WEAK AND STRONG AVERAGES IN PROBABILITY AND
THE THEORY OF NUMBERS

A thesis presented

by

Persi Diaconis

to

The Department of Statistics
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Statistics

Harvard University
Cambridge, Massachusetts
May, 1974
ACKNOWLEDGMENT

For many helpful discussions connected with this study, I thank my advisor, Professor Dennis Hejhal. His careful reading of preliminary drafts is particularly appreciated.

Thanks are due to Professor Stephen Portnoy for several clarifying discussions as well as for being the second reader.

Professor Frederick Mosteller provided needed encouragement throughout this work and Professor Andrew Gleason spent many hours helping me get started.

This research was facilitated by a grant from the National Science Foundation (GS-32327X1). I want to thank Mrs. Holly Grano for cheerfully translating my longhand into a typed manuscript.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NATURAL DENSITY</td>
</tr>
<tr>
<td>1.1 Definitions, Examples, Basic Properties</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Classification of Gap Sets</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>ZETA DENSITY</td>
</tr>
<tr>
<td>2.1 Introduction, Definitions, and Examples</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Properties</td>
<td>10</td>
</tr>
<tr>
<td>2.3 Classification of Gap Sets</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>BINOMIAL DENSITY</td>
</tr>
<tr>
<td>3.1 Definitions</td>
<td>15</td>
</tr>
<tr>
<td>3.2 Examples</td>
<td>16</td>
</tr>
<tr>
<td>3.3 Properties</td>
<td>24</td>
</tr>
<tr>
<td>3.4 A Tauberian Theorem for Binomial Density</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>APPLICATIONS OF TAUBERIAN THEOREMS</td>
</tr>
<tr>
<td>4.1 The Classical Theory and Some Extensions</td>
<td>33</td>
</tr>
<tr>
<td>4.2 Applications</td>
<td>36</td>
</tr>
<tr>
<td>4.2.1 Abel Density</td>
<td>36</td>
</tr>
<tr>
<td>4.2.2 Zeta, Log and the Logarithmic Series Density</td>
<td>37</td>
</tr>
<tr>
<td>4.2.3 Higher Log and Zeta Densities</td>
<td>40</td>
</tr>
<tr>
<td>4.3 Some Densities Implicit in the Theory of Numbers</td>
<td>41</td>
</tr>
<tr>
<td>4.4 Remarks on Differentiating and Integrating in a Parameter</td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td>INFINITE REAVERAGING</td>
</tr>
<tr>
<td>5.1 $H^\infty$ Averages</td>
<td>56</td>
</tr>
<tr>
<td>5.2 Log Density</td>
<td>72</td>
</tr>
<tr>
<td>6</td>
<td>SUMMABILITY AND RANDOM VARIABLES</td>
</tr>
<tr>
<td>6.1 Introduction</td>
<td>80</td>
</tr>
<tr>
<td>6.2 Weighted Weak Laws</td>
<td>84</td>
</tr>
<tr>
<td>6.3 The Strong Law of Large Numbers</td>
<td>99</td>
</tr>
<tr>
<td>6.4 The Central Limit Problem</td>
<td>102</td>
</tr>
</tbody>
</table>

Bibliography | 115 |
INTRODUCTION

The concept of average value is basic in scientific inquiry since individual values of a quantity under study frequently vary without ever settling down. In the work to follow, various averages are critically compared on a large class of problems where several kinds of useful averages have been established.

Given a collection of real numbers \( \{a(i)\}_{i=1}^\infty \), consider the numbers \( \delta(n) = \frac{1}{n} \sum_{i=1}^{n} a(i) \). For large \( n \), it may happen that \( \delta(n) \) tends to a definite limit \( \ell \) in which case we say the numbers \( a(i) \) have natural density \( \ell \). For example, if \( a(i) \) is one or zero as \( i \) is even or not, then the numbers \( a(i) \) do not have a limit as \( i \) goes to infinity, but the corresponding \( \delta(n) \) tend to \( \frac{1}{2} \) since "half of all numbers are even". The notion of density offers one way to represent the order in the sequence of zeros and ones by assigning a single number to the set of values \( \{a(i)\} \) which is somehow "typical" of the set.

We use the single term density to combine two ideas. When the \( a(i) \) are the indicator function of some subset \( A \)
contained in the integers; \( a(i) = 1 \) or 0 as \( i \) is in \( A \) or not, we speak of the density of the set \( A \) and think of

\[
\delta(n) = \frac{1}{n} \sum_{i=1}^{n} a(i)
\]

as the proportion of elements of the set \( A \) less than or equal to \( n \). When \( a(i) \) is an arbitrary sequence of real numbers, the \( \delta(n) \) have the interpretation of average values of the numbers \( \{a(i)\}_{i=1}^{n} \) as an ordered sequence. Since the mathematical arguments used in the two cases are identical, we treat them simultaneously using the term natural density in either case. In the first five chapters, almost all examples are subsets of the integers and to the first interpretation, while in the last chapter the examples concern the general averaging interpretation. The theorems proved throughout apply to both interpretations.

Natural density is so widely accepted as an average that it is valid to ask: Why consider ranges of alternative averages? The following three answers form the central theme of this dissertation.

1. Natural density can easily fail to exist in many real-world problems.
2. Other averages may converge to the same answer more rapidly.

3. The behavior of an alternative average may be the natural problem to consider.

In the first situation, a different averaging process frequently can be found which agrees with natural density whenever natural density exists but extends natural density by being well defined for the problem of interest. Such a density is called here stronger than natural density.

In the second situation, an average may be found which exists for a wide range of problems and, when it exists, converges to the same answer as natural density at a faster rate. Averages with this property often converge for a smaller class of problems than natural density and are here called weaker than natural density.

The third situation occurs when a quantity being studied can be interpreted as an alternative average of other quantities or when an average value is required which does not weight all components equally.

There are other reasons to consider general averages, but the flavor of the subject is captured by the three reasons above, and we proceed to some examples.

Example 1. The distribution of leading digits. Phrased colloquially, the problem is: pick a number at random; what is the probability its left-most digit is a one. A way to
operationalize 'pick a number at random' is to take a newspaper and make a list of all numbers that appear on the front page, then select among these uniformly. Realizing that left-most digits can be one through nine, some people think that for much the same reason, "half of all numbers are even", \( \frac{1}{3} \) of all numbers begin with one. A check of empirical data soon reveals that far more digits begin with low digits than high digits and that about \( \frac{3}{10} \) begin with one.

To put this problem in a mathematical setting, regard all numbers as integers by omitting sign and decimal point. The set of integers with first digit one is the set 
\[ 0 = \{1, 10, 11, \ldots, 19, 100, 101, \ldots, 199, \ldots\} \] . It is not hard to see that this set does not have natural density: letting \( a(i) \) be one or zero as \( i \) is in \( 0 \) or not, the ratios \( \delta(n) = \frac{1}{n} \sum_{i=1}^{n} a(i) \) alternately increase through a run of numbers which begin with one, then decrease as \( n \) ranges through numbers not in \( 0 \). In fact, all points in the interval \( \left[ \frac{1}{9}, \frac{5}{9} \right] \) occur as limit points of the sequence \( \delta(n) \). To see a real-world meaning in the non-existence of this limit, consider the computer generation of pseudo-random numbers.

Most computer random number generators, when generating random integers, try to achieve a uniform distribution over the set \( \{1, 2, \ldots, N\} \) for some large, fixed \( N \). The choice of \( N \) frequently depends on the underlying hardware. For machines that operate base ten, \( 10^8 \) is frequently used as \( N \). For \( N \) any
<table>
<thead>
<tr>
<th>Set</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>493</td>
<td>136</td>
<td>55</td>
<td>52</td>
<td>47</td>
<td>56</td>
<td>55</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>516</td>
<td>137</td>
<td>51</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>40</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>533</td>
<td>121</td>
<td>50</td>
<td>56</td>
<td>44</td>
<td>50</td>
<td>52</td>
<td>48</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>515</td>
<td>124</td>
<td>54</td>
<td>57</td>
<td>47</td>
<td>44</td>
<td>44</td>
<td>57</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>519</td>
<td>119</td>
<td>53</td>
<td>54</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>52</td>
<td>64</td>
</tr>
<tr>
<td>6</td>
<td>545</td>
<td>102</td>
<td>48</td>
<td>59</td>
<td>48</td>
<td>49</td>
<td>41</td>
<td>42</td>
<td>54</td>
</tr>
<tr>
<td>7</td>
<td>545</td>
<td>102</td>
<td>48</td>
<td>59</td>
<td>48</td>
<td>49</td>
<td>41</td>
<td>42</td>
<td>54</td>
</tr>
<tr>
<td>8</td>
<td>545</td>
<td>102</td>
<td>48</td>
<td>59</td>
<td>48</td>
<td>49</td>
<td>41</td>
<td>42</td>
<td>54</td>
</tr>
<tr>
<td>9</td>
<td>523</td>
<td>104</td>
<td>45</td>
<td>58</td>
<td>52</td>
<td>58</td>
<td>50</td>
<td>48</td>
<td>53</td>
</tr>
<tr>
<td>10</td>
<td>529</td>
<td>120</td>
<td>56</td>
<td>62</td>
<td>51</td>
<td>57</td>
<td>45</td>
<td>55</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Observed Frequencies of First Non-zero Digits in 10 Sets of 1000 Pseudo-Random Integers Produced by RAN
power of 10, an easy calculation shows that all leading digits one through nine will appear with common frequency $\frac{1}{9}$. To see that this answer varies from computer to computer, consider the data in Table 1. This shows the frequency of left-most digits one through nine in ten runs of one thousand trials from the in-house Fortran random number generator. In all cases the frequencies vary widely from uniform. The problem is that the modulus $N$ varies from computer to computer and since the set $\Theta$ doesn't have natural density, different values of $N$ lead to different first-digit frequencies. For fixed $N$, the problem of determining the appropriate frequencies is finite and easily solved. For a general formula, see Diaconis [1973] (pg. 35). When this formula was used in reverse on the data of Table 1, it suggested $N = 2^{31}$ as the only plausible modulus and this was later confirmed by examination of the underlying program.

Despite the non-existence of natural density, the right answer is hidden in the set $\Theta$. To get this answer, consider the zeta density of the set $\Theta$: if $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ denotes Riemann's zeta function, consider

$$\frac{1}{\zeta(s)} \sum_{i=1}^{\infty} \frac{a(i)}{i^s} = Z(s).$$

If $Z(s) \to \zeta$ as $s \to 1$, the set $\Theta$, which determines the $a(i)$, is said to have zeta density $\zeta$. It is shown in Chapter 2 that if a set $\Theta$ has natural density, then the limit of $Z(s)$
exists and the two limits are equal. In the case at hand, the zeta limit exists and has value $\log_{10} 2 \approx .301$. The answer is in agreement with huge amounts of empirical data collected by many writers on this problem.

Other averages, considered later, which are shown to exist for this problem are $L'$, $\log$, higher $\log$, and $H^\infty$. All yield the same limiting value. Further references to the problem may be found in Knuth [1969], (pg. 219-229) and Diaconis [1973] (pg. 31-37).

In summary, we have shown an example of a real problem where the non-existence of natural density has externally verifiable consequences and for which the introduction of stronger averages leads to a useful, verifiable answer.

**Example 2. Error bounds and rates of convergence.** It is important to realize that statements concerning average values need say very little about the individual terms being averaged. For example, the statement that a given subset $A$ of the integers has natural density $\frac{1}{2}$ cannot be used to give a bound $N$ such that some element of $A$ will be known to be less than $N$. This problem actually arises in practice, as in Siegel's theorem on the class number (Ayoub [1963], pg. 327), and can be a major source of difficulty.

Statements about natural density become more useful if accompanied by an error term. For example, if $a(i)$ is one or zero as $i$ is even or odd, and $\delta(n) = \frac{1}{n} \sum_{i=1}^{n} a(i)$, the easily
verified statement $\frac{1}{2} \leq \delta(n) \leq \frac{1}{2} + \frac{1}{n}$ says a great deal about
the numbers $a(i)$.

A weaker average than natural density is binomial density, defined as the limit, if it exists, of $B(n) = \frac{1}{2^n} \sum_{i=0}^{n} a(i)(\binom{n}{i})$ as $n \to \infty$. We show in Chapter 3 that if binomial density exists, then so does natural density and the two limits are equal. Thus, proving theorems for binomial density automatically gives results for natural and all stronger densities.

For $a(i)$ the indicator function of the even numbers, we prove $-\frac{c_1}{2^n} \leq B(n) - \frac{1}{2} \leq \frac{c_2}{2^n}$ for $c_1$, $c_2$ explicit constants, so that the binomial average converges to $\frac{1}{2}$ much more rapidly than does the standard average.

As an example of the third theme mentioned above, notice that the binomial average has a direct interpretation as the probability that the total number of heads, in $n$ tosses of a fair coin, will be an even number. It frequently happens that an averaging method has a direct interpretation and then, there may flow a useful interaction between the averaging and real-world views.

A useful way of comparing different averages is by their performance on a standard set of examples. A class of examples carried through this dissertation deals with $A = \bigcup_{k=0}^{\infty} [f(k), g(k)]$ where $f$ and $g$ are any functions. The first digit set $\Theta = \bigcup_{k=0}^{\infty} [10^k, 210^k]$ is an example. We show that any set with
polynomial gaps has natural density (f and g polynomial) but sets with more extreme gap behavior such as exponential gaps (log f, log g polynomials), do not have natural density. For exponential gaps, most of the stronger averages considered exist and give the same answer. In all cases, the answer depends only on the first two coefficients, and the degree of the polynomials involved. An interesting exception to this rule of thumb is seen in the case of infinite iteration of natural density: reaveraging the sequence of partial averages again and again hoping to induce convergence. It is shown in Chapter 5, that such iteration increases the range of the average to include sets with linear exponential gaps, but no further.

Probability theory is another area where averages are extensively used and a large section of this dissertation is devoted to comparing different methods of assigning an average value to random variables $\{X(i)\}_{i=1}^{\infty}$. The problems have quite a different flavor from those above and for this reason Chapter 6 contains a long introductory section.

Following this introduction is a page of notation and then a chapter by chapter summary of results. Most of the individual chapters begin with introductory remarks and examples which may be a useful supplement to this brief overview.
Some Notation

We write \(\mathbb{N}\) for the natural numbers \(\{1, 2, 3, \ldots\}\) and \(\mathbb{R}\) for the real numbers. Frequently, \(a, b, c, \ldots\), will appear as unspecified constants which are not the same from proof to proof. The following order symbols will be used:

(A) \(f(n) = O(g(n))\) or \(f(n) \ll g(n)\) if \(|f(n)| < k|g(n)|\)

for \(n \to \infty\) and some constant \(k\).

(B) \(f(n) = o(g(n))\) if \(|\frac{f(n)}{g(n)}| \to 0\) as \(n \to \infty\).

(C) \(f(n) \sim g(n)\) if \(\frac{f(n)}{g(n)} \to 1\) as \(n \to \infty\).

These order relations will occasionally be quantified to show their dependence on a parameter. For example, \(f(n) = O_k(g(n))\)

means that the implied constant bounding \(|f(n)| / |g(n)|\) is a function of \(k\). One reference to such order relations is Hardy and Wright [1960] (pg. 7 and 8).

If \(\{a(i)\}_{i=1}^{\infty}\) is a sequence of numbers and \(W = \{W(n,k)\}\) is a (possibly infinite) array of weights, let \(W(n) = \sum_{i=1}^{\infty} a(i)w(n,i)\).

Then \(a(i)\) will be said to have \(W\)-limit \(\lambda\) if \(W(n) \to \lambda\) as \(n \to \infty\).

A reference to this usage is Hardy [1949], Chapter 3.

Survey of Results

The first chapter defines natural density of numbers \(a(i)\) as the limit of \(\frac{1}{n} \sum_{i=1}^{n} a(i)\) if it exists. Examples and basic properties of this fundamental notion are documented and then two
classification theorems regarding the existence of natural density for gap sets \( A = \bigcup_{k=0}^{\infty} [f(k), g(k)] \) are derived. Briefly, if \( f \) and \( g \) are polynomials, then \( A \) has natural density while if \( \log f, \log g \) are polynomials, \( A \) cannot have natural density.

This classification is continued in the second chapter which deals with zeta density, the limit of \( \frac{1}{\zeta(s)} \sum_{i=1}^{\infty} \frac{a(i)}{i^s} \). Zeta density is stronger than natural density and exists for sets \( A \) with exponential growth but not for sets with more rapid growth. The first digit set described above can be written as \( \Theta = \bigcup_{k=1}^{\infty} [10^k, 210^k] \) and, as a set with exponential growth, is shown to have zeta density \( \log_{10} 2 \).

Chapter 3 deals with the binomial average \( \frac{1}{2^n} \sum_{i=0}^{n} a(i) \binom{n}{i} \) as \( n \to \infty \). A major example is the set of square-free numbers \( Q \). A number is square free if it has no squared prime divisor. We show the set \( Q \) has binomial density \( \frac{6}{\pi^2} \). This is given two separate proofs. The first proof uses a sieve argument while the second proof uses a new Tauberian theorem for binomial density which allows results for natural density with sufficiently good error term to be translated back to binomial density.

The fourth chapter introduces powerful Tauberian-Abelian theorems. The classical theory is reviewed and extended—the density Tauberian theorems being proved in extended generality. These tools are brought to bear on the averages considered in
previous chapters as well as on many new averaging procedures which are defined and illustrated throughout the chapter. In particular, zeta density is shown completely equivalent to log density. Certain theorems in the theory of numbers are shown to have density interpretations, an example being much of the work on Dirichlet's theorem on primes in arithmetic progressions. The variety and problems associated with weak and strong averages are inherited by conditional averages restricted to the primes. Much of the classification of gap sets is carried over to the primes. In particular, the conditional zeta density of primes with first digit 1 is shown to be $\log_{10} 2$.

Chapter 5 contains a detailed examination of infinite iteration of an averaging process. The idea is that if $\delta(n) = \frac{1}{n} \sum_{i=1}^{n} a(i)$ doesn't have a limit as $n \to \infty$, it may be that reaveraging the $\delta(n)$ may improve matters. It is shown that no finite number of iterations can induce convergence if the first average did not. A properly defined, infinite number of iterations can induce convergence beyond the first average. The classification of the first two chapters is continued here. It is shown that sets of the form $A = \bigcup_{k=1}^{\infty} [10f(k), 10g(k))$ for polynomials $f, g$ have limiting infinite reaverages iff $f(X) = aX + b$, $g(X) = aX + c$ are linear and in this case the limit is $\frac{c-b}{a}$. The theory is extended to a complete analysis of iterations of log averages. The results being parallel to those described above.
The final chapter applies generalized averaging methods to random variables \( X_i \) in place of the numbers \( a(i) \). While much of the notion of weak and strong disappears since the variables \( X_i \) are, in general, unbounded, there is a good deal left to do. The material has two directions. First, there are several real world problems calling for generalized density. For example, the random power series \( \sum_{i=0}^{\infty} X_i z^i \) has an interpretation as a discounted income process where the \( X_i \) represent random costs and payments at time \( i \) in the future and \( X_i z^i \) represents the discounted present day value at discount rate \( z \). The second direction is toward the analytical problems that arise in contrasting the behavior of weighted averages to ordinary averages.

A fairly complete treatment of weak laws of large numbers is given; particular attention is given to the classical example of \( X_i \) taking values \( \pm i^a \), each with probability \( \frac{1}{2^i} \). In considering binomial averages of this sequence, a new asymptotic expansion for negative and fractional moments of the binomial distribution is derived.

There is a section on the strong law which derives almost sure results for binomial and log averages, as well as tools which are used to show that some stronger averages can take variables \( X_i \) which obey the weak law but are known not to obey the strong law and induce almost sure convergence.
The final sections derive central limit theorems and Berry-Esseen theorems for zeta and Abel averages. A typical theorem considers $X_i$ independent and identically distributed with mean 0, variance $\sigma^2$, $E(|X_i|^3) = \rho$. If $F_z(t)$ is the distribution function of the variable

$$\frac{\sqrt{1-z^2}}{\sigma} \sum_{i=0}^{\infty} X_i z^i$$

then, for $0 < z < 1$,

$$\sup_{-\infty < t < \infty} |F_z(t) - \Phi(t)| < \frac{6\rho}{\sigma^3 \sqrt{1-z^2}},$$

where $\Phi$ is the standard normal cumulative.
TABLE 2

A summary of inclusion and equivalence theorems proved between various density relationships for sets of integers.

\[
\begin{align*}
\text{binomial} & \quad \rightarrow \quad \text{Borel} \\
\downarrow & \\
\downarrow & \quad \text{negative binomial} \\
\downarrow & \\
\text{natural} & \quad \leftarrow \quad \text{Abel} \\
\downarrow & \\
\downarrow & \\
\downarrow & \\
\log & \quad \leftarrow \quad \text{zeta} \\
\downarrow & \\
\downarrow & \\
\log & \quad \leftarrow \quad \text{second zeta} \\
\downarrow & \\
\text{higher log} & \quad \leftarrow \quad \text{higher zeta}
\end{align*}
\]

An arrow between densities represents inclusion weak \(\rightarrow\) strong.

The symbol \(\rightarrow\) indicates strict inclusion.
CHAPTER 1

NATURAL DENSITY

1.1 Definitions, Examples, Basic Properties

The most widely used notion of density on infinite collections is the limit of uniform density on finite subcollections. Many people, if pressed, give some version of natural density as their interpretation of the phrase "pick a number at random". Natural density is considered a benchmark in all that follows.

Definition. If \( A = \left\{ a(i) \right\}_{i=1}^{\infty} \) is any real sequence, define \( \delta(A,n) = \frac{1}{n} \sum_{i=1}^{n} a(i) \). \( A \) is said to have natural density \( \ell \), if \( \lim_{n \to \infty} \delta(A,n) = \ell \). As an abuse of notation, if \( A \) is a subset of the natural numbers \( N = \{ 1, 2, 3, \ldots \} \) and \( a(i) \) is one or zero as \( i \) is in \( A \) or not, we write \( \delta(A,n) \) for \( \frac{1}{n} \sum_{i=1}^{n} a(i) \) and say \( A \) has natural density \( \ell \) if \( \delta(A,n) \to \ell \) as \( n \to \infty \).

Examples. The even numbers have natural density \( \frac{1}{2} \), the primes and squares have density 0, and the set of natural numbers with an odd number of digits does not have natural density.

Properties. The density of all natural numbers is 1, finite sets have density 0. A set has density iff its complement has density, natural density is invariant under
translation and addition of any finite number of elements of the set $A$.

The collection of sets having natural density does not form an algebra of sets. For example, take $A = \{\text{even numbers}\}$, $B = \{\text{even numbers with an even number of digits}\} \cup \{\text{odd numbers with an odd number of digits}\}$. A little thought shows that $A$ and $B$ each have density $\frac{1}{2}$ but $A \cap B$, $A \cup B$ do not have natural density.

For a survey of known properties of natural density containing a large bibliography, see Niven [1951].

1.2 Classification of Gap Sets

Examples of sets of integers without density are almost all what might be called "gap" sets with a long run of inclusions followed by a long run of exclusions, the run length going to infinity. One result of this dissertation is a classification of limiting behavior in terms of gap conditions. The following theorem begins this classification and yields a collection of examples used later. Roughly, it states that sets which have gaps with polynomial growth will have natural density.

**Theorem 1.** Let $f(X)$, $g(X)$ be polynomials with real coefficients where, to rule out trivial cases, it may be assumed that $\deg f = \deg g = n$ and that both leading coefficients are equal and greater than 0, thus
\[ f(x) = ax^n + bx^{n-1} + d(n-2)x^{n-2} + \ldots + d(0) \]

\[ g(x) = ax^n + cx^{n-1} + e(n-2)x^{n-2} + \ldots + e(0). \]

Further, it may be assumed that \( 0 < \frac{c-b}{na} < 1 \). Then, the set of integers \( A = \bigcup_{k=1}^{\infty} [f(k), g(k)] \) has natural density \( \frac{c-b}{na} \). In all omitted cases, \( A \) has density 0 or 1.

**Proof.** First rule out the trivial cases. If \( \deg f > \deg g \) then for all large \( k \), the intervals \([f(k), g(k)]\) are empty so \( A \) has density 0. If \( \deg f < \deg g \), then for large \( k \) all the intervals overlap and \( A \) will have density 1. Similar considerations force the two leading coefficients to be positive. Let the leading coefficients be \( a \) and \( a' \). In order that the intervals be nondegenerate past some point, it must be that

\[ (1-1) \quad f(k) < g(k) \]

and

\[ (1-2) \quad g(k-1) < f(k) \]

hold for all \( k \) greater than some fixed number. Writing out inequality (1-2) and expanding the leading term \((k-1)^n\) leads to

\[ ak^n + o(k^n) \leq a'k^n + o(k^n) \]

so that \( a \leq a' \). But \( a' \leq a \) follows from (1-1) so that \( a = a' \). Assuming this, the inequality (1-2) becomes
\[(1-3) \quad a \sum_{j=0}^{n} \binom{n}{j} (-1)^j k^{n-j} + c \sum_{j=0}^{n-1} \binom{n-1}{j} (-1)^j k^{n-1-j} + o(k^{n-1}) \leq a k^n + b k^{n-1} + o(k^{n-1}) \]

or

\[(c-na)k^{n-1} + o(k^{n-1}) \leq b k^{n-1} + o(k^{n-1}) .\]

This last inequality clearly implies that the problem will be non trivial iff

\[c - na < b < c,\]

that is, iff \(0 < \frac{c-b}{na} < 1\). Assuming that the polynomials \(f, g\) satisfy the hypothesis of the theorem,

\[\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} a(i) = \lim_{k \to \infty} \frac{1}{g(k)} \sum_{i=1}^{k} g(k) a(i) .\]

This last sum can be written

\[\sum_{i=1}^{k} [g(i)-f(i)] = \sum_{i=1}^{k} (c-b) i^{n-1} + o(k^{n-1})\]

\[= \frac{(c-b) k^n}{n} + o(k^n)\]

so

\[\lim_{k \to \infty} \frac{1}{g(k)} \sum_{i=1}^{k} g(k) a(i) = \lim_{k \to \infty} \frac{(c-b) k^n}{ng(k)} + o\left(\frac{k^n}{g(k)}\right) = \frac{c-b}{na} .\]

A similar argument applies for the lower limit so the theorem is proved. \\}
Theorem 1 is sharp in the sense that no gap set with exponential gaps has natural density. For example, the integers with an odd number of digits may be represented as the set \( A = \bigcup_{k=0}^{\infty} [10^{2k}, 10^{2k+1}] \) which does not have natural density.

