Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 5, n°2; Décembre/December 2009

www.jehps.net

England and Probability in the Inter-War Years

John Aldrich[1]

Résumé

L'Angleterre et les statistiques pendant l'entre-deux-guerres est un sujet qui a été abondamment traité mais l'Angleterre et les probabilités dans la même période semble avoir été peu considéré. Cela peut sembler paradoxal. Cet article considère la scène probabiliste anglaise et examine les travaux de Turing, Paley et Linfoot, en phase avec la manière de l'Europe continentale. Nous examinons également l'attitude des statisticiens aux travaux continentaux exprimés dans les réactions à la thèse de Turing sur le théorème central de la limite et au traité de Harald Cramér *Random Variables and Probability Distributions*. Quelques documents originaux sont reproduits et fournissent des éléments pour la discussion.

Abstract

England and statistical theory in the inter-war period is a subject that abounds with material but England and probability theory may seem empty. This may seem a paradoxical situation. This paper considers the English probability scene and examines work in the continental manner by Turing, Paley and Linfoot. It also considers the response of statisticians to continental work as manifested in the reactions to Turing's fellowship dissertation on the central limit theorem and Harald Cramér's tract *Random Variables and Probability Distributions*. Some documents from the time are reproduced and they provide the focus for the discussion.

Acknowledgement : I am grateful to two anonymous referees for their comments.

[1] Economics Division, School of Social Sciences. University of Southampton. Southampton, SO17 1BJ . United Kingdom. john.aldrich@soton.ac.uk

1 Introduction

"The theory of probability has been cultivated in England less for what it is than for what it does." These are the words of A. C. Aitken, a New Zealander working in Edinburgh, and his "England" was Britain, or even the British Empire, and was contrasted with "the Continent" or mainland Europe and "America" the United States. Aitken was reviewing Harald Cramér's *Random Variables and Probability Distributions* (1937), the first book in English to present modern continental ideas on the theory of probability. To exemplify what probability *does*, Aitken (1938, p. 193) chose statistical theory and its different standards, "The research of the present century on the theory of estimation and the distribution of statistical coefficients has taken the fundamentals for granted and has sometimes been avowedly non-rigorous." English applied mathematics offered other examples of probability *doing* but Aitken's choice of statistical theory was a good one for the achievement here was substantial and substantially English. Aitken's contrast recurs in the modern historical literature where England is absent from the biggest probability stories, the axiomatisation of probability (see von Plato (1994) and Shafer & Vovk (2006)), the central limit theorem (Le Cam (1986)) and martingale theory (Shafer & Mazliak (2009)) and yet central to the development of modern statistical theory (Hald (1998)).

A search of the English–it is convenient to follow Aitken's usage–journals shows just how seldom the theory of probability was cultivated for "what it is." In the years 1919-39 the journals of the London Mathematical Society, the Edinburgh Mathematical Society, the Cambridge Philosophical Society, the Royal Societies of London and of Edinburgh and the *Quarterly Journal of Mathematics*, the *Philosophical Magazine* and *Messenger of Mathematics* published six papers on continental themes by four English authors. Of course the English could also publish in foreign journals and foreign mathematicians could publish in England but that alters the picture very little. The statistical journals, *Biometrika* and the *Journal of the Royal Statistical Society*, published many articles on probability *doing* but none on what probability *is*.

Sections 2 to 4 and Appendices I to III below present impressions of English probability in the inter-war period. Section 2 examines the small body of work in the continental manner and its place in the spectrum of English probability. Sections 3 and 4 consider two episodes from the 1930s and what they show about English attitudes. Alan Turing wrote his dissertation on the central limit theorem in such splendid isolation that he did not know that there was a modern literature and he repeated Lindeberg's work of a decade earlier. The dissertation is now well-known–thanks to Zabell (1995)–but the reports are also instructive for they were written by the contrasting figures of Abram Besicovitch, a continental analyst and specialist on almost periodic functions, and Ronald Fisher, the great figure of English statistical theory and ultimate authority for the "non-rigorous" approach that Aitken noted. While the Turing episode of Section 3 was a private affair, the episode described in Section 4 was public and a triumph for the continental approach–the publication of Cramér's book to acclaim from the probability-doers, including Aitken. Section 5 is a brief epilogue. Much of the original material is easily accessible on the web but three appendices reproduce Besicovitch's report on Turing, Aitken's review of Cramér and P. J. Daniell's review of an earlier continental contribution from Tornier.

2 People, places and probabilities

In one way or another Cambridge University is central to the story that follows, just as it was central to the life of English mathematics. There were excellent mathematicians elsewhere but there were more in Cambridge and the best students went through Cambridge. In the late nineteenth century Cambridge colleges introduced the fellowship dissertation as a guide to the research potential of would-be fellows; see Aldrich (2005/9) for notes on the system. In November 1934 Alan Turing an aspiring pure mathematician at King's College submitted a dissertation, *On the Gaussian Error Function*, which aimed "to give a rigorous demonstration" of the "limit theorem of the theory of probability." Turing was dissatisfied with the proof that Arthur Eddington (1882-1944), the professor of astronomy, had given in his lectures on the combination of observations. Such lectures had long been part of the Cambridge training in applied mathematics: Ronald Fisher (1890-1962) and Harold Jeffreys (1891-1989) had attended similar ones twenty years before as had Karl Pearson (1857-1936) thirty years before them. Pearson, Fisher and Jeffreys were all trained as applied mathematicians: Pearson worked on elasticity and Jeffreys on geophysics before taking up statistics, while Fisher applied mathematics to biology instead of to physics.

Turing was a product of a new Cambridge. From the beginning of the century there had been a movement to raise the level of pure mathematics, to make it more important relative to applied mathematics and to bring its quality up to that of continental work and by Turing's time this had been successful, at least in analysis. The movement was led by G. H. Hardy (1877-1947) but Hardy's influence extended beyond Cambridge, pervading English mathematics; see the account by Rice and Wilson (2003). Hardy was a great internationalist who worked with foreign mathematicians, visiting them, encouraging them to visit him and settling some, including Besicovitch, in England. There were some major probability figures in Hardy's network: George Pólya (1887-1985) of Zürich, Norbert Wiener (1894-1964) of MIT and Harald Cramér (1893-1985) of Stockholm–appropriately Wiener and Cramér first met when visiting Hardy in 1920. Pólya's substantial work in probability included a paper (1920) on "der zentrale Grenzwertsatz" (a term he coined) but his collaboration with Hardy centred on the Hardy, Littlewood and Pólya volume, *Inequalities* (1933). Wiener began working on Brownian motion in 1919 and he developed links with Paul Lévy, one manifestation of this was a publication (1924) in French on denumerable probabilities; see Aldrich (2007, pp. 18-25) and Bru and Eid (2009, pp. 32-3). The intellectual link between Hardy and Wiener was based on classical analysis and not on probability. Cramér was a number theorist and an actuary and these interests converged on probability. Hardy had no interest in applied mathematics and his link to Cramér was based on number theory. Diaconis (2002, p. 385) writes that Hardy the analytical number theorist had a "genuine antipathy" towards probability but perhaps "indifference" would be more apt. Unlike Borel or Khinchin, Hardy did not use probabilistic arguments but he did not criticise their use of it and he encouraged Cramér to write a book on probability, as we will see in Section 4 below.

