Anatomy of Integers and Random Permutations
Course Lecture Notes

Kevin Ford

Department of Mathematics, 1409 West Green Street, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
Email address: ford@math.uiuc.edu
Contents

Chapter 0. Background material and notation 1
  1. Notation 1
  Number Theory Functions 1
  Permutation notation 1
  Order of magnitude notation (Bachman-Landau, Hardy, Vinogradov) 1
  Probability 1
  General notational conventions 2
  2. Basic summation estimates 2
  3. Arithmetic functions 2
  4. Prime number estimates 3
  5. Inclusion-Exclusion 4
  6. Probability estimates: general 5
  7. Probability estimates: Poisson random variables 5

Chapter 1. The sequence of prime factors of an integer and cycles of a permutation, I 8
  1. Prime factors and divisors 8
  2. Permutations, cycles and fixed sets 9
  3. Cycles and prime factors from intervals: first nibbles 10
  4. Cycles and prime factors from sets: general upper bounds 13
  5. The sequence of cycles and prime factors from intervals 19
  6. Lower bounds on $\mathbb{P}_x\{\omega(n) = k\}$ 20
  7. Prime factors counted with multiplicity 21
  8. Application: Erdős’ multiplication table problem 23
  9. Number of divisors of integers 24
  10. The range of Euler’s function 25
  11. Exercises 27

Chapter 2. Distribution of the largest cycle and largest prime factor 28
  1. Upper bounds 28
  2. Application: large gaps between primes 31
  3. Asymptotic formulas when $u$ is small 32
  4. Exercises 36

Chapter 3. Integers without small prime factors and permutations without small cycles 37
  1. Permutations without small cycles 37
  2. Integers without small prime factors 41
  3. Exercises 45

Chapter 4. Poisson approximation of small cycle lengths and small prime divisors 46
  1. Small cycles of permutations 46
  2. The Kubilius model of small prime factors of integers 47
3. Exercises 51

Chapter 5. Central Limit Theorems 53
  1. Gaussian approximation of Poisson variables 53
  2. Central Limit Theorems for cycles 54
  3. Central Limit theorems for prime factors 56
  4. Exercises 58

Chapter 6. The concentration of divisors of integers and permutations 59
  1. Concentration of divisors 59
  2. A random model of prime factors and cycle lengths 60
  3. Proof of Theorems 6.6, 6.7, and 6.8 62
  4. Proof of Theorem 6.9 64
  5. Exercises 68

Chapter 7. Integers with a divisor in a given interval 69
  1. Exact formulas 69
  2. Easy bounds when $z$ is small or large 69
  3. The critical case $z = 2y$ 71
  4. Some applications of Theorem 7.3 71
  5. A heuristic for $H(x, y, 2y)$ 73
  6. A global-to-local principle 74

Bibliography 79
CHAPTER 0

Background material and notation

1. Notation

Number Theory Functions

- $\tau(n)$ is the number of positive divisors of $n$
- $\omega(n)$ is the number of distinct prime factors of $n$
- $\omega(n; t)$ is the number of distinct prime factors of $n$ which are $\leq t$
- $\Omega(n)$ is the number of prime factors of $n$ counted with multiplicity
- $\mu(n)$ is the Möbius’s function; $\mu(n) = (-1)^{\omega(n)}$ if $n$ is squarefree and $\mu(n) = 0$ otherwise.
- $P^+(n)$ is the largest prime factor of $n$; $P^+(1) = 0$ by convention
- $P^-(n)$ is the smallest prime factor of $n$; $P^-(1) = \infty$ by convention

Permutation notation

- $S_n$ is the permutation group on a set of $n$ objects (we don’t care what the objects are)
- $C_j(\sigma)$ is the number of cycles of length $j$ in the permutation $\sigma$
- $C(\sigma)$ is the total number of cycles in the permutation $\sigma$
- $C_I(\sigma)$ is the number of cycles of length $j \in I$ in the permutation $\sigma$
- $\beta | \sigma$ means that $\beta$ is a divisor of the permutation $\sigma$, i.e. a product of some subset of the cycles of $\sigma$
- A fixed set $I$ of $\sigma$ is a subset of $[n]$ which is itself permuted by $\sigma$. Equivalently, $I$ is the set of indices permuted by a divisor of $\sigma$.
- $|\beta|$ is the size of $\beta$; $\beta$ is a divisor of a permutation.

Order of magnitude notation (Bachman-Landau, Hardy, Vinogradov)

- The notations $f = O(g)$, $f \ll g$ and $g \gg f$ mean that there is a positive constant $C$ so that $|f| \leq Cg$ throughout the domain of $f$. The constant $C$ is independent of any parameters, unless specified by subscripts, e.g. $f(x) = O_x(x^2)$.
- $f \asymp g$ means that both $f \ll g$ and $g \ll f$ hold. Generally makes sense only if $f, g$ are both positive. Equivalently, there are positive constants $C < C'$ such that $Cg \leq f \leq C'g$ throughout the domain of $f$.
- $f \sim g$ as $x \to a$ means $\lim_{x \to a} f(x)/g(x) = 1$. Here $a$ can be finite, $\infty$ or $-\infty$.
- $f = o(g)$ as $x \to a$ means $\lim_{x \to a} f(x)/g(x) = 0$. Here $a$ can be finite, $\infty$ or $-\infty$.

Probability

- $P(X)$ is the probability of the event $X$
- $E(X)$ is the expectation of the event $X$
- $X \overset{d}{=} Y$ means that $X$ has the same distribution as $Y$
- Pois($\lambda$) is a Poisson random variable with parameter $\lambda$
0. BACKGROUND MATERIAL AND NOTATION

General notational conventions

- \( \mathbb{N} = \{1, 2, 3, \ldots\} \), the set of positive integers ("natural numbers")
- \( \mathbb{N}_0 = \{0, 1, 2, 3, \ldots\} = \mathbb{N} \cup \{0\} \)
- \( e = 2.71828182\ldots \) is the base of the natural logarithm
- \( \gamma = 0.57721566\ldots \) is Euler’s constant
- \( \log \) is the natural logarithm
- \( \log_k x \) is the \( k \)-th iterate of the natural logarithm of \( x \)
- \( \lceil x \rceil \) is the least integer which is \( \geq x \)
- \( \lfloor x \rfloor \) is the greatest integer which is \( \leq x \)
- \( \mathbb{I}(S) \) is the indicator function of statement \( S \); \( \mathbb{I}(S) = 1 \) if \( S \) is true, and \( \mathbb{I}(S) = 0 \) if \( S \) is false.
- \( H_n = 1 + 1/2 + \cdots + 1/n \) is the \( n \)-th harmonic sum
- \( H(I) = \sum_{i \in I} 1/i \) is a general harmonic sum, where \( I \subset \mathbb{N} \)
- Variables in boldface type, e.g. \( \mathbf{h} \), usually denote vector quantities.
- A statement for “almost all integers” means that the number of exceptions below \( x \) is \( o(x) \) as \( x \to \infty \).
- The symbols \( p, q \) denote primes unless otherwise noted
- The symbols \( k, l, m, n \) denote integers unless otherwise noted

2. Basic summation estimates

**Harmonic sums.** The harmonic sums \( H_n \) satisfy

(i) \( \log n < H_n < 1 + \log n \);

(ii) \( H_n = \log n + \gamma + O(1/n) \), where \( \gamma = 0.57721566\ldots \) is Euler’s constant.

**Stirling’s formula.** We have the asymptotic (Stirling’s formula)

\[
 n! = \sqrt{2\pi n}(n/e)^n(1 + O(1/n))
\]

and the strict inequalities

\[
 1 < \frac{n!}{\sqrt{2\pi n}(n/e)^n} < e^{1/(12n)} < 1 + \frac{1}{10n} \quad (n \in \mathbb{N}).
\]

**Euler’s summation.** Let \( f \in C^1(y, x) \). Then

\[
 \sum_{y < n \leq x} f(n) = \int_y^x f(t) \, dt + \sum_{t \in \{x\}} f'(t) \, dt + \{y\}f(y) - \{x\}f(x),
\]

where \( \{t\} = t - \lfloor t \rfloor \) is the fractional part of \( t \).

Roughly speaking, the sum of \( f(n) \) is approximated by the corresponding integral of \( f(t) \).

**Abel summation, also called partial summation.** Let \( a_n \) be any sequence of complex numbers with counting function \( A(t) = \sum_{1 \leq n \leq t} a_n \). Let \( 0 < y \leq x \) and suppose \( f \in C^1(y, x) \). Then

\[
 \sum_{y < n \leq x} a_n f(n) = A(x)f(x) - A(y)f(y) - \int_y^x A(t)f'(t) \, dt.
\]

3. Arithmetic functions

A function \( f : \mathbb{N} \to \mathbb{C} \) is called an arithmetic function. An arithmetic function \( f \) is multiplicative if it is not identically zero, and if \( f(mn) = f(m)f(n) \) whenever \( (m, n) = 1 \). Equivalently, if \( n \) has prime factorization \( n = p_1^{c_1} \cdots p_k^{c_k} \), then

\[
 f(n) = f(p_1^{c_1}) \cdots f(p_k^{c_k}).
\]

In particular (the empty product) \( f(1) = 1 \). A function is completely multiplicative if \( f(mn) = f(m)f(n) \) for all \( m, n \in \mathbb{N} \). Important examples include powers \( n^c \) (\( c \) fixed), \( \tau(n) \), the number of positive divisors
of \( n \), Euler’s totient function \( \phi(n) \), and the Möbius function \( \mu(n) \). Clearly, the product and quotient of multiplicative functions is also multiplicative, as is any fixed power of a multiplicative function.

A sum over divisors of \( n \) of a multiplicative function \( f \) may be written as a product:

\[
\sum_{d|n} f(d) = \prod_{p^n|n} \left( 1 + f(p) + f(p^2) + \cdots + f(p^n) \right),
\]

and an “infinite version”

\[
\sum_{d: p|d \Rightarrow p \in T} f(d) = \prod_{p \in T} \left( 1 + f(p) + f(p^2) + \cdots \right),
\]

provided each infinite sum converges, and the product also converges. An important special case is \( f(n) = \frac{1}{n} \), which yields the formula

\[
\sum_{d: p|d \Rightarrow p \in T} \frac{1}{d} = \prod_{p \in T} \left( 1 - \frac{1}{p} \right)^{-1}.
\]

Möbius inversion and the Legendre sieve. The Möbius function encodes a number-theoretic version of inclusion-exclusion. For any finite set of primes with product \( P \), and any set \( A \) of positive integers

\[
\#\{n \in A : (n, P) = 1\} = \sum_{d|P} \mu(d) \#\{n \in A : d|n\}.
\]

An arithmetic function \( f \) is additive if \( f(ab) = f(a) + f(b) \) whenever \( (a, b) = 1 \). A function is completely additive if \( f(ab) = f(a) + f(b) \) for all \( a, b, \in \mathbb{N} \).

Examples include

1. \( f(n) = \omega(n; S) \), the number of distinct prime factors of \( n \) lying in a set \( S \);
2. \( f(n) = \Omega(n) \), the number of prime power divisors of \( n \); this is completely additive;
3. \( f(n) = \log n \), completely additive;
4. \( f(n) = \log g(n) \), where \( g(n) \) is a positive, multiplicative function.

4. Prime number estimates

Throughout, \( p \) denotes a prime number.

**Prime Number Theorem (PNT).**

\[
\pi(x) := \#\{p \leq x\} = \int_2^x \frac{dt}{\log t} + O \left( xe^{-\sqrt{\log x}} \right).
\]

Also,

\[
\sum_{p \leq x} \log p = x + O \left( xe^{-\sqrt{\log x}} \right).
\]

We also have the asymptotic

\[
\int_2^x \frac{dt}{\log t} = \frac{x}{\log x} + O \left( \frac{x}{\log x} \right).
\]

Two consequences are:

**Mertens’ estimates with strong error terms.** For some constants \( c_1, c_2 \) we have

\[
\sum_{p \leq x} \frac{1}{p} = \log \log x + c_1 + O \left( e^{-\sqrt{\log x}} \right),
\]

\[
\sum_{p \leq x} \frac{\log p}{p} = \log x + c_2 + O \left( e^{-\sqrt{\log x}} \right).
\]
and
\begin{equation}
\prod_{p \leq x} \left( 1 - \frac{1}{p} \right) = \frac{e^{-\gamma}}{\log x} \left( 1 + O \left( e^{-\sqrt{\log x}} \right) \right).
\end{equation}

The original Mertens' estimates had error terms $O(1/\log x)$, and this weak form is sufficient in most of our applications. The best-known form of the error terms above take the form \[ \exp\left\{ -c \left( \frac{\log x}{\log 2x} \right)^{3/5} \right\} \] for some constant $c > 0$.

Oftentimes the more crude bounds of Chebyshev suffice: for positive constants $c, c'$ we have
\[ cx \log x \leq \# \{ p \leq x \} \leq c' x \log x \quad (x \geq 2). \]

The Prime Number Theorem also implies that
\[ p_n \sim n \log n \quad (n \to \infty), \]
where $p_n$ is the $n$-th smallest prime.

There are similar bounds for primes in a fixed arithmetic progression. Here we fix $1 \leq a \leq b$ with $\gcd(a, b) = 1$.

**Prime number theorem for arithmetic progressions.** We have
\[ \# \{ p \leq x : p \equiv a \ (\text{mod} \ b) \} \sim \frac{x}{\phi(b) \log x} \quad (x \to \infty) \]

There are results which are uniform in $b$, but they are more complicated to state. We refer the reader to [7] for specifics.

**Mertens’ estimates for primes in arithmetic progressions.** For some constants $c_1(a, b)$ and $c_2(a, b) > 0$ we have
\begin{equation}
\sum_{\substack{p \leq x \\ p \equiv a \ (\text{mod} \ b)}} \frac{1}{p} = \frac{\log \log x}{\phi(b)} + c_1(a, b) + O_b(1/\log x)
\end{equation}
and
\begin{equation}
\prod_{\substack{p \leq x \\ p \equiv a \ (\text{mod} \ b)}} \left( 1 - \frac{1}{p} \right) = \frac{c_2(a, b)}{(\log x)^{1/\phi(b)}} \left( 1 + O_b (1/\log x) \right).
\end{equation}

5. Inclusion-Exclusion

We need a simple version of the inclusion-exclusion principle, with truncation.

**Inclusion-exclusion.** Let $a$ be a non-negative integer. Then, for any $k \in \mathbb{N}$,
\[ 1(a = 0) = \sum_{r=0}^{\infty} (-1)^r \binom{a}{r} = \sum_{r=0}^{k} (-1)^r \binom{a}{r} + (-1)^{k+1} \binom{a-1}{k}. \]

**Proof.** The first equality is trivial from the binomial theorem. For the second, we have
\[ \sum_{r=k+1}^{\infty} (-1)^r \binom{a}{r} = \sum_{r=k+1}^{\infty} (-1)^r \left[ \binom{a-1}{r-1} + \binom{a-1}{r} \right] = (-1)^{k+1} \binom{a-1}{k}. \]

Often, we need to count a reciprocal weighted sum over integers with a given number of prime factors from a given set.
7. Probability estimates: Poisson random variables

**Proposition 0.1.** Let $T$ be a finite set of positive real numbers, and $k \in \mathbb{N}$. Then

$$\frac{1}{k!} \left( H(T)^k - \binom{k}{2} H(T)^{k-2} \sum_{n \in T} \frac{1}{n^2} \right) \leq \sum_{n_1, \ldots, n_k \in T \atop n_1 < \cdots < n_k} \frac{1}{n_1 \cdots n_k} \leq \frac{H(T)^k}{k!}. $$

**Proof.** Evidently,

$$H(T)^k = \sum_{n_1, \ldots, n_k \in T} \frac{1}{n_1 \cdots n_k}. $$

The summands on the right corresponding to distinct, unordered $k$-tuples $(n_1, \ldots, n_k)$ equals

$$k! \sum_{n_1, \ldots, n_k \in T \atop n_1 < \cdots < n_k} \frac{1}{n_1 \cdots n_k},$$

while the summands corresponding to non-distinct $k$-tuples $(n_1, \ldots, n_k)$ have a total sum of at most

$$\binom{k}{2} \sum_{n \in T} \frac{1}{n} H(T)^{k-2}. \quad \Box$$

6. Probability estimates: general

All random variables lie in $\mathbb{R}$, most are non-negative.

**Markov’s inequality.** We have $P(X \geq w) \leq \frac{\mu}{w}$, $\mu = \mathbb{E}X > 0$, $w > 0$.

Two easy consequences are Chebyshev’s inequality for variances and Chernoff’s inequality.

**Chebyshev’s inequality.** If $w > 0$, $\mu = \mathbb{E}X$ and $\mathbb{E}|X - \mu|^2 > 0$, then

$$P(|X - \mu| \geq w \sqrt{\mathbb{E}|X - \mu|^2}) \leq \frac{1}{w^2}. $$

**Chernoff’s inequality.** We have

$$P(X \geq w) \leq \inf_{b \geq 0} \frac{\mathbb{E}e^{bX}}{e^{bw}}. $$

and

$$P(X \leq w) \leq \inf_{b \leq 0} \frac{\mathbb{E}e^{bX}}{e^{bw}}. $$

7. Probability estimates: Poisson random variables

The first Proposition lists basic properties of the Poisson distribution, which are readily verified from the definition.

**Proposition 0.2.** Suppose $X \overset{d}{=} \text{Pois}(\lambda)$. Then

(0.10) $\mathbb{E}X = \lambda$,

(0.11) $\mathbb{E}e^X = e^{(c-1)\lambda}$ $(c > 0)$,

(0.12) $\mathbb{E} \left( \frac{X^m}{m!} \right) = \frac{\lambda^m}{m!}$ $(m \geq 0)$.

If $X_j \overset{d}{=} \text{Pois}(\lambda_j)$, $1 \leq j \leq k$, and $X_1, \ldots, X_k$ are independent, then

(0.13) $X_1 + \cdots + X_k \overset{d}{=} \text{Pois}(\lambda_1 + \cdots + \lambda_k)$.

We also record very useful tail bounds on the Poisson distribution, due to Norton [51].
Now suppose

If

When

is increasing for

is bounded away from 1 we can do better. Suppose \( \alpha < 1 \) and let \( k_0 = \lfloor \alpha \lambda \rfloor \). Then

If \( k_0 = 0 \) then \( Q(\alpha) \leq Q(0) = 1 \) and \( \alpha \lambda < 1 \), hence

Now suppose \( k_0 \geq 1 \). Consecutive summands have ratio \( \geq 1/\alpha \), and thus by (0.1) and the fact that \( \lambda x/(x/e)^x \) is increasing for \( x \leq \lambda \), we have

which concludes the proof of the first inequality.

A second application of Chernoff’s inequality (again with \( b = \log \alpha, \alpha \geq 1 \)) and (0.11) yields

Now assume \( \alpha > 1 \) and write \( k_1 = \lceil \alpha \lambda \rceil \). We have

Consecutive summands have ratio \( \leq 1/\alpha \). By (0.1) and the fact that \( \lambda x/(x/e)^x \) is decreasing for \( x \geq \lambda \),

This concludes the proof of the 2nd inequality. \( \square \)

**Proposition 0.4 (Q bounds).** We have

\[
(0.15) \quad Q(x) = \int_1^x \log t \, dt \quad (\text{all } x) = \sum_{k=2}^{\infty} \frac{(-1)^k}{k(k-1)} (x-1)^k \quad (|x-1| < 1),
\]
and

\[ \frac{x^2}{3} \leq Q(1 + x) \leq x^2 \quad (|x| \leq 1). \]

**Proposition 0.5 (Binomial tails).** Let \( X \) have binomial distribution according to \( n \) trials and parameter \( p \in [0, 1] \); that is, \( \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \). If \( \beta \leq p \) then we have

\[ \mathbb{P}(X \leq \beta n) \leq \exp \left\{ -n \left( \beta \log \frac{\beta}{p} + (1 - \beta) \log \frac{1 - \beta}{1 - p} \right) \right\}. \]

**Proof.** For a proof, see [3, Lemma 4.7.2]. The right side is also an upper bound for \( \mathbb{P}(X \geq \beta n) \) for \( \beta \geq p \) by symmetry. \( \square \)
CHAPTER 1

The sequence of prime factors of an integer and cycles of a permutation, I

Positive integers factor uniquely into a product of prime numbers, and permutations factor uniquely into a product of cycles. Despite this similarity, the two objects, integers and permutations, look very different on the surface. Deeper inspection, however, reveals that the distribution of the two factorizations have many common features, and for much the same underlying reasons.

1. Prime factors and divisors

Given a random number \( n \leq x \) and a prime \( p \), the probability that \( p \mid n \) is very close to \( 1/p \), and moreover these events are close to independent for different \( p \) (as long as \( p \) is “small”, in a sense that will be made precise later). In particular, this heuristic suggests that the number of distinct prime factors \( \omega(n) \) should be about

\[
\sum_{p \leq x} \frac{1}{p} = \log_2 x + O(1)
\]

on average, using Mertens’ sum estimate (0.5). In fact, an easy calculation gives

\[
\frac{1}{x} \sum_{n \leq x} \omega(n) = \frac{1}{x} \sum_{p \leq x} \left\lfloor \frac{x}{p} \right\rfloor = \sum_{p \leq x} 1/p + O(\pi(x)/x) = \log_2 x + O(1).
\]

What is the distribution of \( \omega(n) \) for \( n \leq x \)? One can model the event \( p \mid n \), for a random \( n \leq x \), by a Bernoulli random variable \( X_p \), which equals 1 with probability \( 1/p \) and 0 with probability \( 1 - 1/p \). This in turn is very close to a Poisson random variable \( Z_p \) with parameter \( 1/p \) when \( p \) is large. Thus, \( \omega(n) \) can be modeled by the random variable \( \sum_{p \leq x} Z_p \), which is a Poisson variable with parameter \( \sum_{p \leq x} 1/p = \log_2 x + O(1) \).

The first result in this direction is a classic theorem of Landau from 1900.

**Theorem (Landau, 1900).** For every fixed \( k \),

\[
\#\{n \leq x : \omega(n) = k\} \sim \frac{x}{\log x} \frac{(\log_2 x)^{k-1}}{(k-1)!}.
\]

This already suggests that \( \omega(n) \) has an approximate Poisson distribution, although Landau never wrote this explicitly. It was Hardy and Ramanujan in 1917 who analyzed the behavior of \#\{\( n \leq x : \omega(n) = k\} \) uniformly in \( k \), showing

**Theorem (Hardy-Ramanujan, 1917).** Uniformly for \( x \geq 2 \) and \( k \geq 1 \),

\[
\#\{n \leq x : \omega(n) = k\} \leq C_1 \frac{x}{\log x} \frac{(\log_2 x + C_2)^{k-1}}{(k-1)!},
\]

where \( C_1, C_2 \) are certain absolute constants.

Summing the upper bound for \( k \geq (1 + \varepsilon) \log_2 x \) and \( k \leq (1 - \varepsilon) \log_2 x \), with \( \varepsilon > 0 \) fixed, and using standard bounds on the tail of the Poisson distribution (see (0.3) below), one obtains a sum of \( o(x) \). Consequently, most \( n \leq x \) have close to \( \log_2 x \) distinct prime factors. This result is sometimes referred to as the birth of probabilistic number theory.

Motivated by the fact that the Poisson distribution \( \text{Pois}(\lambda) \) tends to the Gaussian with mean and variance \( \lambda \) as \( \lambda \) tends to infinity, Erdős and Kac proved their celebrated “Central Limit Theorem” for \( \omega(n) \) in 1939:
Theorem (Erdős-Kac, 1939 [25]). For any real $z$,

$$\frac{1}{x} \# \left\{ n \leq x : \frac{\omega(n) - \log_2 x}{\sqrt{\log_2 x}} \leq z \right\} \to \Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} \, dt \quad (x \to \infty).$$

Much further work was done starting in the 1940s, examining the distribution of the entire sequence of prime factors of integers (equivalently, studying the distribution of arbitrary additive functions). Perhaps the most notable was the work of Kubilius, who developed a probabilistic model of integers which provides a kind of meta-tool for studying all kinds of statistical questions about the distribution of prime factors. A key concept in the theory is independence, the idea that if $p$ and $q$ are small primes, then the “events” $p|n$ and $q|n$ are nearly independent, from a probabilistic viewpoint; this idea also played a prominent role in the development of sieve methods. The theory leads to a “Poisson model” of prime factors; namely that the number of prime factors in an interval $(e^{a^*}, e^{b^*})$ has roughly $\text{Pois}(b - a)$ distribution, with disjoint intervals having independent distributions.

The distribution of divisors of integers has also received much attention, beginning in the 1930s. Much of the study was motivated by two fundamental problems:

(a) (Bézout, 1934). Given a quantity $y$, what is the density of integers that have a divisor in $(y, 2y]$?

(b) (Erdős, 1948). Do almost all integers (that is, a set of density 1) have two divisors in some dyadic interval $(z, 2z]$?

Estimates for the density in Problem (a) were given by Erdős and Tenenbaum, with Ford giving the order of magnitude of the order in 2008 [26]. The solution of Problem (b), in the positive, was given by Maier and Tenenbaum in 1984 [47]. Problem (a) is closely related to the Erdős multiplication table problem: How many distinct products are there $ab$ with $1 \leq a \leq N$ and $1 \leq b \leq N$?

2. Permutations, cycles and fixed sets

The classical derangement problem was posed in 1708 by Pierre Raymond de Montmort. The problem asks how many permutations in $S_n$ have no fixed points, that is no 1-cycles. Five years later, he found an exact formula, which is approximately $\frac{1}{n!}$ for large $n$. In the early 1800s, Cauchy introduced the cycle notation and showed that permutations factor uniquely into a product of cycles. He also developed an exact formula for the number of permutations with a given cycle type; that is, the number of cycles of each length.

If $\sigma \in S_n$ has $C_j$ cycles of length $j$ for each $j$, with $\sum_j jC_j = n$, then the number of such permutations equals

$$n! \prod_{j \leq n} \left( \frac{1}{j} \right)^{C_j} \frac{1}{C_j!}.$$

This formula suggests that, for random $\sigma \in S_n$, the quantities $C_1(\sigma), C_2(\sigma), \ldots$ behave like independent Poisson random variables, where $C_j(\sigma)$ has distribution Pois($1/j$). This is not precisely true, because of the condition that $\sum_j jC_j = n$. Goncharov was the first to make such statements rigorous, and in 1944 proved (among other things) the following:

Theorem (Goncharov, 1944 [35]). We have

- For any fixed $j$ and $m$, $\frac{1}{n!} \# \{ \sigma \in S_n : C_j(\sigma) = m \} \to e^{-1/j} \frac{(1/j)^m}{m!} \quad (n \to \infty)$;

- For any real $z$, $\frac{1}{n!} \# \{ \sigma \in S_n : \frac{C(\sigma) - \log n}{\sqrt{\log n}} \leq z \} \to \Phi(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} \, dt \quad (n \to \infty)$.

The (unsigned) Stirling number of the first kind, $S_1(n, m)$, counts the number of permutations $\sigma \in S_n$ with exactly $m$ total cycles; that is, $C(\sigma) = m$. Goncharov used careful asymptotic analysis of Stirling numbers to obtain the second part of the theorem above. The first part was deduced from an exact formula.
which he proved for \( \# \{ \sigma \in S_n : C_j(\sigma) = m \} \) (see Exercise 1.1 below). More recently, the Poisson model has been established in great uniformity: Namely,

\[
(C_1(\sigma), C_2(\sigma), \ldots, C_k(\sigma)) \approx (Z_1, Z_2, \ldots, Z_k)
\]

(meaning the two vectors have distributions which are very close), where \( Z_1, \ldots, Z_k \) are independent and \( Z_j \sim \text{Pois}(1/j) \), provided that \( k = o(n) \) as \( n \to \infty \).

A fixed set of a permutation \( \sigma \) is a subset \( I \subset [n] \) which is itself permuted (i.e., left invariant) by \( \sigma \). A fixed set corresponds to a divisor of \( \sigma \), that is, a product of some subset of the cycles in \( \sigma \) (we include both the empty set and the whole set \([n]\) as fixed sets). For example, if \( \sigma \in S_6 \) has cycle form

\[
\sigma = (1)(24)(356)
\]

then, e.g., \((1)(356)\) is a divisor with corresponding fixed set \( \{1, 3, 5, 6\} \). There play the same role for permutations as divisors do for integers. The existence of fixed sets of a particular size has applications to various questions in statistical group theory, such as generation of \( S_n \) by random permutations, the distribution of transitive subgroups of \( S_n \), and the order of permutations. These in turn have further applications to irreducibility of polynomials over finite fields (with applications to Galois theory) and over global fields.

Since

\[
\sum_{e^k < p \leq e^{k+1}} \frac{1}{p} \approx \log(1 + 1/k) \approx 1/k
\]

when \( k \) is large, we can form an approximate dictionary between statements about random permutations and the analogous statement about random integers.

<table>
<thead>
<tr>
<th>Random permutation</th>
<th>Random integer</th>
</tr>
</thead>
<tbody>
<tr>
<td>cycle of length ( k )</td>
<td>prime factor in ((e^k, e^{k+1}])</td>
</tr>
<tr>
<td>divisor of size ( k )</td>
<td>divisor in ((e^k, e^{k+1}])</td>
</tr>
</tbody>
</table>

### 3. Cycles and prime factors from intervals: first nibbles

With \( n \in \mathbb{N} \) fixed, we consider a random permutation \( \sigma \in S_n \), each permutation occurring with probability \( 1/n! \). Probability and Expectation with respect to this probability space will be denoted \( \mathbb{P}_{\sigma \in S_n} \) and \( \mathbb{E}_{\sigma \in S_n} \). If the value of \( n \) is understood, then often these are abbreviated as \( \mathbb{P}_\sigma \) and \( \mathbb{E}_\sigma \).

Recall that \( C_j(\sigma) \) is the number of cycles of size \( j \) in the permutation \( \sigma \).

The following lemma appears in Watterson [64, Theorem 7].

**Lemma 1.1 (Cycle length lemma).** Let \( m_1, \ldots, m_n \) be non-negative integers with

\[
m_1 + 2m_2 + \cdots + nm_n \leq n.
\]

Then

\[
\mathbb{E}_{\sigma \in S_n} \prod_{j=1}^n \binom{C_j(\sigma)}{m_j} = \prod_{j=1}^n \frac{(1/j)^{m_j}}{m_j!}.
\]

If \( m_1 + 2m_2 + \cdots + nm_n > n \), then the left side is zero.

**Proof.** The second assertion is obvious, since the product on the left side is positive if and only if \( C_j(\sigma) \geq m_j \) for all \( j \), and since \( \sum_j jC_j(\sigma) = n \) this implies that \( \sum jm_j \leq n \). Now assume that \( m_1 + 2m_2 + \cdots + nm_n \leq n \). The product on the left side in the lemma equals the number of ways to choose from \([n]\) a disjoint collection of \( m_1 \) 1-cycles, \( m_2 \) 2-cycles, \ldots, \( m_n \) \( n \)-cycles. The number of ways of choosing from \([n]\) a disjoint collection of \( m_1 \) 1-element sets, \( m_2 \) 2-element sets, \ldots, \( m_n \) \( n \)-element sets is equal to

\[
\frac{n!}{m_1! \cdots m_n!} = \frac{n!}{\prod_{j=1}^n (j!)^{m_j} m_j!}
\]
where $t = n - (m_1 + 2m_2 + \cdots + nm_n)$. The elements of a $k$-element set may be arranged into a cycle in $(k-1)!$ ways. Thus, the number of ways to arrange the elements of these sets into cycles is

$$\prod_{j=1}^{n} (j-1)!^{m_j}.$$ 

Finally, the $t$ elements not used in any of these cycles may be permuted in $t!$ ways. Multiplying these quantities together completes the proof. □

We remark that the RHS in Lemma 1.1 equals

$$\prod_{j=1}^{n} \frac{(1/j)^{m_j}}{m_j!} = \prod_{j=1}^{n} \frac{Z_j}{m_j!},$$

where $Z_j \overset{d}{=} \text{Pois}(1/j)$ and $Z_1, \ldots, Z_n$ are independent. This already suggests that the quantities $C_j(\sigma)$ behave like independent Poisson random variables.

A special case is the well-known formula of Cauchy for the number of permutations with a given cycle type.

**Lemma 1.2 (Cauchy’s formula).** If $m_1 + 2m_2 + \cdots + nm_n = n$, then

$$\mathbb{P}_\sigma \{ C_j(\sigma) = m_j \ (1 \leq j \leq n) \} = \prod_{j=1}^{n} \frac{(1/j)^{m_j}}{m_j!}.$$ 

**Proof.** Apply Lemma 1.1, noting that $\frac{C_j(\sigma)}{m_j!} \neq 0$ for all $j$ if and only if $C_j(\sigma) = m_j$ for every $j$. This implies that

$$\mathbb{P}_\sigma \{ C_j(\sigma) = m_j \ (1 \leq j \leq n) \} = \mathbb{E}_\sigma \prod_{j=1}^{n} \frac{C_j(\sigma)}{m_j!}.$$ 

□

**Corollary 1.3 (Derangements).** We have the exact formula for derangements

$$\mathbb{P}_\sigma(C_1(\sigma) = 0) = \sum_{j=0}^{n} \frac{(-1)^j}{j!}.$$ 

**Proof.** Apply Inclusion-Exclusion with $u = C_1(\sigma)$, followed by the Lemma 1.1. We get

$$\#\{\sigma \in S_n : C_1(\sigma) = 0\} = \sum_{\sigma \in S_n} \sum_{j=0}^{\infty} (-1)^j \binom{C_1(\sigma)}{j}$$

$$= n! \sum_{j=0}^{n} (-1)^j \mathbb{E}_\sigma \binom{C_1(\sigma)}{j} = n! \sum_{j=0}^{n} \frac{(-1)^j}{j!}.$$ 

□

Recall that for a set $T$ of positive integers,

$$C_T(\sigma) = \sum_{j \in T} C_j(\sigma)$$

is the number of cycles in $\sigma$ with length in $T$.

**Corollary 1.4.** For any nonempty set $T \subseteq [n]$, $\mathbb{E}_\sigma C_T(\sigma) = H(T)$. In particular,

$$\mathbb{E}_\sigma C(\sigma) = H_n = \log n + \gamma + O(1/n).$$ 

**Proof.** By Lemma 1.1, $\mathbb{E}_\sigma C_j(\sigma) = 1/j$ for every $j$, and the result follows by linearity of expectation and bounds on harmonic sums. □
**Corollary 1.5.** For any nonempty set \( T \subseteq [n] \),
\[
\mathbb{E}_\sigma(C_T(\sigma) - H(T))^2 \leq H(T)
\]
with equality in the case \( \max T \leq n/2 \).

