Studies in the History of Probability and Statistics XLVI. Measure into Probability: From Lebesgue to Kolmogorov

Author(s): N. H. Bingham

Source: *Biometrika*, Mar., 2000, Vol. 87, No. 1 (Mar., 2000), pp. 145-156

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: https://www.jstor.org/stable/2673568

Biometrika (2000), **87**, 1, *pp.* 145–156

Printed in Great Britain

Studies in the history of probability and statistics XLVI. Measure into probability: From Lebesgue to Kolmogorov

BY N. H. BINGHAM

Department of Mathematics and Statistics, Birkbeck College (University of London), Malet Street, London WC1E 7HX, U.K.

n.bingham@bbk.ac.uk

SUMMARY

Both probability and statistics had developed into major fields by the end of the nineteenth century, with a sizeable body of theory and extensive application, but both disciplines achieved their existing form only this century, by making use of modern ideas. This study traces this evolution, with particular reference to the role of measure theory, from its introduction in definitive form by Lebesgue in 1902 to its successful use in the axiomatisation of probability by Kolmogorov in 1933.

Some key words: History of probability and statistics; Measure theory; Probability.

THE SITUATION AT THE END OF THE NINETEENTH CENTURY

The nineteenth century had begun very promisingly for probability and statistics, with the work of Legendre, Gauss and Laplace on the method of least squares and the role of the normal, i.e. Gaussian or Laplacian, distribution in the theory of errors. Recall that all three of these were giants of mathematics, and that all three were motivated by problems of celestial mechanics and planetary orbits. For a fine account of this heroic period, we refer to Part One, especially Chapter 4, 'The Gauss–Laplace synthesis', of Stigler (1986).

The mid-nineteenth century was very much an heroic period for most of mathematics; the core of the undergraduate curriculum, for example, dates from this period. By contrast, however, there was a long pause in the development of probability and statistics, though for Quetelet and Poisson see Stigler (1986, Part Two), and for Bienaymé see Heyde & Seneta (1977). This ended with the great breakthroughs of the late nineteenth century by the English school of statisticians, beginning with the work of Sir Francis Galton (1822–1911) on regression and correlation (Stigler, 1986, Ch. 8). This is all the more remarkable because Galton was himself a weak mathematician. He was able to go where Gauss himself had not because the questions that motivated him, the laws of heredity and in particular how the child's height depends on the heights of his/her parents, could only be addressed in these terms. Galton's work was continued by F. Y. Edgeworth (1845–1926), to whom we owe the basic theory of the multivariate normal distribution, see Stigler (1986, Ch. 9) and Fitzpatrick (1960, § 11), and by Karl Pearson (1857–1936) (Stigler, 1986, Ch. 10), whose influential book *The Grammar of Science* (Pearson, 1892) made these new ideas accessible to the broader scientific public. Pearson's chi-squared test of 1900 may be regarded as beginning the prelude to the modern era in statistics, and in 1901 Pearson founded *Biometrika*, with W. F. R. Weldon and Francis Galton. For a study in depth of Pearson's work on mathematical genetics, see Magnello (1998).

On the probability side, the most notable practitioners of the time belonged to the French and Russian schools. In France, Henri Poincaré (1854–1912) published his *Calcul des Probabilités* (Poincaré, 1896). In Russia, A. A. Markov (1856–1922), Professor at St. Petersburg (1883–1905) in succession to Tchebychev, published his book at the turn of the century (Markov, 1900).

On the analysis side, Emile Borel (1871–1956), in his *Leçons sur la théorie des fonctions* (Borel, 1898), began the study of measure theory. His work, which introduced the term measure, improved on that of Jordan on content, and paved the way for the later definitive treatment by his pupil Lebesgue; see § 3.

2. THE TURN OF THE CENTURY

When David Hilbert (1862–1943) addressed the International Congress of Mathematicians in Paris in 1900, he gave the famous list of 23 'Hilbert problems', which have so much influenced twentieth-century work. Hilbert's sixth problem was as follows.

Problem 6: *Mathematical treatment of the axioms of physics.* 'The investigations on the foundations of geometry suggest the problem: to treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part; in the first rank are the theory of probabilities and mechanics' (Hilbert, 1902).

One could regard the challenge to axiomatise mechanics as one precursor of the greatest triumph of twentieth-century science, quantum mechanics, which developed from the introduction in 1900 by Max Planck (1858–1947) of a quantum theory of radiation. A study of Hilbert's thinking on the axiomatisation of physics is given by Corry (1997). How the parallel challenge to axiomatise probability was taken up is the main subject of the present paper. Hilbert's description of probability as a physical science, which one can hardly imagine being made today, is striking, and presumably reflects both the progress in statistical mechanics by Maxwell, Boltzmann and Gibbs and the unsatisfactory state of probability theory at that time judged as mathematics.

A referee has contributed the interesting observation that the Royal Society in 1917 classified probability as a branch of algebra, presumably because of Boole's work (Boole, 1854).

Emanuel Czuber (1851–1925) of Vienna gave a report of nearly 300 pages on the state of probability theory at that time, seen from the German-speaking world (Czuber, 1898). This was later made the basis of his influential textbook (Czuber, 1903).

