Chapter 3

Asymptotic Normality in Enumeration

E. Rodney Canfield
Department of Computer Science
University of Georgia

CONTENTS
3.1 Introduction ....................................................... 255
3.2 The normal distribution ........................................... 256
3.3 Method 1: direct approach ......................................... 258
3.4 Method 2: negative roots ......................................... 262
3.5 Method 3: moments ................................................ 266
3.6 Method 4: singularity analysis .................................... 270
3.7 Local limit theorems ................................................. 272
3.8 Multivariate asymptotic normality .............................. 274
3.9 Normality in service to approximate enumeration .......... 276
References ........................................................... 278

3.1 Introduction
The focus of this chapter is the frequent appearance of the normal distribution in the context of combinatorial enumeration. The notion of asymptotic normality is defined, and four methods for establishing its presence are stated and illustrated. The connection of asymptotic normality with other methods of approximate enumeration is briefly explored.
3.2 The normal distribution

A random variable $X$ is said to be **normally distributed** when

$$\text{Prob}\{X \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$ 

We shall denote the function on the right by $\mathcal{N}(x)$. Even though their work is separated by about one hundred years \cite[page 55]{24}, de Moivre and Laplace share credit for the following fundamental theorem concerning independent Bernoulli trials.

**Theorem 3.2.1** Let $0 < p < 1$, let $X_i$, $1 \leq i \leq n$, be a sequence of independent random variables each of which takes the value 1 with probability $p$ and the value 0 with probability $q = 1 - p$, and let $S_n = \sum_{i=1}^{n} X_i$. Then, for any fixed $x$,

$$\text{Prob}(S_n < np + x\sqrt{npq}) \to \mathcal{N}(x), \quad \text{as } n \to \infty.$$ 

We remark that $np$ and $npq$ are the mean and variance, respectively, of $S_n$, and that the assertion of the theorem may be written explicitly in terms of binomial coefficients as

$$\sum_{k<np+x\sqrt{npq}} \binom{n}{k} p^k q^{n-k} \to \mathcal{N}(x), \quad \text{as } n \to \infty,$$

(3.1)

Many textbooks on probability and statistics introduce the normal distribution as an approximation to the binomial distribution, as in the previous equation. Theorem 3.2.1 may be proved by estimating the individual terms via Stirling’s formula and approximating the sum with an integral. Exactly this approach is followed in Section VII.2 of Feller \cite{17}.

The normal distribution is also known as the **Gaussian distribution**. Having achieved fame with his calculation of the orbit of asteroid Ceres (discovered in 1801), Gauss later published a treatise on the subject of celestial orbits. In this work he propounded the method of least squares, and provided theoretical justification by assuming that observational measurements are normally distributed about their mean.

The focus of this article is the frequent appearance of the normal distribution in the context of combinatorial enumeration. The typical situation is that for each positive integer $n$ we have a family of combinatorial objects, such as all permutations of the integers $[n] = \{1, 2, \ldots, n\}$. It is desired to count these with respect to a parameter, for instance how many permutations of $[n]$ have $k$ cycles? This leads to a doubly-indexed array of counts, $a(n,k)$ say. When the objects are given the uniform probability, the parameter can be viewed as a random variable $X_n$. Denoting the mean of $X_n$ as $\mu_n$ and the variance as $\sigma_n^2$, it is often found that the probability of the event $X_n < \mu_n + x\sigma_n$ approaches $\mathcal{N}(x)$ as $n \to \infty$. In this situation, we say the counts $a(n,k)$ are **asymptotically normal**. We repeat the definition.
Definition 3.2.2 Let \( a(n, k) \) be a doubly-indexed sequence of nonnegative numbers. We say that the sequence \( a(n, k) \) is asymptotically normal with mean \( \mu_n \) and variance \( \sigma^2_n \) provided that for the normalized probabilities
\[
p(n, k) = \frac{a(n, k)}{\sum_k a(n, k)}
\]
we have, for each \( x \),
\[
\sum_{k < \mu_n + x \sigma_n} p(n, k) \to \mathcal{N}(x), \quad \text{as } n \to \infty.
\]

In the next few sections we give, with illustrative examples, four methods of proving asymptotic normality: direct, negative roots, moments, and singularity analysis.

The de Moivre-Laplace formulation given above in Theorem 3.2.1 is not always adequate for our task, and so we have ready the following generalization.

Theorem 3.2.3 Let \( X_i, i \geq 1 \), be a sequence of independent 0,1 random variables with \( p_i \) denoting the probability that \( X_i = 1 \), and let \( S_n \) be the nth partial sum \( \sum_{i=1}^n X_i \). Let \( \mu_n, \sigma^2_n \) be the mean and variance of \( S_n \):
\[
\mu_n = \sum_{i=1}^n p_i, \quad \sigma^2_n = \sum_{i=1}^n p_i(1 - p_i).
\]

Then, provided \( \sigma_n \to \infty \),
\[
\text{Prob} \left( S_n < \mu_n + x \sigma_n \right) \to \mathcal{N}(x),
\]
for all \( x \).

Theorem 3.2.1 above is the original, primordial so to speak, central limit theorem. One may dismiss it as a simple application of Stirling’s formula. Consider Theorem 3.2.3, on the other hand. There are some very “combinatorial looking” expressions involved. Namely, one has exactly \( 2^n \) distinct products
\[
A_1 A_2 \cdots A_n, \quad A_i \in \{ p_i, 1 - p_i \}.
\]

Let \( a(n, k) \) be the sum of the \( \binom{n}{k} \) products in which the first alternative \( ( A_i = p_i ) \) is exercised exactly \( k \) times. Theorem 3.2.3 asserts that the numbers \( a(n, k) \) are asymptotically normal, provided the variance becomes infinite. The binomial coefficients have appeared as the number of products making up \( a(n, k) \), but not so prominently as in Equation (3.1). It is not clear that Stirling’s formula can be brought to bear in proving Theorem 3.2.3.

Through the ages mathematicians have sought ways to weaken the hypotheses of Theorem 3.2.1 to fit the latest problem at hand. In Theorem 3.2.3 we still employ
independent 0, 1 variates, but no longer are they assumed to be identically distributed. One may consider other improvements such as variables that assume more than two values, or variables that are weakly dependent, or error estimates for the convergence to normality. The following refinement addresses the first and last of these three.

**Theorem 3.2.4** [4, 15] Let $X_i$, $1 \leq i \leq n$, be independent random variables with means $\mu_i$, variances $\sigma_i^2$, and absolute third central moments $\rho_i = \mathbb{E}|X_i - \mu_i|^3$. With $\mu = \sum_i \mu_i$, $\sigma^2 = \sum_i \sigma_i^2$ denoting the mean and variance of the sum $S_n = \sum_{i=1}^n X_i$, we have

$$\left| \text{Prob} \left( S_n < \mu + x\sigma \right) - \mathcal{N}(x) \right| \leq \frac{C \sum_{i=1}^n \rho_i}{\sigma^3},$$

where $C$ is a universal constant.

This is the Berry-Esseen theorem. It can be used to prove asymptotic normality, and at the same time provide a bound on the error of the approximation. Improved estimates of the universal constant $C$ have been the topic of many papers, even recently. From a purely analytical point of view, requiring the existence of the third moment may be regarded as an excessive assumption. But in combinatorial applications we typically are looking at the limit of finite distributions and the availability of third (and higher) moments is not an issue. Thus, Theorem 3.2.4 can be a widely applicable tool for the combinatorial domain. It is left as an exercise for the reader to deduce Theorem 3.2.3 from 3.2.4.