Theorem 2. Let \( f(X), g(X) \) be polynomials in \( \mathbb{R}[X] \). To avoid trivial cases, assume \( f \) and \( g \) satisfy the hypothesis of theorem 1 above. Then, the set of integers \( A = \bigcup_{k=0}^{\infty} [10^f(k), 10^g(k)] \) does not have natural density. The set of limit points of \( \{\delta(A, n)\}_{n=1}^{\infty} \) form the interval \([0, 1]\) unless \( f \) and \( g \) are linear, \( f(X) = aX + b, g(X) = aX + c \), with \( 0 < \frac{c-b}{a} < 1 \) in which case the limit points of \( \{\delta(A, n)\}_{n=1}^{\infty} \) form the interval

\[
\left[ \frac{10^{c-b} - 1}{10^a - 1}, \frac{10a + (b-c)(10^{c-b} - 1)}{10^{a-1}} \right].
\]

Proof. The trivial cases may be dealt with exactly as in the proof of theorem 1 of this section. Let \( A(n) \) be the number of elements in \( A \leq n \).

Case 1: If \( f(k) = ak + b, g(k) = ak + c \) are linear with \( 0 < \frac{c-b}{a} < 1 \), consider \( j(s, n) \) of the form \( j = s10^a + b \) for \( 1 \leq s \leq 10^{c-b} \).

\[
A(j) = (10^{c-10^b}) + (10^{a+c-10^a+b}) + \ldots + (10^{(n-1)a+c-10(n-1)a+b}) + j - 10^{an+b}
\]

\[
= \frac{10^{c-10^b}}{10^a - 1} \cdot (10^{na} - 1) + j - 10^{an+b}
\]
so that

\[
\frac{A(j)}{j} = 1 - \frac{1}{s} \cdot \left( \frac{10^a - 10^{c-b}}{10^a - 1} \right) + o(1). \]

Clearly,

\[
\lim_{n \to \infty} \frac{A(n)}{n} = \lim_{k \to \infty} \frac{A(j(1,k))}{j(1,k)} = \frac{10^{c-b}-1}{10^a-1}
\]

and

\[
\lim_{n \to \infty} \frac{A(n)}{n} = \lim_{k \to \infty} \frac{A(j(10^{c-b},k))}{j(10^{c-b},k)} = \frac{10^{a+b-c}(10^{c-b}-1)}{10^a-1}
\]

By varying \(s\) in \(1 \leq s \leq 10^{c-b}\) any point in the interval can be achieved as a limit point of \(\frac{A(n)}{n}\).

Case II. If \(f(k) = ak^n + bk^{n-1} + d(n-2)k^{n-2} + \ldots + d(0),\)

\(g(k) = ak^n + bk^{n-1} + e(n-2)k^{n-2} + \ldots + e(0)\) with \(n \geq 2,\)

and \(0 < \frac{c-b}{na} < 1,\) consider \(j\) of the form

\[
j(s,k) = 10^{ak^n+(b+s(n-1))k^{n-1}+(d(n-2)+s(n-2))k^{n-2}+\ldots+(d(0)+s(0))}
\]

where \(s = (s(0), s(1), \ldots, s(n-1)) \in \mathbb{R}^n\) is constrained to lie in the compact set \(0 \leq s(i) \leq e(i) - d(i)\) if \(d(i) - e(i) > 0,\)

\(e(i) - d(i) \leq s(i) \leq 0\) otherwise.

For any such \(j,\)

\[
A(j) = \sum_{i=1}^{k-1} (10^{g(i)} - 10^{f(i)}) + j - 10^f(k).
\]

The largest term in the summation is
\[ 10^g(k-1) = 10^{ak^n + (c-na)k^{n-1} + O(k^{n-2})} \text{ so} \]

\[ \frac{10^g(k-1)}{j} \leq 10^{-k^{n-1}(na+o(1))}. \]

Using this bound for all terms in the sum yields:

\[ \frac{A(j)}{j} = 0(k10^{-k^{n-1}(na+o(1))}) + 1 - \frac{10^f(k)}{j} = 1 - \frac{10^f(k)}{j} + o(1). \]

As before, letting \( \mathbf{0} \) be the zero vector in \( \mathbb{R}^n \),

\[ \lim_{i \to \infty} \frac{A(i)}{i} = \lim_{k \to \infty} \frac{A(j(0,k))}{j(0,k)} = 0 \]

and similarly it can be shown that \( \lim_{i \to \infty} \frac{A(i)}{i} = 1 \). As \( \xi \) varies in the compact rectangle, all values in the unit interval are seen to be attainable as limit points of the sequence \( \frac{A(i)}{i} \). \( \qed \)
CHAPTER 2

ZETA DENSITY

2.1 Introduction, Definitions and Examples

A generalization of natural density is the density induced by Riemann's zeta function. We write \( \zeta(s) = \sum_{i=1}^{\infty} \frac{1}{i^s} \) and define, for \( A = \{a(i)\}_{i=1}^{\infty} \) any real numbers, \( Z(A,s) = \frac{1}{\zeta(s)} \sum_{i=1}^{\infty} \frac{a(i)}{i^s} \); \( a(i) \) is frequently taken as the indicator function of a set \( A \subseteq \{1,2,3,...\} \) with the same notation. As \( s \to 1 \), the distribution of mass becomes diffuse.

While not common, the zeta distribution \((s > 1)\) has been used in a variety of real world contexts. Word length and the number of papers published in a given time period by scientists are two quantities which have been held to have a zeta distribution. A collection of references to usage and statistical properties such as estimates of the parameter \( s \) may be found in the article by Rider (pg. 443) in Patil [1965] or in Chapter 10, Section 3.1 of Johnson and Kotz [1969].

In number theory, use of stronger averages is common. When all attention is restricted to the primes, zeta density is called analytic density or Dirichlet density (see Hasse [1950] (pg. 223-226) or Serre [1973] (pg. 125)). My introduction to the limiting case of the density comes from the interesting note by Golumb [1970].
Definition. \( A = \{ a(i) \}_{i=1}^{\infty} \) has zeta density \( \ell \) if
\[
\lim_{s \to 1} Z(A,s) = \ell \quad \text{as} \quad s \to 1.
\]

Example 1. Let \( M = \{m, 2m, 3m, \ldots \} \) be the set of multiples of \( m \).
\[
Z(M, s) = \frac{1}{\zeta(s)} \sum_{k=1}^{\infty} \frac{1}{(mk)^s} = \frac{1}{m^s}.
\]

Taking the limit as \( s \to 1 \) shows that \( M \) has zeta density \( \frac{1}{m} \) but before taking the limit, this calculation shows that divisibility by \( m \) is independent of divisibility by \( n \) if \( m \) and \( n \) have no common factors since
\[
\frac{1}{m^s} \frac{1}{n^s} = \frac{1}{(mn)^s}.
\]
This is an important difference between zeta and natural density where independence holds only after the limit has been taken. Since for \( s \to 1 \), the zeta distribution is a probability measure on \( \mathbb{N} \), all the tools of the countably additive theory are available before the limit is taken.

Example 2. The set of square free numbers has zeta density \( \frac{6}{\pi^2} \). Recall that a number is square free if it has no square prime factors. For \( s > 1 \), the probability of being divisible by \( p^2 \) is \( \frac{1}{p^{2s}} \). Thus the probability of not being divisible by any squared prime is
\[
\prod_{p} \left( 1 - \frac{1}{p^{2s}} \right) = \frac{1}{\zeta(2s)} \to \frac{1}{\zeta(2)} = \frac{6}{\pi^2}
\]
as \( s \to 1 \). The pairwise independence of the sets which are
multiples of \( p^2 \) implies the pairwise independence of their complements. The passage to the infinite product is valid by the continuity of the probability measure \( Z(\cdot, s) \).

2.2 Properties

1. Zeta density is a finitely additive translation invariant measure on \( \mathbb{N} \). Finite additivity follows from the fact that for \( s > 1 \), \( Z(\cdot, s) \) is countably additive and that limits commute with finite sums. Translation invariance means \( P(A + b) = P(A) \) if \( b \) is any integer positive or negative and \( P(A) \) exists. To see the sets \( A \) and \( A + 1 \) yield the same limiting behavior, consider

\[
Z(A, s) - Z(A+1, s) = \frac{1}{\zeta(s)} \sum_{n \in S} \left\{ \frac{1}{n^s} - \frac{1}{(n+1)^s} \right\}
\]

\[
\leq \frac{1}{\zeta(s)} \sum_{n=1}^{\infty} \left\{ \frac{1}{n^s} - \frac{1}{(n+1)^s} \right\} = \frac{1}{\zeta(s)}.
\]

Letting \( s \to 1 \) shows that \( Z(A, s) \) and \( Z(A+1, s) \) have the same limiting behavior, in particular, both have equal limit if either limit exists. The general result now follows by induction.

It is customary in using finitely additive measures to specify an algebra of subsets as the domain of definition. This algebra may be taken as the algebra of all subsets of the integers by the axiom of choice but it is worth noting that the class of subsets where zeta density exists does not form an algebra. An example can be constructed much as in Chapter 1.
Details of such an example are given in Diaconis [1973] (pg. 25) and will not be repeated here. The problem of extension via the axiom of choice is that it is essentially impossible to specify the 'probability' of any set where the limit did not originally exist. One reason for introducing stronger densities is to go as far as possible using constructive methods before using the axiom of choice.

2. If a set has natural density, then it has zeta density and the two limits are equal. The proof is a standard integration by parts and can be found clearly detailed in Ayoub [1963] (pg. 86). For reference, a formal statement follows.

**Theorem 1.** If \( \{a(i)\}_{i=1}^{\infty} \) are any real numbers, then

\[
\sum_{i=1}^{n} a(i) - \ln n \quad \text{as} \quad n \to \infty
\]

implies

\[
\sum_{n=1}^{\infty} \frac{a(n)}{n^s} \sim \zeta(s) \quad \text{as} \quad s \to 1.
\]

One consequence of this result is to validate a certain kind of probabilistic reasoning in the theory of numbers. Hardy and Littlewood [1922] (pg. 36-37) and later Polya [1959] (pg 382) consider the question: "What is the probability that a number picked at random is prime?" to show a fallacy of this type of reasoning. They reason that a number is prime iff it has no prime factor. Using heuristic independence and the 'fact' that the probability of a number being divisible by \( p \) is
\[ \frac{1}{p} \text{ this becomes: Probability of a number } \leq x \text{ being prime is } \prod_{p \leq \sqrt{x}} (1 - \frac{1}{p}). \] Now a well known theorem of Mertens (Hardy and Wright [1960] (pg. 351)) gives the value of this product as \[ 2e^{-\gamma} \frac{1}{\log x} \] where \( \gamma \) is Euler's constant. On the other hand, the number of primes \( \leq x \) is \[ \frac{x}{\log x} \] and so the 'probability' of a number being prime is \[ \frac{1}{\log x}. \] A contradiction? No, since letting \( x \to \infty \) gives, in both cases, that the 'probability' of a number being prime is zero. The only thing this example demonstrates is that the exact constant in the rate of convergence to zero is in question and we note that the heuristic answer is very close to the answer given by the prime number theorem. The point of refutation is that arguments using independence and probability statements like "the probability of a number being a multiple of \( m \) is \( \frac{1}{m} \)" can be perfectly valid if zeta density is used. The set of primes has zeta density \[ \lim_{s \to 1} \frac{1}{\zeta(s)} \sum \frac{1}{p^s} = 0. \] Theorem 1 of this section can then be cited to assert that if the set in question has natural density, it must agree with the answer given by zeta density.

3. Zeta density exists iff log density exists.

\[ \text{Definition. } A = \left\{ a(i) \right\}_{i=1}^{\infty} \text{ has log density } \lambda \text{ if } \lim_{n \to \infty} \frac{1}{\log n} \sum_{i=1}^{n} \frac{a(i)}{i} = \lambda. \]

Log density is frequently used in number theory (for example, see Halberstam and Roth [1966], (pg. 241)). A
consequence of the classical Tauberian theorems proved in
detail as theorem 6 in Chapter 4 is:

**Theorem.** A \( \subseteq \mathbb{N} \) has log density \( \lambda \) iff \( A \) has zeta density \( \lambda \).

Section 4.2 gives some other equivalent densities.

### 2.3 Classification of Gap Sets

This section delineates a class of sets having zeta dens-
ity but not natural density. Roughly, sets with exponential
gaps have zeta density while sets with more rapidly growing
gaps do not have zeta density.

**Theorem 2.** Let \( f(x) = ax^n + bx^{n-1} + d(n-2)x^{n-2} + \ldots + d(0) \),
\( g(x) = ax^n + cx^{n-1} + e(n-2)x^{n-2} + \ldots + e(0) \) be polynomials
with real coefficients satisfying \( 0 < \frac{c-b}{na} < 1 \) to avoid sets of
density 0 or 1. Let \( A \) be the integers in the union
\( A = \bigcup_{k=0}^{\infty} [10^k f(k), 10^k g(k)] \). Then, \( A \) has zeta density \( \frac{c-b}{na} \) but not
natural density.

**Comment.** A direct proof in the loglinear case \((f, g \text{ linear})\) is possible as given in Diaconis [1973]. For deg \( f \geq 2 \),
a direct proof is not known. A proof using the equivalence of
log and zeta density will now be given.

**Proof:** The theorem is proved if \( \lim_{k \to \infty} \frac{1}{\log k} \sum_{i=1}^{k} \frac{a(i)}{i} = \frac{c-b}{na} \)
is demonstrated:
\[
(2-1) \quad \lim_{n \to \infty} \frac{1}{\log k} \sum_{i=1}^{k} \frac{a(i)}{i} = \lim_{k \to \infty} \frac{1}{(\log 10)g(k)} \sum_{i=1}^{g(k)} \frac{a(i)}{i}.
\]

Writing \( t(x) \) for \( 10^x \) and using the standard analytic fact:
\[
b \sum_{m=a}^{b} \frac{1}{m} = \log \frac{b}{a} + O(\frac{1}{a}),
\]
the last sum in (2-1) is
\[
\sum_{i=1}^{k} \frac{t(g(k))}{m} = \sum_{i=1}^{k} \left\{ \log \frac{t(g(k))}{t(f(k))} + O\left(\frac{1}{t(f(k))}\right) \right\}
\]
\[
= (\log 10) \sum_{i=1}^{k} \{g(k) - f(k)\} + O(1).
\]

Thus the upper limit in (2-1) will be the same as the limit of
\[
\frac{1}{g(k)} \sum_{i=1}^{k} \{g(k) - f(k)\}.
\]

This sum was shown to approach \( \frac{c-b}{na} \) in Theorem 1 of Chapter 1 which shows that the upper limit in (2-1) is \( \frac{c-b}{na} \). A similar argument holds for the lower limit and proves the theorem. ||

**Example.** Theorem 2 shows that the set of numbers with an even number of digits and the first digit set have zeta density \( \frac{1}{2} \) and \( \log_{10} 2 \) respectively. Theorem 2 of Chapter 1 showed these sets did not have natural density.

The results of Theorem 2 can be shown to be sharp in the sense that sets with gaps of the form \([10^{f(k)}, 10^{g(k)}]\) do not have zeta density. The proof can be made to depend on log density and follows from the estimates of Theorem 2 of Chapter 1 in much the same way as Theorem 2 of this chapter followed from the estimates of Theorem 1 in Chapter 1.
CHAPTER 3

BINOMIAL DENSITY

3.1 Definitions

The family of binomial probability measures on the integers is denoted

\[ b(k,n,p) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0,1,2,\ldots,n \text{ and } p \in (0,1) \]

\[ = 0 \quad \text{otherwise.} \]

For fixed p these can be made diffuse by letting n → ∞. The notion of binomial density dates back to Euler who used a roughly equivalent method to sum divergent series. Complete references may be found in Hardy [1949] Chapters 8 and 9. The view taken here, emphasizing summability of sequences and probabilistic aspects has a different flavor from the classical work. The main results of this section are believed to be new.

**Definition.** For \( A = \{a(i)\}_{i=0}^{\infty} \) let \( B(A,n,p) = \sum_{i=0}^{n} a(i) b(i,n,p) \).

A has binomial density \( \lambda \) if \( B(A,n,p) \rightarrow \lambda \) as \( n \rightarrow \infty \). By abuse of notation if \( A \subseteq N \) and \( a(i) \) is the indicator function of \( A \), \( A \) is said to have binomial density \( \lambda \) if \( B(A,n,p) \rightarrow \lambda \).

The statement that \( A \subseteq N \) has binomial density \( \lambda \) has an obvious combinatorial interpretation: Flip a coin \( n \) times, for large \( n \) the probability that the number of heads \( S(n) \in A \) is approximately \( \lambda \). For many sets, binomial density behaves like natural density but the proofs become more difficult. Thus,
the examples that follow are stated as theorems.

3.2 Examples

Theorem 1. The set \( M \) of multiples of \( m \) has binomial density \( \frac{1}{m} \). In fact

\[
\sum_{i \in M} b(i,n,p) = \frac{1}{m} + O(e^{-cn})
\]

for \( c > 0 \) a constant depending on \( p \) and \( m \).

Proof. If \( \omega \) is a primitive \( m^{th} \) root of unity, it is elementary that

\[
1 + \omega h + \omega^2 h + \ldots + \omega^{(m-1)h} = \begin{cases} 
\binom{m}{h} & \text{if } h|m \\
0 & \text{otherwise.}
\end{cases}
\]

Let \( q = 1 - p \), then

\[
(p+q)^n + (\omega p+q)^n + \ldots + (\omega^{m-1}p+q)^n = \sum_{0}^{n} b(j,n,p) + \ldots
\]

\[
+ \sum_{0}^{n} b(j,n,p)\omega^{(m-1)j}
\]

\[
= m \sum_{0}^{n} \binom{n}{mj} p^j (1-p)^{n-j}
\]

so that

\[
(3-1) \quad \frac{1}{m} + \left(\frac{\omega p+q}{m}\right)^n + \ldots + \left(\frac{\omega^{m-1}p+q}{m}\right)^n = \sum_{0}^{n} \binom{n}{mj} p^j (1-p)^{n-j}.
\]

The right side of (3-1) is the required probability (using the convention \( \binom{a}{b} = 0 \) for \( b > a \)). Taking the limit in (3-1) as \( n \to \infty \), the left side is \( \frac{1}{m} + \) terms involving \( \omega \). Each of the
terms involving \( \omega \) is of the form \( a^n \) where \( |a| < 1 \) so that the \( m-i \) terms go to zero as \( n \to \infty \).

It is useful to get more explicit bounds for the error. While better results are possible, we show if

\[
E = E(n, p, m) = \frac{1}{m} \sum_{j=1}^{m-1} \left( \frac{2\pi ij}{m} + q \right)^n
\]

then

\[
|E| \leq e^{-\frac{cnpq}{m^2}} \tag{3-2}
\]

where \( c \) is an absolute constant and \( q = 1-p \). The \( m \)th roots of unity \( \omega^j \) may be taken as \( e^{i\phi} \) with \( \phi = \phi(j) = \frac{2\pi j}{m} \). The term

\[
(q+pe^{i\phi}) = (q + p \cos \phi + ip \sin \phi) = \text{Re}^{i\theta}
\]

with

\[
R^2 = (q + p \cos \phi)^2 + (p \sin \phi)^2 = (1 - 4pq \sin^2(\frac{\phi}{2})).
\]

Thus,

\[
\log R = \frac{1}{2} \log(1 - 4pq \sin^2(\frac{\phi}{2})) = -2pq \sin^2(\frac{\phi}{2}) - \frac{1}{4}(4pq)^2 \sin^4(\frac{\phi}{2}) - \frac{1}{6}(4pq)^3 \sin^6(\frac{\phi}{2}) - \ldots
\]

so \( \log R < -2pq \sin^2(\frac{\phi}{2}) \) or finally, since \( 0 \leq a \leq \frac{\pi}{2} \) implies \( \sin a > \frac{2}{\pi} a \), \( \log R < -c_{pq} \phi^2 \) or \( R < e^{-c_{pq} \phi^2} \) and so the typical term in the sum \( E \) satisfies
\[ \left( q + pe^{\frac{-c'npqj^2}{m^2}} \right)^n \leq \exp[- \frac{c'npqj^2}{m^2}] \]

where \( c' = 4\pi \). Since this is decreasing in \( j \), the claimed upper bound (3-2) for \(|E|\) is found by replacing \( j \) by 1. A slightly more detailed estimate gives

\[ (3-3) \quad |E| \leq (\frac{npq + cm^2}{mnpq}) \exp[- \frac{cnpq}{m^2}] \quad \text{for} \quad c = 4\pi. \]

**Remarks.** The improved bound (3-3) shows that the result of Theorem 1 holds uniformly in \( m = O(\sqrt{n}) \). For fixed \( m \), the theorem shows the error goes to zero exponentially in \( n \) whereas for natural density the error for the set \( M \) is \( O(\frac{1}{n}) \) and for log density the error for \( M \) is \( O(\frac{1}{\log n}) \). The idea seems to be that while binomial density exists for fewer sets, when it does exist we can hope for improved rates of convergence.

It is of interest to note that a probabilistic proof of relation (3-1) in Theorem 1 can be given. Since this technique may prove useful in related problems we sketch a proof.

Let \( a(n,i) \) for \( n = 0,1,2,\ldots \), and \( i = 0,1,2,\ldots, m-1 \) be the probability that in \( n \) flips of a coin the number of heads is congruent to \( i \mod m \). Introducing generating functions

\[ A(s,i) = \sum_{n=0}^\infty a(n,i)s^n \]

for \( i = 0,1,\ldots, m-1 \), one easily shows they satisfy:

\[ A(s,0) = qsA(s,0) + psA(s,m-1) + 1 \]

\[ \vdots \]

\[ A(s,i) = qsA(s,i) + psA(s,i-1) \]

\[ \vdots \]

\[ A(s,m-2) = qsA(s,m-2) + psA(s,m-3) \]
from which it follows that

\[ A(s,0) = \frac{(1-qs)^{m-1}}{(1-qs)^m-(ps)^m}. \]

The denominator as a polynomial in \( s \) has \( m \) distinct zeros at points \( s = \frac{1}{\omega^{p+q}} \) where \( \omega \) is any \( m \)th root of unity. A straightforward partial fraction expansion leads to

\[ a(n,0) = \frac{1}{m} (1 + t_1^n + t_2^n + \ldots + t_{m-1}^n) \]

where

\[ t_j = (q + pe^{2\pi ij/m}). \]

**Theorem 2.** Let \( Q \) be the set of square-free numbers. \( Q \) has binomial density \( \frac{6}{\pi^2} \). In fact,

\[ \frac{1}{2^n} \sum_{i \in Q} \binom{n}{i} = \frac{6}{\pi^2} + O\left(\frac{1}{\log n}\right). \]

**Proof.** If \( \omega \) is a primitive \( m \)th root of unity, then relation (3-1) of Theorem 1 can be written as:

\[ (3-1a) \quad \left(\frac{1}{m} \sum_{j=0}^{m-1} (1+\omega^j)^n\right) - 1 = \binom{n}{m} + \binom{n}{2m} + \ldots + \binom{n}{m\lfloor \frac{n}{m} \rfloor} = c(m,n), \]

where \( \lfloor \cdot \rfloor \) denotes greatest integer. In what follows, \( r \) will denote a number which will eventually be taken as a function of \( n \). We sieve out multiples of squared primes \( p^2 \leq r \) from \( (1+1)^n \) using inclusion-exclusion and derive the following two inequalities:
\[ (3-2a) \sum_{j \in Q_{\lambda}} (n_j^2) \geq 2^n - \sum_{p \leq \sqrt{n}} c(p^2, n) + \sum_{p < q \leq r} c((pq)^2, n) - \ldots \]
\[ \ldots \cdots (-1)^k \sum_{p_1 < \ldots < p_k \leq r} c((p_1 p_2 \ldots p_k)^2, n) \]
\[ + \ldots (-1)^{k(r)} c((p_1 p_2 \ldots p_k(r))^2, n) \]

\[ (3-3a) \sum_{j \in Q_{\lambda}} (n_j^2) \leq 2^n - \sum_{p \leq r} c(p^2, n) + \ldots (-1)^{k(r)} c((p_1 p_2 \ldots p_k(r))^2, n) \]

where \( k(r) \) is the number of primes \( \leq r \). In (3-2a) we first sieve out the coefficients \( \binom{n}{kp^2} \) for \( p \leq \sqrt{n} \) but then put back multiples of \((pq)^2\) which were removed twice and so on. The sieving process is carried out exactly for primes \( p \leq r \) but not so for primes \( r < p \leq \sqrt{n} \) whence the inequality. In (3-3a) the sieving process has removed exactly coefficients \( \binom{n}{kp^2} \) for primes \( p \leq r \); the terms on the right side of (3-3a), not on the left side of (3-3a), are all positive so the inequality follows.

We now examine the right side of (3-2a). The \((-1)\) terms on the left side of (3-1a) which defines \( c(m, n) \) can be treated separately. They are

\[ \sum_{p \leq \sqrt{n}} - \sum_{p < q \leq r} + \ldots (-1)^{k(r)} = \sum_{r < p \leq \sqrt{n}} + \left[ \sum_{p \leq r} - \sum_{p < q \leq r} \right. \]
\[ + \ldots (-1)^{k(r)} \] .

The first sum in square brackets is \( k(r) \), the second is \(-\binom{k(r)}{2}\), and we recognize \((1-1)^{k(r)} - 1\) so that the sum of all \((-1)\) terms
in (3-2a) is \( S(n, r) = \sum_{r < p \leq \sqrt{n}} 1 \). Notice that the left side of (3-1a) can be simplified by separating the first term of the sum (which is \( 2^n \)) and using the relation \( 1 + e^{i\theta} = 2 \cos(\frac{\theta}{2}) e^{\frac{i\theta}{2}} \).

