STUDENT MATHEMATICAL LIBRARY
Volume 71

Asymptopia

Joel Spencer
with Laura Florescu

AMS
American Mathematical Society

Asymptopia

STUDENT MATHEMATICAL LIBRARY
Volume 71

Asymptopia

Joel Spencer
with Laura Florescu

American Mathematical Society
Providence, Rhode Island

2010 *Mathematics Subject Classification.* Primary 05–01, 05A16; Secondary 05C80, 68W40, 11A41, 60C05.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-71

Library of Congress Cataloging-in-Publication Data

Spencer, Joel H., author.
Asymptopia / Joel Spencer with Laura Florescu.
page cm. — (Student mathematical library ; volume 71)
Includes bibliographical references and index.
ISBN 978-1-404-0904-3 (alk. paper)
1. Combinatorial analysis. 2. Combinatorial enumeration problems. 3. Asymptotic expansions. 4. Ramsey numbers—Asymptotic theory. I. Florescu, Laura. II. Title.

QA164.S638 2014
511′.4—dc23

2013049249

10 9 8 7 6 5 4 3 2 1 14 13 12 11 10 09

Contents

Preface

I was 21 years when I wrote this song
I'm 22 now, but I won't be for long
Time hurries on
And the leaves that are green turn to brown
– Paul Simon, *Leaves that Are Green*

1968 was a tumultuous year. America was convulsed by the Vietnam War, nowhere more than on college campuses. The assassinations of Martin Luther King and of Robert Kennedy tore at the nation's heart. The Democratic convention in Chicago was marked by violent riots. America, for many, had become Amerika, the villain. "Do your own thing" was the admonition that resonated so powerfully. Resist authority. Nonconformity was the supreme virtue. For this fledgling mathematician it was a critical juncture. I had left graduate school without a degree. Would my talents find a focus in this chaotic world? My mind swirled with mathematical ideas, but I seemed unable to turn these ideas into a cohesive whole.

Then I met Paul Erdős. Everyone called him *Uncle Paul*.

While others spoke constantly of it, nonconformity was always Uncle Paul's modus operandi. He had no job; he worked constantly. He had no home; the world was his home. Possessions were a nuisance; money a bore. Paul lived on a web of trust, traveling ceaselessly

from center to center spreading his mathematical pollen. "Prove and Conjecture!" was his constant refrain.

Were we, in those halcyon days, *students* of Uncle Paul. I think the word inadequate and inaccurate. Better to say that we were *disciples* of Paul Erdős. We (and the list is long indeed) had energy and talent. Paul, through his actions and his theorems and his conjectures and every fibre of his being, showed us the Temple of Mathematics. The Pages of The Book were there, we had only to open them. Does there exist for all sufficiently large n a triangle free graph on n vertices which does not contain an independent set of size $\sqrt{n \ln n}$? We had no doubts—the answer was either yes or no. The answer was in The Book. Pure thought—our thought—would allow its reading.

I would sit with Uncle Paul and discuss an open problem. Paul would have a blank pad of paper on his lap. "Suppose," he would say in his strong Hungarian accent,[1] "we set

$$p = \sqrt{\frac{\ln n}{n}}."$$

He would write the formula for p on the blank page and nothing else. Then his mind sped on, showing how this particular value of p led to the solution. How, I wondered, did Uncle Paul know which value of p to take?

The final form of mathematics, the form that students see in textbooks, was described by Bertrand Russell:

> Mathematics, rightly viewed, possesses not only truth, but supreme beauty—a beauty cold and austere, like that of sculpture, without appeal to any part of our weaker nature, without the gorgeous trappings of painting or music, yet sublimely pure, and capable of a stern perfection such as only the greatest art can show. The true spirit of delight, the exaltation, the sense of being more than Man, which is the touchstone of the highest excellence, is to be found in mathematics as surely as in poetry.

[1] The documentary film "N is a Number" by George Csicsery [**Csi93**], available on the web, shows Uncle Paul in action.

Doing mathematics is anything but austere. As an undergraduate teacher of mine, Gian-Carlo Rota, put it:

> A mathematician's work is mostly a tangle of guesswork, analogy, wishful thinking and frustration, and proof, far from being the core of discovery, is more often than not a way of making sure that our minds are not playing tricks.

That said, the "guesswork" can be finely honed. Uncle Paul's selection of the right p did not come at random. Brilliance, of course, is more than helpful. But we mortals can also sometimes succeed.

Paul Erdős lived[2] in Asymptopia. Primes less than n, graphs with v vertices, random walks of t steps—Erdős was fascinated by the limiting behavior as the variables approached, but never reached, infinity. Asymptotics is very much an art. In his masterwork, *The Periodic Table*, Primo Levi speaks of the personalities of the various elements. A chemist will feel when atoms want or do not want to bind. In asymptotics the various functions $n \ln n$, n^2, $\frac{\ln n}{n}$, $\sqrt{\ln n}$, $\frac{1}{n \ln n}$ all have distinct personalities. Erdős knew these functions as personal friends. This author had the great privilege and joy of learning directly from Paul Erdős. It is my hope that these insights may be passed on, that the reader may similarly feel which function has the right temperament for a given task.

My decision to write this work evolved over many years, and it was my students who opened my eyes. I would teach courses in discrete mathematics, probability, Ramsey theory, graph theory, the probabilistic method, number theory, and other areas. I would carefully give, for example, Erdős's classic result (Theorem 7.1) on Ramsey numbers: If

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 1,$$

then $R(k,k) > n$. I spent much less time on the asymptotic implication (7.6), that $R(k,k) \geq (1+o(1))\frac{k}{e\sqrt{2}}\sqrt{2}^k$. My students showed me

[2] Erdős's breadth was extraordinary. This refers to only one aspect of his oeuvre.

that Asymptopia deserved its own emphasis. A facility with asymptotic calculations could be taught, could be learned, and was a highly pragmatic element of their mathematical education.

Laura Florescu began her graduate studies at the Courant Institute in the fall of 2012. Almost immediately we began our study of mathematical results, old and new, and began work on open questions. This pursuit happily continues to this day. Early on, with this project still in its nascent phase, Laura graciously offered her assistance. We have discussed ideas for the various chapters and sections together. Some of the ideas, such as giving a proof of the Law of the Iterated Logarithm, originated entirely from her and all of the ideas were jointly discussed. She has written early drafts of many sections. (However, all errors in the final copy, what you are reading now, are my responsibility.) I hope that this project has been as much a learning experience for Laura as it has been for me. With her talents and energy, Laura has a bright future ahead of her. Thank you, Laura.

My editor, Ina Mette, deserves special recognition. We have known each other for many years and I had always wanted to write a book under her editorship. Conversations about this current work took place over a long period of time. Through lunches at Lure in New York, at Central Kávéház in Budapest, through numerous emails and phone calls, the outlines of this current work came into focus. Ina has always been insightful in her suggestions and fully supportive of my ofttimes ill-defined ruminations. Thank you, Ina.

1968 was special for me personally as well as professionally. It was the year I married my wife Mary Ann, whom I wish to thank once again for her assistance, encouragement, and understanding. Without her, this enterprise would have had little meaning.

Joel Spencer
New York
Fall, 2013

A Reader's Guide

> I have never let my schooling interfere with my education.
> – Mark Twain

The Student Mathematical Library is aimed at undergraduate students, but our focus is somewhat broader. We may also envision a graduate student looking for a pragmatic view of asymptotic calculations. We may also envision a high school student learning new mathematical relationships. The common denominator is a love of mathematics. To the largest degree possible, we have strived to make this work self-contained. The reader should be aware, however, of certain assumptions.

Calculus. We do assume a knowledge of first year calculus, as taught in U.S. colleges and, often, high schools. Differentiation and integration is done without proof. The definite integral

$$\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

is assumed; this appears with surprising frequency. We do *not* use differential equations, nor partial differential equations, nor algebra, nor topology. We do *not* use material from the course frequently called (in the U.S.) analysis. In particular, all interchanges of $\lim_n \int f_n(x)dx$ and $\int \lim_n f_n(x)dx$ are done from scratch.

Probability. A number of basic distributions are considered in this work. These include the binomial, the Poisson, and the Gaussian distributions. We have defined these when they appear. Still, some prior knowledge of the notions of random variable, expectation, variance, and independence would be helpful to the student.

Graph Theory. We do not assume a knowledge of graph theory. Still, some prior knowledge of what a graph is, as a set of vertices and edges, would be helpful. We examine the random graph $G(n, p)$. Again, a prior familiarity would be helpful but not necessary.

Number Theory. We expect the reader to know what prime numbers are and to know the unique factorization of positive integers into primes. Otherwise, our presentation of number theory is self-contained.

Algorithms. The mathematical analysis of algorithms is a fascinating subject. Here we give some glimpses into the analyses, but our study of algorithms is self-contained. Certainly, no actual programming is needed.

Our final chapter, Really Big Numbers!, is different in flavor. This author has always been fascinated with big numbers. This chapter is basically a paper written for the *American Mathematical Monthly* three decades ago. Some of the material uses ordinal numbers, like ω^ω, which may be new to the reader.

We sometimes skirt a topic, pulling from it only some asymptotic aspects. This is particularly noticeable in Ramsey theory, one of our favorite topics.

Certain sections are technically quite complicated and are labelled as such. They may be skipped without losing the thread of the argument.

Facility with logarithms is assumed throughout. We use $\ln x$ for natural logarithm and $\lg x$ for the logarithm to the base two.

Asymptopia is a beautiful world. Enjoy!

Chapter 0

An Infinity of Primes

> Truth is on a curve whose asymptote our spirit follows eternally.
> – Léo Errera

We begin with one of the greatest theorems in mathematics.

Theorem 0.1. *There is an infinite number of primes.*

Our proof is not that of Euclid and not better than the proof of Euclid, but it illustrates the theme of this work: looking at the mathematical world through an asymptotic lens.

We begin as Euclid did. Assume Theorem 0.1 is false. Let $p_1, \ldots, p_r$ be a listing of all of the primes. For any nonnegative integer s, the unique factorization theorem states that there is a unique way to express

$$s = p_1^{\alpha_1} p_2^{\alpha_2} \cdot p_r^{\alpha_r}, \tag{0.1}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_r$ are nonnegative integers. We turn this into an *encoding* of the nonnegative integers by creating a map Ψ,

$$\Psi(s) = (\alpha_1, \ldots, \alpha_r). \tag{0.2}$$

Let $n \geq 2$ be arbitrary, though in the application below it shall be large. The integers s, $1 \leq s \leq n$, are each mapped by Ψ to a vector

of length r. How many possibilities are there for the values $\Psi(s)$? We give an *upper bound*. For each $1 \leq i \leq r$ the value α_i must satisfy

$$p_i^{\alpha_i} \leq s \leq n. \tag{0.3}$$

Thus

$$\alpha_i \leq \log_{p_i} n \leq \log_2 n. \tag{0.4}$$

As α_i is a nonnegative integer, there are at most $1+\log_2 n$ possibilities for it. With $n \geq 2$, the number of possibilities is at most $2\log_2 n$. (These kinds of gross upper bounds will appear quite often, and a large part of the *art* of Asymptopia is knowing when to use them and when not to use them.) Thus the number of possible values of $\Psi(s) = (\alpha_1, \ldots, \alpha_r)$ is at most $2^r(\log_2 n)^r$. The vectors $(\alpha_1, \ldots, \alpha_r)$ uniquely determine s by equation (0.1). We have n *different* values $\Psi(s)$. We deduce that

$$2^r(\log_2 n)^r \geq n. \tag{0.5}$$

The above is all true for any $n \geq 2$. But now we apply an asymptotic lens and consider (0.5) asymptotically in n. The left-hand side is a constant times a fixed power of the logarithm function. We know (more on this in §2.4) that any fixed power of $\ln n$ grows slower that any fixed positive power of n, so slower than n itself. This means that for n sufficiently large, (0.5) must fail! We have achieved our *reductio ad absurdum*, the assumption that the number of primes is finite must be false, and Theorem 0.1 is true.

Remark. We do not need the full power of the unique factorization theorem. It suffices to know that every s has *some* representation equation (0.1) as the product of primes to powers. Then for each of the $1 \leq s \leq n$, select arbitrarily one such representation as $\Psi(s)$. One still has n distinct $\Psi(s)$ and at most $2^r(\log_2 n)^r$ possible vectors $(\alpha_1, \ldots, \alpha_r)$.

We have worked out this argument in some detail. For those comfortable with asymptotics, especially Definition 2.8 and §2.4, it would go, informally, something like this: there are n values $\Psi(s)$, $1 \leq s \leq n$ and logarithmically many values for each coordinate α_i; therefore, there are *polylog* many vectors, but polylog grows more slowly than linearly.

We now use this same approach to prove a much stronger result:

Theorem 0.2. *The summation of the reciprocals of the primes diverges.*

Proof. Again, assume not. So

$$\sum_p \frac{1}{p} = C \tag{0.6}$$

for some constant C. (We shall use $\sum_p$ to indicate the sum over all primes p.) Label the primes $p_1, p_2, \ldots$. As equation (0.6) is a convergent sum of positive terms, at some point it reaches $C - \frac{1}{2}$. That is, there exists r such that

$$\sum_{i>r} \frac{1}{p_i} < \frac{1}{2}. \tag{0.7}$$

Call the primes $p_1, \ldots, p_r$ small and the other primes large. Call an integer s *rare* if all of its prime factors are small; otherwise, call s *ordinary*. Again consider the s, $1 \le s \le n$. The rare integers have a factorization (0.1) and so, as with Theorem 0.1, their number is at most most $2^r(\log_2 n)^r$, polylog to the cognoscenti.

What about the ordinary s? For each ordinary s there is some (perhaps several) large prime p dividing it. For a given prime p, the number of elements in $1 \le s \le n$ divisible by it is $\lfloor \frac{n}{p} \rfloor$, which is at most $\frac{n}{p}$. Thus, the total number of ordinary s is at most $\sum \frac{n}{p}$, where p now ranges over the large primes. From (0.7), this is less than $\frac{n}{2}$. Of the n values of s, less than $\frac{n}{2}$ are ordinary, so at least $\frac{n}{2}$ are rare. Thus

$$2^r(\log_2 n)^r \ge \frac{n}{2}. \tag{0.8}$$

As with (0.5), for n sufficiently large (0.5) must fail! We have achieved our *reductio ad absurdum*: the assumption that the sum of the reciprocals of the primes is finite must be false, and Theorem 0.2 is true.

Remark. There was no need to cut large and small primes precisely via (0.7). The same argument works if the sum of the reciprocals of the large primes is less than one.

Chapter 1

Stirling's Formula

> The voyage of discovery lies not in seeking new horizons, but in seeking with new eyes.
> – Marcel Proust

Surely the most beautiful asymptotic formula in all of mathematics is Stirling's formula:

$$n! \sim n^n e^{-n} \sqrt{2\pi n}. \tag{1.1}$$

How do the two most important fundamental constants of mathematics, e and π, find their way into an asymptotic formula for the product of integers? We give two very different arguments (one will not show the full formula) that, between them, illustrate a good number of basic asymptotic methods. The formal language of Asymptopia, such as $o(n)$ and $O(n)$, is deferred to Chapter 2. Two further arguments for Stirling's formula are given in §3.2.3.

1.1. Asymptotic Estimation of an Integral

Consider the integral

$$I_n = \int_0^\infty x^n e^{-x} dx. \tag{1.2}$$

A standard result[1] of freshman calculus, done by integration by parts, is that

$$I_n = n! \tag{1.3}$$

Our problem now is to estimate the integral of (1.2).

- Asymptotically, integrals are often dominated by the largest value of the function being integrated.

Let us set

$$y = y_n(x) = x^n e^{-x} \text{ and } z = z_n(x) = \ln y = n \ln x - x. \tag{1.4}$$

The graph of $y(x)$ when $n = 2$ is unclear, but with $n = 10$ it is looking somewhat like the bell shaped curve. What is going on?

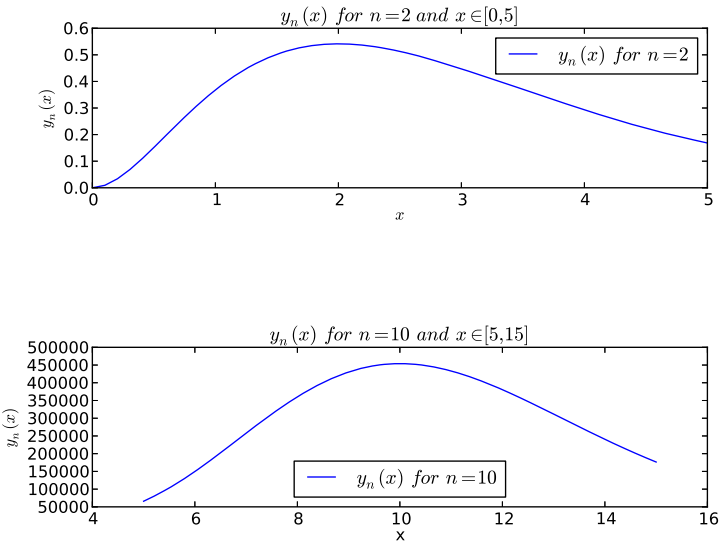


Figure 1. The first plot shows the function $y_n(x) = x^n e^{-x}$ for $n = 2$ and $x \in [0, 5]$, while the second plot shows the same function for $n = 10$ and $x \in [5, 15]$.

[1] We shall assume first-year calculus results throughout this work.

Setting $z' = nx^{-1} - 1 = 0$, we find that $z(x)$ (and hence $y(x)$) has a unique maximum at $x = n$ and that $z(x)$ (and hence $y(x)$) is increasing in $[0, n]$ and decreasing in $[n, \infty)$.

Let us compare $y(n) = n^n e^{-n}$ with values of $y(x)$ when x is "near" n. For example, take $x = 1.1n$.

$$y(1.1n) = (1.1n)^n e^{-1.1n} = y(n)(1.1e^{-0.1})^n. \tag{1.5}$$

But $1.1e^{-0.1} = 0.9953\ldots$. While this number is close to 1, it is a constant less than 1, and so $y(1.1n)$ is exponentially smaller than $y(n)$. Values near $1.1n$ will make a negligible contribution to the integral. Let us move closer and try $x = n + 1$. Now

$$y(n+1) = (n+1)^n e^{-n-1} = y(n)\left(1 + \frac{1}{n}\right)^n e^{-1}. \tag{1.6}$$

As $(1 + \frac{1}{n})^n \sim e$, $y(n+1) \sim y(n)$, and so values near $x = n + 1$ do contribute substantially to the integral.

Moving from $x = n$ in the positive direction (the negative is similar), the function $y = y(x)$ decreases. If we move by 1 (to $x = n+1$), we do not yet "see" the decrease, while if we move by $0.1n$ (to $x = 1.1n$), the decrease is so strong that the function has effectively disappeared. (Yes, $y(1.1n)$ is large in an absolute sense, but it is small relative to $y(n)$.) How do we move out from $x = n$ so that we can effectively see the decrease in $y = y(x)$? This is a question of *scaling*.

- Scaling is the art of asymptotic integration.

Let us look more carefully at $z(x)$ near $x = n$. Note that an additive change in $z(x)$ means a multiplicative change in $y(x) = e^{z(x)}$. We have $z'(x) = nx^{-1} - 1 = 0$ at $x = n$. The second derivative $z''(x) = -nx^{-2}$, so that $z''(n) = -n^{-1}$. We can write the first terms of the Taylor series for $z(x)$ about $x = n$:

$$z(n+\epsilon) = z(n) - \frac{1}{2n}\epsilon^2 + \cdots. \tag{1.7}$$

This gives us a heuristic explanation for our earlier calculations. When $\epsilon = 1$, we have $\frac{1}{2n}\epsilon^2 \sim 0$, so $z(n+\epsilon) = z(n) + o(1)$ and thus $y(n+\epsilon) \sim y(n)$. When $\epsilon = 0.1n$, we have the opposite as $\frac{1}{2n}\epsilon^2$ is large. The middle ground is given when ϵ^2 is on the order of n or

when ϵ is on the order of $\sqrt{n}$. We are thus led to the scaling $\epsilon = \lambda\sqrt{n}$, or

$$x = n + \lambda\sqrt{n}. \tag{1.8}$$

We formally make this substitution in the integral (1.2). Further, we take the factor $y(n) = n^n e^{-n}$ outside the integral so that now the function has maximum value 1. We have scaled both axes. The scaled function is

$$g_n(\lambda) = \frac{y(n+\lambda\sqrt{n})}{y(n)} = (1+\lambda n^{-1/2})^n e^{-\lambda\sqrt{n}}, \tag{1.9}$$

and we find (noting that $dx = \sqrt{n}d\lambda$)

$$I_n = \int_0^\infty x^n e^{-x} dx = n^n e^{-n}\sqrt{n}\int_{-\sqrt{n}}^{+\infty} g_n(\lambda)d\lambda. \tag{1.10}$$

Note that while we have been guided by asymptotic considerations, our calculations up to this point have been exact.

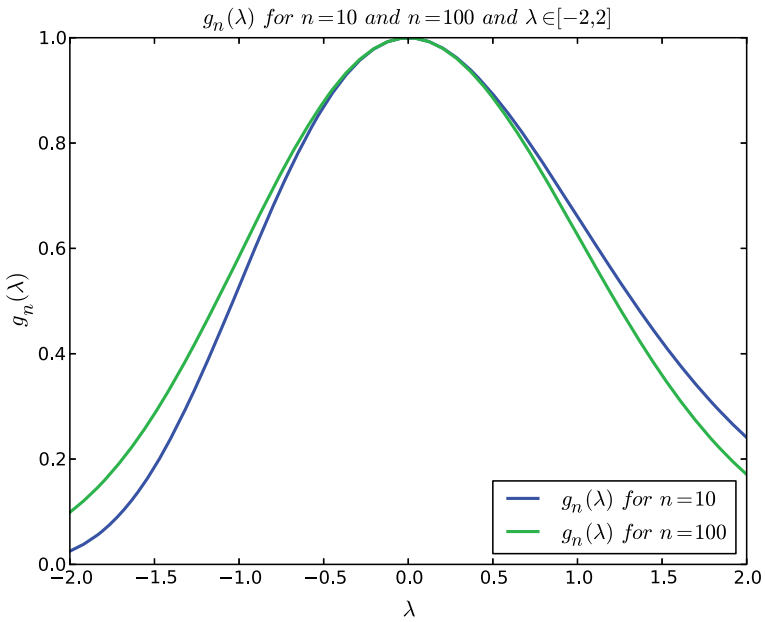


Figure 2. $g_n(\lambda)$ in range $-2 \le \lambda \le +2$ for $n = 10, 100$.

The Taylor series with error term gives

$$\ln(1+\epsilon) = \epsilon - \frac{1}{2}\epsilon^2 + O(\epsilon^3) \tag{1.11}$$

as $\epsilon \to 0$. Let λ be an arbitrary but fixed real number. Then $\lambda n^{-1/2} \to 0$ so that

$$n\ln(1+\lambda n^{-1/2}) - \lambda n^{1/2} = \lambda n^{1/2} - \frac{1}{2}\lambda^2 + o(1) - \lambda n^{1/2} = -\frac{1}{2}\lambda^2 + o(1) \tag{1.12}$$

and

$$g_n(\lambda) \to e^{-\lambda^2/2}. \tag{1.13}$$

That is, when properly scaled, the function $y = x^n e^{-x}$ looks like the bell shaped curve in Figure 2.

Now we would *like* to say

$$\lim_{n\to\infty} \int_{-\sqrt{n}}^{+\infty} g_n(\lambda)d\lambda = \int_{-\infty}^{\infty} e^{-\lambda^2/2}d\lambda = \sqrt{2\pi}\ . \tag{1.14}$$

Justification of the interchange of limits in the integration of a sequence of functions is one of the most basic and most subtle problems discussed in analysis. Here is a sample theorem: *If* $g_n(\lambda)$ are continuous functions on an interval $[a,b]$ *and* $\lim_{n\to\infty} g_n(\lambda) = g(\lambda)$ for all $\lambda \in [a,b]$, then $\lim_{n\to\infty} \int_a^b g_n(\lambda)d\lambda = \int_a^b g(\lambda)d\lambda$.

In our example the $g_n(\lambda)$ are indeed continuous and $\lim_{n\to\infty} g_n(\lambda)$ is given by (1.13). But there are three difficulties:

(1) The left-hand side of the integral in (1.14) is $-\sqrt{n}$.

(2) The right-hand side of the integral in (1.14) is ∞.

(3) We will *not* be assuming results from analysis in this book.

A natural approach is to approximate $g_n(\lambda)$ by $e^{-\lambda^2/2}$. The difficulty is that this approximation is not valid throughout the limits of integration. For example, with $\lambda = \sqrt{n}$, $g_n(\lambda) = (2/e)^n$ is not close to $e^{-\lambda^2/2} = e^{-n/2}$. Let us re-examine (1.12) with the error term from

the Taylor series (1.11). Thus *if* $\lambda n^{-1/2} \to 0$, then

(1.15)
$$n\ln(1+\lambda n^{-1/2}) - \lambda n^{1/2} = \lambda n^{1/2} - \frac{1}{2}\lambda^2 + O(\lambda^3 n^{-1/2}) - \lambda n^{1/2}$$
$$= -\frac{1}{2}\lambda^2 + o(\lambda^2 n^{-1/2}).$$

We now see that the approximation of $g_n(\lambda)$ by $e^{-\lambda^2/2}$ is good as long as $\lambda^2 n^{-1/2} \to 0$, that is, for $\lambda = o(n^{1/4})$. With this in mind let us split the range $[-\sqrt{n}, \infty)$ into a middle range

$$MID = [-L(n), +L(n)],$$

and the two sides

$$LEFT = [-\sqrt{n}, -L(n)]$$

and

$$RIGHT = [L(n), \infty).$$

How should we choose $L(n)$? The middle range should be big enough that most of the integral lies under it but small enough so that the approximation with the bell shaped curve remains valid. The first condition will require that $L(n) \to \infty$ and the second that $L(n) = o(n^{1/4})$. This leaves a lot of room and, indeed, any reasonable $L(n)$ satisfying these criteria would work for our purposes. For definiteness let us set

(1.16)
$$L(n) = n^{1/8}.$$

1.1.1. MID. Let us take the most important region, *MID*, first. Guided by the notion that $g_n(\lambda)$ and $e^{-\lambda^2/2}$ will be close, we define an error[2] function

(1.17)
$$E_n(\lambda) = g_n(\lambda)/e^{-\lambda^2/2}$$

so that we have the exact expression

(1.18)
$$\ln E_n(\lambda) = n\ln(1+\lambda n^{-1/2}) - \lambda\sqrt{n} + \frac{\lambda^2}{2}.$$

As $\lambda n^{-1/2} \to 0$ in *MID*, we can apply the Taylor series to $\ln(1+\epsilon)$ with $\epsilon = \lambda n^{-1/2}$. The first two terms cancel the $\lambda\sqrt{n}$ and $\lambda^2/2$ terms,

[2]Error does not mean mistake!

which is not so surprising as we *designed* the error to be close to one. We employ the Taylor series to two terms with an error term,

$$n\ln(1+\lambda n^{-1/2}) = \lambda\sqrt{n} - \frac{\lambda^2}{2} + n\frac{x^3}{3!}. \tag{1.19}$$

Here x lies somewhere between 0 and $\lambda n^{-1/2}$. As $|\lambda n^{-1/2}| \le n^{-3/8}$, we can bound

$$|n\frac{x^3}{3}| \le \frac{1}{3}n^{-1/8}. \tag{1.20}$$

Thus $|\ln E_n(\lambda)| \le \frac{1}{3}n^{-1/8}$ throughout λ. Critically, this is a *uniform* bound, which holds for all λ in *MID* simultaneously. As $\ln(E_n(\lambda))$ is small, $E_n(\lambda) - 1$ will also be small. Think of $y = \ln(E_n(\lambda))$, with y small $e^y - 1 \sim y$. But to get a rigorous upper bound, let us use a rougher bound $|e^y - 1| \le 2y$, valid when $|y|$ is sufficiently small. For n large, $\frac{1}{3}n^{-1/8}$ will be small and so

$$|E_n(\lambda) - 1| \le \frac{2}{3}n^{-1/8} \tag{1.21}$$

so that

$$|g_n(\lambda) - e^{-\lambda^2/2}| = e^{-\lambda^2/2}|E_n(\lambda) - 1| \le \frac{2}{3}n^{-1/8}e^{-\lambda^2/2} \tag{1.22}$$

and

$$\begin{aligned}\left|\int_{MID} g_n(\lambda) - e^{-\lambda^2/2}d\lambda\right| &\le \int_{MID} |p_n(\lambda) - e^{-\lambda^2/2}|d\lambda \\ &\le \frac{2}{3}n^{-1/8}\int_{MID} e^{-\lambda^2/2}d\lambda.\end{aligned} \tag{1.23}$$

The final integral is less than $\sqrt{2\pi}$, the integral over all λ. The constants are not important, we have bounded the difference in the integrals of $g_n(\lambda)$ and $e^{-\lambda^2/2}$ over *MID* by a constant times $n^{-1/8}$ which in the limit approaches zero.

1.1.2. LEFT. It remains to show that *LEFT* and *RIGHT* give negligible contributions to $\int g_n(\lambda)d\lambda$. Note that we do *not* need asymptotic values of $\int g_n(\lambda)d\lambda$ over *LEFT* or *RIGHT*, only that they approach zero. Thus we can employ a rough (but true) upper bound to $g_n(\lambda)$. The left-hand side is easier. The function $g_n(\lambda)$ is increasing from $-\sqrt{n}$ to $-L(n) = -n^{1/8}$. At $-n^{1/8}$, $g_n(\lambda) \sim e^{-\lambda^2/2} \sim e^{-n^{1/4}/2}$. Since the length of range *LEFT* is less than $\sqrt{n}$, the integral is at

most $\sqrt{n}e^{-n^{1/4}/2}$. The exponential decay dominates the square root growth, and this function goes to zero with n. As this was an upper bound, $\int_{LEFT} g_n(\lambda)d\lambda \to 0$.

1.1.3. RIGHT. The interval *RIGHT* is more difficult for two reasons: The interval has infinite length so that bounding a single value will not be sufficient. More worrisome, the estimate of $\ln(1+\epsilon)$ by $\epsilon - \frac{1}{2}\epsilon^2$ is only valid for ϵ small. We require upper bounds that work for the entire range of ϵ. The following specific bounds ((1.24) and (1.25) are included for completeness) are often useful:

$$\ln(1+\epsilon) \le \epsilon - \frac{1}{2}\epsilon^2 \text{ when } -1 < \epsilon \le 0, \tag{1.24}$$

$$\ln(1+\epsilon) \le \epsilon - \frac{1}{4}\epsilon^2 \text{ when } 0 < \epsilon \le 1, \tag{1.25}$$

$$\ln(1+\epsilon) \le 0.7\epsilon \text{ when } \epsilon > 1. \tag{1.26}$$

We break $RIGHT = [n^{1/8}, \infty)$ into two parts. We set

$$NEARRIGHT = [n^{1/8}, n^{1/2}]$$

and

$$FARRIGHT = [n^{1/2}, \infty),$$

reflecting the ranges for the bounds (1.25) and (1.26) with $\epsilon = \lambda n^{-1/2}$. For *NEARRIGHT* we employ the argument used for *LEFT*. The function $g_n(\lambda)$ is decreasing for λ positive and is $\sim e^{-n^{1/4}/2}$ at $n^{1/8}$. As *NEARRIGHT* has length less than $\sqrt{n}$, $\int g_n(\lambda)d\lambda$ over *NEARRIGHT* is at most $\sqrt{n}e^{-n^{1/4}/2}$ which goes to zero.

In *FARRIGHT*, (1.26) gives that

$$n\ln(1+\lambda n^{-1/2}) - \lambda n^{1/2} \le 0.7\lambda\sqrt{n} - \lambda\sqrt{n} \le -0.3\lambda\sqrt{n}. \tag{1.27}$$

In this interval $g_n(\lambda)$ is thus bounded by the exponentially decaying function $\exp -0.3\lambda\sqrt{n}$. Thus

$$\int_{\sqrt{n}}^{\infty} g_n(\lambda)d\lambda < \int_{\sqrt{n}}^{\infty} e^{-0.3\lambda\sqrt{n}}d\lambda = \frac{1}{0.3\sqrt{n}}e^{-0.3n}, \tag{1.28}$$

and this also goes to zero as $n \to \infty$.

We have shown that the integrals of $g_n(\lambda)$ over *LEFT*, *NEARRIGHT*, and *FARRIGHT* all approach zero and that the integral of $g_n(\lambda)$ over *MID* approached $\sqrt{2\pi}$. Putting it all together, the integral of $g_n(\lambda)$ over $[-\sqrt{n}, \infty)$ does indeed approach $\sqrt{2\pi}$.

Whew! Let us take two general principles from this example:

- Crude upper bounds can be used for negligible terms as long as they stay negligible.
- Terms that are extremely small often require quite a bit of work.

1.2. Approximating Sums by Trapezoids

With this method we will not achieve the full Stirling's formula (1.1) but only

$$n! \sim K n^n e^{-n} \sqrt{n} \tag{1.29}$$

for *some* positive constant K. Our approach follows the classic work [**CR96**] of Richard Courant. We are pleased to reference the eponymous founder of our mathematical home, the Courant Institute.

Our approach is to estimate the logarithm of $n!$ via the formula

$$S_n := \ln(n!) = \sum_{k=1}^{n} \ln(k). \tag{1.30}$$

The notion is that S_n should be close to the integral of the function $\ln(x)$ between $x = 1$ and $x = n$. We set

$$I_n := \int_1^n \ln(x)dx = [x\ln(x) - x]_1^n = n\ln(n) - n + 1. \tag{1.31}$$

Let T_n be the value for the approximation of the integral I_n via the trapezoidal rule using step sizes 1. That is, we estimate $\int_i^{i+1} f(x)dx$ by $\frac{1}{2}(f(i) + f(i+1))$. Summing over $1 \le i \le n-1$,

$$T_n = \frac{1}{2}\ln(1) + \sum_{k=2}^{n-1} \ln(k) + \frac{1}{2}\ln(n) = S_n - \frac{1}{2}\ln(n). \tag{1.32}$$

Set

$$E_n = I_n - T_n \tag{1.33}$$

to be the error when approximating the integral of $\ln(x)$ by the trapezoidal rule. For $1 \le k \le n-1$, let S_k denote the "sliver" of area under the curve $y = \ln(x)$ for $k \le x \le k+1$ but over the straight line between $(k, \ln(k))$ and $(k+1, \ln(k+1))$. The curve is over the straight line as the curve is concave. Then

$$E_n = \sum_{k=1}^{n-1} \mu(S_k), \tag{1.34}$$

where μ denotes the area.

Our goal is to bound the error.

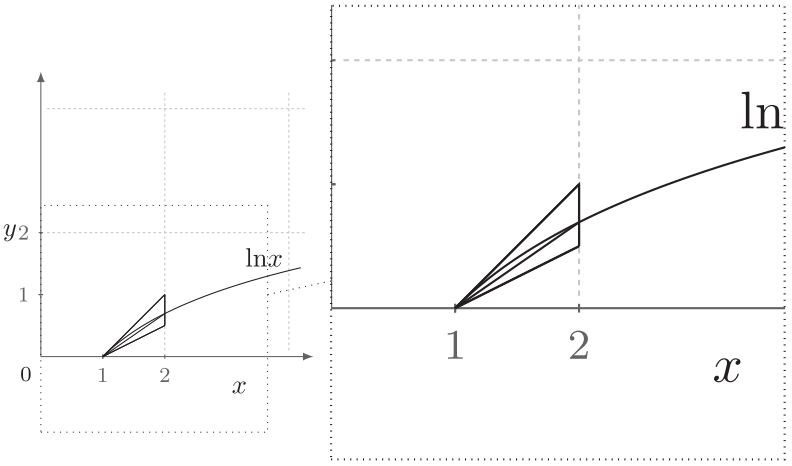


Figure 3. The sliver (shown here with $k = 1$) lies inside the triangle whose upper and lower lines have slopes $\frac{1}{k}$, $\frac{1}{k+1}$, respectively.

Theorem 1.1. *E_n approaches a finite limit c as $n \to \infty$. Equivalently,*

$$\lim_{n\to\infty} \sum_{k=n}^{\infty} \mu(S_k) = 0. \tag{1.35}$$

Assuming Theorem 1.1, (1.30)–(1.32) yield

(1.36)

$$\ln(n!) = T_n + \frac{1}{2}\ln n = I_n - E_n + \frac{1}{2}\ln n = n\ln n - n + 1 - c + o(1) + \frac{1}{2}\ln n.$$

Exponentiating both sides

$$n! \sim n^n e^{-n} \sqrt{n} e^{1-c} \tag{1.37}$$

giving the desired (1.29) with $K = e^{1-c}$.

Now, how do we show Theorem 1.1? We consider $\mu(S_k)$ in Asymptopia, as $k \to \infty$. Roughly,[3] $\mu(S_k)$ is the error between the integral from k to $k+1$ of $f(x) = \ln x$ and the straight line approximation of $f(x)$. This error is caused by the *second* derivative of $f(x)$. (Had the second derivative been zero, the straight line would have been the precise function.) Here, the second derivative $f''(x) = -x^{-2}$ is on the order of k^{-2}, and the interval has length 1, so we feel the error should be on the order of k^{-2}. As k^{-2} is decreasing sufficiently quickly, the infinite sum of $\mu(S_k)$ should converge.

Guided by this intuitive approach, we give an explicit upper bound for $\mu(S_k)$. Observe that it need not be a good upper bound. We still would get convergence of $\sum \mu(S_k)$ even if our upper bound were, say, ten times the actual value.

Here is *one* approach that works. Let $P = (k, \ln k)$, and let $Q = (k+1, \ln(k+1))$. Let C denote the curve $f(x) = \ln x$ in the interval $[k, k+1]$. In the interval $[k, k+1]$, our function $f(x) = \ln x$ has derivative between $\frac{1}{k}$ and $\frac{1}{k+1}$. Let U (upper) be the straight line segment starting at P with slope $\frac{1}{k}$, ending at $x = k+1$. Let L (lower) be the straight line segment starting at P with slope $\frac{1}{k+1}$, ending at $x = k+1$. As the derivative of curve C is always between those of U and L, the curve C is under U and over L. At $x = k+1$, L then is below the curve C, so below the point Q. Thus the straight line PQ lies above the line L. We can then bound $\mu(S_k)$, the area between C and the straight line PQ, by the area between U and L. But this latter area is a triangle. Make the base of the triangle the line from U to L at $x = k+1$ to be the distance from U to L at $x = k+1$, which is precisely the difference of the slopes which is $\frac{1}{k} - \frac{1}{k+1}$. The

[3] An intuitive feel is very useful, but it must be followed up with a rigorous argument!

height of the triangle is then 1, from $x = k$ to $x = k + 1$. We have thus shown

$$\mu(S_k) \leq \frac{1}{2}\left(\frac{1}{k} - \frac{1}{k+1}\right). \tag{1.38}$$

This value is $O(k^{-2})$, and so we achieve convergence. Indeed we have the explicit upper bound

$$\sum_{k=1}^{\infty} \mu(S_k) \leq \sum_{k=1}^{\infty} \frac{1}{2}\left(\frac{1}{k} - \frac{1}{k+1}\right) = \frac{1}{2} \tag{1.39}$$

as the sum telescopes. This yields (1.29), almost Stirling's formula.

1.3. Combining Forces to Estimate the Error

Setting $c = \lim_{n\to\infty} E_n$, define the tail

$$F_n = c - E_n = \sum_{k=n}^{\infty} \mu(S_k). \tag{1.40}$$

Now (1.36) becomes

$$\ln(n!) = n\ln n - n + 1 - c + \frac{1}{2}\ln n + F_n. \tag{1.41}$$

From the proof of Stirling's formula in Section 1.1, we know that $e^{1-c} = \sqrt{2\pi}$. Exponentiating both sides, we may express the result as

$$\frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = e^{F_n}. \tag{1.42}$$

That is, e^{F_n} gives the *error term* in the approximation of Stirling's formula. Since $F_n \to 0$, $e^{F_n} = 1 + F_n(1 + o(1))$ and so

$$\frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = 1 + F_n(1 + o(1)). \tag{1.43}$$

While (1.42) is exact, we do not have a closed form for F_n. Still, we may find it in Asymptopia.

Consider $\mu(S_k)$ more carefully. Parametrizing $y = k+x$, we have[4]

$$\mu(S_k) = \int_0^1 \ln(k+y) - [(1-y)\ln(k) + y\ln(k+1)]dy \tag{1.44}$$

[4]Moving the region of interest to near zero is often times helpful!

as the bracketed term is the equation of the straight line PQ above. From the Taylor series (the asymptotics here are as $k \to \infty$, uniformly over $y \in [0,1]$),

$$\ln(k+y) = \ln k + \frac{1}{k}y - \frac{y^2}{2k^2} + O(k^{-3}). \tag{1.45}$$

As

$$\ln(k+1) = \ln k + \frac{1}{k} - \frac{1}{2k^2} + O(k^{-3}), \tag{1.46}$$

we find

$$(1-y)\ln(k) + y\ln(k+1) = \ln k + \frac{1}{k}y + \frac{y}{2k^2} + O(k^{-3}). \tag{1.47}$$

Subtracting[5]

$$\mu(S_k) = \int_0^1 \frac{1}{2k^2}(y - y^2) + O(k^{-3})dy. \tag{1.48}$$

The main part can be integrated precisely, and

$$\mu(S_k) = \frac{1}{12k^2} + O(k^{-3}). \tag{1.49}$$

This allows us to estimate F_n:

$$F_n = \sum_{k=n}^{\infty} \mu(S_k) \sim \int_n^{\infty} \frac{1}{12z^2}dx = \frac{1}{12n}. \tag{1.50}$$

This gives a more precise approximation for $n!$:

$$\frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = \left(1 + \frac{1+o(1)}{12n}\right). \tag{1.51}$$

Indeed, with considerably more care one can show that

$$\frac{1}{12n+1} \le F_n \le \frac{1}{12n}, \tag{1.52}$$

which yields the remarkably close[6] bounds

$$e^{1/(12n+1)} \le \frac{n!}{n^n e^{-n}\sqrt{2\pi n}} \le e^{1/(12n)}, \tag{1.53}$$

which are valid for *all* n.

[5] Caution! Subtracting in Asymptopia is tricky! Often times main terms cancel and the secondary terms become paramount. Even worse, occasionally the secondary terms also cancel and it is the tertiary terms that are important.

[6] Try it for $n = 10$.

1.4. Estimating the Integral More Accurately

Note. This section gets quite technical and should be considered optional.

Let us begin again with the precise formula

$$n! = n^n e^{-n}\sqrt{n}\int_{-\sqrt{n}}^{\infty} g_n(\lambda)d\lambda. \tag{1.54}$$

Our goal is to replicate (1.51) by more accurately estimating $p_n(\lambda)$. Our previous estimate was $e^{-\lambda^2/2}$. Now, however, we will want the estimate to be within an additive $o(n^{-1})$ term. Our previous definition of *MID* will be too broad. Instead we define

$$L(n) = n^{0.01} \tag{1.55}$$

and

$$MID = [-L(n), +L(n)],$$
$$LEFT = [-\sqrt{n}, -L(n)],$$
$$RIGHT = [L(n), \sqrt{n}].$$

The bounds on $\int p_n(\lambda)d\lambda$ are still (this requires checking!) exponentially small, and thus they are not only $o(1)$ but $o(n^{-1})$. This allows us to concentrate on $\int p_n(\lambda)d\lambda$ over our new *MID*. We have $E_n(\lambda)$ and $\ln(E_n(\lambda))$ as in (1.17) and (1.18). Now, however, we need a more accurate Taylor series estimation for $\ln(1+\epsilon)$ with $\epsilon = \lambda n^{-1/2}$. A priori, it is unclear just how many terms we will need.

- Experimentation is part of the art of Asymptopia.