### 3.3 Method 1: direct approach

Sometimes one can realize the enumeration problem under consideration, suitably normalized, as the distribution of a sum of independent 0, 1 variables. Then, Theorem 3.2.3 is directly applicable. We illustrate this with three examples.

**Example 3.3.1** Let

$$\pi_1 \pi_2 \cdots \pi_n$$

be a permutation of the integers $[n]$. A left-to-right-minimum of $\pi$ is a value $\pi_i$, $1 \leq i \leq n$, satisfying

$$\pi_i = \min \{ \pi_j : j \leq i \}.$$

Let $X_n$ be the number of left-to-right minima in a permutation selected uniformly at random.

**Theorem 3.3.2** [16, 20] With $X_n$ the number of left-to-right minima in a random permutation,

$$\text{Prob} \left( X_n < \log n + x(\log n)^{1/2} \right) \rightarrow \mathcal{N}(x),$$

for each fixed value of $x$. Moreover, the same conclusion holds for the random variable $Y_n$ equal to the number of cycles in a random permutation.
Asymptotic Normality in Enumeration

Proof. The key to seeing that Theorem 3.2.3 is applicable is to build up the permutation \( \pi \) by placing the numbers 1, 2, \ldots, \( n \) one at a time. When 2 is placed, there are 2 possible positions relative to the previously placed element 1; when 3 is placed, there are 3 possible positions relative to the previously placed elements 1 and 2; etc. Letting the \( i \) possible places for the element \( i \) be equally likely means that the final permutation \( \pi \) will have been formed in accordance with the uniform distribution. In the process, we can see that the probability that \( i \) is a left-to-right minimum is \( 1/i \), and that these events are independent. The last assertion holds because whether or not \( i \) will be a left-to-right minimum is unaffected by the order of the first \( i-1 \) placements; it depends only on whether or not \( i \) is placed at the extreme left of the latter. The desired conclusion is now an immediate consequence of Theorem 3.2.3, with \( p_i \) taken to be \( 1/i \), and the well-known formula

\[
\sum_{i=1}^{n} \frac{1}{i} = \log n + O(1).
\]

A bit more combinatorial thinking will allow us to extend the conclusion regarding \( X_n \) to the random variable \( Y_n \), the number of cycles. Suppose a certain permutation \( \pi \) is presented as an unordered set of cycles. Let the smallest element of each cycle serve as the cycle’s representative. Taking the cycles in the order of decreasing representatives, list the elements of each cycle starting with the smallest (the representative) and then following the cyclic order. In this manner, we go from \( \pi \) given as an unordered set of cycles to \( \pi' \) in one line notation, with the number of cycles of \( \pi \) corresponding to the number of left-to-right minima of \( \pi' \). Thus, \( X_n \) and \( Y_n \) have exactly the same distribution. \( \blacksquare \)

Example 3.3.3 This example concerns another permutation statistic, the number of inversions. An inversion of a permutation \( \pi \) of \( [n] \)

\[
\pi_1 \pi_2 \cdots \pi_n
\]

is a pair \((i, j)\) such that \( i < j \) but \( \pi_i > \pi_j \). Again taking the uniform probability distribution on permutations, we define random variable \( Z_n \) by letting \( Z_n(\pi) \) be the number of inversions in permutation \( \pi \).

Theorem 3.3.4 [16, 20] With \( Z_n(\pi) \) the number of inversions of random permutation \( \pi \),

\[
\text{Prob} \left( Z_n < n^2/4 + xn^{3/2}/6 \right) \to \mathcal{N}(x).
\]

Proof. The proof is only a slight embellishment of that given for Theorem 3.3.2. Let the permutation \( \pi \) be built by placing the numbers 1, 2, \ldots, \( n \) as before. When \( i \) is placed, it has a \( 1/i \) probability of creating \( j \) inversions with the previously placed numbers, where \( 0 \leq j < i \). Thus,

\[
Z_n = \sum_{i=1}^{n} Z_n^{(i)},
\]
where the $Z_n^{(i)}$ are independent, and $Z_n^{(i)}$ assumes each of the values $0, 1, \ldots, i-1$ with probability $1/i$. Clearly,

$$E(Z_n^{(i)}) = \frac{i-1}{2}, \quad \text{and} \quad E(Z_n) = \frac{n(n-1)}{4}. $$

Moreover,

$$E\left(\left(Z_n^{(i)}\right)^2\right) = \frac{1}{i} \sum_{j=0}^{i-1} j^2 = \frac{i^3}{3} - \frac{i}{2} + 1/6,$$

whence

$$\text{Var}(Z_n^{(i)}) \sim \frac{i^2}{12}, \quad \text{and} \quad \sigma_n^2 \text{ def } = \text{Var}(Z_n) \sim n^3/36.$$ 

To use Theorem 3.2.4, we need $\rho_i$, the expectation of $|Z_n^{(i)} - i - 1/2|^3$. Clearly, for some $c$,

$$\rho_i \sim c i^3, \quad \text{and} \quad \sum_i \rho_i \sim cn^4/4,$$

whence

$$\frac{\sum_i \rho_i}{\sigma_n^2} \sim 54c n^{4 - 9/2} \to 0.$$ 

By Theorem 3.2.4 the proof is complete. $\blacksquare$

**Example 3.3.5** Both the previous examples are well explained in [17], where citations to the original papers [20] and [16] are given. Our third example is chronologically the earliest of the three, but is presented last due to its greater difficulty. Mark Kac gave a lecture shortly after his arrival in America about independence. One of his topics was divisibility by distinct primes, and he told the audience that there surely must be a central limit theorem awaiting demonstration. A member of the audience, Paul Erdős, “perked up and asked me to explain once again what the difficulty was. Within the next few minutes, even before the lecture was over, he interrupted to announce that he had the solution!” [24], page 90. The Erdős–Kac theorem marks the beginning of probabilistic number theory.

**Theorem 3.3.6** [13] Let $x$ be a fixed real number, and let $f(m)$ be the number of distinct prime divisors of $m$. Define

$$K_n \text{ def } \#\{m \leq n : f(m) < \log \log n + x(\log \log n)^{1/2}\}.$$ 

Then, as $n \to \infty$,

$$\frac{K_n}{n} \to \mathcal{N}(x).$$

**Sketch of proof.** Consider the product $L = p_1 \cdots p_h$ of all the primes less than some fixed bound $H$, and treat the integers from 1 to $L$ as a (uniform) probability space. (Keep in mind throughout this proof sketch the $p_i$ are prime numbers, not probabilities.) Then the event “$m$ is divisible by $p_i$” occurs with probability $1/p_i$, and these
Asymptotic Normality in Enumeration

261

events are (precisely) independent. So, the probability (let us call it \( \alpha_k \)) that an integer in \([1, L]\) has \( k \) distinct prime divisors equals exactly the probability that the sum \( S_h = \sum_{i=1}^{h} X_i \) of independent 0, 1 random variables \( X_i, 1 \leq i \leq h \),

\[
\text{Prob}(X_i = 1) = \frac{1}{p_i},
\]

be \( k \). It is seen that the same equality holds if instead of \( L \) we consider any integral multiple \( qL \) of \( L \). If instead of \( qL \) we consider a very large but arbitrary number \( n \), then the equality becomes not exact, but approximate, and the error can be proven negligible by writing \( n = qL + r \), with \( r \) always limited to \( 0 \leq r < L \). In short, a certain density exists:

\[
\frac{\#\{m \leq n : f_H(m) = k\}}{n} \to \alpha_k,
\]

where \( f_H(m) \) equals by definition the number of distinct prime divisors of \( m \) that are less than or equal to \( H \). It is known (a weak form of Mertens’ Second Theorem)

\[
\mathbb{E}(S_h) = \sum_{p \leq H} \frac{1}{p} \sim \log \log H,
\]

and that \( \text{Var}(S_h) \) is asymptotically the same. Directly by Theorem 3.2.3 above, which Erdős and Kac call the “central limit theorem of the calculus of probability,”

\[
\lim_{H \to \infty} \lim_{n \to \infty} \frac{\#\{m \leq n : f_H(m) < \log \log H + x(\log \log H)^{1/2}\}}{n} = N(x). \tag{3.3}
\]