Pairing conjugate roots of unity, (3-1a) becomes:

\[
\frac{2^n}{m} + \frac{2^{n+1}}{m} \sum_{j=1}^{m-1} (\cos \frac{\pi j}{m})^n \cos \left( \frac{\pi j n}{m} \right) = \frac{2^n}{m} + 2^{n+1} R(m, n) - 1.
\]

With this notation, (3-2a) becomes

\[
(3-4) \sum_{i \in \mathbb{Q}_1} \binom{n}{i} \geq 2^n \left[ 1 - \sum_{p \leq r} \frac{1}{p^2} + \sum_{p < q \leq r} \frac{1}{(pq)^2} - \ldots \frac{(-1)^k(r)}{(p_1 \ldots p_k(r))^2} \right] - 2^n \sum_{r < p \leq \sqrt{n}} \frac{1}{p^2} + S(n, r) - 2^{n+1} \sum_{p \leq \sqrt{n}} R(p^2, n)
\]

\[
+ 2^{n+1} \sum_{p < q \leq r} R((pq)^2, n)
\]

\[
\ldots (-1)^k(r) 2^{n+1} R((p_1 \ldots p_k(r))^2, n) .
\]

The term in square brackets is \( \Pi_1 \left( 1 - \frac{1}{p^2} \right) \) so that (3-4) becomes

\[
(3-5) \sum_{i \in \mathbb{Q}_1} \binom{n}{i} \geq 2^n \Pi_1 \left( 1 - \frac{1}{p^2} \right) - 2^n \sum_{r < p \leq \sqrt{n}} \frac{1}{p^2} + S(n, r)
\]

\[
+ 2^{n+1} \left[ - \sum_{p \leq \sqrt{n}} R(p^2, n) \ldots (-1)^k(r) R((p_1 \ldots p_k(r))^2, n) \right]
\]

Similarly, (3-3a) can be rewritten as
(3-6) \[ \sum_{i \in Q} \binom{n}{i} \leq 2^n \prod_{p \leq r} (1 - \frac{1}{p^2}) - 1 + 2^n \left[ - \sum_{p \leq r} R(p^2, n) + \ldots (-1)^k(r) R((p_1 \ldots p_k(r))^2, n) \right]. \]

We now study the right sides of (3-5) and (3-6). To study the trigonometric sums, we use the easily verified inequality \[ \cos x \leq e^{-\frac{x^2}{2}} \quad \text{for} \quad 0 \leq x \leq \frac{\pi}{2}. \]

Noting that all the \( R(m, n) \) sums have the same form, and using the inequality for \( \cos \),
\[ |R(m, n)| \leq \frac{1}{m} \sum_{j=1}^{m-1} \exp\left[ -\frac{\pi^2 j^2 n}{2m^2} \right] \leq \frac{1}{m} \int_{0}^{\frac{\pi}{m}} \exp\left[ -\frac{\pi^2 t^2 n}{2m^2} \right] dt \]

since the function is decreasing and positive for \( 0 < t < \frac{m}{\pi} \). Making the substitution \( s = \frac{\pi \sqrt{n} t}{m} \), leads to
\[ (3-7) \quad |R(m, n)| \leq \frac{1}{\pi \sqrt{n}} \int_{0}^{\frac{\pi}{\sqrt{n}}} e^{-\frac{s^2}{2}} ds \leq \frac{1}{\pi \sqrt{n}} \int_{0}^{\infty} e^{-\frac{s^2}{2}} ds = \frac{1}{\sqrt{2\pi n}} \]

independent of \( m \). Using (3-7) and the fact that \( a > -|a| \) in (3-5) yields
\[ (3-8) \quad \sum_{j \in Q} \binom{n}{j} \geq 2^n \prod_{p \leq r} (1 - \frac{1}{p^2}) - \left\{ 2^n \sum_{r < p \leq \sqrt{n}} \frac{1}{p^2} + |S(n, r)| \right\} + \frac{2^{n+1}}{\sqrt{2\pi n}} \left[ \sum_{p \leq r} \frac{1}{p} + \sum_{p < q \leq r} \ldots + 1 \right] + \frac{2^{n+1}}{\sqrt{2\pi n}} \sum_{r < p \leq \sqrt{n}} \frac{1}{p} \right\}. \]
There are \( k(r) \) different summations in the square brackets at the right of (3-8). Their sum is \((k(r) + \frac{k(r)}{2}) + \ldots + 1 \leq 2k(r)\). Using this in (3-8) gives

\[
\sum_{j \in Q} \left( \begin{array}{c} n \\ j \end{array} \right) \geq 2^n \prod_{p \leq r} \left( 1 - \frac{1}{p^2} \right) - \left\{ 2^n \sum_{r < p \leq \sqrt{n}} \frac{1}{p^2} + |S(n,r)| + \frac{2^{n+1+k(r)}}{\sqrt{2\pi n}} \right. \\
\left. + \frac{2^{n+1}}{\sqrt{2\pi n}} \sum_{r \leq \sqrt{n}} \frac{1}{p} \right\}.
\]

Now use \( \sum_{p \leq \sqrt{n}} \frac{1}{p} < \frac{\sqrt{n}}{r} \) and \( \sum_{p \leq \sqrt{n}} \frac{1}{p} < \frac{a\sqrt{n}}{\log n} \) for some constant \( a \), recalling the definition of \( S(n,r) \) above, leads to

\[
(3-9) \quad \frac{1}{2^n} \sum_{j \in Q} \left( \begin{array}{c} n \\ j \end{array} \right) \geq \prod_{p \leq r} \left( 1 - \frac{1}{p^2} \right) - \left\{ \frac{r}{2^n} + \frac{bn}{\log n} + \frac{c}{\log n} + \frac{d2^k(r)}{\sqrt{n}} \right\}.
\]

Choosing \( r = \log n \) leads to \( k(r) < \frac{2 \log n}{\log \log n} \), so

\[
\frac{2^k(r)}{\sqrt{n}} \leq \exp\left[ \frac{2 \log n}{\log \log n} - \frac{1}{2} \log n \right] = \exp\left[ \log n \left( \frac{2}{\log \log n} - \frac{1}{2} \right) \right]
\]

\[
< E n^{-1/2 + \epsilon}
\]

for suitable constant \( E \) and any \( \epsilon \). With this choice of \( r \), (3-9) becomes:

\[
\frac{1}{2^n} \sum_{j \in Q} \left( \begin{array}{c} n \\ j \end{array} \right) \geq \prod_{p \leq \log n} \left( 1 - \frac{1}{p^2} \right) - \frac{f}{\log n}.
\]

Since also
\[
\prod_{p \leq \log n} (1 - \frac{1}{p^2}) - \prod_{j > \log n} (1 - \frac{1}{j^2}) \leq \sum_{j > \log n} \frac{1}{j^2} \leq \frac{g}{\log n},
\]

we have (since \(\prod (1 - \frac{1}{p^2}) = \frac{6}{\pi^2}\)):

\[
(3-10) \quad \frac{1}{2^n} \sum_{j \in Q} \left( ^n \! \choose j \right) > \frac{6}{\pi^2} - \frac{h}{\log n}.
\]

Exactly the same estimates applied to (3-6) given the reverse inequality with a bit less work, so the theorem has been proved.

**Remarks.** This gives a new proof of the well known fact (Hardy and Wright [1960], pg. 269) that the square-free numbers have natural density \(\frac{6}{\pi^2}\). The value of having proved Theorem 2 is that it is now known to hold for all stronger densities in particular Borel and natural densities (Borel densities are introduced in Section 3.3 of this chapter). A completely different proof of Theorem 2 based on a Tauberian argument is given as Corollary 2 of this chapter. There the error term is improved to \(O\left( \frac{1}{(\log n)^d} \right)\) for any large \(d\).

3.3 Properties

1. The limiting density is a finitely, additive translation invariant measure on the nonnegative integers, the existence and value of the limit being invariant under translation or removal of any finite number of terms. These results are all elementary. A proof of translation invariance may be
made to depend on Pascal's recurrence: If $B(A,n,p) = \sum_{i \in A} b(i,n,p)$, then $B(A-1,n,p) = \frac{1}{p} B(n+1,A,p) - \frac{q}{p} B(A,n,p)$. Passage to the limit shows that if $A$ has density $\lambda$, then $A-1$ has density $(\frac{1}{p} - \frac{q}{p})\lambda = \lambda$. A similar argument works for $A+1$, invariance follows by induction.

2. **Theorem 3.** If a set of integers has binomial density $\lambda$, then it has natural density $\lambda$ but not conversely.

**Comment.** This Abelian theorem follows from a straightforward combination of classical results and it is surprising not to have found it mentioned in print. Indeed, a popularly cited result (Peyerimhoff [1969], pg. 79) is the non-equivalence of any Cesaro mean with any binomial average. This just underscores the dependence of Theorem 3 on the hypothesis that $a(i) = 0$ or $1$. The most convenient approach to Theorem 3 is to introduce Borel density, the density arising from the limit of Poisson measures on $N$.

**Definition.** For $A = \{a(i)\}_{i=1}^{\infty}$, let $P(A,\lambda) = e^{-\lambda} \sum_{i=0}^{\infty} a(i) \frac{\lambda^i}{i!}$. A has Borel density $\lambda$ if $P(A,\lambda) \to \lambda$ as $\lambda \to \infty$.

**Lemma 1.** For any $A = \{a(i)\}_{i=1}^{\infty}$, if $A$ has binomial density, then $A$ has equal Borel density.

**Proof.** We are given $B(A,n,p) \to \lambda$ as $n \to \infty$. Now

$P(A,\lambda) \to \lambda$ iff $P(A,(p\lambda)) \to \lambda$ as $\lambda \to \infty$
and, writing \( q = 1 - p \),

\[
e^{qx} \sum a(i) \frac{(px)^i}{i!} = \left( \sum \frac{q(x)^i}{i!} \right) \sum \frac{a(i)(px)^i}{i!} = \sum c(n)x^n
\]

where

\[
c(n) = \frac{a(0)q^n}{n!} + \frac{a(1)pq^{n-1}}{n!(n-1)!} + \ldots + \frac{a(n)p^n}{n!} = \frac{1}{n!} B(A,n,p).
\]

Thus,

\[
e^{-px} \sum a(i) \frac{(px)^i}{i!} = e^{-px} qxe^{qx} \sum a(i) \frac{(px)^i}{i!} = e^{-x} \sum B(A,n,p) \frac{x^n}{n!}.
\]

But, since \( B(A,n,p) \to \lambda \) and by the Toeplitz lemma, Borel averages of convergent sequences converge to the same limit, the right side of the last equation \( \to \lambda \) as \( x \to \infty \) showing that \( A \) has Borel density \( \lambda \). |||

**Proof of Theorem 3.** By Lemma 1, the set \( A \) of Theorem 3 is known to have Borel density \( \lambda \) and we can adapt a result from Peyerimhoff [1969] (pg. 76), to our purpose.

**Theorem 4.** If \( A \subset N \) has Borel density \( \lambda \), then \( A \) has Abelian density \( \lambda \).

Finally, from the Hardy-Littlewood Tauberian theorem discussed here as Theorem 4 of Chapter 4, sets of integers have natural density iff they have equal abelian density.

The converse of Theorem 4 fails as shown by a counterexample, constructed using the bounds given by the central limit theorem of probability theory. Such an example is given explicitly in Diaconis [1973] (pg. 19-20), and will not be repeated here. |||
Remark. It is unknown if Borel density is actually equivalent to binomial density. This should be true since, loosely, both concentrate mass within $\pm \sqrt{\lambda}$ of a central point.

3. Compatibility of different binomial densities. It appears that if a set of integers $A$ has binomial density for some $p$, $\lim B(A,n,p)=\lambda$ as $n \to \infty$. Then $B(A,n,p')=\lambda$ for all $p'$. The best that can be proved at present is the truth of this statement for $p'<p$. This will follow from an adaption of a known lemma (Hardy [1949], pg. 179). Consider $E_p$ an infinite lower triangular matrix with rows $b(i,n,p)$ $i=0,1,2,...,n$.

Lemma 2. $E_p E_p^T = E_{pp^1}$.

Proof. The $(i,j)$ entry of $E_p$ is $(i) p^1 q^{i-j}$, thus the $(i,j)$ entry of the product is 0 if $j > i$. For $j \leq i$, it is

$$\sum_{k=j}^{i} \binom{i}{k} p^k q^{i-k} \binom{k}{j} p^j q^{k-j} = \binom{i}{j} p^j \sum_{k=0}^{i-j} \binom{i-j}{k} p^{k+j} q^{j+k} (i-j) = (\binom{i}{j}) (pp_1)^j (pq_1+q)^{i-j} = (\binom{i}{j}) (pp_1)^j (1-pp_1)^{i-j}.$$  

Corollary 1. Let $a$ be an infinite column vector with elements $a(i)$. If

$$\lim_{i \to \infty} (E_p a)_i = \lambda \text{ then } \lim_{i \to \infty} (E_p a) = \lambda$$
for all \( p_1 \leq p \). In particular, if
\[
B(A,n,p) \to \ell \quad \text{then} \quad B(A,n,p_1) \to \ell \quad \text{for} \quad p_1 \leq p.
\]

**Conjecture.** It is known that for general \( a(i) \), the binomial methods strictly increase in strength so that the converse of Corollary 1 fails. We conjecture that for \( a(i) = 0 \) or 1, \( B(A,n,p) \to \ell \) implies \( B(A,n,p_1) \to \ell \) for all \( p_1 \).

3.4 **A Tauberian Theorem for Binomial Density**

Counterexamples such as in Theorem 3 cannot happen if the set in question has natural density with sufficiently good error term. Roughly, good enough means \( o(1/\sqrt{n}) \). A neat statement of this requires the ideas of regular variation. The fine treatment of Feller [1966], Chapter 8, Section 8, serves as a reference for the material needed.

Recall, a real valued function \( L(x) \) varies slowly at \( \infty \) if for all \( t \), \( \frac{L(tx)}{L(x)} \to 1 \) as \( x \to \infty \). A function \( f \) is said to vary regularly at \( \infty \) if \( f(x) \sim x^a L(x) \) where \( L \) varies slowly; \( a \) is called the exponent of variation. The following lemma combines some facts in a form used later.

**Lemma 3.** If \( Z(x) \) varies regularly with exponent \( a \), then
\[
(p+a+1) \sum_{i \leq x} Z(i)i^a \sim Z(x)x^{p+1} \quad \text{as} \quad x \to \infty \quad \text{if} \quad p+a+1 \geq 0.
\]
Proof. A proof with sums replaced by integrals is given in Feller [1966] (pg. 273). The version for sums is an easily verified corollary. |||

Theorem 5. If \( \sum_{i=0}^{n} a(i) = \mathcal{N} + O(f(n)) \) where \( f(n) \) varies regularly with exponent \( a \), \( 0 \leq a \leq \frac{1}{2} \), and \( a(i) \) are arbitrary real numbers, then

\[
(3.11) \quad \sum_{i=0}^{n} a(i) {n \choose i} p^i q^{n-i} = \mathcal{N} + O\left(\frac{f(n)}{\sqrt{n}}\right) \quad \text{for all } p \in (0, 1).
\]

Proof. \( B(A, n, p) = \sum_{i=0}^{n} a(i) {n \choose i} p^i q^{n-i} = q^n \sum_{i=0}^{n} a(i) {n \choose i} r^i \)

with \( r = \frac{p}{q} \). It is convenient to deal with the sum in the second form. Summation by parts gives

\[
(3.12) \quad \sum_{i=0}^{n} a(i) {n \choose i} r^i = A(n) + \sum_{j=0}^{n-1} A(j) \Delta(j)
\]

where \( A(j) = \sum_{i=0}^{j} a(i) \) and

\[
(3.13) \quad \Delta(j) = {n \choose j} r^j - {n \choose j+1} r^{j+1} = {n \choose j} r^j \left\{ \frac{(r+1)j+1-rn}{j+1} \right\}.
\]

Using the hypothesis, (3.12) becomes

\[
(3.14) \quad = 0(n) + \sum_{j=1}^{n-1} j \Delta(j) + \sum_{j=1}^{n-1} O(f(j)) \Delta(j)
\]

\[
= 0(n) + \frac{1}{q^n} + O\left(\sum_{j=1}^{n-1} |\Delta(j)| f(j)\right).
\]

It remains to bound the error term. It follows from (3.13) that \( \Delta(j) \) is of constant sign for \( 0 \leq j \leq np - f \) and for \( np - q < j \leq n \). Taking \( p = -1 \) in Lemma 3 implies
\[
\sum_{i \leq x} \frac{f(i)}{i} = O(f(x)) \quad \text{for all } x > 0.
\]

Let \(v\) be a mode of the underlying binomial distribution:

\[np - q < v \leq (n+1)p.\]

Break the error sum into two. For the lower sum

\[
\sum_{j \leq v} \binom{n}{j} r^j \frac{f(j)}{j} = \left( \sum_{j \leq r} \frac{f(j)}{j} \binom{n}{j} r^j v^v + \sum_{j \leq v-1} \Delta(j) \left( \sum_{i=1}^{j} \frac{f(i)}{i} \right) \right)
\]

\[
= O\left( \frac{f(n)}{q^v n^v} + O\left( \frac{\sum_{i} \Delta(j) f(j)}{n^v} \right) \right).
\]

The first big \(O\) term comes from the standard bound for the maximal term of the binomial distribution (Feller [1968], pg. 151) and the fact that \(O(f(v)) = O(f(n))\) for a regularly varying \(f\).

The sum on the left side of (3-16) can be estimated directly using

\[
\sum_{j \leq m} \binom{n}{j} r^j \leq \frac{p(n-m)}{q^n (n-p-m)^2} \quad \text{for } m < v
\]

(Feller [1968], pg. 151). Break the left side of (3-16) into two parts:

\[
\sum_{j \leq v} \binom{n}{j} r^j \frac{f(j)}{j} \ll \sum_{j \leq m} \binom{n}{j} r^j + \frac{f(m)}{m} \sum_{j=m}^{v} \binom{n}{j} r^j \quad \text{for } m = np - n^\frac{3}{4}.
\]

The first sum at the right, using (3-17), is \(O(\frac{1}{q^n v^n})\) while the second sum is clearly \(O\left( \frac{f(n)}{n^q n^v} \right)\). Using these estimates in (3-16) gives
\begin{align*}
(3-18) \quad \frac{1}{\sqrt{n}} \sum_{j=1}^{\sqrt{n}} \Delta(j) f(j) &= O\left(\frac{f(n)}{nq^{n/\sqrt{n}}} + \frac{f(n)}{nq^n} + \frac{1}{q^n}\right) \\
&= O\left(\frac{f(n)}{q^n}\sqrt{n}\right).
\end{align*}

Very similar arguments using the analog of (3-17) for the upper tail and summation by parts show

\[ O\left( \sum_{\nu<j \leq n} \Delta(j) f(j) \right) = O\left( \frac{f(n)}{nq^n} \right). \]

Combining this with (3-18) shows

\[ O\left( \sum_{j=1}^{n-1} |\Delta(j)| f(j) \right) = O\left( \frac{f(n)}{q^n}\sqrt{n} \right). \]

Using this in (3-14) proves the theorem. \[ \]

**Corollary 2.** The square-free numbers \( Q \) have binomial density \( 6/\pi^2 \). In fact,

\[ \sum_{i \in Q} b(i, n, p) = \frac{6}{\pi^2} + O\left( e^{-c \sqrt{\log n \log \log n}} \right) \]

for \( c \) a constant. If the Riemann hypothesis is true, the error can be improved to \( O\left( \frac{1}{n^{(1/10)-\varepsilon}} \right) \) for arbitrary \( \varepsilon > 0 \).

**Remark.** The improvement here is three fold. The theorem is proved for all \( p \) instead of just \( p = \frac{1}{2} \) (see Theorem 2 of this chapter), with better error and makes contact with standard areas of mathematics. It should be noted that a strong form of the prime number theorem is needed to get the results of
Corollary 2 and as such they are not 'elementary' as that term is used in the theory of numbers.

**Proof of Corollary 2.** The results follow from corresponding results in the natural density problem. Let \( Q(x) = \sum_{1 \leq i \leq x} 1 \) and \( E(x) = Q(x) - x \frac{6}{\pi^2} \).

**Theorem 6.** (Evelyn and Linfoot [1931.])

\[
E(x) = O(\sqrt{x} \ e^{-c \sqrt{\log x \ \log \log x}}).
\]

**Theorem 7.** (Vaidya [1966].) If the Riemann hypothesis is true, then \( E(x) = O(x^{2/5} + \varepsilon) \).

The stated results now follow from Theorem 5 by noting in both the bound for \( E(x) \) varies regularly at \( \infty \) with suitable exponent. |||
CHAPTER 4

APPLICATIONS OF TAUBERIAN THEOREMS

This chapter introduces an important class of tools: the classical Tauberian-Abelian theorems. Three theorems have a direct probabilistic interpretation which can be exploited to give an understanding of many methods of assigning average values. Relations will be derived between the methods introduced thus far: natural, binomial, log, zeta, and a host of new methods: Abel, higher log, negative binomial, and others. The classical Tauberian arguments don't suffice for all purposes and an extension of density Tauberian theorems is derived. The final subsections offer connections between summability results and some of the classical theorems in prime number theory as well as extend much of the discussion to conditional densities.

4.1 The Classical Theory and Some Extensions

We state for reference the Tauberian theorems as given in Feller [1966], Chapter 13.5. Recall, a positive function \( L(x) \) defined on \((0, \infty)\) is slowly varying at \( \infty \) if \( L(tx)/L(t) \to 1 \) as \( t \to \infty \). We follow Feller in introducing the notation \( U(t) \) for the distribution function of a measure concentrated on \((0, \infty)\) with finite Laplace transform \( \omega(t) = \int_0^\infty e^{-tx}U(dx) \). The basic theorem is:
Theorem 1. (Feller [1966], pg. 421.) If \( L \) varies slowly at \( \infty \) and \( 0 \leq a < \infty \), then each of the relations

\[
(4-1) \quad \omega(\tau) \sim \tau^{-a} L\left(\frac{1}{\tau}\right) \quad \text{as} \quad \tau \to 0
\]

\[
(4-2) \quad U(t) \sim \frac{1}{\Gamma(a+1)} t^a L(t) \quad \text{as} \quad t \to \infty
\]

implies the other and this is true if the role of \( \tau \) and \( t \) are interchanged.

If the distribution function \( U \) is differentiable, with density \( u \), there is an important extension which is suggested by differentiation of (4-2) treating \( L(t) \) as constant.

Theorem 2. (Feller [1966], pg. 423.) Using the notation above, for \( 0 < a < \infty \), consider

\[
(4-3) \quad \omega(\tau) \sim \frac{1}{\tau^a} L\left(\frac{1}{\tau}\right) \quad \tau \to 0
\]

\[
(4-4) \quad u(t) \sim \frac{1}{\Gamma(a)} t^{a-1} L(t) \quad \text{as} \quad t \to \infty.
\]

Then (4-4) implies (4-3) with no restriction on \( u \). If \( u \) is ultimately monotone, then (4-3) implies (4-4).

For many purposes, the restriction to monotone density is too stringent. Note that if the density \( u \) itself was differentiable, then monotonicity means \( u'(x) < 0 \) or \( u'(x) > 0 \). A useful extension of Feller's theorem is:

Theorem 3. Notation as above, assume \( u(x) \) is differentiable and satisfies
\[ u'(x) < Ax^{a-2}L(x) \quad \text{or} \quad u'(x) > -Ax^{a-2}L(x) \]

for \( A > 0 \) and all sufficiently large \( x \), where \( 0 \leq a < \infty \). Then, referring to Theorem 2, (4-3) iff (4-4).

Note: If \( a = 0 \), (4-4) becomes \( u(x) = o\left(\frac{L(x)}{x}\right) \) as \( x \to \infty \).

Proof. W.L.O.G. assume \( u'(x) < Ax^{a-2}L(x) \) for large \( x \). Since (4-4) always implies (4-3), we need only show (4-3) implies (4-4). Since (4-3) and (4-2) are equivalent, assume \( U(x) \sim x^aL(x) \) as \( x \to \infty \). Taylor's theorem gives for \( 0 < \delta < 1 \),

\[ (4-5) \quad U(x+\delta x) - U(x) = \delta x u'(x) = \frac{\delta^2 x^2}{2} u'(x+\theta\delta x) \quad \text{for} \quad 0 < \theta < 1. \]

By assumption,

\[ u'(x+\theta\delta x) \leq A(1+\theta\delta)^{a-2}x^{a-2}L(x+\theta\delta x) \]

\[ \leq A(1+\delta)^{a-2}x^{a-2}L(x+\theta\delta x) \quad \text{for} \quad a \geq 2. \]

Or

\[ \leq A(1-\delta)^{a-2}x^{a-2}L(x+\theta\delta x) \quad \text{for} \quad 0 \leq a \leq 2. \]

Take the upper set of signs in (4-5) and dividing by \( \delta U(x) \), gives

\[ (4-6) \quad \frac{U(x+\delta x)-U(x)}{\delta U(x)} \leq \frac{x u(x)}{U(x)} + \frac{\delta Ax^aL(x+\theta\delta x)(1+\delta)^{a-2}}{2U(x)} \]

where the + sign is taken at the extreme right iff \( a > 2 \). Let \( x \to \infty \) in (4-6) and using the condition on \( U(x) \), we find
\[
\frac{(1+\delta)^a - 1}{\delta} \leq \lim_{x \to \infty} \frac{xu(x)}{U(x)} + A'\delta(1+\delta)^{k-2}.
\]

Then let \( \delta \to 0 \) to find \( a \leq \lim_{k \to \infty} \frac{xu(x)}{U(x)} \). Using the lower set of signs in (4-5) and dividing by \( \delta U(x) \), gives

\[
\frac{xu(x)}{U(x)} \leq \frac{U(x) - U(x-\delta x)}{U(x)} + \frac{\delta Ax(L(x + \delta x)(1+\delta)^{a-2}}{2U(x)}
\]

so that

\[
\lim_{x \to \infty} \frac{xu(x)}{U(x)} \leq \frac{1-(1-\delta)^a}{\delta} + \delta A'(1+\delta)^{a-2}.
\]

Letting \( \delta \to 0 \) yields

\[
\lim_{x \to \infty} \frac{xu(x)}{U(x)} \leq a.
\]

These two inequalities show \( \lim_{x \to \infty} \frac{xu(x)}{U(x)} = a \) which is equivalent to (4-4). \( \Box \)

4.2 Applications

4.2.1 Abel Density

The geometric distribution on the integers assigns mass \( p(1-p)^{k-1} \) to the point \( k \) for \( k = 1, 2, \ldots \). This being the probability that a sequence of coin tosses shows its first head on the \( k \)th trial. This measure has mean \( q/p \), and variance \( (q/p)^2 \). A natural way to make the measure diffuse is to let \( p \to 0 \). For historical reasons, when viewed as a summability method, this is associated with the name of Niels Abel.
Definition. Let \( S = \{s(i)\}_{i=1}^{\infty} \) be real numbers, 
\[
A(S, p) = \sum_{i=1}^{\infty} s(i)(1-p)^{i-1}.
\]
\( S \) has Abel density \( \lambda \) if 
\[
\lim_{p \to 0} A(S, p) = \lambda \text{ as } p \to 0.
\]

For sets of integers \( S \), Abel density and natural density are completely equivalent. A proof of this from the Tauberian theorems above is given in Feller [1966] (pg. 423). We state it for reference.

