

SURVEY ARTICLE

HARDY, LITTLEWOOD AND PROBABILITY

N. H. BINGHAM

Abstract

We trace the interplay between the Hardy-Littlewood school of analysis, prominent in the UK in the first half of the last century, and probability theory, barely known in the UK at that time.¹

1. Introduction

This piece is intended as a sequel to the splendid paper by Diaconis, G. H. *Hardy and probability???* [Dia], published in *Bull. LMS* in 2002. There, Diaconis took the view that, despite Hardy's public disdain for applied mathematics, including probability as he then saw it (and as it largely was then), his work has many probabilistic overtones. Littlewood always had a more positive view of both.

We follow Diaconis' ordering of subject-matter: probabilistic number theory (§2), summability and Tauberian theory (§3) and Hardy spaces (§4), before turning to the principal individuals involved.

2. Probabilistic number theory

Recall the *prime divisor functions*: $\omega(n) := \#$ distinct prime divisors n ; $\Omega(n) := \#$ prime divisors n (counted with multiplicity) – *additive* arithmetic functions, as they satisfy $f(mn) = f(m) + f(n)$ if $(m, n) = 1$. For background, see e.g. Tenenbaum [Ten, III.3, III.4], Montgomery and Vaughan [MonV, §§1.3, 2.4, 7.4].

Edmund Landau (1877-1938) proved in 1900 [Lan1] that if $\pi_k(x)$ is the number of $n \leq x$ with k distinct prime factors ($k = 1, 2, \dots$),

$$\pi_k(x) \sim \frac{x}{(k-1)!} \cdot \frac{(\log \log x)^{k-1}}{\log x}.$$

As each n has at least one prime factor, it is better to work with $k+1$ rather than k . Writing

$$\lambda := \log \log x$$

¹MSC: 01, 60

(so $\lambda \rightarrow \infty$ as $x \rightarrow \infty$ – though *extremely slowly*):

$$\frac{1}{x} \pi_{k+1}(x) \sim \frac{(\log \log x)^k}{k! \log x} = \frac{e^{-\lambda} \lambda^k}{k!} \quad (k = 0, 1, 2, \dots) \quad (\lambda, x \rightarrow \infty).$$

Now $\{e^{-\lambda} \lambda^k / k! : k = 0, 1, 2, \dots\}$ forms the *Poisson distribution* $P(\lambda)$ of Probability Theory, with parameter λ (mean λ , variance λ), though Landau made no mention of Poisson or of probability (the extensive 54-page bibliography of his *Handbuch* of 1909 [Lan2] does not cite Poisson). So, although Landau did not do so, one may rephrase his result above as follows: the *proportion* of integers $\leq x$ with $k+1$ distinct prime factors is asymptotically that of a Poisson distribution $P(\lambda)$ with parameter $\lambda := \log \log x$. This first hint of probability theory in analytic number theory is a precursor of probabilistic number theory, our subject here.

Regarding the prime divisor functions, the following result was proved by Hardy and Ramanujan in 1917 ([HarRa]; [HarW, Th. 430]).

$$(i) \sum_{n \leq x} \omega(n) = x \log \log x + C_1 x + O(1/\log x),$$

$$(ii) \sum_{n \leq x} \Omega(n) = x \log \log x + C_2 x + O(1/\log x),$$

where the C_i are specified constants. This result (without the constants or error terms) is summarised by Hardy and Wright as saying that *the average order of both $\omega(n)$ and $\Omega(n)$ is $\log \log n$* . They also treat *normal order*, and show that $\omega(n)$, $\Omega(n)$ have normal order $\log \log n$ [HarW Th. 431]. See also Tenenbaum [Ten, §3.4], where this result is derived from the Turán-Kubilius inequality [Ten §3.2], and Turán’s quantitative form of 1934 [Tur] is also given. Equip the set $\mathbb{N}_N := \{1, 2, \dots, N\}$ with the *uniform* distribution P_N (mass $1/N$ on each member). If $\xi(n) \rightarrow \infty$,

$$P_N(\{n : |\omega(n) - \log \log n| > \xi(n) \sqrt{\log \log n}\}) << 1/\xi^2(N),$$

and similarly for $\Omega(n)$. Tenenbaum remarks that this result may be viewed nowadays as the birth of probabilistic number theory.

If one has the Poisson distribution in mind (which presumably Hardy and Ramanujan did not, any more than Landau did), one sees the $\log \log x$ term on the right as the mean and variance of the Poisson distribution $P(\lambda)$ in Landau’s result above. What this suggests, of course, is a law of large numbers (LLN). And where one has an LLN, one may seek a central limit theorem (CLT) Since the Poisson law $P(\lambda)$ is the n -fold convolution of $P(\lambda/n)$ for each n , the ordinary CLT shows that for large λ $P(\lambda)$ is approximately normal $N(\lambda, \lambda)$ with mean and variance λ .

The CLT for this case followed in 1939, as the Erdős-Kac central limit theorem (completed in 1939, during a seminar given by Kac: Erdős was in the audience, and completed the proof by using *sieve methods*), the other candidate for the birth of probabilistic number theory. Write as usual

$$\Phi(x) := \int_{-\infty}^x \phi(u) du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}u^2} du.$$

Theorem (Erdős-Kac CLT, 1939).

$$P_N(n : \omega(n) \leq \log \log N + x\sqrt{\log \log N}) \rightarrow \Phi(x) \quad (N \rightarrow \infty),$$

and similarly for Ω .

Where one has a central limit theorem, one looks for a Berry-Esseen rate-of-convergence result (following A. C. Berry in 1941 and C.-G. Esseen in 1942). This was found by Rényi and Turán [RenT] in 1958, following earlier work by Turán in 1934 and 1936 on error estimates in the Hardy-Ramanujan theorem. For an integrated treatment of the Erdős-Kac and Rényi-Turán theorems, see [Ten III.4.4]. This uses the Selberg-Delange method (Selberg in 1954, Delange in 1959 and 1971, following work of Sathe in 1953 and 1954; [Ten II.5]). This extends the Landau result – estimates of $\pi_k(x)$ and its analogue with repetitions allowed – from fixed k to $k = O(\log \log x)$ [Ten II.6.1].

