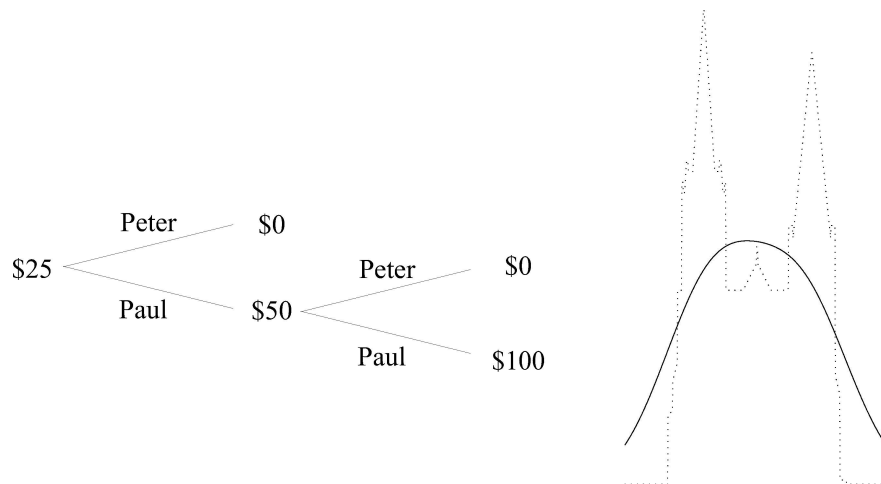# The origins and legacy of Kolmogorov's *Grundbegriffe*

Glenn Shafer
Rutgers School of Business
gshafer@andromeda.rutgers.edu

Vladimir Vovk
Royal Holloway, University of London
v.vovk@rhul.ac.uk

The Game-Theoretic Probability and Finance Project

Working Paper #4

First posted February 8, 2003. Last revised December 29, 2018.

Abstract

April 25, 2003, marked the 100th anniversary of the birth of Andrei Nikolaevich Kolmogorov, the twentieth century's foremost contributor to the mathematical and philosophical foundations of probability. The year 2003 was also the 70th anniversary of the publication of Kolmogorov's *Grundbegriffe der Wahrscheinlichkeitsrechnung.*

Kolmogorov's *Grundbegriffe* put probability's modern mathematical formalism in place. It also provided a philosophy of probability—an explanation of how the formalism can be connected to the world of experience. In this article, we examine the sources of these two aspects of the *Grundbegriffe*—the work of the earlier scholars whose ideas Kolmogorov synthesized.

Contents

1   Introduction

Andrei Kolmogorov's *Grundbegriffe der Wahrscheinlichkeitsrechnung*, which set out the axiomatic basis for modern probability theory, appeared in 1933. Four years later, in his opening address to an international colloquium at the University of Geneva, Maurice Fréchet praised Kolmogorov for organizing and expositing a theory that Émile Borel had created by adding countable additivity to classical probability. Fréchet put the matter this way in the written version of his address (1938b, p. 54):

> It was at the moment when Mr. Borel introduced this new kind of additivity into the calculus of probability—in 1909, that is to say—that all the elements needed to formulate explicitly the whole body of axioms of (modernized classical) probability theory came together.
>
> It is not enough to have all the ideas in mind, to recall them now and then; one must make sure that their totality is sufficient, bring them together explicitly, and take responsibility for saying that nothing further is needed in order to construct the theory.
>
> This is what Mr. Kolmogorov did. This is his achievement. (And we do not believe he wanted to claim any others, so far as the axiomatic theory is concerned.)

Perhaps not everyone in Fréchet's audience agreed that Borel had put everything on the table. But surely many saw the *Grundbegriffe* as a work of synthesis. In Kolmogorov's axioms, and in his way of relating his axioms to the world of experience, they must have seen traces of the work of many others—the work of Borel, yes, but also the work of Fréchet himself, and that of Cantelli, Chuprov, Lévy, Steinhaus, Ulam, and von Mises.

Today, what Fréchet and his contemporaries knew is no longer known. We know Kolmogorov and what came after; we have mostly forgotten what came before. This is the nature of intellectual progress. But it has left many modern students with the impression that Kolmogorov's axiomatization was born full grown—a sudden brilliant triumph over confusion and chaos.

In order to see both the innovation and the synthesis in the *Grundbegriffe*, we need a broad view of the foundations of probability and the advance of measure theory from 1900 to 1930. We need to understand how measure theory became more abstract during those decades, and we need to recall what others were saying about axioms for probability, about Cournot's principle, and about the relation of probability with measure and with frequency. Our review of these topics draws mainly on work by authors listed by Kolmogorov in the *Grundbegriffe*'s bibliography, especially Sergei Bernstein, Émile Borel, Francesco Cantelli, Maurice Fréchet, Paul Lévy, Antoni Łomnicki, Evgeny Slutsky, Hugo Steinhaus, and Richard von Mises. Others enter into our story along the way, but for the most part we do not review the contributions of authors whose foundational work does not seem to have influenced Kolmogorov. We say relatively little, for example, about Harold Jeffreys, John Maynard Keynes, Jan

Łukasiewicz, Paolo Medolaghi, and Frank P. Ramsey. For further information about foundational and mathematical work during the early twentieth century, see Urban (1923), Freudenthal and Steiner (1966), Regazzini (1987a,b), Krengel (1990), von Plato (1994), Benzi (1995), Holgate (1997), Hochkirchen (1999), Bingham (2000), and Bru (2003a).

We are interested not only in Kolmogorov's mathematical formalism, but also in his philosophy of probability—how he proposed to relate the mathematical formalism to the real world. In a 1939 letter to Fréchet, which we reproduce in §A.2, Kolmogorov wrote, "You are also right in attributing to me the opinion that the formal axiomatization should be accompanied by an analysis of its real meaning." Kolmogorov devoted only two pages of the *Grundbegriffe* to such an analysis. But the question was more important to him than this brevity might suggest. We can study any mathematical formalism we like, but we have the right to call it probability only if we can explain how it relates to the empirical phenomena classically treated by probability theory.

Kolmogorov's philosophy was frequentist. One way of understanding his frequentism would be to place it in a larger social and cultural context, emphasizing perhaps Kolmogorov's role as the leading new Soviet mathematician. We will not ignore this context, but we are more interested in using the thinking of Kolmogorov and his predecessors to inform our own understanding of probability. In 1963, Kolmogorov complained that his axioms had been so successful on the purely mathematical side that many mathematicians had lost interest in understanding how probability theory can be applied. This situation persists today. Now, more than ever, we need fresh thinking about how probability theory relates to the world, and there is no better starting point for this thinking than the works of Kolmogorov and his predecessors early in the twentieth century.

We begin by looking at the classical foundation that Kolmogorov's measure-theoretic foundation replaced: equally likely cases. In §2, we review how probability was defined in terms of equally likely cases, how the rules of the calculus of probability were derived from this definition, and how this calculus was related to the real world by Cournot's principle. We also look at some paradoxes discussed at the time.

In §3, we sketch the development of measure theory and its increasing entanglement with probability during the first three decades of the twentieth century. This story centers on Borel, who introduced countable additivity into pure mathematics in the 1890s and then brought it to the center of probability theory, as Fréchet noted, in 1909, when he first stated and more or less proved the strong law of large numbers for coin tossing. But it also features Lebesgue, Radon, Fréchet, Daniell, Wiener, Steinhaus, and Kolmogorov himself.

Inspired partly by Borel and partly by the challenge issued by Hilbert in 1900, a whole series of mathematicians proposed abstract frameworks for probability during the three decades we are emphasizing. In §4, we look at some of these, beginning with the doctoral dissertations by Rudolf Laemmel and Ugo Broggi in the first decade of the century and including an early contribution by Kolmogorov himself, written in 1927, five years before he started work on the *Grundbegriffe*.

In §5, we finally turn to the *Grundbegriffe* itself. Our review of it will confirm what Fréchet said in 1937 and what Kolmogorov himself says in the preface: it was a synthesis and a manual, not a report on new research. Like any textbook, its mathematics was novel for most of its readers. But its real originality was rhetorical and philosophical.

In §6 we discuss how the *Grundbegriffe* was received—at the time and in the following decades. Its mathematical framework for probability, as we know, came to be regarded as fundamental, but its philosophical grounding for probability was largely ignored.

2 The classical foundation

The classical foundation of probability theory, which began with the notion of equally likely cases, held sway for two hundred years. Its elements were put in place by Jacob Bernoulli and Abraham De Moivre early in the eighteenth century, and they remained in place in the early twentieth century. Even today the classical foundation is used in teaching probability.

Although twentieth-century proponents of new approaches were fond of deriding the classical foundation as naive or circular, it can be defended. Its basic mathematics can be explained in a few words, and it can be related to the real world by Cournot's principle, the principle that an event with small or zero probability will not occur. This principle was advocated in France and Russia in the early years of the twentieth century but disputed in Germany. Kolmogorov adopted it in the *Grundbegriffe*.

In this section we review the mathematics of equally likely cases and recount the discussion of Cournot's principle, contrasting the support for it in France with German efforts to replace it with other ways of relating equally likely cases to the real world. We also discuss two paradoxes contrived at the end of the nineteenth century by Joseph Bertrand, which illustrate the care that must be taken with the concept of relative probability. The lack of consensus on how to make philosophical sense of equally likely cases and the confusion engendered by Bertrand's paradoxes were two sources of dissatisfaction with the classical theory.

2.1 The classical calculus

The classical definition of probability was formulated by Jacob Bernoulli in *Ars Conjectandi* (1713) and Abraham De Moivre in *The Doctrine of Chances* (1718): the probability of an event is the ratio of the number of equally likely cases that favor it to the total number of equally likely cases possible under the circumstances.

From this definition, De Moivre derived two rules for probability. The *theorem of total probability*, or the *addition theorem*, says that if A and B cannot

both happen, then

$$\text{probability of } A \text{ or } B \text{ happening}$$
$$= \frac{\# \text{ of cases favoring } A \text{ or } B}{\text{total } \# \text{ of cases}}$$
$$= \frac{\# \text{ of cases favoring } A}{\text{total } \# \text{ of cases}} + \frac{\# \text{ of cases favoring } B}{\text{total } \# \text{ of cases}}$$
$$= (\text{probability of } A) + (\text{probability of } B).$$

The *theorem of compound probability*, or the *multiplication theorem*, says

$$\text{probability of both } A \text{ and } B \text{ happening}$$
$$= \frac{\# \text{ of cases favoring both } A \text{ and } B}{\text{total } \# \text{ of cases}}$$
$$= \frac{\# \text{ of cases favoring } A}{\text{total } \# \text{ of cases}} \times \frac{\# \text{ of cases favoring both } A \text{ and } B}{\# \text{ of cases favoring } A}$$
$$= (\text{probability of } A) \times (\text{probability of } B \text{ if } A \text{ happens}).$$

These arguments were still standard fare in probability textbooks at the beginning of the twentieth century, including the great treatises by Henri Poincaré (1896) in France, Andrei Markov (1900) in Russia, and Emanuel Czuber (1903) in Germany. Some years later we find them in Guido Castelnuovo's Italian textbook (1919), which has been held out as the acme of the genre (Onicescu 1967).

Only the British held themselves aloof from the classical theory, which they attributed to Laplace and found excessively apriorist. The British style of introducing probability, going back to Augustus De Morgan (1838, 1847), emphasized combinatorics without dwelling on formalities such as the rules of total and compound probability, and the British statisticians preferred to pass over the combinatorics as quickly as possible, so as to get on with the study of "errors of observations" as in Airy (1861). According to his son Egon (1990, pp. 13, 71), Karl Pearson recommended to his students nothing more theoretical than books by De Morgan, William Whitworth (1878), and the Danish statistician Harald Westergaard (1890, in German).

The classical theory, which had begun in England with De Moivre, returned to the English language in an influential book published in New York by the actuary Arne Fisher (1915), who had immigrated to the United States from Denmark at the age of 16. Fisher's book played an important role in bringing the methods of the great Scandinavian mathematical statisticians, Jorgen Gram, Thorvald Thiele, Harald Westergaard, and Carl Charlier, into the English-speaking world (Molina 1944, Lauritzen 2002). The Scandinavians stood between the very empirical British statisticians on the one hand and the French, German, and Russian probabilists on the other; they were serious about advancing mathematical statistics beyond where Laplace had left it, but they valued the classical foundation for probability.

After Fisher's book appeared, Americans adopted the classical rules but looked for ways to avoid the classical arguments based on equally likely cases; the two most notable American probability textbooks of the 1920s, by Julian Lowell Coolidge (1925) at Harvard and Thornton C. Fry (1928) at Bell Telephone Laboratories, replaced the classical arguments for the rules of probability with arguments based on the assumption that probabilities are limiting frequencies. But this approach did not endure, and the only American probability textbook from before the second world war that remained in print in the second half of the century was the more classical one by the Petersburg-educated Stanford professor James V. Uspensky (1937).

2.1.1 Geometric probability

Geometric probability was incorporated into the classical theory in the early nineteenth century. Instead of counting equally likely cases, one measures their geometric extension—their area or volume. But probability is still a ratio, and the rules of total and compound probability are still theorems. This is explained clearly by Antoine-Augustin Cournot in his influential treatise on probability and statistics, *Exposition de la théorie des chances et des probabilités*, published in 1843 (p. 29). In his commentary in Volume XI of Cournot's *Œuvres complètes*, Bernard Bru traces the idea back to the mathematician Joseph Fourier and the naturalist George-Louis Leclerc de Buffon.

This understanding of geometric probability did not change in the early twentieth century, when Borel and Lebesgue expanded the class of sets for which we can define geometric extension. We may now have more events with which to work, but we define and study geometric probabilities in the same way as before. Cournot would have seen nothing novel in Felix Hausdorff's definition of probability in the chapter on measure theory in his 1914 treatise on set theory (pp. 416–417).

2.1.2 Relative probability

The classical calculus was enriched at the beginning of the twentieth century by a formal and universal notation for relative probabilities. In 1901, Hausdorff introduced the symbol $p_F(E)$ for what he called the *relative Wahrscheinlichkeit von E, posito F* (relative probability of E given F). Hausdorff explained that this notation can be used for any two events E and F, no matter what their temporal or logical relationship, and that it allows one to streamline Poincaré's proofs of the addition and multiplication theorems. At least two other authors, Charles Saunders Peirce (1867, 1878) and Hugh MacColl (1880, 1897), had previously proposed universal notations for the probability of one event given another. But Hausdorff's notation was adopted by the influential textbook author Emanuel Czuber (1903). Kolmogorov used it in the *Grundbegriffe*, and it persisted, especially in the German literature, until the middle of the twentieth century, when it was displaced by the more flexible $P(E \mid F)$, which Harold

Jeffreys had introduced in his *Scientific Inference* (1931).[1]

Although Hausdorff's relative probability resembles today's conditional probability, other classical authors used "relative" in other ways. For Sylvestre-François Lacroix (1822, p. 20) and Jean-Baptiste-Joseph Liagre (1879, p. 45), the probability of E relative to an incompatible event F was $P(E)/(P(E) + P(F))$. For Borel (1914, pp. 58–59), the relative probability of E was $P(E)/P(\text{not} E)$. Classical authors could use the phrase however they liked, because it did not mark a sharp distinction like the modern distinction between absolute and conditional probability. Nowadays some authors write the rule of compound probability in the form

$$P(A \,\&\, B) = P(A)P(B \mid A)$$

and regard the conditional probability $P(B \mid A)$ as fundamentally different in kind from the absolute probabilities $P(A \,\&\, B)$ and $P(A)$. But for the classical authors, every probability was evaluated in a particular situation, with respect to the equally likely cases in that situation. When these authors wrote about the probability of B "after A has happened" or "when A is known", these phrases merely identified the situation; they did not indicate that a different kind of probability was being considered.

Before the *Grundbegriffe*, it was unusual to call a probability or expected value "conditional" rather than "relative", but the term does appear. George Boole may have been the first to use it, though only casually. In his *Laws of Thought*, in 1854, Boole calls an event considered under a certain condition a conditional event, and he discusses the probabilities of conditional events. Once (p. 261), and perhaps only once, he abbreviates this to "conditional probabilities". In 1887, in his *Metretike*, Francis Edgeworth, citing Boole, systematically called the probability of an effect given a cause a "conditional probability" (Mirowski 1994, p. 82). The Petersburg statistician Aleksandr Aleksandrovich Chuprov, who was familiar with Edgeworth's work, used the Russian equivalent (условная вероятность) in his 1910 book (p. 151). A few years later, in 1917, the German equivalent of "conditional expectation" (bedingte mathematische Erwartung) appeared in a book by Chuprov's friend Ladislaus von Bortkiewicz, professor of statistics in Berlin. We see "conditional probability" again in English in 1928, in Fry's textbook (p. 43).

We should also note that different authors used the term "compound probability" ("probabilité composée" in French) in different ways. Some authors (e.g., Poincaré 1912, p. 39) seem to have reserved it for the case where the two events are independent; others (e.g., Bertrand 1889, p. 3) used it in the general case as well.

---

[1]See the historical discussion in Jeffreys's *Theory of Probability* (1939), on p. 25 of the first or third editions or p. 26 of the second edition. Among the early adopters of Jeffreys's vertical stroke were Jerzy Neyman, who used P{A | B} in 1937, and Valery Glivenko, who used $\boldsymbol{P}(A/B)$ (with only a slight tilt) in 1939.

2.2   Cournot's principle

An event with very small probability is *morally impossible*; it will not happen. Equivalently, an event with very high probability is *morally certain*; it will happen. This principle was first formulated within mathematical probability by Jacob Bernoulli. In his *Ars Conjectandi*, published in 1713, Bernoulli proved a celebrated theorem: in a sufficiently long sequence of independent trials of an event, there is a very high probability that the frequency with which the event happens will be close to its probability. Bernoulli explained that we can treat the very high probability as moral certainty and so use the frequency of the event as an estimate of its probability. This conclusion was later called the law of large numbers.

Probabilistic moral certainty was widely discussed in the eighteenth century. In the 1760s, the French savant Jean d'Alembert muddled matters by questioning whether the prototypical event of very small probability, a long run of many happenings of an event as likely to fail as happen on each trial, is possible at all. A run of a hundred may be metaphysically possible, he felt, but it is physically impossible. It has never happened and never will happen (d'Alembert 1761, 1767; Daston 1979). In 1777, Buffon argued that the distinction between moral and physical certainty was one of degree. An event with probability 9999/10000 is morally certain;. an event with much greater probability, such as the rising of the sun, is physically certain (Loveland 2001).

Cournot, a mathematician now remembered as an economist and a philosopher of science (Martin 1996, 1998), gave the discussion a nineteenth-century cast in his 1843 treatise. Being equipped with the idea of geometric probability, Cournot could talk about probabilities that are vanishingly small. He brought physics to the foreground. It may be mathematically possible, he argued, for a heavy cone to stand in equilibrium on its vertex, but it is physically impossible. The event's probability is vanishingly small. Similarly, it is physically impossible for the frequency of an event in a long sequence of trials to differ substantially from the event's probability (1843, pp. 57, 106).

In the second half of the nineteenth century, the principle that an event with a vanishingly small probability will not happen took on a real role in physics, most saliently in Ludwig Boltzmann's statistical understanding of the second law of thermodynamics. As Boltzmann explained in the 1870s, dissipative processes are irreversible because the probability of a state with entropy far from the maximum is vanishingly small (von Plato 1994, p. 80; Seneta 1997). Also notable was Henri Poincaré's use of the principle in the three-body problem. Poincaré's recurrence theorem, published in 1890, says that an isolated mechanical system confined to a bounded region of its phase space will eventually return arbitrarily close to its initial state, provided only that this initial state is not exceptional. Within any region of finite volume, the states for which the recurrence does not hold are exceptional inasmuch as they are contained in subregions whose total volume is arbitrarily small.

Saying that an event of very small or vanishingly small probability will not happen is one thing. Saying that probability theory gains empirical meaning

only by ruling out the happening of such events is another. Cournot seems to have been the first to say explicitly that probability theory does gain empirical meaning only by declaring events of vanishingly small probability to be impossible:

> ...*The physically impossible event is therefore the one that has infinitely small probability*, and only this remark gives substance— objective and phenomenal value—to the theory of mathematical probability (1843 p. 78).[2]

After the second world war (see §6.2.2), some authors began to use "Cournot's principle" for the principle that an event of very small or zero probability singled out in advance will not happen, especially when this principle is advanced as the means by which a probability model is given empirical meaning.

2.2.1 The viewpoint of the French probabilists

In the early decades of the twentieth century, probability theory was beginning to be understood as pure mathematics. What does this pure mathematics have to do with the real world? The mathematicians who revived research in probability theory in France during these decades, Émile Borel, Jacques Hadamard, Maurice Fréchet, and Paul Lévy, made the connection by treating events of small or zero probability as impossible.

Borel explained this repeatedly, often in a style more literary than mathematical or philosophical (Borel 1906, 1909b, 1914, 1930). According to Borel, a result of the probability calculus deserves to be called objective when its probability becomes so great as to be practically the same as certainty. His many discussions of the considerations that go into assessing the boundaries of practical certainty culminated in a classification more refined than Buffon's. A probability of 10^{-6}, he decided, is negligible at the human scale, a probability of 10^{-15} at the terrestrial scale, and a probability of 10^{-50} at the cosmic scale (Borel 1939, pp. 6–7).

Hadamard, the preeminent analyst who did pathbreaking work on Markov chains in the 1920s (Bru 2003a), made the point in a different way. Probability theory, he said, is based on two basic notions: the notion of perfectly equivalent (equally likely) events and the notion of a very unlikely event (Hadamard 1922, p. 289). Perfect equivalence is a mathematical assumption, which cannot be verified. In practice, equivalence is not perfect—one of the grains in a cup of sand may be more likely than another to hit the ground first when they are thrown out of the cup. But this need not prevent us from applying the principle of the very unlikely event. Even if the grains are not exactly the same, the probability of any particular one hitting the ground first is negligibly small. Hadamard cited Poincaré's work on the three-body problem in this connection, because Poincaré's conclusion is insensitive to how one defines the probabilities for the initial state. Hadamard was the teacher of both Fréchet and Lévy.

---

[2]The phrase "objective and phenomenal" refers to Kant's distinction between the noumenon, or thing-in-itself, and the phenomenon, or object of experience (Daston 1994).

It was Lévy, perhaps, who had the strongest sense of probability's being pure mathematics (he devoted most of his career as a mathematician to probability), and it was he who expressed most clearly in the 1920s the thesis that Cournot's principle is probability's only bridge to reality. In his *Calcul des probabilités* Lévy emphasized the different roles of Hadamard's two basic notions. The notion of equally likely events, Lévy explained, suffices as a foundation for the mathematics of probability, but so long as we base our reasoning only on this notion, our probabilities are merely subjective. It is the notion of a very unlikely event that permits the results of the mathematical theory to take on practical significance (Lévy 1925, pp. 21, 34; see also Lévy 1937, p. 3). Combining the notion of a very unlikely event with Bernoulli's theorem, we obtain the notion of the objective probability of an event, a physical constant that is measured by relative frequency. Objective probability, in Lévy's view, is entirely analogous to length and weight, other physical constants whose empirical meaning is also defined by methods established for measuring them to a reasonable approximation (Lévy 1925, pp. 29–30).

By the time he undertook to write the *Grundbegriffe*, Kolmogorov must have been very familiar with Lévy's views. He had cited Lévy's 1925 book in his 1931 article on Markov processes and subsequently, during his visit to France, had spent a great deal of time talking with Lévy about probability. But he would also have learned about Cournot's principle from the Russian literature. The champion of the principle in Russia had been Chuprov, who became professor of statistics in Petersburg in 1910. Like the Scandinavians, Chuprov wanted to bridge the gap between the British statisticians and the continental mathematicians (Sheynin 1996, Seneta 2001). He put Cournot's principle—which he called "Cournot's lemma"—at the heart of this project; it was, he said, a basic principle of the logic of the probable (Chuprov 1910, Sheynin 1996, pp. 95–96). Markov, Chuprov's neighbor in Petersburg, learned about the burgeoning field of mathematical statistics from Chuprov (Ondar 1981), and we see an echo of Cournot's principle in Markov's textbook (1912, p. 12 of the German edition):

> The closer the probability of an event is to one, the more reason we have to expect the event to happen and not to expect its opposite to happen.
>
> In practical questions, we are forced to regard as certain events whose probability comes more or less close to one, and to regard as impossible events whose probability is small.
>
> Consequently, one of the most important tasks of probability theory is to identify those events whose probabilities come close to one or zero.