Although it was the more ancient foundation, Oxford University was less of a force in English mathematics than Cambridge. However, from 1919 to 1931 Hardy as a professor there and his presence may be part of the explanation of why the first English papers on continental probability came out of Oxford. In 1928 Edward Hubert Linfoot (1905-1982) published two papers on the law of large numbers, extending the result of Khinchin (1925). The technique of Khinchin's paper was accessible to English mathematicians insofar as it derived from Hardy and Littlewood (1914) but there may be a clue to what was behind Linfoot's papers in his (1928a, p. 348 and -28b p. 418) thanks to A. S. Besicovitch for "much valuable assistance and advice." In 1925-6 the Russian Abram Samoilovitch Besicovitch (1891-1970) had been in Oxford–at Hardy's invitation–and Linfoot was one of his first English pupils. There is a nice story in Bell's (1984, p. 52) obituary of Linfoot: Besicovitch had hoped to improve his English but lamented, "I am not learning enough English—he [Linfoot] understands before I explain." Besicovitch's background was in probability: he was a student of Markov, his first paper was on limit theorems in probability (1915) and he had recently (1923-4) written an account of Markov's work on probability. Besicovitch wrote no probability papers in England–from 1927 he was settled in Cambridge–and his main interest was in almost periodic functions on which he wrote a standard work (1932); Besicovitch became an important figure in English mathematics and there are biographies by Burkill (1971) and Taylor (1975). Linfoot moved on to number theory and then out of pure mathematics; Bell (p. 55) writes, "all [Linfoot's papers on pure mathematics] exhibit the mastery of the techniques of classical analysis and penchant for detailed calculation that he was later to bring to his work on optics."

In the early thirties one Cambridge mathematician was doing "international" probability, Raymond Paley (1907-1933); his brief career is recalled by Hardy (1934). Paley was a pupil of Hardy's principal collaborator J. E. Littlewood and, as well as working with Littlewood, he worked with two visitors to Cambridge, Wiener and Antoni Zygmund (1900-1992). Hardy's (1934, p. 78) described the region where Paley and Zygmund (1930-1932) worked as that where "Fourier series and probability come together." Later Paley and Zygmund collaborated with Wiener on a paper–Paley, Wiener and Zygmund (1933)–that brought Wiener's work on Brownian motion to the party. Paley went on to work with Wiener before being killed in an accident in April 1933; Wiener wrote up their joint research as Paley and Wiener (1934). After few years probability was back in Cambridge with Littlewood and Offord (1938) writing on random algebraic equations; their starting point was a paper by Bloch and Pólya (1932).

While probability as pure mathematics was a foreign implant in Cambridge, there was a local probability tradition in the form of a philosophical commentary on the theory of probability, a different take on what probability *is*. Among the luminaries of the tradition were J. Venn, W. E. Johnson, J. M. Keynes and F. P. Ramsey. The great work of the 1920s was the *Treatise on Probability* (1921) of J. M. Keynes (1883-1946) which had its origins in a King's fellowship dissertation of 1907. The philosophical tradition survived Keynes's departure from probability after publishing the *Treatise* and the deaths of Ramsey in 1930 and Johnson in 1931 and it was taken in a new direction by the physicist Harold Jeffreys whose *Theory of Probability* (1939) refounded the statistical theories of Pearson and Fisher on the basis of the Johnson-Keynes conception of probability as degree of rational belief; for this see Aldrich (2005). Jeffreys's conception of probability was rejected by most of those doing statistical theory in England, including Fisher and Aitken: Fisher made his position clear in numerous publications–see Aldrich (2005 and -8)–while Aitken's is evident from his letter to Fisher of February 22nd 1936. Whatever the merits of the thesis, Keynes's *Treatise* provided a wonderful survey of the literature–thus it was the first English work to notice and praise the Russian school–but it was the pre-1914 literature.

Ever since J. C. Maxwell (1831-1879) began writing on the theory of gases in the 1850s English applied mathematics has offered numerous examples of probability *doing*–e.g. from the inter-war period we see Fowler on statistical mechanics, Chapman on Brownian motion, Fisher on branching processes and diffusion, McKendrick and Cormack on epidemics, Rayleigh and Burnside on random flights, etc.–but statistical theory was becoming a great thing and the leading figure in what Aitken called "the theory of estimation and the distribution of statistical coefficients" was R. A. Fisher. When Aitken alluded to research that was "avowedly non-rigorous" he probably had in mind a passage in Fisher's "On the mathematical foundations of theoretical statistics" (1922, p. 323) :

> I should gladly have withheld publication until a rigorously complete proof could have been formulated; but the number and variety of the new results which the method discloses press for publication, and at the same time I am not insensible of the advantage which accrues to Applied Mathematics from the co-operation of the Pure Mathematician, and this co-operation is not infrequently called forth by the very imperfections of writers on Applied Mathematics.

In the event Fisher never showed much appetite for cooperating with pure mathematicians. Fisher's view was that the study of probability is subordinate to that of statistics and he says so very clearly in a letter to Aitken of January 23rd 1936. Other statisticians also found it natural to consider probability from the standpoint of statistical theory.

In the 1930s English statistical theory was beginning to travel, with contributions from, amongst others, Hotelling and Snedecor in America and Darmois in France, but its home was still in England where there were four important centres: University College London, Rothamsted Experimental Station, Edinburgh

University and Cambridge University with University College and Rothamsted far in the lead. Although Cambridge University was slow to adopt modern statistical theory, Cambridge men–Karl Pearson, Edmund Whittaker and Ronald Fisher–had put the other places on the statistical map. University College was the most established centre and its importance went back to 1893 when Karl Pearson, the professor of applied mathematics, first collaborated with Raphael Weldon, the professor of zoology on a subject they called "biometry." There was a second surge in the "English statistical school" associated with R. A. Fisher who went to work at Rothamsted in 1919. Pearson and Fisher were the most important influences on modern statistics and there are large literatures on them and their works; see the guides by Aldrich (2001/9 and 2003/9). When Pearson retired from University College in 1933 Fisher replaced him as Galton Professor of Eugenics and Pearson's son, Egon Pearson, replaced him as head of the Department of Applied Statistics. Soon Neyman joined the younger Pearson's department from Poland. In the mid-1930s University College London was the centre of the statistical universe.

E. T. Whittaker (1873-1956) and his Edinburgh operation are less well known; Whittaker's life and work are recalled by Temple (1956) and Aitken (1958). Whittaker had very broad interests in pure and applied mathematics and "was conversant with a range of mathematics which no other Briton has encompassed" according to Temple (2004). Whittaker had been a Cambridge fellow before moving to the University of Edinburgh. In 1914 he set up a Mathematical Laboratory and his lectures on the mathematics of treating numerical data were published ten years later as the *Calculus of Observations*. The chapter in the *Calculus* on "normal frequency distributions" covered asymptotic normality and the series developments of Gram, Charlier and Brus. The main technique was the Fourier transform, a topic Whittaker treated in his book on complex analysis, *Course of Modern Analysis*, a pioneering work in the English context. Whittaker was a great reader: he knew more of the old French literature than Lévy but he knew nothing of Lévy for his reading also seems to have stopped in 1914. Alexander Aitken (1895-1967) was Whittaker's prize student, colleague and successor in the Edinburgh chair; Aitken's life and work is recounted by Whittaker & Bartlett (1968). Aitken's main field was algebra but he inherited his master's interests in numerical methods and actuarial mathematics. Aitken's first statistical publication was on generating functions (1931) but he is best remembered today for his (1935) matrix formulation of least squares and "Aitken's generalised least squares"–see Farebrother (1997). In 1931 Aitken and Fisher began corresponding; at first they discussed interpolation and fitting polynomials but later their exchanges extended to statistical theory and probability theory, especially when Aitken started teaching those subjects. By December 1935 Aitken was asking Fisher for copies of his papers.