After possibly a number of false starts, examine the Taylor series out to four terms with the error term. From Theorem 2.18

$$\ln(1+\epsilon) = \epsilon - \frac{1}{2}\epsilon^2 + \frac{1}{3}\epsilon^3 - \frac{1}{4}\epsilon^4 + \frac{1}{5}x^5 \tag{1.56}$$

with $|x| \le \epsilon$. Applying this to (1.18), the first two terms cancel as before and

$$\ln(E_n(\lambda)) = \frac{1}{3}\lambda^3 n^{-1/2} - \frac{1}{4}\lambda^4 n^{-1} + \frac{1}{5}n^{-3/2}x^5. \tag{1.57}$$

With our *MID* now narrower, $|\frac{1}{5}n^{-3/2}x^5| \le n^{-1.45}$ which will be negligible for our purposes here. With $y = \ln E_n(\lambda)$ we want to go

from y to $e^y - 1$. Because we need greater accuracy (and after some experimentation!), we bound

$$e^y - 1 = y + \frac{y^2}{2} + O(y^3). \tag{1.58}$$

Thus (1.57) becomes

$$E_n(\lambda) = 1 + \frac{1}{3}\lambda^3 n^{-1/2} + \frac{1}{18}\lambda^6 n^{-1} - \frac{1}{4}\lambda^4 n^{-1} + O(n^{-1.41}). \tag{1.59}$$

(The O term in (1.59) contains several terms of which the largest is $(\lambda^3 n^{-1/2})^3$.)

This gives us a good estimate for $\int g_n(\lambda)d\lambda$ over *MID*:

$$\begin{aligned}&\int_{-L(n)}^{L(n)} g_n(\lambda)d\lambda \\ &\quad = \int_{-L(n)}^{L(n)} e^{-\lambda^2/2}[1 + \frac{1}{3}\lambda^3 n^{-1/2} + \frac{1}{18}\lambda^6 n^{-1} \\ &\qquad\qquad - \frac{1}{4}\lambda^4 n^{-1} + O(n^{-1.41})].d\lambda\end{aligned} \tag{1.60}$$

The contribution of the $O(n^{-1.41})$ term to the integral is $o(n^{-1})$. This is an acceptable error, so we rewrite

$$\begin{aligned}&\int_{-L(n)}^{L(n)} g_n(\lambda)d\lambda \\ &\quad = o(n^{-1}) + \int_{-L(n)}^{L(n)} e^{-\lambda^2/2}[1 + \frac{1}{3}\lambda^3 n^{-1/2} + \frac{1}{18}\lambda^6 n^{-1} - \frac{1}{4}\lambda^4 n^{-1}]d\lambda.\end{aligned} \tag{1.61}$$

We want to replace the limits of integration to $\pm\infty$, but we must pause for a moment as we require an accuracy of $o(n^{-1})$.

Let us give some very rough upper bounds on $\int_{L(n)}^{\infty} \lambda^4 e^{-\lambda^2/2}d\lambda$, as the other side and the smaller powers are similar. We bound $\lambda^4 \le e^{\lambda}$, certainly true for $\lambda \ge L(n)$. Then

$$\int_{L(n)}^{\infty} \lambda^4 e^{-\lambda^2/2}d\lambda \le \int_{L(n)}^{\infty} e^{\lambda - \frac{1}{2}\lambda^2}d\lambda = e^{1/2}\int_{L(n)-1}^{\infty} e^{-y^2/2}dy \tag{1.62}$$

by substituting $y = \lambda - 1$.[7] Here we substitute $y = L(n) - 1 + z$ and bound $\frac{1}{2}y^2 \geq \frac{1}{2}(L(n)-1)^2 + z(L(n)-1)$ so that

$$\int_{L(n)-1}^{\infty} e^{-y^2/2}dy \leq e^{-(L(n)-1)^2/2}\int_0^{\infty} e^{-z(L(n)-1)}dz = e^{-(L(n)-1)^2/2)(L(n)-1)^{-1}}. \tag{1.63}$$

This is exponentially small in n and, so, certainly $o(n^{-1})$. (Note however that it was important to let $L(n)$ increase fast enough. Had we tried, say, $L(n) = \ln\ln n$, the bound would be $o(1)$ and not the desired $o(n^{-1})$.)

Returning to (1.61) we now have

$$\int_{-L(n)}^{L(n)} g_n(\lambda)d\lambda = o(n^{-1}) + \int_{-\infty}^{\infty} e^{-\lambda^2/2}[1 + \frac{1}{3}\lambda^3 n^{-1/2} + \frac{1}{18}\lambda^6 n^{-1} - \frac{1}{4}\lambda^4 n^{-1}]d\lambda. \tag{1.64}$$

Fortunately $\int_{-\infty}^{+\infty} \lambda^i e^{-\lambda^2/2}d\lambda$ can be found precisely for each non-negative integer i by elementary[8] calculus. For odd i the integral is zero and for $i = 0, 4, 6$ the integrals are $\sqrt{2\pi}$, $3\sqrt{2\pi}$, and $15\sqrt{2\pi}$, respectively, so that

$$\int_{-L(n)}^{L(n)} p_n(\lambda)d\lambda = o(n^{-1}) + \sqrt{2\pi}\left(1 + \frac{15}{18n} - \frac{3}{4n}\right), \tag{1.65}$$

which is the promised $1 + \frac{1}{12n}$ term.

Remark. We need not stop here. One can take the Taylor series for $\ln(1+\epsilon)$ out further and redefine *MID* to be narrower. With a considerable amount of effort, one can show

$$n! = n^n e^{-n}\sqrt{2\pi n}[1 + \frac{1}{12n} + \frac{1}{288n^2} + o(n^{-2})], \tag{1.66}$$

and, indeed, one gets an infinite sequence of such approximations.

[7] The tail of the normal distribution is more carefully studied in §3.1.
[8] Elementary does not mean easy! Use integration by parts.

1.5. An Application to Random Walks

Here we will apply Stirling's formula to yield a classical result in the study of random walks.

Let G be an arbitrary graph for which each vertex has at least one, but only a finite number of neighbors. Let s (source) be some specified vertex of G. A simple random walk on G begins at s. Each time unit it moves uniformly from its current position v to one of the neighbors of v.

The study of random walks was begun by George Pólya around 1920. There is an essential dichotomy. A random walk is called *recurrent* if with probability 1 it returns to its beginning, here s. Otherwise, the random walk is called *transient*. In this case, while the walk *might* return to s, there is a positive probability that it will *never* return to s. Let $p(t)$ denote the probability (dependent on G and s) that the random walk will be at s at time t. Pólya showed that the dichotomy depended on the decay of $p(t)$. He showed:

Theorem 1.2. *If $\sum_{t=1}^{\infty} p(t)$ is finite, then the random walk is transient, and if $\sum_{t=1}^{\infty} p(t)$ is infinite, then the random walk is recurrent.*

Proof. Suppose there is a probability α that the random walk ever returns to s. Once it returns, it is again beginning a random walk. Hence the probability that it returns at least j times would be α^j. The expected number of times it returns would then be $\sum_{j=1}^{\infty} \alpha^j$. This expected number is also $\sum_{t=1}^{\infty} p(t)$. If $\alpha < 1$, then the sum is finite. If $\alpha = 1$ the sum is infinite. □

Now let us restrict ourselves to the grid $\mathbb{Z}^d$. The vertices are the vectors $\vec{v} = (a_1, \ldots, a_d) \in \mathbb{Z}^d$, and the neighbors of $\vec{v}$ are those $\vec{w}$ which agree with $\vec{v}$ in all but one coordinate and are one away from $\vec{v}$ in that coordinate. (This is the usual grid for $\mathbb{Z}^d$.) By symmetry, the start matters little so we consider walks beginning at the origin $\vec{0}$. In $\mathbb{Z}^2$, for example, from each (a, b) we move randomly either North $(a, b+1)$, East $(a+1, b)$, South $(a, b-1)$, or West $(a-1, b)$. We

continue forever, giving a sequence $\vec{0} = \vec{w_0}, \vec{w_1}, \ldots$, where $\vec{w_t}$ denote the position at time t.

Is the random walk in $\mathbb{Z}^d$ recurrent or transient? George Pólya gave the surprising solution:

Theorem 1.3. *The random walk in Z^d is recurrent if $d = 1$ or $d = 2$ and is transient if $d \geq 3$.*

From parity considerations one can only return to $\vec{0}$ after an even number of steps. Thus the nature of the random walk depends on whether $\sum_{t=1}^{\infty} p(2t)$ is finite. In Asymptopia we shall find the asymptotics of $p(2t)$ (note that $p(2t)$ depends on the dimension d).

In one dimension we want the probability that out of $2t$ steps precisely t are $+1$ (to the right). This has the formula

$$p(2t) = 2^{-2t}\binom{2t}{t}. \tag{1.67}$$

Applying Stirling's formula,

$$p(2t) \sim 2^{-2t}\frac{(2t)^{2t}e^{-2t}\sqrt{2\pi(2t)}}{[t^t e^{-t}\sqrt{2\pi t}]^2} \sim \sqrt{\frac{1}{\pi t}}. \tag{1.68}$$

As $\sum t^{-1/2}$ diverges, the random walk in $\mathbb{Z}^2$ is recurrent.

For dimension two there is a clever way to compute $p(2t)$. Change the coordinate system[9] to basis $\vec{v_1} = (\frac{1}{2}, \frac{1}{2})$, $\vec{v_2} = (\frac{1}{2}, -\frac{1}{2})$. Now North, South, East, and West have coordinates $(1, -1)$, $(-1, +1)$, $(1, 1)$, and $(-1, -1)$, respectively. One returns to the origin at time $2t$ if and only if each coordinate is zero. The new coordinates are now *independent*, and so we find the closed formula

$$p(2t) = [2^{-2t}\binom{2t}{t}]^2. \tag{1.69}$$

From (1.68) we now find

$$p(2t) \sim \frac{1}{\pi t}. \tag{1.70}$$

As this series diverges, the random walk in $\mathbb{Z}^2$ is recurrent.

[9] Effectively, tilt your head at a 45 degree angle!

Remark. While $\sum \frac{1}{\pi t}$ diverges, it *barely* diverges in the sense that the sum up to t grows only logarithmically. This makes the random walk in $\mathbb{Z}^2$ a strange beast. For example, it is possible to prove that the expected time until the first return to the origin is infinite.

In the remainder we assume that the dimension $d \geq 3$. (The methods apply also to the cases $d = 1, 2$, but there the exact formulae make things easier.) These cases are all quite similar, and the reader may concentrate on $d = 3$. For dimension $d \geq 3$, there does not exist a closed form[10] for $p(2t)$. In Asymptopia, however, that is hardly a stumbling block. In our asymptotics below, $d \geq 3$ is arbitrary but fixed, and $t \to \infty$.

Each step in the random walk is in one of $2d$ directions, each equally likely. However, we split the choice of directions into two parts. First, we decide for each step in which dimension it is moving. Let X_i denote the number of steps in dimension i. Then X_i has binomial distribution $BIN[2t, \frac{1}{d}]$. Note, however, that as $\sum_{i=1}^{d} X_i = 2t$, the X_i are not independent!

Theorem 1.4. *The probability that all X_i, $1 \leq i \leq d$, are even is $2^{-d+1} + o(1)$.*

On an intuitive level, each X_i has probability roughly $1/2$ of being even. However, once $X_1, \ldots, X_{d-1}$ are even, X_d is automatically even. We use a result on binomial distributions which is of independent interest.

Theorem 1.5. *Let $\epsilon > 0$ be arbitrary but fixed. Let $p = p(n)$ with $\epsilon \leq p \leq 1 - \epsilon$ for all n. Let $X = X_n$ have binomial distribution $BIN[n, p(n)]$. Then, the probability that X_n is even approaches $1/2$.*

Proof. The binomial formula gives

(1.71)
$$(px + (1-p)y)^n = \sum_{i=1}^{n} \binom{n}{i} (px)^i ((1-p)y)^{n-i} = \sum_{i=1}^{n} \Pr[X = i] x^i y^{n-i}.$$

[10] The often used phrase "closed form" does not have a precise definition. We do not consider an expression involving a summation to be in closed form.

Set $x = -1$, $y = 1$. Then

$$(1.72)\quad (p-(1-p))^n = \sum_{i=1}^{n} \Pr[X = i](-1)^i = \Pr[X \text{ even}] - \Pr[X \text{ odd}].$$

As $1 = \Pr[X \text{ even}] + \Pr[X \text{ odd}]$,

$$(1.73)\qquad \Pr[X \text{ even}] = \frac{1}{2} + \frac{1}{2}(1-2p)^n.$$

With p not close to either 0 or 1, $(1-2p)^n \to 0$ and $\Pr[X \text{ even}] = \frac{1}{2} + o(1)$. □

We could also prove Theorem 1.5 in Asymptopia. Here is the rough[11] idea. The distribution $BIN[n,p]$ takes, with probability $1 - o(1)$, values $i \sim pn$. For such i one computes

(1.74)

$$\frac{\Pr[BIN[n,p] = i]}{\Pr[BIN[n,p] = i+1]} = \frac{\binom{n}{i} p^i (1-p)^{n-i}}{\binom{n}{i+1} p^{i+1}(1-p)^{n-i-1}} = \frac{(i+1)(1-p)}{(n-i)p}.$$

For those $i \sim pn$ this ratio is nearly 1. Supposing i even for convenience, the contribution to $\Pr[BIN[n,p] = i]$ to $BIN[n,p]$ being even is asymptotically the same as the contribution $\Pr[BIN[n,p] = i+1]$ to $BIN[n,p]$ being odd.

Note, however, that the proof of Theorem 1.5 gives a much stronger result. The probability that $BIN[n,p]$ is even minus the probability that $BIN[n,p]$ is odd is exponentially small—something Asymptopia does not yield. Furthermore, though we do not use it here, the proof of Theorem 1.5 shows that the result holds even under the much weaker hypothesis that $p \gg n^{-1}$ and $1 - p \gg n^{-1}$. While we aim for proofs from Asymptopia in this work, we should remain mindful that other techniques can sometimes be even more powerful!

Proof of Theorem 1.4. Formally we show for $1 \le i \le d-1$ that the probability that $X_1, \ldots, X_i$ is all even is $2^{-i} + o(1)$. The case $i = 1$ is precisely Theorem 1.5 with $p = \frac{1}{d}$. By induction suppose the result for i. Set $m = 2t - (X_1 + \cdots + X_i)$. With probability $1 - o(1)$ all $X_1, \ldots, X_i \sim \frac{2t}{d}$ so that $m \sim 2t(1 - \frac{i}{d})$. Thus, with probability $2^{-i} + o(1)$ all $X_1, \ldots, X_i$ are even *and* $m \sim 2t\frac{d-i}{d}$. Conditional on

[11] Try to write it in detail yourself!

these values X_{i+1} has distribution $BIN[m, \frac{1}{d-i}]$. From Theorem 1.5, the conditional probability that X_{i+1} is even is $\frac{1}{2} + o(1)$.

In particular, $X_1, \ldots, X_{d-1}$ are all even with probability $2^{1-d} + o(1)$. As $X_1 + \cdots + X_d = 2t$ is even, X_d is then even tautologically. □

Theorem 1.6. *Let $p_d(2t)$ denote the probability that a random walk on $\mathbb{Z}^d$ beginning at the origin is at the origin at time $2t$. Then, for $d \geq 3$ fixed and $t \to \infty$,*

$$p_d(2t) \sim 2^{1-d} \left(\sqrt{\frac{d}{t\pi}} \right)^d . \tag{1.75}$$

Proof. As above, let X_i denote the number of steps in direction i. With probability 2^{1-d} all X_i are even. From large deviation bounds given later, in particular Theorem 8.3, there is a K so that with probability $1 - o(t^{-d/2})$ all X_i are within $K\sqrt{t \ln t}$ of $\frac{t}{d}$. The $o(t^{-d/2})$ term will not affect (1.75). (We actually only need that the X_i are $\frac{t}{d} + o(t)$.) Condition on the X_i all being even and having values $2s_i$ with $2s_i \sim \frac{2t}{d}$. Now in each dimension the probability that the $+1$ and -1 steps balance is the probability that a random walk in $\mathbb{Z}$ of time $2s_i$ ends at the origin which is from (1.68) $\sim (1/s_i\pi)^{1/2} \sim (d/t\pi)^{1/2}$. Conditioning on the $X_1, \ldots, X_d$ the events that each dimension i balances between $+1$ and -1 are mutually independent, and so the probability that they all balance—precisely what we need to return to the origin—is $\sim ((2d/t\pi)^{1/2})^d$ as claimed. □

We can now easily complete the proof of Polya's Theorem 1.3. When $d \geq 3$, $p_d(2t) = O(t^{-d/2})$. In these cases $d/2 > 1$ and so $\sum_{t=1}^{\infty} p_d(2t)$ is finite, and Theorem 1.2 gives that the random walk is transient.

Chapter 2

Big Oh, Little Oh, and All That

> Although this may seem a paradox, all exact science is dominated by the idea of approximation.
> – Bertrand Russell

Functions can be complicated. We may encounter creatures such as

$$f(n) = 7n^{5/2} + 18n^2 \ln^3 n + 84. \tag{2.1}$$

In Asymptopia this is not a problem. We are interested in the behavior of $f(n)$ as n approaches infinity. As such, the leading term (here $7n^{5/2}$) dominates. Thus we will write (formal definitions below)

$$f(n) \sim 7n^{5/2} \tag{2.2}$$

and ignore the lower order terms.

In Asymptopia we want our functions to be simple. Here is what we strive for:

Definition 2.1. A function $g(n)$ is said to be in *standard form* if it is the product of terms of the following types:

(1) Constants such as $\sqrt{2\pi}, 6, e^{-2}$.
(2) Constant powers of n such as $n, \sqrt{n}, n^{5/2}, n^{-3}$.
(3) Constant powers of $\ln n$ such as $\ln n, \sqrt{\ln n}, \frac{1}{\ln n}$.

(4) Exponentials such as $2^n, e^{-n}, 2^{n/2}$.

(5) n^{cn} for constant c, such as n^n.

One need not have all of the types below. One important example that does is

$$n^n e^{-n}\sqrt{2\pi n}. \tag{2.3}$$

This is the asymptotic formula for $n!$, called Stirling's formula, that we studied in depth in Chapter 1.

2.1. The Language of Asymptotics

In our applications we imagine $f(n)$ as a complicated function and $g(n)$ in standard form.

Definition 2.2. We write $f(n) \sim g(n)$ and say $f(n)$ is asymptotic to $g(n)$ when

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1. \tag{2.4}$$

Definition 2.3. We write $f(n) = O(g(n))$ and say $f(n)$ is big oh of $g(n)$ when there is a positive constant C such that for all sufficiently large n

$$f(n) \le Cg(n). \tag{2.5}$$

Equivalently,

$$\limsup_{n\to\infty} \frac{f(n)}{g(n)} < \infty. \tag{2.6}$$

Definition 2.4. We write $f(n) = \Omega(g(n))$ and say $f(n)$ is omega of $g(n)$ when there is a positive constant ϵ such that for all sufficiently large n,

$$f(n) \ge \epsilon g(n). \tag{2.7}$$

Equivalently,

$$\liminf_{n\to\infty} \frac{f(n)}{g(n)} > 0. \tag{2.8}$$

Definition 2.5. We write $f(n) = \Theta(g(n))$ and say $f(n)$ is theta of $g(n)$ when there exist positive constants C, ϵ so that for n sufficiently large

$$\epsilon g(n) \leq f(n) \leq Cg(n). \tag{2.9}$$

Equivalently,

$$f(n) = O(g(n)) \text{ and } f(n) = \Omega(g(n)). \tag{2.10}$$

Definition 2.6. We write $f(n) = o(g(n))$ and say $f(n)$ is little oh of $g(n)$ if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0. \tag{2.11}$$

We shall also sometimes write

$$f(n) \ll g(n) \tag{2.12}$$

and say that $f(n)$ is *negligible* compared to $g(n)$.

Definition 2.7. We write $f(n) = \omega(g(n))$ and say $f(n)$ is little omega of $g(n)$ if

$$\lim \frac{f(n)}{g(n)} = \infty. \tag{2.13}$$

We shall also sometimes write

$$f(n) \gg g(n) \tag{2.14}$$

and say that $f(n)$ grows faster than $g(n)$.

Powers of $\ln n$ come up frequently in Asymptopia, so much so that we give them a special name:

Definition 2.8. A function $f(n)$ is said to be *polylog* if $f(n) = \Theta(\ln^c n)$ for some positive constant c.

2.2. . . . and How to Use It

Certainly $\sim$ (Definition 2.2) is the best of all possible worlds. Quite frequently, however, we are studying a function $f(n)$ and we do not know a standard form $g(n)$ with $f(n) \sim g(n)$. In those cases we try to get upper bounds and lower bounds on $f(n)$ in standard form. These are represented by O (Definition 2.3) and Ω (Definition 2.4),

respectively. When we find (we are not always successful!) a $g(n)$ with upper bound $f(n) = O(g(n))$ and lower bound $f(n) = \Omega(g(n))$, we deduce $f(n) = \Theta(g(n))$. We can then say that we have found $f(n)$ *up to a constant factor*. Sometimes we can sandwich a function between an upper bound $UB(n)$ and a lower bound $LB(n)$:

Proposition 2.9. *Suppose there are functions $UB(n)$ and $LB(n)$ with $LB(n) \leq f(n) \leq UB(n)$. Suppose further than*

$$LB(n) = g(n)(1 + o(1))$$

and $UB(n) = g(n)(1 + o(1))$. Then $f(n) = g(n)(1 + o(1))$.

Except for $\sim$ our asymptotic language is oblivious to constants. That is, $f(n) = O(g(n))$ if and only if $f(n) = O(10g(n))$ if and only if $f(n) = O(\frac{1}{10}g(n))$. The same holds for $\Omega, \Theta, o, \omega$. As such, there is no point in placing constants in $g(n)$. We avoid writing $f(n) = O(10n^{3/2})$ and instead write the simpler $f(n) = O(n^{3/2})$. This notion that "constants do not matter" may be mysterious at first but it often makes life simpler, as the following results show.

Theorem 2.10. *Let $f(n) = f_1(n) + f_2(n)$. Suppose $f_1(n) = O(g(n))$ and $f_2(n) = O(g(n))$. Then $f(n) = O(g(n))$.*

Proof. There are constants C_1, C_2 so that $f_1(n) \leq C_1 g(n)$ for n sufficiently large and $f_2(n) \leq C_2 g(n)$ for n sufficiently large. Thus $f(n) \leq Cg(n)$ for n sufficiently large where we set $C = C_1 + C_2$.

Proposition 2.11. *Theorem 2.10 holds with O replaced by o, Ω, Θ, or ω. It holds in all these cases when $f(n) = f_1(n) + \cdots + f_r(n)$ for any constant r.*

The proofs are similar. Effectively, Theorem 2.10 and Proposition 2.11 allow us to replace the sum of two (or any bounded number) functions by their maximum. For the lower bounds ω, Ω we have stronger statements. If $f_1(n) = \Omega(g(n))$ and $f(n) \geq f_1(n)$, then $f(n) = \Omega(g(n))$. If $f_1(n) = \omega(g(n))$ and $f(n) \geq f_1(n)$, then $f(n) = \omega(g(n))$.

It is worth empasizing that O, o, Ω, ω only give bounds in one direction. Consider, for example, the statement

$$7n^{5/2} + 18n^2 \ln^3 n + 84 = O(n^4). \tag{2.15}$$

This is correct! It is also correct to say that you weigh less than one thousand pounds, just not very edifying. Still, sometimes we are struggling with a function $f(n)$, and we manage to prove $f(n) = O(n^4)$. It may turn out later that $f(n) = \Theta(n^{5/2})$ but, for the moment, we have a result.

2.3. Little Oh One

When $f(n) = o(1)$, it means that $f(n) \to 0$ as $n \to \infty$. Of particular use in Asymptopia is the factor $1 + o(1)$. This is a term that approaches one. (Warning: The $o(1)$ term may be negative here. If $h(n) = 1 + o(1)$, then with arbitrarily small positive ϵ we must have $1 - \epsilon \le h(n) \le 1 + \epsilon$ for n sufficiently large.) Thus

$$f(n) = g(n)(1 + o(1)) \quad \text{if and only if } f(n) \sim g(n). \tag{2.16}$$

Often times we express $f(n)$ as a main term $f_1(n)$ and a minor term $f_2(n)$. The following result allows us to ignore the minor term in Asymptopia.

Proposition 2.12. *Let $f(n) = f_1(n) + f_2(n)$. Suppose $f_1(n) \sim g(n)$ and $f_2(n) = o(g(n))$. Then $f(n) \sim g(n)$.*

2.4. Little Fleas and Littler Fleas: The Strange Hierarchy of Asymptopia

Comparing functions in standard form is not difficult, but the rules may appear strange at first. There is a natural ordering of the basic types of functions:

(1) constants,

(2) constant positive powers of $\ln n$,

(3) constant positive powers of n,

(4) exponentials c^n, $c > 1$,

(5) n^{cn} for constant positive c, such as n^n.

Each type below grows slower than the following ones. Even stronger, any positive power, however large, of one type of function grows slower than any positive power, however small, of a function of one of the following types. In particular

Theorem 2.13. *For all positive K (however large) and ϵ (however small),*

$$\ln^K n \ll n^\epsilon, \tag{2.17}$$

$$n^K \ll (1+\epsilon)^n, \tag{2.18}$$

$$K^n \ll n^{\epsilon n}. \tag{2.19}$$

The most useful of these is (2.17). The logarithm function grows to infinity extremely slowly. As an example consider a race between $\ln^5 n$ and $\sqrt{n}$. For n moderately small $\ln^5 n$ is in the lead. But take $n = e^{1000}$. Then $\ln^5 n = 10^{15}$. This is quite large in absolute terms but trivial compared to $\sqrt{n} = e^{500}$, which is bigger than a googol (10^{100}). Similarly, the exponential function grows to infinity extremely quickly. As an example consider a race between n^5 and $(1.1)^n$. When n is a million, n^5 is a mere nonillion, while $(1.1)^n$ has more than 41000 *digits*.

Proof. First set $f(n) = n^K$, $g(n) = (1+\epsilon)^n$, and consider (2.18). We compare growth rates:

$$\lim_{n\to\infty} f(n+1)/f(n) = \lim_{n\to\infty} (1+n^{-1})^K = 1^K = 1. \tag{2.20}$$

Fix any c (e.g., $c = 1+\frac{\epsilon}{2}$) with $1 < c < 1+\epsilon$. There exists n_0 so that for $n > n_0$, $f(n+1)/f(n) < c$. Hence

$$\frac{f(n_0+m)}{g(n_0+m)} \le \frac{c^m f(n_0)}{(1+\epsilon)^m g(n_0)} \to 0 \tag{2.21}$$

as $m \to \infty$. For (2.17) we parametrize $n = e^m$ so that it becomes $m^K \ll (e^\epsilon)^m$, which is (2.18) with e^ϵ replacing $1+\epsilon$. For (2.19) fix any $c > n$ and n_0 with $n_0^\epsilon \ge c$. For $n \ge n_0$, $n^{\epsilon n} \ge c^n \gg K^n$. □

Now suppose both $g_1(n)$ and $g_2(n)$ are in standard form. Which is bigger in Asymptopia? First compare the terms n^{cn}. The one with the larger c is bigger. If they have equal c (or, more often, if these terms do not appear), go on to the exponentials c^n. Again, the one with the larger c is bigger. If it is a tie, go to the powers n^c. Again the one with larger c is bigger. If it is a tie, go to the powers $\ln^c n$, and again the one with the larger c is bigger. If all of these are equal,

then $g_1(n)$ and $g_2(n)$ are equal up to a constant factor. In practice, this is quite easy. Some examples:

$$n \ln^2 n \ll n^{3/2}, \tag{2.22}$$

$$n^2 2^n \ll 3^n, \tag{2.23}$$

$$n^{n/2} \ll n^n e^{-5n}. \tag{2.24}$$

The last example (2.24) shows that these rules apply even when the values c in n^{cn} or n^c or $\ln^c n$ are negative. Other examples:

$$n^{0.9} \ll n \ln^{-2} n, \tag{2.25}$$

$$\frac{2^n}{\sqrt{n}} \gg 1.99^n. \tag{2.26}$$

These also apply to c^n when $0 < c < 1$. A final example:

$$0.9^n \ln^3 n \ll 1. \tag{2.27}$$

2.5. Little Oh One in the Exponent

How can we express that $f(n)$ is somewhat near n^2 (or any other constant power) when we do not know that it is very near n^2? A useful terminology is to place a $1 + o(1)$ factor in the exponent and to write $f(n) = n^{2(1+o(1))}$. This means

(1) for any fixed positive ϵ, if n is sufficiently large, then $f(n) > n^{2(1-\epsilon)}$;

(2) for any fixed positive ϵ, if n is sufficiently large, then $f(n) < n^{2(1+\epsilon)}$.

How accurately do we know $f(n)$? Not very! It could be that $f(n) \sim n^2$. But, using our characterization of fleas and little fleas, it could be that $f(n) = n^2 \ln n$ or $n^2 \ln^5 n$ or $n^2 \ln^{-3} n$. Why would we use such a rough formulation? Sometimes we simply do not have a more accurate view of $f(n)$, and so this describes how well we understand it. Other times we do not need a more accurate view of $f(n)$ for later calculations. It is important not to confuse the coarser $1+o(1)$ in the exponent with the more accurate $1 + o(1)$ factor "downstairs" as in (2.16).

In the above example we could also have written $f(n) = n^{2+o(1)}$. The expression $2 + o(1)$ represents a function approaching two. The expression $2(1 + o(1))$ represents twice a function approaching one. These are equivalent! Which form one uses is largely a matter of taste.

As a further example, consider $g(n) = 2^{n(1+o(1))}$. This means

(1) for any fixed positive ϵ, if n is sufficiently large, then $g(n) > 2^{n(1-\epsilon)}$;

(2) for any fixed positive ϵ, if n is sufficiently large, then $g(n) < 2^{n(1+\epsilon)}$.

A flea smaller than exponential will not affect this. If, for example, $g(n) = 2^{n(1+o(1))}$, then after multiplication or division of $g(n)$ by n^{10} this still holds. So, for example,

$$\binom{n}{n/2} = 2^{n(1+o(1))} \tag{2.28}$$

is a correct statement, though we can and usually should do better.

2.6. Inverting Functions

Consider a relation such as $y = x \ln x$ for $x \geq 1$. This is an increasing function of x and so there is a unique inverse function $x = f(y)$ for $y \geq 0$. There is no compact way to write $f(y)$ precisely. In Asymptopia (as $x, y \to \infty$) the problem disappears. For x large $\ln y = \ln x + \ln \ln x \sim \ln x$. Thus

$$x = \frac{y}{\ln x} \sim \frac{y}{\ln y}. \tag{2.29}$$

The general result is similar.

Theorem 2.14. *Let a, c be positive reals, and let b be any real number, possibly zero, positive or negative. Set $f(x) = cx^a \ln^b x$ defined for $x > 1$. For y sufficiently large (dependent on a, b) there is a unique x with $y = f(x)$. Write $x = g(y)$ for such y. Asymptotically in y*

$$x \sim dy^{1/a}(\ln y)^{-b/a}, \tag{2.30}$$

where $dc^{1/a}a^{-a/b} = 1$.

As the x^a term dominates the polylogarithmic term, $y = f(x)$ is eventually increasing, and so the inverse function exists. To show (2.30), we calculate $f(x)$, noting that $\ln y \sim a \ln x$.

Often times we are not concerned with precise constants, and the following simpler form is useful:

Theorem 2.15. *If $y = \Theta(x^a \ln^b x)$, then $x = \Theta(y^{1/a} \ln^{-b/a} y)$.*

Example. In Chapter 7, $R(3,k)$ is defined as the least n with a certain property. It is sometimes convenient to reverse variables and let $f(n)$ be the biggest k with $R(3,k) \leq n$. The functions $R(3,k)$ and $f(n)$ are then inverses. To show (as indeed is the case) that $R(3,k) = \Theta(k^2 \ln^{-1} k)$, one can equivalently show $f(n) = \Theta(\sqrt{n}\sqrt{\ln n})$.

2.7. Taylor Series

When $f(x)$ is an appropriately nice function, $f(x)$ is nicely approximated around $x = x_0$ using Taylor series. Generally speaking, the approximation becomes better as you use more terms. Let $TAY_k(x_0, \epsilon)$ denote the Taylor series approximation to $f(x_0 + \epsilon)$, using the Taylor series around x_0 out to the k-th derivative. Specifically,

$$TAY_k(x_0, \epsilon) = f(x_0) + f'(x_0)\epsilon + \frac{f^{(2)}(x_0)}{2}\epsilon^2 + \cdots + \frac{f^{(k)}(x_0)}{k!}\epsilon^k, \tag{2.31}$$

where $f^{(s)}$ denotes the s-th derivative of f. How close is the approximation $TAY_n(x_0, \epsilon)$ to the actual value $f(x_0 + \epsilon)$? The general result is

Theorem 2.16. *Let $f(x)$ be a function whose first $k+1$ derivatives are continuous at $x = x_0$. Then*

$$f(x_0 + \epsilon) = TAY_k(x_0, \epsilon) + O(\epsilon^{k+1}), \tag{2.32}$$

where O is understood as $\epsilon \to 0$.

The most used case is $k = 1$. Then

$$f(x_0 + \epsilon) = f(x_0) + f'(x_0)\epsilon + O(\epsilon^2). \tag{2.33}$$

The special cases e^x, $\ln(1+x)$, $\ln(1-x)$ (all around $x_0 = 0$) come up frequently. For convenience we give some of these special cases:

$$e^\epsilon = \epsilon + O(\epsilon^2) \text{ and } e^\epsilon = \epsilon + \frac{\epsilon^2}{2} + O(\epsilon^3), \tag{2.34}$$

$$\ln(1+\epsilon) = \epsilon + O(\epsilon^2) \text{ and } \ln(1+\epsilon) = \epsilon - \frac{\epsilon^2}{2} + O(\epsilon^3). \tag{2.35}$$

The function $\ln(1-x)$ is particularly nice as all of the derivatives are negative. We get an inequality

$$\ln(1-\epsilon) < -\epsilon - \frac{\epsilon^2}{2} - \cdots - \frac{\epsilon^k}{k!}, \tag{2.36}$$

which is valid for all $0 < \epsilon < 1$ and all $k \geq 1$. Note, however, that the function $\ln(1+\epsilon)$ has derivatives of alternate sign. For that reason, inequalities for $\ln(1+\epsilon)$ that hold for all positive ϵ are messier.

A second approach gives the error explicitly. We omit the proof, which may be found in calculus texts.

Theorem 2.17. *Let $\epsilon > 0$ be arbitrary. Let $f(x)$ be a function whose first $k+1$ derivatives are continuous in the region $[x_0, x_0+\epsilon]$. Then*

$$f(x_0+\epsilon) = TAY_k(x_0, \epsilon) + \frac{f^{(k+1)}(z)}{(k+1)!}\epsilon^{k+1} \tag{2.37}$$

for some $x_0 \leq z \leq x_0+\epsilon$. Further, when $\epsilon < 0$ equation (2.37) still holds but now the first $k+1$ derivatives are continuous in $[x_0+\epsilon, x_0]$ and $x_0+\epsilon \leq z \leq x_0$.

In applications we often have a uniform bound on the $f^{(k+1)}(x)$.

Theorem 2.18. *Let $\epsilon > 0$ be arbitrary. Let $f(x)$ be a function whose first $k+1$ derivatives are continuous in the region $[x_0, x_0+\epsilon]$. Suppose $|f^{(k+1)}(z)| \leq A$ for all $z \in [x_0, x_0+\epsilon]$. Then*

$$|f(x_0+\epsilon) - TAY_k(x_0, \epsilon)| \leq A\frac{\epsilon^{k+1}}{(k+1)!}. \tag{2.38}$$

Further, when $\epsilon < 0$, equation (2.38) still holds, but now we assume that the first $k+1$ derivatives of $f(x)$ are continuous in $[x_0+\epsilon, x_0]$ and that $|f^{(k+1)}(z)| \leq A$ for all $z \in [x_0+\epsilon, x_0]$.

Chapter 3

Integration in Asymptopia

> The search for truth is more precious than its possession.
> – Albert Einstein

Finding a definite integral in Asymptopia is quite different from first year calculus. We look at where the function is largest and examine the area near that maximum.

When properly scaled, the graph of the function near its maximum will often look like the bell shaped curve.

When the maximum is on the boundary of the region, often times the graph will look like the exponential function e^{-w}. There is usually technical work to ensure that the bulk of the area lies near the maximum, in the region we call *MID*. But the "picture" of the function in Asymptopia is quite simple and informative.

3.1. The Gaussian Tail

Here we examine the asymptotics of the tail of the Gaussian distribution

$$GT(a) = \int_a^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \tag{3.1}$$

as $a \to \infty$. Set

$$f(x) = (2\pi)^{-1/2} e^{-x^2/2}$$

so that we seek $\int_a^\infty f(x)dx$. As $f(x)$ is decreasing its maximum over $[a, \infty)$ is at $x = a$. Set

$$z = \ln(f(x)) = -x^2/2 - \frac{1}{2}\ln(2\pi).$$

Then $z'(a) = -a$ so that the Taylor series for z around $x = a$ begins as

$$z(a + \epsilon) = z(a) - a\epsilon + \cdots . \tag{3.2}$$

This leads to the natural scaling $\epsilon = \frac{1}{a}\lambda$.

With this as background (somewhat similarly to §1.1), we try the substitution

$$x = a + \frac{\lambda}{a} \tag{3.3}$$

so that

$$GT(a) = \frac{1}{\sqrt{2\pi}} \frac{1}{a} \int_{\lambda=0}^\infty e^{-(a+\frac{\lambda}{a})^2/2} d\lambda. \tag{3.4}$$

We take out the factor of $e^{-a^2/2}$ to give

$$GT(a) = \frac{1}{\sqrt{2\pi}} \frac{1}{a} e^{-a^2/2} \int_{\lambda=0}^\infty e^{-\lambda} e^{-\frac{\lambda^2}{2a^2}} d\lambda. \tag{3.5}$$

We can now make a good guess (the *art* of Asymptopia) for the answer. By the "time" the $e^{-\lambda^2/2a^2}$ term becomes substantially smaller than 1, the $e^{-\lambda}$ term is so small that it does not matter. If we ignore the second term, we get $\int_0^\infty e^{-\lambda} d\lambda = 1$. As the second term always lies in $[0, 1]$, we quickly get the upper bound

$$\int_0^\infty e^{-\lambda} e^{-\frac{\lambda^2}{2a^2}} d\lambda \le \int_0^\infty e^{-\lambda} d\lambda = 1. \tag{3.6}$$

For the lower bound we split the range into

$$MID = [0, L(a)]$$

and

$$RIGHT = [L(a), \infty).$$

(In this problem there is no *LEFT*.) We want to select $L = L(\lambda)$ so that $\lambda^2/2a^2$ is negligible throughout *MID*, which forces $L = o(a)$. But we also want the integral over *MID* to be almost 1 which forces $L(a) \to \infty$. For definiteness let us take $L(a) = \sqrt{a}$. Throughout *MID*, $e^{-\lambda^2/2a^2} \geq e^{-1/2a}$. Thus

$$\int_0^{\sqrt{a}} e^{-\lambda} e^{\frac{-\lambda^2}{2a^2}} d\lambda \geq e^{-1/2a} \int_0^{\sqrt{a}} e^{-\lambda} d\lambda = e^{-1/2a}[1 - e^{-\sqrt{a}}]. \tag{3.7}$$

As $a \to \infty$, $e^{-1/2a}[1 - e^{-\sqrt{a}}] \to e^0[1 - 0] = 1$. Since here we are only looking for a lower bound, we can ignore the integral over *RIGHT*.

(*Caution.* If the function were sometimes negative, we could not do this!) We have sandwiched the integral (3.6) between 1 and a function of a approaching 1, and therefore the limit is 1. Plugging back into (3.5) gives the desired asymptotics:

$$GT(a) \sim \frac{1}{a}\sqrt{\frac{1}{2\pi}} e^{-a^2/2}. \tag{3.8}$$

The weight of the Gaussian tail from a to ∞ is nearly all very close to a. We can make this more precise:

Theorem 3.1. *Let N denote the standard Gaussian distribution. Let $a < b$ approach ∞ with $b - a \gg a^{-1}$.*

$$\frac{\Pr[N \geq b]}{\Pr[N \geq a]} \to 0. \tag{3.9}$$

The proof is a simple plug-in to (3.8). The case $b = a + 1$ leads to surprising results. Suppose that IQ scores[1] are given by a Gaussian distribution with mean 100 and standard deviation 15. Suppose Frisbee University selects 1000 students randomly from those with IQ above 145, meaning three standard deviations above the mean. How many of those students will have IQ above 160, meaning four standard deviations above the mean? The values $a = 3, b = 4$ are

[1] Whether IQ scores have any validity whatsoever is a separate question!

constants so we can check tables for the Gaussian tail. $\Pr[N \geq 3] = 0.001350\cdots$, maybe one student in a good sized high school class would be three standard deviations above the mean. But $\Pr[N \geq 4] = 0.00003167\cdots$ is *far* smaller. Of these 1000 elite students only $1000 \cdot 0.00003167/0.001350$ or about 23 of them would be in the higher category, and probably none of them would be five standard deviations above the mean.

3.2. High Trigonometric Powers

We examine here the asymptotics of the integral

$$S(n) = \int_0^\pi \sin^n x dx \tag{3.10}$$

as $n \to \infty$. The behavior is remarkably similar to $\int x^n e^{-x} dx$ discussed in Chapter 1. Set $y = \sin^n x$ and $z = \ln y = n \ln \sin x$. At $x = \frac{\pi}{2}$, $\sin x$ is maximized so that y and z are also maximized. We compute $z'(x) = n \cot(x)$ so that $z''(x) = -n \csc^2(x)$. At the maximal point $z = \frac{\pi}{2}$ we compute $z'(\pi/2) = 0$ and $z''(\pi/2) = -n$. The Taylor series for $z(x)$ about $\pi/2$ begins as

$$z\left(\frac{\pi}{2} + \epsilon\right) = -n\epsilon^2 + \cdots . \tag{3.11}$$

This leads to the $x = \frac{\pi}{2} + \frac{1}{\sqrt{n}} w$. Set

$$f(w) = \sin^n \left(\frac{\pi}{2} + \frac{1}{\sqrt{n}} w\right) \tag{3.12}$$

and

$$g(w) = \ln f(w) = n \ln \sin(\frac{\pi}{2} + \frac{1}{\sqrt{n}} w) \tag{3.13}$$

so that

$$S(n) = \frac{1}{\sqrt{n}} \int f(w) dw, \tag{3.14}$$

where the integral now goes from $-\frac{2}{\pi}\sqrt{n}$ to $+\frac{2}{\pi}\sqrt{n}$.

For w fixed, $\lim_{n\to\infty} g(w) = -w^2/2$ and $\lim_{n\to\infty} f(w) = e^{-w^2/2}$. (Like $x^n e^{-x}$, $\sin^n x$, where properly scaled, looks like the bell shaped curve!)

We split $[0, \pi]$ into a middle region

$$MID = [-L(n), +L(n)],$$

in which $g(w) \sim -w^2/2$, and left and right regions

$$LEFT = [-\frac{2}{\pi}\sqrt{n}, -L(n)]$$

and

$$RIGHT = [L(n), \frac{2}{\pi}\sqrt{n}]$$

on which $\int f(w)$ will be appropriately negligible. We must select $L(n)$ large enough so that *MID* contains most of the integral and small enough so that $g(w) \sim -w^2/2$ holds uniformly throughout *MID*. A wide range of $L(n)$ will work; let us use $L(n) = n^{\cdot 1}$.