The Russian statistician Evgeny Slutsky discussed Chuprov's views in his influential article on limit theorems, published in German in 1925. Kolmogorov included Lévy's book and Slutsky's article in his bibliography, but not Chuprov's book. An opponent of the Bolsheviks, Chuprov was abroad when they

9

seized power, and he never returned home. He remained active in Sweden and Germany, but his health soon failed, and he died in 1926, at the age of 52.

2.2.2 Strong and weak forms of Cournot's principle

Cournot's principle has many variations. Like probability, moral certainty can be subjective or objective. Some authors make moral certainty sound truly equivalent to absolute certainty; others emphasize its pragmatic meaning.

For our story, it is important to distinguish between the strong and weak forms of the principle (Fréchet 1951, p. 6; Martin 2003). The strong form refers to an event of small or zero probability that we single out in advance of a single trial: it says the event will not happen on that trial. The weak form says that an event with very small probability will happen very rarely in repeated trials.

Borel, Lévy, and Kolmogorov all enunciated Cournot's principle in its strong form. In this form, the principle combines with Bernoulli's theorem to produce the unequivocal conclusion that an event's probability will be approximated by its frequency in a particular sufficiently long sequence of independent trials. It also provides a direct foundation for statistical testing. If the empirical meaning of probability resides precisely in the non-happening of small-probability events singled out in advance, then we need no additional principles to justify rejecting a hypothesis that gives small probability to an event we single out in advance and then observe to happen (Bru 1999).

Other authors, including Chuprov, enunciated Cournot's principle in its weak form, and this can lead in a different direction. The weak principle combines with Bernoulli's theorem to produce the conclusion that an event's probability will *usually* be approximated by its frequency in a sufficiently long sequence of independent trials, a general principle that has the weak principle as a special case. This was pointed out by Castelnuovo in his 1919 textbook (p. 108). Castelnuovo called the general principle the *empirical law of chance* (la legge empirica del caso):

> In a series of trials repeated a large number of times under identical conditions, each of the possible events happens with a (relative) frequency that gradually equals its probability. The approximation usually improves with the number of trials. (Castelnuovo 1919, p. 3)

Although the special case where the probability is close to one is sufficient to imply the general principle, Castelnuovo preferred to begin his introduction to the meaning of probability by enunciating the general principle, and so he can be considered a frequentist. His approach was influential at the time. Maurice Fréchet and Maurice Halbwachs adopted it in their textbook in 1924. It brought Fréchet to the same understanding of objective probability as Lévy: it is a physical constant that is measured by relative frequency (1938a, p. 5; 1938b, pp. 45–46).

The weak point of Castelnuovo and Fréchet's position lies in the modesty of their conclusion: they conclude only that an event's probability is *usually* approximated by its frequency. When we estimate a probability from an observed

frequency, we are taking a further step: we are assuming that what usually happens has happened in the particular case. This step requires the strong form of Cournot's principle. According to Kolmogorov (1956, p. 240 of the 1965 English edition), it is a reasonable step only if "we have some reason for assuming" that the position of the particular case among other potential ones "is a regular one, that is, that it has no special features".

2.2.3 British indifference and German skepticism

The mathematicians who worked on probability in France in the early twentieth century were unusual in the extent to which they delved into the philosophical side of their subject. Poincaré had made a mark in the philosophy of science as well as in mathematics, and Borel, Fréchet, and Lévy tried to emulate him. The situation in Britain and Germany was different.

In Britain there was little mathematical work in probability proper in this period. In the nineteenth century, British interest in probability had been practical and philosophical, not mathematical (Porter 1986, p. 74ff). British empiricists such as Robert Leslie Ellis (1849) and John Venn (1888) accepted the usefulness of probability but insisted on defining it directly in terms of frequency, leaving little meaning or role for the law of large numbers and Cournot's principle (Daston 1994). These attitudes, as we noted in §2.1, persisted even after Pearson and Fisher had brought Britain into a leadership role in mathematical statistics. The British statisticians had little interest in mathematical probability theory and hence no puzzle to solve concerning how to link it to the real world. They were interested in reasoning directly about frequencies.

In contrast with Britain, Germany did see a substantial amount of mathematical work in probability during the first decades of the twentieth century, much of it published in German by Scandinavians and eastern Europeans. But few German mathematicians of the first rank fancied themselves philosophers. The Germans were already pioneering the division of labor to which we are now accustomed, between mathematicians who prove theorems about probability and philosophers, logicians, statisticians, and scientists who analyze the meaning of probability. Many German statisticians believed that one must decide what level of probability will count as practical certainty in order to apply probability theory (von Bortkiewicz 1901, p. 825; Bohlmann 1901, p. 861), but German philosophers did not give Cournot's principle a central role.

The most cogent and influential of the German philosophers who discussed probability in the late nineteenth century was Johannes von Kries, whose *Principien der Wahrscheinlichkeitsrechnung* first appeared in 1886. Von Kries rejected what he called the orthodox philosophy of Laplace and the mathematicians who followed him. As von Kries saw it, these mathematicians began with a subjective concept of probability but then claimed to establish the existence of objective probabilities by means of a so-called law of large numbers, which they erroneously derived by combining Bernoulli's theorem with the belief that small probabilities can be neglected. Having both subjective and objective probabilities at their disposal, these mathematicians then used Bayes's theorem to

reason about objective probabilities for almost any question where many observations are available. All this, von Kries believed, was nonsense. The notion that an event with very small probability is impossible was, in von Kries's eyes, simply d'Alembert's mistake.

Von Kries believed that objective probabilities sometimes exist, but only under conditions where equally likely cases can legitimately be identified. Two conditions, he thought, are needed:

- Each case is produced by equally many of the possible arrangements of the circumstances, and this remains true when we look back in time to earlier circumstances that led to the current ones. In this sense, the relative sizes of the cases are *natural*.

- Nothing besides these circumstances affects our expectation about the cases. In this sense, the Spielräume[3] are *insensitive*.

Von Kries's *principle of the Spielräume* was that objective probabilities can be calculated from equally likely cases when these conditions are satisfied. He considered this principle analogous to Kant's principle that everything that exists has a cause. Kant thought that we cannot reason at all without the principle of cause and effect. Von Kries thought that we cannot reason about objective probabilities without the principle of the Spielräume.

Even when an event has an objective probability, von Kries saw no legitimacy in the law of large numbers. Bernoulli's theorem is valid, he thought, but it tells us only that a large deviation of an event's frequency from its probability is just as unlikely as some other unlikely event, say a long run of successes. What will actually happen is another matter. This disagreement between Cournot and von Kries can be seen as a quibble about words. Do we say that an event will not happen (Cournot), or do we say merely that it is as unlikely as some other event we do not expect to happen (von Kries)? Either way, we proceed as if it will not happen. But the quibbling has its reasons. Cournot wanted to make a definite prediction, because this provides a bridge from probability theory to the world of phenomena—the real world, as those who have not studied Kant would say. Von Kries thought he had a different way of connecting probability theory with phenomena.

Von Kries's critique of moral certainty and the law of large numbers was widely accepted in Germany (Kamlah, 1983). Czuber, in the influential textbook we have already mentioned, named Bernoulli, d'Alembert, Buffon, and De Morgan as advocates of moral certainty and declared them all wrong; the concept of moral certainty, he said, violates the fundamental insight that an event of ever so small a probability can still happen (Czuber 1903, p. 15; see also Meinong 1915, p. 591).

---

[3]In German, Spiel means "game" or "play", and Raum (plural Räume) means "room" or "space". In most contexts, Spielraum can be translated as "leeway" or "room for maneuver". For von Kries, the Spielraum for each case was the set of all arrangements of the circumstances that produce it.

This wariness about ruling out the happening of events whose probability is merely very small does not seem to have prevented acceptance of the idea that zero probability represents impossibility. Beginning with Wiman's work on continued fractions in 1900, mathematicians writing in German had worked on showing that various sets have measure zero, and everyone understood that the point was to show that these sets are impossible (see Felix Bernstein 1912, p. 419). This suggests a great gulf between zero probability and merely small probability. One does not sense such a gulf in the writings of Borel and his French colleagues; as we have seen, the vanishingly small, for them, was merely an idealization of the very small.

Von Kries's principle of the Spielräume did not endure, for no one knew how to use it. But his project of providing a Kantian justification for the uniform distribution of probabilities remained alive in German philosophy in the first decades of the twentieth century (Meinong 1915; Reichenbach 1916). John Maynard Keynes (1921) brought it into the English literature, where it continues to echo, to the extent that today's probabilists, when asked about the philosophical grounding of the classical theory of probability, are more likely to think about arguments for a uniform distribution of probabilities than about Cournot's principle.

2.3 Bertrand's paradoxes

How do we know cases are equally likely, and when something happens, do the cases that remain possible remain equally likely? In the decades before the *Grundbegriffe*, these questions were frequently discussed in the context of paradoxes formulated by Joseph Bertrand, an influential French mathematician, in a textbook that he published in 1889 after teaching probability for many decades (Bru and Jongmans 2001).

We now look at discussions by other authors of two of Bertrand's paradoxes: Poincaré's discussion of the paradox of the three jewelry boxes, and Borel's discussion of the paradox of the great circle.[4] The latter was also discussed by Kolmogorov and is now sometimes called the "Borel-Kolmogorov paradox".

2.3.1 The paradox of the three jewelry boxes

This paradox, laid out by Bertrand on pp. 2–3 of his textbook, involves three identical jewelry boxes, each with two drawers. Box A has gold medals in both drawers, Box B has silver medals in both, and Box C has a gold medal in one and a silver medal in the other. Suppose we choose a box at random. It will be Box C with probability 1/3. Now suppose we open at random one of the drawers in the box we have chosen. There are two possibilities for what we find:

---

[4]In the literature of the period, "Bertrand's paradox" usually referred to a third paradox, concerning two possible interpretations of the idea of choosing a random chord on a circle. Determining a chord by choosing two random points on the circumference is not the same as determining it by choosing a random distance from the center and then a random orientation.

- We find a gold medal. In this case, only two possibilities remain: the other drawer has a gold medal (we have chosen Box A), or the other drawer has a silver medal (we have chosen Box C).

- We find a silver medal. Here also, only two possibilities remain: the other drawer has a gold medal (we have chosen Box C), or the other drawer has a silver medal (we have chosen Box B).

Either way, it seems, there are now two cases, one of which is that we have chosen Box C. So the probability that we have chosen Box C is now 1/2.

Bertrand himself did not accept the conclusion that opening the drawer would change the probability of having Box C from 1/3 to 1/2, and Poincaré gave an explanation (1912, pp. 26–27). Suppose the drawers in each box are labeled (where we cannot see) α and β, and suppose the gold medal in Box C is in drawer α. Then there are six equally likely cases for the drawer we open:

1. Box A, Drawer α: gold medal.

2. Box A, Drawer β: gold medal.

3. Box B, Drawer α: silver medal.

4. Box B, Drawer β: silver medal.

5. Box C, Drawer α: gold medal.

6. Box C, Drawer β: silver medal.

When we find a gold medal, say, in the drawer we have opened, three of these cases remain possible: case 1, case 2, and case 5. Of the three, only one favors our having our hands on Box C. So the probability for Box C is still 1/3.

2.3.2   The paradox of the great circle

This paradox, on pp. 6–7 of Bertrand's textbook, begins with a simple question: if we choose at random two points on the surface of a sphere, what is the probability that the distance between them is less than $10'$?

By symmetry, we can suppose that the first point is known. So one way of answering the question is to calculate the proportion of a sphere's surface that lies within $10'$ of a given point. This is 2.1×10^{-6}.[5] Bertrand also found a different answer using a different method. After fixing the first point, he said, we can also assume that we know the great circle that connects the two points, because the possible chances are the same on great circles through the first

---

[5]The formula Bertrand gives is correct, and it evaluates to this number. Unfortunately, he then gives a numerical value that is twice as large, as if the denominator of the ratio being calculated were the area of a hemisphere rather than the area of the entire sphere. (Later in the book, on p. 169, he considers a version of the problem where the point is drawn at random from a hemisphere rather than from a sphere.) Bertrand composed his book by drawing together notes from decades of teaching, and the carelessness with which he did this may have enhanced the sense of confusion that his paradoxes engendered.
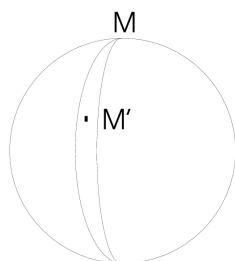
Borel's Figure 13.

point. There are 360 degrees—2160 arcs of $10'$ each—in this great circle. Only the points in the two neighboring arcs are within $10'$ of the first point, and so the probability sought is $2/2160$, or 9.3×10^{-4}. This is many times larger than the first answer. Bertrand suggested that both answers were equally valid, the original question being ill posed. The concept of choosing points at random on a sphere was not, he said, sufficiently precise.

In his own probability textbook, published in 1909 (pp. 100–104), Borel explained that Bertrand was mistaken. Bertrand's first answer, obtained by assuming that equal areas on the sphere have equal chances of containing the second point, is correct. His second answer, obtained by assuming that equal arcs on a great circle have equal chances of containing it, is incorrect. Writing M and M′ for the two points to be chosen at random on the sphere, Borel explained Bertrand's mistake as follows:

> . . . The error begins when, after fixing the point M and the great circle, one assumes that the probability of M′ being on a given arc of the great circle is proportional to the length of that arc. If the arcs have no width, then in order to speak rigorously, we must assign the value zero to the probability that M and M′ are on the circle. In order to avoid this factor of zero, which makes any calculation impossible, one must consider a thin bundle of great circles all going through M, and then it is obvious that there is a greater probability for M′ to be situated in a vicinity 90 degrees from M than in the vicinity of M itself (*fig. 13*).

To give this argument practical content, Borel discussed how one might measure the longitude of a point on the surface of the earth. If we use astronomical observations, then we are measuring an angle, and errors in the measurement of the angle correspond to wider distances on the ground at the equator than at the poles. If we instead use geodesic measurements, say with a line of markers on each of many meridians, then in order to keep the markers out of each other's way, we must make them thinner and thinner as we approach the poles.

2.3.3   Appraisal

Poincaré, Borel, and others who understood the principles of the classical theory were able to resolve the paradoxes that Bertrand contrived. Two principles emerge from the resolutions they offered:

- The equally likely cases must be detailed enough to represent new information (e.g., we find a gold medal) in all relevant detail. The remaining equally likely cases will then remain equally likely.

- We may need to consider the real observed event of non-zero probability that is represented in an idealized way by an event of zero probability (e.g., a randomly chosen point falls on a particular meridian). We should pass to the limit only after absorbing the new information.

Not everyone found it easy to apply these principles, however, and the confusion surrounding the paradoxes was another source of dissatisfaction with the classical theory.

Modern theories have tried to solve the problem by representing explicitly the possibilities for new information. This Kolmogorov did this using a partition (see p. 45 below). Other authors have used filtrations (Doob 1953), event trees (Shafer 1996) and game protocols (Shafer and Vovk 2001). These devices may be helpful, but they have not put the paradoxers out of business. Puzzles like Bertrand's paradox of the three jewelry boxes still flourish (Bar-Hillel and Falk 1982, Shafer 1985, Barbeau 1993, Halpern and Tuttle 1993).

3   Measure-theoretic probability before the *Grundbegriffe*

A discussion of the relation between measure and probability in the first decades of the twentieth century must navigate many pitfalls, for measure theory itself evolved, beginning as a theory about the measurability of sets of real numbers and then becoming more general and abstract. Probability theory followed along, but since the meaning of *measure* was changing, we can easily misunderstand things said at the time about the relation between the two theories.

The development of theories of measure and integration during the late nineteenth and early twentieth centuries has been studied extensively (Hawkins 1975, Pier 1994a). Here we offer only a bare-bones sketch, beginning with Borel and Lebesgue (§3.1) and touching on those steps that proved most significant for the foundations of probability. We discuss the work of Carathéodory, Radon, Fréchet, and Nikodym, who made measure primary and the integral secondary (§3.2), as well as the contrasting approach of Daniell, who took integration to be basic (§3.4).

We dwell more on the relation of measure and integration to probability. We discuss perceptions of the relevance of Fréchet's work to probability (§3.3) before turning to Wiener's theory of Brownian motion. Then we discuss Borel's

strong law of large numbers, which focused attention on measure rather than on integration (§3.5). After looking at Steinhaus's axiomatization of Borel's denumerable probability and its relation to the Polish work on independent functions, we turn to Kolmogorov's use of measure theory in probability in the 1920s. Kolmogorov's work in probability began in collaboration with Khinchin, who was using Steinhaus's picture to develop limit theorems, but he quickly dropped Steinhaus's picture in favor of Fréchet's integral, which he brought to prominence in probability theory with his 1931 article on Markov processes (§3.6).

3.1 The invention of measure theory by Borel and Lebesgue

Émile Borel is usually considered the founder of measure theory. Whereas Peano and Jordan had extended the concept of length from intervals to a larger class of sets of real numbers by approximating the sets inside and out with finite unions of intervals, Borel used countable unions. His motivation came from complex analysis. In his doctoral dissertation in 1894 (published in 1895), Borel studied certain series that were known to diverge on a dense set of points on a closed curve and hence, it was thought, could not be continued analytically into the region bounded by the curve. Roughly speaking, Borel discovered that the set of points where divergence occurred, although dense, can be covered by a countable number of intervals with arbitrarily small total length. Elsewhere on the curve—almost everywhere, we would say now—the series does converge, and so analytic continuation is possible (Hawkins 1975, §4.2). This discovery led Borel to a new theory of measurability for subsets of $[0, 1]$, which he published in 1898.

Borel's innovation was quickly seized upon by Henri Lebesgue, who made it the basis for the powerful theory of integration that he first announced in 1901. We now speak of Lebesgue measure on the real numbers R and on the n-dimensional space R^n, and of the Lebesgue integral in these spaces.

We need not review Lebesgue's theory, but we should mention one theorem, the precursor of the Radon-Nikodym theorem: any countably additive and absolutely continuous set function on the real numbers is an indefinite integral. This result first appeared in Lebesgue's 1904 book (Hawkins 1975, p. 145; Pier 1994, p. 524). He generalized it to R^n in 1910 (Hawkins 1975, p. 186).

We should also mention a note published in 1918 by Wacław Sierpiński on the axiomatic treatment of Lebesgue measure. In this note, important to us because of the use Hugo Steinhaus later made of it, Sierpiński characterized the class of Lebesgue measurable sets as the smallest class K of sets satisfying the following conditions:

I For every set E in K, there is a nonnegative number $\mu(E)$ that will be its measure and will satisfy conditions II, III, IV, and V.

II Every finite closed interval is in K and has its length as its measure.

III K is closed under finite and countable unions of disjoint elements, and μ is finitely and countably additive.

IV If $E_1 \supset E_2$ and E_1 and E_2 are in K, then $E_1 \setminus E_2$ is in K.

V If E is in K and $\mu(E) = 0$, then any subset of E is in K.

An arbitrary class K satisfying these five conditions is not necessarily a field; there is no requirement that an intersection of two of K's elements also be in K.[6]

3.2 Abstract measure theory from Radon to Saks

Abstract measure and integration began in the second decade of the twentieth century and came into full flower in the fourth.

The first and most important step was taken by Johann Radon, in a celebrated article published in 1913. Radon unified Lebesgue and Stieltjes integration by generalizing integration with respect to Lebesgue measure to integration with respect to any countably additive set function (*absolut additive Mengenfunktion*) on the Borel sets in R^n. The generalization included a version of the theorem of Lebesgue's we just mentioned: if a countably additive set function g on R^n is absolutely continuous with respect to another countably additive set function f, then g is an indefinite integral with respect to f (Hawkins 1975, p. 189).

Constantin Carathéodory was also influential in drawing attention to measures on Euclidean spaces other than Lebesgue measure. In 1914, Carathéodory gave axioms for outer measure in a q-dimensional space, derived the notion of measure, and applied these ideas not only to Lebesgue measure on Euclidean spaces but also to lower-dimensional measures on Euclidean space, which assign lengths to curves, areas to surfaces, etc. (Hochkirchen 1999). Carathéodory also recast Lebesgue's theory of integration to make measure even more fundamental; in his 1918 textbook on real functions, he defined the integral of a positive function on a subset of R^n as the $(n + 1)$-dimensional measure of the region between the subset and the function's graph (Bourbaki, 1994, p. 228).

It was Fréchet who first went beyond Euclidean space. In 1915, Fréchet observed that much of Radon's reasoning does not depend on the assumption that one is working in R^n. One can reason in the same way in a much larger space, such as a space of functions. Any space will do, so long as the countably additive set function is defined on a σ-field of its subsets, as Radon had required. One thus arrives at the abstract theory of integration on which Kolmogorov based probability. As Kolmogorov put it in the preface to his *Grundbegriffe*,

---

[6]Recall that a *field* of sets is a collection of sets closed under relative complementation and finite union and intersection. A field of sets closed under denumerable union and intersection is a *Borel* field. A field that has a largest element is called an *algebra*, and a Borel field that has a largest element is called a *σ-algebra*. Although *algebra* and *σ-algebra* are now predominant in probability theory, *field* and *Borel field* were more common in mathematical work before the second world war.

> ... After Lebesgue's investigations, the analogy between the measure of a set and the probability of an event, as well as between the integral of a function and the mathematical expectation of a random variable, was clear. This analogy could be extended further; for example, many properties of independent random variables are completely analogous to corresponding properties of orthogonal functions. But in order to base probability theory on this analogy, one still needed to liberate the theory of measure and integration from the geometric elements still in the foreground with Lebesgue. This liberation was accomplished by Fréchet.

Fréchet did not, however, manage to generalize Radon's theorem on absolute continuity to the fully abstract framework. This generalization, now called the *Radon-Nikodym theorem*, was obtained by Otton Nikodym in 1930.

It should not be inferred from Kolmogorov's words that Fréchet used "measure" in the way we do today. In his 1915 articles and in his treatise on probability, cited in the *Grundbegriffe* but not published until 1937–1938, Fréchet used *fonction additive d'ensembles* for what we now call a measure. He makes this comment on p. 6 of the treatise:

> We should note a tendency, in recent years, to model probability theory on measure theory. In fact, it is not the notion of the measure of a set, but rather the notion of an *additive set function* that is appropriate, because of the theorem (or postulate) of total probability, for representing a probability, either continuous or discrete.

Kolmogorov was similarly old-fashioned in the text of the *Grundbegriffe*, using *vollständig additive Mengenfunktion*. Fréchet may have liberated the theory of measure and integration from its geometric roots, but both Fréchet and Kolmogorov continued to reserve the word measure for geometric settings. As Stanisław Ulam explained in 1943, measure has two characteristic properties: additivity for disjoint sets and equality of measure for sets that are congruent or otherwise considered equivalent (Ulam 1943, p. 601). We do find early examples in which "measure" is used in reference to an additive set function on a set not necessarily endowed with a congruence or equivalence relation: Ulam himself in German (1930, 1932) and Eberhard Hopf in English (1934). But the usage seems to have become standard only after the second world war. Doob's example is instructive. He began using "measure function" in a completely abstract context in his articles in the late 1930s and then abbreviated it, more and more often during the 1940s, to "measure". The full phrase "measure function" still surfaces occasionally in Doob's 1953 book, but by then the modern usage had been cemented in place by his student Paul Halmos in *Measure Theory*, published in 1950.

Nikodym's theorem was the beginning of a torrent of work on abstract measure and integration in the 1930s. We can gain some perspective on what happened by looking at the two editions of Stanisław Saks's textbook on integration. The first, which appeared in French almost simultaneously with Kolmogorov's

Grundbegriffe (the preface is dated May 1933), discusses the Perron and Denjoy integrals as well as the Lebesgue integral, but stays, throughout the eleven chapters of the main text, within Euclidean space. We find abstract integration only in a fifteen-page appendix, entitled "Intégrale de Lebesgue dans les espaces abstraits", which opens with a bow to Radon, Fréchet, the Radon-Nikodym theorem, and Ulam's 1932 announcement in 1932 concerning the construction of product measures. In the second edition four years later, in 1937, the abstract Lebesgue integral comes at the beginning, as the topic of Chapter I, now with bows to Radon, Daniell, Nikodym, and Jessen. There is again an appendix on the Lebesgue integral in abstract spaces, but this one is written by Stefan Banach, for whom an integral was a linear operator.