**Proof.** We have
\[
C_T(\sigma)^2 = \sum_{i,j \in T, i \neq j} C_i(\sigma)C_j(\sigma) + 2 \sum_{i \in T} \left( \frac{C_i(\sigma)}{2} \right) + \sum_{i \in T} C_i(\sigma).
\]
In the double sum over \( i, j \), the summands with \( i + j > n \) are zero. Applying Lemma 1.1, we get
\[
\mathbb{E}_\sigma C_T(\sigma)^2 \leq \sum_{i \in T} \frac{1}{i} + \sum_{i \in T} \left( \frac{1}{i^2} + \frac{1}{i} \right) = H(T)^2 + H(T),
\]
with equality if \( \max T \leq n/2 \). Using Corollary 1.4, we conclude that
\[
\mathbb{E}_\sigma(C_T(\sigma) - H(T))^2 = \mathbb{E}_\sigma C_T(\sigma)^2 - 2H(T)\mathbb{E}_\sigma C_T(\sigma) + H(T)^2 \leq H(T).
\]

Taking \( T = [n] \), so that \( H(T) = H_n = \log n + O(1) \), and applying Chebyshev’s inequality we find that
\[
\mathbb{P}_\sigma \left( C(\sigma) - H_n \geq \xi \sqrt{\log n} \right) \leq \frac{\mathbb{E}_\sigma |C(\sigma) - H_n|^2}{\xi^2 \log n} \ll \frac{1}{\xi^2} \quad (\xi \geq 1).
\]
That is, \( C(\sigma) \) is very close to \( H_n \) for most \( \sigma \).

The situation with primes is more complicated (see the lower bound in Proposition 0.1), but we so have a clean upper bound of the same type. **In what follows, we denote \( \mathbb{P}_x \) and \( \mathbb{E}_x \) the probability and expectation with respect to a random integer in \([1, x] \). Here \( x \) may be any real number.**

**Lemma 1.6.** Let \( T_1, \ldots, T_k \) be nonempty, disjoint subsets of the primes in \([2, x] \), and let \( m_1, \ldots, m_k \geq 0 \). Then
\[
\mathbb{E}_x \prod_{j=1}^{k} \left( \frac{\omega(n; T_j)}{m_j} \right) \leq \prod_{j=1}^{k} \frac{H(T_j)^{m_j}}{m_j!}.
\]

**Proof.** We first write
\[
\sum_{n \leq x} \left( \frac{\omega(n; T_j)}{m_j} \right) = \sum_{\substack{p_{j,1}, \ldots, p_{j,m_j} \in T_j \\ (1 \leq j \leq k) \\ p_{j,1} < \cdots < p_{j,m_j}}} \# \{ n \leq x \colon p_{1,1} \cdots p_{k,m_k} | n \} = \sum_{\substack{p_{j,1}, \ldots, p_{j,m_j} \in T_j \\ (1 \leq j \leq k) \\ p_{j,1} < \cdots < p_{j,m_j}}} \left\lfloor \frac{x}{p_{1,1} \cdots p_{k,m_k}} \right\rfloor.
\]
Using \( [y/n] \leq [y]/n \) for real \( y \) and \( n \in \mathbb{N} \), Proposition 0.1 then gives
\[
\mathbb{E}_x \prod_{j=1}^{k} \left( \frac{\omega(n; T_j)}{m_j} \right) \leq \prod_{j=1}^{k} \sum_{\substack{p_{j,1} \cdots p_{j,m_j} \in T_j \\ (1 \leq j \leq k) \\ p_{j,1} < \cdots < p_{j,m_j}}} \frac{1}{p_{j,1} \cdots p_{j,m_j}} \leq \prod_{j=1}^{k} \frac{H(T_j)^{m_j}}{m_j!}.
\]

A corollary is a theorem of Turán from 1934 [63]:

**Corollary 1.7 (Turán’s variance theorem).** We have
\[
\mathbb{E}_x (\omega(n) - \log_2 x)^2 \ll \log_2 x.
\]
PROOF. We directly compute
\[ E_x(\omega(n) - \log_2 x)^2 = E_x \left( 2 \left( \frac{\omega(n)}{2} \right) + \omega(n)(1 - 2 \log_2 x) + (\log_2 x)^2 \right). \]
Let \( T = T_1 \) be the set of all primes in \([2, x]\), so that \( H(T) = \log_2 x + O(1) \) by Mertens’ theorem (0.5). Using Lemma 1.6 with \( k = 1 \) and \( m_1 = 2 \), together with (1.1), we obtain
\[
E_x(\omega(n) - \log_2 x)^2 \leq H(T)^2 + (1 - 2 \log_2 x)E_x(\omega(n) + (\log_2 x)^2
= O(\log_2 x). \]
We can interpret Corollary 1.7 probabilistically. If \( n \leq x \) is chosen at random, then Corollary 1.7 gives an upper bound on the variance of \( \omega(n) \), telling us that \( \omega(n) \) is concentrated near \( \log_2 x \). In fact, it follows immediately from Chebyshev’s inequality that uniformly for \( \xi \geq 1 \),
\[
P_x \left( |\omega(n) - \log_2 x| \geq \xi \sqrt{\log_2 x} \right) \leq \frac{1}{\xi^2}.
\]
An immediate corollary is the following famous result of Hardy and Ramanujan [42] from 1917. We note that for \( \sqrt{2} < n \leq x \), \( \log_2 x = \log_2 n + O(1) \). That is, we can say that for most \( n \), \( \omega(n) \) is close to \( \log_2 n \).

**Theorem 1.8 (Hardy-Ramanujan).** The function \( \omega(n) \) has normal order \( \log \log n \), meaning that for any \( \varepsilon > 0 \) we have
\[
(1 - \varepsilon) \log \log n \leq \omega(n) \leq (1 + \varepsilon) \log \log n
\]
for almost all integers \( n \).

4. **Cycles and prime factors from sets: general upper bounds**

**Theorem 1.9 (Cycles in sets theorem).** Let \( T_1, \ldots, T_r \) be arbitrary disjoint, nonempty subsets of \([n]\) and \( k_1, \ldots, k_r \geq 0 \). Then
\[
P_{\sigma} \left( C_{T_1}(\sigma) = k_1, \ldots, C_{T_r}(\sigma) = k_r \right) \leq e \prod_{j=1}^{r} \left( \frac{H(T_j)^{k_j}}{k_j!} e^{-H(T_j)} \right) \cdot \left( 1 + \frac{k_1}{H(T_1)} + \cdots + \frac{k_r}{H(T_r)} \right).
\]

**Proof.** Evidently
\[
n\#\{\sigma \in S_n : C_{T_1}(\sigma) = k_1, \ldots, C_{T_r}(\sigma) = k_r \} = \sum_{\alpha \in S_n} \sum_{\alpha | \sigma} \sum_{\alpha \ in \ a \ cycle} |\alpha|.
\]
Write \( \sigma = \alpha \beta \) and let \( h = |\alpha| \). Either \( h \in T_j \) for some unique \( j \), or \( h \notin T_1 \cup \cdots \cup T_r \). Thus, for some \( t \), \( 0 \leq t \leq r \), we have
\[
(C_{T_1}(\beta), \ldots, C_{T_r}(\beta)) = (m_{t,1}, \ldots, m_{t,r}),
\]
where
\[
m_{t,i} = k_i - 1 (i = t \geq 1),
\]
It is permissible to think of \( \beta \in S_{n-h} \) and thus
\[
n\#\{\sigma \in S_n : C_{T_1}(\sigma) = k_1, \ldots, C_{T_r}(\sigma) = k_r \} = \sum_{t=0}^{r} \sum_{h=1}^{n} \sum_{\alpha \ in \ a \ cycle} h \sum_{\beta \ in \ S_{n-h}} 1
\]
\[
= \sum_{t=0}^{r} \sum_{h=1}^{n} \frac{n!}{(n-h)!} \sum_{\beta \ in \ S_{n-h}} 1.
\]
By the multinomial theorem,

\[ n \# \{ \sigma \in S_n : C_T(\sigma) = k_j \ (1 \leq j \leq r) \} = n! \sum_{t=0}^{r} \sum_{h=1}^{n} \frac{1}{b_i!} \frac{1}{b_j!} \prod_{b_i \geq 0} \sum_{i \in T_j} \frac{1}{b_i!} \frac{1}{b_j!} \prod_{b_i \geq 0} \sum_{i \in T_j} \frac{1}{b_i!} \frac{1}{b_j!} \]

uniformly for \( m_i, i = 1, \ldots, n \) and \( k_1, \ldots, k_r \), using Cauchy’s formula (Lemma 1.2) to count such permutations for each type. It follows that

\[ n! \sum_{t=0}^{r} \sum_{h=1}^{n} \frac{1}{b_i!} \frac{1}{b_j!} \prod_{b_i \geq 0} \sum_{i \in T_j} \frac{1}{b_i!} \frac{1}{b_j!} \prod_{b_i \geq 0} \sum_{i \in T_j} \frac{1}{b_i!} \frac{1}{b_j!} \]

say. Let \( T = T_1 \cup \cdots \cup T_r \) and \( T_0 = [n] \setminus T \), and separately consider the summation over \( i \in T \) and \( i \in T_0 \).

By the multinomial theorem,

\[ Y = n! \sum_{t=0}^{r} \sum_{h=1}^{n} \frac{1}{b_i!} \frac{1}{b_j!} \prod_{b_i \geq 0} \sum_{i \in T_j} \frac{1}{b_i!} \frac{1}{b_j!} \prod_{b_i \geq 0} \sum_{i \in T_j} \frac{1}{b_i!} \frac{1}{b_j!} \]

\[ = n! \prod_{t=0}^{r} \prod_{i \in T_j} H(T_j)^k \prod_{i \in T_0} e^{H(T_i)} \]

The claimed bound now follows using harmonic sum bounds in the form

\[ H(T_0) = H_n - \sum_{j=1}^{r} H(T_j) \leq \log n + 1 \sum_{j=1}^{r} H(T_j). \]

**Remark 1.10.** Whenever \( r \) is bounded, and \( k_j = O(H(T_j)) \) for each \( j \), the right side is

\[ \ll P(Z_1 = k_1, \ldots, Z_r = k_r). \]

where \( Z_j \overset{d}{=} \text{Pois}(H(T_j)) \) for each \( j \), and \( Z_1, \ldots, Z_r \) are independent. Thus, Theorem 1.9 gives an upper bound for counts of cycle lengths in sets \( T_1, \ldots, T_r \) of the expected order (up to a constant factor) according to the Poisson model. This is a useful tool for showing that the actual cycle counts cannot vary too much from the expected means.

In the special case \( r = 1 \), Theorem 1.9 implies that for any \( T \subset [n] \) and \( k \geq 0 \),

\[ P_{\sigma} (C_T(\sigma) = k) \leq e^{1-H(T)} \left( \frac{H(T)^{k-1}}{(k-1)!} + \frac{H(T)^k}{k!} \right). \]

Specializing to the case of cycle lengths in a single interval \([m] \), we obtain the following very useful corollary:

**Corollary 1.11 (Cycles in intervals).** Uniformly for \( 1 \leq m \leq n \) and \( 0 \leq \lambda \leq 1 \), we have

\[ P_{\sigma \in S_n} (C_m(\sigma) \leq \lambda \log m) \leq em^{-Q(\lambda)}. \]

Uniformly for \( 1 \leq m \leq n \) and \( 1 \leq \lambda \), we have

\[ P_{\sigma \in S_n} (C_m(\sigma) \geq \lambda \log m) \leq e^{\lambda^2 m^{-Q(\lambda)}}. \]

In particular, uniformly for \( 1 \leq m \leq n \) and \( 0 \leq \psi \leq \sqrt{\log m} \), we have

\[ P_{\sigma \in S_n} \left( |C_m(\sigma) - \log m| > \psi \sqrt{\log m} \right) \leq e^{-\frac{1}{2} \psi^2}. \]
more than 1. Theorem 1.9 implies a right-tail bound of $H_2$. The final estimate follows by taking $\lambda^*$, and, since $\lambda^* < 1$ then the probability in question is $\leq 1$ trivially, $\lambda = 1 + O(1/H_m)$ and

$$m^{-Q(\lambda)} = m^{-O(1/H_m^2)} > 1,$$

so the desired bound holds. Now suppose that $\lambda^* \geq 1$. We have $\lambda \log m \leq \lambda(H_m - 1) - 1 = \lambda^* H_m$ and thus by Theorem 1.9, together with Proposition 0.3, we get that

$$P_{\sigma \in S_n} \left( C_{[n]}(\sigma) \geq \lambda \log m \right) \leq e^{1-H_m} \sum_{k \geq \lambda \log m+1} \frac{H_k}{k!} \leq e^{1-H_m} \sum_{k \geq \lambda^* H_m} \frac{H_k}{k!} \leq e^{1-H_m} Q(\lambda^*).$$

Now $1 \leq \lambda^* \leq \lambda$, and $Q'(u) = \log u$ implies that

$$Q(\lambda^*) \geq Q(\lambda) - \frac{\lambda + 1}{H_m} \log \lambda,$$

and, since $\log m \leq H_m \leq \log m + 1$ we conclude that

$$P_{\sigma \in S_n} \left( C_{[n]}(\sigma) \geq \lambda \log m \right) \leq e^{1-H_m} Q(\lambda) + (\lambda + 1) \log \lambda \leq e^{\lambda - Q(\lambda) - 1}.$$

The final estimate follows by taking $\lambda = 1 + \psi/\sqrt{\log m}$ and using the bound (0.16) for $Q(u)$. Here $0 \leq \lambda \leq 2$.

In particular, taking $m = n$, we see that $C(\sigma)$ usually does not vary more that $\sqrt{\log n}$ from its mean $H_n$.

The same proof yields a much more general result:

**Corollary 1.12.** Let $T \subset [n]$. Uniformly for $0 < \xi \leq \sqrt{H(T)}$, we have

$$P_{\sigma} \left( \left| C_T(\sigma) - H(T) \right| > \xi \sqrt{H(T)} \right) \ll e^{-\frac{1}{4}\xi^2}.$$

This is not very useful when $H(T) < 1$, however. In this case, we expect that $C_T(\sigma)$ will rarely be much more than 1. Theorem 1.9 implies a right-tail bound of

$$P_{\sigma} \left( C_T(\sigma) = k \right) \ll \frac{H(T)^{k-1}}{(k-1)!},$$

whereas the Poisson model predicts that the right side should be smaller, namely $H(T)^k/k!$; but see Homework Exercise 1.2 below.

**Theorem 1.13 (Prime factors in sets).** Let $T_0, T_1, \ldots, T_r$ be a partition of the primes in $[2,x]$ with $T_1, \ldots, T_r$ nonempty. Define

$$H_j = H(\{p-1 : p \in T_j\}) = \sum_{p \in T_j} \frac{1}{p-1} \quad (1 \leq j \leq r).$$
Let \( k_1, \ldots, k_r \geq 0 \) and such that if \( T_0 = \emptyset \) then \( k_1, \ldots, k_r \) are not all zero. Then
\[
\mathbb{P}_x \{ \omega(n; T_j) = k_j \ (1 \leq j \leq r) \} \ll \prod_{j=1}^r \left( \frac{H_{k_j}}{k_j!} e^{-H_j} \right) \left( \eta + \frac{k_1}{H_1} + \cdots + \frac{k_r}{H_r} \right) \\
\leq \prod_{j=1}^r \left( \frac{(H(T_j) + 2)^{k_j}}{k_j!} e^{-H(T_j)} \right),
\]
where \( \eta = 0 \) if \( T_0 \) is empty, \( \eta = 1 \) otherwise.

**Proof.** Let
\[
N = \# \{ n \leq x : \omega(n; T_j) = k_j \ (1 \leq j \leq r) \},
\]
Define \( m_{t,j} = k_j - 1 \) and let
\[
L_t(x) = \sum_{\omega(h; T_j) = m_{t,j} \ (1 \leq j \leq r)} \frac{1}{h} \quad (0 \leq t \leq r).
\]
We use the unique factorization of integers into primes, (the “Wirsing trick”), starting with
\[
\log x \ll \log n = \sum_{p \mid n} \log p^a \quad (x^{1/3} \leq n \leq x).
\]
It follows that
\[
(log x) N \ll \sum_{n \leq x^{1/3}} \log x + \sum_{n \leq x \omega(h; T_j) = k_j \ (1 \leq j \leq r)} \log p^a.
\]
In the first sum, \( \log x \ll \frac{x^{1/3} \log x}{n} \ll \frac{x^{1/2}}{n} \), hence the sum is at most \( \ll x^{1/2} L_0(x) \). In the double sum, let \( n = p^a h \) with \( p \in T_j \). If \( t \geq 1 \) then \( \omega(h; T_j) = k_j - 1 \) and \( \omega(h; T_j) = k_j \) otherwise. That is, we have \( \omega(h; T_j) = m_{t,j} \) for \( 1 \leq j \leq t \). Also, \( p \in T_0 \) is only possible if \( \eta = 1 \). Hence
\[
(log x) N \ll x^{1/2} L_0(x) + \sum_{t=1}^r \sum_{\omega(h; T_j) = m_{t,j} \ (1 \leq j \leq r)} \log p^a.
\]
Using Chebychev’s Estimate for primes or the Prime Number Theorem, the innermost sum over \( p^a \) is \( O(x/h) \) and thus the double sum over \( h, p^a \) is \( O(L_t(x)) \). Also, if \( k_j = 0 \) then there is the sum corresponding to \( t = j \) is empty. This gives
\[
(1.5) \quad \mathbb{P}_x \left( \omega(n; T_j) = k_j \ (1 \leq j \leq r) \right) \ll \frac{1}{\log x} \left( (\eta + x^{-1/2}) L_0(x) + \sum_{1 \leq t \leq r; k_t > 0} L_t(x) \right).
\]
Now we fix \( t \) and bound the sum \( L_t(x) \); if \( t \geq 1 \) we may assume that \( k_t \geq 1 \). Write the denominator \( h = h_1 \cdots h_r h_0 \), where, for \( 1 \leq j \leq r \), \( h_j \) is composed only of primes from \( T_j \) and \( \omega(h_j; T_j) = m_{t,j} \), and further that \( h_0 \) is composed of primes in \( T_0 \). For \( 1 \leq j \leq r \) we have
\[
\sum_{h_j} \frac{1}{h_j} \leq \prod_{p \in T_j} \left( 1 - \frac{1}{p} \right)^{-1} \ll (\log x) \prod_{p \in T_1 \cup \cdots \cup T_r} \left( 1 - \frac{1}{p} \right).
\]
and, using Mertens’ product estimate (0.7),
\[
\sum_{h} \frac{1}{H^r} \leq \prod_{p \in T_0} \left( 1 - \frac{1}{p} \right)^{-1} \ll (\log x) \prod_{p \in T_1 \cup \cdots \cup T_r} \left( 1 - \frac{1}{p} \right).
\]
Thus,
\[
L_t(x) \ll (\log x) \prod_{j=1}^r \frac{H_{m_{t,j}}}{m_{t,j}!} \prod_{p \in T_1 \cup \cdots \cup T_r} \left( 1 - \frac{1}{p} \right).
\]
Using the elementary inequality $1 + y \leq e^y$, we see that
\[
\prod_{p \in T_1 \cup \ldots \cup T_r} \left(1 - \frac{1}{p}\right) \leq e^{-H(T_1) - \cdots - H(T_r)}
\]
and we obtain
\[
(1.6) \quad L_t(x) \ll (\log x) \prod_{j=1}^r \left(\frac{H_{m_{t,j}}^{n_{t,j}}}{m_{t,j}!} e^{-H(j)}\right)
\]
Combining estimates (1.5) and (1.6), we conclude that
\[
N \ll x \left(\eta + x^{-1/2} + \sum_{j=1}^r k_j \frac{H_j}{H_j} \prod_{j=1}^r \left(1 + 1/H_j\right)^{k_j}\right)
\]
By hypothesis, either $\eta = 1$ or $k_j/H_j \geq 1/H_j \gg 1/\log \log x$ for some $j$, and hence the additive term $x^{-1/2}$ may be omitted. This proves the first claim.

Next,
\[
\prod_{j=1}^r \frac{H_j^{k_j}}{k_j!} \left(\eta + \sum_{j=1}^r \frac{k_j}{H_j}\right) \leq \prod_{j=1}^r \frac{H_j^{k_j}}{k_j!} \prod_{j=1}^r \left(1 + 1/H_j\right)^{k_j} = \prod_{j=1}^r \frac{(H_j + 1)^{k_j}}{k_j!}
\]
and we have
\[
H_j = H(T_j) + \sum_{p \in T_j} \frac{1}{p^\omega} \leq H(T_j) + 1.
\]
This proves the final inequality. \qed

**Remark.** A version of Theorem 1.13 is stated in [62, Theorem 2], where only the case $r = 1$ is proved. The full proof may also be found in [30].

**Corollary 1.14 (Hardy-Ramanujan inequality, 1917).** Uniformly for $x \geq 3$ and $k \in \mathbb{N}$ we have
\[
\mathbb{P}_x\{\omega(n) = k\} \ll \frac{(\log \log x + O(1))^{k-1}}{(k-1)!}.
\]
**Proof.** Let $T_1$ consist of all primes $\leq x$. Then $\eta = 0$ and $H_1 = \log \log x + O(1)$ by Mertens’ sum estimate (0.5), and we obtain the desired bound. \qed

Taking as a single set the primes in an interval, we obtain the following very useful corollary.

**Corollary 1.15 (Prime factors in intervals).** Uniformly for $3 \leq t \leq x$ and $0 \leq \lambda \leq 1$, we have
\[
\mathbb{P}_x\{\omega(n, t) \leq \lambda \log \log t\} \ll (\log t)^{-Q(\lambda)}.
\]
Let $\lambda_0 > 1$. Uniformly for $3 \leq t \leq x$ and $1 \leq \lambda \leq \lambda_0$, we have
\[
\mathbb{P}_x\{\omega(n, t) \geq \lambda \log \log t\} \ll_{\lambda_0} (\log t)^{-Q(\lambda)}.
\]
In particular, uniformly for $3 \leq t \leq x$ and $0 \leq \psi \leq \sqrt{\log \log t}$, we have
\[
\mathbb{P}_x\{\omega(n, t) - \log \log t > \psi \sqrt{\log \log t}\} \ll e^{-\frac{1}{2} \psi^2}.
\]
**Proof.** The proof is identical to the proof of Corollary 1.11, using $T = T_1$ as the set of primes in $[2, t]$, $H(T) = \log \log t + O(1)$ from Mertens’ bound (0.5), and Theorem 1.13. \qed

Taking as a special case $t = n$, we recover a strong form of Theorem 1.8.

We can also analyze the distribution of integers composed only of primes factors from a given set.
Corollary 1.16. Let $T$ be a subset of the primes $\leq x$, and let $\mathcal{N}(T)$ denote the set of integers $\leq x$ composed only of primes from $T$. For all $k \geq 1$,
\[
\#\{n \in \mathcal{N}(T) : \omega(n) = k\} \ll \frac{x}{\log x} \cdot \frac{(H(T) + 1)^{k-1}}{(k-1)!}.
\]

Proof. Apply Theorem 1.13 with $T_1 = T$ and $T_2$ being the set of all primes $\leq x$ that are not in $T$, $k_1 = k$ and $k_2 = 0$. Then $\eta = 0$, $H_1 \leq H(T) + 1$ and the result follows. \qed

Remarks. Applying Theorem 1.13 with $T_1$ the set of all primes $\leq x$ that are not in $T$, and $k_1 = 0$, we see that
\[
|\mathcal{N}(T)| \ll e^{-H(T)} x \approx \frac{x}{\log x}
\]
since $H(T) = \log_2 x + O(1)$ by Mertens’ estimate (0.5). Oftentimes we have a corresponding lower bound
\[(1.7) \quad |\mathcal{N}(T)| \gg e^{H(T)} \frac{x}{\log x},
\]
and this allows us to conclude that, conditionally on $n \in \mathcal{N}(T)$, that $\omega(n)$ has an approximate Poisson distribution with parameter $H(T) + O(1)$. That is, combining (1.7) with Corollary 1.16, we obtain
\[
\mathbb{P}_x(\omega(n) = k | n \in \mathcal{N}(T)) \ll e^{-H(T)} \frac{(H(T) + 1)^{k-1}}{(k-1)!},
\]
For example, let $T$ be the set of primes $\leq x$ that are 1 mod 4; in particular, such numbers are the sum of two squares. Then (1.7) follows from a theorem of Landau, and we have $H(T) = \frac{1}{2} \log_2 x + O(1)$ by Mertens’ theorem for arithmetic progressions (0.8). We then conclude that
\[
\mathbb{P}_x(\omega(n) = k | n \in \mathcal{N}(T)) \ll \frac{(\frac{1}{2} \log_2 x + O(1))^{k-1}}{(k-1)! \sqrt{\log x}},
\]
what is, conditional on $n \in \mathcal{N}(T)$, $\omega(n)$ has roughly a Poisson distribution with parameter $\frac{1}{2} \log_2 x$.

If we condition on $\omega(n) = k$, the $r = 2$ case of Theorem 1.13 supplies tail bounds for $\omega(n,T)$. If $X,Y$ are independent Poisson random variables with parameters $\lambda_1, \lambda_2$, respectively, then $X + Y \overset{d}= \text{Pois}(\lambda_1 + \lambda_2)$ and hence, for $0 \leq \ell \leq k$, we have
\[
\mathbb{P}(X = \ell | X + Y = k) = \frac{\mathbb{P}(X = \ell \land Y = k - \ell)}{\mathbb{P}(X + Y = k)} = \frac{e^{-\lambda_1 - \lambda_2}(\lambda_1^\ell / \ell!) (\lambda_2^{k-\ell} / (k-\ell)!)}{e^{-\lambda_1 - \lambda_2}(\lambda_1 + \lambda_2)^k / k!} = \left(\frac{k}{\ell}\right) \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^\ell \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{k-\ell}.
\]
Thus, conditional on $\omega(n) = k$ we expect that $\omega(n,T)$ will have roughly a binomial distribution with parameter $\alpha = H(T)/H(S)$, where $S$ is the set of all primes in $[2,x]$.

Theorem 1.17. Fix $A > 1$ and suppose that $1 \leq k \leq A \log \log x$. Let $T$ be a nonempty subset of the primes in $[2,x]$ and define let $\alpha = H(T)/H(S)$. For any $0 \leq \psi \leq \sqrt{\alpha k}$ we have
\[
\mathbb{P}_x\left(\omega(n,T) - \alpha k \geq \psi \sqrt{\alpha (1-\alpha)k} \bigg| \omega(n) = k\right) \ll_A e^{-\frac{1}{4}\psi^2},
\]
the implied constant depending only on $A$.

We leave the proof as an exercise. It requires the lower bound
\[
\mathbb{P}_x(\omega(n) = k) \gg_A \left(\frac{\log_2 x}{k-1}! \log x\right)^{k-1},
\]
which follows, e.g., from the Sathe-Selberg theorem; see also Theorem 1.23 below.
5. The sequence of cycles and prime factors from intervals

In this section, we take a first look at the random sequence $C_{[m]}(\sigma) (1 \leq m \leq n)$ for $\sigma \in S_n$, and random function $\omega(n, t) (1 \leq t \leq x)$ for integers $n \leq x$. As long as $m$ and $t$ are not too small, it is relatively easy to deduce from Corollaries 1.11 and 1.15 that $C_{[m]}(\sigma)$ is uniformly close to $\log m$ for most $\sigma \in S_n$ and $\omega(n, t)$ is uniformly close to $\log_2 t$ for most $n \leq x$.

**Theorem 1.18.** Let $3 \leq \xi \leq x$. With probability $1 - O((\log \log \xi)^{-1/3})$, we have

$$|\omega(n, t) - \log_2 t| < 2\sqrt{\log_2 t \log_3 t} \quad (\xi \leq t \leq x).$$

**Theorem 1.19.** Let $2 \leq \xi \leq n$. With probability $1 - O(1/(\log \xi)^{1/3})$, we have

$$|C_{[m]} - \log m| < 2\sqrt{\log m \log_2 m} \quad (\xi \leq m \leq n).$$

**Remark 1.20.** When $t$ is bounded, $\omega(n, t)$ has a discrete distribution and we cannot say anything about almost all $n$; in fact it takes every possible value with positive probability; e.g. $\omega(n, 3)$ takes the values $0, 1, 2$ with probabilities (as $x \to \infty$) $\frac{1}{3}, \frac{2}{3}, \frac{1}{6}$, respectively. The same is true for $C_{[m]}$ when $m$ is bounded; see Exercise 1.1.

**Proof.** The proofs of Theorems 1.18 and 1.19 are nearly identical, the latter being simpler due to the discrete nature of the sequence of values of $m$ in question. Thus, we show full details only for Theorem 1.18. Let

$$k_1 = \lfloor \log_2 \xi \rfloor + 1, \quad k_2 = \lfloor \log_2 x \rfloor,$

and for $k_1 \leq k \leq k_2$, let $t_k = e^{\psi k}$. Put $t_{k-1} = \xi$ and $t_{k+1} = x$. For each $k$, $k_1 - 1 \leq k \leq k_2 + 1$, let $F_k$ be the event

$$F_k = \{ |\omega(n, t_k) - \log_2 t_k| \geq 2\sqrt{(k-1)(k-1)-1} \}.$$

As $\log_2 t_k = k + O(1)$ for all $t_k$ (including the endpoints),

$$2\sqrt{(k-1)(k-1)-1} = \psi \sqrt{\log_2 t_k}, \quad \psi = 2\sqrt{\log k + O(1/\sqrt{k})}.$$

Hence, by the third part of the Prime Factors in Intervals Corollary (Cor. 1.15),

$$\mathbb{P} F_k \ll e^{-\frac{\psi^2}{2k}} \ll \frac{1}{k^{4/3}}.$$

Summing over $k$, we see that $F_k$ holds for some $k$ with probability $O(1/k_2^{1/3})$. Now suppose that $F_k$ fails for every $k$ in the range $k_1 - 1 \leq k \leq k_2 + 1$. Let $\xi \leq t \leq x$ and suppose that $t_k < t \leq t_{k+1}$. Evidently,

$$\omega(n, t_k) \leq \omega(n, t) \leq \omega(n, t_{k+1}).$$

By the failure of $F_k$ at every $k$,

$$\omega(n, t) \geq \log_2 t_k + 1 - 2\sqrt{(k-1)(k-1)} \geq \log_2 t - 2\sqrt{\log_2 t \log_3 t}$$

and

$$\omega(n, t) \leq \log_2 t_{k+1} - 1 + 2\sqrt{k \log k} \leq \log_2 t + 2\sqrt{\log_2 t \log_3 t}. \quad \square$$

Theorems 1.18 and 1.19 also tell us about the normal behavior of $p_j(n)$, the $j$-th smallest (distinct) prime factor of $n$, and $D_j(\sigma)$, the length of the $j$-th smallest cycle of $\sigma$ (note that $D_j(\sigma) = D_{j+1}(\sigma)$ for some $j$ when $\sigma$ has cycles of the same length). Since a typical integer has about $\log t$ prime factors $\leq t$, we expect $p_j(n) \approx e^j$. Likewise, a typical permutation $\sigma \in S_n$ has about $\log m$ cycles of length $\leq m$, thus we expect that $D_j(\sigma) \approx e^j$.

**Theorem 1.21 (j-th smallest prime factor).** Let $1 \leq \theta \leq \log_2 x$. For all but $O(x/\theta^{1/3})$ integers $n \leq x$, we have

$$|\log_2 p_j(n) - j| < 3\sqrt{j \log j} \quad (\theta \leq j \leq \omega(n)).$$
Theorem 1.22 (j-th smallest cycle). Let $1 \leq \theta \leq \log n$. With probability $1 - O(\theta^{-1/3})$, we have

$$\left| \log D_j(\sigma) - j \right| < 3\sqrt{j \log j} \quad (\theta \leq j \leq C(\sigma)).$$

Proof. The proof of Corollaries 1.21 and 1.22 are nearly identical, and so we provide details only for Theorem 1.22. We may suppose that $\theta \geq \theta_0$, where $\theta_0$ is a sufficiently large, absolute constant, for otherwise the conclusion of the Corollary is trivial if the implied constant is large enough. Let $\xi = [e^{(2/3)\theta}]$. By Theorem 1.19, with probability $1 - O(1/\theta^{1/3})$, we have

$$|C_m(\sigma) - \log m| < 2\sqrt{\log m \log_2 m} \quad (\xi \leq m \leq n).$$

Also, by Exercise 1.2 (b), with probability $1 - O(1/\xi)$ all the cycles of $\sigma$ of length $\geq \xi$ have distinct lengths. Now suppose that $\sigma$ is a permutation satisfying (1.9), and such that the cycles of $\sigma$ with lengths $\geq \xi$ have distinct lengths. We suppose that $\theta_0$ is so large that the right side of the inequality in (1.9) is at most $\frac{1}{2} \log m$ when $m \geq \xi$. In particular,

$$C_{\lfloor m \rfloor}(\sigma) < \frac{3}{2} \log \xi < \theta,$$

that is, $D_0(\sigma) > \xi$. Thus, we may apply (1.9) with $m = D_j(\sigma)$ for all $\theta \leq j \leq C(\sigma)$. As the cycle lengths $\geq \xi$ are distinct, we have $j = C_{\lfloor m \rfloor}(\sigma) > \frac{1}{2} \log D_j(\sigma)$ and hence

$$|j - \log D_j(\sigma)| < 2\sqrt{\log D_j(\sigma) \log_2 D_j(\sigma)} < 2\sqrt{2j \log(2j)} < 3\sqrt{j \log j}$$

provided that $\theta_0$ is large enough (and hence $j$ is large enough).

Slightly better bounds than those in Theorems 1.18 and 1.19 are attainable, based on ideas stemming from the ‘Law of the Iterated Logarithm’ from probability theory. Essentially one can replace the factor $\log_3 t$ (or $\log_2 m$) with $\log t$ (or $\log_3 m$), and this is best possible. This is deducible, e.g., from the Kubilius model of integers and the analog for permutations; see Section 4 below.

