SOME TAUBERIAN THEOREMS RELATED TO COIN TOSSING

BY

PERSI DIACONIS and CHARLES STEIN

TECHNICAL REPORT NO. 87
JUNE 1976

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MPS74-21416

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SOME TAUBERIAN THEOREMS RELATED TO COIN TOSSING

BY

PERSI DIACONIS and CHARLES STEIN

TECHNICAL REPORT NO. 87
JUNE 1976

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MPS74-21416

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
1. INTRODUCTION AND STATEMENT OF RESULTS

A subset $A$ of the set $\mathbb{N}$ of all non-negative integers is said to have Euler density $\lambda$ (with parameter $p \in (0,1)$) if

$$\lim_{n \to \infty} \sum_{i \in A} b(i,n,p) = \lambda$$

where

$$b(i,n,p) = \begin{cases} \binom{n}{i} p^i (1-p)^{n-i} & \text{for } i \in \{0, \ldots, n\} \\ 0 & \text{otherwise.} \end{cases}$$

We shall also say that $A$ has $E_p$ density $\lambda$ if (1-1) holds. This notion was introduced by Euler in order to manipulate divergent series. Modern references are Hardy (1949) and Peyerimhoff (1969). The principal result of this note is that the existence and value of $E_p$ density does not depend on the value of $p$. In greater detail, we have

**Theorem 1:** For any $A \subseteq \mathbb{N}$ and $p \in (0,1)$ the following assertions are equivalent:

(1-2) $A$ has $E_p$ density $\lambda$,

(1-3) $\lim_{\lambda \to \infty} e^{-\lambda} \sum_{i \in A} \frac{\lambda^i}{i!} = \lambda$,

(1-4) for all $\varepsilon > 0$, $\lim_{n \to \infty} \frac{1}{\varepsilon \sqrt{n}} \sum_{i \in A} 1 = \lambda$.  

\[ \Box \]
This was conjectured independently by Erdős and Gleason, at least in part. The proof is a straightforward application of Wiener's Tauberian theorem after an appropriate variance-stabilizing transformation has been made and the normal approximation to the binomial used. Variance-stabilizing transformations, as described, for example, in Anscombe (1948), are familiar tools to statisticians. The summability methods considered here can be applied to sequences \( \{a_i\}_{i=0,1,2,...} \) rather than sets, which can be thought of as sequences of 0s and 1s. For example, such a sequence is said to have \( E_p \) density \( \lambda \) if

\[
\lim_{n \to \infty} \frac{b(i,n,p)}{n^{\lambda/\lambda_i}} = \lambda.
\]

The extension of Theorems 2 and 3 to bounded sequences is straightforward. For arbitrary unbounded sequences it is known that the \( E_p \) summability methods are inequivalent (Hardy (1949), pg. 180).

An example will be useful in comparing rates of convergence of different densities.

**Example 2:** If \( A \) is the set of multiples of an integer \( a \), then

\[
\left| \sum_{i \in A} b(i,n,p) - \frac{1}{a} \right| \leq e^{-8np(1-p)/a^2}.
\]

Thus \( A \) has \( E_p \) density \( \frac{1}{a} \) for any fixed \( p \), \( 0 < p < 1 \).
Often the most readily available information about a set of integers is that it has Cesaro density \( C_1 \) density with a given rate of convergence. The next theorem asserts that if this rate is better than \( 1/\sqrt{n} \) the set has \( E_p \) density. A positive, measurable, real-valued function \( L \) is **slowly varying at infinity** if for any \( a > 0 \), \( \lim_{x \to \infty} L(ax) = L(a) \). A measurable real valued function is said to vary **regularly at infinity** with exponent \( \rho, -\infty < \rho < \infty \), if \( f(x) = x^\rho L(x) \) for \( 0 < x < \infty \). Seneta (1976) contains the basic facts about regular variation. We can now state:

**Theorem 3:** If \( A \subseteq \mathbb{N} \) has \( E_p \) density \( \ell \) then \( A \) has \( C_1 \) density \( \ell \).

