G.H. Hardy and Probability ????

Persi Diaconis *

August 20, 2001

Abstract

Despite a true antipathy to the subject Hardy contributed deeply to modern probability. His work with Ramanujan begat probabilistic number theory. His work on Tauberian theorems and divergent series has probabilistic proofs and interpretations. Finally Hardy spaces are a central ingredient in stochastic calculus. This paper reviews his prejudices and accomplishments through these examples.

*Departments of Mathematics and Statistics, Stanford University, Stanford, CA 94305

1 Hardy

G.H. Hardy was a great hard analyst who worked in the first half of the twentieth century. He was born on February seventh, 1877, in Surrey and died on December first, 1947 at Cambridge. C.P. Snow's biographical essay on Hardy which opens modern editions of Hardy's *A Mathematician's Apology* gives a good non-mathematical picture of Hardy. Titchmarsh's obituary gives a bit of a mathematical overview. It is reprinted in his collected works which comprise seven thick volumes containing Hardy's 350 papers.

Hardy had a close connection to the London Mathematical Society, serving as its President in 1926-1928 and almost continuously on its council. He left his estate to the Society and his personal library now resides in the LMS headquarters in Russell Square.

Hardy is most celebrated today for his work in analytical number theory. He was the first to prove that Riemann's zeta function had infinitely many zeros on the critical line. Working jointly with Ramanujan, he developed the circle method for extracting information on the coefficients of generating functions with complicated singularities. They used the circle method to give extremely accurate approximations to the number of partitions of n. In work with Littlewood, the circle method gave definitive answers for many instances of 'Waring's problem' on the representation of integers as sums of nth powers. These methods have been refined and amalgamated into modern number theory; A splendid accessible development of the classical work appears in Ayoub (1963). Modern developments may be found in Vaughn (1997).

2 Hardy and Probability

I want to argue that Hardy had no knowledge of probability theory and indeed had a genuine antipathy to the subject. To begin with, Hardy loved clear rigorous argument. At the time he worked, the mathematical underpinnings of probability were a vague mess; random variables were defined as the observed values of random quantities. Mark Kac (1985, Chapter 3) has given a splendid picture of what it was like to work in probability at this time. It was only in 1933 that Kolmogorov gave a measure theoretic interpretation of probability; a random variable was defined as a measurable function. Then one could see that early workers in probability; Bernoulli, Laplace, Gauss, Chebychev, Markov were doing mathematics after all.

Probability would surely have been part of applied mathematics for Hardy. He had very little use for applications writing 'very little of mathematics is useful and that little is comparatively dull'. His student Titchmarsh wrote "...His instinct was for the purest of mathematics. I worked on the theory of Fourier integrals under his guidance for a good many years before I discovered for myself that his theory has applications in applied mathematics, if the solution of certain differential equations can be called applied. I never heard him refer to these applications."

Hardy wasn't alone in his lack of interest in probability. Alan Turing did his fellowship thesis at Cambridge in 1933. (Turing was at Kings College, Hardy at Trinity.) Turing gave an elegant proof of the classical central limit of probability; this theorem had in fact been proved in at least three ways before. Turing rediscovered Lindeberg's proof of 1922. So little of probability was known in mathematical circles of Oxford and Cambridge that the thesis was granted. Further details and background may be found in Zabell (1995).

Hardy's collaborator Littlewood did have a feel for probability as a perusal of his 'Mathematicians Miscellany' shows. Indeed, late in life he developed a curiosity about parapsychology and published refined tail bounds for the binomial distribution (Littlewood 1969). As far as I know, Ramanujan also had no feel for probability. My only evidence for this is that very refined estimates which can be seen as properties of the Poisson distribution are developed by Ramanujan without note of their probabilistic content. See Diaconis (1983) for further details.

Curiously, Hardy's most well known work outside mathematics has probabilistic underpinnings. This is his celebrated letter to *Science* (July 10, 1908) on what is now called 'Hardy-Weinberg Equilibrium'. This addressed the problem of why dominant traits don't just take over. In the simplest setting, the trait is determined by a single allele which can be of type A or a with A dominant. Children inherit one copy of each of their parents' alleles and so can be $\{A, A\}$, $\{A, a\}$, $\{a, a\}$, the first two possibilities producing the dominant trait. If the original proportions of these three pairings in a large population are $p, 2q, r$ with $p+2q+r = 1$, under random mating (where the pairs are split apart and randomly combined) the new proportions are $(p+q)^2, 2(p+q)(q+r), (q+r)^2$. Hardy points out that these proportions are stable if $q^2 = pr$ and further, for any starting proportions, stability is achieved after one iteration. Hardy's 'back of the envelope calculation' is carefully worded with plenty of sensible caveats.

There is one other instance of insightful probabilistic thinking in Hardy's work with Littlewood on the Goldbach and prime k-tuples conjecture. This is a generalization of the well known (and still open) twin prime conjecture which posits infinitely many primes p with $p + 2$ also prime; it is natural to ask also for the proportion of twin primes up to x. Hardy's papers give a sophisticated development of conjectured asymptotics which suggest that at least one of the authors was quite familiar with probabilistic heuristics. Here, Hardy was certainly aided by earlier heuristic investigations of Sylvester and others. Hardy refers to what I would call 'probabilistic reasoning' as "a priori judgment of common sense" in his expository account (Hardy (1922, page 2). Mosteller (1972) gives a motivated development and compares the conjecture to actual counts.

Lest the reader be fooled by the last two examples, let me report a direct link with Hardy which underscores my main point. Paul Erdös knew Hardy and was a great friend to me. My interests in Hardy and probability started when Erdös said "You know, if Hardy had known anything of probability. He certainly would have proved the law of the iterated logarithm. He was an amazing analyst and only missed the result because he didn't have any feeling for probability." Let me explain: Borel (1909) had proved the strong law of large numbers. This studies $S_n(x)$ - the number of ones in the first n places of the binary expansion of the real number x. Borel proved

$$\lim_{n\to\infty} \frac{S_n(x) - n/2}{n} = 0 \tag{2.1}$$

for Lebesgue almost all x in $[0, 1]$. That is, almost all numbers have half their binary digits one and half their digits zero. While not posed as a probability problem, the binary digits of a point chosen at random from the unit interval are a perfect mathematical model of independent tosses of a fair coin.

It was also known that $(S_n(x) - n/2)/\sqrt{n}$ did not have a limit for almost all x. The question arose of the exact rate of growth. In 1914 Hardy and Littlewood showed n could be replaced by $\sqrt{n \log n}$ in (2.1). They missed the right result which was found in 1924 by Khintchin; for almost all x in $[0, 1]$.

$$\overline{\lim} \frac{S_n(x) - n/2}{\sqrt{2n \log \log n}} = 1. \tag{2.2}$$

Erdös had the deepest understanding of this result; he gave definitive, fascinating extensions (See Erdös (1942) or Bingham (1986). The analytic component needed to prove (2.2) would have been straightforward for Hardy, it was probabilistic thinking that was missing.

One more tale to underscore my point. I said above that Hardy's personal library now resides in the LMS headquarters. Not all of his books made it there. Twenty years ago, Steve Stigler, the great historian of probability and statistics, was able to purchase Hardy's copies of Jeffreys' *Theory of Probability* and a probability book by Borel from a London bookseller. The first was unopened, the second uncut.

In the next three sections I develop three detailed examples where Hardy's work had intimate, delicate connections with probability. The problems are the number of prime divisors of a typical integer, Tauberian Theorems (along with Hardy's lifelong love affair with divergent series) and finally the Hardy spaces H^p. The last section tries to make sense of the evidence and reach some conclusions.