6. Lower bounds on $P_x(\omega(n) = k)$

Theorem 1.23. Fix $A \geq 1$. For some large constant $x_0(A)$, we have uniformly for $x \geq x_0(A)$ and $1 \leq k \leq A \log_2 x$ that

$$P_x(\omega(n) = k) \asymp A \frac{(\log_2 x)^{k-1}}{(k-1)! \log x}.$$

Proof. The upper bound follows from the Hardy-Ramanujan inequality (Theorem 1.14), since

$$(\log_2 x + O(1))^{k-1} \ll_A (\log_2 x)^{k-1} \quad (k \leq A \log_2 x).$$

For the lower bound, WLOG suppose that $A$ is sufficiently large. Let $Q = 10A^2 + 1$, $R = x^{1/100A}$ and $T$ the set of all primes in $[Q, R]$. By Mertens’ theorem (0.5),

$$H := H(T) = \log_2 x + O_A(1).$$

We assume that $x_0$ is large enough so that for $x \geq x_0(A)$ we have

$$H \geq \frac{\log_2 x}{2}.$$

Let $\mathcal{N}$ be the set of integers of the form $p_1 \cdots p_k \leq x$ with $Q \leq p_1 < p_2 < \ldots < p_{k-1} \leq R < p_k$ and such that $p_1 \cdots p_{k-1} \leq x^{1/2}$. Clearly $\omega(n) = k$ for each such $n \in \mathcal{N}$. Given $p_1, \ldots, p_{k-1}$ with product $\leq x^{1/2}$, by the Prime Number Theorem the number of choices for $p_k$ is

$$\pi\left(\frac{x}{p_1 \cdots p_{k-1}}\right) - \pi(R) \gg \frac{x}{p_1 \cdots p_{k-1}}.$$

Thus,

$$P_x(\omega(n) = k) \gg \frac{S_1 - S_2}{\log x},$$

where $S_1$ and $S_2$ are...
Using (1.10), we have $k \leq 2AH$ and hence

$$S_1 \geq \frac{H^{k-1}}{(k-1)!} \left(1 - \frac{k^2}{20H^2A^2}\right) \geq 0.8 \frac{H^{k-1}}{(k-1)!}.$$ 

To bound $S_2$, we use the fact that integers composed of primes below $R$ that are $> x^{1/2}$ are very rare; we will devote the next sections to this type of problem. Let $\alpha = \frac{1}{\log R} = \frac{100A}{\log x}$. Using Proposition 0.1 again,

$$S_2 \leq \sum_{Q \leq p_1 < \ldots < p_{k-1} \leq R} \frac{1}{x^{\alpha/2}(p_1 \cdots p_{k-1})^{1-\alpha}} \leq e^{-50A} \frac{(H')^{k-1}}{(k-1)!}, \quad H' = \sum_{Q \leq p \leq R} \frac{1}{p^{1-\alpha}}.$$ 

Using the inequality $e^x \leq 1 + 2x$ for $0 \leq x \leq 1$ and Mertens’ estimate (0.6),

$$H' \leq \sum_{Q \leq p \leq R} \frac{1 + 2\alpha \log p}{p} \leq H + 2\alpha(\log R - \log Q + O(1)) \leq H + 2$$

if $A$ is large enough. Using (1.10) again yields

$$(H')^{k-1} \leq H^{k-1}(1 + 2/H)^{2AH} \leq e^{4A}H^{k-1}$$

which implies that

$$S_1 - S_2 \geq \frac{H^{k-1}}{(k-1)!} \left(0.8 - e^{4A-50A}\right) \geq \frac{H^{k-1}}{2(k-1)!} = \frac{(\log x - O_A(1))^{k-1}}{2(k-1)!}$$

Again, using that $x \geq x_0(A)$, we conclude that

$$\mathbb{P}_x(\omega(n) = k) \gg A \left(\frac{\log x}{k-1}\right)^{k-1}(k-1)! \log x,$$

as required.

Much stronger bounds are known for $\mathbb{P}_x(\omega(n) = k)$ in a wide range of uniformity in $k$. The method are complex-analytic and Chapter II.6 in [61] is devoted to this subject.

### 7. Prime factors counted with multiplicity

When prime factors of an integer are counted with multiplicity, that is, counted by means of the function $\Omega(n)$, the normal behavior is the same as for the function $\omega(n)$. That is, for integers $n \leq x$, $\Omega(n)$ is tightly concentrated near $\log_2 x$. However, the behavior changes “out in the right tail” region, owing to the influence of large powers of small primes.

Here, we provide a general use utility for analyzing $\Omega(n)$ and other functions. It is often convenient to use $\Omega(n)$, rather than $\omega(n)$, in applications because $\Omega(n)$ is completely additive ($\Omega(ab) = \Omega(a) + \Omega(b)$ for every $a, b$). It is based on the method of parameters, used to capture tails of the distribution of a random variable (cf. Chernoff’s inequality), sometimes referred to as “Rankin’s trick” in the literature.

**Lemma 1.24 (Halberstam-Richert).** Let $f$ be a non-negative, real valued multiplicative function and define

(a) $A(x) = \frac{1}{x} \sum_{p \leq x} f(p) \log p \quad (x \geq 1)$;
1. THE SEQUENCE OF PRIME FACTORS OF AN INTEGER AND CYCLES OF A PERMUTATION, I

\[ B(x) = \sum_{p} \sum_{k \geq 2} \frac{f(p^k)}{p^k} \log p^k. \]

Then, for all \( x > 1 \) we have

\[
\sum_{n \leq x} f(n) \leq (A + B(x) + 1) \frac{x}{\log x} \sum_{n \leq x} \frac{f(n)}{n} \quad A := \max_{y \leq x} A(y)
\]

\[
\leq (A + B(x) + 1) e^{B(x)} \frac{x}{\log x} \exp \left( \sum_{p \leq x} \frac{f(p)}{p} \right).
\]

**Proof.** Fix \( x \geq 1 \), let \( A = \max_{y \leq x} A(y) \), \( B = B(x) \) and also define

\[
M(x) = \sum_{n \leq x} f(n), \quad L(x) = \sum_{n \leq x} \frac{f(n)}{n}.
\]

We begin in a similar way to the proof of The Prime Factors in Sets Theorem (Thm. 1.13). Since \( \log u \leq u \),

\[
M(x) \log x = \sum_{n \leq x} f(n) \log(x/n) + \sum_{n \leq x} f(n) \sum_{p^k \mid n} \log p^k \quad (n = p^k h)
\]

\[
\leq xL(x) + \sum_{p^k \leq x} (\log p^k) f(p^k) \sum_{h \leq x/p} f(h)
\]

\[
\leq xL(x) + \sum_{p^k \leq x} (\log p^k) f(p^k) \frac{x}{p^k} \sum_{h \leq x/p} f(h) + \sum_{p \leq x} f(p) \log p \sum_{h \leq x/p} f(h).
\]

Recalling (b), the first double sum over \( p^k \) and \( h \) is bounded by \( BxL(x) \). Invoking (a),

\[
\sum_{p \leq x} f(p) \log p \sum_{h \leq x/p} f(h) = \sum_{h \leq x/p} f(h) \sum_{p \leq x/h} f(p) \log p \leq A(x/h)x \sum_{h \leq x} \frac{f(h)}{h} \leq AxL(x).
\]

We obtain

\[
M(x) \log x \leq (1 + B + A)xL(x),
\]

which completes the proof of the first asserted inequality. For the second, we invoke (b) again, using (0.3),

\[
L(x) \leq \sum_{P^{+}(n) \leq x} \frac{f(n)}{n} = \prod_{p \leq x} \left( 1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \cdots \right)
\]

\[
\leq \exp \left( \sum_{p \leq x} \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \cdots \right)
\]

\[
\leq \exp \left( B + \sum_{p \leq x} \frac{f(p)}{p} \right).
\]

**Corollary 1.25.** Let \( T \) be a subset of the primes in \([2, x]\), and let \( 1 \leq y_0 < \min T \). Uniformly for \( 1 \leq y \leq y_0 \) we have

\[
\sum_{n \leq x} y^{\Omega(n, T)} \ll_{y_0} xe^{(y - 1)H(T)}.
\]

**Proof.** The function \( f(n) = y^{\Omega(n, T)} \) is multiplicative, with \( f(p^k) = 1 \) for \( p \not\in T \) and \( f(p^k) = y^k \) if \( p \in T \). Thus,

\[
\frac{1}{u} \sum_{p \leq u} f(p) \log p \leq \frac{y}{u} \sum_{p \leq u} \log p \ll y.
\]
by the Prime Number Theorem. Thus, $A(u) \ll y_0 \ll y_0$. Also,

$$B(x) \ll \sum_{p} \sum_{k \geq 2} \frac{f(p^k)}{p^k} \log p^k = \sum_{p \leq T} \sum_{k \geq 2} \log p^k + \sum_{p \leq T} \sum_{k \geq 2} \frac{y^k \log p^k}{p^k}$$

$$\ll 1 + \frac{\sum_{p \leq T} (\log p) \sum_{k=2}^\infty k \left(\frac{y}{p}\right)^k}{\sum_{k=2}^\infty}$$

$$= 1 + \frac{\sum_{p \leq T} (\log p) \left(\frac{y}{p}\right) + (\frac{y}{p})^2 - \left(\frac{y}{p}\right)^3}{\sum_{p \leq T} (\log p) \left(\frac{y}{p}\right)^2}$$

$$\ll 1 + \frac{\sum_{p \geq \min T} \log p}{\sum_{p \leq \min T} p(p - y)}$$

$$\ll y_0 \frac{1}{\min T - y} \ll y_0$$

since $y \leq y_0 < \min T$. Hence, the hypotheses of Lemma 1.24 hold, with bounded $A(x), B(x)$, the bounds depending on $y_0$. We conclude that

$$\sum_{n \leq x} g^{\Omega(n, T)} \ll y_0 \frac{x}{\log x} \exp \left(\sum_{p \leq x} \frac{f(p)}{p}\right) = \frac{x}{\log x} \exp \left(\sum_{p \leq x} \frac{1}{p} \sum_{p \leq T} \frac{1}{p}\right)$$

$$\ll \frac{x}{\log x} \exp \left(\sum_{p \leq x} \frac{y - 1}{p} \sum_{p \leq T} \frac{1}{p}\right)$$

$$\ll \frac{x}{\log x} \exp \left((y - 1)H(T) + \log_2 x + O(1)\right),$$

since $y \geq 1$, and where Mertens’ bound (0.5) was used in the last step. 

**Corollary 1.26.** Let $1 < \lambda_0 < 2$. Uniformly for $1 \leq \lambda \leq \lambda_0$ and $3 \leq t \leq x$,

$$P_{x, \lambda} \{\Omega(n, t) \geq \lambda \log_2 t\} \ll_{\lambda_0} (\log t)^{-Q(\lambda)}.$$

**Proof.** Let $T$ be the set of primes in $[2, t]$. By Mertens’ bound (0.5), $H(T) = \log_2 t + O(1)$. By Corollary 1.25,

$$\#\{n \leq x : \Omega(n, t) \geq \lambda \log_2 t\} \leq \sum_{n \leq x} \lambda^{\Omega(n, t) - \lambda \log_2 t}$$

$$\ll_{\lambda_0} \lambda^{-\lambda \log_2 t} x e^{(\lambda - 1)H(T)} \ll_{\lambda_0} x (\log t)^{-Q(\lambda)}.$$

**Remark 1.27.** When $\lambda \geq 2$, the behavior of the quantity in Corollary 1.26 is different than that of the quantity in Corollary 1.15. This is due to the behavior of powers of small prime factors, most important being powers of 2, and in fact

$$\#\{n \leq x : \Omega(n, t) \geq \lambda \log_2 t\} \approx x (\log t)^{-Q(2) - (\log 2)(\lambda - 2)} = \frac{x \log t}{2^{\lambda \log_2 t}}.$$  

8. Application: Erdős’ multiplication table problem

In 1955, Erdős [22] posed the following problem: Estimate the number, $A(N)$, of distinct products of the form $ab$ with $a \leq N$, $b \leq N$. Erdős proved that $A(N) = o(N^2)$, and later in 1960 [23] refined the
estimates to prove that $A(N) = N^2(\log N)^{-E + o(1)}$, where 

$$E = Q \left( \frac{1}{\log 2} \right) = 1 - \frac{1 + \log\log 2}{\log 2} = 0.08607\ldots.$$ 

**Theorem 1.28.** We have $A(N) \ll N^2(\log N)^{-E}$.

**Proof.** Let $k_0 = \frac{\log_2 N}{\log 2}$. By Corollary 1.26 (with $t = N$), the number of distinct products with $\Omega(ab) \geq k_0$ is bounded above by 

$$\#\{m \leq N^2 : \Omega(m) \geq k_0\} \ll N^2(\log N)^{-Q(1/\log 2)} = N^2(\log N)^{-E}.$$ 

There are $O(N)$ products with $a = 1$ or $b = 1$. If $a > 1, b > 1$ and $\Omega(ab) < k_0$, then $\omega(a) = h, \omega(b) = j$ with $h \geq 1, j \geq 1$ and $h + j = k < k_0$. The number of pairs $a, b$ with a fixed $h, j$ is, by Theorem 1.23,

$$\ll \frac{N^2(\log^2 N)^{h+j-2}}{(\log N)^2(h-1)!(j-1)!}.$$ 

Summing first over all $j, h$ with $h + j = k$ using the binomial theorem, and then over $k < k_0$ we obtain an upper bound for the total number of pairs $a, b$ with $\Omega(ab) < k_0$ of

$$\ll \frac{N^2}{\log^2 N} \sum_{k < k_0} \frac{(2 \log_2 N)^{k-2}}{(k-2)!} \ll N^2(\log^2 N)^{-Q(1/\log 4)} = N^2(\log N)^{-E}$$

upon invoking Proposition 0.3. \hfill \Box

**Remarks.** The choice of $k_0$ is motivated by $\lambda = \frac{1}{\log 2}$ being the unique solution of $2Q(\lambda/2) = Q(\lambda)$.

### 9. Number of divisors of integers

The number, $\tau(n)$, of positive divisors of $n$, is closely related to the distribution of $\omega(n)$. From the formula

$$\tau(n) = \prod_{p^a || n} (a + 1)$$

and the elementary inequality $2 \leq a + 1 \leq 2^a$, it follows that

$$2^{\omega(n)} \leq \tau(n) \leq 2^{\Omega(n)}.$$ 

By a classical theorem of Dirichlet, $E_x \tau(n) \sim \log x$ as $x \to \infty$, so the average (mean) of $\tau(n)$ for $n \leq x$ is about $\log x$.

What is $\tau(n)$ for a “typical” $n \leq x$? By Theorem 1.15 and Corollary 1.26, if $\xi = \xi(x) \to \infty$ as $x \to \infty$ then for almost all $n \leq x$ we have

$$\log_2 x - \xi \sqrt{\log_2 x} \leq \omega(n) \leq \Omega(n) \leq \log_2 x + \xi \sqrt{\log_2 x},$$

and therefore for such $n$ it follows that

$$\tau(n) = (\log x)^{\log 2} \exp\{O(\xi \sqrt{\log_2 x})\} = (\log x)^{\log 2 + o(1)}$$

as $x \to \infty$.

Hence, the mode is much smaller than the mean.

**Further analysis of the sum $\sum_{n \leq x} \tau(n)$:** As we have just seen, this sum must be dominated by unusual integers, those with an abnormally large number of prime factors. But how large? Heuristically, most integers have few repeated prime factors (see Exercise 1.5), so that $\tau(n) \approx 2^{\omega(n)}$. The number of $n \leq x$ with $\omega(n) = k$ has order about $x^{\log_2 x^k k!}$, so we get

$$\sum_{n \leq x} \tau(n) \approx \sum_{k} 2^k x^{\log_2 x}^{k} = \frac{x}{\log x} \sum_{k} \frac{(2 \log_2 x)^k}{k!}.$$
the sum over \( k \) has a peak around \( k = 2 \log_2 x \), so we expect that the sum is dominated by integers with \( \omega(n) \sim 2 \log_2 x \). Moreover, the distribution is roughly Poisson with parameter \( 2 \log_2 x \), which is well-approximated by a Gaussian (Proposition 5.1). This motivates the next result.

**Theorem 1.29.** Let \( 1 \leq \psi \leq \sqrt{\log_2 x} \). Then
\[
\sum_{n \leq x, |\omega(n) - 2 \log_2 x| > \psi \sqrt{\log_2 x}} \tau(n) \ll (x \log x) e^{-\frac{1}{12} \psi^2}.
\]

**Proof.** Let \( \lambda = \frac{\psi}{\sqrt{\log_2 x}} \in [0, 1] \). Let \( 1 \leq t \leq 2 \). Then
\[
\sum_{\omega(n) \geq (2 + \lambda) \log_2 x} \tau(n) \ll \sum_{n \leq x} \tau(n) t^{\omega(n) - (2 + \lambda) \log_2 x} = t^{-(2 + \lambda) \log_2 x} \sum_{n \leq x} \tau(n) t^{\omega(n)}.
\]
The summand is multiplicative, and satisfies the conditions of Lemma 1.24. Hence
\[
\sum_{n \leq x} \tau(n) t^{\omega(n)} \ll \frac{x}{\log x} \exp\left( \sum_{p \leq x} \frac{2t}{p} \right) \ll x (\log x)^{2t - 1}
\]
and therefore
\[
\sum_{\omega(n) \geq (2 + \lambda) \log_2 x} \tau(n) \ll x (\log x)^{2t - 1 - (2 + \lambda) \log t}.
\]
The optimum value of \( t \) to minimize the right side is \( t = 1 + \frac{\lambda}{2} \), and then the exponent of \( \log x \) is
\[
1 - Q(1 + \frac{\lambda}{2}) \leq 1 - \frac{1}{12} \lambda^2
\]
using (0.16).

Similarly, taking \( t = 1 - \frac{\lambda}{2} \), we obtain with a second application of Lemma 1.24 the estimate
\[
\sum_{\omega(n) \leq (2 - \lambda) \log_2 x} \tau(n) \ll t^{-(2 - \lambda) \log_2 x} \sum_{n \leq x} \tau(n) t^{\omega(n)} \ll x (\log x)^{-2Q(1 - \lambda/2)} \ll x (\log x)^{1 - \frac{1}{12} \lambda^2}.
\]
Finally, \( (\log x)^{\lambda^2} = e^{\psi^2} \) and the proof is complete. \( \square \)

**10. The range of Euler’s function**

Let \( \phi(n) \) be Euler’s “totient” function, i.e., the number of integers \( m \in [n] \) that are relatively prime to \( n \). Let \( V \) be the image of \( \phi \), i.e. \( V = \{1, 2, 4, 6, 8, 10, 12, 16, \ldots\} \), and let \( V(x) \) be the number of elements of \( V \) that are \( \leq x \), e.g. \( V(15) = 7 \). Since \( \phi(p) = p - 1 \) for all primes \( p \), we have \( V(x) \geq \pi(x + 1) \gg x / \log x \) by the Prime Number Theorem. Here we show an upper bound which is very close to this.

**Theorem 1.30 (Erdős, 1935 [20]).** We have \( V(x) = x (\log x)^{-1 + o(1)} \).
Proof. Fix an integer $m \geq 10$, let $M = p_1 p_2 \cdots p_m$ be the product of the first $m$ primes. Let

$$J_m = \{1 \leq j \leq M : (j, M) = 1, \ (j - 1, M) \geq \log m\}.$$ 

For each $j \in [M]$ with $(j, M) = 1$ let $P_j = \{p \leq x : p \equiv j \pmod{M}\}$ and

$$P = \bigcup_{j \in J_m} P_j.$$ 

We first show that $|J_m|/\phi(M) \to 1$ as $m \to \infty$. Let $2|\ell|/M$. Then

$$\#\{1 \leq j \leq M : (j, M) = 1, (j - 1, M) = \ell\} = \prod_{p|M/\ell} (p - 2) \equiv \phi(M) \prod_{p|M/\ell} \frac{p - 2}{p - 1} \prod_{p > 2} \frac{1}{p - 2}.$$ 

Since $p - 2 > \sqrt{p}$ for $p \geq 5$, the inner product is $\ll \sqrt{3/\ell}$. The first product is $\ll 1/\log m$ by Mertens. Hence, summing over $\ell < \log m$ we have

$$\phi(M) - |J_m| = \#\{1 \leq j \leq M : (j, M) = 1, (j - 1, M) < \log m\} \ll \frac{\phi(M)}{\log m} \sum_{\ell < \log m} \frac{1}{\ell} \ll \frac{\phi(M)}{(\log m)^{1/2}}.$$ 

Hence,

$$|J_m| \sim \phi(M) \quad (m \to \infty).$$

Fix $\varepsilon > 0$. By an elementary estimate, for some large constant $C$, if $\phi(n) \leq x$ then $n \leq y := Cx \log_2 x$. Let $c_m = |J_m|/\phi(M)$. By Mertens theorem in arithmetic progressions (0.8),

$$H(P) = c_m \log_2 x + O_m(1).$$

Let $\delta = 2/\log_2 m$ and suppose $m$ is large enough so that $0 < \delta < 1/2$. Also, $\delta \to 0$ as $m \to \infty$. By Theorem 1.13, followed by (0.3) and Stirling’s formula,

$$\#\{n \leq y : \omega(n; P) < \delta \log_2 x\} \ll \frac{y}{(\log x)^{1-c_m}} \sum_{k < \delta \log_2 x} \frac{(1 - c_m) \log_2 x + O_m(1)}{k!} \ll \frac{x \log_2 x}{(\log x)^{1-c_m-\delta \log(e/\delta)}}.$$ 

If $m$ is large enough, the exponent of $\log x$ is $> 1 - \varepsilon$.

Now let $W$ be the number of values of Euler’s function which have the form $\phi(n)$, where $\omega(n; P) \geq L := \lfloor \delta \log_2 x \rfloor$. Suppose that $K \geq L$ and $\{k_j : j \in J_m\}$ is a vector of non-negative integers with sum $K$. If $\omega(n; P_j) = k_j$ for all $j \in J_m$ then $\phi(n)$ is divisible by $\prod_{j \in J_m} (j - 1, M)^{k_j}$. Thus,

$$W \ll \sum_{K \geq L} \frac{x}{(\log m)^K} \sum_{\sum k_j = K} \prod_{j \in J_m} (j - 1, M)^{k_j}$$

$$\ll \sum_{K \geq L} \frac{x}{(\log m)^K} \sum_{\sum k_j = K} \frac{x}{(\log m)^K}$$

$$\ll \frac{x(K + 1)^{|J_m|}}{(\log m)^K} \ll \frac{x(K + 1)^{|J_m|}}{(\log m)^K} \ll \frac{x(\log_2 x)^{|J_m|}}{(\log x)^{\delta \log_2 m}} \ll \frac{x(\log_2 x)^{|J_m|}}{(\log x)^{\delta \log_2 m}} \ll \frac{x}{\log x}.$$ 

This completes the proof. $\square$
11. Exercises

**Exercise 1.1.** (a) Derive the following general inclusion-exclusion formula, valid for any non-negative integer \( u \):

\[
I(u = m) = \sum_{j=m}^{\infty} (-1)^{j-m} \binom{j}{m} \binom{u}{j}.
\]

(b) Let \( 1 \leq j \leq n \) and \( 0 \leq m \leq n/j \). Using part (a), derive an exact formula for the number of permutations \( \sigma \in S_n \) with \( C_j(\sigma) = m \).

(c) With \( j, m \) fixed, evaluate

\[
\lim_{n \to \infty} \mathbb{P}_{\sigma \in S_n} (C_j(\sigma) = m).
\]

**Exercise 1.2.** (a) Show that if \( T \) is a nonempty subset of \([n]\), and \( k \geq 0 \), then

\[
\mathbb{P}_\sigma (C_T(\sigma) \geq k) \leq \frac{H(T)^k}{k!}.
\]

(this is sometimes stronger than Theorem 1.9, especially if \( H(T) \) is small).

(b) Show that the probability that a permutation \( \sigma \in S_n \) has two cycles of the same length \( \geq \ell \), is \( O(1/\ell) \).

**Exercise 1.3.** Show that \( \mathbb{E} Z = n + 1 \). Contrast this with the behavior of \( 2^{C(\sigma)} \) for most \( \sigma \in S_n \).

**Exercise 1.4.** Let \( 1 \leq k \leq n \). Show that if \( T \) is a nonempty subset of \([n]\) with \( \max T \leq n/k \), then

\[
\mathbb{E}_\sigma C_T(\sigma)^k = \mathbb{E} Z^k,
\]

where \( Z \overset{d}{=} \text{Pois}(H(T)) \).

**Exercise 1.5.** (a) Show that if \( T \) is a nonempty subset of the primes in \([2, x]\), and \( k \geq 0 \), then

\[
\mathbb{P}_x \{ \omega(n, T) \geq k \} \leq \frac{H(T)^k}{k!}.
\]

(this is sometimes stronger than Theorem 1.13, especially if \( H(T) \) is small).

(b) Show that the number of \( n \leq x \) that have two prime factors in some dyadic interval of the form \((z, 2z] \) with \( z > y \), is \( O(x/\log y) \).

**Exercise 1.6.** (a) Prove that \( \mathbb{E}_{\sigma \in S_n} \frac{1}{C(\sigma)} \sim \frac{1}{\log n} \) as \( n \to \infty \).

(b) Prove that \( \mathbb{E}_x \left( \frac{1}{\omega(n)} \mathbb{I}(n \geq 2) \right) \sim \frac{1}{\log_2 x} \) as \( x \to \infty \).

**Exercise 1.7.** Provide full details for the proof of Corollary 1.12.

**Exercise 1.8.** Starting with Lemma 1.24, prove Corollary 1.15 using the method used to prove Corollary 1.26.

**Exercise 1.9.** Prove Theorem 1.17 using Lemma 0.5 and Theorem 1.23.

**Exercise 1.10.** Provide full details of the proof of Theorem 1.19.

**Exercise 1.11.** Provide full details of the proof of Theorem 1.21.
CHAPTER 2

Distribution of the largest cycle and largest prime factor

1. Upper bounds

Theorem 1.9 implies that

\[ \nu(n,m) := \Pr_{\sigma \in S_n}(C_{m,n}(\sigma) = 0) \leq e^{1-H_n+H_m} \leq e^{2m/n} \quad 1 \leq m \leq n. \]

When \( m \) is small, however, the number of cycles is at least \( n/m \) and this is extremely rare by Theorem 1.11. We can argue heuristically as follows: \( C_1(\sigma), \ldots, C_m(\sigma) \) behaves like a set of independent Poisson variables \( Z_1, \ldots, Z_m \) with \( Z_j \overset{d}{=} \text{Pois}(1/j) \). Thus, the event \( C_{m,n}(\sigma) = 0 \) can be modeled by the event \( Z_1 + 2Z_2 + \cdots + mZ_m = n \). Using the ideas behind Chernoff’s inequality, and Proposition 0.2, for any \( w \geq 1 \) we have

\[ \Pr_{\sigma}(Z_1 + 2Z_2 + \cdots + mZ_m = n) \leq \mathbb{E}_{\sigma} w^{Z_1+2Z_2+\cdots+mZ_m-n} = w^n \exp \left\{ \frac{w-1}{1} + \frac{w^2-1}{2} + \cdots + \frac{w^m-1}{m} \right\}. \]

Optimizing the choice of \( w \) will show that the RHS is decaying very rapidly as a function of \( u = n/m \).

We utilize this idea to show the following.

**Theorem 2.1 (No large cycles).** Uniformly for \( 1 \leq m \leq n \) we have

\[ \nu(n,m) \leq e^{-u \log u + u - 1}, \quad u = n/m. \]

**Proof.** Let \( w = u^{1/m} \). Following the heuristic above, we first write

\[ \nu(n,m) \leq \mathbb{E}_{\sigma \in S_n} w^{C_1(\sigma)+2C_2(\sigma)+\cdots+mC_m(\sigma)-n}. \]

For each \( j \in [m] \), write \( w^j = 1 + (w^j - 1) \). By the binomial theorem and Lemma 1.1,

\[ \nu(n,m) \leq w^{-n} \mathbb{E}_{\sigma \in S_n} \prod_{j=1}^{m} \left( \sum_{k_j=0}^{\infty} (w^j - 1)^{k_j} \binom{C_j(\sigma)}{k_j} \right) \]

\[ = w^{-n} \sum_{k_1,\ldots,k_m \geq 0} (w-1)^{k_1} \cdots (w^m-1)^{k_m} \mathbb{E}_{\sigma \in S_n} \binom{C_1(\sigma)}{k_1} \cdots \binom{C_m(\sigma)}{k_m} \]

\[ \leq w^{-n} \sum_{k_1,\ldots,k_m \geq 0} (w-1)^{k_1} \cdots (w^m-1)^{k_m} \prod_{j=1}^{m} \frac{(1/j)^{k_j}}{k_j!} \]

\[ = w^{-n} \exp \left\{ \frac{w-1}{1} + \frac{w^2-1}{2} + \cdots + \frac{w^m-1}{m} \right\}. \]

The mean value theorem implies that \( w^j = u^{j/m} \leq 1 + (u-1)j/m \) for \( 1 \leq j \leq m \) and hence

\[ w - 1 + \frac{w^2-1}{2} + \cdots + \frac{w^m-1}{m} \leq \sum_{j=1}^{m} \frac{(u-1)j/m}{j} = u - 1. \]

We conclude that

\[ \nu(n,m) \leq u^{-n/m} e^{-u-1} = e^{-u \log u + u - 1}. \]
Remarks. The upper bound in Theorem 2.1 is reasonably sharp throughout the range on \( n, m \). For example, if \( m = 1 \) then
\[
A(1, n) = \frac{1}{n!} = \frac{1}{u!} = e^{-u \log u + u - (1/2) \log u + O(1)}
\]
by Stirling’s formula.

Let \( \Psi(x, y) = \# \{ n \leq x : P^+(n) \leq y \} \). These are known as \( y \)-smooth, or \( y \)-friable numbers. Applying Theorem 1.13 with \( T_1 \) being the set of all primes in \( (y, x] \), and \( k_1 = 0 \), we have \( H(T_1) = \log_2 x - \log_2 y + O(1) \) by Mertens’ estimate (0.5), and hence
\[
\Psi(x, y) \ll x \frac{\log y}{\log x}.
\]
When \( \log y \) is much smaller than \( \log x \), one can do substantially better using the ideas behind the proof of Theorem 2.1.

**Theorem 2.2.** Uniformly for \( x \geq 10 \) and \( \log x \leq y \leq x \) we have
\[
\Psi(x, y) \leq xe^{-u \log u + O(u)}, \quad u = \frac{\log x}{\log y},
\]
Remarks. There is a change of behavior around \( y = \log x \), due to the fact that for smaller \( y \), if \( \prod_{p \leq y} p \approx x \), then some of the exponents of primes dividing \( n \) must be large.

We note some special cases which we will find useful for applications:

\[
\Psi(x, \log x) \ll \frac{x^{\alpha(1)}}{\log x}, \quad 0 < \alpha < \frac{2}{3}, (x \to \infty),
\]
\[
\Psi(x, (\log x)^c) \ll x^{1-1/c + o(1)} (x \to \infty)
\]
for any fixed \( c \geq 1 \), and
\[
\Psi(x, x^{c(\log_3 x - \log_2 x)} / (\log x)^{c+o(1)}) \ll \frac{x}{(\log x)^{c+o(1)}} (x \to \infty).
\]

**Proof of Theorem 2.2.** Define
\[
\alpha = 1 - \frac{\log u}{\log y}
\]
By our hypothesis that \( \log x \leq y \leq x \),
\[
1 \leq u \leq \frac{\log x}{\log_2 x}, \quad \frac{\log_3 x}{\log_2 x} \leq \alpha \leq 1, \quad x^{1-\alpha} = e^{u \log u}.
\]
Define
\[
S := \sum_{P^+(n) \leq y} \frac{1}{n^\alpha} = \prod_{p \leq y} \left( 1 + \frac{1}{p^\alpha} + \frac{1}{p^{2\alpha}} + \cdots \right) = \prod_{p \leq y} \left( 1 + \frac{1}{p^\alpha - 1} \right) \leq \exp \left( \sum_{p \leq y} \frac{1}{p^\alpha - 1} \right).
\]
In the case \( 0 < \alpha < 2/3 \), we use the simple bound
\[
\Psi(x, y) \leq \sum_{n \leq x, P^+(n) \leq y} \left( \frac{x}{n} \right)^\alpha \leq x^\alpha S = xe^{-u \log u} S.
\]
We have \( y \leq u^3 \leq \log^3 x \) and \( u = \frac{\log x}{\log_2 x} \). When \( p \leq 2^{1/\alpha} \) we have \( p^\alpha - 1 \gg \alpha \log p \), and when \( p > 2^{1/\alpha} \) we have \( p^\alpha - 1 \geq \frac{1}{2} p^\alpha \). Thus, using (2.5) again and ignoring that \( p \) is prime,

\[
\log S \leq \sum_{2 \leq n \leq 2^{1/\alpha}} O(1) + \sum_{n \geq 2} \frac{2 \log n}{n^\alpha} \ll \frac{1}{\alpha} \int_2^{2^{1/\alpha}} \frac{dt}{\log t} + \int_1^y \frac{dt}{t^\alpha} \\
\ll 2^{1/\alpha} + \frac{y^{1-\alpha}}{1-\alpha} \\
\ll (\log x)^{\alpha(1)} + u \ll u.
\]

This completes the proof in the case \( 0 < \alpha < 2/3 \).

Next, assume that \( \alpha \geq 2/3 \). For all \( w > 0 \), \( \log w \leq w \) and thus for \( n \leq x \), \( \log(x/n) = \alpha^{-1} \left( \log(x/n) \right)^\alpha \leq \alpha^{-1} (x/n)^\alpha \). Hence,

\[
(\log x) \Psi(x, y) = \sum_{n \leq x \atop P^+(n) \leq y} \log(x/n) + \log n \\
\ll x^\alpha S + \sum_{n \leq x \atop P^+(n) \leq y} \log P
\tag{2.7}
\]

In the double-sum on the right side of (2.7), let \( n = p^k m \), and separate into cases depending on \( k = 1 \) or \( k > 1 \). The \( k = 1 \) terms contribute

\[
\leq \sum_{m \leq x \atop P^+(m) \leq y} \log P \ll \sum_{m \leq x \atop P^+(m) \leq y} \min(y, x/m) \ll \sum_{m \leq x \atop P^+(m) \leq y} y^{1-\alpha} \left( \frac{x}{m} \right)^\alpha = u x^\alpha S.
\]

The terms with \( k \geq 2 \) contribute

\[
\leq \sum_{p \leq y \atop 2 \leq k \leq \frac{\log x}{\log p}} \log P \sum_{m \leq x \atop P^+(m) \leq y \atop p^k \leq m} 1 \leq \sum_{p \leq y \atop 2 \leq k \leq \frac{\log x}{\log p}} \log P \sum_{P^+(m) \leq y \atop p^k \leq m} \left( \frac{x}{p^k m} \right)^\alpha \ll x^\alpha S.
\]

Therefore, by (2.5),

\[
\Psi(x, y) \ll \frac{u}{\log x} x^\alpha S = \frac{x e^{-u \log u}}{\log y} S.
\tag{2.8}
\]

It remains to bound \( S \). Since \( \alpha \geq 2/3 \), \((p^\alpha - 1)^{-1} = p^{-\alpha} + O(1/p^{4/3}) \). For any \( 0 \leq x \leq 1 \), the mean value theorem implies that \( u^x \leq 1 + (u - 1)x \), hence

\[
\frac{1}{p^\alpha} = \frac{1}{p} \frac{\log x}{\log y} \ll \frac{1}{p} \left( 1 + (u - 1) \frac{\log p}{\log y} \right)
\]

Thus, by Mertens bounds (0.5) and (0.6),

\[
\log S \leq O(1) + \sum_{p \leq y} \frac{1}{p^\alpha} \\
\leq O(1) + \sum_{p \leq y} \frac{1}{p} + \frac{u - 1}{\log y} \sum_{p \leq y} \frac{\log p}{p} \\
\leq \log_2 y + O(u + 1).
\]

Thus, \( S \ll (\log y)e^{O(u)} \). Combining this with (2.8), we get the claimed bound in the case \( \alpha \geq 2/3 \). \(\square\)
When \( y \leq \log x \), simpler bounds are possible, since \( P^+(n) \leq y \) implies that \( n = \prod_{p \leq y} p^{a_p} \). Therefore, we have the equality

\[
\Psi(x, y) = \#\left\{ (a_p)_{p \leq y} : \sum_{p \leq y} a_p \log p \leq \log x : 0 \leq a_p (p \leq y) \right\},
\]

and good bounds can be arrived at by combinatorial counting; see, e.g. Section III.5.2 in \[61\].