Conversely, if

\[
(1-5) \quad \left| \frac{1}{n} \sum_{i \in A} 1 \right| \leq g(n)
\]

where \( g(x) \) varies regularly at infinity with exponent \( \rho, -1 < \rho < -\frac{1}{2} \), then

\[
(1-6) \quad \left| \sum_{i \in A} b(i,n,p) - \ell \right| \leq \frac{k g(n)}{\sqrt{n}}
\]

for some constant \( k \).

The set \( A \), which does not include \( 0 \), includes \( 1, 2, 3 \), does not include \( 4 \) through \( 7 \), includes the next five integers, and so on, is easily seen to have \( C_1 \) density \( \frac{1}{2} \) but is not \( E_{1/2} \) summable to any limit. Similarly, it is easy to construct examples of sets \( A \) with arbitrarily slow rate of convergence in (1-5) which have \( E_p \) density. Example 2 shows that the information provided by (1-6) can be far from
best possible. Here is an example where the rate of convergence given by Theorem 3 is the best rate available. An integer is square-free if it has no squared prime factors. Let \( Q \subseteq \mathbb{N} \) be the set of square-free numbers. Using a known result due to Walfisz along with Theorem 3 yields

**Corollary 4:**

\[
\left| \sum_{i \in Q} b(i, n, p) - \frac{6}{\pi^2} \right| \leq E(n)
\]

where

\[
E(n) \leq C_1 \exp\left\{ -C_2 \log^{3/5} n (\log \log n)^{-1/5} \right\}
\]

for \( C_1, C_2 \) constants. If the Riemann Hypothesis is true, then

\[
E(n) \leq C_3 n^{-1/10 + \varepsilon}
\]

for any \( \varepsilon > 0 \).

Probabilistically, both binomial and uniform densities have been used as models for the random division of a set of counters into two piles. Laplace, in a controversy reported in Todhunter (1965, pg. 200, 465), argued that the binomial model was more appropriate in determining if the number of counters in one pile was odd or even. Gardner (1973) discusses the "random" division of a pile of sticks in connection with the randomization mechanism used in the I-Ching. Theorem 3 says that if \( n \) is reasonably large the choice of model will not make much practical difference.
2. PROOF OF THEOREMS

Proof of Theorem 1: That (1-2) implies (1-3) is Theorem 119 in Hardy (1949). The main steps of the remaining parts of the proof will be stated as a sequence of auxiliary lemmas. Letting \( A \subset \mathbb{N} \) be a fixed set throughout the proof, define the real valued step function:

\[
(2-1) \quad f(x) = \begin{cases} 
1 & \text{if } 2\sqrt{2i} < x < 2\sqrt{2(i+1)} \text{ for some } i \in A \\
0 & \text{otherwise.}
\end{cases}
\]

**Lemma 5:** \( A \) has \( \mathbb{P} \) density \( \lambda \) if and only if

\[
(2-2) \quad \lim_{t \to \infty} \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} f(x+t) e^{-x^2/2\sigma^2} \, dx = \lambda \quad \text{with } \sigma^2 = 2(1-\rho).
\]

**Lemma 6:** (1-3) holds if and only if (2-2) holds with \( \sigma^2 = \sqrt{2} \).

**Lemma 7:** (1-4) holds if and only if for any \( \varepsilon > 0 \)

\[
(2-3) \quad \lim_{t \to \infty} \frac{1}{\varepsilon} \int_{0}^{\varepsilon} f(x+t) \, dx = \lambda.
\]

Since the Fourier transform of the function

\[
\frac{1}{\sqrt{2\pi \sigma^2}} e^{-x^2/2\sigma^2}
\]

does not vanish, Wiener's Tauberian theorem (see, for example, Hardy (1949), Chapter 12) implies that if the limit in (2-2) exists for any \( \sigma^2 > 0 \) it exists for all \( \sigma^2 > 0 \). Thus (1-2) is equivalent to (1-3).
Further, Wiener's Tauberian theorem implies that if the limit (2-2) exists then the limit (2-3) exists. Thus (1-2) or (1-3) imply (1-4). That (1-4) implies (1-2) requires a separate argument.

