Short Course on Asymptotics
Mathematics Summer REU Programs
University of Illinois
Summer 2015

A.J. Hildebrand

7/21/2015

Contents

Notations and Conventions

$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the set of complex numbers
$\mathbb{N}$	the set of positive integers
$x, y, t, \dots$	real numbers
$h, k, n, m, \dots$	integers (usually positive)
$[x]$	the greatest integer $\leq x$ (floor function)
$\{x\}$	the fractional part of x, i.e., $x - [x]$
$p, p_i, q, q_i, \dots$	primes
$\sum_{n \leq x}$	summation over all **positive** integers $\leq x$
$\sum_{0 \leq n \leq x}$	summation over all **nonnegative** integers $\leq x$ (i.e., including $n = 0$)
$\sum_{p \leq x}$	summation over all primes $\leq x$

Convention for empty sums and products: An empty sum (i.e., one in which the summation condition is never satisfied) is defined as 0; an empty product is defined as 1.

Chapter 1

Introduction: What is asymptotics?

1.1 Exact formulas versus approximations and estimates

Much of undergraduate mathematics is about *exact* formulas and identities. For example, in calculus we are taught how to evaluate (in terms of "elementary" functions) integrals of certain classes of functions, such as all rational functions or powers of trigonometric functions; in discrete mathematics we learn to evaluate (in "closed form") certain sums such as $\sum_{k=0}^{n} 2^k$, $\sum_{k=0}^{n} k$, or $\sum_{k=0}^{n} \binom{n}{k}$; in combinatorics we learn formulas for various combinatorial quantities such as the total number of subsets of a set; in number theory we learn how to evaluate various number-theoretic functions such as the Euler phi function; and in differential equations we learn how to *exactly* solve certain types of differential equations.

In the real world, however, the situations where an exact solution or formula exists are the exception rather than the rule. In many problems in combinatorics, number theory, probability, analysis, and other areas, one encounters quantities (functions) that arise naturally and are worth studying, but for which no (simple) exact formula is known, and most likely no such formula exists.[1]

[1] In some cases it can even be *proved* that no such formula exists. For example, there is a remarkable result ("Risch theorem") that shows that a certain class of functions cannot be integrated in terms of elementary functions. This class of functions includes the functions $(\log t)^{-1}$ and e^{-t^2} whose integrals define the logarithmic integral, and the error function integral, respectively.

What one often can do in such cases is to derive so-called *asymptotic estimates* for these quantities, i.e., approximations of the quantities by "simple" functions, which show how these quantities behave "asymptotically" (i.e., when the argument, x or n, goes to infinity). In applications such approximations are often just as useful as an exact formula would be.

Asymptotic analysis, or "asymptotics", is concerned with obtaining such asymptotic estimates; it provides an array of tools and techniques that can be helpful for that purpose. In some cases obtaining asymptotic estimates is straightforward, while in others this can be quite difficult and involve a fair amount of ad-hoc analysis.

1.2 A sampler of applications of asymptotics

Perhaps the best way the answer the questions "What is asymptotics?," and "What is asymptotics good for?" is by describing some of the many problems to which asymptotic methods can be successfully applied. The problems we have selected are all natural and important questions, taken from a variety of areas in mathematics. The goal of this course is to develop the tools to approach these and similar problems.

Combinatorics: Estimating the factorial function, $n!$. How fast does the sequence of factorials, $n! = 1 \cdot 2 \cdots n$, grow as $n \to \infty$? It is easy to see that $n!$ grows at a faster rate than a^n, for any *fixed* a. On the other hand, since $n! \le n^n$, it does not grow faster than n^n. But what is the "real" rate of growth of the factorial sequence? The answer is given by *Stirling's formula*, which says that, for large n, $n! \approx (n/e)^n\sqrt{2\pi n}$. Stirling's formula is one of the most famous asymptotic estimates, and it is key to many other asymptotic estimates.

Probability: Approximating birthday probabilities. The famous "birthday paradox" says that in a group of 23 people there is a better than 50% chance of finding two people with a common birthday. The probability of getting at least one matching birthday in a group of n people can be explicitly written down, but the formula is unwieldy and doesn't provide any insight into the nature of this paradox. In particular, why is the "cutoff value" $n = 23$ so small compared to the total number (i.e., 365) of possible birthdays? Asymptotic analysis provides an answer to such questions: Using asymptotic techniques one can derive simple approximations for various types of "birthday probabilities" that show clearly how these probabilities behave as a function of the group size n, and that allow one to easily calculate appropriate cutoff values.

Probability: The normal distribution. Given n tosses with a fair coin, the probability of getting k heads is given by $\binom{n}{k}2^{-n}$. These probabilities, when plotted as a function of k, have a characteristic "bell curve" shape of the normal (Gaussian) distribution. This is special case of the *Central Limit Theorem*, one of the most famous results in probability. Asymptotic analysis enables one to prove this result rigorously, and also provides some insight into why the normal distribution is so ubiquitous in probability and in real life.

Combinatorics/probability: Estimating the harmonic numbers: The numbers $H_n = \sum_{k=1}^{n} 1/k$, which are the partial sums of the harmonic series $\sum_{k=1}^{\infty} 1/k$, are called harmonic numbers. These numbers arise naturally in a variety of combinatorial and probabilistic problems. There is no exact formula for H_n, but with asymptotic methods one can get remarkably accurate estimates for H_n.

Analysis/probability: The error function integral. The integral $E(x) = \int_x^\infty e^{-t^2} dt$ is of importance in probability and statistics as it represents (up to a constant factor) the tail probability in a normal distribution. The integral representing $E(x)$ cannot be evaluated exactly, but methods of asymptotic analysis lead to quite accurate approximations to $E(x)$ in terms of elementary functions.

Number theory: the n-th prime. Because of the erratic nature of the distribution of primes there exists no exact "closed" formula for p_n, the n-th prime. However, using the Prime Number Theorem (which concerns the behavior of $\pi(x)$, the *number* of primes below x) and asymptotic techniques one can derive *approximate* formulas such as $p_n \sim n \log n$. Such "asymptotic estimates" are less precise than an exact formula, but they still provide useful information about the size of p_n; for instance, the above estimate implies that the series $\sum_{n=1}^{\infty} 1/p_n$ diverges, while the series $\sum_{n=1}^{\infty} 1/(p_n \log n)$ converges.

Number theory/analysis: The logarithmic integral. The function $\mathrm{Li}(x) = \int_2^x (\log t)^{-1} dt$ is called the logarithmic integral. It arises in number theory as the "best" approximation to the prime counting function $\pi(x)$, the number of primes up to x. However, while $\mathrm{Li}(x)$ is a smooth function with nice analytic properties such as monotonicity and infinite differentiability, it itself is a bit of a mystery as it cannot be evaluated in terms of elementary functions. Nonetheless, the behavior of this function can be described quite well by methods of asymptotic analysis.

Analysis: The behavior of the W-function: Trying to solve the equation $w^w = x$ for w leads to the "W-function", $w = W(x)$, a function that arises in many contexts. The W-function cannot be expressed in terms of elementary function. However, using asymptotic techniques such as one can give good approximations for $W(x)$.

Analysis: Estimation of oscillating integrals: The integrals $S(x) = \int_0^x \sin(t^2)dt$ and $C(x) = \int_0^x \cos(t^2)dt$ are called "Fresnel integral," they arise in mathematical physics and other areas, and they have interesting geometric properties. For example, the graph of $(S(x), C(x))$, as x ranges over positive values, has a distinctive spiral shape. As in the case of the error function integral $E(x)$ defined above, these integrals cannot be evaluated directly in terms of elementary functions. However, with techniques from asymptotic analysis, one can derive approximations of these integrals that help explain the geometric features.

Chapter 2

Asymptotic notations

2.1 Big Oh

Among the various notations in asymptotics, by far the most important and most common is the so-called "Big Oh" notation $O(\dots)$.

We define this notation first in the simplest and most common case encountered in asymptotics, namely the behavior of functions of a real or integer variable as the variable approaches infinity. Given two such functions $f(x)$ and $g(x)$, defined for all sufficiently large real numbers x, the notation

$$f(x) = O(g(x))$$

means the following: *There exist constants* x_0 *and* c *such that*

$$|f(x)| \le c|g(x)| \quad (x \ge x_0). \tag{2.1}$$

If this holds, we say that $f(x)$ **is of order** $O(g(x))$, and we call the above estimate a O**-estimate** ("Big Oh estimate") for $f(x)$. The constant c is called the O**-constant**, and the range $x \ge x_0$ the **range of validity** of the O-estimate.

If an explicit range is given in an O-estimate, the corresponding inequality is understood to hold in this range. For example,

$$f(x) = O(g(x)) \quad (0 \le x \le 1)$$

means that the inequality $|f(x)| \le c|g(x)|$ holds, with some constant c, in the range $0 \le x \le 1$ (instead of $x \ge x_0$ for some x_0).

In most applications we are interested in estimates that are valid when the variable involved is sufficiently large, or tends to infinity. We thus make the convention that **if no explicit range is given, a O-estimate is assumed to hold in a range of the form $x \geq x_0$, with a suitable x_0.**

The value of the O-constant c is usually not important; all that matters is that such a constant exists. In fact, in many situations it would be quite tedious (though, in principle, possible) to work out an explicit value for c, even if we are not interested in getting the best-possible value. **The beauty of the O-notation is that it allows us to express, in a succinct and suggestive manner, the existence of such a constant without having to write down the constant.**

If f or g depend on some parameter λ, then the notation "$f(x) = O_\lambda(g(x))$" indicates that the constants x_0 and c implicit in the estimate may depend on λ. If these constants can be chosen independently of λ, for λ in some range, then the estimate is said to hold **uniformly** in that range.