Cambridge received the new Fisherian statistics from John Wishart (1898-1956), a man whose career joined all four centres: Wishart had studied with Whittaker and worked for Karl Pearson and for Fisher. When Wishart gave his first Cambridge lectures on mathematical statistics in 1932 the audience included M. S. Bartlett (1910-2002) who became Wishart's first research student.

Bartlett also heard Eddington lecture on the combination of observations and the theoretical physicist Ralph Fowler lecture on statistical mechanics where, Bartlett (1982, p. 43) recalled, probabilities were introduced "most discreetly" as "weights." Just as Turing was one of the best undergraduate pure mathematicians, so Maurice Bartlett was one of the best of the applied. After Cambridge Bartlett went to Egon Pearson's department at University College as an assistant lecturer but after a year he left for the Imperial Chemical Industries research establishment, returning to Cambridge in 1938. See Bartlett (1982), Olkin (1989) and Whittle (2004) for Bartlett's life and work.

In this period English universities did not put on specialised courses in probability theory: combinatorial probability was taught as part of school algebra and continuous distributions were taught in university courses on statistical mechanics and the theory of errors. There was a book resembling a university textbook, Burnside's posthumous *Theory of Probability* (1928), but this did not come of his teaching. William Burnside (1852-1927) the noted group theorist only developed an interest in probability and projected a book after he retired from teaching naval cadets ballistics; Burnside's work in probability (and statistics) is discussed by Aldrich (2009). Burnside knew the works of Bertrand (1888) and Poincaré (1912) and parts of his *Theory* read like a critical commentary on the latter. He did not refer to any recent French works or to any German literature and so his *Theory*, like Keynes's *Treatise* and Whittaker's *Calculus*, reflected a past age. The English journals followed the foreign literature to a certain degree. The main periodical for reviews of mathematics books was the *Mathematical Gazette* and it published reviews of books by Borel and Lévy, among others: the reviews by J. Marshall appear typical: the reviews–see e.g. Marshall (1926a and -26b)–provide adequate summaries of the books but they give no sense of perspective and the reviewer is an observer rather than a participant–I have found no original works by Marshall. The statisticians more or less ignored probability: Keynes's *Treatise* was the *only* probability book reviewed in the *Journal of the Royal Statistical Society* between 1912, when Keynes reviewed Poincaré, Bachelier and Markov, and 1938, when Bartlett reviewed Cramér. Attitudes in statistics had actually begun to change a little earlier for the third in the surveys of "Recent advances in mathematical statistics" (1935) by Oscar Irwin (1898-1982) contains a section on "Recent developments in probability theory.–probability and the theory of measure."

3 Turing's dissertation

The first of our episodes from the 1930s centres on Alan Turing (1912-1954) and his fellowship dissertation, "On the Gaussian Error Function;" the dissertation is available online at Hodge (2002-3), the Turing Digital Archive. Supervising the election was the Provost (head) of King's College, a classicist, John Sheppard, and the referees chosen were Fisher and Besicovitch. Turing's candidature had the support of Keynes, by now the world-famous economist and the most powerful figure in the college. Keynes was a great talent-spotter and there do not seem to be any intellectual links between his probability and Turing's.

In his biography of Turing Hodge (1985, pp. 87-8) describes how Turing came to write the dissertation and Zabell (1995) reviews the mathematical argument and so I will offer only a few remarks to place the work in the general probability landscape. The wider setting of the problem is indicated in Besicovitch's report (Appendix I below) and there are thorough treatments of the older literature in Hald (1996, ch. 17) and of the modern in Le Cam (1986).

Turing started from Eddington's proof of the central limit theorem. Eddington had taught the proof for many years and it is reproduced in Brunt's textbook, *The Combination of Observations* (1917, pp. 15-7). There it is called a "generalised form of Hagen's proof" referring to Hagen's (1837) derivation of the error function which was based on the notion of "elementary" errors. Brunt (p. 15) defines the probability that the resultant error due to n elementary errors will lie between x and $x + dx$ as

$$f(n, x).dx.$$

The analysis begins with the relationship between $f(n+1, x)$ and $f(n, x)$

$$f(n+1, x) = \int_{-\infty}^{\infty} f(n, x - \epsilon).g(\epsilon)d\epsilon$$

where g is the density of an elementary error. Both sides of the equation are expanded using Taylor's theorem, negligible terms are neglected and n is replaced by a continuous variable t until the equation

$$\frac{\partial f}{\partial t} = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}$$

is obtained. The solution f is the Gaussian error function. Turing (1935, p. i) remarked that the proof had been given by "Czuber, Morgan Crofton and others." There are no specifics but the references were most likely Czuber's 1891 book on the theory of errors and Crofton's *Encyclopedia Britannica* article of 1885 to which Czuber (1891, p. 97) refers; Stigler (1979, p. 294) found a similar idea in Edgeworth (1883). The proof also appears in one of Czuber's later books, one that reviewed by Keynes in 1911.

In the dissertation Turing (1935, pp. 41-3) presented his own version of the argument and concluded by commenting:

> This method has a very strong appeal to the intuition. Hence presumably its popularity. It would seem moreover that it would be made rigorous by long but not difficult analysis. I shall follow this up to some extent; I shall show that difficulties do arise and why they arise. I cannot of course show that it is impossible to build up a proof on these lines but I shall at least show that the "intuitive appeal" is a fraud.

Appendix B of the dissertation considers a second "popular" proof using the characteristic function and which Turing (1935, pp. 47-50) associates with Lyapunov. One wonders whether Turing looked at Whittaker and Robinson, the text recommended for the course, where the same method is used but with less attention to rigour. Reading seems to have played little part in what Hodge calls Turing's "self-contained" way of working. Turing's proof appears in §§9-13 (pp. 18-33) of the dissertation. Hodge (p. 88) quotes from a letter Turing wrote in April 1934, "I am sending some research I did last year to Czuber in Vienna, not having found anyone in Cambridge who is interested in it." Emanuel Czuber (1851-1925) was dead, as Turing half-expected, but there was a bigger setback, as he reports in the Preface to the dissertation: "when considering publishing it I was informed that an almost identical proof had been given by Lindeberg." Turing does not indicate the source of the information–most probably Besicovitch–and, in fact, he makes no personal acknowledgments of any kind. Candidates were required to acknowledge any debts and Turing was declaring, in effect, that he had none–beyond the "elementary matter forming part of one's general mathematical education." Of course, there is a reference to Lindeberg (1922) and to the Liapunov papers that Lindeberg cites; Liapunov's papers from 1901-2 were in Keynes's (1921) bibliography though they were not discussed in the text. Turing's (p. 33) only other references were for proofs of the completeness of the Hermite functions–Turing had his own–and these were to Wiener's *Fourier Integral*, von Neumann's *Quantenmechanik* and Courant and Hilbert's *Methoden der Mathematischen Physik*. None of these were probability works but they underline the international character of Cambridge mathematics.

There is a very useful account of the dissertation by Zabell (1995, pp. 486-489) goes through it section by section comparing Turing's work did what Lindeberg did and what subsequent writers have done. Turing's work is strikingly different from that of Eddington: about all they have in common is that they start from the convolution although Turing works with distribution functions and not with densities. Turing found it necessary to define all the concepts he needed, including that of the distribution function. Although Turing said that "an almost identical proof" had been produced by Lindeberg, that was quite accurate. Besicovitch (see Appendix I below) said only that "idea" was the same and Zabell (p. 488) elaborates the point. Zabell also emphasises that Turing stopped work on the dissertation and on the subject when he learnt of Lindeberg's work and that the dissertation contained some material that was not yet in the literature.