The Poisson $P(\lambda)$ and normal laws occur together here, as above (though in discrete situations such as Landau’s result the Poisson is perhaps more natural). Poisson and normal approximation can both be handled by the Chen-Stein method [BarHJ] (see §3 below); this method has recently been applied to give a short proof of the Erdős-Kac theorem by Harper [Harp].

Diaconis tellingly says [Dia, 387]: ‘My interests in Hardy and probability started when Erdős said: ‘You know, if Hardy had known anything of probability, he certainly would have proved the law of the iterated logarithm². He was an amazing analyst, and only missed the result because he didn’t have any feeling for probability’ (which could also be said of Landau). This also delayed the start of probabilistic number theory, but the subject is now flourishing; see e.g. [Kub], [Ten], [Ell].

Diaconis [Dia, §§2, 3] covers much of the material here, including the

²Hardy and Littlewood obtained a partial result in 1914, following Hausdorff in 1913. The complete result was obtained by Khinchin in 1924.

Turán-Kubilius inequality, and Hildebrand’s use of it and sieve methods to prove the Prime Number Theorem [Hil]; as he says, these results played a major role in his mathematical development.

This fascinating interplay between probabilistic methods on the one hand, and the distribution of the primes (as ‘God-given’ and deterministic as anything could be) on the other is neatly summarised by Kac in the chapter heading of Ch. 4 in his delightful book [Kac1]: ‘Primes play a game of chance’. We shall call this *Kac’s dictum*. Another way of putting this is due to the analytic number theorist R. C. Vaughan (1945–): ‘It’s obvious that the primes are randomly distributed – we just don’t know what that means yet’ – *Vaughan’s dictum*.³

3. Summability and Tauberian theory

The connections between probability and summability were also influential on Diaconis’ early work [Dia, §4].

The most basic summability method is the Cesàro method of order 1, C_1 , which maps a sequence (s_n) into its sequence of arithmetic means. This preserves convergence, and limits, when present, but may well produce convergence when not. The second-order method C_2 is obtained by applying C_1 twice, etc.; the family C_α for $\alpha > 0$ is obtained by replacing all factorials by Gamma functions. See e.g. [Har3, V, VI] for background and details.

The next summability method one usually meets is the *Abel method* A , with geometrically decreasing weights. See [Har3, VII] for details, and connections between the Cesàro and Abel methods.

Next, one has the *Borel method*, with *Poisson* weights $e^{-x}x^k/k!$ (and one takes limits as $x \rightarrow \infty$), and the *Euler methods* E_p ($0 < p < 1$), with weights those of the *binomial* distribution $B(n, p)$, namely $\binom{n}{k}p^k(1-p)^{n-k}$, and takes limits as $n \rightarrow \infty$. For details, see [Har3, VIII, IX].

In classical summability theory, the interest of such methods is largely in the possibility they offer to make divergent sequences converge. In particular, in complex analysis they may be used to extend the domain of analyticity of a power series, and this is largely the view taken in [Har3].

Laws of large numbers (LLNs).

In probability theory, however, one has another view – that of laws of

³My wife’s instant response to this:
(*Cecilie*) *Bingham’s dictum*: Primes play a game of chance – we just don’t know the rules yet.

large numbers. If $X, X_1, \dots, X_n, \dots$ are independent and identically distributed (iid) random variables, then

$$\frac{1}{n} \sum_1^n X_k \rightarrow c \quad (n \rightarrow \infty), \quad \text{i.e.} \quad X_n \rightarrow c \quad (C_1)$$

for some constant c if and only if

$$E[|X|] < \infty, \quad \text{and then} \quad E[X] = c.$$

This is *Kolmogorov's strong law of large numbers (SLLN)* of 1933, one of the key results in his pioneering *Grundbegriffe* [Kol1]. Incidentally, this result completes a large number of preliminary results of this type, going back to *Bernoulli's theorem* (Jakob Bernoulli (1654-1705), in his posthumous *Ars Conjectandi* (AC) of 1713 – weak LLN for Bernoulli trials). It is the final mathematical form of the folklore idea of a "Law of Averages", and underlies the interpretation (it is no more than that) of probability as a limiting frequency. (One can weaken independence, and identical distribution, and work on more general spaces, but we keep to the real iid case here for simplicity.) *Probability and summability.*

Remarkably, Kolmogorov's result holds verbatim if the summability method C_1 in it is replaced by C_α for each (and so all) $\alpha \geq 1$, and by the Abel method A . This result is due to T.-L. Lai in 1974 [Lai]. This gives a sense in which the *Abel-Cesàro* family of summability methods is the natural family in probability theory in the L_1 case – when the *mean* or *first moment* $E[X]$ exists.

There is a complement to this. Y.-S. Chow [Cho] showed in 1973 that

$$\sum_1^\infty \frac{e^{-x} x^k}{k!} X_k \rightarrow c \quad (x \rightarrow \infty), \quad \text{i.e.} \quad X_n \rightarrow c \quad (B)$$

for some constant c if and only if

$$E[|X|^2] < \infty, \quad \text{and then} \quad E[X] = c.$$

The same result holds if the Borel method B is replaced by the Euler method E_p for some (and so for all) $p \in (0, 1)$. This makes precise a sense in which the *Euler-Borel* family of summability methods is the natural family in probability in the L_2 case – when the *variance* or *second moment* exists.

It turns out that Kolmogorov's SLLN is reflected in a discontinuity of

functional form in the result for C_α as α passes through 1 (Bingham and Tenenbaum [BinT], cf. [Bin7]):

Theorem (LLN for C_α).

(i) For $0 < \alpha \leq 1$,

$$X_n \rightarrow \mu \quad (C_\alpha) \quad \Leftrightarrow \quad E[|X|^{1/\alpha}] < \infty \quad \& \quad E[X] = \mu.$$

(ii) For $\alpha \geq 1$,

$$X_n \rightarrow \mu \quad (C_\alpha) \quad \Leftrightarrow \quad E[|X|] < \infty \quad \& \quad E[X] = \mu.$$

Case (ii) is Lai's result above. For case (i), the case $\alpha \in (\frac{1}{2}, 1)$ is due to G. G. Lorentz in 1955 [Lor], for $\alpha \in (0, \frac{1}{2})$ to Chow and Lai in 1973 [ChoL]; the case $\alpha = \frac{1}{2}$ is in Dénél and Derriennic [DenD] in 1988.

Riesz means.