**Theorem 4.** (Hardy-Littlewood). \( S \subseteq \{1, 2, 3, \ldots\} \) has natural density iff \( S \) has Abel density and then the two limits are equal.

### 4.2.2 Zeta, Log and the Logarithmic Series Density

Zeta and log densities have been introduced in sections 2.1 and 2.2. A third density which arises naturally in this context is the density induced by the logarithmic series distribution. This puts mass \( \frac{\theta^k}{k} \) on \( k = 1, 2, \ldots; \theta < 1 \); \( a = \frac{1}{\log(1-\theta)} \).

For a discussion of statistical properties for \( \theta \) fixed in this range, see Johnson and Kotz [1969], Chapter 7. Letting \( \theta \to 1 \) makes the distribution diffuse.

**Definition.** Let \( A = \{a(i)\}_{i=1}^{\infty} \) be real numbers such that 
\[
\text{LS}(A, \theta) = \frac{1}{\log(1-\theta)} \sum_{i=1}^{\infty} a(i) \frac{\theta^i}{i}
\]
exists for \( 0 < \theta < 1 \). \( A \) has \( \text{LS density} \) \( \lambda \) if \( \text{LS}(A, \theta) \to \lambda \) as
\( \theta + 1 \). Note that the terms of this distribution come from formally integrating the terms of the geometric distribution. Thinking of LS density as integrated Abel can be useful since a routine application of L'hôpital's theorem shows:

**Lemma 1.** Abel density is weaker than LS density. That is, if \( A \subset N \) has Abel density, then \( A \) has equal LS density.

**Proof.** Consider

\[
\lim_{\theta \to 1} \frac{\sum_{i=1}^{\infty} a(i) \theta i}{\log(1-\theta)} = \lim_{\theta \to 1} \frac{\int_{0}^{\theta} \sum_{i=1}^{\infty} a(i)x^i \, dx}{-\log(1-\theta)},
\]

the change of orders of integration being valid since the terms of the series are all positive. Now by L'hôpital's rule, this last limit is equal to \( \lim_{\theta \to 1} (1-\theta) \sum_{i=1}^{\infty} a(i) \theta^i = \lambda \) if \( A \) has Abel density \( \lambda \).

The first new application of the theorems of Section 4.1 is to prove the equivalence of the three densities.

**Theorem 5.** For sets of integers \( A \), the log, zeta, and LS densities are equivalent.

**Proof.** Assume first \( \sum_{i \leq n} a(i) \sim \lambda \log n \) as \( n \to \infty \) where \( a(i) \) is the indicator function of \( A \). Define a measure on the right-hand half line with mass \( \frac{a(i)}{i} \) at the integer \( i \). The distribution function of this measure, \( U(x) \), satisfies
\[ U(x) \sim k \log x \text{ as } x \to \infty. \] The Laplace transform of \( U \) is
\[ \omega(t) = \int_0^\infty e^{-tx} dU(x) = \sum_{i=1}^\infty e^{-ti} \frac{a(i)}{i}, \]
Since \( \log \) is slowly varying, Theorem 1 says
\[ U(x) \sim k \log x \iff \omega(t) \sim k \log \left( \frac{1}{t} \right) \text{ as } x \to \infty, t \to 0. \]
That is,
\[ \sum_{i=1}^\infty e^{-ti} \frac{a(i)}{i} \sim k \log \left( \frac{1}{t} \right), \]
or, letting \( e^{-t} = y \),
\[ \sum_{i=1}^\infty y^i \frac{a(i)}{i} \sim k \log \left( \frac{1}{1-y} \right) \text{ as } y \to 1, \]
iff
\[ \sum_{i \leq x} \frac{a(i)}{i} \sim k \log x \text{ as } x \to \infty, \]
as was to be shown.

For the relations concerning zeta density, define a measure on the numbers \( \log i \) with mass \( \frac{a(i)}{i} \). The Laplace transform is
\[ \omega(t) = \sum_{i=1}^\infty \frac{a(i)}{i} e^{-t \log i} = \sum_{i=1}^\infty \frac{a(i)}{i^{t+1}}. \]
As \( t \to 0 \), \( \omega(t) \sim \frac{k}{t} \) iff \( U(x) = \sum_{\log i \leq x} \frac{a(i)}{i} \sim kx \text{ as } x \to \infty. \)
Let \( s = t+1 \), \( x = \log y \), this becomes
\[ \sum_{i=1}^\infty \frac{a(i)}{i^s} \sim \frac{k}{s-1} \text{ as } s \searrow 1 \iff \sum_{i \leq y} \frac{a(i)}{i} \sim k \log y \text{ as } y \uparrow \infty. \]
That is, A has zeta density \( \lambda \) iff A has log density \( \lambda \). Here we have used the well-known result \( \zeta(s) \sim \frac{1}{s-1} \) as \( s \searrow 1 \) (see Ingham [1932], pg. 21).

4.2.3 Higher Log and Zeta Densities

A straightforward generalization of the densities in Section 4.2.2 above is to higher log and zeta densities.

**Definition.** For \( A = \{a(i)\}_{i=2}^{\infty} \) real numbers, let

\[
LL(A,n) = \frac{1}{\log \log n} \sum_{i=2}^{n} \frac{a(i)}{i \log i}.
\]

A has second log density \( \lambda \) if \( \lim LL(A,n) = \lambda \) as \( n \to \infty \).

Arguments similar to those above show that if A has log density, then A has equal second log density but not conversely. A counter example being \( A = \bigcup_{k=0}^{\infty} [10^{10^k}, 10^{2 \cdot 10^k}] \). This set has second log density \( \log_{10} 2 \) but not log density. A is the set of integers \( i \) such that \( \log i \) has leading digit 1. Theorem 5 shows this set does not have zeta density. The corresponding generalization of zeta density is to the limit of the measure which puts mass \( \frac{1}{\log \zeta(s)} \frac{1}{i (\log i)^5} \) on the number \( a(i) \). Arguments which parallel exactly the arguments in Section 4.2.2 show this second zeta density is equivalent to second log density for sets \( A \subset \mathbb{N} \). These densities are just the first steps in a chain of higher log and zeta densities each of which
properly generalizes the one before. The existence of this chain is interesting in showing there is nothing final in taking the step from natural to zeta density.

4.3 Some Densities Implicit in the Theory of Numbers

A large portion of the work in analytic number theory is concerned with the distribution of prime numbers with various properties. To answer questions of this sort, it is natural to consider densities conditional on being a prime. Since the primes have density zero, a bit of care is called for. The basic notion is again that of natural density.

Definition. For \( A \subseteq \mathbb{N} \), let \( \pi(x) \) be the number of primes \( \leq x \) and \( \pi(A,x) \) be the number of primes \( \leq x \) in \( A \). \( A \) has (conditional) natural density \( \lambda \) if \( \lim_{x \to \infty} \frac{\pi(A,x)}{\pi(x)} = \lambda \).

Most other densities considered above have analogs in the primes, for example:

Definition. \( A \subseteq \mathbb{N} \) has (conditional) zeta density \( \lambda \) if

\[
\lim_{s \to 1} \frac{1}{\log \zeta(s)} \sum_{p \in A} \frac{1}{p^s} = \lambda
\]

where the sums are over primes. It is easily verified that \( \log \zeta(s) \sim \sum_{p} \frac{1}{p^s} \) as \( s \to 1 \), so that \( A \) has conditional zeta density \( \lambda \) iff \( \frac{1}{\log \zeta(s)} \sum_{p \in A} \frac{1}{p^s} \to \lambda \) as \( s \to 1 \).
Using a standard fact: \( \sum_{p \leq x} \frac{1}{p} \sim \log \log x \) (Lemma 2 below), we can define a conditional log density.

**Definition.** A \( \subset N \) has (conditional) log density \( \lambda \) if

\[
\lim_{x \to \infty} \frac{1}{\log \log x} \sum_{p \in A} \frac{1}{p} = \lambda
\]

These densities have found widespread use, frequently without recognition of their being densities. For example, given two relatively prime integers \( a, m \), let \( A \) be the prime \( p \equiv a(m) \). In the first third of the 19th century, Dirichlet proved that there were infinitely many such primes and that in the language above:

**Theorem 6.** (Dirichlet, see e.g. Serre [1973], pg. 73.) A has conditional zeta density \( 1/\phi(m) \) where \( \phi \) is Euler's function.

Thirty or forty years later, Mertens proved:

**Theorem 7.** (Mertens, see Davenport [1967], pg. 61)

\[
\sum_{\substack{p \equiv a(m) \\ p \leq x}} \frac{1}{p} = \frac{1}{\phi(m)} \log \log (x) + c + O\left(\frac{1}{\log x}\right).
\]

This says the primes \( p \equiv a(m) \) have log density \( 1/\phi(m) \). One point of the present discussion is that Mertens' result (without the error term) can be easily deduced from Dirichlet's theorem through a Tauberian argument.
Theorem 8. Conditional log and zeta density are equivalent for subsets of primes. Both densities offer a proper generalization of conditional natural density.

Proof. A word for word repetition of the proof of Theorem 8 above using a measure which puts mass \( \frac{a(p)}{p} \) on points \( \log p \).

It is important to realize that all of the problems with limiting relations encountered for subsets of integers have analogs conditional on being prime. Consider the following quote from Serre [1973]: "One can prove that, if \( A \) has natural density \( k \), the analytic (zeta) density of \( A \) exists and is equal to \( k \). On the other hand, there exist sets having an analytic density but no natural density. It is the case, for example, of the set \( p' \) of prime numbers whose first digit (in the decimal system, say) is equal to 1. One sees easily, using the prime number theorem, that \( p' \) does not have natural density and on the other hand Bombieri has shown me a proof that the analytic density of \( p' \) exists (it is equal to \( \log_{10} 2 = 0.3010300... \)).

We proceed now to give an original proof of this fact in a more general form.

Theorem 9. Let \( f(x) \), \( g(x) \) be polynomials of the same degree, with the same positive leading coefficient:

\[
f(x) = ax^n + bx^{n-1} + ... \quad \text{where } 0 < \frac{c-b}{na} < 1.
g(x) = ax^n + cx^{n-1} + ...
\]
Consider the set \( S = \bigcup_{k=0}^{\infty} [10^f(k), 10^g(k)] \) and let \( A = S \cap P \) where \( P \) is the set of prime integers. Then

a) The set \( A \) does not have conditional natural density.

b) The set \( A \) has conditional log (zeta) density \( \frac{c-b}{na} \).

**Proof.** In both parts it is easiest to use the device of taking limits along the two subsequences \( 10^f(k), 10^g(k) \) where sequence of partial sums achieves minimum and maximum values.

The proof of (a) is very close to the proof of Theorem 2 of Chapter 1, but uses the bounds given by the prime number theorem in the form \( \pi(x) = \frac{x}{\log x} + O\left(\frac{x}{(\log x)^2}\right) \). While details will be omitted, it is of interest to note that the set of limit points of the partial averages is just as in Theorem 2 of Chapter 2: \([0,1]\) for \( \deg f \geq 2 \) and \([\frac{1}{g}, \frac{5}{g}]\) for the first digit set.

For the proof of (b), the average considered is

\[
L(x) = \frac{1}{\log \log x} \sum_{p \leq x} \frac{1}{p}.
\]

The sequence \( L(x) \) is largest at points \( x \) of the form \( x = 10^g(k) \), assuming \( x \) of this form; writing \( 10^z = t(z) \),

\[
(4-7) \quad \sum_{p \leq x} \frac{1}{p} = \sum_{p \in A} \sum_{i=1}^{k} \frac{1}{p} = \sum_{p=t(f(i))}^{t(g(i))} \frac{1}{p}.
\]

Using \( \sum_{p \leq x} \frac{1}{p} = \log \log x + c + O\left(\frac{1}{(\log x)^3}\right) \) to be proved at the finish, (4-7) becomes
\[(4-8) \quad \sum_{i=1}^{k} \{ \log \log 10^g(k) - \log \log 10^f(k) \} + O \left( \sum_{i=1}^{k} \frac{1}{(\log t(f(k))^3} \right) \]

\[= \sum_{i=1}^{k} \{ \log g(i) - \log f(i) \} + O(1) \]

\[= \log \prod_{i=1}^{k} \frac{g(i)}{f(i)} + O(1) \]

We have assumed neither \( f \) nor \( g \) have integral roots. If either does have integral roots, then by arbitrarily small perturbation of the coefficient \( a \), say, the roots can be made non-integral, the proof carried through, and then the argument finished by noting that the answer in part (b) is a continuous function of \( a \).

We thus assume that the complex roots of \( f \) are \( \{ r(j) \}_{j=1}^{n} \) and those of \( g \) are \( \{ s(j) \}_{j=1}^{n} \) counted with their multiplicity, so that

\[ f(X) = a \prod_{i=1}^{n} (X - r(i)) \quad g(X) = a \prod_{i=1}^{n} (X - s(i)) \]

The product in (4-8) can be rewritten as

\[(4-9) \quad \prod_{j=1}^{k} \frac{g(j)}{f(j)} = \prod_{j=1}^{n} \frac{p_j(k)}{r(j)} \quad \text{for} \quad p_j(k) = \prod_{m=1}^{m-s(j)} \frac{m-s(j)}{m-r(j)} \]

Recall that the Gamma function can be written:

\[ \Gamma(z) = \lim_{n \to \infty} \frac{n^zn!}{(1+x)(2+x)\ldots(n+x)} \quad \text{(Bromwich [1931], pg. 112).} \]

Using this, since the products \( p_j(k) \) can be written as
\[ p_j(k) = \frac{k^{-r(j)}k! k^{r(j)}}{(1-r(j))(2-r(j)) \ldots (k-r(j))} \times \frac{k^{-s(j)}k! k^{s(j)}}{(1-s(j))(2-s(j)) \ldots (k-s(j))} \]

we see \( p_j(k) \sim k^{r(j)}s(j)c(j) \) as \( k \to \infty \) where \( c(j) = \frac{\Gamma(-r(j))}{\Gamma(-s(j))} \).

Using this in (4-9) and recalling \( \sum_{j=1}^{n} r(j) = -\frac{b}{a}, \quad \sum_{j=1}^{n} -s(j) = -\frac{c}{a} \),
yields:

\[ \frac{k}{a} \sum_{i=1}^{q(i)} \frac{c-b}{f(i)} \sim \frac{c-b}{a} \log k + o(\log k) \quad \text{(4-10)} \]

for \( d \) a constant, thus

\[ \log \frac{k}{a} \sum_{i=1}^{q(i)} \frac{c-b}{f(i)} = \frac{c-b}{a} \log k + o(\log k) \cdot \]

Finally,

\[ \log \log 10^g(k) = \log(\log 10 \ g(k)) - n \log k \]

as \( k \to \infty \), so that

\[ \lim_{y \to \infty} \lim_{k \to \infty} L(y) = \lim_{y \to \infty} L(t(g(k))) = \frac{c-b}{na}. \]

The same ingredients combine to give the same limit along points \( x \) of the form \( 10^f(k) \) so that \( \lim_{y \to \infty} L(y) = \frac{c-b}{na} \).

It remains only to prove the refined estimate

\[ \sum_{p \leq x} \frac{1}{p} = \log \log x + c + O\left(\frac{1}{(\log x)^3}\right). \]

Note this is better than the usual estimate as in Hardy and Wright [1960] (pg. 427). The following proof is based on a
strong form of the prime number theorem (for a related proof, see Prachar [1957] pg. 71).

**Lemma 2.** \( \sum_{p \leq x} \frac{1}{p} = \log \log x + B_1 + O(R(x)) \) where \( R(x) = \frac{1}{(\log x)^d} \) for any large \( d \).

**Proof.** \( \sum_{p \leq x} \frac{1}{p} = \int_1^x \frac{1}{x} d\pi(x) = \frac{\pi(x)}{x} + \int_1^x \frac{\pi(t)}{t^2} dt \). Now \( \pi(x) = \xi(x) + O(xe^{-a(\log x)^{\frac{3}{2}}}) \)

(Ingham [1932], pg. 65). So

(4-11) \( \sum_{p \leq x} \frac{1}{p} = \frac{\xi(x)}{x} + \int_1^x \frac{\xi(t)}{t^2} dt + O(\exp(-a(\log x)^{\frac{3}{2}})) \)

\[ + \left( \int_1^x \frac{R_1(t)}{t^2} dt \right) \]

where \( R_1(t) \) is \( \pi(t) - \xi(t) \). Now,

\[ \int_1^x \frac{R_1(t)}{t^2} dt = \int_1^\infty \frac{R_1(t)}{t^2} dt - \int_x^\infty \frac{R_1(t)}{t^2} dt = c - \int_x^\infty \frac{R_1(t)}{t^2} dt. \]

Further,

\[ \frac{\xi(t)}{x} + \int_1^x \frac{\xi(t)}{t^2} dt = \int_1^x \frac{1}{t} d\xi(t) = \int_1^x \frac{1}{t \log t} dt = \log \log x. \]

Putting this in (4-7) gives

\[ \sum_{p \leq x} \frac{1}{p} = \log \log x + c + O\left( \int_x^\infty \frac{R_1(t)}{t^2} dt \right). \]
Finally, the error is easily seen to be

$$0((\exp - a(\log x)^{1/2})(\log x)^{1/2})$$

by L'hôpital's rule and this easily gives the statement of the lemma.

A classical elementary result in number theory is the analytical fact:

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1)$$

(Hardy and Wright [1960], pg. 348), which can be translated as: The numbers \(a(i) = \log i\) if \(i\) is prime, 0 otherwise, have log density 1. To see that the passage from log to natural density is non-trivial, note that the statement that the \(a(i)\) have natural density 1 is equivalent to the prime number theorem. Gelfond and Linnik [1965] (pg. 52), in their rendition of Selberg's 'elementary' proof of Dirichlet's theorem, prove:

**Theorem 10.** \(\sum_{\substack{p \leq x \leq \phi(m) \equiv a(m) \equiv 0}} \frac{\log p}{p} = \frac{\log x}{\phi(m)} + O(1)\).

Looking at these results suggests investigation of a new density.

**Definition.** For \(A = \{a(i)\}_{i=1}^{\infty}\), let \(N(x) = \sum_{i \leq x} \frac{\log i}{i}\), and

$$L'(x) = \frac{1}{N(x)} \sum_{i=1}^{x} \frac{a(i) \log i}{i}.$$
A has \( L' \) density \( \ell \) if \( \lim_{x \to \infty} L'(A,x) = \ell \). Note: An integration by parts and use of \( \sum_{i \leq x} \frac{1}{i} = \log x + \gamma + O\left(\frac{1}{x}\right) \) gives

\[
N(x) = \sum_{i \leq x} \frac{\log i}{i} = \sum_{i \leq x} \frac{1}{i} \log x - \int_{1}^{x} \frac{\log t + \gamma + O\left(\frac{1}{t}\right)}{t} \, dt
\]

\[
= \frac{1}{2} (\log x)^2 + c + O\left(\frac{\log x}{x}\right)
\]

for some constant \( c \). Thus \( A \) has \( L' \) density \( \ell \) iff

\[
\frac{2}{\log^2 x} \sum_{i \leq x} \frac{a(i) \log i}{i} \to \ell \quad \text{as} \quad x \to \infty.
\]

**Example 1.** The two theorems immediately above say that the primes \( p = a(m) \) have conditional \( L' \) density \( 1/\phi(m) \). We will show that this is a stronger statement than Mertens theorem.

Investigation of the \( L' \) density reveals it is equivalent to what may be called a differentiated zeta density and that both of these densities lie properly between natural and log density. The details follow.

Recall \( A \) has zeta density if \( \sum_{i=1}^{\infty} \frac{a(i)}{i^s} \sim \frac{\ell}{(s-1)} \) as \( s \to 1^+ \). Formal differentiation of both sides of this relation leads to:

**Definition.** \( A \) has \( Z' \) density \( \ell \) if \( \sum_{i=1}^{\infty} \frac{a(i) \log i}{i^s} \sim \frac{\ell}{(s-1)^2} \) as \( s \to 1^+ \).

**Proof.** \( Z' \) density means \( \sum_{i=1}^{\infty} \frac{a(i) \log i}{ii^{s-1}} \sim \frac{\ell}{(s-1)^2} \) as \( s \to 1^+ \).
or that \( \sum_{i=1}^{\infty} \frac{a(i) \log i}{it} \sim \frac{\lambda}{t^2} \) as \( t \to 0 \). Define a measure on the positive reals with mass \( \frac{a(i)}{i} \log i \) at \( \log i \). This has Laplace transform

\[
\omega(t) = \sum_{i=1}^{\infty} \frac{a(i) \log i}{i} e^{-t \log i} = \sum_{i=1}^{\infty} \frac{a(i) \log i}{it}
\]

and distribution function

\[
U(t) = \sum_{\log i \leq t} \frac{a(i) \log i}{i}.
\]

Theorem 1 of this chapter says \( \omega(\tau) \sim \frac{\lambda}{\tau^2} \) as \( \tau \to 0 \) iff \( U(t) \sim \frac{\lambda}{2} t^2 \) as \( t \to \infty \). Let \( t = \log x, \tau = s-1 \), and this becomes

\[
\sum_{i=1}^{\infty} \frac{a(i) \log i}{i^s} \sim \frac{\lambda}{(s-1)^2} \quad \text{iff} \quad \sum_{i \leq x} \frac{a(i) \log i}{i} \sim \frac{\lambda \log^2(x)}{2}.
\]

**Corollary 1.** Both \( Z' \) and \( L' \) densities are weaker than log density.

**Proof.** Since \( Z' \) is differentiated zeta density, \( Z' \) is weaker than zeta density by L'hôpital's rule. Since \( \log \) is equivalent to zeta density after Theorem 6 of Chapter 2, the result follows. |||

**Note.** In Section 4.2.2 LS density was defined and shown to be equivalent to \( \log \) and zeta densities. There is a modification of LS density, namely,
\[
\sum_{i=1}^{\infty} \frac{\theta^i \log \frac{i}{\theta}}{i} a(i) \sim \frac{\log^2(1-\theta)}{2} \quad \text{as } \theta \to 1^{-}
\]

which may be shown equivalent to \(Z\)' and \(L\)' densities by a similar application of the Tauberian theorems.

The next result shows these new densities are strictly stronger than natural density.

**Theorem 12.** a) If real numbers \(\{a(i)\}_{i=1}^{\infty}\) have natural density \(\lambda\), then

\[
\lim_{n \to \infty} \frac{2}{(\log n)^2} \cdot \sum_{i=1}^{n} \frac{a(i) \log i}{i} = \lambda.
\]

b) The inclusion is strict, for the integers with lead digit 1 have \(L\)' density \(\log_{10} 2\) but not natural density.

**Proof.** a) Consider the sum

\[
\sum_{i=1}^{n} \frac{a(i) \log i}{i} = A(n) \frac{\log n}{n} + \sum_{i=1}^{n-1} A(i) \Delta(i)
\]

where \(A(j) = \sum_{i=j}^{n} a(j)\), \(\Delta(j) = \frac{\log j}{j} - \frac{\log (j+1)}{j+1}\), since \(\frac{\log x}{x}\) is decreasing for \(x \geq 1\), \(\Delta(j) \geq 0\) for \(j \geq 3\). Assume \(A(j) = 2j + o(j)\), then

\[
\sum_{i=1}^{n} \frac{a(i) \log i}{i} = 0(\log n) + \lambda \sum_{i=1}^{n-1} i \Delta(i) + o\left(\sum_{i=1}^{n-1} i \Delta(i)\right)
\]

\[
= 0(\log n) + \lambda \sum_{i=1}^{n} \frac{\log i}{i} + o\left(\sum_{i=1}^{n-1} \frac{\log i}{i}\right)
\]

\[
= \frac{\lambda}{2}(\log n)^2 + o((\log n)^2)
\]

so the \(a(i)\) have \(L\)' density \(\lambda\).
b) If \( a(i) \) is the indicator function of the set
\[
A = \bigcup_{k=0}^{\infty} [10^k, 2 \cdot 10^k],
\]
the relation
\[
\sum_{i=1}^{n} \frac{\log i}{i} = (\log n) \frac{1}{2} + c + O\left(\frac{\log n}{n}\right)
\]
proved in this section yields
\[
\sum_{i=a}^{b} \frac{\log i}{i} = \frac{1}{2}(\log^2 b - \log^2 a) + O\left(\frac{\log a}{a}\right).
\]
In considering the sum
\[
L'(n) = \frac{2}{\log^2 n} \sum_{i=1}^{n} \frac{a(i) \log i}{i}
\]
note that it rises and falls periodically achieving maximal values at \( n = 2 \cdot 10^k \) and minimal values at \( n = 10^k \).

\[
(4-12) \sum_{i \leq 2 \cdot 10^k} \frac{a(i) \log i}{i} = \sum_{i=1}^{k} \frac{210_i}{i} \sum_{j=10^i}^{210_i} \frac{\log j}{j}
\]

\[
= \sum_{i=1}^{k} \frac{1}{2}(\log^2 (2 \cdot 10^i) - \log^2 (10^i)) + O\left(\frac{\log 10^i}{10^i}\right).
\]
The sum in the error is \( O(k) \) while the term in curly brackets
is \((\log 2)^2 + 2 \log 2 \log 10 \) i so that (4-12) equals
\[
\log 2 \log 10 \frac{k^2}{2} + O(k).
\]
Dividing by \( \frac{\log 2^{10^k}}{2} = \frac{(\log 10^k)^2}{2} \), yields
\[
\lim_{n \to \infty} E'(n) \leq \frac{\log 2}{\log 10} = \log_{10} 2.
\]

A very similar computation leads to \(\log_{10} 2 \leq \lim_{n \to \infty} L'(n)\) which proves the set has \(L'\) density \(\log_{10} 2\). Since \(A\) is a 'log-linear' set, Theorem 2 of Chapter 1 determined that \(A\) didn't have natural density proving (b) of the theorem. 

Remark. It is possible to define a family of densities by considering sums of the form \(\frac{1-b}{n^{1-b}} \sum_{i=1}^{n} a(i)\) where \(0 \leq b \leq 1\).

The limiting cases given log density \((b=1)\) and natural density \((b=0)\). It is not hard to show that for each \(b\), the density is stronger than natural density. However, the only \(b\) such that the first digit set has a density is \(b = 1\). The \(L'\) is easily seen stronger than the members of this family for \(0 \leq b < 1\) and so represents a subtle turning point.

A computation much like the one used to prove part (b) of Theorem 12 shows that the \(L'\) density exists for any set with polynomial gaps, agreeing with the zeta density determined for such sets in Theorem 2 of Chapter 2.