The hitch of getting from here to the final result, Kac tells the not necessarily mathematically trained readers of his autobiography, is “one of a common variety having to do with the difficulty of justifying the interchange of taking limits.” The number theoretic tool which Kac was lacking at the time of his lecture, and which Erdős supplied, is the theorem of Brun [6]: provided

\[
\frac{\log n}{\log H_n} \to \infty,
\]

we have

\[
\#\{m \leq n : m \text{ is not divisible by any prime less than } H_n\} = \left(e^{-\gamma} + o(1)\right) \frac{n}{H_n}.
\]

The rest of the proof consists in defining \( H = H_n \) as a function of \( n \), and in lieu of the iterated limit in (3.3) considering instead

\[
\lim_{n \to \infty} \frac{\#\{m \leq n : f_{H_n}(m) < \log \log H_n + x(\log \log H_n)^{1/2}\}}{n}.
\]

Provided \( H_n \) grows sufficiently slowly with \( n \), the limiting behavior seen in (3.3) is preserved; provided \( H_n \) grows sufficiently quickly with \( n \), the two sets

\[
\{m \leq n : f_{H_n}(m) < \log \log H_n + x(\log \log H_n)^{1/2}\}
\]

© 2015 by Taylor & Francis Group, LLC
and
\[ \{m \leq n : f(m) < \log \log n + x(\log \log n)^{1/2} \} \]
will be sufficiently close in size. Brun’s theorem given above, plus two other asymptotic formulas due to Mertens [28],
\[ \sum_{p \leq n} \frac{\log p}{p} = \log n + O(1), \]
and
\[ \prod_{p \leq n} \left( 1 - \frac{1}{p} \right) \sim \frac{e^{-\gamma}}{\log n}, \]
constitute the necessary tools. The reader may enjoy attempting the proof. Fair warning, despite Erdős’ having found the solution in “minutes,” Kac reports (as the one designated to write up the proof) spending considerably more time in “assimilating Erdős’ proof.”

### 3.4 Method 2: negative roots

By the fundamental theorem of algebra a monic polynomial \( P(x) \) can be factored
\[ P(x) = \prod_{r \in R} (x + r), \]
where \( R \) is the multiset of the negatives of the (complex) roots. The coefficient of \( x^k \) in \( P(x) \) is given by the familiar formula
\[ [x^k]P(x) = \sum_{|R| - k \text{ elements}} \prod_{r \in R} r, \]
the sum of the products of the elements of \( R \) taken \( |R| - k \) at the time. Comparing this with the earlier appearing expression (3.2) leads to an interesting conclusion: if \( r_i \geq 0, 1 \leq i \leq d \), are nonnegative real numbers then
\[ [x^k] \prod_{i=1}^{d} \frac{x + r_i}{1 + r_i} = \text{Prob} \left( \sum_{i=1}^{d} X_i = k \right), \]
where \( X_i, 1 \leq i \leq d \), are independent 0, 1 random variables with
\[ \text{Prob}(X_i = 1) = \frac{1}{1 + r_i}. \]
(The product of \((1 + r_i)\) on the left is needed to normalize the coefficients into probabilities.) In brief, when a polynomial \( P(x) \) has real and nonpositive roots, we have
Asymptotic Normality in Enumeration

a probabilistic interpretation of its coefficients. Theorem 3.2.4 is applicable. A bit of algebra shows that the mean and variance of \( \sum X_i \) are given by the following “root-free” expressions

\[
\mu = \frac{P'(1)}{P(1)}, \quad \sigma^2 = \frac{P(1)(P'(1) + P''(1)) - P'(1)^2}{P(1)^2}.
\]

Adapting Theorem 3.2.4 to this situation yields the following.

**Theorem 3.4.1** [21] Let

\[ P_n(x) = \sum_k c(n,k)x^k \]

be a sequence of polynomials with real nonpositive roots, and let \( \mu_n, \sigma_n^2 \) be the associated means and variances. Then, provided \( \sigma_n \to \infty \),

\[
\frac{1}{P_n(1)} \sum_{k < \mu_n + x\sigma_n} c(n,k) \to \mathcal{N}(x),
\]

as \( n \to \infty \).

**Example 3.4.2** This technique was exploited by Lévy [26], and has been independently discovered at least one other time: Combinatorialists trace it to [21]. In that work, Harper studies the Stirling numbers of the second kind, \( S(n,k) \), which is the number of partitions of an \( n \)-set into \( k \) nonempty and pairwise disjoint subsets. Harper’s proof bears repeating.

**Theorem 3.4.3** [21] The polynomials \( P_n(x) = \sum_k S(n,k)x^k \) have all real and nonpositive roots, and the Stirling numbers of the second kind are asymptotically normal.

**Proof.** By a simple combinatorial argument

\[
S(n+1,k) = kS(n,k) + S(n,k-1).
\]

(The two terms on the right count the partitions in which \( n+1 \) is not, and is, respectively, in a block by itself.) Multiplying both sides by \( x^k \) and summing on \( k \), this translates into

\[
P_{n+1}(x) = x\left(P'_n(x) + P_n(x)\right).
\]

From here, one finds that the roots of these polynomials, (other than the common root \( x = 0 \) which they all have), exhibit a nice **interlacing property**. Assume, inductively, that \( P_n(x) \) has the roots

\[ r_{n-1} < r_{n-2} < \cdots < r_1 < r_0 = 0. \]

It is clear that \( P'_n(x) \) will be alternately positive, then negative, at the points 0, \( r_1, \ldots \). Then by (3.5) \( P_{n+1}(x) \) is alternately positive, then negative, at the points \( r_1, r_2, \ldots \). We have found roots for \( P_{n+1}(x) \) at \( x = 0 \), and in each of the \( n-1 \) intervals \( (r_1,r_0),(r_2,r_1),\ldots,(r_{n-1},r_{n-2}) \). That’s \( 1 + (n-1) = n \) real and nonpositive roots so
far. The sign of \( P_{n+1}(x) \) at \( r_{n-1} \) is opposite that of \( P'_n(x) \), that is, \((-1) \times (-1)^{n-1} = (-1)^n\). But \( P_{n+1}(x) \), being of degree \( n+1 \), has sign \((-1)^{n+1}\) as \( x \to -\infty\). So, there is an \((n+1)\)th root in the interval \((-\infty, r_{n-1})\). By induction then, each polynomial \( P_n(x) \) has \( n \) roots in the interval \((-\infty, 0]\), and these roots do indeed exhibit an interlacing phenomenon.