Banach's appendix was one of the early signs of a split between two schools of thought concerning the integral, which Bourbaki (1994, p. 228) traces back to Carathéodory's 1918 textbook. One school, following Carathéodory, has made measure ever more abstract, axiomatized, and basic. The other, following Young, Daniell, and Banach, takes integration as basic and puts more emphasis on the topological and geometric structures that underlie most instances of integration.

Kolmogorov was a vigorous participant in the discussion of measure and integration in the late 1920s and early 1930s. In 1933, Saks cited three of Kolmogorov's articles, including one in which Kolmogorov advanced a novel theory of integration of his own (1930a). In 1937, Saks cited these same articles again but took no notice of the *Grundbegriffe*.

3.3 Fréchet's integral

In an interview in 1984, Fréchet's student Jean Ville recalled how Fréchet had wanted him to write a dissertation in analysis, not in probability. In Paris in the 1930s, Ville explained, probability was considered merely an honorable pastime for those who had already distinguished themselves in pure mathematics (Crépel 1984, p. 43). Fréchet's own career, like that of his older colleagues Borel and Castelnuovo, had followed this pattern. His dissertation, completed in 1906 under Jacques Hadamard, can be said to have launched general topology.[7] He continued to concentrate on general topology and linear functionals until 1928, when, at the behest of Borel, he moved from Strasbourg to Paris and turned his main attention to probability and statistics. His stature as a mathematician assured him a leadership role. In 1941, at the age of 63, he succeeded to Borel's chair in Calculus of Probabilities and Mathematical Physics at the University

---

[7]The term "general topology" seems to have come into common use only after the second world war. Earlier names included "point-set theory", "analysis situ", and "general analysis". The topological spaces that interested Fréchet and his colleagues were spaces of functions, and the objects of greatest interest were real-valued functions on these spaces. Following a suggestion by Hadamard that has now been universally adopted, Fréchet called such functions "fonctionnelles", and he often called general topology "le calcul fonctionnel" (Taylor 1982, pp. 250–251). Nowadays we speak of "functional analysis" or "the theory of functions".

of Paris.[8]

When Fréchet generalized Radon's integral in 1915, he was explicit about what he had in mind: he wanted to integrate over function space. In some sense, therefore, he was already thinking about probability. An integral is a mean value. In a Euclidean space this might be a mean value with respect to a distribution of mass or electrical charge, but we cannot distribute mass or charge over a space of functions. The only thing we can imagine distributing over such a space is probability or frequency.

Why did Fréchet fail at the time to elaborate his ideas on abstract integration, connecting them explicitly with probability? One reason was the war. Mobilized on August 4, 1914, Fréchet was at or near the front for about two and a half years, and he was still in uniform in 1919. Thirty years later, he still had the notes on abstract integration that he had prepared, in English, for the course he had planned to teach at the University of Illinois in 1914–1915.

We should also note that Fréchet was not always enthusiastic about axioms. In a lecture delivered in 1925 (Fréchet 1955, pp. 1–10), he argued that the best principles for purely mathematical work are not necessarily best for practical science and education. He felt that too much emphasis had been put on axioms in geometry, mechanics, and probability; for some purposes these topics should be de-axiomatized.

Regardless of Fréchet's opportunities and inclinations, the time was not yet ripe in 1915 for general theorizing about probability in function space. The problem, as the American mathematician Theophil Hildebrandt pointed out in 1917 (p. 116), was the lack of interesting examples. Fréchet's integral, he said,

> . . . depends upon the existence, in a general class of elements, of an absolutely additive function v whose range is a class of subclasses of the fundamental class, i.e., a function such that
>
> $$v(\Sigma_n E_n) = \Sigma_n v(E_n),$$
>
> the E_n being mutually distinct and finite or denumerably infinite in number. The examples of this which have been given for the general space are trivial in that they reduce either to an infinite sum or an integral extended over a field in a finite number of dimensions. There is still lacking a really effective and desirable absolutely additive function for the higher type of spaces. . . .

---

[8]In his 1956 autobiography (p. 50), Norbert Wiener recalled that in 1920 he would not have been surprised had Fréchet turned out "to be the absolute leader of the mathematicians of his generation". That things turned out differently was due, Wiener thought in hindsight, to the excessive abstract formalism of Fréchet's work. Others have pointed to the fact that Fréchet contributed, both in general topology and in probability, more definitions than theorems (Taylor 1982, 1985, 1987).

Borel and Hadamard first proposed Fréchet for the Académie des Sciences in 1934, but he was not elected until 1956, when he was 77. Harald Cramér included an uncharacteristically negative comment about Fréchet's work in probability and statistics in his own scientific memoir (Cramér 1976, p. 528). Cramér was a professor of actuarial science rather than mathematics, but he contributed more than Fréchet to mathematical statistics, and he may have been irritated that Fréchet had never invited him to Paris.

In his presidential address to the American Mathematical Society on New Year's Day in 1915, Edward Van Vleck had worried even about the examples that Hildebrandt dismissed as trivial. According to Van Vleck, Poincaré, Borel, and Felix Bernstein had clarified the problem of mean motion by showing that exceptional cases have only measure zero. But

> ...care must be taken inasmuch as measure is not an invariant of analysis situ and hence may be dependent on the parameters introduced. This application of measure is as yet prospective rather than actual. ... (p. 337)

The motions of classical physics are functions of time and hence belong in function space, but the laws of classical physics are deterministic, and so we can put probability only on the initial conditions. In the examples under discussion, the initial conditions are finite-dimensional and can be parameterized in different ways. The celebrated ergodic theorems of Birkhoff and von Neumann, published in 1932, (Zund 2002), did provide a rationale for the choice of parameterization, but they were still concerned with Lebesgue measure on a finite-dimensional space of initial conditions.

The first nontrivial examples of probability in function space were provided by Daniell and Wiener.

3.4   Daniell's integral and Wiener's differential space

Percy Daniell, an Englishman working at the Rice Institute in Houston, Texas, introduced his integral in a series of articles in the *Annals of Mathematics* from 1918 to 1920. Although he built on the work of Radon and Fréchet, his viewpoint owed more to earlier work by William H. Young.

Like Fréchet, Daniell considered an abstract set E. But instead of beginning with an additive set function on subsets of E, he began with what he called an integral on E—a linear operator on some class T_0 of real-valued functions on E. The class T_0 might consist of all continuous functions (if E is endowed with a topology), or perhaps of all step functions. Applying Lebesgue's methods in this general setting, Daniell extended the linear operator to a wider class T_1 of functions on E, the summable functions. In this way, the Riemann integral is extended to the Lebesgue integral, the Stieltjes integral to the Radon integral, and so on (Daniell 1918). Using ideas from Fréchet's dissertation, Daniell also gave examples in infinite-dimensional spaces (Daniell 1919a,b).

In a remarkable but unheralded 1921 article in the *American Journal of Mathematics*, Daniell used his theory of integration to analyze the motion of a particle whose infinitesimal changes in position are independently and normally distributed. Daniell said he was studying *dynamic probability*. We now speak of *Brownian motion*, with reference to the botanist Robert Brown, who described the erratic motion of pollen in the early nineteenth century (Brush 1968). Daniell cited work in functional analysis by Vito Volterra and work in probability by Poincaré and Pearson, but he appears to have been unaware of the history of

his problem, for he cited neither Brown, Poincaré's student Louis Bachelier (Bachelier 1900, Courtault and Kabanov 2002), nor the physicists Albert Einstein and Marian von Smoluchowski (Einstein 1905, 1906; von Smoluchowski 1906). In retrospect, we may say that Daniell's was the first rigorous treatment, in terms of functional analysis, of the mathematical model that had been studied less rigorously by Bachelier, Einstein, and von Smoluchowski. Daniell remained at the Rice Institute in Houston until 1924, when he returned to England to teach at Sheffield University. Unaware of the earlier work on Brownian motion, he seems to have made no effort to publicize his work on the topic, and no one else seems to have taken notice of it until Stephen Stigler spotted it in 1973 in the course of a systematic search for articles related to probability and statistics in the *American Journal of Mathematics* (Stigler 1973).

The American ex-child prodigy and polymath Norbert Wiener, when he came upon Daniell's 1918 and July 1919 articles, was in a better position than Daniell himself to appreciate and advertise their remarkable potential for probability (Wiener 1956, Masani 1990, Segal 1992). As a philosopher (he had completed his Ph.D. in philosophy at Harvard before studying with Bertrand Russell in Cambridge), Wiener was well aware of the intellectual significance of Brownian motion and of Einstein's mathematical model for it. As a mathematician (his mathematical mentor was G. H. Hardy at Cambridge), he knew the new functional analysis, and his Cincinnati friend I. Alfred Barnett had suggested that he use it to study Brownian motion (Masani 1990, p. 77).

In November 1919, Wiener submitted his first article on Daniell's integral to the *Annals of Mathematics*, the journal where Daniell's four articles on it had appeared. This article did not yet discuss Brownian motion; it merely laid out a general method for setting up a Daniell integral when the underlying space E is a function space. But by August 1920, Wiener was in France to explain his ideas on Brownian motion to Fréchet and Lévy (Segal, 1992, p. 397). Fréchet, he later recalled, did not appreciate its importance, but Lévy showed greater interest, once convinced that there was a difference between Wiener's method of integration and Gâteax's (Wiener 1956, p. 64). Wiener followed up with a quick series of articles: a first exploration of Brownian motion (1921a), an exploration of what later became known as the Ornstein-Uhlenbeck model (1921b, Doob 1966), a more thorough and later much celebrated article on Brownian motion ("Differential-Space") in 1923, and a final installment in 1924.

Because of his work on cybernetics after the second world war, Wiener is now the best known of the twentieth-century mathematicians in our story—far better known to the general intellectual public than Kolmogorov. But he was not well known in the early 1920s, and though the most literary of mathematicians—he published fiction and social commentary—his mathematics was never easy to read, and the early articles attracted hardly any immediate readers or followers. Only after he became known for his work on Tauberian theorems in the later 1920s, and only after he returned to Brownian motion in collaboration with C. A. B. Paley and Antoni Zygmund in the early 1930s (Paley, Wiener, and Zygmund 1933) do we see the work recognized as central and seminal for the emerging theory of continuous stochastic processes. In 1934, Lévy finally de-

monstrated that he really understood what Wiener had done by publishing a generalization of Wiener's 1923 article.

Wiener's basic idea was simple. Suppose we want to formalize the notion of Brownian motion for a finite time interval, say $0 \leq t \leq 1$. A realized path is a function on $[0, 1]$. We want to define mean values for certain functionals (real-valued functions of the realized path). To set up a Daniell integral that gives these mean values, Wiener took T_0 to consist of functionals that depend only on the path's values at a finite number of time points. One can find the mean value of such a functional using Gaussian probabilities for the changes from each time point to the next. Extending this integral by Daniell's method, he succeeded in defining mean values for a wide class of functionals. In particular, he obtained probabilities (mean values for indicator functions) for certain sets of paths. He showed that the set of continuous paths has probability one, while the set of differentiable paths has probability zero.

It is now commonplace to translate this work into Kolmogorov's measure-theoretic framework. Kiyoshi Itô, for example, in a commentary published along with Wiener's articles from this period in Volume 1 of Wiener's collected works, writes as follows (p. 515) concerning Wiener's 1923 article:

> Having investigated the differential space from various directi-
> ons, Wiener defines the Wiener measure as a σ-additive probability
> measure by means of Daniell's theory of integral.

It should not be thought, however, that Wiener defined a σ-additive probability measure and then found mean values as integrals with respect to that measure. Rather, as we just explained, he started with mean values and used Daniell's theory to obtain more. This Daniellian approach to probability, making mean value basic and probability secondary, has long taken a back seat to Kolmogorov's approach, but it still has its supporters (Whittle 2000, Haberman 1996).

Today's students sometimes find it puzzling that Wiener could provide a rigorous account of Brownian motion before Kolmogorov had formulated his axioms—before probability itself had been made rigorous. But Wiener did have a rigorous mathematical framework: functional analysis. As Doob wrote in his commemoration of Wiener in 1966, "He came into probability from analysis and made no concessions." Wiener set himself the task of using functional analysis to describe Brownian motion. This meant using some method of integration (he knew many, including Fréchet's and Daniell's) to assign a "mean" or "average value" to certain functionals, and this included assigning "a measure, a probability" (Wiener 1924, p. 456) to certain events. It did not involve advancing an abstract theory of probability, and it certainly did not involve divorcing the idea of measure from geometry, which is deeply implicated in Brownian motion. From Wiener's point of view, whether a particular number merits being called a "probability" or a "mean value" in a particular context is a practical question, not to be settled by abstract theory.

Abstract theory was never Wiener's predilection. He preferred, as Irving Segal put it (1992, p. 422), "a concrete incision to an abstract envelopment". But

we should mention the close relation between his work and Lévy's general ideas about probability in abstract spaces. In his 1920 paper on Daniell's integral (p. 66), Wiener pointed out that we can use successive partitioning to set up such an integral in any space. We finitely partition the space and assign probabilities (positive numbers adding to one) to the elements of the partition, then we further partition each element of the first partition, further distributing its probability, and so on. This is an algorithmic rather than an axiomatic picture, but it can be taken as defining what should be meant by a probability measure in an abstract space. Lévy adopted this viewpoint in his 1925 book (p. 331), with due acknowledgement to Wiener. Lévy later called a probability measure obtained by this process of successive partitioning a *true probability law* (1970, pp. 65–66).

3.5 Borel's denumerable probability

Impressive as it was and still is, Wiener's work played little role in the story leading to Kolmogorov's *Grundbegriffe*. The starring role was played instead by Émile Borel.

In retrospect, Borel's use of measure theory in complex analysis in the 1890s already looks like probabilistic reasoning. Especially striking in this respect is the argument Borel gave in 1897 for his claim that a Taylor series will usually diverge on the boundary of its circle of convergence. In general, he asserted, successive coefficients of the Taylor series, or at least successive groups of coefficients, are independent. He showed that each group of coefficients determines an arc on the circle, that the sum of lengths of the arcs diverges, and that the Taylor series will diverge at a point on the circle if it belongs to infinitely many of the arcs. The arcs being independent, and the sum of their lengths being infinite, a given point must be in infinitely many of them. To make sense of this argument, we must evidently take "in general" to mean that the coefficients are chosen at random and "independent" to mean probabilistically independent; the conclusion then follows by what we now call the Borel-Cantelli Lemma. Borel himself used probabilistic language when he reviewed this work in 1912 (Kahane 1994), and Steinhaus spelled the argument out in fully probabilistic terms in 1930 (Steinhaus 1930a). For Borel in the 1890s, however, complex analysis was not a domain for probability, which was concerned with events in the real world.

In the new century, Borel did begin to explore the implications for probability of his and Lebesgue's work on measure and integration (Bru 2001). His first comments came in an article in 1905, where he pointed out that the new theory justified Poincaré's intuition that a point chosen at random from a line segment would be incommensurable with probability one and called attention to Anders Wiman's work on continued fractions (1900, 1901), which had been inspired by the question of the stability of planetary motions, as an application of measure theory to probability.

Then, in 1909, Borel published a startling result—his strong law of large numbers (Borel 1909a). This new result strengthened measure theory's connection both with geometric probability and with the heart of classical proba-

bility theory—the concept of independent trials. Considered as a statement in geometric probability, the law says that the fraction of ones in the binary expansion of a real number chosen at random from $[0, 1]$ converges to one-half with probability one. Considered as a statement about independent trials (we may use the language of coin tossing, though Borel did not), it says that the fraction of heads in a denumerable sequence of independent tosses of a fair coin converges to one-half with probability one. Borel explained the geometric interpretation, and he asserted that the result can be established using measure theory (§I.8). But he set measure theory aside for philosophical reasons and provided an imperfect proof using denumerable versions of the rules of total and compound probability. It was left to others, most immediately Faber (1910) and Hausdorff (1914), to give rigorous measure-theoretic proofs (Doob 1989, 1994; von Plato 1994).

Borel's discomfort with a measure-theoretic treatment can be attributed to his unwillingness to assume countable additivity for probability (Barone and Novikoff 1978, von Plato 1994). He saw no logical absurdity in a countably infinite number of zero probabilities adding to a nonzero probability, and so instead of general appeals to countable additivity, he preferred arguments that derive probabilities as limits as the number of trials increases (1909a, §I.4). Such arguments seemed to him stronger than formal appeals to countable additivity, for they exhibit the finitary pictures that are idealized by the infinitary pictures. But he saw even more fundamental problems in the idea that Lebesgue measure can model a random choice (von Plato 1994, pp. 36–56; Knobloch 2001). How can we choose a real number at random when most real numbers are not even definable in any constructive sense?

Although Hausdorff did not hesitate to equate Lebesgue measure with probability, his account of Borel's strong law, in his *Grundzüge der Mengenlehre* in 1914 (pp. 419–421), treated it as a theorem about real numbers: the set of numbers in $[0, 1]$ with binary expansions for which the proportion of ones converges to one-half has Lebesgue measure one. But in 1916 and 1917, Francesco Paolo Cantelli rediscovered the strong law (he neglected, in any case, to cite Borel) and extended it to the more general result that the average of bounded random variables will converge to their mean with arbitrarily high probability. Cantelli's work inspired other authors to study the strong law and to sort out different concepts of probabilistic convergence.

By the early 1920s, it seemed to some that there were two different versions of Borel's strong law—one concerned with real numbers and one concerned with probability. In 1923, Hugo Steinhaus proposed to clarify matters by axiomatizing Borel's theory of denumerable probability along the lines of Sierpiński's axiomatization of Lebesgue measure. Writing A for the set of all infinite sequences of ρs and ηs (ρ for "rouge" and η for "noir"; now we are playing red or black rather than heads or tails), Steinhaus proposed the following axioms for a class $\mathfrak{K}$ of subsets of A and a real-valued function μ that gives probabilities for the elements of $\mathfrak{K}$:

I $\mu(E) \geq 0$ for all $E \in \mathfrak{K}$.

II   1 For any finite sequence e of ρs and ηs, the subset E of A consisting of all infinite sequences that begin with e is in $\mathfrak{K}$.

      2 If two such sequences e_1 and e_2 differ in only one place, then $\mu(E_1) = \mu(E_2)$, where E_1 and E_2 are the corresponding sets.

      3 $\mu(A) = 1$.

III $\mathfrak{K}$ is closed under finite and countable unions of disjoint elements, and μ is finitely and countably additive.

IV If $E_1 \supset E_2$ and E_1 and E_2 are in $\mathfrak{K}$, then $E_1 \setminus E_2$ is in $\mathfrak{K}$.

V If E is in $\mathfrak{K}$ and $\mu(E) = 0$, then any subset of E is in $\mathfrak{K}$.

Sierpiński's axioms for Lebesgue measure consisted of I, III, IV, and V, together with an axiom that says that the measure $\mu(J)$ of an interval J is its length. This last axiom being demonstrably equivalent to Steinhaus's axiom II, Steinhaus concluded that the theory of probability for an infinite sequence of binary trials is isomorphic with the theory of Lebesgue measure.

In order to show that his axiom II is equivalent to setting the measures of intervals equal to their length, Steinhaus used the Rademacher functions—the nth Rademacher function being the function that assigns a real number the value 1 or -1 depending on whether the nth digit in its dyadic expansion is 0 or 1. He also used these functions, which are independent random variables, in deriving Borel's strong law and related results. The work by Rademacher (1922) and Steinhaus marked the beginning of the Polish school of "independent functions", which made important contributions to probability theory during the period between the wars (Holgate 1997).

Steinhaus cited Borel but not Cantelli. The work of Borel and Cantelli was drawn together, however, by the Russians, especially by Evgeny Slutsky in his wide-ranging article in the Italian journal *Metron* in 1925. Cantelli, it seems, was not aware of the extent of Borel's priority until he debated the matter with Slutsky at the International Congress of Mathematicians at Bologna in 1928 (Seneta 1992, Bru 2003a).

The name "strong law of large numbers" was introduced by Khinchin in 1928. Cantelli had used "uniform" instead of "strong". The term "law of large numbers" had been introduced originally by Poisson (1837) and had come to be used as a name for Bernoulli's theorem (or for the conclusion, from this theorem together with Cournot's principle, that the frequency of an event will approximate its probability), although Poisson had thought he was naming a generalization (Stigler 1986, p. 185).

3.6   Kolmogorov enters the stage

Although Steinhaus considered only binary trials in his 1923 article, his reference to Borel's more general concept of denumerable probability pointed to generalizations. We find such a generalization in Kolmogorov's first article on probability, co-authored by Khinchin (Khinchin and Kolmogorov 1925), which

showed that a series of discrete random variables $y_1 + y_2 + \cdots$ will converge with probability one when the series of means and the series of variances both converge. The first section of the article, due to Khinchin, spells out how to represent the random variables as functions on $[0, 1]$: divide the interval into segments with lengths equal to the probabilities for y_1's possible values, then divide each of these segments into smaller segments with lengths proportional to the probabilities for y_2's possible values, and so on. This, Khinchin notes with a nod to Rademacher and Steinhaus, reduces the problem to a problem about Lebesgue measure. This reduction was useful because the rules for working with Lebesgue measure were clear, while Borel's picture of denumerable probability remained murky.

Dissatisfaction with this detour into Lebesgue measure must have been one impetus for the *Grundbegriffe* (Doob 1989, p. 818). Kolmogorov made no such detour in his next article on the convergence of sums of independent random variables. In this sole-authored article, dated 24 December 1926 and published in 1928, he took probabilities and expected values as his starting point. But even then, he did not appeal to Fréchet's countably additive calculus. Instead, he worked with finite additivity and then stated an explicit ad hoc definition when he passed to a limit. For example, he defined the probability P that the series $\sum_{n=1}^{\infty} y_n$ converges by the equation

$$P = \lim_{\eta \to 0} \lim_{n \to \infty} \lim_{N \to \infty} \mathfrak{W} \left[\mathrm{Max} \left| \sum_{k=n}^{p} y_k \right|_{p=n}^{N} < \eta \right],$$

where $\mathfrak{W}(E)$ denotes the probability of the event E. (This formula does not appear in the Russian and English translations provided in Kolmogorov's collected works; there the argument has been modernized so as to eliminate it.) This recalls the way Borel proceeded in 1909: think through each passage to the limit.

It was in his seminal article on Markov processes (Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, dated 26 July 1930 and published in 1931) that Kolmogorov first explicitly and freely used Fréchet's calculus as his framework for probability. In this article, Kolmogorov considered a system with a set of states $\mathfrak{A}$. For any two time points t_1 and t_2 ($t_1 < t_2$), any state $x \in \mathfrak{A}$, and any element $\mathfrak{E}$ in a collection $\mathfrak{F}$ of subsets of $\mathfrak{A}$, he wrote

$$P(t_1, x, t_2, \mathfrak{E}) \tag{1}$$

for the probability, when the system is in state x at time t_1, that it will be in a state in $\mathfrak{E}$ at time t_2. Citing Fréchet, Kolmogorov assumed that P is countably additive as a function of $\mathfrak{E}$ and that $\mathfrak{F}$ is closed under differences and countable unions and contains the empty set, all singletons, and $\mathfrak{A}$. But the focus was not on Fréchet; it was on the equation that ties together the transition probabilities (1), now called the Chapman-Kolmogorov equation. The article launched the study of this equation by purely analytical methods, a study that kept probabilists occupied for fifty years.

As many commentators have noted, the 1931 article makes no reference to probabilities for trajectories. There is no suggestion that such probabilities are needed in order for a stochastic process to be well defined. Consistent transition probabilities, it seems, are enough. Bachelier (1900, 1910, 1912) is cited as the first to study continuous-time stochastic processes, but Wiener is not cited. We are left wondering when Kolmogorov first became aware of Wiener's success in formulating probability statements concerning Brownian trajectories. He certainly was aware of it in 1934, when he reviewed Lévy's generalization of Wiener's 1923 article in the *Zentralblatt*. He then called that article well known, but so far as we are aware, he had not mentioned it in print earlier.