4.4 Remarks on Differentiating and Integrating in a Parameter

It is worthwhile summarizing some of the results proved in this chapter from the point of view of formal integration or differentiation of asymptotic relations. There are two obvious rules of thumb:

A) Integrating a density relation in a parameter makes a stronger density. This follows from \(L'\) hospital's rule under regularity conditions.
Example 2. Formally integrating the limiting density relationship for Abel density

\[ \sum_{i=1}^{\infty} a(i)x^i - \frac{\xi}{1-x} \quad \text{as} \quad x \to 1 \]

leads to considering

\[ \sum_{i=1}^{\infty} a(i) \frac{x^i}{i} - \xi \log \frac{1}{1-x} \quad \text{as} \quad x \to 1. \]

This new density was shown equivalent to log density in Theorem 5 of this chapter.

B) Differentiating a density in a parameter makes a weaker density.

Example 3. Differentiating the limiting relation for zeta density

\[ \sum_{i=1}^{\infty} \frac{a(i)}{i^s} - \frac{\xi}{s-1} \quad \text{as} \quad s \to 1 \]

leads to

\[ \sum_{i=1}^{\infty} \frac{a(i)\log(i)}{i^s} - \frac{\xi}{(s-1)^2} \quad \text{as} \quad s \to 1. \]

Caution is called for in interchanging equivalent densities. For example, since integrated Abel density is equivalent to zeta density (Theorem 5), it might be supposed that Abel density was equivalent to differentiated zeta density. This is certainly false (Theorems 11 and 12).
Example 4. The negative binomial distribution on the positive integers puts mass \( \binom{r+k-1}{k} p^r (1-p)^{k-r} \) on the integer \( k \), with \( p \in (0,1) \), \( r \in \mathbb{N} \). It is natural to make this diffuse by letting \( p \to 0 \). For \( r = 1 \), this is Abel density and for \( r > 1 \), it becomes Abel density differentiated \( r-1 \) times. Thus by the general remarks above, it follows that as \( r \) increases this is a sequence of weaker and weaker densities.
CHAPTER 5

INFINITE REAVERAGING

To deal with the problem as non-existence of natural density in the first digit problem, B. J. Flehinger [1966] has suggested the following: for \( A \subset \mathbb{N} \) form \( \delta(A, n, 1) \) as \( \frac{1}{n} \sum_{i=1}^{n} a(i) \). If \( \lim \delta(A, n, 1) \) exists as \( n \to \infty \), there is no problem. If not, form \( \delta(A, n, 2) = \frac{1}{n} \sum_{i=1}^{n} \delta(A, n, 1) \), trying to "average out the anomalous behavior". Flehinger succeeded in showing that if this reaveraging procedure is repeated infinitely often, it settles down to the right answer in the first digit problem. This leads to considering the infinite iteration of a general summability method for general sequences \( a(i) \). The work presented here does not go that far, but we do extend Flehinger's work in several directions giving a detailed analysis of infinite iteration of natural density (\( H^\infty \) density).

We determine for what kinds of sets this process exists and the relative place of \( H^\infty \) density in the class of methods we work with. A similar analysis is carried through in detail for infinite iterations of log averages.

5.1 \( H^\infty \) Averages

With notation as above, it is easily checked that

\[
\lim_{n \to \infty} \delta(A, n, i) \geq \lim_{n \to \infty} \delta(A, n, i-1)
\]

and, similarly, that the upper limits are decreasing in \( i \). This leads to the following
definition:

**Definition.** A has $H^\infty$ density if

$$\lambda = \lim_{i \to \infty} \lim_{n \to \infty} \delta(A, i, n) = \lim_{i \to \infty} \lim_{n \to \infty} \delta(A, i, n).$$

A recurrent theme in the subject is that it is actually necessary to iterate infinitely often. If the reaveraging process settles down after any finite number of iterations, then the set in question had natural density.

**Lemma 1.** If $A \subset N$, $\delta(A, n, i)$ is as defined above for $i, n = 1, 2, \ldots$, if $\lim \delta(a, n, i) \to \lambda$ as $n \to \infty$ for some $i$, then $\lim \delta(A, n, i) \to \infty$ as $n \to \infty$ for every $i$.

**Proof.** Assuming Cauchy's theorem: $C(i) \to \lambda$, then

$$\frac{1}{n} \sum C(i) \to \lambda,$$

we need only show $\delta(A, n, 1) \to \lambda$ to show all averages exist and are equal. Inductively, we need only show $\delta(A, n, k) \to \lambda$ gives us $\delta(A, n, k-1) \to \lambda$ as $n \to \infty$. The neatest way to do this is to invoke the "corrected converse" of Cauchy's theorem:

**Theorem 1.** (Bromwich [1931], pg. 423). $b(n)$ real, $n(b(n) - b(n+1)) < M$, then $C(n) = \frac{1}{n}(b(1) + \ldots + b(n)) \to \lambda$ implies $b(n) \to \lambda$.

We need only verify the Tauberian condition. In our example, $b(i) = \frac{1}{i}(a(1) + \ldots + a(i))$ for some $a(i)$, so
\[ n(b(n) - b(n+1)) = n(a(1)\left(\frac{1}{n} - \frac{1}{n+1}\right) + \ldots + a(n)\left(\frac{1}{n} - \frac{1}{n+1}\right) - \frac{a(n+1)}{n+1}) \]

\[ = n\left(\sum_{i=1}^{n} \frac{a(i)}{n(n+1)} - \frac{a(n+1)}{n+1}\right) \]

\[ = \frac{1}{n+1} \sum a(i) - \frac{n}{n+1} a(n+1). \]

In our example, the \( a(i) \) are \( 0 \leq a(i) \leq 1 \) so the quantity above is bounded and we are done. \[ ||| \]

**Remarks.** Several alternate proofs of the above theorem depend on the Tauberian theorem, e.g., Hardy [1949], Theorems 55 and 92. The proof of Lemma 1 shows the need for the type of improvement presented in the section on density Tauberian theorems (Section 4.1). The integral version of Lemma 1 would be:

\[ \int_{1}^{X} f(t)dt = o(x) \quad \text{and} \quad xf'(x) < k \implies f(x) = o(1). \]

The classical density version (as in Theorem 2, Chapter 4) asserts the conclusion under the stronger constraint that \( f \) is monotone, i.e., that \( f' < 0 \). In terms of \( a(i) \), that \( a(i) \) is monotone decreasing. This would not be satisfied by \( a(i) = 0 \) or 1 as \( i \in A \) or not and so the improved result is needed.

Lemma 1 shows that if a set has density, then it has equal \( H^\infty \) density. As the next step in classification, we show that \( H^\infty \) is weaker than log density.

**Theorem 2.** \( H^\infty \) is weaker than \( L' \) which is weaker than log density.
Proof. Recall the $L'$ density, defined in Section 4.3, is the limiting value of \( \frac{2}{\log^2 x} \sum \frac{a(i) \log i}{i} \). Corollary 1 of Chapter 4 showed $L'$ was weaker than log density, so it remains to show the first inclusion. Let $a(i)$ be any bounded sequence, consider

\[
(5-1) \sum_{i=M}^{N} \frac{a(i) \log i}{i} = \sum_{i=M}^{N} \frac{A(i) \log i}{i} + O(\log N),
\]

where $A(i) = \sum_{j=1}^{i} a(j)$. To prove (5-1), write

\[
\sum_{i=M}^{N} \frac{a(i) \log i}{i} = \sum_{i=M}^{N} A(i) \frac{\log i}{i} - \frac{\log(i+1)}{i+1} + O(\log N)
\]

\[
= \sum_{i=M}^{N} \frac{A(i)}{i} \left( \log i - i \log(1 + \frac{1}{i}) \right) + O(\log N).
\]

Since $i \log(1 + \frac{1}{i}) = O(1)$, this last quantity

\[
= \sum_{i=M}^{N} \frac{A(i)}{i} \log i + O(\log n) + O(\log N)
\]

\[
= \sum_{i=M}^{N} \frac{A(i)}{i} \log i + \sum_{i=M}^{N} A(i) \log(i) \left( \frac{1}{1+i} - \frac{1}{i} \right) + O(\log n)
\]

\[
= \sum_{i=M}^{N} \frac{A(i)}{i} \log i + O(\log N).
\]

Inductively, from (5-1) for each fixed $k,$
\[
(5-2) \quad \sum_{i=M}^{N} \frac{a(i) \log i}{i} = \sum_{i=M}^{N} \delta(A, i, k) \frac{\log i}{i} + O_k(\log N)
\]

\[
\geq \inf_{i \geq M} \delta(A, i, k) \sum_{i=M}^{N} \frac{\log i}{i} + O_k(\log N)
\]

\[
= \inf_{i \geq M} \delta(A, i, k) \left\{ \frac{\log^2 N}{2} + \frac{\log^2 M}{2} + O\left(\frac{\log M}{M}\right) \right\}
\]

\[+ O_k(\log N).\]

Divide both sides of (5-2) by \(\frac{\log^2 N}{2}\) and let \(N \to \infty\) to get:

\[
\lim_{N \to \infty} \frac{2}{\log^2 N} \sum_{i=1}^{N} \frac{a(i) \log i}{i} = \lim_{N \to \infty} \frac{2}{\log^2 N} \sum_{i=M}^{N} \frac{a(i) \log i}{i}
\]

\[
\geq \inf_{i \geq M} \delta(A, i, k).
\]

Finally, let \(M \to \infty\) to get

\[
\lim_{N \to \infty} \frac{2}{\log^2 N} \sum_{i=1}^{N} \frac{a(i) \log i}{i} \geq \lim_{i \to \infty} \delta(A, i, k).
\]

A similar relation is true for the upper limit, so that

\[
\ldots \lim_{n \to \infty} \delta(A, n, k) \leq \lim_{n \to \infty} \delta(A, n, k+1) \ldots \leq \lim_{n \to \infty} \frac{2}{\log^2 n} \sum_{i=1}^{n} \frac{a(i) \log i}{i}
\]

\[
\leq \lim_{n \to \infty} \frac{2}{\log^2 n} \sum_{i=1}^{n} \frac{a(i) \log i}{i} \leq \ldots
\]

\[
\leq \lim_{i \to \infty} \delta(A, i, k) \leq \ldots
\]

which proves the theorem. |||
The next theorem shows that $H^\infty$ density gives density to any log linear set $S = \bigcup_{k=0}^{\infty} [10^{ak+b}, 10^{ak+c})$. The basic idea of the proof uses ideas similar to those in Flehinger [1966]. Since results are done in added generality and since an incorrect version of Flehinger's result has appeared in print (see Knuth, Vol. 2 [1971], pg. 219-228, and the remarks following Theorem 5), it seems proper to include details.

In what follows, $S = \bigcup_{k=0}^{\infty} [10^{ak+b}, 10^{ak+c})$ and $\delta(i,0) = \delta(S,i,0) = 1$ or 0 as $i \in S$ or not, $\delta(x,k) = \frac{1}{x} \sum_{i \leq x} \delta(i,k-1)$ for $k = 1, 2, \ldots$.

**Lemma 2.** For all $k \geq 1$ and $\varepsilon > 0$, there exist functions $Q_k(s)$, $R_k(s)$, and an integer $N(k,\varepsilon)$ such that for $n > N(k,\varepsilon)$ and $1 \leq s \leq 10^a$,

$$|\delta(s)Q_{10^an+b,k} - Q_k(s)| < \varepsilon \quad \text{if} \quad s \leq 10^{c-b}$$

$$|\delta(s10^an+b,k) - (Q_k(s) + R_k(s))| < \varepsilon \quad \text{for} \quad 10^{c-b} \leq s \leq 10^a,$$

The functions $Q_k$, $R_k$ satisfy the recursive relations:

$$Q_k(s) = \frac{1}{s} \left( \frac{1}{\alpha} \int_{1}^{g} Q_{k-1}(y)dy + \int_{1}^{s} Q_{k-1}(y)dy + \frac{1}{\alpha} \int_{r}^{g} R_{k-1}(y)dy \right)$$

$$R_k(s) = \frac{1}{s} \int_{r}^{s} R_{k-1}(y)dy, \quad Q_0(s) \equiv 1, \quad R_0(s) \equiv -1,$$

where $g = 10^a$, $r = 10^{c-b}$, $\alpha = 10^a - 1$.

**Remark.** This shows that the partial averages $\delta(x,k)$ oscillate with period $10^a$ as $x \to \infty$. Calling
\[ S_k(y) = \begin{cases} 
    Q_k(y) & y \leq r, \\
    Q_k(y) + R_k(y) & y > r,
\end{cases} \]

this shows that \( \delta(x,k) \) can be uniformly well approximated by the relatively simple function \( S_k(x) \) which only depends on where in the period \((1,10^a)\) \( x \) lies. Note that \( S_k \) satisfies the recursive relation

\[ S_k(y) = \frac{1}{y} \left( \frac{1}{\alpha} \int_1^y S_{k-1}(x) \, dx + \int_1^y S_{k-1}(x) \, dx \right). \]

To simplify notation, we write \( 10^x \) as \( t(x) \).

**Proof.** We use induction on \( k \). For \( k = 1 \), let \( J = \text{st}(am+b) \) for some integer \( m \) and \( 1 \leq s \leq t(a) \).

If \( 1 \leq s \leq t(c-b) \), then

\[ \sum_{i \leq J} \delta(i,0) = (t(c)-t(b)) + (t(a+c)-t(a+b)) + \ldots \]

\[ + (t(m-1)a+c) - t((m-1)a+b) + \text{st}(am+b) - t(am+b) \]

\[ = \frac{t(c)-t(b)}{10^a-1} \left( t(ma) - 1 + \text{st}(am+b) - t(am+b) \right) \]

so that

\[ \delta(J,1) = 1 - \frac{1}{s} \left( \frac{t(a)-t(c-b)}{t(a)-1} \right) + 0(10^{-m}) = 1 - \frac{1}{s} \left( \frac{a-r}{\alpha} \right) + 0(10^{-m}). \]

For \( 10^{c-b} \leq s \leq 10^a \), it follows similarly that
\[ \delta(J,1) = \frac{1}{s} \left\{ t(c) - t(b) \right\} t(a-b) + O(10^{-m}) \]

\[ = 1 - \frac{1}{s} \frac{g-r}{\alpha} + \frac{1}{s}(r-s) + O(10^{-m}). \]

Computing from the recursive definitions in the statement of the lemma,

\[ Q_1(s) = \frac{1}{s} \left\{ \frac{1}{\alpha}(g-1) + (s-1) - \frac{1}{\alpha}(g-r) \right\} = \frac{1}{s} \left\{ (s-1) + \frac{r-1}{\alpha} \right\} \]

\[ = 1 - \frac{1}{s} \frac{g-r}{\alpha}, \]

\[ R_1(s) = \frac{1}{s}(r-s). \]

The computation above shows

\[ |\delta(st(\alpha m+b),1) - S_1(s)| = O(10^{-m}) \]

which is more than asserted by the lemma for \( k = 1. \)

For \( k > 1, \)

\[ \delta(st(\alpha m+b),k) = \frac{1}{st(\alpha m+b)} \sum_{i=0}^{st(\alpha m+b)} \delta(i,k-1) \]

(5-3)

\[ = \frac{1}{s} \left\{ \sum_{0 \leq j < m} \frac{1}{t(a(m-j))} \sum_{i=t(aj+b)}^{t(a(j+1)+b)} \frac{\delta(i,k-1)}{t(a_j+b)} \right\} + O(t(-m)) \]

where the error represents neglected terms less than \( t(b). \)

The inner sums in (5-3) can be well approximated using the inductive step. That is, the difference
(5-4) \[
\left| \sum_{i=t(a_j+b)}^{t(i, k-1)} \frac{1}{t(a_j+b)} - \sum_{i=t(a_j+b)}^{t(1)} \frac{1}{t(a_j+b)} S_{k-1}(\frac{1}{t(a_j+b)}) \right| \\
\leq (q-1)\varepsilon
\]

for any \(1 \leq q \leq t(a)\) if \(j > N(k-1, \varepsilon)\). Further, by the definition of integration as a limit of Riemann sums,

(5-5) \[
\left| \sum_{i=t(a_j+b)}^{t(j, k-1)} S_{k-1}(\frac{1}{t(a_j+b)}) - \int_1^q S_{k-1}(y)dy \right| < \varepsilon
\]

for all \(j\) sufficiently large. Combining (5-3), (5-4), and (5-5) gives

(5-6) \[
\left| \delta(st(a(m+b)), k) - \frac{1}{s} \left( \sum_{0 \leq j < m} \frac{1}{t(a(m-j))} \right) \int_1^q S_{k-1}(x)dx \\
+ \int_1^s S_{k-1}(x)dx \right| \\
\leq \sum_{0 \leq j < N} \frac{M}{t(a(m-j))} + \sum_{N \leq j} \frac{qe}{t(m-j)} + ge
\]

where \(M\) is an upper bound for the expressions in absolute value on the left side of (5-4) and (5-5) which holds for all \(j\) and \(N\) is chosen so large that inequalities (5-4) and (5-5) hold for all \(j\) past \(N\). Note as \(m \to \infty\), the bound on the right side of (5-6) becomes \(2ge\). Finally,

\[
\sum_{0 \leq j < m} \frac{1}{t(a(m-j))} = \frac{1}{t(a)} \frac{t(-ma)-1}{t(-a)-1} = \frac{1}{t(a)-1} + O(t(-m))
\]
so that

\[ |\delta(st(am+b),k) = \frac{1}{s}(t(a)-1) \left( \int_1^g S_{k-1}(x)dx + \int_1^S S_{k-1}(x)dx \right) | < \varepsilon \]

for \( m \) sufficiently large. Comparing this with the recursive definition of \( S_k \), we have the proof of the lemma. |||

**Corollary 1.** The set of limit points of \( \delta(A,i,1) \) as \( i \to \infty \), form the interval

\[ \left[ \frac{10^{c-b}-1}{10^a-1}, \frac{10^a+(b-c)(10^{c-b}-1)}{10^a-1} \right] . \]

Lemma 2 provides the basic tools needed to prove the final result.

**Theorem 3.** The set \( S \) has \( H^\infty \) density \( \frac{c-b}{a} \).

**Proof.** We show:

(5-7) \quad \{ Q_m(s) + \frac{c-b}{a} \} \quad \text{as} \quad m \to \infty \quad \text{uniformly in} \quad s \in [1,10^a].

(5-8) \quad R_m(s) \to 0

This shows \( S_m(s) + \frac{c-b}{a} \) uniformly in \( s \). To see this implies the theorem, choose \( M \) so large that for \( m > M \),

\[ \sup_{1 \leq s \leq 10^a} |S_m(s) - \frac{c-b}{a}| < \varepsilon . \]

By Lemma 2,

\[ \frac{c-b}{a} - \varepsilon \leq \inf_{s} S_m(s) \leq \lim_{x \to \infty} \delta(x,m) + \varepsilon \leq \sup_{s} S_m(s) + 2\varepsilon \leq \frac{c-b}{a} + 3\varepsilon . \]
for all $m$ past some point, so clearly $\lim_{k \to \infty} \lim_{x \to \infty} \delta(x, k) = \frac{c - b}{a}$.

Similarly, $\lim_{k \to \infty} \lim_{x \to \infty} \delta(x, k) = \frac{c - b}{a}$ so that (5-3) and (5-4) prove Theorem 3.

To prove (5-4), consider the sequence $R_m(s)$:

$$R_0(s) = -1, \quad R_1(s) = -1 + \frac{r}{s}, \quad R_2(s) = -1 + \frac{r}{s}(1 + \log \frac{s}{r})$$

and in general

$$R_m(s) = -1 + \frac{r}{s}(1 + \frac{1}{1!} \log \frac{s}{r} + \frac{1}{2!} \log^2 \frac{s}{r} + \ldots$$

$$+ \frac{1}{(m-1)!} \log^{m-1} \frac{s}{r}).$$

Since $s$ is in a compact set, this converges uniformly to $-1 + \frac{r}{s} \exp(\log \frac{s}{r}) = 0$, proving (5-4).

To prove (5-3), write the recursion for $Q_m(s)$ as

$$Q_m(s) = \frac{1}{s} \left( c_m + 1 + \int_1^s Q_{m-1}(x) dx \right)$$

where

$$\delta_m = \frac{1}{\alpha} \left\{ \int_1^g Q_{m-1}(x) dx + \int_r^g R_{m-1}(x) dx \right\} - 1.$$

The reason for this choice is that

(5-8) $Q_m(s) = 1 + \frac{1}{s} \left( c_m + \frac{1}{1!} c_{m-1} \log s + \frac{c_{m-2}}{2!} (\log s)^2 + \ldots$$

$$+ \frac{c_{1}}{(m-1)!} (\log s)^{m-1} \right).$
This is true for \( m = 1 \) since
\[
Q_1(s) = 1 + \frac{1}{s} \left\{ \left( \int_1^g Q_{m-1} + \int_r^g Q_{m-1} \right) - 1 \right\}
\]
is true by definition. The general result follows by an easy induction. We will show
\[
(5-10) \quad c_m + (\log g r) - 1 \quad \text{as} \quad m \to \infty.
\]
The \( \log \) is to base \( g \).

One way to see that (5-5) and (5-6) combine to give (5-4) is to consider a sequence \( b_m + \lambda \) as \( m \to \infty \). We claim that
\[
\sum_{i=0}^{m} b_{m-i} \frac{x^i}{i!} + \lambda e^x
\]
uniformly for \( x \in C \), a compact set. To see this, let \( A = \sup_{x \in C} e|x| \),
\( B = \sup_m |b_{m-2}| \). Choose \( N \) so large that for \( m > N \), both
\[
|b_{m-2}| < \frac{\varepsilon}{A} \quad \text{and} \quad \sum_{i=m}^{\infty} \frac{|x|^i}{i!} < \min(\frac{\varepsilon}{B}, \frac{\varepsilon}{2})
\]
uniformly in \( C \). Then, for \( m > 2N \),
\[
\left| \sum_{i=0}^{m} b_{m-i} \frac{x^i}{i!} + \lambda e^x \right| \leq \frac{\varepsilon}{A} e|x| + \frac{\varepsilon}{B} \sum_{i>N} \frac{|x|^i}{i!} + |\lambda| \sum_{i>m} \frac{|x|^i}{i!} \leq 3\varepsilon
\]
uniformly for \( x \in C \). In the situation of the theorem,
\( c_m + (\log g r) - 1, x = \log s \), so the computation shows
\[
\sum_{i=0}^{m} c_{m-i} \frac{x^i}{i!} + ((\log g r) - 1)e^{\log s} = s((\log g r) - 1) = s\left(\frac{c-b}{a} - 1\right).
\]
Putting this in (5-5) gives (5-4).

It remains to prove (5-6). From the recursion

\[(5-11) \quad c_m = \frac{1}{\alpha} \left\{ \int_1^g Q_{m-1}(x) \, dx + \int_r^g R_{m-1}(x) \, dx \right\} - 1, \]

calculate:

\[c_1 = \frac{r-1}{\alpha} - 1 = \frac{r-g}{\alpha} \]

and inductively

\[c_{m+1} = \frac{1}{\alpha} \left\{ c_m \log g + c_{m-1} \frac{(\log g)^2}{2!} + \ldots + \frac{c_1 (\log g)^m}{m!} \right\} \]

\[+ r \left\{ 1 + \log \frac{g}{r} + \ldots + \frac{(\log \frac{g}{r})^m}{m!} \right\} - g \, . \]

Let \( c(z) = c_1 z + c_2 z^2 + c_3 z^3 + \ldots \) be the generating function
of the \( c_i \). Using

\[g^z = 1 + z \log g + \frac{(z \log g)^2}{2!} + \ldots \]

and

\[c_{m+1} = \frac{1}{g} c_m + \frac{g-1}{g} c_{m+1} = \frac{1}{g} \left\{ c_{m+1} + c_m \log g + \ldots + \frac{c_1 (\log g)^m}{m!} \right\} \]

\[+ r \left\{ 1 + \ldots + \frac{(\log \frac{g}{r})^m}{m!} \right\} - 1 \, , \]

we see \( c_{m+1} \) must be the coefficient of \( z^{m+1} \) in

\[\frac{1}{g} c(z) g^z + \frac{rz}{g} \frac{g}{r} z \left( \frac{1}{1-z} \right) - \frac{z}{1-z} \, . \]

That is
\[ c(z) = \frac{z}{1-z} \left[ \frac{(\frac{g}{r})^{z-1} - 1}{g^{z-1} - 1} \right]. \]

As \( z \to 1 \), the quantity in square brackets approaches
\[
\frac{\log(\frac{g}{r})}{\log g} = 1 - \log g \quad r. \]

Thus \( c(z) \) has a simple pole at \( z = 1 \), with residual \( 1 - \log g \quad r. \) By elementary considerations, it follows that \( c_m = \log g \quad r - 1 \)
as \( m \to \infty \) proving (5-4). 

It is worthwhile seeing what has not been proven in
Theorem 3. Knuth [1971] (pg. 219-228), used a computation
similar to the one above in the special case of the first
digit problem to assert a strong type of uniform convergence
to the limit \( \ell \). Using the notation above, he claimed that
there is \( N(\varepsilon) \) so that for \( i, k > N, |\delta(i,k) - \frac{C-b}{a}| < \varepsilon \). It
is interesting to see that this is a non-rectifiable error and
that no result like this need be true. To this end we prove

**Lemma 3.** Let \( \delta(1,0), \delta(2,0), \ldots, \delta(m,0) \) be any real numbers,
define
\[
\delta(i,n) = \frac{1}{i} \sum_{j=1}^{i} \delta(j,n-1). \]

For each \( j, 1 \leq j \leq m, \lim_{i \to \infty} \delta(j,i) = \delta(1,0). \)

**Proof.** Let \( \delta \) be the column vector
\( \delta' = (\delta(1,0), \delta(2,0), \ldots, \delta(m,0)) \). Let \( A \) be the matrix of the
averaging operator
\[
(A^m\delta)_{ij} = \begin{cases} 
\frac{1}{n} & j = 1,2,\ldots,i \\
0 & \text{otherwise.}
\end{cases}
\]

Thus, \((A^m\delta)' = (\delta(1,n),\delta(2,n),\ldots,\delta(m,n))\). The operator \(A\) clearly has distinct eigen values \(1,\frac{1}{2},\ldots,\frac{1}{m}\). A standard theorem in linear algebra states that there exists a basis of eigen vectors for \(A\), call them \(x_1, x_2, \ldots, x_m\). It is not difficult to write down the \(x_1\):

\[
x_1 = (1,1,\ldots,1)
\]

\[
x_2 = (0,1,\frac{2}{1},\ldots,\frac{m-1}{1})
\]

\[
\vdots
\]

\[
x_k = (0,0,\ldots,0,\frac{k-1}{k-1},\frac{k-1}{k-1},\ldots,\frac{m-1}{k-1}).
\]

(The first \(k-1\) entries being 0.) Suppose that with respect to the basis \(x_1, \delta = \sum_{i=1}^{m} q_i x_i\). Then, \(A^m\delta = \sum_{i=1}^{m} q_i \frac{1}{i^n} + q_1 x_1\) as \(n \to \infty\). To finish the proof, note that \(q_1 = \delta(1,0)\).