If only we knew \( \sigma_n \to \infty \), the Stirling numbers of the second kind, \( S(n, k) \), would be proven asymptotically normal. Let \( B_n \), the \( n \)th Bell number, equal \( \sum_k S(n, k) \), the total number of partitions of \([n]\). Using (3.4), as found by Harper,

\[
\mu_n = B_{n+1}/B_n - 1, \quad \sigma_n^2 = B_{n+2}/B_n - (B_{n+1}/B_n)^2 - 1. \tag{3.6}
\]

Some careful asymptotic estimates for \( B_n \) complete Harper’s proof. ■

In colorful language Harper says that traditionally proofs of asymptotic normality proceed by “torturing the characteristic function until it converges to \( e^{-t^2/2}\)…” (See Section 3.5). But in the present proof, one needs a “hat from which to pull a rabbit,” and that hat is the central limit theorem. With no intention to diminish the magical aura of the proof, might not one nevertheless seek a definite combinatorial meaning for the probabilities \( 1/(1 + r_i) \)? That is, perhaps we can define some independent random variables \( Y_i \) with \( \sum_i Y_i \) having the desired distribution, and apply Theorem 3.2.3 directly. Think back to the first example of the direct method: the number of cycles in a random permutation. There we let each cycle be represented by its smallest element, and produced a listing of numbers whose left-to-right minima corresponded to the cycles. A similar process suggests itself for partitions of \([n]\). Let the partition \( \pi \) be given as an unordered collection of subsets, usually referred to as blocks. Represent each block by its smallest element, and in this way the blocks may be numbered \( 1, 2, \ldots, k \), \( k \) being the number of blocks. Now define a function \( f_\pi : [n] \to [k] \) by letting all members of the \( i \)th block be mapped by \( f_\pi \) to \( i \). Make a list of the values of this function

\[
f_\pi(1) \ f_\pi(2) \ \cdots \ \ f_\pi(n). 
\]

It will be seen that \( f_\pi \) is a restricted growth function, meaning

\[
f_\pi(i) \leq 1 + \max\{ f_\pi(j) : j < i \},
\]

and that the number of blocks of partition \( \pi \) is exactly the number of left-to-right maxima of the associated restricted growth function \( f_\pi \). Let the set of all restricted growth functions on \([n]\) have the uniform probability, and define the \( 0, 1 \) random variables \( Y_i \) to be the indicators of the event “a left-to-right maximum for \( f \) occurs at position \( i \).” We have just seen that

\[
\text{Prob} \left( \sum_i Y_i = k \right) = \frac{S(n, k)}{B_n},
\]

so the sum \( \sum_i Y_i \) has the desired distribution. Unfortunately, the \( Y_i \) are not independent, and the theorems available to us do not apply. (The one-line notation for set partitions, widely used now, appears first in [32, page 96].)
**Example 3.4.4** We close this section with an example taken from the theory of matchings in graphs. Let \( G = (V, E) \) be a simple graph; a subset \( M \subseteq E \) is called a *matching* in \( G \) provided no two distinct edges \( e, e' \in M \) have a vertex in common. Let \( \mathcal{M}(G) \) be the set of all matchings in graph \( G \), and define the matching polynomial \( P_G(x) \) by

\[
P_G(x) \equiv \sum_{M \in \mathcal{M}(G)} x^{|M|}.
\]

Thus, \([x^k]P_G(x)\) equals the number of matchings in \( G \) of size \( k \). In addition to its graph theoretical origins, this concept has been a long-time interest in statistical physics where the terminology is that of *monomer-dimer coverings*. The remarkable fact that \( P_G(x) \) always has all its roots real and negative is due to this latter community. See the seminal paper of Heilmann and Lieb [23]. In that work Heilmann and Lieb prove a more general assertion using different polynomials, namely

\[
Q_G(x) \equiv \sum_{M \in \mathcal{M}(G)} (-1)^{|M|} W(M) x^{n-2|M|},
\]

where whenever used as a subscript of the symbol “\( Q \)” \( G \) is a complete graph on \( n \) vertices whose edges carry positive weights, and \( W(M) \) is the product of the weights of the edges comprising \( M \). For a vertex \( i \) in the edge-weighted complete graph \( G \), partition \( \mathcal{M}(G) \) into those \( M \) which, respectively, do and do not cover vertex \( i \); one is led to the recursion

\[
Q_G(x) = xQ_{G-\{i\}}(x) - \sum_{j \in [n]-i} w_{ij} Q_{G-\{i\}-j}(x).
\]

(\( w_{ij} \) is the weight on edge \( i.j \).) From here an argument resembling that given earlier for \( S(n, k) \) establishes by induction that the roots of \( Q_{G-\{i\}} \) are real, distinct, and interlace those of \( Q_G \). By taking a limit in which all edge weights are 1 and 0, one finds that \( x^k P_G(-1/x^2) \) (\( G \) now arbitrary) has all real, albeit not necessarily distinct, roots. Whence the matching polynomial \( P_G(x) \) has all roots real and nonpositive.

Godsil [18] looked up this topic and, building on Harper’s work, found the following version of Theorem 3.4.1 for the matching polynomials.

**Theorem 3.4.5** [18] Let \( G_n \) be a sequence of graphs, \( X_n \) be the size of a matching \( M \) drawn uniformly at random from \( \mathcal{M}(G_n) \), \( \mu_n = \mathbb{E}(X_n) \), and \( \sigma_n^2 = \text{Var}(X_n) \). Then, provided \( \sigma_n^2 \to \infty \),

\[
\text{Prob}(X_n < \mu_n + x\sigma_n) \to \mathcal{N}(x).
\]

Two applications are noted: \( G_n \) a sequence of regular graphs of degree \( r \), with the number of vertices becoming infinite; and \( G_n \) equal to the complete graph, \( K_n \). Later, Kahn [25] continued the study of this topic, and gave, among other things, a formulation of the condition \( \sigma_n \to \infty \) in purely graph theoretical notions.
Theorem 3.4.6 [25] Let $G_n$ be a sequence of graphs, $X_n$ be the size of a matching $M$ drawn uniformly at random from $\mathcal{M}(G_n)$, $\mu_n = \mathbb{E}(X_n)$, and $\nu_n$ the size of the largest matching in $\mathcal{M}(G_n)$. Then

$$\operatorname{Prob}(X_n < \mu_n + x\sigma_n) \to \mathcal{N}(x)$$

if and only if

$$\nu_n - \mu_n \to \infty.$$

3.5 Method 3: moments

For a random variable $X$, the associated moment generating function is defined by

$$M_X(\tau) \overset{\text{def}}{=} \mathbb{E}e^{\tau X}.$$

No question of convergence arises when $\tau$ is pure imaginary, but $\tau = 0$ could be the only real value for which $M_X(\tau)$ is defined. For $X$ distributed as standard normal, $M_X(\tau)$ is the entire function $e^{\tau^2/2}$. When $M_X(\tau)$ is analytic near $\tau = 0$, it has the power series expansion

$$M_X(\tau) = \sum_{k=0}^{\infty} \mathbb{E}X^k \frac{\tau^k}{k!},$$

hence the nomenclature “moment generating function.” The following theorem finds frequent combinatorial application.

Theorem 3.5.1 [11] Suppose $X_n$ is a sequence of random variables with distribution functions $F_n(x)$ such that

$$\mathbb{E}e^{\tau X_n} \to e^{\tau^2/2}$$

for all $\tau$ in a nonempty real interval $(-a, +a)$. Then, for all $x$,

$$F_n(x) \to \mathcal{N}(x).$$

Similar limit theorems are known for the characteristic function $\mathbb{E}e^{i\tau X}$ (Lévy’s Continuity Theorem, [5], Theorem 26.3) and for moments (the Fréchet-Shohat Theorem, [5], Theorem 30.2). The theorem for moments states that if each member of sequence $X_n$ has moments of all orders, and for each $k$ the $k$th moments of $X_n$ converge to the $k$th moment of the standard normal, then the $X_n$ are asymptotically normal with mean $0$ and variance $1$.