Besicovitch's report is very efficient. It recounts the history of the problem from Laplace through Markov and Liapunov to Lindeberg and then compares Turing's proof with Lindeberg's, explaining why Besicovitch is "completely confident" that Turing's effort was done in genuine ignorance of Lindeberg's work. Besicovitch gave his estimate of the magnitude of Turing's achievement, "If the paper were published fifteen years ago it would be an important event in the mathematical literature of that year."

The correspondence between Sheppard, who was in charge of the process, and Fisher, the other referee, is in the Fisher archive at the University of Adelaide and the two letters in which Fisher gave his impressions of the dissertation are available online. Fisher's opinion of the value of the research was very different from Besicovitch's. In his first letter he told Sheppard,

> The subject is one which I have always thought decidedly unattractive and has been worked over, from various points of view, by Continental and especially Scandinavian writers to the point of making it positively repellent.

Fisher probably had in mind the older Scandinavians, Charlier, Gram and Thiele: they were mentioned by Whittaker and Robinson and a summary of their work was available in Arne Fisher's book *Mathematical Theory of Probabilities* (1922). Although Ronald Fisher found the subject "unattractive," it was linked to his theory of estimation, just as it was linked to the old theory of errors. In his "Mathematical foundations" Fisher (1922, pp. 323-30) established the large-sample normality of maximum likelihood estimates without any reference to the conditions under which the result would hold.

Turing submitted the dissertation early in November 1934; naturally it reflected the discovery of Lindeberg's earlier work. On the 14th Sheppard invited Fisher to serve as referee and Fisher sent his first report on the 27th. It is quite thin and Keynes thought something more substantial was needed; on December 14th Sheppard was writing to Fisher:

> Mr. Keynes has told me of his conversation with you about Turing's dissertation, and your very kind and ready assent to his suggestion, that if I asked you, you would send me for the guidance of the Electors a somewhat more detailed account of the personal qualities shown by Turing, of the nature and difficulty of his task and of the way in which he tackled it...

Fisher obliged with a second letter, although it was not much more detailed.

In this second letter Fisher mentions the appearance of the characteristic function in the dissertation, wishing that it had been more thoroughly discussed. This was a mathematical technique that Fisher knew. In his early work on distribution theory he had used his trademark geometric method but more recently he had been using transforms or generating functions. In his review of Whittaker and Robinson, Fisher (in Fisher and Jackson (1924, p. 156)) praised the "elegant development of the theory of the distribution of a linear function of deviations." A year later he (1925c) was noting Romanovsky's use of "la fonction génératrice des moments" to find the distribution of the correlation coefficient (which Fisher had found a decade earlier using a geometric argument) and by 1929 he was using the moment-generating function in his own work.

Sheppard's main concern was with "the intellectual promise and qualities" shown in the dissertation and in both letters Fisher reported favourably: his final word was the concluding sentence of the second letter:

> in reading through his paper I form a very high opinion of his taste, virtuosity would not be too strong a term, in the art of framing conclusive mathematical demonstrations.

The reference in this letter to Turing having "already worked under Kramér [sic]" is puzzling and may have been a misunderstanding of something Keynes had said. Of course, Besicovitch was also impressed with Turing:

> we see a display of very exceptional abilities at the very start of his research work, which makes me to recommend him as a very strong candidate for Fellowship.

A committee of fellows assessed the dissertation. The mathematician on the committee was the analyst Albert Edward Ingham (1900-1967). Ingham had a brush with statistical theory when he (1933) inverted the characteristic function of the Wishart distribution for Wishart and Bartlett (1933)–the original derivation in Wishart (1928) had used a Fisher-inspired geometric argument. The career of another King's fellow further illustrates the smallness of the mathematical world in those days: Philip Hall, the group-theorist, had worked briefly for Karl Pearson, himself a fellow of King's in the 1880s. In March 1935 the competition for fellowships was stiff–there were 25 candidates–but Keynes's motion that Turing be elected was carried unanimously. I owe this information to Patricia McGuire, the King's librarian, who also identified Besicovitch as a referee.

4 Cramér's tract

Linfoot, Paley and Turing were young English mathematicians who worked in continental probability, although Turing only realised this afterwards. I have described Fisher's response but other established mathematicians knew about continental work and we now turn to their reaction to it. I will use book reviews although these were only part of the reaction for some books, including von Mises (1929 and -31) and Kolmogorov's *Grundbegriffe*, were read and cited but never reviewed in the English journals. Three reviews from the 1930s will be discussed, two from the *Mathematical Gazette* (reproduced in Appendices 2 and 3 below) and one from the *Journal of the Royal Statistical Society.*

Percy Daniell (1889-1946) of the University of Sheffield made a very interesting observer for he was established in several fields and could have responded in any number of ways to Erhard Tornier's (1936) book on probability and general integration theory. Daniell was the English analyst whose work was most closely related to continental interests in probability but he was also an applied mathematician with an interest in the foundations of mathematics; for Daniell generally see Aldrich (2007) and for further aspects of the interplay between analysis and probability theory see Bru and Eid (2009, passim). Daniell could not recommend Tornier's book to a British audience: he disliked both the analysis and probability components and did not care for the package. Daniell (1937, p. 67) had no sympathy for the book's approach to probability:

> Professor Tornier has apparently never heard of J. M. Keynes or other critics of fundamental notions. He writes glibly of obtaining probability in some cases "näherungsweise", though he does not say how this can be done without a circular use of Bayes' formula or else by the often disproved limit theories, such as Venn's. Professor Tornier merely assumes that the probability is a number satisfying the postulates for a general mass.

J. L. Doob also reviewed Tornier but from the perspective of one familiar with the recent continental probability literature: he (1937, p. 317) began by observing, "In the last few years the theory of probability has been more and more influenced by the modern theories of measure." There are further comments on Daniell's review of Tornier in Aldrich (2007, pp. 29-30).

Tornier's book passed almost without notice and sank without trace but Cramér's *Random Variables and Probability Distributions* made an impression, being favourably reviewed in the *Gazette* by Aitken, in *JRSS* by Bartlett and it soon became a standard reference. This was not so for the works from the 1920s that were reviewed in the *Gazette.* Cramér's book was in English which probably helped although professional mathematicians of the time were expected to know German and French. The origins of the book were actually in England as Cramér (1976, p. 516) recalled:

> 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.

A book eventually materialised as one of the *Tracts on Mathematics and Mathematical Physics*, the well-regarded series Hardy that edited for Cambridge University Press and which was founded in 1905 by Whittaker. Cramér (1976, p. 529) adds that when he visited Hardy in 1938 the latter "expressed his satisfaction with my tract which was written on his initiative."

Cramér's *Random Variables* is a short book of 120 pages in three parts. "Principles" introduces the Kolmogorov axioms, confining the discussion to Euclidean space. The core of the book is "Distributions in R_1" while the third part, "Distributions in R_k," is a brief epilogue. (There is a detailed description in Aitken's review, Appendix 3 below.) As Aitken and Bartlett emphasise, Cramér offers probability as pure mathematics, not probability as physics or as statistics. However they liked the book and were ready to accept it on its own terms–something Daniell would not grant Tornier. Again, while Aitken and Bartlett recognised Fisher as a great statistician, they did not share his aversion to mathematical work on the central limit theorem. By the late 1930s Fisher was no longer the undisputed leader of English statistical theory. His rival, Jerzy Neyman (1894-1981) at University College from 1934 to 1938, had a distinctly continental outlook. Neyman had contributed to the *Comptes Rendus*–see his (1926)–and he kept abreast of developments abroad: his 'Outline" (1937) has references to Kolmogorov (1933) and to Polish probability work and he contributed to the translation of von Mises (1928). Neyman also pointed his colleagues in the continental direction, so that when David (1938) was developing a finite population central limit theorem she looked to Lévy and not to the domestic products of Bowley or Isserlis.