4 Hilbert's sixth problem

At the beginning of the twentieth century, many mathematicians were dissatisfied with what they saw as a lack of clarity and rigor in the probability calculus. The whole calculus seemed to be concerned with concepts that lie outside mathematics: event, trial, randomness, probability. As Henri Poincaré wrote, "one can hardly give a satisfactory definition of probability" (1912, p. 24).

The most celebrated call for clarification came from David Hilbert. The sixth of the twenty-three open problems that Hilbert presented to the International Congress of Mathematicians in Paris in 1900 was to treat axiomatically, after the model of geometry, those parts of physics in which mathematics already played an outstanding role, especially probability and mechanics (Hilbert 1902). To explain what he meant by axioms for probability, Hilbert cited Georg Bohlmann, who had labeled the rules of total and compound probability axioms rather than theorems in his lectures on the mathematics of life insurance (Bohlmann 1901). In addition to a logical investigation of these axioms, Hilbert called for a "rigorous and satisfactory development of the method of average values in mathematical physics, especially in the kinetic theory of gases".

Hilbert's call for a mathematical treatment of average values was answered in part by the work on integration that we discussed in the preceding section, but his suggestion that the classical rules for probability should be treated as axioms on the model of geometry was an additional challenge. Among the early responses, we may mention the following:

- In his dissertation, in 1904 in Zürich, Rudolf Laemmel discussed the rules of total and compound probability as axioms. But he stated the rule of compound probability only in the case of independence, a concept he did not explicate. Schneider reprinted excerpts from this dissertation in 1988 (pp. 359–366).

- In a 1907 dissertation in Göttingen, directed by Hilbert himself, Ugo Broggi gave only two axioms: an axiom stating that the sure event has probability one, and the rule of total probability. Following tradition, he then defined probability as a ratio (a ratio of numbers of cases in the discrete setting; a ratio of the Lebesgue measure of two sets in the geometric

setting) and verified his axioms. He did not state an axiom corresponding to the classical rule of compound probability. Instead, he gave this name to a rule for calculating the probability of a Cartesian product, which he derived from the definition of geometric probability in terms of Lebesgue measure. Again, see Schneider (pp. 367–377) for excerpts. Broggi mistakenly claimed that his axiom of total probability (finite additivity) implied countable additivity; see Steinhaus 1923.

- In an article written in 1920, published in 1923, and listed in the bibliography of the *Grundbegriffe*, Antoni Łomnicki proposed that probability should always be understood relative to a density ϕ on a set $\mathcal{M}$ in R^r. Łomnicki defined this probability by combining two of Carathéodory's ideas: the idea of p-dimensional measure and the idea of defining the integral of a function on a set as the measure of the region between the set and the function's graph (see §3.1 above). The probability of a subset m of $\mathcal{M}$, according to Łomnicki, is the ratio of the measure of the region between m and ϕ's graph to the measure of the region between $\mathcal{M}$ and this graph. If $\mathcal{M}$ is an r-dimensional subset of R^r, then the measure being used is Lebesgue measure on R^{r+1}; if $\mathcal{M}$ is a lower-dimensional subset of R^r, say p-dimensional, then the measure is the $(p+1)$-dimensional Carathéodory measure. This definition covers discrete as well as continuous probability; in the discrete case, $\mathcal{M}$ is a set of discrete points, the function ϕ assigns each point its probability, and the region between a subset m and the graph of ϕ consists of a line segment for each point in m, whose Carathéodory measure is its length—i.e., the point's probability. The rule of total probability follows. Like Broggi, Łomnicki treated the rule of compound probability as a rule for relating probabilities on a Cartesian product to probabilities on its components. He did not consider it an axiom, because it holds only if the density itself is a product density.

Two general tendencies are notable here: an increasing emphasis on measure, and an attendant decline in the role of compound probability. These tendencies are also apparent in the 1923 article by Steinhaus that we have already discussed. Steinhaus did not mention compound probability.

We now discuss at greater length responses by Bernstein, von Mises, Slutsky, Kolmogorov, and Cantelli.

4.1 Bernstein's qualitative axioms

In an article published in Russian in 1917 and listed by Kolmogorov in the *Grundbegriffe*'s bibliography, Sergei Bernstein showed that probability theory can be founded on qualitative axioms for numerical coefficients that measure the probabilities of propositions.

Bernstein's two most important axioms correspond to the classical axioms of total and compound probability:

- If A and A_1 are equally likely, B and B_1 are equally likely, A and B

are incompatible, and A_1 and B_1 are incompatible, then (A or B) and (A_1 or B_1) are equally likely.

- If A occurs, the new probability of a particular occurrence α of A is a function of the initial probabilities of α and A.

Using the first axiom, Bernstein concluded that if A is the disjunction of m out of n equally likely and incompatible propositions, and B is as well, then A and B must be equally likely. It follows that the numerical probability of A and B is some function of the ratio m/n, and we may as well take that function to be the identity. Using the second axiom, Bernstein then finds that the new probability of α when A occurs is the ratio of the initial probability of α to that of A.

Bernstein also axiomatized the field of propositions and extended his theory to the case where this field is infinite. He exposited his qualitative axioms again in a probability textbook that he published in 1927, but neither the article nor the book were ever translated out of Russian into other languages. John Maynard Keynes included Bernstein's article in the bibliography of the 1921 book where he developed his own system of qualitative probability. Subsequent writers on qualitative probability, most prominently Bernard O. Koopman in 1940 and Richard T. Cox in 1946, acknowledged a debt to Keynes but not to Bernstein. The first summary of Bernstein's ideas in English appeared only in 1974, when Samuel Kotz published an English translation of Leonid E. Maistrov's history of probability.

Unlike von Mises and Kolmogorov, Bernstein was not a frequentist. He had earned his doctorate in Paris, and the philosophical views he expresses at the end of his 1917 article are in line with those of Borel and Lévy: probability is essentially subjective and becomes objective only when there is sufficient consensus or adequate confirmation of Cournot's principle.

4.2 Von Mises's Kollektivs

The concept of a Kollektiv was introduced into the German scientific literature by Gustav Fechner in the 1870s and popularized in his posthumous *Kollektivmasslehre*, which appeared in 1897 (Sheynin 2004). The concept was quickly taken up by Georg Helm (1902) and Heinrich Bruns (1906).

Fechner wrote about the concept of a Kollektiv-Gegenstand (collective object) or a Kollektiv-Reihe (collective series). It was only later, in Meinong (1915) for example, that we see these names abbreviated to Kollektiv. As the name Kollektiv-Reihe indicates, a Kollektiv is a population of individuals given in a certain order; Fechner called the ordering the Urliste. It was supposed to be irregular—random, we would say. Fechner was a practical scientist, not concerned with the theoretical notion of probability. But as Helm and Bruns realized, probability theory provides a framework for studying Kollektivs.

Richard von Mises was a well established applied mathematician when he took up the concept of a Kollektiv in 1919. His contribution was to realize that the concept can be made into a mathematical foundation for probability

theory. As von Mises defined it, a Kollektiv is an infinite sequence of outcomes satisfying two axioms:

1. the relative frequency of each outcome converges to a real number (the probability of the outcome) as we look at longer and longer initial segments of the sequence, and

2. the relative frequency converges to the same probability in any subsequence selected without knowledge of the future (we may use knowledge of the outcomes so far in deciding whether to include the next outcome in the subsequence).

The second property says we cannot change the odds by selecting a subsequence of trials on which to bet; this is von Mises's version of the "hypothesis of the impossibility of a gambling system", and it assures the irregularity of the Urliste.

According to von Mises, the purpose of the probability calculus is to identify situations where Kollektivs exist and the probabilities in them are known, and to derive from probabilities for other Kollektivs from these given probabilities. He pointed to three domains where probabilities for Kollektivs are know:

1. Games of chance, where devices are carefully constructed so the axioms will be satisfied.

2. Statistical phenomena, where the two axioms can sometimes be confirmed, to a reasonable degree.

3. Theoretical physics, where the two axioms play the same hypothetical role as other theoretical assumptions.

(See von Mises 1931, pp. 25–27.)

Von Mises derived the classical rules of probability, such as the rules for adding and multiplying probabilities, from rules for constructing new Kollektivs out of an initial one. He had several laws of large numbers. The simplest was his definition of probability: the probability of an event is the event's limiting frequency in a Kollektiv. Others arose as one constructed further Kollektivs.

Von Mises's ideas were taken up by a number of mathematicians in the 1920s and 1930s. Kolmogorov's bibliography includes an article by Arthur Copeland (1932, based on his 1928 article) that proposed founding probability theory on particular rules for selecting subsequences in von Mises's scheme, as well as articles by Karl Dörge (1930), Hans Reichenbach (1932), and Erhard Tornier (1933) arguing for alternative schemes. But the most prominent mathematicians of the time, including the Göttingen mathematicians (MacLane 1995), the French probabilists, and the British statisticians, were hostile or indifferent.

After the publication of the *Grundbegriffe*, Kollektivs were given a rigorous mathematical basis by Abraham Wald (1936, 1937, 1938) and Alonzo Church (1940), but the claim that they provide a foundation for probability was refuted by Jean Ville (1936, 1939). Ville pointed out that whereas a Kollektiv in von Mises's sense will not be vulnerable to a gambling system that chooses a subsequence of trials on which to bet, it may still be vulnerable to a more clever

gambling system, which also varies the amount of the bet and the outcome on which to bet. Ville called such a system a "martingale". Ville's work inspired Doob's measure-theoretic martingales. Our own game-theoretic foundation for probability (Shafer and Vovk 2001) combines Ville's idea of a martingale with ideas about generalizing probability that go back to Robert Fortet (1951, pp. 46–47).

4.3   Slutsky's calculus of valences

In an article published in Russian in 1922, Evgeny Slutsky presented a viewpoint that greatly influenced Kolmogorov. As Kolmogorov said on the occasion of Slutsky's death in 1948, Slutsky was "the first to give the right picture of the purely mathematical content of probability theory".

How do we make probability theory purely mathematical? Markov had claimed to do this in his textbook, but Slutsky did not think Markov had succeeded, for Markov had retained the subjective notion of equipossibility, with all its subjectivity. The solution, Slutsky felt, was to remove both the word "probability" and the notion of equally likely cases from the theory. Instead of beginning with equally likely cases, one should begin by assuming merely that numbers are assigned to cases, and that when a case assigned the number α is further subdivided, the numbers assigned to the subcases should add to α. The numbers assigned to cases might be equal, or they might not. The addition and multiplication theorems would be theorems in this abstract calculus. But it should not be called the probability calculus. In place of "probability", he suggested the unfamiliar word валентность, or "valence". (He may have been following Laemmel, who had used the German "valenz", which can be translated into English as "weight".) Probability would be only one interpretation of the calculus of valences, a calculus fully as abstract as group theory.

Slutsky listed three distinct interpretations of the calculus of valences:

1. Classical probability (equally likely cases).

2. Finite empirical sequences (frequencies).

3. Limits of relative frequencies. (Slutsky remarks that this interpretation is particularly popular with the English school.)

Slutsky did not think probability could be reduced to limiting frequency, because sequences of independent trials have properties that go beyond their possessing limiting frequencies. He gave two examples. First, initial segments of the sequences have properties that are not imposed by the eventual convergence of the relative frequency. Second, the sequences must be irregular in a way that resists the kind of selection discussed by von Mises (he cites an author named Умов, not von Mises).

In his 1925 article on limit theorems, written in German, Slutsky returned briefly to his calculus of valences, now saying that it should follow the lines laid out by Bernstein and that it should be called *Disjunktionsrechnung*, a name Ladislaus von Bortkiewicz had suggested to him in a letter. *Disjunktionsrechnung*

was to be distinguished from *Stochastik*, the science concerned with the application of *Disjunktionsrechnung* to random phenomena.

The word "stochastic" was first associated with probability by Jacob Bernoulli. Writing in Latin (1713, p. 213), Bernoulli called the art of combining arguments and data "*Ars Conjectandi* sive *Stochastice*"—the art of conjecturing or guessing. Greek in origin, "stochastic" may have seemed as erudite and mysterious to Bernoulli as it does to us. It was not used widely until it was revived and promoted by von Bortkiewicz in 1917.[9] As von Bortkiewicz's saw it, the probability calculus is mathematics, and *Stochastik* is the science of applying it to the real world. The term was adopted, with acknowledgement to von Bortkiewicz, by Chuprov in 1922 and Du Pasquier in 1926. Slutsky entitled his 1925 article "Über stochastische Asymptoten und Grenzwerte", stochastic limits being the limiting frequencies and averages we extract from statistical data. In 1923, Chuprov's student Jacob Mordukh called exchangeability "stochastic commutativity". In 1934, Khinchin gave "stochastic process" the meaning it has today.[10] The prestige of the Moscow school assured the international adoption of the term, but von Bortkiewicz's gloss on "stochastic" was lost. In Stalin's time, the Russians were not interested in suggesting that the applications of probability require a separate science. So stochastic became merely a mysterious but international synonym for random, aléatoire, zufällig, casuale, and случайный.

4.4 Kolmogorov's general theory of measure

As we have seen, Kolmogorov had used Fréchet's integral hesitantly in his article on the convergence of sums of random variables in 1926, but forthrightly in his article on Markov processes in 1930. Between these two mathematical contributions, he wrote a thought piece about how probability should fit into a broader abstract theory. Dated 8 January 1927 and published in 1929, this piece was neither a survey (it did not cite Fréchet, Steinhaus, or Slutsky) nor a report on research, and it included no proofs. It was philosophical—a sketch of a program.

By all appearances, the piece was inspired by Slutsky. But whereas Slutsky was a statistician, Kolmogorov was already established as a mathematician's

---

[9]On p. 3, von Bortkiewicz wrote:

> The consideration of empirical quantities, oriented to probability theory but based on the "law of large numbers", may be called S t o c h a s t i k ($\sigma\tau o\chi\acute{\alpha}\zeta\epsilon\sigma\theta\alpha\iota$ = aim, conjecture). Stochastik is not merely probability theory, but rather probability theory in its application... "Stochastik" also signified applied probability theory for Bernoulli; only he represented an antiquated viewpoint with respect to the applications considered.

For one appearance of "stochastic" between Bernoulli and von Bortkiewicz, see Prevost and Lhuilier (1799).

[10]In 1931, Kolmogorov had called what we now call a "Markov process" a "stochastically definite process" and had used "stochastic process" casually. In addition to fixing the precise meaning of "stochastic process", Khinchin's 1934 article introduced "Markov process" (Dynkin (1989). "Markov chain" was already in use in the 1920s (Bru 2003a).

mathematician. In 1922, at the age of 19, Kolmogorov had constructed an integrable function whose Fourier series diverges almost everywhere (1923). In 1925 he had published two notes on the theory of integration in the *Comptes rendus*. Slutsky had mentioned frequencies as an alternative interpretation of a general calculus. Kolmogorov pointed to more mathematical examples: the distribution of digits in the decimal expansions of irrationals, Lebesgue measure in an n-dimensional cube, and the density of a set A of positive integers (the limit as $n \to \infty$ of the fraction of the integers between 1 and n that are in A).

The abstract theory Kolmogorov sketches is concerned with a function M that assigns a nonnegative number $M(E)$ to each element E of class of subsets of a set A. He calls $M(E)$ the measure (мера) of E, and he calls M a measure specification (мероопределение). So as to accommodate all the mathematical examples he has in mind, he assumes, in general, neither that M is countably additive nor that the class of subsets to which it assigns numbers is a field. Instead, he assumes only that when E_1 and E_2 are disjoint and M assigns a number to two of the three sets E_1, E_2, and $E_1 \cup E_2$, it also assigns a number to the third, and that

$$M(E_1 \cup E_2) = M(E_1) + M(E_2)$$

then holds (cf. Steinhaus's Axioms III and IV). In the case of probability, however, he does suggest (using different words) that M should be countably additive and that class of subsets to which it assigns numbers should be a field, for only then can we uniquely define probabilities for countable unions and intersections, and this seems necessary to justify arguments involving events such as the convergence of random variables.

He defines the abstract Lebesgue integral of a function f on A, and he comments that countable additivity is to be assumed whenever such an integral is discussed. He writes $M_{E_1}(E_2) = M(E_1 E_2)/M(E_1)$ "by analogy with the usual concept of relative probability". He defines independence for partitions, and he comments, no doubt in reference to Borel's strong law and other results in number theory, that the notion of independence is responsible for the power of probabilistic methods within pure mathematics.

The mathematical core of the *Grundbegriffe* is already here. Many years later, in his commentary in Volume II of his collected works (p. 520 of the English edition), Kolmogorov said that only the set-theoretic treatment of conditional probability and the theory of distributions in infinite products was missing. Also missing, though, is the bold rhetorical move that Kolmogorov made in the *Grundbegriffe*—giving the abstract theory the name probability.

4.5 The axioms of Steinhaus and Ulam

In the 1920s and 1930s, the city of Lwów in Poland[11] was a vigorous center of mathematical research, led by Hugo Steinhaus. In 1929 and 1930, Steinhaus's work on limit theorems intersected with Kolmogorov's, and his approach promoted the idea that probability should be axiomatized in the style of measure theory.

As we saw in §3.5, Steinhaus had already, in 1923, formulated axioms for heads and tails isomorphic to Sierpiński's axioms for Lebesgue measure. This isomorphism had more than a philosophical purpose; Steinhaus used it to prove Borel's strong law. In a pair of articles written in 1929 and published in 1930 (Steinhaus 1930a,b), Steinhaus extended his approach to limit theorems involving an infinite sequence of independent draws $\theta_1, \theta_2, \ldots$ from the interval $[0, 1]$. Here are the axioms he gave (Steinhaus 1930b, pp. 23–24; we translate loosely from the French):

1. Probability is a nonnegative finite number.

2. The probability that $\theta_i \in \Theta_i$ for $i = 1, \ldots, n$, where the Θ_i are measurable subsets of $[0, 1]$, is
$$|\Theta_1| \cdot |\Theta_2| \cdots |\Theta_n|,$$
where $|\Theta_i|$ is the Lebesgue measure of Θ_i.

3. If $E_1, E_2, \ldots$ is a sequence of disjoint subsets of $[0, 1]^\infty$, and for every k the probability $\mu(E_k)$ for the sequence $\theta_1, \theta_2, \ldots$ being in E_k is determined, then the probability $\mu(\cup_{k=1}^\infty E_k)$ is determined and is equal to $\sum_{k=1}^\infty \mu(E_k)$.

4. If $E_1 \supset E_2$ and $\mu(E_1)$ and $\mu(E_2)$ are determined, then $\mu(E_1 \setminus E_2)$ is determined.

5. If $\mu(E_1) = 0$ and $E_1 \supset E_2$, then $\mu(E_2)$ is determined.

6. If $\mathfrak{K}^*$ is a class of E such that it is possible to define $\mu(E)$ for all the E in $\mathfrak{K}$ while respecting postulates 1–5, and if $\mathfrak{K}$ is the intersection of all the $\mathfrak{K}^*$ with this property, then $\mu(E)$ is defined only for E in $\mathfrak{K}$.

Except for the second one, these axioms are identical with Steinhaus's axioms for heads and tails (see p. 26). In the case of heads and tails, the second axiom specified probabilities for each initial finite sequence of heads and tails. Here it specifies probabilities for $\theta_1, \theta_2, \ldots, \theta_n$.

Steinhaus presented his axioms as a "logical extrapolation" of the classical axioms to the case of an infinite number of trials (Steinhaus 1930b, p. 23). They were more or less tacitly used, he asserted, in all classical problems, such as the problem of the gambler's ruin, where the game as a whole—not merely finitely many rounds—must be considered (Steinhaus 1930a, p. 409).

---

[11]Though it was in Poland between the two world wars, this city is now in Ukraine. Its name is spelled differently in different languages: Lwów in Polish, Lviv in Ukrainian, and Lvov in Russian. When part of Austria-Hungary and, briefly, Germany, it was called Lemberg. Some articles in our bibliography refer to it as Léopol.

As in the case of heads and tails, Steinhaus showed that there are probabilities uniquely satisfying his axioms by setting up an isomorphism with Lebesgue measure on $[0, 1]$, this time using a sort of Peano curve to map $[0, 1]^\infty$ onto $[0, 1]$. He used the isomorphism to prove several limit theorems, including one formalizing Borel's 1897 claim concerning the circle of convergence of a Taylor's series with randomly chosen coefficients.

Steinhaus's axioms were measure-theoretic, but they were not yet abstract. His words suggested that his ideas should apply to all sequences of random variables, not merely ones uniformly distributed, and he even considered the case where the variables were complex-valued rather than real-valued, but he did not step outside the geometric context to consider probability on abstract spaces. This step was taken by Stanisław Ulam, one of Steinhaus's junior colleagues at Lwów. At the International Congress of Mathematicians in Zürich in 1932, Ulam announced that he and another Lwów mathematician, Zbigniew Łomnicki (a nephew of Antoni Łomnicki), had shown that product measures can be constructed in abstract spaces (Ulam 1932).

Ulam and Łomnicki's axioms for measure were simple. They assumed that the class of $\mathfrak{M}$ of measurable sets on space X satisfy four conditions:

I. $X \in \mathfrak{M}$, and $\{x\} \in \mathfrak{M}$ for all $x \in \mathfrak{M}$.

II. If $M_i \in \mathfrak{M}$ for $i = 1, 2, \ldots$, then $\cup_{i=1}^\infty M_i \in \mathfrak{M}$.

III. If $M, N \in \mathfrak{M}$, then $M \setminus N \in \mathfrak{M}$.

IV. If $M \in \mathfrak{M}$, $m(M) = 0$, and $N \subset M$, then $N \in \mathfrak{M}$.

And he assumed that the measures $m(M)$ satisfy three conditions:

1. $m(X) = 1$; $m(M) \geq 0$.

2. $m(\cup_{i=1}^\infty M_i) = \sum_{i=1}^\infty m(M_i)$ when $M_i \cap M_j = \emptyset$ for $i \neq j$.

3. If $m(M) = 0$ and $N \subset M$, then $m(N) = 0$.

In today's language, this says that m is a probability measure on a σ-algebra that is complete (includes all null sets) and contains all singletons. Ulam announced that from a countable sequence of spaces with such probability measures, one can construct a probability measure satisfying the same conditions on the product space.

We do not know whether Kolmogorov knew about Ulam's announcement when he wrote the *Grundbegriffe*. Ulam's axioms would have held no novelty for him, but he would presumably have found the result on product measures interesting. Łomnicki and Ulam's article, which appeared only in 1934, cites the *Grundbegriffe*, not Ulam's earlier announcement, when it lists its axioms. Kolmogorov cited the article in 1935.

4.6   Cantelli's abstract theory

Like Borel, Castelnuovo, and Fréchet, Francesco Paolo Cantelli turned to probability after distinguishing himself in other areas of mathematics. His first publications on probability, expositions of the classical theory in a teaching journal, appeared in 1905 and 1906, when he was about 40. His work on the strong law of large numbers came ten years later.

It was only in the 1930s, about the same time as the *Grundbegriffe* appeared, that Cantelli introduced his own abstract theory of probability. This theory, which has important affinities with Kolmogorov's, is developed most clearly in "Una teoria astratta del calcolo delle probabilità", published in 1932 and listed in the *Grundbegriffe*'s bibliography, and in "Considérations sur la convergence dans le calcul des probabilités", a lecture Cantelli gave in 1933 at the Institut Henri Poincaré in Paris and subsequently published in 1935.

In the 1932 article, Cantelli argued for a theory that makes no appeal to empirical notions such as possibility, event, probability, or independence. This abstract theory, he said, should begin with a set of points having finite nonzero measure. This could be any set for which measure is defined, perhaps a set of points on a surface. He wrote $m(E)$ for the area of a subset E. He noted that $m(E_1 \cup E_2) = m(E_1) + m(E_2)$ and $0 \leq m(E_1 E_2)/m(E_i) \leq 1$ for $i = 1, 2$. He called E_1 and E_2 *multipliable* when $m(E_1 E_2) = m(E_1)m(E_2)$. Much of probability theory, he pointed out, including Bernoulli's law of large numbers and Khinchin's law of the iterated logarithm, could be carried out at this abstract level.