Example 3.5.2 Let $q(n,k)$ be the number of partitions of $n$ into exactly $k$ distinct parts. For example, $q(8,3) = 2$, the two partitions being $5 + 2 + 1$ and $4 + 3 + 1$. The following theorem is stated without proof in [14], excluding the explicit formula for the variance.
Asymptotic Normality in Enumeration

Theorem 3.5.3 [14] Define

\[ \mu_n = \frac{2\sqrt{3}\ln(2)}{\pi} n^{1/2} \]

\[ \sigma_n^2 = \frac{\sqrt{3}}{\pi} \left( 1 - \left( \frac{2\sqrt{3}\ln(2)}{\pi} \right)^2 \right) n^{1/2}. \]

Then the numbers \( q(n,k) \) are asymptotically normal with mean \( \mu_n \) and variance \( \sigma_n^2 \).

Proof. Define the polynomials \( Q_n(y) = \sum k \cdot q(n,k)y^k \), and the generating function

\[ f(x,y) = \sum_{n=0}^\infty x^n Q_n(y) = \prod_{m=1}^\infty (1 + yx^m). \]

By Theorem E it suffices to show

\[ e^{-\tau \mu_n / \sigma_n} \frac{Q_n(e^{\tau / \sigma_n})}{Q_n(1)} \rightarrow e^{\tau^2 / 2}. \] (3.7)

Using contour integration around a circle of radius \( e^{-t}, t > 0 \), we have

\[ e^{-nt} Q_n(e^{i\tau}) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(e^{-it}e^{i\theta}, e^{i\tau}) e^{-ni\theta} d\theta. \]

Throughout the remainder of the proof, asymptotic formulas hold uniformly for \( n \rightarrow \infty \) and \( t, \tau \rightarrow 0 \) with \( t = (c_2/n)^{1/2} \) and \( \tau = O(t^{1/2}) \). (The constants \( c_0, c_1, c_2 \) are respectively \( 1/2, \ln 2, \pi^2 / 12 \).)

Using

\[ f(e^{-it}e^{i\theta}, e^{i\tau}) e^{-ni\theta} \sim f(e^{-it}, e^{i\tau}) \exp \left( c_1 \tau^{-2i\theta} - c_2 t^{-3} \theta^2 \right), \quad |\theta| \leq t^{7/5}, \] (3.8)

we find

\[ \int_{-t^{7/5}}^{+t^{7/5}} f(e^{-it}e^{i\theta}, e^{i\tau}) e^{-ni\theta} d\theta \sim \sqrt{\frac{\pi}{c_2}} f(e^{-it}, e^{i\tau}) t^{3/2} \exp \left( -\frac{c_1^2}{4c_2^2} \tau^2 t^{-1} \right). \]

Using

\[ |f(e^{-it}e^{i\theta}, e^{i\tau})| \leq f(e^{-it}, e^{i\tau}) e^{-c|\theta|^{1/7}}, \quad c > 0, \quad t^{7/5} \leq |\theta| \leq \pi, \] (3.9)

we find that the integral over \( t^{7/5} \leq |\theta| \leq \pi \) is negligible by comparison. This gives an asymptotic for \( e^{-ni} Q_n(e^{i\tau}) \), which we use twice to find

\[ \frac{Q_n(e^{i\tau})}{Q_n(1)} \sim \frac{f(e^{-it}, e^{i\tau})}{f(e^{-it}, 1)} \exp \left( -\frac{c_1^2}{4c_2^2} \tau^2 t^{-1} \right). \]

Finally, using

\[ \frac{f(e^{-it}, e^{i\tau})}{f(e^{-it}, 1)} \sim \exp \left( c_1 \tau^{-1} + \frac{c_0}{4} \tau^2 t^{-1} \right), \] (3.10)
we reach the desired conclusion (3.7). The three key formulas (3.8), (3.9), and (3.10) can be proven either by Euler-Maclaurin summation, or by the Cahier-Mellin formula; see [8], for a similar calculation.

The value for $\sigma_n^2$ appears in [31]. In the latter work, Szekeres proves a formula equivalent to
\[
q(n,k) \sim \frac{e^{-x^2/2}}{\sigma_n(2\pi)^{1/2}} Q_n(1), \quad k = \mu_n + x\sigma_n,
\]
for $k - \mu_n = o(n^{1/3})$. That is, he proves a local limit theorem; see Section 3.7. In [14] Erdős and Lehner concentrate on $p(n,k)$, the analogous quantity for unrestricted partitions. These numbers are not asymptotically normal, and it would be interesting to follow the methodology of Theorem 6 to see where the difference arises. Erdős and Lehner’s derivation of the limiting distribution of $p(n,k)$ is a brilliant application of inclusion/exclusion—reminiscent of Brun’s sieve in Example 3. Another theorem about partitions, similar in conclusion to Theorem 6, has been given by Goh and Schmutz [19]: The number of distinct parts in a random unrestricted partition is asymptotically normal with mean \( (\sqrt{6}/\pi) n^{1/2} \) and variance \( (\sqrt{6}/2\pi - \sqrt{54}/\pi^3) n^{1/2} \).

**Example 3.5.4** Let $g_i$, $i \geq 1$, be a sequence of nonnegative integers, and define a doubly-indexed array $a(n,k)$ by the equation

\[
a(n,k) \overset{\text{def}}{=} \frac{1}{k!} \sum_{(i_1, \ldots, i_k)} \binom{n}{i_1 \cdots i_k} g_{i_1} \cdots g_{i_k}, \tag{3.11}
\]

Combinatorially, $a(n,k)$ is the number of ways to partition $[n]$ into $k$ blocks, and to build a $g$-object on each block, $g_i$ being the number of $g$-objects on an $i$-set. If, for instance, we let $g_i$ equal 1 for all $i$, then the above becomes a familiar formula for $S(n,k)$, the number of partitions of $[n]$ into $k$ blocks. As another illustration, if $g_i = (i-1)!$, the number of full cycles on an $i$-set, then the above $a(n,k)$ is the number of permutations of $[n]$ having $k$ cycles. Letting $g(u)$ be the exponential generating function of the sequence $g_i$,

\[
g(u) = \sum_{i \geq 1} g_i \frac{u^i}{i!},
\]

the generating function equivalent of (3.11) is

\[
\sum_{n=0}^{\infty} \frac{u^n}{n!} \left( \sum_{k} a(n,k) y^k \right) = \exp(g(u)).
\]

The polynomials $P_n(y) = \sum_k a(n,k) y^k$ are polynomials of binomial type [30].

**Theorem 3.5.5** [7] Let

\[
g(u) = \sum_{j=1}^{d} g_j u^j / j! = \sum_{j=1}^{d} c_j u^j
\]
be a real polynomial with nonnegative coefficients, at least two of which are nonzero, and such that
\[ \gcd \{ j : c_j \neq 0 \} = 1. \]

Let the doubly-indexed array \( a(n,k) \) be defined by (3.11). Then the numbers \( a(n,k) \) are asymptotically normal.

**Sketch of proof.** As in the prior example, relying on Theorem 3.5.1, it suffices to define \( \mu_n, \sigma^2_n \) that satisfy
\[
e^{-\tau \mu_n / \sigma_n} P_n(e^{\tau / \sigma_n}) \frac{P_n(1)}{P_n(y)} \to e^{\tau^2 / 2}.\]

(3.12)

Again, we begin with contour integration around a circle to extract \( P_n(y) / n! \) as the coefficient of \( u^n \) in \( \exp(y g(u)) \), and, as before, the integral is dominated by a small arc near the positive real axis, provided the radius \( \rho \) is chosen by the saddlepoint equation
\[
\rho g'(\rho) = \frac{n}{y}.
\]

Let \( r(X) \) denote the inverse function of \( X g'(X) \), so the latter radius is \( \rho = r(n/y) \).