No form of Knuth's answer can make sense since by simply changing \(\delta(1,0)\), we can completely change the "limiting density".

The next theorem shows that \(H^\infty\) density does not exist if the gaps have a quadratic or higher order polynomial in the exponent. We construct an example and proof in the quadratic case.

**Lemma 4.** The set \(S = \bigcup_{k=0}^{\infty} [10k^2, 10(k+1)^2] \) does not have \(H^\infty\) density, rather, for each \(k\), \(\lim_{x \to \infty} \delta(s,x,k) = 0\), \(\lim_{x \to \infty} \delta(s,x,k) = 1\).
Proof. Recall we write $10^y = t(y)$. Consider $x$ of the form $x = t(n + \frac{s}{2})^2$ where $0 < \gamma_1 < s \leq 1$. Writing $\delta(x,k)$ for $\delta(s,x,k)$:

$$\delta(x,1) = \frac{1}{x} \left\{ \sum_{k=1}^{n-1} \left( t((k+\frac{s}{2})^2) - t(k^2) \right) + t((n + \frac{s}{2})^2 - t(n^2)) \right\}.$$ 

The largest term in the sum, when divided by $x$, is

$$t\{(n-\frac{s}{2})^2 - (n + \frac{s}{2})^2\} = o(t(-n)),$$

where the implied constant is independent of $s$ and $n$. Thus, for $\gamma_1 < s \leq 1$, $\delta(x,1) = 1 + o(nt(-n)) + o(t(-n)) = 1 + o(1)$ where the implied constant may depend on $\gamma_1$, but is independent of $s$ and $n$. This proves $\lim_{x \to \infty} \delta(x,1) = 1$. Assume inductively, $0 \leq \gamma_1 \leq \gamma_2 \ldots \leq \gamma_j < 1$ have been found such that for $\gamma_1 < s \leq 1$, $\delta(t((n + \frac{s}{2})^2),i) = 1 + o(1)$ as $n \to \infty$. We now show for any $\epsilon > 0$, $\gamma_j + \epsilon < s \leq 1$ implies

$$\delta(t(n + \frac{s}{2})^2,j+1) = 1 + o(1),$$

as $n \to \infty$.

$$\delta(t((n + \frac{s}{2})^2),j+1) = \frac{1}{t((n+\frac{s}{2})^2)} \left[ \sum_{k=1}^{t((n + \frac{s}{2})^2)} \delta(k,j) + \sum_{k=1}^{t((n + \frac{s}{2})^2)} \delta(k,j) \right].$$

In the first sum, since $0 \leq \delta(k,j) \leq 1$, dividing by $t((n+\frac{s}{2})^2)$ shows the first sum is $o(t((\gamma_j + \epsilon - s)n)) = o(1)$, where the
implied constant may depend on $\varepsilon$ but not on $n, s$. All terms in the second sum are $1 + o(1)$. Making this substitution, the second sum is $Y + o(Y)$ where

$$Y = \frac{t((n+\frac{s}{2})^2) - t((n + \frac{y_j + \varepsilon}{2})^2)}{t((n+\frac{s}{2})^2)} = 1 + o(1).$$

Combining these estimates gives

$$\delta(t((n+\frac{s}{2})^2), j+1) = 1 + o(1) + o(1 + o(1)) + o(1) = 1 + o(1)$$

as was to be shown. The result for $\lim_{x \to \infty} \delta(x, i)$ now follows by taking $\gamma_1 = \varepsilon$, $\gamma_2 = \gamma_1 + \frac{\varepsilon}{2}$, $\gamma_3 = \gamma_2 + \frac{\varepsilon}{4}$, ... Essentially, the same estimates give the same results for $\lim_{x \to \infty} \delta(x, 1) = 0(t((1-s)n)) = o(1)$. The rest of the proof following similarly. 

It is possible, but notationally awkward, to extend the above example to any polynomials $f(x), g(x)$ of degree $\geq 2$. The set $S$ of Lemma 4 is $S = \bigcup_{k=0}^{\infty} [10^k, 10^{k+1}]$ and by the results of Chapter 2, has zeta density $\frac{1-\gamma}{2} = \frac{1}{2}$.

5.2 Log Density

We proceed now to develop a similar classification for infinite iteration of log averages. Throughout, $A \subseteq \mathbb{N}$ is a subset of the integers. $a(i) = 1$ or $0$ as $i$ is in $A$ or not. $L(A, n, 1) = \frac{1}{\log n} \sum_{i=1}^{n} \frac{a(i)}{i}$ and inductively
L(A,n,k) = \frac{1}{\log n} \sum_{i=1}^{n} \frac{L(A,i,k-1)}{i}. \text{ As before,}

\lim_{n \to \infty} L(A,n,k) \leq \lim_{n \to \infty} L(A,n,k-1) \text{ and similarly, applying \log \n\rightarrow \infty \text{ averages increases the lower limits.}

\text{Definition. A has } L^\infty \text{ density } \ell \text{ if}

\lim_{k \to \infty} \lim_{n \to \infty} L(a,n,k) = \ell = \lim_{k \to \infty} \lim_{n \to \infty} L(A,n,k).

The first result for this procedure shows that no finite number of iterations beyond the first can induce convergence.

\textbf{Theorem 4.} Let } a(i) \text{ be real numbers with } \sum_{i=1}^{n} a(i) = o(\log n). \text{ If } n \log n|a(n)-a(n-1)| < k, \text{ then } a(i) \to 0 \text{ as } i \to \infty.

\textbf{Proof.}

\begin{align*}
o(\log x) &= o(\log x^{1+\delta}) + o(\log x) = \sum_{x \leq i \leq x^{1+\delta}} \frac{a(i)}{i} \\
&= a(\lceil x^{1+\delta} \rceil) \left( \sum_{x \leq i \leq x^{1+\delta}} \frac{1}{i} \right) + \sum_{i=1}^{x^{1+\delta} - x} \left( \sum_{j=x}^{x+1} \frac{1}{j} \right) a(\lceil x + i \rceil) \\
&\quad - a(\lceil x + i + 1 \rceil) \\
&= a(\lceil x^{1+\delta} \rceil) \log \left( \frac{x^{1+\delta}}{x} \right) + O\left( \frac{a(\lceil x^{1+\delta} \rceil)}{x} \right) \\
&\quad + \sum_{i=1}^{x^{1+\delta} - x} \left( (\log(x+i) - \log x) + O\left( \frac{1}{x} \right) \right) \left( a(\lceil x + i \rceil) - a(\lceil x + i + 1 \rceil) \right)
\end{align*}
(5-12) \leq \delta a([x^{1+\delta}]) \log x + \sum_{i=1}^{x^{1+\delta}-x} \frac{\log(x+i)-\log x}{(x+i)\log(x+i)} \\
+ O\left(\frac{1}{x} \sum_{i=1}^{x^{1+\delta}-x} \frac{1}{(x+i)\log(x+i)}\right) + O\left(\frac{a[x^{1+\delta}]}{x}\right),

where we have written \([x]\) for the integral part of \(x\). The sum in the error term is \(\frac{1}{x}(\log\log x^{1+\delta} - \log\log x) = \frac{1}{x}\). Breaking the first sum into two pieces, consider

\[
\sum_{i=1}^{x^{1+\delta}-x} \frac{1}{x+i} - \log x - \sum_{i=1}^{x^{1+\delta}-x} \frac{1}{(x+i)\log(x+i)}
\]

\[
= \log \frac{x^{1+\delta}}{x} - \log x\{\log\log x^{1+\delta} - \log\log x\} + O\left(\frac{1}{x}\right)
\]

\[
= \log x\{\delta - \log(1+\delta)\}.
\]

From this and (5-12), we have

(5-13) \quad 0 \leq \lim_{x \to \infty} \delta a([x^{1+\delta}]) + \{\delta - \log(1+\delta)\},

assuming \(a[x^{1+\delta}] = o(x\log x)\) which follows from \(\sum_{i<x} \frac{a(i)}{i} = o(\log x)\).

Dividing (5-13) by \(\delta\), gives \(0 \geq \lim_{\delta \to 0} \lim_{x \to \infty} a([x^{1+\delta}]),\) so that \(\lim_{i \to \infty} a(i) > 0\).

Using \(a(n) - a(n+1) \geq \frac{k}{n \log n}\) in (5-12) gives the opposite inequality for the upper limit which proves the theorem.

Defining kth log density in the obvious way yields the following.

**Corollary 2.** A set \(A \subseteq N\) has kth log density iff it has log density and then the two limits are equal.
Proof. Since the log summability method is regular, if log density exists, then all iterates exist and give an equal limit. Conversely, if \( \frac{1}{\log n} \sum_{i=1}^{n} \frac{b(i)}{i} \rightarrow l \) where \( b(i) = \frac{1}{\log i} \sum_{j=1}^{n} a(j) \), then we need only check the Tauberian condition in Theorem 4.

\[
 b_i - b_{i+1} = \left( \sum_{j=1}^{i} \frac{a(j)}{j} \right) \left[ \frac{1}{\log i} - \frac{1}{\log (i+1)} \right] - \frac{a(i+1)}{i \log i} \\
 \leq \frac{k \log i \log (i+1)}{\log i \log (i+1)} - \frac{a(i+1)}{i \log i} < \frac{k'}{i \log i} .
\]

A similar argument shows \( b_{i+1} - b_i \leq \frac{k''}{i \log i} \), so the Tauberian condition is satisfied and Theorem 4 implies \( a(i) \rightarrow l \). The argument clearly extends to any bounded sequence \( a(i) \) which proves the corollary. \( \| \)

The \( L^\infty \) density extends log density in much the same way as \( H^\infty \) density extends natural density. In particular, \( L^\infty \) is weaker than loglog(\( LL \)) density defined in Section 4.2.3 as the limit of \( \frac{1}{\log \log n} \sum_{i=1}^{n} \frac{a(i)}{i \log i} \) if it existed.

Theorem 5. \( L^\infty \) density is weaker than \( LL \) density.

Proof. The theorem follows from the relation, valid for any bounded sequence \( a(i) \),

\[
(5-14) \quad \sum_{i=m}^{n} \frac{a(i)}{i \log i} = \sum_{i=m}^{n} \frac{L(A,n,1)}{i \log i} + O(1)
\]

where
\[ L(A,i,1) = \frac{1}{\log i} \sum_{j=1}^{i} \frac{a(j)}{j}. \]

To prove this, write
\[
\sum_{i=m}^{n} \frac{a(i)}{i \log i} = O(1) + \sum_{i=m}^{n} \left[ \left( \sum_{j=1}^{i} \frac{a(j)}{j} \right) \left( \frac{1}{\log i} - \frac{1}{\log (i+1)} \right) \right]
= O(1) + \sum_{i=m}^{n} L(A,i,1) \frac{\log(1+\frac{1}{i})}{\log(i+1)}
= O(1) + \sum_{i=m}^{n} \frac{L(A,i,1)}{i \log i}.
\]

The last equality used \( \log(1+\frac{1}{i}) = \frac{1}{i} + O(\frac{1}{i^2}) \). Iterating relation (5-14), gives
\[
\sum_{m}^{n} \frac{a(i)}{i \log i} = \sum_{i=m}^{n} \frac{L(A,i,k)}{i \log i} + O_k(1)
\]
so that
\[
(5-15) \quad \sum_{m}^{n} \frac{a(i)}{i \log i} \leq \sup_{i \geq m} L(A,i,k)(\log\log n - \log\log m) + O(1).
\]

Dividing both sides of (5-15) by \( \log\log n \) and letting \( n \to \infty \), gives \( \lim_{n \to \infty} LL(A,n) \leq \lim_{n \to \infty} L(A,n,k) \) for each \( k \). A similar relation holds for the lower limits and the lemma follows from the definition of \( L^\infty \) density just as in the proof of Theorem 2 of this chapter. \[ \]

The theorem that follows shows that loglog linear sets have \( L^\infty \) density. The proof is similar to Theorem 3 above and we can make direct use of the results of Section 5.1. We continue to write \( t(y) \) for \( 10^y \).
Lemma 5. For all $k$, $\epsilon > 0$, the functions introduced in Lemma 2 have the property that there exists $N(k, \epsilon)$ such that for $1 \leq s \leq 10^a$,

$$|L(t(st(an+b)), k) - Q_k(s)| < \epsilon \quad \text{if} \quad 0 \leq s \leq 10^{c-b}$$

$$|L(t(st(an+b)), k) - (Q_k(s)+R_k(s))| < \epsilon \quad \text{if} \quad 10^{c-b} \leq s \leq 10^a.$$  

Proof. For $k = 1$, let $n = t(st(am+b))$ for some integer $m$, $1 \leq s \leq 10^a$. If $1 \leq s \leq 10^{c-b}$, then

$$\sum_{i \leq n} \frac{a(i)}{i} = \sum_{j=0}^{m-1} \frac{t(a(j+1)+b)}{tt(a(j+b))} \frac{1}{i} + \sum_{j=tt(a(j+b))}^{n} \frac{1}{i} + O(1)$$

$$= \sum_{i=0}^{m-1} \{\log(tt(a(j+c)) - \log(tt(a(j+b))) + O(\frac{1}{tt(a(j+b))})$$

$$+ \log n - \log tt(am+b) + O(1)$$

$$= \log 10\left[\sum_{i=1}^{m-1} (t(a(j+c)) - t(a(j+b))) + O(1) + st(am+b) - t(am+b)\right] + O(1)$$

$$= \log 10[\{t(c) - t(b)\} \frac{t(am)-1}{t(a)-1} + (s-1)t(am+b)] + O(1).$$

Dividing by $\log n = (\log 10)st(am+b)$, we find

$$L(n, 1) = 1 - \frac{1}{s} \left\{\frac{t(a)-t(c-b)}{t(a)-1}\right\} + O(10^{-am}).$$

Similarly for $10^{c-b} \leq s \leq 10^a$, 

- 78 -

\[ L(n,1) = \frac{1}{s} \sum_{i=1}^{n} \frac{L(i,k-1)}{t(a)} \cdot 10^{a-b} + O(10^{-am}). \]

Combining these results, leads to

\[ |L(st((am+b),1) - S_1(s)| = O(10^{-am}) \]

which proves the lemma for \( k = 1 \). For \( k > 1 \), \( n \) again of the form \( t(st(a(\cdot+b)) \) with \( 1 \leq s \leq 10^a \):

\[
\begin{align*}
L(n,k) &= \frac{1}{\log n} \sum_{i=1}^{n} \frac{L(i,k-1)}{t(a)} = \frac{1}{s \log 10} \\
&\quad \times \left\{ \sum_{0 \leq j < m} \frac{1}{t(a^i)} \right\} \sum_{i=t(a^i)} \frac{L(i,k-1)}{t(a^i)} \right\} + O(t(-am)) \right) + o(t(b)).
\end{align*}
\]

where the error term comes from neglected terms \( \leq tt(b) \). The inner sums can be well approximated using the inductive step:

\[
\left| \sum_{i=t(a^j)}^{L(i,k-1)} \frac{L(i,k-1)}{t(a^j)} - \sum_{i=t(a^j)}^{S_k-1} \frac{\log i}{t(a^j)} \right| \leq \varepsilon(q-1)
\]

for \( 1 \leq q \leq 10^a \) and \( j > N(k-1,\varepsilon) \) where \( \lambda = \log 10 \). The second sum in absolute value is a Riemann sum for the integral \( \int_1^q S_k(y)dy \). To see this, note that multiplying and dividing by \( d = tt(a^j) \) and letting \( c = \frac{1}{\lambda t(a^j)} \), leads to
\[
\left| c \sum_{i=d}^{\infty} \frac{tq(t^i(a+b)) S_{k-1}(1+\log(\frac{i}{d}))}{i} \right| - c \int_1^q t((q-1)t(a+b)) \times \frac{S_{k-1}(1+c \log y)}{y} \, dy < \epsilon
\]

for \(j\) large. Making the substitution \(1 + c \log y = z\) in the integral, shows that the sum is a Riemann sum for \(\int_1^q S_{k-1}(y)dy\) as asserted. Combining these estimates leads to a proof of Lemma 5 in a word for word repetition of the proof of Lemma 2. ||

The required theorem now follows easily.

**Theorem 6.** \(S = \bigcup_{k=0}^{\infty} [10^{10^a k+b}, 10^{10^a k+c}]\) with \(0 \leq \frac{c-b}{a} \leq 1\) has \(L^\infty\) density \(\frac{c-b}{a}\).

**Proof.** From Lemma 5, it suffices to show \(R_k(s) \to 0\), \(Q_k(s) + \frac{c-b}{a}\) uniformly in \(s\) as \(k \to \infty\). This was proved in Theorem 3 above. ||

The construction of counter examples for sets with non-linear polynomial terms in the second exponent goes over in a straightforward manner and will not be repeated here.
CHAPTER 6

SUMMABILITY AND RANDOM VARIABLES

6.1 Introduction

The classical limit theorems of probability theory deal with averages of random variables \( X_i \), asserting results like \( \frac{1}{n} \sum_{i=1}^{n} X_i \to 0 \), almost surely as \( n \to \infty \). The results in this chapter relate to weighted averages of variables \( X_i \) in much the same way as previous chapters considered generalized averages of real numbers \( a(i) \). For example, consider independent and identically distributed (iid) variables \( X_i \) with mean \( \mu \). The binomial average of the \( X_i \) is a sequence of variables \( B(n) = \frac{1}{2^n} \sum_{i=0}^{n} \binom{n}{i} X_i \). A theorem typical of this chapter is that \( B(n) \to \mu \) in probability as \( n \to \infty \).

If the \( X_i \) were all normal variables with mean \( 0 \), variance \( 1 \), then the random power series \( A(z) = \sum_{i=0}^{\infty} z^i X_i \) would be almost surely well defined and have a normal distribution with mean \( 0 \), variance \( \sigma^2(z) = \sum_{i=0}^{\infty} z^{2i} = \frac{1}{1-z^2} \) so that \( \sqrt{1-z^2} A(z) \) is again standard normal. If the \( X_i \) are different from normal but iid with mean \( 0 \) and variance \( 1 \). We can ask what happens to the sum \( \sqrt{1-z^2} A(z) \) as \( z \to 1 \). We show that such sums obey a central limit theorem, converging to a standard normal distribution as \( z \to 1 \).

The general theme is that weighted sums, properly normalized, behave very much like sums of independent variables with the
classical norming up to regularity conditions. This theme works in two directions. First, in situations where weighted sums arise in a practical context, standard procedures can deal with the problem. Second, consideration of weighted sums can shed light on the regularity conditions underpinning the classical theorems.

The classical norming \( \frac{1}{n} \sum X_i \) treats all the \( X_i \) equally whereas binomial averages weight the \( X_i \) with \( i \) close to \( \frac{n}{2} \) most favorably while paying little attention to \( X_i \) in either tail. We now consider two situations where such discounting of the data is natural.

**Example 1: Quality control.** An important applied problem arises when observing a manufacturing process, trying to detect deviations from the norm. The most widely used method, that of Shewhart style control charts, considers observations \( X_i \) and stops the process when an observation appears "out of line". Instead of just looking at a single observation, it makes sense to look at the entire history of the sequence of trial observations, weighting the most recent observations as most important. For example, the statistic \( \frac{1}{\log n} \sum_{i=0}^{n-1} \frac{X_i}{n-i} \) pays less attention to observations which occurred at the beginning of the process. This statistic does consider all of the data and might make earlier detection of a trend possible. Details of this special scheme of weighting may be
found in Lai [1973]. Many other real world examples could be added where a related procedure makes sense. Consider the problem confronting a doctor in reviewing medical history (for example, blood pressure). Clearly the most recent tests are most important but tests made in the recent past may also have significance as might the entire history of the process. The limit theorems that follow allow consideration of procedures like the reverse log averages above.

**Example 2: Rate of return.** Economists discount future costs and earnings by a method which can be approximated by random power series models. Let $X_i$ be the income from a process at time $i$. If the program being considered was building a new college dormitory, then perhaps $X_i$ would be negative while the building was going up and then positive into the future as the rewards were being realized. For $0 < z < 1$, the amount $X_i z^i$ in hand at the present time could be reinvested $i$ times at interest rate $\frac{1}{z}$ so that at time $i$, the total amount on hand would be $X_i$. In this sense, the present value of having amount $X_i$ at time $i$ is $X_i z^i$. This leads to consideration of the present value of the entire process, $A(z) = \sum_{i=0}^{\infty} X_i z^i$. If the process were finite, then $A(1)$ has the interpretation of actual costs and profits of the process. The limit theorems dealing with power series contain results such as
\[(1-z)A(z) \sim (1-z) \sum_{i=0}^{\infty} \mu_i z^i \quad \text{as} \quad z \rightarrow 1,\]

in probability, where \(\mu_i\) is the mean of \(X_i\).

Further details and references to the economics literature in the case of random polynomials may be found in Fairley [1968]. The discounting interpretation has also been mentioned by Whitt [1972].

There has been work done on limit theorems with weighted averages but the field is small enough so that even though many of these references were discovered after research reported in this chapter was completed, the results reported here, as well as the overall approach to the problem, seem to be new. Chapter 2, Section 10 of Revesz [1968], Kahane [1968], papers by Chow and Lai [1973], Chow and Teicher [1973], Whitt [1972] and the survey paper of Hanson [1970] all contain results related to those proved here and many other references.

The work in this chapter is largely devoted to weak convergence of weighted sums. The second section gives results for the law of large numbers for many of the averages considered in previous chapters. The fourth section is on the central limit problem where attention is largely confined to extension of classical results to the case of infinite row sums in triangular arrays. Analogs of the theorems of Laplace, Liapounov, and Lindeberg are presented for Abelian averages and Berry-Esseen type theorems are given for zeta and Abelian
averages. Section three deals with strong convergence and presents theorems for a class of weighted means including log and higher log averages.

Some interesting questions not pursued here are large deviations, the law of the iterated logarithm, and extension to dependent variables.

6.2 Weighted Weak Laws

The problem considered here is convergence in probability of sums
\[ Y_n = \sum_{i=0}^{k(n)} w(n,i)X_i \] as \( n \to \infty \). Here \( w(n,i) \) is a matrix of nonnegative weights with
\[ \sum_{i=0}^{k(n)} w(n,i) = 1 \] and we permit \( k(n) = \infty \). Since the variables \( X_i \) may not have means in particular cases, we adopt the following.

**Definition.** Variables \( Y_i \) converge in probability if there exists a sequence \( a(n) \) such that \( P(|Y_n - a(n)| > \epsilon) \to 0 \) as \( n \to \infty \).

We will derive necessary and sufficient conditions for convergence in probability from the known results about convergence in triangular arrays. Before doing this, it is useful to formulate Chebyshev's inequality in this setting and consider some examples.

Let \( X_i \) be independent variables with means \( \mu_i \) and variances \( \sigma_i^2 \). Assume \( B(n) = \sum_{k=1}^{k(n)} w(n,k)^2 \sigma_k^2 < \infty \) for all \( n \).
This guarantees that the variables $Y_n = \frac{k(n)}{\sum_{k=1}^{k(n)} w(n,k)(X_k - \mu_k)}$ exist almost surely if $k(n) = \infty$ (Kolmogorov [1956], pg. 67).

The classical Chebyshev inequality yields

$$P(\left| Y_n \right| \leq \varepsilon) \geq 1 - \frac{B(n)}{\varepsilon^2}$$

so that

**Theorem 1.** Notation as above, if $B(n) \to 0$, then the $W$ average of the $X_i$ differs from the $W$ average of the $\mu_i$ by less than $\varepsilon$ with probability arbitrarily close to 1.

**Example 3.** Consider binomial averages, $w(n,k) = \binom{n}{k} 2^{-n}$.

Let $X_i$ be i.i.d. with mean 0, variance $\sigma^2$.

$$B(n) = \frac{\sigma^2}{2^2} \sum_{i=0}^{n} \binom{n}{i}^2 = \frac{\sigma^2}{2^2} \frac{\binom{2n}{n}}{2^{-n}} = O(\frac{1}{\sqrt{n}})$$

by straightforward computation, so Chebyshev's inequality implies

$$P(\left| \frac{1}{2^n} \sum_{i=0}^{n} \binom{n}{i}X_i \right| > \varepsilon) \to 0 \text{ as } n \to \infty .$$

Under the conditions of Example 3, a similar proof yields weak convergence for Borel averages; that is:

$$P(|e^{-\lambda} \sum_{i=0}^{\infty} X_i \frac{\lambda^i}{i!}| > \varepsilon) \to 0 \text{ as } \lambda \to \infty .$$

**Example 4.** $X_i$ as in Example 3, consider $(1-x) \sum_{i=0}^{\infty} x^i X_i$.

This converges with probability 1 for $x < 1$ and
\[ B(x) = (1-x)^2 \sigma^2 \sum_{i=0}^{\infty} x^i = \sigma^2 \frac{(1-x)^2}{1-x^2} \to 0 \]

as \( x \to 1 \), so that

\[ P(|(1-x)\sum x^i X_i| > \varepsilon) \to 0 \quad \text{as} \quad x \to 1. \]

The classical necessary and sufficient conditions for convergence in probability of triangular arrays specialize to give necessary and sufficient conditions for the weak law for weighted sums. The results must be stated separately for weighting arrays with \( w(n,i) = 0 \) for \( i > k(n) \) with \( k(n) < \infty \), and for arrays which have infinite rows. The difference occurs because of problems concerning the (almost sure) existence of the "partial averages" \( \sum_{i=0}^{k(n)} w(n,i) X_i \).

Suppose first all \( k(n) < \infty \) and \( \sum_{i=1}^{k(n)} w(n,i) = 1 \) for all \( n \).

Let \( m(k) \) be a median of the variable \( X_k \) and let \( m(n,k) = w(n,k)m(k) \) be a median of \( w(n,k)X_k \).