3 Additive Number Theory

Hardy and Ramanujan sometimes regarded numbers playfully as when Hardy reported his taxi number - 1729 - as dull and Ramanujan said 'no Hardy, no Hardy, 1729 is the smallest number which is the sum of two cubes in two different ways'. Properties such as prime and 'almost prime' are notable in their own right. Hardy and Ramanujan studied

$$\omega(N) = \#\text{Distinct Prime Divisors of } N,$$

(so $\omega(12) = 2$) and called numbers round if $\omega(N)$ was large. It is easy to see that $1 \leq \omega(N) \leq \frac{\log N}{\log\log N}$, the upper bound being achieved for N a product of the first primes. they asked how large $\omega(N)$ is for typical N and showed the answer is $\log\log N$; now $\log\log N$ is practically constant for any N humans encounter (e.g. $\log\log 10^{100} \doteq 5$) so most numbers encountered are fairly flat. To make this precise, Hardy and Ramanujan (1920) proved

$$\textit{THEOREM} \qquad \frac{1}{x}\left| N \leq x : -\psi_N < \frac{\omega(N) - \log\log N}{\sqrt{\log\log N}} \leq \psi_N \right| \to 1 \tag{3.1}$$

As x tends to infinity for any function ψ_N increasing to infinity.

Their reasoning is interesting. Numbers with $\omega(N) = 1$ are primes and powers of primes. By the prime number theorem, the proportion of the first x numbers of this form is asymptotically $\frac{1}{\log x}$. Assuming the prime number theorem, Landau (following Gauss and followed by Sathe and Selberg) had shown that for fixed j the proportion of the first x numbers with $j + 1$ distinct prime divisors is asymptotic to

$$\frac{e^{-\lambda}\lambda^j}{j!} \quad \text{with} \quad \lambda = \log\log x$$

Plotting $e^{-\lambda}\lambda^j/j!$ as a function of j gives a discrete version of the familiar bell shaped curve. The plot peaks at $\log\log x$ and most of the mass is concentrated around the peak within a few multiples of $\sqrt{\log\log x}$. Hardy and Ramanujan made this argument rigorous by replacing Landau's asymptotics by useful upper and lower bounds.

Impressive as the argument is, to a probabilist, the project seems out of focus; they are proving the weak law of large numbers by using the local central limit theorem. If all that is wanted is their theorem, there are much easier arguments. With all their work, one could reach much stronger conclusions.

The easier argument was found by Turan (1934). Write

$$\omega(N) = \sum_{p|N} X_p(N) \quad \text{with} \quad X_p(N) = \begin{cases} 1 & \text{if } p \text{ divides } N. \\ [2mm]0 & \text{otherwise.} \end{cases}$$

Let the average value $\bar{w}_x = \frac{1}{x}\sum_{w\leq x} w(N)$. Clearly, the average of w is the sum of the averages of the X_p and these are $\frac{1}{x}\sum_{w\leq x} X_p(N) = \frac{1}{x}\lfloor\frac{x}{p}\rfloor$.
Thus

$$\bar{w}_x = \frac{1}{x}\sum_{p\leq x}\left\lfloor\frac{x}{p}\right\rfloor = \frac{1}{x}\sum_{p\leq x}\left(\frac{x}{p}+0(1)\right) = \sum_{p\leq x}\frac{1}{p}+0\left(\frac{1}{\log x}\right).$$

Now an elementary estimate due to Mertens gives

$$\sum_{p\leq x}\frac{1}{p} = \log\log x + \beta_1 + 0(\log x) \quad \text{where} \quad \beta_1 = \gamma + \sum_p \left(\log\left(1-\frac{1}{p}\right)\right) + \frac{1}{p} \doteq 0.2615$$

The mean value is not enough to prove the concentration result (3.1).
Turan also computed the variance

$$\sigma_x^2 = \frac{1}{x}\sum_{N\leq x}(\omega(N)-\bar{w}_x)^2 = \log\log x + \beta_2 + 0\left(\frac{\log\log x}{\log x}\right),$$

$$\beta_2 = \beta_1 - \left(\frac{\pi^2}{6}\right) - \sum_p \frac{1}{p^2} \doteq -1.8357. \tag{3.2}$$

The argument is not much more complex than the argument above for $\bar{w}_x$. From here, Chebychev's elementary inequality gives

$$\frac{1}{v}\Big|N\leq x:\ \Big|\frac{w(N)-\bar{w}_x}{\sigma_x}\Big|\leq\psi\Big| \leq \frac{1}{\psi^2}.$$

Hardy used Turan's proof in his number theory book (rewritten to remove all traces of probability). In his lectures on Ramanujan (Chapter 3) he gave Turan's argument but augmented it with the heuristics underlying the bell-shaped curve discussed above.

The Hardy Ramanujan paper had a large follow up. To begin with, Erdös and Kac (1940) made the central limit theorem explicit, proving

$$\frac{1}{x}\Big|N\leq x:\ \frac{w(N)-\bar{w}_x}{\sigma_x}\leq t\Big| \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-z^2/2}dz.$$

The Erdös-Kac paper is the first appearance in modern probability of what is called the *invariance principle*. A universality result that has grown into the modern weak convergence machine. Billingsley (1974, 1999) gives beautiful descriptions of these developments both in what is now called probabilistic number theory and probability more generally.

A number theoretic function f is called *additive* if $f(NM) = f(M) + f(N)$ for M, N relatively prime. Thus $w(N)$ is additive as is the total number of prime divisors or $F(N) = \sum_{p|N} p$. This last function arises in analyzing the average running time of the discrete Fourier transform on Z/NZ using the Cooley-Tukey fast Fourier transform; Diaconis (1980).

Study of the typical behavior of additive number theoretic functions has developed into a large area called probabilistic number theory. Basically, the $X_p(N)$ behave like independent random variables and very refined probabilistic limit theorems have been proved. (There is a wonderful introduction to this subject in Kac (1954).) The survey by Billingsly (1974). Books by Elliott (1979, 1980, 1985), and Tenenbaum (1995) are excellent sources.

Here is one exciting contribution of probabilistic number theory to main stream number theory; a crucial tool, abstracting Turan's basic argument above, is called the Turan-Kubilius inequality; using this, and the large sieve, Hildebrand (1986) has given a new elementary proof of the prime number theorem.

To close this section on a personal note; the expression for the variance of $\omega(N)$ at (3.2) was derived in my first published paper (done jointly with Fred Mosteller and Hironari Onishi). When I arrived at Harvard's statistics department I met Fred Mosteller. He was a statistician who studied the primes as data. He was fascinated by the Hardy-Ramanujan, Erdös-Kac results but wanted a more accurate expression for the variance (Turan had not derived terms past the initial log log) when $x = 10^6$, the exact value for $\sigma_x^2 = .9810$ while the fitted term $\log\log x$ is 2.626 Here the corrected approximation $\log\log x + \beta_2 \doteq .7401$ makes a big numerical difference.