The estimate resulting from the contour integration is
\[
\frac{P_n(y)}{P_n(1)} \sim \frac{\exp(y g(r(n/y)))}{(2\pi B)^{1/2}}, \quad n \to \infty, \ y \to 1,
\]

with \( B = d^2 c_d r(n)^d \) (recall \( d \) is the degree of \( g(u) \)). In Example 3.5.2, the next step was to form the quotient, but here a nice feature of Example 3.5.2 is missing: Even for \( y \to 1 \) we must keep up with both \( r(n) \) and \( r(n/y) \), and we cannot use the same radius for both \( P_n(1) \) and \( P_n(y) \). However, the common terms, \( B \) and \( n! \), do drop out of the quotient leaving
\[
\frac{P_n(y)}{P_n(1)} \sim \frac{\exp(y g(r(n/y)))}{\exp(g(r(n)))} \cdot \left( \frac{r(n)}{r(n/y)} \right)^n, \quad n \to \infty, \ y \to 1.
\]

The next step is to recognize the right side of this formula as
\[
\exp \left( \int_1^y g(r(n/\beta)) d\beta \right).
\]

Looking back at (3.12), we need to define \( \mu_n, \sigma^2_n \) so that
\[
-\frac{\tau \mu_n}{\sigma_n} + \int_1^y g(r(n/\beta)) d\beta \to \frac{\tau^2}{2}.
\]

The expansion of \( g(r(n/\beta)) \) about \( \beta = 1 \) begins
\[
g(r(n/\beta)) = g(r(n)) - \frac{n^2}{B} (\beta - 1) + \cdots
\]
and so we are led to define

\[ \mu_n = g(r(n)) \]
\[ \sigma_n^2 = g(r(n)) - \frac{n^2}{B}. \]

The rest of the proof involves looking carefully at the higher-order derivatives of the function \( \beta \mapsto g(r(n/\beta)) \). The number of derivatives that come into play, call it \( K \), must satisfy \( KJ/2 > d \), where \( J \) is the second highest index (after \( d \)) such that \( c_J \neq 0 \). Details can be found in [7].

Applications of Theorem 3.5.5 include partitions into blocks of bounded size, permutations with cycles of bounded size, and permutations whose order divides a given \( d > 1 \).

### 3.6 Method 4: singularity analysis

The next theorem furnishes an example of a simply stated criterion on the bivariate generating function \( f(x, y) \), which implies asymptotic normality for the coefficients \( a(n, k) \). In this case, the criterion is that the function have only one singularity on the circle of convergence, and that it be well behaved.

**Theorem 3.6.1** [3] Let \( f(x, y) = \sum_{n,k} a(n, k)x^ny^k \), with \( a(n, k) \geq 0 \). Suppose there exist

(i) a function \( A(s) \) continuous and nonzero near 0,

(ii) a function \( r(s) \) with bounded third derivative near 0,

(iii) a nonnegative integer \( m \), and

(iv) positive numbers \( \varepsilon \) and \( \delta \) such that

\[
\left( 1 - \frac{x}{r(s)} \right)^m f(x, e^s) - \frac{A(s)}{1-x/r(s)}
\]

is analytic and bounded for \( |x| < \varepsilon, |x| < r(0) + \delta \).

Put \( \mu = -r'(0)/r(0) \) and \( \sigma^2 = \mu^2 - r''(0)/r(0) \). If \( \sigma^2 \neq 0 \), then the numbers \( a(n, k) \) are asymptotically normal with mean \( \mu n \) and variance \( \sigma^2 n \).

**Sketch of proof.** Let

\[ f(z, e^s) = \sum \phi_n(s)z^n. \]
Asymptotic Normality in Enumeration

If a power series $\sum a_n x^n$ has finite radius of convergence $r$, and $f(z) = A/(1-z/r)^m$ is analytic in a larger circle, then $a_n \sim A n^{m-1} r^{-n} / (m-1)!$. By this sort of analysis it is shown
\[ \phi_n(s) \sim A(s)n^m r(s)^{-n} / m!, \]
uniformly in $s$. The characteristic function is
\[ e^{-\mu_n/\sigma_n} \frac{\phi(it/\sigma_n)}{\phi(1)} \sim \left( \frac{r(it/\sigma_n)}{r(0)} \right)^{-n}, \]
and by expanding
\[ \frac{r(s)}{r(0)} = \exp \left( \frac{r'(0)}{r(0)} s + \left( \frac{r''(0)}{r(0)} - \left( \frac{r'(0)}{r(0)} \right)^2 \right) (s^2/2) + \cdots \right) \]
the proof is completed. \[ \Box \]

Example 3.6.2 Say that a permutation $\pi$ of $[n]$ has a rise at position $i$, $1 \leq i < n$, if $\pi_i < \pi_{i+1}$. For $n \geq 1$ we follow the (fairly) standard practice of declaring there also to be a rise at position $i = 0$. Let $A(n,k)$, $1 \leq k \leq n$, be the number of permutations of $[n]$ having $k$ rises. These are called the Eulerian numbers.

Theorem 3.6.3 [3, 12] The Eulerian numbers $A(n,k)$ are asymptotically normal with mean $\mu_n = n/2$ and variance $\sigma^2_n = n/12$.

Proof. We start with the two-variable exponential generating function as given by David and Barton [12]
\[ f(x,y) \overset{\text{def}}{=} \sum_{n,k \geq 0} E_n(y) \frac{x^n}{n!} = \frac{1-y}{1-y e^{-x y / x - 1}}. \] (3.13)

By solving for the vanishing of the denominator, we find that the radius of convergence for $f(x,e^s)$ is
\[ r(s) = \frac{e^s}{e^s - 1} = 1 - s/2 + s^2/12 + \cdots. \]

Theorem 3.6.1 is applicable, and $r(0), r'(0), r''(0)$ equal $1, -1/2, 1/6$ respectively. Hence, $\mu = 1/2$ and $\sigma^2 = (1/2)^2 - 1/6 = 1/12$. \[ \Box \]

Remarks. The asymptotic normality of the Eulerian numbers can be, and has been, obtained in other ways. Let $E_n(y) = \sum_k A(n,k) y^k$ be the associated polynomials. With the convention that $E_0 = 1$, the following recursion can be deduced by combinatorial reasoning:
\[ A(n,k) = kA(n-1,k) + (n-k+1)A(n-1,k-1), \quad n \geq 1. \]

From here we find
\[ E_n(y) = y(1-y)E'_{n-1}(y) + nyE_{n-1}, \quad n \geq 1. \] (3.14)
By taking derivatives with respect to $y$, evaluating at $y = 1$, and solving the resulting recursions, one can obtain the two exact formulas

\[
\mu_n = \frac{n+1}{2}, \\
\sigma^2_n = \frac{n+1}{2}
\]

for the mean and variance. Using the technique illustrated in Section 3.4 one can prove from (3.14) that the polynomials $E_n(y)$ have interlacing real negative roots. Hence, by Theorem 3.4.1, the Eulerian numbers $A(n,k)$ are asymptotically normal. This is the approach taken in [9].

Another approach is that taken by David and Barton [12]. Working with the two-variable exponential generating function (3.13), they prove asymptotic normality using a variant of the method of moments, the method of cumulants.