Chow's 'delayed sums' [Cho] ('moving averages' in [Bin7]) involved averages of the form $\frac{1}{c\sqrt{n}} \sum_{n \leq k < n+c\sqrt{n}} X_k$. It turns out that these already had a long pedigree in classical summability theory, as *typical means*, or *Riesz means*. These were introduced by Marcel Riesz (1886-1969) in 1909, as they are ideally suited to analytic continuation of Dirichlet series, as ubiquitous in analytic number theory as power series are in complex analysis. They are used for this purpose in the book by Hardy and Riesz [HarRi]⁴; there is a monograph account in [ChaM].

The theorem above deals with the powers, or L_p -spaces, but Riesz means enable one to tie general moments accurately to the length of the interval along which one takes the moving average. Take a function $\lambda(\cdot) \uparrow \infty$, and write $\lambda_n = \lambda(n)$. For $s_n = \sum_0^n a_k$, write

$$s_n \rightarrow s \quad R(\lambda, 1)$$

(or just $R(\lambda)$ if, as here, one uses only order $k = 1$) for

$$\frac{1}{x} \int_0^x \left\{ \sum_{n: \lambda_n \leq y} a_n \right\} dy \rightarrow s, \quad \text{i.e.} \quad \frac{1}{\lambda(x)} \int_0^x s(y) d\lambda(y) \rightarrow s \quad (x \rightarrow \infty)$$

⁴This book was published in 1915; Riesz was Hungarian. The title page contains a memorable dedication ending *Auctores hostes idemque amici*: The authors, enemies and yet friends.

(cf. Karamata [Kar]). It turns out that for $p \geq 1$

$$R_p := R(\exp\{\int_1^n x^{-1/p} dx\}, 1)$$

gives a summability method equivalent to the method M_p obtained from moving averages over intervals $[n, n + cn^{1/p}]$ ([BinT], [Bin7]), and that each of R_p, M_p is interchangeable with $C_{1/p}$ in Case (i) of the theorem above.

Self-neglecting functions.

More generally, call a (continuous) function ϕ *self-neglecting*, or *Beurling slowly varying*, $\phi \in SN$, if

$$\phi(x + t\phi(x))/\phi(x) \rightarrow 1 \quad (x \rightarrow \infty) \forall t \in \mathbb{R}$$

(see [BinGT, §2.11], [BinO1], [BinO2] for some of the extensive background on such functions). For $\phi(\cdot) \uparrow \infty$, write

$$\Phi(x) := \int_1^x dt/\phi(t), \quad \lambda(x) := \exp\{\Phi(x)\}, \quad \psi := \phi^{\leftarrow}.$$

Then [BinG, Th. 1] moving averages over intervals of lengths $t\phi(x)$ converge to s ($s_n \rightarrow s$ ($M(\phi)$) iff $s_n \rightarrow s$ ($R(\lambda)$). In the probabilistic context, one needs the condition that $\psi := \phi^{\leftarrow} \in BI$ (is of *bounded increase*: [BinGT, §2.1.2]). Then $X_n \rightarrow \mu$ ($M(\phi)$) iff $E[\phi(X)] < \infty$ and $E[X] = \mu$ [BinG, Th. 3]. The condition of bounded increase is needed to preserve LLN-type behaviour, in which the details of the distribution are lost in the limit, leaving only the mean. If ψ grows too fast (e.g., exponentially), the *entire distribution* survives the passage to the limit (*Erdős-Rényi law*, or almost-sure *non-invariance principle*: see e.g. [BinG] for details and references).

Random-walk and circle methods.

My own entry into this field was triggered by writing the Mathematical Reviews to two papers, by Diaconis and Stein [DiaS] and by Schmaal, Stam and de Vries [Sch]. For the Euler and Borel methods, their weights are probabilities (of Poisson or binomial distributions). For matrix methods of summation $A = (a_{nk})$, call A a *random-walk method* if

$$a_{nk} = P(S_n = k)$$

for some random walk $S_n = X_1 + \dots + X_n$, with X_n iid. Quite a lot can be said about these methods; see e.g. Korevaar [Kor] VI, [Bin1-7].

If one uses normal (Gaussian) approximation in the weights above, one obtains a (discrete version of) a *Valiron method*. These, together with *Taylor methods* and *Meyer-König methods*, are included in Meyer-König's family of *circle methods* or *Kreisverfahren* [MeyK]. Some of this material is included in [Har3, VIII], though not all: Hardy died in 1947 and both appeared in 1949. The motivation here is analytic continuation of power series (hence 'circle', for the circle of convergence). See [Bin3] and [BinT] for more detail here.

Gap theorems.

The Borel method involves a *continuous* passage to the limit, the Euler (and random-walk) methods a *discrete* passage to the limit. It turns out that there are subtle differences. The Borel and Euler methods have *gap* or *high-indices* theorems [Har3, §7.13]: if $s_n = \sum_1^n a_k$ with the a_n *vanishing* off some suitably sparse subsequence (n_k) , then *no* order condition is needed in the relevant Tauberian theorem. The 'Euler gap theorem' for the Euler method ($n_{k+1} - n_k \geq h\sqrt{n_k}$ for some $h > 0$) is due to Meyer-König and Zeller in 1956 [MeyK-Z1] (cf. [MeyK-Z3]); the corresponding 'Borel gap' theorem is due to Gaier in 1965 [Gai] (following partial results by Erdős and others). That there is *no* (pure) gap theorem for the discrete Borel method was shown by Meyer-König and Zeller in 1960 [MeyK-Z2] by functional-analytic methods. For more detail and further references, see [Bin6, §3.1].

The Chen-Stein method.

A powerful way of approximating by the normal distribution was introduced by Charles Stein in 1970 (unpublished), and extended to Poisson approximation by Chen in 1976 [Che]. A monograph on the *Chen-Stein method* is given by Barbour, Holst and Janson [BarHJ]. This is clearly highly relevant to the probabilistic number theory of §2; see Harper [Harp].

Law of the iterated logarithm (LIL) and of the single logarithm (LSL).

We have discussed LIL above, in connection with Erdős' comment to Diaconis about Hardy. It concerns averages, and so is a strong limit theorem for C_1 , holding in L_1 . There are analogues for $C_{1/\alpha}$, holding in L_α , for $\alpha > 1$, but with only a single rather than an iterated logarithm; see e.g. [Bin5,6,7].