3.2.1. MID. We expand $g(w)$ in a Taylor series about 0. We have $g(0) = n \ln \sin(\frac{\pi}{2}) = 0$, $g'(0) = nn^{-1/2} \cot(\frac{\pi}{2}) = 0$, and $g''(0) = -\sec^2(\frac{\pi}{2})$ so the series begins $-w^2/2$ as expected. Now we need a *uniform* upper bound on the third derivative over *MID*. We find

$$g^{(3)}(w) = 2n^{-1/2} \cot(\theta) \csc^2(\theta) \tag{3.15}$$

with $\theta = \frac{\pi}{2} + \frac{1}{\sqrt{n}}w$. With $|w| \le L(n)$, θ is close to $\frac{\pi}{2}$ so that $g^{(3)}(w)$ is close to $2n^{-1/2}$. We use the rough upper bound $|g^{(3)}(w)| \le 3n^{-1/2}$ for all $|w| \le L(n)$. When $w \in MID$ and z lies between 0 and w, then $|g^{(3)}(z)| \le 3n^{-1/2}$ as well. Theorem 2.18 gives the Taylor series for $g(w)$ with error bound,

$$|g(w) + \frac{w^2}{2}| \le 3n^{-1/2}\frac{w^3}{6} \le n^{-0.2} \tag{3.16}$$

for all $w \in MID$. Exponentiating, and using the rough bound $|e^\delta - 1| \le 2\delta$, valid for δ small,

$$e^{-w^2/2}(1 - 6n^{-0.2}) \le f(w) \le e^{-w^2/2}(1 + 6n^{-0.2}). \tag{3.17}$$

We have sandwiched $f(w)$ so that

$$\int_{MID} f(w)dw = O(n^{-0.2}) + \int_{MID} e^{-w^2/2}dw. \tag{3.18}$$

As $L(n) \to \infty$,

$$\int_{MID} e^{-w^2/2}dw = \sqrt{2\pi} + o(1). \tag{3.19}$$

Thus,

$$\int_{MID} f(w)dw = \sqrt{2\pi} + o(1). \tag{3.20}$$

3.2.2. RIGHT, LEFT. To show that *RIGHT* contributes negligibly (by symmetry, *LEFT* is the same) we can use that $f(w)$ and $g(w)$ are decreasing functions. At $w = L(n)$, $g(w) = -w^2/2 + O(n^{-0.2}) \sim -n^{.2}/2$ so that $f(w)$ is exponentially small. The interval *RIGHT* has length less than $\sqrt{n}$, the exponential dominates the polynomial and so $\int_{RIGHT} f(w)dw$ is exponentially small, and hence $o(1)$. The integral of (3.14) is therefore $\sqrt{2\pi} + o(1)$ and

$$S(n) = \int_0^\pi \sin^n x dx \sim \sqrt{\frac{2\pi}{n}}. \tag{3.21}$$

3.2.3. Stirling Redux. $S(n)$ may be calculated precisely using calculus. An ingenious integration by parts yields the recursion

$$S(n) = \frac{n-1}{n} S(n-2) \tag{3.22}$$

with initial conditions $S(0) = \pi$, $S(1) = 2$. For notational simplicity we set

$$A = 2^{-2m} \frac{(2m)!}{m!m!}. \tag{3.23}$$

When n is even, setting $n = 2m$

$$S(2m) = \pi \prod_{i=1}^{m} \frac{2i-1}{2i} = A\pi. \tag{3.24}$$

When n is odd, setting $n = 2m + 1$,

$$S(2m+1) = 2 \prod_{i=1}^{m} \frac{2i}{2i+1} = \frac{2}{2m+1} A^{-1}. \tag{3.25}$$

These formulae will yield *two* new arguments for the constant $\sqrt{2\pi}$ in Stirling's formula. We assume (1.29), that $n! \sim Kn^n e^{-n}\sqrt{n}$ for some constant K. Then

$$A \sim 2^{-2m} \frac{K(2m)^{2m} e^{-2m}\sqrt{2m}}{(Km^m e^{-m}\sqrt{m})^2} \sim \frac{\sqrt{2}}{K\sqrt{m}}. \tag{3.26}$$

Applying (3.24) (we could similarly apply (3.25)),

$$S(2m) \sim \frac{\sqrt{2}\pi}{K\sqrt{m}}. \tag{3.27}$$

Comparing to the previously found asymptotics (3.21), we may solve for K, yielding $K = \sqrt{2\pi}$.

A second argument, modifying slightly the approach [**CR96**] of Richard Courant, avoids asymptotic evaluation of $S(n)$. As $\sin^n x$ is decreasing in n, $S(2m) \geq S(2m+1) \geq S(2m+2)$. The recursion (3.22) gives $\lim_m S(2m+1)/S(m) = 1$. As $S(2m+1)$ is sandwiched between these two terms, $\lim_m S(2m)/S(2m+1) = 1$. Thus $A\pi \sim m^{-1}A^{-1}$ and $A \sim (m\pi)^{-1/2}$. We may now solve (3.26) for K, again yielding $K = \sqrt{2\pi}$.

3.3. An Easy Integral

Here we examine the asymptotics of

$$F(n) = \int_0^1 x^n dx \tag{3.28}$$

as $n \to \infty$. Of course, we know the exact value, but it provides a good example. Further, Asymptopia gives new insight into what the function "looks like" on $[0, 1]$.

Set $y(x) = x^n$ and $z(x) = \ln y(x) = n \ln x$ so that y and z are maximized (over $[0, 1]$) at the endpoint 1. The Taylor series for $z(x)$ about $x = 1$ begins $z(1+\epsilon) = n\epsilon$. This leads us to the parametrization

$$x = 1 + \frac{w}{n}. \tag{3.29}$$

Set

$$f(w) = \left(1 + \frac{w}{n}\right)^n \tag{3.30}$$

and

$$g(w) = \ln f(w) = n \ln(1 + \frac{w}{n}). \tag{3.31}$$

With this change of variables,

$$F(n) = n^{-1} \int_{-n}^0 f(w)dw. \tag{3.32}$$

With w fixed, $\lim_{n\to\infty} g(w) = -w$. Now we split into a middle term *MID* of the form $[-L(n), 0]$ and a left term *LEFT*, $[-n, -L(n)]$. (There is no *RIGHT* interval here.) We must select $L(n)$ large enough that *MID* contains most of the integral and small enough that $g(w) \sim -w$ holds uniformly throughout *MID*. A wide range of $L(n)$ will work; let us use $L(n) = n^{\cdot 1}$. For w in *MID* the Taylor series with error (2.38) gives

$$g(w) = -w - \frac{1}{n}(1 + \frac{z}{n})^{-2}\frac{w^2}{2} \tag{3.33}$$

for some z, $w \le z \le 0$. As $\frac{z}{n}$ is small in *MID*, $(1 + \frac{z}{n})^{-2} \le 2$ so that

$$|g(w) + w| \le \frac{w^2}{n} \le n^{-0.8}. \tag{3.34}$$

Exponentiating and using that $|e^\delta - 1| \le 2\delta$ for δ small,

$$|f(w) - e^{-w}| \le 2n^{-0.8}. \tag{3.35}$$

With the length of *MID* only $L(n) = n^{0.1}$,

$$\int_{MID} f(w)dw = \int_{MID} e^{-w}dw + O(n^{-0.7}). \tag{3.36}$$

As $L(n) \to \infty$,

$$\int_{MID} e^{-w}dw = o(1) + \int_0^\infty e^{-w}dw = 1 + o(1). \tag{3.37}$$

To show that *LEFT* contributes negligibly, we note that $f(w), g(w)$ are increasing functions of w. At $w = -L(n) = -n^{0.1}$, inequality (3.34) gives $g(w) = -w + o(1)$ so that $f(w)$ is an exponentially small function of n. The interval *LEFT* has length less than n, so the exponentially small term dominates the polynomial factor and $\int_{LEFT} f(w)dw$ is $o(1)$. Thus the integral of (3.32) has value $1 + o(1)$ and $F(n) \sim \frac{1}{n}$, as we knew all along!

3.4. Integrals with logs

Sometimes when the integrand has logarithmic factors there is no precise integration. In Asymptopia this is not a problem. Consider an integral involving an $\ln x$ factor, where the limits of integration are from some small value to n. The notion will be that for "most" x in that region, $\ln x$ will be very close to $\ln n$. In Asymptopia the $\ln x$

factor becomes a constant $\ln n$ factor which can be taken out of the integral. A basic example of this is

$$Li(n) := \int_e^n \frac{dx}{\ln x}. \tag{3.38}$$

The lower limit $x = e$ is quite arbitrary, but one wants to avoid the pole at $x = 1$. The intuitive notion is that one can replace $\ln x$ with $\ln n$ giving

$$Li(n) \sim \frac{1}{\ln n} \int_e^n dx \sim \frac{n}{\ln n}. \tag{3.39}$$

As $\frac{1}{\ln x} \geq \frac{1}{\ln n}$, we immediately get the lower bound

$$Li(n) \geq \frac{1}{\ln n} \int_e^n dx \sim \frac{n}{\ln n}. \tag{3.40}$$

For the upper bound we split the range $[e, n]$ into *LEFT* and *MID* at some $u(n)$. We need take $u(n)$ large enough that $\ln x \sim \ln n$ throughout *MID* and yet small enough that the integral over *LEFT* will be negligible compared to the proposed answer $\frac{n}{\ln n}$. The first condition requires $u = u(n) = n^{1-o(1)}$; let us take $u(n) = n \ln^{-10} n$ for definiteness. Now we take the rough but accurate upper bound $\frac{1}{\ln x} \leq 1$ so that

$$\int_{LEFT} \frac{dx}{\ln x} \leq \int_{LEFT} dx \leq u(n), \tag{3.41}$$

which is indeed $o(\frac{n}{\ln n})$. Now throughout *MID*, $\ln(x) \geq \ln(u(n)) \geq \ln n - 10 \ln \ln n \sim \ln n$ so that

$$\int_{MID} \frac{dx}{\ln x} \sim \frac{1}{\ln n} \int_{MID} dx \sim \frac{n}{\ln n} \tag{3.42}$$

so that

$$L(n) = \int_{LEFT} \frac{dx}{\ln x} + \int_{MID} \frac{dx}{\ln x} = o(\frac{n}{\ln n}) + (1 + o(1)) \frac{n}{\ln n} \sim \frac{n}{\ln n} \tag{3.43}$$

as desired.

Comment. The function $Li(n)$ is a better estimate for $\pi(n)$, the number of primes $p \le n$, than the simpler $\frac{n}{\log n}$.

As a further example consider

$$F(n) := \int_e^n \frac{s}{\ln s} ds. \tag{3.44}$$

Replacing $\ln s$ with the constant $\ln n$ leads us to conjecture $F(n) \sim \frac{n^2}{2}\frac{1}{\ln n}$. As $\frac{s}{\ln s} \ge \frac{s}{\ln n}$, we quickly get the lower bound

$$F(n) \ge \frac{1}{\ln n}\int_e^n s \cdot ds \sim \frac{n^2}{2\ln n}. \tag{3.45}$$

For the lower bound we split the range $[e, n]$ into *LEFT* and *MID* and $u(n) = n\ln^{-10} n$ again works well. In *MID*, $\frac{1}{\ln s} = (1 - o(1))\frac{1}{\ln n}$, so

$$\int_{MID} \frac{s}{\ln s} ds = (1 - o(1)) \int_{MID} \frac{s}{\ln n} ds = (1 - o(1))\frac{n^2}{2\ln n}. \tag{3.46}$$

As the integrand is always positive, this provides a lower bound for $F(n)$. We have sandwiched $F(n)$ giving (3.45).

Further, let us consider briefly

$$G(n) := \int_1^n s(\ln s) ds. \tag{3.47}$$

We use the same split into *LEFT* and *MID* as with (3.44), $\ln s \sim \ln n$ in *MID*, and

$$G(n) \sim (\ln n)\int_{MID} s \cdot ds \sim \frac{n^2 \ln(n)}{2}. \tag{3.48}$$

For this problem we did not need Asymptopia: the precise definite integral is $G(n) = \frac{n^2}{2}\ln(n) - \frac{n^2}{4} + \frac{1}{4}$. But even here one may feel that Asymptopia gives one insight into the integral.

Chapter 4

From Integrals to Sums

> Simplicity is the highest goal, achievable when you have overcome all difficulties. After one has played a vast quantity of notes and more notes, it is simplicity that emerges as the crowning reward of art.
> – Fryderyk Chopin

The creation of the integral calculus by Newton and Leibnitz in the 17th century was surely one of the great advances in mathematics. It was, however, only with the development of the Riemann integral by Bernhard Riemann in the 19th century that calculus approached its modern form.

At its heart is the definition of the area under the curve $y = f(x)$ between $x = a$ and $x = b$. For any n the interval $[a, b]$ is broken into n equal pieces by finding $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$. Values $z_i \in [x_{i-1}, x_i]$, $1 \leq i \leq n$, are selected. One sets $A_n = \sum_{i=1}^{n}(x_i - x_{i-1})f(z_i)$. Geometrically, A_n represents the total area of n rectangles which, for n large, hug the region under the curve $y = f(x)$. The area under the curve $y = f(x)$ is then *defined* to be the limit of this process as $n \to \infty$. (One requires also that the interval lengths $x_i - x_{i-1}$ approach zero uniformly as $n \to \infty$.) Riemann demonstrated that this limit would exist (and not depend on the choices of the x_i and the z_i) for a wide class of functions. In particular, splitting the

interval into n equal pieces and always taking the right-hand side of the interval for z_i, we would have

$$A_n = \sum_{i=1}^{n} \frac{b-a}{n} f(a + i\frac{b-a}{n}). \tag{4.1}$$

Consider the area under the curve $y = x^2$ between $x = 0$ and $x = 1$. This area was first found by Archimedes, in *The Quadrature of the Parabola* , written in the third century B.C. Archimedes used an ingenious decomposition of the area into triangles. From Riemann's vantage point, using the particular choice of x_i, z_i above, we have

$$A_n = \frac{1}{n} \sum_{i=1}^{n} (i/n)^2 = \frac{1}{n^3} \sum_{i=1}^{n} i^2. \tag{4.2}$$

In this case the precise formula

$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6} \tag{4.3}$$

immediately yields $\lim_{n\to\infty} A_n = \frac{1}{6}$.

In Asymptopia we turn this story on its head. We have a sum, and we seek standard form asymptotic values. We *know* first year calculus, so we know that $\int_0^1 x^2 dx = \frac{1}{3}$. Thus, we can *deduce* that

$$\sum_{i=1}^{n} i^2 \sim \frac{n^3}{6}. \tag{4.4}$$

Frequently, we will be presented with a sum in the form $\sum_{i=a}^{b} f(i)$. We would like to say that the precise sum is close to the corresponding integral $\int_a^b f(x)dx$.

Some caution is needed. Consider $\sum_{i=0}^{n} 2^i$, which has the precise value $2^{n+1} - 1$. The corresponding integral $\int_0^n 2^x dx$ works out to $\frac{1}{\ln 2}[2^n - 1]$. The terms -1 may be considered negligible but still the two values are off by a constant factor. That said, for a wide variety of sums, the approximation via an integral *is* quite accurate.

4.1. Approximating Sums by Integrals

Theorem 4.1. *Let $a < b$ be integers. Let $f(x)$ be an integrable function in $[a-1, b+1]$. Set $S = \sum_{i=a}^{b} f(i)$.*

(1) *If f is an increasing function on $[a-1, b+1]$, then*

$$\int_{a-1}^{b} f(x)dx \le S \le \int_{a}^{b+1} f(x)dx. \tag{4.5}$$

(2) *If f is an decreasing function on $[a-1, b+1]$, then*

$$\int_{a-1}^{b} f(x)dx \ge S \ge \int_{a}^{b+1} f(x)dx. \tag{4.6}$$

Proof. For $a \le i \le b$ consider the rectangle with base $[i-1, i]$ and height $f(i)$. When f is increasing, the region given by these rectangles lies over the curve $y = f(x)$ from $x = a-1$ to $x = b$ and therefore has greater area, giving the first inequality of (4.5). The other inequalities are similar. □

The following formulation is particularly useful for our applications.

Theorem 4.2. *Let $a < b$ be integers. Let $f(x)$ be an integrable function in $[a-1, b+1]$. Set $S = \sum_{i=a}^{b} f(i)$ and $I = \int_a^b f(x)dx$. Let M be such that $|f(x)| \le M$ for all $a-1 \le x \le b+1$. Suppose f is either an increasing function or a decreasing function on $[a-1, b+1]$. Then*

$$|S - I| \le M. \tag{4.7}$$

Proof. The bounds in the four cases of (4.5) and (4.6) are off from I by an integral of $f(x)$ over a unit interval. □

In practice one often uses a rough upper bound for M. Some examples:

$$\sum_{i=1}^{n} i^{\alpha} = \frac{n^{\alpha+1}}{\alpha+1} + O(n^{\alpha}) \text{ for any } \alpha > 0, \tag{4.8}$$

$$\sum_{i=n}^{2n} \frac{1}{i} = \ln(2) + O(\frac{1}{n}). \tag{4.9}$$

Occasionally, the function $f(x)$ will not be defined at $x = a-1$. The simple solution: remove the $x = a$ term! We can use this idea to bound $\ln(n!) = \sum_{i=1}^{n} \ln(i)$. The function $\ln(x)$ is not defined at $x = 0$,

but the value $\ln(1) = 0$, so instead we write $\ln(n!) = \sum_{i=2}^{n} \ln(i)$. Now Theorem 4.1 applies and

$$\ln(n!) \geq \int_1^n \ln(x)dx = n\ln n - n + 1 \tag{4.10}$$

so that

$$n! \geq e(n/e)^n \geq (n/e)^n. \tag{4.11}$$

While this is certainly not as powerful as the approximation via trapezoids in §1.2, it is quite handy and holds for all $n \geq 1$.

The assumption that $f(x)$ is monotone in Theorem 4.2 is rarely an obstacle. The result below may be improved, but it suffices for nearly all applications. When the number r of turning points of $f(x)$ is bounded, one still finds $|S - I| = O(M)$.

Theorem 4.3. *Let $a < b$ be integers. Let $f(x)$ be an integrable function in $[a-1, b+1]$. Set $S = \sum_{i=a}^{b} f(i)$ and $I = \int_a^b f(x)dx$. Let M be such that $|f(x)| \leq M$ for all $a - 1 \leq x \leq b + 1$. Suppose $[a-1, b+1]$ can be broken up into at most r intervals such that $f(x)$ is monotone on each. Then*

$$|S - I| \leq 6rM. \tag{4.12}$$

Proof. Let $a - 1 = x_0 < x_1 < \cdots < x_r = b + 1$ be such that $f(x)$ is monotone on each $[x_{j-1}, x_j]$. We split the integers correspondingly, letting u_j, v_j denote the first and last integers in $[x_{j-1}, x_j]$. Set S_j equal the sum of the $f(i)$ from $u_j + 1$ to $v_j - 1$, and let I_j denote the integral of $f(x)$ over $[u_j + 1, v_j - 1]$. From Theorem 4.2, each $|S_j - I_j| \leq M$. Hence,

$$\left| \sum_{j=1}^{r} S_j - \sum_{j=1}^{r} I_j \right| \leq rM. \tag{4.13}$$

We have deleted $2r$ integers so that S is at most $2rM$ from $\sum_{j=1}^{r} S_j$. The integrals miss the $r - 1$ interior ranges $[v_j - 1, u_{j+1} + 1]$ and also cover regions $[a - 1, a]$ and $[b, b + 1]$ not in $[a, b]$. Thus I is within

$3rM$ of $\sum_{j=1}^{r} I_j$. Finally,

$$|S-I| \le \left|S - \sum_{j=1}^{r} S_j\right| + \left|\sum_{j=1}^{r} S_j - \sum_{j=1}^{r} I_j\right| + \left|I - \sum_{j=1}^{r} I_j\right| \le 6rM. \tag{4.14}$$

□

A typical example of Theorem 4.15 is the evaluation of

$$A_n = \sum_{i=0}^{n} i(n-i) \tag{4.15}$$

in Asymptopia. The function $f(x) = x(n-x)$ is differential and so can have a turning point only when the derivative is zero. Here this occurs at $x = \frac{n}{2}$ and frequently it occurs at only a bounded number r of points. We evaluate

$$I_n = \int_0^n x(n-x)dx = n^3 \int_0^1 y(1-y)dy = \frac{n^3}{6} \tag{4.16}$$

and note that $f(x) = O(n^2)$ in $[-1, n+1]$ so that

$$A_n = \frac{n^3}{6} + O(n^2). \tag{4.17}$$

4.2. The Harmonic Numbers

The n-th *harmonic number*, denoted by H_n, is given by the sum

$$H_n = \sum_{i=1}^{n} \frac{1}{i}. \tag{4.18}$$

4.2.1. Asymptotics. Theorem 4.2 gives the lower bound

$$H_n \ge \int_1^{n+1} \frac{dx}{x} = \ln(n+1). \tag{4.19}$$

For the upper bound, as x^{-1} is not defined at $x = 0$, we remove the first term

$$H_n - 1 = \sum_{i=2}^{n} \frac{1}{i} \le \int_1^n \frac{dx}{x} = \ln(n) \tag{4.20}$$

so that

$$\ln n < \ln(n+1) \le H_n \le \ln(n) + 1. \tag{4.21}$$

This gives a useful expression

$$H_n = \ln(n) + \Theta(1). \tag{4.22}$$

4.2.2. Better Asymptotics. Applying trapezoidal approximations, we can do even better. We follow the arguments of §1.2 almost verbatim. We set

$$I_n = \int_1^n \frac{dx}{x} = \ln n. \tag{4.23}$$

Let T_n be the value for the approximation of the integral I_n via the trapezoidal rule using step sizes 1. That is, we estimate $\int_i^{i+1} f(x)dx$ by $\frac{1}{2}(f(i) + f(i+1))$. Summing over $1 \le i \le n-1$,

$$T_n = \frac{1}{2} \cdot 1 + \sum_{k=2}^{n-1} \frac{1}{k} + \frac{1}{2} \cdot \frac{1}{n} = H_n - \frac{1}{2} - \frac{1}{n}. \tag{4.24}$$

Set

$$E_n = T_n - I_n \tag{4.25}$$

to be the error when approximating the integral of x^{-1} by the trapezoidal rule. For $1 \le k \le n-1$, let S_k denote the "sliver" of area above the curve $y = x^{-1}$ for $k \le x \le k+1$ but over the straight line between $(k, \frac{1}{k})$ and $(k+1, \frac{1}{k+1})$. The straight line is over the curve as the curve is convex. Then

$$E_n = \sum_{k=1}^{n-1} \mu(S_k), \tag{4.26}$$

where μ denotes the area.

Our goal is to bound the error.

Theorem 4.4. *E_n approaches a finite limit c as $n \to \infty$. Equivalently,*

$$\lim_{n\to\infty} \sum_{k=n}^{\infty} \mu(S_k) = 0. \tag{4.27}$$

Assuming Theorem 4.4, equations (4.23), (4.24), and (4.25) yield

$$H_n = T_n + \frac{1}{2} + \frac{1}{2n} = \ln n + c + o(1) + \frac{1}{2}. \tag{4.28}$$

For historic reasons we set

$$\gamma = c + \frac{1}{2}, \tag{4.29}$$

where γ is called Euler's constant. It first appeared in a 1734 paper by Leonhard Euler. Its approximate value is 0.577. γ has been very well studied; in the pantheon of mathematical constants one could put it just under π and e in importance. Unlike π and e, surprisingly little is known about it. For example, it is not (yet!) known whether or not γ is irrational. Using γ, we get a simple expression for H_n:

$$H_n = \ln n + \gamma + o(1). \tag{4.30}$$

Now, how do we show Theorem 4.4? We consider $\mu(S_k)$ in Asymptopia as $k \to \infty$. Roughly, $\mu(S_k)$, the error between the integral from k to $k+1$ of $f(x) = x^{-1}$ and the straight line approximation of $f(x)$. This error is caused by the *second* derivative of $f(x)$. (Had the second derivative been zero, the straight line would have been the precise function.) Here the second derivative $f''(x) = -x^{-3}$ is on the order of k^{-3} and the interval has length 1, so we feel the error should be on the order of k^{-3}. As k^{-3} is decreasing sufficiently quickly, the infinite sum of $\mu(S_k)$ should converge.

We have an exact expression:

$$\mu(S_k) = \frac{1}{2}\left[\frac{1}{k} + \frac{1}{k+1}\right] - \int_k^{k+1} \frac{dx}{x} = \frac{1}{2}\left[\frac{1}{k} + \frac{1}{k+1}\right] - \ln\left(1 + \frac{1}{k}\right). \tag{4.31}$$

Now we take Taylor series $\ln(1 + k^{-1}) = k^{-1} - \frac{1}{2}k^{-2} + \frac{1}{3}k^{-3} + O(k^{-4})$ and $\frac{1}{k+1} = k^{-1} - k^{-2} + k^{-3} + O(k^{-4})$. The k^{-1} and k^{-2} terms cancel and

$$\mu(S_k) = \frac{1}{2}k^{-3} - \frac{1}{3}k^{-3} + O(k^{-4}) = \frac{1}{6}k^{-3} + O(k^{-4}) \tag{4.32}$$

so that, as expected, $\mu(S_k) = O(k^{-3})$, and Theorem 4.4 is proven. Following §1.3, we can estimate the error in (4.30). As $\sum_{k=1}^{\infty} \mu(S_k) = c = \gamma - \frac{1}{2}$, we can write $E_n = \gamma - \frac{1}{2} - \sum_{k=n}^{\infty} \mu(S_k)$, giving the expression

$$H_n = \ln(n) + \gamma + \frac{1}{2n} - \sum_{k=n}^{\infty} \mu(S_k). \tag{4.33}$$

From (4.32)

$$\sum_{k=n}^{\infty} \mu(S_k) \sim \sum_{k=n}^{\infty} \frac{1}{6} k^{-3} + O(k^{-4}) = \frac{1}{12n^2} + O(n^{-3}), \tag{4.34}$$

giving the more accurate estimation

$$H_n = \ln(n) + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + O(n^{-3}). \tag{4.35}$$

4.2.3. $2n$ **Prisoners.** We now present an application of the approximation of the harmonic numbers to an amusing puzzle, the Prisoners Game.[1] The story goes as follows: We have $2n$ prisoners who have numbers $1, \ldots, 2n$. (As a puzzle $2n$ is often taken as 100. Also the prisoners have names. We delete these unnecessary aspects here.) Their numbers $1, \ldots, 2n$ are each written on a sheet of paper in $2n$ boxes (one number in each), and then the boxes are placed in random order on a table in a room. We let $\sigma(j)$ denote the number placed in box j. As the boxes were ordered randomly, we may consider σ to be a random permutation on $\{1, \ldots, n\}$.

Each prisoner is taken to the box room and may look in at most n boxes, attempting to find his own number. He must leave the room as it was found and is not allowed communication with the other prisoners once the game started. However, the prisoners can, and will, agree on a strategy before the game begins. The rule is draconian: If *any* prisoner fails to find his name, then *all* of the prisoners are executed. If *all* prisoners find their names, then they are all freed.

Each prisoner has probability $\frac{1}{2}$ of finding his name. If the prisoners work independently, the chance that they find their number would be a mere 2^{-n}. But we give a strategy for which the probability they are all freed is more than 30%.

Spoiler Alert: *Here Is the Strategy.* Prisoner j looks inside box j. He finds number $\sigma(j)$. If $\sigma(j) = j$, he is finished. Otherwise, he looks in the box associated to the number he just found, box $\sigma(j)$. He finds number $\sigma(\sigma(j))$. If $\sigma(\sigma(j)) = j$, he is finished. Otherwise he looks in the box associated to the number he just found, box $\sigma(\sigma(\sigma(j)))$.

[1] We have heard about this puzzle from Peter Winkler's *Seven Puzzles You Think You Must Not Have Heard Correctly* [**Win**]. The idea originally appears in [**GM03**].

Prisoner j continues in this fashion until he either finds his number or has opened n boxes.

The reason that this strategy works is that each prisoner is following a cycle of the permutation σ of the $2n$ numbers. Prisoner j succeeds if $\sigma^u(j) = j$ for some $1 \leq u \leq n$. If the permutation has no cycles of length greater than n, then each prisoner will find his number! We will see that the probability that a uniformly random permutation of the numbers from 1 to $2n$ does not have a cycle of length greater than n is about 30%.

Consider a cycle C of length exactly k, and count how many such cycles can exist in a given permutation. We have $\binom{2n}{k}$ possibilities for the entries in C, $(k-1)!$ possibilities for their order, giving $(2n)_k/k$ potential C. Each C is a cycle with probability $1/(2n)_k$ as k values of σ are determined. The expected number of such C is then $\frac{1}{k}$. Let A denote the expected number of cycles of length greater than n. Then

$$A = \sum_{k=n+1}^{2n} \frac{1}{k} = H_{2n} - H_n. \tag{4.36}$$

It is not possible that there is *more* than one cycle of length greater than n. Thus A also represents the probability that there is a cycle of length greater than n. We apply the asymptotics (4.35) to (4.36) giving

$$A = (\ln(2n) - \gamma + o(1)) - (\ln(n) - \gamma + o(1)) = \ln(2) + o(1). \tag{4.37}$$

The probability $1 - A$ that all the prisoners are freed then approaches $1 - \ln 2 = 0.31182$.

Chapter 5

Asymptotics of Binomial Coefficients $\binom{n}{k}$

> Many persons who have not studied mathematics confuse it with arithmetic and consider it a dry and arid science. Actually, however, this science requires great fantasy.
> – Sofia Kovalevskaya

What is a good asymptotic formula for the binomial coefficient $\binom{n}{k}$? The simple answer is: It depends!

When k is fixed, it is easy:

$$\binom{n}{k} \sim \frac{n^k}{k!} \text{ for } k \text{ fixed.} \tag{5.1}$$

But now suppose k also goes to infinity. There are different answers depending on the asymptotic relation between n and k. We introduce the useful notation

$$(n)_k = n(n-1)(n-2)(n-3)\cdots(n-k+1) \tag{5.2}$$

and read $(n)_k$ as "n lower k".

5.1. k Relatively Small

In this section we restrict ourselves to the cases when $k = o(n)$.

We start by writing

$$\binom{n}{k} = A\frac{n^k}{k!}, \tag{5.3}$$

where we set

$$A = A(n,k) := \frac{(n)_k}{n^k} = \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right). \tag{5.4}$$

As Stirling's formula gives the asymptotics for $k!$, an asymptotic formula for $A(n,k)$ leads directly to an asymptotic formula for $\binom{n}{k}$. Clearly $A \le 1$. Our approach to A is via its logarithm

$$\ln A = \sum_{i=1}^{k-1} \ln\left(1-\frac{i}{n}\right). \tag{5.5}$$

The main estimate is that as $x \to 0$, $\ln(1-x) \sim x$.

Case 1. $k = o(\sqrt{n})$. Then

$$\sum_{i=1}^{k-1} \ln\left(1-\frac{i}{n}\right) \sim \sum_{i=1}^{k-1} -\frac{i}{n} \sim -\frac{k^2}{2n} = o(1). \tag{5.6}$$

Thus $A \sim 1$. Thus the asymptotic formula (5.1) still holds. With $k \to \infty$, employing Stirling gives

$$\binom{n}{k} \sim \frac{n^k}{k!} \sim \left(\frac{ne}{k}\right)^k (2\pi k)^{-1/2}. \tag{5.7}$$

Case 2. $k \sim c\sqrt{n}$ with c a fixed positive real. Then the argument above gives $\ln A \sim \frac{c^2}{2}$ and so

$$\binom{n}{k} \sim e^{-c^2/2}\frac{n^k}{k!}. \tag{5.8}$$

For bigger k, let us use the Taylor series expansion $\ln(1-x) = -x+O(x^2)$. Applying this to the formula for $\ln A$, we have $\ln(1-\frac{i}{n}) = -\frac{i}{n}+O(\frac{i^2}{n^2})$. Summing over $1 \le i < k$ gives $\frac{k^2}{2n}$ from the first term and the $O(k^3n^{-2})$ for the error term. We see that the error term is $o(1)$ as long as $k = o(n^{2/3})$. This gives

Case 3. $k = o(n^{2/3})$. Then

$$\binom{n}{k} \sim e^{-k^2/2n}\frac{n^k}{k!}. \tag{5.9}$$

As k gets bigger, we use the more precise $\ln(1-x) = -x - \frac{x^2}{2} + O(x^3)$. Applying this to the formula for $\ln A$, we have $\ln(1-\frac{i}{n}) = -\frac{i}{n} - \frac{i^2}{2n^2} + O(i^3 n^{-3})$. Summing over $1 \le i < k$ gives $\frac{k^2}{2n}$ from the first term, $\frac{k^3}{6n^2}$ for the second term, and $O(k^4 n^{-3})$ for the error. As long as $k = o(n^{3/4})$ the error term is $o(1)$. This gives

Case 4. $k = o(n^{3/4})$. Then

$$\binom{n}{k} \sim e^{-k^2/2n} e^{-k^3/6n^2} \frac{n^k}{k!}. \tag{5.10}$$

Indeed, we get an infinite sequence of formulae as we take the expansion of $\ln(1-x)$ out to more and more terms. For each integer $j \ge 2$ we have a formula for when $k = o(n^{1-(1/j)})$. In practice, use of even Case 4 is extremely rare.

Case 5. $k = o(n)$. All $1 \le i \le k-1$ have $\frac{i}{n} = o(1)$ so that $\ln(1-\frac{i}{n}) \sim -\frac{i}{n}$ and therefore $\ln A \sim -\frac{k^2}{2n}$. Thus

$$A(n,k) = e^{-\frac{k^2}{2n}(1+o(1))} \tag{5.11}$$

and

$$\binom{n}{k} \sim \left(\frac{ne}{k}\right)^k (2\pi k)^{-1/2} e^{-\frac{k^2}{2n}(1+o(1))}. \tag{5.12}$$

When $k \gg \sqrt{n \ln n}$, the $k^{-1/2}$ factor may be absorbed into the $o(1)$ term in the exponent so we get the simpler expression

$$\binom{n}{k} = \left(\frac{ne}{k}\right)^k e^{-\frac{k^2}{2n}(1+o(1))} \text{ for } \sqrt{n \ln n} \ll k \ll n. \tag{5.13}$$

With the $1 + o(1)$ factor in the exponent, (5.13) is not as accurate as the previous bounds. It is still a very useful estimate in many applications.

An often very handy upper bound, applying (4.11), is

$$\binom{n}{k} \le \frac{n^k}{k!} \le \left(\frac{ne}{k}\right)^k. \tag{5.14}$$

This is true for *all* n, k, not just asymptotically. When k is linear in n, there are other estimates (such as (5.31)) which are preferable. But when $k = o(n)$ and $k \to \infty$, (5.14) is pretty good in the sense that the k-th root of $\binom{n}{k}$, applying (5.12), is asymptotic to $\frac{ne}{k}$. Again,

this is not as strong as a truly asymptotic estimate, but is still very useful in many applications.

5.2. Some Exercises

Working asymptotically with binomial coefficients takes some practice. *Warning*: The solutions immediately follow the problems!

Exercise 5.1. Find an asymptotic formula for

$$f(n) = \sum_{k=1}^{n} A(n,k) \text{ with } A(n,k) = \frac{(n)_k}{n^k}. \tag{5.15}$$

Solution. The $A(n,k)$ go from near one to near zero when $k = \Theta(\sqrt{n})$ so we are led to the scaling $k = c\sqrt{n}$ of Case 2. In that region

$$f(n) \sim \sum_k e^{-c^2/2} \sim \int_{k=0}^{\infty} e^{-c^2/2} dk \sim \int_{c=0}^{\infty} e^{-c^2/2} \sqrt{n} dc = \sqrt{n}\sqrt{\pi/2}. \tag{5.16}$$

Exercise 5.2. Find an asymptotic formula for

$$g(n) = \sum_{k=1}^{n} \frac{A(n,k)}{k} \quad \text{with } A(n,k) = \frac{(n)_k}{n^k}. \tag{5.17}$$

Solution. One is tempted to scale $k = c\sqrt{n}$ as above but that makes life too complicated. The harmonic series $\sum k^{-1}$ does not see the fine gradations in the $\Theta(\sqrt{n})$ range. Rather let us split the sum into a small, medium, and large range. The medium range should include $k = \Theta(\sqrt{n})$. We have a lot of leeway but, for definiteness, let $\sqrt{n}\ln^{-1} n < k < \sqrt{n}\ln n$ be the medium range. The small and large ranges are then $k < \sqrt{n}\ln^{-1} n$ and $k > \sqrt{n}\ln n$, respectively. Over the small region $A(n,k) \sim 1$, so the summation of $A(n,k)k^{-1}$ is asymptotic to the summation of k^{-1}, the harmonic series, which is $\sim \ln(\sqrt{n}\ln^{-1}(n)) \sim \frac{1}{2}\ln n$. In the large range the $A(n,k) = o(1)$ and the harmonic series sums over that range to $\sim \frac{1}{2}\ln n$, so the summation of $A(n,k)k^{-1}$ is $o(\ln n)$. In the middle range $A(n,k)$ has a complex behavior as described in Case 2. But since $A(n,k) \le 1$, the summation of $A(n,k)k^{-1}$ over the middle range is at most the summation of k^{-1} over the middle range which is only $\sim 2\ln(\ln n)$, or $o(\ln n)$. Thus $g(n) \sim \frac{1}{2}\ln n$.

Exercise 5.3. Find an asymptotic formula for

$$g(\epsilon) = \sum_{k=1}^{\infty} k^2(1-\epsilon)^k \tag{5.18}$$

as $\epsilon \to 0^+$.

Solution. The exponential term starts to dip when $k = \Theta(\epsilon^{-1})$ so we are led to the parametrization $k = c\epsilon^{-1}$ for which $(1-\epsilon)^k \sim e^{-c}$. Thus

$$g(\epsilon) \sim \int_{c=0}^{\infty} c^2\epsilon^{-2}e^{-c}\epsilon^{-1}dc = \epsilon^{-3}\int_{c=0}^{\infty} c^2e^{-c}dc = 2\epsilon^{-3}. \tag{5.19}$$

Remark. The exact formula is

$$g(\epsilon) = 2(1-\epsilon)^2\epsilon^{-3} + (1-\epsilon)\epsilon^{-2}. \tag{5.20}$$

Exercise 5.4. Set

$$f(n,p) = \sum_{k=1}^{n} k^2(n)_k p^{k+1}. \tag{5.21}$$

Find a parametrization of $p = p(n)$ so that we can "see" $f(n,p)$ go from near zero to infinity.

Solution. Let's first use the upper bound $(n)_k \le n^k$ and parametrize $p = \frac{c}{n}$ so that $f(n,p) \le p\sum k^2c^k$. When $c < 1$, the summation of k^2c^k converges so $f(n,p) = O(p) = o(1)$. What about when $c = 1$? We want the reverse inequality. From Case 1 $(n)_k \sim n^k$ when (say) $k \le n^{0.49}$. Thus

$$f\left(n, \frac{1}{n}\right) \ge (1-o(1))\frac{1}{n}\sum_{k\le n^{0.49}} k^2 \sim \frac{1}{n}\frac{1}{3}(n^{0.49})^3, \tag{5.22}$$

and so $f(n,\frac{1}{n}) \to \infty$. The change in $f(n,p)$ thus takes place between $\frac{c}{n}$ and $\frac{1}{n}$, where c is an arbitrary (but *fixed*) real less than one. We have to look more closely. What if $p = p(n) \sim \frac{1}{n}$ but $p(n) < \frac{1}{n}$, for example, $p(n) = \frac{1}{n} - n^{-10}$? To examine these, we parametrize $p = \frac{1-\epsilon}{n}$ where $\epsilon = \epsilon(n) \to 0$. We rewrite

$$f(n,p) \sim n^{-1}\sum_{k=1}^{n} k^2(np)^k A(n,k) = n^{-1}\sum_{k=1}^{n} k^2(1-\epsilon)^k A(n,k). \tag{5.23}$$

As $A(n,k) \le 1$, we have the upper bound

$$f(n,p) \le (1+o(1))n^{-1}\sum_k k^2(1-\epsilon)^k = (2+o(1))n^{-1}\epsilon^{-3}, \tag{5.24}$$

where we have employed the previous exercise. When $\epsilon \gg n^{-1/3}$, $n^{-1}\epsilon^{-3} \to 0$ for $f(n,p) = o(1)$.

What about smaller ϵ? Suppose $\epsilon = Kn^{-1/3}$ for some constant K. Then $\sum_{k=1}^{\infty} k^2(1-\epsilon)^k \sim 2\epsilon^{-3} = 2K^{-3}n$. But further, the bulk of that sum occurs when $k = \Theta(\epsilon^{-1}) = \Theta(n^{1/3})$. In that range the extra factor $A(n,k)$ is still asymptotically one. Let's split the sum for $f(n,p)$ somewhere between $\Theta(n^{1/3})$ and $\Theta(n^{1/2})$, say at $n^{0.4}$. The sum for $k > n^{0.4}$ is negligible as the $(1-\epsilon)^k$ term is exponentially small. For $k \le n^{0.4}$, the factor $A(n,k) \sim 1$, and so the terms are asymptotically $k^2(1-\epsilon)^k$ and so their sum is asymptotically $2K^{-3}n$. Thus $f(n,p) \sim 2K^{-3}$. So we've actually found the fine behavior of $f(n,p)$: when $p = \frac{1}{n} - Kn^{-4/3}$, $f(n,p) \to 2K^{-3}$. The function $2K^{-3}$ approaches infinity and zero as K approaches zero and infinity, respectively. Therefore:

- If $p(n) = \frac{1-\epsilon(n)}{n}$ and $\epsilon(n) \gg n^{-4/3}$, $f(n,p) \to 0$.
- If $p(n) = \frac{1-\epsilon(n)}{n}$ and $\epsilon(n) \ll n^{-4/3}$, $f(n,p) \to \infty$.
- If $p(n) = \frac{1-\epsilon(n)}{n}$ and $\epsilon(n) \sim Kn^{-4/3}$, $f(n,p) \to 2K^{-3}$.

The parametrization $p = \frac{1}{n} - Kn^{-4/3}$, $K \in (0,\infty)$, gives the critical window when $f(n,p)$ is increasing from "zero to infinity".

Exercise 5.5. Find an asymptotic formula for

$$h(n,k) = \binom{n}{k} k^{k-2}p^{k-1}(1-p)^{k(n-k)+\binom{k}{2}-(k-1)} \tag{5.25}$$

$$\text{with } k \sim cn^{2/3} \text{ and } p = n^{-1}.$$

Remark. $h(n,k)$ is the expected number of trees of size k in the random graph $G(n,p)$.

Solution. This is very delicate! From Case 4 we have

$$\binom{n}{k} \sim e^{-k^2/2n}e^{-c^3/6}\frac{n^k}{k!} \sim e^{-k^2/2n}e^{-c^3/6}\frac{n^k e^k}{k^k\sqrt{2\pi k}}. \tag{5.26}$$

As

$$k(n-k)+\binom{k}{2}-(k-1)=kn-\frac{1}{2}k^2+O(k)$$

and

$$\ln(1-p)=-n^{-1}+O(n^{-2}),$$

$$\ln[(1-p)^{k(n-k)+\binom{k}{2}-(k-1)}]=-k+\frac{k^2}{2n}+o(1). \tag{5.27}$$

Now many of the terms cancel:

$$\begin{aligned} h(n,k) &\sim e^{-k^2/2n}e^{-c^3/6}\frac{n^k e^k}{k^k\sqrt{2\pi k}}k^{k-2}n^{1-k}e^{-k}e^{k^2/2n} \\ &\sim e^{-c^3/6}nk^{-5/2}(2\pi)^{-1/2} \\ &\sim e^{-c^3/6}n^{-2/3}c^{-5/2}(2\pi)^{-1/2}. \end{aligned} \tag{5.28}$$

Exercise 5.6. With $h(n,k)$ as in (5.25) and $0<a<b$ constants, find an asymptotic formula for

$$H(n,a,b):=\sum_{an^{2/3}<k<bn^{2/3}} h(n,k). \tag{5.29}$$

Remark. This quantity is the expected number of tree components of size between $an^{2/3}$ and $bn^{2/3}$ in $G(n,p)$.

Solution. We approximate $h(n,k)$ by (5.28). The parametrization $k=cn^{2/3}$ turns the sum into the integral

$$\begin{aligned} H(n,a,b) &\sim \int_k e^{-c^3/6}n^{-2/3}c^{-5/2}(2\pi)^{-1/2}dk \\ &= \int_a^b e^{-c^3/6}c^{-5/2}(2\pi)^{-1/2}dc. \end{aligned} \tag{5.30}$$

Note that the answer depends only on a and b, the number of vertices n has scaled out.

5.3. k Linear in n

Here we assume $k\sim pn$ where $0<p<1$. Now we write $\binom{n}{k}=\frac{n!}{(k!(n-k)!)}$ and apply Stirling's formula to the three terms. (By the way, this is not a good idea to try when $k=o(n)$—the formula for the $(n-k)!$ gets very messy, and you *definitely* should use the formulae of Cases 1–5 when they apply.)

Let us throw the square root terms into the form $2^{o(n)}$. One finds that

$$\binom{n}{k} = 2^{n(H(p)+o(1))} \text{ when } k \sim pn, \tag{5.31}$$

where H is the entropy function

$$H(p) = -p \lg p - (1-p) \lg(1-p). \tag{5.32}$$

The entropy function is defined for $0 < p < 1$, is symmetric about $p = \frac{1}{2}$, has $\lim_{p\to 0+} H(p) = \lim_{p\to 1-} H(p) = 0$ (indeed, it's handy to define $H(0) = H(1) = 0$), and has a maximum at $p = \frac{1}{2}$ where $H(1/2) = 1$, as can be seen in Figure 1. We do note that this is not the full asymptotics since the $o(1)$ is in the exponent multiplied by n. More accurate formulae are available, but this one is quite handy.

Now suppose $p > \frac{1}{2}$ and consider the probability that the binomial distribution $BIN[n, \frac{1}{2}]$ is at least pn. This is precisely $2^{-n} \sum \binom{n}{k}$ where the sum is over $k \geq pn$. For rougher approximations we may finesse the summation. The binomial coefficients decrease for $k \geq \frac{n}{2}$. The sum is therefore at least the largest value and at most (as there are fewer than n terms) n times the largest value. When we allow a $1 + o(1)$ in the exponent, a factor of n "downstairs" does not matter.