3. THE FRENCH SCHOOL

The foundation for the transformation of probability theory characteristic of this century was provided by Henri Lebesgue (1875–1941) in his thesis *Intégrale, longueur, aire* (Lebesgue, 1902). Lebesgue's key contribution was the systematic use of countably additive set-functions $\mu(.)$, so that

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$$

for disjoint sets A_n in some suitable class, that is the σ-algebra of measurable sets; for μ

nonnegative these are called 'measures'; he linked this with his construction of an integral, the Lebesgue integral, based on such a measure. Lebesgue's new measure theory provided a mathematically satisfactory treatment of length, area, volume, hypervolume, mass, electrostatic charge, etc.; his new integration theory both greatly extended the Riemann integration theory then in use and greatly simplified the taking of limits in integrals by providing widely applicable convergence theorems, the theorems of dominated convergence and of monotone convergence.

Maurice Fréchet (1878–1973) introduced the now ubiquitous axioms for metric spaces in his thesis (Fréchet, 1906). Fréchet's work begins the trend towards abstraction and axiomatisation characteristic of modern analysis. In particular, it led to the freeing of Lebesgue's measure theory from its Euclidean origins, particularly in the later work of C. Carathéodory (1873–1950), and to the development of general topology. The seminal role of Fréchet's work was acknowledged by Kolmogorov early in his classic *Grundbegriffe* (Kolmogorov, 1933); see § 10. For Fréchet's later work, see Fréchet (1928).

Emile Borel (1871–1956), with his pupil Lebesgue one of the founders of measure theory, published in 1909 his *Les probabilités dénombrables et leurs applications arithmétiques* (Borel, 1909). This influential study is notable for several major results, Borel's Normal Number Theorem that almost all numbers are strongly normal to all bases simultaneously, the Borel–Cantelli lemma and Borel's theorem on continued fractions. The corollary to the Normal Number theorem, that such numbers exist, is striking; no specific example is known! This is a classic example of a non-constructive existence proof. Unfortunately, Borel's work is non-rigorous, and, oddly, hardly mentions measure theory. A detailed critique of his work is provided by Novikoff & Barone (1977), and Barone & Novikoff (1978), who also showed that he was partially anticipated by the American mathematician Van Vleck (Van Vleck, 1908).

It is the height of irony that the initiator of measure theory should have failed to see the natural application of his own creation to probability theory. Presumably, Borel's view of measure theory at that time was still tied to the Euclidean and geometric setting. Again, he was presumably led by the fact that his work mainly involved only countable sequences of events, 'les probabilitiés dénombrables' of his title, to infer that its natural context was that of purely atomic measure spaces, where, as nontrivial null sets do not exist, the measure theory as such is trivial. Be that as it may, this crucial oversight delayed the necessary evolution of probability into a measure-theoretic formulation until the work of Steinhaus (1923), see § 4, and Kolmogorov (1933), see § 10.

Paul Lévy (1886–1971), one of the great probabilists of the century, wrote the first probability textbook based on measure-theoretic ideas, *Calcul des Probabilités* (Lévy, 1925), where he introduced and studied characteristic functions. He also created the theory of stochastic processes with stationary independent increments, now known as Lévy processes, and the Lévy–Khinchin formula for them, and gave in his later work a deep study of the sample path and martingale properties of Brownian motion. For background, see his obituary by Loève (1973).

In his autobiography (Lévy, 1970, pp. 67–8), Lévy writes poignantly of realising too late, on first seeing Kolmogorov (1933), the opportunity of writing such a book that he himself had missed. He also writes here on his differences in outlook from Borel:

> He felt that probability had to be based on rigorous analysis, while Borel felt that it should be studied from the point of view of applications. Thus, Borel may actually have delayed the acknowledgement of probability as a proper mathematical discipline'
>
> (Krengel 1977, p. 23).

Though out of chronological order, because they stand to one side of the main line of development above, two other French contributions deserve mention. In his thesis *Théorie de la spéculation* (Bachelier, 1900), written under Poincaré, Louis Bachelier (1870–1946) introduced probability theory into economics by proposing Brownian motion as a model for the fluctuations of stock prices; for details and background, see Brush (1968). This pioneering work is commemorated in the title of the Bachelier Finance Society, founded in 1996. For Bachelier's work in the context of modern mathematical finance, see e.g. Bingham & Kiesel (1998, §§ 5.3, 5.6.1).

In 1908, Paul Langevin (1872–1946) published his work on Brownian motion (Langevin, 1908), which, though non-rigorous, is notable for its introduction of a stochastic differential equation. This development was decades ahead of its time; the rigorous treatment of Brownian motion came in the twenties with Wiener, see § 8, and the forties with Lévy. The Langevin equation was studied later by Ornstein & Uhlenbeck (1930), and, rigorously, by Doob (1942). For the early history of Brownian motion, see Haas-Lorentz (1913) and, for later developments, Bingham (1998).