The logarithmic method.

Above, one has summability methods requiring *more* than the Cesàro method (or in probabilistic terms, one deals with a *refinement* of the Kolmogorov strong law). The classical logarithmic method is a prime example of where one requires *less* (or deals with coarsenings of the Kolmogorov strong law). This is useful in a number of areas, particularly probability theory and analytic number theory ([Har3, §§4.16, 5.16]; Bingham and Gashi [BinGa];

[Bin15]).

4. Hardy spaces.

Hardy spaces grew out of a series of papers by Hardy on functions analytic on the unit disk \mathbb{D} , and the boundary values (on the unit circle \mathbb{T}) that they have under suitable growth restrictions. The first monograph treatment, by Duren in 1970 [Dur], cites three papers by Hardy and eight by Hardy and Littlewood, over the period 1913-1941 (distributed over Volumes II, III and IV of Hardy's *Collected Papers*). Hardy spaces H_p were named and studied by F. Riesz in 1923. The focus originally was on individual functions, but with the development of functional analysis (a milestone here being Banach's 1932 book [Ban]), spaces of functions began to be studied for their own sake – by A. E. Taylor and by S. S. Walters in 1950, by W. Rudin and by R. P. Boas in 1955, and others. In addition to [Dur], other useful books are by Hoffman in 1962 [Hof], Garnett in 1981 [Gar], and Rosenblum and Rovnyak in 1985 [RosR]. The general area of spaces of holomorphic functions is nowadays called *holomorphic spaces* for short; see e.g. the survey by Sarason in 1998 [Sar] and the book containing it, by Axler et al. [Axl].

Hardy spaces were studied (with the notation H_p and with references to Hardy, but without the name Hardy spaces) by Marcinkiewicz and Zygmund in 1940 [MarZ]. For background on the links between the Polish school and probability, see e.g. Holgate [Hol], [Bin9], [Bin11]. For modern textbook accounts, see e.g. Durrett [Durr, Ch. 6,7], Bass [Bas, Ch. IV].

Another probabilistic strand grew largely out of the work of Gabor Szegő (1895-1985), in papers from the period 1915-21 (so a little after Hardy's work began) and later of 1952. An important development was Kolmogorov's major paper of 1941 on stationary sequences in Hilbert space [Kol2]. This paper introduced the *Kolmogorov isomorphism theorem (KIT)*, which (for a stationary stochastic process $X = (X_n)$ in discrete time) enables one to pass at will between the *time domain*, in which one focusses explicitly on time n , and the *frequency domain*, in which one focusses on the *spectral measure* μ on \mathbb{T} . The next year, Cramér [Cra2] showed that X is the sequence of Fourier coefficients of a random measure Y on \mathbb{T} , Y and μ being linked by

$$E[(dY(\theta))^2] = d\mu(\theta).$$

This shows clearly the link with Itô calculus of 1944, where for Brownian motion $B = (B_t)$ one has $E[(dB_t)^2] = dt$. For textbook expositions making the links here clear, see e.g. Janson [Jan] and Kallenberg [Kal].

In 1948, Beurling [Beu] continued the use of Hilbert-space methods to study stationary discrete-time stochastic processes, and showed that the Lebesgue decomposition $\mu(d\theta) = w(\theta)d\theta + \mu_s(d\theta)$ of μ has an interpretation in terms of invariant subspaces of Hilbert space (under the shift, $n \mapsto n+1$), the decomposition of μ corresponding to the role of the *remote past*. The nice situation is that where the remote past is forgotten (the tail σ -field at $-\infty$ is trivial). For this, one needs *Szegő's condition* ($\log w \in L_1$). This visibly involves entropy, and is in fact related to the Gibbs variational principle of statistical mechanics. One also needs $\mu_s = 0$. There is a great deal more to be said, but for brevity we refer here to the writer's recent surveys on this subject, [Bin12] (one dimension), [Bin13] (higher dimensions), [Bin14] (statistics), to [BinIK], [KasB], [BinM], to the books on orthogonal polynomials on the unit circle and Szegő's theorem by Simon [Sim1,2], and the books by Nikolskii [Nik1,2] on the shift operator and spectral function theory.

5. Hardy.

Hardy was a pure mathematician through and through, famous (or notorious, depending on one's point of view) for his lack of interest in applied mathematics; he left a fine account of his views of mathematics in the autobiographical *A mathematician's apology* [Har2]. It is not quite true that Hardy showed no interest in probability: he wrote on (what is now known as) the Hardy-Weinberg law in Mendelian genetics in 1908 ([Har1]; the result was obtained independently by Weinberg [Wei], also in 1908; [Fel, V.5]). Indeed, J. M. Hammersley (1920-2004)⁵ was fond of saying that it was for the Hardy-Weinberg law that Hardy would ultimately be principally remembered. Hardy appears in §2 in his work of 1917 with Ramanujan, in the middle of the transitional period between the introduction of measure theory by Lebesgue in 1901 and its successful harnessing to the service of probability theory by Kolmogorov in 1933 (see e.g. [Bin9] for an account of this interesting transitional period). But Williams' description of probability theory pre-measure theory as a shambles [Wil, §1.3] is fair, and it is understandable that Hardy chose not to divert his energies from his many interests in pure mathematics to study probability theory.

Hardy began publication in 1899, before Lebesgue's work. His series *Notes on some points in the integral calculus, I-LXIV* (NIC in Hardy's works)

⁵Hammersley taught the present writer when he was an undergraduate at Trinity College, Oxford, 1963-66)

covered 1901-29, beginning with the Riemann integral and later using the Lebesgue integral when needed. Hardy's views on the Lebesgue integral are available directly, e.g. in [HarLP, §6.1] of 1934. This passage is quoted by Burkill in the Preface to his Cambridge Tract [Bur] of 1951, the first standard work in English on the Lebesgue integral (though Burkill there acknowledges his debt to the treatment [Tit, X-XII] in the 1932 book by Titchmarsh, a pupil of Hardy); here Burkill (writing in 1949, after Hardy's death in 1947) writes 'I wish to record that one of my many debts to G. H. Hardy lay in his encouragement to write this tract'. But Hardy's views on the Lebesgue integral are also available vividly indirectly, through his influence on Wiener (§7 below).