3.7 Local limit theorems

Suppose a doubly-indexed array $a(n,k)$ is asymptotically normal with mean $\mu_n$ and variance $\sigma^2_n$. This may tell us nothing about an individual number $a(n,k)$. For instance, in Example 3, concerning the number, $f(m)$, of prime divisors of a random integer $m$, we dare not hope for an asymptotic formula for the number of prime divisors of $m$. Indeed, we suspect that integers with one prime divisor may very well be adjacent to an integer with an extraordinary number of prime divisors. Just knowing the frequency with which $f(m)$ assumes various values, the very content of asymptotic normality, is a surprising and pleasing amount of information for such a noisy function.

On the other hand, suppose for a given $n$ the numbers $a(n,k)$ do not vary chaotically like the divisor counting function; suppose instead they vary smoothly. Let $\mathcal{D}_x$ be the discrete set of real $t$ in the interval $(-\infty, x]$ with the property that $\mu_n + t\sigma_n$ is an integer. The $t \in \mathcal{D}_x$ are spaced at a distance $1/\sigma_n$, and if $\sigma_n \to \infty$,

\[
\sum_{t \in \mathcal{D}_x} a(n, \mu_n + t\sigma_n) \to \mathcal{N}(x).
\]

But by asymptotic normality

\[
\sum_{t \in \mathcal{D}_x} \frac{a(n, \mu_n + t\sigma_n)}{\sum_k a(n,k)} \to \mathcal{N}(x).
\]

Here we have two smoothly varying sequences of numbers whose partial sums are asymptotic; optimistically, one hopes

\[
\frac{a(n,k)}{\sum_k a(n,k)} \sim \frac{e^{-t^2/2}}{\sigma_n \sqrt{2\pi}} \left|_{t=(k-\mu_n)/\sigma_n} \right.
\]
Asymptotic Normality in Enumeration

This prompts the next definition.

**Definition 3.7.1** We say the doubly indexed sequence \( a(n, k) \) satisfies a local limit theorem on the set \( S \) of real numbers provided

\[
\sup_{x \in S} \left| \frac{\sigma_n a(n, \lfloor \mu_n + x\sigma_n \rfloor)}{\sum_k a(n, k)} - \frac{e^{-x^2/2}}{\sqrt{2\pi}} \right| \to 0.
\]

**Theorem 3.7.2** [3] Suppose that \( a(n, k) \) are asymptotically normal, and \( \sigma_n^2 \to \infty \). If for each \( n \) the sequence \( a(n, k) \) is unimodal in \( k \), then \( a(n, k) \) satisfy a local limit theorem on the set \( \{ x : |x| \geq \epsilon \} \), any \( \epsilon > 0 \). If for each \( n \) the sequence \( a(n, k) \) is log concave in \( k \), then \( a(n, k) \) satisfy a local limit theorem on the set \( \mathbb{R} \).

Unimodality,

\[
a(n, 1) \leq \cdots \leq a(n, K) \geq a(n, K + 1) \geq \cdots,
\]

and log concavity,

\[
a(n, k)^2 \geq a(n, k+1)a(n, k-1),
\]

are features that arise often in combinatorics. Both of these constitute “smoothness” adequately enough to pass from a central to local limit theorem, as stated in Theorem 3.7.2.

**Example 3.7.3** If \( P_n(y) = \sum_k a(n, k)y^k \) has all its roots real and nonpositive, then its coefficients are log concave. Thus, from Examples 3.4.2 and 3.6.2, we derive genuine asymptotic formulas for Stirling numbers of the second kind \( S(n, k) \) and Eulerian numbers \( A(n, k) \) when \((k - \mu_n)/\sigma_n \) is bounded.

**Theorem 3.7.4** [21, 29] Let \( \mu_n = n/\ln n \) and \( \sigma_n^2 = n/(\ln n)^2 \). Then, uniformly for \( k - \mu_n = O(\sigma_n) \),

\[
\frac{S(n, k)}{B_n} \sim \frac{e^{-x_n^2/2}}{\sigma_n(2\pi)^{1/2}}, \quad x_n = (k - \mu_n)/\sigma_n.
\]

**Proof.** To obtain the formulas for \( \mu_n, \sigma_n^2 \) we use Equation (3.6) and an asymptotic estimate of the Bell number \( B_n \). Then, Theorem 3.7.2 is applicable.

**Theorem 3.7.5** [9] Let \( \mu_n = n/2 \) and \( \sigma_n^2 = n/12 \). Then, uniformly for \( k - \mu_n = O(\sigma_n) \),

\[
\frac{A(n, k)}{n!} \sim \frac{e^{-x_n^2/2}}{\sigma_n(2\pi)^{1/2}}, \quad x_n = (k - \mu_n)/\sigma_n.
\]

**Proof.** Derivation of the formulas for \( \mu_n, \sigma_n^2 \) has been indicated earlier. Then, Theorem 3.7.2 is applicable.
Remark 3.7.6 Harper [21] states Theorem 9.1 as a corollary to his central limit theorem for $S(n,k)$, giving a quite succinct and informal justification. The first asymptotic formula for $S(n,k)$ covering all $k$ was provided earlier by Moser and Wyman [29]. In Moser and Wyman’s formula the connection to the central and local limit theorems is not at all apparent unless one specializes $k$ to be near $\mu_n$ and does the necessary algebra. In their work [9] Carlitz et al. acknowledge their debt to Harper, and utilize the Berry-Esseen theorem in deriving an asymptotic formula for $A(n,k)$ with explicit error bound. Their proof, which Riordan complements in his math review, can be read as a careful fleshing out of Harper’s succinct proof. Theorem G, due to Bender, can be read as an adaptation of the local limit theorem in [9] that does not use the Berry-Esseen theorem and that holds in perfect generality.

3.8 Multivariate asymptotic normality

With the intent of providing the reader with the basic and most widely used principles of our topic, the discussion thus far has been totally one-dimensional. Of course, there is such a thing as a $d$-dimensional Gaussian distribution, and multivariate distributions are ubiquitous in combinatorial settings.

Definition 3.8.1 Let $k, i$ be $d$-dimensional vectors of integers, and $a(n,k)$ an array of nonnegative numbers. We say that $a(n,k)$ is asymptotically normal with mean $m_n$ and covariance matrix $B_n$ if

$$\sup_u \left| \frac{1}{\sum a(n,k)} \sum a(n,i) - \frac{1}{(2\pi)^{d/2}|B_n|^{1/2}} \int_{x \leq u} \exp \left( -\frac{1}{2} x B_n^{-1} x^T \right) dx \right| \to 0.$$

For a $d$-dimensional normal $Z$ with mean $0$ and covariance matrix $B$ (the probability density in the second operand above), the moment generating function is found to be

$$E(e^{\tau \cdot Z}) = \exp \left( \frac{1}{2} \tau B \tau^T \right).$$

As in the one-dimensional case, one may prove convergence to normality by the method of moments. For instance, let each of $\xi_1, \xi_2, \ldots, \xi_\ell$ be a $d$-dimensional real vector, and consider the random variable $X = (x_1, \ldots, x_d)$ given by

$$X = \xi_i \text{ with probability } c_i, \quad 1 \leq i \leq \ell, \quad \sum_i c_i = 1.$$

Form the $\ell \times d$ matrix $M$ using $(c_i)^{1/2} \xi_i$ as the $i$th row:

$$(M)_{ij} = (c_i)^{1/2} (\xi_i)_j.$$
then, the $(j,k)$ entry of the product $M^TM$ is $E(x_jx_k)$. With $\mathbf{m}$ the vector of means, we find
\[
Ee^{\tau^T} = \exp \left( \mathbf{m} \cdot \tau + \frac{1}{2} \tau^T (M^TM - N) \tau + O(\sum |\tau_j|^3) \right),
\]
where $(N)_{jk} = \langle \mathbb{E}x_j \rangle \langle \mathbb{E}x_k \rangle$. Now $B = M^TM - N$ is the covariance matrix for $X$, and it follows that the sums $\sum_{k=1}^{n} X_k$ (with the $X_i$ independent and each distributed like $X$) are asymptotically normal with mean $n\mathbf{m}$ and covariance matrix $nB$. So, we have a $d$-dimensional analog of Theorem 3.2.3 courtesy of the method of moments, and we might reasonably anticipate products $g(x)^n$ to be likely sources of normality.