4. THE POLISH SCHOOL

There was a great flowering of Polish mathematics between the two world wars, particularly centred on analysis, foundational questions and the journals *Studia Mathematica* and *Fundamenta Mathematicae*. A number of Polish mathematicians of this period worked on probability, beginning with Hugo Steinhaus (1887–1972), who in 1923 gave an early attempt at axiomatisation (Steinhaus, 1923). Steinhaus worked later, from 1936 on, in collaboration with Mark Kac (1914–1984). The other important Polish collaboration of this period was that between Jozef Marcinkiewicz (1910–1940) and Antoni Zygmund (1900–1992). Leading Polish contributions from this time include the Marcinkiewicz–Zygmund inequalities and the Marcinkiewicz–Zygmund law of the iterated logarithm of 1937, and Steinhaus' work on the natural boundary of a random power series. A detailed study of this Polish work, which centred on independent functions as a bridge between probability and analysis, was given by Holgate (1997). The present study is intended as a complement to this. Much interesting historical background to this Polish work is given by Kahane & Lemarié-Rieusset (1995, Part I, Ch. 9).

A wealth of interesting background on Polish mathematics during this period is given in Kuratowski (1980). Of particular interest is the Lvov School. Its greatest member was Stefan Banach (1892–1945), the founder of modern functional analysis with his classic book (Banach, 1932); other members included Steinhaus, S. Mazur, S. Ulam and J. Schauder. They met to eat, drink and do mathematics in the Szkocka, or Scottish Café, where they wrote mathematics on the marble table-tops. Eventually, Banach's wife produced a book for this purpose, which became the famous 'Scottish Book'. This was presented to the S. Banach International Mathematical Centre in Warsaw on the day it opened, 13 January 1972, by Banach's son. For background on the Scottish Book and its associations, see Kuratowski (1980, Ch. 2), Ulam (1976, Ch. 2) and Cieselski (1987, 1988).

5. THE RUSSIAN SCHOOL

We turn now to the work which definitively succeeded in harnessing measure theory to the service of probability theory, that of the Russian school.

Sergei Natanovich Bernstein (1880–1968), professor at Kharkov, wrote extensively on probability and analysis, including Bernstein polynomials and their application to the law of large numbers, Bernstein's theorem on completely monotone functions and his book (Bernstein, 1911). For an appreciation of Bernstein's work, see Kolmogorov & Sarmanov (1960).

Alexander Yakovlevich Khinchin (1894–1956) was second only to Kolmogorov in the Russian school of probability. His greatest achievement is, arguably, his law of the iterated logarithm of 1924, completing, with the law of large numbers and the central limit theorem, the trio of classical limit theorems in probability theory. Other important results include his definitive form of the weak law of large numbers of 1929, the Birkhoff–Khinchin ergodic theorem of 1931, the Lévy–Khinchin formula and joint work with Kolmogorov of 1925. For a detailed appreciation, see the obituary article by Gnedenko & Kolmogorov (1960).

Andrei Nikolaevich Kolmogorov (1903–1987) wrote some dozen papers on probability between his first work in this area, with Khinchin in 1925, mentioned above, and his epoch-making *Grundbegriffe* (Kolmogorov, 1933), discussed in § 10, mainly on limit theorems and the Lévy–Khinchin formula. For background and details, see e.g. Shiryaev (1986), Bingham (1989) and Kendall (1990).

6. The Italian School

Francesco Paolo Cantelli (1906–1985) is remembered for two principal contributions. The first, of 1917, is his part in the Borel–Cantelli lemmas. The second, of 1933, is his work on the Glivenko–Cantelli theorem, or Fundamental theorem of Statistics, that the sample determines the population in the limit. The work of Vladimir Glivenko (1897–1940) was also of 1933. For background, see e.g. Cifarelli & Regazzini (1996) and Seneta (1992).

Bruno de Finetti (1906–1985) contributed to probability theory through his work on stochastic processes with independent increments, or Lévy processes in modern terminology, and the Lévy–Khinchin formula for them, in 1929. He also worked on exchangeability, not then so called, from 1933 onwards. However, perhaps de Finetti's most significant work was his early advocacy of the Bayesian approach to statistics, culminating in his book (de Finetti, 1970). This includes de Finetti's provocative and famous dictum that probability does not exist. A critique of this book is given by Good (1978). For a full appreciation of de Finetti's work, see Cifarelli & Regazzini (1996) and the obituary article by Lindley (1986).

A full account of work on probability theory in Italy between the wars is given by Regazzini (1987).

7. The German School

Hilbert, in a footnote to this problem list of 1900, refers to work of G. Bohlmann on axiomatisation of probability; Bohlmann's work was published in the *Encyclopädia der Mathematischen Wissenschaften* of 1901, and in Bohlmann (1908). In particular, Bohlmann gave the first rigorous definition of independence, the product rule for independence having been regarded as a theorem rather than a definition previously, and distinguished between independence and pairwise independence. For background here, and much else besides, see Krengel (1997).

Felix Hausdorff (1868–1942) contributed to probability theory in his early work of 1913, in the context which later became that of the law of the iterated logarithm. He obtained the bound $O(n^{\frac{1}{2}+\varepsilon})$, later refined to $O\{(n \log n)^{\frac{1}{2}}\}$ by G. H. Hardy and J. E. Littlewood in 1914, and to $O\{(n \log \log n)^{\frac{1}{2}}\}$ by Khinchin in 1924. Hausdorff also gave the first rigorous proof of the strong law of large numbers for Bernoulli trials in his book *Grundzüge der Mengenlehre* (Hausdorff, 1914), which continued the work of Borel, Lebesgue and Fréchet by treating families of sets, rather than of points, as objects of study. This led to the development of general topology, which has fed back into probability theory in more modern developments. For example, it is crucial for my own first love, weak convergence and functional central limit theorems, and is the key ingredient missing from the otherwise still magisterial Banach (1932).