A good way to remember the definition of Big Oh is to think of $O(\dots)$ as an abbreviation for the phrase *"a function that, in absolute value, is bounded by a constant times ..., in the relevant range"*. A term $O(\dots)$ appearing in an equation should be interpreted in this manner. Thus, for example, a statement such as

$$\log(1 + x) = x + O(x^2) \quad (|x| \leq 1/2)$$

translates to the following:

> "$\log(1+x) = x + f(x)$, *where* $f(x)$ *is a function that satisfies* $|f(x)| \leq cx^2$ *for some constant* c *and all* x *in the range* $|x| \leq 1/2$."

Comparing the latter statement, which does not use O-notation, with the former clearly shows the power and elegance of this notation.

Examples. Here are some examples of Big Oh estimates, along with selected proofs. Recall our convention that if no explict range is given in an O-estimate the estimate is understood to hold for all large enough values of the variable, i.e., in a range of the form $x \geq x_0$, with a suitable x_0.

(1) $x = O(e^x)$.

(2) $\log(n^2 + 1) = O(\log n)$.

(3) $\log(1+x) = x - x^2/2 + O(x^3)$ for $|x| \leq 1/2$. More generally, this holds for $|x| \leq c$, for any constant $c < 1$, with the O-constant now depending on c.

(4) $1/(1+x) = 1 - x + O(x^2)$ for $|x| \le 1/2$. (Again, this remains valid for $|x| \le c$, for any fixed constant $c < 1$.)

(5) $\log x = O_\epsilon(x^\epsilon)$ for any $\epsilon > 0$.

The proofs of such estimates are usually straightforward exercises at the calculus level. As an illustration, we give detailed proofs of (1) and (3), with explicit values for the constants involved.

Proof of (1). By the definition of a O-estimate, we need to show that there exist constants c and x_0 such that $x \le ce^x$ for all $x \ge x_0$, or equivalently,

$$\sup_{x \ge x_0} \frac{x}{e^x} \le c.$$

To prove this, consider the function $q(x) = xe^{-x}$. The derivative of $q(x)$ is $q'(x) = e^{-x}(1-x)$, which is negative for $x > 1$ and positive for $x < 1$. Hence $q(x)$ increases for $x < 1$, decreases for $x > 1$, and has a unique maximum at $x = 1$, with maximal value $q(1) = e^{-1}$. Hence $x/e^x = q(x) \le q(1) = e^{-1}$ for *all* $x \in \mathbb{R}$, so the desired inequality holds with $c = e^{-1}$ and any x_0, for example, $x_0 = 0$. In fact, the above argument shows that the value $c = e^{-1}$ is best possible in the range $x \ge 0$. □

Proof of (3). By the Taylor expansion of $\log(1+x)$ we have, for any x with $|x| < 1$,

$$\log(1+x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1} x^n}{n}.$$

If we now suppose that $|x| \le 1/2$, this implies

$$\begin{aligned}\left| \log(1+x) - x + \frac{x^2}{2} \right| &\le \sum_{n=3}^{\infty} \frac{|x|^n}{n} \\ &\le \frac{1}{3} \sum_{n=3}^{\infty} |x|^n = \frac{|x|^3}{3(1-|x|)} \\ &\le \frac{|x|^3}{3(1-(1/2))} = \frac{2}{3}|x|^3,\end{aligned}$$

and so

$$\log(1+x) - x + \frac{x^2}{2} = O(|x|^3),$$

or, equivalently,

$$\log(1+x) = x - \frac{x^2}{2} + O(|x|^3),$$

This proves (3) with O-constant $2/3$ for $|x| \leq 1/2$; the same argument gives (3) with O-constant $3/(1-c))$ in the range $|x| \leq c$, where c is any constant < 1. □

2.2 Small oh and asymptotic equivalence

The notation

$$f(x) = o(g(x)) \quad (x \to \infty)$$

means that $g(x) \neq 0$ for sufficiently large x and

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0.$$

If this holds, we say that $f(x)$ **is of smaller order than** $g(x)$. This is equivalent to having a O-estimate $f(x) = O(g(x))$ with a constant c that can be chosen arbitrarily small (but positive) and a range $x \geq x_0(c)$ depending on c. Thus, a o-estimate is stronger than the corresponding O-estimate.

A closely related notation is that of asymptotic equivalence:

$$f(x) \sim g(x) \quad (x \to \infty)$$

means that $g(x) \neq 0$ for sufficiently large x and

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

If this holds, we say that $f(x)$ **is asymptotic (or "asymptotically equivalent") to** $g(x)$ **as** $x \to \infty$**.**

By an **asymptotic formula** for a function $f(x)$ we mean a relation of the form $f(x) \sim g(x)$, where $g(x)$ is a "simple" function.

Examples

(1) $\log x = o(x)$.

(2) $\log(x+1) \sim \log x$.

(3) $\sqrt{n+1} \sim \sqrt{n}$.

All of these relations are easy to prove, e.g., using l'Hopital's Rule.

2.3 Working with Big Oh estimates

Some useful O-estimates. We begin with a list of some basic O-estimates that can serve as building blocks for more complicated estimates. The estimates are straightforward to prove (e.g., via Taylor expansions). Here, α denotes an arbitrary real number, and C an arbitrary positive constant. The constant $1/2$ in the range $|z| \leq 1/2$ can be replaced by any fixed constant $c < 1$ (with the O-constant then depending on the value of this constant).

$\frac{1}{1+z} = 1 + O(\lvert z\rvert)$	$(\lvert z\rvert \leq 1/2)$
$(1+z)^\alpha = 1 + O_\alpha(\lvert z\rvert)$	$(\lvert z\rvert \leq 1/2)$
$(1+z)^\alpha = 1 + \alpha z + O_\alpha(\lvert z\rvert^2)$	$(\lvert z\rvert \leq 1/2)$
$\log(1+z) = O(\lvert z\rvert)$	$(\lvert z\rvert \leq 1/2)$
$\log(1+z) = z + O(\lvert z\rvert^2)$	$(\lvert z\rvert \leq 1/2)$
$\log(1+z) = z - \frac{z^2}{2} + O(\lvert z\rvert^3)$	$(\lvert z\rvert \leq 1/2)$
$e^z = 1 + O_C(\lvert z\rvert)$	$(\lvert z\rvert \leq C)$
$e^z = 1 + z + O_C(\lvert z\rvert^2)$	$(\lvert z\rvert \leq C)$
$e^z = 1 + z + \frac{z^2}{2} + O_C(\lvert z\rvert^3)$	$(\lvert z\rvert \leq C)$

Table 2.1: Some basic O-estimates.

Properties of O-estimates. We next state some general properties of O-estimates that are useful when working with expressions involving multiple O-terms and that can significantly simplify such expressions.

- **Constants inside O-terms.** If C is a positive constant, then the estimate $f(x) = O(Cg(x))$ is equivalent to $f(x) = O(g(x))$. In particular, the estimate $f(x) = O(C)$ is equivalent to $f(x) = O(1)$. Thus, there is no point in keeping constants inside O-terms; for example, the estimate $f(x) = O(2e^x)$ can be rewritten in equivalent, but simpler, form as $f(x) = O(e^x)$.

- **Adding and multiplying O-terms.** O-estimates can be added and multiplied in the following sense: If $f_1(x) = O(g_1(x))$ and $f_2(x) = O(g_2(x))$, then $f_1(x) + f_2(x) = O(|g_1(x)| + |g_2(x)|)$, and $f_1(x)f_2(x) = O(|g_1(x)g_2(x)|)$. Analogous relations hold for the sum and product of any *fixed* number of O-terms.

- **Differences of O-terms.** Differences of O-terms satisfy the same upper bound as sums: That is, if $f_1(x) = O(g_1(x))$ and $f_2(x) = O(g_2(x))$, then $f_1(x) - f_2(x) = O(|g_1(x)| + |g_2(x)|)$. Note that the right-hand side involves the *sum* of the (absolute values of the) upper-bound functions $g_1(x)$ and $g_2(x)$, not the difference. **In particular, O-terms can never cancel each other out.**

- **Sums of an arbitrary number of O-terms.** The sum property extends to sums of an arbitrary finite or infinite number of O-terms provided the individual O-estimates hold *uniformly*, i.e., with the same c- and x_0-constants: Suppse there exists constants c and x_0, *independent of n*, such that $|f_n(x)| \le c|g_n(x)|$ for all $x \ge x_0$ and all $n \in \mathbb{N}$). Then $\sum_{n=1}^{\infty} f_n(x) = O(\sum_{n=1}^{\infty} |g_n(x)|)$. In other words, one can pull O-constants out of sums and write write $\sum_n O(g_n(x)) = O(\sum_n |g_n|)$, provided the individual O-estimates hold uniformly in n.

- **Distributing O-terms.** O-terms can be distributed in the following sense: If $f_1(x) = O(g_1(x))$ and $f_2(x) = O(g_2(x))$, then for any function $h(x)$, we have $h(x)(f_1(x) + f_2(x)) = O(|h(x)|(|g_1(x)| + |g_2(x)|)$. This property extends in an obvious manner to sums of any *fixed* number of terms.

Proofs. All of these properties are easily derived from the definition of an O-estimate. As an illustration, we give a formal proof of the sum property:

($*$) If (1) $f_1(x) = O(g_1(x))$ and (2) $f_2(x) = O(g_2(x))$, then (3) $f_1(x) + f_2(x) = O(|g_1(x)| + |g_2(x)|)$.