In the 1935 article, Cantelli explains how his abstract theory should be related to frequencies in the world. The classical calculus of probability, he says, should be developed for a particular class of events in the world in three steps:

1. Study experimentally the equally likely cases (check that they happen equally frequently), thus justifying experimentally the rules of total and compound probability.

2. Develop an abstract theory based only on the rules of total and compound probability, without reference to their empirical justification.

3. Deduce probabilities from the abstract theory, and use them to predict frequencies.

His own abstract theory, Cantelli explains, is precisely the theory one obtains in the second step. The theory can begin with cases that are not equally likely. But application of the theory, in any case, involves initial verification and subsequent prediction of frequencies. Cantelli reviews earlier Italian discussion of the relation between probability and frequency and quotes with approval his friend and collaborator Guido Castelnuovo, who had explained that limiting frequency should be taken not as a basis for the logical construction of the calculus of probabilities but rather as a way of connecting the calculus to its applications.

A fair evaluation of the importance of Cantelli's role is clouded by the cultural differences that separated him from Kolmogorov, who represented a younger generation, and even from his contemporaries Bernstein, Fréchet, and Slutsky (Benzi 1988, Bru 2003a). Cantelli belonged to an older mathematical culture that had not absorbed the new theory of functions, and so from Fréchet and Kolmogorov's point of view, one could question whether he understood even his own discoveries. Yet he was quick to contrast his own mathematical depth with the shallowness of others; he offended Slutsky with his disdain for Chuprov (Seneta 1992).

We cannot really say that Cantelli's 1932 article and 1933 lecture were sources for the *Grundbegriffe*. The theory in Kolmogorov's 1929 article already went well beyond anything Cantelli did in 1932, in both degree of abstraction (instead of developing an abstract measure theory, Cantelli had simply identified events with subsets of a geometric space for which measure was already defined) and mathematical clarity. The 1933 lecture was more abstract but obviously came too late to influence the *Grundbegriffe*. On the other hand, we may reasonably assume that when Cantelli prepared the 1932 article he had not seen Kolmogorov's 1929 article, which was published in Russian at a time when Russian mathematicians published their best work in German, French, and Italian. So we may say that Cantelli developed independently of Kolmogorov the project of combining a frequentist interpretation of probability with an abstract axiomatization that stayed close to the classical rules of total and compound probability. This project was in the air.

5 The *Grundbegriffe*

When Kolmogorov sat down to write the *Grundbegriffe*, in a rented cottage on the Klyaz'ma River in November 1932, he was already being hailed as the Soviet Euler. Only 29 years of age, his accomplishments within probability theory alone included definitive versions of the law of the iterated logarithm (Kolmogorov 1929b) and the strong law of large numbers (Kolmogorov 1930b), as well as his pathbreaking article on Markov processes. He was personally acquainted with many of the leading mathematicians of his time; his recent trip abroad, from June 1930 to March 1931, had included a stay at Hilbert's department in Göttingen and extensive discussions with Fréchet and Lévy in France (Shiryaev 1989, 2000, 2003a).

The *Grundbegriffe* was an exposition, not another research contribution. In his preface, after acknowledging that Fréchet had shown how to liberate measure and integration from geometry, Kolmogorov said this:

> In the pertinent mathematical circles it has been common for some time to construct probability theory in accordance with this general point of view. But a complete presentation of the whole system, free from superfluous complications, has been missing (though a book by Fréchet, [2] in the bibliography, is in preparation).

Kolmogorov aimed to fill this gap, and he did so brilliantly and concisely, in just 62 pages. Fréchet's much longer book, which finally appeared in two volumes in 1937 and 1938 (Fréchet 1937–1938), is remembered only as a footnote to Kolmogorov's achievement.

Fréchet's own evaluation of the *Grundbegriffe*'s contribution, quoted at the beginning of this article, is correct so far as it goes. Borel had introduced countable additivity into probability in 1909, and in the following twenty years many authors, including Kolmogorov, had explored its consequences. The *Grundbegriffe* merely rounded out the picture by explaining that nothing more was needed: the classical rules together with countable additivity were a sufficient basis for what had been accomplished thus far in mathematical probability. But it was precisely Kolmogorov's mathematical achievement, especially his definitive work on the classical limit theorems, that had given him the grounds and the authority to say that nothing more was needed.

Moreover, Kolmogorov's appropriation of the name *probability* was an important rhetorical achievement, with enduring implications. Slutsky in 1922 and Kolmogorov himself in 1927 had proposed a general theory of additive set functions but had relied on the classical theory to say that probability should be a special case of this general theory. Now Kolmogorov proposed axioms for probability. The numbers in his abstract theory were probabilities, not merely valences or меры. As we explain in §5.3 below, his philosophical justification for proceeding in this way so resembled the justification that Borel, Chuprov, and Lévy had given for the classical theory that they could hardly raise objections.

It was not really true that nothing more was needed. Those who studied Kolmogorov's formulation in detail soon realized that his axioms and definitions were inadequate in a number of ways. Most saliently, his treatment of conditional probability was not adequate for the burgeoning theory of Markov processes, to which he had just made so important a contribution. And there were other points in the monograph where he could not obtain natural results at the abstract level and had to fall back to the classical examples—discrete probabilities and probabilities in Euclidean spaces. But these shortcomings only gave impetus to the new theory, for the project of filling in the gaps provided exciting work for a new generation of probabilists.

In this section, we take a fresh look at the *Grundbegriffe*. In §5.1 we review broadly the contents of the book and the circumstances of its publication. Then, in §5.2, we review the basic framework—the six axioms—and two ideas that were novel at the time: the idea of constructing probabilities on infinite-dimensional spaces (which lead to his celebrated consistency theorem), and the definition of conditional probability using the Radon-Nikodym theorem. Finally, in §5.3, we look at the explicitly philosophical part of the monograph: the two pages in Chapter I where Kolmogorov explains the empirical origin of his axioms.

5.1   An overview

In a letter to his close friend Pavel Sergeevich Aleksandrov, dated November 8, 1932 (Shiryaev 2003b, Vol. 2, pp. 456–458), Kolmogorov explained that he

planned to spend November, December, and January writing the *Grundbegriffe*. The preface is dated Easter, 1933, and the monograph appeared that same year in a series published by Springer, *Ergebnisse der Mathematik und Ihrer Grenzgebiete*.

In 1936, when Communist leaders began to complain, in the context of a campaign against Kolmogorov's teacher Nikolai Luzin, about the Moscow mathematicians' publishing their best work in the West (Demidov and Levshin 1999; Vucinich 2000; Lorentz 2002, p. 205), Kolmogorov had his student Grigory Bavli translate the *Grundbegriffe* into Russian. An English translation, by Nathan Morrison, was published in 1950, with a second edition in 1956. The second Russian edition, published in 1974, modernized the exposition substantially and is therefore less useful for those interested in Kolmogorov's viewpoint in 1933. We have made our own translations from the German original, modernizing the notation in only three minor respects: $\cup$ instead of $+$, $\cap_n$ instead of $\mathfrak{D}_n$, and $\emptyset$ instead of 0 for the empty set.

The monograph has six chapters:

I. **The elementary probability calculus.** This chapter deals with the case where the sample space is finite. Even in this elementary case, Kolmogorov treats conditional probabilities as random variables.

II. **Infinite probability fields.** Here Kolmogorov introduces countable additivity and discusses the two classical types of probability measures, as we now call them: discrete probability measures and probability measures on Euclidean space specified by cumulative distribution functions.

III. **Random variables.** Here Kolmogorov proves his consistency theorem: Given a Cartesian product R^M, where R is the real numbers and M is an arbitrary set, and given consistent distribution functions for all the finite-dimensional marginals, the resulting set function on the field in R^M formed by finite-dimensional cylinder sets is countably additive and therefore extends, by Carathéodory's theorem, to a probability measure on a Borel field. Kolmogorov then discusses convergence in probability, a concept he attributes to Bernoulli and Slutsky.

IV. **Mathematical expectations.** This chapter begins with a discussion, without proofs, of the abstract Lebesgue integral. After explaining that the integral of a random variable is called its mathematical expectation, Kolmogorov discusses Chebyshev's inequality and some criteria for convergence in probability. In a final section, he gives conditions for interchanging the expectation sign with differentiation or integration with respect to a parameter.

V. **Conditional probabilities and expectations.** This is the most novel chapter; here Kolmogorov defines conditional probability and conditional expectation using the Radon-Nikodym theorem.

VI. Independence. Law of large numbers. After defining independence and correlation, Kolmogorov reviews, without proofs, his own and Khinchin's results on the law of large numbers.

There is also an appendix on the zero-one law.

In his preface, after saying that he aimed to present the abstract viewpoint without superfluous complications, Kolmogorov indicated that there was nevertheless some novelty in the book:

> I would also like to call attention here to the points in the further presentation that fall outside the circle of ideas, familiar to specialists, that I just mentioned. These points are the following: probability distributions in infinite-dimensional spaces (Chapter III, §4), differentiation and integration of mathematical expectations with respect to a parameter (Chapter IV, §5), and above all the theory of conditional probabilities and expectations (Chapter V). . . .

We will not discuss the differentiation and integration of expectations with respect to a parameter, for Kolmogorov's results here, even if novel to specialists in probability, were neither mathematically surprising nor philosophically significant. The other two results have subsequently received much more attention, and we take a look at them in the next section.

5.2   The mathematical framework

Kolmogorov's six axioms for probability are so familiar that it seems superfluous to repeat them but so concise that it is easy to do so. We do repeat them (§5.2.1), and then we discuss the two points just mentioned: the consistency theorem (§5.2.2) and the treatment of conditional probability and expectation (§5.2.3). As we will see, a significant amount of mathematics was involved in both cases, but most of it was due to earlier authors—Daniell in the case of the consistency theorem and Nikodym in the case of conditional probabilities and expectations. Kolmogorov's contribution, more rhetorical and philosophical than mathematical, was to bring this mathematics into his framework for probability.

5.2.1   The six axioms

Kolmogorov began with five axioms concerning a set E and a set $\mathfrak{F}$ of subsets of E, which he called *random events*:

I $\mathfrak{F}$ is a field of sets.[12]

II $\mathfrak{F}$ contains the set E.

III To each set A from $\mathfrak{F}$ is assigned a nonnegative real number $\mathsf{P}(A)$. This number $\mathsf{P}(A)$ is called the probability of the event A.

---

[12]For definitions of *field of sets* and *Borel field*, see p. 18.

IV $P(E) = 1$.

V If A and B are disjoint, then

$$P(A \cup B) = P(A) + P(B).$$

He then added a sixth axiom, redundant for finite $\mathfrak{F}$ but independent of the first five axioms for infinite $\mathfrak{F}$:

VI If $A_1 \supseteq A_2 \supseteq \cdots$ is a decreasing sequence of events from $\mathfrak{F}$ with $\bigcap_{n=1}^{\infty} A_n = \emptyset$, then $\lim_{n \to \infty} P(A_n) = 0$.

This is the *axiom of continuity*. Given the first five axioms, it is equivalent to countable additivity.

The six axioms can be summarized by saying that P is a nonnegative additive set function in the sense of Fréchet with $P(E) = 1$.

In contrast with Fréchet, who had debated countable additivity with de Finetti a few years before (Fréchet 1930, de Finetti 1930, Cifarelli and Regazzini 1996), Kolmogorov did not try to make a substantive argument for it. Instead, he made this statement (p. 14):

> ... Since the new axiom is essential only for infinite fields of probability, it is hardly possible to explain its empirical meaning. ... In describing any actual observable random process, we can obtain only finite fields of probability. Infinite fields of probability occur only as idealized models of real random processes. *This understood, we limit ourselves arbitrarily to models that satisfy Axiom VI.* So far this limitation has been found expedient in the most diverse investigations.

This echoes Borel, who adopted countable additivity not as a matter of principle but because he had not encountered circumstances where its rejection seemed expedient (1909a, §I.5). But Kolmogorov was much clearer than Borel about the purely instrumental significance of infinity.

5.2.2 Probability distributions in infinite-dimensional spaces

Suppose, using modern terminology, that $(E_1, \mathfrak{F}_1), (E_2, \mathfrak{F}_2), \ldots$ is a sequence of measurable spaces. For each finite set of indices, say $i_1, \ldots, i_n$, write $\mathfrak{F}^{i_1, \ldots, i_n}$ for the induced σ-algebra in the product space $\prod_{j=1}^{n} E_{i_j}$. Write E for the product of all the E_i, and write $\mathfrak{F}$ for the algebra (not a σ-algebra) consisting of all the cylinder subsets of E corresponding to elements of the various $\mathfrak{F}^{i_1, \ldots, i_n}$. Suppose we define consistent probability measures for all the marginal spaces $(\prod_{j=1}^{n} E_{i_j}, \mathfrak{F}^{i_1, \ldots, i_n})$. This defines a set function on $(E, \mathfrak{F})$. Is it countably additive?

In general, the answer is negative; a counterexample was given by Erik Sparre Andersen and Børge Jessen in 1948. But as we noted in §4.5, Ulam had announced in 1932 a positive answer for the case where all the measures on the

marginal spaces are product measures. Kolmogorov's consistency theorem, in §4 of Chapter III of the *Grundbegriffe*, answered it affirmatively for another case, where each E_i is a copy of the real numbers and each $\mathfrak{F}_i$ consists of the Borel sets. (Formally, though, Kolmogorov had a slightly different starting point: finite-dimensional distribution functions, not finite-dimensional measures.)

In his September 1919 article (Daniel 1919b), Daniell had proven a closely related theorem, using an infinite-dimensional distribution function as the starting point. In Bourbaki's judgment (1994, p. 243), the essential mathematical content of Kolmogorov's result is already in Daniell's. But Kolmogorov did not cite Daniell in the *Grundbegriffe*, and even 15 years later, in 1948 (§3.1), Gnedenko and Kolmogorov ignored Daniell while claiming the construction of probability measures on infinite products as a Soviet achievement. What are we to make of this?

In a commemoration of Kolmogorov's early work, published in 1989, Doob hazards the guess that Kolmogorov was unaware of Daniell's result when he wrote the *Grundbegriffe*. This may be true. He would not have been the first author to do repeat Daniell's result; Børge Jessen presented the result to the seventh Scandinavian mathematical conference in 1929 and became aware of Daniell's priority only in time to acknowledge it in a footnote to the published version (Jessen 1930). Saks did not even mention Daniell's integral in his first edition in 1933. This integral became more prominent later that year and the next, with the increased interest in Wiener's work on Brownian motion,[13] but the articles by Daniell that Wiener had cited did not include the September 1919 article.

In 1935, Wiener and Jessen both called attention to Daniell's priority. So it seems implausible that Gnedenko and Kolmogorov would have remained unaware of Daniell's construction in 1948. But their claim of priority for Kolmogorov may appear more reasonable when we remember that the *Grundbegriffe* was not meant as a contribution to pure mathematics. Daniell's and Kolmogorov's theorems seem almost identical when they are assessed as mathematical discoveries, but they differed in context and purpose. Daniell was not thinking about probability, whereas the slightly different theorem formulated by Kolmogorov was about probability. Neither Daniell nor Wiener undertook to make probability into a conceptually independent branch of mathematics by establishing a general method for representing it measure-theoretically.

Kolmogorov's theorem was more general than Daniell's in one respect—Kolmogorov considered an index set of arbitrary cardinality, rather than merely a denumerable one as Daniell had. This greater generality is merely formal, in two senses; it involves no additional mathematical complications, and it does not seem to have any practical application. The obvious use of a non-denumerable index would be to represent continuous time, and so we might conjecture that Kolmogorov is thinking of making probability statements about trajectories in

---

[13]One small indication that Kolmogorov may have been aware of Wiener's and Daniell's articles before the end of 1933 is the fact that he submitted an article on Brownian motion to the *Annals of Mathematics*, the American journal where Daniell and Wiener had published, in 1933; it appeared in 1934 with the legend "Received September 9, 1933".

continuous time, as Wiener had done in the 1920s but Kolmogorov had not done in 1931. This might have been motivated not only by Bachelier's work on Brownian motion, which Kolmogorov had certainly studied, but also by the work on point processes by Filip Lundberg, Agnar Erlang, and Ernest Rutherford and Hans Geiger in the first decade of the twentieth century (Cramér 1976, p. 513). But Kolmogorov's construction does not accomplish anything in this direction. The σ-algebra on the product obtained by the construction contains too few sets; in the case of Brownian motion, it does not include the set of continuous trajectories. It took some decades of further research were required in order to develop general methods of defining σ-algebras, on suitable function spaces, rich enough to include the infinitary events one typically wants to discuss (Doob 1953; Schwartz 1973; Dellacherie and Meyer 1975; Bourbaki 1994, pp. 243–245). The generally topological character of these richer extensions and the failure of the consistency theorem for arbitrary Cartesian products remain two important caveats to the *Grundbegriffe*'s main thesis—that probability is adequately represented by the abstract notion of a probability measure.

It is by no means clear that Kolmogorov even did prefer to start the study of stochastic processes with unconditional probabilities on trajectories. Even in 1935, he recommended the opposite (Kolmogorov 1935, pp. 168–169 of the English translation). In the preface to the second Russian edition of the *Grundbegriffe* (1974), he acknowledged Doob's innovations with a rather neutral comment: "Nowadays people prefer to define conditional probability with respect to an arbitrary algebra $\mathcal{F}' \subseteq \mathcal{F}$"....

5.2.3   Experiments and conditional probability

In the case where A has nonzero probability, Kolmogorov defined $\mathsf{P}_A(B)$ in the usual way. He called it "bedingte Wahrscheinlichkeit", which translates into English as "conditional probability".

His general treatment of conditional probability and expectation was novel. It began with a set-theoretic formalization of the concept of an *experiment* (*Versuch* in German). Here Kolmogorov had in mind a subexperiment of the grand experiment defined by the conditions $\mathfrak{S}$. This subexperiment might give only limited information about the outcome ξ of the grand experiment. If the subexperiment is well-defined, Kolmogorov reasoned, it should define a partition $\mathfrak{A}$ of the sample space E for the grand experiment: its outcome should amount to specifying which element of $\mathfrak{A}$ contains ξ. He formally identified the subexperiment with $\mathfrak{A}$. Then he introduced the idea of conditional probability relative to $\mathfrak{A}$. As he explained, first in the case of finite fields of probability (Chapter I) and then in the general case (Chapter V), this is a random variable, not a single number:

- In the finite case, he wrote $\mathsf{P}_{\mathfrak{A}}(B)$ for the random variable whose value at each point ξ of E is $\mathsf{P}_A(B)$, where A is the element of $\mathfrak{A}$ containing ξ, and he called this random variable the "conditional probability of B after the experiment $\mathfrak{A}$" (p. 12). This random variable is well defined for all the ξ

in elements $\mathfrak{A}$ that have positive probability, and these ξ form an event that has probability one.

- In the general case, he represented the partition $\mathfrak{A}$ by a function u on E that induces it, and he wrote $\mathsf{P}_u(B)$ for any random variable that satisfies

$$\mathsf{P}_{\{u \subset A\}}(B) = \mathsf{E}_{\{u \subset A\}} \mathsf{P}_u(B) \tag{2}$$

for every set A of possible values of u such that the subset $\{\xi \,|\, u(\xi) \in A\}$ of E (this is what he meant by $\{u \subset A\}$) is measurable and has positive probability (p. 42). By the Radon-Nikodym theorem (only recently proven by Nikodym), this random variable is unique up to a set of probability zero. Kolmogorov called it the "conditional probability of B with respect to (or knowing) u". He defined $\mathsf{E}_u(y)$, which he called "the conditional expectation of the variable y for a known value of u", analogously (p. 46).

Because many readers will be more familiar with Doob's slightly different definitions, it may be wise to add a few words of clarification. The mapping u can be any mapping from E to another set, say F, that represents $\mathfrak{A}$ in the sense that it maps two elements of E to the same element of F if and only if they are in the same element of $\mathfrak{A}$. The choice of u has no effect on the definition of $\mathsf{P}_u(B)$, because (2) can be rewritten without reference to u; it says that the conditional probability for B given $\mathfrak{A}$ is any random variable z such that $\mathsf{E}_C(z) = \mathsf{P}_C(B)$ for every C in $\mathfrak{F}$ that is a union of elements of $\mathfrak{A}$ and has positive probability.

Kolmogorov was doing no new mathematics here; the mathematics is Nikodym's. But Kolmogorov was the first to prescribe that Nikodym's result be used to define conditional probability, and this involves a substantial philosophical shift. The rule of compound probability had encouraged mathematicians of the eighteenth and nineteenth centuries to see the probability of an event A "after B has happened" or "knowing B" as something to be sought directly in the world and then used together with B's probability to construct the joint probability of the two events. Now the joint probability distribution was consecrated as the mathematical starting point—what we abstract from the real world before the mathematics begins. Conditional probability now appears as something defined within the mathematics, not something we bring directly from outside.[14]

5.2.4   When is conditional probability meaningful?

To illustrate his understanding of conditional probability, Kolmogorov discussed Bertrand's paradox of the great circle, which he called, with no specific reference, a "Borelian paradox". His explanation of the paradox was simple but formal. After noting that the probability distribution for the second point conditional on a particular great circle is not uniform, he said (p. 45):

---

[14]George Boole had emphasized the possibility of deriving conditional probabilities from unconditional probabilities, and Freudenthal and Steiner (1966, p. 189) have pointed to him as a precursor of Kolmogorov.

This demonstrates the inadmissibility of the idea of conditional probability with respect to a given isolated hypothesis with probability zero. One obtains a probability distribution for the latitude on a given great circle only when that great circle is considered as an element of a partition of the entire surface of the sphere into great circles with the given poles.

This explanation has become part of the culture of probability theory, even though it cannot completely replace the more substantive explanations given by Borel.

Borel insisted that we explain how the measurement on which we will condition is to be carried out. This accords with Kolmogorov's insistence that a partition be specified, for a procedure for measurement will determine such a partition. Kolmogorov's explicitness on this point was a philosophical advance. On the other hand, Borel demanded more than the specification of a partition. He demanded that the measurement be specified realistically enough that we see real partitions into events with nonzero probabilities, not merely a theoretical limiting partition into events with zero probabilities.

Borel's demand that the theoretical partition into events of probability zero be approximated by a partition into events of positive probability seems to be needed in order to rule out nonsense. This is illustrated by an example given by Lévy in 1959. Suppose the random variable X is distributed uniformly on $[0, 1]$. For each $x \in [0, 1]$, let $C(x)$ be the denumerable set given by

$$C(x) := \{x' \,|\, x' \in [0, 1] \text{ and } x - x' \text{ is rational}\}.$$

Let $\mathfrak{A}$ be the partition

$$\mathfrak{A} := \{C(x) \,|\, x \in [0, 1]\}.$$

What is the distribution of X conditional on, say, the element $C(\frac{\pi}{4})$ of the partition $\mathfrak{A}$? We cannot respond that X is now uniformly distributed over $C(\frac{\pi}{4})$, because $C(\frac{\pi}{4})$ is a denumerable set. Kolmogorov's formal point of view gives an equally unsatisfactory response; it tells us that conditionally on $C(\frac{\pi}{4})$, the random variable X is still uniform on $[0, 1]$, so that the set on which we have conditioned, $C(\frac{\pi}{4})$, has conditional probability zero (see §A.3 below). It makes more sense, surely, to take Borel's point of view, as Lévy did, and reject the question. It makes no practical sense to condition on the zero-probability set $C(\frac{\pi}{4})$, because there is no real-world measurement that could have it, even in an idealized way, as an outcome.

The condition that we consider idealized events of zero probability only when they approximate more down-to-earth events of nonzero probability again brings topological ideas into the picture. This issue was widely discussed in the 1940s and 1950s, often with reference to an example discussed by Jean Dieudonné (1948), in which the conditional probabilities defined by Kolmogorov do not even have versions (functions of ξ for fixed B) that form probability measures (when considered as functions of B for fixed ξ). Gnedenko and Kolmogorov (1949) and David Blackwell (1956) addressed the issue by formulating more or

less topological conditions on measurable spaces or probability measures that rule out pathologies such as those adduced by Dieudonné and Lévy. For modern versions of these conditions, see Rogers and Williams (2000).

