional hypersurfaces. Thus, he points out, "Even rational multivariate functions are not easy to deal with." Secondly, the multivariate analogue of the one-dimensional residue theorem is the considerably more difficult theory of Leray [Ler50]. This theory was later fleshed out by Aizenberg and Yuzhakov, who spend a few pages [AY83, Section 23] on generating functions and combinatorial sums. Further progress in using multivariate residues to evaluate coefficients of generating functions was made by Bertozzi and McKenna [BM93], though at the time of Odlyzko's survey none of the papers based on multivariate residues such as [Lic91, BM93] had resulted in any kind of systematic application of these methods to enumeration.

The focus of these lecture notes is a recent vein of research, begun in [PW02] and continued in [PW04, BP04, Lla03, Lla05, PW07, BP08] and in several manuscripts in progress. This research extends ideas that are present to some degree in [Lic91, BM93], using complex methods that are genuinely multivariate to evaluate coefficients via the multivariate Cauchy formula

$$a_{\mathbf{r}} = \left(\frac{1}{2\pi i}\right)^d \int_T \mathbf{z}^{-\mathbf{r}-\mathbf{1}} F(\mathbf{z})\, d\mathbf{z}\,. \tag{1.4}$$

By avoiding symmetry-breaking decompositions such as $F = \sum f_n(z_1, \ldots, z_{d-1}) z_d^n$, one hopes the methods will be more universally applicable and the formulæ more canonical. In particular, the results of Bender *et al.* and the results of Bertozzi and McKenna are seen to be two instances of a more general result estimating the Cauchy integral via topological reductions of the cycle of integration. These topological reductions, while not fully automatic, are algorithmically decidable in large classes of cases. An ultimate goal, stated in [PW02, PW04], is to develop software to automate all of the computation.

I can by no means say that the majority of multivariate generating functions fall prey to these new techniques. The class of functions to which the methods described in these lecture notes may be applied is larger than the class of rational functions, but similar in spirit: the function must have singularities, and the dominant singularity must be a pole. This translates to the requirement that the function be meromorphic in a neighborhood of a certain polydisk (see the remark following [PW07, theorem 3.16] for exact hypotheses), which means that it has a representation, at least locally, as a quotient of analytic functions. Nevertheless, as illustrated in [PW07] and in the present lecture notes, meromorphic functions cover a good number of combinatorially interesting examples.

Throughout these notes, I will reserve the variable names

$$F = \frac{G}{H} = \sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$$

for the meromorphic function F expressed as the quotient of locally analytic functions G and H. We assume this representation to be in lowest terms, so G and H do not have a common factor vanishing anywhere that the series for F converges. The variety $\{\mathbf{z} : H(\mathbf{z}) = 0\}$ at which the denominator H vanishes is called the **singular variety** and is denoted by $\mathcal{V}$.

More details will be provided in Chapter 9 but the method is briefly described as follows.

(i) Use the multidimensional Cauchy integral (1.4) to express $a_{\mathbf{r}}$ as an integral over a d-dimensional torus T in $\mathbb{C}^d$.

(ii) Observe that T may be replaced by any cycle homologous to $[T]$ in $H_d(\mathcal{M})$, where $\mathcal{M}$ is the domain of holomorphy of the integrand.

(iii) Use Morse theoretic methods to find a homology basis of cycles that are "pushed down" as far as possible, in the sense that the maximum modulus of the integrand over the cycle cannot be made smaller.

(iv) Use algebraic methods to find **critical** points in $\mathcal{V}$ depending on $\overline{\mathbf{r}}$ that are saddle points for the magnitude of the integrand.

(v) Use topological methods to pick one of the critical points, call it $\mathbf{z}(\overline{\mathbf{r}})$ and replace the integral over T by an integral over a *quasi-local* cycle $\mathcal{C}$ near $\mathbf{z}(\overline{\mathbf{r}})$.

(vi) Evaluate the integral over $\mathcal{C}$ by a combination of residue and saddle point techniques.

1.4 Outline of the remaining chapters

My primary concern in these notes is the derivation of asymptotics once a generating function is known. Nevertheless, some discussion is required of how generating functions are obtained and what meaning can be read into them. One reason to include this is to make these notes into a somewhat self-contained reference. Another is that in obtaining asymptotics, one must sometimes return to the derivation for a new form of the generating function, turning an intractable generating function into a tractable one by changing variables, re-indexing, aggregating and so forth. Consequently, Chapter 2 pulls together a short course in the derivation of univariate and multivariate generating functions. Chapter 3 is a review of univariate asymptotics. Much of this material serves as mathematical background for the multivariate case. While some excellent sources are available, I have not seen the essential techniques briefly summarized in this way and it seems almost certain that someone trying to understand the main subject of these notes will also profit from a review of the essentials of univariate asymptotics.