Proof of ($*$). Assume that (1) and (2) hold. By definition, this means that there exist constants c_i and x_i such that, $|f_i(x)| \le c_i|g_i(x)|$ for $x \ge x_i$ for each $i = 1, 2$. Then for $x \ge \max(x_1, x_2)$ we have

$$\begin{aligned} |f_1(x) + f_2(x)| &\le |f_1(x)| + |f_2(x)| \\ &\le c_1|g_1(x)| + c_2|g_2(x)| \\ &\le \max(c_1, c_2)(|g_1(x)| + |g_2(x)|) \end{aligned}$$

Hence, setting $c = \max(c_1, c_2)$ and $x_0 = \max(x_1, x_2)$, we have

$$|f_1(x) + f_2(x)| \le c(|g_1(x)| + |g_2(x)|) \quad (x \ge x_0).$$

This proves (3) with the O-constants c and x_0 as above. □

Example: Simplifying O-estimates. To illustrate the use of the above properties in manipulating and simplifying O-expressions, we now show that

$$\left(1 + \frac{1}{x} + O\left(\frac{1}{x^2}\right)\right)^2 = 1 + \frac{2}{x} + O\left(\frac{1}{x^2}\right). \tag{2.2}$$

The proper interpretation of a relation of this type (i.e., one involving O-terms on both the left and the right side) is that *any function $f(x)$ that is estimated by the expression on the left-hand side is also estimated by the expression on the right-hand side.*

To prove (2.2), we start out with the expression on the left, and then use repeatedly the distributive and additive properties of O-estimates to expand and simplify the expression:

$$\begin{aligned}
\left(1 + \frac{1}{x} + O\left(\frac{1}{x^2}\right)\right)^2 &= \left(1 + \frac{1}{x} + O\left(\frac{1}{x^2}\right)\right)^2 \\
&= \left(1 + \frac{1}{x} + O\left(\frac{1}{x^2}\right)\right) + \frac{1}{x}\left(1 + \frac{1}{x} + O\left(\frac{1}{x^2}\right)\right) \\
&\quad + O\left(\frac{1}{x^2}\right)\left(1 + \frac{1}{x} + O\left(\frac{1}{x^2}\right)\right) \\
&= \left(1 + \frac{1}{x} + O\left(\frac{1}{x^2}\right)\right) + \left(\frac{1}{x} + \frac{1}{x^2} + O\left(\frac{1}{x^3}\right)\right) \\
&\quad + \left(O\left(\frac{1}{x^2}\right) + O\left(\frac{1}{x^3}\right) + O\left(\frac{1}{x^4}\right)\right) \\
&= 1 + \frac{2}{x} + O\left(\frac{1}{x^2}\right).
\end{aligned}$$

Asymptotic tricks. We conclude this section with some useful tricks when dealing with asymptotic expressions.

- **The log trick:** When dealing with an expression involving products of many terms, or high powers, it often helps to consider instead the *logarithm* of the given expressions, which turns products into sums and powers into scalar multiples. The latter can usually easily be simplified using the above properties of O-estimates. Exponentiating the resulting estimate for the logarithm then yields an estimate for the original expression. The estimates for $\log(1+z)$ and e^z given in Table 2.1 are useful in converting between logarithms and exponentials.

- **The factor trick:** Another useful trick is to identify a dominating term in the given expression, factor out this term so that the expression is of the form $f(x)(1+e(x))$, where $f(x)$ is the dominating term and $e(x)$ represents the *relative* error. For example, when trying to estimate $(x+1+1/x)^x$, a natural first step is to factor out the term x inside the parentheses to get $x^x(1+1/x+1/x^2)^x$, and then deal with the second factor, $(1+1/x+1/x^2)^x$, e.g., using the log trick.

- **The max term trick:** *A sum of finitely many terms is bounded* from above *by the number of terms times the maximal term.* This simple observation can yield useful estimates with little or no effort. For example, the sum $\sum_{k=1}^n \sqrt{\log k}$ is seen to be $O(n\sqrt{\log n})$ by this argument.

- **The single term trick:** *A sum of nonnegative terms is bounded* from below *by any single term in this sum.* This is another "trivial" observation that can be surprisingly effective. For example, the sum $\sum_{k=1}^n k^k$ is seen to be $\geq n^n$, and one can show that this trivial lower bound is very close to best-possible.

2.4 A case study: Comparing $(1+1/n)^n$ with e

As is well-known, $(1+1/n)^n$ converges to the Euler constant e as $n \to \infty$, but how quickly does $(1+1/n)^n$ approach its limit, e? For example, does the difference

$$\epsilon_n = e - \left(1 + \frac{1}{n}\right)^n$$

approach 0 fast enough so that $\sum_{n=1}^\infty e_n$ converges?

Using asymptotic techniques we can derive precise estimates for e_n that enable us to answer questions of this type. Specifically, we will show that

$$\left(1 + \frac{1}{n}\right)^n = e - \frac{e}{2n} + O\left(\frac{1}{n^2}\right) \quad (n \geq 2). \tag{2.3}$$

As a consequence of (2.3), we have

$$\sum_{n=2}^{\infty} \epsilon_n = \sum_{n=2}^{\infty} \left(\frac{e}{2n} + O\left(\frac{1}{n^2}\right) \right) = \frac{e}{2} \sum_{n=2}^{\infty} \frac{1}{n} + O\left(\sum_{n=2}^{\infty} \frac{1}{n^2} \right).$$

Since $\sum_{n=2}^{\infty} 1/n$ diverges, while $\sum_{n=2}^{\infty} 1/n^2$ converges, it follows that the series $\sum_{n=2}^{\infty} e_n$ diverges.

Proof of (2.3). We use the log trick, i.e., we first estimate the logarithm of the expression on the left of (2.3), and then exponentiate the resulting estimate.

Let $n \geq 2$. Applying the estimate $\log(1+z) = z - z^2/2 + O(|z|^3)$ ($|z| \leq 1/2$) from Table 2.1 with $z = 1/n$, we get

$$\begin{aligned} \log\left(1 + \frac{1}{n}\right)^n &= n \log\left(1 + \frac{1}{n}\right) \\ &= n \left(\frac{1}{n} - \frac{1}{2n^2} + O\left(\frac{1}{n^3}\right) \right) \\ &= 1 - \frac{1}{2n} + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Exponentiating, we get

$$\begin{aligned} \left(1 + \frac{1}{n}\right)^n &= \exp\left\{ 1 - \frac{1}{2n} + O\left(\frac{1}{n^2}\right) \right\} \\ &= e \cdot e^{-1/(2n)} \cdot e^{O(1/n^2)}. \end{aligned} \tag{2.4}$$

We can simplify the latter two exponentials using the estimates $e^z = 1 + z + O(|z|^2)$ and $e^z = 1 + O(|z|)$ from Table 2.1:

$$\begin{aligned} e^{-1/(2n)} &= 1 - \frac{1}{2n} + O\left(\frac{1}{(2n)^2}\right), \\ e^{O(1/n^2)} &= 1 + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Substituting these estimates into (2.4) and simplifying, we get

$$\begin{aligned} \left(1 + \frac{1}{n}\right)^n &= e \left(1 - \frac{1}{2n} + O\left(\frac{1}{(2n)^2}\right) \right) \left(1 + O\left(\frac{1}{n^2}\right) \right) \\ &= e \left\{ \left(1 - \frac{1}{2n} + O\left(\frac{1}{(2n)^2}\right) \right) + O\left(\frac{1}{n^2}\right) \left(1 - \frac{1}{2n} + O\left(\frac{1}{(2n)^2}\right) \right) \right\} \\ &= e \left(1 - \frac{1}{2n} + O\left(\frac{1}{n^2}\right) \right), \end{aligned}$$

which proves (2.3). □

More precise estimates can be obtained by using additional terms in the Taylor expansion of $\log(1+z)$; in fact, in principle, one could derive an estimate for $(1+1/n)^n$ with error term $O(1/n^k)$, for any fixed integer k.

2.5 Remarks and extensions

O-expressions as terms in equations. While allowing O-expressions to appear in equations is very convenient, one has to be careful not to mis-use equations involving O's and one cannot manipulate O-terms like ordinary arithmetic expressions. For example, if $f(x)$ and $g(x)$ satisfy $f(x) = x + O(1/x)$ and $g(x) = x + O(1/x)$, it would be wrong to equate the two O-expressions and conclude that the two functions must be equal. The reason is that the two O-terms stand for functions that are bounded in absolute value by a constant times $1/x$, but need not be the same. Thus, the only thing one can conclude about the relation between $f(x)$ and $g(x)$ is that $f(x) = g(x) + O(1/x)$.

Vinogradov notations. An sometimes convenient alternative to the O-notation is given by the **Vinogradov notations** $\ll$, $\gg$, and $\asymp$. These are defined as follows:

- "$f(x) \ll g(x)$" means the same as $f(x) = O(g(x))$;
- "$f(x) \gg g(x)$" means the same as $g(x) \ll f(x)$);
- "$f(x) \asymp g(x)$" means that both $f(x) \ll g(x)$ and $f(x) \gg g(x)$ hold.

A good way to remember these notations is to think of $\ll$ and $\gg$ as standing for "less than or equal to a constant times" or "greater than or equal to a (positive) constant times", after placing absolute values on each side of the estimate. For example, with this interpretation, the estimate $f(x) \asymp g(x)$, or equivalently $f(x) \ll g(x) \ll f(x)$, means that there exist positive constants c_1 and c_2 and a constant x_0 such that

$$c_1|g(x)| \leq |f(x)| \leq c_2|g(x)| \quad (x \geq x_0).$$

In this case we say that f and g have the same **order of magnitude**.

Table 2.2 below summarizes the various asymptotic notations.

Terminology	**Notation**	**Definition**
Big oh notation	$f(s) = O(g(s) \quad (s \in S)$	There exists a constant c such that $\|f(s)\| \leq c\|g(s)\|$ for all $s \in S$
Vinogradov notation	$f(s) \ll g(s) \quad (s \in S)$	Equivalent to "$f(s) = O(g(s)\ (s \in S)$"
Order of magnitude estimate	$f(s) \asymp g(s) \quad (s \in S)$	Equivalent to "$f(s) \ll g(s)$ and $g(s) \ll f(s)$ $(s \in S)$".
Small oh notation	$f(s) = o(g(s)) \quad (s \to s_0)$	$\lim_{s \to s_0} f(s)/g(s) = 0$
Asymptotic equivalence	$f(s) \sim g(s) \quad (s \to s_0)$	$\lim_{s \to s_0} f(s)/g(s) = 1$

Table 2.2: Overview of asymptotic terminology and notation. In these definitions S denotes a set of real or complex numbers contained in the domain of the functions f and g, and s_0 denotes a (finite) real or complex number of $\pm\infty$. The most common case is when s is a real variable x and restricted to a range S of the form $x \geq x_0$ for some constant x_0. If no explicit range is given, the range is understood to be of this form.