5.3   The empirical origin of the axioms

Kolmogorov devoted about two pages of the *Grundbegriffe* to the relation between his axioms and the real world. These two pages, the most thorough explanation of his frequentist philosophy he ever provided, are so important to our story that we quote them in full. We then discuss how this philosophy was related to the thinking of his predecessors.

5.3.1   In Kolmogorov's own words

Section 2 of Chapter I of the *Grundbegriffe* is entitled "Das Verhältnis zur Erfahrungswelt". It is only two pages in length. This subsection consists of a translation of the section in its entirety.

<div align="center">

The relation to the world of experience

</div>

The theory of probability is applied to the real world of experience as follows:

1. Suppose we have a certain system of conditions $\mathfrak{S}$, capable of unlimited repetition.

2. We study a fixed circle of phenomena that can arise when the conditions $\mathfrak{S}$ are realized. In general, these phenomena can come out in different ways in different cases where the conditions are realized. Let E be the set of the different possible variants $\xi_1, \xi_2, \ldots$ of the outcomes of the phenomena. Some of these variants might actually not occur. We include in the set E all the variants we regard *a priori* as possible.

3. If the variant that actually appears when conditions $\mathfrak{S}$ are realized belongs to a set A that we define in some way, then we say that the event A has taken place.

   Example. The system of conditions $\mathfrak{S}$ consists of flipping a coin twice. The circle of phenomena mentioned in point 2 consists of the appearance, on each flip, of head or tails. It follows that there are four possible variants (*elementary events*), namely

<div align="center">

heads—heads, heads—tails, tails—heads, tails—tails.

</div>

Consider the event A that there is a repetition. This event consists of the first and fourth elementary events. Every event can similarly be regarded as a set of elementary events.

4. Under certain conditions, that we will not go into further here, we may assume that an event A that does or does not occur under conditions $\mathfrak{S}$ is assigned a real number $\mathsf{P}(A)$ with the following properties:

A. One can be practically certain that if the system of conditions $\mathfrak{S}$ is repeated a large number of times, n, and the event A occurs m times, then the ratio m/n will differ only slightly from $\mathsf{P}(A)$.

B. If $\mathsf{P}(A)$ is very small, then one can be practically certain that the event A will not occur on a single realization of the conditions $\mathfrak{S}$.

Empirical Deduction of the Axioms. Usually one can assume that the system $\mathfrak{F}$ of events $A, B, C \ldots$ that come into consideration and are assigned definite probabilities form a field that contains E (Axioms I and II and the first half of Axiom III—the existence of the probabilities). It is further evident that $0 \leq m/n \leq 1$ always holds, so that the second half of Axiom III appears completely natural. We always have $m = n$ for the event E, so we naturally set $\mathsf{P}(E) = 1$ (Axiom IV). Finally, if A and B are mutually incompatible (in other words, the sets A and B are disjoint), then $m = m_1 + m_2$, where m, m_1, and m_2 are the numbers of experiments in which the events $A \cup B$, A, and B happen, respectively. It follows that

$$\frac{m}{n} = \frac{m_1}{n} + \frac{m_2}{n}.$$

So it appears appropriate to set $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$.

Remark I. If two assertions are both practically certain, then the assertion that they are simultaneously correct is practically certain, though with a little lower degree of certainty. But if the number of assertions is very large, we cannot draw any conclusion whatsoever about the correctness of their simultaneous assertion from the practical certainty of each of them individually. So it in no way follows from Principle **A** that m/n will differ only a little from $\mathsf{P}(A)$ in every one of a very large number of series of experiments, where each series consists of n experiments.

Remark II. By our axioms, the impossible event (the empty set) has the probability $\mathsf{P}(\emptyset) = 0$. But the converse inference, from $\mathsf{P}(A) = 0$ to the impossibility of A, does not by any means follow. By Principle **B**, the event A's having probability zero implies only that it is practically impossible that it will happen on a particular unrepeated realization of the conditions $\mathfrak{S}$. This by no means implies that the event A will not appear in the course of a sufficiently long series of experiments. When $\mathsf{P}(A) = 0$ and n is very large, we can only say, by Principle **A**, that the quotient m/n will be very small—it might, for example, be equal to $1/n$.

5.3.2 The philosophical synthesis

The philosophy set out in the two pages we have just translated is a synthesis, combining elements of the German and French traditions in a way not previously attempted.

By his own testimony, Kolmogorov drew first and foremost from von Mises. In a footnote, he put the matter this way:

> . . . In laying out the assumptions needed to make probability theory applicable to the world of real events, the author has followed in large measure the model provided by Mr. von Mises. . . .

The very title of this section of the *Grundbegriffe*, "Das Verhältnis zur Erfahrungswelt", echoes the title of the section of von Mises's 1931 book that Kolmogorov cites, "Das Verhältnis der Theorie zur Erfahrungswelt". But Kolmogorov does not discuss Kollektivs. As he explained in his letter to Fréchet in 1939 (translated in Appendix A.2), he thought only a finitary version of this concept would reflect experience truthfully, and a finitary version, unlike the infinitary version, could not be made mathematically rigorous. So for mathematics, one should adopt an axiomatic theory, "whose practical value can be deduced directly" from a finitary concept of Kollektivs.

Although Kollektivs are in the background, Kolmogorov starts in a way that echoes Chuprov more than von Mises. He writes, as Chuprov did (1910, p. 149), of a system of conditions (Komplex von Bedingungen in German, комплекс условий in Russian). Probability is relative to a system of conditions $\mathfrak{S}$, and yet further conditions must be satisfied in order for events to be assigned a probability under $\mathfrak{S}$. Kolmogorov says nothing more about these conditions, but we may conjecture that he was thinking of the three sources of probabilities mentioned by von Mises: chance devices designed for gambling, statistical phenomena, and physical theory.

Like the German objectivists, Kolmogorov did not think that every event has a probability. Later, in his article on probability in the second edition of the *Great Soviet Encyclopedia*, published in 1951, he was even more explicit about this:

> Certainly not every event whose occurrence is not uniquely determined under given conditions has a definite probability under these conditions. The assumption that a definite probability (i.e. a completely defined fraction of the number of occurrences of an event if the conditions are repeated a large number of times) in fact exists for a given event under given conditions is a *hypothesis* which must be verified or justified in each individual case.

Where do von Mises's two axioms—probability as a limit of relative frequency and its invariance under selection of subsequences—appear in Kolmogorov's account? Principle **A** is obviously a finitary version of von Mises's axiom identifying probability as the limit of relative frequency. Principle **B**, on the other hand, is the strong form of Cournot's principle (see §2.2.2 above). Is it a finitary version of von Mises's principle of invariance under selection? Perhaps so. In a Kollektiv, von Mises says, we have no way of singling out an unusual infinite subsequence. One finitary version of this is that we have no way of singling out an unusual single trial. So when we do select a single trial (a single realization of the conditions $\mathfrak{S}$, as Kolmogorov puts it), we should not expect anything unusual. In the special case where the probability is very small, the usual is that the event will not happen.

Of course, Principle **B**, like Principle **A**, is only satisfied when there is a Kollektiv—i.e., under certain conditions. Kolmogorov's insistence on this point is confirmed by his comments in 1956, quoted on p. 11 above, on the importance and nontriviality of the step from "usually" to "in this particular case".

As Borel and Lévy had explained so many times, Principle **A** can be deduced from Principle **B** together with Bernoulli's law of large numbers, which is a consequence of the axioms (see our discussion of Castelnuovo and Fréchet's use of this idea on p. 10). But in the framework that Kolmogorov sets up, the deduction requires an additional assumption; we must assume that Principle **B** applies not only to the probabilities specified for repetitions of conditions $\mathfrak{S}$ but also to the corresponding probabilities (obtaining by assuming independence) for repetitions of n-fold repetitions of $\mathfrak{S}$. It is not clear that this additional assumption is appropriate, not only because we might hesitate about independence (see Shiryaev's comments in the third Russian edition of the *Grundbegriffe*, p. 120), but also because the enlargement of our model to n-fold repetitions might involve a deterioration in its empirical precision, to the extent that we are no longer justified in treating its high-probability predictions as practically certain. Perhaps these considerations justify Kolmogorov's presenting Principle **A** as an independent principle alongside Principle **B** rather than as a consequence of it.

Principle **A** has an independent status in Kolmogorov's story, however, even if we do regard it as a consequence of Principle **B** together with Bernoulli's law of large numbers, because it comes into play at a point that precedes the adoption of the axioms and hence precedes the derivation of the law of large numbers: it is used to motivate (empirically deduce) the axioms (cf. Bartlett 1949). The parallel to the thinking of Hadamard and Lévy is very striking. In their picture, the idea of equally likely cases motivated the rules (the rules of total and compound probability), while Cournot's principle linked the resulting theory with reality. The most important change Kolmogorov makes in this picture is to replace equally likely cases with frequency; frequency now motivates the axioms of the theory, but Cournot's principle is still the essential link with reality.

In spite of the obvious influence of Borel and Lévy, Kolmogorov cites only von Mises in this section of the *Grundbegriffe*. Philosophical works by Borel and Lévy, along with those by Slutsky and Cantelli, do appear in the *Grundbegriffe*'s bibliography, but their appearance is explained only by a sentence in the preface: "The bibliography gives some recent works that should be of interest from a foundational viewpoint." The emphasis on von Mises may have been motivated in part by political prudence. Whereas Borel and Lévy persisted in speaking of the subjective side of probability, von Mises was an uncompromising frequentist. Whereas Chuprov and Slutsky worked in economics and statistics, von Mises was an applied mathematician, concerned more with aerodynamics than social science, and the relevance of his work on Kollektivs to physics had been established in the Soviet literature by Khinchin in 1929 (see also Khinchin, 1961, and Siegmund-Schultze, 2004). (For more information on the political context, see Blum and Mespoulet 2003; Lorentz 2002; Mazliak 2003; Seneta 2003;

Vucinich 1999, 2000, 2002.) Certainly, however, Kolmogorov's high regard for von Mises's theory of Kollektivs was sincere. It emerged again in the 1960s, as a motivation for Kolmogorov's thinking about the relation between algorithmic complexity and probability (Cover et al. 1989, Vovk and Shafer 2003).

The evident sincerity of Kolmogorov's frequentism, and his persistence in it throughout his career, gives us reason to think twice about the proposition, which we discussed on p. 44 that he had in mind, when he wrote the *Grundbegriffe*, a picture in which a Markov process is represented not by transition probabilities, as in his 1931 article, but by a probability space in which one can make probability statements about trajectories, as in Wiener's work. The first clause in Kolmogorov's frequentist credo is that the system of conditions 𝕾 should be capable of unlimited repetition. If trajectories are very long, then this condition may be satisfied by transition probabilities (it may be possible to repeatedly start a system in state x and check where it is after a certain length of time) without being satisfied to the same degree by trajectories (it may not be practical to run through more than a single whole trajectory). So transition probabilities might have made more sense to Kolmogorov the frequentist than probabilities for trajectories.

6 Reception

Mythology has it that the impact of the *Grundbegriffe* was immediate and revolutionary. The truth is more ordinary. Like every important intellectual contribution, Kolmogorov's axiomatization was absorbed slowly. Established scholars continued to explore older approaches and engage in existing controversies, while younger ones gradually made the new approach the basis for new work. The pace of this process was influenced not only by the intellectual possibilities but also by the turmoil of the times. Every mathematical community in Europe, beginning with Germany's, was severely damaged by Hitler's pogroms, Stalin's repression, and the ravages of the second world war.[15]

We cannot examine these complexities in depth, but we will sketch in broad strokes how the *Grundbegriffe* came to have the iconic status it now enjoys. In §6.1 we look at how Kolmogorov's axiomatization became the accepted framework for mathematical research in probability, and in §6.2 we look at how the philosophical interpretation of the framework evolved.

---

[15]Many of the mathematicians in our story were direct victims of the Nazis. Halbwachs, Antoni Łomnicki, and Saks were murdered. Doeblin and Hausdorff committed suicide to escape the same fate. Bavli died at the front. Banach died from his privations just after the war ended. Loéve, arrested and interned at Drancy, was saved only by the liberation of Paris. Castelnuovo, Lévy, Schwartz, and Steinhaus survived in hiding. Others, including Łukasiewicz, Nikodym, Sierpiński, and Van Dantzig, merely lost their jobs during the war. Those who fled Europe included Felix Bernstein, Carnap, Einstein, Feller, Hadamard, Popper, Rademacher, Reichenbach, von Mises, and Wald.

6.1 Acceptance of the axioms

Although the *Grundbegriffe* was praised by reviewers, its abstract measure-theoretic approach continued to co-exist, during the 1930s, with other approaches. The controversy concerning von Mises's concept of Kollektivs continued, and although most mathematicians weighed in against using Kollektivs as the foundation of probability, Kolmogorov's axioms were not seen as the only alternative. Only after the second world war did the role of the abstract measure-theoretic framework become dominant in research and advanced teaching.

6.1.1 Some misleading reminiscences

Reminiscences of mathematicians active in the 1930s sometimes give the impression that the acceptance of Kolmogorov's framework was immediate and dramatic, but these reminiscences must be taken with a grain of salt and a large dose of context.

Particularly misleading is a passage (p. 67–68) in the memoir Lévy published in 1970, when he was in his eighties:

> Starting in 1924, I gradually became accustomed to the idea that one should not consider merely what I called the true probability laws. I tried to extend a true law. I arrived at the idea, arbitrary as it might be, of a law defined in a certain Borel field. I did not think of saying to myself that this was the correct foundation for the probability calculus; I did not have the idea of publishing so simple an idea. Then, one day, I received A. Kolmogorov's tract on the foundations of the probability calculus. I realized what a chance I had lost. But it was too late. When would I ever be able to figure out which of my ideas merited being published?

In evaluating these words, one must take account of the state of mind Lévy displays throughout the memoir. Nothing was discovered in probability in his lifetime, it seems, that he had not thought of first.

In fact, Lévy was slow, in the 1930s, to mention or use the *Grundbegriffe*. Barbut, Locker, and Mazliak (2004, pp. 155–156) point out that he seems to have noticed its zero-one law only in early 1936. Although he followed Fréchet in paying lip service to Kolmogorov's axiomatization (Lévy 1937, 1949) and ruefully acknowledged the effectiveness of Doob's use of it (Lévy 1954), he never used it in his own research (Schwartz 2001, p. 82).

A frequently quoted recollection by Mark Kac about his work with his teacher Hugo Steinhaus can also give a misleading impression concerning how quickly the *Grundbegriffe* was accepted:

> ...Our work began at a time when probability theory was emerging from a century of neglect and was slowly gaining acceptance as a respectable branch of pure mathematics. This turnabout came as a

result of a book by the great Soviet mathematician A. N. Kolmogorov on foundations of probability theory, published in 1933. (Kac 1985, pp. 48–49)

This passage is followed by a sentence that is quoted less frequently: "It appeared to us awesomely abstract." On another occasion, Kac wrote,

> . . . The appearance in 1933 of the book by Kolmogorov on foundations of probability theory was little noticed in Lwów, where I was a student. I do remember looking at the book, but I was frightened away by its abstractness, and it did not seem to have much to do with what I had been working on. I was wrong, of course. . . (Kac 1982, p. 62)

It is hard to guess from Kac's recollections what his more mature colleagues, Steinhaus and Ulam, thought about the importance of the *Grundbegriffe* at the time.

6.1.2   First reactions

The most enthusiastic early reviews of the *Grundbegriffe* appeared in 1934: one by Willy Feller, of Kiel, in the *Zentralblatt* and one by Henry L. Rietz, of Iowa, in the *Bulletin of the American Mathematical Society*. Feller wrote:

> The calculus of probabilities is constructed axiomatically, with no gaps and in the greatest generality, and for the first time systematically integrated, fully and naturally, with abstract measure theory. The axiom system is certainly the simplest imaginable. . . . The great generality is noteworthy; probabilities in infinite dimensional spaces of arbitrary cardinality are dealt with. . . . The presentation is very precise, but rather terse, directed to the reader who is not unfamiliar with the material. Measure theory is assumed.

Rietz was more cautious, writing that "This little book seems to the reviewer to be an important contribution directed towards securing a more logical development of probability theory."

Other early reviews, by the German mathematicians Karl Dörge (1933) and Gustav Doetsch (1935) were more neutral. Both Dörge and Doetsch contrasted Kolmogorov's approach with Tornier's, which stayed closer to the motivating idea of relative frequency and so did not rely on the axiom of countable additivity.

Early authors to make explicit use of Kolmogorov's framework included the Polish mathematicians Zbigniew Łomnicki and Stanisław Ulam, the Austrian-American mathematician Eberhard Hopf, and the American mathematician Joseph L. Doob. We have already mentioned Łomnicki and Ulam's 1934 article on product measures. Hopf used Kolmogorov's framework in his work on the method of arbitrary functions, published in English in 1934. Doob, in two articles

in 1934, used the framework in an attempt to make Ronald Fisher and Harold Hotelling's results on maximum likelihood rigorous.

The French probabilists did not take notice so quickly. The first acknowledgement may have been in comments by Fréchet in a talk at the International Congress of Mathematicians in Oslo in 1936:

> The foundations of probability theory have changed little. But they have been enriched by particulars about the additivity of the probabilities of denumerable sets of incompatible events. This revision was completely and systematically exposited for the first time by A. Kolmogorov. But on this point we mention, on the one hand, that one would find the elements of such an exposition scattered about in the previous publications of many of the authors who have written about the calculus of probabilities in recent years. And on the other hand, if this revision was needed, it was because of the revolution brought about by Émile Borel, first in analysis by his notion of measure, and then in probability theory by his theory of denumerable probabilities.[16]

6.1.3 The situation in 1937

In the years immediately following the publication of the *Grundbegriffe*, the burning issue in the foundation of the probability was the soundness and value of von Mises's concept of a Kollektiv. Although von Mises had been forced to leave Berlin for Istanbul when Hitler took power in 1933, he had continued to argue that any satisfactory mathematical foundation for probability would have to begin with Kollektivs. In 1936, in the second edition of his *Wahrscheinlichkeit, Statistik und Wahrheit*, he specifically criticized Kolmogorov's way of avoiding them (pp. 124ff).

The rejection of von Mises's viewpoint by most mathematicians, especially the French, became clear at the colloquium on mathematical probability held in Geneva in October 1937 (Wavre 1938–1939).[17] Wald presented his demonstration of the existence of Kollektivs at this colloquium, and it was accepted. But the mathematicians speaking against Kollektivs as a foundation for probability included Cramér, de Finetti, Feller, Fréchet, and Lévy. No one, it seems, rose to defend von Mises's viewpoint.[18] Fréchet summed up, in a contribution rewritten after the colloquium, with the words we quoted at the beginning of this article; the classical axioms, updated by including countable additivity, remained the

---

[16]These comments are on pp. 153–154 of the *Les mathématiques et le concret*, a collection of essays that Fréchet published in 1955. The essay in which they appear is not in the proceedings of the Oslo meeting, but it is identified in the 1955 collection as having been presented at that meeting.

[17]The colloquium was very broad, both in topics covered and in nationalities represented. But its location dictated a French point of view. Borel was the honorary chair. Most of the written contributions, including those by Feller and Neyman, were in French.

[18]Neither von Mises nor Kolmogorov were able to attend the Geneva meeting, but both made written contributions, including, in von Mises's case, comments on the foundational issues.

best starting point for probability theory, and Kolmogorov had provided the best exposition.

Willy Feller, now working with Harald Cramér in Stockholm, was the most enthusiastic adherent of the *Grundbegriffe* at Geneva. Nearly 30 years younger than Fréchet, even a few years younger than Kolmogorov himself, Feller praised the *Grundbegriffe* without stopping to discuss how much of it had already been formulated by earlier authors. Kolmogorov's well known axiomatization was, Feller declared, the point of departure for most modern theoretical research in probability (Feller 1938, p. 8).[19] In Feller's view, Kolmogorov had demonstrated the sufficiency of additivity for probability theory, something that had been doubted not so long before. General opinion held, Feller said, that Kolmogorov's axiomatization covered too vast a field. But could anyone demonstrate the existence of a relation between probabilities that the axiomatization did not take into account (p. 13)?

Others at the colloquium paid less attention to Kolmogorov's axiomatization. Jerzy Neyman, for example, prefaced his contribution, which was concerned with statistical estimation, with an exposition of the modernized classical theory that did not mention Kolmogorov. He provided his own resolution of the problem of defining P(A | B) when B has probability zero. In a similar article read to the Royal Society in 1937 (Neyman 1937), Neyman did mention the *Grundbegriffe*, calling it "a systematic outline of the theory of probability based on that of measure", but also cited Borel, Lévy, and Fréchet, as well as Łomnicki and Ulam. Rejection of Kollektivs did not, evidently, necessarily mean a thorough acceptance of Kolmogorov's framework. Many who rejected Kollektivs could agree that Kolmogorov had modernized the classical theory by showing how it becomes abstract measure theory when countable additivity is added. But no real use was yet being made of this level of abstraction.

The Italians, for whom Cantelli represented the modernized classical alternative to von Mises, paid even less attention to the *Grundbegriffe*. Throughout the 1930s, Cantelli debated von Mises, developed his own abstract theory, and engaged other Italian mathematicians on the issues it raised, all without citing Kolmogorov (Regazzini 1987b, pp. 15–16).

Especially striking is the absence of mention, even by enthusiasts like Feller, of Kolmogorov's treatment of conditional probability. No one, it appears, knew what to do with this idea in 1937. Neyman did something simpler, and the French generally preferred to stick with the classical approach, taking the rule of compound probability as an axiom (Lévy 1937, Ville 1939). Paul Halmos recalls that Kolmogorov's approach to conditional probability was puzzling and unconvincing to him (Halmos 1985, p. 65), and this was probably the dominant

---

[19]Feller's contribution to the Geneva meeting, was mainly devoted to reconciling Tornier's theory, which followed von Mises in defining probability as a limiting relative frequency, with Kolmogorov's axiomatization. Feller had collaborated in Kiel with Tornier, who became an ardent Nazi (Segal 2003, p. 150). According to Abraham Frankel (1967, p. 155), Tornier forced Feller out of his job by revealing Feller's Jewish ancestry. Doob (1972), on the other hand, reports that Feller lost his job by refusing to sign a loyalty oath. Feller spent about five years with Cramér in Stockholm before emigrating to the United States in 1939.

attitude, even among the most able mathematicians.

6.1.4 Triumph in the textbooks

The first textbook author to sign on to Kolmogorov's viewpoint was the Swedish statistician Harald Cramér, who was ten years older than Kolmogorov. Cramér explicitly subscribed to Kolmogorov's system in his *Random Variables and Probability Distributions*, published in 1937, and his *Mathematical Methods of Statistics*, written during the second world war and published in 1946. In these books, we find the now standard list of axioms for measure, expressed in terms of a set function $P(S)$, and this can be contrasted with the old fashioned discussion of the classical rules for probabilities $\Pr.(E)$ that we still find in the second edition (1950) of Fréchet's treatise and in the textbook, also published in 1950, of his student Robert Fortet. In substance, however, Cramér made no use of the novelties in the *Grundbegriffe*. As a statistician, he had no need for the strong laws, and so his books made no use of probability distributions on infinite-dimensional spaces. He also had no need for Kolmogorov's treatment of conditional probability; in 1937 he did not even mention conditional probability, and in 1946 he treated it in an old-fashioned way.

In the Soviet Union, it was Boris Gnedenko, who became Kolmogorov's student in 1934, who brought the spirit of the *Grundbegriffe* to an introductory textbook in probability and statistics for advanced students. Gnedenko's Курс теории вероятностей, published in 1950, treated the strong laws as well as statistics and followed the *Grundbegriffe* closer than Cramér had. With the translation of the *Grundbegriffe* itself into English in 1950, the appearance of Loève's *Probability Theory* in 1955, and the translation of Gnedenko's second edition into English in 1962, Kolmogorov's axioms became ubiquitous in the advanced teaching of probability in English.

6.1.5 Making it work

The irresistibility of an axiomatic presentation, even a superficial one, for the advanced teaching of probability was part and parcel of the continuing advance of the spirit of Hilbert, which called for the axiomatization of all mathematics on the basis of set theory. Driven by this same spirit, a younger generation of mathematicians undertook, in the late 1930s and on through the war years, to correct the shortcomings of Kolmogorov's framework.