**Theorem 2.** In order that the variables obey the W law of large numbers, it is necessary and sufficient that

A) \[ \sum_{k=0}^{k(n)} \int_{|x| > 1/(w(n,k))} dF_k(x+m(n,k)) \to 0 \]

B) \[ \sum_{k=0}^{k(n)} (w(n,k))^2 \int_{|x| < 1/(w(n,k))} x^2 dF_k(x+m(n,k)) \to 0 \]

as \( n \to \infty \). Here \( F_k \) is the distribution function of \( X_k \). In
this case, \( a(n) \) in the definition of weak convergence can be chosen as

\[
a(n) = \sum_{k=1}^{k(n)} w(n,k) \left\{ \int_{|x|<1/(w(n,k))} x dF_k(x+m(n,k))+m(k) \right\}
\]

**Proof.** This follows from the classical theorem for triangular arrays as in Gnedenko and Kolmogorov [1954] Chapters 4 and 5, in particular, Section 22 and remark 1 in Section 27. The triangular array considered is \( Y_{n,k} = w(n,k)X_k \). Conditions A and B are translations of the cited remark 1. |||

The notion of weak and strong average takes on a more limited character when the quantities being averaged are allowed to be unbounded in both directions. The following example of variables without means which obey the weak law and the weak binomial law shows the conditions in Theorem 2 can be verified in particular cases. More easily applied sufficient conditions will also be given.

**Example 5.** Let the i.i.d. variables \( X_i \) take on values \( \{\pm 2, \pm 3, \ldots\} \) with \( P(X_j=i) = P(X_j=-i) = \frac{c}{i^2 (\log i)^b} \) where \( c \) is a constant making the density sum to 1. The variables \( X_i \) are symmetric about 0 but don't have a finite mean for \( b \leq 1 \). The medians \( m(k) \) and \( m(n,k) \) are zero and conditions A and B of Theorem 2 become.
(6-1) \[
\sum_{k=0}^{k(n)} \sum_{i > w(n,k)} \frac{1}{i^2 (\log i)^b} \to 0
\]

(6-2) \[
\sum_{k=0}^{k(n)} \frac{1}{(w(n,k))^2} \sum_{i = 2}^{\frac{1}{w(n,k)}} \frac{1}{(\log i)^b} \to 0.
\]

Since \(\log^b(x)\) varies slowly, the inner, upper tail sum in (6-1) is asymptotically \(\frac{w(n,k)}{\log^b(w(n,k))}\) (see Feller [1966], pg. 273, Theorem 1). Similarly, the inner, lower tail sum in (6-2) is asymptotically \(\frac{1}{w(n,k)\log^b(w(n,k))}\). Thus, both conditions (6-1) and (6-2) will be satisfied if

(6-3) \[
\sum_{k=0}^{k(n)} \frac{w(n,k)}{\log^b(w(n,k))} \to 0 \quad \text{as} \quad n \to \infty.
\]

Consider first the case of natural density \(w(n,k) = \frac{1}{n}\) for \(k = 1, 2, \ldots, n\). Condition (6-3) becomes

\[
\frac{1}{n(\log n)^b} \sum_{i=1}^{n} \frac{1}{(\log n)^b} \to 0 \quad \text{if} \quad b > 0.
\]

In this case, the centering constants of Theorem 2 become

\[
a(n) = \sum_{i=1}^{n} \frac{1}{-n i (\log i)^b} = 0
\]

so we have proved \(P\left(\frac{1}{n} \sum_{i=1}^{n} X_i > \varepsilon\right) \to 0\) as \(n \to \infty\).

Using binomial weights \(w(n,k) = \frac{\binom{n}{k}}{2^n}\) in (6-3) along with the inequality \(w(n,k) \ll \frac{1}{\sqrt{n}}\), shows that for these weights (6-3) becomes
$$0(\frac{1}{(\log(n))^b}) = o(1) \quad \text{for} \quad b > 0.$$ 

Again, the centering constants vanish by symmetry so that

$$P\left(\left|\frac{1}{2^n} \sum_{i=0}^{n} \binom{n}{i} X_i \right| > \varepsilon\right) \to 0 \text{ as } n \to \infty.$$ 

We now derive analogs of Khintchine's weak law and a theorem of Markov in the case of binomial averages.

**Theorem 3.** Let $X_i$ be i.i.d. with mean $\mu$, then $X_i$ obey the binomial weak law, converging weakly to $\mu$.

**Proof.** W.L.O.G. we may assume that $X_i$ has median 0. Conditions A and B of Theorem 2 become

$$(6-4) \quad \sum_{k=0}^{n} \int_{|x| \geq \frac{2^n}{\binom{n}{k}}} dF(k) \leq \sum_{k=0}^{n} \binom{n}{k} \int_{|x| \geq \frac{2^n}{\binom{n}{k}}} x \, dF(x) \leq \sum_{k=0}^{n} \binom{n}{k} \int_{|x| \geq 4\sqrt{n}} x \, dF(x) + 0 \quad \text{as } x \to \infty.$$ 

$$(6-5) \quad \sum_{k=0}^{n} \frac{(\binom{n}{k})^2}{2^{2n}} \int_{|x| \leq \frac{2^n}{\binom{n}{k}}} x^2 \, dF(x) \leq \sum_{k=0}^{n} \frac{(\binom{n}{k})^2}{2^{2n}} \int_{|x| \leq c} x^2 \, dF(x) + \sum_{k=0}^{n} \frac{(\binom{n}{k})}{2^n} \int_{c \leq |x| \leq \frac{2^n}{\binom{n}{k}}} |x| \, dF(x),$$

where $c$ is any positive constant. The integral in the first sum on the right of (6-5) is bounded. Using $\sum_{k=0}^{n} (\binom{n}{k})^2 = \binom{2n}{n} = o\left(\frac{2^{2n}}{\sqrt{n}}\right)$,
the first sum is \( O_c \left( \frac{1}{\sqrt{n}} \right) \). Choose \( c \) so that \( \int_{|x| \leq c} |x| dF(x) \leq \varepsilon \), so the second sum in (6-5) is \( \leq \varepsilon \) and the inequality (6-5) becomes

\[
\sum_{k=0}^{n} \frac{n^2}{2^{2n}} \int_{|x| \leq 2^{n/k}} x^2 dF(x) \leq O_c \left( \frac{1}{\sqrt{n}} \right) + \varepsilon < 2\varepsilon
\]

for large \( n \). Theorem 2 states the binomial averages of the \( X_i \) converge in probability to

\[
a(n) = \sum_{k=0}^{n} \frac{n^2}{2^{2n}} \int_{|x| \leq 2^{n/k}} x dF(x) .
\]

Since the integral + \( \mu \) uniformly in \( k \) as \( n \to \infty \), the Toeplitz lemma implies \( a(n) \to \mu \), so that clearly

\[
P\left( \left| \frac{1}{2^n} \sum_{i=1}^{n} X_i - \mu \right| > \varepsilon \right) \to 0 \text{ as } n \to \infty . \]

**Theorem 4.** If independent variables \( X_i \) have uniformly bounded \( 1+\delta \) moments, then \( X_i \) obey the weak binomial law.

**Proof.** Since we need only prove there are centering constants for weak convergence, there is no loss of generality in assuming \( m(k) = 0 \). The first condition in Theorem 2 becomes:

\[
(6-6) \quad \sum_{k=0}^{n} \int_{|x| \leq 2^{n/k}} dF_k(x) \leq \sum_{k=0}^{n} \frac{n^{1+\delta}}{2^n (1+\delta)} \int_{|x| \leq 2^{n/k}} |x|^{1+\delta} F_k(x) \leq \frac{c}{n^\delta} \to 0
\]

as \( n \to \infty \). The estimate on the sums \( \sum_{k=0}^{n} x_k^{1+\delta} \) comes from Lemma 1 proved below. The second condition of Theorem 2 is:
\[
\sum_{k=0}^{n} \frac{(\frac{n}{k})^2}{2^{2n}} \int_{|x| \leq \frac{2^n}{(\frac{n}{k})}} x^2 dF_k(x) \ll \sum_{\frac{n}{k} \leq \frac{n}{(1+\delta)}} \left\{ \frac{(\frac{n}{k})^{1+\delta}}{2^{n(1+\delta)}} \int_{|x| \leq \frac{n}{(1+\delta)}} x^{1+\delta} dF_k(x) \right\} \to 0
\]

as above. 

The next pair of theorems give necessary and sufficient conditions for weighted weak laws when \( k(n) = \infty \). The presentation becomes greatly simplified if attention is restricted to variables \( X_i \) with finite means \( \mu_i \) and weighting schemes \( w(n,i) \) which have the property that \( w(n,i)X_i \) are infinitesimal, that is, such that \( \sup_{1 \leq k \leq \infty} P(|w(n,k)X_k| \geq \varepsilon) \to 0 \) as \( n \to \infty \). In this case \( W.L.O.G. \mu_i = 0 \) and we have the following theorems.

**Theorem 5.** Notation as above, let \( S(n,k) = \{|x| \leq \frac{\varepsilon}{w(n,k)}\} \), \( CS \) be the complement of \( S \). In order that the series \( \sum_{i=1}^{\infty} w(n,i)X_i \) converge with probability 1, it is necessary and sufficient that for each \( \varepsilon > 0 \):

a) \( \sum_{k=1}^{\infty} \int_{CS(n,k)} dF_k(x) < \infty \)

b) \( \sum_{k=1}^{\infty} w(n,k) \int_{S(n,k)} x dF_k(x) < \infty \)

c) \( \sum_{k=1}^{\infty} (w(n,k))^2 \left\{ \int_{S(n,k)} x^2 dF_k(x) - \left( \int_{S(n,k)} x dF_k(x) \right)^2 \right\} < \infty \)

for each \( n \). Assuming convergence, it is necessary and sufficient for the \( W \) average of the \( X_i \) to converge weakly to zero that the three series (a), (b), (c) converge to zero as \( n \to \infty \).
Proof. The first part of the theorem is the classical three series theorem (Gnedenko and Kolmogorov [1954], pg. 137). The second part of the theorem results from the corresponding result for triangular arrays by noticing that the proof of the theorem stated as remark three in Gnedenko and Kolmogorov [1954] (pg. 134), remains valid if \( k(n) \) is replaced throughout by \( \infty \). The details are entirely straightforward and are omitted. 

The same methods used in the proof of Theorems 3 and 4 above, give analogs of the theorems of Khintchine and Markov for Abel and zeta density.

The final problem to be considered in this section is the distribution of sums with components \( X_i \) taking only the two values \( \pm i^a \) each with probability \( \frac{1}{2} \). This problem is due to S. N. Bernstein, who used these variables as an example of variables which obey the central limit theorem but not the weak law. The behavior of these variables with respect to various averaging processes, sheds light on the strengths and weaknesses of the average considered. The first example goes back to Bernstein.

**Example 6.** The variables \( X_i \) satisfy the weak law iff \( a < \frac{1}{2} \).

**Proof.** Asymptotically, the sum of the variances of \( X_1, \ldots, X_n \) is proportional to \( n^{2a+1} \). For \( a < \frac{1}{2} \), Chebyshev's inequality implies that
\[ p\left( \left| \frac{x_1 + \ldots + x_n}{n} \right| > \epsilon \right) \geq 1 - \frac{cn^{2a+1}}{en^2} \to 1. \]

For \( a \geq \frac{1}{2} \), use of the central limit theorem shows that

\[ \frac{1}{n^{(2a+1)/2}} \sum_{i=1}^{n} X_i \]

tends to a nondegenerate normal distribution so that it is impossible that \( \frac{1}{n} \sum_{i=1}^{n} X_i \) converge in probability to a constant.

Details may be found in Feller [1968] (pg. 255).

**Example 7.** The variables \( X_i \) satisfy the weak binomial law for \( a < \frac{1}{4} \) but not for \( a \geq \frac{1}{4} \).

For a proof, we need bounds on the variance of the weighted sum \( \sum_{i=0}^{n} \binom{n}{i} X_i \) which is \( \sum_{i=1}^{n} \binom{n}{i} i^{2a} \). The following lemma gives an asymptotic expression for sums of this type.

**Lemma 1.** For \( a > 0 \), all \( b \),

\[ \sum_{i=1}^{n} \binom{n}{i} a_i^b \sim c(a,b) 2^{a} n b^{+\frac{1}{2}} (1-b) \]

as \( n \to \infty \) where

\[ c(a,b) = \frac{2^{\frac{3}{2}(a-1)-b}}{\rho^{\frac{1}{2}(a-1)}} \]

is a constant depending only on \( a \) and \( b \).

**Proof.** To simplify notation, assume \( n = 2v \) is even and rewrite the sums as
\[(6-7) \quad \frac{v}{k=-(v-1)} (v^{n+k})^a (v+k)^b.\]

The general term in this sum is:

\[(6-8) \quad \left\{ \frac{n!}{(v+k)!(v-k)!} \right\}^a (v+k)^b = \left\{ \frac{n^{1/2}}{\sqrt{2\pi(v+k)}} \frac{1}{v+k+\frac{1}{2}a(v-k)} \right\}^a E(n,k)\]

where \(E(n,k) \leq \exp \left\{ \frac{1}{12n} - \frac{1}{12(v+k)+1} - \frac{1}{12(v-k)+1} \right\} \). If \(k\) only varies between

\[(6-9) \quad -\frac{n^{5/8}}{2} \quad \text{and} \quad \frac{n^{5/8}}{2},\]

we see \(E(n,k) \leq \exp \left( \frac{ac}{n} \right)\), for some constant \(c\), uniformly in \(k\).

To bound the upper tail,

\[
\sum_{i=r}^{n} \binom{n}{i} a \binom{i}{r} b \leq \binom{n}{r} (n-r) n^b \quad \text{if} \quad b > 0
\]

\[
= O(\exp a(\log n)(n + \frac{1}{2} - \frac{b}{a})
- (\log(n-r))(n - r + \frac{1}{2} - \frac{1}{a}) - (\log r)(r + \frac{1}{2}))
\]

Setting \(r = \frac{n}{2} + \frac{n^{5/8}}{2} \), this becomes

\[
O(\exp a(\log n)(n + \frac{1}{2} + \frac{b}{a}) - (\log(\frac{n}{2} - \frac{n^{5/8}}{2}))(\frac{n}{2} - \frac{n^{5/8}}{2} + \frac{1}{2} - \frac{1}{a})
- \log(\frac{n}{2} + \frac{n^{5/8}}{2} + \frac{1}{2}))
\]

\[
= O(\exp a\{n \log 2 - [(\log(1 - n^{-3/8}))(\frac{n}{2} - \frac{n^{5/8}}{2})
+ (\log(1 + n^{-3/8}))(\frac{n}{2} + \frac{n^{5/8}}{2})] + O(\log n)\}).
\]

Using \(\log(1+x) = x - \frac{x^2}{2} + O(|x|^3)\) for \(-\frac{1}{2} < x < \frac{1}{2}\), the term in square brackets in the exponential, becomes
\[-95.-\]

\[
(-n^{-3/8} - \frac{n^{-6/8}}{2} + 0(n^{-9/8}))(\frac{n^{5/8}}{2} - \frac{n^{5/8}}{2})
+ (n^{-3/8} - \frac{n^{-6/8}}{2} + 0(n^{-9/8}))(\frac{n^{5/8}}{2} + \frac{n^{5/8}}{2})
= (\frac{n^{5/8}}{2} - \frac{n^{2/8}}{4} + 0(n^{-1/8})) + (\frac{n^{2/8}}{2} + \frac{n^{-1/8}}{4} + 0(n^{-4/8}))
+ (\frac{n^{5/8}}{2} - \frac{n^{2/8}}{4} + 0(n^{-1/8})) + (\frac{n^{2/8}}{2} - \frac{n^{-1/8}}{4} + 0(n^{-4/8}))
= \frac{n^{1/4}}{2} + 0(1)
\]

so that for \(r = \frac{n}{2} + \frac{n^{5/8}}{2}\)

\[(6-10) \quad \sum_{i=r}^{n} (\frac{n}{i})^{a_i b} = 0(2^{an_e}(-an^{\frac{1}{8}})/3) .\]

The same bound holds for the upper tail if \(b \leq 0\) and similarly for the lower tail:

\[(6-11) \quad \sum_{i=1}^{\frac{n}{2} - \frac{n^{5/8}}{2}} (\frac{n}{i})^{a_i b} = 0(2^{an_e}(-an^{\frac{1}{8}})/3)\]

for all fixed \(b\). Consider the main term in \((6-8)\)

\[(6-12) \quad \left\{ \frac{n^{n^{1/2}}}{\sqrt{2\pi (v+k)c(v-k)d}} \right\}^a = \frac{n^{a(n^{1/2})}}{(2\pi)^{a/2} v^a(c+d)} \left\{ \exp \frac{-a(c \log(1 + \frac{k}{v})}{(2\pi)^{a/2} v^a(c+d)} \exp \frac{d \log(1 - \frac{k}{v})}{(2\pi)^{a/2} v^a(c+d)} \right\} ,\]

where \(c = v + k + \frac{1}{2} - \frac{b}{a}, d = v - k + \frac{1}{2}\). Now \(|\frac{k}{v}| \leq n^{-3/8}\), so that the following 3-term expansion is valid for all \(k\) in

\[-\frac{n^{5/8}}{2} \leq k \leq \frac{n^{5/8}}{2}, \log(1 \pm \frac{k}{v}) = \pm \frac{k}{v} - \frac{1}{2}(\frac{k}{v})^2 + 0(|\frac{k}{v}|^2) .\] Using
this, the exponential term in (6-12) becomes
\[
(v+k)(k_v^2 - \frac{1}{2}(k_v^2)^2 + 0(|k_v|^3)) - (v-k)(k_v^2 + \frac{1}{2}(k_v^2)^2 + 0(|k_v|^3)) + 0(k_v)
\]
\[
= 2k_v^2 - k_v^2 + 0(|k_v|^3) + 0(k_v^3) + 0(k_v) = k_v^2 + O(n^{-1/8}).
\]
So (6-12) can be written as
\[
(6-13) \quad \frac{2^{an}}{\pi^{a/2}} v^{-a/2} \left\{ (\exp -a(k_v^2))(1 + O(n^{-1/8})) \right\}.
\]
Combining (6-13) and (6-9) gives, for \( -\frac{n^{5/8}}{2} \leq k \leq \frac{n^{5/8}}{2} \),
\[
(6-14) \quad \left[ \frac{n!}{(v+k)!(v-k)!} \right]^{a} (v+k)^b = \frac{2^{an}}{\pi^{a/2}} v^{-a/2} \left\{ \exp -a(k_v^2) \right\} (1 + O(n^{-1/8}))
\]
Combining (6-10), (6-11), and (6-14) in (6-7) yields
\[
(6-15) \quad \sum_{i=1}^{n} (n_i)^a b^i = \frac{2^{an} v^{b-a/2}}{\pi^{a/2}} \sum_{k=v-n^{5/8}/2}^{v+n^{5/8}/2} \left\{ \exp(-\frac{ak_v^2}{v}) \right\} (1 + O(n^{-1/8}))
\]
\[
+ O(2an_\epsilon^\epsilon \epsilon^{-\epsilon/3} / 3).
\]
The difference
\[
\left| \sum_{k=v-n^{5/8}/2}^{v+n^{5/8}/2} \frac{1}{\sqrt{v}} \exp(-\frac{ak_v^2}{v}) - \int_{-\infty}^{\infty} e^{-ax^2} dx \right| \to 0 \text{ as } n \to \infty.
\]
Since the integral has value \( \sqrt{\frac{\pi}{a}} \), we have finally:
\[
\sum_{i=1}^{n} (n_i)^a b^i \approx 2^{an} b^{-\frac{a}{2}} (a-1) c(a,b)
\]
with
\[ c(a,b) = \frac{2^{\frac{1}{2}}(a-1)-b}{\sqrt{a} \pi^{\frac{1}{2}}(a-1)} \]

**Remark.** The constant checks in the easily verified cases \( a = 1, b = 0; a = 2, b = 0 \). The lemma seems to be a useful result which does not seem to appear in the literature in this generality. For references to some special cases (mostly \( a = 1, b \) a negative integer), see Chao [1972]. We note without proof that basically the same approach can be used to show \( E(f(n)) - f(np) \) as \( n \) goes to \( \infty \) where \( f \) varies regularly at infinity and the expectation is with respect to the binomial measure \( b(i,n,p) \).

To prove the claims made in Example 7 above, we use Theorem 1 of this chapter: The weak binomial law holds if \( B(n) \to 0 \) as \( n \to \infty \). For variables \( X_i = \pm i^a \),
\[ B(n) = \frac{1}{2^{2n}} \sum_{i=1}^{n} \binom{n}{i} i^{2a} < < n^{2a-\frac{1}{2}} + 0 \]

for \( a < \frac{1}{4} \) by Lemma 1. To show that the weak binomial law fails for \( a > \frac{1}{4} \), we need the central limit theorem for triangular arrays. Let \( Y(n,k) = \binom{n}{k}X_k \) for \( k = 0,1,2,\ldots,n \). Let
\[ S^2(n) = \sum_{k=1}^{n} \binom{n}{k}^2 k^{2a} \]. The theorem states that if
\[ \sum_{k=1}^{n} \int_{\{|Y(n,k)| \geq S(n)\}} Y^2(n,k) dp \to 0 \]
as \( n \to \infty \) for each fixed \( \varepsilon > 0 \), then
the standard normal distribution. The convergence is in distribution (see Billingsley [1968], pg. 42). In the case at hand,

$$\max_{0 \leq k \leq n} |Y(n,k)| \leq \frac{c2^n}{\sqrt{n}} n^a = c2^n n^{a - \frac{1}{2}}$$

for some constant c. By the estimate in Lemma 1, $S^2(n) \sim c'2^n n^{2a - \frac{1}{2}}$ so $S(n) \sim c''2^n n^{a - \frac{1}{2}}$. For fixed $\varepsilon > 0$, the set $\{ |Y(n,k)| \geq \varepsilon S_n \} \subset \{ |Y(n,k)| \geq \varepsilon c''2^n n^{a - \frac{1}{2}} \}$. But, for large n, $|Y(n,k)| \leq c2^n n^{a - \frac{1}{2}} \leq c''2^n n^{a - \frac{1}{2}}$, so these sets are empty for n large. Then the Lindeberg condition (6-16) is satisfied so that

$$\text{Prob}(a \leq \frac{1}{S(n)} \sum_{k=0}^{n} (\binom{n}{k} X_k \leq b) + \phi(b) - \phi(a),$$

where $\phi$ is the standard normal cumulative function. Since $S(n)$ is asymptotically proportional to $2^n n^{a - \frac{1}{2}}$, this shows that it is not possible for

$$\text{Prob}(\frac{1}{2^n} \sum_{i=0}^{n} (\binom{n}{i} X_i | > \varepsilon) \to 0 \quad \text{if} \quad a \geq \frac{1}{4}.$$

It would appear that the use of stronger averages on the variables $X_i$ under consideration should lead to larger ranges of the parameter a for which weak convergence was possible. We now show that using log averages improves the range of a, but only slightly.
Example 8. The variables $X_i$ taking values $\pm \sqrt{\log i}^a$ each with probability $\frac{1}{2}$, obey the strong log law for $a < \frac{1}{2}$ but do not obey the weak log law for $a \geq \frac{1}{2}$. Taking $a = 0$ shows that use of log averages has added the single point $a = \frac{1}{2}$ to the problem considered in Example 6 above.

To prove the claims, consider variables $Y_i = \frac{X_i}{i \log i}$. These have variance $V_i = \frac{1}{i (\log i)^{2-2a}}$ so that for $a < \frac{1}{2}$, $\sum_{i=2}^{n} V_i < \infty$, by the Kolmogorov 3 series theorem this implies $\sum_{i=2}^{n} Y_i$ converges almost surely. Now, the Kronecker lemma implies $\lim_{n \to \infty} \frac{1}{\log n} \sum_{i=2}^{n} \frac{X_i}{i} = 0$ almost surely as was claimed. For $a \geq \frac{1}{2}$, an argument much the same as in the previous example, using the central limit theorem, shows that these variables don't obey the weak law.

6.3 The Strong Law of Large Numbers

This section presents strong laws for a fairly restricted class of weighting schemes (which include log averages) as well as discusses examples which show that weighted averages can change the basic character of convergence. The averages considered are called weighted averages by Peyerimhoff:

$$\frac{1}{W(n)} \sum_{i=1}^{n} w(i)X_i \text{ where } W(n) = \sum_{i=1}^{n} w(i).$$
Theorem 6. Let $X_i$ be independent with mean 0, variance $\sigma_i^2$. The convergence of $\sum_{i=1}^{\infty} (\frac{w(i)}{W(i)})^2 \sigma_i^2$ is a sufficient condition for the almost sure convergence of $\frac{1}{W(n)} \sum_{i=1}^{n} w(i)X_i$ to zero. The condition is also necessary in the sense that if $\sum_{i=1}^{\infty} (\frac{w(i)}{W(i)})^2 \sigma_i^2 = \infty$, there exist variables $Y_i$ with variance $\sigma_i^2$ and mean 0 such that the $Y_i$ don't satisfy the W strong law.

Proof. Sufficiency: The convergence of $\sum_{i=1}^{\infty} (\frac{w(i)}{W(i)})^2 \sigma_i^2$ implies the almost sure convergence of $\sum_{i=1}^{\infty} \frac{w(i)}{W(i)} X_i$ by the three series theorem. This and the Kronecker lemma imply $\frac{1}{W(n)} \sum_{i=1}^{n} w(i)X_i \to 0$ as $n \to \infty$ almost surely.

To prove necessity, note that if $\frac{1}{W(n)} \sum_{i=1}^{n} w(i)X_i \to 0$, almost surely then

$$\frac{w(i)}{W(i)} X_i = \frac{1}{W(i)} \sum_{j=1}^{i} w(j)X_j - \frac{w(i-1)}{W(i-1)} \sum_{j=1}^{i-1} w(j)X_j \to 0$$

so that $\frac{w(i)}{W(i)} X_i \to 0$ as $i \to \infty$ almost surely is a necessary condition for the strong W convergence of $X_i$ to 0. Suppose that $\sum_{i=1}^{\infty} (\frac{w(i)}{W(i)})^2 \sigma_i^2 = \infty$. Define $Y_i$ as independent variables taking values $\pm \frac{w(i)}{W(i)}$ and 0 with probability $\frac{1}{2} (\frac{w(i)}{W(i)})^2 \sigma_i$ and $1 - (\frac{w(i)}{W(i)})^2 \sigma_i$ in the case $\frac{w(i)}{W(i)} < 1$. If $\frac{w(i)}{W(i)} \geq 1$, let $Y_i$ take values $\pm \sigma_i$ each with probability $\frac{1}{2}$. $Y_i$ has mean 0, variance $\sigma_i^2$ and

$$\sum_{i=1}^{\infty} P(|\frac{w(i)}{W(i)} Y_i | \geq 1) = \sum_{i=1}^{\infty} \min(1, (\frac{w(i)}{W(i)})^2 \sigma_i^2) = \infty.$$
And the Borel-Cantelli lemma implies that \( \frac{W(i)}{W(i)} Y_i \geq 1 \) infinitely often with probability 1. |||

**Corollary 1.** The convergence of the series \( \sum_{i=1}^{\infty} \frac{\sigma_r^2}{(k \log k)^2} \) is a sufficient condition for the strong log law. That is, with probability one

\[
\lim_{n \to \infty} \frac{1}{\log n} \sum_{i=1}^{n} \frac{X_i - \mu_i}{i} = 0.
\]

**Example 9.** There exist independent variables \( X_i \) which satisfy the weak law, don't satisfy the strong law but such that the strong log law holds.