In what follows, $k$ denotes an unspecified constant which need not be the same from equation to equation. We use the $O$, $o$ notation as described in Hardy (1949), pg. xvi.

Proof of Lemma 5: Using bounds for the normal approximation to the binomial measure, as given in Feller (1968, Chapter 7) or Hardy (1949, Theorem 138), we see that

\[
(2-4) \quad \sum_{i \in A} \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i \in A} \left(2np(1-p)\right)^{-1/2} \exp \left\{ \frac{-(i-mp)^2}{2np(1-p)} \right\} + R_{n,i}
\]

where $R_{n,i}$ satisfies

\[
(2-5) \quad \sum_{i=0}^{\infty} |R_{n,i}| \leq \frac{k}{\sqrt{n}}.
\]

Next we compute, writing $\sigma^2 = 2(1-p)$,

\[
(2-6) \quad \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(x + 2\sqrt{2np}) e^{-x^2/2\sigma^2} dx
\]

\[
= \sum_{i \in A} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{2\sqrt{2i+1}-2\sqrt{2np}}^{2\sqrt{2i+1}} e^{-x^2/2\sigma^2} dx
\]

\[
= \sum_{i \in A} \frac{e^{-2(\sqrt{2i+1}-\sqrt{2np})^2/\sigma^2}}{\sqrt{\pi\sigma^2}} + R'_{n,i}
\]

6
where \( R'_{n,i} \) satisfies a condition analogous to (2-5). In comparing the sums in (2-4) and (2-6) we are free to only consider \( i \) satisfying

\[
(2-7) \quad i \in S_n \quad \text{where} \quad S_n = \{ i : |i-np| < \sqrt{n \log^2 n} \}
\]

since well known bounds on the tails of these sums (Feller (1968), pg. 151) show they are negligible for large \( n \). Thus, the difference between (2-4) and (2-6) is

\[
(2-8) \quad \sum_{i \in S_n} \frac{e^{-(i-np)^2/2np(1-p)}}{\sqrt{2\pi np(1-p)}} \left[ 1 - \sqrt{np} e^{f(i,n,p)} \right] + o(1),
\]

where

\[
f(i,n,p) = \left( \frac{(i-np)^2}{2np(1-p)} - \frac{1}{(1-p)} \right) (\sqrt{2i} - \sqrt{2np})^2 = \frac{(\sqrt{2i} - \sqrt{2np})^3}{8np(1-p)} \{ \sqrt{2i} + 3\sqrt{2np} \}.
\]

A straightforward argument shows that, for \( i \in S_n \),

\[
(2-9) \quad f(i,n,p) = O \left( \frac{(\log n)^8}{\sqrt{n}} \right),
\]

while clearly, for \( i \in S_n \),

\[
\sqrt{np} = 1 + O \left( \frac{(\log n)^2}{\sqrt{n}} \right).
\]

Using this and (2-9) in (2-8) shows that the sum in (2-8) goes to zero as \( n \) goes to \( \infty \). It remains to relate the integral on the left of expression (2-6) to the integral (2-2). Toward this end we calculate for \( u \) and \( t \) real numbers,
(2-10) \[ \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} \left| f(x+t) - f(x+u) \right| e^{-x^2/2\sigma^2} \ dx \]

\[ \leq \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} \left| e^{-(x-t)^2/2\sigma^2} - e^{-(x-u)^2/2\sigma^2} \right| \ dx \]

\[ = \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-|u-t|/2}^{1/2} e^{-x^2/\sigma^2} \]

\[ - |u-t|/2 \]

\[ \leq \frac{1}{\sqrt{2\pi \sigma^2}} |u-t| . \]

For given \( t \) define \( n(t) = \lfloor t^2/8p \rfloor \) where \( \lfloor x \rfloor \) denotes the greatest integer less than or equal to \( x \). Then clearly \( 2\sqrt{2\pi n(t)} = t + O\left( \frac{1}{t} \right) \) as \( t \to \infty \). This and (2-11) show that

\[ \lim_{t \to \infty} \int_{-\infty}^{\infty} f(x+t)e^{-x^2/2} \ dx = \lim_{n \to \infty} \int_{-\infty}^{\infty} f(x + 2\sqrt{2\pi n})e^{-x^2/2} = \lim_{n \to \infty} \sum_{i \in \mathbb{A}} b(i,n,p) . \]

This completes the proof of Lemma 5.