Hausdorff became a victim of the Nazi period: he, his wife and her sister took their own lives on 26 January 1942 to escape deportation to Theresienstadt from the prison camp of Endenich. The day before, he wrote his own immortal epitaph:

'auch Endenich
Ist noch vielleicht das Ende nich!'

(Eichhorn, 1994). For background on Hausdorff, see Brieskorn (1996).

Richard von Mises (1833–1953), in his book *Wahrscheinlichkeit, Statistik und Wahrheit* (von Mises, 1928), introduced a theory of what he called collectives. In effect, this was an attempt to turn the Strong Law of Large Numbers from a theorem into an axiom. Lévy remarked that it is as impossible to build a satisfactory theory in this way as it is to square a circle. Von Mises' approach attempts to build a theory of random sequences sequence by sequence, in contrast to the Kolmogorov measure-theoretic approach, which makes statements about the totality of all such sequences, rather than about individual sequences. We return to later developments here in § 11. Von Mises is the originator of the method of differentiable statistical functionals, also known as the von Mises calculus or the infinite-dimensional delta-method. This goes back to his early work of 1931 on what is now called the Cramér–von Mises statistic for testing goodness of fit; for details, see e.g. Cramér (1946, § 30.8). For the general theory see von Mises (1947), or, for a modern textbook account, van der Vaart (1988, Ch. 20).

U. Krengel, in 'Bernoulli News' volume 4 number 2, gives an excellent and thorough account of probability in the German-speaking world during the period 1890–1990.

Since he wrote in German, we include here mention of the probabilistic work of the Hungarian mathematician George Pólya (1887–1985); see Chung (1987) for background and details. Pólya was a pupil of Fejer; he worked in ETH Zürich, and became a naturalised Swiss. In Pólya (1920), he coined the term 'central limit theorem', in German. In Pólya (1921), he gave his theorem on transience and recurrence of the simple random wall, whose influence is still widely felt; it inspired the recent book by Doyle & Snell (1984), for instance. The term 'central limit theorem', which is standard nowadays, had not become established by the time of Lindeberg (1922), where the Lindeberg condition for the central limit theorem for independent, not necessarily identically distributed random variables is given. Lindeberg proved sufficiency of his condition; necessity was proved in 1935 by Lévy and by Feller. For details, see Zabell (1995); we return to Lindeberg in connection with Turing in § 9 below.

8. The American School

Norbert Wiener (1894–1964) wrote on probability in a series of papers between 1920 and 1923. In particular, Wiener was the first to construct a rigorous mathematical model for the physical phenomenon of Brownian motion. It is in recognition of Wiener's work here, and in particular his construction of Wiener measure on the space of continuous functions, path space, that Brownian motion is often referred to as the Wiener process. Wiener's work culminated in his book with R. E. A. C. Paley (1907–1933), Paley & Wiener (1934); see also Paley, Wiener & Zygmund (1933). For historical background, see Kahane (1997, §§ 8, 9) and Kahane & Lemarié-Rieusset (1995, Part I, § 9.1). In addition, Wiener's work on Fourier integrals was highly influential in convincing the analysis community to change from the Riemann to the measure-theoretic Lebesgue integral (Wiener, 1930). His work on generalised harmonic analysis (Wiener, 1933) laid the basis for the rigorous treatment of spectral analysis in time series; the subject was initiated in 1898 by Sir Arthur Schuster in his work on geomagnetism (Schuster, 1900).

An important stimulus to American work in probability and statistics was the founding in 1930 of the journal *The Annals of Mathematical Statistics.* In its early volumes are to be found work of J. L. Doob (1910–), who wrote on probability from 1937 on, S. S. Wilks, Harold Hotelling, J. Neyman, the Polish mathematician who built up the Department of Statistics at Berkeley, and many others. For background on Doob, see Snell (1997).

William Feller (1906–1970), born in Yugoslavia, became one of the many prominent mathematicians who left Europe and went to the U.S.A. during the Nazi period. Originally an analyst, his interest in probability theory dates from his time in Kiel (1928–33); his published work on probability theory dates from 1935. An interesting obituary of Feller is given by Doob (1972). Here,

> 'in order to set Feller's work in context, the author is led to give a concise but penetrating history of the subject during the 20th century. Feller (1906–1970) made original and profound contributions to probability theory over a period (from 1935 to his death) during which it was transformed from a poor relation into a central branch of mathematics'.

9. The English School

The work of the French, Polish and Russian schools is pure probability and highly mathematical. This forms an interesting contrast to the work of the English school, whose roots were partly statistical, in the work of Pearson and others at the turn of the century, and partly philosophical.

John Maynard Keynes (1883–1946), later Lord Keynes, the economist, wrote *A Treatise on Probability* (Keynes, 1921) as a Fellowship Dissertation for King's College, Cambridge. This is a philosophical treatise, with little mathematical content. Keynes was a pupil of the Cambridge philosopher, W. E. Johnson. For a recent critique of the book, written from the point of view of a financial practitioner, see Bernstein (1996, Ch. 13).