Chapter 3

Applications to probability

3.1 Probabilities in the birthday problem.

The well-known birthday problem deals with the probability of having two people with the same birthday in a group of a given size. It turns out that this probability is much larger what one would naively expect: for a group of size 23 there is already a greater than even chance of finding two matching birthdays. To see this, note that the probability that in a group of k people there are *no* matching birthdays is given by

$$\frac{365 \cdot 364 \cdots (365 - k + 1)}{365^k} \tag{3.1}$$

(assuming 365 possible birthdays, each equally likely). Calculating these probabilities numerically for successive values of k yields 23 as the "cutoff value" at which the probability of getting two matching birthdays exceeds 0.5. However, this brute force approach does not provide any insight into why this value is so small, nor does it extend to more general questions.

Now consider the case when there are n possible "birthdays" instead of 365. In analogy to (3.1), the probability that in a group of k people there are *no* matching "birthdays" is given by

$$P(n, k) = \frac{n(n-1)\cdots(n-k+1)}{n^k} = \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right). \tag{3.2}$$

As a natural generalization of the original birthday problem, we can ask for the smallest group size, k, needed to ensure a better than even chance of finding two matching "birthdays" in the group, or equivalently, the minimal value $k = k^* = k^*(n)$ for which $P(n, k) < 0.5$. How does $k^*(n)$ behave as

$n \to \infty$? In this section, we will derive an asymptotic estimate for $P(n,k)$ that allows us to answer such questions theoretically.

Proposition 3.1. *Let $P(n,k)$ be the generalized birthday probability defined above, i.e., $P(n,k)$ is the probability that in a group of k people there are* no *matching birthdays, assuming n possible birthdays, each equally likely. Then, for $k \le n/2$, we have*

$$P(n,k) = \exp\left\{ \frac{-k^2}{2n} + O\left(\frac{k}{n}\right) + O\left(\frac{k^3}{n^2}\right) \right\}. \tag{3.3}$$

Corollary 3.2. *Suppose n and k tend to infinity. Then:*

- *If $k \sim \sqrt{2\log 2}\sqrt{n}$, then $P(n,k) \to 1/2$, i.e., the probability of a matching birthday approaches $1/2$. Thus, the cutoff value $k^*(n)$ defined above satisfies $k^*(n) \sim \sqrt{2\log 2}\sqrt{n}$ as $n \to \infty$.*

- *If $k = o(\sqrt{n})$, then $P(n,k) \to 1$, i.e., the probability of a matching birthday approaches 0.*

- *If $k/\sqrt{n} \to \infty$, then $P(n,k) \to 0$, i.e., the probability of a matching birthday approaches 1.*

In particular, for the classical birthday problem (i.e., the case $n = 365$), the above approximation gives $\sqrt{2(\log 2) \cdot 365} = 22.4944\ldots$ as approximation to the cutoff value $k^*(365)$, which is fairly close to the "true" cutoff value 23.

Proof of the Corollary. Suppose $k \sim \sqrt{2\log 2}\sqrt{n}$. Then the terms $O(k/n)$ and $O(k^3/n^2)$ on the right-hand side of (3.3) are of order $O(1/\sqrt{n})$ and hence tend to 0, while $k^2/(2n)$ converges to $\log 2$. Hence $P(n,k)$ converges to $\exp(-\log 2) = 1/2$, as claimed. The other assertionss follow in the same way. $\square$

Proof of Proposition 3.1. For $k \le n/2$, we have

$$\begin{aligned}
\log P(n,k) &= \sum_{i=0}^{k-1} \log\left(1 - \frac{i}{n}\right) \\
&= \sum_{i=0}^{k-1} \left(-\frac{i}{n} + O\left(\frac{i^2}{n^2}\right)\right) \\
&= -\frac{1}{n} \cdot \frac{k(k-1)}{2} + O\left(k \cdot \frac{k^2}{n}\right) \\
&= -\frac{k^2}{2n} + O\left(\frac{k}{n}\right) + O\left(\frac{k^3}{n^2}\right).
\end{aligned}$$

Exponentiating, we obtain the desired estimate (3.3). □

3.2 Poisson approximation to the binomial distribution.

The binomial distribution with parameters p and n, where $0 < p < 1$ and n is a positive integer, is given by

$$b_p(n,k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

The quantity $b_p(n,k)$ can be interpreted as the probability of getting exactly k heads in n coin tosses performed with a biased coin that produces heads with probability p.

The Poisson distribution with parameter $\lambda > 0$ is given by

$$P_\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

A familiar result in probability theory says that, under “suitable” conditions, $b_p(n,k)$ is well approximated by $P_\lambda(k)$, with $\lambda = np$. Most elementary probability texts are not very precise on the conditions, and simply state that p should be small, and ideally of size $1/n$, and that k should be small compared to n.

Here we will prove a precise estimate for $b_p(n,k)$ in terms of $P_\lambda(k)$ (with $\lambda = nk$) which exhibits clearly the conditions under which the Poisson approximation is valid.

Proposition 3.3. *For $0 < p \le 1/\sqrt{n}$ and $0 \le k \le \sqrt{n}$, we have*

$$b_p(n,k) = P_\lambda(k)\left(1 + O\left(\frac{k^2}{n}\right) + O(p^2 n)\right), \tag{3.4}$$

where $\lambda = np$.

Note that the error terms in (3.4) are of order $O(1/n)$ if $k = O(1)$ and $p = O(1/n)$, and of order $o(1)$ (i.e., tend to 0) if $k = o(\sqrt{n})$ and $p = o(1/\sqrt{n})$. Thus we have the following corollary.

Corollary 3.4. *Let $n \to \infty$.*

- *If $k = O(1)$ and $p = O(1/n)$, then*

$$b_p(k,n) = P_\lambda(k)\left(1 + O\left(\frac{1}{n}\right)\right).$$

- *If $k = o(\sqrt{n})$ and $p = o(1/\sqrt{n})$, then*

$$b_p(k,n) \sim P_\lambda(k).$$

The first set of conditions, $k = O(1)$ and $p = O(1/n)$, represents in a way the best-case scenario for Poisson approximation, while the second set of conditions, $k = o(\sqrt{n})$ and $p = o(1/\sqrt{n})$, represent the minimal conditions under which the binomial and Poisson probabilities are asymptotically equal.

Proof of Proposition 3.3. We have

$$\begin{aligned} b_p(n,k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{(np)^k}{k!} (1-p)^{n-k} \\ &= P(n,k) P_{nk}(k) e^{nk} (1-p)^{n-k}, \end{aligned} \tag{3.5}$$

where $P(n,k)$ is given by (3.3). By Proposition 3.1 and the assumption $k \le \sqrt{n}$, we have

$$\begin{aligned} P(n,k) &= \exp\left(\frac{-k^2}{2n} + O\left(\frac{k}{n}\right) + O\left(\frac{k^3}{n^2}\right)\right) \\ &= \exp\left(O\left(\frac{k^2}{n}\right)\right). \end{aligned}$$

Moreover,

$$\begin{aligned}(1-p)^{n-k} &= \exp\{(n-k)\log(1-p)\}\\ &= \exp\{(n-k)(-p+O(p^2))\}\\ &= e^{-np}\exp\{O(np^2)+O(kp)\}.\end{aligned}$$

Substituting these estimates into (3.5), we obtain

$$b_p(n,k) = P_{nk}(k)\exp\left\{O\left(\frac{k^2}{n}\right)+O(np^2)+O(kp)\right\}. \tag{3.6}$$

Now, using the elementary inequality

$$2|ab| \le a^2+b^2 \quad (a,b\in\mathbb{R})$$

(which follows, for example, by observing that $a^2+b^2\pm 2ab = (a\pm b)^2 \ge 0$), we have

$$kp = \frac{k}{\sqrt{n}}\cdot p\sqrt{n} \le \frac{1}{2}\left(\left(\frac{k}{\sqrt{n}}\right)^2+(p\sqrt{n})^2\right) = O\left(\frac{k^2}{n}\right)+O(np^2),$$

so the third error term on the right of (3.6) can be absorbed by the first two. We then get

$$\begin{aligned}&\exp\left\{O\left(\frac{k^2}{n}\right)+O(np^2)+O(kp)\right\}\\ &\qquad= \exp\left\{O\left(\frac{k^2}{n}\right)+O(np^2)\right\}\\ &\qquad= 1+O\left(\frac{k^2}{n}\right)+O(np^2).\end{aligned}$$

where in the final step we used the estimate $e^z = 1+O(z)$ for $|z|\le z_0$, and our assumptions $p\le 1/\sqrt{n}$ and $k\le\sqrt{n}$ (which ensures that both of the O-terms in the exponential are bounded). Substituting this into (3.6) yields the desired estimate (3.4). □

3.3 The center of the binomial distribution

The probability of getting exactly n heads in $2n$ tosses with a fair coin is

$$P(n \text{ heads in } 2n \text{ tosses}) = \binom{2n}{n}2^{-2n}. \tag{3.7}$$

This probability is the largest among the probabilities for getting a specific number of heads in $2n$ coin tosses, it represents the probability of a "tie" in a sequence of $2n$ fair games, and it arises naturally in a number of problems in probability and combinatorics, in particular, the "drunkard's walk problem," which we will describe below. It is therefore of interest to study its behavior as $n \to \infty$.

We begin with an asymptotic estimate for the central binomial coefficient, $\binom{2n}{n}$:

Proposition 3.5. *We have*

$$\binom{2n}{n} = \frac{2^{2n}}{\sqrt{\pi n}} \left(1 + O\left(\frac{1}{n}\right)\right). \tag{3.8}$$

Proof. We have $\binom{2n}{n} = (2n)!/n!^2$. To estimate the factorials on the right-hand side, we will use *Stirling's formula*, one of the most famous asymptotic estimates, which states that

$$n! = \sqrt{2\pi n}\, n^n e^{-n} \left(1 + O\left(\frac{1}{n}\right)\right). \tag{3.9}$$

We will prove Stirling's formula in a later chapter. Here we assume the formula and apply it to derive the desired estimate for $\binom{2n}{n}$.