### 2. Application: large gaps between primes

Let \( p_n \) denote the \( n \)th prime, and let

\[
G(X) := \max_{p_{n+1} \leq X} (p_{n+1} - p_n)
\]
denote the the maximum gap between consecutive primes less than \( X \).

In 1931, Westzynthius \[65\] proved that infinitely often the gap between consecutive prime numbers can be an arbitrarily large multiple of the average gap, which is \( \sim \log X \) by the PNT. After improvements by Ricci in 1934 and Erdős in 1935, in 1938 Rankin \[54\] proved that

\[
G(X) \geq \frac{\log X \log_2 X \log_4 X}{(\log_3 X)^2}.
\]

This was not improved until August 2014, in two independent papers of Ford-Green-Konyagin-Tao and Maynard. Later, Ford, Green, Konyagin, Maynard, and Tao \[31\] established the current world record

**Theorem 2.3 (Ford-Green-Konyagin-Maynard-Tao \[31, Theorem 1\]).** We have

\[
G(X) \geq \frac{\log X \log_2 X \log_4 X}{\log_3 X}
\]

for sufficiently large \( X \).

Here we prove Rankin’s bound (2.9) using a very simple argument.

**Idea #1.** Let \( x \) be the largest integer such that \( P(x) \leq X/3 \), where \( P(x) \) is the product of primes below \( x \). By the PNT, \( x \sim \log X \). Let \( J(x) \) be the largest gap between numbers that are coprime to \( P(x) \); the set of such numbers is periodic modulo \( P(x) \), so \( J(x) \) exists. Such a gap occurs between \( P(x) \) and \( 3P(x) \), that is, below \( X \). Each number in the gap has a prime factor \( \leq x < P(x) \), thus these numbers are composite. Therefore, \( G(X) \geq J(x) \).

**Idea #2.** Suppose that the integers coprime to \( P(x) \) have a gap \([u, u+y]\) of length \( y \). For each prime \( p \leq x \), let \( a_p \) be the residue class \(-u \mod p \). Then the set of residue classes \( a_p \), for \( p \leq x \), cover all integers in \([1, 2, \ldots, y-1]\); for all \( 1 \leq j \leq y-1 \), there is a \( p \leq x \) with \( p|(u+j) \), hence \( j \equiv a_p \mod p \). Conversely, if we can find residue classes \( a_p \), one for each prime \( p \leq x \), that cover \([1, y-1]\), then \( G(X) \geq J(x) \geq y \).

**Rankin’s argument, based on earlier work of Westzynthius and Erdős.** Suppose that \( x < y < x\log x \), let

\[
z = y^{\left(\frac{\log x}{\log_3 x}\right)},
\]

\[
P_1 = \{ p : p \leq 2 \log x \text{ or } z < p \leq x/2 \},
\]

\[
P_2 = \{ p : 2 \log x < p \leq z \},
\]

\[
P_3 = \{ p : x/2 < p \leq x \}.
\]

First, we set \( a_p = 0 \mod p \) for all \( p \in P_1 \). These \( a_p \) cover all integers in \([1, y-1]\) that have a prime factor from \( P_1 \). Let \( S_0 \) be the set of uncovered integers \( n \in [1, y-1] \). Such \( n \) satisfy either \( P^+(n) \leq z \), or they have a prime factor \( > x/2 \). In the latter case, as \( n \) has no prime factor \( < 2 \log x \) and \( n \leq y \leq x \log x \), we conclude that \( n \) is prime. Let \( u = \frac{5 \log x}{\log_3 x} \). Then \( u \log u > (5 - o(1)) \log_2 x \). By Theorem 2.2 and the PNT,

\[
|S_0| \leq \Psi(y, z) + \pi(y) \ll \frac{y}{\log^4 x} + \frac{y}{\log x} \ll \frac{y}{\log x}.
\]
Suppose that distinct primes in by
Assume we have equality instead of a choice of (2.10)

\[ \rho \] (2.12)

differential-delay equation are
and this proves (2.9). Therefore, if \( c > 0 \) is small enough, and we take
\[ y = c \frac{x(\log x) \log_3 x}{(\log_3 x)^2}, \]
then \( |S_k| \lesssim \frac{x}{10 \log x} \leq \pi(x) - \pi(x/2) \), again using the PNT. Finally, the elements of \( S_k \) can be mapped to distinct primes in \( P_3 \). Thus, if \( \ell \in S_k \) maps to \( p \), take \( a_p = \ell \mod p \) to cover \( \ell \).

In conclusion, if \( y \) is given by (2.10), then \( G(X) \geq J(x) \geq y \). As \( x \sim \log X \),
\[ y \gg (\log X)(\log_2 X)(\log_4 X) \]
and this proves (2.9).

3. Asymptotic formulas when \( u \) is small

The idea behind the asymptotic formula is to first develop a recurrence formula. For \( 1 \leq \ell \leq m \), there are \( \binom{n}{\ell} \ell! \) ways to form an \( \ell \)-cycle from \( [n] \). Hence
\[
\nu(n,m) = \frac{1}{n!} \sum_{\sigma \in S_n \atop C_{(m,n)}(\sigma) = 0} \frac{1}{n} \sum_{\tau \in \sigma \atop \text{a cycle}} |\tau| = \frac{1}{n \cdot n!} \sum_{\ell=1}^{m} \ell \binom{n}{\ell} (\ell-1)! (n-\ell)! \nu(n-\ell,m)
\]
(2.11)

Heuristic. Suppose that \( \nu(n,m) \approx f(u) \), where \( u = n/m \) and \( f \) is continuous. By (2.11),
\[
f(u) \approx \frac{1}{n} \sum_{k=n-m}^{n-1} f(k/m) \approx \frac{1}{n} \int_{n-m}^{n} f(t/m) dt = \frac{1}{u} \int_{u-1}^{u} f(v) dv.
\]
Assume we have equality instead of \( \approx \). Differentiation gives \( u f'(u) = -f(u-1) \). This is known as a differential-delay equation. If we add the natural initial conditions \( f(u) = 1 \) for \( 0 \leq u \leq 1 \), then there is a unique continuous solution. This motivates the definition of the Dickman function \( \rho(u) \).

**Definition 2.4.** The Dickman function \( \rho : [0, \infty) \to \mathbb{R} \) is the unique continuous solution of
\[
\rho(u) = 1 \quad (0 \leq u \leq 1); \quad u \rho'(u) = -\rho(u-1) \quad (u > 1).
\]

**Lemma 2.5.** We have
(a) \( u \rho(u) = \int_{u-1}^{u} \rho(v) dv \) for \( u \geq 1 \);
(b) \( \rho(u) > 0 \) for all \( u \geq 0 \);
(c) \( \rho(u) \) is decreasing for \( u \geq 0 \);
(d) For \( u \geq 1 \), \( -\frac{\rho'(u)}{\rho(u)} \ll 1 + \log u \).

**Proof.** (a) follows by integrating (2.12) from \( u = 1 \) to \( u = v \) with \( v \geq 1 \). To prove (b), assume that \( \tau = \min\{u : \rho(u) = 0\} \) exists. Since \( \rho(u) = 1 - \log u \) for \( 1 \leq u \leq 2 \), \( \tau > 2 \). By (a),
\[
0 = \tau \rho(\tau) = \int_{\tau-1}^{\tau} \rho(v) \, dv > 0,
\]
3. Asymptotic formulas when \( u \) is small

Figure 1. Dickman’s function from \( 0 \leq u \leq 15 \).

a contradiction. That \( \rho(u) \) is decreasing is clear from (b) and (2.12). This proves (c). From (2.12) and (a),

\[
(2.13) \quad \frac{-\rho'(u)}{\rho(u)} = \frac{\rho(u-1)}{\int_{u-1}^{u} \rho(v) \, dv}.
\]

Let \( B_k = \max_{1 < v \leq k/2} (-\rho'(v)/\rho(v)) \). We have

\[
B_4 = \max_{1 < v \leq 2} \frac{1}{v(1 - \log v)} = \frac{1}{2(1 - \log 2)} = 1.629 \ldots.
\]

If \( k \geq 4 \) and \( k/2 < u \leq (k + 1)/2 \) then the denominator on the right side of (2.13) is at least

\[
\int_{u-1}^{u-1/2} \rho(v) \, dv \geq \rho(u-1) \int_{u-1}^{u-1/2} e^{-B_k(v-u+1)} \, dv = \frac{\rho(u-1)(1 - e^{-\frac{1}{2}B_k})}{B_k}.
\]

Using that \( e^{-\frac{1}{2}B_k} \leq e^{-\frac{1}{2}B_4} < 1/2 \), we infer that

\[
B_{k+1} \leq \frac{B_k}{1 - e^{-\frac{1}{2}B_k}} \leq B_k \left( 1 + 2e^{-\frac{1}{2}B_k} \right).
\]

The function \( x(1 + 2e^{-x/2}) \) is increasing for \( x \geq 0 \), hence if \( C \) is large and \( B_k \leq C \log k \) then \( B_{k+1} \leq (C \log k)(1 + 2/k^{C/2}) \leq C \log(k + 1) \). Therefore, \( B_k \leq \log k \) and (d) follows.

In Figure 3, we plot the Dickman function on a log-scale, and it is evident that \( \rho \) decreases rapidly. In fact, \( \rho(u) = e^{-u \log u - u \log_2(2u) + O(u)} \); see [61], Ch. III.5.4 for further asymptotics and proofs.

**Theorem 2.6.** For all \( n \geq m \geq 1 \) we have

\[
(2.14) \quad \rho\left( \frac{n}{m} \right) \leq \nu(n,m) \leq \rho\left( \frac{n+1}{m+1} \right).
\]
Consequently, for $\sqrt{n \log n} \leq m \leq n$ we have

\begin{equation}
\nu(n, m) = \rho(u) \left( 1 + O \left( \frac{u \log(u + 1)}{m} \right) \right), \quad u = n/m.
\end{equation}

**Remarks.** Inequality (2.15) recovers Theorem 4 of [50], with a much shorter proof, and provides an asymptotic formula for $\nu(n, m)$ as long as $n = o(m^2 / \log m)$. When $n \gg m^2 / \log m$, $\nu(n, m) \neq \rho(n/m)$, the asymptotic having a different shape; see [53], Theorem 2.4 or [50] for details. Thus the final conclusion is best-possible.

**Proof.** Suppose $m \leq n \leq 2m$. As there is at most one cycle of length $> m$, Lemma 1.1 implies

\begin{equation}
\nu(n, m) = 1 - \sum_{k=m+1}^{n} \frac{1}{k} \lesssim \sum_{k=m+1}^{n} \int_{k-1}^{k} \frac{dt}{t} = \log \frac{n}{m} = 1 - \rho \left( \frac{n}{m} \right)
\end{equation}

and

\begin{equation}
H_n - H_m \geq \sum_{k=m+1}^{n} \int_{k}^{k+1} \frac{dt}{t} = \int_{m+1}^{n+1} \frac{dt}{t} = \log \left( \frac{n+1}{m+1} \right) = 1 - \rho \left( \frac{n+1}{m+1} \right),
\end{equation}

the bounds (2.14) hold when $m \leq n \leq 2m$.

Now fix $m \geq 1$, let $N \geq 2m + 1$ and assume that (2.14) holds when $m \leq n \leq N - 1$. Using (2.11) followed by Lemma 2.5 (a–c),

\begin{equation}
\nu(N, m) = \frac{1}{N} \sum_{k=N-m}^{N-1} A(k, m) > \frac{1}{N} \sum_{k=N-m}^{N-1} \rho(k/m) \geq \frac{1}{N} \sum_{k=M-n}^{N-1} \int_{k}^{k+1} \rho(t/m) dt
\end{equation}

\begin{equation}
= \frac{1}{N} \int_{N-m}^{N} \rho(v/m) dv = \frac{1}{N/m} \int_{N/m-1}^{N/m} \rho(v) dv = \rho(N/m)
\end{equation}

and

\begin{equation}
\nu(N, m) \lesssim \frac{1}{N} \sum_{k=N-m}^{N-1} \rho \left( \frac{k+1}{m+1} \right) \lesssim \frac{1}{N} \sum_{k=N-m}^{N-1} \int_{k-1}^{k} \rho \left( \frac{t+1}{m+1} \right) dt
\end{equation}

\begin{equation}
= \frac{m+1}{N} \int_{N/m}^{N} \rho(v) dv
\end{equation}

\begin{equation}
= \frac{m+1}{N} \int_{N/m}^{N+1} \rho(v) dv - \frac{m+1}{N} \int_{N/m}^{N} \rho(v) dv
\end{equation}

\begin{equation}
= \frac{N+1}{N} \rho \left( \frac{N+1}{m+1} \right) - \frac{m+1}{N} \int_{N/m}^{N+1} \rho(v) dv.
\end{equation}

The final integral on the right side is $\gg \frac{1}{m+1} \rho \left( \frac{N+1}{m+1} \right)$ and thus $\nu(n, m) \leq \rho \left( \frac{N+1}{m+1} \right)$. The claimed bounds (2.14) now follow by induction on $n$.

Now we have

\begin{equation}
\frac{n}{m} - \frac{n+1}{m+1} = \frac{n-m}{m(m+1)} \leq \frac{n}{m^2}.
\end{equation}

Thus, by Lemma 2.5 (d),

\begin{equation}
\rho \left( \frac{n+1}{m+1} \right) \leq \rho \left( \frac{n}{m} \right) e^{O((n/m^2) \log(2u))} \leq \rho \left( \frac{n}{m} \right) e^{O(u \log(u+1)/m)}
\end{equation}
When $\sqrt{n \log n} \leq m \leq n$, $u \log(u + 1)/m \ll 1$ and (2.15) follows.

Comparing Theorems 2.1 and 2.6, we immediately conclude that

\begin{equation}
\rho(u) \ll e^{-u \log u + u}.
\end{equation}

The "100 prisoners problem"

Imagine a prison holding 100 prisoners. They are offered to play a game, the reward being freedom for all if they win; but they all must win in order for any to go free. The prisoners are numbered 1 to 100. Inside a room are 100 boxes, and the numbers 1 through 100 are placed in these boxes in random order. One by one, the prisoners are led into the room and allowed to open 50 boxes. If a prisoner finds his own number in one of the boxes, he wins. The prisoners are allowed to discuss strategy before the game begins, but then are separated and allowed no communication whatsoever (e.g., one prisoner cannot mark the boxes indicating which number is inside). Is there a strategy that allows all of them to win with large probability?

Naively, if each prisoner chooses 50 boxes at random, then each has a 1/2 chance of winning, but there is only a $1/2^{100}$ chance that they all win, and go free. There is a much better strategy, based on observing that the number inside the boxes form a permutation of $[100]$ (we can think of the boxes as lying in a row, 1st, 2nd, ...). The strategy for each prisoner number $k$ is thus: first open the $k$-th box. If the number is $k$, he wins. Otherwise, if box $k$ contains the number $m$, next open box $m$. Continue in this manner until either he finds his own number (win) or has opened 50 boxes without finding his own number (lose). Under what conditions does prisoner $k$ win with this strategy? He is essentially “following the cycle containing $k$”, and if the cycle length is $\leq 50$ he will win. Thus, if there are no cycles of length 50 or more, then everyone wins! The likelihood of this is $A(100, 50)$, and by (2.16) this equals

$$1 - H_{100} + H_{50} \approx 0.31$$

Thus, the prisoners have a 31% chance of all going free.

Next we develop a recursive formula for $\psi(x, y)$ analogous to (2.11) and based on an idea of Hildebrand.

**Lemma 2.7.** For $x \geq y \geq 2$ we have

$$\Psi(x, y) = \frac{1}{\log x} \sum_{p \leq y} (\log p) \Psi\left(\frac{x}{p}, y\right) + O\left(\frac{x}{\log x}\right).$$

**Proof.** We start with

$$\Psi(x, y) \log x = \sum_{n \leq x} \log(x/n) + \sum_{n \leq x} \sum_{p \mid n \log p}. \sum_{p \mid n} \log p.$$

The first sum is $\ll \sum_{n \leq x} (x/n)^{1/2} \ll x$. In the second sum, let $n = mp^k$, so that $P^+(m) \leq y$ and $m \leq x/p^k$. The terms with $k = 1$ have sum

$$\sum_{p \leq y} (\log p) \sum_{m \leq x/p} 1 = \sum_{p \leq y} (\log p) \Psi\left(\frac{x}{p}, y\right).$$

Likewise, the terms with $k > 1$ contribute

$$\sum_{p \leq y} \sum_{k \geq 2} \frac{x}{p^k} \ll x.$$

**Theorem 2.8.** For $x \geq y \geq 3$ we have $\Psi(x, y) = x \rho(u) + O(x/\log y)$, where $u = \frac{\log x}{\log y}$.

The proof of Theorem 2.8 is Exercise 2.1 below.

Theorem 2.8 provides an asymptotic formula for $\Psi(x, y)$ as long as $1/\log y = o(\rho(u))$. By Lemma 2.5 (f), this happens only for $u \ll \frac{\log x}{\log y}$. In fact, the asymptotic $\Psi(x, y) \sim x \rho(u)$ is true in a large range of $x, y$; see [43] or [61, Ch. III.5] for specific statements.
4. Exercises

Exercise 2.1. For \( x \geq y \geq 2 \) let \( u = \frac{\log x}{\log y} \).

(a) Show that \( \Psi(x, y) = x\rho(u) + O(x/\log x) \) for \( y \leq x \leq y^2 \).

(b) Define \( \Delta(x, y) \) by \( \Psi(x, y) = x(\rho(u) + \Delta(x, y)) \). With \( y \) fixed, let \( \Delta_k := \max_{y \leq x \leq y^2} |\Delta(x, y)| \). Use Lemma 2.7 to prove
\[
\Delta_k \leq \frac{\log y + O(1)}{\log(y^{2k-1})} \Delta_{k-1} + O\left(\frac{1}{\log(y^{2k-1})}\right).
\]

(c) Use (a) and (b) to prove Theorem 2.8.

Exercise 2.2. Let
\[
G = 1 - \int_1^\infty \frac{\rho(u)}{u^2} \, du = 0.624329988\ldots
\]

\( G \) is known as the “Golomb-Dickman constant”, although it was first written down by de Bruijn.

(a) Let \( C^+(\sigma) \) denote the length of the largest cycle in \( \sigma \). Show that \( E_n C^+(\sigma) \sim Gn \) as \( n \to \infty \).

(b) Show that \( E_x \log P^+(n) \sim G \log x \) as \( x \to \infty \).

Exercise 2.3. Define a function \( \rho_2(u) \) by \( \rho_2(u) = 1 \) for \( 0 \leq u \leq 1 \) and
\[
\rho_2(u) = \rho(u) + \int_0^{u-1} \frac{\rho(w)}{u-w} \, dw \quad (u \geq 1).
\]

(a) For a permutation \( \sigma \in S_n \), let \( k_2(\sigma) \) denote the length of the 2nd largest cycle (it may equal the length of the largest cycle), and let \( k_2(\sigma) = 0 \) if \( \sigma \) has only one cycle. Show that, uniformly for \( 1 \leq m \leq n \) and \( u = n/m \), that
\[
P_\sigma(k_2(\sigma) \leq m) = \rho_2(u) + O\left(\frac{1 + \log u}{m}\right).
\]

(b) Let \( q_2(n) \) denote the 2nd largest prime factor of an integer \( n \) (it may equal the largest prime factor, if \( P^+(n)^2 \mid n \)), and define \( q_2(n) = 0 \) if \( n \) is prime. Show that, uniformly for \( 2 \leq y \leq x \) and \( u = \frac{\log x}{\log y} \),
\[
P_x(q_2(n) \leq y) = \rho_2(u) + O\left(\frac{1 + \log u}{\log y}\right).
\]

(c) Show that \( \rho_2(u) \sim c/u \) as \( u \to \infty \), where \( c = \int_0^\infty \rho(w) \, dw \).
CHAPTER 3

Integers without small prime factors and permutations without small cycles

1. Permutations without small cycles

For \(1 \leq m \leq n\), let
\[
U_{n,m} = \mathbb{P}_{\sigma \in \mathcal{S}_n}(C_j(n) = 0 \ (1 \leq j \leq m)).
\]
In particular, \(U_{n,0} = 1\). Based on our heuristic model, \(U_{n,m}\) should be (for \(m\) of moderate size) about the probability that \(Z_1 = \cdots = Z_m = 0\), where \(Z_j \overset{d}{=} \text{Pois}(1/j)\) and \(Z_1, \ldots, Z_m\) are independent. This probability equals \(e^{-H_m} \approx e^{-\gamma m}\). This cannot be expected to hold for large \(m\), for example \(U_{n,m} = 1/n\) if \(m \geq n/2\) (permutations lacking cycles of length \(\leq n/2\) must be \(n\)-cycles). A special case of Corollary 1.11 (with \(\lambda = 0\)) implies that
\[
(3.1) \quad U_{n,m} \ll \frac{1}{m}
\]
uniformly for \(1 \leq m \leq n\). Our aim in this section is to prove strong asymptotics for \(U_{n,m}\) throughout the range \(1 \leq m \leq n\).

As a first attempt, we’ll use inclusion-exclusion, obtaining for any \(\ell \geq 1\) the formula
\[
(3.2) \quad U_{n,m} = \mathbb{E}_{\sigma} 1(C_{[m]}(\sigma) = 0) = \sum_{r=0}^{\ell} (-1)^r \mathbb{E}_{\sigma} (C_{[m]}(\sigma))^r \prod_{r \in I} \left( \frac{1}{r} \right)^k + O \left( \binom{C_{[m]}(\sigma)}{\ell + 1} \right).
\]
To evaluate the right side of (3.2) we derive a generalization of Lemma 1.1.

**Lemma 3.1.** Let \(I \subseteq [n]\) and let \(k \geq 0\).
\[
\mathbb{E}_{\sigma} \binom{C_I(\sigma)}{k} \leq \frac{H(I)^k}{k!},
\]
with equality if and only if \(k \max I \leq n\).

**Proof.** For non-negative integers \(x_1, \ldots, x_t, k\) we have
\[
\binom{x_1 + \cdots + x_t}{k} = \sum_{i_1 + \cdots + i_t = k} \prod_{j=1}^{t} \binom{x_j}{i_j}.
\]
Thus, since \(C_I(\sigma) = \sum_{r \in I} C_r(\sigma)\),
\[
\mathbb{E}_{\sigma} \binom{C_I(\sigma)}{k} = \sum_{\sum_{r \in I} k_r = k} \mathbb{E}_{\sigma} \prod_{r \in I} \binom{C_r(\sigma)}{k_r}.
\]
We apply Lemma 1.1 to the expectation on the right side, followed by the multinomial theorem, obtaining
\[
\mathbb{E}_{\sigma} \binom{C_I(\sigma)}{k} \leq \sum_{\sum_{r \in I} k_r = k} \prod_{r \in I} \frac{(1/r)^{k_r}}{k_r!} = \frac{H(I)^k}{k!},
\]
with equality if and only if \( \sum_{r \in I} rk_r \leq n \) for all choices of the \( k_r \). This latter condition clearly holds if \( k(\max I) \leq n \). On the other hand, if \( k(\max I) > n \) then there are terms with \( \sum_{r \in I} rk_r > n \), e.g. taking \( k_{\max I} = k \), \( k_r = 0 \) for other \( r \).

**Theorem 3.2.** Let \( 1 \leq m \leq n \) and set \( u = n/m \). Then

\[
U_{n,m} = e^{-H_m} + O\left(\left(\frac{e(\log m + 1)}{u}\right)^u\right).
\]

**Proof.** The bound is trivial if \( u < e(\log m + 1) \), thus we may assume that \( u \geq e(\log m + 1) \). Let \( \ell = \lfloor u \rfloor = \lfloor n/m \rfloor \). For \( 0 \leq r \leq \ell \), \( r \max I_1 \leq n \), and thus applying Lemma 3.1 to (3.2), we obtain

\[
U_{n,m} = \mathbb{E}_\sigma \sum_{r=0}^\ell (-1)^r \binom{C[m]}{r}_\sigma + O\left(\binom{C[m]}{\ell+1}\right)
\]

\[
= \sum_{r=0}^\ell (-1)^r \frac{H_m^r}{r!} + O\left(\frac{H_m^{\ell+1}}{(\ell+1)!}\right).
\]

Since \( \ell + 1 > e(\log m + 1) \geq eH_m \), the sum equals \( e^{-H_m} + O(H_m^{\ell+1}/(\ell+1)! \). Finally, since \( H_m \leq \log m + 1 \) and \( (\ell + 1)! \geq ((\ell + 1)/e)^{\ell+1} \geq (u/e)^u \) we obtain the claimed bound.

The inclusion-exclusion identity (3.2) corresponds to the original Brun sieve technique. The bound in Theorem 3.2 is nontrivial only for \( u \gg \log m \), that is, when \( m \ll n/\log n \). When \( u \) is large, however, the error term is very tiny compared to the main term \( e^{-H_m} \gg 1/m \).

When \( u \) is bounded, the behavior of \( U_{n,m} \) is more complex. We will derive a recurrence for \( U_{n,m} \) in (3.6) below, and use it to obtain an asymptotic for \( U_{n,m} \) when \( u \) is small. We first must introduce the Buchstab function \( \omega(u) \), defined recursively for \( u \geq 1 \) by

\[
\omega(u) = \frac{1}{u} \quad (1 \leq u \leq 2), \quad u \omega(u) = 1 + \int_1^u \omega(v) \, dv \quad (u > 2).
\]

An easy induction argument shows that \( \omega(u) \) is continuous, differentiable except at the point \( u = 2 \), and satisfies \( 1/2 \leq \omega(u) \leq 1 \) for all \( u \geq 1 \). Differentiating the integral equation in (3.3) yields

\[
\omega'(u) = \frac{\omega(u-1) - \omega(u)}{u} = -\frac{1}{u} \int_{u-1}^u \omega'(v) \, dv \quad (u > 2).
\]

**Lemma 3.3.** We have \( \omega'(u) \ll 1/|u|! \). Consequently, for some \( C \in [1/2, 1] \) we have

\[
\omega(u) = C + O(1/|u|!).
\]

**Proof.** By \( 1/2 \leq \omega(u) \leq 1 \) for all \( u \), (3.4) gives \( |\omega'(u)| \leq 1/(2u) \) for \( u > 2 \). By induction on \( k \geq 2 \) we have

\[
|\omega'(u)| \leq \frac{1}{2u(u-1)\cdots(u-k+2)} \quad (u \geq k).
\]

For any \( u \), let \( k = |u| \leq u \). Then \( |\omega'(u)| \leq 1/k! \). The second claim follows from the first and

\[
\omega(u) - \omega(v) = \int_v^u \omega'(w) \, dw = \frac{1}{|v|!} \quad (2 \leq v \leq u).
\]

Much more is known about \( \omega(u) \), in fact we have \( \omega(u) = e^{-\gamma} + O(1/|u|!) \), and \( \omega(u) = e^{-\gamma} \) changes sign infinitely many times (Maiol). We will derive such an asymptotic in a indirect way using the next Theorem.

**Theorem 3.4.** Suppose that \( m, n \) are integers with \( 1 \leq m \leq n - 1 \). Then

\[
\frac{1}{2m+1} \leq U_{n,m} \leq \frac{1}{m+1}
\]
Lemma 3.3, the first sum equals absolute value by (3.5). Thus, (3.5) follows by induction on $n$.

The second part provides an asymptotic $U_{n,m} \sim \omega(n)/m$ as long as $m \to \infty$ as $n \to \infty$.

**Proof.** We follow the method of Granville [36, Theorem 2.2], beginning with an analog of the recursion (2.11). If $\sigma \in S_n$ has no cycles of length $\leq m$, then either $\sigma$ is an $n$-cycle, or all cycles in $\sigma$ have length in $[m+1, n-m-1]$. Following the proof of (2.11), we start with

$$U_{n,m} = \frac{1}{n} + \frac{1}{n \cdot n!} \sum_{\sigma \in S_n} \sum_{\alpha|\sigma, C_{\alpha}(\sigma) = 0} |\alpha|.$$

With $\ell = |\alpha|$ fixed, there are $\binom{n}{\ell}(\ell - 1)!$ ways to choose $\alpha$. Writing $\sigma = \alpha \beta$, there are $(n - \ell)!U(n - \ell, m)$ ways to choose $\beta$. Letting $k = n - \ell$, we thus we obtain

$$U_{n,m} = \frac{1}{n} + \frac{1}{n \cdot n!} \sum_{m+1 \leq k \leq n-m-1} U_{k,m}. \tag{3.6}$$

If $m+1 \leq n \leq 2m+1$, then $U_{n,m} = 1/n$ and thus $\frac{1}{2m+1} \leq U_{n,m} \leq \frac{1}{m+1}$. Now suppose that $N \geq 2m+2$ and that (3.5) holds for $m+1 \leq n \leq N-1$. By (3.6),

$$NU_{N,m} \leq 1 + \sum_{m+1 \leq k \leq N-m-1} \frac{1}{m+1} = 1 + \frac{N - 2m - 1}{m+1} \leq \frac{N}{m+1}$$

and

$$NU_{N,m} \geq 1 + \sum_{m+1 \leq k \leq N-m-1} \frac{1}{2m+1} = \frac{N}{2m+1}.$$

Thus, (3.5) follows by induction on $n$.

The second inequality is true when $1 \leq u = n/m \leq 2$ since then $U_{n,m} = 1/n = \omega(n)/m$. We now proceed by induction on $N = |u|$ with $m$ fixed. Let

$$\Delta(n,m) = nU_{n,m} - \omega(n)/m, \quad \Delta_N := \max_{m+1 \leq n \leq mN} |\Delta(n,m)|$$

In particular, $\Delta_2 = 0$. Now suppose that $N \geq 2$ and that $mN < n \leq m(N+1)$. By (3.6),

$$\omega(n/m) + \Delta(n,m) = \frac{m}{n} + \frac{1}{n} \sum_{k=m+1}^{n-m-1} \omega \left( \frac{k}{m} \right) + \frac{1}{n} \sum_{k=m+1}^{n-m-1} \Delta(k,m). \tag{3.7}$$

If $m+1 \leq k \leq n-m-1$, then $n-k \leq mN$, and thus the second sum on the right side is bounded in absolute value by $(n-2m-1)\Delta_N$. Writing $u = n/m$ and using Euler’s summation formula together with Lemma 3.3, the first sum equals

$$\sum_{k=m+1}^{n-m-1} \omega \left( \frac{k}{m} \right) = -\omega(u-1) + \sum_{k=m+1}^{n-m} \omega \left( \frac{k}{m} \right)$$

$$= -\omega(u-1) + \int_m^{n-m} \omega(t/m) \, dt + \int_m^{n-m} \frac{t - \lfloor t/m \rfloor}{m} \omega'(t/m) \, dt$$

$$= O(1) + m \int_1^{n-1} \omega(v) \, dv.$$

By (3.3), the final integral equals $u \omega(u) - 1$, and hence

$$|\Delta(n,m)| \leq \left( 1 - \frac{2}{N+1} \right) \Delta_N + O\left( \frac{1}{n} \right).$$
and thus
\[ \Delta_{N+1} \leq \left(1 - \frac{2}{N+1}\right) \Delta_N + O\left(\frac{1}{mN}\right). \]
Iterating this gives
\[ \Delta_N \ll \frac{1}{m} \quad (N \geq 2) \]
and the second claim in the theorem follows. \(\square\)

Combining Theorems 3.2 and 3.4, we derive a strong asymptotic for \(\omega(u)\) and a strong uniform asymptotic for \(U_{n,m}\). In figure 1 is a graph of \(\omega(u)\) for \(1 \leq u \leq 4\). The rapid convergence to \(e^{-\gamma}\) is evident.

**Theorem 3.5.** We have
\[ \omega(u) = e^{-\gamma} + O(e^{-u/5} + O(u)). \]

**Proof.** In light of Lemma 3.3 it suffices to evaluate \(C = \lim_{u \to \infty} \omega(u)\). Let \(m\) be large, \(n = m^2, u = m\).

By Theorems 3.2 and 3.4, we have
\[ U_{m^2,m} = \frac{\omega(m)}{m} \left(1 + O\left(\frac{1}{m}\right)\right) = e^{-H_m} \left(1 + O\left(m \left(\frac{e(1 + \log m)}{m}\right)^m\right)\right) \]
ad it follows that \(C = e^{-\gamma}\). \(\square\)

**Theorem 3.6.** For any \(1 \leq m \leq n - 1\) we have
\[ U_{n,m} = e^{-H_m} (1 + O(e^{-u/5})) \quad (u = n/m). \]

**Proof.** When \(u > 5 \log m\), Theorems 3.2 and 3.5 gives
\[ U_{n,m} = e^{-H_m} (1 + O(e^{-u/4})). \]

When \(u \leq 5 \log m\), Theorems 3.4 and 3.5 imply
\[ U_{n,m} = \frac{\omega(u)}{m} \left(1 + O\left(\frac{1}{m}\right)\right) = e^{-H_m} (1 + O(e^{-u/5})). \]
2. Integers without small prime factors

Let \( \Phi(x, z) \) denote the number of positive integers \( n \leq x \) that have no prime factor \( \leq z \). Again, a simple heuristic suggests that for small \( z \) we should have \( \Phi(x, z) \approx x \prod_{p \leq z} (1 - 1/p) \), and this is what we will in fact demonstrate below. A special case of Corollary 1.15 (with \( \lambda = 0 \)) implies that

\[
(3.8) \quad \Phi(x, z) = \# \{ n \leq x : \omega(n, z) = 0 \} \ll x (\log z)^{-Q(0)} = \frac{x}{\log z} \times \prod_{p \leq z} \left( 1 - \frac{1}{p} \right),
\]

uniformly for \( 2 \leq z \leq x \). Here \( \omega(n, z) \) is the number of distinct prime factors of \( n \) that are \( \leq z \).