Bru (2002, pp. 25–26) recounts the development of this enterprise in the Borel Seminar in Paris in the late 1930s. Though officially run by Borel, who retained his chair until 1941, the seminar was actually initiated in 1937 by Ville, Borel's assistant, in the image of Karl Menger's seminar in Vienna, where Ville had developed his ideas on martingales. The initial theme of the seminar was the same as Menger's theme when Ville had visited there in 1934–1935: the foundations of the calculus of probabilities. It was the younger people in the seminar, especially Ville, Wolfgang Doeblin, Robert Fortet, and Michel Loève, who argued, against Borel and Lévy's resistance, for Kolmogorov's axiomatiza-

tion. In 1938, the Borel Seminar turned to the fundamental mathematical issue, which also occupied Doob and the Russian school: how to extend the axiomatic theory to accommodate continuous-time processes.

The French work on Kolmogorov's framework continued through and after the war, under the leadership of Fortet, who was handicapped, as Bru explains, by the disdain of Bourbaki:

> ... Bourbaki having decreed that the measurable structure—neither sufficiently algebraic nor sufficiently topological—was unacceptable, Kolmogorov's axiomatization was hardly taken seriously, and one knows how damaging this was for the development of probability theory in France in the fifties and sixties. Fortet felt a certain bitterness about this, though he formulated it discreetly and always with humor.

The approach favored by Bourbaki can be seen in the treatise *Radon Measures on Arbitrary Spaces and Cylindrical Measures*, published in 1973 by Laurent Schwartz, Lévy's son-in-law and one of the members of the Bourbaki group most engaged with probability (see also Weil 1940, Schwartz 1997).

It was Doob who finally provided the definitive treatment of stochastic processes within the measure-theoretic framework, in his *Stochastic Processes* (1953). Not only did Doob show how to construct satisfactory probability spaces for stochastic processes; he also enriched the abstract framework with the concept of a filtration and with a measure-theoretic version of Ville's concept of a martingale, and he gave conditions for the existence of regular conditional distributions. It would go far beyond the scope of this article to analyze Doob's achievement. Suffice it to say that although Bourbaki was never persuaded, leaving probability still outside the mainstream of pure mathematics, Lévy eventually acknowledged defeat (in the second edition of his *Théorie de l'addition des variables aléatoires*, 1954), and Doob's elaboration of Kolmogorov's framework became the generally accepted mathematical foundation for probability theory.

We do not know what Kolmogorov thought of Doob's innovations. In the preface to the second Russian edition of the *Grundbegriffe* (1974), he acknowledged Doob's innovation with a rather neutral comment: "Nowadays people prefer to define conditional probability with respect to an arbitrary σ-algebra $\mathcal{F}' \subseteq \mathcal{F}$"....

6.2 The evolution of the philosophy of probability

Whereas Kolmogorov's axiomatization became standard, his philosophy of probability showed little staying power. We do find a version of his philosophy in a textbook by Cramér. We also find Cournot's principle, in forms advocated by Borel, Lévy, and Chuprov, flourishing into the 1950s. But aside from Cramér's imitation, we find hardly any mention of Kolmogorov's philosophy in the subsequent literature, and even Cournot's principle faded after the 1950s.

6.2.1 Cramér's adaptation of Kolmogorov's philosophy

Harald Cramér, who felt fully in tune with Kolmogorov's frequentism, repeated the key elements of his philosophy in his 1946 book (pp. 148–150). Cramér expressed Kolmogorov's caution that the theory of probability applies only under certain conditions by saying that only certain experiments are random. In the context of a random experiment $\mathfrak{E}$, Cramér stated Kolmogorov's Principle **A** in this way:

> *Whenever we say that the probability of an event E with respect to an experiment $\mathfrak{E}$ is equal to P, the concrete meaning of this assertion will thus simply be the following: In a long series of repetitions of $\mathfrak{E}$, it is practically certain that the frequency of E will be approximately equal to P. — This statement will be referred to as the frequency interpretation of the probability P.*

He stated Kolmogorov's Principle **B** as a principle applying to an event whose probability is very small or zero:

> *If E is an event of this type, and if the experiment $\mathfrak{E}$ is performed one single time, it can thus be considered as practically certain that E will not occur.* — This particular case of the frequency interpretation of a probability will often be applied in the sequel.

The final sentence of this passage shows Cramér to be a less careful philosopher than Kolmogorov, for it suggests that Principle **B** is a particular case of Principle **A**, and this is not strictly true. As we noted when discussing Castelnuovo's views in §2.2.2, the weak form of Cournot's principle is indeed a special case of Principle **A**. But Principle **B** is the strong form of Cournot's principle, and this is not merely a special case of Principle **A**.

Cramér's textbook was influential across the world; it was even translated into Russian. But its influence was not enough to keep Kolmogorov's Principles **A** and **B** alive. Gnedenko's textbook emphasized that probability is relative to a certain complex of conditions $\mathfrak{S}$, but it did not mention the two principles. Aside from Cramér's book, we do not know of any other textbook or philosophical discussion where Kolmogorov's formulation of these principles is imitated or discussed.

This disappearance of interest in Kolmogorov's philosophical formulation among mathematicians working in probability theory may be due in part to a lack of interest in philosophical issues altogether. Doob, for example, explained in a debate with von Mises in 1941 that he rejected all philosophical explication; the meaning of a probability statement should be left to those who actually use probability in the real world. It is also of significance, however, that Kolmogorov's philosophical formulation quickly encounters difficulties once we follow Doob in taking the probability space Ω to be the set of trajectories for a stochastic process. In many applications, contrary to the assumption with which Kolmogorov begins, the experiment that produces the trajectory may not be capable of repetition.

6.2.2 Cournot's principle after the *Grundbegriffe*

As we explained in §2.2.1, Borel and Lévy had a simpler philosophy of probability than Kolmogorov's; they used only Principle **B**, Cournot's principle. Their viewpoint flourished in the first couple of decades after the *Grundbegriffe*.

Borel, probably the most influential advocate of the principle, sharpened his statement of it in the 1940s. In earlier years, he wrote frequently about the practical meaning of probabilities very close to zero or one, but it is hard to discern in these writings the philosophical principle, which we do find in Hadamard and Lévy, that interpreting a very small probability as impossibility is the unique way of bringing probability theory into contact with the real world. But in the 1940s, we find the principle articulated very clearly. In his 1941 book, *Le jeu, la chance et les théories scientifiques modernes*, he calls it the "fundamental law of chance" (la loi fondamentale du hasard). Then, in 1943, on the first page of the text of his "Que sais-je?" volume, *Les probabilités et la vie*, he finally coined the name he used thereafter: "the only law of chance" (la loi unique du hasard). This name appears again in the 1948 edition of *Le Hasard* and the 1950 edition of *Éléments de la théorie des probabilités* (see also Borel 1950). It was also popularized by Robert Fortet, in his essay in François Le Lionnais's *Les grands courants de la pensée mathématique*, first published in 1948.

Borel's books from this period, as well as Le Lionnais's, were translated into English. They may have influenced Marshall Stone, one of the few American mathematicians to echo the idea that probability theory makes contact with reality only by the prediction that events of small probability will not happen (Stone 1957). Per Martin-Löf has told us that he also learned the idea from reading Borel (see Martin-Löf 1969, p. 616; 1966–1967, p. 9; 1969–1970, p. 8).

Borel never used the name "Cournot's principle"; nor did Lévy or Kolmogorov. The name seems to be due to Fréchet, but the reference to Cournot can be traced back to Chuprov, who considered Cournot the founder of the philosophy of modern statistics (Sheynin 1996, p. 86). As we mentioned earlier, Chuprov called the principle Cournot's "lemma". This enthusiasm for Cournot and the principle was brought from Russian into German by Chuprov's student, Oskar Anderson, who spent the 1930s in Sofia and then moved to Munich in 1942. Anderson called the principle the "Cournotsche Lemma" or the "Cournotsche Brücke"—a bridge between mathematics and the world. We find both phrases already in Anderson's 1935 book, but the book may have been less influential than an article Anderson contributed to a special issue of the Swiss philosophy journal *Dialectica* in 1949, alongside articles by Borel and Lévy revisiting their versions of Cournot's principle. Fréchet took these articles as one of his themes in his presiding report at the session on probability at the Congrès International de Philosophie des Sciences at Paris that same year (Fréchet 1951), where he accepted the attribution of the principle to Cournot ("bien qu'il semble avoir été déjà plus ou moins nettement indiqué par d'Alembert") but suggested the appellation "principe de Cournot", reserving "lemma" as a purely mathematical term. It was normal for Fréchet to legislate on terminology; from 1944

to 1948 he had led the effort by the Association Française de Normalisation to standardize probability terminology and notation, putting in place appellations such as Borel-Cantelli, Kolmogorov-Smirnov, etc. (Bru 2003b, Pallez 1949). Fréchet had second thoughts about giving so much credit to Cournot; when he reprinted his 1949 report as a chapter in a book in 1955, he replaced "principe de Cournot" with "principe de Buffon-Cournot". But here no one else appears to have followed his example.

Both Anderson and the Dutch mathematical statistician David Van Dantzig argued for using Cournot's principle as the foundation for statistical testing: Anderson in *Dialectica* (Anderson 1949), and Van Dantzig at the meeting in Paris (Van Dantzig 1951). Neyman found this view of statistical testing incomprehensible; at the same meeting in Paris he said Anderson was the "only contemporary author I know who seems to believe that the inversion of the theorem of Bernoulli is possible" (Neyman 1951, p. 90). The German mathematical statistician Hans Richter, also in Munich, emphasized Cournot's principle in his own contributions to *Dialectica* (Richter 1954; von Hirsch 1954) and in his probability textbook (Richter 1956), which helped bring Kolmogorov's axioms to students in postwar Germany. As a result of Richter's book, the name "Cournotsche Prinzip" is fairly widely known among probabilists in Germany.

The name "Cournot's principle" was brought into English by Bruno de Finetti (1951), who ridiculed it in several different languages as "the so-called principle of Cournot". De Finetti participated in the 1949 Paris conference, and he presumably remembered that Fréchet had coined the phrase. His disdain for the name may have been influenced, however, by the similarity between the names Cournot and Carnot. When mentioning the principle in his 1970 book (p. 181), de Finetti cites a note, dating back to 1930, in which Borel mentions "Carnot's principle"—the second law of thermodynamics—as an example where a small probability should be interpreted as an impossibility. In spite of his tone, de Finetti has his own way of making sense of the principle. As he explained in a note written in 1951 for Fréchet's *Les mathématiques et le concret*, he did not really disagree with the statement that one should act as if an event with a very small probability should not happen. Rather he took the principle as a tautology, a consequence of the subjective definition of probability, not a principle standing outside probability theory and relating it to the real world (de Finetti 1955, p. 235; see also Dawid 2004).

6.2.3   The new philosophy of probability

Why did no one discuss Kolmogorov's philosophy after Cramér, and why did discussion even of Cournot's principle fade after the 1950s?

Those who do not consider Cournot's principle useful or coherent may consider this explanation enough for its fading, but there are also sociological explanations. The discussions of Cournot's principle that we have just reviewed were discussions by mathematicians. But since the publication of the *Grundbegriffe*, philosophy has been more and more a matter for professionals. As we noted in §2.2.3, this was already true in Germany before the *Grundbegriffe*. German

and Austrian philosophers had worked on probability for over 50 years, from the 1880s to the 1930s; they had endlessly debated subjective and objective interpretations, the meaning of possibility, and the relation of probability to logic and to induction. This tradition was rich and autonomous; it had left the mathematicians and their preoccupations behind. In particular, since von Kries it had had little truck with Cournot's principle.

Because of its mathematical austerity, the *Grundbegriffe* was scarcely noticed by philosophers when it appeared. According to Ville (1985), it had received little attention at the Menger seminar in Vienna, the mathematical satellite of the Vienna circle, where Popper and other philosophers rubbed shoulders with mathematicians such as Carathéodory, Gödel, von Mises, and Wald (Hacohen 2000).

When Kolmogorov finally is mentioned in the philosophical literature, in Ernest Nagel's *Principles of the Theory of Probability* in 1939, he is merely listed as one of many mathematicians who have given axioms for probability. No mention is made of the philosophy he spelled out in the *Grundbegriffe*. For Nagel, Kolmogorov's work was pure mathematics, and it was his job to propose different interpretations of it—frequentist, subjectivist, or whatever (Nagel 1939, pp. 40–41):

> ...Abstract sets of postulates for probability have been given by Borel, Cantelli, Kolmogoroff, Popper, Reichenbach, and several other writers. ...From an abstract mathematical point of view, the probability calculus is a chapter in the general theory of measurable functions, so that the mathematical theory of probability is intimately allied with abstract point-set theory. This aspect of the subject is under active investigation and has been especially cultivated by a large number of French, Italian, Polish, and Russian mathematicians.

Nagel then listed nine possible interpretations. Kolmogorov would have agreed, of course, that his axioms have many interpretations. He said this on p. 1 of the *Grundbegriffe*:

> As we know, every axiomatic (abstract) theory can be interpreted in an unlimited number of ways. The mathematical theory of probability, in particular, has numerous interpretations in addition to those from which it grew. Thus we find applications of the mathematical theory of probability to areas of research that have nothing to do with the ideas of chance and probability in their concrete senses.[20]

But for Kolmogorov, the concrete sense of chance and probability was frequentist. The other applications were about something else (areas or volumes, for example), not probability.

---

[20]The next sentence states that an appendix is devoted to such applications, but there is no such appendix, and the sentence was omitted from the 1936 translation into Russian and subsequently from the 1950 translation into English.

The philosophers of probability most prominent in the United States after the second world war, Carnap and Reichenbach, were themselves refugees from Germany and Austria, even more steeped in the German tradition. They believed that they had new things to say, but this was because they had new philosophical tools, not because they had new things to learn from mathematicians. As a consequence, the philosophical tradition lost what the mathematicians of Kolmogorov's generation had learned about how to make sense of frequentism. They lost Borel's, Lévy's and Kolmogorov's understanding of Cournot's principle, and they also missed Ville's insight into the concept of a Kollektiv.

The one post-war philosopher who might have been expected to make Cournot's principle central to his philosophy of probability was Karl Popper. Popper taught that in general scientific theories make contact with reality by providing opportunities for falsification. Cournot's principle tells us how to find such opportunities with a probabilistic theory: single out an event of very small probability and see if it happens.

Popper is in fact one of the few English-speaking post-war philosophers who discussed Cournot's principle, and on the whole he seems to have subscribed to it. This already seems clear, for example, in §68 of his celebrated *Logik der Forschung* (1935). The picture is muddied, however, by Popper's own youthful and enduring ambition to make his mark by axiomatizing probability. He had tried his hand at this in 1938, before he was even aware of Kolmogorov's work (Popper 1938). By 1959, when he published the expanded English edition of *Logik der Forschung*, under the title *The Logic of Scientific Discovery*, he knew the importance that Kolmogorov's axioms had assumed, and he included an extensive discussion of the axioms themselves, finding reasons to disagree with them and propose alternatives. This makes it difficult to keep track of the issue of interpretation (see, however, the footnote on p. 191). When he returns to the issue in *Realism and the Aim of Science* in 1983, he speaks explicitly about Cournot's principle, but again his views are very complex (see especially p. 379).

6.2.4 Kolmogorov's reticence

The scant attention paid to Kolmogorov's own philosophy is due in part to his own reticence. Given his great prestige, people would have paid attention to if he had written about it at length, but he did not do so. In the Soviet Union, it would have been very dangerous to do so.

Kolmogorov's discussion of theoretical statistics at the Tashkent conference in 1948 (Kolmogorov 1948b) shows that he was interested in the controversies about Bayesian, fiducial, and confidence methods and that he wanted to develop a predictive approach. He was careful, however, to treat statistical theory as a branch of mathematics rather than a philosophical enterprise using mathematics as a tool, and this did not permit him to engage in dialogue with the schools of thought that were debating each other in the West.

Stalin's death led to only slightly greater breathing space for mathematical statistics. The crucial step was a conference held jointly in 1954 by the Academy of Sciences, the Central Statistical Bureau, and the Ministry of Higher

Education. At this meeting, Kolmogorov was still careful to criticize the notion that there is a universal general theory of statistics; all that is needed, he said, is mathematical statistics, together with some techniques for collecting and processing data. But there are some problems, he said, in domains such as physics, actuarial work, and sampling, in which probability can legitimately be applied to statistics (Kolmogorov 1954; Kotz 1965; Lorentz 2002; Sheynin 1996, p. 102). This formulation permitted him to publish a still very concise exposition of his frequentist viewpoint, which appeared in English, as a chapter in *Mathematics, Its Content, Methods, and Value* (1956).

In the 1960s, when Kolmogorov did return to the foundations of probability (Kolmogorov 1963, 1965, 1968), he still emphasized topics far from politics and philosophy: he was applying information theory to linguistics and the analysis of literature. The definition of randomness he produced used new ideas concerning complexity but was inspired by the idea of a finitary Kollektiv, which he had previously thought impossible to treat mathematically (Li and Vitányi 1997, Vovk and Shafer 2003). It could be said that the new definition brought Cournot's principle back in a new and perhaps more convincing form. Instead of ruling out events of small probability just because they are singled out in advance, we rule out sets of outcomes that have simple enough a description that we can in fact single them out in advance. But Kolmogorov said nothing this philosophical. Vladimir Arnol'd tells us that like most people who had lived through Stalin's terror, Kolmogorov was afraid of the authorities until his last day (Arnol'd 1993, p. 92).

Kolmogorov's reticence complemented Doob's explicit rejection of philosophy, helping to paint for many in the West a picture of mathematical statistics free from philosophy's perils and confusions. In fact he did have a philosophy of probability—the one he presented in the *Grundbegriffe*.

7   Conclusion

The great achievement of the *Grundbegriffe* was to seize the notion of probability from the classical probabilists. In doing so, Kolmogorov made space for a mathematical theory that has flourished for seven decades, driving ever more sophisticated statistical methodology and feeding back into other areas of pure mathematics. He also made space, within the Soviet Union, for a theory of mathematical statistics that could prosper without entanglement in the vagaries of political or philosophical controversy.

Kolmogorov's way of relating his axioms to the world has received much less attention than the axioms themselves. But Cournot's principles has re-emerged in new guises, as the information-theoretic and game-theoretic aspects of probability have come into their own. We will be able to see this evolution as growth—rather than mere decay or foolishness—only if we can see the *Grundbegriffe* as a product of its own time.

A   Appendix

A.1   Ulam's contribution to the Zürich Congress

Stanisław Ulam's contribution to the International Congress of Mathematicians in Zürich (Ulam 1932), was entitled "Zum Massbegriffe in Produkträumen". We here provide a translation into English. We have updated the notation slightly, using $\{x\}$ instead of (x) for the set containing a single element x, $\cup$ instead of $\sum$ for set union, $\emptyset$ instead of 0 for the empty set, and $x_1, x_2, \ldots$ instead of $\{x_i\}$ for a sequence.

The article by Łomnicki and Ulam appeared in 1934, in Volume 23 of *Fundamenta Mathematicae* rather than in Volume 20 as they had hoped.

On the idea of measure in product spaces
By St. Ulam, Lwów

The content of the report consists of the presentation of some results that were obtained in collaboration with Mr. Z. Lomnicki.

Let X, Y be two abstract spaces, on which a measure $m(M)$ is defined for certain sets M (the measurable sets). No topological or group-theoretic assumption is made about the nature of the spaces. Now a measure will be introduced on the space $X \times Y$ (i.e., the space of all ordered pairs (x, y), where $x \in X$ and $y \in Y$). This corresponds to problems in probability, which are concerned with finding probabilities of combined events from probabilities of individual (independent) events.

We make the following assumptions about the measure on X (and analogously on Y) and the class $\mathfrak{M}$ of measurable sets there:

   I. The whole space X is measurable: $X \in \mathfrak{M}$. Also the singleton set $\{x\}$ is in $\mathfrak{M}$ for each $\{x\} \in X$.

  II. From $M_i \in \mathfrak{M}$ for $i = 1, 2, \ldots$, it follows that $\cup_{i=1}^{\infty} M_i \in \mathfrak{M}$.

 III. From $M, N \in \mathfrak{M}$, it follows that $M \setminus N \in \mathfrak{M}$.

 IV. If $M \in \mathfrak{M}$, $m(M) = 0$, and $N \subset M$, then N is also in $\mathfrak{M}$.

   1. $m(X) = 1$; $m(M) \geq 0$.

   2. $m(\cup_{i=1}^{\infty} M_i) = \sum_{i=1}^{\infty} m(M_i)$ when $M_i \cap M_j = \emptyset$ for $i \neq j$.

   3. From $m(M) = 0$ and $N \subset M$, it follows that $m(N) = 0$.[21]

With these assumptions, we obtain

Theorem. One can introduce a measure on the space $X \times Y$ such that all sets of the form $M \times N$ are measurable and have the measure $m(M) \cdot m(N)$ (here M and N denote measurable subsets of X and Y, respectively) and all our assumptions remain satisfied.

---

[21]This condition follows from the previous one and was omitted from the 1934 article.

An analogous theorem holds for a countable product—i.e., a set

$$\prod X_i = X_1 \times X_2 \times \cdots$$

consisting of all sequences $x_1, x_2, \ldots$, where $x_i \in X_i$.

Some of the properties of these measures will be studied in connection with the formulation of questions in probability theory. A detailed presentation will appear shortly in *Fundamenta Mathematicae* (presumably Vol. 20).

A.2 A letter from Kolmogorov to Fréchet

The Fréchet papers in the archives of the Academy of Sciences in Paris include a letter in French to Fréchet, in which Kolmogorov elaborates briefly on his philosophy of probability. Here is a translation.

<div align="center">

Moscow 6, Staropimenovsky per. 8, flat 5
3 August 1939

</div>

Dear Mr. Fréchet,

I thank you sincerely for sending the proceedings of the Geneva Colloquium, which arrived during my absence from Moscow in July.

The conclusions you express on pp. 51–54 are in full agreement with what I said in the introduction to my book:

> In the pertinent mathematical circles it has been common for some time to construct probability theory in accordance with this general point of view. But a complete presentation of the whole system, free from superfluous complications, has been missing. . .

You are also right to attribute to me (on p. 42) the opinion that the formal axiomatization should be accompanied by an analysis of its real meaning. Such an analysis is given, perhaps too briefly, in the section "The relation to the world of experience" in my book. Here I insist on the view, expressed by Mr. von Mises himself (Wahrscheinlickeitsrechnung 1931, pp. 21–26), that "collectives" are <u>finite</u> (though very large) in real practice.

One can therefore imagine three theories:

A A theory based on the notions of "very large" finite "collectives", "approximate" stability of frequencies, etc. This theory uses ideas that cannot be defined in a purely formal (i.e., mathematical) manner, but it is the only one to reflect experience truthfully.

B A theory based on infinite collectives and limits of frequencies. After Mr. Wald's work we know that this theory can be developed in a purely formal way without contradictions. But in this case its relation to experience cannot have any different nature than for any other axiomatic theory. So in agreement with Mr. von Mises, we should regard theory B as a certain "mathematical idealization" of theory A.

C An axiomatic theory of the sort proposed in my book. Its practical value can be deduced directly from the "approximate" theory A without appealing to theory B. This is the procedure that seems simplest to me.

Yours cordially,
A. Kolmogoroff

A.3   A closer look at Lévy's example

Corresponding to the partition $\mathfrak{A}$ considered by Lévy is the σ-algebra $\mathfrak{G}$ consisting of the Borel sets E in $[0, 1]$ such that

$$\left.\begin{array}{r} x \in E \\ x' - x \text{ is rational} \end{array}\right\} \implies x' \in E. \tag{3}$$

Each of Lévy's $C(x)$ is in $\mathfrak{G}$, and conditional expectation with respect to the partition $\mathfrak{A}$ in Kolmogorov's sense is the same as conditional expectation with respect to $\mathfrak{G}$ in Doob's sense.

Lemma 1. *For any rational number $a = k/n$ (k and n are integers, $k \geq 0$, and $n \geq 1$),*

$$\mathsf{P}([0, a] \mid \mathfrak{G}) = a \quad a.s.,$$

where P is uniform on $[0, 1]$.