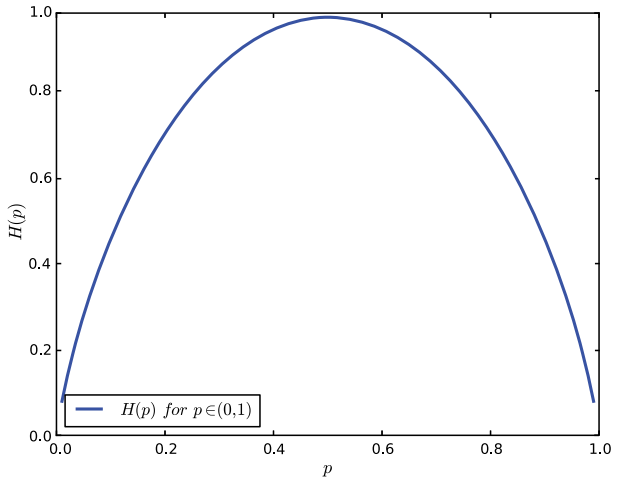


Figure 1. Entropy function $H(p)$ for $p \in (0, 1)$

Therefore (5.31) gives the formula for the summation of $\binom{n}{k}$ over all $k \geq pn$. Dividing by 2^n

$$\Pr[BIN[n, \frac{1}{2}] \geq pn] = 2^{n(H(p)-1+o(1))}. \tag{5.33}$$

More accurate formulae are available, but this one is quite handy.

Strangely, we may use (5.33), a result asymptotic in n, to give a bound for all n.

Theorem 5.7. *For any* $1 > p > \frac{1}{2}$ *and* any $n \geq 1$,

$$\Pr[BIN[n, \frac{1}{2}] \geq pn] \leq 2^{n(H(p)-1)}. \tag{5.34}$$

Proof. Assume not. Let a be such that (5.34) fails. We can write

$$\Pr[BIN[a, \frac{1}{2}] \geq pa]^{1/a} = 2^{H(p)-1+\epsilon}, \tag{5.35}$$

where ϵ is now a positive constant. Let $s \geq 1$, integral, and bound from below the probability that $BIN[as, \frac{1}{2} \geq p(as)]$. Break up the as coin flips into s blocks of a. If in each block there are at least pa heads, then there are at least $p(as)$ heads in total. Thus

$$\Pr[BIN[as, \frac{1}{2} \geq p(as)] \geq \Pr[BIN[a, \frac{1}{2}] \geq pa]^s = 2^{(as)(H(p)-1+\epsilon)}. \tag{5.36}$$

As $s \to \infty$ and setting $n = as$, this contradicts the asymptotics of (5.33). □

5.3.1. Exercise.

Exercise 5.8. In what ranges of k is $\binom{n}{k}$ bigger, smaller, and roughly the same as 8^k?

Solution. When $k \sim pn$, we have

$$\binom{n}{k} = 2^{n(H(p)+o(1))} \text{ and } 8^k = 2^{3n(p+o(1))}. \tag{5.37}$$

The curves $y = H(p)$ and $y = 3p$ intersect at a unique value $p = p_0$. Let ϵ be arbitrarily small but fixed. For $k = pn$ with $p < p_0 - \epsilon$, $H(p) > 3p$, and so $\binom{n}{k}$ is much bigger. When $p > p_0 + \epsilon$, $H(p) < 3p$, and so 8^k is much bigger. The graphs of the two functions will cross at $k_0 \sim p_0 n$.

5.4. At and Near the Middle Binomial Coefficient

Here we look at the region when k is very close to $\frac{n}{2}$. (When n is odd, look at $\lfloor n/2 \rfloor$ and all the asymptotics will be the same.)

Direct application of Stirling's formula gives

$$\binom{n}{\frac{n}{2}} \sim \sqrt{\frac{2}{\pi n}} 2^n. \tag{5.38}$$

(Note this means the probability of having precisely half heads with n flips of a fair coin is $\sim \sqrt{\frac{2}{\pi n}}$.) Now let us write $k = \frac{n+i}{2}$. One then has

$$B := \binom{n}{k} / \binom{n}{n/2} = \frac{(n/2)!(n/2)!}{k!(n-k)!} = \prod_{j=1}^{i/2} \frac{\frac{n}{2} - j + 1}{\frac{n}{2} + j}, \tag{5.39}$$

and therefore

$$\ln B = \sum_{j=1}^{i/2} \ln\left(\frac{\frac{n}{2} - j + 1}{\frac{n}{2} + j}\right) = \sum_{j=1}^{i/2} -\frac{4j}{n} + O\left(\frac{j^2}{n^2}\right) = -\frac{i^2}{2n} + O(i^3 n^{-2}), \tag{5.40}$$

and so

$$\binom{n}{k} \sim \binom{n}{\frac{n}{2}} e^{-(i^2/2n)} \tag{5.41}$$

as long as $i = o(n^{2/3})$. In particular, if you parametrize (good idea!)

$$k = \frac{n + c\sqrt{n}}{2}, \tag{5.42}$$

then we have the asymptotics

$$\binom{n}{k} \sim \binom{n}{\frac{n}{2}} e^{-(c^2/2)}, \tag{5.43}$$

and this will be valid for any constant c and even $c = c(n) \to \infty$ as long as $c(n) = o(n^{1/6})$.

One particular case is often handy. If we set (another good idea!) $c = 100\sqrt{\ln n}$, then we have

$$\binom{n}{k} = \Theta(2^n n^{-1/2} n^{-5000}). \tag{5.44}$$

The binomial coefficients increase to the middle and then decrease, and there are only $O(n)$ terms, so we get

$$\sum_{s\geq k}\binom{n}{s} = \Theta(n2^n n^{-1/2} n^{-5000}). \tag{5.45}$$

In many applications this makes an excellent cut, as this sum is so small.

5.4.1. Exercise.

Exercise 5.9. Find the asymptotics of

$$S := \sum_{s=0}^{n}\binom{n}{k}^3. \tag{5.46}$$

Solution. We see that for $k = \frac{n+100\sqrt{\ln n}}{2}$ as above,

$$\binom{n}{k}^3 = O(\binom{n}{\frac{n}{2}}^3 n^{-15000}) \tag{5.47}$$

and so

$$\sum_{s\geq k}\binom{n}{s}^3 = O(\binom{n}{\frac{n}{2}}^3 n^{-14999}) \tag{5.48}$$

so that all these terms together are negligible compared to the biggest (the middle) term. By symmetry the same holds for the sum over $s \leq k^-$ where $k^- = n - k$. So the asymptotics of the sum over all s is the same as the asymptotics of the sum over the limited range $k^- \leq s \leq k$. In that limited range we use the parametrization (5.42). Then, from (5.43)

$$\begin{aligned} S &\sim \binom{n}{n/2}^3 \sum_k e^{-3c^2/2} \sim \binom{n}{n/2}^3 \int_{c=-\infty}^{+\infty} e^{-3c^2/2}\frac{\sqrt{n}}{2}dc \\ &\sim \binom{n}{n/2}^3 \sqrt{n\pi/6}. \end{aligned} \tag{5.49}$$

Finally, the asymptotics (5.38) for the middle binomial coefficients give

$$S \sim \left(2^n\sqrt{\frac{2}{\pi n}}\right)^3 \sqrt{n\pi/6} \sim 8^n \cdot \frac{\sqrt{4/3}}{\pi n}. \tag{5.50}$$

5.5. The Binomial Distribution

Let n be a positive integer and $0 \le p \le 1$. The binomial distribution with parameters n, p, denoted $BIN[n,p]$, is often the first distribution studied in a probability course. It is the number of heads after n independent flips of a coin, where the probability that the coin comes up heads is p. It has the precise expression

$$\Pr[BIN[n,p] = k] = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } 0 \le k \le n. \tag{5.51}$$

$BIN[n,p]$ is known[1] to have expection pn and variance $np(1-p)$. A basic and very general tail bound is given by Chebyshev's inequality:

Theorem 5.10. *Let X be any distribution with mean μ and variance σ^2. Let λ be a positive real. Then*

$$\Pr[|X - \mu| \ge \lambda\sigma] \le \lambda^{-2}. \tag{5.52}$$

Proof. This follows immediately from the inequality

$$\sigma^2 = E[(X-\mu)^2] \ge (\lambda\sigma)^2 \Pr[|X - \mu| \ge \lambda\sigma]. \tag{5.53}$$

□

In the special case of the binomial distribution

$$\Pr[|BIN[n,p] \ge \lambda\sqrt{np(1-p)}] \le \lambda^{-2}. \tag{5.54}$$

In most cases inequality (5.54) is quite weak compared to the actual probability. Nonetheless it can be quite useful.

5.6. The Binomial Tail

Let S_n denote the number of heads minus the number of tails after n flips of a fair coin. If there are t heads, there are $n - t$ tails so $S_n = t - (n-t) = n - 2t$. S_n is then directly connected to the binomial distribution by

$$\Pr[S_n = a] = \Pr[BIN[n, \frac{1}{2}] = (n+a)/2] = \binom{n}{(n+a)/2} 2^{-n}. \tag{5.55}$$

[1] We omit the proofs for these basic results.

Using the parametrization (5.42),

$$\Pr[S_n \geq c\sqrt{n}] = \sum_{b \geq c\sqrt{n}} \Pr[BIN[n, \frac{1}{2}] = (n+b)/2]. \tag{5.56}$$

The Central Limit Theorem (see §8.5) gives the asymptotics of $\Pr[S_n \geq c\sqrt{n}]$ when c is fixed. Using Chernoff bounds, Theorem 8.2 gives an upper bound on $\Pr[S_n \geq c\sqrt{n}]$ for any $c = c(n)$. Here we find the asymptotics for $\Pr[S_n \geq c\sqrt{n}]$ when $c = c(n)$ approaches infinity appropriately slowly. Let N denote the standard normal distribution.

Theorem 5.11. *Let $c = c(n) \to +\infty$ so that $c(n) = o(n^{1/6})$. Then*

$$\Pr[S_n \geq c(n)\sqrt{n}] \sim \Pr[N \geq c(n)] = \int_{c(n)}^{\infty} \sqrt{\frac{1}{2\pi}} e^{-t^2/2} dt. \tag{5.57}$$

From our results (3.8) on the Gaussian tail this is equivalent to

$$\Pr[S_n \geq c(n)\sqrt{n}] \sim \frac{1}{c(n)} \frac{1}{\sqrt{2\pi}} e^{-c(n)^2/2}. \tag{5.58}$$

Remark. The application of Chebyshev's inequality (5.54) with $p = \frac{1}{2}$ yields

$$\Pr[BIN[n, \frac{1}{2}] \geq \frac{n}{2} + c(n)\sqrt{n}] \leq [2c(n)]^{-2}. \tag{5.59}$$

The exponential bound (5.58) is thus far stronger when $c(n)$ is large.

Upper Bound. Set $f(k) = \binom{n}{k} 2^{-n}$. For $0 \leq k < n$ set

$$g(k) = \frac{f(k+1)}{f(k)} = \frac{n-k}{k+1}. \tag{5.60}$$

For $k > \frac{n}{2}$, $g(k) < 1$ and for all $0 \leq k < n$, the function $g(k)$ is decreasing. (That is, the binomial distribution $BIN[n, \frac{1}{2}]$ is a concave function of k which hits a maximum at $k = \frac{n}{2}$ and then decreases.) Set $k = \lceil (n + c(n)\sqrt{n})/2 \rceil$. As g is decreasing $f(k+i) \leq f(k)g(k)^i$ for all $i \geq 0$. We bound the tail by a geometric series

$$\sum_{l=k}^{n} \Pr[BIN[n, \frac{1}{2}] = l] \leq \sum_{i=0}^{\infty} f(k)g(k)^i = \frac{f(k)}{1 - g(k)}. \tag{5.61}$$

From (5.60) $(1 - g(k))^{-1} \sim \sqrt{n}/(2c(n))$. Combining (5.38) and (5.43) gives $f(k) \sim \sqrt{2/n\pi} \exp[-c(n)^2/2]$ so that that the upper bound (5.61) asymptotically matches (5.58).

Lower Bound. We bound $\Pr[S_n \geq c(n)\sqrt{n}]$ from below by

$$\Pr[c(n)\sqrt{n} \leq S_n \leq 2c(n)\sqrt{n}],$$

which is $\sum_L^U f(k)$ with $L = \lceil (n + c(n)\sqrt{n})\rceil$ and $U = \lfloor (n + 2c\sqrt{n})\rfloor$. For $L \leq k \leq U$, parametrizing $k = (n + t\sqrt{n})/2$, $f(k) \sim h(k)$ where $h(k) \sim \sqrt{2/n\pi}e^{-t^2/2}$. We apply Theorem 4.1 to approximate $\sum h(k)$ by the integral

$$\begin{aligned} \sum_L^U h(k) &\sim \int_L^U h(k)dk \\ &= \frac{\sqrt{n}}{2}\int_{c(n)} 2c(n)\sqrt{\frac{2}{n\pi}}e^{-t^2/2}dt \\ &= \frac{1}{\sqrt{2\pi}}\Pr[c(n) \leq N \leq 2c(n)]. \end{aligned} \tag{5.62}$$

From Theorem 3.1, $\Pr[N \geq 2c(n)] \ll \Pr[N \geq c(n)]$. Thus $\Pr[c(n) \leq N \leq 2c(n)] \sim \Pr[N \geq c(n)]$ and the lower bound (5.62) also asymptotically matches (5.58).

Remark. Another approach to the lower bound is to show that $f(k + i)$ is well approximated by $f(k)g(k)^i$ until the terms become appropriately negligible.

Chapter 6

Unicyclic Graphs

> Comstock grins and says, "You sound awfully sure of yourself, Waterhouse! I wonder if you can get me to feel that same level of confidence."
> Waterhouse frowns at the coffee mug. "Well, it's all in the math," he says. "If the math works, why then you *should* be sure of yourself. That's the whole point of math."
> – Neal Stephenson, *Cryptonomicon*

A unicyclic graph on n vertices is a connected graph with precisely n edges. Such graphs have a clear shape. There is a cycle of vertices and then trees "sprouting out" from the vertices on the cycle. (Figure 1 near the end of this chapter gives a typical picture.) We shall let $UN(n)$ denote the number of labeled unicyclic graphs on n vertices. For convenience, we shall consider the vertices labeled $1, \ldots, n$.

Our goal in this chapter is to first find an exact formula for $U(n)$ and then an asymptotic formula. We begin, however, with results on trees that are most interesting in their own right.

6.1. Rooted Trees

A rooted tree[1] T consists of three parts:

(1) A finite set V of vertices.

(2) A designated vertex $r \in V$ which is called the *root*.

(3) A function π which sends vertices other than the root into vertices, which may or may not be the root. Formally, $\pi : V - \{r\} \to V$. When $\pi(v) = w$, we say that w is the *parent* of v and that v is a *child* (there may be others) of w.

For T to be a rooted tree it must satisfy one of the following:

(1) Beginning at any nonroot v, if one repeatedly applies π, one eventually reaches the root r.

(2) There are no "cycles" in π. More explicitly, there are no distinct $v_0, \ldots, v_s$ with $\pi(v_i) = v_{i+1}$ for $0 \leq i < s$ and $\pi(v_s) = v_0$. This includes the case $s = 0$, as there is no v_0 which is its own parent.

The two properties are equivalent. If there were a cycle $v_0, \ldots, v_r$, then beginning at v_0 and applying π repeatedly one would only reach $v_0, \ldots, v_r$ and not the root r. Inversely, suppose that starting at $v = v_0$ and continually applying π, one never reached the root r. Define $v_{i+1} = \pi(v_i)$, a valid function as this would be then defined for all i. As V is finite, there would be a repetition which means there would be a first repetition, $i < j$ with $v_i = v_j$. In that case $v_i, v_{i+1}, \ldots, v_{j-1}$ would form a cycle.

Definition 6.1. A vertex v of a rooted tree is a *leaf* if v is not a root *and* there is no vertex w with $\pi(w) = v$.

We picture the root r at the bottom of T. When $\pi(x) = y$, we draw a directed edge from x to y. Following the paths always leads us to the bottom node r. The leaves v are nonroots at which the tree ends.

[1] There are many equivalent definitions of rooted trees in the literature. The definition we use is motivated by data structures used in computer science.

Consider rooted trees on a vertex set $V = \{1, \ldots, n\}$ with root $r = 1$. Our first goal is to give a remarkable formula due to Cayley for the number of these trees.

Theorem 6.2 (Cayley's Formula for Rooted Trees)**.** *The number of rooted trees on $V = \{1, \ldots, n\}$ with root $r = 1$ is precisely n^{n-2}.*

A *tree* on $V = \{1, \ldots, n\}$ is a connected graph with no cycles. Cayley's formula is more famous as the following:

Theorem 6.3 (Cayley's Formula for Trees)**.** *The number of labeled trees on $V = \{1, \ldots, n\}$ is precisely n^{n-2}.*

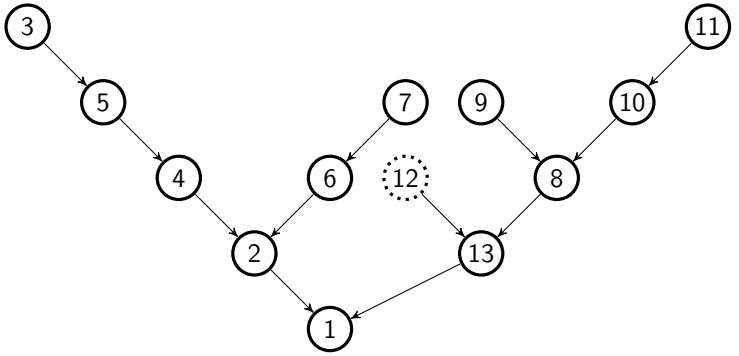
There is an easy bijection between rooted trees and trees. Given the rooted tree, define a tree by "erasing the arrows" and looking at the graph with edges $\{x, \pi(x)\}$. Conversely, let a tree T be given on $V = \{1, \ldots, n\}$, and let $r \in V$ be designated. For any $v \neq r$, there is a unique path from v to r. Let w be the vertex directly after v in that path. (If the path has length 1, it will be r.) Now define $\pi(v) = w$. Beginning at v and applying π, one then goes through the path, eventually reaching the root r. Thus T corresponds to a rooted tree with root r, and so Theorems 6.2 and 6.3 are equivalent. We shall give a proof of Theorem 6.2. Actually, many proofs are available in the literature. The proof we give *encodes* rooted trees by means of what are called Prüfer sequences. A tree's Prüfer sequence shall be a sequence of length $n - 1$ of vertices v where the final term is the root r. Vertices v may, and often do, repeat in the sequence.

6.2. Rooted Trees to Prüfer Sequences

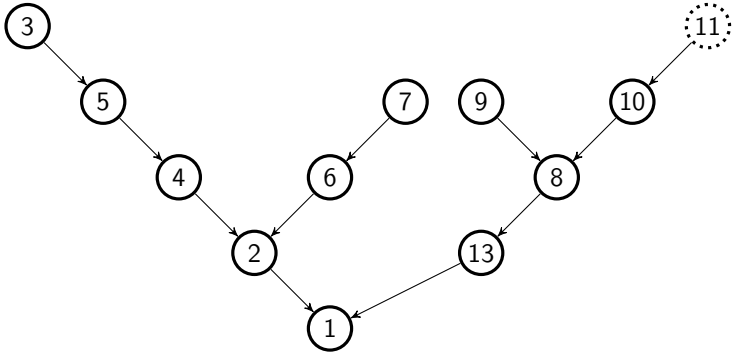
We begin with a rooted tree on $V = \{1, \ldots, n\}$ with root $r = 1$.

The Prüfer sequence is found sequentially. For the first step, find that leaf w with maximal value. Then $\pi(w)$ (*not w itself!*) becomes the first element of the Prüfer sequence. Now delete w from the tree. (As w was a leaf, this will not cause trouble with other $\pi(v)$.) Now iterate. At each step we find the leaf w on the remaining tree with maximal value, add $\pi(w)$ to the Prüfer sequence, and delete w from the tree. We end after $n - 1$ steps. After $n - 2$ steps, only the root r and some other vertex v remain in the tree. On the final step, that

v will be the leaf with maximal value (indeed, the only leaf) so that $\pi(v) = r$ will be the final element of the Prüfer sequence. Below we see this algorithm[2] in action:

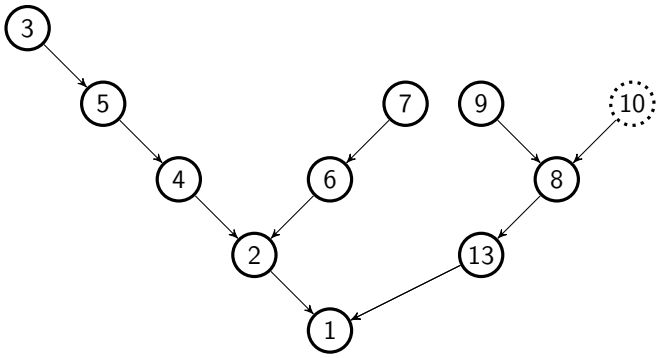


Sequence: 13

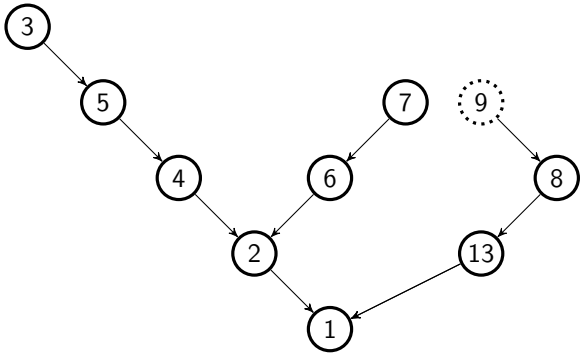


Sequence: 13, 10

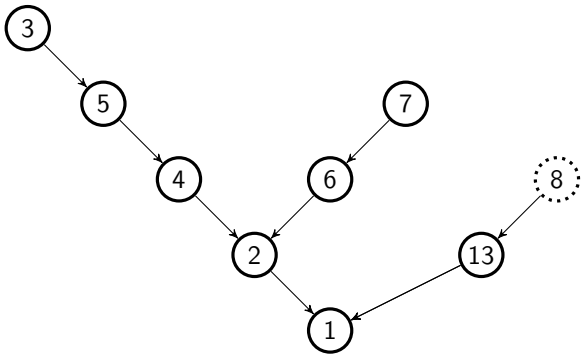
[2]Efficient implementation of this algorithm is an interesting problem in computer science.



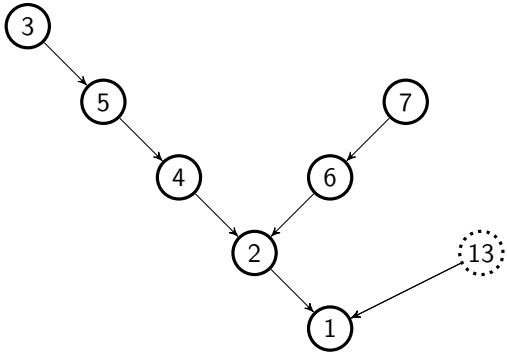
Sequence: 13, 10, 8



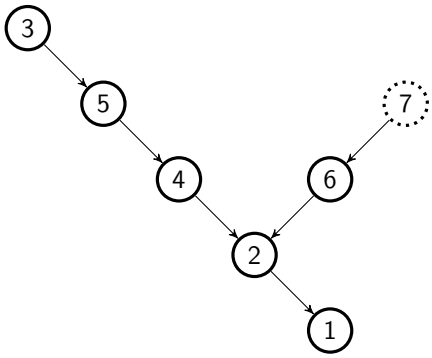
Sequence: 13, 10, 8, 8



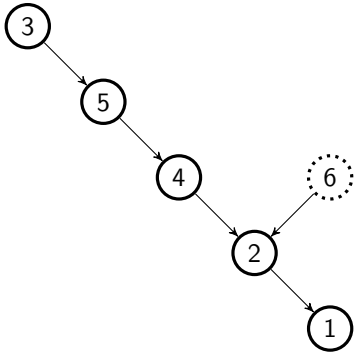
Sequence: 13, 10, 8, 8, 13



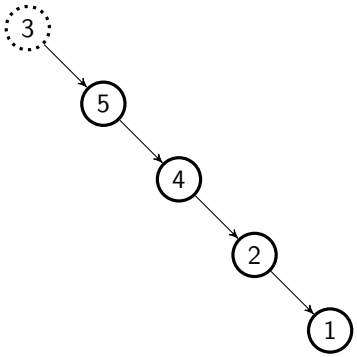
Sequence: 13, 10, 8, 8, 13, 1



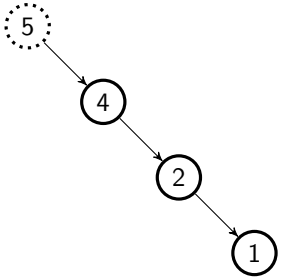
Sequence: 13, 10, 8, 8, 13, 1, 6



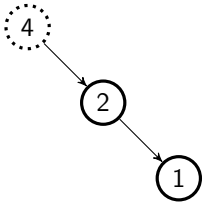
Sequence: 13, 10, 8, 8, 13, 1, 6, 2



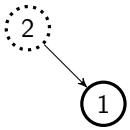
Sequence: 13, 10, 8, 8, 13, 1, 6, 2, 5



Sequence: 13, 10, 8, 8, 13, 1, 6, 2, 5, 4



Sequence: 13, 10, 8, 8, 13, 1, 6, 2, 5, 4, 2



End Sequence: 13, 10, 8, 8, 13, 1, 6, 2, 5, 4, 2, 1

Now, in order to completely characterize the bijection between Prüfer sequences and spanning trees, we have to argue why the sequence is unique to its tree. The inverse algorithm is given in the next section.

6.3. Prüfer Sequences to Rooted Trees

We begin with a Prüfer sequence $p_1, \ldots, p_{n-1}$, a sequence of length $n-1$ whose values are $v \in \{1, \ldots, n\}$ and whose last value $p_{n-1} = 1$. Our goal is to define $\pi(v) \in V$ for all $v \in V - \{1\}$. We will do this in $n-1$ steps.

Let v be the first element of the Prüfer sequence. Call $w \in V$ available if it is not a root and it does not appear in the Prüfer sequence. Let w be the largest available vertex. (The Prüfer sequence has at most $n-2$ nonroots so some w will be available.) Set $\pi(w) = v$. We delete w from the list of available vertices and delete the first element from the Prüfer sequence and iterate.

More precisely, on the t-th step, a $w \in V$ is available if

(1) w is not a root,

(2) we have not previously set $\pi(w)$,

(3) w does not appear in the remaining Prüfer sequence $p_t, \ldots, p_{n-1}$.

Precisely $t-1$ values $\pi(w)$ have been set, there is 1, the root, as the last term, and the remaining Prüfer sequence (remember that the last term is the root) has at most $n-1-t$ nonroots, so some w will be available. Let w be the maximal available vertex. Set $\pi(w) = p_t$.

Once $\pi(w)$ is set, it is no longer available. Thus $\pi(w)$ will be set for $n-1$ distinct values and therefore for every nonroot w.

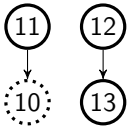
We further claim that the π created by this algorithm has no cycles. On the t-th step p_t is not available, so we never set $\pi(w) = w$. Suppose there were a cycle of length $s+1 \geq 1$, a sequence $v_0, \ldots, v_s$ with $\pi(v_i) = v_{i+1}$ for $0 \leq i < s$ and $\pi(v_s) = v_0$. In the creation of π there would be a *first* $w \in \{v_0, \ldots, v_s\}$ for which $\pi(w)$ was set. Say $w = v_i$. At that point $\pi(v_{i-1})$ has not yet been set. (When $w = v_0$, $\pi(v_s)$ has not yet been set.) But once $\pi(w)$ is set, w is no longer available. Thus we cannot set $\pi(v_{i-1}) = w$ at some later step.

In summary, given a Prüfer sequence $p_1, \ldots, p_{n-1}$, this algorithm creates a rooted tree by creating the parent function π.

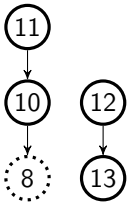
Below we recreate this algorithm

Prüfer sequence: 13, 10, 8, 8, 13, 1, 6, 2, 5, 4, 2, 1
List of available labels: 12, 11, 9, 7, 3
$\pi(12) = 13$

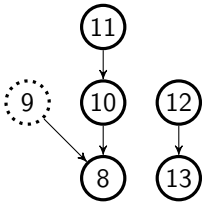
Prüfer sequence: 10, 8, 8, 13, 1, 6, 2, 5, 4, 2, 1
List of available labels: 11, 9, 7, 3
$\pi(11) = 10$



Prüfer sequence: 8, 8, 13, 1, 6, 2, 5, 4, 2, 1
List of available labels: 10, 9, 7, 3
$\pi(10) = 8$

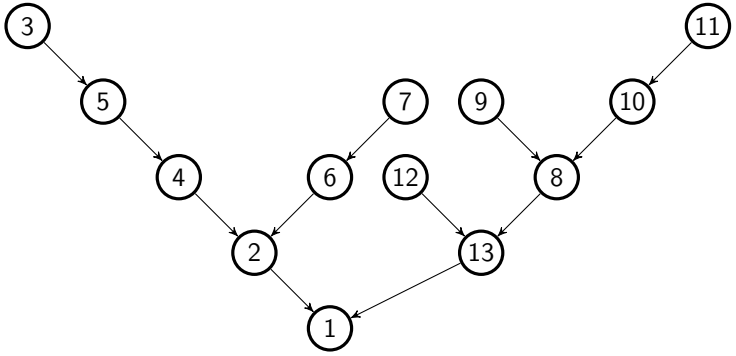


Prüfer sequence: 8, 13, 1, 6, 2, 5, 4, 2, 1
List of available labels: 9, 7, 3
$\pi(9) = 8$



Prüfer sequence: 13, 1, 6, 2, 5, 4, 2, 1
List of available labels: 3
$\pi(3) = 13$

And the process continues in this fashion until we arrive at



6.4. Proof of Bijection

We have one algorithm $R \to P$ that sends rooted trees to Prüfer sequences and another $P \to R$ that sends Prüfer sequences to rooted trees. Now we further claim that these algorithms give a *bijection* between the Prüfer sequences and the rooted trees. We outline the argument.

Suppose $R \to P$ sends a rooted tree T to a Prüfer sequence $p_1, \ldots, p_{n-1}$. Suppose that in the first step, w is the maximal leaf so that $p_1 = \pi(w)$. Let v be a leaf of T. Then v cannot appear in the Prüfer sequence as only parents are placed in the Prüfer sequence. Inversely, suppose v is not a leaf. In T this v has some child w with $\pi(w) = v$. During the algorithm, all nonleaves are eliminated so at some step w is eliminated and at that step $v = \pi(w)$ is added to the Prüfer sequence. That is, the leaves of T are precisely the nonroots v which do not appear in the Prüfer sequence. Thus in the first step of the algorithm $P \to R$, the maximal available vertex is actually the maximal leaf w of T and the algorithm sets $\pi(w) = p_1$. The inductive step is similar, and so algorithm $P \to R$ reconstructs the original tree T.

Similarly, suppose $P \to R$ sends Prüfer sequence $p_1, \ldots, p_{n-1}$ to a rooted tree T. In T the parents will be precisely $p_1, \ldots, p_{n-1}$, so the leaves will be precisely those nonroots which do not appear in the Prüfer sequence. Suppose that in the first step of $P \to R$, we set $\pi(w) = p_1$. This is done for that maximal available vertex w, but this is precisely the maximal leaf w of T. Thus in the first step of $R \to P$, we do set $\pi(w) = p_1$. The inductive step is similar, and so the algorithm $R \to P$ reconstructs the original Prüfer sequence.

There are n choices for each of the first $n-2$ values of the Prüfer sequence and one choice (the root) for the final value. Thus there are precisely n^{n-2} Prüfer sequences. As we have a bijection between Prüfer sequences and rooted trees, there are precisely n^{n-2} rooted trees, proving Theorem 6.2.

6.5. Rooted Forests

On an intuitive level a rooted forest is simply a collection of disjoint rooted trees. Formally, a rooted forest F consists of three parts:

(a) a finite set V of vertices,

(b) a designated nonempty set $R \subset V$. The $r \in R$ are called *roots*,

(c) a function π which sends vertices other than a root into vertices, which may or may not be roots. Formally, $\pi : V - R \to V$. When $\pi(v) = w$, we say that w is the parent of v and that v is a child (there may be others) of w.

For T to be a rooted forest, it must satisfy one of the following:

(a) Beginning at any nonroot v, if one continually applies π one eventually reaches the root r.

(b) There are no "cycles" in π. More explicitly, there are no distinct $v_0, \ldots, v_s$ with $\pi(v_i) = v_{i+1}$ for $0 \leq i < s$ and $\pi(v_s) = v_0$. This includes the case $s = 0$, there is no v_0 which is its own parent.

As before the two properties are equivalent. The definition of leaf does not change.

Definition 6.4. A vertex v of a rooted forest is a *leaf* if v is not a root *and* there is no vertex w with $\pi(w) = v$.

We will give a powerful extension of Theorem 6.2.

Theorem 6.5 (Cayley's Formula for Rooted Forests)**.** *Let $R \subset V$ with $|R| = r$, $|V| = n$. The number of labeled rooted forests on V with designated roots R is precisely rn^{n-r-1}.*

Our argument will extend the notion of Prüfer sequences. Let a rooted forest F be given with vertices $V = \{1, \ldots, n\}$ and r roots.

6.6. Prüfer Sequences to Rooted Forests

We begin with a Prüfer sequence $p_1, \ldots, p_{n-r}$, a sequence of length $n - r$ whose values are $v \in \{1, \ldots, n\}$ and whose last value $p_{n-r} \in R$.

(Note that V and R are given in advance.) Our goal is to define $\pi(v) \in V$ for all $v \in V - \{1\}$. We will do this in $n - r$ steps.

Let v be the first element of the Prüfer sequence. Call $w \in V$ available if it is not a root and it does not appear in the Prüfer sequence. Let w be the largest available vertex. (The Prüfer sequence has at most $n - r - 1$ nonroots so some w will be available.) Set $\pi(w) = v$. We delete w from the list of available vertices, delete the first element from the Prüfer sequence, and iterate.

More precisely, on the t-th step, a $w \in V$ is available if

(1) w is not a root,

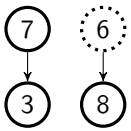
(2) we have not previously set $\pi(w)$,

(3) w does not appear in the remaining Prüfer sequence $p_t, \ldots, p_{n-1}$.

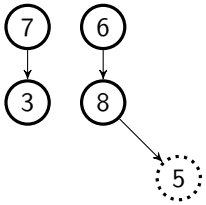
We select the largest available w and set $\pi(w) = p_t$. We omit the argument that π will be a rooted forest, which is almost identical to that of rooted trees. We illustrate the algorithm with $V = \{1, \ldots, 8\}$, $R = \{1, 2\}$, and the Prüfer sequence 385521.

Sequence: 3, 8, 5, 5, 2, 1
List of available labels: 7, 6, 4
$\pi(7) = 3$

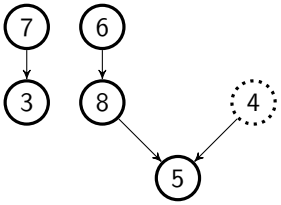
Sequence: 8, 5, 5, 2, 1
List of available labels: 6, 4, 3
$\pi(6) = 8$



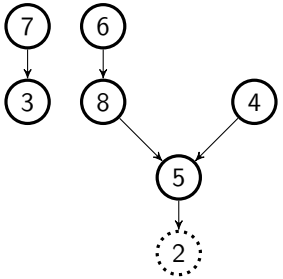
Sequence: 5, 5, 2, 1
List of available labels: 3, 4, 8
$\pi(8) = 5$



Sequence: 5, 2, 1
List of available labels: 3, 4
$\pi(4) = 5$

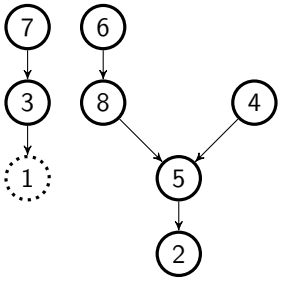


Sequence: 2, 1
List of available labels: 3, 5
$\pi(5) = 2$



Sequence: 1
List of available labels: 3
$\pi(3) = 1$

End forest:

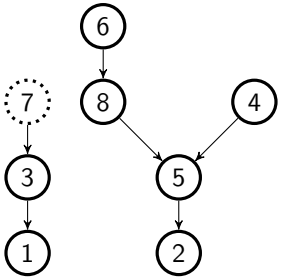


6.7. . . . and Back Again

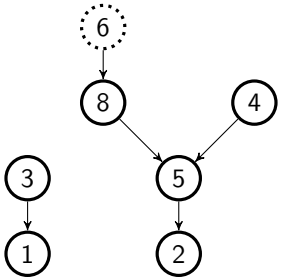
We begin with a rooted forest with $V = \{1, \ldots, n\}$ and R given.

The Prüfer sequence is found sequentially. For the first step, find that leaf w with maximal value. Then $\pi(w)$ (*not w itself!*) becomes the first element of the Prüfer sequence. Now delete w from the forest. (As w was a leaf, this will not cause trouble with other $\pi(v)$.) Now iterate. At each step we find the leaf w on the remaining tree with maximal value, add $\pi(w)$ to the Prüfer sequence, and delete w from the tree. We end after $n-r$ steps. After $n-r-1$ steps only the root R and some other vertex v remain in the tree. On the final step, that v will be the only leaf and hence the leaf with maximal value. Thus $\pi(v)$, the final element, will be a root.

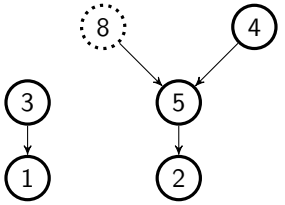
We illustrate this algorithm with the forest given in the previous section.



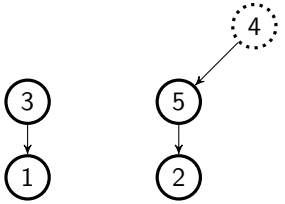
Prüfer sequence: 3



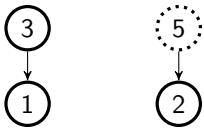
Prüfer sequence: 3, 8



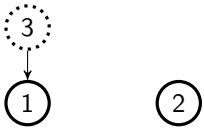
Prüfer sequence: 3, 8, 5



Prüfer sequence: 3, 8, 5, 5



Prüfer sequence: 3, 8, 5, 5, 2



End Prüfer sequence: 3, 8, 5, 5, 2, 1

Note that we returned to the Prüfer sequence we started with. Indeed, the two algorithms give a bijection between Prüfer sequences and rooted forests. Again we omit the details, which mirror the arguments for rooted trees. There are n choices for each of the first $n-r-1$ elements of the Prüfer sequence and r choices for the final element. Thus there are rn^{n-r-1} Prüfer sequences. The bijection thus shows Theorem 6.5.

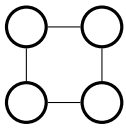
6.8. An Exact Formula for Unicyclic Graphs

Here we will find the exact formula

$$UN(n) = \sum_{k=3}^{n} \frac{(n)_k}{2k} n^{n-k-1} k \tag{6.1}$$

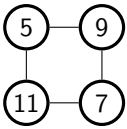
for the number of labeled unicyclic graphs on $V = \{1, \ldots, n\}$. We are constructing a unicyclic graph in three steps.

First, we choose the length k of the unique cycle in the graph and we draw it:

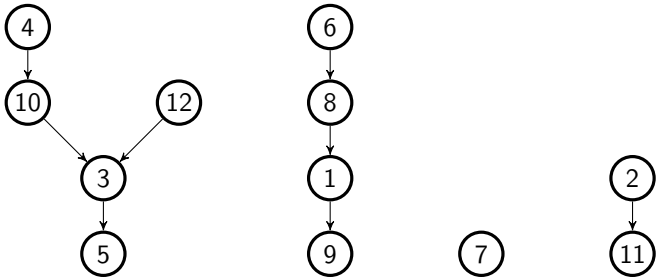


The value k can be anywhere in the range $3 \le k \le n$. Thus we first write $\sum_{k=3}^{n}$.

Second, we label the vertices of the k-cycle with distinct vertices. For the first vertex we have n choices, for the second $n-1$, and so on, up until the last step when we have $n-k+1$ choices giving a total of $(n)_k$ choices (recall the notation in (5.2)). However, there are $2k$ such sequences that give the same labeled cycle, as the cycle may start anywhere and may go in either way. (In algebra nomenclature, the automorphism group of the k-cycle has size $2k$.) Thus there is an overcount by a factor of $2k$. This gives the factor $(n)_k/2k$.



Third, create a rooted forest on the V, with R consisting of the vertices in the k-cycle, as shown in the example below:



This gives the factor kn^{n-k-1} by Theorem 6.5.

We now attach the trees in the rooted forest to their corresponding locations in the cycle and erase the arrows, giving the unicyclic graph shown in Figure 1.

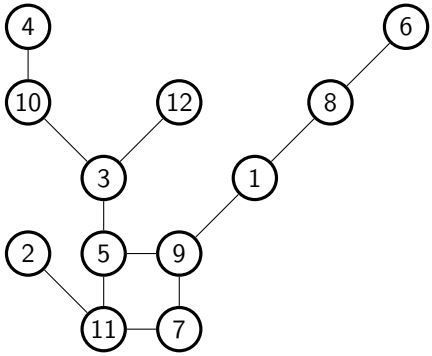


Figure 1. Unicyclic graph on twelve vertices

6.9. Counting Unicyclic Graphs in Asymptopia

While formula (6.1) is precise, it is not very satisfactory in Asymptopia. The terms in the sum are each reasonably nice but it is not clear *a priori* which terms are the important ones. We begin by taking a common term out of the sum so that

$$UN(n) = \frac{n^{n-1}}{2} \sum_{k=3}^{n} \frac{(n)_k}{n^k}. \tag{6.2}$$

This addend is now precisely the ratio $A(n,k)$ studied in Chapter 5 and formulae (5.15) and (5.16) gives that $\sum_{k=1}^{n} A(n,k) \sim \sqrt{n}\sqrt{\pi/2}$. In (6.2) the sum only goes from $k = 3$ to n. However, the values at $k = 1, 2$ are both $1 + o(1)$ and so are $o(\sqrt{n})$. Thus

$$\sum_{k=3}^{n} \frac{(n)_k}{n^k} \sim \sqrt{n}\sqrt{\pi/2}. \tag{6.3}$$

Plugging back into (6.2) gives an interesting and informative result:

Theorem 6.6.

$$UN(n) \sim \sqrt{\frac{\pi}{8}} n^{n-\frac{1}{2}}. \tag{6.4}$$

Comment. The ratio of $UN(n)$ to the number n^{n-2} of trees is $cn^{3/2}$ with $c = \sqrt{\pi/8}$. From a tree we can create a unicyclic graph by adding any edge. This can be done in $\binom{n}{2} - (n-1) \sim n^2/2$ ways. There are $\sim n^n/2$ pairs of a tree T and an edge e which together give a unicyclic graph G. However, there is serious overcounting. Let G be a unicyclic graph with cycle of length k. For any edge e in the cycle, $G-e = T$ is a tree and G can be expressed as $T+e$. Thus such a G has been counted k times. But the cycle lengths k vary so this does not tell us what the ratio is. One would hardly guess *a priori* that the ratio would be $\Theta(n^{3/2})$.

Chapter 7

Ramsey Numbers

> Working with Paul Erdős was like taking a walk in the hills. Every time when I thought that we had achieved our goal and deserved a rest, Paul pointed to the top of another hill and off we would go.
> – Fan Chung

The **Ramsey number** $R(k, l)$ is the smallest integer n such that in any two-coloring of the edges of a complete graph on n vertices K_n by red and blue, either there is a complete subgraph on k vertices, all of whose edges are colored red, K_k, or there exists a blue K_l. The existence of such an n is a consequence of Ramsey's theorem, a classic and important result but one that will not concern us here. Rather, our results concern lower bounds to $R(k, l)$. Unraveling the definition, $R(k, l) > n$ means that there *exists* a two-coloring with neither complete subgraph K_k nor K_l a monochromatic graph.

We will first focus on $R(k, k)$.