Following what we did with permutations, we first prove a bound for \( \Phi(x, z) \) which is very strong for small \( z \) and then by another method work out a bound which is good for large \( z \).

**Theorem 3.7.** Uniformly for \( x \geq 2, 2 \leq z \leq x^{1/(4.5 \log_2 x)} \) we have

\[
\Phi(x, z) = x \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) + O(xe^{-u/2}) \quad u = \frac{\log x}{\log z};
\]

**Proof.** We may assume that \( x \) is sufficiently large. Following the proof of Lemma 3.2, we apply “Brun’s pure sieve”. Let \( \ell \geq 1 \), and apply inclusion-exclusion to obtain

\[
\mathbb{I}(\omega(n, z) = 0) = \sum_{r=0}^{\ell} (-1)^r \left( \frac{\omega(n, z)}{r} \right) + O \left( \left( \frac{\omega(n, z)}{\ell + 1} \right) \right).
\]

We sum over \( x \) and use

\[
\sum_{n \leq x} \left( \frac{\omega(n, z)}{r} \right) = \sum_{p_1 < \cdots < p_r \leq z} \left( \frac{x}{p_1 \cdots p_r} \right) = \sum_{p_1 < \cdots < p_r \leq z} \left( \frac{x}{p_1 \cdots p_r} + O(1) \right).
\]

The totality of the \( O(1) \) terms is

\[
\ll \sum_{r=0}^{\ell} \# \{ d \leq z^r : \omega(d) = r \} \leq z^\ell.
\]

Let \( T \) be the set of primes \( \leq z \), and let \( H = H(T) \). By Mertens’ sum estimate (0.5), \( H(T) = \log_2 z + O(1) \).

By Theorem 1.6,

\[
\sum_{n \leq x} \left( \frac{\omega(n, z)}{\ell + 1} \right) \leq x \frac{H^{\ell+1}}{(\ell + 1)!}.
\]

We get

\[
\Phi(x, z) = x \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) + \sum_{r=0}^{\ell} (-1)^r \sum_{p_1 < \cdots < p_r \leq z} \frac{1}{p_1 \cdots p_r} + O \left( z^\ell + x \frac{H^{\ell+1}}{(\ell + 1)!} \right).
\]

Extending the sum on \( r \) to all non-negative integers, we have using (0.4)

\[
\sum_{r=0}^{\ell} (-1)^r \sum_{p_1 < \cdots < p_r \leq z} \frac{1}{p_1 \cdots p_r} = \sum_{d \leq z} \frac{\mu(d)}{d} \prod_{p \leq z} \left( 1 - \frac{1}{p} \right).
\]

Now assume that \( \ell \geq 2H \). Using (0.1), we have

\[
\left| \sum_{r=\ell+1}^{\infty} (-1)^r \sum_{p_1 < \cdots < p_r \leq z} \frac{1}{p_1 \cdots p_r} \right| \leq \sum_{r=\ell+1}^{\infty} \frac{H^r}{r!} \ll \frac{H^{\ell+1}}{(\ell + 1)!}.
\]

We conclude that

\[
\Phi(x, z) = x \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) + O \left( z^\ell + x \frac{H^{\ell+1}}{(\ell + 1)!} \right).
\]
If \( 2 \leq z \leq e^{\sqrt{\log x}} \), equivalently \( \sqrt{\log x} \leq u \leq \frac{\log x}{\log 2} \), take \( \ell = \lceil u/5 \rceil \) so that \( \ell \geq 2H \) if \( x \) is large enough. Then
\[
S \leq x^{1/5} \leq xe^{-u/2},
\]
and
\[
\frac{H^{\ell+1}}{\ell!} \leq \left( \frac{5eH}{u} \right)^{u/5} \ll e^{-u/2}.
\]
If \( e^{\sqrt{\log x}} \leq z \leq x^{1/(4.5 \log 2 \cdot x)} \), equivalently \( 4.5 \log 2 \cdot x \leq u \leq \sqrt{\log x} \), we take \( \ell = \lfloor u - 1 \rfloor \). Then
\[
\ell \geq u - 2 \geq 4.5 \log 2 \cdot x - 2 \geq 4.5H + 2
\]
for large \( x \). Here \( z^\ell \leq x/z \leq xe^{-u} \) and
\[
\frac{H^{\ell+1}}{\ell!} \leq \left( \frac{eH}{u - 1} \right)^{u-1} \ll \left( \frac{e}{4.5} \right)^{u-1} e^{-u/2}.
\]
When \( z \) is large, the asymptotics of \( \Phi(x, z) \) are related to the Buchstab function \( \omega(u) \) defined in (3.3).

**Theorem 3.8.** Uniformly for \( 2 \leq z \leq x \) we have
\[
\Phi(x, z) = \frac{x\omega(u) - z}{\log z} + O \left( \frac{x}{\log^2 x} \right), \quad u = \log \frac{x}{\log z}.
\]

This provides an asymptotic formula \( \Phi(x, z) \sim \omega(u) \frac{x}{\log z} \) as long as \( z \to \infty \) and \( x = o(x) \) as \( x \to \infty \).

**Proof.** When \( z \leq x \leq z^2 \), the Prime Number Theorem gives
\[
\Phi(x, z) = 1 + \pi(x) - \pi(z) = \frac{x}{\log x} - \frac{z}{\log z} + O \left( \frac{x}{\log^2 x} \right) = \frac{x\omega(u) - z}{\log z} + O \left( \frac{x}{\log^2 x} \right),
\]
proving (i). Here we used that \( \omega(u) = 1/u \) for \( 1 \leq u \leq 2 \).

Now suppose \( 2 \leq z \leq x^{1/2} \). We start with the fundamental theorem of arithmetic in the form
\[
\log x = \log \frac{x}{n} + \log n = \frac{x}{n} + \sum_{p^k | n} \log p.
\]

Thus,
\[
(\log x)\Phi(x, z) = \sum_{n \leq x} \frac{\log x}{n} + \sum_{n \leq x} \sum_{p^k | n, \, P^-(n) > z} \log p.
\]

In \( S_1 \), break up the summands into intervals \( xe^{-j} < n \leq xe^{1-j} \) for integers \( j \geq 1 \). Using (3.8)
\[
S_1 \leq \sum_{j \geq 1} j\Phi(xe^{-j}, z) \ll \sum_{j \geq 1} \frac{x e^{1-j}}{\log z} \ll \frac{x}{\log z}.
\]

We have
\[
S_2 = \sum_{n \leq x \atop p^k | n} \log p = \sum_{p^k \leq x \atop p > z} (\log p)\Phi(x/p^k, z).
\]
Using (3.8) again, the terms with \( k \geq 2 \) contribute
\[
\ll \sum_{\substack{p^k \leq x \\ p > z}} (\log p) \frac{x}{p^k \log z} \ll \frac{x}{\log z} \sum_{p > z} \frac{\log p}{p^2 - p} \ll \frac{x}{z \log z}.
\]
The terms with \( k = 1 \) give, by the Prime Number Theorem,
\[
\sum_{z < p \leq x/z} (\log p) \Phi(x/p, z) + \sum_{x/z < p \leq x} \log p = \sum_{z < p \leq x/z} (\log p) \Phi(x/p, z) + x + O \left( \frac{x}{\log z} \right).
\]
This completes the proof of (ii). □

Now we argue that Lemma 3.9 (ii) is analogous to the recurrence (3.6). Let
\[
V_{n,m} = \frac{\Phi(e^n, e^m)}{e^n}.
\]
Assuming that \( \Phi(x, z)/x \) is slowly varying in \( x \) and in \( z \), we get from Lemma 3.9 (ii)
\[
n V_{n,m} \approx \sum_{k=m+1}^{n-m} V_{n-k,m} \sum_{e^{k-1} < p \leq e^k} \frac{\log p}{p} \approx \sum_{k=m+1}^{n-m} V_{n-k,m}
\]
using Mertens’ estimate (0.6).

**Proof of Theorem 3.8.** Let \( z_0 \) be a sufficiently large constant. The statement is trivial if \( z \leq z_0 \) and also it follows from (3.8) when \( z > x^{1/2} \), thus we may assume that \( z_0 \leq z \leq x^{1/2} \).

We iterate the recurrence in Lemma 3.9 (ii) in a manner similar to the way we analyzed (3.6), however there are more delicate error terms to analyze. Define \( \Delta(x, z) \) by
\[
\frac{\log z}{x} \Phi(x, z) = \omega(u) - \frac{z}{x} + \Delta(x, z),
\]
where \( u = \frac{\log x}{\log z} \). With \( z \) fixed, define
\[
\Delta^*_N = \max_{z \leq x \leq z^N} |\Delta(x, z)| \quad (N = 2, 3, \ldots).
\]
By Lemma 3.9 (i), we have
\[
\Delta^*_2 \ll \frac{1}{\log z}.
\]
Now suppose that \( N \geq 3 \) and \( z^{N-1} < x \leq z^N \). Divide Lemma 3.9 (ii) by \( ux \), obtaining
\[
\frac{\log z}{x} \Phi(x, z) = \frac{1}{u} + O \left( \frac{1}{u \log z} \right) + \frac{1}{u \log z} \sum_{z < p \leq x/z} \log p \left( \frac{x}{p} \Phi(x/p, z) \right).
\]
Since \( z/x \leq 1/\sqrt{x} < \frac{1}{u \log z} \) by our assumptions on \( z \) and \( u \), we get
\[
\omega(u) + \Delta(x, z) = \frac{1}{u} + O \left( \frac{1}{u \log z} \right) + \frac{1}{u \log z} \sum_{z < p \leq x/z} \log p \left( \omega \left( \frac{\log(x/p)}{\log z} \right) - \frac{zp}{x} + \Delta \left( \frac{x}{p}, z \right) \right).
\]
In (3.9), the summands \(-zp/x\) contribute, by the Prime Number Theorem,
\[
\ll \frac{z}{u \log z} \sum_{z < p \leq x/z} \log p \ll \frac{1}{u \log z}.
\]
Since \( x/p < x/z < z^{N-1} \), the summands \( \Delta(x/p, z) \) contribute an amount which in absolute value does not exceed
\[
\frac{\Delta^{N-1}}{u \log z} \sum_{z < p \leq x/z} \log p = \frac{\Delta^{N-1}}{\log x} (\log x - 2 \log z + O(1)) \leq (1 - 1/N) \Delta^{N-1}.
\]
if \( z \) is large enough, where we used Mertens’ estimate (0.6). Using (0.6), partial summation and the rapid decay of \( \omega'(u) \) (Theorem 3.3), we obtain

\[
\sum_{z < p \leq x/z} \frac{\log p}{p} \omega \left( \frac{\log(x/p)}{\log z} \right) = \int_{z}^{x/z} \frac{1}{t} \omega \left( \frac{\log(x/t)}{\log z} \right) dt + \int_{z}^{x/z} O(e^{-\sqrt{\log t}}) \omega' \left( \frac{\log(x/t)}{\log z} \right) \frac{dt}{t \log z} + O(1)
\]

\[
= (\log z) \int_{1}^{u-1} \omega(v) dv + O(1).
\]

Inserting all of these estimates into (3.9), we find that

\[
|\Delta(x, z)| \leq \left| -\omega(u) + \frac{1}{u} + \frac{1}{u} \int_{1}^{u-1} \omega(v) dv \right| + \left( 1 - \frac{1}{N} \right) \Delta_{N-1} + O \left( \frac{1}{u \log z} \right)
\]

\[
= \left( 1 - \frac{1}{N} \right) \Delta_{N-1} + O \left( \frac{1}{N \log z} \right)
\]

using the recurrence (3.3) and that \( N - 1 \leq u \leq N \). Taking a maximum over all \( z^{N-1} < x \leq z^N \) we get

\[
\Delta_{N} \leq \max \left( \Delta_{N-1}, \left( 1 - \frac{1}{N} \right) \Delta_{N-1} + O \left( \frac{1}{N \log z} \right) \right).
\]

Iterating this gives

\[
\Delta_{N} \ll \frac{1}{\log z} \quad (N = 2, 3, \ldots).
\]

From the definition of \( \Delta_{N} \) we conclude that

\[
\Delta(x, z) \ll \frac{1}{\log z} \quad (u \geq 2).
\]

This completes the proof. \( \square \)

An unusual application of these theorems is the evaluation of the constant in the Mertens’ product formula (0.7). It is rather straightforward to obtain a version of (0.7) where \( e^{-\gamma} \) is replaced by some unspecified constant, and evaluation of the constant is not that easy.

**Corollary 3.10.** Assume a weak form of Mertens’ product formula

\[
\prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \leq \frac{1}{\log z} \quad (z \geq 2).
\]

Then

\[
\prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \sim \frac{e^{-\gamma}}{\log z} \quad (z \to \infty).
\]

**Proof.** This proof was suggested by Granville. It is easy to check that in the proof of Theorems 3.7 and 3.8, we did not use (0.7) anywhere in the strong form. Fix \( z \) and let \( x = z^{10 \log z} \). Then \( u = \frac{\log z}{\log x} = 10 \log z > 5 \log z \) if \( z \) is large enough. Using (3.10) and Theorem 3.7, we have

\[
\Phi(x, z) = x \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \left( 1 + O(e^{-u/\log z}) \right) = x \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \left( 1 + O(1/\log^4 z) \right).
\]

On the other hand, using Theorem 3.8, followed by Theorem 3.5,

\[
\Phi(x, z) = \frac{x \omega(u)}{\log z} \left( 1 + O \left( \frac{1}{\log z} \right) \right) = x e^{-\gamma} \left( 1 + O \left( \frac{1}{\log z} \right) \right).
\]

The claim follows. \( \square \)
Theorem 3.11. Uniformly for $2 \leq z \leq x$ we have
\[ \Phi(x, z) = x \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \left( 1 + O(e^{-u/5}) \right). \]

Proof. When $u \geq 4.8 \log_2 x$, equivalently, $z \leq x^{1/4.8 \log_2 x}$, we use Theorem 3.7 and get
\[ \Phi(x, z) = x \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \left( 1 + O(e^{-u/2 \log z}) \right) = x \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \left( 1 + O(e^{-u/5}) \right). \]
When $1 \leq u \leq 2$, the claim follows from (3.8). Now suppose that $2 \leq u \leq 4.8 \log_2 x$. By Theorem 3.8, followed by Mertens' estimate (0.7), we have
\[ \Phi(x, z) = x \log z \left( 1 + O \left( \frac{1}{\log z} \right) \right) = xe^\gamma \omega(u) \prod_{p \leq z} \left( 1 - \frac{1}{p} \right) \left( 1 + O \left( \frac{1}{\log z} \right) \right). \]
Now
\[ \frac{1}{\log z} = \frac{u}{\log x} \ll \frac{u}{e^{u/4.8}} \ll e^{-u/5} \]
and the claim follows from Theorem 3.5. □

3. Exercises

Exercise 3.1. Show that $\Phi(x, z) \asymp x/\log z$ uniformly for $x \geq 2z \geq 4$. Be careful with the case of small $x$.

Exercise 3.2. Show that $\Psi(x, y) \ll xe^{-u}$ uniformly for all $x \geq y \geq 3$, where $u = \frac{\log x}{\log y}$.

Exercise 3.3 ([41], Exercise 03). (Integers with a large smooth part). For an integer $n$ and $y \geq 2$ let $n_y$ be the product of all prime powers dividing $n$ with the prime $\leq y$. Define, for $y \leq z \leq x$ the function
\[ \Theta(x, y, z) = \# \left\{ n \leq x : n_y > z \right\}. \]
(a) Show that $\Theta(x, y, z) \leq \sum_{z < a \leq x/y} \Phi(x/a, y) + \Psi(x, y)$.
(b) Show that, for any $2 \leq y \leq z \leq x$ we have
\[ \Theta(x, y, z) \ll x \exp \left\{ - \frac{\log z}{2 \log y} \right\}. \]

Exercise 3.4. (a) Show that for $1 \leq m \leq n$,
\[ 1 = \nu(n, m) + \sum_{k=0}^{n-m-1} U_{n-k, m} \nu(k, m), \]
where $\nu(k, m) = 1$ if $m > k$.
(b) Combine (a) with Theorems 2.6 and 3.4 to deduce that
\[ \rho(u) + \int_0^{u-1} \rho(v) \omega(u - v) \, dv = 1 \quad (u \geq 1). \]
Remark: This provides a ‘combinatorial proof’ of a purely analytic statement.
CHAPTER 4

Poisson approximation of small cycle lengths and small prime divisors

1. Small cycles of permutations

Let $1 \leq k \leq n$ and consider the problem of modeling

$$C_k = (C_1(\sigma), \ldots, C_k(\sigma))$$

by the random vector

$$Z_k = (Z_1, \ldots, Z_k), \quad Z_j \overset{d}{=} \text{Pois}(1/j).$$

We especially desire a good approximation when $k$ is large, as opposed to bounded (ref. Theorem 1.9). We express our results in terms of the Total Variational Distance $d_{TV}(X, Y)$ between two random variables $X$ and $Y$ taking values in a discrete space $\Omega$, defined by

$$d_{TV}(X, Y) := \sup_{U \subseteq \Omega} \mathbb{P}(X \in U) - \mathbb{P}(Y \in U).$$

The supremum occurs when $U = \{ \omega \in \Omega : \mathbb{P}(X = \omega) > \mathbb{P}(Y = \omega) \}$, hence

$$d_{TV}(X, Y) = \sum_{\omega \in \Omega} \max \left( 0, \mathbb{P}(X = \omega) - \mathbb{P}(Y = \omega) \right).$$

Replacing $U$ by $\Omega \setminus U$, we see that $d_{TV}(X, Y) = d_{TV}(Y, X)$.

In comparing $C_k$ and $Z_k$, the space of values is $\Omega = \mathbb{N}_0^k$.

**Lemma 4.1.** We have

$$d_{TV}(C_k, Z_k) = \sum_{h \in \mathbb{N}_0^k} \prod_{j=1}^k \frac{1}{h_j !} \max \left( 0, e^{-H_k} - U_{n', k} \right),$$

where $n' = n'(h) = n - \sum_{j=1}^k jh_j$.

**Proof.** From (4.2) we have

$$d_{TV}(C_k, Z_k) = \mathbb{E} \left[ \max \left( 0, \mathbb{P}(Z_k = h) - \mathbb{P}(C_k = h) \right) \right].$$

Clearly,

$$\mathbb{P}(Z_k = h) = e^{-H_k} \prod_{j=1}^k \frac{(1/j)^{h_j}}{h_j !}.$$

Now fix $h$, write $g = h_1 + 2h_2 + \cdots + kh_k$ and consider $\mathbb{P}(C_k = h)$. If $g > n$, then $\mathbb{P}(C_k = h) = 0$. Now suppose that $g \leq n$. Write $\sigma = \sigma_1 \sigma_2$, where $\sigma_1$ is the product of the cycles of length at most $k$ and permutes a subset $I$ of $[n]$ of size $g$, and $\sigma_2$ is the product of the cycles of length greater than $k$ and permutes $[n] \setminus I$ of size $n' = n - g$. By Cauchy’s formula (Theorem 1.2), applied to $\sigma_1$, it follows that

$$\mathbb{P}(C_k = h) = U_{n', k} \prod_{j=1}^k \frac{(1/j)^{h_j}}{h_j !}.$$
and the lemma follows.

**Theorem 4.2 (Poisson distribution of small cycles).** Let $1 \leq k \leq n$. Then
\[ d_{TV}(C_k, Z_k) \ll e^{-n/(5k)}. \]

**Proof.** Consider a generic vector $h = (h_1, \ldots, h_k) \in \mathbb{N}_0^k$ and let $n' = n - (h_1 + 2h_2 + \cdots + kh_k)$. If $n' > k$ then By Theorem 3.6,
\[ U_{n',k} = e^{-H_k} \left(1 + O\left(e^{-n/(5k)}\right)\right). \]
If $n' \leq k$ we’ll use the trivial bound $\max(0, e^{-H_k} - U_{n',k}) \leq e^{-H_k}$, and thus for all $h$ we have
\[ \max\left(0, e^{-H_k} - U_{n',k}\right) \ll e^{-H_k - n'/(5k)}. \]
Therefore, and thus
\[
\begin{align*}
\sum_{h \in H_k} \frac{1}{h_j!} \prod_{j=1}^k \frac{(1/j)h_j}{h_j!} \max\left(0, e^{-H_k} - U_{n',k}\right) &
\ll e^{-H_k - n/(5k)} \sum_{h \in \mathbb{N}_0^k} \prod_{j=1}^k \frac{(1/j)h_j e^{h_j/(5k)}}{h_j!} \\
&= \exp\left\{-H_k - \frac{n}{5k} + \sum_{j=1}^k \frac{e^{j/(5k)}}{j}\right\} \\
&\ll e^{-n/(5k)},
\end{align*}
\]
using (2.1) in the last step with $u = e^{1/5}$ and $w = e^{1/(5k)}$. The theorem now follows from Lemma 4.1. □

**Remarks.** By a more sophisticated sieve method than that used to prove Theorem (2.1), see [29], it is possible to prove that
\[ d_{TV}(C_k, Z_k) \ll e^{-f(n/k)}, \]
where $f(x) \sim x \log x$ as $x \to \infty$. This is the true order (the asymptotics of the logarithm of the left side), and a result of Arratia and Tavaré [2]. Sharper bounds are known, and are expressed in terms of the Dickman and Buhstab functions (see [53]).

We almost immediately obtain the following corollary, by grouping together integers into sets.

**Corollary 4.3.** Let $k \leq n$. Then, for any subset $T \subseteq [k]$ and $A \subseteq \mathbb{N}_0$, we have
\[ \mathbb{P}(C_T(\sigma) \in A) = \mathbb{P}(Z_T \in A) + O(e^{-n/(5k)}). \]
where $Z_T \overset{d}{=} \text{Pois}(H(T))$.

As long as $k = o(n)$ as $n \to \infty$, the error term is $o(1)$ and this establishes, in a very strong form, the validity of the Poisson model for $C_k$.

### 2. The Kubilius model of small prime factors of integers

We will make formal a probabilistic interpretation of various results about the distribution of integers which have been stated in earlier sections. Consider a randomly chosen integers $n \in [1, x]$. Such an integer $n$ has a canonical prime factorization as
\[ n = \prod_{p \leq x} p^{v_p}. \]
We regard each of the exponents $v_p$ as random variables (they depend on $p$ and also on $x$). We compute exactly
\[ \mathbb{P}_x(v_p = k) = \frac{1}{x} \left(\frac{x}{p^k} - \left\lfloor \frac{x}{p^{k+1}} \right\rfloor\right) = \frac{1}{p^k} - \frac{1}{p^{k+1}} + O\left(\frac{1}{x}\right), \]
the error term being relatively small when $p^k$ is small. Moreover, the variables $v_p$ are quasi-independent; that is, the correlations are small, again provided that the primes are small. The variables $v_p$ corresponding to large $p$ are very dependent on each other, for example the event $(v_p > 0, v_q > 0)$ is impossible if $pq > x$.
The model of Kubilius is a sequence of idealized random variables which remove the error terms above, and is thus easier to compute with. For each prime $p$, define the random variable $X_p$ that has domain $\mathbb{N}_0$ and such that
\[ P(X_p = k) = \frac{1}{p^k} - \frac{1}{p^{k+1}} = \frac{1}{p^k} \left( 1 - \frac{1}{p} \right) \quad (k = 0, 1, 2, \ldots). \]
Furthermore, the variables $X_p$ are all independent. If $y$ is small compared with $x$, we expect that the random vector
\[ X_y = (X_p : p \leq y) \]
has distribution close to that of the random vector
\[ V_y = (v_p : p \leq y). \]
Again, $V_y$ depends on $x$ as well.

Recall the definition (4.1) of the total variation distance and the basic identity (4.2).

**Lemma 4.4.** We have
\[ d_{TV}(X_y, V_y) = \sum_{P^+(m) \leq y} \max \left( 0, \frac{\zeta_y}{m} - \frac{1}{\lfloor x \rfloor} \Phi \left( \frac{x}{m}, y \right) \right), \quad \zeta_y = \prod_{p \leq y} \left( 1 - \frac{1}{p} \right). \]

**Proof.** (cf. Tenenbaum [59]). Fix $u = (u_p : p \leq y)$ and write $m = \prod_{p \leq y} p^{v_p}$. Then
\[ P(X_y = u) = \prod_{p \leq y} P(X_p = u_p) = \prod_{p \leq y} \frac{1}{p^{v_p}} \left( 1 - \frac{1}{p} \right) = \frac{\zeta_y}{m} \]
and
\[ P_x(V_y = u) = \frac{1}{\lfloor x \rfloor} \# \{ \ell \in \mathbb{N} : m\ell \leq x, P^{-}(\ell) > y \} = \frac{1}{\lfloor x \rfloor} \Phi \left( \frac{x}{m}, y \right). \quad \square \]
The lemma follows from (4.2).

**Theorem 4.5 (Kubilius Model Approximation).** Let $2 \leq y \leq x$. Then
\[ d_{TV}(X_y, V_y) \ll \exp \left\{ -\frac{\log x}{5 \log y} \right\}. \]

**Proof.** Let $\delta = 1/(5 \log y)$, so that $0 < \delta \leq 1/3$. Let $m$ satisfy $P^+(m) \leq y$. If $m \leq x/y^2$ then Theorem 3.11 implies that
\[ \Phi \left( \frac{x}{m}, y \right) = \frac{x}{m} \zeta_y \left( 1 + O(e^{-\delta \log(x/m)}) \right). \]
For $m > x/y^2$ we’ll just use the trivial bound
\[ \max \left( 0, \frac{\zeta_y}{m} - \frac{1}{\lfloor x \rfloor} \Phi \left( \frac{x}{m}, y \right) \right) \leq \frac{\zeta_y}{m} \]
Thus, for all $m$ we have
\[ \max \left( 0, \frac{\zeta_y}{m} - \frac{1}{\lfloor x \rfloor} \Phi \left( \frac{x}{m}, y \right) \right) \ll \frac{\zeta_y}{m} e^{-\delta \log(x/m)} = \frac{\zeta_y}{x^2 m^{1-\delta}}. \]
By Lemma 4.4,
\[ d_{TV}(X_y, V_y) \ll \zeta_y x^{-\delta} \sum_{P^+(m) \leq y} m^{-1+\delta} \]
\[ = \zeta_y x^{-\delta} \prod_{p \leq y} \left( 1 + \frac{1}{p^{1-\delta}} + \frac{1}{p^{2-2\delta}} + \cdots \right) \]
\[ \ll \frac{x^{-\delta}}{\log y} \exp \left\{ \sum_{p \leq y} \frac{p^\delta}{p} \right\}. \]
Since \( p^\delta = 1 + O(\delta \log p) \), Mertens’ estimates (0.5) and (0.6) imply that the final sum on \( p \) is \( \log y + O(1) \). The theorem follows.

We next use the Kubilius model to show that prime factors have an approximate Poisson distribution. There are two complications. First, as with permutations, large prime factors (those \( > x^c \) for some fixed \( c > 0 \)) cannot be Poisson distributed because they are highly dependent on each other, and the number of such factors is limited (trivially bounded by \( 1/c \)). Secondly, and unlike the case of permutations, the small prime factors also cannot be Poisson distributed (that is, as \( x \to \infty \)). Take the case \( \omega(n, 2) \), which equals 0 or 1, each with probability tending to \( \tfrac{1}{2} \) as \( x \to \infty \). Likewise, for fixed \( t \), \( \omega(n, t) \) takes only finitely many values and thus cannot approach a Poisson limit as \( x \to \infty \). Hence, in the result stated below, the Poisson approximation reveals itself only when “intermediate prime factors” of \( n \) are dominant, that is, those in an interval \([y, z]\) where \( y \to \infty \) and \( \frac{\log z}{\log x} \to 0 \) as \( x \to \infty \). For a set \( T \) of primes, denote

\[
U_T = \sum_{p \in T} \mathbb{1}(X_p \geq 1), \quad W_T = \sum_{p \in T} X_p
\]

which, in the Kubilius model, are model for \( \omega(n; T) \) and \( \Omega(n; T) \), respectively. Since \( \mathbb{E} \mathbb{1}(X_p \geq 1) = 1/p \) and \( \mathbb{E} X_p = 1/(p-1) \) we have

\[\mathbb{E} U_T = H(T), \quad \mathbb{E} W_T = H'(T) := \sum_{p \in T} \frac{1}{p-1}.\]

Define also

\[H''(T) := \sum_{p \in T} \frac{1}{p^2}.\]

**Theorem 4.6.** Let \( T \) be a finite subset of the primes, and suppose either \( Y = U_T \) or \( Y = W_T \). Let \( H = \mathbb{E} Y \), using the formulas (4.3). Let \( Z \overset{d}{=} \text{Pois}(H) \). Then

\[
\mathbb{P}(Y = k) - \mathbb{P}(Z = k) \ll \begin{cases} 
H''(T)^k e^{-H} \left( \frac{1}{k+1} + \left( \frac{k-H}{H} \right)^2 \right) & 0 \leq k \leq 1.9H \\
H''(T)^{0.9H} (1.9)^{-k} & k > 1.9H.
\end{cases}
\]

**Proof.** Write \( H'' = H''(T) \). When \( k = 0 \), \( \mathbb{P}(Z = 0) = e^{-H} \) and

\[
\mathbb{P}(Y = 0) = \mathbb{P}(\forall p \in T : X_p = 0) = \prod_{p \in T} \left( 1 - \frac{1}{p} \right) = e^{-H + O(H'')} = e^{-H (1 + O(H''))},
\]

and the desired inequality follows.

For \( k \geq 1 \), we work with moment generating functions. For any complex \( s \), (0.11) implies

\[(4.4) \quad \mathbb{E} s^Z = e^{(s-1)H}.
\]

If \( Y = U_T \), then \( H = H(T) \), and uniformly for complex \( s \) with \( |s| \leq 2 \) we have

\[
\mathbb{E} s^{U_T} = \prod_{p \in T} \mathbb{E} s^{1(X_p \geq 1)} = \prod_{p \in T} \left( 1 + \frac{s-1}{p} \right) = e^{(s-1)H + O(|s-1|^2 H'')} \left( 1 + O(|s-1|^2 H''(T)) \right)
\]

If \( Y = W_T \) then \( H = H'(T) \) and uniformly for \( |s| \leq 1.9 \) we have

\[
\mathbb{E} s^{W_T} = \prod_{p \in T} \mathbb{E} s^{X_p} = \prod_{p \in T} \left( 1 + \frac{s-1}{p-s} \right) = \prod_{p \in T} \left( 1 + \frac{s-1}{p-1} + \frac{(s-1)^2}{(p-1)(p-s)} \right) = e^{(s-1)H + O(|s-1|^2 H'')} \left( 1 + O(|s-1|^2 H''(T)) \right).
\]
Then, for any $0 < r \leq 1.9$, (4.4), (4.5) and (4.6) imply
\[
\mathbb{P}(Y = k) - \mathbb{P}(Z = k) = \frac{1}{2\pi i} \int_{|s| = r} \frac{E s^Y - E s^Z}{s^{k+1}} \, dz
\]
\[
= \frac{1}{r^k} \int_0^1 e^{-2\pi i k\theta} \left[ E(\text{re}^{2\pi i\theta})^Y - E(\text{re}^{2\pi i\theta})^Z \right] \, d\theta
\]
\[
\ll \frac{H''}{r^k} \int_0^{1/2} |\text{re}^{2\pi i\theta} - 1|^2 e^{r\cos(2\pi\theta) - 1)H} \, d\theta.
\]
Now, for $0 \leq \theta \leq \frac{1}{2}$,
\[
r \cos(2\pi\theta) - 1 = r - 1 - 2r \sin^2(\pi\theta) \ll r - 1 - 8r\theta^2
\]
and
\[
|\text{re}^{2\pi i\theta} - 1|^2 = (r - 1 - 2r \sin^2(\pi\theta))^2 + \sin^2(2\pi\theta) \ll (r - 1)^2 + \theta^2,
\]
so we obtain
\[
\mathbb{P}(Y = k) - \mathbb{P}(Z = k) \ll \frac{H''e^{(r-1)H}}{r^k} \int_0^{1/2} (|r - 1|^2 + \theta^2)e^{-8r\theta^2H} \, d\theta
\]
\[
\ll \frac{H''e^{(r-1)H}}{r^k} \left( \frac{|r - 1|^2}{\sqrt{1 + rH}} + \frac{1}{(1 + rH)^{3/2}} \right).
\]
When $1 \leq k \leq 1.9H$, we take $r = k/H$ in (4.7) and obtain, using Stirling’s formula,
\[
\mathbb{P}(Y = k) - \mathbb{P}(Z = k) \ll H'' \frac{H^k e^{-H}}{k!} \left( \frac{|k/H - 1|^2}{k^{1/2}} + \frac{1}{k^{3/2}} \right)
\]
\[
\ll H'' \frac{e^{-H}}{k!} \left( \frac{k - H}{H} \right)^2 + \frac{1}{k}.
\]
This completes the proof when $k \leq 1.9H$.

When $k > 1.9H$ we take $r = 1.9$ and the result follows in this case as well from (4.7). \qed

**Theorem 4.7.** Let $T$ be a finite subset of the primes. Then
\[
d_{TV}(U_T, \text{Pois}(H(T))) \ll \frac{H''(T)}{1 + H(T)}
\]
and
\[
d_{TV}(W_T, \text{Pois}(H'(T))) \ll \frac{H''(T)}{1 + H(T)}.
\]

**Proof.** Let $Y \in \{U_T, W_T\}$. If $Y = U_T$, let $H = H(T)$ and if $Y = W_T$, let $H = H'(T)$. Let $Z \overset{d}{=} \text{Pois}(H)$. Again, write $H'' = H''(T)$. From (4.2),
\[
d_{TV}(Y, Z) \leq \sum_{k=0}^{\infty} |\mathbb{P}(Z = k) - \mathbb{P}(Y = k)|.
\]
Consider two cases. First, if $H \leq 2$, we have by Theorem 4.6,
\[
d_{TV}(Y, Z) \ll H'' + \sum_{k > 1.9H} H''(1.9)^{-k} \ll H''.
\]
If $H > 2$, Theorem 4.6 likewise implies that
\[
\sum_{k > 1.9H} |\mathbb{P}(Y = k) - \mathbb{P}(Z = k)| \ll H'' \sum_{k > 1.9H} e^{0.9H} \frac{(1.9)^k}{k} \ll H'' e^{-0.3H}
\]
and also
\[
\sum_{k \leq 1.9H} |\mathbb{P}(Y = k) - \mathbb{P}(Z = k)| \ll H'' e^{-H} \sum_{k=0}^{\infty} \frac{H^k}{k!} \left[ \frac{1}{k+1} + \frac{k(k-1) + k - 2kH + H^2}{H^2} \right] \\
= H'' e^{-H} \left[ \sum_{k=0}^{\infty} \frac{H^k}{(k+1)!} + \sum_{k=2}^{\infty} \frac{H^{k-2}}{(k-2)!} + \frac{1}{H} \sum_{k=1}^{\infty} \frac{H^{k-1}}{(k-1)!} \right] \\
- 2 \sum_{k=1}^{\infty} \frac{H^{k-1}}{(k-1)!} + \sum_{k=0}^{\infty} \frac{H^k}{k!} \right] \\
= H'' e^{-H} \left[ e^H - 1 \right] + e^H \left[ e^H - 2e^H + e^H \right] \\
\ll \frac{H'}{H}.
\]

This proves the bound when \( H > 2 \), upon noting that \( H'(T) \asymp H(T) \). \( \square \)

We can use Theorem 4.6 to deal with prime factors in an arbitrary collection of subsets, by a simple combinatorial device. The following is a consequence of Exercise 4.1. Here \( Z_{T_i} \overset{d}{=} \text{Pois}(H(T_i)) \), and \( Z_{T_1}, \ldots, Z_{T_m} \) are independent.