In his *JRSS* review of Cramér Bartlett (1938, p. 207) emphasised the links to "modern analysis" and argued that, "To some readers the beginning may seem too sophisticated but from the point of view of mathematical analysis it is a fairly natural and logical one." Bartlett did not know the continental work on probability: five years earlier he (1933) tried to negotiate a truce between the probabilities of Fisher and Jeffreys and, like them, he made no reference to continental work. Aitken, however, had been following the continental literature for some time. In January 1936 he told Fisher:

> Several interesting books in the last three years have appeared founding probability very prettily on the theory of measure and integration of sets of points; but refraining from mentioning what these sets are in such cases as the tossing of an inhomogeneous, irregular and biassed die, etc. No subject is so perennially interesting, or uselessly controversial.

Presumably one of those books was Kolmogorov's *Grundbegriffe*.

Unlike Daniell, neither Aitken nor Bartlett were exercised by the issue of interpretation. Bartlett (1938, p. 207) observed quite neutrally:

> Avoiding the attempt by von Mises to define probability in terms of an unlimited sequence of values, he merely associates a number with an event and considers reasonable rules for operating with such numbers.

In one respect Cramér's tract was much more congenial to English taste than the work of Tornier and–one assumes–of Kolmogorov. The tract's second, main part could be regarded as a rigorous treatment and extension of the material in Whittaker and Robinson. Also, as noted above, Cramér's chief tool, the characteristic function, was used by English mathematical statisticians to obtain exact distributions; Bartlett's first paper–Wishart & Bartlett (1932)– used the characteristic function to re-derive distributions originally obtained by geometric methods.

Cramér even made some contact with English statistical thought though the works he referred to were either marginal to that tradition or incidental to the book's main purpose. Edgeworth (1905) was one of the standard references on asymptotic expansions but it was outside the English mainstream and, in any case, Cramér (1938, p. 86) was not much impressed, only commenting, "The formal definition of this series was given by Edgeworth." Cramér (p. 48) referred to Student (1908) to illustrate his theorem on the distribution of a ratio of independent random variables. While Student's paper was central to the Fisherian stream of English mathematical statistics, Cramér seems to have come to it through Rider's (1930) survey of small sample work in the *Annals* *of Mathematics*. Cramér had not yet got to the heart of English statistics but Bartlett (1938, p. 208) appreciated that, if he did, he could do useful work there:

> Statisticians may rather regret that theorems relating to the convergence of "likelihood" estimates could not have been included, as they might conveniently be associated with the Central Limit Theorem.

That association would be one of the highlights of Cramér's *Mathematical Methods of Statistics* (1946).

Cramér's tract was also noted in the science weekly *Nature*. The joint review "Recent developments in probability theory" by "A. v. Z." also described three publications from Gauthier-Villars, Fréchet (1937), Lévy (1937) and Bachelier (1937) The feature of the Fréchet and Lévy works that most struck the reviewer was their insistence on the relationship between probability theory and "general analysis." When A. v. Z. considered Cramér's tract, he emphasised the restricted nature of the assumptions–finite-dimensional spaces and independent random variables.

5 Sequels

What became of the people and their ideas? Turing established himself in mathematical logic, leaving Cambridge to work with von Neumann and Church in Princeton, and on the outbreak of the Second World War he was back in Britain where he worked at the Government Code and Cypher School at Bletchley Park breaking German codes; see Good (1979) and Hodge (ch. 3-5). At the end of the war Turing was one of the major figures in British computing. The war changed other Cambridge pure mathematicians into statisticians so that the post-war statistical community with people like G. A. Barnard, I. J. Good and D. V. Lindley was more receptive to probability theory.

Cramér's book was immediately recognised as authoritative but it did not generate any local research. Jeffreys's *Theory of Probability* (1939, p. 78) gave a Whitaker and Robinson proof of the central limit theorem but added a reference to Cramér for its greater "attention to mathematical detail." Cramér's book was also a reference for the foundational parts of Maurice Kendall's *Advanced Theory of Statistics* (1943). Other influences are harder to pin down but may have been more important. David Kendall, perhaps the most influential English probabilist of the post-war period, was a student when Cramér's book came out. Kendall bought it when his tutor told him, "I can't say anything about probability, but Cramér is all right." (See Bingham (1996, p. 169).)

Aitken did not do anything himself with those "pretty ideas" and his interest in probability found its chief expression in teaching and its most visible product was his undergraduate textbook, *Statistical Mathematics* (1939). This was too elementary a work to incorporate the sophisticated developments that had taken place around him but Aitken at least mentioned them.

Cramér became interested in the work of the English statistical school and he met Fisher in 1938 and Bartlett in 1939. Cramér was impressed by Fisher and he (1976, p. 529) recalled an interesting exchange:

> I had expressed my admiration for his geometrical intuition in dealing with probability distributions in multidimensional spaces, and received the somewhat acid reply: "I am sometimes accused of intuition as a crime!"

In 1946 Cramér published *Mathematical Methods of Statistics*, which brought together English statistics and continental probability. One of the notable features of the book was its treatment (pp. 500ff) of the large-sample distribution of the maximum likelihood estimator–something for which Bartlett had appealed for in 1938. Fisher described Cramér's book in a letter to a friend (Bennett (1990, pp. 330-1)).

> I recently received a very highbrow treatise of a first class Swede, namely Cramer, purporting to deal with mathematical statistics, of which I think a full half was a comprehensive introduction to a theory of point sets I suppose in view of the didactic manner in which the theory of probability is approached in Russia, for example, and in France, and in recent years in the United States, this sort of thing must seem a quite essential clarification of the subject.

If anything in later years Fisher became more hostile to "this sort of thing" and regularly complained that statisticians in mathematics departments were not well enough trained *as scientists*.

After the war Bartlett worked on stochastic processes and married English applied mathematics and Continental pure mathematics in his own way. In his *Introduction to Stochastic Processes*: *with Special Reference to Methods and Applications* he (1955, p. xiii) recognised the "important theoretical contributions" made by "American, French, Russian and Swedish writers" but his real interest was in the "methods and applications"–with what probability *does*.

5.1 Appendix I

So far it has proved impossible to trace the owner of the copyright of this report. If you have any information please contact the editor of this journal.

Mr. A. S. Besicovitch's Report on Mr. A. M. Turing's Dissertation 1935 On the Gaussian Error Function.

The main problem of the dissertation is to prove the Gaussian law of distribution of errors from the Laplace principle.

According to the Laplace principle an error of observation is the total of a large number of small independent errors, so that analytically the problem consists in finding the law of distribution of values of the sum of n variables

$$x_1 + x_2 + ... + x_n$$

under very general conditions for the law of distribution of each of the variables.

The dissertation is not to be judged from the point of view of its scientific value, as its main results were established long ago and even the fundamental idea of the method is not new. But the dissertation gives very strong grounds for judging its author's mathematical abilities.

The importance of the problem is well known. The difficulty of the problem was very great. At the beginning of the last century Laplace himself and later Poisson developed a method for its solution. But their solution was lacking an accuracy required by the mathematical analysis of the nineteenth century, and in the course of that century we see numerous attempts either to make the Laplace Poisson proof rigorous or to give a new one, and these attempts were crowned by success only in the period 1898-1908 by the works of Markoff and Liapounoff.