Despite his openness to the Lebesgue integral, Hardy seems to have taken little interest in functional analysis. It fell to Frank Smithies (1912-2002), whose Cambridge PhD of 1936 on integral equations was written under Hardy's supervision [Smi], to introduce the then still new functional analysis to the UK in the late 1930s (from France – Laurent Schwartz and co-workers)⁶.

As noted in §4, Hardy spaces are a prime example of the intermingling of complex analysis and functional analysis – or, as they are sometimes called, 'hard' (real or complex) and 'soft' (abstract or functional) analysis. Hardy is often quoted as making comments in which his preference for the first is clear.

6. Littlewood.

J. E. Littlewood (1885-1977) was the younger partner (to G. H. Hardy (1877-1947) in the Hardy-Littlewood collaboration (95 papers, from 1912 to 1948) – the longest and most famous in mathematical history. Both wrote excellent and readable autobiographical books, [Har2] and [Lit1]. Their styles were complementary: Hardy normally did the initiating and writing-up; Littlewood bore the brunt of the hard problem-solving.

Volume I of Littlewood's *Collected Papers* [Lit2] contains his collaboration with R. E. A. C. Paley (1907-33): three papers, 'Theorems on Fourier series and power series', a short one of 1931 and two long ones of 1937 after Paley's death (recall that the book by Paley and Wiener [PalW] was also posthumous). The resulting Littlewood-Paley theory has connections with martingales, and is important in both analysis and probability. See e.g. Stein

⁶Smithies lectured to the present writer on functional analysis while he was a research student at Cambridge, 1966-69.

[Ste] for analysis, Varopoulos [Var] for probability.

Volume II of [Lit2] (§5, Probabilistic analysis, p.1578) contains five papers with his pupil A. C. (Cyril) Offord (1906-2000) on real roots of random polynomials. This subject was continued, by Kac [Kac2, I], by Offord's pupil J. E. A. Dunnage, and by Dunnage's pupil K. Farahmand.

7. Wiener.

The American Norbert Wiener (1894-1964) was a wonderfully broad and prolific mathematician. In addition to a number of excellent mathematics book, and four volumes of collected papers (1976-1985), Wiener wrote a two-volume autobiography, [Wie2,3]. In [Wie2, XIV] (Emancipation: Cambridge, June 1913 – April, 1914) Wiener gives (as well as his observations on the Cambridge and the England of that time) an account of what a revelation Hardy's lectures (on the Lebesgue integral) were. In [Wie3, Ch. 7] (An unofficial Cambridge don, 1931-2), he describes his dealings with a much older Hardy and with Littlewood; this led to his writing his book [Wie1] on the Fourier integral.⁷ He also comments on how Hardy teased him, saying that he was really a pure mathematician at heart, and only did applicable work to curry favour with his engineering colleagues at MIT. In [Wie3, Ch. 8] (Back home, 1932-33) Wiener describes his short but productive collaboration with Paley, before Paley's untimely death in a skiing accident and the posthumous book [PalW].

8. Pitt.

In 1937-38 H. R. Pitt (1914-2005; Sir Harry Pitt from 1978), a pupil of Hardy, visited MIT and worked with Wiener, mainly on Tauberian theory. This year had a transformative effect on Pitt, and led to Pitt's form of Wiener's (Tauberian) theorem (or Wiener-Pitt theorem). For details, see [BinH].

Pitt was primarily an analyst, best known for his work on Tauberian theorems, but he retained the interest in probability theory that he acquired from Wiener during his *annus mirabilis* of 1937-8. Pitt later supervised C. W. J. Granger (1934-2009; Nobel Prize in Economics, 2003; Sir Clive Granger, 2005). Granger's work has been extremely influential in econometrics and

⁷This book, incidentally, makes the case for teaching the Fourier integral via the Lebesgue integral so convincingly that, despite the extra difficulties that learning measure theory and the measure-theoretic integral presents, I find it both saddening and surprising that it is so often not taught to UK undergraduates.

time series; see e.g. [Bin12, §1].

9. Cramér.

The Swedish mathematician Harald Cramér (1893-1985) took his PhD at Stockholm in 1917 on Dirichlet series, under Marcel Riesz, who had excellent links with Hardy ([HarRi]; cf. [Bin14]). Cramér worked originally on analytic number theory (see [Cra5, I]), but moved to probability and statistics as he had a family to support and took up a job as an actuary. From his autobiographical account [Cra4, §3.1, §3.2]: he visited Cambridge in 1920, where he worked under Hardy and met Wiener [Wie2, 64]. ‘During a visit to England in 1927, I saw my old teacher and friend G. H. Hardy. When I told him that I had become interested in probability theory, he said that there was no mathematically satisfactory book in English on this subject, and encouraged me to write one.’ This task actually took a decade. Cramér was much influenced by Kolmogorov’s *Grundbegriffe* of 1933. His book finally appeared as a Cambridge Tract in 1937 [Cra1]. It was the first clear treatment of measure-theoretic probability theory in English, and it owes its existence to Hardy’s encouragement. Cramér followed this up in 1946 with [Cra3], the first book to synthesise both the measure-theoretic probability of Kolmogorov and the pioneering work of Fisher in statistics.

10. Daniell.

P. J. Daniell (1889-1946) was Professor of Mathematics at Sheffield from 1923 (and the last person to be Senior Wrangler in Cambridge, in 1909). He is remembered for the Daniell integral (work of 1918), and the Daniell-Kolmogorov theorem, the foundational existence theorem for a stochastic process (Daniell, 1919; Kolmogorov, *Grundbegriffe*, 1933). A fascinating account of Daniell’s life, times and work is given by Aldrich [Ald]. His paper has many points of contact with this and with [Dia], and is strongly recommended to all readers of this piece.

11. Others.

Perhaps the other two people most relevant to British probability before WWII were Sir Harold Jeffreys (1891-1989) and R. A. (Sir Ronald) Fisher (1890-1962). Jeffreys was primarily a geophysicist. He was the first to realise that the Earth’s core is molten iron⁸. He wrote a book *Theory of probability*

⁸(but was a scientific conservative: he was a vehement opponent of Continental Drift)

[Jef], largely devoted to Bayesian statistics (on which see e.g. [Bin10]). Fisher was both the greatest statistician ever and a geneticist of the first rank. His theory of ‘fiducial probability’ is often described as Fisher’s great mistake. Both had a traditional background in applied mathematics, and (perhaps as a result) both actively disliked measure-theoretic probability. Both were in Cambridge for much of their careers, but in different areas (Fisher was Professor of Genetics, Jeffreys worked in mathematics, then geophysics, and was then Professor of Astronomy), and did not get on⁹.