**Corollary 4.8.** Let \( T_1, \ldots, T_m \) be disjoint sets of primes. Then
\[
d_{TV} \left( \left( U_{T_1}, \ldots, U_{T_m} \right), \left( Z_{T_1}, \ldots, Z_{T_m} \right) \right) \ll \sum_{j=1}^{m} \frac{H''(T_j)}{\max(1, H(T_j))}.
\]

Combining this Corollary with the Kubilius model (Thm. 4.5), we conclude that the “intermediate” (not too small and not too large) prime factors of an integer are “Poisson distributed”.

**Theorem 4.9.** Let \( 2 \leq y \leq x \), and let \( T_1, \ldots, T_m \) be disjoint sets of primes \( \leq y \). Then
\[
d_{TV} \left( \left( \omega(n; T_1), \ldots, \omega(n; T_m) \right), \left( Z_{T_1}, \ldots, Z_{T_m} \right) \right) \ll \sum_{j=1}^{m} \frac{H''(T_j)}{\max(1, H(T_j))} + O \left( e^{-\log \frac{y}{\log y}} \right).
\]

**Proof.** Let \( \omega = (\omega(n; T_1), \ldots, \omega(n; T_m)) \), \( U = (U_{T_1}, \ldots, U_{T_m}) \) and \( Z = (Z_{T_1}, \ldots, Z_{T_m}) \) By the triangle inequality for \( d_{TV} \), which follows easily from the definition (4.1), we have
\[
d_{TV}(\omega, Z) \leq d_{TV}(\omega, U) + d_{TV}(U, Z) \\
\leq d_{TV}(V_y, X_y) + d_{TV}(U, Z).
\]

The theorem now follows by combining Theorem 4.5 with Corollary 4.8. \( \square \)

### 3. Exercises

**Exercise 4.1.** (a) Prove that if \( X_1, \ldots, X_m \) are independent discrete random variables, and \( Y_1, \ldots, Y_m \) are independent discrete random variables (with \( Y_j \) having the same domain as \( X_j \)), then
\[
d_{TV}((X_1, \ldots, X_m), (Y_1, \ldots, Y_m)) \leq \sum_{j=1}^{m} d_{TV}(X_j, Y_j).
\]

(b) Let \( X_j \overset{d}{=} \text{Pois}(\lambda_j) \) for \( 1 \leq j \leq m \), where \( 0 < \lambda_j \leq 1 \) for each \( j \). Also suppose that \( Y_j \) is a Bernoulli random variable, with \( \mathbb{P}(Y_j = 0) = 1 - \lambda_j, \mathbb{P}(Y_j = 1) = \lambda_j \) for each \( j \). Show that
\[
d_{TV}((X_1, \ldots, X_m), (Y_1, \ldots, Y_m)) \ll \sum_{j=1}^{m} \lambda_j^2.
\]
Exercise 4.2. For each $j \in \mathbb{N}$, let $Z_j$ be Poisson with parameter $1/j$, and $Z_1, Z_2, \ldots$ independent.

(a) Show that $P(Z_j \leq 1 \forall j) = e^{-\gamma}$.

(b) Let $A_n$ be the probability that a random $\sigma \in S_n$ has distinct cycle sizes. Prove that $\lim_{n \to \infty} A_n = e^{-\gamma}$.

(This is a result of Lehmer from 1972).
CHAPTER 5

Central Limit Theorems

1. Gaussian approximation of Poisson variables

It is well-known that, as \(\lambda \to \infty\) that \(\text{Pois}(\lambda)\) approaches a Gaussian distribution. This is a special case of the Central Limit Theorem. Below we record a quantitative version with explicit error term, and provide an elementary proof.

**Proposition 5.1 (Poisson CLT).** Uniformly for real \(\lambda \geq 1\), \(X \sim \text{Pois}(\lambda)\), and real \(z\), we have
\[
\mathbb{P}(X \leq \lambda + z\sqrt{\lambda}) = \Phi(z) + O\left(\frac{1}{\lambda^{3/2}}\right),
\]
where
\[
\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} \, dt
\]
is the distribution function of the standard Gaussian distribution.

**Proof.** We may assume that \(\lambda\) is sufficiently large. Let \(h^* = 3\sqrt{\lambda \log(1 + \lambda)}\). First observe that by the Poisson Tails Proposition 0.3 and the crude bounds for \(Q(x)\) (0.16), we have
\[
\mathbb{P}(|X - \lambda| > h^*) \leq 2e^{-3\log(1+\lambda)} = \frac{2}{(1 + \lambda)^3}.
\]
Likewise,
\[
\int_{|t| > 3\sqrt{\log(1+\lambda)}} e^{-\frac{1}{2}t^2} \, dt \ll \frac{1}{(1 + \lambda)^3}.
\]
Consequently, we may assume that \(|z| \leq h^*\), and deduce
\[
\mathbb{P}(X \leq \lambda + z\sqrt{\lambda}) = e^{-\lambda} \sum_{\lambda - h^* \leq k \leq \lambda + z\sqrt{\lambda}} \frac{\lambda^k}{k!} + O\left(\frac{1}{\lambda^{3/2}}\right).
\]
For \(|k - \lambda| \leq h^*\), Stirling’s formula implies that
\[
k! = \left(\frac{e}{\lambda}\right)^k \sqrt{2\pi\lambda} \left(1 + O\left(\frac{|k - \lambda|}{\lambda}\right)\right).
\]
Write \(k = \lambda + u\). Then, for \(|u| \leq h^*\), we have
\[
e^{-\lambda} \frac{\lambda^k}{k!} = \frac{1 + O\left(\frac{|u|+1}{\lambda}\right)}{\sqrt{2\pi\lambda}} e^{-\lambda} \left(\frac{e\lambda}{\lambda + u}\right)^{\lambda+u} = \frac{1 + O\left(\frac{|u|+1}{\lambda}\right)}{\sqrt{2\pi\lambda}} \frac{e^u}{(1 + u/\lambda)^{\lambda+u}}
\]
\[
= \frac{1 + O\left(\frac{|u|+1}{\lambda}\right)}{\sqrt{2\pi\lambda}} \exp\left\{ u - (\lambda + u) \left(\frac{u}{\lambda} - \frac{1}{2} \left(\frac{u}{\lambda}\right)^2 + O\left(\frac{(u/\lambda)^3}{\lambda}\right)\right) \right\}
\]
\[
= \left(1 + O\left(\frac{1 + |u|}{\lambda} + \frac{|u|^3}{\lambda^2}\right)\right) e^{-\frac{u^2}{2\lambda}} \frac{1}{\sqrt{2\pi\lambda}}.
\]
It follows that
\[ e^{-\lambda} \sum_{\lambda-h^* \leq k \leq \lambda+z\sqrt{\lambda}} \frac{\lambda^k}{k!} = M + E, \]
where
\[ M = \frac{1}{\sqrt{2\pi \lambda}} \sum_{\lambda-h^* \leq k \leq \lambda+z\sqrt{\lambda}} e^{-\frac{(k-\lambda)^2}{2\lambda}} \]
and
\[ E \ll \frac{1}{\sqrt{\lambda}} \sum_k \left( \frac{1}{\lambda} \left[ 1 + \frac{|k-\lambda|}{\lambda} + \frac{|k-\lambda|^3}{\lambda^2} \right] e^{-\frac{|k-\lambda|^2}{2\lambda}}. \]

For some integer \( a \geq 1 \) we have \((a-1)\sqrt{\lambda} \leq |k-\lambda| \leq a\sqrt{\lambda}\). Summing over all \( a \) gives
\[ E \ll \sum_{a=1}^{\infty} \left( \frac{a + a^3}{\sqrt{\lambda}} \right) e^{-(a-1)^2/2} \ll \frac{1}{\sqrt{\lambda}}. \]

By Euler summation,
\[ M = \frac{1}{\sqrt{2\pi \lambda}} \left[ \int_{\lambda-h^*}^{\lambda+z\sqrt{\lambda}} e^{-\frac{(t-\lambda)^2}{2\lambda}} \, dt - \int_{\lambda-h^*}^{\lambda+z\sqrt{\lambda}} \{t\} \left( \frac{t-\lambda}{\lambda} \right) e^{-\frac{(t-\lambda)^2}{2\lambda}} \, dt + O(1) \right]. \]

The integral involving \( \{t\} \) is \( O(1) \). The first equals, by (5.1),
\[ \sqrt{\lambda} \int_{-3\sqrt{\log(1+\lambda)}}^2 e^{-\frac{1}{2}u^2} \, du = \sqrt{\lambda} \int_{-\infty}^2 e^{-\frac{1}{2}u^2} \, du + O(\lambda^{-5/2}), \]
and hence
\[ M = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^2 e^{-\frac{1}{2}u^2} \, du + O \left( \frac{1}{\sqrt{\lambda}} \right) = \Phi(z) + O \left( \frac{1}{\sqrt{\lambda}} \right). \]

The proof is complete. \( \square \)

2. Central Limit Theorems for cycles

Combining Theorem 4.3 with the Central Limit Theorem for Poisson variables (Theorem 5.1 below) establishes a Central Limit Theorem for the count of cycles whose lengths lie in an arbitrary set \( I \subset [n] \).

**Theorem 5.2.** Let \( I \subset [n] \). Uniformly for all \( I \) and any real \( w \),
\[ \Pr_{\sigma \in \mathcal{S}_n} \left( C_I(\sigma) \leq H(I) + w\sqrt{H(I)} \right) = \Phi(w) + O \left( \frac{\log(2H(I))}{\sqrt{H(I)}} \right). \]

**Proof.** Let \( H = H(I) \). We may assume that \( H \geq 100 \), otherwise the claim is trivial. If \(|w| \geq \sqrt{3 \log H} \) then the result follows from Corollary 1.12, since the left side is thus \( O(1/H) = \Phi(w) + O(1/H) \) if \( w \leq -\sqrt{\log H} \) and is \( 1 - O(1/H) = \Phi(w) + O(1/H) \) if \( w \geq \sqrt{3 \log H} \). Suppose now that \(|w| < \sqrt{3 \log H} \), let
\[ A = H + w\sqrt{H}, \quad m = \left[ \frac{n}{5 \log H} \right], \quad J = I \cap [m]. \]

Because
\[ H(I \setminus J) = \sum_{\substack{m < k \leq n \\ k \in I}} \frac{1}{k} \leq H((m, n] \cap \mathbb{N}) \leq \log \log H + O(1) \]
we have \( H(J) = H + O(\log \log H) \). Thus,
\[ A = H(J) + w'\sqrt{H(J)}, \quad w' = w + O \left( \frac{\log \log H}{\sqrt{H}} \right). \]
Let $Y$ be a Poisson random variable with parameter $H(J)$. Thus, by Theorem 4.3 and Proposition 5.1,
\[
\mathbb{P}_\sigma(C_I(\sigma) \leq A) = \mathbb{P}(Y \leq A) + O(e^{-n/5m})
\]
\[
= \Phi(w') + O \left( \log \frac{H(J)}{5} \right)
\]
\[
= \Phi(w) + O \left( \log \frac{H}{\sqrt{H}} \right).
\]

We also have
\[
A - 5 \log H = H(J) + w'' \sqrt{H(J)}, \quad w'' = w + O \left( \frac{\log H}{\sqrt{H}} \right)
\]
and it follows that
\[
\mathbb{P}_\sigma(C_I(\sigma) \leq A) \geq \mathbb{P}_\sigma(C_J(\sigma) \leq A - 5 \log H) \quad \text{and} \quad C_{I \setminus J}(\sigma) \leq 5 \log H,
\]
since $\min(I \setminus J) \geq m \geq n/(5 \log H)$ implies that $C_{I \setminus J}(\sigma) \leq 5 \log H$ always. Hence, by Theorem 4.3 and Lemma 5.1,
\[
\mathbb{P}_\sigma(C_I(\sigma) \leq A) \geq \Phi(w'') + O \left( \frac{\log H}{\sqrt{H}} \right)
\]
\[
= \Phi(w) + O \left( \frac{\log H}{\sqrt{H}} \right).
\]
The theorem follows by combining the upper and lower bounds for $\mathbb{P}(C_I(\sigma) \leq A)$. 

The special case $I = [n]$ was established by Goncharov [35], without a specific rate of convergence. Goncharov analyzed carefully the asymptotics of the Stirling number of the first kind, $s(n,m)$, the absolute value of which counts the number of permutations $\sigma \in S_n$ with $C(\sigma) = m$. Since $H_n = \log n + O(1)$ and $\Phi$ has bounded derivative, we quickly arrive at the following.

**Theorem 5.3.** Let $n \geq 100$ and $w$ be real. Then
\[
\mathbb{P}_{\sigma \in S_n} \left( C(\sigma) \leq \log n + w \sqrt{\log n} \right) = \Phi(w) + O \left( \frac{\log_2 n}{\sqrt{\log n}} \right).
\]

**Proof.** Letting $\log n + w \sqrt{\log n} = H_n + w' \sqrt{H_n}$, we have
\[
w - w' \leq \frac{1}{\sqrt{\log n}}.
\]
Hence,
\[
\mathbb{P}_{\sigma} \left( C(\sigma) \leq \log n + w \sqrt{\log n} \right) = \Phi(w') + O \left( \frac{\log H_n}{\sqrt{H_n}} \right) = \Phi(w) + O \left( \frac{\log_2 n}{\sqrt{\log n}} \right).
\]

The big-$O$ term in Theorem 5.2 cannot be made smaller than $1/\sqrt{H(I)}$ since $C_I(\sigma)$ is integer valued, and thus the left side is constant in intervals of $w$ of length $1/\sqrt{H(I)}$, while $\Phi'(w) \gg 1$ if $w$ is bounded. We remark that when $H(I)$ is bounded, $C_I(\sigma)$ is expected to have Poisson distribution with small parameter, and this cannot be approximated by a Gaussian.

We also derive that the $j$-th smallest cycle of $\sigma$, denoted $D_j(\sigma)$ (with ties allowed), also obeys the Gaussian law, refining Theorem 1.22.
**Theorem 5.4.** Uniformly for $j$ in the range

$$1 \leq j \leq \log n - \sqrt{\log n \log \log n}$$

and for any real $w$,

$$P_{\sigma \in S_n} \left( \log D_j(\sigma) \leq j + w\sqrt{j} \right) = \Phi(w) + O \left( \frac{\log(2j)}{\sqrt{j}} \right).$$

**Proof.** We may assume that $j \geq 10$ and that $n$ is sufficiently large, the statement being trivial otherwise. We may also assume that $|w| \leq \sqrt{\log j}$, since the statement for $w$ outside this range follows from the monotonicity of $P(\log D_j(\sigma) \leq j + w\sqrt{j})$, as a function of $w$, the statement for the two points $w = \pm \sqrt{\log j}$ and the fact that $\Phi(-\sqrt{\log j}) \lesssim 1/\sqrt{j}$ and $\Phi(\sqrt{\log j}) = 1 - O(1/j^{1/2})$.

Let $k = \left[ e^{j+w\sqrt{j}} \right]$, so by hypothesis,

$$\log k \leq j + \sqrt{j} \log j \leq j + (\log n) \log \log n \leq \log n.$$ 

Then $D_j(\sigma) \leq k$ is equivalent to $C_{[k]}(\sigma) \geq j$. As $H_k = \log k + O(1)$ and $\sqrt{H_k} = \sqrt{j} + O(|w| + 1)$, we have

$$j - 1 = H_k - u\sqrt{H_k}, \quad \text{where} \quad u = w + O \left( \frac{w^2 + 1}{\sqrt{j}} \right).$$

By Theorem 5.2,

$$P_{\sigma} (D_j(\sigma) \leq k) = P_{\sigma} (C_{[k]}(\sigma) \geq j) = 1 - P_{\sigma} (C_{[k]}(\sigma) \leq j - 1) = 1 - \Phi(-u) + O \left( \frac{\log H_k}{\sqrt{H_k}} \right) = \Phi(u) + O \left( \frac{\log(2j)}{\sqrt{j}} \right).$$

Also,

$$\Phi(u) = \Phi(w) + O \left( \frac{w^2 + 1}{\sqrt{j}} \right) = \Phi(w) + O \left( \frac{\log(2j)}{\sqrt{j}} \right)$$

and the proof is complete. \qed

**3. Central Limit theorems for prime factors**

**Theorem 5.5 (Prime factors CLT).** Suppose that $T$ is a subset of the primes in $[2, x]$. For any real $w$,

$$P_x \left( \omega(n; T) \leq H(T) + w \sqrt{H(T)} \right) = \Phi(w) + O \left( \frac{\log(2H(T))}{\sqrt{H(T)}} \right).$$

**Proof.** Let $H = H(T)$. Assume $H \geq 100$, else the conclusion is trivial. As in the proof of Theorem 5.2, the conclusion in the case $|w| \geq 3 \log H$ follows from Theorem 1.13 and Proposition 0.3.

Suppose now that $|w| < \sqrt{3 \log H}$, let

$$A = H + w\sqrt{H}, \quad y = x^{1/(5 \log H)}, \quad J = T \cap [2, y].$$

Because

$$H(T \setminus J) \leq \sum_{y < p \leq x} \frac{1}{p} \leq \log H + O(1)$$

we have $H(J) = H + O(\log \log H)$. Thus,

$$A = H(J) + w\sqrt{H(J)}, \quad w' = w + O \left( \frac{\log \log H}{\sqrt{H}} \right).$$
Let $Y$ be a Poisson random variable with parameter $H(J)$. Thus, by Theorem 4.9 and Proposition 5.1,
\[
P_x(\omega(n; T) \leq A) \leq P_x(\omega(n; J) \leq A) \\
= P(Y \leq A) + O(1/H) \\
= \Phi(w') + O\left(\frac{1}{\sqrt{H}}\right) \\
= \Phi(w) + O\left(\frac{\log \log H}{\sqrt{H}}\right).
\]

We also have
\[
A - 5 \log H = H(J) + w'' \sqrt{H(J)}, \quad w'' = w + O\left(\frac{\log H}{\sqrt{H}}\right)
\]
and it follows that
\[
P_x(\omega(n; T) \leq A) \geq P_x(\omega(n; J) \leq A - 5 \log H \text{ and } \omega(n; T \setminus J) \leq 5 \log H) \\
= P_x(\omega(n; J) \leq A - 5 \log H),
\]
since $\min(T \setminus J) \geq y$ implies that $\omega(n; T \setminus J) \leq 5 \log H$ always. Hence, by Theorem 4.9 and Lemma 5.1,
\[
P_x(\omega(n; T) \leq A) \geq \Phi(w'') + O(1/\sqrt{H}) \\
= \Phi(w) + O\left(\frac{\log H}{\sqrt{H}}\right).
\]
The theorem follows by combining the upper and lower bounds for $P(\omega(n; T) \leq A)$.

\[\Box\]

We remark that an error term $H(T)^{-1/2}$ is best possible; in fact, the error term may be replaced by an asymptotic expansion in powers of $H(T)^{-1/2}$; this is a consequence of a general theory of CLT-type expansions for sums of random variables; see [34].

As with cycles of permutations, we may extend this to handle the distribution of all prime factors of integers, or any function $\omega(n, T)$ where the large prime factors contribute ‘negligibly’. The following is a quantitative version of the famous theorem of Erdős and Kac [25].

**Theorem 5.6 (CLT for all prime factors).** For any real $w$,
\[
P_x\left(\omega(n) \leq \log_2 x + w \sqrt{\log_2 x}\right) = \Phi(w) + O\left(\frac{\log_3 x}{(\log_2 x)^{1/2}}\right).
\]

**Proof.** Let $P$ be the set of all primes $\leq x$, $H = H(P)$, and
\[
\log_2 x + w \sqrt{\log_2 x} = H + w' \sqrt{H}.
\]
Since $H = \log_2 x + O(1)$, we have $w - w' \ll (\log_2 x)^{-1/2}$. Thus, by Theorem 5.5,
\[
P_x\left(\omega(n) \leq \log_2 x + w \sqrt{\log_2 x}\right) = \Phi(w') + O\left(\frac{\log H}{H}\right) = \Phi(w) + O\left(\frac{\log_3 x}{(\log_2 x)^{1/2}}\right).
\]

An error term of $O((\log_2 x)^{-1/2})$ in Theorem 5.6 is best possible (cf. work of Rényi-Turán, Delange and Kubilius in the late 1950s/early 1960s).

We also derive that the $j$-th smallest prime factor of $n$, denoted $p_j(n)$ (with ties allowed), also obeys the Gaussian law, refining Theorem 1.21.

**Theorem 5.7.** Uniformly for $j$ in the range
\[
1 \leq j \leq \log_2 x - \sqrt{(\log_2 x) \log_3 x}
\]
and for any real $w$, we have
\[
P_x\left(\log_2 p_j(n) \leq j + w \sqrt{j}\right) = \Phi(w) + O\left(\frac{\log(2j)}{\sqrt{j}}\right).
\]
A version of this, without a rate of convergence, was proved by Galambos [33].

4. Exercises

**Exercise 5.1.** Prove Theorem 5.7.

**Exercise 5.2.** (a) Let $T_1, T_2$ be disjoint sets of primes $\leq x$. Show that, uniformly for all real $w_1, w_2$,
\[
P_x \left( \omega(n; T_1) \leq H(T_1) + w_1 \sqrt{H(T_1)}, \omega(n; T_1) \leq H(T_1) + w_2 \sqrt{H(T_1)} \right) = \Phi(w_1)\Phi(w_2) + O \left( \frac{\log(2^x \sqrt{H(T_1)})}{\sqrt{H(T_1)}} \right).
\]

(b) Let $T_1$ denote the set of primes $\leq x$ that are $1 \mod 4$, and let $T_2$ denote the set of primes $\leq x$ that are $3 \mod 4$. Show that
\[
P_x \left( \omega(n; T_1) \leq \frac{1}{2} \log_2 x + w_1 \sqrt{\frac{1}{2} \log_2 x}, \omega(n; T_2) \leq \frac{1}{2} \log_2 x + w_2 \sqrt{\frac{1}{2} \log_2 x} \right) = \Phi(w_1)\Phi(w_2) + O \left( \frac{\log x}{\sqrt{\log_2 x}} \right).
\]

**Exercise 5.3.** (Galambos, 1976 [33]). Fix $\varepsilon > 0$ and $z \in \mathbb{R}$. Show that if $j = j(x) \to \infty$ as $x \to \infty$ and $j(x) \leq (1 - \varepsilon) \log_2 x$, then
\[
P_x \left( \log q_{j+1}(n) - \log q_j(n) \leq z \right) \to 1 - e^{-z} \quad (x \to \infty).
\]
Here $q_1(n) \leq q_2(n) \leq \cdots$ are the prime factors of $n$, with duplicates allowed. Hint: for a given prime $p$, consider the event
\[
\{ \omega(n, p - 1) = j - 1, p | n, \omega(n; (p, p^{\varepsilon x}] \cap \mathbb{N}) \geq 1 \}.
\]

**Exercise 5.4.** Fix $\varepsilon > 0$ and $z \in \mathbb{R}$. Show that if $j = j(n) \to \infty$ as $n \to \infty$ and $j(n) \leq (1 - \varepsilon) \log n$, then
\[
P_{\sigma \in S_n} \left( \log D_{j+1}(\sigma) - \log D_j(\sigma) \leq z \right) \to 1 - e^{-z} \quad (n \to \infty).
\]
This matches the spacing distribution of points in a Poisson process.
CHAPTER 6

The concentration of divisors of integers and permutations

1. Concentration of divisors

Erdős conjectured [21] in 1948 that almost all integers have two divisors $d$ and $d'$ with $d < d' < 2d$. This may seem counterintuitive, given that we have already shown (cf. Theorem 1.21) that a typical integer has about $2^i$ divisors less than $e^i$, and hence the $k$-th smallest divisor of a typical integer, for $k$ large, is about $\exp\{k^{1/\log 2}\}$. Hence, one may be led to believe that $d_k+1/d_k \to \infty$ for a typical $n$, as long as $k \to \infty$ (and $\tau(n) - k \to \infty$ by symmetry). This, however, is faulty reasoning, as (i) as we see in Theorem 5.7, there is a very large “normal” range of deviation for each prime $\tau$ (and about $\exp\{e^{1/\log 2}\}$). Therefore, one may be led to believe that $n$.

We define the Erdős-Hooley $\eta$-function $\Delta(n) := \max \#\{d \in \mathbb{N} : \sigma(d) \leq \log n\}$, that is to say the maximum number of divisors $n$ has in any interval of logarithmic length 1. Its normal order (almost sure behavior) has proven quite mysterious. Work on the distribution of $\Delta(n)$ began with Erdős [15], Erdős and Nicolas [17, 18] and Hooley [44] in the 1970s. Further work on the normal and average behavior of $\Delta(n)$ can be found in works of Tenenbaum [57, 58], Hall and Tenenbaum [38, 39, 40], Maier and Tenenbaum [47, 48, 49], and most recently Ford, Green and Koukoulopoulos [32]. See also [41, Ch. 5,6,7]. Tenenbaum’s survey paper [60, p. 652–658] includes a history of the function $\Delta(n)$ and description of many applications in number theory.

The best bounds for $\Delta(n)$ for “normal” $n$ currently known were obtained in papers of Maier and Tenenbaum [49] (upper bound) and Ford, Green and Koukoulopoulos [32] (lower bound). For almost all $n$ we have

$$ (\log_2 n)^{1-o(1)} \leq \Delta(n) \leq (\log_2 n)^{\log 2+o(1)}, $$

where $\eta = 0.35332277270132346711 \ldots$ is a specific constant.

In this section, we prove a weaker lower bound, recovering the bound proved by Maier and Tenenbaum [47].
Theorem 6.2. Fix \( \varepsilon > 0 \). For almost all \( n \) we have \( \Delta(n) \geq (\log \log n)^{n_1 - \varepsilon} \), where

\[
\eta_1 = \frac{\log 2}{\log \left( \frac{\log 3}{\log 5} \right)} = 0.2875404895 \ldots
\]

The same argument will provide a measure of the concentration of divisors of random permutations.

Theorem 6.3. For a permutation \( \sigma \) on \( S_n \), denote by

\[
\Delta(\sigma) := \max_r \#\{ \tau | \sigma : |\tau| = r \}.
\]

Then, for all but \( o(n!) \) of the permutations \( \sigma \in S_n \), we have

\[
\Delta(\sigma) \geq (\log n)^{n_1 - o(1)}.
\]

Here the terms \( o(1) \) refer to functions that \( \to 0 \) as \( n \to \infty \).

2. A random model of prime factors and cycle lengths

As we have seen by Theorem 4.2, the number of cycles of size \( i \) in a random permutation is well-approximated by a \( \text{Pois}(1/i) \) variable. Also, by Theorem 4.9, \( \omega(n; T) \) is well approximated by \( \text{Pois}(H(T)) \) for a set \( T \) of primes which are \( \leq x^{o(1)} \) and not “too thin”. In particular, if \( i \) is much larger than \( K \) and \( T \) is the set of primes in \( (e^{i/K}, e^{i+1/K}] \) then by Mertens’ Theorem (0.5), \( H(T) \approx 1/i \) and hence \( \omega(n; T) \) is well-approximate by \( \text{Pois}(1/i) \). In turn, when \( i \) is large, by Exercise 4.1 (b), \( \text{Pois}(1/i) \) is well-approximated by a Bernouilli random variable \( Y \) with \( \mathbb{P}(Y = 1) = 1/i \). This motivates the following random model.

Definition 6.4. We define \( A \) to be the random set of positive integers such that \( \mathbb{P}(i \in A) = 1/i \) for each \( i \), and the events \( i \in A \) are independent for different values of \( i \). That is, if \( Y_1, Y_2, \ldots \) are independent Bernouilli random variables with \( \mathbb{P}(Y_i = 1) = 1/i \) for each \( i \), then \( A = \{ i : Y_i = 1 \} \).

The property that an integer has \( k \) close divisors can be modeled by the event that \( A \) has \( k \) equal subset sums.

Definition 6.5. Let \( k \geq 2 \) be an integer. Let \( \beta_k \) be the supremum of all exponents \( c < 1 \) for which the following is true with probability \( \to 1 \) as \( D \to \infty \): there are distinct sets \( A_1, \cdots, A_k \subset A \cap [D^c, D] \) with

\[
\sum_{a \in A_1} a = \cdots = \sum_{a \in A_k} a.
\]

It is not a priori obvious that \( \beta_k \) exists. We will establish this later in Theorem 6.9. In particular, we have \( 0 < \beta_k < 1/10 \) for all \( k \).

Define

\[
(6.2) \quad \zeta_k = \frac{\log k}{\log(1/\beta_k)}.
\]

Theorem 6.6 ([32]). Fix \( k \geq 2 \) and \( \varepsilon > 0 \). For almost every positive integer \( n \), we have

\[
\Delta(n) \geq (\log \log n)^{\zeta_k - \varepsilon}.
\]

Theorem 6.7 ([32]). Fix \( k \geq 2 \) and \( \varepsilon > 0 \). Then, for all but \( o_{n \to \infty}(n!) \) of the permutations \( \sigma \in S_n \), we have

\[
\Delta(\sigma) \geq (\log n)^{\zeta_k - \varepsilon}.
\]

We also prove a more general version of Theorem 6.1. Define \( \alpha_k \) be the supremum of all real numbers \( \alpha \) such that for almost every \( n \in \mathbb{N} \), \( n \) has \( k \) divisors \( d_1 < \cdots < d_k \) with \( d_k \leq d_1(1 + (\log n)^{-\alpha}) \). In 1964, Erdős [14] conjectured that \( \alpha_2 = \log 3 - 1 \), and this was confirmed by Erdős and Hall [16] (upper bound) and Maier and Tenenbaum [47] (lower bound). Maier and Tenenbaum [49] showed that

\[
\alpha_k \leq \frac{\log 2}{k + 1} \quad (k \geq 3)
\]
and (this is not stated explicitly in [49])
\begin{equation}
\alpha_k \geq \frac{(\log 3 - 1)^{m-1} 3^{m-1}}{(3 \log 3 - 1)^{m-1}} \quad (2^{m-1} < k \leq 2^m, m \in \mathbb{N}).
\end{equation}

See also [60, p. 655–656]¹. In particular, it is not known if \( \alpha_3 > \alpha_4 \), although Tenenbaum [60] conjectures that the sequence \( (\alpha_k)_{k \geq 2} \) is strictly decreasing. Ford, Green and Koukoulopoulos connected \( \alpha_k \) to the constant \( \beta_k \) from the random model.

It is easy to see that for any \( k \geq 2 \), \( \alpha_k \leq \alpha_2 \leq 1 \). Indeed, \( \alpha_k \leq \alpha_2 \) is obvious. Let \( \delta = (\log x)^{1-\varepsilon} \) for arbitrary fixed \( \varepsilon > 0 \). If \( n \) has two divisors \( d, d' \) with \( d < d' \leq d(1 + \delta) \), then WLOG there exist two such divisors with \( (d, d') = 1 \), hence \( dd'/n \). Thus, the number of \( n \leq x \) divisible by two such divisors is at most
\[
\sum_{1/\sqrt{x} \leq d \leq \sqrt{x}} x \frac{x}{dd'} \leq x \sum_{1/\sqrt{x} < d < (1+\delta)} \frac{d\delta}{d^2} \ll \delta x \log x = o(x)
\]
as \( x \to \infty \). This proves \( \alpha_2 \leq 1 \).

**Theorem 6.8 ([32]).** For all \( k \geq 2 \), \( \alpha_k \leq \beta_k/(1 - \beta_k) \).

The authors in [32] conjecture the corresponding upper bound \( \alpha_k \leq \beta_k/(1 - \beta_k) \). The methods of [32] allow one to compute \( \beta_k \) for any \( k \) using a finite procedure. In particular, we have
\[
\beta_3 = \frac{\log 3 - 1}{\log 3 + \frac{1}{\xi}} = 0.02616218797316965133\ldots
\]
and
\[
\beta_4 = \frac{\log 3 - 1}{\log 3 + \frac{1}{\xi} + \frac{1}{\lambda}} = 0.01295186091360511918\ldots
\]
where
\[
\xi = \frac{\log 2 - \log(e - 1)}{\log(3/2)}, \quad \lambda = \frac{\log 2 - \log(e - 1)}{1 + \log 2 - \log(e - 1) - \log(1 + 2^{1-\xi})}.
\]

In [32] it is also proved that
\[
\sup_{k \geq 2} \xi_k \geq \eta = 0.3533227727\ldots
\]
which, combined with Theorem 6.6, gives the lower bound in (6.1). The authors of [32] conjecture that \( \limsup_{k \to \infty} \xi_k = \eta \). Most of the paper [32] is devoted to relating \( \beta_k \) to a certain combinatorial optimization problem, which grows very complex as \( k \) increases.

In these notes we will concentrate on the easiest case \( k = 2 \) and prove

**Theorem 6.9.** We have \( \beta_2 = 1 - \frac{1}{\log 3} = 0.0897607\ldots \) and, for any \( k \geq 2 \),
\[
\beta_k \geq \frac{\lfloor \log k \rfloor}{\log 2}.
\]
Trivially, \( \beta_k \leq \beta_2 \) for all \( k \). Thus, we see that \( \beta_k \) exists for all \( k \geq 2 \) and \( 0 < \beta_k \leq \beta_2 < 1/11 \) for all \( k \). Combining Theorem 6.9 with Theorems 6.6 and 6.7 gives Theorems 6.2 and 6.3. Also, combining Theorems 6.8 and 6.9 gives Theorem 6.1.