In later work (Diaconis, 1982), I derived asymptotic expansions such as

$$\begin{aligned}
\bar{w}_x &= \log\log x + \beta_1 + \sum_{j=1}^{n} \frac{a_j}{(\log x)^j} + 0\left(\frac{1}{(\log x)^{n+1}}\right) \qquad \text{with e.g. } a_1 = \gamma - 1 \\
\sigma_x^2 &= \beta_2 + \log\log x \sum_{k=0}^{n} \frac{b_k}{(\log x)^k} + \sum_{k=1}^{n} \frac{c_k}{(\log x)^\ell} + 0\left(\frac{\log\log x}{(\log x)^{n+1}}\right) \\
\text{with } b_o &= 1, b_1 = 0,\ c_1 = 3\gamma - 1 + 2\sum_p \frac{\log p}{p(p-1)}.
\end{aligned}$$

Just recently, the computer scientist Don Knuth (2000, pg. 303-339) found use for these in the analysis of factoring algorithms.

To summarize, the Hardy-Ramanujan paper gave birth to probabilistic number theory, the invariance principle, several applications in analysis of algorithms and gave me my first entry to probability and number theory.

4 Tauberian Theorems and Divergent Series

Tauberian Theorems and various averaging or summability techniques were a favorite topic of Hardy. He wrote about them early and often. His last book, the posthumously published *Divergent Series*, was devoted to this subject: I find it curious that this large intellectual undertaking has more or less disappeared from modern mathematics, or been so transformed so as to be unrecognizable. In what follows, I will introduce the subject from the beginning, give some of Hardy's theorems, find some fascinating probability hidden in Hardy's proofs, and pose some open problems.

4.1 First Example

Suppose you take today's *New York Times*, look at the front page, and count what proportion of the numbers that appear begin with 'one'. Many people are surprised to find that this is often about 30%. Lead digits aren't uniformly distributed. Empirically, in chemical handbooks and other compilations of data, the proportion of numbers that begin with i is about $\log_{10}\left(1+\frac{1}{i}\right)$ (and $\log_{10}\left(1+\frac{1}{1}\right) = \log_{10} 2 \doteq .301$). This is quite different from the distribution of final digits where about $\frac{1}{10}$ of numbers end in $i, 0 \leq i \leq 9$.

A subset S of $\{1, 2, 3, \ldots\}$ has density ℓ if $\frac{1}{I}\sum_{i \leq I} s_i \to \ell$ where s_i is one or zero as i is in S or not. The set of even numbers has density $\frac{1}{2}$, the set of multiples of three has density $\frac{1}{3}$ and so on. What about $S = \{i \text{ with lead digit } 1\}$. Here $S = \{1, 10, 11, 12, \ldots, 19, 100, 101, \ldots, 199, \ldots\}$. The proportion up to 9 is $\frac{1}{9}$ but the proportion up to 19 is $\frac{11}{19}$, up to 99 we get $\frac{11}{99} = \frac{1}{9}$ again, and so on. The limiting proportions osscilate between a lower limit of $\frac{1}{9}$ and in upper limit of $\frac{5}{9}$. Thus S doesn't have a limiting density.

With s_i as above, we have been looking at the average $s_I^{(1)} = \frac{1}{I}\sum_{i \leq I} s_i$. The averaging didn't tame the original sequence s_i; it is natural to try averaging the $s_i^{(1)}$ forming $s_I^{(2)} = \frac{1}{I}\sum_{j \leq I} s_j^{(1)}$. And then perhaps $s_I^{(3)}$ and so on. Hardy proved that if a sequence $\{s_i^0\}_{i=1}^{\infty}$ is bounded and any average $s_I^{(k)} \underset{\infty}{\overset{I}{\to}} \ell$ for $k \geq 2$, then already $s_I^{(1)} \underset{\infty}{\overset{I}{\to}} \ell$. Thus no sequence of such repeated averages can assign the first digit set a limit. This is an example of a simple Tauberian theorem.

More generally, Tauberian theorems relate different methods of averaging: A typical Tauberian theorem says if one method of averaging converges then a second method converges for all sequences satisfying an additional Tauberian condition (here boundedness).

Returning to the first digit example, it is not hard to see that s_i is *logarithmically summable* to the right answer:

$$\frac{1}{\log I}\sum_{i \leq I} \frac{s_i}{i} \to \log_{10}(2).$$

At first logarithmic averages seem unusual; they weigh earlier numbers more heavily. We will see in a moment that if $s_i \to \ell$, $\frac{1}{\log I}\sum_{i \leq I} \frac{s_i}{I} \to \ell$. Indeed, for bounded sequences if $\frac{1}{I}\sum s_i \to \ell$ then logarithmic averages also converge to ℓ.

There is much else to say about the first digit problem. See Diaconis (1977), Raimi (1976) or Hill (1995). We will be content having found the empirical answer .301 in an average of the sequence of first digits.

4.2 Basic Summability

A matrix summability method is defined by an infinite matrix $A = \{A_{ij}\}_{1 \leq i,j < \infty}$. A sequence s_i is A-summable to limit ℓ if

$$\sum_j A_{ij} s_j \underset{\infty}{\overset{i}{\to}} \ell.$$

for example the basic Cesaro averages have $A_{ij} = \begin{cases} \frac{1}{i} & 1 \leq j \leq i \\ 0 & \text{else} \end{cases}$. Similarly the logarithmic method is a matrix method. I do not know if every reasonable method is equivalent to a matrix method. For example, referring to the iterated averaging method of section 4.1, the lower limits of successive averages are increasing. The upper limits decreasing. If these meet the sequence is H_∞ summable $\left(\text{thus } \lim_k \underline{\lim}_I\, s_I^{(k)} = \lim_k \overline{\lim}_I\, s_I^{(k)}\right)$ Flehlinger (1966) showed that the first digit indicators s_i are H_∞ summable to $\log_{10}(2)$. Is H_∞ equivalent to a matrix method?

Summability theory begins with the Toeplitz Lemma. This says that the matrix method A assigns limit ℓ to all sequences converging to ℓ if and only if A satisfies, for all i,

$$A_{ij} \xrightarrow[\infty]{j} 0, \ \sum_j |A_{ij}| < \infty, \ \sum_j A_{ij} \xrightarrow[\infty]{i} 1 \tag{4.1}$$

this applies to Cesaro and logarithmic methods. Methods satisfying (4.1) are called regular. The Toeplitz Lemma is a kind of Tauberian theorem comparing the matrix method A to the identity matrix. When such a comparison is available with no restriction on the sequences the method is called an Abelian Theorem. Perhaps this is because of a special case due to Abel: If $s_i \to \ell$ then $\lim_{x\to 1}(1-x)\sum s_i x^i \to \ell$.

4.3 Probability and Tauberian Theorems Hidden in Hardy

Given two matrix methods of summability A, B, call A *weaker* than B if any A summable sequence is B summable to the same limit.

Example 1 Cesaro is weaker than Abel.

Proof In a moment we will verify the identity

$$(1-x)\sum_{i=0}^{\infty} s_i x^i = (1-x)^2 \sum_{i=1}^{\infty} \left(\frac{1}{i+1} \sum_{j=0}^{i} s_i \right) (j+1) x^i \tag{4.2}$$

Assuming this, suppose s_i is Cesaro summable to ℓ. Thus $\bar{s}_I = \frac{1}{I+1}\sum_{i=0}^{I} s_i \to \ell$. Since the method $(1-x)^2 \sum \bar{s}_i (i+1) x^i$ is regular, the limit of the right side is ℓ, so the limit of the left side is ℓ. For complete rigor, note that Cesaro summability to ℓ implies $s_i = 0(i)$ so the infinite sums converge for $0 \leq x < 1$.