R. A. Fisher (1890–1962) wrote on statistics from 1916 onwards, doing his pioneering work on likelihood, sufficiency, estimation of parameters, analysis of variance and design of experiments while at Rothamsted Experimental Station in Harpenden, Hertfordshire during the 1920s. Though generally regarded as the greatest of all statisticians, Fisher's work on probability, from 1930 onwards, was less successful. He wrote mainly on 'inverse probability' and the 'fiducial argument'. Zabell writes that 'fiducial inference stands as R. A. Fisher's one great failure' (Zabell, 1992), though Sprott questions this in his review (Sprott, 1994). For background on Fisher's extensive contributions to statistics and genetics, we refer to the biography by Box (1978), and to his obituary in *Biographical Memoirs of the Royal Society* Volume **9** in 1963.

The first successful textbook synthesis of Fisher's ideas, which essentially mark the birth of modern statistics, with Kolmogorov's, which mark the birth of modern probability, is Cramér (1946). The opening passage of the preface of this remarkable book is still worth quoting, and reads as follows.

> 'During the last 25 years, statistical science has made great progress, thanks to the brilliant schools of British and American statisticians, among whom the name of Professor R. A. Fisher should be mentioned in the foremost place. During the same time, largely owing to the work of French and Russian mathematicians, the classical calculus of probability has developed into a purely mathematical theory satisfying modern standards of rigour.
>
> The purpose of the present work is to join these two lines of development into an exposition of the mathematical theory of modern statistical methods, in so far as these are based on the concept of probability. A full understanding of the theory of these methods requires a fairly advanced knowledge of pure mathematics.'

Cramér proceeds to develop measure theory in his First Part and measure-theoretic probability in his Second Part, before turning to statistical inference in his Third Part.

Frank Plumpton Ramsey (1904–1930), in his posthumus book (Ramsey, 1931), laid the foundations for modern decision theory in statistics. Ramsey's work led to ideas of coherence in decision theory, and the principle 'minimise expected loss, maximise expected utility'. This viewpoint has been highly influential. For instance, Lindley (1971, § 4.8) contains a remarkable tribute, likening Ramsey's ideas to the 'Newton's laws of motion of the social sciences'.

Alan Matheson Turing (1912–1954) obtained a proof of the central limit theorem in 1934, again a Fellowship Dissertation at King's College, Cambridge, although he was anticipated by Lindeberg (1922); see Zabell (1995). Turing went on to head the statistical work by the cryptographers at Bletchley Park, who cracked the German military radio codes in World War II (Hodges, 1983), and to become one of the pioneers of the computer.

Sir Harold Jeffreys (1891–1989), originally a geophysicist, wrote on probability from 1919 onwards. In his book *Theory of Probability* (Jeffreys, 1939), Jeffreys comments on Keynes and Ramsey, and acknowledges his own inspiration, namely Pearson (1892). Despite its title, Jeffreys' book is more a book on Bayesian statistics than on probability theory. It is a difficult book to assess, as it is partly mathematical and partly not. It contains a violent attack on measure-theoretic probability and its practitioners; see e.g. p. 373. As an indication of its tone, the book fails to mention the greatest of all probabilists, Kolmogorov, but cites 154 other authors by name.

10. KOLMOGOROV'S GRUNDBEGRIFFE OF 1933

In his classic book *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Kolmogorov (1933) inaugurated the modern era in probability theory. Kolmogorov here successfully places probability theory on a rigorous axiomatic footing, harnessing the full power of measure theory, regarding a probability measure as a positive measure of mass one. In particular, Kolmogorov completes here the solution of Hilbert's Sixth Problem, on the axiomatisation of probability.

For a full appreciation of the *Grundbegriffe*, we must refer the reader to the obituary articles cited in § 5. Two highlights, however, deserve special mention here. One is the modern definition of conditioning, using σ-fields and the Radon–Nikodým theorem, due to Radon in 1913 in the Euclidean setting and to Nikodým in 1930 in the general case. This is, for example, the basis for the now-ubiquitous concept of a martingale. Williams

(1991, §§ 9.1, 9.2) rightly calls this 'the central definition of modern probability'. The second is the following final form of the Strong Law of Large Numbers.

THEOREM. *For $X, X_1, X_2, \ldots$ independent and identically distributed random variables, and $\mu \in \mathbb{R}$, the following are equivalent.*

(i) *The mean of X exists and is μ, that is $E|X| < \infty$, $EX = \mu$.*

(ii) *The sample mean converges to μ with probability one, i.e.*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to \mu$$

almost surely as $n \to \infty$.

This completes the line of work begun by Jakob Bernoulli (1654–1705) in his posthumous classic, *Ars Conjectandi* of 1713, i.e. Bernoulli's theorem, the weak law of large numbers for Bernoulli trials. This is the precise form of the folklore idea of the Law of Averages, and shows convincingly that the Kolmogorov axiomatics successfully capture the essence of probability.

An idea of the impact of the *Grundbegriffe* is conveyed by the speed with which its ideas permeated the literature. One sees this clearly in the passage from the preface to Cramér (1946) quoted above. Again, the Introduction to the classic book Doob (1953) on stochastic processes states baldly that 'probability is simply a branch of measure theory, with its own special emphasis and field of application, and no attempt has been made to sugar-coat this fact'.