7.1. Initial Erdős Argument

In 1947 Paul Erdős in [**Erd47**] gave a new method to prove the existence of an object (here, a coloring) with some desired properties. He considered an appropriately defined random object and showed that

the probability of the random object having the desired properties was positive. From that he concluded that an object with those properties *must* exist. This method is called Erdős Magic (more formally, The Probabilistic Method) as it does not actually present the desired object. His 1947 paper [**Erd47**] (only three pages long!) dealt with the Ramsey number $R(k,k)$. Recall $R(k,k) > n$ means that there exists a coloring of K_n with no monochromatic K_k.

Theorem 7.1. *If*

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 1, \tag{7.1}$$

then $R(k,k) > n$.

Proof. Color edges of K_n either red or blue uniformly at random. Now consider a set M on k vertices of K_n and the probability that all its edges have the same color, $\Pr(A_M)$. Since there are two choices for it to be monochromatic and $2^{\binom{k}{2}}$ total choices (two choices for each edge and $\binom{k}{2}$ of them), then $\Pr(A_M) = 2^{1-\binom{k}{2}}$. There are $\binom{n}{k}$ possible choices for M. The probability of the disjunction of the A_M is at most the sum of the $\Pr(A_M)$, which is $\binom{n}{k}2^{1-\binom{k}{2}}$. By hypothesis this is less than one. The complement thus has positive probability. The complement is that the coloring has precisely the desired property, that there is no monochromatic K_k. By Erdős Magic, such a coloring must exist. □

We defer the asymptotic consequences of Theorem 7.1 to §7.4.

7.2. Deletion

In a refinement of Erdős Magic we consider a random coloring and then delete the blemishes.

Theorem 7.2. *For any integer* m

$$R(k,k) > m - \binom{m}{k} 2^{1-\binom{k}{2}}. \tag{7.2}$$

Proof. Color again the edges of K_m with two colors uniformly at random. For a set M of k vertices, let X_M be the indicator function for M

being monochromatic. That is, $X_M = 1$ if M is monochromatic, else $X_M = 0$. Let $X = \sum X_M$, the sum over all sets M of k vertices. Thus X is the number of monochromatic sets M. Then $E[X_M] = 2^{1-\binom{k}{2}}$, the probability that M is monochromatic. The expectation of a sum is the sum of the expectations so $E[X] = \binom{m}{k}2^{1-\binom{k}{2}}$. For a discrete random variable X, there is always a positive probability (another form of Erdős Magic) that $X \le E[X]$. Thus there is a coloring of K_m with at most $\binom{m}{k}2^{1-\binom{k}{2}}$ monochromatic k-sets. Fix such a coloring. For each monochromatic k-set remove one vertex. (A vertex may be removed several times; this only helps.) The number of remaining vertices is at least $m-\binom{m}{k}2^{1-\binom{k}{2}}$. On the remaining set, the blemishes having been removed, there are no monochromatic k-sets. □

We defer the asymptotic consequences of Theorem 7.2 to §7.4.

7.3. Lovász Local Lemma

The Lovász local lemma in [**EL**] is a powerful probability result. Let A_α, $\alpha \in I$, be events. Let G be a graph on the index set I. We say that a graph G is a dependency graph for the events if, for each $\alpha \in I$, the event A_α is mutually independent of the events A_β, β not adjacent to α.

Theorem 7.3 (Lovász Local Lemma)**.** *Let A_α, $\alpha \in I$ be events with dependency graph G. Suppose all $P(A_\alpha) \le p$. Suppose further that each $\alpha \in I$ is adjacent to at most d other $\beta \in I$ in the dependency graph. Suppose further that $4dp \le 1$. Then the conjunction of the complements of the A_α is not empty, that is,* $\Pr[\bigwedge \overline{A_\alpha}] > 0$.

We do not give the proof of the Lovász local lemma in this work, as it may be found in many works. Our own book, *The Probabilistic Method* (with Noga Alon [**AS08**]) gives proofs of Theorems 7.1, 7.2, 7.3, 7.4, 7.5, 7.6—and much more! We apply it to give a bound on $R(k,k)$.

Theorem 7.4. *If*

$$4\left(\binom{k}{2}\binom{n}{k-2}\right)2^{1-\binom{k}{2}} < 1, \tag{7.3}$$

then $R(k,k) > n$.

Proof. As usual, consider a uniformly random two-coloring of K_n. As usual, for each set M of k vertices, let A_M be the event that the complete graph on M is monochromatic. As above, $\Pr(A_M) = 2^{1-\binom{k}{2}}$. We create a dependency graph by letting M, M' be adjacent if they overlap in at least two vertices. When M, M' are not adjacent, they have no common pairs so the events $A_M, A_{M'}$ are independent. Moreover, A_M will be mutually independent of all $A_{M'}$ for which M, M' are not adjacent. Each A_M is then adjacent to d' other $A_{M'}$, where d' is the number of M' overlapping M in at least two vertices. The precise calculation of d' is awkward and turns out not to have asymptotic significance. Rather, we find all such M' by selecting two vertices from M and then $k-2$ other vertices. This will be a multiple counting when M', M overlap in more than two vertices. Still, it gives the upper bound $d' \leq d$ with $d = \binom{k}{2}\binom{n}{k-2}$. From Theorem 7.3 the conjunction of the complements of the A_M is not empty, which again means that there *exists* a coloring of K_n with no monochromatic K_k. □

We defer the asymptotic consequences of Theorem 7.4 to §7.4.

7.4. Computations for $R(k, k)$

For Theorems 7.1, 7.2, and 7.4 we want to deduce asymptotic lower bounds for $R(k, k)$.

First, consider Theorem 7.1. Let n_0 denote the maximal n satisfying (7.1). From Chapter 5 we know that the approximation of $\binom{n}{k}$ splits into cases, depending on the relationship between n and k. When $k = o(\sqrt{n})$ or, equivalently, $n \gg k^2$, (5.7) applies. As this is not an atypical problem, we first give a quite coarse lower bound on $A(k)$. Bounding $\binom{n}{k} < n^k$, we see that if $n^k 2^{1-\binom{k}{2}} \leq 1$, then n satisfies (7.1). Taking k-th roots, we may take $n = (1+o(1))2^{(k-1)/2}$. We have an exponential lower bound on n and so certainly $n \gg k^2$. Applying (5.7), we want n with

$$n^k = (1 + o(1))k!2^{\binom{k}{2}-1}. \tag{7.4}$$

Taking k-th roots of (7.4) and noting that from Stirling's formula (1.1) $k!^{1/k} \sim k/e$,

$$n_0 = (1+o(1))\frac{k}{e}2^{(k-1)/2}. \tag{7.5}$$

(*Caution*: We could not estimate $\binom{k}{2}$ by $k^2/2$ since it was in the exponent.) Hence Erdős's 1947 paper [**Erd47**] yields the lower bound

$$R(k,k) \geq (1+o(1))\frac{k}{e\sqrt{2}}\sqrt{2}^k. \tag{7.6}$$

Now consider Theorem 7.2 and, following (7.2), set

$$f(m) = m - \binom{m}{k}2^{1-\binom{k}{2}}. \tag{7.7}$$

We parametrize $m = yn_0$, with $y \geq 1$, and n_0 as defined above.

- Solutions to basic problems often lead to good parametrizations for more elaborate problems.

Rather than recalculating from scratch, we compare $f(m)$ with $f(n_0)$. As a function of m the negative term in (7.7) has asymptotically an m^k factor. At $y = 1$ it is one, so at $y \geq 1$ it will be asymptotically y^k. Thus

$$f(yn_0) \sim yn_0 - y^k. \tag{7.8}$$

Now, rather than the more precise (7.5), let us simply use that

$$n_0 = (\sqrt{2}+o(1))^k. \tag{7.9}$$

Consider any fixed positive ϵ. With $y = \sqrt{2}-\epsilon$, $y^k = o(n_0)$, and so $f(y) \sim y$. If $y = \sqrt{2}+\epsilon$, then $y^k \gg yn_0$ so $f(y)$ is negative. We optimize at $y = \sqrt{2}+o(1)$. The negative term in $f(m)$ (selecting y slightly smaller than $\sqrt{2}$) is negligible, and $F(yn_0) \sim yn_0$. Thus Theorem 7.2 improves Theorem 7.1 by a factor of $\sqrt{2}$:

$$R(k,k) \geq (1+o(1))\frac{k}{e}\sqrt{2}^k. \tag{7.10}$$

Now consider Theorem 7.4. With n_0 as above, we now parametrize $n = zn_0$. Following (7.3), we set

$$g(n) = 4\binom{k}{2}\binom{n}{k-2}2^{1-\binom{k}{2}}. \tag{7.11}$$

Comparing (7.1) and (7.3),

$$4\binom{k}{2}\binom{n}{k-2} \sim \frac{2k^4}{n^2}\binom{n}{k}. \tag{7.12}$$

Thus at $z = 1$, $g(n) = g(n_0) \sim 2k^4 n_0^{-2}$. From the rough (7.9), we write $g(n_0) = (2 + o(1))^{-k}$. As $\binom{n}{k} \sim n^k/k!$, replacing n_0 by $n_0 z$ gives an extra z^{k-2} factor. Thus $g(n_0 z) \sim g(n_0) z^{k-2}$. We may take $z = 2 - \epsilon$ for any fixed ϵ and $g(n_0 z) \ll 1$. Theorem 7.2 therefore gives an improved Theorem 7.1 by a factor of 2:

$$R(k,k) \geq (1 + o(1))\frac{k\sqrt{2}}{e}\sqrt{2}^k. \tag{7.13}$$

Bound (7.13) is the best known asymptotic lower bound on the Ramsey number $R(k,k)$. The best known upper bound (which we do not deal with in this work) is of the form $(4 + o(1))^k$. The gap between the upper and lower bounds on this most basic of all Ramsey functions remains a vexing open problem.

7.5. Asymmetrical Ramsey Numbers

A powerful extension of Erdős's Theorem 7.1 is given by examining a random coloring, but one in which the probability of red is given by an appropriate p. As we often do not know which p to choose, we make p a parameter, later to be optimized.

Theorem 7.5. *If there exists p, $0 \leq p \leq 1$ such that*

$$\binom{n}{k}p^{\binom{k}{2}} + \binom{n}{l}(1-p)^{\binom{l}{2}} < 1, \tag{7.14}$$

then $R(k,l) > n$.

Proof. Color edges of K_n either red or blue independently, giving each edge the color red with proability p. Now consider a set M on k vertices of K_n and the probability that all its edges are red, $\Pr(A_M)$. Further, consider a set N on l vertices of K_n and the probability that all its edges are blue, $\Pr(B_N)$. Then $\Pr(A_M) = p^{\binom{k}{2}}$ and $\Pr(B_N) = (1-p)^{\binom{l}{2}}$. There are $\binom{n}{k}$ possible choices for M and $\binom{n}{l}$ possible choices for N. The probability of the disjunction of the A_M and B_N is at most the sum of the $\Pr(A_M)$, and the $\Pr(B_N)$ which is precisely

the left-hand side of (7.14). By hypothesis this is less than one. The complement thus has positive probability. The complement is that the coloring has precisely the desired property, that there is neither red K_k nor blue K_l. By Erdős Magic, such a coloring must exist. □

For each $p \in [0,1]$, Theorem 7.5 gives a lower bound for $R(k,l)$. We want to make this lower bound as large as possible. This leads to a problem in *asymptotic calculus.* Let $F(k,l)$ denote the maximal n for which there exists $p \in [0,1]$ with satisfying (7.14). The various asymptotics of $F(k,l)$ can be quite challenging. As a representative case, we restrict in this section to $l = 2k$, looking for a lower bound for $R(k,2k)$.

We simplify[1] the problem, replacing $\binom{n}{k}$ and $\binom{n}{2k}$ by n^k, n^{2k}, respectively, $\binom{k}{2}$ and $\binom{2k}{2}$ by $k^2/2$ and $2k^2$, respectively, and < 1 by ≤ 1. Now let $G(k)$ denote the maximal n (we will no longer restrict n to be integral) for which there exists $p \in [0,1]$ with

$$n^k p^{k^2/2} + n^{2k}(1-p)^{2k^2} \leq 1. \tag{7.15}$$

- In Asymptopia $A + B$ is often well approximated by $\max(A,B)$.

We replace (7.15) by the weaker system of inequalities:

$$n^k p^{k^2/2} \leq 1 \quad \text{and} \quad n^{2k}(1-p)^{2k^2} \leq 1. \tag{7.16}$$

The first inequality implies $n \leq (p^{-1/2})^k$ and the second that $n \leq ((1-p)^{-1})^k$. These bounds are increasing and decreasing, respectively, in p and so their minimum is maximized when they cross. This occurs when $p^{-1/2} = (1-p)^{-1}$, so $p = (3-\sqrt{5})/2$ and $n = \phi^k$ where, serendipitously, $\phi = (1+\sqrt{5})/2$ is the famous Golden Ratio. As (7.16) is a weakening of (7.15), this provides an upper bound $G(k) \leq \phi^k$.

- Upper bounds lead to lower bounds, and vice versa.

With $p = (3-\sqrt{5})/2$, the value $n_0 = \phi^k$ fails (7.15), but this is easily fixed. For any fixed positive ϵ set, $n = n_0(1-\epsilon)$. Now n^k has been lowered by a factor $(1-\epsilon)^k$ and n^{2k} by a factor of $(1-\epsilon)^{2k}$,

[1] Part of the *art* of Asymptopia is finding those simplifications that retain the essential elements of the problem.

both of which approach zero as $k \to \infty$. Thus $n^k p^{k^2/2} \le \frac{1}{2}$ and $n^{2k}(1-p)^{2k^2/2} \le \frac{1}{2}$, and (7.15) is now satisfied. Thus $G(k) \sim \phi^k$.

Returning to $R(k, 2k)$, we seek $F(k, 2k)$, the maximal n for which there exists p satisfying (with $l = 2k$) (7.14). The $k!$ and $(2k)!$ factors in (7.14) make the condition easier to meet than (7.15). Taking $n = \phi^n$ and $p = (3 - \sqrt{5})/2$, (7.14) is easily satisfied and so $R(k, 2k) \ge F(k, 2k) \ge \phi^k$. A more careful analysis, including the $k!$, $(2k)!$ factors, yields $F(k, k) = \Theta(k\phi^k)$, but this we leave to the reader!

7.6. Application to $R(3, l)$

We combine the deletion method of Theorem 7.2 with the parametrization method of Theorem 7.4.

Theorem 7.6. *For every positive integer m and $p \in [0, 1]$,*

$$R(k,l) \ge m - \binom{m}{k} p^{\binom{k}{2}} + \binom{m}{l}(1-p)^{\binom{l}{2}}. \tag{7.17}$$

Proof. Color the edges of K_m with red and blue, each edge red with probability p, blue with probability $1 - p$, the choice of edge colors being mutually independent. For each set M of k vertices, let X_M be the indicator function for M being red. Then $E[X_M] = p^{\binom{k}{2}}$ as $\binom{k}{2}$ edges need to be red. For each set N of l vertices, let Y_N be the indicator function for N being blue. Then $E[Y_N] = (1-p)^{\binom{l}{2}}$ as $\binom{l}{2}$ edges need to be red. Let $X = \sum X_M + \sum Y_N$, the sum over all sets M of k vertices and all sets N of l vertices. From linearity of expectation

$$E[X] = \binom{m}{k} p^{\binom{k}{2}} + \binom{m}{l}(1-p)^{\binom{l}{2}}. \tag{7.18}$$

For a discrete random variable X there is always a positive probability (another form of Erdős Magic) that $X \le E[X]$. Fix a coloring of K_m with $X \le E[X]$. For each red k-set and for each blue l-set, remove one vertex. (A vertex may be removed several times; this only helps.) The number of remaining vertices is at least $m - E[X]$. On the remaining set, the blemishes having been removed, there are no monochromatic k-sets. □

The general application of Theorem 7.6 can be quite challenging. Here we consider the special case $k=3$. Thus

$$R(3,l) \geq m - \binom{m}{3}p^3 - \binom{m}{l}(1-p)^{\binom{l}{2}}. \tag{7.19}$$

If either of the negative terms is m or more, then the right-hand side will be negative. On the other side, if both negative terms are $\frac{m}{4}$ or less, then (7.19) gives $R(3,l) \geq \frac{m}{2}$. Our sense is that these two sets of conditions are quite similar. We replace (7.19) with the two conditions

$$\binom{m}{3}p^3 \leq \frac{m}{4} \quad \text{and} \quad \binom{m}{l}(1-p)^{\binom{l}{2}} \leq \frac{m}{4}. \tag{7.20}$$

We bound $1-p \leq e^{-p}$ and $\binom{m}{l} \leq^l$. We replace the second condition by the weaker

$$\left(me^{-p(l-1)/2}\right)^l \leq \frac{m}{4}. \tag{7.21}$$

Because the quantity in parenthesis is being taken to a high power requiring that it is $\leq \frac{m}{4}$, the constraint is similar to requiring that it is ≤ 1. By setting $p = 2.01(\ln m)/l$, the exponent is bigger than $\ln m$ and so the quantity in parenthesis is ≤ 1, and thus (7.21) is satisfied. The first condition is basically $\frac{m^2p^3}{6} \leq \frac{m}{4}$, which is satisfied (letting c_1, c_2, c_3 be positive absolute constants below) for $p = c_1 m^{-2/3}$. We have set two conditions on p, and we now take m so that they are the same. That is, we set $2.01(\ln m)/l = c_1 m^{-2/3}$ so that $l = c_2 m^{2/3}(\ln m)$. Inverting the function (Theorem 2.15), $m = c_3 l^{3/2} \ln^{-3/2} l$. Thus

$$R(3,l) \geq \frac{m}{2} = \Omega(l^{3/2} \ln^{-3/2} l). \tag{7.22}$$

Remark. The quest for the asymptotics of $R(3,l)$ has a long history. In a classic 1935 paper [**ES35**], Paul Erdős and George Szekeres show that $R(3,l) \leq \binom{l+1}{2}$. In 1957 Erdős showed $R(3,l) \geq l^{1+\epsilon}$ for a fixed (small) ϵ. (At this early date Theorem 7.6 was not known.) Improvements to both the upper and the lower bounds followed over the decades. Finally, in 1995 in [**Kim95**] Jeong-Han Kim showed that $R(3,l) = \Theta(l^2/(\ln l)$. Even today the search continues to find a constant c so that $R(3,l) \sim cl^2/(\ln l)$.

Chapter 8

Large Deviations

> Any new possibility that existence acquires, even the least likely, transforms everything about existence.
> – Milan Kundera, *Slowness*

If we flip a fair coin a million times it is very unlikely that we will get more than six hundred thousand heads. But how unlikely? The study of large deviations concerns the estimation of such extremely small probabilities.

8.1. The Chernoff Bound

Let X be any random variable. Our goal is to give an upper estimate (often called a *tail bound*) on $\Pr[X \geq a]$ and, similarly, $\Pr[X \leq a]$. In Asymptopia our emphasis will be on those cases where these probabilities are very small.

Definition 8.1. The Laplace transform of X is the function $f(\lambda)$ defined by

$$f(\lambda) = E[e^{\lambda X}]. \tag{8.1}$$

In particular, if X is a discrete random variable taking on values $a_1, \ldots, a_s$ with probabilities $p_1, \ldots, p_s$, respectively, then

$$f(\lambda) = \sum_{i=1}^{s} p_i e^{\lambda a_i}. \tag{8.2}$$

We shall be particularly interested in cases in which X can be expressed as a sum $X = \sum_{i=1}^{n} X_i$ where the X_i are mutually independent random variables. In that case the variables $\exp[\lambda X_i]$ are mutually independent. The expectation of a product of mutually independent random variables is the product of their expectations, so the Laplace transform of X is the product of the Laplace transforms of the X_i. That is,

$$E[e^{\lambda X}] = \prod_{i=1}^{n} E[e^{\lambda X_i}]. \tag{8.3}$$

Often the X_i are quite simple, in which case the Laplace transform has a nice form.

The Laplace transform and its complex cousin the Fourier transform are centerpieces[1] of modern probability theory. In many cases full knowledge of the Laplace transform will give one full knowledge of the random variable X. For the Chernoff bounds we use only *one* value of the Laplace transform—but it needs to be a good one! For *any* positive λ, $e^{\lambda X}$ is a positive random variable. Therefore,

$$\Pr[X \geq a] = \Pr[e^{\lambda X} \geq e^{\lambda a}] \leq \frac{E[e^{\lambda X}]}{e^{\lambda a}}. \tag{8.4}$$

The Chernoff bound is (8.4). But the key to the Chernoff bound is the choice of λ. As each λ gives an upper bound, we want to choose λ so that the upper bound is minimal. Often times we do not get the absolute minimum and near minimality works well. For this bound to be effective, we must be able to evaluate or, more often, find a reasonable upper bound on $f(\lambda) = E[e^{\lambda X}]$.

A small alteration allows us to consider $\Pr[X \leq a]$:

$$\Pr[X \leq a] = \Pr[e^{-\lambda X} \geq e^{-\lambda a}] \leq \frac{E[e^{-\lambda X}]}{e^{-\lambda a}} \text{ for all } \lambda > 0. \tag{8.5}$$

[1] A knowledge of Laplace transforms is not needed in this work.

8.2. The Gaussian Tail

Let X be the standard Gaussian distribution, with probability density function $(2\pi)^{-1/2}e^{-x^2/2}$. The Laplace transform[2]

$$\begin{aligned} E[e^{\lambda X}] &= \int_{-\infty}^{+\infty} \sqrt{\frac{1}{2\pi}} e^{-x^2/2} e^{\lambda x} dx \\ &= e^{\lambda^2/2} \int_{-\infty}^{+\infty} \sqrt{\frac{1}{2\pi}} e^{-y^2/2} dy = e^{\lambda^2/2} \end{aligned} \tag{8.6}$$

so that (8.6) becomes

$$\Pr[X \geq a] \leq e^{\frac{\lambda^2}{2} - \lambda a}. \tag{8.7}$$

This is minimized when $\lambda = a$, giving

$$\Pr[X > a] \leq e^{-a^2/2}. \tag{8.8}$$

The precise asymptotics of $\Pr[X \geq a]$ were given by (3.8). For a large (and here we are interested in large deviations) the bound (8.8) is quite good.

When Y is a Gaussian with mean μ and variance σ^2, the Laplace transform is $E[e^{\lambda Y}] = e^{\lambda\mu} e^{\lambda^2\sigma^2/2}$. Now reparametrize and consider $\Pr[Y \geq \mu + a\sigma]$:

$$\Pr[Y \geq \mu + a\sigma] \leq e^{\lambda\mu} e^{\lambda^2\sigma^2/2} e^{-\lambda(\mu + a\sigma)} = e^{\lambda^2\sigma^2/2} e^{-a\lambda\sigma}. \tag{8.9}$$

This is minimized when $\lambda = \frac{a}{\sigma}$, giving

$$\Pr[Y \geq \mu + a\sigma] \leq e^{-a^2/2}. \tag{8.10}$$

8.3. The Gaussian Paradigm I

Suppose we are given a random variable X that we *think* behaves like a Gaussian with zero mean and standard deviation σ^2, and we want to bound $\Pr[X \geq a\sigma]$. We can hope that setting $\lambda = \frac{a}{\sigma}$ in the Chernoff bound (8.4) will yield an upper bound on $\Pr[X \geq a\sigma]$ which is close to $\exp[-a^2/2]$. Assuming finiteness of all terms,

$$E[e^{\lambda X}] = E[1 + \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} X^i] = 1 + \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} E[X^i]. \tag{8.11}$$

[2] by the substitution $y = x + \lambda$

This sequence begins $1+\frac{1}{2}a^2$. *If* there were no other terms, this would be bounded from above by $\exp[a^2/2]$. Then (8.4) *would* yield

$$\Pr[X \geq a\sigma] \leq e^{a^2/2}/e^{a^2} = e^{-a^2/2}. \tag{8.12}$$

Roughly, the study of tail distributions breaks into three regimes:

Small Deviations. In this regime a is a constant. If the Central Limit Theorem applies, then $\Pr[X \geq a\sigma]$ is approximately $\Pr[N \geq a]$, where N is the standard normal distribution. While the methods for large deviations below also work for small deviations and do give an upper bound, they do not get this fine a result. The Central Limit Theorem is at the core of traditional probability courses, but in Asymptopia[3] the small deviation cases rarely arise (see §8.5).

Large Deviations (Our main emphasis). In this regime $a \to \infty$, but $\frac{a}{\sigma} \to 0$. In many cases, as detailed in §8.7, the Chernoff bounds give an upper bound on $P[X \geq \mu + a\sigma]$ that is close to $\exp[-a^2/2]$. Critically, $\lambda = \frac{a}{\sigma} = o(1)$. The terms in the sum (8.11) for $i \geq 3$ have extra powers of λ, and we can often use that to show that they are negligible.

Very Large Deviations. In this regime $a = \Omega(\sigma)$. Now λ is a constant and the later terms of (8.11) are no longer negligible. One can often still use Chernoff bounds, but the calculation of optimal λ can be challenging.

The Poisson distribution provides a good example as its Laplace transform has a relatively simple closed form. Let P_n denote[4] the Poisson distribution with mean n, as given by (12.37). Then

$$E[e^{\lambda P_n}] = e^{-n} \sum_{i=0}^{\infty} \frac{n^i e^{\lambda i}}{i!} = e^{ne^{\lambda}-n}. \tag{8.13}$$

We switch to zero mean by setting $Y_n = P_n - n$ so that

$$E[e^{\lambda Y_n}] = e^{ne^{\lambda}-n-n\lambda}. \tag{8.14}$$

Set $\sigma_n = \sqrt{n}$ so that $\mathrm{Var}[Y_n] = n = \sigma_n^2$.

[3] At least, in this work!

[4] While P_n is defined for all *real* positive n, in this section we consider n integral.

We wish to bound $\Pr[Y_n \geq a_n\sigma_n]$. Applying the Chernoff bound (8.4) with $\lambda_n = a_n/\sigma_n$ gives

$$\Pr[Y_n \geq a_n\sigma_n] \leq e^{-n\lambda_n - n + ne^{\lambda_n}} e^{-a_n^2}. \tag{8.15}$$

Now assume $a_n = o(\sigma_n)$. Then $\lambda_n \to 0$. Therefore

$$ne^{\lambda_n} = n[1 + \lambda_n + \frac{1}{2}\lambda_n^2 + o(\lambda_n^2)] = n + n\lambda_n + (1 + o(1))\frac{a_n^2}{2}, \tag{8.16}$$

so that

$$\Pr[Y_n \geq a_n\sigma_n] \leq e^{(1+o(1))a_n^2/2} e^{-a_n^2} = e^{-(1+o(1))a_n^2/2}. \tag{8.17}$$

This bound applies in both the large deviation and small deviation regimes. However, in the small deviation regime one gets a better bound by application of the Central Limit Theorem. As the bound is dropping exponentially in the square of a_n, one can get very small tail bounds for relatively moderate a_n. Take, for example, $a_n = 10\sqrt{n}$. Now

$$\Pr[Y_n \geq \sqrt{10n \ln n}] \leq e^{-(1+o(1))50 \ln n} = n^{-50(1+o(1))}. \tag{8.18}$$

Thus, for example, for n sufficiently large, returning to the original P_n,

$$\Pr[P_n \geq n + \sqrt{10n \ln n}] \leq n^{-49}. \tag{8.19}$$

8.4. Heads Minus Tails

Let X_i, $1 \leq i \leq n$, be mutually independent with $\Pr[X_i = +1] = \Pr[X_i = -1] = \frac{1}{2}$, and set $S_n = \sum_{i=1}^n X_i$. We may think of S_n as the number of heads minus the number of tails after n flips of a fair coin. Alternately, we can imagine a particle, initially at the origin, taking a random walk on Z for time n. Then S_n represents the particle's location at time n. As S_n has mean $\mu = 0$ and variance $\sigma^2 = n$, we shall parametrize as in §8.3 and bound $\Pr[S_n \geq a\sqrt{n}]$.

For each $1 \leq i \leq n$, the variable $e^{\lambda X_i}$ is e^{λ} or $e^{-\lambda}$, each with probability one-half. From the mutual independence, (8.3) gives

$$E[e^{\lambda S_n}] = \prod_{i=1}^n E[\lambda X_i] = \cosh^n \lambda \tag{8.20}$$

with hyperbolic cosine

$$\cosh(\lambda) := \frac{e^{\lambda} + e^{-\lambda}}{2}, \tag{8.21}$$

and thus

$$\Pr[S_n \geq a\sqrt{n}] \leq \cosh^n(\lambda) e^{-a\lambda\sqrt{n}}. \tag{8.22}$$

When λ is small (not always the case, see below) we will use the upper bound

$$\cosh(\lambda) \leq e^{\lambda^2/2}. \tag{8.23}$$

The Taylor series for both sides of (8.23) begin with $1 + \frac{1}{2}\lambda^2$. There are no odd powers of λ on either side, and the coefficient of λ^{2t}, $t \geq 2$, is always smaller for $\cosh(\lambda)$. This shows the inequality, and it also indicates that if λ is small the two sides are pretty close. Using (8.23), (8.22) becomes

$$\Pr[S_n \geq a\sqrt{n}] \leq e^{n\lambda^2/2} e^{-a\lambda\sqrt{n}}. \tag{8.24}$$

Setting $\lambda = \frac{a}{\sqrt{n}}$ (as indicated in §8.3) gives

Theorem 8.2.

$$\Pr[S_n \geq a\sqrt{n}] \leq e^{-a^2/2}. \tag{8.25}$$

As S_n is symmetric,

$$\Pr[|S_n| \geq a\sqrt{n}] \leq 2e^{-a^2/2}. \tag{8.26}$$

The binomial distribution $BIN[n, \frac{1}{2}]$, counting the number of heads in n flips of a fair coin, is one of the most basic probability distributions. Comparing to S_n the count of $+1, -1$ is replaced by a count of $1, 0$. Equivalently, one divides by two and then adds one-half. Therefore,

$$BIN[n, \frac{1}{2}] = \frac{n}{2} + \frac{S_n}{2}. \tag{8.27}$$

Bound (8.26) then becomes the useful

$$\Pr[|BIN[n, \frac{1}{2}] - \frac{n}{2}| \geq a\sqrt{n}] \leq 2e^{-2a^2}. \tag{8.28}$$

While the binomial is arguably more basic than S_n, it is usually mathematically more convenient to work with a distribution that has zero mean.

While Theorem 8.2 is always correct, its accuracy wanes as λ becomes bigger. Here $\lambda = an^{-1/2}$. Thus, even for, say, $\Pr[S_n > n^{0.9}]$, we will have λ small. Now consider the "very large deviation" problem of bounding $\Pr[S_n \geq bn]$, where $b \in (0,1)$ is fixed. Bound (8.22) becomes

$$\Pr[S_n \geq bn] \leq (\cosh(\lambda)e^{-b\lambda})^n. \tag{8.29}$$

The n in the exponent does not affect the optimal λ, and we are left with the calculus problem of finding λ to minimize $\cosh(\lambda)e^{-b\lambda}$. The solution[5] is

$$\lambda = \tanh^{-1}(b) = \frac{1}{2}[\ln(1+b) - \ln(1-b)], \tag{8.30}$$

and $(\cosh(\lambda)e^{-b\lambda})^n$ matches (after some algebraic manipulation) the correct asymptotics as given by (5.33) in Chapter 5.

8.5. ... and the Central Limit Theorem

The Central Limit Theorem tells us that $n^{-1/2}S_n$ approaches the standard limit. That is, for any real a,

$$\lim_{n\to\infty} \Pr[S_n \geq a\sqrt{n}] = \int_a^\infty \sqrt{\frac{1}{2\pi}} e^{-x^2/2} dx. \tag{8.31}$$

Bounds (8.31) and (8.25) have different strengths. For a fixed and $n \to \infty$, the Central Limit Theorem (8.31) gives a stronger result. But when a goes to infinity as a function of n, we cannot use the Central Limit Theorem at all. The real strength of (8.25) is that it holds for *all* n, a, not just asymptotically. Suppose, as does occur, we want to know when the tail $\Pr[S_n > a\sqrt{n}]$ becomes smaller than n^{-10}. Solving $\exp[-a^2/2] = n^{-10}$, we set $a = \sqrt{20 \ln n}$, and now it holds.

8.6. The Binomial Distribution

The binomial distribution $BIN[n,p]$ is the number of heads thrown in n tosses of a coin, where the probability of heads is p for each toss.

[5] Hyperbolic functions seem to have gone out of fashion. When was the last time, if at all, that you saw an inverse hyperbolic tangent?

We have an exact formula

$$\Pr[BIN[n,p]=i]=\binom{n}{i}p^i(1-p)^{n-i}. \tag{8.32}$$

In Asymptopia this is *too* precise. We feel that $BIN[n,p]$ is usually near pn, and we want to bound the probability that it is some distance from pn. In this section we let $p\in(0,1)$ be arbitrary but fixed and consider the asymptotics as $n\to\infty$.

It is convenient to shift to a zero mean. Basically, we count heads but we subtract p for each coin flip. Technically, define X_i, $1\le i\le n$ by $\Pr[X_i=1-p]=p$ and $\Pr[X_i=-p]=1-p$. Set $X=\sum_{i=1}^n X_i$, where the X_i are assumed mutually independent. Then $X=BIN[n,p]-np$. Set $\mu=E[X]$ and $\sigma^2=\mathrm{Var}[X]$ so that $\mu=0$ and $\sigma=(np(1-p))^{1/2}$. This includes the case $p=\frac12$, with $X=\frac12 S_n$, which was covered in §8.4. Our object now is to bound $\Pr[X\ge a\sigma]$ and $\Pr[X\le a\sigma]$ by something like $\exp[-a^2/2]$. Following the Gaussian paradigm we will apply the Chernoff bounds (8.4) and (8.5) with $\lambda=a/\sigma$. We sometimes succeed.

We let $f(\lambda)$ denote the Laplace transform of X_i so that we have the precise

$$f(\lambda):=E[e^{\lambda X_i}]=pe^{(1-p)\lambda}+(1-p)e^{-p\lambda}. \tag{8.33}$$

In the Taylor series for $f(\lambda)$, the coefficient of $\lambda^u/u!$ is $E[X_i^u]$. X_i has mean zero and variance $p(1-p)$ so that

$$f(\lambda):=1+p(1-p)\frac{\lambda^2}{2}+\cdots. \tag{8.34}$$

We would like to bound $f(\lambda)$ from above by $\exp[p(1-p)\lambda^2/2]$. This does not work exactly (though it does for $p=\frac12$) as the coefficients of λ^3 and higher powers are not in the right order. However, when $\lambda=o(1)$ we can handle this in Asymptopia. Now $f(\lambda)=1+(1+o(1))p(1-p)\frac{\lambda^2}{2}$, and therefore

$$f(\lambda)\le\exp[(1+o(1))p(1-p)\frac{\lambda^2}{2}] \tag{8.35}$$

and

$$E[e^{\lambda X}] = E[e^{\lambda X_i}]^n \le \exp[(1+o(1))np(1-p)\frac{\lambda^2}{2}] \\ = \exp[(1+o(1))\sigma^2\lambda^2/2]. \tag{8.36}$$

Then, (8.4) gives, with $\lambda = a/\sigma$,

$$\Pr[X \ge a\sigma] \le e^{(1+o(1))\sigma^2\lambda^2/2}e^{-a\lambda\sigma} \\ = e^{(1+o(1))\frac{a^2}{2}}e^{-a^2} = e^{-(1+o(1))a^2/2}. \tag{8.37}$$

The same argument applies on the other side, and

$$\Pr[X \le a\sigma] \le e^{-(1+o(1))a^2/2}. \tag{8.38}$$

As $\lambda = a/\sigma$ this works only when $a = o(\sigma) = o(\sqrt{n})$. In particular, (8.36) does not apply when $a\sigma = \Omega(n)$, but it does apply if, for example, $a\sigma = n^{0.99}$. Equations (8.36) and (8.38) often work in reverse when we have a desired upper bound on the tail probability and want to say that X lies close to its mean. Suppose, for example, we want $\Pr[X \ge a\sigma] \le n^{-10}$. Solving $\exp[-a^2/2] = n^{-10}$ gives $a = (20\ln n)^{1/2}$. Let $\epsilon > 0$ be arbitrary but fixed, and set $a = (1+\epsilon)(20\ln n)^{1/2}$. Then certainly $a = o(\sqrt{n})$ and so (8.36) gives $\Pr[X \ge a\sigma] \le n^{-10}$. Note that the power 10 is transformed into a constant in the calculation of a. Further, the particular value of p is only reflected in a constant factor in σ. Putting these together yields a useful result:

Theorem 8.3. *For all $p \in (0,1)$ and $c > 0$, there exists a K so that*

$$\Pr[|BIN[n,p] - np| \ge K\sqrt{n\ln n}] = o(n^{-c}). \tag{8.39}$$

Effectively, Theorem 8.3 tells us that the binomial distribution is quite concentrated about its mean, and moving out a relatively small distance from the mean, the tail probability becomes extremely small.

8.7. The Gaussian Paradigm II

Continuing §8.3, here we give some general conditions that will imply, roughly, a Gaussian tail. Suppose that for each n we are given a sum

$$Y_n = \sum_{i=1}^{n} X_{i,n}. \tag{8.40}$$

Critically, we assume that for each n the variables $X_{i,n}$ are mutually independent. Assume for technical convenience that all means $E[X_{i,n}] = 0$, so all $E[Y_n] = 0$. Set $\sigma_{i,n}^n = \text{Var}[X_{i,n}]$ and $\sigma_n^2 = \sum_{i=1}^n \sigma_{i,n}^2$ so that $\sigma_n^2 = \text{Var}[Y_n]$. We wish to bound $\Pr[Y_n \geq a_n\sigma_n]$. Assume that $\lim_{n\to\infty} a_n/\sigma_n = 0$. Set $\lambda_n = a_n/\sigma_n$ such that $\lim_{n\to\infty} \lambda_n = 0$.

Theorem 8.4. *With the above assumptions, consider further that*

$$E[e^{\lambda_n X_{i,n}}] \leq e^{(1+o(1))\lambda_n^2\sigma_{i,n}^2/2}. \tag{8.41}$$

Then

$$\Pr[Y_n \geq a_n\sigma_n] \leq e^{-(1+o(1))a_n^2/2}. \tag{8.42}$$

In (8.41) the $o(1)$ term must be uniform over all $1 \leq i \leq n$. Precisely, for all $\epsilon > 0$ there exists n_0 such that for $n \geq n_0$ and *all* $1 \leq i \leq n$,

$$E[e^{\lambda_n X_{i,n}}] \leq e^{(1+\epsilon)\lambda_n^2\sigma_{i,n}^2/2}. \tag{8.43}$$

Proof. Let $\epsilon > 0$ be arbitrary, let n_0 be such that (8.43) is satisfied, and let $n \geq n_0$. As the $X_{i,n}$ are mutually independent, the variables $\exp[\lambda_n X_{i,n}]$ are mutually independent. The expectation of a product of mutually independent random variables is the product of their expectations. Thus

$$E[e^{\lambda_n Y_n}] = E[\prod_{i=1}^n e^{\lambda_n X_{i,n}}] = \prod_{i=1}^n E[e^{\lambda_n X_{i,n}}]. \tag{8.44}$$

Applying (8.43) and observing that the exponents add,

$$E[e^{\lambda_n Y_n}] \leq e^{(1+\epsilon)\lambda_n^2\sigma_n^2/2}. \tag{8.45}$$

We now apply the Chernoff bound (8.4), and

$$\Pr[Y_n \geq a_n\sigma_n] \leq E[e^{\lambda_n Y_n}]e^{-\lambda_n a_n\sigma_n} \leq e^{(1+\epsilon)\lambda_n^2\sigma_n^2/2}e^{-\lambda_n a_n\sigma_n}. \tag{8.46}$$

As $\lambda_n = a_n/\sigma_n$ the σ_n terms cancel, and

$$\Pr[Y_n \geq a_n\sigma_n] \leq e^{(1+\epsilon)a_n^2/2}e^{-a_n^2} = e^{-(1-\epsilon)a_n^2/2}. \tag{8.47}$$

As $n \to \infty$ we can make ϵ arbitrarily small positive yielding (8.42). □

We can give a simple sufficient condition for (8.41) to hold. Let Z be any random variable with $E[Z] = 0$ and set $\sigma^2 = \mathrm{Var}[Z]$. From (8.11)

$$E[e^{\lambda Z}] = 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{i=3}^{\infty} \frac{\lambda^i E[Z^i]}{i!}. \tag{8.48}$$

Suppose that $|Z| \leq K$ always. Then for all $i \geq 3$, $|Z^i| \leq K^{i-2}Z^2$ always so $E[Z^i] \leq E[|Z^i|] \leq K^{i-2}\sigma^2$. Thus

$$\sum_{i=3}^{\infty} \frac{\lambda^i E[Z^i]}{i!} \leq K\lambda^3\sigma^2 \sum_{i=3}^{\infty} \frac{\lambda^{i-3}K^{i-3}}{i!} \leq K\lambda^3\sigma^2 e^{K\lambda}. \tag{8.49}$$

When K is fixed and $\lambda \to 0$, this is $o(\lambda^2\sigma^2)$ and so condition (8.41) holds.

Theorem 8.5. *Let $X_{i,n}, Y_n$, $\sigma_{i,n}$, σ_n be as in Theorem* 8.4. *Let $a_n = o(\sigma_n)$. Assume an absolute constant K such that all $|X_{i,n}| \leq K$ always. Then*

$$\Pr[Y_n \geq a_n\sigma_n] \leq e^{-(1+o(1))a_n^2/2}. \tag{8.50}$$

This follows immediately from Theorem 8.4 as (8.49) gives that the conditions of that theorem hold.

As an example, consider $Y = BIN[n, p(n)] - np(n)$. Y is the sum of n identically distributed random variables which have value $1-p(n)$ or $-p(n)$ so certainly have absolute value at most 1. Assume $p(n) \leq \frac{1}{2}$. Set $\sigma^2 = \mathrm{Var}[Y] = np(n)(1-p(n))$, which is between $np(n)/2$ and $np(n)$. Suppose $a_n = o(\sqrt{np(n)})$. The conditions of Theorem 8.5 then apply, and $\Pr[Y_n \geq a_n\sigma_n]$ is bounded from above by $\exp[(1+o(1))a_n^2/2]$.

Chapter 9

Primes

> The weak points of [Alan Turing's] argument were essentially the weaknesses of the analytic scientific method when applied to the discussion of human beings. Concepts of objective truth that worked so well for the prime numbers could not so straightforwardly be applied by scientists to other people.
> – Andrew Hodges, *The Enigma*

Primes would seem to be the ultimate in precision. A number 317 is either prime or it is not (this one is!), and there is no approximation to its primality. Nonetheless, Asymptopia is the proper place to examine primes in the aggregate.

Definition 9.1. For $n \geq 2$, $\pi(n)$ denotes the number of primes p with $2 \leq p \leq n$.

Our goal in this chapter is to approach one of the great theorems of mathematics.

Theorem 9.2 (The Prime Number Theorem)**.**

$$\pi(n) \sim \frac{n}{\ln n}. \tag{9.1}$$

This result was first conjectured in the early nineteenth century. (While the conjecture is sometimes attributed to Gauss, the history

is murky.) It was a central problem for that century, finally being proven independently by Hadamard in [**Had96**] and Vallée-Poussin in 1898 in [**dlVP96**]. Their proofs involved complex variables, and a long search continued for an elementary proof. This was finally obtained in 1949 by Selberg in [**Sel49**] and Erdős in [**Erd49**]. Still, a full proof of (9.1) is beyond the limits of this work. We shall come close to it with the following results:

Theorem 9.3. *There exists a positive constant c_1 such that*

$$(c_1 + o(1))\frac{n}{\ln n} \leq \pi(n). \tag{9.2}$$

That is, $\pi(n) = \Omega(n/\ln n)$.

Theorem 9.4. *There exists a positive constant c_2 such that*

$$\pi(n) \leq (c_2 + o(1))\frac{n}{\ln n}. \tag{9.3}$$

That is, $\pi(n) = O(n/\ln n)$.

Together, Theorems 9.3 and 9.4 yield

$$\pi(n) = \Theta\left(\frac{n}{\ln n}\right). \tag{9.4}$$

With more effort we shall show

Theorem 9.5. If *there exists a positive constant c such that*

$$\pi(n) \sim c\frac{n}{\ln n}, \tag{9.5}$$

then $c = 1$.

9.1. Fun with Primes

A Break! No asymptotics in this section!

How many factors of the prime 7 are there in 100!? The numbers $7, 14, \ldots, 98$ all have a factor of 7 so that gives $\frac{98}{7} = 14$ factors. *And,* 49 and 98 have a second factor of 7 which gives an additional $\frac{98}{49} = 2$ factors. In total there are $16 = 14 + 2$ factors of 7.