Substituting (3.9) into $\binom{2n}{n} = (2n)!/n!^2$, we obtain

$$\begin{aligned}
\binom{2n}{n} &= \frac{(2n)!}{n!^2} \\
&= \frac{\sqrt{2\pi(2n)}(2n)^{2n} e^{-2n} \left(1 + O(1/n)\right)}{\left(\sqrt{2\pi n} n^n e^{-n} \left(1 + O(1/n)\right)\right)^2} \\
&= \frac{2^{2n}}{\sqrt{\pi n}} \left(1 + O\left(\frac{1}{n}\right)\right) \left(1 + O\left(\frac{1}{n}\right)\right)^{-2} \\
&= \frac{2^{2n}}{\sqrt{\pi n}} \left(1 + O\left(\frac{1}{n}\right)\right),
\end{aligned}$$

where we have used the estimate $(1 + O(z))^{-2} = 1 + O(z)$ $(|z| \le z_0)$ and basic properties of O-estimates. This proves the desired estimate (3.8). □

As an immediate application we obtain an estimate for the probabilities (3.7):

Corollary 3.6. *The probability for getting exactly n heads and n tails in $2n$ coin tosses with a fair coin satisfies*

$$(3.10)\qquad P(n \text{ heads in } 2n \text{ tosses}) = \frac{1}{\sqrt{\pi n}}\left(1 + O\left(\frac{1}{n}\right)\right).$$

In practice, this approximation is surprisingly accurate, even for relatively small values of n. For example, when $n = 10$, the approximation gives $1/\sqrt{\pi \cdot 10} = 0.178412\ldots\ldots$, while the exact value for the probability on the left is $\binom{20}{10}2^{-20} = 0.17619\ldots$, giving a relative error of around 1.25%. For $n = 100$, the relative error decreases to around 0.125%, and for $n = 1000$, it is approximately 0.0125%. This behavior is consistent with the rate of decrease of the relative error, $O(1/n)$, in (3.10).

Our second application is to the so-called *drunkard's walk problem*: In a d-dimensional drunkard's walk, a "drunkard" walks along a grid in d-dimensional space, with moves given by d-dimensional vectors $(\pm 1, \ldots, \pm 1)$, selected at random with uniform probability. For example, for $d = 1$, the possible moves are $+1$ and -1, each chosen with probability $1/2$, while for $d = 2$, the possible moves are $(1,1), (1,-1), (-1,1), (-1,-1)$, each selected with probability $1/4$.

The drunkard's walk problem is the question whether or not the drunkard is "guaranteed" (with probability 1) to eventually get back to his or her starting location. The answer hinges on the behavior of the probabilities

$$P_n = P(\text{return to start in exactly } 2n \text{ steps}).$$

More specifically, one can show the following;

- If $\sum_{n=1}^{\infty} P_n$ diverges, then with probability 1 the drunkard returns to the starting point.
- If $\sum_{n=1}^{\infty} P_n$ converges, then there is a positive probability that the drunkard never returns to the starting point.
- In either case, the expected number of returns to the origin is given by the sum of the series $\sum_{n=1}^{\infty} P_n$.

We will show:

Corollary 3.7. *Consider a d-dimensional random walk as described above, and let P_n be the probability of return to the starting point in $2n$ steps. Then $\sum_{n=1}^{\infty} P_n$ diverges for dimensions $d = 1, 2$, and converges for dimensions $d \geq 3$. Hence, in dimensions 1 and 2 the drunkard returns to the starting point with probability 1, while in higher dimensions there is a positive probability of never returning to the starting point.*

Proof. We begin by expressing the probabilities P_n in terms of binomial probabilities. To this end, note that a d-dimensional random walk returns to the starting point after $2n$ moves if and only if the $2n$ vectors representing these moves sum to the zero vector. This in turn occurs if and only if, in each component of these vectors, exactly n of the entries are $+1$ and the other n are -1. Since the movement vectors are chosen uniformly among all possible ± 1 tuples, the distribution of $+1$'s and -1's in each component is the same as that of $2n$ coin tosses with a fair coin. Thus, the probability of that, *in a given component*, exactly n entries $+1$ and n are -1 is given by $\binom{2n}{n}2^{-2n}$ (cf. (3.7)).

Since the entries in different components are chosen independently, the probability that this holds for *all d components* of the movement vectors, and hence the probability, P_n of returning to the origin in $2n$ steps, is given by

$$P_n = \left(\binom{2n}{n}2^{-2n}\right)^d.$$

By Proposition 3.5, we have

$$P_n \sim \left(\frac{1}{\sqrt{\pi n}}\right)^d = \frac{\pi^{-d/2}}{n^{d/2}} \quad (n \to \infty).$$

Since the series $\sum_{n=1}^{\infty} n^{-d/2}$ converges if $d > 2$ and diverges if $d \leq 2$, it follows that $\sum_{n=1}^{\infty} P_n$ converges if $d \geq 3$ and diverges when $d = 1, 2$, as claimed. □

3.4 Normal approximation to the binomial distribution

One of the most famous results in probability is the De Moivre-Laplace theorem, which says that, under suitable conditions, the binomial distribution is well approximated by an appropriately scaled normal (Gaussian) distribution. Specifically, let

$$b_{p,n}(k) = \binom{n}{k}p^k(1-p)^{n-k},$$

denote the (discrete) binomial density with parameters p and n, and let

$$\phi_{\sigma,\mu}(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(t-\mu)^2/(2\sigma^2)}$$

the (continuous) density of a normal distribution with mean μ and standard deviation σ. Then, under appropriate conditions, we have

$$b_{p,n}(k) \approx \phi_{\sigma,\mu}(k) \quad \text{where } \mu = np,\ \sigma = \sqrt{p(1-p)n}. \tag{3.11}$$

This is a special case of the Central Limit Theorem in probability, one of the most famous results in probability theory.

In the case of $2n$ tosses with a fair coin (i.e., $p = 1/2$), $\mu = (1/2)(2n) = n$, $\sigma = \sqrt{(1/2)(1/2)(2n)} = \sqrt{n/2}$, and setting $k = n + h$, (3.11) takes the particularly simple form

$$b_{1/2,2n}(n+h) \approx \frac{1}{\sqrt{n\pi}} e^{-h^2/n}. \tag{3.12}$$

In this section we will prove a precise form of (3.12) (see Corollary 3.9). The key is the following estimate for binomial coefficients:

Proposition 3.8. *For $|h| \le n^{2/3}$ we have*

$$\binom{2n}{n+h} = \binom{2n}{n} e^{-h^2/n} \left(1 + O\left(\frac{|h|^3}{n^2}\right)\right). \tag{3.13}$$

Proof. Let $q(h) = \binom{2n}{n+h}/\binom{2n}{n}$. By the symmetry of binomial coefficients, we have $q(h) = q(-h)$, so it suffices to consider positive values of h.

Expanding the binomial coefficients $\binom{2n}{n+h}$ and $\binom{2n}{n}$ gives

$$\begin{aligned} q(h) &= \frac{(2n)(2n-1)\cdots(n-h+1)}{1\cdot 2\cdots(n+h)} \left(\frac{(2n)(2n-1)\cdots(n+1)}{1\cdot 2\cdots n}\right)^{-1} \\ &= \frac{n(n-1)\cdots(n-h+1)}{(n+1)\cdots(n+h)} = \prod_{i=1}^{h} \frac{n-i+1}{n+i} \\ &= \prod_{i=1}^{h} \left(1 - \frac{2i-1}{n+i}\right). \end{aligned}$$

Taking logarithms, it follows that

$$\log q(h) = \sum_{i=1}^{h} \log\left(1 - \frac{2i-1}{n+i}\right). \tag{3.14}$$

Since $h \le n^{2/3}$, we have, for $1 \le i \le h$,

$$0 \le \frac{2i-1}{n+i} \le \frac{2h}{n} \le 2n^{-1/3},$$

and the latter expression is at most $1/2$ provided n is sufficiently large (e.g., $n \geq 2^6$). We can therefore apply the estimate $\log(1+z) = z + O(|z|^2)$ ($|z| \leq 1/2$) to the terms on the right of (3.14) and obtain

$$\begin{aligned}
\log q(h) &= \sum_{i=1}^{h} \left\{ -\frac{2i-1}{n+i} + O\left(\frac{(2i-1)^2}{n^2}\right) \right\} \\
&= \sum_{i=1}^{h} \left\{ -\frac{2i-1}{n} \left(1 + O\left(\frac{i}{n}\right)\right) + O\left(\frac{i^2}{n^2}\right) \right\} \\
&= -\frac{1}{n} \sum_{i=1}^{h} (2i-1) + O\left(\frac{h^3}{n^2}\right) \\
&= -\frac{1}{n} \left(2\frac{h(h+1)}{2} - h\right) + O\left(\frac{h^3}{n^2}\right) \\
&= -\frac{h^2}{n} + O\left(\frac{h^3}{n^2}\right),
\end{aligned}$$

where we have estimated the contribution of the O-terms trivially by the number of terms, h, times the size of the maximal O-term, $O(h^2/n^2)$.