Then in 1922 Lindeberg gave an entirely new and incomparably simpler method. Mr. Turing's proof is based on two fundamental ideas, one of which coincides with the idea of Lindeberg's method, but failing to see that this idea is sufficient for the complete solution of the problem Mr Turing completes his method by means of another idea. These are the main steps of the proof: the author first considers the deviations from the Gauss law of the law of variation of the sum $Y + x$ where Y varies according the Gauss law and variation of x is restricted only by conditions of a very general kind. This enables the author to prove that if values of Y are of the same order of magnitude as those of $x_1 + x_2 + ... + x_n$, then the law of variation of the sum $Y + x_1 + x_2 + ... + x_n$ approaches uniformly the Gauss law. From this the author concludes finally that the law of variation of $x_1 + x_2 + ... + x_n$ also approaches uniformly the Gauss law.

Having a common idea with the Lindeberg proof the development of Mr. Turing's method is very much different from that of Lindeberg, which makes me completely confident that the work has been done in a genuine ignorance of Lindeberg's work.

Mr Turing's proof is somewhat more complicated than the Lindeberg proof, but all the same it is an excellent success and it would be so not only for a beginner but also for a fully developed scientist.

If the proof were published fifteen years ago it would be an important even in the mathematical literature of that year.

In Mr Turing's case we see a display of very exceptional abilities at the very start of his research works, which makes me to recommend him as a very strong candidate for a Fellowship.

5.2 Appendix II

This review is reproduced with permission from the *Mathematical Gazette*, **21**, 67-68.

Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie. By E. TORNIER. Pp. vi, 160. RM. 9. 1936. (Teubner)

Let no one buy this book hoping to obtain an account of Probability as it is usually understood. More than half the book is devoted to an abstract theory of the measure of sets of a general character and of the corresponding integrals. This is done without any reference to anyone, not even Fréchet, de la Vallée Poussin or Carathéodory. The style is forbidding, and there is no attempt to help the reader see what is intended. The only derivative considered is essentially a net-derivative, and it is pointed out that any *two* such net-derivatives are equally nearly everywhere, but that there are more than countable infinity of possible nets, and the measure of the sum of the sets where the net-derivatives are not equal to a particular one may be positive. In the case of n-dimensional Euclidean differentiation this difficulty is overcome by Vitali's theorem or by such a masterly method as that used by de la Vallée Poussin Professor Tornier does not discuss it even so far as to point out the difficulty of bringing in the essential concept of "shape" in abstract spaces.

On page 101 we come to a discussion of probability proper, but Professor Tornier has apparently never heard of J. M. Keynes or other critics of the fundamental notions. He writes glibly of obtaining probability in some cases "näherungsweise", though he does not say how this can be done without a circular use of Bayes' formula or else by the often disproved limit theories, such as Venn's. Professor Tornier merely assumes that the probability is a number satisfying the postulates for a general mass. With this he is able to obtain the usual general theorems on probability, including Chebyshef's. (why, by the way, do we spell Russian names as if they were first turned into German ?) Within 25 pages of the end the Gauss error function occurs, but only in connection with the more searching asymptotic approximations for total probabilities (Khinchine).

There seems little to recommend this book to a British audience. It is astonishing, however, that a book of so abstract a character should be accepted in modern Germany, where, according to publications received by many of us, mathematics is to be made concrete and "nordic".

P. J. D.

5.3 Appendix III

This review is reproduced with permission from the *Mathematical Gazette*, **21**, 67-68.

Random Variables and Probability Distributions. By HARALD CRAMÉR. Pp. 120, 6s. 6d. 1937. Cambridge Tracts, 36. (Cambridge)

The theory of probability has been cultivated in England less for what it is than for what it does. The research of the present century on the theory of estimation and the distribution of statistical coefficients has taken the fundamentals for granted and has sometimes been avowedly non-rigorous. On the Continent, however, there has been strong dissatisfaction with the axiomatic basis of probability, which has been manifested in different ways. Von Mises has essayed to reformulate the frequency theory of probability by a new theory of admissible sequences of events; Lévy, Fréchet, Kolmogoroff, Cantelli, Tornier and others have replaced the old *a priori* definition of probability as ratio of favourable to possible cases by a more rigorous theory based on sets of points and Lebesgue measure; Hostinsky has developed Poincaré's theory of arbitrary functions.

The tract which we review has come at exactly the time when it was required, as a presentation to English readers of the *a priori* standpoint of Lévy, Fréchet, Kolmogoroff and Cramér himself. The tendency of these writers is towards mathematical abstraction, and the theory that emerges is a branch of pure mathematics, a part of general analysis. An event is described by a set of coordinates, a point or vector, in a space R_k of k dimensions. Possible events form a point-set or aggregate; the probability of an event is a completely additive set function of the associated point-set, and the treatise begins by developing the properties of such functions. The approach thus resembles that of Kolmogoroff's *Grundbegriffe* of 1933 (Springer, Berlin), but is somewhat more specialised in that the event-space considered is a Euclidean space R_k of a finite number k of dimensions.

It will be seen from these brief indications that no time is wasted over discussions of "equal likeliness". In fact the question does not arise, since to specify the point-set and the Lebesgue measure of subsets is in fact to say what subsets are equally likely. The question of "equal likeliness", that is, of proper choice of the event-space, thus devolves on the *applied* statistician; and the well-known "paradoxes" of geometrical probability show how different choices of event-space lead to different conclusions.

Chapters I and II are devoted to introductory remarks, the definition of Borel sets S in R_k, of completely additive set functions $P(S)$, point functions $g(X)$, and the Lebesgue-Stieltjes integral of $g(X)$ with respect to a set function $P(S)$. The probability function $P(S)$ for a set S is defined as a non-negative completely additive set function such that $P(R_k) = 1$. For a set comprising all vectors having components less than those of X, the total probability function $F(X)$, a point function as contrasted with a set function, is called the distribution function of X. Independent random variables are defined as those for which the compound distribution function, in the product space of all the variables, is the product of the several distribution functions.

Chapter II considers distributions in R_1, mean values, moments and absolute moments, and inequalities. Chapters IV and V introduce the characteristic function defined by

$$f(t) = \int_{-\infty}^{\infty} \exp(itX)dF(x),$$

with important theorems on the necessary and sufficient conditions for convergence of distribution functions to ensue from uniform convergence of characteristic functions. Chapter V considers the addition of random variables by composition or convolution of distribution functions and multiplication of characteristic functions. This is followed by a theory of convergence in probability, with well-chosen examples on the binomial, Poisson, Pearson Type III and Cauchy distributions. There is also a very useful theorem (Th. 16) on a quotient of random variables.

Chapters VI and VII bring us to the core of the book, the "central limit theorem", according to which the sum of a large number of independent random variables is distributed approximately according to a normal law. The necessary and sufficient condition of Lindeberg is given, and the sufficient condition of Liapounoff. Chapter VIII gives Liapounoff's theorem on the order of the remainder term in the approach to normal approximation and the author's own asymptotic expansion (first given formally by Edgeworth), which differs in the arrangement of terms from the so-called Gram-Charlier, Bruns or Type A Series.

Chapters VIII briefly considers what has been called the *homogeneous random process*, in which the addition of random variables (which we may suppose carried out at discontinuous intervals of time $\triangle\tau$) is replaced by integration with respect to a continuous time parameter τ, and Chapters IX and X extend the important theorems of earlier chapters from R_1 to R_k. The work concludes with a bibliography, confined almost entirely to memoirs later than 1900, and for the most part later than 1925.

As a piece of exposition, the tract must be given the highest praise for its clarity and for the excellent arrangement of its material. It is really the first book to give an adequate account in English of the contemporary tidying up of the purely mathematical side of probability, and should be studied by everyone interested in the postulational basis of the subject.