Definition 9.6. For $n \geq 1$ and p prime, $v_p(n)$ denotes the number of factors p in n. Equivalently, $v_p(n)$ is that nonnegative integer a such that p^a divides n, but p^{a+1} does not divide n.

Theorem 9.7. *For any $n \geq 1$ and p prime*

$$v_p(n!) = \sum_{i=1}^{\infty} \lfloor \frac{n}{p^i} \rfloor. \tag{9.6}$$

Equivalently,

$$v_p(n!) = \sum_{i=1}^{s} \lfloor \frac{n}{p^i} \rfloor \text{ with } s = \lfloor \log_p n \rfloor. \tag{9.7}$$

When $i > \lfloor \log_p n \rfloor$, $n < p^i$, so the addend in (9.6) is zero, and thus the seemingly infinite sum is in fact a finite sum. The argument with $p = 7, n = 100$ easily generalizes. For any $i \leq s$ there are $\lfloor np^{-i} \rfloor$ numbers $1 \leq j \leq n$ that have (at least) i factors of p. We count each such i and j once, as then an i with precisely u factors of p will be counted precisely u times.

We apply Theorem 9.7 to study binomial coefficients . Let $n = a + b$, and set $C = \binom{n}{a} = \frac{n!}{a!b!}$. Applying (9.7),

$$v_p(C) = v_p(n!) - v_p(a!) - v_p(b!) = \sum_{i=1}^{s} \lfloor \frac{n}{p^i} \rfloor - \lfloor \frac{a}{p^i} \rfloor - \lfloor \frac{b}{p^i} \rfloor \tag{9.8}$$

with $s = \lfloor \log_p n \rfloor$ as in (9.7).

Theorem 9.8. *With $n = a + b$, p prime, and $C = \binom{n}{a}$,*

$$0 \leq v_p(C) \leq \lfloor \log_p n \rfloor. \tag{9.9}$$

Proof. Set $\alpha = ap^{-i}$, $\beta = bp^{-i}$. Then the addend in (9.8) is

$$\lfloor \alpha + \beta \rfloor - \lfloor \alpha \rfloor - \lfloor \beta \rfloor. \tag{9.10}$$

This term is zero if the fractional parts of α, β sum to less than one, and it is one if they sum to one or more. The sum (9.8) consists of $s = \lfloor \log_p n \rfloor$ terms, each one or zero, and so lies between 0 and s. □

Remark. With $n = a + b$ there are two arguments why $a!b!$ divides $n!$. One: The proof of Theorem 9.8 gives that, for all primes p, $v_p(n!) \geq v_p(a!) + v_p(b!) = v_p(a!b!)$, and thus $a!b!$ divides $n!$. Two: The quotient $\frac{n!}{a!b!} = \binom{n}{a}$ *counts* the a-subsets of n-sets and hence must be a nonnegative integer. Which proof one prefers is an aesthetic

question,[1] but it is frequently useful to know more than one proof of a theorem.

There is an amusing way of calculating $v_p(C)$ with $C = \binom{n}{a}$ and $a+b=n$, first shown in [**E.E52**]. Write a, b is base p. Add them (in base p) so that you will get n in base p.

Theorem 9.9. *$v_p(C)$ is the number of carries when you add a, b getting n, all in base p.*

For example, let $a = 33$, $b = 25$ so $n = 58$ (written in decimal), and set $p = 7$. In base 7, $a = 45$, $b = 34$. When we add them[2]

```
   45
+  34
 ----
  112
```

There were two carries, and $v_p(\binom{45}{34}) = 2$.

We indicate the argument. For each $1 \le i$, we get a carry from the $(i-1)$-st place (counting from the right, starting at 0) to the i-th place if and only if the fractional parts of ap^{-i} and bp^{-i} add to at least one, and that occurs if and only if term (9.10) is one.

9.2. Prime Number Theorem—Lower Bound

Let n be even (n odd will be similar). The upper *and* lower bounds come from examining the prime factorization of binomial coefficients. Set $r = \pi(n)$, let $p_1, \ldots, p_r$ denote the primes up to n, and write

$$\binom{n}{n/2} = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}. \tag{9.11}$$

(There might not be a factor of p_i. In that case we simply write $\alpha_i = 0$.) We rewrite the upper bound of Theorem 9.8 as

$$p_i^{\alpha_i} \le n. \tag{9.12}$$

[1] These authors like the "counts" argument.

[2] To paraphrase the wonderful songwriter Tom Lehrer, base seven is just like base ten—if you are missing three fingers!

Thus

$$\binom{n}{n/2} \le n^r. \tag{9.13}$$

Stirling's formula gives an asymptotic formula for $\binom{n}{n/2}$, but here we use only the weaker $\binom{n}{n/2} = 2^{n(1+o(1))}$. Taking ln of both sides of (9.13) and dividing gives

$$\pi(n) = r \ge \frac{\ln \binom{n}{n/2}}{\ln n} = \frac{n}{\ln n}(\ln 2)(1 + o(1)). \tag{9.14}$$

What if n is odd? In Asymptopia we simply apply (9.14) to the even $n-1$. Thus

$$\pi(n) \ge \pi(n-1) \ge \frac{\ln \binom{n-1}{(n-1)/2}}{\ln(n-1)}, \tag{9.15}$$

which is again $\frac{n}{\ln n}(\ln 2)(1 + o(1))$.

9.3. Prime Number Theorem—Upper Bound

Again assume n is even. There are $\pi(n) - \pi(n/2)$ primes p with $\frac{n}{2} < p < n$. Each of them appears in $\binom{n}{n/2}$ to the first power. (They appear once in the numerator as a factor of p and never in the denominator.) Thus, with the product over these primes,

$$\prod p \le \binom{n}{n/2}. \tag{9.16}$$

We again do not need a more precise estimate, and here we simply bound $\binom{n}{n/2} \le 2^n$. Each factor p is a factor of at least $\frac{n}{2}$. Thus

$$(\frac{n}{2})^{\pi(n)-\pi(\frac{n}{2})} \le 2^n. \tag{9.17}$$

Taking ln of both sides gives

$$\pi(n) - \pi(\frac{n}{2}) \le \frac{n}{\ln(n/2)}(\ln 2). \tag{9.18}$$

For $n = 2k+1$ odd, we apply the same argument to $\binom{n}{k}$ getting an upper bound on $\pi(n) - \pi(k+1)$. We combine the even and odd cases by writing

$$\pi(n) - \pi(\lceil \frac{n}{2} \rceil) \le \frac{n}{\ln(n/2)}(\ln 2). \tag{9.19}$$

Turning (9.19) into an upper bound on $\pi(n)$ is a typical problem in Asymptopia. Set $x_0 = n$ and $x_{i+1} = \lceil \frac{x_i}{2} \rceil$. This sequence decreases until finally reaching $x_s = 1$. Applying (9.19) to $n = x_0, \ldots, x_{s-1}$ and adding, we get

$$\pi(n) \leq \sum_{i=0}^{s-1} \frac{x_i}{\ln(x_i/2)}(\ln 2). \tag{9.20}$$

In the exact world this would be a daunting sum. In Asymptopia we know that we are aiming for $\Theta(\frac{n}{\ln n})$, so let us stop the sum when x_i gets somewhat lower than that. For definiteness (but there is a lot of room here) let u be the first index with $x_u \leq n \ln^{-2} n$. Applying (9.19) only down to x_{u-1} and adding, we get

$$\pi(n) - \pi(x_u) \leq \sum_{i=0}^{u-1} \frac{x_i}{\ln(x_i/2)}(\ln 2). \tag{9.21}$$

Now we use the trivial bound $\pi(x_u) \leq x_u \leq n \ln^{-2} n$. While this is a "bad" bound for $\pi(x_u)$, it is a negligible value for us and

$$\pi(n) \leq o(\frac{n}{\ln n}) + \sum_{i=0}^{u-1} \frac{x_i}{\ln(x_i/2)}(\ln 2). \tag{9.22}$$

In the range $0 \leq i < u$, $n \geq x_i \geq n \ln^{-2}(n)$. The smallest $\ln(x_i/2)$ would be $\ln(n) - 2\ln\ln(n) - \ln(2)$. But this is $\sim \ln(n)$. Thus all $\ln(x_i/2)$ terms are $\sim \ln(n)$ and

$$\pi(n) \leq o(\frac{n}{\ln n}) + (1 + o(1)) \sum_{i=0}^{u-1} \frac{x_i}{\ln(n)}(\ln 2). \tag{9.23}$$

The x_i form a near geometric series that sums to less than $2n$, so

$$\pi(n) \leq (1 + o(1))n\frac{2\ln 2}{\ln n}, \tag{9.24}$$

giving Theorem 9.4.

9.4. Prime Number Theorem with Constant

Note. This section gets quite technical and should be considered optional.

Here we show Theorem 9.5. That is, we *assume* that there is a constant c such that $\pi(n) \sim c(n/(\ln n))$ and then show that c must be 1. It is a big *if*. *A priori,* from Theorems 9.3 and 9.4 the ratio of $\pi(n)$ to $n/(\ln n)$ could oscillate between two positive constants, never approaching a limit. Indeed, despite their similar appearances, the proof of Theorem 9.5 is a long way from the proof of the Prime Number Theorem (9.1) itself.

We consider the factorization (9.11) more carefully. Our goal will be to show that if $c \neq 1$, then the left-hand and right-hand sides cannot match. We split the primes from 1 to n into intervals. We shall let K be a large but fixed constant. (More about just how large later.) For $1 \leq i < K$, let P_i denote the set of primes p with

$$\frac{n}{i+1} < p \leq \frac{n}{i}, \tag{9.25}$$

and let SP (small primes) denote the set of primes p with $p < \frac{n}{K}$. Let V_i, $1 \leq i < K$ denote the contribution of the $p \in P_i$ to the factorization (9.11). That is, V_i is the product of $p_j^{\alpha_j}$ in (9.11), where p_j is restricted to P_i. Similarly, let V_{SP} denote the contribution of the $p \in SP$ to the factorization (9.11). That is, V_i is the product of $p_j^{\alpha_j}$ in (9.11), where p_j is restricted to SP.

We first show that SP makes a relatively small contibution to (9.11). There are $\leq \pi(n/K)$ primes $p \in SP$ and each (9.12) contributes at most a factor of n so that $V_{SP} \leq n^{\pi(n/K)}$.

From (9.24) $\pi(n/K) < ((2\ln 2) + o(1))(n/K)/\ln(n/K)$. With K fixed, $\ln(n/K) \sim \ln(n)$ so that $\pi(n/K) < (\ln 2 + o(1))(n/K)/\ln(n)$. Thus (9.25),

$$V_{SP} < n^{(2\ln 2 + o(1))(n/K)/\ln(n)} = 2^{(2n/K)(1+o(1))}, \tag{9.26}$$

so that

$$\ln(V_{SP}) < \frac{2n\ln 2}{K}(1 + o(1)). \tag{9.27}$$

While this is not a small number in absolute terms, it will be relatively small compared to the total contribution which is $2^{n(1+o(1))}$.

For $1 \leq i < K$, we now look at V_i. As all primes considered have $p > \frac{n}{K}$ and K is fixed, they have $p > \sqrt{n}$. Thus, the sum of Theorem

9.7 has only one term. Theorem 9.8 with $a = n/2$ is then simply

$$v_p\left(\binom{n}{n/2}\right) = \lfloor n/p \rfloor - 2\lfloor n/2p \rfloor. \tag{9.28}$$

This is either zero or one, and it is one precisely when $\lfloor n/p \rfloor$ is odd. We have *designed* P_i so that $\lfloor n/p \rfloor = i$ for $p \in P_i$. When i is even, no primes $p \in P_i$ appear in the factorization (9.11) (or, the same thing, they appear with exponent zero), and so $V_i = 1$. (For example, with $\frac{n}{7} < p \le \frac{n}{6}$, $n!$ has six factors of p and $(n/2)!^2$ has twice three factors of p, and they all cancel.)

Now suppose $1 \le i < K$ is odd. Then V_i is simply the product of all primes $p \in P_i$. Each such prime p lies between $\frac{n}{K}$ and n, and so can be considered $p = n^{1+o(1)}$. The number of such primes is $\pi(n/i) - \pi(n/(i+1))$. In this range $\ln(n/i) \sim \ln n$. Our assumption for Theorem 9.5 then gives that $\pi(n/i) \sim c\frac{n}{i \ln n}$ and that $\pi(n/(i+1)) \sim c\frac{n}{(i+1)\ln n}$. We deduce[3] that the number of primes is $\sim c\frac{n}{\ln n}(\frac{1}{i} - \frac{1}{i+1})$.

Thus

$$V_i = n^{c(1+o(1))(n/(\ln n))(\frac{1}{i} - \frac{1}{i+1})} \tag{9.29}$$

and

$$\ln(V_i) \sim cn(\frac{1}{i} - \frac{1}{i+1}). \tag{9.30}$$

From the factorization (9.11) we have

$$\ln\left(\binom{n}{n/2}\right) = \ln V_{SP} + \sum \ln(V_i). \tag{9.31}$$

For convenience, assume $K = 2T$ is even so we can write the odd $i < K$ as $2j-1$, $1 \le j \le T$. From (2.28) the left-hand side of (9.31) is asymptotically $n \ln 2$. Thus

$$(1 + o(1))n \ln 2 = cn(1 + o(1)) \sum_{j=1}^{T} \left(\frac{1}{2j-1} - \frac{1}{2j}\right) + \ln V_{SP}. \tag{9.32}$$

[3] *Caution.* Subtraction in Asymptopia is dangerous! It is critical here that $i \le K$ and that K is a fixed constant, so $\frac{1}{i}$ and $\frac{1}{i+1}$ are positive constants. Where, for example, $K = \ln \ln n$, we could not do the subtraction. With $i \sim (\ln \ln n)/2$, for example, the asymptotics of $\pi(n/i)$ and $\pi(n/(i+1))$ would be the same, and so one could *not* deduce the asymptotics of their difference!

Dividing by n,

$$(9.33) \qquad (1+o(1))(\ln 2) = c(1+o(1)) \sum_{k=1}^{2T-1} \frac{(-1)^{k+1}}{k} + \frac{1}{n} \ln V_{SP}.$$

We need[4] the fact that

$$(9.34) \qquad \ln 2 = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{i} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots .$$

We can now see the idea. The $\ln(V_{SP})$ will be negligible, and (9.33) becomes $\ln 2 = c(\ln 2)$. The actual argument consists of eliminating all $c \neq 1$.

Suppose $c > 1$. Select $K = 2T$ so that $c \sum_{k=1}^{2T-1} \frac{(-1)^{k+1}}{k} > \ln 2$. As $\ln V_{SP} \geq 0$ the right-hand side of (9.33) would be bigger than the left-hand side.

Suppose $c < 1$. Applying the upper bound (9.27), the right-hand side of (9.33) would be at most $c \sum_{k=1}^{2T-1} \frac{(-1)^{k+1}}{k} + \frac{2 \ln 2}{K}$. As $K \to \infty$, this sum approaches $c \ln 2$, which is less than $\ln 2$. Thus, we may select K so that this sum is less than $\ln 2$.[5] But now the right-hand side of (9.33) would be smaller than the left-hand side.

Both assumptions led to a contradiction, and since we *assumed* that c existed, it must be that $c = 1$.

9.5. Telescoping

Suppose we have a reasonable function $f(x)$ and we wish to asymptotically evaluate $\sum_{p \leq n} f(p)$. We assume the Prime Number Theorem (9.1), giving the asymptotics of $\pi(s)$ as $s \to \infty$. On an intuitive level, we think of $1 \leq s \leq n$ as being prime with "probability" $\pi(s)/s \sim 1/(\ln s)$. Then s, $1 \leq s \leq n$, would contribute $f(s)/(\ln s)$ to the sum, and $\sum_{p \leq n} f(p)$ would be roughly $\sum_{s \leq n} f(s)/(\ln s)$. This is not a proof: integers are either prime or they are not, yet surprisingly we can often get this intuitive result. The key is called telescoping.

[4] Again, from calculus!

[5] A subtle wrinkle here: while we examine the behavior as $K \to \infty$, we select K a constant, dependent only on c.

We write

$$\sum_{p\le n} f(p) = \sum_{s=2}^{n} f(s)(\pi(s) - \pi(s-1)). \tag{9.35}$$

Reversing sums (and noting $\pi(1) = 0$),

$$\sum_{s=2}^{n} f(s)(\pi(s)-\pi(s-1)) = f(n)\pi(n)+\sum_{s=2}^{n-1} \pi(s)(f(s)-f(s+1)). \tag{9.36}$$

While (9.36) is correct, its effectiveness depends on our ability to asymptotically calculate the sum. An important success is when $f(s) = \frac{1}{s}$, we ask for the asymptotics of

$$F(n) = \sum_{p\le n} \frac{1}{p}. \tag{9.37}$$

The first term of (9.36) is then $\sim \frac{1}{n}\frac{n}{\ln n} = o(1)$. The sum is asymptotically $\sum \frac{s}{\ln s}\frac{1}{s(s+1)} \sim \sum \frac{1}{s\ln s}$, the sum from $s = 2$ to $n-1$. From Chapter 4,

$$\sum_{s=2}^{n-1} \frac{1}{s\ln s} \sim \int_2^n \frac{dx}{x\ln x} \sim \ln\ln n. \tag{9.38}$$

That is, $F(n) \sim \ln\ln n$, strengthening Theorem 0.2. For another example, take $f(s) = s$ so that $F(n) = \sum_{p\le n} p$. Then

$$F(n) = n\pi(n) - \sum_{s=2}^{n-1} \pi(s) \sim \frac{n^2}{\ln n} - \int_2^{n-1} \frac{s}{\ln s} ds. \tag{9.39}$$

This integral was handled in Chapter 3 where it is shown in (3.44) to be $\sim \frac{n^2}{2\ln n}$. Thus $F(n) \sim \frac{n^2}{2\ln n}$.

Chapter 10

Asymptotic Geometry

> Who could ever calculate the path of a molecule?
> How do we know that the creations of worlds are
> not determined by falling grains of sand?
> — Victor Hugo, *Les Misérables*

The beauty of geometry lies in its precision: the square of the hypotenuse *is* the sum of the squares of the other two sides. Asymptopia is a universe of approximation, of sometimes quite coarse estimates. Asymptotic geometry is a meld.

10.1. Small Triangles

Let P, Q, R be independently and uniformly selected in the unit square $[0,1]^2$. Let $\mu(PQR)$ denote the area of the triangle PQR. For $\epsilon > 0$ define

$$f(\epsilon) = \Pr[\mu(PQR) \leq \epsilon]. \tag{10.1}$$

A precise formula for $f(\epsilon)$ would be quite challenging. Here we search for the asymptotics of $f(\epsilon)$ as ϵ approaches zero. Of course, $f(\epsilon)$ will approach zero. But at what rate? Is it order of ϵ, of ϵ^2, of $\epsilon \ln(\epsilon^{-1})$? We shall show

Theorem 10.1. *The probability that the area of the triangle PQR is less than ϵ is approximately ϵ:*

$$f(\epsilon) = \Theta(\epsilon). \tag{10.2}$$

We will *not* attempt to get the right constant. This allows us to be quite cavalier with constant factors.

10.1.1. The Upper Bound. First we select P uniformly. Now we select Q and consider the distance $|PQ|$. The probability that the distance is between r and $r + \Delta r$ is at most $\pi(r + \Delta r)^2 - \pi r^2$, the area of the annulus around P. (It might be less as some of the annulus might lie outside the unit square.) Thus the density function for r is at most $2\pi r \cdot dr$. Further $0 \le r \le \sqrt{2}$ tautologically.

We now condition on $|PQ| = r$ and ask when can R be so that $\mu(PQR) \le \epsilon$? The altitude from R to PQ must be at most $\frac{2\epsilon}{r}$. Extend the line PQ in both directions and create a band with width $\frac{4\epsilon}{r}$, from $\frac{2\epsilon}{r}$ above the line PQ to $\frac{2\epsilon}{r}$ below the line PQ. The intersection of the band with the square can be at most $\sqrt{2}$ long and so its area is at most $\frac{4\sqrt{2}\epsilon}{r}$.

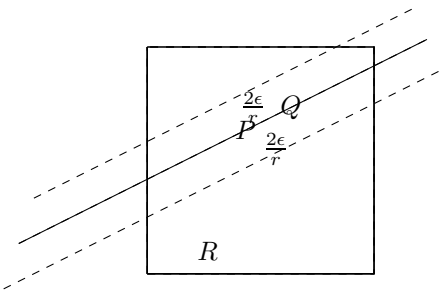


Figure 1. Extending the line PQ and creating bands above and below it.

This gives an upper bound on the probability. Thus

$$f(\epsilon) \le \int_0^{\sqrt{2}} \frac{4\sqrt{2}\epsilon}{r} 2\pi r \cdot dr = 16\pi\sqrt{2}\epsilon. \tag{10.3}$$

10.1.2. The Lower Bound. Let us try to use the ideas of the upper bound argument, turned around to give a lower bound of the same order of magnitude. (This does not always work. Indeed, sometimes when it does not work despite your best efforts, it is an indication that your upper bound is too coarse.) To avoid edge effects, let us select P in the disk of radius .01 around the center $(.5, .5)$. This gives a factor $10^{-4}\pi$, which is not a problem since we are not worrying about constants. Now select Q in the disk of radius .01 around P. Another constant factor of $10^{-4}\pi$. As P, Q are both near the center, the line PQ, when extended to the edges of the square, has length at least 1. As $|PQ| \leq 0.01$, the band has width $4\epsilon/|PQ| \geq 400\epsilon$, so (as we are dealing with a lower bound) replace it with a band of width 400ϵ. Now the area of the intersection of the band with the square is at least 400ϵ. (When ϵ is large, this will not make sense, as then the band would be largely outside the square; however, as $\epsilon \to 0$, it is correct.) Thus the probability of R being in the band, so that $\mu(PQR) \leq \epsilon$, is at least 400ϵ. The total lower bound is then $(10^{-4}\pi)^2 \cdot 400\epsilon$. The constants are ridiculously small (surely you can do better!), but it is $\Omega(\epsilon)$.

10.1.3. The Heilbronn Triangle Problem. Here is a beautiful question due to the mathematician Hans Heilbronn that remains substantially open. Let $S = \{P_1, \ldots, P_n\}$ be a set of n vertices in the unit square. For each three $P, Q, R \in S$, let $\mu(PQR)$ denote the area of triangle PQR, considered zero if P, Q, R are collinear. Let $m(S)$ denote the *minimum* such area, that is, the $\min \mu(PQR)$ over all $P, Q, R \in S$. Heilbronn asked how large this can be?

That is, defining

$$\Delta(n) = \min_S \mu(S), \tag{10.4}$$

where S ranges over all n-sets, what can one say asymptotically about $\Delta(n)$?

So far not very much! But let's combine asymptotic geometry and Erdős Magic to get a lower bound. That is, we want to "find" n vertices where all of the triangles have area at least ϵ, where we make ϵ as large as we can.

For now, let ϵ be a parameter. We call triangle PQR *large* if its area is at least ϵ. We will pick points P_i at random in the unit square. Instead of selecting n vertices (which is what we want in the end), we shall select some $m \geq n$ vertices at random and set $S^+ = \{P_1, \ldots, P_m\}$. For each small triangle $P_iP_jP_k$, we shall delete one of the vertices. There are $\binom{m}{3} \leq m^3/6$ triples $P, Q, R \in S^+$, and each (using the upper bound above) is large with probability at most $16\pi\sqrt{2}\epsilon$. Removing one vertex from each large triangle giving a set S. Tautologically, S has no large triangle. We want S to have at least n points. We started with m points and removed an expected number $\leq m^3(8\pi\sqrt{2}/3)\epsilon$ vertices. (Some vertices may have been removed more than once but that is only in our favor.) The expected number of vertices remaining is then at least $f(m)$, where

$$f(m) = m - m^3(8\pi\sqrt{2}/3)\epsilon = m - cm^3\epsilon \tag{10.5}$$

and $c = \frac{8}{3}\pi\sqrt{2}$.

We want $|S| \geq n$. This gives an asymptotic calculus problem: What (as an asymptotic function of n) is the largest ϵ such that $f(m) \geq n$ for some m? Given ϵ, $f(m)$ hits a minimum when $f'(m) = 1 - 3cm^2\epsilon = 0$, when $m = (3c)^{-1/2}\epsilon^{-1/2}$. (There are some technical issues as m is not necessarily integral. As n is large and $m \geq n$, it is not hard to show that taking the nearest integer to m above has negligible effect.) At the minimum $f(m) = 2m/3$, so now $f(m) = \frac{2}{3}(3c)^{-1/2}\epsilon^{-1/2}$. We want to take n so that $f(m) = n$. Reversing the function, we find $\epsilon = c_1 n^{-2}$ with c_1 an absolute (though quite small!) constant. Looking back, we start with $m = \frac{3m}{2}$ vertices and select ϵ so that there are at most $\frac{n}{2}$ small triangles. We have shown

Theorem 10.2. *The minimum area over the n-sets is at least n^{-2}:*

$$\Delta(n) \geq c_1 n^{-2}. \tag{10.6}$$

How close is this to the actual answer? Using quite sophisticated techniques, mathematicians Miklós Ajtai, János Komlós, and Endre Szemerédi were able to improve this to $\Delta(n) \geq c_2 n^{-2} \ln n$ in [**AKS83**]. But the best known upper bound is roughly on the order $n^{-8/7}$. So the gap is quite large!

10.2. The Convex Hull of n Random Points

Let Ω denote the unit disk, centered at the origin, in R^2. Let $P_1,\ldots,P_n$ be n vertices selected uniformly and independently from Ω. Consider the convex hull of those n points. By a vertex on the convex hull, we refer to it being on the boundary of it, and by a vertex in the convex hull, we imply that it is in the interior of it. We are now interested in the following question:

On average, how many vertices will be on that convex hull?

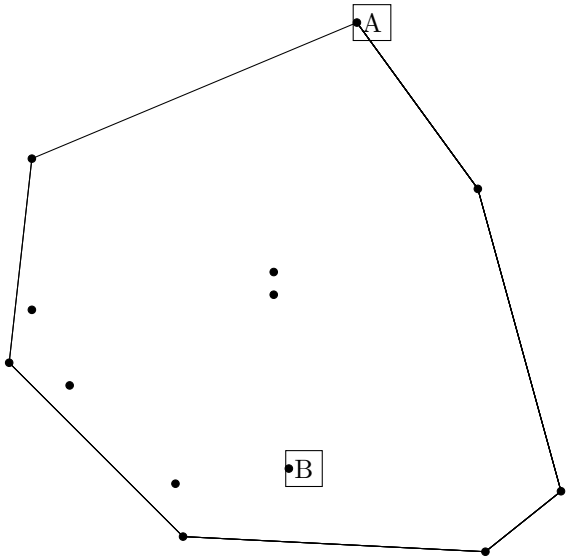


Figure 2. The convex hull of points in the plane. Point A is *on* the convex hull, while point B is *in* the convex hull.

We let A_n denote the expected number of vertices on the convex hull. A moment's reflection convinces one that the points on the convex hull are highly likely to be near the border of Ω, and so most of the n points will not be on the convex hull. It should not be surprising that $A_n = o(n)$. But what is the order of A_n? Is it $\ln(n)$? $n\ln^{-10} n$? $n^{1/4}$? We shall see in this section that

$$A_n = \Theta(n^{1/3}). \tag{10.7}$$

For convenience let us relabel the vertices $P, P_1, \ldots, P_{n-1}$. Let B_n denote the event that P is a vertex on the convex hull. Equivalently, B_n is the event that P is not on the convex hull of $P_1, \ldots, P_{n-1}$. As all vertices have the same probability for being a vertex on the convex hull

$$A_n = n \Pr[B_n]. \tag{10.8}$$

We first select P. From the symmetry of Ω, the only important aspect about P is its distance, call it R, from the origin. As P is selected uniformly, we have $\Pr[R \leq r] = r^2$ for $r \in [0,1]$, and hence R has density function $2r$, $r \in [0,1]$. Let $B_{n,r}$ be the event B_n conditional on P being distance r from the origin. Thus

$$\Pr[B_n] = \int_0^1 \Pr[B_{n,r}] 2r dr. \tag{10.9}$$

Still, finding $\Pr[B_{n,r}]$ looks to be (and is!) a fearsome task. We approach it by *giving ground*. Let L be that line through P perperdicular to the line from P to the origin. L splits Ω into two parts; let S_P denote the smaller part.

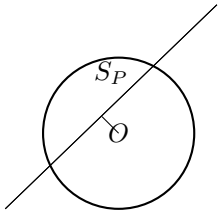


Figure 3. Visualizing S_P

Call P extremal if none of the $P_1, \ldots, P_{n-1}$ lie in S_P, and let $C_{n,r}$ be the event that P is extremal. An extremal point will necessarily lie on the convex hull. Thus $\Pr[B_{n,r}] \geq \Pr[C_{n,r}]$, and so finding the asymptotics of $\Pr[C_{n,r}]$ will give us a *lower bound* on $\Pr[B_{n,r}]$ and thence A_n.

- Seemingly intractible calculations can be approached asymptotically by giving ground. The art lies in finding the right way to give ground so that the problem is now tractible *and* that not too much ground has been given.

10.2.1. The Lower Bound. What is $\Pr[C_{n,r}]$? Let $g(r)$ denote the area of S_P as defined above. As Ω has area π, each P_i has probability $g(r)/\pi$ of being in S_P. As the P_i are chosen independently,

$$\Pr[C_{n,r}] = \left(1 - \frac{g(r)}{\pi}\right)^{n-1}. \tag{10.10}$$

Thus we have the exact formula

$$\Pr[C_n] = \int_0^1 \left(1 - \frac{g(r)}{\pi}\right)^{n-1} 2r dr. \tag{10.11}$$

Our rough discussion led us to believe that only points near the boundary would contribute significantly to A_n. This corresponds to values of r near one. We reparametrize by setting $1 - r = s$ and $h(s) = g(1 - s)$ so that

$$\Pr[C_n] = \int_0^1 (1 - \frac{h(s)}{\pi})^{n-1} 2(1 - s) ds. \tag{10.12}$$

To find $h(s)$, it is convenient to assume (by symmetry) that $P = (0, s - 1)$ so that we have the exact formula

$$h(s) = \int_{-\sqrt{2s-s^2}}^{+\sqrt{2s-s^2}} [s + \sqrt{1 - x^2} - 1] dx. \tag{10.13}$$

This integral can be evaluated exactly, but that is not our style. Rather, we consider the asymptotics of $h(s)$ for s small. Consider the rectangle bounded by $x = \pm\sqrt{2s - s^2}$, $y = s - 1$ and $y = -1$ crossed by the unit circle. We may then visualize $h(s)$ as that area of the rectangle above the unit circle. The entire rectangle has area $2s\sqrt{2s - s^2} \sim 2\sqrt{2}s^{3/2}$. The circle in this tiny region is effectively a parabola! (See Figure 4.)

That is, $1 - \sqrt{1 - x^2} \sim \frac{1}{2}x^2$ as s, and hence x, is small. The limits of integration are $\sim \pm\sqrt{2s}$. The area under the circle is then given asymptotically by

$$= \int_{-\sqrt{2s-s^2}}^{+\sqrt{2s-s^2}} [1 - \sqrt{1 - x^2}] dx \sim \int_{-\sqrt{2s}}^{+\sqrt{2s}} \frac{1}{2}x^2 dx = \frac{4}{3}s^{3/2}. \tag{10.14}$$

We therefore have

$$h(s) \sim Ks^{3/2} \text{ with } K = 2\sqrt{2} - \frac{4}{3}. \tag{10.15}$$

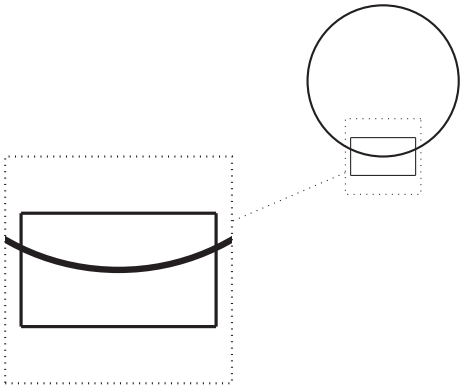


Figure 4. The circle is nearly a parabola for $s = 0.1$!

At this stage we can find the correct scaling for s. The key term in (10.12) is $(1 - \pi^{-1}h(s))^{n-1}$. This is a decreasing function in $h(s)$. When $h(s) \ll n^{-1}$, the term is asymptotically one, and when $h(s) \gg n^{-1}$, the term is asymptotically zero. The "critical region" is when $h(s)$ is of the order n^{-1}. Now looking at (10.15), this will occur when s is of the order $n^{-2/3}$. We are therefore led to the *scaling*

$$s = zn^{-2/3}. \tag{10.16}$$

(Note that the constants π, K do not affect the choice of scaling, though they do affect the constants in our answer below.) For z constant, $h(s) \sim Kz^{3/2}n^{-1}$ so that

$$(1 - \frac{h(s)}{\pi})^{n-1} \sim e^{-(K/\pi)z^{3/2}}. \tag{10.17}$$

We make the change of variables (10.16) in (10.12), noting that $2(1 - s) \sim 2$ when s is small, giving

$$\Pr[C_n] \sim 2n^{-2/3} \int_0^\infty e^{-(K/\pi)z^{3/2}} dz. \tag{10.18}$$

The integral converges and so $\Pr[C_n] = \Theta(n^{-2/3})$, and we find the lower bound

$$A_n = \Omega(n^{1/3}). \tag{10.19}$$

Formal Justification of (10.19). As the particular constants do not concern us, we only need to show $\Pr[C_n] \geq (c + o(1))n^{2/3}$ for some

(small) positive constant c. Since we only need the lower bound, let's restrict the region to $0 \le z \le 1$. We then have the precise lower bound

$$\Pr[C_n] \ge n^{-2/3} \int_0^1 (1 - \frac{h(zn^{-2/3})}{\pi})^{n-1} 2(1 - zn^{-2/3}) dz. \tag{10.20}$$

For n large, $2(1 - zn^{-2/3}) \ge 1$ (say) for all $z \in [0, 1]$ so

$$\Pr[C_n] \ge n^{-2/3} \int_0^1 g_n(z) dz \quad \text{with } g_n(z) = (1 - \frac{h(zn^{-2/3})}{\pi})^{n-1}. \tag{10.21}$$

For any fixed z, $g_n(z) \to \exp[-(K/\pi)z^{3/2}]$ and $0 \le g_n(z) \le 1$ tautologically. Hence

$$\int_0^1 g_n(z) dz \to \int_0^1 e^{-(K/\pi)} z^{3/2} dz, \tag{10.22}$$

which is a positive constant.

10.2.2. The Upper Bound. (*Note*: This section gets quite technical and should be considered optional.)

- Examination of lower bound arguments can provide vital clues toward upper bound arguments, and vice versa.

Our goal is now to show $A_n = O(n^{1/3})$ for which we need $\Pr[B_n] = O(n^{-2/3})$, where $\Pr[B_n]$ is given by (10.9) or, equivalently,

$$\Pr[B_n] = n^{-2/3} \int_0^{n^{2/3}} \Pr[B_{n,1-zn^{-2/3}}] 2(1 - zn^{-2/3}) dz, \tag{10.23}$$

using the parametrization $r = 1 - s$ with $s = zn^{-2/3}$.

- Chip away at the easy regions using crude bounds.

Suppose P is very close to the boundary. There let us simply bound $\Pr[B_{n,r}] \le 1$. Effectively, we place all points near the boundary in the upper bound for the convex hull. Consider, say, the region $0 \le z \le 100$. The contribution to (10.23) of that region is that bounded by $n^{-2/3} \int_0^{100} 2dz = 200z^{-2/3}$. As our goal is an overall bound of $O(n^{-2/3})$, this is an acceptable amount.

Suppose P is fairly far from the boundary. Let O_1, O_2, O_3 be three points at distance $\frac{s}{2}$ with all angles $O_i P O_j$ being 120 degrees.

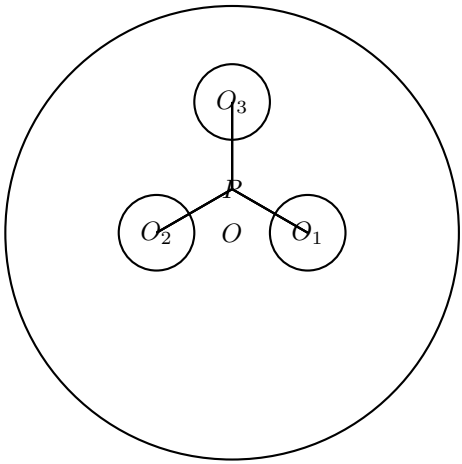


Figure 5. Visualizing the upper bound argument

Let D_1, D_2, D_3 be disks with centers O_1, O_2, O_3, respectively, with radius, say, $\frac{s}{10}$. (See Figure 5.)

These disks then lie entirely inside Ω. For any choice of points $Q_1 \in D_1$, $Q_2 \in D_2$, $Q_3 \in D_3$, P will lie inside the triangle $Q_1Q_2Q_3$. Thus for P to not be in the convex hull of $P_1, \ldots, P_{n-1}$, one of the D_1, D_2, D_3 would have to have none of the $P_1, \ldots, P_{n-1}$. Therefore, we can bound $\Pr[B_{n,r}]$ from above by $3(1 - \frac{s^2}{100})^{n-1}$. Suppose $s \geq n^{-0.49}$. This bound would then be $\exp[-\Omega(n^{0.02})]$, so the contribution to (10.23) would be $o(n^{-2/3})$ by a wide margin.

We are left with the "difficult" range $100n^{-2/3} \leq s \leq n^{-0.49}$ or, equivalently, $100 \leq z \leq n^{\frac{2}{3}-0.49}$. We know from our lower bound arguments (especially (10.17)) that S_P is very likely to have one of the P_i. Indeed, it is very likely to have several of the P_i. Our idea is that if S_P contains a P_i to the "left" of P and a P_j to the "right" of P, then the line segment P_iP_j will "undercut" P and then P would be very likely to lie in the convex hull of the $P_1, \ldots, P_{n-1}$. Here is one argument (out of many) that makes this explicit.

Recall we are asumming $P = (0, s-1)$ with $s = zn^{-2/3}$. As $s = o(1)$, we note $\sqrt{2s - s^2} \sim \sqrt{2s}$. Let R_1 denote the rectangle with width $\frac{1}{5}\sqrt{2s}$, height $\frac{1}{5}s$, and center $(-\frac{1}{2}\sqrt{2s}, \frac{1}{2}s - 1)$. Let R_2 denote the rectangle of the same dimensions with center $(+\frac{1}{2}\sqrt{2s}, \frac{1}{2}s - 1)$. Let

Q_1 be any point in R_1. The line Q_1P would then be nearly horizontal and so, when continued beyond P, would intersect the unit circle at (x, y) with y very close to -1, certainly $y \leq 0$. Similarly, let Q_2 be any point in R_2. The line Q_2P, when continued beyond P, would intersect the unit circle at (x, y) with $y \leq 0$. Thus if Q_3 is any point with $y \geq 0$, the triangle $Q_1Q_2Q_3$ would contain P. Therefore, the probability that P is not in the convex hull of $P_1, \ldots, P_{n-1}$ is at most the probability that either R_1 or R_2 or the semicircle $y \geq 0$ have none of the $P_1, \ldots, P_{n-1}$. R_1 has none of these points with probability $(1 - Ks^{3/2})^{n-1}$ where $K = \frac{1}{25}\sqrt{2}$. The same holds for R_2. The semicircle has half the area, so the probability that it contains none of the P_i is exceptionally small, 2^{1-n}, certainly smaller than the probability R_1 contains none of those points. Thus,

$$\Pr[B_{n,r}] \leq 3(1 - Ks^{3/2})^{n-1} \leq 3e^{-Kz^{3/2}\frac{n-1}{n}}. \tag{10.24}$$

Annoyingly, though not atypically, the term $\frac{n-1}{n}$ is not quite one. An easy way out is to assume $n \geq 2$ (we could even assume $n \geq 10^{10}$ if we wished as our object is the asymptotics of A_n) so that $\frac{n-1}{n} \geq \frac{1}{2}$. Set $K' = \frac{K}{2}$. Then

$$\Pr[B_{n,r}] \leq 3e^{-K'z^{3/2}}. \tag{10.25}$$

The contribution to (10.23) of this intermediate region is therefore at most

$$n^{-2/3} \int_{z=100}^{\infty} 3e^{-K'z^{3/2}} dz = O(n^{-2/3}), \tag{10.26}$$

as the integral is finite. The total of the three contributions to (10.23) is therefore $O(n^{-2/3})$, and so the expected number of vertices on the convex hull is $O(n^{1/3})$.

- Ugly proofs can be cleaned up.

Were all of the above arguments necessary? Not really. The argument for intermediate points can be applied directly to those points near the boundary. Indeed, with a bit more care, these arguments can be applied to points far from the boundary. It is important and proper to clean up a proof. But the first thing is to get a proof, and the clearing out of easy cases is often very helpful in that pursuit.

Chapter 11

Algorithms

> If creativity were anything but random, someone would have figured out the algorithm by now.
> – Scott Adams, *Dilbert*

Asymptopia is a natural setting in which to study the running times of algorithms. One describes running times of algorithms as $O(n^2)$ or $\Theta(n)$ or the ubiquitous $O(n \ln n)$. Generally, one has an algorithm with a parameter n (for example, sorting n objects), and one wants the time as a function of n. Moreover,[1] the interest is not in a particular n but in the rate of growth of the time as n grows. Time is an elusive concept; some computers are faster than others. One instead would like to count steps, but that is elusive as well. Whether a command such as $X \leftarrow X + 1$ should be counted as one or two steps is uncertain. Employing the $O, o, \Theta, \ldots$ language of Chapter 2 allows one to sweep the constants under the rug.

11.1. Recurrences

Let a, b be positive integers. Let f be a nonnegative function on the integers. Let c (less important) be a nonnegative integer. We examine the recurrence

$$T(n) = aT(n/b) + f(n) \text{ with initial condition } T(1) = c. \tag{11.1}$$

[1] at least on the theoretical side!

In applications, $T(n)$ will be the running time of an algorithm. The algorithm will be recursive, calling itself a times on data with parameter n/b. Further there will be other steps taking time $f(n)$. We refer to $f(n)$ as the *overhead*. We assume $f(n) \geq 0$. To avoid trivialities, we shall assume that either the initial value c or the overhead function $f(n)$ is not always zero. To avoid floors and ceilings, we will, for the moment, restrict n to be a power of b, $n = b^s$. $T(n)$ is then determined. Increasing the overhead function $f(n)$ can only increase the time function $T(n)$, but the relationship between the asymptotics of $f(n)$ and the asymptotics of $T(n)$ is an unexpected one.

A basic, and instructive, case is given by zero overhead. Consider the recurrence

$$T(n) = aT(n/b) \text{ with initial condition } T(1) = 1. \tag{11.2}$$

Now $T(b) = aT(1) = a$, $T(b^2) = aT(b) = a^2$, and, in general, $T(b^s) = a^s$. What is $T(n)$? With $n = b^s$,

$$a^s = (b^{\log_b a})^s = (b^s)^{\log_b a} = n^\gamma, \tag{11.3}$$

where we set

$$\gamma = \log_b a. \tag{11.4}$$

In some instances γ has a nice value. The recurrence $T(n) = 4T(n/2)$ with $T(1) = 1$ has solution $T(n) = n^2$. The recurrence $T(n) = 2T(n/2)$ with $T(1) = 1$ has solution $T(n) = n$. In other instances γ will be a not so nice real number. Note, though, that it will be a constant, independent of n.

Adding overhead $f(n)$ will increase $T(n)$. We analyze the general (11.1) by comparing $T(n)$ to the zero overhead solution (11.3). To that end we parametrize

$$S(n) = \frac{T(n)}{n^\gamma}. \tag{11.5}$$

Dividing (11.1) by n^γ gives

$$\frac{T(n)}{n^\gamma} = a\frac{T(n/b)}{n^\gamma} + \frac{f(n)}{n^\gamma}. \tag{11.6}$$

The left-hand side is simply $S(n)$. This special value of γ has $b^\gamma = a$ so that

$$S(n/b) = \frac{T(n/b)}{(n/b)^\gamma} = \frac{T(n/b)}{an^\gamma}, \tag{11.7}$$

and hence $S(n/b) = aT(n/b)n^{-\gamma}$. In terms of $S(n)$, (11.6) becomes

$$S(n) = S(n/b) + f(n)n^{-\gamma} \text{ with initial condition } S(1) = c. \tag{11.8}$$

Now we define

$$g(n) = f(n)n^{-\gamma} \tag{11.9}$$

and call $g(n)$ the *normalized overhead*. When $n = b^s$,

$$S(n) = \sum_{i=0}^{b-1} g(nb^{-i}) + c \tag{11.10}$$

so that

$$T(n) = n^\gamma \left[\sum_{i=0}^{s-1} g(nb^{-i}) + c \right]. \tag{11.11}$$

With important exceptions, the exact formula (11.11) is difficult to use because of the difficulty in evaluating the sum. In Asymptopia, however, sums are often approximated by their largest term. This depends on whether the normalized overhead is increasing or decreasing.