It remains to prove Theorems 6.6, 6.7, 6.8, and 6.9.

**Further Remarks.** The average order of \( \Delta(n) \) is also rather mysterious, the best known bounds being
\[
\log \log x \leq \sum_{n \leq x} \Delta(n) \leq e^{\sqrt{\log \log x}},
\]
for some \( c > 0 \). The lower bound is due to Hall and Tenenbaum (see [41, Theorem 60]) and the upper bound is from an unpublished manuscript of Koukoulopoulos which slightly refines a bound of Tenenbaum [57] (see also [41, Theorem 70]).

---

¹The factor \( 3^{m-1} \) is missing in the stated lower bounds for \( \alpha_k \) in [60].
3. Proof of Theorems 6.6, 6.7, and 6.8

In this section we assume Theorem 6.9 and deduce Theorems 6.6, 6.7, and 6.8. In particular, \( 0 < \beta_k < 1/10 \) for all \( k \).

We begin with a simple combinatorial argument, first used in a related context in the work of Maier-Tenenbaum [47], which shows how to use equal subsums in multiple intervals \( (u^\varepsilon, u] \) to create many more common subsums in \( A \). For any finite subset \( S \) of positive integers, denote by \( \Sigma S \) the sum of the elements of \( S \).

**Lemma 6.10.** Let \( k \geq 1 \) be an integer and fix \( \varepsilon > 0 \). Let \( C, D \) be parameters with \( \log u \leq (\log v)^{o(1)} \) and \( u \to \infty \) as \( v \to \infty \). Then, with probability \( \to 1 \) as \( v \to \infty \), there are distinct \( A_1, \ldots, A_M \subset A \cap (u, v] \) with \( \sum A_1 = \cdots = \sum A_M \) and \( M \geq (\log v)^{\varepsilon_k - \varepsilon} \).

**Proof.** Fix \( \delta \in (0, \beta_k) \), depending on \( \varepsilon \), set \( \alpha := \beta_k - \delta \) and assume that \( v \) is sufficiently large in terms of \( \varepsilon, \delta \). Set

\[
m := \left\lfloor \frac{\log \log v - \log \log u}{-\log \alpha} \right\rfloor
\]

and consider the intervals \( J_i := (u^{\alpha^{i+1}}, u^{\alpha^i}] \), \( i = 0, 1, \ldots, m - 1 \). Due to the choice of \( m \), these all lie in \( (u, v] \). For each \( i \in \{0, 1, \ldots, m - 1\} \), let \( E_i \) be the event that there are distinct \( A_{i,1}, \ldots, A_{i,k} \subset J_i \) with equal sums. Then, by the definition of \( \beta_k \), if \( v \) is large enough then \( \mathbb{P}(E_i) \geq 1 - \delta \), uniformly in \( i \). These events \( E_i \) are all independent. The Law of Large Numbers then implies that, with probability at least \( 1 - \delta \), at least \( (1 - 2\delta)m \) of them occur. Suppose we are in this event, suppose that \( E_i \) occurs for \( i \in I \), where \( |I| \geq (1 - 2\delta)m \). Then there are numbers \( s_i, i \in I \), and sets \( A_{i,1}, \ldots, A_{i,k} \subset J_i \) all with sum \( s_i \).

It is now clear that for each \( j = \{j_i : i \in I\} \in [k]^I \), the set

\[
B_j = \bigcup_{i \in I} A_{i,j_i}
\]

lies in \( (u, v] \) and has sum \( \sum_{i \in I} s_i \). Moreover, these sets are distinct. Let \( M = k^{|I|} \), then

\[
M \geq k^{(1 - \varepsilon_k)m} \geq k^{-1} \left( \frac{\log v}{\log u} \right)^{(1 - 2\varepsilon_k)/(\log k)}.
\]

Recall the definition (6.2) of \( \zeta_k \). By our assumption on \( u \), and if \( \varepsilon \) is small enough, the right side is \( \geq (\log v)^{\varepsilon_k - \varepsilon} \), as required. \( \square \)

**Lemma 6.11.** Let \( x \) be a large parameter, suppose that

\[
1 \leq K \leq (\log x)^{1/2},
\]

and let \( I = (u, v] \cap \mathbb{N} \), where

\[
u = \left\lfloor 100K(\log_2 x)^2 \right\rfloor, \quad v = \left\lfloor \frac{K \log x}{\log \log x} \right\rfloor - 1.
\]

For \( i \in I \), let \( T_i \) be the set of primes in \( (\varepsilon^{i/K}, \varepsilon^{(i+1)/K}] \), and define the random set

\[
\mathcal{B} = \{i \in (u, v] : \omega(u; T_i) \geq 1\}.
\]

Uniformly for any collection \( \mathcal{J} \) of subsets of \( I \), we have

\[
\mathbb{P}(A \cap I \in \mathcal{J}) = \mathbb{P}_x(\mathcal{B} \in \mathcal{J}) + O(1/\log_2 x).
\]

**Proof.** For \( u < i < v \), let \( \omega_i = \mathbb{1}(\omega(u; T_i) \geq 1) \), \( P_i \overset{d}{=} \text{Pois}(H(T_i)) \) (these being independent for different \( i \)) and let \( Z_i = \mathbb{1}(P_i \geq 1) \). Each \( T_i \) is contained in \( \lfloor \log x, y \rfloor \), where \( y = x^{1/5 \log_3 x} \). Hence, by Theorem 4.9 and (6.5),

\[
d_{TV}(\omega, Z) \ll \frac{1}{\log_2 x} + \sum_{i \in I} H''(T_i) \ll \frac{1}{\log_2 x}.
\]
3. Proof of Theorems 6.6, 6.7, and 6.8

By the strong form of Mertens’ estimate (0.5),
\[ H(T_i) = \log \left( \frac{i + 1}{K} \right) - \log \left( \frac{i}{K} \right) + O(e^{-(i/K)^{1/2}}) = \frac{1}{i} + O \left( \frac{1}{i^2} \right). \]

Hence, if \( Y_i \) is a Bernoulli variable with \( \mathbb{P}(Y_i = 1) = 1/i \), then by Exercise 4.1 (b),
\[ d_{TV}(Z, Y) \leq \sum_{i \in I} d_{TV}(Z_i, Y_i) \ll \sum_{i \in I} \frac{1}{i^2} \ll \frac{1}{\log_2 x}. \]

By the triangle inequality,
\[ d_{TV}(\omega, Y) \leq d_{TV}(\omega, Z) + d_{TV}(Z, Y) \ll \frac{1}{\log_2 x} \]
and the claim follows. \( \square \)

**Proof of Theorem 6.6.** Fix \( \varepsilon > 0 \), let \( x \) be large, let \( K = (\log_2 x)^2 \) and define \( u, v \) by (6.5). Define \( B \) and sets \( T_i \) as in Lemma 6.11. Throughout the proof, \( o(1) \) means a function \( \to 0 \) as \( x \to \infty \). By Lemma 6.10, with probability \( 1 - o(1) \), there are distinct sets \( A_1, \ldots, A_M \) of \( A \cap (u, v) \) with equal sums and \( M \geq (\log_2 x)^{\varepsilon k - \varepsilon} \). We also have that with probability \( 1 - o(1) \), \( |A \cap (u, v)| \leq 2 \log v \). Indeed, by Markov’s inequality,
\[ \mathbb{P}(|A \cap (u, v)| > 2 \log v) = \mathbb{P}(2|A \cap (u, v)| > 2^{2 \log v}) \leq v^{- \log 4} \mathbb{E}[2|A \cap (u, v)|] = v^{- \log 4} \prod_{j = u + 1}^{D} \left( 1 - \frac{1}{i} + \frac{2}{i} \right) = (v/u)v^{- \log 4} = o(1) \quad (v \to \infty). \]

Let \( F \) be the event that \( A \cap (u, v) \) has at most \( 2 \log v \) elements and has \( M \) distinct subsets with equal sums. By the above discussion, \( \mathbb{P}(F) = 1 + o(1) \). By Lemma 6.11, the corresponding event \( F' \) for the random set \( B \) also holds with probability \( 1 - o(1) \); that is, \( F' \) is the event that \( |B \cap (u, v)| \leq 2 \log v \) and that there are distinct subsets \( B_1, \ldots, B_M \) of \( B \) with equal sums. If we are in the event \( F' \) and \( n \) is divisible by \( \prod_{b \in B} p_b \), where \( p_b \in T_b \) for each \( b \in B \), then let \( d_i = \prod_{b \in B} p_b \) for each \( i \leq M \). For \( 1 \leq i < j \leq M \) we have
\[ |\log d_i - \log d_j| = \left| \sum_{b \in B_1} \log p_b - \sum_{b \in B_2} \log p_b \right| \leq \frac{|B_i| + |B_j|}{K} \sum_{b \in B_1} b - \sum_{b \in B_2} b \]
\[ = \frac{|B_i| + |B_j|}{K} \leq \frac{4 \log v}{K} \ll \frac{1}{\log_2 x}. \]

Thus, there are \( M \) divisors \( d_i \) of \( n \) whose logarithms all lie in a single interval of length \( O(1/\log_2 x) < 1 \). It follows that \( \mathbb{P}_x(\Delta(n) \geq M) = 1 - o(1) \), as required for Theorem 6.6. \( \square \)

**Proof of Theorem 6.7.** Fix \( \varepsilon > 0 \). Let \( u = \log n \) and \( v = n/\log n \). For each \( j \), let \( Z_j \overset{d}{=} \text{Pois}(1/j) \), with \( Z_1, Z_2, \ldots \) independent. For a random permutation \( \sigma \in S_n \), let \( C = \{ j : C_j(\sigma) \geq 1 \} \), and define the random set \( \tilde{A} = \{ j : Z_j \geq 1 \} \). By Theorem 4.2, Exercise 4.1 (b), and the triangle inequality, we have
\[ d_{TV}(A \cap (u, v), C \cap (u, v)) \leq d_{TV}(A \cap (u, v), \tilde{A} \cap (u, v)) + d_{TV}(\tilde{A} \cap (u, v), C \cap (u, v)) = o(1) \]
as \( n \to \infty \). By Lemma 6.10, with probability \( \to 1 \) as \( n \to \infty \), \( A \cap (u, v) \) has \( M \) distinct subsets \( A_1, \ldots, A_M \) with equal sums, where \( M \geq (\log v)^{\varepsilon k - \varepsilon} \). Hence, \( C \) has distinct subsets \( S_1, \ldots, S_M \) with equal sums with probability \( \to 1 \) as \( n \to \infty \). Each subset \( S_j \) corresponds to a distinct divisor of \( \sigma \), the size of the divisor being the sum of elements of \( S_j \). As \( \varepsilon > 0 \) is arbitrary, the result follows. \( \square \)
6. THE CONCENTRATION OF DIVISORS OF INTEGERS AND PERMUTATIONS

Proof of Theorem 6.8. Fix $0 < c < \frac{\beta_k}{-\beta_k}$, so that $c < 1/10$. Let $x$ be large and set $K = (\log x)^c$.

Define $u, v$ by (6.5), let $D = v$ and define $c'$ by $D^{c'} = u$. By assumption,
\[ c' \sim \frac{c}{c + 1} \quad (x \to \infty) \]
and therefore there is a $\delta > 0$ so that $c' \leq \beta_k - \delta$ for sufficiently large $x$. Let $n$ be a random integer chosen uniformly in $[1,x]$. With probability $\rightarrow 1$ as $x \to \infty$, $\omega(n) \leq 2 \log x$ (e.g., Theorem 1.8). By the definition of $\beta_k$ and Lemma 6.11, with probability $1 - o(1)$, the set $B$ defined in that Lemma has $k$ distinct subsets $B_1, \ldots, B_k$ with equal sums. For each $b \in B$, let $p_b$ be a prime factor of $n$ lying in $T_b$. Let $d_i = \prod_{b \in B_i} p_b$, for $1 \leq i \leq k$. We have
\[ \left| \sum_{b \in B_j} \log p_b - \sum_{b \in B_j} \log p_b \right| \leq \frac{|B_j| + |B_j|}{K} \leq \frac{4 \log_2 x}{(\log x)^c}. \]
Thus,
\[ \max(d_j) \leq \min(d_j) \exp\left\{ O\left( \frac{\log_2 x}{(\log x)^c} \right) \right\} = \min(d_j) \left( 1 + O\left( \frac{\log_2 x}{(\log x)^c} \right) \right). \]
Since $c$ is arbitrary subject to $c < \beta_k/(1 - \beta_k)$, we conclude that $\alpha_k \geq \beta_k/(1 - \beta_k)$.

It remains to prove Theorem 6.9, which we accomplish in the next section.

4. Proof of Theorem 6.9

Our proof use the method introduced by Maier and Tenenbaum in [47], adapted to the random set $A$. This can be thought of as a 'global-to-local' principle, where the local behavior of divisor ratios is deduced from a result about the global distribution of ratios. A similar principle will be utilized in the study of integers with a divisor in a given interval.

It is easy to see the claimed lower bound on $\beta_2$ given that $\beta_2$ exists, following the idea in Lemma 6.10. Indeed, fix very small $\varepsilon > 0$ and $m \in \mathbb{N}$, and set $\alpha = \beta_2 - \varepsilon$. Let $D$ be large and set $I_i = (D^{\alpha'}, D^{\alpha'+1})$ for $1 \leq i \leq m$. By the definition of $\beta_2$, with probability $\rightarrow 1$ as $D \to \infty$ there are distinct sets $A_i, A_j \in I_i, 1 \leq i \leq m, 1 \leq j \leq 2$ such that
\[ \Sigma A_{i,1} = \Sigma A_{i,2} \quad (1 \leq i \leq m). \]
It follows that the $2^m$ sets $B_j = \bigcup_{i=1}^m A_{i,j}$, where $j = (j_1, \ldots, j_m) \in \{1,2\}^m$, all have equal sums. It follows that $\beta_{2^m} \geq \alpha^m$. As $\varepsilon$ is arbitrary, $\beta_{2^m} \geq \beta_{2^m}^0$. The claim now follows from the obvious fact that $(\beta_k)_{k=2}^\infty$ is a decreasing sequence ($k+1$ equal subset sums implies $k$ equal subset sums).

It thus remains to establish the value of $\beta_2$. First, we show that $|A \cap I| \approx H(I)$ with high probability.

**Lemma 6.12.** Let $I$ be a finite set of positive integers, and $1 \leq \theta \leq \sqrt{H(I)}$. Then
\[ \mathbb{P}\left(\left| |A \cap I| - H(I) \right| \geq \theta \sqrt{H(I)} \right) \leq 2 e^{-\frac{1}{4} \theta^2}. \]

**Proof.** Let $H = H(I)$. By hypothesis, $H \geq 1$. For any $\lambda > 0$ we have
\[ \mathbb{E} \lambda^{|A \cap I|} = \prod_{j \in I} \left( 1 + \frac{\lambda - 1}{j} \right) \leq e^{(\lambda - 1)H}. \]
Take $\lambda = 1 + \theta/\sqrt{H}$. Then
\[ \mathbb{P}\left( |A \cap I| \geq H + \theta \sqrt{H} \right) \leq \mathbb{E} \lambda_1^{|A \cap I|-H-\theta \sqrt{H}} \leq \lambda_1^{-H-\theta \sqrt{H}} e^{(\lambda_1 - 1)H} = e^{-Q(\lambda_1)} \leq e^{-\frac{1}{4} \theta^2} \]
using (0.4) at the last step. Similarly, if $\lambda_2 = 1 - \theta/\sqrt{H}$ then
\[ \mathbb{P}\left( |A \cap I| \leq H - \theta \sqrt{H} \right) \leq \mathbb{E} \lambda_2^{|A \cap I|-H+\psi \sqrt{H}} \leq e^{-Q(\lambda_2)} \leq e^{-\frac{1}{4} \theta^2}. \]
**Lemma 6.13.** Given $1 \leq C < D$, $\psi \geq 1$ and $0 < \varepsilon \leq 1$, the probability that both
\[
\left| \#(A \cap (C, v)) - \log(v/C) \right| \leq \varepsilon \log(v/C) \quad (Ce^\psi \leq v \leq D),
\]
\[
\left| \#(A \cap (u, D]) - \log(D/u) \right| \leq \varepsilon \log(D/u) \quad (C \leq u \leq De^{-\psi})
\]
is \(\geq 1 - O\varepsilon(e^{-(1/3)}\varepsilon^2)^\psi\).

**Proof.** The probability in question is at least 1 minus the probability that
\[
(6.6) \quad \left| \#(A \cap (e^k, e^l]) - (l - k) \right| \geq \varepsilon(l - k) - 4
\]
for some $k, l \in \mathbb{N}$ with either $k = \lfloor \log C \rfloor$ and $k + \psi - 1 \leq l \leq \log D + 1$ or $l = \lfloor \log D \rfloor + 1$ and $\log C - 1 \leq k \leq l - \psi + 2$. This comes from the monotonicity of $\#(A \cap (C, v))$ as a function of $v$, and the monotonicity of $\#(A \cap (u, D])$ as a function of $u$ (cf., the proof of Theorem 1.18). By Lemma 6.12 with $\theta = \varepsilon(l - k)^{1/2} + O((l - k)^{-1/2})$, the probability that (6.6) holds is $O(e^{-\frac{1}{3}\varepsilon^2(k-l)})$. Summing on $k, l$ gives the lemma. \qed

To establish the lower bound on $\beta_2$, we first analyze the global distribution of divisor ratios. For $1 \leq C < D$ define the random set
\[
\lambda(C, D) = \left\{ \Sigma A_1 - \Sigma A_2 : A_1 \subseteq A \cap (C, D), A_2 \subseteq A \cap (C, D), A_1 \cap A_2 = \emptyset, A_1 \neq \emptyset, A_2 \neq \emptyset \right\},
\]
where $\Sigma A$ is the sum of the elements of $A$.

**Lemma 6.14.** Fix $\varepsilon$ satisfying
\[
0 < \varepsilon < \frac{1}{200}.
\]
Suppose that $1 \leq C < D$ and
\[
C \leq D^{1 - \frac{1}{120} - \varepsilon}.
\]
For any $10 \leq \xi \leq \log D$, with probability $1 - O\varepsilon(1/\log \xi)$ we have
\[
\#\lambda(C, D) \geq D/\xi \quad \text{and} \quad \Sigma(A \cap [1, D]) \leq \xi D.
\]

**Proof.** We may assume that $\xi \geq e^{20}$, else the lemma is trivial. Firstly, $E\Sigma(A \cap [1, D]) = D$, hence by Markov’s inequality,
\[
(6.7) \quad \mathbb{P}(\Sigma(A \cap [1, D])) > \xi D \leq \frac{1}{\xi}.
\]
We will use a second moment argument. Let $V$ denote the number of quadruples $(A_1, A_2, A_3, A_4)$ with each $A_i$ a nonempty subset of $A'$, $A_1 \cap A_2 = A_3 \cap A_4 = \emptyset$ and
\[
(6.8) \quad \Sigma A_1 - \Sigma A_2 = \Sigma A_3 - \Sigma A_4.
\]
Our main task is to show that with probability $1 - O\varepsilon(1/\log \xi)$ we have
\[
(6.9) \quad V \leq \xi \cdot \frac{3^{2|A'|}}{2D} \quad \text{and} \quad |A'| \geq \frac{\log D}{\log 3}.
\]
Assuming (6.9), let
\[
g_m = \#\{(A_1, A_2) : \emptyset \neq A_1 \subseteq A', \emptyset \neq A_2 \subseteq A', A_1 \cap A_2 = \emptyset, \Sigma A_1 - \Sigma A_2 = m\}.
\]
In this notation, $V = \sum_m g_m^2$ and $\lambda(C, D) = \{m : g_m > 0\}$. By Cauchy’s inequality,
\[
(3^{2|A'|} - 2|A'| + 1)^2 \leq \left( \sum_m g_m^2 \right)^2 \leq \#\lambda(C, D) \sum_m g_m^2 = \#\lambda(C, D) V.
\]
The lemma now follows from (6.9), since $(3^{2|A'|} - 2|A'| + 1)^2 \geq \frac{1}{2}3^{2|A'|}$ for large enough $D$. 

4. Proof of Theorem 6.9
To prove (6.9), we first separate off those solutions of (6.8) with \( A_1 = A_3 \) and \( A_2 = A_4 \). Thus,

\[
V \leq 3^{A'|} + V^* ,
\]

where \( V^* \) counts solutions with \( A_1 \neq A_3 \) or \( A_2 \neq A_4 \). Let \( \psi = \frac{\log \xi}{3} \). Let \( E \) be the event that

\[
|A \cap (B, D)| \geq (1 - \varepsilon) \log(D/B) \quad (C \leq B \leq De^{-\psi}).
\]

This implies that

(6.10) \[
|A \cap (B, D)| \geq (1 - \varepsilon) \log(D/B) - \psi \quad (C \leq B \leq D).
\]

By Lemma 6.13, \( \mathbb{P}(E) \ll e^{-(1/3)\varepsilon^2} \ll_{\varepsilon} 1/ \log \xi \). Assume now that we are in event \( E \). In particular, since \( \log(D/C) \geq \left( \frac{1}{\log 3} + \varepsilon \right) \log D \), we have

\[
3^{A'} \geq 3^{(1-\varepsilon)\log(D/C) - \psi} \geq D
\]

if \( D \) is large enough, giving the second part of (6.9). Hence,

(6.11) \[
V \leq \frac{\xi}{4} \cdot \frac{3^{2|A'|}}{D} + V^*.
\]

Our next task is to show that

(6.12) \[
T := \mathbb{E} \left( \frac{V^* 1(E)}{3^{2|A'|}} \right) \ll \frac{\xi^{1/2}}{D}.
\]

Assuming (6.12), Markov's inequality gives

\[
\mathbb{P}(V^* \geq \frac{\xi}{4} \cdot \frac{3^{2|A'|}}{D} \text{ and } E) \leq \frac{T}{(\xi/4)D^{-1}} \ll \xi^{-1/2},
\]

which implies that (6.9) holds with probability \( 1 - O_{\varepsilon}(1/ \log \xi) \), as required.

It remains to prove (6.12). Assuming \( E \), consider solutions of (6.8) with \( A_1 \neq A_3 \) or \( A_2 \neq A_4 \). Elements in \( A_1 \cap A_3 \) and in \( A_2 \cap A_4 \) cancel each other. Let \( n^* \) be the largest uncancelled element, and write

\[
A' = Q \cup \{n^*\} \cup Q', \quad \max Q < n^* < \min Q'.
\]

In particular, all elements of \( Q' \) cancel out in (6.8), that is,

(6.13) \[
\Sigma A_1 \cap (Q \cup \{n^*\}) - \Sigma A_2 \cap (Q \cup \{n^*\}) = \Sigma A_3 \cap (Q \cup \{n^*\}) - \Sigma A_4 \cap (Q \cup \{n^*\}).
\]

Also, by (6.10), if \( M = \max Q \) then

(6.14) \[
|Q'| = |A \cap (M, D)| - 1 \geq (1 - \varepsilon) \log(D/M) - \psi - 1 =: k(M).
\]

Given \( Q \), the intersections \( A_i \cap Q \) for \( 1 \leq i \leq 4 \), and one of the \( O(1) \) possibilities for which sets \( A_i \) contain \( n^* \) (there are 6 possibilities, namely \( n^* \) may lie in a single set \( A_i \), or in \( A_1 \) and \( A_4 \) or in \( A_2 \) and \( A_3 \)), the element \( n^* \) is uniquely determined by (6.13). We will compute \( T \) as follows:

- Fix \( M \in (C, D) \) and \( Q \subset (C, M) \) with \( M \in Q \); compute \( \mathbb{P}(A \cap (C, M] = Q) \);
- Fix the sets \( A_i \cap Q \), \( i = 1, 2, 3, 4 \);
- Fix one of the possibilities for \( n^* \); compute \( \mathbb{P}(A \cap (M, n^*)] = \{n^*\}) \);
- Fix \( Q' \subseteq \{n^*, D\} \) for which (6.14) holds; compute \( \mathbb{P}(A \cap (n^*, D] = Q') \);
- there are then \( 3^{|Q'|} \) ways to form the sets \( A_i \cap Q' \), \( 1 \leq i \leq 4 \), namely each element of \( Q' \) lies in \( A_1 \cap A_3 \), or in \( A_2 \cap A_4 \) or neither (it cannot be in both since \( A_1, A_2 \) are disjoint).
- By (6.14), we have

\[
\frac{3^{|Q'|}}{32|A'|} = \frac{1}{3^{2+2|Q'|+|Q'|}} \leq \frac{1}{3^{2+k(M)+2|Q'|}}.
\]
Following this process, we have
\[
T \leq \sum_{C < M \leq D} \frac{1}{3^{2k(M)}} \sum_{Q \subseteq (C, M)} \sum_{M \in Q} \left( \prod_{A_1 \cap Q, \ldots, A_4 \cap Q} \sum_{n^*} \mathbb{P}(A \cap (M, n^*) \{n^*\}) \times \prod_{Q' \subseteq (n^*, D)} \mathbb{P}(A \cap (n^*, D) = Q').
\]

The inner sum over \(Q'\) is clearly \(\leq 1\). For each \(n^*\),
\[
\mathbb{P}(A \cap (M, n^*) \{n^*\}) \leq \frac{1}{n^*} \leq \frac{1}{M}.
\]

Also, with \(Q\) fixed there are \(3^{2|Q|}\) ways to choose \(A_1 \cap Q, \ldots, A_4 \cap Q\). Thus,
\[
T \leq \sum_{C < M \leq D} \frac{1}{M \cdot 3^{2k(M)}} \sum_{Q \subseteq (C, M)} \mathbb{P}(A \cap (C, M) = Q).
\]

The inner sum on \(Q\) equals \(\mathbb{P}(M \in A) = 1/M\). Recalling (6.14), we see that
\[
T \ll \varepsilon 3^{\psi} \sum_{C < M \leq D} \frac{1}{M^2(D/M)^{(1-\varepsilon)\log^3}} = 3^{\psi} D^{-(1-\varepsilon)\log^3} \sum_{C < M \leq D} M^{(1-\varepsilon)\log^3 - 2} \ll \frac{3^{\psi}}{D},
\]
as required for (6.12). This completes the proof of the lemma. \(\square\)

**Proof that \(\beta_2 \geq 1 - \frac{1}{\log 3}\).** We use a device from [47] to build up a solution of \(\Sigma A_1 = \Sigma A_2\), starting with Lemma 6.14. Fix \(\varepsilon \in (0, \frac{1}{300})\), let \(N\) be large and let \(\xi = \xi(N) = \log_2 N\), let \(D_0 = N^{1-\varepsilon}\) and \(D_j = D_0(3\xi)^j\) for \(j \geq 1\). Let \(C = D_0^{1-\frac{\varepsilon}{\log 3}}\) and put \(\lambda_j = \lambda(C, D_j)\) for \(j \geq 0\). Let \(E_j\) be the event that \(0 \notin \lambda_j\), and let \(E'_j\) be the event that \(E_j\) holds and also that \(|\lambda_j| \geq D_j / \xi\) and \(\Sigma(A \cap (C, D_j)) \leq \xi D_j\). By Lemma 6.14,
\[
0 \leq \mathbb{P}(E_j) - \mathbb{P}(E'_j) \ll 1 / \log \xi.
\]

Let \(H_j\) be the event that there are integers \(n, n' \in A\) with \(\xi D_j < n < n' \leq 3\xi D_j\) and \(n - n' \in \lambda_j\), and define \(F_j = E'_j \land H_j\). Since \(F_j\) implies that \(0 \in \lambda_{j+1}\) we have
\[
\mathbb{P}(E_{j+1}) \leq \mathbb{P}(E_j) - \mathbb{P}(F_j).
\]

With \(A \cap (C, D_j)\) fixed such that \(E'_j\) holds we have
\[
\mathbb{P}(H_j) \geq \sum_{m \in \lambda_j} \sum_{m > 0} \mathbb{P}(A \cap (\xi D_j, 3\xi D_j) \{n, n + m\})
\]
\[
= \sum_{m \in \lambda_j} \sum_{m > 0} \sum_{\xi D_j < n \leq 2\xi D_j} \frac{1}{n(n + m)} \prod_{\xi D_j < h \leq 3\xi D_j} \left(1 - \frac{1}{h}\right)
\]
\[
\geq \sum_{m \in \lambda_j} \sum_{m > 0} \sum_{\xi D_j < n \leq 2\xi D_j} \frac{1}{n(n + m)}
\]
\[
\gg |\lambda_j| \xi D_j \gg \frac{1}{\xi^2}.
\]

It follows that
\[
\mathbb{P}(F_j) = \mathbb{P}(E'_j \mid F_j) \gg \xi^{-2} \mathbb{P}(E'_j).
\]
Hence, for some constants \( c > 0 \) and \( c' > 0 \), which depend only on \( \varepsilon \),
\[
\mathbb{P}(E_{j+1} \leq \mathbb{P}(E_j) - c\xi^{-2}\mathbb{P}(E_j) = (1 - c/\xi^2)\mathbb{P}(E_j) + \frac{c'}{\log \xi} \frac{e}{\xi^2}.
\]
Iterating this, starting with \( \mathbb{P}(E_0) \leq 1 \), gives
\[
\mathbb{P}(E_j) \leq (1 - c/\xi^2)^j + \frac{c'c}{\xi^2 \log \xi} \sum_{h=0}^{j-1} (1 - c/\xi^2)^h.
\]
Taking \( J = \lceil \xi^3 \rceil \), we conclude that \( \mathbb{P}(E_J) \leq 1/\log \xi \). Since \( D_J < N \), it follows that with probability \( 1 - O(1/\log \xi) \), there are distinct sets \( A_1, A_2 \subseteq A \cap (C, N) \) with \( \Sigma A_1 = \Sigma A_2 \). As \( \varepsilon > 0 \) is arbitrary, this proves that
\[
\beta_2 \geq 1 - \frac{1}{\log 3}.
\]

**Proof of the upper bound** \( \beta_2 \leq 1 - \frac{1}{\log 3} \). Fix \( \varepsilon > 0 \) small, let \( N \) be large and put \( C = N^{1 - \frac{1}{\log 3} + \varepsilon} \).

It suffices to show that with probability \( \to 0 \) as \( N \to \infty \), \( A \cap (C, N) \) has two disjoint, nonempty subsets with equal sums. Let \( E \) be the event that
\[
|A \cap (C, B)| \leq \log(B/C) + \frac{\varepsilon}{2} \log N \quad (C \leq B \leq N).
\]
If \( N \) is large enough, this occurs if we have
\[
|A \cap (C, B)| \leq (1 + \varepsilon/4) \log(B/C) \quad (C \log N \leq B \leq N).
\]
By Lemma 6.13 with \( \psi = \log \log N \), \( \mathbb{P}(E) = o(1) \) as \( N \to \infty \). The probability that there exist distinct nonempty sets \( A_1, A_2 \in A \cap (C, N) \) with \( \Sigma A_1 = \Sigma A_2 \) is at most \( \mathbb{P}(E) + \mathbb{E} S \cdot \mathbb{I}(E) \), where \( S \) is the number of pairs \( A_1, A_2 \) of distinct subsets of \( A \cap (C, N) \) with equal sums. For counting \( S \), WLOG let \( M = \max A_1 > \max A_2 \), \( A_1' = A_1 \setminus \{M\} \) and \( M' = \max(A_1' \cup A_2) \). Given \( A_1' \) and \( A_2 \), \( M \) is unique determined (if it exists; it must also satisfy \( M > M' \)). Also, with \( M' \) fixed, (6.15) implies that
\[
|A \cap (C, M')| \leq \log(M'/C) + \frac{\varepsilon}{2} \log N =: k(M').
\]
Then
\[
\mathbb{E} S \cdot \mathbb{I}(E) \leq 2 \sum_{C < M' \leq N} \sum_{A_1' \cap (C, M') = A'} \sum_{M' \in A'} \mathbb{P}(M \in A \cap (M', N)).
\]
The innermost probability is \( \frac{1}{M} \leq \frac{1}{M'} \), while the number of pairs \( (A_1', A_2) \) equals \( 3^{k(A')} \leq 3^{k(M')} \). Thus,
\[
\mathbb{E} S \cdot \mathbb{I}(E) \leq 2 \sum_{C < M' \leq N} \frac{3^{k(M')}}{M'} \sum_{A_1' \in (C, M')} \mathbb{P}(A \cap (C, M') = A')
\]
\[
\leq 2 \sum_{C < M' \leq N} \frac{3^{k(M')}}{(M')^2}
\]
\[
\ll \frac{3^{(\varepsilon/2) \log N}}{C \log 3} \sum_{C < M' \leq N} (M')^{\log 3 - 2} \ll \varepsilon 3^{(\varepsilon/2) \log N} N^{-\varepsilon \log 3} \ll N^{-\varepsilon/2}.
\]

**5. Exercises**

**Exercise 6.1.** Let \( 2 \leq \ell \leq k \). Show that \( \beta_k \geq \beta_{\ell}^{\left\lceil \frac{\log \ell}{\log k} \right\rceil} \).
CHAPTER 7

Integers with a divisor in a given interval

1. Exact formulas

For \(0 < y < z\), let \(\tau(n; y, z)\) be the number of divisors \(d\) of \(n\) which satisfy \(y < d \leq z\). Define \(H(x, y, z)\) to be the number of positive integers \(n \leq x\) with \(\tau(n; y, z) > 0\), and \(H_r(x, y, z)\), the number of \(n \leq x\) with \(\tau(n; y, z) = r\). By inclusion-exclusion,

\[
H(x, y, z) = \sum_{k \geq 1} (-1)^{k-1} \sum_{y < d_1 < \cdots < d_k \leq z} \left\lfloor \frac{x}{\text{lcm}[d_1, \ldots, d_k]} \right\rfloor,
\]

but this is not useful for estimating \(H(x, y, z)\) unless \(z - y\) is small. With \(y\) and \(z\) fixed, however, this formula implies that the set of positive integers having at least one divisor in \((y, z]\) has an asymptotic density, i.e.

\[
\varepsilon(y, z) := \lim_{x \to \infty} \frac{H(x, y, z)}{x} = \sum_{k \geq 1} (-1)^{k-1} \sum_{y < d_1 < \cdots < d_k \leq z} \frac{1}{\text{lcm}[d_1, \ldots, d_k]}.
\]

In these notes we primarily focus on the case \(y \leq x^{3/4}\). Since \(d|n\) if and only if \((n/d)|n\), if \(n \approx x\) then we expect that

\[
H(x, y, z) \approx H(x, x/z, x/y).
\]

This has been proved in [26], although the details are rather messy due to the fact that many \(n\) are significantly smaller than \(x\), and we actually require short interval versions, that is, estimating \(H(x, y, z) - H(x', y, z)\) from below.