The identity (4.2) has a simple probabilistic proof and interpretation. For $0 \leq x < 1$, $(1-x)x^i$ is the chance that an x-coin shows i heads before the first tail. And $(1+i)(1-x)^2 x^i$ is the chance that an x coin shows i heads before the second tail. Now it is an elementary calculation that the conditional distribution of the first tail, given that the second tail occurs at $i+2$, is uniform over $\{1, 2, \ldots, i+1\}$. This is true for all values of x. Now (4.2) is seen as the basic probabilistic identity $E(f(X)) = E(E(f(X)|Y = y))$, with X the number of heads before the first tail, Y the number of heads before the second tail and $f(i) = s_i$. The expectations refer to average with respect to the underlying measures.

One point of this section is there are dozens of such elementary probability identities hidden in Hardy's comparison proofs. There are also many further probability identities which give comparisons. Later in this section I will conjecture (and partially prove) an equivalence.

Example 2. A sequence is EULER(p) summable to ℓ if $\lim_{n\to\infty} \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} s_i \to \ell$. The Euler method was a basic tool used to derive analytic continuations in classical work (Hardy (1949), Chapters 8,4). Of course, the Euler method is just an average with respect to p-coin tossing. Here are some applications of elementary coin tossing facts to Euler summability.

Let X be the number of heads if a p-coin is tossed n times. Let Y be the number of heads if a p' coin is tossed X times. Here Y depends on X. The unconditional distribution of Y is the distribution of the number of heads if a pp' coin is tossed n times. This is easy to see analytically. It also follows from the following probabilistic considerations: picture a flower with n seeds; the seeds drop to the ground or not independently with probability p. Then, they germinate with probability

p'. Clearly, the number of seeds that germinate has the same distribution as the number of heads when a pp' coin flipped n times.

This probabilistic argument may be rewritten as

$$\sum_{i=0}^{n} \binom{n}{i} (pp')^i (1-pp')^{n-i} s_i = \sum_{i=0}^{n} \left\{ \sum_{j=0}^{i} \binom{i'}{j} (p')^j (1-p')^{i-j} s_j \right\} \binom{n}{i} p^j (1-p)^{n-i}.$$

From this it follows that the Euler(p) methods increase in strength as p decreases.

Similarly, the probabilistic relation: "Pick X from a Poisson-distribution; Let Y be the number of heads if a p-coin is flipped X times, then Y has a Poisson(λp) distribution" shows that the Euler(p) method is weaker than the Borel method:

$$\lim_{\lambda \to \infty} \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} s_j \to \ell.$$

The identities above amount to a factorization of summability methods. Consider A, B, C regular summability methods. If $CA = B$ then clearly A is weaker than B. It is natural to ask about a converse: If A is weaker than B must there be a factorization? The following proposition shows that there is at least some sense to this question.

Proposition Let A and B be regular lower triangular matrix methods with A weaker than B. If $A_{ii} \neq 0$ $1 \leq i < \infty$ then there is a regular lower triangular method C such that $CA = B$.

Proof Define $C = BA^{-1}$. Here A^{-1} exists and the product is well defined because of triangularity. To see that C is regular, Let $s_i \to \ell$. Let $t_i = (A^{-1}s)_i$ then $t_i =$ is A-summable to ℓ and thus B summable ℓ so that BA^{-1} is regular □.

I have found it difficult to formulate a neat generalization of the proposition to the non-triangular case. Changing the first few rows of a method doesn't change the set of sequences assigned a limit but does foul up algebraic identities. A plausible conjecture is this: A is weaker than B if and only if there are equivalent A', B' and C such that $CA' = B'$.

4.4 Remarks on Probability and Summability.

There are many points of contact between Hardy's work on divergent series and classical probability. To begin, recall the classical Hardy-Littlewood Tauberian Theorem: Given a sequence s_i, let $\Delta_i = s_i - s_{i-1}$. If $|\Delta_i| \leq \frac{K}{i}$ for some K, and $\frac{1}{n}\sum_{i=1}^{n} s_i \to \ell$ then $s_i \to \ell$. This refined the original condition of Tauber who reached the same conclusion assuming $|\Delta_i| = o\left(\frac{1}{i}\right)$. The Hardy-Littlewood Theorem was abstracted to a theorem relating the tails of a measure on $[0, \infty)$ to the behavior of its Laplace transform at 0. Karamata gave a soft proof of the result using weak compactness. Feller (1971, Chapter 8) gives a wonderful account of these developments and many probabilistic applications. Bingham Goldie, and Teugels (1987) present a definitive account.

On a personal note, my interest in Hardy's work started with the first digit problem and extended to developing basic probability theorems such as the law of large numbers for a variety of other averages. If X_i $1 \leq i < \infty$ are independent random variables the sum $\sum_{i=0}^{\infty} X_i x^i$ may be interpreted as a discounted return. Its distribution for x close to 1 is of interest in financial applications. See Lai (1973), Diaconis (1974), or Bingham et al. (1986, Sec. 15) for much further development.

It is natural for a probabilist to try to replace the signed measures in the rows of the Toeplitz Lemma by honest probability measures. If A is a regular summability method let $\widetilde{A}$ have all negative entries replaced by zero and the rows normalized sum to one. Does this give an equivalent method? Alas, no; The regular method $2s_i - s_{i+1}$ sums 2^n to zero but its probabilistic counterpart is the identity. For bounded sequences one can show if $2s_i - s_{i+1} \to \ell$ then $s_i \to \ell$. Are A and $\widetilde{A}$ equivalent for bounded sequences? Again, no: The regular method $s_i - s_{i-1} + s_{i-2}$ 'probabilizes' to $(s_n + s_{n-1})/2$. This sums the sequence $1, 1, -1, -1, 1, 1, -1, -1, \ldots$ to zero but the original method doesn't assign a limit to this sequence. I do not believe we can avoid negative measures in summability theory.

To summarize, Hardy's work on summability, divergent series and Tauberian Theorems is intimately connected to probability theory. It has had great applications therein and conversely, probabilistic techniques have given extensions and unification to Hardy's work. Our next section shows similar contact in a completly different context.

5 Hardy Spaces and Probability

The theory of Hardy's H^p spaces can be roughly described as what happens when analytic function theory meets up with real variables L^p theory. It turns out that these theories are intimately connected with probabilistic ideas such as Brownian motion and martingales. A competition developed between probabilists and analysts over whose techniques prove better results. The two sides have merged into a novel part of modern analysis.

This section begins with harmonic functions and the H^p spaces outlining some of what Hardy and his coworkers did and pointing to highlights. Next, Brownian motion and martingales enter the fray. Sharp connections through the work of Burkholder-Gurdy-Silverstein are followed by modern triumphs: The dual of H^1 as BMO and the corona problem. The section concludes with a sketch of the legacy of Hardy's work: H^p martingales and their applications to finance, Dirichlet spaces and their applications to rates of convergence of Markov chains.

Technical treatment of these topics has wonderful literature. The classical book of Kellogg (1953) is still a readable account of harmonic functions. Axler et al. (2001) gives a modern treatment. Duren (2000), Koosis (1999) and Garnett (1981) give a clear treatment of the H^p spaces. Petersen (1977) and Durrett (1984) describe the interaction between probability and Hardy spaces. The two books by Bass (1995, 1998) give a thorough treatment of the interactions between probability and analysis. The remarkable series by Dellacherie-Meyer (1975, 1980, 1983) show how far probabilistic potential theory has developed. The encyclopedic account by Doob (1984) has much detailed discussion.