Kolmogorov's approach in the *Grundbegriffe* took root so rapidly because it was an idea whose time had come. With the benefit of hindsight, it is perhaps surprising that this successful reformulation of probability theory in terms of measure theory did not happen much earlier. It is the main purpose of this study to explore the background, in the hope of casting some light on this crucial and interesting period.

It may be of some historical interest to note the actual stimulus that triggered the writing of the *Grundbegriffe*, namely that Kolmogorov's dacha needed a new roof, according to a personal communication from A. N. Shiryaev.

11. POSTSCRIPT: 'KOLMOGOROV MARK II'

The measure-theoretic approach to probability theory outlined above, 'Kolmogorov Mark I', has been extremely successful, and has become standard. It is a mark of Kolmogorov's supreme versatility and intellectual curiosity that he returned in the 1960s to foundational questions in probability theory, from quite another point of view.

Recall from § 7 that von Mises sought, unsuccessfully, a theory of random sequences sequence by sequence. The Kolmogorov axiomatics makes statements about the totality of all such sequences, and Kolmogorov saw that this left more to be said. If we replicate a thousand tosses of a fair coin, for instance, the sample space contains 2^{1000} points, each with equal probability 2^{-1000}, by symmetry. Classical probability theory does not distinguish one of these points from another, the one point containing 1000 zeros, say, from a particular 'typical' outcome. Yet these two differ dramatically from one viewpoint: it takes 1000 bits of information to encode a typical outcome, but two bits to encode 'all noughts'. Kolmogorov saw that the essence of randomness here is maximal complexity. He developed this approach in the 1960s, leading to the modern subject of algorithmic information theory. This approach, 'Kolmogorov Mark II', is the subject of Volume 3 of his *Selected Works* (Kolmogorov, 1987); see also Kolmogorov & Uspensky (1987) and Vovk (1987). The subject has been pursued by Gregory Chaitin, Jorma Rissanen and others. One may regard this work as, in some sense, bringing von Mises' programme to a successful conclusion.

ACKNOWLEDGEMENT

I thank Ulrich Krengel and the referees for their comments.

REFERENCES

BACHELIER, L. (1900). Théorie de la spéculation. *Ann. Sci. Ecole Norm. Sup.* **17**, 21–86.

BANACH, S. (1932). *Théorie des Opérations Linéaires.* Warsaw: Monografje Matematyczne.

BARONE, J. & NOVIKOFF, A. (1978). A history of the axiomatic formulation of probability from Borel to Kolmogorov: Part I. *Arch. Hist. Exact. Sci.* **18**, 123–90.

BERNSTEIN, P. (1996). *Against the Gods. The Remarkable Story of Risk.* New York: Wiley.

BERNSTEIN, S. N. (1911). *Probability Theory* (in Russian). St. Petersburg: Academy of Sciences (2nd ed. 1927, 3rd ed. 1934, 4th ed. 1946).

BINGHAM, N. H. (1989). The work of A. N. Kolmogorov on strong limit theorems. *Theory Prob. Applic.* **34**, 129–39.

BINGHAM, N. H. (1998). Fluctuations. *Math. Scientist* 323, 63–73.

BINGHAM, N. H. & KIESEL, R. (1998). *Risk-Neutral Valuation. Pricing and Hedging of Financial Derivatives.* London: Springer-Verlag.

BOHLMANN, G. (1908). Die Grundbegriffe der Wahrscheinlichkeitsrechnung in ihrer Anwendung auf die Lebensversicherung. *Atti del IV Congresso Int. Mat. Roma, III*, Sessione IIIb, 244–78.

BOOLE, G. (1854). *Investigation of The Laws of Thought* (reprinted 1916 as Volume II in his *Collected Logical Works.* Chicago: Open Court).

BOREL, E. (1898). *Leçons sur la Théorie des Fonctions.* Paris: Gauthier-Villars (2nd ed. 1914).

BOREL, E. (1909). Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo* **27**, 247–71.

BOX, J. F. (1978). *R. A. Fisher. The Life of a Scientist.* London: Wiley.

BRIESKORN, E. (Ed.) (1996). *Felix Hausdorff zum Gedächtnis. Band I.* Braunschweig: Friedrich Vieweg & Sohn.

BRUSH, S. G. (1968). A history of random processes. 1: Brownian movement from Brown to Perrin. *Arch. Hist. Exact Sci.* 15, 1–36.

CHUNG, K. L. (1987). Pólya's work in probability. *Bull. Lond. Math. Soc.* **19**, 570–6.

CIESELSKI, K. (1987). Lost legends of Lvov 1: the Scottish Café. *Math. Intelligencer* **9**, no. 4, 36–7.

CIESELSKI, K. (1988). Lost legends of Lvov 2: Banach's grave. *Math. Intelligencer* **10**, no. 1, 50–1.

CIFARELLI, D. M. & REGAZZINI, E. (1996). De Finetti's contributions to probability and statistics. *Statist. Sci.* **11**, 253–82.

CORRY, L. (1997). David Hilbert and the axiomatization of physics. *Arch. Hist. Exact Sci.* **51**, 83–198.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton: Princeton Univ. Press.

CZUBER, E. (1898). Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen. *Jahresb. Deutscher Mathematiker-Vereinigung* **7.**

CZUBER, E. (1903), *Wahrscheinlichkeitsrechnung und ihrer Anwendung auf Fehlerausgleichungsrechnung, Statistik und Lebensversicherung.*

DOOB, J. L. (1942). The Brownian movement and stochastic equations. *Ann. Math.* **43**, 351–69 (reprinted 1954 in *Selected Papers on Noise and Stochastic Processes*, Ed N. Wax. New York: Dover).