Exponentiating this estimate, we get

$$\begin{aligned}
q(h) &= e^{-h^2/n} \exp\left\{ O\left(\frac{|h|^3}{n^2}\right) \right\} \\
&= e^{-h^2/n} \left(1 + O\left(\frac{|h^3|}{n^2}\right)\right),
\end{aligned}$$

since, by our assumption $|h| \leq n^{2/3}$, the error term $O(h^3/n^2)$ is of order $O(1)$ and hence can be estimated using $e^z = 1 + O(|z|)$ ($|z| \leq z_0$). This is the asserted estimate. □

We can now prove the De Moivre–Laplace theorem in the following precise form:

Corollary 3.9. *The probability for getting exactly $n+h$ heads and n tails in $2n$ coin tosses with a fair coin satisfies, If $|h| \leq n^{2/3}$,*

$$P(n+h \text{ heads in } 2n \text{ tosses}) = \frac{e^{-h^2/n}}{\sqrt{\pi n}} \left(1 + O\left(\frac{1}{n}\right) + O\left(\frac{|h|^3}{n^2}\right)\right). \tag{3.15}$$

Proof. Combining Propositions 3.8 and 3.5 we get

$$
\begin{aligned}
P(n+h \text{ heads in } 2n \text{ tosses}) &= \frac{1}{2^{2n}}\binom{2n}{n} e^{-h^2/n}\left(1+O\left(\frac{|h|^3}{n^2}\right)\right) \\
&= \frac{1}{\sqrt{\pi n}}\left(1+O\left(\frac{1}{n}\right)\right)\left(1+O\left(\frac{|h|^3}{n^2}\right)\right) \\
&= \frac{1}{\sqrt{\pi n}}\left(1+O\left(\frac{1}{n}\right)+O\left(\frac{|h|^3}{n^2}\right)\right).
\end{aligned}
$$

This proves (3.15) □

Chapter 4

Integrals

4.1 The error function integral

In this chapter we discuss some methods to obtain asymptotic estimates for integrals. Our first example is the integral

$$E(x) = \int_x^\infty e^{-t^2} dt,$$

which represents (up to scaling) a tail probability in the Gaussian distribution. As noted in the last chapter, the "bell curve function" e^{-t^2} cannot be integrated in elementary terms, so there is no explicit formula for the integral $E(x)$, and we have to use numerical or asymptotic methods to (approximately) evaluate the integral.

We are interested in the behavior of $E(x)$ when x is large. We will prove:

Theorem 4.1. *We have*

$$E(x) = e^{-x^2}\left(\frac{1}{2x} + O\left(\frac{1}{x^3}\right)\right). \tag{4.1}$$

Proof. Observe that the integrand, e^{-t^2}, takes its maximal value at the lower endpoint, $t = x$, and decays rather rapidly as t gets larger. Thus, we expect that the bulk of the contribution to the integral comes from a small neighborhood of the lower endpoint x. To make this idea precise, we make the change of variables $t = x + s$ and express the integrand in terms of its maximal value, e^{-x^2}, times a decay factor:

$$e^{-t^2} = e^{-x^2} e^{-2sx-s^2}.$$

The integral $E(x)$ then becomes

$$E(x) = e^{-x^2} \int_0^\infty e^{-2sx} e^{-s^2} ds = e^{-x^2} I(x), \tag{4.2}$$

say, and it remains to deal with the integral $I(x)$.

To obtain the desired estimate (4.1) we need to show that $I(x)$ satisfies

$$I(x) = \frac{1}{2x} + O\left(\frac{1}{x^3}\right). \tag{4.3}$$

At this point we could obtain a simple upper bound by dropping the factor e^{-s^2} in the integral $I(x)$ and integrating e^{-2sx} over the range $[0, \infty)$, giving

$$I(x) \le \int_0^\infty e^{-2xs} ds = \frac{1}{2x},$$

which is in fact the main term in the desired estimate (4.3).

To improve this upper bound to an asymptotic estimate, we need to estimate the error we made in replacing the factor e^{-s^2} by 1 in the integral. To this end, we split the range of integration at the point $s = 1$ and write

$$I(x) = \int_0^1 e^{-2sx} e^{-s^2} ds + \int_1^\infty e^{-2sx} e^{-s^2} ds = I_1(x) + I_2(x),$$

say. Such a split is motivated by the rapid decay of the factor e^{-sx} in the integral $I(x)$ (for large x), which suggests that the bulk of the contribution to $I(x)$ comes from a small interval near the lower endpoint 0. Thus, we expect that $I_1(x)$ gives the main term in (4.3).

In the integral $I_1(x)$, we can estimate the factor e^{-s^2} using the Taylor expansion of the exponential function as

$$e^{-s^2} = 1 + O(s^2) \quad (0 \le s \le 1),$$

so that

$$I_1(x) = \int_0^1 e^{-2sx} e^{-s^2} ds = \int_0^1 e^{-2sx} \left(1 + O(s^2)\right) ds.$$

Making the change of variables $u = 2sx$ we obtain

$$\begin{aligned} I_1(x) &= \frac{1}{2x} \int_0^{2x} e^{-u} \left(1 + O\left(\frac{u^2}{x^2}\right)\right) du \\ &= \frac{1}{2x} \left((1 - e^{-2x}) + O\left(\frac{1}{x^2} \int_0^\infty e^{-u^2} u^2 du\right)\right) \\ &= \frac{1}{2x} + O\left(\frac{1}{x^3}\right), \end{aligned} \tag{4.4}$$

which is exactly the right-hand side of the desired estimate (4.3).

For the second integral, $I_2(x)$, a crude upper bound is sufficient:

$$(4.5) \qquad \begin{aligned} I_2(x) &= \int_1^\infty e^{-2sx} e^{-s^2} ds \\ &\le \int_1^\infty e^{-2sx} ds = \frac{e^{-2x}}{2x}, \end{aligned}$$

which is of the order of the error term $O(1/x^3)$ in (4.3) (with plenty to spare, since $e^{-2x} = O_A(1/x^A)$ for any fixed constant A).

Combining the estimates (4.4) and (4.5) yields the desired estimate (4.3) for $I(x)$, and hence completes the proof of the theorem. □

The technique used in this argument can be refined and generalized in a variety of ways. We mention here two such extensions:

Expansions in powers of $1/x$. By taking more terms in the Taylor expansion of e^{-s^2} in the above argument, one can refine the theorem, replacing the error term $O(x^{-3})$ by an expansion of the form $\sum_{i=1}^k a_i x^{-2i-1} + O(x^{-2k-3})$, with suitable coefficients a_i.

More general integrands. Integrals of the form $\int_x^\infty e^{-t^2} f(t) dt$, where $f(t)$ is a slowly varying function ("slow" relative to the rate of variation in the exponential factor e^{-t^2}), can be treated in much the same way, the result being an estimate with main term $f(x)e^{-x^2}/(2x)$. For example, one can show that

$$\int_x^\infty e^{-t^2} \cos t dt = \frac{e^{-x^2}}{2x} \left(\cos x + O\left(\frac{1}{x}\right)\right).$$

4.2 The logarithmic integral

The "logarithmic integral" is defined by

$$\mathrm{Li}(x) = \int_2^x (\log t)^{-1} dt \quad (x \ge 2).$$

This integral, which cannot be evaluated in terms of elementary functions, is the best known approximation to the number of primes $\le x$. It is a much more precise approximation than the approximation $x/\log x$ provided by the prime number theorem in its simplest form.

The following theorem gives an "expansion" for $\mathrm{Li}(x)$ in terms of powers of $1/\log x$. In particular, this result shows that $\mathrm{Li}(x)$ differs from $x/\log x$ by a term of order of magnitude $x/(\log x)^2$.

Theorem 4.2. *For any fixed positive integer k, we have*

$$\mathrm{Li}(x) = \frac{x}{\log x}\left(\sum_{i=0}^{k-1} \frac{i!}{(\log x)^i} + O_k\left(\frac{1}{(\log x)^k}\right)\right) \quad (x \geq 2). \tag{4.6}$$

Note that, because of the factor $i!$, the series in the main term diverges if one lets $k \to \infty$. The resulting infinite series $\sum_{i=0}^{\infty} i!(\log x)^{-i}$ is another example of an "asymptotic series", a "faux" Taylor series, which diverges everywhere, but which, when truncated at some level k, behaves like an ordinary truncated Taylor expansion, in the sense that the error introduced by truncating the series at the k-th term has an order of magnitude equal to that of the $(k+1)$-st term in the series.

To prove the theorem we require a crude estimate for a generalized version of the logarithmic integral, namely

$$\mathrm{Li}_k(x) = \int_2^x (\log t)^{-k} dt \quad (x \geq 2),$$

where k is a positive integer (so that $\mathrm{Li}_1(x) = \mathrm{Li}(x)$). This result is of independent interest, and its proof is quite instructive.

Lemma 4.3. *For any fixed positive integer k, we have*

$$\mathrm{Li}_k(x) \ll_k \frac{x}{(\log x)^k} \quad (x \geq 2).$$

Proof. First note that the bound holds trivially in any range of the form $2 \leq x \leq x_0$ (with the O-constant depending on x_0). We therefore may assume that $x \geq 4$. In this case, we have $2 \leq \sqrt{x} \leq x$, so that we may split the range $2 \leq t \leq x$ into the two subranges $2 \leq t \leq \sqrt{x}$ and $\sqrt{x} \leq t \leq x$. In the first subrange the integrand is bounded by $1/(\log 2)^k$, so the integral over this range is $\leq (\log 2)^{-k}(\sqrt{x} - 2) \ll_k \sqrt{x}$, which is of the desired order of magnitude.