Proof of Lemma 6: The normal approximation to the Poisson measure as in Feller (1968, pg. 194) or Hardy (1949, Theorem 137) implies that

\[ e^{-\lambda} \sum_{i \in \mathbb{A}} \frac{\lambda^i}{i!} = \sum_{i \in \mathbb{A}} e^{-(i-\lambda)^2/2\lambda} + R_{1,\lambda} \]

where \( \sum_{i=0}^{\infty} |R_{i,\lambda}| < \frac{k}{\sqrt{\lambda}} \). From here, the proof of Lemma 5 holds essentially word for word.
Proof of Lemma 7: We easily compute that

\[
\frac{1}{\varepsilon} \int_0^\varepsilon f(x+t)dx = \frac{1}{t} \sum_{t < 2\sqrt{2i} < t + \varepsilon} (2\sqrt{2(i+1)} - 2\sqrt{2i}) + \mathcal{O}(\frac{1}{t})
\]

\[
= \frac{1}{\varepsilon t} \sum_{t^2/8 < i < (t+\varepsilon)^2/8} 1 + \mathcal{O}(\frac{1}{t})
\]

\[
= \frac{1}{\varepsilon t} \sum_{t^2/8 < i < t^2/8 + \varepsilon t/4} 1 + \mathcal{O}(\frac{1}{t})
\]

\[
= \frac{1}{\varepsilon \sqrt{x/2}} \sum_{i \in A} 1 + \mathcal{O}(\frac{1}{\sqrt{x}})
\]

where \( x = t^2/8 \). The last sum is (1-4) with \( \varepsilon \) replaced by \( \varepsilon/\sqrt{2} \).

This completes the proof of Lemma 7.

\[\square\]

It only remains to show that (1-4) implies (1-2). Let

\[
b_i = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{otherwise.} \end{cases}
\]

Let \( a_i = b_i - i \). Condition (1-4) becomes

(2-11) for any \( \varepsilon > 0 \), \( \delta > 0 \), there is an \( N \) so that for \( n > N \),

\[
| \sum_{n \leq i < n + \varepsilon \sqrt{n}} a_i | \leq \delta \sqrt{n}.
\]
We first show that if (2-11) holds, then it holds uniformly in \( \varepsilon \).

Specifically,

\[
(2-12) \quad \text{for any positive real numbers } \delta < a < b, \text{ there is an } N \text{ so that for } n > N, \quad \left| \sum_{n \leq i \leq n + s \sqrt{n}} a_i \right| < \delta \sqrt{n} \quad \text{for } s \in [a, b].
\]

To prove (2-12), find \( N \) so large that \( n > N \) implies \( b / \sqrt{n} < 1 \) and

\[
(2-13) \quad \left| \sum_{n \leq i \leq n + r \sqrt{n}} a_i \right| \leq t \sqrt{n}
\]

with \( r = \delta / 8, t = \delta^2 / 16b \). Then, for \( n > N \), let \( x_0 = n \), \( x_1 = n + r \sqrt{n} \), and inductively, \( x_{i+1} = x_i + r \sqrt{x_i} \). Let \( I \) be the index such that \( x_i \leq n + s \sqrt{n} < x_{i+1} \). Thus \( x_{I+1} - x_I \leq 2r \sqrt{n} \) and

\[
\left| \sum_{n \leq i \leq n + s \sqrt{n}} a_i \right| \leq \sum_{i=0}^{I} \left| \sum_{x_i < j < x_{i+1}} a_j \right| + 2r \sqrt{n}
\]

\[
\leq \sum_{i=0}^{I} t \sqrt{x_i} + 2r \sqrt{n}
\]

\[
\leq \frac{ts}{r} \sqrt{n} + 2r \sqrt{n}
\]

\[
\leq \delta \sqrt{n}.
\]

This completes the proof of (2-12).