Theorem 11.1. *Let $T(n)$ be given by recursion* (11.1) *with γ and the normalized overhead $g(n)$ defined by* (11.4) *and* (11.9)*, respectively.*

(1) *If $\sum_{j=0}^{\infty} g(b^j)$ is finite, then $S(n) = \Theta(1)$ and $T(n) = \Theta(n^\gamma)$. We call this the* low overhead *regime. In particular, if there exists a positive $\epsilon > 0$ such that $g(n) = O(n^{-\epsilon})$, then $T(n) = \Theta(n^\gamma)$.*

(2) *If $g(n) = \Theta(1)$, then $S(n) = \Theta(\ln n)$ and $T(n) = \Theta(n^\gamma \ln n)$. We call this the* just right overhead *regime.*

(3) *If there exists a positive $\epsilon > 0$ such that $g(n) \geq (1+\epsilon)g(n/b)$ for all sufficiently large n, then $T(n) = \Theta(f(n))$. We call this the* high overhead *regime.*

Theorem 11.1 is strikingly simple to apply. If we are in the *low overhead* regime (for example, $f(n) = n^{\gamma-\kappa}$ for some positive κ),

then the solution, up to constant factor, is the same as if we had no overhead at all. If we are in the *high overhead* regime (for example, $f(n) = n^{\gamma+\kappa}$ for some positive κ), then the solution, up to constant factor, is the overhead. If we are in the *just right overhead* regime (for example, $f(n) = n^{\gamma}$), the process is more delicate and $T(n)$ has an extra logarithmic factor above the zero overhead case. There are examples of $f(n)$ which do not fall into any of these three regimes, but they are relatively rare. Examples of the high overhead regime generally lead to algorithms with poor running time and so are rarely considered. Our examples will come from the low overhead and just right overhead regimes.

In all three cases the arguments follow quickly from (11.10). In the low overhead case $\sum_{i=0}^{b-1} g(nb^{-i}) \leq \sum_{j=1}^{\infty} g(b^j)$ which is finite so $S(n)$ is bounded. Effectively, $g(b)$ is a positive proportion of the sum. In the high overhead case $\sum_{i=0}^{s-1} g(nb^{-i})$ is between $g(n)$ and $g(n)(1+\epsilon)/\epsilon$, and so is $\Theta(g(n))$. Effectively, $g(n)$ is a positive proportion of the sum. In the just right overhead case, all $g(nb^{-s}) = \Theta(1)$. There are $s = \log_b n = \Theta(\ln n)$ terms (as b is a constant), so $\sum_{i=0}^{s-1} g(nb^{-i}) = \Theta(\ln n)$. Effectively, all $g(nb^{-i})$ contribute to the sum.

What happens when n is not a power of b? Often times in a recursive algorithm one splits n into b parts which are all either $\lfloor n/b \rfloor$ or $\lceil n/b \rceil$. Suppose that in the recursion (11.1), the $aT(n/b)$ addend was replaced by any value between $aT(\lfloor n/b \rfloor)$ and $aT(\lceil n/b \rceil)$. For example, we might have the recursion $T(n) = T(\lfloor n/2 \rfloor) + T(\lceil n/2 \rceil) + n$. When $n = b^s$, we would get the solution of Theorem 11.1. The function $T(n)$, by a simple inductive argument, will be nondecreasing. Thus when $b^{s-1} \leq n \leq b^s$, we bound $T(b^{s-1}) \leq T(n) \leq T(b^s)$. In all cases of interest to us, the application of Theorem 11.1 will give $T(n) = \Theta(h(n))$ for some $h(n) = n^{\alpha} \ln^{\beta} n$. As b is a constant, $h(b^{s-1}), h(n), h(b^s)$ all lie within constants of each other. $T(n)$ is then sandwiched, and $T(n) = \Theta(h(n))$ as well.

11.2. Multiplying Large Numbers

We are given two large numbers $x, y < 2^n$, so that each has n or fewer binary[2] digits. We wish to find the product $z = xy$. We assume that multiplication of single digit numbers each take one step. We also assume that addition and subtraction of single digit numbers takes one step. From that, addition and subtraction of two n digit numbers takes $O(n)$ steps. (There is "carrying" in addition and "borrowing" in subtraction, but to ease the presentation we will assume that these operations are free.) We also assume that any multiplication of the form $w2^s$ is free. (Effectively it consists of moving the string w to the left s places.)

We all learned multiplication at a young age. Each digit in x is multiplied by each digit in y. This is $O(n^2)$ multiplications. Then we must add n numbers, each with n digits, which give $O(n^2)$ additions. The total number of steps is $O(n^2)$.

Surprisingly, there is a faster algorithm! We give Karatsuba's algorithm from [**KO**], which, while not the fastest way[3] to multiply large numbers, well illustrates the use of recurrences. Assume that n is a power of two. Considering x as a string of length n (possibly with zeroes on the left), split the string in half, giving values x_L, x_R each of $n/2$ digits. (For example, when $x = 11000111$, $x_L = 1100$, $x_R = 0111$.) Similarly, split y into y_L, y_R. Then

$$x = 2^{n/2}x_L + x_R \text{ and } y = 2^{n/2}y_L + y_R. \tag{11.12}$$

Karatsuba's Algorithm.

(1) Compute $x_L y_L$.

(2) Compute $x_R y_R$.

(3) (!) Add $x_L + x_R$.

(4) (!) Add $y_L + y_R$.

(5) (!!) Compute $(x_L + x_R)(y_L + y_R)$.

[2] Other bases are similar.

[3] The fastest algorithm is known as the *fast Fourier transform* and takes time $O(n \ln n)$.

(6) Subtract the first two products giving

$$(x_L + x_R)(y_L + y_R) - x_L y_L - x_R y_R = x_L y_R + x_R y_L. \tag{11.13}$$

(7) Multiply by $2^n, 2^{n/2}$ and add to give

$$z = xy = x_L y_L 2^n + (x_L y_R + x_R y_L)2^{n/2} + x_L x_R. \tag{11.14}$$

Karatsuba's algorithm is a recursive algorithm: the computations of products (other than by powers of two, which are free) are done recursively. Let $T(n)$ denote the total number of steps. The first two parts are products of $n/2$-digit numbers, and so each takes time $T(n/2)$. The additions in parts (3) and (4) each take $O(n)$. Part (5) has a technical aspect in that the sums may have $\frac{n}{2} + 1$ digits. In that case (other cases are easier) express $x_L + x_R = 2^{n/2} + v$ and $y_L + y_R = 2^{n/2} + w$. Their product is then $2^n + 2^{n/2}(v + w) + vw$. The product vw is done recursively, taking $T(n/2)$ steps. The sums each take $O(n)$ steps, so the total is $T(n/2) + O(n)$ steps. The two subtractions in part (6) each take $O(n)$ steps. The two additions in part (7) each take $O(n)$ steps. All the $O(n)$ terms add to $O(n)$. There were three (parts (1), (2), and (5)) recursive calls to Karatsuba's algorithm, taking $3T(n/2)$ steps. The recursion is then

$$T(n) = 3T(n/2) + O(n). \tag{11.15}$$

The initial value $T(1) = 1$ does not affect the result. Recursion (11.15) is in the low overhead regime of Theorem 11.1. Thus

$$T(n) = O(n^\gamma) \quad \text{with } \gamma = \log_2 3 = 1.5850\cdots. \tag{11.16}$$

11.3. Multiplying Large Matrices

We are given two large square matrices A, B, of the same size n by n. We wish to find the product $C = AB$. We assume that addition, multiplication, and subtraction of real numbers each take one step.

We all know how to multiply matrices. Each value of C is determined by the addition of n numbers, each given by a multiplication. There are n^2 entries and therefore $O(n^3)$ steps.

Surprisingly, there is a faster algorithm—the Strassen algorithm from [**Str69**]. Assume that n is a power of two. (When $2^{s-1} < n \le 2^s$, we can pad A, B with zeroes and this will not affect the asymptotic

result.) Split each matrix into four equal parts and multiply them accordingly.

Thus, let

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

and

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}.$$

Now, of course,

$$\begin{aligned} C_{11} &= A_{11}B_{11} + A_{12}B_{21}, \\ C_{12} &= A_{11}B_{12} + A_{12}B_{22}, \\ C_{21} &= A_{21}B_{11} + A_{22}B_{21}, \\ C_{22} &= A_{21}B_{12} + A_{22}B_{22}. \end{aligned}$$

To speed up computation, we define[4] the following:

$$\begin{aligned} X_1 &= (A_{11} + A_{22})(B_{11} + B_{22}), \\ X_2 &= (A_{21} + A_{22})B_{11}, \\ X_3 &= A_{11}(B_{12} - B_{22}), \\ X_4 &= A_{22}(B_{21} - B_{11}), \\ X_5 &= (A_{11} + A_{12})B_{22}, \\ X_6 &= (A_{21} - A_{11})(B_{11} + B_{12}), \\ X_7 &= (A_{12} - A_{22})(B_{21} + B_{22}). \end{aligned}$$

Now we notice that we can compute the matrix C faster:

$$\begin{aligned} C_{11} &= X_1 + X_4 - X_5 + X_7, \\ C_{12} &= X_3 + X_5, \\ C_{21} &= X_2 + X_4, \\ C_{22} &= X_1 - X_2 + X_3 + X_6. \end{aligned}$$

The number of operations required for this algorithm is

$$T(n) = 7T(n/2) + O(n^2). \tag{11.17}$$

[4]No, it is not at all obvious why these X_i are so defined—except that it works!

The first term represents the time needed to compute $X_1, \ldots, X_7$, which all involve multiplying matrices of size $(n/2) \times (n/2)$, and the second term represents the additions required for each application of the algorithm.

Recursion (11.17) is in the low overhead regime of Theorem 11.1. Thus

$$T(n) = O(n^{\log_2 7}). \tag{11.18}$$

11.4. Merge Sort

Sorting is both immensely practical and mathematically fascinating. The input is n real numbers $x_1, \ldots, x_n$. The output is an array of the same n numbers but in increasing order. We take as the basic step a *comparison*, determining which of x_i, x_j is smaller. Given an algorithm for sorting n numbers, $T(n)$ will denote the number of such comparisons made. A cautionary note: this does *not* always correspond to the time when the algorithm is implemented.

The heart of the Merge Sort algorithm is a merge algorithm. Here the input is two arrays $y_1, \ldots, y_m$ and $z_1, \ldots, z_m$ which are already sorted. That is, we assume $y_1 \leq y_2 \leq \cdots \leq y_m$ and $z_1 \leq z_2 \leq \cdots \leq z_m$. The output is an array $w_1 \leq w_2 \leq \cdots \leq w_{2m}$ of these $2m$ numbers in increasing order. The merge algorithm takes at most $2m - 1$ comparisons:

- Compare the smallest elements of the two arrays, y_1, z_1.
- Make the smallest w_1 and remove it from its array.
- Iterate.

We always take the smaller of the smallest elements from the remaining arrays and add it to the merged array, deleting it from its own array. When one array becomes empty, the remaining elements of the other array are added to the merged array. In the worst case this occurs when $y_1, \ldots, y_{m-1}$ and $z_1, \ldots, z_{m-1}$ have been placed in the merged array and then y_m, z_m are compared. Each comparison adds one element to the merged array, and the last element is free so the number of comparisons is at most $2m - 1$.

We define the Merge Sort when $n = 2^t$. The initial case is $t = 0$, $n = 1$, which requires zero comparisons. The input is a sequence $x_1, \ldots, x_n$ in arbitrary order.

Merge Sort.

(1) Sort $x_1, \ldots, x_{n/2}$ giving $y_1 \leq \cdots \leq y_{n/2}$.

(2) Sort $x_{n/2+1}, \ldots, x_n$ giving $z_1 \leq \cdots \leq z_{n/2}$.

(3) Merge $y_1, \ldots, y_{n/2}$, $z_1 \leq \cdots z_{n/2}$ giving $w_1 \leq \cdots \leq w_n$.

(4) Output $w_1 \leq \cdots \leq w_n$.

Merge Sort is a recursive algorithm; the sorting in parts one and two are done recursively. Let $T(n)$ denote the total number of comparisons. The first two parts each take $T(n/2)$ comparisons. Part (3), the merge, takes $2n - 1$ comparisons. Thus

$$T(n) = 2T(n/2) + n - 1 \tag{11.19}$$

with initial value $T(1) = 0$. Recursion (11.19) is in the just right overhead regime of Theorem 11.1. Thus

$$T(n) = O(n \ln n). \tag{11.20}$$

Following the proof of Theorem 11.1 gives a more precise result. As $\log_2 2 = 1$, we set $S(n) = T(n)/n$ so that (11.19) becomes

$$S(n) = S(n/2) + 1 - \frac{1}{n} \text{ with } S(1) = 0, \tag{11.21}$$

and (with $n = 2^t$)

$$T(n) = nS(n) = n \sum_{j=1}^{t} (1 - 2^{-j}). \tag{11.22}$$

The ones sum to $t = \lg n$. The 2^{-j} sum to $1 - 2^{-t} = 1 - n^{-1}$. Thus

$$T(n) = n \lg n - n + 1. \tag{11.23}$$

11.5. The Sorting Game

Paul and Carole[5] play a mathematical game with parameters n, q. Carole selects n distinct real numbers $x_1, \ldots, x_n$. There are q rounds. Each round Paul selects i, j and asks if $x_i < x_j$. Paul wins if at the

[5] anagram of Oracle!

end of the q rounds he knows the order of the $x_1, \ldots, x_n$. The values n, q are known to both Paul and Carole. The actual values $x_1, \ldots, x_n$ are immaterial, only their order is pertinent.

Carole, however, does not actually select the order in advance. Rather, she plays what is called an *adversary strategy*. She answers each question in a way that will best thwart Paul. Now the Sorting Game is a perfect information game. There are no draws so that with perfect play (for a given n, q) either Paul or Carole will win always.

Consider first a mathematical variant of the popular game Twenty Questions. Carole thinks of an integer x from 1 to n. There are q rounds. Each round Paul asks any Yes/No question about x. Paul wins if at the end of the q rounds he knows the number x. This game has a precise solution. If $n \leq 2^q$, then Paul has a strategy that wins. Each round he asks a question that splits the remaining possibilities as evenly as possible. After q rounds, regardless of Carole's responses, there can be only one possibility x remaining. If $n > 2^q$, then Carole has an adversary strategy that wins. At the beginning of each round, there is a set S of answers x that are still possible. Paul's question partitions S into S_Y, S_N, where S_Y is the set of $x \in S$ for which the answer would be Yes, and S_N the set of $x \in S$ for which the answer would be No. If $|S_Y| \geq |S_N|$ Carole responds Yes, otherwise Carole responds no. After each round, regardless of Paul's question, at least half of the x that were in S are still in S. After q rounds at least 2^{-q} of the x that were originally in S are still in S. The original S had size n so the final size has size at least $n2^{-q}$. With $n2^{-q} > 1$ the final S has at least two x and so Carole has won.

The adversary strategy used by Carole in Twenty Questions carries over[6] to the Sorting Game. At the beginning of each round, S is the set of orderings σ so that $x_{\sigma(1)} \leq \cdots \leq x_{\sigma(n)}$ is still viable—that is, that σ agrees with all previous questions and answers. As with Twenty Questions, Paul's question partitions S into S_Y, S_N, where S_Y is the set of $x \in S$ for which the answer would be Yes, and S_N the set of $x \in S$ for which the answer would be No. If $|S_Y| \geq |S_N|$ Carole responds Yes; otherwise, Carole responds No. After each round,

[6]Indeed, in any game in which Paul is to determine one of m possibilities via Yes/No questions, he needs at least $\lceil \lg m \rceil$ questions. This is frequently referred to as the *information-theoretic lower bound*.

regardless of Paul's question, at least half of the x that were in S are still in S. After q rounds at least 2^{-q} of the σ that were originally in S are still in S. The original S had size $n!$, so the final size has size at least $n!2^{-q}$. With $n!2^{-q} > 1$, the final S has at least two x, and so Carole has won.

Let $T(n)$ be the minimal q for which Paul wins the Sorting Game with parameters n, q. We think of $T(n)$ as the smallest number of comparisons needed to sort n objects. The adversary Carole is a personification of worst case. For Paul to win, Carole cannot win and so it is necessary that $n!2^{-q} \leq 1$. Hence,

Theorem 11.2. $T(n) \geq \lceil \lg(n!) \rceil$.

In Twenty Questions, Paul could always find a question that would split S into two equal (or off by one when $|S|$ was odd) parts. In the Sorting Game, however, Paul's questions are more restrictive. For a given set S of σ there might not be a question of the form "Is $x_i < x_j$?" that evenly splits S. Instead, we give a strategy for Paul that gives an upper bound on $T(n)$.

The strategy, or algorithm if you will, is called *Insertion Sort*. The values are $y_1, \ldots, y_{m-1}$ and z. Paul knows at the outset that $y_1 \leq y_2 \leq \cdots \leq y_{m-1}$. His goal is to order $y_1, \ldots, y_{m-1}$ and z. Equivalently, his goal is to place z into the ordering. The Insertion algorithm takes $\lceil \log_2 m \rceil$ comparisons. Suppose $m = 2^t$. Paul compares z to the median of the y's. If Carole says z is smaller, then Paul needs to place z in the ordering of the first $(m/2) - 1$ values y_i. If Carole says z is bigger, then Paul needs to place z in the ordering of the last $(m/2) - 1$ values y_i. Either way, Paul has halved the possible places for z so that with t comparisons the place for z is determined. For other m Paul can still employ this median strategy, choosing either of the two medians when there are an even number of y's.

The Insertion Sort begins with $y_1, \ldots, y_n$ in arbitrary order. For $j = 2$ to n the Insertion Algorithm is applied to $y_1, \ldots, y_{j-1}$ and y_j, rearranging so that now $y_1 \leq \cdots \leq y_j$.

The j-th step takes $\lceil \lg j \rceil$ steps. Hence

Theorem 11.3. $T(n) \leq \sum_{j=2}^{n} \lceil \lg(j) \rceil$.

As $\lceil \lg(j) \rceil < \lg(j) + 1$, we can combine bounds to give

Theorem 11.4. $\lg(n!) \leq T(n) \leq \lg(n!) + n - 1$.

Or, in the language of Asymptopia, $T(n) = \lg(n!) + O(n)$.

Caution: Our definition of step does not always correspond to time when the algorithm is implemented. Merge Sort can be efficiently implemented and is an $O(n \ln n)$ algorithm for sorting. Insertion Sort, however, does not have an efficient implementation.

11.6. Quicksort

Quicksort is a popular sorting algorithm, easy to implement and interesting to analyze. Like its cousin Merge Sort, Quicksort is a recursive algorithm, splitting the objects into two groups. Let S be the set of objects to be sorted. When S has zero or one elements, there is nothing to do. Otherwise, an element $x \in S$, called the *pivot*, is selected. x is compared to all other $y \in S$. These comparisons partition $S - \{x\}$ into L, those $y < x$, and R, those $y > x$. The two groups L, R are then sorted recursively. This gives the full ordering of S as all elements in L are to the left of x which is to the left of all elements of R.

The efficiency of Quicksort depends on the position of the pivot. A pivot x near the middle of S is best as then L, R will be around half the size of S. We select the pivot x randomly from S. Now Quicksort is a randomized algorithm,[7] and we let $T(n)$ denote the expected number of comparisons to sort n objects.

Let S have n elements, and let $a \in S$. Let T_a denote the expected number of comparisons during Quicksort on S in which a is *not* the pivot. As every comparison consists of one pivot and one nonpivot element,

$$T(n) = \sum_{a \in S} E[T_a]. \tag{11.24}$$

We give a coarse upper bound for $E[T_a]$. When $a \in U$ and Quicksort is applied to U, we call the choice x of pivot *good* if x is in the second or third quartile. As $x \in U$ is chosen uniformly, it is good with

[7] This variant is sometimes called *randomized Quicksort*.

probability roughly one-half. When x is good, both $|L| \leq 3n/4$ and $|R| \leq 3n/4$. That is, each good choice of pivot x is either a itself or $|U|$ goes down to at most $\frac{3}{4}|U|$. The latter can occur at most $\log_{4/3} n = O(\ln n)$ times. As choices are good with probability at least one-half, the expected number of bad pivots is $O(\ln n)$, and so $E[T_a] = O(\ln n)$. From (11.24),

$$T(n) = O(n \ln n). \tag{11.25}$$

We can be more precise. The initial pivot x on n elements is compared to all $n-1$ other y. When it has the i-th position amongst the elements, $|L| = i-1$ and $|R| = n-i$, so the expected number of further comparisons is $T(i-1)+T(n-1)$. As x is chosen uniformly, the value i is uniform in $1 \leq i \leq n$. Hence, we find the precise recursion

$$T(n) = n-1+\frac{1}{n}\sum_{i=1}^{n} T(i-1)+T(n-i) = n-1+\frac{2}{n}\sum_{i=1}^{n-1} T(i) \tag{11.26}$$

with initial conditions $T(0) = T(1) = 0$. We rewrite this as

$$nT(n) = n(n-1) + 2\sum_{i=1}^{n-1} T(i). \tag{11.27}$$

Replace n by $n-1$ to give

$$(n-1)T(n-1) = (n-1)(n-2) + 2\sum_{i=1}^{n-2} T(i). \tag{11.28}$$

Subtracting and combining terms gives the simpler recurrence

$$nT(n) = (n+1)T(n-1) + 2(n-1). \tag{11.29}$$

Dividing by $n(n+1)$ and setting

$$S(n) = \frac{T(n)}{n+1}, \tag{11.30}$$

the recurrence now becomes

$$S(n) = S(n-1) + \frac{2(n-1)}{n(n+1)} \tag{11.31}$$

with initial condition $S(0) = T(0)/2 = 0$. Decomposing the fraction into partial fractions,

$$S(n) = \sum_{i=1}^{n} \frac{2(i-1)}{i(i+1)} = \sum_{i=1}^{n} \left(\frac{4}{i+1} - \frac{2}{i} \right). \tag{11.32}$$

In terms of the harmonic number H_n defined in (4.18),

$$S(n) = 4(H_n + \frac{1}{n+1} - 1) - 2H_n = 2H_n + \frac{4}{n+1} - 4 \tag{11.33}$$

and

$$T(n) = (n+1)S(n) = 2(n+1)H_n - 4n. \tag{11.34}$$

The asymptotic formula (4.22) then yields

$$T(n) \sim 2n \ln n. \tag{11.35}$$

Chapter 12

Potpourri

> Deep in the human consciousness is a pervasive need for a logical universe that makes sense. But the real universe is always one step beyond logic.
> – Frank Herbert, *Dune*

12.1. The Law of the Iterated Logarithm

Let $X_i = \pm 1$ be independent random variables that take the values 1 and -1, each with probability $1/2$ and set, as usual, $S_n = X_1 + \cdots + X_n$. Here, however, X_i and S_n are defined for all i, n. We have already studied the distribution of S_n via the Central Limit Theorem (8.31), Chernoff bounds (Theorem 8.2) and the Binomial Tail (Theorem 5.11 and (5.58). The probability is that, say, $S_n \geq 6\sqrt{n}$ is quite small. But there are an infinite number of n, so it should not be surprising (and we shall prove in Theorem 12.4) that with probability one $S_n \geq 6\sqrt{n}$ for an infinite number of n. In this section we basically ask how exceptional an S_n are we likely to find. We look for $f(n)$ with two properties:

(1) For every $\epsilon > 0$, with probability one $S_n \geq f(n)(1-\epsilon)$ for infinitely many n.

(2) For every $\epsilon > 0$, with probability one $S_n \leq f(n)(1+\epsilon)$ for all but finitely many n.

The Law of the Iterated Logarithm is that the function

$$f(n) = \sqrt{2}\sqrt{\ln \ln n}\sqrt{n} \tag{12.1}$$

has the above properties. Rather than jumping to the answer, we will take a more leisurely approach, finding partial results using many of the techniques in Asymptopia.

12.1.1. Two Infinite Results. (*Caution*: We assume some familiarity with infinite probability spaces in this section.) In the first result we use that if $\Pr[E] \leq \epsilon$ for all $\epsilon > 0$, then $\Pr[E] = 0$. In the second result we use that if E_n is a countable sequence of events, each with zero probability, then their disjunciton has zero probability. The reader unfamiliar with these notions can simply assume Theorems 12.1 and 12.2 and continue to the next section.

Let A_n be events in a probability space, defined for all positive integers n. Set $p_n = \Pr[A_n]$. Let *INF* denote the event that infinitely many of the A_n occur. Two results determine $\Pr[INF]$ in many (not all!) cases. The first result is called the Borel–Cantelli lemma.

Theorem 12.1. *Suppose* $\sum_{n=1}^{\infty} p_n = K$ *is finite. Then* $\Pr[INF] = 0$.

Proof. Let B_m be the disjunction of the A_n, $n \geq m$. That is, B_m is the event that some A_n holds, $n \geq m$. For any $\epsilon > 0$, the convergence of $\sum_{n\geq 1} p_n$ to K implies the existence of an m such that $\sum_{n<m} p_n \geq K - \epsilon$ so that $\sum_{n\geq m} p_n \leq \epsilon$. The probability of the disjunct B_m is at most the sum of the probabilities so $\Pr[B_m] \leq \epsilon$. But $INF \subset B_m$, so $\Pr[INF] \leq \epsilon$. As this holds for all $\epsilon > 0$, $\Pr[INF] = 0$. □

Theorem 12.2. *Suppose* $\sum_{n=1}^{\infty} p_n$ *is infinite. Suppose further that the* A_i *are mutually independent events. Then* $\Pr[INF] = 1$.

Proof. For $m \leq M$, let $C_{m,M} = \bigwedge_{n=m}^{M} \overline{A_n}$. That is, $C_{m,M}$ is the event that no A_n holds, $m \leq n \leq M$. From the mutual independence of the A_n, $\Pr[C_{m,N}] = \prod_{m\leq n\leq M}(1 - p_n)$. Let $C_m = \bigwedge_{n\geq m} \overline{A_n}$, that no A_n holds, $n \geq m$. Fix m. For every K the divergence of $\sum p_n$ implies the existence of M so that $\sum_{m\leq n\leq K} p_n \geq K$. As $1 - x \leq e^{-x}$ for all $0 \leq x \leq 1$, for such K

$$\Pr[C_{m,N}] = \prod_{m=n}^{M} (1 - p_n) \leq \prod_{m=n}^{M} e^{-p_n} \leq e^{-K}. \tag{12.2}$$

As $C_m \subset C_{m,N}$, $\Pr[C_m] \le e^{-K}$. As K can be arbitrarily large, $\Pr[C_m] = 0$. The disjunction of a countable number of events, each with probability zero, has probability zero so $\Pr[\vee_{m\ge 1} C_m] = 0$. The complement of this event then has probability one. But the complement is that for all $m \ge 1$ there is some $n \ge m$ with A_n, and that is precisely the event *INF*. □

12.1.2. A Weak Upper Bound.

Theorem 12.3. *Set $f(n) = c\sqrt{n}\sqrt{\ln n}$ with $c > \sqrt{2}$. With probability one $S_n \le f(n)$ for all but finitely many n.*

Proof. Let A_n be the event $S_n > f(n)$. The Chernoff bound (Theorem 8.2) gives that

$$\Pr[A_n] \le e^{-f(n)^2/2n} = e^{-(c^2/2)\ln n} = n^{-c^2/2}. \tag{12.3}$$

With $c > \sqrt{2}$, $c^2/2 > 1$, and so $\sum \Pr[A_n]$ converges. By Theorem 12.1, the Borel–Cantelli lemma with probability one only a finite number of A_n hold. □

While Theorem 12.3 is correct, it is far from the $f(n)$ that we seek. When n, m are close together, the values S_n, S_m are highly correlated. To improve this in §12.1.6, we will take a carefully chosen increasing function $g(u)$ and examine S_n on the values $n = g(u)$. As there are fewer[1] values n the analogue of Theorem 12.3 will work for a smaller $f(n)$. Then we will need to argue that the values between $g(u-1)$ and $g(u)$ are also reasonably small. There is an interesting tradeoff here: the faster $g(u)$ increases, the fewer values $n = g(u)$ there are, and the smaller function $f(n)$ we can choose. However, the further apart $g(u-1), g(u)$ are, the more difficulty we have with the values $g(u-1) < n < g(u)$.

12.1.3. A Weak Lower Bound. Values S_n, S_m are highly correlated when n, m are close together and are quite weakly correlated when n, m are far apart. Set, for any $n < m$,

$$S_{n,m} = \sum_{i=n+1}^{m} X_i = S_m - S_n. \tag{12.4}$$

[1] The words "fewer" and "more" must be taken with a grain of salt as there will always be an infinite number of values n.

Let $g(u)$ be an increasing integer valued function, and set

$$W_u = S_{g(u-1),g(u)} = S_{g(u)} - S_{g(u-1)}. \tag{12.5}$$

W_u depends on X_i, $i \in (g(u-1), g(u)]$. These sets are disjoint over u and hence, critically, the W_u are mutually independent. We want information about $S_{g(u)}$ which differs from W_u by $S_{g(u-1)}$. By letting g grow quickly, we can ensure that $S_{g(u-1)}$ will have a negligible effect.

In this section we set $g(0) = 2$ and $g(u+1) = g(u)^2$ so that $g(u) = 2^{2^u}$.

Theorem 12.4. *Let λ be an arbitrarily large fixed real number. Set $f(n) = \lambda\sqrt{n}$. With probability one, $S_n \geq f(n)$ for infinitely many n.*

Proof. For $u \geq 1$, let A_u be the event that

$$W_u \geq (\lambda + 1.1)\sqrt{g(u) - g(u-1)}.$$

From the Central Limit Theorem (see §8.5)

$$\Pr[W_u] = \Pr[N \geq (\lambda + 1.1)] + o(1),$$

where N is the standard normal distribution. Taking any $p < \Pr[N \geq (\lambda + 1.1)]$, $\Pr[W_u] \geq p$ for all sufficiently large u. Thus $\sum_u \Pr[A_u]$ is infinite. The A_u, being dependent only on W_u, are mutually independent. From Theorem 12.2 with probability one A_u holds for infinitely many u. As $g(u) \sim g(u) - g(u-1)$, A_u implies $W_u \geq (\lambda+1)\sqrt{g(u)}$ for u sufficiently large. Tautologically, $|S_{g(u-1)}| \leq g(u-1) = \sqrt{g(u)}$. Thus $W_u \geq (\lambda+1)\sqrt{g(u)}$ implies

$$S_{g(u)} \geq (\lambda+1)\sqrt{g(u)} - \sqrt{g(u)} = \lambda\sqrt{g(u)}. \tag{12.6}$$

That is, $S_n \geq \lambda\sqrt{n}$ for infinitely many values $n = g(u)$. □

One can extend Theorem 12.4 to allow $\lambda = \lambda(n)$ to grow slowly to infinity.

Theorem 12.5. *Let $c < \sqrt{2}$, and let*

$$\lambda = \lambda(n) = c\sqrt{\ln\ln\ln n}. \tag{12.7}$$

Set $f(n) = \lambda(n)\sqrt{n}$. With probability one $S_n \geq f(n)$ for infinitely many n.

We outline the argument. Let A_u be the event

$$W_u \geq (\lambda(n) + 1.1)\sqrt{g(u) - g(u-1)}.$$

As $\lambda(n) \to \infty$, we cannot use the Central Limit Theorem directly. Our results on the Binomial Tail, especially Theorem 5.11, give that

$$\Pr[A_u] = e^{-(1+o(1))\lambda(n)^2/2} = (\ln\ln n)^{-(1+o(1))c^2/2}. \tag{12.8}$$

As $n = 2^{2^u}$, $\ln\ln(n) \sim u \ln 2$. Letting $c^2/2 = 1 - \delta$, $\Pr[A_u] = u^{\delta - 1 - o(1)}$. Once again $\sum_u \Pr[A_u]$ is infinite, and the remainder of the proof is as before.

While Theorem 12.5 is correct, it is far from the $f(n)$ that we seek. With $g(u) = g(u-1)^2$ we have assured ourselves that W_u is quite close to $S_{g(u)}$. But this has been overkill. Most of the time $S_{g(u-1)}$ will be nowhere near $-g(u-1)$. Again we have an interesting tradeoff in the choice of $g(u)$. The slower $g(u)$ grows, the more values $n = g(u)$ we have, and so we can get $W_u \geq f(n)$ infinitely often for a larger $f(n)$. However, the slower $g(u)$ grows, the more difficulty we have with the difference between W_u and $S_{g(u)}$. In §12.1.5 we will examine a function $g(u)$ that does well in both respects.

12.1.4. A Pretty Good Lower Bound.

Theorem 12.6. *Let $c < \sqrt{2}$, and let*

$$\lambda = \lambda(n) = c\sqrt{\ln\ln n}. \tag{12.9}$$

Set $f(n) = \lambda(n)\sqrt{n}/2$. With probability one $|S_n| \geq f(n)$ for infinitely many n.

Proof. Select c_1 with $c < c_1 < \sqrt{2}$. Let K be a very large integer. K will be dependent only on c, c_1 in ways we shall soon specify. For $u \geq 0$, set $g(u) = K^u$. That is, we have replaced the double exponential growth of $g(u)$ with singly exponential growth. As in (12.5) we set

$$W_u = S_{g(u-1),g(u)} = S_{g(u)} - S_{g(u-1)}. \tag{12.10}$$

As $g(u-1) = K^{-1}g(u)$, W_u has distribution S_m with $m = (1 - K^{-1})g(u)$. Let A_u be the event that $W_u \geq c_1\sqrt{n}\sqrt{\ln\ln n}$ with $n = g(u) = K^u$. Here $\sqrt{n} = (1 - K^{-1})^{-1/2}\sqrt{m}$. Our results on the Binomial Tail, Theorem 5.11, now give

$$\Pr[A_u] = e^{-(1+o(1))(1-K^{-1})^{-1}c_1^2(\ln\ln n)/2}. \tag{12.11}$$

As $c_1^2/2 < 1$, by selecting K sufficiently large $c_1^2(1-K^{-1})^{-1}/2 < 1$, so that

$$\Pr[A_u] = (\ln n)^{-1+\delta+o(1)} \tag{12.12}$$

for some positive δ. As $n = K^u$,

$$\Pr[A_u] = (u\ln(K))^{-1+\delta+o(1)}. \tag{12.13}$$

Regardless of the choice of constant K, $\sum_u \Pr[A_u]$ is infinite. As before the A_u are mutually independent. As before, from Theorem 12.2 with probability one, A_u holds for infintely many u. With $n = g(u) = K^u$ and A_u, we have $S_n - S_{n/u} \geq c_1\sqrt{n}\sqrt{\ln\ln n}$. For the moment we lose a factor of two and simply say that either $|S_n|$ or $|S_{n/u}|$ is at least half that size. □

12.1.5. A Very Good Lower Bound.

Theorem 12.7. *Let $c < \sqrt{2}$, and let*

$$\lambda = \lambda(n) = c\sqrt{\ln\ln n}. \tag{12.14}$$

Set $f(n) = \lambda(n)\sqrt{n}$. With probability one $|S_n| \geq f(n)$ for infinitely many n.

Proof. We gain the further factor of two of Theorem 12.7 by showing that it is unlikely that $S_{g(u-1)}$ will substantially affect the large W_u we have already guaranteed. Continuing the notation of Theorem 12.7, we let B_u be the event that

$$S_{g(u-1)} \leq -(c_1 - c)\sqrt{n}\sqrt{\ln\ln n} \tag{12.15}$$

with $n = g(u) = K^u$ so that $g(u-1) = K^{-1}n$. (This "gain" of a K factor is critical as it will turn up in the exponent for the large deviation!) The Chernoff bound of Theorem 8.2 bounds

$$\Pr[B_u] \leq \exp[-K(c_1-c)^2(\ln\ln n)] = (\ln n)^{-a} \tag{12.16}$$

with $a = K(c_1 - c_2)^2$. We select K sufficiently large so that $a > 1$. Now, as $n = K^u$,

$$\sum_u \Pr[B_u] = \sum_u (u\ln K)^{-a}, \tag{12.17}$$

which, regardless of K, is finite. From Theorem 12.1 with probability one, B_u occurs only finitely often.

When A_u occurs and B_u does not occur, setting $n = g(u) = K^u$, we have

$$S_{g(u)} \geq c_1\sqrt{n}\sqrt{\ln\ln n} - (c_1 - c)\sqrt{n}\sqrt{\ln\ln n} = c\sqrt{n}\sqrt{\ln\ln n}. \quad (12.18)$$ □

12.1.6. A Very Good Upper Bound. As a preliminary we use a beautiful and precise result called the Reflection Principle. Consider a walk of length m. More formally, let $X_i = \pm 1$ be independent, set $W_0 = 0$ and $W_i = W_{i-1} + X_i$ for $1 \leq i \leq m$. Set $MAX = \max_{0\leq i\leq m} W_i$, and let u be a positive integer.

Theorem 12.8. *With the above notations*

$$\Pr[MAX \geq u] = \Pr[W_m \geq u] + \Pr[W_m > u]. \quad (12.19)$$

Proof. The choices of $X_1, \ldots, X_m$ with $MAX \geq u$ fall into three disjoint categories:

(1) $W_m = u$,

(2) $W_m > u$,

(3) $W_m < u$ and $MAX \geq u$.

Take any choice of $X_1, \ldots, X_m$ in the second category. As the walk moves in steps of ± 1, there will be some (maybe many) s with $W_s = u$. Let s be the largest integer for which $W_s = u$. Now *reflect* the walk from s on. Formally, define $X'_1, \ldots, X'_m$ by $X'_i = X_i$, $1 \leq i \leq s$, and $X'_i = -X_i$, $s < i \leq m$. The walk $X'_1, \ldots, X'_m$ is in the third category. We can reverse this. Take any choice of $X'_1, \ldots, X'_m$ in the third category. As the walk moves in steps of ± 1, there will be some (maybe many) s with $W_s = u$. Let s be the largest integer for which $W_s = u$. Now *reflect* the walk from s on. Formally, define $X_1, \ldots, X_m$ by $X_i = X'_i$, $1 \leq i \leq s$, and $X_i = -X'_i$, $s < i \leq m$. The reflection gives a bijection between the walks in the second and third categories, and hence they must have equal cardinalities and (as all choices have the same probability 2^{-m}) equal probabilities. Thus,

$$\begin{aligned} \Pr[MAX \geq u] &= \Pr[W_m = u] + 2\Pr[W_m > u] \\ &= \Pr[W_m > u] + \Pr[W_m \geq u]. \end{aligned} \quad (12.20)$$ □

The Chernoff bound of Theorem 8.2 then gives

$$\Pr[MAX \geq u] \leq 2\Pr[W_m \geq u] \leq 2e^{-m^2/2u}. \tag{12.21}$$

While this bound can be improved, it will suffice for our purposes.

Theorem 12.9. *Let $c > \sqrt{2}$, and let*

$$\lambda = \lambda(n) = c\sqrt{\ln\ln n}. \tag{12.22}$$

Set $f(n) = \lambda(n)\sqrt{n}$. With probability one $|S_n| \geq f(n)$ for only finitely many n.

Proof. Select c_1, c_2 with $c > c_2 > c_1 > \sqrt{2}$. As in the upper bound argument of Theorem 12.7, we shall split the integers into intervals $(g(u-1), g(u)]$. Now, however, we shall let α be a real number barely (as made explicit soon) larger than one, and we set $g(u) = \lfloor \alpha^u \rfloor$. For $u \geq 0$, let B_u be the event

$$S_n \geq c_1\sqrt{n}\sqrt{\ln\ln n} \tag{12.23}$$

with $n = g(u)$. The Chernoff bound of Theorem 8.2 bounds

$$\Pr[B_u] \leq \exp[-c_1^2(\ln\ln n)/2] = (\ln n)^{-a} \tag{12.24}$$

with $a = c_1^2/2 > 1$. As $\ln n = u\ln(\alpha) + o(1)$,

$$\sum_u \Pr[B_u] \leq (1 + o(1)) \sum_u (u\ln(\alpha))^{-a}, \tag{12.25}$$

which is finite. The Borel–Cantelli lemma (Theorem 12.1) gives that with probability one, B_u fails for all but finitely many u.

Now we must examine where the walk goes between $g(u-1)$ and $g(u)$. Let C_u be the event

$$S_n - S_{g(u-1)} \geq (c_2 - c_1)\sqrt{g(u)}\sqrt{\ln\ln g(u)} \quad \text{for some } g(u-1) \leq n \leq g(u). \tag{12.26}$$

We have a walk $W_n = S_n - S_{g(u-1)}$ of length $g(u) - g(u-1) = (\alpha - 1)g(u) + O(1)$, where the $O(1)$ refers only to the round-offs of α^{u-1}, α^u. Applying the Reflection Principle in the form (12.18) and setting $m = g(u)$,

$$\Pr[C_u] \leq 2\exp\left[-\frac{(c_2 - c_1)^2(\ln\ln m)}{2(\alpha - 1)^2}\right] = 2(\ln m)^{-a} \tag{12.27}$$

with $a = (c_2 - c_1)^2(\alpha - 1)^{-2}/2$. We select α just barely above one so that $a > 1$. Now $\Pr[C_u] = O((\ln m)^{-a}) = O(u^{-a})$. With $a > 1$, $\sum \Pr[C_u]$ converges and by the Borel–Cantelli lemma (Theorem 12.1) with probability one, C_u fails for all but finitely many u.

For all u for which B_u and C_u fail and all $g(u-1) \le n \le g(u)$,

(12.28)
$$S_n \le c_1\sqrt{m}\sqrt{\ln\ln m} + (c_2 - c_1)\sqrt{m}\sqrt{\ln\ln m} \le c_2\sqrt{m}\sqrt{\ln\ln m}$$

with $m = g(u)$. Finally, we select α just barely above one so that $c_2\alpha^{1/2} < c$. Then as $n \ge \alpha^{-1}m$ so that for u sufficiently large $c_2\sqrt{m}\sqrt{\ln\ln m} \le c\sqrt{n}\sqrt{\ln\ln n}$. As B_u and C_u fail for only finitely many u, all but finitely many n then have the property that

(12.29)
$$S_n \le c\sqrt{n}\sqrt{\ln\ln n},$$

as desired. □

12.1.7. Reflections. The Law of the Iterated Logarithm is one of the most celebrated results in mathematics. There is a naturalness to the question, We know that a random walk tends to be on the order of $\sqrt{n}$ from the origin at time n, but just how much more than that will it be in exceptional cases? It has a remarkably precise answer, including the constant $\sqrt{2}$. It has the iterated logarithm, $\ln\ln n$, which mathematicians always find very appealing.

Was it pure serendipity that the upper and lower bounds matched so well? We think not. At the heart is the estimate of $\Pr[S_n \ge \alpha\sqrt{n}]$ by $\exp[-\alpha^2/2]$. When α is increased by a factor $1+\epsilon$, it results in the estimate being changed by a $(1+\epsilon)^2$ factor *in the exponent*. Because of the Borel–Cantelli lemma, one examined whether various sums of the probabilities are finite and this involves sums of terms of the form $u^{-\beta}$. The value $\beta = 1$ is the critical one, the knife-edge between convergence and divergence, between appearance infinitely often and appearance only finitely often.

12.2. The Amazing Poisson Distribution

12.2.1. Inclusion-Exclusion.

A Break! No asymptotics in this section!

Let $A_1, \ldots, A_n$ be any events in a finite probability space Ω. Let $S_0 = 1$, and for $1 \le k \le n$ let

$$S_k = \sum_{\{i_1,\ldots,i_k\}} \Pr[A_{i_1} \wedge A_{i_2} \wedge \cdots \wedge A_{i_k}]. \tag{12.30}$$

That is, S_k is the sum of all the probabilities of the conjunctions of precisely k of the events. In particular $S_1 = \sum_{i=1}^{n} \Pr[A_i]$ is the sum of the probabilities of the events. The following result, known as the *Inclusion-Exclusion Principle*, is widely used.