DOOB, J. L. (1953). *Stochastic Processes.* New York: Wiley.

DOOB, J. L. (1972). William Feller and twentieth century probability. *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, **2**, pp. xv–xx. Berkeley, CA: Univ. California Press (*Math. Rev.* **52** #7795).

DOYLE, P. G. & SNELL, J. L. (1984). *Random Walks and Electrical Networks*, Carus Math. Monographs **22**. Washington, DC: Math. Assoc. America.

EICHHORN, E. (1994). In memoriam Felix Hausdorff (1868–1942). Ein biographischer Versuch. *Vorlesungen zum Gedenken an Felix Hausdorf*, 1–88. *Berliner Studienreihe Math.* **5**. Berlin: Heldermann (Math. Rev. 95f:01030).

FINETTI, B. DE (1970). Teoria delle probabilità: sintesi introduttiva con appendice critica. Turin: Nuova Bibl. Sci. Einaudi, G. Einaudi Editore (trans. 1974, 1975 as *Theory of Probability: A Critical Introduction*, **1**, **2**. London: Wiley).

FITZPATRICK, P. J. (1960). Leading British statisticians of the nineteenth century. *J. Am. Statist. Assoc.* **55**, 38–70 (reprinted 1977 in *Studies in the History of Probability and Statistics*, **2**, Ed. Sir Maurice Kendall and R. L. Plackett. London: Charles Griffin).

FRÉCHET, M. (1906). Sur quelques points du calcul fonctionnel. *Red. Circ. Mat. Palermo* **22**, 1–74.

FRÉCHET, M. (1928). *Les Espaces Abstraits.* Paris: Gauthier-Villars.

GNEDENKO, B. V. & KOLMOGOROV, A. N. (1960). Obituary, Alexander Yakovlevich Khinchin. *Russian Math. Surveys* **15**, 93–106.

GOOD, I. J. (1978). Review of de Finetti (1970). *Math. Rev.* **55**, #13,514a,b.

HAAS-LORENTZ, G. L. DE (1913). *Die Brownsche Bewegung und einige verwandte Erscheinungen.* Braunschweig: Teubner.

HAUSDORFF, F. (1914). *Grundzüge der Mengenlehre*, Leipzig (3rd ed. 1935, Berlin, reprinted 1944, New York: Dover).

HEYDE, C. C. & SENETA, E. (1977). *I. J. Bienaymé. Statistical Theory Anticipated.* New York: Springer-Verlag.

HILBERT, D. (1902). Mathematical problems (Transl. from the German by M. W. Newson). *Bull. Am. Math. Soc.* (2) **8**, 437–9.

HODGES, A. (1983). *Alan Turing: The Enigma of Intelligence.* London: Counterpoint.

HOLGATE, P. (1997). Studies in the history of probability and statistics XLV. Independent functions: probability and analysis in Poland between the Wars. *Biometrika* **84**, 159–73.

JEFFREYS, H. (1939). *Theory of Probability.* Oxford: Oxford Univ. Press (3rd ed. 1983).

KAHANE, J.-P. (1997). A century of interplay between Taylor series, Fourier series and Brownian motion. *Bull. Lond. Math. Soc.* **29**, 257–79.

KAHANE, J.-P. & LEMARIÉ-RIEUSSET, P.-G. (1995). *Fourier Series and Wavelets*, Studies in the Development of Modern Mathematics **3**. Luxembourg: Gordon and Breach.

KENDALL, D. G. (1990). Obituary, Andrei Nikolaevich Kolmogorov (1903–1987). *Bull. Lond. Math. Soc.* **22**, 31–100.

KEYNES, J. M. (1921). *A Treatise on Probability.* London: Macmillan.

KOLMOGOROV, A. N. (1933). *Grundgebriffe der Wahrscheinlichkeitsrechnung.* Berlin: Springer-Verlag (English trans. 1950 *Foundations of the Theory of Probability.* New York: Chelsea. 2nd ed. 1974 in German and Russian).

KOLMOGOROV, A. N. (1987). *Theory of Information and Theory of Algorithms. Selected Works*, **3** (in Russian). Moscow: Nauka.

KOLMOGOROV, A. N. & SARMANOV, O. V. (1960). The work of S. N. Bernstein on probability theory (on his eightieth birthday). *Theory Prob. Appl.* **4**, 197–203.

KOLMOGOROV, A. N. & USPENSKY, V. A. (1987). Algorithms and randomness. *Theory Prob. Applic.* **32**, 389–412.

KRENGEL, U. (1990). Wahrscheinlichkeitstheorie. In *Ein Jahrhundert Mathematik, 1890–1990.* Festschrift zum Jubiläum der DMV, Ed. G. Fischer et al., pp. 457–89. Brauschweig/Wiesbaden: Fr. Vieweg.

KURATOWSKI, K. (1980). *A Half-Century of Polish Mathematics. Remembrances and Reflections.* Warsaw: PWN and Oxford: Pergamon Press.

LANGEVIN, P. (1908). Sur la théorie du mouvement brownien. *Comptes Rendus Acad. Sci. Paris* **146**, 530–3.