5.1 Harmonic Functions and the H^p Spaces

A harmonic function in the plane is a solution of Laplace's equation $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$. For planar regions, this equation characterizes the flow of heat in homogeneous media, the flow of ideal fluids and electrical currents. Kellogg (1953) gives a readable development of the appearance of harmonic functions in physical problems. If $f(z) = u(x, y) + i\widetilde{u}(x, y)$ is analytic, its real and imaginary parts are harmonic and there is a close connection between properties of analytic functions and harmonic functions. In this section, analytic notation will be used.

Hardy studied analytic function f on the open unit disc $D = \{|z| < 1\}$. Define mean values

$$m_p\{r, f\} = \left\{ \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta \right\}^{\frac{1}{p}} \quad 0 < r < 1$$
$$m_\infty(r, f) = \sup_{0 \le \theta < 2\pi} |f(re^{i\theta})|$$

Define H^p, $0 < p \le \infty$ as the set of analytic functions f and with $m_p(r, f)$ bounded as $r \to 1$. Thus H^∞ is the set of bounded analytic functions and H^2 is the set of functions of form $\Sigma_{j=0}^\infty a_j \, z^j$ with $\Sigma|a_j|^2 < \infty$. In a paper regarded as the start of the subject, Hardy (1915) showed that $m_p(r, f)$ is monotone increasing in r. Two starting facts are crucial:

First, $f \in H^p$ has boundary limits almost surely. The boundary function $\widetilde{f}$ is in L^p of the circle and determines f. For $0 < p < \infty$, boundary functions $\widetilde{f}$ are exactly all possible L^p limits of trigonometric polynomials $\Sigma_{j=0}^n \, a_j e^{ij\theta}$. They may also be described as functions in L^p of form $\Sigma_{j=0}^\infty \, a_j e^{ij\theta}$.

A second basic ingredient is the Hardy-Littlewood maximal inequality. For $0 < p \le \infty$ if $f \in H^p$ and $f_*(\theta) = \sup_{0<r<1} |f(re^{i\theta})|$ then $f_* \in L^p[0, 2\pi)$ and $\|f_*\|_p \le A_p \|\widetilde{f}\|_p$ for some constant A_p independent of f. This can be used to show boundary limits exist.

These two ingredients will reappear in martingale theory: a class of functions with limits at the boundary and with sup-norm controlled by boundary behavior.

The seemingly mild requirements: analytic in the disc and bounded in L^p give rise to a surprisingly rich structure theory. There is a canonical form which parameterizes $f \in H^p$ in terms of a countable set of zeros, an absolutely continuous and a singular measure on the unit circle. These give f as a product of a Blaschke product and an inner and outer function. These were developed by Beurling who applied them to give a complete description of the subspaces of ℓ^2 invariant under the one sided shift. This use of H^p theory in a natural problem with no apparent connection to H^p shows its strengths. A completely different example is a theorem of F. and M. Riesz: A probability measure μ on $[0, 1]$ with $\int_0^1 e^{2\pi ij\theta} \mu(d\theta) = 0$ for $j = -1, -2, -3, \ldots$, is absolutely continuous.

The H^p spaces come with a surprisingly useful extremal and interpolation theories. These have been widely developed and applied by control theorists with numerous books and conferences devoted to H^∞ control. A typical extremal problem is to find, for Q in the dual of H^p,

$$\|Q\| = \sup\{|Q(f)| : \; f \in H^p, \; \|f\|p \le 1\}$$

along with a description of the extremal f's. For appropriate Q this translates into maximizing $f(x_o)$ or $f^{(n)}(x_o)$. A typical interpolation problem is find conditions for the existence of $f \in H^p$ with $f(z_k) = w_k$. The H^p spaces have enough structure that these problems can often be usefully or explicitly solved.

5.2 A Bit of Probability

Two of the great achievements of twentieth century probability are Brownian motion and martingales. Complex Brownian motion may be described as a probability measure μ on $C([0, \infty], \mathbf{C})$ which is normalized to paths that start at zero, has independent increments and assigns Gauss measure to the end point of the path from zero at time $t \in [0, \infty)$. If you don't know what these words mean, think of a particle starting at zero and wiggling around in the complex plane as time evolves. We will write B_t for the position of the particle at time t.

Two fundamental theorems connect Brownian motion to complex function theory: Paul Lévy proved that if f is a nonconstant analytic function with $f(0) = 0$ then $f(B_t)$ is again Brownian motion with time rescaled to $\int_0^t |f'(B_s)|^2 ds$. This is the basis of Brownian proofs of a host of basic theorems in complex variables; from Liouville's theorem to the Picard theorems. See e.g. Durrett (1984), Bass (1995), or Petersen (1977).

The second fundamental theorem is Kakutani's Brownian solution of the Dirichlet problem. For simplicity, state this on the unit disc D. Let $B_{z,t}$ be Brownian motion started at $z \in D$. Let T be the random time this Brownian path first hits the boundary. Let g be a continuous real valued function on the boundary. Define $f(z) = \int g(B_{z,T})P(dw)$. Kakutani showed that f is a harmonic function with boundary values g. This is the start of a huge subject, probabilistic potential theory to which we return in Section 5.4.

The second arm of modern probability we need is martingale theory. This is a product of the twentieth century. Originating with the notion of fair game, it was developed by Lévy and (chiefly) Doob in the period 1940-1960. It is a mainstay of any standard graduate probability course, see Williams (1991) or Neveu (1972).

Mathematically, a martingale is a family of measurable functions $X_t(\omega)$ $0 \le t < \infty$. On a probability space $(\Omega, \mathcal{F}, P)$ which satisfies the projection identity, for $u \ge t$

$$E(X_u | X_s, s \le t) = X_t.$$

I will not define the conditional expected value; think of it as the average value, given what happened up to time t.

Brownian motion is a martingale. Thus analytic functions of Brownian motion are martingales. More to the point, under mild restrictions, the martingale convergence theorem says that martingales have limits $X_\infty = \lim_{t\to\infty} X_t$ and Doob's maximal inequality says that for $0 < p < \infty$ the values of $M_t = \sup_{0 \le s \le t} |X_s|$ are bounded by X_∞ in e.g. L^p. This is more than a similarity. Modern work has shown that essentially all of the results of H^p theory can be derived from martingale theory. Moreover, using techniques of real variables shows the way to extending to higher dimensions and very general domains. We turn to these developments next.

5.3 Probability and H^p Spaces

A real breakthrough occurred when Burkholder-Gundy-Silverstein (1971) used Brownian motion techniques to solve problems that had frustrated complex analysts. Among many other things, they showed that for $1 \le p < \infty$ $f = u + i\tilde{u}$ is in H^p if and only if the maximal function of u is in L^p. This real variables characterization of H^p pointed the way to new results in the classical theory. It also allowed generalizations of much of the elegant H^p theory to higher dimensions. One of the great results is the Fefferman-Stein (1971) result identifying the dual space of H^1 with functions of bounded mean oscillation (BMO) as introduced earlier by John and Nirenberg. The argument of Fefferman and Stein is filled with a mix of probability and analysis. All of their results are proved for domains in higher dimensions.

The success of probabilistic methods in classical contexts encouraged probabilists to extract the essentials and create notions of H^p-martingales ($\lim_{t\to\infty} X_t \in L^p$) and BMO martingales. The H^p restriction serves as a useful compactness criteria allowing extensions of the Heine-Borel characterization (in $\mathcal{R}^k$ compact = closed and bounded) to function spaces. A nice survey of this appears in Delbaen and Schachermayer (1999). They give applications to finance, deriving the existence of hedging strategies for European and American style contingent claims for incomplete security markets. See Kramov (1996) for the financial background.