Most of the key figures in UK probability (and statistics) after WWII were much affected by the War itself. Turing led the attack on Enigma at Bletchley Park; Good was Turing’s statistical assistant. D. G. Kendall was an analyst at Oxford before the War, with leanings towards astronomy; he learned statistics from Bartlett for his war work (on rockets, with Bartlett), and became the leading UK probabilist after the war, first in Oxford (where he was a colleague of J. M. Hammersley, who was in the Royal Artillery in WWII), and then in Cambridge (see e.g. [Bin8]). Daniell’s health was badly damaged by his war work. Barnard worked under Bartlett on the then new area of Operational Research.

For an account of the vexed passage from Lebesgue’s work on measure theory to Kolmogorov’s *Grundbegriffe* on probability theory, with particular reference to the Russian, Polish, French, German, Italian and American schools, see [Bin9]. The contributions of the UK were mainly statistical and philosophical rather than mathematical (e.g. Keynes and Ramsey, in addition to the names above).

For an entertaining glimpse into probability in Cambridge, 1957-67, see Kingman [Kin]: it was a shambles. Fortunately, UK probability is in better health nowadays.

12. Conclusion.

The Hardy-Littlewood school of analysis in general, and G. H. Hardy in particular, emerge from all this not as probabilists, but as analysts aware of and sympathetic to probability. In particular, Hardy’s benign and creative influence on Wiener and Cramér, and on Pitt and others within the UK, deserves our thanks.

This piece is offered as a sequel and complement to Diaconis’ *G. H. Hardy*

⁹Both men worked on remanent magnetism – study of the Earth’s history as locked in by the magnetisation of rocks. This has clear implications for Continental Drift.

and probability??? [Dia]. I hope that all readers will read (or re-read) this first.

References

- [Ald] J. Aldrich, But you have to remember P. J. Daniell of Sheffield. *Electronic J. Hist. Prob. Stat.* **3.2** (2007), 58p.
- [Axl] S. J. Axler, J. E. McCarthy and D. Sarason (ed.), *Holomorphic spaces*. Cambridge Univ. Press, 1998.
- [Ban] S. Banach, *Théorie des opérations linéaires*. Monografie Matematyczne **1**, Warszawa, 1932.
- [BarHJ] A. D. Barbour, L. Holst and S. Janson, *Poisson approximation*. Oxford Univ. Press, 1992.
- [Bas] R. F. Bass, *Probabilistic techniques in analysis*. Springer, 1995.
- [Beu] A. Beurling, On two problems concerning linear transformations in Hilbert space, *Acta Math.* **81** (1948), 239-255 (reprinted in *The Collected Works of Arne Beurling*, Volume 1, Birkhäuser, 1989, p.147-163).
- [Bin1] N. H. Bingham, Tauberian theorems and the central limit theorem. *Ann. Probab.* **9** (1981), 221-231.
- [Bin2] N. H. Bingham, On Euler and Borel summability. *J. London Math. Soc.* (2) **29** (1984), 141-146.
- [Bin3] N. H. Bingham, On Valiron and circle convergence. *Math. Z.* **186** (1984), 273-286.
- [Bin4] N. H. Bingham, Tauberian theorems for summability methods of random-walk type. *J. London Math. Soc.* (2) **30** (1984), 281-287.
- [Bin5] N. H. Bingham, Variants on the law of the iterated logarithm. *Bull. London Math. Soc.* **18** (1986), 433-467.
- [Bin6] N. H. Bingham, Tauberian theorems for Jakimovski and Karamata-Stirling methods. *Mathematika* **35** (1988), 216-224.
- [Bin7] N. H. Bingham, Moving averages. *Almost everywhere convergence* (ed. G. A. Edgar and L. Sucheston), Academic Press, 1989.
- [Bin8] N. H. Bingham, A conversation with David Kendall. *Statistical Science* **11** (1996), 159-188; Editor's note, *ibid.* **12** (1997), 220.
- [Bin9] N. H. Bingham, Studies in the history of probability and statistics **XLVI**. Measure into probability: from Lebesgue to Kolmogorov. *Biometrika* **87** (2000), 145-156.
- [Bin10] N. H. Bingham, Finite additivity versus countable additivity. *Electronic J. History of Probability and Statistics* **6.1** (2010), 35p.

- [Bin11] N. H. Bingham, Józef Marcinkiewicz: Analysis and probability. *Proc. Józef Marcinkiewicz Centenary Conference* (Poznań, 2010), Banach Centre Publications **95** (2011), 27-44, Inst. Math., Polish Acad. Sci., Warszawa.
- [Bin12] N. H. Bingham, Szegő's theorem and its probabilistic descendants. *Probability Surveys* **9** (2012), 287-324.
- [Bin13] N. H. Bingham, Multivariate prediction and matrix Szegő theory. *Probability Surveys* **9** (2012), 325-339.
- [Bin14] N. H. Bingham, Modelling and prediction of financial time series. *Communications in Statistics: Theory and Methods* **43** (2014), 1-11.
- [Bin15] N. H. Bingham, Riesz means and Beurling moving averages. Preprint.
- [BinGa] N. H. Bingham and B. Gashi, Logarithmic moving averages. *J. Math. Anal. Appl.* **421** (2015), 1790-1802.
- [BinG] N. H. Bingham and C.M. Goldie, Riesz means and self-neglecting functions. *Math. Z.* **199** (1988), 443-454.
- [BinGT] N. H. Bingham, C. M. Goldie and J. L. Teugels, *Regular variation*, Cambridge Univ. Press, 1987.
- [BinH] N. H. Bingham and W. K. Hayman, Sir Harry Raymond Pitt, 1914-2005. *Biogr. Memoirs Fell. R. Soc.* **54** (2008), 257-274 (reprinted in *Bull. London Math. Soc.* **42** (2010)).
- [BinIK] N. H. Bingham, A. Inoue and Y. Kasahara): An explicit representation of Verblunsky coefficients. *Statistics and Probability Letters* **82.2** (2012), 403-410.
- [BinM] N. H. Bingham and Badr Missaoui, Aspects of prediction. *J. Applied Probab.* **51A** (2014), 189-201.
- [BinO1] N. H. Bingham and A. J. Ostaszewski, Beurling slow and regular variation. *Transactions London Math. Soc.* **1** (2014), 29-56.
- [BinO2] N. H. Bingham and A. J. Ostaszewski, Beurling moving averages and approximate homomorphisms. arXiv:1407.4093.
- [BinT] N. H. Bingham and G. Tenenbaum, Riesz and Valiron means and fractional moments. *Math. Proc. Cambridge Phil. Soc.* **99** (1986), 143-149.
- [Bur] J. C. Burkill, *The Lebesgue integral*, Cambridge Tracts **40**, Cambridge Univ. Press, 1951.
- [ChaM] K. Chandrasekharan and S. Minakshisundaram, *typical means*. Oxford Univ. Press, 1952.
- [Che] L. H. Y. Chen, Poisson approximation for dependent trials. *Ann. Prob.* **3** (1975), 534-545.
- [Cho] Y.-S. Chow, Delayed sums and Borel summability of independent identically distributed random variables. *Bull. Inst. Math. Acad. Sinica* **1**