In the remaining range $\sqrt{x} \leq t \leq x$, the integrand is bounded by $\leq (\log \sqrt{x})^{-k} = 2^k (\log x)^{-k}$, so the integral over this range is at most $2^k(\log x)^{-k}(x - \sqrt{x}) \ll_k x(\log x)^{-k}$, which again is of the desired order of magnitude. □

Proof of Theorem 4.2. Integration by parts shows that, for $i = 1, 2, \ldots$,

$$\begin{aligned}\mathrm{Li}_i(x) &= \frac{x}{(\log x)^i} - \frac{2}{(\log 2)^i} - \int_2^x t \frac{-i}{(\log t)^{i+1} t} dt \\ &= \frac{x}{(\log x)^i} - \frac{2}{(\log 2)^i} + i\,\mathrm{Li}_{i+1}(x).\end{aligned}$$

Applying this identity successively for $i = 1, 2, \ldots, k$ (or, alternatively, using induction on k) gives

$$\mathrm{Li}(x) = \mathrm{Li}_1(x) = O_k(1) + \sum_{i=1}^{k} \frac{(i-1)!x}{(\log x)^i} + k!\,\mathrm{Li}_{k+1}(x).$$

(Here the term $O_k(1)$ absorbs the constant terms $2(\log 2)^{-i}$ that arise when using the above estimate for each $i = 1, 2, \ldots, k$.) Since $\mathrm{Li}_{k+1}(x) \ll_k x(\log x)^{-k-1}$ by Lemma 4.3, we get the desired estimate (4.6). □

4.3 The Fresnel integrals

As a final example, we consider the integrals

$$C(x) = \int_0^x \cos(t^2) dt, \quad S(x) = \int_0^x \sin(t^2) dt. \tag{4.7}$$

These integrals are known as Fresnel integrals, and they arise in physics and other areas. Plotting $(C(x), S(x))$ for $0 \le x < \infty$ reveals an interesting spiral type behavior: As $x \to \infty$, the point $(C(x), S(x))$ approaches a limit point, $(\sqrt{\pi/8}, \sqrt{\pi/8})$, and it does so in a spiral-like fashion; these spirals are called *Cornu spirals* or *Euler spirals.*

We will prove the following result, which gives an asymptotic estimate for $C(x)$ and $S(x)$ and explains the spiral shape of the graph of $(C(x), S(x))$:

Theorem 4.4. *We have*

$$C(x) = \sqrt{\frac{\pi}{8}} + \frac{\sin x^2}{2x} + O\left(\frac{1}{x^3}\right), \quad S(x) = \sqrt{\frac{\pi}{8}} - \frac{\cos x^2}{2x} + O\left(\frac{1}{x^3}\right). \tag{4.8}$$

In particular, the theorem shows, that as $x \to \infty$, $(C(x), S(x))$ approaches the point $(\sqrt{\pi/8}, \sqrt{\pi/8})$ and that the difference $(C(x), S(x)) - (\sqrt{\pi/8}, \sqrt{\pi/8})$ is, up to a term $O(1/x^3)$, given by $((\sin x^2)/(2x), -(\cos x^2)/(2x))$. The latter represents a spiral movement, with radius decreasing at the rate $1/(2x)$.

Proof. Let

$$E_1(x,y) = \int_x^y e^{it^2} dt = \int_x^y \cos(t^2) dt + i \int_x^y \sin(t^2) dt. \tag{4.9}$$

Making the change of variables $s = t^2$, and integrating by parts, we get

$$\begin{aligned} E_1(x,y) &= \int_{x^2}^{y^2} \frac{e^{is}}{2\sqrt{s}} ds \\ &= \left[\frac{e^{is}}{2i\sqrt{s}} \right]_{x^2}^{y^2} - \int_{x^2}^{y^2} \frac{e^{is}}{-4is^{3/2}} ds \\ &= \frac{e^{iy^2}}{2iy} - \frac{e^{ix^2}}{2ix} + \int_{x^2}^{y^2} \frac{e^{is}}{4is^{3/2}} ds \\ &= \frac{ie^{ix^2}}{2x} + O\left(\frac{1}{y}\right) - \left[\frac{e^{is}}{4s^{3/2}} \right]_{x^2}^{y^2} + \int_{x^2}^{y^2} \frac{(-3/2)e^{is}}{4s^{5/2}} ds \\ &= \frac{ie^{ix^2}}{2x} + O\left(\frac{1}{y}\right) + O\left(\frac{1}{(x^2)^{3/2}}\right). \end{aligned}$$

Letting $y \to \infty$ we see that the limit $\lim_{y\to\infty} E_1(x,y)$ exists and is equal to

$$\int_x^\infty e^{it^2} dt = \lim_{y\to\infty} E_1(x,y) = \frac{ie^{ix^2}}{2x} + O\left(\frac{1}{x^3}\right).$$

Taking real and imaginary parts, we conclude that

$$\begin{aligned} \int_x^\infty \cos(t^2) dt &= -\frac{\sin x^2}{2x} + O\left(\frac{1}{x^3}\right), \\ \int_x^\infty \sin(t^2) dt &= \frac{\cos x^2}{2x} + O\left(\frac{1}{x^3}\right). \end{aligned}$$

To complete the proof of (4.8), it remains to show that the infinite Fresnel integrals satisfy

$$S(\infty) = \int_0^\infty \sin(t^2) dt = \sqrt{\frac{\pi}{8}}, \quad C(\infty) = \int_0^\infty \cos(t^2) dt = \sqrt{\frac{\pi}{8}}. \tag{4.10}$$

This can be derived from the well-known formula for the Gaussian integral,

$$\int_{-\infty}^\infty e^{-t^2} dt = \sqrt{\pi}. \tag{4.11}$$

by a standard complex integration argument: Let R be a large number, and consider the integral of e^{-z^2} over the straight-line path 0 to R, followed by the circular arc from R to $Re^{i\pi/4}$, and the straight-line path from $Re^{i\pi/4}$ to 0. Since the function e^{-z^2} is analytic everywhere, the integral over the entire closed path is 0, while the integral over the circular arc tends to 0 as $R \to \infty$. Using (4.11) it follows that

$$\begin{aligned}
\frac{\sqrt{\pi}}{2} &= \int_0^\infty e^{-t^2} dt = \int_0^{e^{i\pi/4}\infty} e^{-z^2} dz \\
&= \int_0^\infty e^{-(te^{i\pi/4})^2} e^{i\pi/4} dt = \frac{1+i}{\sqrt{2}} \int_0^\infty e^{-it^2} dt \\
&= \frac{1+i}{\sqrt{2}} \int_0^\infty (\cos(t^2) - i\sin(t^2)) dt \\
&= \frac{1+i}{\sqrt{2}} (C(\infty) - iS(\infty)) \\
&= \frac{1}{\sqrt{2}} (C(\infty) + S(\infty)) + \frac{i}{\sqrt{2}} (C(\infty) - S(\infty)) .
\end{aligned}$$

This implies $C(\infty) = S(\infty) = \sqrt{\pi}/(2\sqrt{2})$, which proves (4.10) □

Chapter 5

Sums

5.1 Euler's summation formula

The familiar integral test for the convergence of a series is based on the fact that when $f(x)$ is a decreasing function of x, then

$$\int_N^\infty f(x)dx \le \sum_{n=N}^\infty f(n) \le \int_{N-1}^\infty f(x)dx.$$

These inequalities can be seen geometrically by interpreting the sum as the combined area of rectangles with base $[n, n+1]$ and height $f(n)$, $n = N, N+1, N+2, \ldots$, and comparing this area with the area below the curve $y = f(x)$.

Euler's summation formula (also referred to as "Euler-MacLaurin summation formula") can be thought of as a more sophisticated and more precise version of this simple observation. Rather than just providing a bound, it gives an exact formula for the sum that takes into account the difference between the sum and the corresponding integral. Moreover, the formula does not require f to be monotone.

Euler's summation formula can be stated in many different versions. Here we give two simple and easy-to-apply versions of the formula, a first degree form, in which the difference between sum and integral is expressed as an integral over the first derivative of the function f, and a second degree form, which involves an integral over the second derivative of f. The most general version of Euler's formula is an analogous, though more complicated, identity involving the k-th derivative of f. The first degree form is particularly simple and easy to remember and often sufficient in applications, but for some applications (e.g., Stirling's formula) we need the second degree form. Higher degree forms are rarely needed in applications.

Theorem 5.1 (Euler's summation formula)**.** *Let $a < b$ be integers and suppose $f(t)$ is a function that is defined on the interval $[a, b]$ and has a continuous derivative there. Then*

$$\sum_{n=a+1}^{b} f(n) = \int_a^b f(t)dt + \int_a^b \{t\} f'(t)dt, \tag{5.1}$$

where $\{t\}$ denotes the fractional part of t, i.e., $\{t\} = t - [t]$. If, in addition, $f(t)$ also has a continuous second derivative on the interval $[a, b]$, then

$$\sum_{n=a+1}^{b} f(n) = \int_a^b f(t)dt + \frac{1}{2}f(b) - \frac{1}{2}f(a) + \int_a^b \psi(t) f''(t)dt, \tag{5.2}$$

where

$$\psi(t) = \frac{1}{2}\left(\{t\} - \{t\}^2\right).$$

Before proving this result, we derive two corollaries, obtained by estimating the last integrals in (5.1) and (5.2) trivially.

Corollary 5.2 (Euler's summation formula, O-version)**.** *Let $a < b$ be integers and suppose $f(t)$ is a function that is defined on the interval $[a, b]$ and has continuous first and second derivatives there. Then*

$$\sum_{n=a+1}^{b} f(n) = \int_a^b f(t)dt + O\left(\int_a^b |f'(t)|dt\right), \tag{5.3}$$

$$\sum_{n=a+1}^{b} f(n) = \int_a^b f(t)dt + \frac{1}{2}f(b) - \frac{1}{2}f(a) + O\left(\int_a^b |f''(t)|dt\right). \tag{5.4}$$

The second corollary is a corresponding result for infinite series, obtained by letting $b \to \infty$ in (5.3) and (5.4).

Corollary 5.3 (Euler's summation formula, infinite series version)**.** *Let a be an integers and suppose $f(t)$ is a function that is defined on the interval $[a, \infty)$ and has continuous first and second derivatives there. Suppose, moreover, that the series $\sum_{n=a+1}^{\infty} f(n)$ converges. Then*

$$\sum_{n=a+1}^{\infty} f(n) = \int_a^\infty f(t)dt + O\left(\int_a^\infty |f'(t)|dt\right), \tag{5.5}$$

$$\sum_{n=a+1}^{\infty} f(n) = \int_a^\infty f(t)dt - \frac{1}{2}f(a) + O\left(\int_a^\infty |f''(t)|dt\right). \tag{5.6}$$

As an immediate application of the latter result, we can derive a precise estimate for the partial sums of the series $\sum_{n=1}^{\infty} 1/n^2$. The infinite series converges, and by a famous result of Euler its sum equals $\pi^2/6$. Thus, the partial sums, $\sum_{n=1}^{N} 1/n^2$, converge to $\pi^2/6$, as $N \to \infty$.