Theorem 12.10. *With A_i, S_k as above,*

$$\Pr[\bigwedge_{i=1}^{n} \overline{A_i}] = S_0 - S_1 + S_2 - \cdots \pm S_n = \sum_{k=0}^{n} (-1)^k S_k. \tag{12.31}$$

Proof. Let $u \in \Omega$, and suppose there are precisely r events $A_{j_1}, \ldots, A_{j_r}$ which u satisfies. Then $\Pr[u]$ will appear as an addend precisely $\binom{r}{k}$ times in S_k (that is, for all $\{i_1, \ldots, i_k\} \subseteq \{j_1, \ldots, j_r\}$). This includes the case $k = 0$, as $S_0 = 1 = \sum_u \Pr[u]$. When $k > r$, $\binom{r}{k} = 0$, and $\Pr[u]$ does not appear. When $r = 0$ (that is, u satisfies none of the A_i) $\Pr[u]$ appears only in S_0 and so with a weight of one on the right-hand side. When $r > 0$, it appears with a weight of

$$\sum_{k=0}^{n} (-1)^k \binom{r}{k} = \sum_{k=0}^{r} (-1)^k \binom{r}{k} = 0. \tag{12.32}$$

Thus the right-hand side is $\sum \Pr[u]$ over those u for which $r = 0$, which is precisely the left-hand side. □

The condition that Ω is finite is not necessary. Indeed, any n events $A_1, \ldots, A_n$ have a "Venn diagram" with at most 2^n parts and Ω can effectively be replaced by a set of size at most 2^n.

12.2.2. The Bonferroni Inequalities.

Still no asymptotics!

The Inclusion-Exclusion Principle is too precise to be of much value in Asymptopia. As k gets large, the values S_k often become difficult to estimate, much less compute exactly. However, one can

sometimes estimate S_k quite well for fixed k. The Bonferroni inequalities state that the alternating sum for the Inclusion-Exclusion Principle alternately overestimates and underestimates the actual value. More precisely,

Theorem 12.11. *With A_i, S_k as in §12.2.1, for t even,*

$$\Pr[\bigwedge_{i=1}^{n} \overline{A_i}] \leq S_0 - S_1 + S_2 - \cdots + S_t = \sum_{k=0}^{t} (-1)^k S_k, \tag{12.33}$$

while for t odd

$$\Pr[\bigwedge_{i=1}^{n} \overline{A_i}] \geq S_0 - S_1 + S_2 - \cdots - S_r = \sum_{k=0}^{t} (-1)^k S_k. \tag{12.34}$$

Proof. Set

$$f_t(r) = \sum_{k=0}^{t} (-1)^k \binom{r}{k} = 1 - \binom{r}{1} + \binom{r}{2} - \cdots \pm \binom{r}{t}. \tag{12.35}$$

We claim that for t even, $f_t(r) \geq 0$ for all positive r and for t odd $f_t(r) \leq 0$ for all positive r. For $t = 1$, $f_1(r) = 1 - r$ and this is clear. We use a double induction, first on t and then on r. Assume the result for $t - 1$ and all r. For $r \leq t$, $f_t(r)$ is the alternating sum and so is zero. We apply the identity $\binom{r+1}{t} = \binom{r}{t} + \binom{r}{t-1}$. Hence,

$$f_t(r+1) = \sum_{k=0}^{t} (1)^k \binom{r}{k} - \sum_{k=1}^{t} (-1)^{k-1} \binom{r}{k-1} = f_t(r) - f_{t-1}(r). \tag{12.36}$$

Suppose t is even. By induction, all $f_{t-1}(r) \leq 0$. As $f_t(t) = 0$, $f_t(r) \geq 0$ for all $r > t$. Similarly, suppose t is odd. By induction, all $f_{t-1}(r) \geq 0$. As $f_t(t) = 0$, now $f_t(r) \leq 0$ for all $r > t$, completing the claim.

Now we apply the argument used in the proof of Theorem 12.10. Let $u \in \Omega$ satisfy precisely r of the A_i. When $r = 0$, the addend $\Pr[u]$ appears once in both $\Pr[\bigwedge \overline{A_i}]$ and S_0. When $r > 0$, the addend $\Pr[u]$ appears $f_t(r)$ times in the right-hand sum of (12.33) and (12.34). For t even, this gives a positive contribution and hence (12.33), while for t negative this gives a negative contribution and hence (12.34). □

12.2.3. The Poisson Paradigm. The Poisson distribution is one of the most basic and interesting distributions in probability. Let μ be an arbitrary positive real. The Poisson distribution of mean μ, denoted by P_μ, is a distribution over the nonnegative integers given by the formula

$$\Pr[P_\mu = i] = e^{-\mu}\frac{\mu^i}{i!}. \tag{12.37}$$

A simple calculation gives that P_μ does indeed have mean μ. The formal definition masks the character of the Poisson distribution. Informally, suppose we are given a large number of events, and let X be the number of those events that occur. Suppose further that none of the events are very likely but that the expected number of events occurring is approximately μ, which is in the constant range. Suppose further that the events are either mutually independent or nearly mutually independent. The *Poisson Paradigm* then states[2] that X will have distribution roughly P_μ.

A standard example is $BIN[n, \frac{\mu}{n}]$, the number of heads in n coin flips, where the probability of heads is $\frac{\mu}{n}$. With μ fixed and $n \to \infty$,

$$\lim_{n\to\infty} \Pr[BIN[n, \frac{\mu}{n}] = i] = \lim_{n\to\infty} \binom{n}{i} \left(\frac{\mu}{n}\right)^i \left(1 - \frac{\mu}{n}\right)^{n-i} = e^{-\mu}\frac{\mu^i}{i!}. \tag{12.38}$$

Theorem 12.12. *For each n, let $A_1^{(n)}, \ldots A_n^{(n)}$ be n events. For $0 \le k \le n$, let $S_k^{(n)}$ be given by* (12.30)*. Let μ be a positive constant, independent of n. Suppose that for each positive k,*

$$\lim_{n\to\infty} S_k^{(n)} = \frac{\mu^k}{k!}. \tag{12.39}$$

Then

$$\lim_{n\to\infty} \Pr\left[\bigwedge_{i=1}^{n} \overline{A_i^{(n)}}\right] = e^{-\mu}. \tag{12.40}$$

It is instructive[3] to give an *incorrect* argument. From Theorem 12.10 the desired probability is $\sum_{k=0}^{n}(-1)^k S_k^{(n)}$. Each $S_k \to \mu^k/k!$, so that sum approaches $\sum_{k=0}^{\infty}(-1)^k\mu^k/k! = e^{-\mu}$. **WRONG!** The

[2]Being a paradigm, the terms are not tightly defined.
[3]though pedagogically dangerous!

limit of an infinite sum is *not necessarily* the sum of the limits. For example, let $f_n(i) = \frac{1}{n}$ for $1 \le i \le n$ and $f_n(i) = 0$ for $i > n$. For each i, $\lim_{n\to\infty} f_n(i) = 0$ so that $\sum_{i=1}^{\infty} \lim_{n\to\infty} f_n(i) = 0$. For each n, $\sum_{i=1}^{\infty} f_n(i) = n(1/n) = 1$ and $\lim_{n\to\infty} \sum_{i=1}^{\infty} f_n(i) = 1$. Two different answers! Indeed, the Inclusion-Exclusion Principle is not sufficient, the correct argument below uses the more powerful Bonferroni inequalities.

We also note informally why (12.39) occurs in many situations. Let us restrict to the symmetric case where all $\Pr[A_i] = p$ and $\mu \sim np$. When the events $A_1, \ldots, A_n$ are close to being independent, the k-fold conjunctions $A_{i_1} \wedge \cdots \wedge A_{i_k}$ all have probability near p^k. There are $\binom{n}{k} \sim n^k/k!$ addends, so it is natural to expect (but still requires proof!) that $S_k \sim (n^k/k!)p^k \sim \mu^k/k!$.

Proof. Let $\epsilon > 0$ be arbitrarily small but fixed. As μ (important!) is a constant and $e^{-\mu}$ has the infinite sum $\sum_{k=0}^{\infty}(-1)^k\mu^k/k!$, the sum will eventually stay within ϵ of its limit. Let t be an even number (for convenience) such that

$$\left|\sum_{k=0}^{t'}(-1)^k\frac{\mu^k}{k!} - e^{-\mu}\right| \le \frac{\epsilon}{2} \tag{12.41}$$

for $t' = t$ and $t' = t+1$. Now consider the *finite* number of sequences $S_0^{(n)}, \ldots, S_t^{(n)}$. For each $0 \le k \le t$, $S_k^{(n)} \to \mu^k/k!$. This implies there is an n_k so that for $n \ge n_k$,

$$\left|S_k^{(n)} - \frac{\mu^k}{k!}\right| \le \frac{\epsilon}{2(t+1)}. \tag{12.42}$$

Let N be the maximum of $n_0, \ldots, n_t$. For $n \ge N$, (12.42) holds for all $0 \le i \le t$ simultaneously. Alternately adding and subtracting,

$$\left|\sum_{k=0}^{t}(-1)^k S_k^{(n)} - \sum_{k=0}^{t}(-1)^k\frac{\mu^k}{k!}\right| \le \sum_{k=0}^{t}\frac{\epsilon}{2(t+1)} = \frac{\epsilon}{2}. \tag{12.43}$$

Combining (12.41) and (12.43) gives

$$\left|\sum_{k=0}^{t}(-1)^k S_k^{(n)} - e^{-\mu}\right| \le \epsilon. \tag{12.44}$$

As t was even, the Bonferroni inequality (12.33) gives the *upper bound*

$$\Pr[\bigwedge_{i=1}^{n} \overline{A_i^{(n)}}] \le \sum_{k=0}^{t} (-1)^k S_k^{(n)} \le e^{-\mu} + \epsilon \tag{12.45}$$

for n sufficiently large. We now apply the same argument, replacing t by the odd $t+1$, giving the *lower bound*

$$\Pr[\bigwedge_{i=1}^{n} \overline{A_i^{(n)}}] \ge \sum_{k=0}^{t+1} (-1)^k S_k^{(n)} \ge e^{-\mu} - \epsilon. \tag{12.46}$$

The desired probability is, for arbitrarily small ϵ, eventually sandwiched between $e^{-\mu} - \epsilon$ and $e^{-\mu} + \epsilon$. Hence it must approach $e^{-\mu}$. □

While Theorem 12.12 only gives the probability that none of the events hold, it quickly generalizes into the probability that precisely u of the events hold.

Theorem 12.13. *Using the notation of Theorem* 12.12*, assume* (12.39) *and* (12.40)*. Assume further that*

$$\lim_{n\to\infty} \max_i \Pr[A_i^{(n)}] = 0. \tag{12.47}$$

Let $X^{(n)}$ denote the distribution of the number of $A_1^{(n)}, \ldots, A_n^{(n)}$ holding. Then $X^{(n)}$ approaches the Poisson distribution with mean μ. That is, for every fixed u,

$$\lim_{n\to\infty} \Pr[X^{(n)} = u] = \frac{\mu^u}{u!}. \tag{12.48}$$

Proof. Fix u and consider $\Pr[X^{(n)} = u]$. Let $\{i_1, \ldots, i_u\} \subset \{1, \ldots, n\}$. The probability that no A_j, $1 \le j \le n$, holds is $e^{-\mu} + o(1)$ by Theorem 12.12. The probability that any of $A_{i_1}, \ldots, A_{i_u}$ hold is $o(1)$ by assumption (12.47). Hence, the probability that no A_j, $j \ne i_1, \ldots, i_u$, holds is still $e^{-\mu} + o(1)$. Now sum, over all $\{i_1, \ldots, i_u\} \subset \{1, \ldots, n\}$ the probability that $A_{i_1}, \ldots, A_{i_u}$ hold times the probability that no other A_j holds. The second term is always $e^{-\mu} + o(1)$. The sum of the first terms is $S_u^{(n)} = (\mu^k/k!) + o(1)$. Hence, the sum is $e^{-\mu}\mu^k/k! + o(1)$, as desired. □

12.3. The Coupon Collector Problem

There are n coupon types, call them $1, \ldots, n$. For each time unit you receive a coupon that is randomly chosen from 1 to n. How long will it be until you receive at least one of each coupon type?

12.3.1. The Expected Value. Let T_i be the number of time units between the moment you have $i-1$ different coupon types and the moment you have i different coupons types. Then $T = \sum_{i=1}^{n} T_i$ is the total time to receive all the coupon types. Set $p = (n-i)/n$, which is the probability that a coupon received is of a new type. Then $\Pr[T_i = u] = (1-p)^{u-1}p$ as $u-1$ times you receive a coupon of a type already received and then you receive a coupon of a new type. T_i has what is called the *geometric distribution* with parameter p. As a classical result,

$$E[T_i] = \sum_{u=1}^{\infty} pu(1-p)^{u-1} = \frac{1}{p}. \tag{12.49}$$

Here $\frac{1}{p} = \frac{n}{n-i}$. Linearity of expectation then gives the exact value

$$E[T] = \sum_{i=1}^{n} E[T_i] = \sum_{i=1}^{n} \frac{n}{n-i} = n \sum_{j=1}^{n} \frac{1}{j} = nH_n, \tag{12.50}$$

where H_n is the harmonic number discussed in §4.2. The asymptotic formula (4.30) for H_n then gives

$$E[T] = n \ln n + n\gamma + o(n), \tag{12.51}$$

where $\gamma = 0.577\cdots$ is Euler's constant.

12.3.2. The Fine Behavior. Suppose that m coupons have been received, each randomly chosen from coupon types $1, \ldots, n$. Let $p(m,n)$ denote the probability that every coupon type has been received. Clearly, for n fixed, $p(m,n)$ is an increasing function of m which is 0 for $m < n$ and approaches 1 as $m \to \infty$. In Asymptopia we search for the asymptotic parametrization of $m = m(n)$ so that we can "see" $p(m,n)$ going from near one to near zero.

Given n, m, let A_i, $1 \le i \le n$, be the event that no coupon of type i has been received. Then $\Pr[A_i] = (1-\frac{1}{n})^m$. Set $\mu(m,n) = n(1-\frac{1}{n})^m$

so that $\mu(m,n)$ is the expected number of coupon types that have not been received.

We look for a parametrization of $m = m(n)$ for which μ varies through positive constant values. Let's approximate $(1-\frac{1}{n})^m$ by $e^{-m/n}$. Then $\mu(m,n)$ would be approximated by $ne^{-m/n}$. Setting $m = n\ln n$ would give μ approximately 1. Adding or subtracting a constant to m/n would multiply $ne^{-m/n}$ by a constant. This leads us to the parametrization

$$m = m(n) = n\ln n + cn + o(n). \tag{12.52}$$

For c any real constant we then have

$$\mu(m,n) \sim ne^{-m/n} = \mu + o(1) \tag{12.53}$$

with $\mu = e^{-c}$.

Wait a minute! While this looks right, we now have to check that secondary terms are negligible. Let $m = m(n)$ be given by (12.52). From the Taylor series with error (2.35), $\ln(1-\frac{1}{n}) = -\frac{1}{n} - O(n^{-2})$. Thus $m\ln(1-\frac{1}{n}) = -\frac{m}{n} + o(1)$ as $O(mn^{-2}) = o(1)$ in this range. (Note that this would *not* be accurate if, say, $m = n^2$.) The $o(1)$ addend becomes a $1 + o(1)$ factor upon exponentiation and, indeed, $n(1-\frac{1}{n})^m \sim ne^{-m/n}$.

Now let k be an arbitrary fixed positive integer, and let S_k be given by (12.30). For any $i_1, \ldots, i_k$ the probability that none of those k coupon types is received is $(1-\frac{k}{n})^m$. Thus

$$S_k = \binom{n}{k}\left(1-\frac{k}{n}\right)^m. \tag{12.54}$$

With $m = m(n)$ given by (12.52) the analysis above[4] gives $(1-\frac{k}{n})^m \sim e^{-km/n}$ so that

$$S_k \sim \frac{n^k}{k!}(e^{-m/n})^k \sim \frac{(ne^{-m/n})^k}{k!} \sim \frac{\mu^k}{k!}. \tag{12.55}$$

The conditions for Theorem 12.12 are met, so the Poisson paradigm does hold. We state the result in particularly striking form:

[4]Critically, k is fixed and $n \to \infty$.

Theorem 12.14. *Let m coupons each be uniformly and independently chosen from n coupon types. Let $p(m,n)$ denote the probability that at least one of every coupon type has been chosen. Let $m = m(n) = \frac{\ln n}{n} + \frac{c}{n} + o(\frac{1}{n})$. Then*

$$\lim_{n\to\infty} p(m,n) = e^{-e^{-c}}. \tag{12.56}$$

12.4. The Threshold of Connectivity

12.4.1. The Erdős–Rényi Random Graph. The random graph $G(n,p)$ is a probability space over graphs on vertex set $V = \{1,\ldots,n\}$. Informally, each unordered pair of vertices $\{i,j\}$ flips a coin to decide if i,j are adjacent. The coin comes up heads with probability p. We require n a positive integer and $0 \le p \le 1$.

For $p = 0$, $G(n,p)$ has no edges while for $p = 1$ it has all $\binom{n}{2}$ edges. As p goes from 0 to 1 the random graph *evolves* from empty to full. Critically, we consider p as a function of n, $p = p(n)$. For many natural properties A of graphs, Erdős and Rényi found a parametrization $p = p(n)$ in [**ER59**] for which $\Pr[A]$ moved from asymptotically zero to asymptotically one. None was so striking as their results on connectivity; see Theorem 12.16 below.

For $1 \le i \le n$, let A_i be the event that vertex i is isolated, that there are no edges $\{i,j\} \in G$. Let $NOI = \bigwedge_{i=1}^{n} \overline{A_i}$, the event that there are no isolated vertices.

Theorem 12.15. *Let $p = p(n)$ satisfy*

$$p = \frac{\ln n}{n} + \frac{c}{n} + o(n). \tag{12.57}$$

Then[5]

$$\lim_{n\to\infty} \Pr[G(n,p(n)) \models NOI] = e^{-e^{-c}}. \tag{12.58}$$

The proof is quite similar to the Coupon Collector result, Theorem 12.14.

[5]The notation $G \models A$ is read "G models A" and denotes the event that random structure G has property A.

Proof. For each i

$$\Pr[A_i] = (1-p)^{n-1} \sim e^{-pn} \sim \frac{\mu}{n} \tag{12.59}$$

with $\mu = e^{-c}$. For $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$,

$$\Pr[A_{i_1} \wedge \cdots \wedge A_{i_k}] = (1-p)^{k(n-1)-K}, \tag{12.60}$$

where we set $K = \binom{k}{2}$. (For each of the $i_1, \ldots, i_k$, there are $n-1$ non-edges, but the K internal edges $\{i_r, i_s\}$ have been counted twice.) For k fixed, K is fixed and, as $1-p \sim 1$, $(1-p)^K \sim 1$. Informally, the dependence between the A_{i_r} is asymptotically negligible. With S_k as in (12.30),

$$S_k = \binom{n}{k}(1-p)^{k(n-1)-K} \sim \frac{(n(1-p)^{n-1})^k}{k!} \sim \frac{\mu^k}{k!}, \tag{12.61}$$

so that the conditions for the Poisson paradigm, Theorem 12.12, are met and

$$\lim_{n\to\infty} \Pr[\bigwedge_{i=1}^{n} \overline{A_i}] = e^{-\mu} = e^{-e^{-c}}. \tag{12.62}$$

□

Let CON be the event that G is connected.

Theorem 12.16. *Let $p = p(n)$ satisfy*

$$p = \frac{\ln n}{n} + \frac{c}{n} + o(n). \tag{12.63}$$

Then

$$\lim_{n\to\infty} \Pr[G(n, p(n)) \models CON] = e^{-e^{-c}}. \tag{12.64}$$

Proof. The properties CON and NOI are similar but not the same. If G is connected, then it tautologically must have no isolated vertices. Thus $\Pr[CON] \geq \Pr[NOI]$. However, G may have no isolated vertices and still not be connected. To show Theorem 12.16 from Theorem 12.15, we will show that this occurs with probability $o(1)$. If G has no isolated vertices and is not connected, then G must have a component of some size r, $2 \leq r \leq \lfloor \frac{n}{2} \rfloor$. Let $F(n, p, r)$ denote the expected

number of such components in $G(n,p)$. $F(n,p,r)$ is an upper bound on the probability of having any such component. Then

$$\Pr[G(n,p) \models \mathit{NOI}] - \Pr[G(n,p) \models \mathit{CON}] \le \sum_{r=2}^{\lfloor n/2 \rfloor} F(n,p,r). \tag{12.65}$$

Our goal is to show that the right-hand sum above is $o(1)$. The case $r = 2$ is special. We choose the two vertices, require them to be adjacent, and require no adjacencies between them and the other vertices. Thus

$$F(n,p,2) = \binom{n}{2} p(1-p)^{2(n-2)}. \tag{12.66}$$

As $\binom{n}{2} \sim n^2/2$ and $(1-p)^{-2} \sim 1$,

$$F(n,p,2) \sim \frac{n^2 p}{2}(1-p)^{2(n-1)} \sim \frac{p}{2}(n(1-p)^{n-1})^2 \sim \frac{p}{2}\mu. \tag{12.67}$$

As $\mu = e^{-c}$ is constant, $F(n,p,2) = O(p) = o(1)$. For general r we choose a set S of r vertices, choose (using Cayley's formula (6.3)) a tree on S, require all the edges in the tree to be adjacent, and require no adjacencies between S and its complement. We may have further adjacencies inside S. Graphs for which S contains more than one tree are multiply counted, so we obtain an upper bound for $F(n,p,r)$:

$$F(n,p,r) \le \binom{n}{r} r^{r-2} p^{r-1} (1-p)^{r(n-r)}. \tag{12.68}$$

We bound $\binom{n}{r} \le (ne/r)^r$ (5.14), $r^{r-2} \le r^r$ (for convenience), $1 - p \le e^{-p}$ and write $p^{r-1} = p^r/p$. Further, as $r \le \frac{n}{2}$, we bound $n - r \ge \frac{n}{2}$. This allows us to take out an r-th power,

$$F(n,p,r) \le p^{-1} \left[(ne/r) r p e^{-pn/2} \right]^r. \tag{12.69}$$

Let $A = nepe^{-pn/2}$ denote the bracketed term. Here $nep \sim \ln n$ and $e^{-pn/2} = \Theta(n^{-1/2})$, so $A = O(n^{-1/2} \ln n)$. Thus,

$$\sum_{r=3}^{n/2} F(n,p,r) \le p^{-1} \sum_{r=3}^{n/2} A^r \sim p^{-1} A^3 = o(1). \tag{12.70}$$

Hence, $\Pr[G(n,p) \models \mathit{CON}]$ and $\Pr[G(n,p) \models \mathit{NOI}]$ have the same asymptotic probability. □

12.5. Tower and Log Star

The tower function, here denoted $T(n)$, is a fast growing function defined by initial value $T(0) = 1$ and recursion $T(n+1) = 2^{T(n)}$. Thus $T(1) = 2$, $T(2) = 4$, $T(3) = 2^4 = 16$, $T(4) = 2^{16} = 65536$, and $T(5)$ makes a googol (10^{100}) look small.[6] The log star function, written $\log^*(m)$, the inverse of the tower function. $\log^*(m)$ is that n such that $T(n-1) < m \le T(n)$. Thus $\log^*(1000) = 4$ and $\log^*(10^{100}) = 5$. Alternatively, begin with m and recursively take the lg stopping when the result is less than one. (For example, 1000, $9.9\cdots$, $3.31\cdots$, $1.72\cdots$, $0.78\cdots$.) $\log^*(m)$ is one less than the number of times lg was applied. $\log^*(m)$ approaches infinity exceptionally slowly. Compare $\log^*(m)$ with $\lg(\lg(m))$. Set $m = 2^{2^s}$ so that $\lg\lg(m) = s$. Then $\log^*(m) = 2 + \log^*(s)$, which grows more slowly. Indeed, $\log^*(m)$ grows more slowly that the r-times iterated logarithm for any constant r.

12.5.1. Robustness. The tower function was defined by its initial value and the recursion. Changing these, appropriately viewed, has only small effect.

Theorem 12.17. *Let $1 < \alpha, \beta$. Let $a, b > 0$. Let $A(n)$ be defined by initial value $A(0) = a$ and recursion $A(n+1) = \alpha^{A(n)}$. Let $B(n)$ be defined by initial condition $B(0) = b$ and recursion $B(n+1) = \beta^{B(n)}$. Then there exist integer constants c_1, c_2 so that*

$$A(n) \le B(n + c_1) \text{ and } B(n) \le A(n + c_2) \tag{12.71}$$

for all sufficiently large n.

Proof. Assume, by symmetry, that $\alpha \le \beta$. Select c_1 such that $a = A(0) \le B(c_1)$. Then $A(n) \le B(n + c_1)$ by induction on n.

The other side is not so easy. The natural induction will not work as A is growing slower than B. Instead, we select K with $\alpha^K \ge \beta$. We then select Y with $(\alpha^K/\beta)^Y \ge K$. Now suppose $x \ge Ky$ and $y \ge Y$. Then

$$\alpha^x \ge (\alpha^K)^y \ge \beta^y(\alpha^K/\beta)^Y \ge K\beta^y. \tag{12.72}$$

[6] Chapter 13 gives functions that make the tower function look slow.

We select n_0 so that $B(n_0) \geq Y$ and then select c_2 such that $A(n_0 + c_2) \geq KB(n_0)$. By induction, applying (12.72), $A(n + c_2) \geq KB(n)$ for all $n \geq n_0$. □

Theorem 12.18. *Let $1 < \alpha, \beta$. Let $a, b > 0$. Let $A^*(m)$ denote the number of times, beginning with m, that one can take the log to the base α before becoming less than a. Let $B^*(m)$ denote the number of times, beginning with m, that one can take the log to the base β before becoming less than b. Then*

$$A^*(m) = B^*(m) + \Theta(1). \tag{12.73}$$

Theorem 12.18 is a reformulation of Theorem 12.17 as A^* and B^* are the inverse functions of A, B, respectively. Taking $\beta = 2$, $b = 1$, $B^*(m) = \log^*(m)$. This indicates the robustness of the log star function. Changing the base of the log and changing the finishing point only affects the function by an additive constant.

12.5.2. Long Chains. Let $T = T_t$ be the full binary tree of depth t. That is, there is a root at level 0. Each node at level i, $0 \leq i < t$, has two children. There are 2^s nodes at level s for $0 \leq s \leq t$. We select a random subset $S \subset T$ as follows. At each level s we select uniformly precisely one node and place it in S. A set $C \subset T_n$ is called[7] a *chain* if, given two distinct $x, y \in C$, one of them is a descendant of the other. Let $M = M(S)$ be the maximal size $|C|$ of a chain $C \subset S$.

Theorem 12.19. $E[M] = \log^* t + \Theta(1)$.

Proof. We shall break the levels into sections. Basically we want a section from level j to level 2^j. The robustness of the log star function allows us plenty of room to increase or decrease the function 2^j so as to increase or decrease the number of nodes of C in each section without significantly altering the number of sections.

Lower Bound. Set $a(1) = 1$ and $a(i + 1) = 4^{a(i)}$. Let u be the maximal integer with $a(u) \leq t$. (As this is a lower bound, we may consider $t = a(u)$.) By section i, $1 \leq i < u$, we mean those nodes at level s, $a(i) < s \leq a(i + 1)$. We create C as follows: Put the root in C and ignore (for simplicity) level one. For $1 \leq i < u$ suppose the

[7] Our thanks to Yuval Peres for calling this problem to our attention.

elements in C through section $i-1$ have been determined. Let w be that element at lowest level. Consider all the elements of S in section i. If none of these elements are descendants of w, then do nothing and go on to section $i+1$. If any of these elements are descendants of w, then select one (arbitrarily!) and add it to C. The section is so thick that this is very likely to happen.

Set $j = a(i)$ and $J = a(i+1)$ so that $J = 4^j$. The worst case is when the element w is at level j. There are 2^j nodes at level j. When a node is selected uniformly at level s, $j < s \le 4^j$, it has probability 2^{-j} of being a descendant of w. The elements on different levels are chosen independently. Hence, the probability that none of the nodes of S in section i are descendants of w is

$$(1-2^{-j})^{J-j} \le e^{-2^{-j}J+o(1)} \sim e^{-2^j}, \tag{12.74}$$

which is extremely small. As $\sum_j \exp[-2^j]$ converges the total expected number of sections in which no vertex is selected is $O(1)$. Hence, the expected size of C is $u - O(1)$. From Theorem 12.18 $u = \lg^*(t) - O(1)$, so that the expected size of C is $\lg^*(t) - O(1)$.

Upper Bound. Reset $a(1) = 4$ and $a(i+1) = 2^{a(i)/2}$. Let u now be the minimal integer with $a(u) \ge t$. (As this is an upper bound, we may consider $t = a(u)$.) By section i, $1 \le i < u$, we now mean those nodes at level s, $a(i) < s \le a(i+1)$. Set $j = a(i)$ and $J = a(i+1)$ so that $J = 2^{j/2}$. Let $Z = Z_i$ denote the number of pairs $v, w \in S$, both in section i, with w a descendant of v. When $k < l$, a random node at level l has probability 2^{-k} of being a descendant of a given node at level k. Thus

$$E[Z] = \sum_{j<k<l\le J} 2^{-k} \le \sum_{j<k\le J} 2^{-k}\sqrt{2}^k = O(\sqrt{2}^{-j}), \tag{12.75}$$

which is very small. Let Y denote the number of nodes in the largest chain in S in section i. Then $Z \ge Y - 1$. (Indeed, $Z \ge \binom{Y}{2}$ but this inequality will suffice.) Therefore, $E[Y] \le E[Z]+1$. The longest chain C can have at most five nodes in levels zero through four and $1 + Z_i$ nodes from section i, giving an upper bound of $5 + u - 1 + \sum Z_i$. As $\sum_i E[Z_i]$ converges, the expected value of this upper bound is $u + O(1)$. From Theorem 12.18 $u = \lg^*(t) + O(1)$. □

Chapter 13

Really Big Numbers!

> "Yes, please," said Milo. "Can you show me the biggest number there is?"
> "I'd be delighted," [the Mathemagician] replied, opening one of the closet doors. "We keep it right here. It took four miners just to dig it out."
> Inside was the biggest
>
> **3**
>
> Milo had ever seen. It was fully twice as high as the Mathemagician.
> — Norton Juster, *The Phantom Tollbooth*

"Describe, on a 3×5 card, as large a positive integer as you can."[1]

Many mathematicians have at some time played the game above, either solitaire or in competition. My solutions in the second, sixth and twelfth grades, respectively, are shown in the first three figures.

WARP 0

[1] This chapter is a slightly revised version of [**Spe83**]. Reprinted with permission of the Mathematical Association of America.

$1,000,000,000,000,000,$
$000,000,000,000,000,000$
WARP 1

One Googolplexplexplexplex-
plexplexplexplexplexplex-
plexplexplexplexplexplexplex
WARP 2

The last needs a word of explanation. Since googol is 10^{100} and googolplex is 10^{googol}, let us define N-plex as 10^N. Actually, by twelfth grade I could write

one googolplexplexplex...
with a googol plexes,

and even some more elaborate variants. These were at best WARP 2.2. The next level is:

Let $f_1(x) = 2x$
and $f_{n+1}(x) = f_n^{(x)}(1)$
$f_9(9)$
WARP 3

Here $f^{(x)}$ represents the x-th iterate of f. Iterated doubling is exponentiation, $f_2(x) = 2^x$. Iterated exponentiation (as discussed in §12.5) is the tower function, $f_3(x)$ is 2 to the 2 to the $2 \cdots$ to the 2 with x twos. My WARP 2 solution is approximated $f_3(21)$, one for each plex and five to get to a googol. There is no word for f_4.[2] $f_4(4) = f_3(f_3(f_3(f_3(1)))) = f_3(f_3(4)) = f_3(65536)$ is already WARP 2.1.

Three ideas help us create large numbers. First, we concentrate on constructing rapidly growing functions. The numbers will then be the value of the function $f(x)$ for some reasonably small x. Second, we use iteration to build a larger function from a given one. Third, we use diagonalization. Having defined the functions f_n above, we define a diagonal function, called f_ω, by

$$f_\omega(n) = f_n(n).$$

[2]Today f_4 is sometimes referred to as the WOW function.

This is called the Ackermann function. (There are several similar formulations.) The Ackerman function does occasionally appear in "real" mathematics. For example, van der Wærden proved in 1927 in [**dW27**] that to all n there exists $W(n)$ such that if the integers from 1 to $W(n)$ are divided into two classes, then there exists an arithmetic progression of length n in one of the classes. His proof gave a $W(n)$ roughly equal to $f_\omega(n)$. (It is possible that far smaller $W(n)$, even of exponential order, will suffice, and this remains an open problem.[3])

Once $f_\omega(n)$ is defined, there is no reason to stop. We define a new function, let us call it $f_{\omega+1}(n)$, by $f_{\omega+1}(n) = f_\omega^{(n)}(1)$. Having defined $f_{\omega+1}$ we may define $f_{\omega+2}$, $f_{\omega+3}$,.... When faced with ellipses, we resort to diagonalization. We define a new function, called $f_{2\omega}$, by $f_{2\omega}(n) = f_{\omega+n}(n)$

$$f_{2\omega}(9)$$
WARP 3.2

We are defining here a hierarchy of functions in which each function has an immediate successor and where limit functions are defined by the diagonalization of an appropriate subsequence. The usual representation for ordinal numbers provides a perfect framework in which to do this. The ordinals $\alpha < \omega^\omega$ have a simple representation. Each such α may be uniquely written

$$\alpha = a_1\omega^{s_1} + a_2\omega^{s_2} + \cdots + a_r\omega^{s_r} \ (\omega > s_1 > s_2 > \cdots > s_r \geq 0,$$

where the a_i are positive integers. (We write $a\omega^s$ instead of the more customary $\omega^s a$ for convenience of expression.) The limit ordinals are those α with $s_r > 0$. For those we define a specific "natural" sequence $\alpha(n)$ of ordinals approaching α by

$$\alpha(n) = a_1\omega^{s_1} + a_2\omega^{s_2} + \cdots + (a_r - 1)\omega^{s_r} + n\omega^{s_r - 1}.$$

For example, if $\alpha = 2\omega^4 + 3\omega^3$, then $\alpha(n) = 2\omega^4 + 2\omega^3 + n\omega^2$. We define the natural sequence approaching ω^ω by

$$\omega^\omega(n) = \omega^n.$$

[3]Since original publication of his paper, Saharon Shelah [**She88**] has shown an upper bound for $W(n)$ roughly of order $f_4(n)$. Whether $W(n)$ is of exponential order remains a vexing open question.

Now we define $f_\alpha(n)$ for each $\alpha \leq \omega^\omega$ using transfinite induction by

(1) $f_{\alpha+1}(n) = f_\alpha^{(n)}(1)$,

(2) $f_\alpha(n) = f_{\alpha(n)}(n)$ when α is a limit ordinal,

(3) initial value $f_1(n) = 2n$.

$$f_{\omega^\omega}(9)$$
WARP 3.5

Let us emphasize that though we are using the language of infinite ordinals, the functions f_α are recursive functions and the values $f_\alpha(t)$ are well defined integers. The infinite ordinals are, in one sense, merely finite sequences of positive integers being manipulated in particular ways. A recursive program for computing $f_\alpha^{(t)}(n)$ could take the following form.

```
FUNCTION F(α, N, T)
BEGIN
IF T > 1,
      SET X = F(α, N, T − 1)
      RETURN F(α, X, 1)
IF T = 1 and α = 1
      RETURN 2N
IF T = 1 and LIMITORDINAL(α)
      RETURN F(α(N), N, 1)
IF T = 1 and NOT LIMITORDINAL(α)
      RETURN F(α − 1, 1, N)
END
```

The representation of α, the predicate LIMITORDINAL(α), and the functions $\alpha - 1$ and $\alpha(N)$ need to be defined explicitly, though we do not do so here.

We continue the ordinals a half-WARP further. Set

$$\omega_1 = w, \omega_2 = \omega^\omega, \ldots, \omega_{s+1} = \omega^{\omega_s}, \ldots,$$

and set ϵ_0 equal the limit of the ω_s. (We emphasize that ω_1 is *not* the first uncountable ordinal. All ordinals in this chapter are countable.) Each ordinal $\alpha < \omega_{s+1}$ is uniquely represented as

$$\alpha = a_1\omega^{\beta_1} + a_2\omega^{\beta_2} + \cdots + a_r\omega^{\beta_r} \ (\omega_s > \beta_1 > \beta_2 > \cdots > \beta_r \geq 0),$$

the a_i positive integers. A "typical" ordinal is

$$7\omega^{\omega^{2\omega+1}} + 14\omega^{3\omega^{\omega+8}+5\omega^{\omega}}.$$

Now for limits. We say $n\omega^{\beta}$ is the natural sequence approaching $\omega^{\beta+1}$. If β itself is a limit ordinal, then its limit sequence $\beta(n)$ has already been defined, and we call $\omega^{\beta(n)}$ the natural sequence approaching ω^{β}. For sums we keep all but the smallest term fixed and take the limit sequence approaching that smallest term. Thus

$$7\omega^{\omega^{2\omega+1}} + 13\omega^{3\omega^{\omega+8}+5\omega^{\omega}} + \omega^{3\omega^{\omega+8}+4\omega^{\omega}+\omega^{n}}$$

is the natural sequence for the ordinal above. Finally, ϵ_0 has the natural sequence $\epsilon_0(n) = \omega_n$. Now the hierarchy f_α defined above may be extended to all $\alpha < \epsilon_0 + \omega$. We have a big number:

$$f_{\epsilon_0+9}(9)$$
WARP 4

This should win the game against any nonlogician!

Bibliography

[AKS83] M. Ajtai, J. Komlós, and E. Szemerédi, *An* $0(n \log n)$ *sorting network*, Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing (New York, NY, USA), STOC '83, ACM, 1983, pp. 1–9.

[AS08] Noga Alon and Joel Spencer, *The probabilistic method*, 3 ed., New York: Wiley-Interscience, 2008.

[CR96] Richard Courant and Herbert Robbins, *What is mathematics?*, Oxford University Press, 1996.

[Csi93] George Csicsery, *N is a number*, 1993.

[dlVP96] C.-J. de la Vallée Poussin, *Recherches analytiques la théorie des nombres premiers*, Ann. Soc. Scient. Bruxelles **20** (1896), 183–256.

[dW27] B. L. van der Waerden, *Beweis einer baudetschen vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212–216.

[E.E52] E. E. Kummer, *Über die ergänzungssätze zu den allgemein reciprocitätsgesetzen*, Journal für die reine und angewandte Mathematik (1852), 93–146.

[EL] Paul Erdös and László Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions*, Infinite and Finite Sets (to Paul Erdős on his 60th birthday).

[ER59] Paul Erdös and A. Rényi, *On random graphs*, Publicationes Mathematicae **6** (1959), 290–297.

[Erd47] Paul Erdős, *Some remarks on the theory of graphs*, Bull. Amer. Math. Soc. **53** (1947), 292–294.

[Erd49] Paul Erdös, *Démonstrations élémentaire du théorème sur la distribution des nombers premiers*, Scriptum 1, Centre Mathématique (1949).

[ES35] Paul Erdős and George Szekeres, *A combinatorial problem in geometry*, Compositio Mathematica **2** (1935), 463–470.

[GM03] Anna Gál and Peter Bro Miltersen, *The cell probe complexity of succinct data structures*, ICALP, 2003, pp. 332–344.

[Had96] J. Hadamard, *Sur la distribution des zéros de la fonction $\xi(s)$ et ses conséquences arithmétiques*, Bull. Soc. Math. France **24** (1896), 199–220.

[Kim95] Jeong Han Kim, *The Ramsey number $r(3,t)$ has order of magnitude $t^2/\log t$*, Random Structures and Algorithms (1995), 173–207.

[KO] A. Karatsuba and Yu. Ofman, *Multiplication of many digital numbers by automatic computers*, Proceedings of the USSR Academy of Sciences **145** (1964), 293–294.

[Sel49] A. Selberg, *An elementary proof of the prime number theorem*, Ann. Math. **50** (1949), 305–313.

[She88] Saharon Shelah, *Primitive recursive bounds for van der Waerden numbers*, Journal of the AMS **1** (1988), no. 3, 683–697.

[Spe83] Joel Spencer, *Large numbers and unprovable theorems*, Amer. Math. Monthly **90** (1983), 669–675.

[Str69] Volker Strassen, *Gaussian elimination is not optimal*, Numer. Math. **13** (1969), 354–356.

[Win] Peter Winkler, *Seven puzzles you think you must not have heard correctly*. Unpublished manuscript.

Index

Selected Published Titles in This Series

71 **Joel Spencer,** Asymptopia, 2014

70 **Lasse Rempe-Gillen and Rebecca Waldecker,** Primality Testing for Beginners, 2014

69 **Mark Levi,** Classical Mechanics with Calculus of Variations and Optimal Control, 2014

68 **Samuel S. Wagstaff, Jr.,** The Joy of Factoring, 2013

67 **Emily H. Moore and Harriet S. Pollatsek,** Difference Sets, 2013

66 **Thomas Garrity, Richard Belshoff, Lynette Boos, Ryan Brown, Carl Lienert, David Murphy, Junalyn Navarra-Madsen, Pedro Poitevin, Shawn Robinson, Brian Snyder, and Caryn Werner,** Algebraic Geometry, 2013

65 **Victor H. Moll,** Numbers and Functions, 2012

64 **A. B. Sossinsky,** Geometries, 2012

63 **María Cristina Pereyra and Lesley A. Ward,** Harmonic Analysis, 2012

62 **Rebecca Weber,** Computability Theory, 2012

61 **Anthony Bonato and Richard J. Nowakowski,** The Game of Cops and Robbers on Graphs, 2011

60 **Richard Evan Schwartz,** Mostly Surfaces, 2011

59 **Pavel Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina,** Introduction to Representation Theory, 2011

58 **Álvaro Lozano-Robledo,** Elliptic Curves, Modular Forms, and Their L-functions, 2011

57 **Charles M. Grinstead, William P. Peterson, and J. Laurie Snell,** Probability Tales, 2011

56 **Julia Garibaldi, Alex Iosevich, and Steven Senger,** The Erdős Distance Problem, 2011

55 **Gregory F. Lawler,** Random Walk and the Heat Equation, 2010

54 **Alex Kasman,** Glimpses of Soliton Theory, 2010

53 **Jiří Matoušek,** Thirty-three Miniatures, 2010

52 **Yakov Pesin and Vaughn Climenhaga,** Lectures on Fractal Geometry and Dynamical Systems, 2009

51 **Richard S. Palais and Robert A. Palais,** Differential Equations, Mechanics, and Computation, 2009

50 **Mike Mesterton-Gibbons,** A Primer on the Calculus of Variations and Optimal Control Theory, 2009

For a complete list of titles in this series, visit the AMS Bookstore at **www.ams.org/bookstore/stmlseries/**.

Photo by Mary Ann Spencer

Asymptotics in one form or another are part of the landscape for every mathematician. The objective of this book is to present the ideas of how to approach asymptotic problems that arise in discrete mathematics, analysis of algorithms, and number theory. A broad range of topics is covered, including distribution of prime integers, Erdős Magic, random graphs, Ramsey numbers, and asymptotic geometry.

The author is a disciple of Paul Erdős, who taught him about Asymptopia. Primes less than n, graphs with v vertices, random walks of t steps—Erdős was fascinated by the limiting behavior as the variables approached, but never reached, infinity. Asymptotics is very much an art. The various functions $n \ln n$, n^2, $\frac{\ln n}{n}$, $\sqrt{\ln n}$, $\frac{1}{n \ln n}$ all have distinct personalities. Erdős knew these functions as personal friends. It is the author's hope that these insights may be passed on, that the reader may similarly feel which function has the right temperament for a given task. This book is aimed at strong undergraduates, though it is also suitable for particularly good high school students or for graduates wanting to learn some basic techniques.

Asymptopia is a beautiful world. Enjoy!

"This beautiful book is about how to estimate large quantities — and why. Building on nothing more than first-year calculus, it goes all the way into deep asymptotical methods and shows how these can be used to solve problems in number theory, combinatorics, probability, and geometry. The author is a master of exposition: starting from such a simple fact as the infinity of primes, he leads the reader through small steps, each carefully motivated, to many theorems that were cutting-edge when discovered, and teaches the general methods to be learned from these results."

—László Lovász, Loránd Eötvös University

"This is a lovely little travel guide to a country you might not even have heard about — full of wonders, mysteries, small and large discoveries,... and in Joel Spencer you have the perfect travel guide!"

—Günter M. Ziegler, Freie Universität Berlin, coauthor of "Proofs from THE BOOK"

ISBN: 978-1-4704-0904-3

9 781470 409043

STML/71

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-71