A. C. A.

6 References

Turing's dissertation is available on the Alan Turing website created by Andrew Hodge http://www.turingarchive.org/. Many of Fisher's letters, as well as the volume of correspondence edited by Bennett (1990), are available from the R. A. Fisher Digital Archive at the University of Adelaide

http://digital.library.adelaide.edu.au/coll/special//fisher/index.html.

Aitken, A. C. (1931) Some Applications of Generating Functions to Normal Frequency, *Quarterly Journal of Mathematics*, **2**, 131-135.
_________ (1935) On Least Squares and Linear Combinations of Observations, *Proceedings of the Royal Society of Edinburgh*, **55**, 42-48.
_________ (1938) Review of *Random Variables and Probability Distributions.* by H. Cramér, *Mathematical Gazette*, **22**, 193-194.
_________ (1939) *Statistical Mathematics*, Edinburgh: Oliver & Boyd.
_________ (1958) The Contributions of E. T. Whittaker to Algebra and Numerical Analysis, *Proceedings of the Edinburgh Mathematical Society*, **11**, 31-38.
Aldrich, J. (2001/9) *Karl Pearson: A Reader's Guide*, website http://www.economics.soton.ac.uk/staff/aldrich/kpreader.htm
_________ (2003/9) *A Guide to R. A. Fisher*, website http://www.economics.soton.ac.uk/staff/aldrich/fisherguide/rafreader.htm
_________ (2005) The Statistical Education of Harold Jeffreys, *International Statistical Review*, **73**, 289-308.
_________ (2005/9) *The Mathematics PhD in the Unitd Kingdom*, website http://www.economics.soton.ac.uk/staff/aldrich/PhD.htm
_________ (2007) "But You Have to Remember P. J. Daniell of Sheffield", *Journal Electronique d'Histoire des Probabilités et de la Statistique*, **3**, 58pp.
_________ (2008) Keynes among the Statisticians, *History of Political Economy*, **63** (1), 265-316.
_________ (2009) Burnside's Engagement with "Modern Statistical Theory", *Archive for History of Exact Sciences*, **40** (2), 51-79.
Bartlett, M. S. (1933) Probability and Chance in the Theory of Statistics, *Proceedings of the Royal Society of London, A*, **141**, 518-534.
_________ (1938) Review of *Random Variables and Probability Distributions* by H. Cramér, *Journal of the Royal Statistical Society*, **101**, 51-79.
_________ (1955) *An Introduction to Stochastic Processes: with Special Reference to Methods and Applications*, Cambridge: Cambridge University Press.
_________ (1982) Chance and Change in J. Gani (ed.) (1982) *The Making of Statisticians*, New York: Springer-Verlag.
Bell, J. L. (1984) Obituary: Edward Hubert Linfoot, *Bulletin of the London Mathematical Society*, **16**, 52-58.
Bennett, J. H. (1990) (ed) *Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher*, Oxford, University Press. The text is available on the website http://digital.library.adelaide.edu.au/coll/special//fisher/stat_inf/index.html
Bertrand, J. (1888) *Calcul des Probabilités*, Paris: Gauthier-Villars.
Besicovitch, A. S. (1915) New Proof of the Theorem on the Limit of the Probability in the General Case of Independent Trials, *Bulletin of the Academy of Science* (Petrograd) (6), **9**, 719-748. (Russian).
_________ (1923-4) A. A. Markov's Contribution to Probability Theory, *Bulletin of the Academy of Science* (Leningrad) (6), **17**, 45-52. (Russian).
_________ (1932) *Almost Periodic Functions*, Cambridge: Cambridge University Press.
_________ (1935) Report on Mr. A. M. Turing's Dissertation 1935 On the Gaussian Error Function, King's College Cambridge Library KCAC/4/11/2/6/5.

Bingham, N. H. (1996) A Conversation with David Kendall, *Statistical Science*, **11**, 159-188.
Bloch, A. and G. Pólya (1932) On the Roots of Certain Algebraic Equations, , *Proceedings of the London Mathematical Society*, **33**, 102-114.
Bru, B. and S. Eid (2009) Jessen's Theorem and Lévy's Lemma: A Correspondence, *Journal Electronique d'Histoire des Probabilités et de la Statistique*, **5**, (1).
Brunt, D. (1917) *The Combination of Observations*. Cambridge University Press, Cambridge.
Burkill, J. C. (1971) Abram Samoilovitch Besicovitch, *Biographical Memoirs of Fellows of the Royal Society of London*, **17**, 1-16.
Burnside, W. (1928) *Theory of Probability*, Cambridge: Cambridge University Press.
Courant, R. and D. Hilbert (1924) *Methoden der mathematischen Physik*, Berlin: Springer.
Cramér, H. (1937) *Random Variables and Probability Distributions*, Cambridge: Cambridge University Press.
_________ (1946) *Mathematical Methods of Statistics*, Princeton University Press, London.
_________ (1976) Half a Century with Probability Theory: Some Personal Recollections, *Annals of Probability*, **4**, 509-546.
Crofton, M. W. (1885) Probability, *Encyclopaedia Britannica,* ninth edition, vol. , 768-788. London.
Czuber, E. (1891) *Theorie der Beobachtungsfehler*, Leipzig: Teubner.
Daniell, P. J. (1937) Review of *Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie* by E. Tornier. *Mathematical Gazette*, **21**, 67-68.
David, F. N. (1938) Limiting Distributions Connected with Certain Methods of Sampling Human Populations, *Statistical Research Memoirs*, **2**, 69-90.
Diaconis, P. (2002) G. H. Hardy and Probability???, *Bulletin of the London Mathematical Society, 34,* 385-402.
Doob, J. L. (1937) Review of *Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie* by E. Tornier. *Bulletin of the American Mathematical Society*, **43**, 317-318.
Edgeworth, F. Y. (1883) The Law of Error, *Philosophical Magazine*, **16**, 300-309.
_________ (1905) The Law of Error, *Proceedings of the Cambridge Philosophical Society*, **20**, 36-141.
Farebrother, R. W. (1997) A. C. Aitken and the Consolidation of Matrix Theory, *Linear Algebra and its Applications*, **264**, 3-12.
Fisher, A. (1922) *The Mathematical Theory of Probabilities and its Application to Frequency Curves and Statistical Methods*, second edition, New York: Macmillan.
Fisher, R. A. (1922) On the Mathematical Foundations of Theoretical Statistics, *Philosophical Transactions of the Royal Society, A*, **222**, 309-368.
_________ (1925) Sur la Solution de l'Équation Intégrale de M. V. Romanovsky, *Comptes Rendus*, **180**, 1897-1899.

_________ (1929) Moments and Product Moments of Sampling Distributions, *Proceedings of the London Mathematical Society*, **30**, 199-238.

Fisher, R. A. and J. Jackson (1924) Publications: *The Calculus of Observations. A Treatise on Numerical Mathematics* (E. T. Whittaker and G. Robinson), *The Observatory*, **47**, 154-157.

Fisher-Aitken correspondence 1931-1951 (part of the Adelaide Fisher digital archive) at
http://digital.library.adelaide.edu.au/coll/special//fisher/corres/aitken/index.html

Fisher-Sheppard correspondence 1934 (part of the Adelaide Fisher digital archive) at
http://digital.library.adelaide.edu.au/coll/special//fisher/corres/turing/index.html

Good, I. J. (1979) Studies in the History of Probability and Statistics. XXXVII A. M. Turing's Statistical Work in World War II, *Biometrika*, **66**, 393-396.

Hagen, G. (1837) *Grundzüge der Wahrscheinlichkeits-Rechnung*, Berlin:Dümmler.