To estimate the difference between these partial sums and their limit, we apply (5.6) with $f(t) = t^{-2}$ (so that $f'(t) = (-2)t^{-3}$ and $f''(t) = 6t^{-4}$) and $a = N$, to get

$$\begin{aligned}\sum_{n=N+1}^{\infty} \frac{1}{n^2} &= \int_N^\infty \frac{1}{t^2} dt - \frac{1}{2N^2} + O\left(\int_N^\infty \frac{1}{t^4} dt\right) \\ &= \frac{1}{N} - \frac{1}{2N^2} + O\left(\frac{1}{N^3}\right).\end{aligned}$$

Hence we have

$$\sum_{n=1}^{N} \frac{1}{n^2} = \frac{\pi^2}{6} - \frac{1}{N} + \frac{1}{2N^2} + O\left(\frac{1}{N^3}\right).$$

Proof of Theorem 5.1. We begin by proving (5.1). We start out by evaluating the second integral on the right of this equation over an interval $[n-1, n)$, where $a < n \leq b$. Note that for t in this interval we have $\{t\} = t - n + 1$. Therefore, using integration by parts we get

$$\begin{aligned}\int_{n-1}^{n} \{t\} f'(t) dt &= \int_{n-1}^{n} (t - n + 1) f'(t) dt \\ &= (t - n + 1) f(t) \Big|_{n-1}^{n} - \int_{n-1}^{n} f(t) dt \\ &= f(n) - \int_{n-1}^{n} f(t) dt.\end{aligned}$$

Adding these equations for $n = a + 1, a + 2, \ldots, b$ we get

$$\int_a^b \{t\} f'(t) dt = \sum_{n=a+1}^{b} \int_{n-1}^{n} \{t\} f'(t) dt = \sum_{n=a+1}^{b} f(n) - \int_a^b f(t) dt,$$

which proves (5.1).

The proof of (5.2) is similar, but a bit more involved. Again we start with the second integral on the right of this equation, restricted to an interval of the type $[n-1, n)$, where $a < n \leq b$. Setting $t = n - 1 + s$, we have, for

t in such an interval, $\{t\} = s$, and hence $\psi(t) = (1/2)(s - s^2)$. Using two integrations by parts we get

$$\begin{aligned}
\int_{n-1}^{n} \psi(t) f''(t)dt &= \int_0^1 \frac{1}{2}(s - s^2) f''(n - 1 + s)ds \\
&= \frac{1}{2}(s - s^2) f'(n - 1 + s)\Big|_0^1 - \int_0^1 \frac{1}{2}(1 - 2s) f'(n - 1 + s)ds \\
&= -\int_0^1 \frac{1}{2}(1 - 2s) f'(n - 1 + s)ds \\
&= -\frac{1}{2}(1 - 2s) f(n - 1 + s)\Big|_0^1 - \int_0^1 f(n - 1 + s)ds \\
&= \frac{1}{2} f(n) + \frac{1}{2} f(n - 1) - \int_{n-1}^{n} f(t)dt.
\end{aligned}$$

The desired equation (5.2) follows upon adding these equations for $n = a + 1, a + 2, \ldots, b$:

$$\begin{aligned}
\int_a^b \psi(t) f''(t)dt &= \sum_{n=a+1}^{b} \frac{1}{2}(f(n) + f(n - 1)) - \int_a^b f(t)dt \\
&= \sum_{n=a+1}^{b} f(n) - \frac{1}{2} f(b) + \frac{1}{2} f(a) - \int_a^b f(t)dt. \qquad \square
\end{aligned}$$

5.2 Harmonic numbers

As a second illustration of Euler's formula we consider the partial sums

$$H_N = \sum_{n=1}^{N} \frac{1}{n}$$

of the harmonic series. The "harmonic numbers" H_N arise naturally in a variety of combinatorial and probabilistic problems.[1]

We will prove the following result, which gives an estimate for H_N with absolute error $O(1/N)$.

[1] One such problem is the "coupon problem": Given N different coupons and assuming that each cereal box contains a coupon chosen at random from these N coupons, what is the expected number (in the sense of probabilistic expectation) of cereal boxes you need to buy in order to obtain a complete set of coupons? The answer is NH_N.

Proposition 5.4. *We have*

$$H_N = \log N + \gamma + O\left(\frac{1}{N}\right),$$

where $\gamma = 0.57721\ldots$ *is a constant, the so-called Euler constant.*

Proof. We are going to apply Euler's summation formula in the form (5.1) to the function $f(t) = 1/t$ (so that $f'(t) = (-1)t^{-1}$) and the interval $[a, b] = [1, N]$. Since the term corresponding to $n = a = 1$ (namely $1/1 = 1$) is not included in the sum on the left of (5.1), we need to add this term back in to obtain the full sum H_N. With this in mind, we get

$$\begin{aligned} H_N &= \int_1^N \frac{1}{t} dt + \int_1^N \{t\} \frac{-1}{t^2} dt + 1 \\ &= \log N - I(N) + 1, \end{aligned}$$

where

$$I(N) = \int_1^N \{t\} t^{-2} dt.$$

The latter integral cannot be evaluated exactly (the fractional part function in the integrand messes things up), so one might think the best one can do is estimate the integrand by $\{t\}t^{-2} = O(t^{-2})$, leading to the bound

$$I(N) = \int_1^N O(t^{-2}) dt = O\left(\int_1^N t^{-2} dt\right),$$

which gives nothing better than $I(N) = O(1)$,

However, we can do better by employing the following "tail-swapping" trick: Given an integral of the form $\int_a^N \ldots$, where a is fixed and N is the relevant variable, we write $\int_a^N = \int_a^\infty - \int_N^\infty$ and, instead of estimating $\int_a^N$ directly, we estimate the "tail integral", $\int_N^\infty$. This trick can be employed whenever the integral converges when the upper limit is replaced by infinity, and in this case the infinite integral, $\int_a^\infty$, is simply a constant that is independent of N. (Usually, this constant cannot be explicitly evaluated.)

In the case of the integral $I(N)$, the integrand is order $O(t^{-2})$, and hence the integral converges when extended to infinity. The tail-swapping trick can therefore be applied and we get

$$\begin{aligned} I(N) &= \int_1^\infty \frac{\{t\}}{t^2}dt - \int_N^\infty \frac{\{t\}}{t^2}dt \\ &= I + O\left(\int_N^\infty \frac{1}{t^2}dt\right) \\ &= I + O\left(\frac{1}{N}\right), \end{aligned}$$

where

$$I = \int_1^\infty \frac{\{t\}}{t^2}dt.$$

This gives the result, with the constant γ defined by

$$\gamma = 1 - I = 1 - \int_1^\infty \frac{\{t\}}{t^2}dt. \qquad \square$$

5.3 Proof of Stirling's formula with unspecified constant

As a final illustration of Euler's summation formula, we will prove Stirling's formula, i.e., the estimate

$$n! = \sqrt{2\pi n}\, n^n e^{-n}\left(1 + O\left(\frac{1}{n}\right)\right).$$

but with an unspecified constant C instead of $\sqrt{2\pi}$.

We first restate the formula in a notation (namely, writing N in place of n) that is more convenient for our purposes.

Proposition 5.5 (Stirling's formula)**.** *There exists a positive constant C such that, for any positive integer N,*

$$N! = C\sqrt{N}\, N^N e^{-N}\left(1 + O\left(\frac{1}{N}\right)\right). \tag{5.7}$$

Proof. Since $N!$ is defined as a product, we use the "log trick": Instead of estimating $N!$ directly, we estimate the logarithm of $N!$, which can be written as

$$\log N! = \log \prod_{n=1}^N n = \sum_{n=1}^N \log n.$$

We will show that

$$(5.8) \qquad \sum_{n=1}^{N} \log n = N \log N - N + \frac{1}{2} \log N + c + O\left(\frac{1}{N}\right),$$

for a suitable constant c. The estimate (5.7) follows from this (with $C = e^c$) upon exponentiating both sides and applying the estimate $e^z = 1 + O(z)$ ($|z| \le z_0$).

Let

$$S(N) = \sum_{n=1}^{N} \log n = \sum_{n=2}^{N} \log n$$

denote the sum on the left of (5.8). We apply Euler's summation formula in the second degree form (5.2) with $a = 1$, $b = N$, and the function $f(t) = \log t$ (so that $f'(t) = t^{-1}$, $f''(t) = (-1)t^{-2}$) to get

$$(5.9) \qquad \begin{aligned} S(N) &= I_1(N) + \frac{1}{2} \log N - \frac{1}{2} \log 1 + I_2(N) \\ &= I_1(N) + \frac{1}{2} \log N + I_2(N), \end{aligned}$$

with

$$I_1(N) = \int_1^N (\log t) dt, \quad I_2(N) = \int_1^N \psi(t)(-1)t^{-2} dt.$$

The first integral can be explicitly evaluated:

$$(5.10) \qquad I_1(N) = t(\log t - 1)\Big|_1^N = N \log N - N + 1.$$

It therefore remains to deal with the integral $I_2(N)$. Since the integrand is of order $O(t^{-2})$, this integral converges when extended to infinity, so we can estimate it by the tail-swapping trick:

$$(5.11) \qquad \begin{aligned} I_2(N) &= \int_1^\infty \psi(t)(-1)t^{-2} dt - \int_N^\infty \psi(t)(-1)t^{-2} dt \\ &= I + O\left(\int_N^\infty t^{-2} dt\right) = I + O(1/N), \end{aligned}$$

where

$$I = \int_1^\infty \psi(t)(-1)t^{-2} dt$$

is a finite constant independent of N.

Inserting (5.10) and (5.11) and into (5.9) gives

$$\begin{aligned} S(N) &= I_1(N) + \frac{1}{2}\log N + I_2(N) \\ &= N\log N - N + 1 + \frac{1}{2}\log N + I + O\left(\frac{1}{N}\right). \end{aligned}$$

This is the desired estimate (5.8) with constant $c = 1 + I$. □