We must show that

\[
(2-14) \quad \left| \frac{1}{2^n} \sum_{i=0}^{n} \binom{n}{i} a_i \right| < \delta.
\]
Using the boundedness of the sequence \( a_i \) and the Central Limit Theorem (Feller (1968), Chapter 7), we first choose numbers \( N_1 \), \( w \), and \( z \), such that

\[
(2-15) \quad \left| \frac{1}{2^n} \sum_{i=0}^{n} (\binom{n}{i}) a_i \right| < \frac{\delta}{10} + \left| \frac{1}{\sqrt{n}} \sum_{\frac{n}{2} - w\sqrt{n} \leq i \leq \frac{n}{2} + z\sqrt{n}} \binom{n}{i} a_i \right|.
\]

Let \( N_2 > N_1 \) be so large that for \( n > N_2 \),

\[
(2-16) \quad \left| \sum_{i \in S(s)} a_i \right| < \frac{\delta}{10} \sqrt{n}
\]

uniformly for \( 0 < s < z - w \) where \( S(s) = \{ i : \frac{n}{2} - w\sqrt{n} \leq i \leq \frac{n}{2} - w\sqrt{n} + s\sqrt{n} \} \).

Summation by parts now shows that the sum on the right side of (2-15) can be written as

\[
(2-17) \quad \frac{1}{2^n} \sum_{i \in S(s)} \left( (\binom{n}{i}) - (\binom{n}{i+1}) \right) A(i) + g(n)
\]

where

\[
A_i = \frac{1}{\binom{n}{i}} \sum_{j=n/2-w\sqrt{n}}^{i} a_j,
\]

\[
|g(n)| < \frac{\delta}{10} \sqrt{n}.
\]

Using (2-16), the sum in (2-17) is bounded by

\[
(2-18) \quad \frac{\sqrt{n}}{2^n} \frac{\delta}{10} \sum_{i \in S(s)} \left| (\binom{n}{i}) - (\binom{n}{i+1}) \right| + \frac{\delta}{10}.
\]

We also clearly have

\[
\sum_{i \in S(s)} \left| (\binom{n}{i}) - (\binom{n}{i+1}) \right| \leq \sum_{i=0}^{n} \left| (\binom{n}{i}) - (\binom{n}{i+1}) \right| < \frac{k2^n}{\sqrt{n}}.
\]

Using this in (2-18) completes the proof of (2-14) and Theorem 1.
Proof of Example 2: If \( w = e^{2\pi i/a} \) is a primitive \( a \)th root of unity, it is elementary that for integers \( h \),

\[
\sum_{j=0}^{a-1} w^{jh} = \begin{cases} 
a & \text{if } a \text{ divides } h \\
0 & \text{otherwise.}
\end{cases}
\]

From this we have an identity of C. Ramus (Knuth (1973), pg. 70):

\[
\frac{1}{a} \sum_{j=0}^{a-1} (w^j p + (1-p))^n = \sum_{j>0} \binom{n}{n-j} p^j (1-p)^{n-j}
\]

(using the convention that \( \binom{b}{c} = 0 \) for integers \( c > b \)). The left-hand side of (2.19) is \( \frac{1}{a} + E(n,p,a) \) where the error term

\[
E(n,p,a) = \frac{1}{a} \sum_{j=1}^{a-1} \left( p e^{2\pi i j/a} + (1-p) \right)^n.
\]

Writing \( \phi = \frac{2\pi i}{a} \), define \( R \) and \( \theta \) by \( ((1-p) + p e^{i\phi}) = Re^{i\theta} \). Thus

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

A Taylor expansion and use of \( \sin x \geq \frac{2}{\pi} x \) for \( 0 \leq x \leq \frac{\pi}{2} \) yield

\[
\left| (1-p) + p e^{2\pi i j/a} \right|^n = \left| (1-p) + p e^{2\pi i (a-j)/a} \right|^n \leq e^{-\delta n p(1-p)j^2/a^2}.
\]

The required upper bound for \( E(n,p,a) \) follows by replacing \( j \) by 1.