Part II collects some mathematical background that is, to varying degrees, independent of the specific goals of multivariate asymptotics. Chapters 4, 5 and 7 contain general mathematical background that most students will have seen in their undergraduate or early graduate education, though perhaps not in the depth required for multivariate asymptotics. Specifically, Chapter 4 is a refresher on differential forms and integrals on manifolds. A basic knowledge of calculus on real and complex manifolds is needed: enough that one can speak of integrating holomorphic forms over manifolds with boundaries. One also needs enough algebraic topology to understand homology groups and compute bases for homology groups of certain algebraic varieties. This chapter sums up a semester or two of basic material, so it is not completely self-contained, but serves to collect all the definitions and results that will be called upon in later chapters. Chapter 5 covers one-dimensional saddle point integrals. Chapter 6 discusses saddle point integrals in more than one variable. Most of the results in these chapters can be found in a reference such as [BH86]. The treatment here differs from the usual sources in that Fourier and Laplace type integrals are treated as instances of a single complex-phase case. Working in the analytic category, analytic techniques (contour deformation) are used whenever possible, after which comparisons are given to the corresponding C^∞ approach (which uses integration by parts in place of contour deformation). The last two chapter in Part II concern algebraic geometry. Chapter 7 covers techniques in computational algebra, such as may be found in [CLO98] and other books on Gröbner bases. Chapter 8 then gives an introduction to amoebas of polynomials (after [GKZ94]), covering just a few basics that are necessary for a solid understanding of Laurent series.

Part III is devoted to results that are relatively recent. The first chapter, Chapter 9, is a conceptual guide and overview. The first two sections of the chapter are devoted to a synoposis of stratified Morse theory as it relates to the estimation of coefficients $a_{\mathbf{r}}$ by reducing the multivariate Cauchy integral (1.4) to a recognizable saddle-point integral. As with the topological section of Chapter 4, this is a very brief summary of a substantial theory; this time, the theory is not available in easy-to-read sources, so a longer development is included in the appendices. The third section of this chapter gives algebraic equations for the critical points height function in the Cauchy integral. When the generating function F is a rational function, the critical point equations are algebraic, which is where the computational apparatus of Chapter 7 is used. The last section of Chapter 9 shows how the topology looks when the critical point is on the boundary of the domain of convergence of the series $\sum_{\mathbf{r}} a_{\mathbf{r}} \mathbf{z}^{\mathbf{r}}$ (in the previous literature, these are called *minimal points*).

Having reduced the computation of $a_{\mathbf{r}}$ to a saddle point integral with known parameters, plugging in results on saddle point integration yields theorems for the end-user. These break into several types. Chapter 10 discusses the case where $\mathbf{z}$ is a smooth point of $\mathcal{V}$ and multivariate residue theory is not needed. Chapter 11 discusses the case where $\mathcal{V}$ is the intersection of smooth hypersurfaces near $\mathbf{z}$ and simple multivariate residues are needed. Both of these cases are reasonably well understood. A final case that arises for rational functions is a singularity with nontrivial monodromy. In this case our knowledge is limited, but some known results are discussed in Chapter 12. Finally, Chapter 13 discusses further points. Asymptotics in a regime such as $s \sim \sqrt{r}$ or $s \sim \lambda r + o(r)$ where a change of phase occurs as $s/r = \lambda$ present more difficulties and are only partially understood. Some results of M. Lladser [Lla03] on this topic are presented. Secondly, Chapter 13 discusses some results that are known when the dominant singularity is not a pole.

The final chapter contains a number of worked examples.

Notes

The viewpoint in Section 1.1 is borrowed from the introduction to [Sta97]. The two, very different, motivating problems, alluded to in Section 1.3, that brought me to study multivariate asymptotics were the hitting time generating function from [LL99] and the Aztec Diamond random tiling generating function from [JPS98].

Exercises

Exercise 1.1. Find an asymptotic expansion $f \sim \sum_{j=0}^{\infty} g_j$ for a function f as $x \downarrow 0$ such that $\sum_{j=0}^{\infty} g_j(x)$ is not convergent for any $x > 0$. Conversely, suppose that $f(x) = \sum_{j=0}^{\infty} g_j(x)$ for $x > 0$ and $g_{j+1} = o(g_j)$ as $x \downarrow 0$; does it follow that $\sum_{j=0}^{\infty} g_j$ is an asymptotic expansion of f?

Exercise 1.2. Prove or give a counterexample: if g is a continuous function and for each λ we have $a_{rs} = g(\lambda) + O(r+s)^{-1}$ as $r, s \to \infty$ with $r/s \to \lambda$, then

$$a_{rs} \sim g(r/s) \text{ as } r, s \to \infty$$

and λ varies over a compact interval in $\mathbb{R}^+$.