- (1973), 207-220.
- [ChoL] Y.-S. Chow and T.-L. Lai, Limiting behaviour of weighted sums of independent random variables. *Ann. Prob.* **1** (1973), 810-824.
- [Cra1] H. Cramér, *Random variables and probability distributions*. Cambridge Tracts in Math. **36** (1937), Cambridge Univ. Press (2nd ed. 1962, 3rd ed. 1970).
- [Cra2] H. Cramér, On harmonic analysis in certain function spaces. *Ark. Mat. Astr. Fys.* **28B** (1942), 1-7 (reprinted in *Works II*, 941-947).
- [Cra3] H. Cramér, *Mathematical methods of statistics*. Princeton Univ. Press, 1946.
- [Cra4] H. Cramér, Half a century with probability theory. Some personal recollections. *Ann. Prob.* **4** (1976), 509-546 (*Works II*, 1352-1386).
- [Cra5] H. Cramér, *Collected Works*, I, II, Springer, 1994.
- [DenD] Y. Déniel and Y. Derriennic, Sur la convergence presque sure, au sens de Cesàro d'ordre α , $0 < \alpha < 1$, des variables aléatoires indépendantes et identiquement distribuées. *Prob. Th. Rel. fields* **79** (1988), 629-636.
- [Dia] P. Diaconis, G. H. Hardy and probability???. *Bull. London Math. Soc.* **34** (2002), 385-402.
- [DiaS] P. Diaconis and C. Stein, Tauberian theorems for coin-tossing. *Ann. Prob.* **6** (1978), 483-490.
- [Dur] P. L. Duren, *Theory of H_p spaces*. Academic Press, 1970.
- [Durr] R. Durrett, *Brownian motion and martingales in analysis*. Wadsworth, Belmont CA, 1984.
- [Ell] P. D. T. A. Elliott, *Probabilistic number theory I, II*, Springer, 1979, 1980.
- [Gai] D. Gaier, Der allgemeine Lückenumkehrsatz für das Borel-Verfahren. *Math. Z.* **88** (1965), 410-417.
- [Gar] J. B. Garnett, *Bounded analytic functions*. Academic Press, 1981.
- [Har1] G. H. Hardy, Mendelian proportions in a mixed population, *Science* **28** (1908), 49-50 (*Works VII*, 2(c), 477-478).
- [Har2] G. H. Hardy, *A mathematician's apology*, Cambridge Univ. Press, 1940 (repr. 1979);
- [Har3] G. H. Hardy, *Divergent series*, Oxford Univ. Press, 1949.
- [HarLP] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge Univ. Press, 1934 (2nd ed. 1952).
- [HarRa] G. H. Hardy and S. Ramanujan, The normal order of prime factors of a number n . *Quart. J. Math.* **48** (1917), 76-92 [Hardy, *Collected Papers* Vol. II, 100-113, Oxford Univ. Press, 1967].

- [HarRi] G. H. Hardy and M. Riesz, *The general theory of Dirichlet's series*. Cambridge Tracts in Math. **18**, Cambridge Univ. Press, 1915.
- [HarW] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 6th ed. (rev. D. R. Heath-Brown and J. H. Silverman), Oxford Univ. Press, 2008.
- [Harp] A. J. Harper, Two new proofs of the Erdős–Kac theorem, with bound on the rate of convergence, by Stein's method for distributional approximations. *Math. Proc. Camb. Phil. Soc.* **147** (2009), 95–114.
- [Hil] A. Hildebrand, The prime number theorem via the large sieve. *Mathematika* **33** (1986), 23–30.
- [Hof] K. Hoffman, *Banach spaces of analytic functions*. Prentice-Hall, Englewood Cliffs NJ, 1962.
- [Hol] P. Holgate, Studies in the history of probability and statistics **XLV**. The late Philip Holgate's paper 'Independent functions: Probability and analysis in Poland between the Wars' (Introduction, Bingham, 159–160; text, Holgate, 161–173). *Biometrika* **84** (1997), 159–173.
- [Jan] S. Janson, *Gaussian Hilbert spaces*. Cambridge Tracts in Math. **129**, Cambridge Univ. Press, 1997.
- [Jef] H. Jeffreys, *Theory of probability*, Oxford Univ. Press, 1939 (2nd ed. 1960, 3rd ed. 1983).
- [Kac1] M. Kac, *Statistical independence in probability, analysis and number theory*. Carus Math. Monog. **12**, Math. Assoc. America, 1959.
- [Kac2] M. Kac, *Probability and related topics in physical sciences*. Interscience, New York NY, 1959.
- [Kal] O. Kallenberg, *Foundations of modern probability*, 2nd ed., Springer, 2002.
- [Kar] J. Karamata, *Sur les théorèmes inverses des procédés de sommabilité*. Actualités Scientifiques et Industrielles **450**, Hermann, Paris, 1937.
- [KasB] Y. Kasahara and N. H. Bingham, Verblunsky coefficients and Nehari sequences. *Trans. Amer. Math. Soc.* **366** (2014), 1363–1376.
- [Kin] J. F. C. Kingman, A fragment of autobiography. *Probability and Mathematical Genetics: Papers in Honour of Sir John Kingman* (ed. N. H. Bingham and C. M. Goldie), LMS LNS **378** (2010), 17–34, Cambridge Univ. Press.
- [Kol1] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Erg. Math. **3**, Springer, 1933 [*Foundations of the theory of probability*, Chelsea, New York, 1956].
- [Kol2] A. N. Kolmogorov, Stationary sequences in Hilbert space. *Bull. Moskov.*