LEBESGUE, H. (1902). Intégrale, longueur, aire. *Annali di Mat.* (3) **7**, 231–59.

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

LÉVY, P. (1970). *Quelques Aspects de la Pensée d'un Mathématicien.* Paris: Gauthier-Villars.

LINDEBERG, J. W. (1922). Eine neue Herleitung des Exponentialgesetz in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15**, 211–25.

LINDLEY, D. V. (1971). *Taking Decisions.* London: Wiley (2nd ed. 1985).

LINDLEY, D. V. (1986). Obituary, Bruno de Finetti, 1906–1985. *J. R. Statist. Soc.* A **149**, 252.

LOÈVE, M. (1973). Obituary, Paul Lévy. *Ann. Prob.* **1**, 1–18.

MAGNELLO, M. E. (1998). Karl Pearson's mathematicization of inheritance: from ancestral heredity to Mendelian genetics (1895–1909). *Ann. Sci.* **55**, 35–94.

MARKOV, A. A. (1900). *Calculus of Probabilities* (in Russian). St. Petersburg: Academy of Sciences (2nd ed. 1908, trans. (1912) as *Wahrscheinlichkeitsrechnung.* Leipzig: B. G. Teubner.

MISES, R. VON (1928). *Wahrscheinlichkeit, Statistik und Wahrheit* (3rd ed. 1951; trans. 1939, reprinted 1981, as *Probability, Statistics and Truth*, New York: Dover).

MISES, R. VON (1947). On the asymptotic distribution of differentiable statistical functionals. *Ann. Math. Statist.* **18**, 309–48.

NOVIKOFF, A. & BARONE, J. (1977). The Borel Law of Large Numbers, the Borel Zero-One Law, and the work of Van Vleck. *Historia Mathematica* **4**, 43–65.

ORNSTEIN, L. S. & UHLENBECK, G. E. (1930). On the theory of Brownian motion. *Phys. Rev.* **36**, 823–41 (reprinted 1954 in *Selected Papers in Noise and Stochastic Processes*, Ed. N. Wax. New York: Dover).

PALEY, R. E. A. C. & WIENER, N. (1934). *Fourier Transforms in the Complex Domain.* Providence, RI: American Mathematical Society.

PALEY, R. E. A. C., WIENER, N. & ZYGMUND, A. (1933). Notes on random functions. *Math. Zeit.* **37**, 647–68.

PEARSON, K. (1892). *The Grammar of Science*. London: Walter Scott (2nd ed. 1900, 3rd ed. 1911).
POINCARÉ, H. (1896). *Calcul des Probabilités*. Paris: Gauthier-Villars (2nd ed. 1912).
PÓLYA, G. (1920). Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem. *Math. Z.* **8**, 171–81.
PÓLYA, G. (1921). Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz. *Math. Ann.* **84**, 149–60.
RAMSEY, F. P. (1931). *The Foundations of Mathematics, and Other Logical Essays*, Ed. R. B. Braithwaite. London: Kegan Paul.
REGAZZINI, E. (1987). Probability theory in Italy between the two world wars. *Metron* **44**, 5–42.
SCHUSTER, A. (1900). The periodogram of magnetic declination. *Trans. Camb. Phil. Soc.* **18**, 107–35.
SENETA, E. (1992). On the history of the strong law of large numbers and Boole's inequality. *Historia Mathematica* **19**, 24–39.
SHIRYAEV, A. N. (1986). Kolmogorov—life and creative activities. *Ann. Prob.* **17**, 866–944 (publications, 945–64).
SNELL, J. L. (1997). A conversation with Joe Doob. *Statist. Sci.* **12**, 301–11.
SPROTT, D. A. (1994). Review of Zabell (1992). *Math. Rev.* 94d:62006.
STEINHAUS, H. (1923). Les probabilités dénombrables et leur rapport à la théorie de la mesure. *Fund. Math.* **4**, 286–310.
STIGLER, S. M. (1986). *The History of Statistics. The Measurement of Uncertainty before 1900*. Harvard: Belknap, Harvard Univ. Press.
ULAM, S. (1976). *Adventures of a Mathematician*. New York: Charles Scribner's Sons.
VAART, A. W. VAN DER (1998). *Asymptotic Statistics*. Cambridge: Cambridge Univ. Press.
VAN VLECK, E. (1908). On non-measurable sets of points, with an example. *Trans. Am. Math. Soc.* **9**, 237–44.
VOVK, V. G. (1987). Law of the iterated logarithm for random Kolmogorov, or chaotic, sequences. *Theory Prob. Appl.* **32**, 413–25.
WIENER, N. (1930). Generalised harmonic analysis. *Acta Math.* **55**, 117–258 (reprinted, with 'Tauberian theorems', 1964, Cambridge, MA: MIT Press).
WIENER, N. (1933). *The Fourier Integral and Certain of its Applications*. Cambridge: Cambridge University Press (reissued 1988 Cambridge Math. Library).
WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge: Cambridge University Press.
ZABELL, S. L. (1992). R. A. Fisher and the fiducial argument. *Statist. Sci.* **7**, 369–87.
ZABELL, S. L. (1995). Alan Turing and the central limit theorem. *Am. Math. Mon.* **102**, 493–94.

[*Received May* 1999. *Revised August* 1999]