Probabilists and analysts see each other's results as challenges; almost always, each side has succeeded in proving the other's results. For a case in point, one of the most spectacular results of the analytical theory is Carleson's (1962) proof of the corona theorem. This involves the maximal ideals in the algebra H^∞. For $z \in D$ the set $M_z = \{f : f(z) = 0\}$ is a maximal ideal and Carleson showed these were dense in all maximal ideals. The heart of the argument is a simple to state result. If $f_1, f_2, \ldots f_n$ are in H^∞ and $|f_1(z)| + \ldots + |f_n(z)| > \delta$ for some $\delta > 0$ and all $z \in D$ then there exists $g_1, g_2, \ldots, g_n$ in H^∞ such that $f_1(z)g_1(z) + \ldots + f_n(z)g_n(z) \equiv 1$. Carleson's proof is a tour de force introducing new techniques that have been widely applied. In 1980, N. Varopoulos gave a Brownian motion proof of the corona theorem, Bass (1995) gives a nice treatment of this. In fairness, it must be added that Wolf gave an easier analytic proof, see e.g. Koosis (1999).

The rivalry between probabilists and analysts has continued in healthy course. The solutions of general elliptic partial differential equations can be studied using a generalization of Brownian motion called diffusions, see Bass (1998). Non-linear equations such as $\nabla^2 u = u^2$ can be studied using 'super-Brownian motion' a collection of Brownian particles interacting with each other in time and space. Of course, super Brownian motion can be studied using the non-linear equation. For a lovely entry to the work of Dynkin and le Gall in this area see le Gall (1999).

There are many further connections between probability and function theory. One remarkable line of results constructs counter-examples; you can't tell if a function is in H^1 in terms of the growth of Fourier coefficients: putting random $\pm$ signs on the Fourier coefficients of a function in H^1 almost surely gives a function not in H^p for any p in $[0, \infty]$. See Kahane (1985) for a thorough treatment. The Hardy-Littlewood maximal theorem is the first maximal inequality. Variations due to Kolmogorov, Hopf, Garsia and many others are crucial in modern proofs of the martingale convergence theorem and the ergodic theorem. See Garsia (1973). We will content ourselves with one final development which again shows the rich interplay between analysis, probability and applications.

5.4 Dirichlet Spaces and Rates of Convergence

The success of the blend of analysis and probability sparked remarkable abstractions called abstract or probabilistic potential theory. One version of this revolves around an abstraction of Brownian motion called a *Markov process.* This is a family of random variables X_t taking values in a general space $\mathcal{X}$ satisfying the Markov property: for all $t < u$

$$P\{X_u \in B | X_s,\ s < t\} = Q_{t-u}(X_t, B)$$

for a fixed Kernel $Q_s(x, dy)$. Under fairly minimal conditions one can attach a boundary $\partial\mathcal{X}$ to $\mathcal{X}$ and emulate many of the classical results on the unit disc. This subject is described in Meyer (1966) and the books by Dellacherie-Meyer (1975, 1980, 1983).

The spaces $\mathcal{X}$ that arise in applications vary from finite spaces (classical Markov chains) to infinite products such as u_3^L with u_3 the unitary group and $L = Z^3$. (Lattice-Gauge theory). It can be difficult to describe $Q(x, dy)$ and several alternative descriptions of Markov processes have arisen. Most widely used is the infinitesimal generator through the Hille-Yoshida theorem (see e.g. Ethier-Kurtz, 1986).

In 1964 a celebrated analyst (A. Beurling) and a probabilist (J. Deny) combined forces to give a useful, very general description of a Markov process through an associated quadratic form, now called a Dirichlet form defined on a class of functions on $\mathcal{X}$. The entire theory (path properties through boundary theory), for quite general spaces, has been developed in this language. I think it is fair to say that in its modern form, as exposited by Fukushima et al. (1994) the theory is essentially completely inaccessible to a working probabilist or analyst. This is a pity, because the theory has amazing accomplishments.

In joint work with Dan Stroock and Laurent Saloff-Coste I have been translating and applying some of the tools of Dirichlet forms to problems in applied probability such as how many times should a deck of cards be shuffled to thoroughly mix it. Mathematicians interested in theoretical computer science have applied Dirichlet form techniques to algorithmic questions which are provably sharp P-complete but allow Monte-Carlo Markov Chain algorithms to give provably accurate approximations with probability arbitrarily close to one. These stochastic algorithms work in reasonable times and are beginning to be used by working scientists in applied problems. Nice overviews of the theoretical developments are given by Saloff-Coste (1997). For practical applications, the wonderful book by Liu (2001) is recommended.

It is a stretch of the historical record to lay these developments on Hardy's doorstep (he might well turn up his nose at them). Nonetheless, as an active worker in the field I am sure that Hardy's work on H^p spaces sends dense threads to modern developments. To shine a clear light on this, I conclude with a modern result where the debt to Hardy is clear.

The story begins with what is now called Hardy's inequality. In one form (Hardy-Littlewood-Polya ((1967), p. 239-243). This says if $p > 1, a_n \geq 0$ and $A_n = a_1 + a_2 + \ldots + a_n$, then

$$\sum_{n=1}^{\infty} \left(\frac{A_n}{n}\right)^p \leq \left(\frac{p}{p-1}\right)^p \sum_{n=1}^{\infty} a_n^p. \tag{5.1}$$

Hardy originally proved this to give an elementary proof of Hilbert's double series theorem.

If you are like me, the inequality (5.1) seems mysterious. What is it good for? I was delighted to find a probabilistic application in Miclo (1999). Following extensions by Hardy and others, Miclo proved a weighted version of 5.1: Let μ and ν be positive functions on N_+. Let A be the smallest positive constant such that

$$\sum_{j=1}^{\infty} \mu(n) \left(\sum_{0<h\leq j} f(h)\right)^2 \leq A \sum_{j=1}^{\infty} \nu(j) f^2(j)$$

for all $f \in \ell^2(\nu)$. Let

$$B = \sup_{k\geq 1} \left(\sum_{j=1}^{k} \frac{1}{\nu(j)}\right) \left(\sum_{j\geq k} \mu(j)\right). \tag{5.2}$$

Miclo shows that $B \leq A \leq 4B$, so B is a good surrogate for A. To get Hardy's inequality with $p = 2$ from (5.2) take $\mu(j) = \frac{1}{j^2}$, $\nu(j) = 1$. Then $B = \sup_{k\geq 1} k \sum_{j\geq k} \frac{1}{j^2}$ is bounded. Elegant though this may be, it is still not clear what it is good for!