2. Easy bounds when \(z\) is small or large

When \(z - y\) is small compared with \(y\), it is very rare to have more than one divisor in \((y, z]\).

**Theorem 7.1.** For \(2 \leq y + 1 \leq x^{3/4}\) with \(z = (1 + \eta)y \in [y + 1, 2y]\). Then

\[
H(x, y, z) = x \sum_{y < d \leq z} \frac{1}{d} + O(\eta y + \eta^2 x \log(2y)).
\]

Consequently, if \(y \to \infty\), \(z - y \to \infty\), \(z = o(x)\) and \(\eta = o(1/\log(2y))\) as \(x \to \infty\), then

\[
H(x, y, z) \sim \eta x.
\]

**Remarks.** If \(y \leq \sqrt{x}\) and \(z - y = o(y/\log y)\), then the second conclusion follows. Using more sophisticated methods, Tenenbaum [56] showed that \(H(x, y, z) \sim \eta x\) in the range \(z - y \leq y/(\log y)^{\log 4 - 1+\varepsilon}\) for any fixed \(\varepsilon > 0\), and the constant \(\log 4 - 1\) cannot be replaced by a smaller number. See also [41, p. 38–39].

**Proof.** We start with a simple truncated form of inclusion-exclusion, which implies that

\[
\sum_{y < d \leq z} \left\lfloor \frac{x}{d} \right\rfloor - \sum_{y < d_1 < d_2 \leq z} \left\lfloor \frac{x}{d_1, d_2} \right\rfloor \leq H(x, y, z) \leq \sum_{y < d \leq z} \left\lfloor \frac{x}{d} \right\rfloor.
\]

The first sum over \(d\) equals

\[
O(\eta y) + x \sum_{y < d \leq z} \frac{1}{d}.
\]
In the sum over \(d_1, d_2\), let \(m = (d_1, d_2)\), \(t_i = d_i/m\) for \(i = 1, 2\). We also have \(m \leq d_2 - d_1 \leq z - y = \eta y\). Therefore,

\[
\sum_{y < d_1 < d_2 \leq z} \left\lfloor \frac{x}{d_1, d_2} \right\rfloor \leq x \sum_{m \leq \eta y} \frac{1}{m} \sum_{t_1 < t_2 \leq z/m} \frac{1}{t_1 t_2} \\
\leq x \sum_{m \leq \eta y} \frac{1}{m} \left( \log \frac{z}{y} + O \left( \frac{m}{y} \right) \right)^2 \\
\ll x \eta^2 \sum_{m \leq \eta y} \frac{1}{m} \\
\ll x \eta^2 \log(2y).
\]

The first assertion follows.

To achieve the second conclusion, we use

\[
\sum_{y < d \leq z} \frac{1}{d} = \log \frac{z}{y} + O(1/y) = \eta + O(\eta^2 + 1/y).
\]

Under the hypotheses, \(\eta \rightarrow 0\) and \(\log \min(2y, 2x/y) \rightarrow \infty\), and we deduce that \(H(x, y, z) \sim \eta x\). \(\square\)

**Theorem 7.2.** (a) Fix \(c > 0\). For some \(y_0(c)\), if \(y \geq y_0(c)\) and \(y^{1+c} \leq z \leq x\) then \(H(x, y, z) \gg_c x\).

(b) For all \(2 \leq y \leq z \leq x\), we have

\[
H(x, y, z) = x \left( 1 + O \left( \frac{\log y}{\log z} \right) \right).
\]

**Proof.** For (a), we will show more, that there are \(\gg_c x\) integers \(n \leq x\) with exactly one prime divisor in \((y, z]\). Let \(T\) denote the set of primes in \((y, z]\). We may assume without loss of generality that \(c \leq 1\) and that \(z = y^{1+c}\). We have, by Mertens’ estimate (0.5),

\[
H(T) = \log(1 + c) + O(1/\log y) \in \left[ \frac{1}{2} \log(1 + c), 1 \right]
\]

if \(y_0(c)\) is large enough. By Exercise 1.5,

\[
P_x(\omega(n, T) \geq 1) \geq E_x \left[ \omega(n, T) - \left( \omega(n, T) \right)^2 \right] \\
\geq \frac{1}{x} \sum_{y < p \leq z} \left( \frac{x}{p} - \frac{H(T)^2}{2} \right) \\
\geq H(T) - \frac{H(T)^2}{2} - O \left( \frac{\pi(z)}{x} \right) \\
\geq H(T) - \frac{H(T)^2}{2} - O \left( \frac{\pi(z)}{x} \right) \\
\geq \frac{H(T)}{3} \gg_c 1
\]

if \(y_0(c)\) is large enough.

The upper bound in (b) is trivial. For the lower bound, we also consider integers with at least one prime factor in \((y, z]\). By Theorem 1.13,

\[
P_x(\tau(n, y, z) = 0) \leq P_x(\omega(n, T) = 0) \ll e^{-H(T)} \ll \frac{\log y}{\log z}
\]

and (b) follows. \(\square\)
Remarks. In general, the error term $O\left(\frac{\log x}{\log z}\right)$ cannot be improved. For example, note that $\tau(n, y, z) = 0$ if $n = mh$ with $m \leq y$ and $P^-(h) > z$. If $z \leq \sqrt{x}$ then by Exercise 3.1,

$$\#\{n \leq x : \tau(n, y, z) = 0\} \geq \sum_{m \leq y^{1/2}} \Phi\left(\frac{x}{m}\right) \geq \sum_{m \leq y^{1/2}} \frac{x}{m \log z} \asymp \frac{x}{\log y} \log z.$$  

3. The critical case $z = 2y$

Besicovitch [4] showed in 1934 that

$$\liminf_{y \to \infty} \varepsilon(y, 2y) = 0,$$

and used this to construct an infinite set $\mathcal{A}$ of positive integers such that its set of multiples $\mathcal{B}(\mathcal{A}) = \{am : a \in \mathcal{A}, m \geq 1\}$ does not possess asymptotic density. Erdős in 1935 [19] showed $\lim_{y \to \infty} \varepsilon(y, 2y) = 0$ and in 1960 [23] gave the further refinement

$$\varepsilon(y, 2y) = (\log y)^{-E + o(1)} \quad (y \to \infty),$$

where

$$E = 1 - \frac{1 + \log \log 2}{\log 2} = 0.086071 \ldots.$$ 

In 1984, Tenenbaum [56] refined the bounds to

$$\frac{x}{(\log y)^E \exp\{c\sqrt{\log y \log_2 y}\}} \ll H(x, y, 2y) \ll \frac{x}{(\log y)^E (\log_2 y)^{1/2}},$$

valid for $100 \leq y \leq \sqrt{x}$, where $c > 0$ is a constant. Hall and Tenenbaum's book *Divisors* [41, Ch. 2] gives a simpler proof of Tenenbaum's theorem.

**Theorem 7.3 (Ford [26]).** Uniformly for $4 \leq y \leq x^{1/2}$ we have

$$H(x, y, 2y) \asymp \frac{x}{(\log y)^E (\log_2 y)^{3/2}}.$$  

Remarks. Theorem 1 of [26] establishes the order of $H(x, y, z)$ for all $x, y, z$, improving upon cruder estimates of Tenenbaum [56].

Our proof of the lower bound implicit in Theorem 7.3 gives a somewhat stronger conclusion, where we restrict to integers in an interval which are squarefree and free of small prime factors. We will leave the details as an exercise, Exercise ?? below.

**Theorem 7.4.** Fix $c' < 1 < c$, and $w \geq 1$. We have, for $4 \leq y \leq x^{1/2}$,

$$\#\{c' x < n \leq x : \mu^2(n) = 1, \tau(n, y, cy) \geq 1\} \gg_c c' H(x, y, z).$$

Remarks. In Theorem 7.4, we consider $c, c', w$ all fixed. In [28], the order of magnitude of $\#\{n \leq x : P^-(n) > w, \tau(n, y, 2y) \geq 1\}$ was determined for all $x, y, w$. The results change behavior depending on the relative size of $y, w$.

4. Some applications of Theorem 7.3

1. Distinct products in a multiplication table, a problem of Erdős from 1955 ([22], [23]). Let $A(x)$ be the number of positive integers $n \leq x$ which can be written as $n = m_1 m_2$ with each $m_i \leq \sqrt{x}$. Earlier, see Theorem 1.28, we showed that $A(x) \ll x/(\log x)^E$.

**Theorem 7.5.** We have

$$A(x) \asymp \frac{x}{(\log x)^E (\log \log x)^{3/2}}.$$
7. INTEGERS WITH A DIVISOR IN A GIVEN INTERVAL

Proof. Evidently \( A(x) \) is at least the number of \( n \leq x/4 \) with a divisor in \( (\frac{1}{4} \sqrt{x}, \frac{1}{2} \sqrt{x}) \). Also, if \( n = m_1 m_2 \) with \( m_1 \leq \sqrt{x}, m_2 \leq \sqrt{x} \) then for some integer \( k \geq 0 \), \( 2^{-k-1} \sqrt{x} < m_1 \leq 2^{-k} \sqrt{x} \), and hence \( n \leq 2^{-k} x \). Thus,

\[
H \left( \frac{x}{4}, \frac{\sqrt{x}}{4}, \frac{\sqrt{x}}{2} \right) \leq A(x) \leq \sum_{k \geq 0} H \left( \frac{x}{2^k}, \frac{\sqrt{x}}{2^k+1}, \frac{\sqrt{x}}{2^k} \right).
\]

The theorem now follows quickly from Theorem 7.3. For the right side, use Theorem 7.3 when \( k \leq 10 \log_2 x \) and the trivial upper bound \( x/2^k \) when \( k > 10 \log_2 x \). \( \square \)

2. Distribution of Farey gaps (Cobeli, Ford, Zaharescu [6]).

**Theorem 7.6.** Let \( (\frac{q}{Q}, \frac{1}{Q}, \ldots, \frac{Q-1}{Q}, \frac{1}{Q}) \) denote the sequence of Farey fractions of order \( Q \), and let \( N(Q) \) denote the number of distinct gaps between successive terms of the sequence. Then

\[
N(Q) \approx \frac{Q^2}{(\log Q)^{\varepsilon}(\log \log Q)^{3/2}}.
\]

Proof. The distinct gaps are precisely those products \( qq' \) with \( 1 \leq q, q' \leq Q \). \( (q, q') = 1 \) and \( q + q' > Q \). Thus, \( \max(q, q') > Q/2 \), so \( N(Q) \approx H(Q^2, Q/2, Q) \) and the upper bound follows from Theorem 7.3. For the lower bound, consider squarefree \( 0.3Q^2 < n \leq 0.36Q^2 \) with a divisor in \( (0.5Q, 0.6Q] \). The complementary divisor then lies in \( (0.5Q, 0.72Q] \). Hence

\[
N(Q) \geq \frac{1}{2} \# \{0.3Q^2 < n \leq 0.36Q^2 : \mu^2(n) = 1, \tau(n, 0.5Q, 0.6Q) \geq 1\}
\]

and the lower bound follows from Theorem 7.4. \( \square \)

3. Density of unions of residue classes. Given moduli \( m_1, \ldots, m_k \), let \( \delta_0(m_1, \ldots, m_k) \) be the minimum, over all possible residue classes \( a_1 \mod m_1, \ldots, a_k \mod m_k \), of the density of integers which lie in at least one of the classes. By a theorem of Rogers (see [37, p. 242-244]), the minimum is achieved by taking \( a_1 = \cdots = a_k = 0 \) and thus \( \delta_0(m_1, \ldots, m_k) \) is the density of integers possessing a divisor among the numbers \( m_1, \ldots, m_k \). When \( m_1, \ldots, m_k \) consist of the integers in an interval \( (y, z] \), then \( \delta_0(m_1, \ldots, m_k) = \varepsilon(y, z) \).

4. Partial Möbius divisor sums, which was first studied by Erdős and Hall [24]. Define

\[
M(n, y) = \sum_{d \mid n, d \leq y} \mu(d).
\]

**Theorem 7.7 (K. Ford, unpublished).** Let \( 10 \leq y \leq \sqrt{x} \). The number of integers \( n \leq x \) with \( M(n, y) \neq 0 \) is

\[
x \approx \frac{x}{(\log y)^{\varepsilon}(\log_2 y)^{3/2}}.
\]

This requires versions of \( H(x, y, 2y) \) which count integers free of prime factors \( \leq w \), uniformly in \( w \), as well as a version counting integers with exactly one divisor in \( (y, 2y] \). The former is dealt with in [28] and the latter in [26].

Here we argue more crudely and show that

\[
(7.2) \quad \# \{n \leq x : M(n, y) \neq 0\} \ll \frac{x}{(\log y)^{3/2}} \quad (y \leq x^{1/3}).
\]

Let \( w = \exp\{(\log y)^{\varepsilon/2}\} \). By Theorem 3.8,

\[
\# \{n \leq x : P^{-}(n) > w\} = \Phi(x, w) \ll \frac{x}{(\log y)^{3/2}}.
\]

Now consider \( n \leq x \) with \( p = P^{-}(n) \leq w \). If \( p^a|n \) with \( p^a > w^2 \) then for some \( d \in \mathbb{N} \) with \( d > w, d^2|n \). The number of such \( n \) is at most

\[
\ll \sum_{d > w} \frac{x}{d^2} \ll \frac{x}{w} \ll \frac{x}{\log y}.
\]
Now consider \( n \leq x \) where \( p = P^-(n) \leq w, \ p^a \mid n, \ p^a \leq w^2 \), \( n = p^a m \). If \( \tau(n, y, py) = 0 \) then for any squarefree \( d \mid m, \ pd \leq y \) and hence
\[
M(n, y) = \sum_{d \mid m, \ d \leq y} \mu(d) + \mu(pd) = 0.
\]
The number of such \( n \) is at most
\[
\sum_{p \leq w} H \left( \frac{x}{p^n}, y, py \right) \ll \sum_{p \leq w} \sum_{0 \leq k \leq \frac{\log p}{\log 2}} H \left( \frac{x}{p^n}, 2^k y, 2^{k+1} y \right) \ll \sum_{p \leq w} \frac{x \log p}{p^a (\log y)^2} \ll \frac{x}{(\log y)^{2/2}}
\]
by Mertens’ estimate (0.6). This proves (7.2).

5. A heuristic for \( H(x, y, 2y) \)

Write \( n = n'n'' \), where \( n' \) is composed only of primes \( \leq 2y \) and \( n'' \) is composed only of primes \( > 2y \). For simplicity, assume \( n' \) is squarefree and \( n'' \leq y^{100} \). Assume for the moment that the set \( D(n') = \{ \log d : d \mid n' \} \) is uniformly distributed in \([0, \log n']\). If \( n' \) has \( k \) prime factors, then the expected value of \( \tau(n', y, 2y) \) should be about \( \frac{2^k \log^2}{\log n'} \approx \frac{2^k}{\log y} \). This is \( \gg 1 \) precisely when \( k \geq k_0 + O(1) \), where \( k_0 := \left\lfloor \frac{\log \log y}{\log 2} \right\rfloor \). Using the fact that, e.g. Theorem 1.13 for the upper bound, the number of \( n \leq x \) with \( n'' \) having \( k \) prime factors is of order
\[
\frac{x}{\log y} \frac{(\log \log y)^k}{k!},
\]
we obtain a heuristic estimate for \( H(x, y, 2y) \) of order
\[
\frac{x}{\log y} \sum_{k \geq k_0 + O(1)} \frac{(\log \log y)^k}{k!} \gg \frac{x (\log \log y)^{k_0}}{k_0! \log y} \times \frac{x}{(\log y)^{k_0} (\log \log y)^{1/2}}.
\]
This is slightly too big, and the reason stems from the uniformity assumption about \( D(n') \). In fact, for most \( n' \) with about \( k_0 \) prime factors, the set \( D(n') \) is far from uniform, possessing many clusters of divisors and large gaps between clusters. This substantially decreases the likelihood that \( \tau(n', y, 2y) \geq 1 \). If we write \( n' = p_1 \cdots p_k \), where \( p_1 < p_2 < \ldots < p_k \), then we expect \( \log \log p_j \approx \frac{\log \log y}{k_0} = j \log 2 + O(1) \) for each \( j \).

The Central Limit Theorem for prime factors \( \leq 2y \) (e.g. Theorem 5.7) tell us that with high probability there is a \( j \) for which \( \log \log p_j \leq j \log 2 - c\sqrt{\log \log y} \), where \( c \) is a small positive constant. Thus, the \( 2^j \) divisors of \( p_1 \cdots p_j \) will be clustered in an interval of logarithmic length about \( \ll \log p_j \leq 2^j e^{-c\sqrt{\log \log y}} \). On a logarithmic scale, the divisors of \( n' \) will then lie in \( 2^k e^{-c\sqrt{\log \log y}} \) translates of this cluster, the total length of the clusters being \( \ll 2^k e^{-c\sqrt{\log \log y}} \). A measure of the degree of clustering of the divisors of an integer \( a \) is given by
\[
\mathcal{Z}(a) = \bigcup_{d \mid a} [-\log 2 + \log d, \log d] \quad L(a) = \text{meas} \mathcal{Z}(a).
\]
The probability that \( \tau(n', y, 2y) \geq 1 \) should then be about \( L(n')/\log y \). Making this precise leads to the upper and lower bounds for \( H(x, y, 2y) \) given below in Proposition 7.8. The upper bound for \( L(a) \) given in Lemma 7.9 (iii) below quantifies how small \( L(a) \) must be when there is a \( j \) with \( \log \log p_j \) considerably smaller than \( j \log 2 \).

What we really need to count is \( n \) for which \( n' \) has about \( k_0 \) prime factors \( \text{and} \ L(n') \gg \log n' \). This roughly corresponds to asking for \( \log \log p_j \geq j \log 2 - O(1) \) for all \( j \). The analogous problem from statistics theory is to ask for the likelihood than given \( k_0 \) random numbers in \([0, 1]\), there are \( \leq k_0 x + O(1) \) of them
which are \( \leq x \), uniformly in \( 0 \leq x \leq 1 \). Later, we will see that this probability is about \( 1/k_0 \approx 1/\log \log y \) and this leads to the correct order of \( H(x, y, 2y) \) given in Theorem 7.3.

6. A global-to-local principle

In this section, we estimate \( H(x, y, 2y) \) in terms of an average over \( L(a) \), as defined in (7.3). As \( L(a) \) captures the global distribution of divisors of \( a \), we call this a ‘global-to-local’ principle. Introduce the notation

\[
\mathcal{P}(x) = \{n \in \mathbb{N} : \mu^2(n) = 1, P^+(n) \leq x\}.
\]

**Proposition 7.8.** If \( y_0 \) is sufficiently large and \( y_0 \leq y \leq \sqrt{x} \), then

\[
H(x, y, 2y) \sim \frac{x}{\log^2 y} \sum_{a \in \mathcal{P}(y)} \frac{L(a)}{a}.
\]

We first show some basic inequalities for \( L(a) \) and then relate sums of the type on the RHS in Proposition 7.8.

**Lemma 7.9.** We have

(i) \( L(a) \leq \min(\tau(a) \log 2, \log(2a)) \);

(ii) If \( (a, b) = 1 \), then \( L(ab) \leq \tau(b)L(a) \);

(iii) If \( p_1 < \cdots < p_k \), then

\[
L(p_1 \cdots p_k) \leq \min_{0 \leq j \leq k} 2^{k-j}(\log(2p_1 \cdots p_j)).
\]

**Proof.** Part (i) is immediate, since \( \mathcal{L}(a) \) is the union of \( \tau(a) \) intervals of length \( \log 2 \), all contained in \([-\log 2, \log a]\). Part (ii) follows from

\[
\mathcal{L}(ab) = \bigcup_{d|b} \{u + \log d : u \in \mathcal{L}(a)\}.
\]

Combining parts (i) and (ii) with \( a = p_1 \cdots p_j \) and \( b = p_{j+1} \cdots p_k \) yields (iii). \( \square \)

**Lemma 7.10.** Let \( w_2 \geq w_1 \geq 2 \). Then

\[
\sum_{a \in \mathcal{P}(w_2)} \frac{L(a)}{a} \ll \left(\frac{\log w_2}{\log w_1}\right)^2 \sum_{a \in \mathcal{P}(w_1)} \frac{L(a)}{a}.
\]

**Proof.** Given \( a \in \mathcal{L}(w_2) \), write \( a \) uniquely as \( a = a_1 a_2 \) where \( P^+(a_1) \leq w_1 < P^-(a_2) \). By Lemma 7.9 (ii), \( L(a) \leq \tau(a_2)L(a_1) \). Thus,

\[
\sum_{a \in \mathcal{P}(w_2)} \frac{L(a)}{a} = \sum_{a_1 \in \mathcal{P}(w_1)} \frac{L(a_1)}{a_1} \sum_{p_2|a_2 \Rightarrow w_1 < p_2 \leq w_2} \frac{\tau(a_2)\mu^2(a_2)}{a_2}.
\]

By Mertens’ product estimate (0.7), the sum on \( a_2 \) equals

\[
\prod_{w_1 < p \leq w_2} \left(1 + \frac{2}{p}\right) \leq \prod_{w_1 < p \leq w_2} (1 - 1/p)^{-2} \ll \left(\frac{\log w_2}{\log w_1}\right)^2.
\]

For the upper bound we also need the following technical lemma, which is a special case of a result of Kououlouopoulos [45, Lemma 2.2].

**Lemma 7.11.** We have

\[
\sum_{a \in \mathcal{P}(2y)} \frac{L(a)}{a} \log^2(P^+(a) + y^{2/3}/a) \ll \frac{1}{(\log y)^2} \sum_{a \in \mathcal{P}(2y)} \frac{L(a)}{a}.
\]
6. A global-to-local principle

Proof. Let \( \mathcal{P}_1 = \{ a \in \mathcal{P}(y^{1/4}) : a > y^{1/2} \} \). Then clearly

\[
\sum_{a \in \mathcal{P}(2y)} \frac{L(a)}{a \log^2(P(a) + y^{2/3}/a)} \ll \sum_{a \in \mathcal{P}_1} \frac{L(a)}{a \log^2 P(a)} + \frac{1}{(\log y)^2} \sum_{a \in \mathcal{P}(2y)} \frac{L(a)}{a}.
\]

For \( a \in \mathcal{P}_1 \), let \( p = P(a) \) and \( a = pb \), so \( b > y^{1/4} \). By Lemma 7.9 (ii), \( L(ab) \leq 2L(b) \) and thus

\[
\sum_{a \in \mathcal{P}_1} \frac{L(a)}{a \log^2 P(a)} \leq 2 \sum_{p \leq y^{1/4}} \frac{1}{p \log^2 p} \sum_{b \in \mathcal{P}(p)} \frac{L(b)}{b} \leq 2 \sum_{p \leq y^{1/4}} \frac{1}{p \log^2 p} \log y \sum_{b \in \mathcal{P}(p)} \frac{L(b) \log^3 b}{b}.
\]

Next,

\[
\sum_{b \in \mathcal{P}(p)} \frac{L(b) \log^3 b}{b} = \sum_{b \in \mathcal{P}(p)} \frac{L(b)}{b} \sum_{p_1 \mid b, p_2 \mid b, p_3 \mid b} (\log p_1)(\log p_2)(\log p_3) \leq 8 \sum_{p_1, p_2, p_3 \leq p} (\log p_1)(\log p_2)(\log p_3) \sum_{t \in \mathcal{P}(p)} \frac{L(t)}{t},
\]

where we have written \( b = [p_1, p_2, p_3]t \) and used Lemma 7.9 (ii) again. Considering separately the three cases \( p_1 = p_2 = p_3 \), two of the \( p_i \) equal, all \( p_i \) distinct), we find that

\[
\sum_{p_1, p_2, p_3 \leq p} (\log p_1)(\log p_2)(\log p_3) \ll (\log p)^3
\]

by Mertens’ estimate. Extending the range of \( t \) to \( t \in \mathcal{P}(2y) \), we get

\[
\sum_{a \in \mathcal{P}_1} \frac{L(a)}{a \log^2 (P(a) + y^{2/3}/a)} \ll \sum_{p \leq y^{1/4}} \frac{1}{p \log^2 p} \frac{(\log p)^3}{(\log y)^3} \sum_{t \in \mathcal{P}(2y)} \frac{L(t)}{t}.
\]

A final application of Mertens’ estimate concludes the proof. \( \square \)

We are now ready to embark on the proof of Proposition 7.8.

We begin with the lower bound, which is easier. Consider integers \( n = ap_1p_2p_3 \leq x \) with \( p_1 \) and \( p_2 \) prime,

\[
a \leq y^{1/5} < p_1 < p_2 < \frac{1}{4} y^{4/5} < P^{-}(b),
\]

and with \( \log(y/p_1p_2) \in \mathcal{L}(a) \). The last condition implies that there is a divisor \( d \mid a \) with

\[
\frac{d}{2} < \frac{y}{p_1p_2} < \frac{d}{2},
\]

which is equivalent to \( y < dp_1p_2 \leq 2y \). Thus, for such \( n \), \( \tau(n, y, 2y) \geq \tau(ap_1p_2, y, 2y) \geq 1 \). In particular, \( y^{4/5} \leq y/a < p_1p_2 \leq 2y \), so that \( x/ap_1p_2 \geq x/(2y^{4/5}) \geq \frac{1}{2} y^{4/5} \). Thus, by Exercise 3.1, for each triple \( a, p_1, p_2 \), the number of possible \( b \) is

\[
\geq \Phi \left( \frac{x}{ap_1p_2} ; \frac{1}{4} y^{4/5} \right) \gg \frac{x}{ap_1p_2 \log y}.
\]

Now \( \mathcal{L}(a) \) is the disjoint union of intervals of length \( \geq \log 2 \), all contained in \( [-\log 2, \frac{1}{2} \log y] \). For each such interval \( [u, v] \) we have by Mertens’ estimate (0.5)

\[
\sum_{u \leq \log(y/p_1p_2) \leq v} \frac{1}{p_1p_2} \gg \sum_{8y^{1/3} < p_1 \leq y^{2/3}} \frac{1}{p_1} \sum_{p_2 \leq \log y} \frac{1}{p_2} \gg \frac{v - u}{\log y}.
\]
Thus, with $a$ fixed, we have
\[ \frac{1}{p_1 p_2} \sum_{\log(y/p_1 p_2) \in \mathcal{L}(a)} L(a) \log y \gtrsim \ \frac{L(a)}{\log y}. \]

Hence,
\[ H(x, y, 2y) \gtrsim \frac{x^2}{\log y} \sum_{x, y \in \mathcal{P}(1/5)} L(a) \frac{a}{a}. \]

Next, we replace the sum over a more convenient set, starting with
\[ \sum_{a \leq y^{1/5}} L(a) \frac{a}{a} \gtrsim \sum_{a \in \mathcal{P}(y^{1/15})} \frac{L(a)}{a} \sum_{p|a} \log p. \]

Next,
\[ \sum_{a \in \mathcal{P}(y^{1/15})} \frac{L(a) \log a}{a} = \sum_{a \in \mathcal{P}(y^{1/15})} \frac{L(a)}{a} \sum_{p|a} \log p \]
\[ = \sum_{p \leq y^{1/15}} \frac{\log p}{p} \sum_{b \in \mathcal{P}(y^{1/15})} \frac{L(pb)}{b} \]
\[ \leq \sum_{p \leq y^{1/15}} \frac{2 \log p}{p} \sum_{b \in \mathcal{P}(y^{1/15})} \frac{L(b)}{b} \]
\[ = (2 \log(y^{1/15}) + O(1)) \sum_{b \in \mathcal{P}(y^{1/15})} \frac{L(b)}{b} \]

using the relation $L(pb) \leq 2L(b)$ from Lemma 7.9 (ii) and Mertens’ estimate (0.6). Thus,
\[ \sum_{a \leq y^{1/5}} \frac{L(a)}{a} \gtrsim \sum_{a \in \mathcal{P}(y^{1/15})} \frac{L(a)}{a} \left(1 - \frac{2 \log(y^{1/15}) + O(1)}{\log(y^{1/5})}\right) \gtrsim \sum_{a \in \mathcal{P}(y^{1/15})} \frac{L(a)}{a}. \]

Applying Lemma 7.10 with $w_1 = y^{1/15}$ and $w_2 = y$ concludes the proof of the lower bound in Proposition 7.8.

We now prove the upper bound in Proposition 7.8. We will first show that
\[ H(x, y, 2y) \ll x \sum_{a \in \mathcal{P}(2y)} \frac{L(a)}{a \log^2(y^{2/3}/a + P^+(a))}. \]

We may assume that $y$ is sufficiently large (say $y \geq y_0$), as (7.4) is trivial for small $y$. First, we relate $H(x, y, 2y)$ to $H^*(x, y, z)$, the number of squarefree integers $n \leq x$ with $\tau(n, y, z) \geq 1$. Let $\delta$ be a fixed, small constant. Write $n = n'n''$, where $n'$ is squarefree, $n''$ is squarefull and $(n', n'') = 1$. The number of $n \leq x$ with $n'' > y^\delta$ is
\[ \leq x \sum_{n'' > y^\delta} \frac{1}{n''} \ll \frac{x}{y^{\delta/2}}, \]

since the number of squarefull integers below $w$ is $O(\sqrt{w})$. If $n'' \leq y^\delta$, then for some $f|n''$, $n'$ has a divisor in $(y/f, 2y/f)$, hence
\[ H(x, y, 2y) \leq \sum_{n'' \leq y^\delta} \sum_{f|n''} H^* \left( \frac{x}{n''}, \frac{2y}{f}, \frac{y}{f} \right) + O \left( \frac{x}{y^{\delta/2}} \right). \]
For any \( n'' \) and \( f, y/f \geq y^{1-\delta} \). Next, we show that
\[
H^*(x_1, y_1, 2y_1) - H^*(\frac{1}{2}x_1, y_1, 2y_1) \ll x_1 \sum_{a \in \mathbb{P}(2y_1)} \frac{L(a)}{a \log^2 \left( \frac{y_1^{3/4}}{a + P^+(a)} \right)} \quad (y^{1-\delta} \leq y_1 \leq x_1^{\frac{1}{2}+\delta}).
\]
Consider squarefree \( n \in (\frac{1}{2}x_1, x_1] \) with a divisor in \((y_1, 2y_1]\). Put \( z_1 = 2y_1, y_2 = \frac{x_1}{y_1}, z_2 = \frac{x_1}{y_2} \). Then \( n = m_1m_2 \) with \( y_1 < m_1 \leq z_1 \) (\( i = 1, 2 \)). For some \( j \in \{1, 2\} \) we have \( p = P^+(m_j) < P^+(m_{3-j}) \). Write
\[
n = ap, \quad P^+(a) < p < P^-(b).
\]
Then \( b > p \) and this is crucial to our argument. Since \( \tau(ap, y_j, z_j) \geq 1 \) and
\[
y_2 \geq \frac{1}{4}x_1^{1-\delta} \geq \frac{1}{4} \frac{1/2 + \delta}{y_1^{1/2}} \geq y_1^{3/4}
\]
if \( \delta \) is small enough, we have
\[
p \geq y_j/a \geq y_1^{3/4}/a.
\]
Thus, \( p \geq Q(a) := \max(P^+(a), y_1^{3/4}/a) \). We also have \( p \leq \min(z_1, z_2) \leq 2y_1 \). As noted earlier, \( b > p \) and so \( ap \leq x_1/p \). Hence, by \((3.8)\), given \( a \) and \( p \), the number of possible \( b \) is
\[
\leq \Phi \left( \frac{x_1}{ap} \right) \leq \frac{x_1}{ap \log p} \leq \frac{x_1}{ap \log Q(a)}.
\]
Since \( a \) has a divisor in \((y_j/p, z_j/p]\), we have \( \log(y_j/p) / \log(y_1) \) or \( \log(2y_j/p) / \log(y_1) \) (the latter case is needed if \( j = 2 \) since \( z_2 = 4y_2 \)). Since \( \mathcal{L}(a) \) is the disjoint union of intervals of length \( \geq \log 2 \) with total measure \( L(a) \), by repeated use of Mertens’ sum estimate \((0.5)\), we obtain
\[
\sum_{\log(cy_j/p) / \log(y_1)} \frac{1}{p} \ll \frac{L(a)}{\log Q(a)} \quad (c = 1, 2),
\]
and \((6.7)\) follows upon summing over all possible \( a \) and over \( j = 1, 2 \).

Write \( x_2 = x/n'', \ y_1 = y/f \). Each \( n \in \{x_2/y_1^2, x_2\} \) lies in an interval \((2^{-r-1}x_2, 2^{-r}x_2] \) for some integer \( 0 \leq r \leq \frac{\delta \log y_1}{\log 2} \). We note that \( y_1 \geq y_1^{1-\delta} \) since \( f \leq n'' \leq y_1 \) and also
\[
x_1 = 2^{-r}x_2 \geq x_2y_1^{-\delta} \geq xy^{-2\delta} \geq y^{2-2\delta} \geq y_1^{2-2\delta},
\]
which implies \( y_1 \leq x_1^{1/1-2\delta} \leq x_1^{\delta/\delta} \). Applying \((6.7)\) with \( x_1 = 2^{-r}x_2 \) and summing over \( r \) we find that
\[
H^*(x_2, y_1, 2y_1) \ll \frac{x_2}{y_1^2} + x_2 \sum_{a \in \mathbb{P}(2y_1)} \frac{L(a)}{a \log^2 \left( \frac{y_1^{3/4}}{a + P^+(a)} \right)}.
\]
The first term \( x_2/y_1^2 \) may be ignored because \( L(1) = \log 2 \) and thus the term \( a = 1 \) is \( \gg 1/\log^2 y_1 \). Thus, by \((7.5)\),
\[
H(x, y, 2y) \ll \frac{x}{y^{\delta/2}} + x \sum_{n'' \leq y^3} \frac{1}{n''} \sum_{f(\gcd(2y, f))} \sum_{a \in \mathbb{P}(2y)} \frac{L(a)}{a \log^2 \left( (y/f)^{3/4} + P^+(a) \right)}.
\]
Again, the term \( x/y^{\delta/2} \) is negligible and may be omitted. We have \((y/f)^{3/4} \geq (y^{1-\delta})^{3/4} \geq y^{2/3} \) for any pair \((n'', f)\) and
\[
\sum_{n''} \frac{\tau(n'')}{n''} = \prod_p \left( 1 + \frac{3}{p^2} + \frac{4}{p^3} + \cdots \right) \ll 1.
\]
This completes the proof of \((7.4)\).
Combining \((7.4)\) with Lemma 7.11, we find that
\[
H(x, y, 2y) \ll \frac{1}{2 \log^2 x} \sum_{a \in \mathbb{P}(2y)} \frac{L(a)}{a}.
\]
Finally, applying Lemma 7.10 with \( w_1 = y \) and \( w_2 = 2y \) completes the proof of the upper bound in Proposition 7.8.
Bibliography