Hardy, G. H. (1934) Raymond Edward Alan Christopher Paley, *Journal of the London Mathematical Society*, **9**, 76-80.

Hardy, G. H. and J. E. Littlewood (1914) Some Problems of Diophantine Approximation, *Acta Mathematica*, **37**, 151-191, 193-239.

Hodge, A. (1985) *Alan Turing: The Enigma of Intelligence*, London: Harper-Collins.

_________ (2002-3) *The Turing Digital Archive*
http://www.turingarchive.org/.

Ingham, A. E. (1933) An Integral which Appears in Statistics, *Proceedings of the Cambridge Philosophical Society*, **29**, 260-270.

Irwin, J. O. (ed.) (1935) Recent Advances in Mathematical Statistics (1933), *Journal of the Royal Statistical Society*, **98**, 714-769.

Jeffreys, H. (1939) *Theory of Probability*, Oxford: University Press.

Kendall, M. G. (1943) *The Advanced Theory of Statistics, vol. 1*, London: Griffin.

Keynes, J. M. (1911) Review of *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik, und Lebensversicherung* by Emanuel Czuber, *Journal of the Royal Statistical Society*, **74**, 643-647.

_________ (1921) *A Treatise on Probability*, London: Macmillan.

Khinchin, A. (1926) Über das Gesetz der Großen Zahlen, *Mathematische Annalen*, **96**, 152-168.

Kolmogorov, A. N. (1933) *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Berlin: Springer.

Le Cam, L. (1986) The Central Limit Theorem Around 1935, *Statistical Science*, **1**, 78-91.

Lévy, P. (1925) *Calcul des Probabilités*, Paris: Gauthier-Villars.

_________ (1937) *Théorie de l'addition des variables aléatoires*, (Fascicule I de la Collection des monographies des probabilités), Paris, Gauthier-Villars.

Lindeberg, J. W. (1922) Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift*, **15**, 211-225.

Linfoot, E. H. (1928) On the "Law of Large Numbers", *Proceedings of the London Mathematical Society*, **28**, 347-367.

_________ (1928) On the Law of Large Numbers, *Philosophical Transactions, A*, **227**, 417-451.
Littlewood, J. E. & A. C. Offord (1938) On the Number of Real Roots of a Random Algebraic Equation, I, *Journal of the London Mathematical Society*, **13**, 288-295.
Marshall, J. (1926a) Review of *Calcul des Probabilités* by E. Borel. *Mathematical Gazette*, **13**, 36-37.
_________ (1926b) Review of *Calcul des Probabilités* by P. Lévy. *Mathematical Gazette*, **13**, 214.
Mises, R. von (1928) *Wahrscheinlichkeit Statistik und Wahrheit*, Vienna: Springer. English translation by J. Neyman, D. Sholl and E. Rabinowitsch as *Probability Statistics and Truth* (1939).
_________ (1931) *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Leipzig: Deuticke.
Neumann, J. von (1932) *Mathematische Grundlagen der Quantenmechanik*, Berlin: Springer.
Neyman, J. (1926) Sur les lois de probabilité qui tendent vers la loi de Gauss, *Comptes Rendus*, **182**, 1590-1592.
_________ (1937) Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability, *Philosophical Transactions, A*, **236**, 333-380.
Olkin, I. (1989) A Conversation with Maurice Bartlett, *Statistical Science*, **4**, 151-163.
Paley R. E. A. C. and N. Wiener (1934) *Fourier Transforms in the Complex Domain*, American Mathematical Society Colloquium Publications no. 19.
Paley R. E. A. C., N. Wiener and A. Zygmund (1933) Note on Random Functions, *Mathematische Zeitschrift*, **37**, 647-668.
Paley R. E. A. C., and A. Zygmund (1930-1932) On Some Series of Functions (1), (2), (3), *Proceedings of the Cambridge Philosophical Society*, **26**, 337-357, 458-474, **28**, 190-205.
Plato, J. von (1994) *Creating Modern Probability*, Cambridge: Cambridge University Press.
Poincaré, J. H. (1912) *Calcul des Probabilités*, Paris: Gauthier-Villars.
Pólya, G. (1920) Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem, *Mathematische Zeitschrift*, **8**,, 171-181.
Rice, A. C. and R. J. Wilson (2003) The Rise of British Analysis in the Early 20th Century: the Role of G. H. Hardy and the London Mathematical Society, *Historia Mathematica*, **30**, 173-194.
Rider, P. R. (1930) A Survey of the Theory of Small Samples, *Annals of Mathematics*, **31** (4), 577-628.
Shafer, G. & V. Vovk (2006) The Sources of Kolmogorov's *Grundbegriffe*, *Statistical Science*, **21**, 70-98.
Siegmund-Schultze, R. (2001) *Rockefeller and the Internationalization of Mathematics Between the Two World Wars*, Basel: Birkhäuser.
Stigler, S. M. (1978) Francis Ysidro Edgeworth, Statistician, (with discussion) *Journal of the Royal Statistical Society*, **141** (3), 287-322.
Student (1908) The Probable Error of a Mean, *Biometrika*, **6**, 1-25.

Taylor, S. J. (1975) Abram Samoilovitch Besicovitch, *Bulletin of the London Mathematical Society*, **7**, 191-210.
Temple, G. F. J. (1956) Edmund Taylor Whittaker, *Biographical Memoirs of Fellows of the Royal Society of London*, **2**, 299-325.
_________ (2004) Whittaker, Sir Edmund Taylor (1873–1956), [rev. I. Grattan-Guinness], *Oxford Dictionary of National Biography*, Oxford University Press.
Tornier, E. (1936) *Wahrscheinlichkeitsrechnung und allgemeine Integrationstheorie*, Leipzig and Berlin: Teubner.
Turing, A. (1935) *On the Gaussian Error Function*, King's College Fellowship dissertation. The original is in the library of King's College Cambridge but a copy is available online at
http://www.turingarchive.org/browse.php/C/28
Venn, J. (1866). *The Logic of Chance: An Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to its Application to Moral and Social Science*, 3rd edition 1888. London: Macmillan.
Whittaker, E. T. & G. Robinson (1924) *The Calculus of Observations: A Treatise on Numerical Mathematics*, London: Blackie.
Whittaker, E. T. & G. S. Watson (1915) *A Course of Modern Analysis*, (First edition 1903 by Whittaker alone) Cambridge: Cambridge University Press.
Whittaker, J. M. & M. S. Bartlett (1968) Alexander Craig Aitken, 1895-1967, *Biographical Memoirs of Fellows of the Royal Society*, **24**, 1-14.
Whittle, P. (2004) Maurice Stevenson Bartlett, *Biographical Memoirs of Fellows of the Royal Society*, **50**, 15-33.
Wiener, N. (1924) The Average Value of a Functional, *Proceedings of the London Mathematical Society*, **22**, 454-467.
_________ (1924) Un problème de probabilités dénombrables, *Bull. Soc. Math. France*, 11, 569-578.
_________ (1933) *The Fourier Integral and Certain of its Applications*, Cambridge: Cambridge University Press.
Wishart, J. & M. S. Bartlett (1932) Distribution of Second-order Moment Statistics in a Normal System, *Proceedings of the Cambridge Philosophical Society*, **28**, 455-459.
_________ (1933) The Generalized Product Moment Distribution in a Normal System, *Proceedings of the Cambridge Philosophical Society*, **29**, 260-270.
Zabell, S. (1995) Alan Turing and the Central Limit Theorem, *American Mathematical Monthly*, **102**, 483-494.
[A. v. Z.] (1938) Recent Developments in Probability Theory, *Nature*, **141**, 55-56.