Miclo applies his result to prove spectral gap bounds for birth and death chains. These are Markov chains on N_+ which move up or down by at most one each step. Moving up corresponds to a birth, moving down to a death. The position of the chain represents the size of the population. Let $b(i)$, $i \geq 1$ be the chance of a birth from i. Let $d(i)$ $i \geq 1$ be the chance of a death from i. Under mild conditions on $b(i), d(i)$ the chain has a stationary probability $\mu(i)$ satisfying the reversibility condition $\mu(i)b(i) = \mu(i+1)d(i+1)$ for all i. The spectral gap of the chain may be defined as

$$\lambda = \inf_{f\in\ell^2(\mu)} \frac{\mathcal{E}(f|f)}{\mu(f-\mu(f))^2}$$

where the Dirichlet form is given by

$$\mathcal{E}(f|f) = \sum_{i=1}^{\infty} (f(i+1) - f(i))^2 \mu(i) b(i),$$

and $\mu(g) = \sum_{i=1}^{\infty} \mu(i) g(i)$. The spectral gap controls the rate of convergence to stationarity. If $K_x^\ell(y)$ is the chance of being at y after ℓ steps starting at x, then $\|K_x^\ell - \pi\| \leq (1-\lambda)^\ell / \sqrt{\pi(x)}$. One tries to bound the spectral gap away from 0 if possible. Miclo derives good bounds on λ in terms of a much more tractable quantity; let

$$B_+(i) = \sup_{k \geq i} \left(\sum_{j=i+1}^{k} \frac{1}{\mu(j) d(j)} \right) \left(\sum_{j \geq k} \mu(j) \right)$$

$$B_-(i) = \sup_{k < i} \left(\sum_{i=k}^{i-1} \frac{1}{\mu(j) b(j)} \right) \left(\sum_{j \leq k} \mu(j) \right)$$

Let $B^* = \inf_i B_+(i) \vee B_-(i)$. Miclo proves

$$\frac{1}{4B^*} \leq \lambda \leq \frac{2}{B^*} \tag{5.3}$$

One practical point is that B^* can often be approximated in natural problems. Miclo demonstrates this through examples. The point for this article is that Miclo proves the crucial part of (5.3) (the lower bound) by a straightforward application of his weighted Hardy inequality. I find all of this tantalizing; there are variations of the spectral gap formula called Sobolev inequalities; these require bounds on ℓ^p norms. I have high hope I can vary Miclo's argument along the lines of Hardy's inequality (5.1) and get useful results. For completeness, I must add that Hardy's inequality is used for other things such as interpolating between various L^p spaces. See Stein and Weiss (1975, pg. 196). .

To summarize this long section, Hardy introduced H^p spaces to study harmonic and analytic functions on the disc. This led to a rich, seemingly esoteric, development in pure mathematics. Probabilistic interpretations were found and used as tools to sharpen Hardy's analysis. This was so successful that pure probabilistic analogs of H^p were developed, far from function theory. These probabilistic analogs have become an important ingredient in stochastic calculus. The give and take that has become probabilistic potential theory, a huge healthy area with applications in finance, computer science, biology and elsewhere. Finally, people are still studying and using Hardy's old, magical inequalities.

6 Conclusions

When I prepared my Hardy lecture, I was a bit nervous. Even though Hardy is one of my true heroes, there is a critical tone underlying some of my remarks and I was afraid of offending the LMS audience. The talk uncovered some strong feelings about Hardy. A senior algebraist present made the following comments "Hardy was great but he really fouled up my field. His *Pure Mathematics* and other writings set the tone for generations of British mathematicians. There is no mention of algebra or geometry. Those of us interested in such subjects have had to fight for our place among the analysts. Pure mathematics indeed". Some of the applied mathematicians present had stronger opinions.

Hardy's insistence on pure mathematics, free of applications rankles me. Yet a look at the picture I have painted above shows much merit in his approach. Proceeding in his own way he created subjects with deep roots that have nurtured rich applied fields. When I began my review of the H^p spaces they seemed (to me!) a funny fringe of mathematics that had more or less shriveled. I hope I have shown that this work plays a crucial part in much current mathematics, both theoretical and applied. In contrast, I see much current applied work, while of immediate interest, as not having a very long shelf life.

Based on the developments I have sketched we can take issue with Hardy's famous apology "I have never done anything 'useful'. No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the activity of the world".

In retrospect, I don't think I would have been able to convince Hardy to learn any probability. Rather, he has convinced me to learn more of his kind of mathematics.

Acknowledgments

This is a version of the Hardy lecture given June 22, 2001, at the London Mathematical Society. The author thanks Rosemary Bailey, Ben Garling, and Susan Oakes for enabling his tour as Hardy Fellow. Thanks too for help from Arnab Chacabourty, Izzy Katznelson, Steve Evans, Charles Stein, Laurent Saloff-Coste, Steve Stigler and David Williams.

References

Axler, S., Bourdow, P. and Ramey, W. (2001), *Harmonic Function Theory*, 2nd Ed., Springer, New York.

Ayoub, R. (1963), "An Introduction to the Analytic Theory of Numbers". *Amer. Math. Soc.*, Providence.

Bass, R. (1995), *Probabilistic Techniques in Analysis.* Springer, New York.

Bass, R. (1998), *Diffusions and Elliptic Operators.* Springer, New York.

Beurling, A. and Deny, J. (1958), "Espaces de Dirichlet I, le cas elementaire". *Acta. Math.* **99**, 203-224.

Billingsley, P. (1974), "The Probability of Additive Arithmetic Functions". *Ann. Probab.* **2**, 749-791.

Billingsley, P. (1999), *Convergence of Probability Measures.* 2nd Ed., Wiley, New York.

Bingham, N. (1986), "Variants on the Law of the Iterated Logarithm". *Bull. Lond. Math. Soc.* **18**, 433-467.

Bingham, N., Goldie, C. and Teugels, J. (1987), *Regular Variation.* Cambridge University Press, Cambridge.

Borel, E. (1904), "Les Probabilités Denombrables et Leurs Applications Arithmetiques". *Rendiconti del Circold Mathematico de Palermo* **27**, 247-271.

Burkholder, D., Gundy, R. and Silverstein, M. (1971), "A Maximal Function Characterization of the Class H^p". *Trans. Amer. Math. Soc.* **157**, 137-153.

Carleson, L. (1962), "Interpolations by Bounded Analytic Functions and the Corona Problem". *Annals Math.* **76**, 547-559.

Davis, B. (1975), "Picard's theorem and Brownian Motion". *Trans. Amer. Math. Soc.* **213**, 353-362.

Delbaen, F. and Schachermayer, W. (1999), "A Compactness Principle for Bounded Sequences of Martingales with Applications". *Progress in Probability* **45**, 137-173.

Diaconis, P,. (1974), *Weak and Strong Averages in Probability and the theory of Numbers.* Ph.D. Thesis, Department of Statistics, Harvard University.

Diaconis, P., Mosteller, F. and Onishi, H. (1977), "Second-Order Terms for the Variances and Covariances of the Number of Prime Factors – Including the Square Free Case". *Jour. Number Th.* **9**, 187-202.

Diaconis, P. (1977), "Examples for the Theory of Infinite Iteration of Summability Methods". *Can. J. Math* **29**, 489-497.

Diaconis, P. (1977), "The Distribution of Leading Digits and Uniform Distribution Mod 1". *Ann. Probab.* **5**, 72-81.

Diaconis, P. and Stein, C. (1978), "Some Tauberian Theorems Related to Coin Tossing". *Annals Probab.* **6**, 483-490.

Diaconis, P. (1980), "Average Running Time of the Fast Fourier Transform". *J. Algorithms* **1**, 187-208.

Diaconis, P. (1982), *Asymptotic Expansions for the Mean and Variance of the Number of Prime Factors of a Number n.* Technical Report, Department of Statistics, Stanford University.

Diaconis, P. (1983), *Projection Pursuit for Discrete Data.* Technical Report, Department of Statistics, Stanford University.

Doob, J. (1953), *Stochastic Processes.* Wiley, New York.

Doob, J. (1954), "Semimartingales and Subharmonic Functions". *Trans. Amer. Math. Soc.* **77**, 86-121.

Doob, J. (1984), *Classical Potential Theory and its Probabilists Counterparts.* Springer, New York.