In [1] it is shown that if the coefficients $\phi_n(x)$ in the ordinary generating function
\[
\sum_n a(n, k)z^n x^k = \sum_n \phi_n(x)z^n
\]
behave asymptotically like a product $\alpha_n h(x)g(x)^n$, then under suitable general conditions the $a(n, k)$ will satisfy a central limit theorem. Generalizing Theorem 3.6.1, Bender and Richmond prove in [1] that this principle applies to generating functions satisfying the conditions the $a(n, k)$ will satisfy a central limit theorem.

As another example, consider $a(n, k_1, k_2)$ equal to the number of matchings in the product graph $[n] \times [2]$ consisting of $k_1$ horizontal and $k_2$ vertical edges. Then the generating function is given by
\[
\sum_{n,k_1, k_2} a(n, k_1, k_2) z^{n,k_1,k_2} = \sum_{n\geq0} \begin{pmatrix} 1 & x_1 & 1 & 1 \\ x_2 & 0 & x_2 & 0 \\ x_2^2 & 0 & 0 & 0 \end{pmatrix}^n = \begin{pmatrix} 1 & x_1 & 1 & 1 \\ x_2 & 0 & x_2 & 0 \\ x_2^2 & 0 & 0 & 0 \end{pmatrix} (1,1)
\]

They deduce that $a(n, k_1, k_2)$ are asymptotically normal with mean equal to $n \mathbf{m}$ and covariance matrix equal to $nB$ where
\[
\mathbf{m} = (1/4, 1/4), \quad B = \begin{bmatrix} 1/8 & -1/12 \\ -1/12 & 1/8 \end{bmatrix}.
\]
(indicating the extraction of element $(1,1)$ from the matrix) [2]. For a great variety of such matrix recursions, the authors of [2] show that both central and local limit theorems hold. The mean and variance are again proportional to $n$, but the details of how to calculate these are not easy and the interested reader is asked to consult the original paper.

### 3.9 Normality in service to approximate enumeration

Let $g(u) = \sum g_j u^j$ be a power series with real, nonnegative coefficients having radius of convergence $R$, $0 < R \leq \infty$. For each $r < R$ we may consider the random variable $X_r$ which assumes the value $j$ with probability $g_j r^j / g(r)$. Define

$$a(r) \overset{\text{def}}{=} rg'(r) / g(r), \quad b(r) \overset{\text{def}}{=} r a'(r);$$

it will be seen that these are the mean and variance of $X_r$. Assuming $X_r$ to be asymptotically normal gives

$$\frac{1}{g(r)} \sum_{j \leq a(r) + \sqrt{b(r)}} g_j r^j \to \mathcal{N}(x), \quad r \to R. \tag{3.15}$$

Let the sequence $r_n$ be defined by the equation

$$a(r_n) = n.$$

If $r_n \to R$, and if passage from central to local limit theorem is permitted, we have (after isolating $g_n$ on one side)

$$g_n \sim \frac{g(r_n)}{r_n^a (2\pi b(r_n))^{1/2}}. \tag{3.16}$$

$n$ being $a(r_n) + 0 \times b(r_n)^{1/2}$ and $\mathcal{N}(0)$ being $(2\pi)^{-1/2}$. In [22] Hayman has axiomatized, so to speak, a large class of power series $g(u)$ for which the conclusions (3.15) and (3.16) hold true. Hayman calls these functions $g(u)$ admissible. We present the exact definition in a moment, but first the prototypical example: $g(u) = e^u$. Then, $a(r) = r, b(r) = r, r_n = n$, and from (3.16)

$$\frac{1}{n!} \sim \frac{e^n}{n^a (2\pi n)^{1/2}}.$$

Stirling’s formula (upside down). Hence the title of [22]. Now for the definition.

Observe that the characteristic function of $X_r$ is $g(re^{i\theta}) / g(r)$; by Section 3.5 a necessary and sufficient condition for $X_r$ to be asymptotically normal is

$$e^{-i\theta a(r)/b(r)^{1/2}} \frac{g(re^{i\theta}/b(r)^{1/2})}{g(r)} \to e^{-\theta^2/2}.$$
Asymptotic Normality in Enumeration

**Definition 3.9.1** A power series \( g(z) \) convergent for \( |z| < R \), with \( 0 < R \leq \infty \), is admissible provided we have a function \( \delta(r) \), \( 0 < \delta(r) < \pi \), defined for \( R_0 < r < R \) such that

\[
g(re^{i\theta}) \sim g(r) \exp(i\theta a(r) - (1/2)\theta^2 b(r)), \quad \text{as } r \to R,
\]

uniformly for \( |\theta| \leq \delta(r) \), while uniformly for \( \delta(r) \leq |\theta| \leq \pi \).

It is also required that

\[
b(r) \to +\infty, \quad \text{as } r \to R.
\]

Deriving (3.15) and (3.16) for an admissible function is a simple application of the circle method. The value of Hayman’s paper lies in its characterization of a large class of power series to which the circle method is applicable. Numerous properties of admissible functions are proven, as well as closure of the class under a number of operations. The paper provides a template of how to use the circle method effectively. For example, the proofs in Section 3.5 are found by making simple modifications of this template. The later and deeper sections of the paper have not been explored by combinatorialists whose immediate intent is applications to enumeration.

In later years McKay, jointly with a variety of co-authors, has developed a new method of asymptotic enumeration that follows this same general pattern. The seminal example is the generating function

\[
f(x) = \prod_{1 \leq i < j \leq n} (1 + x_i x_j),
\]

treated by McKay and Wormald in [27]. The coefficient

\[
[x_1^{d_1} \cdots x_n^{d_n}] f(x)
\]

is the number of simple, labeled graphs on \( n \) vertices having degree sequence \( d_1, \ldots, d_n \). For certain degree sequences the coefficient can be estimated by the multi-dimensional circle method. For the multi-dimensional integral a set of radii \( r_1, \ldots, r_n \) is needed, and these are found by solving a system of saddlepoint equations. The main contribution to the integral might arise from \( x_i = r_i e^{i\theta} \) with \( |\theta| \) all small, but in some instances a host of small contributing regions is identified. Around the points that do contribute, the shape of the integrand is found to be asymptotically normal, now a multivariate normal, and the estimation of the primary integral involves diagonalizing the Hermitian covariance matrix. Complications arise due to the fact that the number of variables is growing along with \( n \). This means that the Taylor series used in the primary integrals cannot stop with a linear and quadratic, but must go at least, typically, to quintic. Estimating the integral of these quintic terms is a new challenge. The necessary argument that the contribution to the integral from outside the primary region is negligible becomes quite complicated and combinatorial in nature. Although similar issues recur in each new application of this method, no useful
general systematization has yet been proposed. The intermediate algebra and analysis can become too large to handle without computer-based methods, yet the final result is typically a succinct and elegant asymptotic formula. These formulas strikingly agree with formulas found for other ranges of the parameters and proven by totally different methods.

References


References


References