**Proof.** Consider \( X_i \) independent with mean 0, variance \( \sigma_i^2 = \frac{1}{\log i} \). Since \( \sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} = \sum_{i=1}^{\infty} \frac{1}{i \log i} \) diverges, applying Theorem 6 to the case of natural density, we may assume the strong law doesn't hold for \( X_i \). Since \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu_i}{i} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{i}{\log i} \to 0 \) as \( n \to \infty \), Theorem 1 of this chapter implies the \( X_i \) satisfy the weak law. Finally,

\[
\sum_{i=2}^{\infty} \frac{\sigma_i^2}{(i \log i)^2} = \sum_{i=2}^{\infty} \frac{1}{i (\log i)^3} < \infty,
\]

so that the strong log law holds for \( X_i \) by Corollary 1.

Even at this late date, the strong law is poorly understood. For example, satisfactory necessary and sufficient conditions for the strong law are unknown (Chung [1951]). Consideration of summability methods should prove a useful tool. In view of
the example above and other preliminary work, we offer the following conjecture: If $X_i$ are independent variables such that $\frac{1}{n} \sum_{i=1}^{n} X_i \to 0$ in probability as $n \to \infty$, then $(1-z) \sum_{i=1}^{\infty} X_i z^i \to 0$ almost surely as $z \to 1$.

As the final result of this section, we prove a theorem which suggests, in principle, much of the theory may be carried over to strong convergence for binomial averages.

**Theorem 7.** Let $X_i$ be independent variables, uniformly bounded by $k$, with mean 0. Then, if

$$B(n) = \frac{1}{2^n} \sum_{i=0}^{n} \binom{n}{i} X_i, \quad \lim_{n \to \infty} B(n) = 0$$

with probability 1.

**Proof.** $E((B(n))^6) \ll \frac{1}{2^n \binom{n}{i}^2} + \frac{1}{2^n \binom{n}{i}^3}$.

By Lemma 1, this is

$$\ll \frac{1}{2n \sqrt{n}} \frac{2^{2n}}{n} + \frac{1}{2n \binom{3n}{n}^2} \ll \frac{1}{n^{3/2}}.$$

Then, for $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} P(|B(n)| > \varepsilon) = \sum_{n=1}^{\infty} P(|B(n)|^3 > \varepsilon^3) \leq \sum_{n=1}^{\infty} \frac{E(|B(n)|^6)}{\varepsilon^6} < \infty.$$ 

So, by the Borel-Cantelli lemma, $B(n) \to 0$ with probability 1.

### 6.4 The Central Limit Problem

The problem of convergence in distribution of properly normalized weighted sums of random variables can be considered
as a special case of the classically solved problem of convergence of general triangular arrays. The only new element is the possibility of an array with infinite rows. We confine our attention to this problem. Since methods in particular cases are similar, the principal problem considered will be convergence of Abel means. We prove a Lindeberg type theorem for variables $X_i$ with variances $\sigma_i^2$ and then a Berry-Essen type theorem in the nonidentical case. These theorems are new but may be regarded as extensions of work reported by Whitt [1972].

Let $X_i$ be independent variables with mean 0, variance $\sigma_i^2$. Let $s^2(z) = \sum_{i=0}^{\infty} z^{2i} \sigma_i^2$ exist for all $z < 1$. This implies that $S(z) = \sum_{i=0}^{\infty} z^i X_i$ is almost surely well defined. We write $N$ for the standard normal random variable with distribution function $\Phi(x)$.

**Theorem 8.** \[ \frac{1}{s^2(z)} \sum_{i=0}^{\infty} z^{2i} \int_{\{|x_i| > \frac{\epsilon s(z)}{z_i}\}} x_i^2 \, dp \rightarrow 0 \]
as $z \rightarrow 1$, for each $\epsilon > 0$ is a sufficient condition for convergence in distribution of $\frac{1}{s(z)} S(z)$ to the standard normal distribution as $z \rightarrow 1$.

**Proof.** We follow the classical proof of Levy as presented in Billingsley [1968], (pg. 43). It is sufficient to show that for any bounded $C^\infty$ function $f$ all of whose derivatives are
bounded, \( E(f(S(z))/S(z)) \to E(f(N)) \) as \( z \to 1 \). Given any such \( f \), let
\[
g(h) = \sup_{-\infty < x < \infty} |f(x+h) - f(x) - f'(x)h - f''(x)\frac{h^2}{2}|.
\]
It is clear that there exists a constant \( k \) depending only on \( f \) such that \( g(h) \leq k \min(h^2, |h|^3) \). With this definition, clearly
\[
(6-17) \quad |\{ f(x+h_1) - f(x+h_2) \} - \{ f'(x)(h_1-h_2) \} + \frac{1}{2} f''(x)(h_1^2-h_2^2) | \leq g(h_1) + g(h_2).
\]
Let \( \{ N(k) \}_{k=0}^\infty \) be \( N(0, \sigma_k^2) \) random variables independent of the \( X_i \) and each other. Let \( T(k,z) = \sum_{0 \leq i < k} z^i X_i + \sum_{k < i < \infty} z^i N(i) \).
There exist almost surely for each \( k, z < 1 \). Note that for fixed \( z \),
\[
(6-18) \quad f(N) - f(S(z)/S(z)) = \sum_{i=0}^\infty \{ f(T(i,z)+z^i N(i)) - f(T(i,z)+z^i X_i) / S(z) \}.
\]
To see this, note that \( T(0,z)+N(0) = \sum_{i=0}^\infty z^i N(i) / S(z) = N \) almost surely and that the sum telescopes. Since \( f \) is continuous, we need only show that \( \lim_{i \to \infty} T(i+1,z) + z^{i+1} N(i+1) = S(z) \) almost surely, but this is obvious from the definitions. Taking expected values of both sides of (6-18), leads to
\[
(6-19) \quad |E(f(S(z)/S(z))) - E(f(N))| \leq \sum_{i=0}^\infty |E(f(T(i,z)+z^i N(i))) - E(f(T(i,z)+z^i X_i)) - E(f(T(i,z)+z^i N(i)))|,
\]
where the interchange of sum and expectation is justified since \( f \) is bounded. By independence of \( T(i, z), N(i) \) and \( X_i \), we have

\[
E\{f'(\frac{T(i, z)}{s(z)}z_i(X_i - N(i))\} = E\{f''(\frac{T(i, z)}{s(z)} \frac{z_i^2}{2} (X_i^2 - N^2(i))\} = 0
\]

so that (6-17) and (6-19) combine to give:

\[
(6-20) \quad |E\{f(\frac{S(z)}{s(z)})\} - E\{f(N)\}| \leq \sum_{i=0}^{\infty} E\{g(\frac{X_i z_i}{s(z)}) + g(\frac{N(i) z_i}{s(z)})\}.
\]

And we need only show that the two sums at the right of (6-20) + 0 as \( z \to 1 \). Given \( \varepsilon > 0 \), let \( R(i) = \{|X_i| \leq \frac{\varepsilon s(z)}{z}\} \).

\[
E\{g(\frac{X_i z_i}{s(z)})\} = \int_{R(i)} g(\frac{X_i z_i}{s(z)}) dp + \int_{CR(i)} g(\frac{X_i z_i}{s(z)}) dp
\]

\[
\leq k \int_{R(i)} |X_i z_i|^3 dp + k \int_{CR(i)} |X_i z_i|^2 dp
\]

\[
\leq \frac{ekz^{2i} \sigma_i^2}{s^2(z)} + \frac{kz^{2i}}{s^2(z)} \int_{CR(i)} |X_i|^2 dp.
\]

Summing this leads to

\[
\sum_{i=0}^{\infty} E\{g(\frac{X_i z_i}{s(z)})\} \leq k \varepsilon + \frac{kz}{s^2(z)} \sum_{i=0}^{\infty} z^{2i} \int_{CR(i)} |X_i|^2 dp + k \varepsilon
\]

as \( z \to 1 \). Since \( \varepsilon \) is arbitrary, the first sum on the right of (6-20) + 0 as \( z \to 1 \). Since this analysis is valid when \( X_i \) is replaced by \( N(i) \), to show the second sum in (6-20) + 0 as \( z \to 1 \), we need only show that
\[
\frac{1}{s^2(z)} \sum_{i=0}^{\infty} \frac{z^{2i}}{\epsilon s(z)} \int_{CR'(i)} |N(i)|^2 dp + 0
\]

where

\[
R'(i) = \{ |N(i)| \leq \frac{\epsilon s(z)_i}{z^i} \}.
\]

This sum is smaller than

\[
(6-21) \quad \frac{1}{s^2(z)} \sum_{i=0}^{\infty} \frac{z^{2i}}{\epsilon s(z)} \int_{-\infty}^{\infty} \frac{|N(i)|^3}{\epsilon s(z)} dp = \frac{E(|N|^3)}{\epsilon s^3(z)} \sum_{i=0}^{\infty} z^{3i} \sigma_i^3.
\]

Now

\[
\frac{z^{2i} \sigma_i^2}{s^2(z)} = \frac{z^{2i}}{s^2(z)} \int_{R(i)} |X_i|^2 dp + \frac{z^{2i}}{s^2(z)} \int_{CR(i)} |X_i|^2 dp
\]

\[
\leq \epsilon^2 + \frac{z^{2i}}{s^2(z)} \int_{CR(i)} |X_i|^2 dp
\]

\[
\leq \epsilon^2 + \frac{1}{s^2(z)} \sum_{j=0}^{\infty} z^{2j} \int_{CR(j)} |X_j|^2 dp + 0
\]

uniformly in \( i \) as \( z \to 1 \). Thus, \( \sup_{0 \leq i \leq \infty} \frac{z^i \sigma_i}{s(z)} \to 0 \) as \( z \to 1 \). So for arbitrary \( \delta \), we can find \( A \) so that for \( A \leq z \leq 1 \),

\[
\sum_{i=0}^{\infty} z^{3i} \sigma_i^3 \leq \delta s(z) \sum_{i=0}^{\infty} z^{2i} \sigma_i^2 = \delta s^3(z).
\]

Using this in (6-21) shows that

\[
\frac{E(|N|^3)}{\epsilon s^3(z)} \sum_{i=0}^{\infty} z^{3i} \sigma_i^3 + 0 \quad \text{as} \quad z \to 1.
\]
Corollary 2. If the $X_i$ have $2+\delta$ moments $d(i)$, then a sufficient condition for convergence in distribution as above is

$$\frac{1}{s^{2+\delta}(z)} \sum_{i=0}^{\infty} z^i (2+\delta) d(i) \to 0 \quad \text{as} \quad z \to 1.$$  

Proof.

$$\frac{1}{s^2(z)} \sum_{i=0}^{\infty} z^{2i} \int \left\{ \frac{|X_i|^{2+\delta}}{s^2(z)} \right\} \leq \frac{1}{s^2(z)} \sum_{i=0}^{\infty} z^{2i} \frac{i^{2+\delta}}{s^2(z)} \int_{-\infty}^{\infty} |X_i|^{2+\delta} dp \to 0 \quad \text{as} \quad z \to 1.$$  

Corollary 3. If $X_i$ are independent and identically distributed with mean 0, variance $\sigma^2$, then

$$\frac{\sqrt{1-z^2}}{\sigma} \sum_{i=0}^{\infty} X_i z^i \overset{d}{\to} N(0,1) \quad \text{as} \quad z \to 1.$$  

Proof.

$$\frac{1}{s^2(z)} \sum_{i=0}^{\infty} z^{2i} \int \left\{ \frac{|X_i|^{2+\delta}}{s^2(z)} \right\} \leq \frac{1-z^2}{\sigma^2} \sum_{i=0}^{\infty} z^{2i} \int \left\{ \frac{|X_i|^{2+\delta}}{\sqrt{1-z^2}} \right\} \leq \frac{1}{\sigma^2} \int \frac{|X_i|^{2+\delta}}{\sqrt{1-z^2}} dp \to 0 \quad \text{as} \quad z \to 1.$$  

Remark. A straightforward modification of the proofs above yield similar results for zeta and Borel averages of random variables. Instead of giving details, we proceed to the corresponding Berry-Esseen results.
The first theorem proved is a central limit theorem with rate of convergence for zeta averages of i.i.d. variables with mean 0, variance $\sigma^2$, and $E(|X_i|^3) = \rho$. The infinite convolution $\sum_{i=1}^{\infty} \frac{X_i}{i^s}$ converges almost surely for all $s > \frac{1}{2}$ since the corresponding variance sum is $\sum_{i=1}^{\infty} \frac{\sigma^2}{i^{2s}} = \sigma^2 \zeta(2s)$.

Let

$$\frac{1}{\sigma \sqrt{\zeta(2s)}} \sum_{k=1}^{\infty} \frac{X_k}{k^s}$$

have distribution function $F_s(x)$.

Theorem 9. $\sup_{-\infty < x < \infty} |F_s(x) - \phi(x)| \leq \frac{c_0}{\sigma^3} \frac{1}{\sqrt{\zeta(2s)}}$

for all $s > \frac{1}{2}$. Here $c_0$ is a universal constant independent of the distribution of the $X_i$. In particular, since $\zeta(2s) \to \infty$ as $s \to \frac{1}{2}$, the normalized zeta average of the $X_i$ converge in distribution to the normal distribution.

Proof. The proof requires standard inequalities which will be quoted rather than rederived. The page references that follow are to Feller [1966]. If $f(t)$ is the characteristic function of the underlying distribution of $X_i$, let $g(s,t)$ be the characteristic function of the distribution $F_s(x)$ so

$$g(s,t) = \frac{1}{\sigma \sqrt{\zeta(2s)}} \sum_{k=1}^{\infty} f\left(\frac{t}{\sigma \sqrt{\zeta(2s)} k^s}\right).$$

The standard smoothing lemma (pg. 512) states:
\begin{equation}
|F_s(x) - \phi(x)| \leq \frac{1}{\pi} \int_{-T}^{T} \frac{1}{|y|} \left| g(s,y) - e^{-\frac{y^2}{2}} \right| dy + \frac{24m}{\pi T}
\end{equation}

for all \( x \) and \( T > 0 \), with \( m = \frac{1}{\sqrt{2\pi}} \). Also, pg. 516 gives the bound

\begin{equation}
|\log f(t) - \frac{\sigma^2 t^2}{2}| \leq \frac{5}{12} \rho |t|^3 \quad \text{for} \quad |t| < \frac{\sigma^2}{\rho}.
\end{equation}

Thus

\begin{equation}
|\log(g(s,y)) - \frac{y^2}{2}| \leq \sum_{k=1}^{\infty} |\log f\left( \frac{y}{\sigma \sqrt{\zeta(2s)} k^2} \right) - \frac{y^2}{2 \zeta(2s) k^2}| \leq \frac{5}{12} \frac{\rho |y|^3 \zeta(3s)}{(\zeta(2s))^{3/2} \sigma^3}
\end{equation}

valid for \( |y| \leq \frac{\sigma^3}{\rho} \sqrt{\zeta(2s)} k^5 \) or, for \( |y| \leq \frac{\sigma^3}{\rho} \sqrt{\zeta(2s)} \). Taking \( T = \frac{\sigma^3}{\rho} \sqrt{\zeta(2s)} \) in (6.22), we have the integrand bounded by

\begin{equation}
\frac{1}{|y|} \exp\left[ \frac{5}{12} \frac{\rho |y|^3 \zeta(3s)}{\sigma^3 (\zeta(2s))^{3/2}} \right] - 1 |.
\end{equation}

Since \( |e^{t-1}| \leq |t| e^{|t|} \), this last expression is bounded by

\begin{equation}
\frac{5}{12} \frac{\rho \sigma^3}{\sigma^3 (\zeta(2s))^{3/2}} \exp\left[ -\frac{y^2}{2} + \frac{5}{12} \frac{\rho |y|^3 \zeta(3s)}{\sigma^3 (\zeta(2s))^{3/2}} \right] \leq \frac{5}{12} \frac{\rho |y|^2 \zeta(3s)}{(\zeta(2s))^{3/2}} \exp\left[ -\frac{y^2}{2} + \frac{5}{12} \frac{\zeta(3s)}{\sqrt{\zeta(2s)}} \right].
\end{equation}

Using (6.25) in (6.22), extending the limits of integration to \(-\infty\) to \(\infty\) and evaluating the integral, leads to
\[ |F_s(x) - \phi(x)| \leq \frac{5}{24} \frac{\rho}{\sigma^3} \frac{\zeta(3s)}{(\zeta(2s))^{3/2}} \left( \frac{1}{2} - \frac{5}{12} \frac{\sqrt{\pi}}{\zeta(3s)} \right)^{3/2} + \frac{24\rho}{\pi \sqrt{2\pi} \sigma^3 \sqrt{\zeta(2s)}} \]

\[ = \frac{\rho}{\sigma^3} \frac{1}{\sqrt{\zeta(2s)}} [c(s) + \frac{24}{\pi \sqrt{2\pi}}] \]

where \( c(s) \to 0 \) as \( s \to 1 \). This clearly proves the result stated and we omit a computation of numerical bounds for the term in square brackets. |||

As a final theorem, let \( X_k \) be independent with first 3 moments \( \mu(k) \), \( \sigma^2(k) \), and \( \rho(k) = E(|X_k|^3) \). Let \( s^2(z) = \sum_{i=0}^{\infty} \sigma_i^2 z^{2i} \) and \( r(z) = \sum_{i=0}^{\infty} \rho(i)z^{3i} \) exist for \( 0 \leq z < 1 \). The existence of \( s^2(z) \) implies that \( \frac{1}{s(z)} \sum_{i=0}^{\infty} (X_i - \mu(i))z^i \) is almost surely well defined for \( 0 \leq z < 1 \), let \( F(z,x) \) be the distribution function of this last sum.

**Theorem 10.** \[ \sup_{-\infty < x < \infty} |F(z,x) - \phi(x)| \leq \frac{cr(z)}{s^3(z)} \] where \( c \leq 6 \), for all \( z \).

**Proof.** Here again, page references are to Feller [1966]. If \( f(t) \) is the characteristic function of a distribution with 3 moments, the expansion of \( e^{it} \) easily yields the inequality

\[ |f(t) - 1 + \frac{1}{2} \sigma^2 t^2| \leq \frac{1}{6} \rho|t|^3 \]

or

(6-26) \[ |f(t)| \leq 1 - \frac{1}{2} \sigma^2 t^2 + \frac{1}{6} \rho|t|^3 \quad \text{if} \quad \frac{1}{2} \sigma^2 t^2 \leq 1. \]

If \( f_k(t) \) are the characteristic functions of the components \( X_k - \mu(k) \), the characteristic function of \( F(z,x) = f(z,t) \)

\[ = \prod_{i=1}^{\infty} f_k \left( \frac{t z^i}{s(z)} \right). \]

The smoothing inequality (pg. 512) says
(6-27) \[ |F(z,x) - \Phi(x)| \leq \frac{1}{\pi} \int_{-T}^{T} |f(z,y) - e^{-\frac{1}{2}y^2}| \frac{dy}{|y|} + \frac{24}{\pi \sqrt{2} \pi T} \]

for all \( T > 0 \). \( T \) will be taken as \( \frac{8}{7} \frac{s^3(z)}{r(z)} \). To bound the integrand in (6-27), we use the equality:

(6-28) \[ \sum_{i=0}^{\infty} a(i) - \sum_{b=0}^{\infty} b(i) = \sum_{k=0}^{\infty} \sum_{i=0}^{k-1} (a(k) - b(k)) \sum_{i=0}^{\infty} a(i) \sum_{i=k+1}^{\infty} b(i) \]

or, if \( |a(i)|, |b(i)| \leq |c(i)| \),

(6-29) \[ \sum_{i=0}^{\infty} a(i) - \sum_{i=0}^{\infty} b(i) \leq \sum_{k=0}^{\infty} |a(k) - b(k)| \sum_{i=0}^{\infty} |c(i)|. \]

Here, \( a(k) = f_k(tz^k) \), \( b(k) = \exp[-\frac{1}{2} \frac{(\sigma(k)z^k t^2)}{s(z)}] \). We proceed to bound the terms \( a(k), b(k) \) after which it will be clear that the products and sums in (6-28) and (6-29) converge.

From (6-26) above, it follows that

(6-30) \[ |f_k(tz^k)| \leq 1 - \frac{1}{2} \frac{(t z^k \sigma(k))^2}{s(z)} + \frac{\rho(k) z^{3k} t^3}{6 s(z)^3} \]

\[ \leq \exp[-\frac{1}{2} \frac{(\sigma(k)z)^2}{s(z)} + \frac{\rho(k) z^{3k} t^3}{6 s(z)^3}] \]

holds for \( \frac{\sigma(k)z^k}{s(z)} T \leq \sqrt{2} \).

Let

\[ c(k) = \exp[-\frac{1}{2} \frac{(\sigma(k)z^k)^2}{s(z)} + \frac{3}{8} \frac{\rho(k) z^{3k} t^3}{s(z)^3}] \cdot \]

Clearly \( |b(k)| \leq c(k) \). Using (6-30), we see \( |a(k)| \leq c(k) \) if
\[ \frac{\sigma(k)z^k}{s(z)} T \leq \frac{4}{3} < \sqrt{2} \text{. If } \frac{\sigma(k)z^k}{s(z)} T > \frac{4}{3}, \text{ then we shall see } c(k) \geq 1 \text{ and since } |f_k(y)| \leq 1, \text{ this would give } |a(k)| \leq c(k). \text{ Consider the exponent in the definition of } c(k). \text{ Using the now assumed inequality for } T, \text{ this is:} \]

\[ -\frac{1}{2}\left(\frac{\sigma(k)z^k}{s(z)}\right)^2 + \frac{3}{8} \frac{\sigma(k)z^{3k}T}{s(z)^3} \geq -\frac{1}{2}\left(\frac{\sigma(k)z^k}{s(z)}\right)^2 + \frac{\sigma(k)}{2} \frac{s(z)z^{3k}}{\sigma(k)s(z)^3} z^k \]

\[ \geq 0 \]

iff \( \sigma(k)^3 \leq \sigma(k) \), but this last always holds by the classical moment inequality, so \(|b(k)| \leq c(k)\) for all \( k, z \).

The theorem as stated is trivially true for \( \frac{6r(z)}{s(z)^3} > 1 \), or recalling \( T = \frac{8}{9} \frac{s(z)^3}{r(z)} \), for \( T = \frac{16}{3} \). Consider only values of \( z \) such that \( T > \frac{16}{3} \). Use of \( \sigma(k)^3 \leq \sigma(k) \) in the expression for \( c(k) \) leads to

\[ c(k) \geq \exp\left[\frac{1}{2}\left(\frac{t\sigma(k)z^k}{s(z)}\right)^2 \left[ -1 + \frac{3}{4} \frac{\sigma(k)z^kT}{s(z)} \right] \right]. \]

c(\( k \)) will be smallest for those \( k \) such that

\[(6.31) \quad \frac{3}{4} \frac{\sigma(k)}{s(z)} T < 1 \text{ or } \frac{\sigma(k)z^k}{s(z)} < \frac{4}{3} < \frac{1}{4}. \]

We also have the obvious bound

\[ c(k) \geq \exp(-\frac{t^2}{2}(\frac{\sigma(k)z^k}{s(z)})^2). \]

Combining this with (6.31), we see \( c(k) \geq \exp(-\frac{t^2}{2}(\frac{1}{4})^2) = e^{\frac{-t^2}{32}}. \)
Using this and the definition of $c(k)$:

\[
(6-32) \quad \sum_{k=1}^{\infty} c(k) \leq \exp(t^2(-\frac{1}{2} + \frac{3}{8} \frac{r(z)}{s(z)^3} T + \frac{1}{32}))
\]

\[
= \exp(t^2(-\frac{1}{2} + \frac{1}{3} + \frac{1}{32})) < \exp\left(\frac{t^2}{32}\right).
\]

Also,

\[
\sum_{k=0}^{\infty} |a(k)-b(k)| \leq \sum_{k=0}^{\infty} \left( |f_k^t(t^kz) - 1 + \frac{1}{2}(t^k\sigma(k))^2| + |1 - \frac{1}{2}(t^k\sigma(k))^2 - \exp(-\frac{1}{2}(\frac{\sigma(k)z^k}{s(z)})^2)| \right)
\]

\[
\leq \sum_{k=0}^{\infty} \frac{1}{6} \frac{\sigma(k)z^{3k}}{s(z)} \frac{t^3}{3} + \frac{t^4}{8s(z)^4} \sum_{k=0}^{\infty} \sigma(k)^4 z^{4k}
\]

\[
= \frac{|t|^3 r(z)}{6s(z)^3} + \frac{t^4}{8s(z)^4} \sum_{k=0}^{\infty} \sigma(k)^4 z^{4k}
\]

The moment inequality implies

\[
\sigma(k)^4 z^{4k} \leq \rho(k)^{4/3} z^{4k} = (\rho(k)z^{3k})^{1/3} \rho(k)z^{3k}
\]

\[
\leq r(z)^{1/3} \rho(k)z^{3k},
\]

so that

\[
(6-33) \quad \sum_{k=0}^{\infty} |a(k)-b(k)| \leq \frac{|t|^3 r(z)}{6s(z)^3} + \frac{t^4}{8s(z)^3} \left(\frac{r(z)}{s(z)}\right)^{\frac{4}{3}}
\]

\[
\leq \frac{|t|^3 r(z)^{5/3}}{6s(z)^3} + \frac{t^4}{8s(z)^3} \left(\frac{r(z)}{s(z)}\right)^{\frac{5}{9}}.
\]

Using (6-32) and (6-33) in (6-29), shows that the integrand in (6-27) is bounded by
\[
\frac{8}{9} \frac{t^2}{6} + \frac{5}{72} t^3 e^{-\frac{t^2}{8}}.
\]

Use of
\[
\int_0^\infty x^3 e^{-ax^2} \, dx = \frac{1}{2a^2}, \quad \int_0^\infty x^2 e^{-ax^2} \, dx = \frac{1}{4a} \sqrt{\pi}
\]
yields
\[
|F(z,x) - \Phi(x)| \leq c \frac{r(z)}{s(z)^3}
\]

with \( c \leq 6 \) as was to be shown. |||
BIBLIOGRAPHY


Davenport, H. Multiplicative Number Theory, Markham, Chicago, 1967.