Dellacherie, C. and Meyer, P.A. (1975), *Probabilités et Potentiel*, **1**, Herman, Paris.

Dellacherie, C. and Meyer, P.A. (1980), *Probabilités et Potentiel: Théorie des Martingales.* Herman, Paris.

Dellacherie, C. and Meyer, P.A. (1983), *Probabilités et Potentiel: Théorie Discrète du Potentiel.* Herman, Paris.

Duren, P. (2000), *Theory of H^p Spaces.* Dover, New York.

Durrett, R. (1984), *Brownian Motion and Martingales in Analysis.* Wadsworth, Belmont, California.

Elliott, P. (1979), *Probabilistic Number Theory I.* Springer, New York.

Elliott, P. (1980), *Probabilistic Number Theory II.* Springer, New York.

Elliott, P. (1985), *Arithmetic Functions and Integer Products.* Springer, New York.

Erdös, P. (1942), "On the Law of the Iterated Logarithm". *Ann. Math.* **43**, 419-436.

Erdös, P. and Kac, M. (1940), "The Gaussian Law of Errors in the Theory of Additive Number-theoretic Functions". *Amer. Jour. Math.* **62**, 738-742.

Ethier, S. and Kurtz, T. (1986), *Markov Processes: Characterization and Convergence.* Wiley, New York.

Fefferman, C. and Stein, E. (1972), "H^p Spaces of Several Variables". *Acta. Math.* **129**, 137-193.

Feller, W. (1971), *An Introduction to Probability and its Application* **2**, 2nd Ed., Wiley, New York.

Flehinger, B. (1966), "On the Probability a Random Integer has Initial Digit A". *Amer. Math. Monthly* **73**, 1056-1061.

Fukushima, M., Oshima, Y. and Takeda, M. (1994), *Dirichlet Forms and Symmetric Markov Processes.* de Gruyter, Berlin.

Garsia, A. (1973), *Martingale Inequalities*, Benjamin, Reading, Mass.

Garnett, J. (1981), *Bounded Analytic Functions*, Academic Press, New York.

Hardy, G.H. (1915), "The Mean Value of the Modulus of an analytic Function". *Proc. Lond. Math. Soc.* **14**, 269-277.

Hardy, G.H. (1928), "Notes on Some Points in the Integral Calculus LXIV: Further Inequalities Between Integrals". *Messenger of Mathematics* **5**, 12-16.

Hardy, G.H. (1928), *A Course of Pure Mathematics.* 6th Ed., Cambridge University Press, Cambridge.

Hardy, G.H. (1949), *Divergent Series.* Oxford University Press, Oxford.

Hardy, G.H. (1954), *Ramanujan.* Chelsea, New York.

Hardy, G.H. (1967), *A Mathematicians Apology, with a Forward by C.P. Snow.* Cambridge University Press, Cambridge.

Hardy, G.H. and Littlewood, J.E. (1930), "A Maximal Theorem with Function-Theoretic Applications. *Acta. Math.* **54**, 81-116.

Hardy, G.H. and Ramanujan, S. (1920), "The Normal Number of Prime Factors of a Number n". *Quaterly Jour. Math.* **48**, 76-92.

Hardy, G.H. and Wright, E.M. (1960), *Introduction to the Theory of Numbers.* 4th Ed., Oxford University Press.

Hildebrand, A. (1986), "The Prime Number Theorem Via the Large Sieve". *Mathematika* **33**, 23-30.

Hill, T. (1995),"A Statistical Derivation of the Significant-Digit Law". *Statistical Science* **10**, 354-363.

Kac, M. (1954), "Statistical Independence in Probability Analysis and Number Theory". *Math. Assoc. Amer.*, Washington.

Kac, M. (1985), *Enigmas of Chance: An Autobiography.* Harper and Row, New York.

Kahane, J.P. (1985), *Some Random Series of Functions.* Cambridge University Press, Cambridge.

Kakutani, S. (1945), "Two-Dimensional Brownian Motion and Harmonic Functions". *Proc. Imp. Acad: Tokyo* **20**, 706-714.

Kellogg, O. (1953), *Foundations of Potential Theory.* Dover, New York.

Khintchine, A. (1923), "Über Dyadische Brüche". *Math. Zeit.* **18**, 109-116.

Knuth, D. (2000), *Selected Papers on analysis of Algorithms.* CLSI Lecture Notes, Stanford, California.

Koosis, P. (1999), *Introduction to H^p Spaces 2nd Ed..* Cambridge University Press, Cambridge.

Kramkov, D. (1996), "Optional Decomposition of Supermartingales and Hedging Contingent Claims in Incomplete Security Markets". *Probability Theory and Related Fields* **105**, 459-479.

Lai, T.L. (1973), "Optional Stopping and Sequential Tests which Maximize the Maximum Expected Sample Size". *Ann. Statist.* **1**, 659-673.

le Gall, J.F. (1994), *Spatial Branching Processes, Random Snakes and Partial Differential Equations*, Birkhauser, Basel.

Littlewood, J.E. (1957), *A Mathematicians Miscellany.* Methuen, London.

Littlewood, J.E. (1969), "On the Probability in the Tail of a Binomial Distribution". *Adv. Appl. Probab.* **1**, 43-72.

Liu, J. (2001), *Monte Carlo Strategies in Scientific Computing*, Springer, New York.

Meyer, P.A. (1966), *Probability and Potentials.* Blaisdell, Waltham, Mass.

Meyer, Y. (1992), *Wavelets and Operators.* Cambridge University Press, Cambridge.

Miclo, L. (1999), "An Example of Application of Discrete Hardy's Inequality". *Markov Process Related Fields* **5**, 319-330.

Mosteller, F. (1972), "A Data Analytic Look at Goldbach Counts". *Statistsica Neerlandica* **26**, 227-242.

Neveu, J. (1972), *Martingales* à *Temps Discret.* Paris, Masson.

Petersen, K. (1977), *Brownian Motion, Hardy Spaces and Bounded Mean Oscillation.* Cambridge University Press, Cambridge.

Raimi, R. (1976), "The First Digit Problem". *Amer. Math. Monthly* **83**, 521-538.

Revuz, D. and Yor, M. (1999), *Martingales and Brownian Motion.* 3rd Ed., Springer, New York.

Saloff-Coste, L. (1997), *Lectures on Finite Markov Chains. Springer Lecture Notes in Math* **1665**, 361-413, Springer Verlag, Berlin.

Sathe, L. (1954), "On a Problem of Hardy on the Distributions of Integers Having a Given Number of Prime Factors". *J. Ind. Math. Soc.* **18**, 27-81.

Stein, E. and Weiss, G. (1975), *Introduction to Fourier Analysis on Euclidian Spaces*, Princeton University Press, Princeton.

Tenenbaum, J. (1995), *Introduction to Analytic and Probabilistic Number Theory.* Cambridge University Press, Cambridge.

Turan, P. (1934), "On a Theorem of Hardy and Ramanujan". *Jour. Lond. Math. Soc.* **9**, 274-276.

Varopoulos, N. (1980), "The Helson-Szegö Theorem and A_p Functions for Brownian Motion and Several Variables". *J. Functional Anal.* **39**, 85-121.

Vaughn, R. (1997), *The Hardy-Littlewood Method.* 2nd Ed., Cambridge University Press, Cambridge.

Williams, D. (1991), *Probability with Martingales.* Cambridge University Press, Cambridge.

Zabell, S. (1995), "Alan Turing and the Central Limit Theorem". *Amer. Math. Monthly* **102**, 483-494.