- Gos. Univ. Mat.* **2** (1941), 1-40 (in Russian; reprinted, *Selected works of A. N. Kolmogorov, Vol. 2: Theory of probability and mathematical statistics*, Nauka, Moskva, 1986, 215-255).
- [Kor] J. Korevaar, *Tauberian theorems: A century of developments*. Grundle. math. Wiss. **329**, Springer, 2004.
- [Kub] J. Kubilius, *Probabilistic methods in the theory of numbers*. Transl. Math. Mono. 11, Amer. Math. Soc., 1964 (Russian, 1962).
- [Lai] T.-L. Lai, Summability methods for independent, identically distributed random variables. *Proc. Amer. Math. Soc.* **45**, 253-261.
- [Lan1] E. Landau, Sur quelques problèmes relatifs à la distribution des nombres premiers. *Bull. Soc. Math. France* **28** (1900), 25-38.
- [Lan2] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, 2nd ed., Chelsea, New York, 1953 (1st ed. 1909).
- [Lit1] J. E. Littlewood, *A mathematician's miscellany*, 1953 (2nd ed., *Littlewood's Miscellany*, ed. B. Bollobás, 1986), Cambridge Univ. Press.
- [Lit2] J. E. Littlewood, *Collected Papers, I, II*. Oxford Univ. Press, 1982.
- [Lor] G. G. Lorentz, Borel and Banach properties of methods of summation. *Duke Math. J.* **22** (1955), 129-141.
- [MarZ] J. Marcinkiewicz and A. Zygmund, A theorem of Lusin. *Duke Math. J.* **4** (1938), 473-485 (A. Zygmund (ed.), *Józef Marcinkiewicz, Collected Papers*, PWN, Warszawa, 1964, 428-443).
- [MeyK] W. Meyer-König, Untersuchungen über einige verwandte Limitierungsverfahren. *Math. Z.* **52** (1949), 257-304.
- [MeyK-Z1] W. Meyer-König and K. Zeller, Lückenumkehrsätze und Lückenperfektheit. *Math. Z.* **66** (1956), 203-224.
- [MeyK-Z2] W. Meyer-König and K. Zeller, On Borel's method of summability. *Proc. Amer. Math. Soc.* **11** (1960), 307-314.
- [MeyK-Z3] W. Meyer-König and K. Zeller, Tauber-Sätze und M -Perfektheit. *Math. Z.* **177** (1981), 257-266.
- [MonV] H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory I. Classical theory*. Cambridge Stud. Adv. Math. **97**, Cambridge Univ. Press, 2007.
- [Nik1] N. K. Nikolskii, *Treatise on the shift operator: Spectral function theory*. Grundle. math. Wiss. **273**, Springer, 1986.
- [Nik2] N. K. Nikolskii, *Operators, functions and systems: an easy reading. Volume 1: Hardy, Hankel and Toeplitz; Volume 2: Model operators and systems*. Math. Surveys and Monographs **92, 93**, Amer. Math. Soc., 2002.
- [PalW] R. E. A. C. Paley and N. Wiener, *Fourier transforms in the complex*

- domain*. AMS Colloquium Publications XIX, American Math. Soc., 1934.
- [RenT] A. Rényi and P. Turán, On a theorem of Erdős and Kac. *Acta Arith.* **4** (1958), 71-84.
- [RosR] M. Rosenblum and J. Rovnyak, *Hardy classes and operator theory*. Oxford Univ. Press, 1985; Dover, 1997.
- [Sar] D. Sarason, Holomorphic spaces: a brief and selective survey. In [AxM-cCS], p.1-34.
- [Sch] A. Schmaal, A. J. Stam and T. de Vries, Tauberian theorems for limitation methods admitting a central limit theorem, *Math. Z.* **150** (1976), 75-82.
- [Sim1] B. Simon, *Orthogonal polynomials on the unit circle. Part 1: Classical theory; Part 2: Spectral theory*. AMS Colloquium Publications 54.1, 54.2, Amer. Math. Soc., Providence RI, 2005.
- [Sim2] B. Simon, *Szegő's theorem and its descendants: Spectral theory for L^2 perturbations of orthogonal polynomials*. Princeton Univ. Press, 2011.
- [Smi] F. Smithies, *Integral equations*, Cambridge Tracts **49**, Cambridge Univ. Press, 1965.
- [Ste] E. M. Stein, *Topics in harmonic analysis related to the Littlewood-Paley theory*. Ann. Math. Studies **63**, Princeton Univ. Press, 1970.
- [Ten] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, 2nd ed. Cambridge Studies Adv. Math. 46, Cambridge Univ. Press, 1995 (3rd French ed., Belin, 2008).
- [Tit] E. C. Titchmarsh, *The theory of functions*, Oxford Univ. Press, 1932 (2nd ed. 1939).
- [Tur] P. Turán, On a theorem of Hardy and Ramanujan, *J. London Math. Soc.* **9** (1934), 274-276.
- [Var] N. T. Varopoulos, Aspects of probabilistic Littlewood-Paley theory. *J. Functional Analysis* **38** (1980), 25-60.
- [Wei] W. Weinberg, Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* **64** (1908), 368-382.
- [Wie1] N. Wiener, *The Fourier integral and certain of its applications*. Cambridge Univ. Press, 1933 (reprinted, Cambridge Mathematical Library, 1988).
- [Wie2] N. Wiener, *Ex-prodigy: My childhood and youth*. MIT Press, Cambridge MA, 1953.
- [Wie3] N. Wiener, *I am a mathematician: The later life of a prodigy*. MIT Press, Cambridge MA, 1956.
- [Wil] D. Williams, *Weighing the odds: A course in probability and statistics*,

Cambridge Univ. Press, 2001.

Mathematics Department, Imperial College, London SW7 2AZ. n.bingham@ic.ac.uk