A slightly more detailed estimate gives

\[
E(n,p,a) \leq \left( \frac{np(1-p) + ca^2}{anp(1-p)} \right) \exp \left[ -\frac{cnp(1-p)}{a^2} \right]
\]

with \( c = 4\pi \).
Proof of Theorem 3. If $A \subset N$ has $E_p$ density $\lambda$, then it is well known that $\lim_{x \to \infty} \frac{1}{x} \sum_{i \in A} x_i = \lambda$. (See Peyerimhoff (1969), pg. 25, for a short proof.) The Hardy-Littlewood Tauberian theorem (Peyerimhoff, pg. 81) implies that $A$ has $C_1$ density $\lambda$. For the converse, let

$$a_i = \begin{cases} 1 & i \in A \\ 0 & \text{otherwise.} \end{cases}$$

Assume that $\sum_{i=0}^{n} a_i = \lambda n + O(f(n))$ where $f(n)$ varies regularly at infinity with exponent $\rho$, $0 < \rho < \frac{1}{2}$. Throughout the proof let $p$, $0 < p < 1$, be fixed and let $q = 1 - p$. We must show that

$$(2-20) \quad \sum_{i=0}^{n} a_i p^i q^{n-i}(\frac{n}{i}) = \lambda + O\left(\frac{f(n)}{\sqrt{n}}\right).$$

Write the left side of (2-20) as

$$\sum_{i=0}^{n} a_i p^i q^{n-i}(\frac{n}{i}) = q^n \sum_{i=0}^{n} a_i (\frac{n}{i})^i$$

with $r = \frac{p}{q}$. It is convenient to deal with the sum in the second form.

Summation by parts gives

$$(2-21) \quad \sum_{i=0}^{n} a_i (\frac{n}{i})^i = A(n) + \sum_{j=0}^{n-1} A(j)\Delta(j)$$

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

$$(2-22) \quad \Delta(j) = \binom{n}{j}r^j - \binom{n}{j+1}r^{j+1} = \binom{n}{j}r^j \left\{ \frac{(r+1)j + 1 - rn}{j+1} \right\}.$$
Using the hypothesis, (2-21) becomes

\[(2-23) \quad = 0(n) + \sum_{j=1}^{n-1} j \Delta(j) + \sum_{j=1}^{n-1} o(f(j)\Delta(j))\]
\[= 0(n) + \frac{q}{q^n} + o \left(\sum_{j=1}^{n-1} |\Delta(j)||f(j)| \right).\]

We now bound the sum on the right side of (2-23).

\[(2-24) \quad \sum_{j=1}^{n-1} |\Delta(j)||f(j)| \leq \max_{1 \leq j \leq n} f(j) \left\{ \sum_{i=1}^{n-1} |\Delta(i)| \right\}
\[\leq knf(n) \sum_{i=1}^{n-1} |\Delta(i)| \leq \frac{k'f(n)}{q^n \sqrt{n}}.\]

In the next to last inequality we have used the fact that

\[f(n) \sim \sup_{1 \leq x \leq n} f(x) \quad \text{as on page 19 of Seneta (1976). From (2-22) the } \Delta(j)
\]

are of constant sign for \(0 < j \leq np - 1\) and for \(np - q < j \leq n\). Thus

the last inequality in (2-24) follows from the standard bound for the

maximal term of the binomial distribution (Feller (1968), page 151).

Using (2-24) in (2-23) completes the proof of Theorem 3.

Proof of Corollary 4: Let

\[Q(x) = \sum_{i \in x \cap Q} 1\]

and

\[E(x) = Q(x) - x \frac{6}{\pi^2} .\]
Walfisz (1962) proved that

\[ E(x) = O \left( \sqrt{x} \exp \left\{ -c \log^{3/5} x \left( \log \log x \right)^{-1/5} \right\} \right) \]

while Vaidya (1966) proved that if the Riemann hypothesis is true, then

\[ E(x) = O(x^{2/5} + \epsilon) \]

The stated results now follow from Theorem 3 by noting in both cases the bound for \( E(x) \) varies regularly at \( \infty \) with suitable exponent.
References