Proof. We must show that

$$\int_E a \, dx = \mathsf{P}([0, a] \cap E) \tag{4}$$

for all $E \in \mathfrak{G}$. Consider the sets

$$E_1 := E \cap \left[0, \frac{1}{n}\right],$$

$$E_2 := E \cap \left[\frac{1}{n}, \frac{2}{n}\right],$$

$$\vdots$$

$$E_n := E \cap \left[\frac{n-1}{n}, 1\right].$$

Since E is invariant with respect to adding $\frac{1}{n}$ (modulo 1), these sets are congruent. So the left-hand side of (4) is $an\mathsf{P}(E_1)$ while the right-hand side is $k\mathsf{P}(E_1)$; since $k = an$, these are equal. □

Lemma 2. *For any bounded Borel function $f : [0, 1] \to R$,*

$$\mathsf{E}(f \mid \mathfrak{G}) = \mathsf{E}(f) \quad a.s. \tag{5}$$

Proof. Equation (5) means that

$$\int_E f \, dx = \mathsf{E}(f)\mathsf{P}(E) \tag{6}$$

for any $E \in \mathfrak{G}$. By Luzin's theorem (Kolmogorov and Fomin 1999), for any $\epsilon > 0$ there exists a continuous function $g : [0,1] \to R$ such that

$$\mathsf{P}\{x \mid f(x) \neq g(x)\} < \epsilon;$$

so it suffices to prove (6) for continuous functions. We can easily do this using Lemma 1. □

As a special case of Lemma 2, we have

$$\mathsf{P}(B \mid \mathfrak{G}) = \mathsf{P}(B) \quad \text{a.s.}$$

for any Borel set $B \subseteq [0,1]$. In other words, we can take X's conditional distribution in Lévy's example to be uniform, just like its unconditional distribution, no matter which $C(x)$ we condition on. This is exceedingly unnatural, because the uniform distribution gives the set on which we are conditioning probability zero.

Theorem II.(89.1) of Rogers and Williams (1995, p. 219) tells us that if a σ-algebra $\mathfrak{G}$ is countably generated, then conditional probabilities with respect to $\mathfrak{G}$ will almost surely give the event on which we are conditioning probability one and will be equally well behaved in many other respects. But the σ-algebra $\mathfrak{G}$ characterized by (3) is not countably generated.

Acknowledgments

Rutgers's Dana Library, and Mitchell Brown, at Princeton's Fine Library, have both been exceedingly helpful in locating other references.

We take full responsibility for our translations into English of various passages from French, German, Italian, and Russian, but in some cases we were able to consult previous translations.

Although this article is based almost entirely on published sources, extensive archival material is available to those interested in further investigation. The Accademia dei Lincei at Rome has a Castelnuovo archive (Gario 2001). There is an extensive Fréchet archive at the Académie des Sciences in Paris. Lévy's papers were lost when his Paris apartment was ransacked by the Nazis, but his extant correspondence includes letters exchanged with Fréchet (Barbut, Locker, and Mazliak 2004), Kai Lai Chung (in Chung's possession), and Michel Loève (in Bernard Bru's possession). Additional correspondence of Lévy's is in the library of the University of Paris at Jussieu and in the possession of his family. The material at Kolmogorov and Aleksandrov's country home at Komarovka is being cataloged under the direction of Albert N. Shiryaev. Doob's papers, put into order by Doob himself, are accessible to the public at the University of Illinois, Champagne-Urbana.

Volumes 1–195 of the *Comptes rendus hebdomadaires des séances de l'Académie des Sciences* (1885–1932) are available free on-line at the Bibliothèque Nationale de France (http://gallica.bnf.fr/).

References

George B. Airy. *On the algebraical and numerical theory of errors of observations and the combination of observations*. Macmillan, London, 1861. Second edition 1875, third 1879.

Aleksandr D. Aleksandrov, Andrei N. Kolmogorov, and Mikhail A. Lavrent'ev, editors. *Математика, ее содержание, методы и значение*. Nauka, Moscow, 1956. The Russian edition had three volumes. The English translation, *Mathematics, Its Content, Methods, and Meaning*, was first published in 1962 and 1963 in six volumes by the American Mathematical Society, Providence, RI, and then republished in 1965 in three volumes by the MIT Press, Cambridge, MA. Reprinted by Dover, New York, 1999. 86

Erik Sparre Andersen and Børge Jessen. On the introduction of measures in infinite product sets. *Det Kongelige Danske Videnskabernes Selskab, Matematisk-Fysiske Meddelelser*, XXV(4):1–8, 1948.

Oskar Nikolaevich Anderson. *Einführung in die mathematische Statistik*. Springer, Vienna, 1935.

Oskar Nikolaevich Anderson. Die Begründung des Gesetzes der grossen Zahlen und die Umkehrung des Theorems von Bernoulli. *Dialectica*, 3(9/10):65–77, 1949.

Oskar Nikolaevich Anderson. *Probleme der statistischen Methodenlehre in den Sozialwissenschaften.* Physica, Würzburg, 1954.

Anonymous. Account in Russian of a conference on statistics in Moscow in 1954. *Вестник статистики (Bulletin of Statistics)*, 5:39–95, 1954. 85

Vladimir Igorevich Arnol'd. Об А. Н. Колмогорове (On A. N. Kolmogorov). In Shiryaev (1993), pages 144–172. Two different English translations have appeared, one on pp. 129–153 of Zdravkovska and Duren (1993), and one on pp. 89–108 of Flower (2000).

Louis Bachelier. Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure, 3e série*, 17:21–86, 1900. This was Bachelier's doctoral dissertation. Reprinted in facsimile in 1995 by Éditions Jacques Gabay, Paris. An English translation, by A. James Boness, appears on pp. 17–78 of *The Random Character of Stock Market Prices*, edited by Paul H. Cootner, MIT Press, Cambridge, MA, 1964.

Louis Bachelier. Les probabilités à plusieurs variables. *Annales scientifiques de l'École Normale Supérieure, 3e série*, 27:339–360, 1910.

Louis Bachelier. *Calcul des probabilités.* Gauthier-Villars, Paris, 1912.

Maya Bar-Hillel and Ruma Falk. Some teasers concerning conditional probabilities. *Cognition*, 11:109–122, 1982.

Ed Barbeau. The problem of the car and goats. *The College Mathematics Journal*, 24:149–154, 1993.

Marc Barbut, Bernard Locker, and Laurent Mazliak. *Paul Lévy–Maurice Fréchet, 50 ans de correspondance en 107 lettres.* Hermann, Paris, 2004.

Jack Barone and Albert Novikoff. A history of the axiomatic formulation of probability from Borel to Kolmogorov: Part I. *Archive for History of Exact Sciences*, 18:123–190, 1978.

Maurice S. Bartlett. Probability and logic, mathematics and science. *Dialectica*, 3(9/10):104–113, 1949.

Raymond Bayer, editor. *Congrès International de Philosophie des Sciences, Paris, 1949; IV: Calcul des Probabilités.* Number 1146 in *Actualités Scientifiques et Industrielles.* Hermann, Paris, 1951. 79, 80, 90, 96

Heinrich Behnke, Günter Bertram, and Robert Sauer, editors. *Grundzüge der Mathematik. Band IV. Praktische Methoden und Anwendungen der Mathematik (Geometrie und Statistik).* Vanderhoeck & Ruprecht, Göttingen, 1966. 81

Margherita Benzi. Un "probabilista neoclassico": Francesco Paolo Cantelli. *Historia Mathematica*, 15:53–72, 1988.

Margherita Benzi. Dubbiezze e controversie: Il dibattito su logica e probabilità in Italia nei primi anni del Novecento. *Historia Mathematica*, 22:43–63, 1995.

Jacob Bernoulli. *Ars Conjectandi*. Thurnisius, Basel, 1713. This pathbreaking work appeared eight years after Bernoulli's death. A facsimile reprinting of the original Latin text is sold by Éditions Jacques Gabay. A German translation appeared in 1899 (*Wahrscheinlichkeitsrechnung von Jakob Bernoulli. Anmerkungen von R. Haussner*, Ostwald's Klassiker, Nr. 107–108, Engelmann, Leipzig), with a second edition (Deutsch, Frankfurt) in 1999. A Russian translation of Part IV, which contains Bernoulli's law of large numbers, appeared in 1986: Я. Бернулли, О законе больших чисел, Nauka, Moscow. It includes a preface by Kolmogorov, dated October 1985, and commentaries by other Russian authors. Bing Sung's translation of Part IV into English, dated 1966, remains unpublished but is available in several university libraries in the United States. Oscar Sheynin's English translation of Part IV, dated 2005, can be downloaded from www.sheynin.de.

Felix Bernstein. Über eine Anwendung der Mengenlehre auf ein aus der Theorie der säkularen Störungen herrührendes Problem. *Mathematische Annalen*, 71: 417–439, 1912.

Sergei N. Bernstein. Опыт аксиоматического обоснования теории вероятностей (On the axiomatic foundation of the theory of probability). *Сообщения Харьковского математического общества (Communications of the Kharkiv mathematical society)*, 15:209–274, 1917. Reprinted on pp. 10–60 of Bernstein's Собрание сочинений, Volume IV, Nauka, Moscow, 1964.

Sergei N. Bernstein. *Теория вероятностей (Theory of Probability)*. Государственное Издательство (State Publishing House), Moscow and Leningrad, 1927. Second edition 1934, fourth 1946.

Joseph Bertrand. *Calcul des probabilités*. Gauthier-Villars, Paris, 1889. Some copies of the first edition are dated 1888. Second edition 1907. Reprinted by Chelsea, New York, 1972.

Nicholas H. Bingham. Measure into probability: From Lebesgue to Kolmogorov. *Biometrika*, 87:145–156, 2000.

David Blackwell. On a class of probability spaces. In Jerzy Neyman, editor, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 1–6. University of California Press, Berkeley and Los Angeles, 1956.

Alain Blum and Martine Mespoulet. *L'Anarchie bureaucratique. Statistique et pouvoir sous Staline*. La Découverte, Paris, 2003.

Georg Bohlmann. Lebensversicherungs-Mathematik. In *Encyklopädie der mathematischen Wissenschaften, Bd. I, Teil 2*, pages 852–917. Teubner, Leipzig, 1901.

George Boole. *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities.* Macmillan, London, 1854. Reprinted by Dover, New York, 1958.

Émile Borel. Sur quelques points de la théorie des fonctions. *Annales scientifiques de l'École Normale Supérieure, 3ᵉ série*, 12:9–55, 1895.

Émile Borel. Sur les séries de Taylor. *Acta Mathematica*, 20:243–247, 1897. Reprinted in Borel (1972), Volume 2, pp. 661–665.

Émile Borel. *Leçons sur la théorie des fonctions.* Gauthier-Villars, Paris, 1898.

Émile Borel. Remarques sur certaines questions de probabilité. *Bulletin de la Société mathématique de France*, 33:123–128, 1905. Reprinted in Borel (1972), Volume 2, pp. 985–990.

Émile Borel. La valeur pratique du calcul des probabilités. *Revue du mois*, 1: 424–437, 1906. Reprinted in Borel (1972), Volume 2, pp. 991–1004.

Émile Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo*, 27:247–270, 1909a. Reprinted in Borel (1972), Volume 2, pp. 1055–1079.

Émile Borel. *Éléments de la théorie des probabilités.* Gauthier-Villars, Paris, 1909b. Third edition 1924. The 1950 edition was translated into English by John E. Freund and published as *Elements of the Theory of Probability* by Prentice-Hall in 1965.

Émile Borel. *Notice sur les travaux scientifiques.* Gauthier-Villars, Paris, 1912. Prepared by Borel to support his candidacy to the Académie des Sciences. Reprinted in Borel (1972), Volume 1, pp. 119–190.

Émile Borel. *Le Hasard.* Alcan, Paris, 1914. The first and second editions both appeared in 1914, with later editions in 1920, 1928, 1932, 1938, and 1948.

Émile Borel. Sur les probabilités universellement négligeables. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 190:537–540, 1930. Reprinted as Note IV of Borel (1939).

Émile Borel. *Valeur pratique et philosophie des probabilités.* Gauthier-Villars, Paris, 1939. Reprinted in 1991 by Éditions Jacques Gabay. 72

Émile Borel. *Le jeu, la chance et les théories scientifiques modernes.* Gallimard, Paris, 1941.

Émile Borel. *Les probabilités et la vie.* Presses Universitaires de France, Paris, 1943. Second edition 1946, third 1950, fourth 1958, sixth 1967. The fourth edition was translated into English by Maurice Baudin and published as *Probabilities and Life* by Dover, New York, in 1962.

Émile Borel. Probabilité et certitude. *Dialectica*, 3(9/10):24–27, 1949.

Émile Borel. *Probabilité et certitude*. Presses Universitaires de France, Paris, 1950. An English translation, *Probability and Certainty*, was published in 1963 by Walker, New York.

Émile Borel. *Œuvres de Émile Borel*. Centre National de la Recherche Scientifique, Paris, 1972. Four volumes. 72

Nicolas Bourbaki (pseudonym). *Elements of the History of Mathematics*. Springer, Berlin, 1994. Translated from the 1984 French edition by John Meldrum.

Marcel Brissaud, editor. *Ecrits sur les processus aléatoires: Mélanges en hommage à Robert Fortet*. Lavoisier, Paris, 2002. 73

Ugo Broggi. *Die Axiome der Wahrscheinlichkeitsrechnung*. PhD thesis, Universität Göttingen, 1907. Excerpts reprinted on pp. 359–366 of Schneider (1988).

Bernard Bru. Borel, Lévy, Neyman, Pearson et les autres. *MATAPLI*, 60: 51–60, octobre 1999.

Bernard Bru. Émile Borel. In Heyde and Seneta (2001), pages 287–291.

Bernard Bru. Présentation de l'œuvre de Robert Fortet. In Brissaud (2002), pages 19–48.

Bernard Bru. Souvenirs de Bologne. *Journal de la Société Française de Statistique*, 144(1–2):134–226, 2003a. Special volume on history.

Bernard Bru. Personal communication, 2003b.

Bernard Bru and François Jongmans. Joseph Bertrand. In Heyde and Seneta (2001), pages 185–189.

Heinrich Bruns. *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*. Teubner, Leipzig and Berlin, 1906. Free on-line at http://historical.library.cornell.edu.

Stephen G. Brush. A history of random sequences. I: Brownian movement from Brown to Perrin. *Archive for History of Exact Sciences*, 5:1–36, 1968.

George-Louis Buffon. Essai d'arithmétique morale. In *Supplément à l'Histoire naturelle*, volume 4, pages 46–148. Imprimerie Royale, Paris, 1777.

Francesco Paolo Cantelli. Sui fondamenti del calcolo delle probabilità. *Il Pitagora. Giornale di matematica per gli alunni delle scuole secondarie*, 12:21–25, 1905. In issue 1–2, for October and November 1905.

Francesco Paolo Cantelli. Calcolo delle probabilità. *Il Pitagora. Giornale di matematica per gli alunni delle scuole secondarie*, 12:33–39 and 68–74, 1905–1906. In issue 3–4, for December 1905 and Janaury 1906, continued in issue 5–6–7, for February, March, and April 1906.

Francesco Paolo Cantelli. La tendenza ad un limite nel senso del calcolo delle probabilità. *Rendiconti del Circolo Matematico di Palermo*, 41:191–201, 1916a. Reprinted in Cantelli (1958), pp. 175–188.

Francesco Paolo Cantelli. Sulla legge dei grandi numeri. *Memorie, Accademia dei Lincei*, V, 11:329–349, 1916b. Reprinted in Cantelli (1958), pp. 189–213.

Francesco Paolo Cantelli. Sulla probabilità come limite della frequenza. *Atti Reale Accademia Nazionale dei Lincei. Rendiconti*, 26:39–45, 1917. Reprinted in Cantelli (1958), pp. 214–221.

Francesco Paolo Cantelli. Una teoria astratta del calcolo delle probabilità. *Giornale dell'Istituto Italiano degli Attuari*, 8:257–265, 1932. Reprinted in Cantelli (1958), pp. 289–297.

Francesco Paolo Cantelli. Considérations sur la convergence dans le calcul des probabilités. *Annales de l'Institut Henri Poincaré*, 5:3–50, 1935. Reprinted in Cantelli (1958), pp. 322–372.

Francesco Paolo Cantelli. *Alcune Memorie Matematiche*. Giuffrè, Milan, 1958. 74

Constantin Carathéodory. Über das lineare Mass von Punktmengen—eine Verallgemeinerung des Längenbegriffs. *Nachrichten der Akademie der Wissenschaften zu Göttingen. II. Mathematisch-Physikalische Klasse*, 4:404–426, 1914.

Constantin Carathéodory. *Vorlesungen über reelle Funktionen*. Teubner, Leipzig and Berlin, 1918. Second edition 1927.

Guido Castelnuovo. *Calcolo delle probabilitá*. Albrighi e Segati, Milan, Rome, and Naples, 1919. Second edition in two volumes, 1926 and 1928. Third edition 1948.

Aleksandr Aleksandrovich Chuprov. Очерки по теории статистики (*Essays on the theory of statistics*). Sabashnikov, Saint Petersburg, second edition, 1910. The first edition appeared in 1909. The second edition was reprinted by the State Publishing House, Moscow, in 1959.

Aleksandr Aleksandrovich Chuprov. Das Gesetz der großen Zahlen und der stochastisch-statistische Standpunkt in der modernen Wissenschaft. *Nordisk statistik tidskrift*, 1(1):39–67, 1922.

Alonzo Church. On the concept of a random sequence. *Bulletin of the American Mathematical Society*, 46:130–135, 1940.

Donato Michele Cifarelli and Eugenio Regazzini. De Finetti's contribution to probability and statistics. *Statistical Science*, 11:253–282, 1996.

Julian Lowell Coolidge. *An Introduction to Mathematical Probability*. Oxford University Press, London, 1925.

Arthur H. Copeland Sr. Admissible numbers in the theory of probability. *American Journal of Mathematics*, 50:535–552, 1928.

Arthur H. Copeland Sr. The theory of probability from the point of view of admissible numbers. *Annals of Mathematical Statistics*, 3:143–156, 1932.

Antoine-Augustin Cournot. *Exposition de la théorie des chances et des probabilités*. Hachette, Paris, 1843. Reprinted in 1984 as Volume I (B. Bru, editor) of Cournot (1973–1984).

Antoine-Augustin Cournot. *Œuvres complètes*. Vrin, Paris, 1973–1984. 10 volumes, with an eleventh to appear. 75

Jean-Michel Coutault and Youri Kabanov, editors. *Louis Bachelier: Aux origines de la finance mathématique*. Presses Universitaires Franc-Comtoises, distributed by C*i*D, Paris, 2002.

Thomas M. Cover, Peter Gács, and Robert M. Gray. Kolmogorov's contributions to information theory and algorithmic complexity. *Annals of Probability*, 17:840–865, 1989.

Richard T. Cox. Probability, frequency, and reasonable expectation. *American Journal of Physics*, 14:1–13, 1946.

Harald Cramér. *Random Variables and Probability Distributions*. Cambridge University Press, Cambridge, 1937.

Harald Cramér. *Mathematical Methods in Statistics*. Princeton University Press, Princeton, NJ, 1946.

Harald Cramér. Half a century with probability theory: Some personal recollections. *Annals of Probability*, 4:509–546, 1976.

Pierre Crépel. Quelques matériaux pour l'histoire de la théorie des martingales (1920–1940). Technical report, Seminaires de Probabilités, Université de Rennes, 1984.

Emanuel Czuber. *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung*. Teubner, Leipzig, 1903. Second edition 1910, third 1914.

Jean d'Alembert. Réflexions sur le calcul des probabilités. In *Opuscule mathématiques*, volume 2, pages 1–25. 1761.

Jean d'Alembert. Doutes et questions sur le calcul des probabilités. In *Mélanges de littérature, d'histoire, et de philosophie*, volume 5, pages 275–304. 1767.

Percy John Daniell. A general form of integral. *Annals of Mathematics*, 19: 279–294, June 1918.

Percy John Daniell. Integrals in an infinite number of dimensions. *Annals of Mathematics*, 20:281–288, July 1919a. July.

Percy John Daniell. Functions of limited variation in an infinite number of dimensions. *Annals of Mathematics*, 21:30–38, September 1919b. September.

Percy John Daniell. Further properties of the general integral. *Annals of Mathematics*, 21:203–220, March 1920.

Percy John Daniell. Integral products and probability. *American Journal of Mathematics*, 43:143–162, July 1921.

Lorraine Daston. D'Alembert's critique of probability theory. *Historia Mathematica*, 6:259–279, 1979.

Lorraine Daston. How probabilities came to be objective and subjective. *Historia Mathematica*, 21:330–344, 1994.

A. P. Dawid. Probability, causality and the empirical world: A Bayes–de Finetti–Popper–Borel synthesis. *Statistical Science*, 19:44–57, 2004.

Bruno de Finetti. A proposito dell'estensione del teorema delle probabilità totali alle classi numerabili. *Rendiconti del Reale Istituto Lombardo di Scienze e Lettere*, 63:901–905, 1063–1069, 1930.

Bruno de Finetti. *Compte rendu critique du colloque de Genève sur la théorie des probabilités*. Number 766 in *Actualités Scientifiques et Industrielles*. Hermann, Paris, 1939. This is the eighth fascicle of Wavre (1938–1939).

Bruno de Finetti. Recent suggestions for the reconciliation of theories of probability. In Jerzy Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 217–225. University of California Press, Berkeley and Los Angeles, 1951.

Bruno de Finetti. Notes de M. B. de Finetti sur le "Rapport général". In *Les mathématiques et le concret* Fréchet (1955), pages 232–241.

Bruno de Finetti. *Teoria Delle Probabilità*. Einaudi, Turin, 1970. An English translation, by Antonio Machi and Adrian Smith, was published as *Theory of Probability* by Wiley (London) in two volumes in 1974 and 1975.

Abraham De Moivre. *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play*. Pearson, London, 1718. Second edition 1738, third 1756.

Augustus De Morgan. *An Essay on Probabilities, and on their application to Life Contingencies and Insurance Offices*. Longman, Orme, Brown, Green & Longmans, London, 1838. Reprinted by Arne Press, New York, 1981.

Augustus De Morgan. *Formal Logic: or, The Calculus of Inference, Necessary and Probable*. Taylor and Walton, London, 1847.

Claude Dellacherie and Paul-André Meyer. *Probabilités et Potentiel*. Hermann, Paris, 1975.

Sergei S. Demidov and Boris V. Levshin, editors. *Дело академика Николая Николаевича Лузина (The Affair of Academician Nikolai Nikolaevich Luzin)*. РХГИ (Russian Christian Humanitarian Institute), Saint Petersburg, 1999.

Jean Dieudonné. Sur le théorème de Lebesgue-Nikodym. III. *Annales de l'Université de Grenoble*, 23:25–53, 1948.

Gustav Doetsch. Review of Kolmogorov (1933). *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 45:153, 1933.

Joseph L. Doob. Probability and statistics. *Transactions of the American Mathematical Society*, 36:759–775, 1934a.

Joseph L. Doob. Stochastic processes and statistics. *Proceedings of the National Academy of Sciences of the United States*, 20:376–379, 1934b.

Joseph L. Doob. Stochastic processes depending on a continuous parameter. *Transactions of the American Mathematical Society*, 42:107–140, 1937.

Joseph L. Doob. Stochastic processes with an integral-valued parameter. *Transactions of the American Mathematical Society*, 44:87–150, 1938.

Joseph L. Doob. Regularity properties of certain families of chance variables. *Transactions of the American Mathematical Society*, 47:455–486, 1940a.

Joseph L. Doob. The law of large numbers for continuous stochastic processes. *Duke Mathematical Journal*, 6:290–306, 1940b.

Joseph L. Doob. Probability as measure. *Annals of Mathematical Statistics*, 12:206–214, 1941. This article originated as a paper for a meeting of the Institute of Mathematical Statistics in Hanover, New Hampshire, in September 1940. It was published together with an article by von Mises and comments by Doob and von Mises on each other's articles.

Joseph L. Doob. Topics in the theory of Markoff chains. *Transactions of the American Mathematical Society*, 52:37–64, 1942.

Joseph L. Doob. Markoff chains—denumerable case. *Transactions of the American Mathematical Society*, 58:455–473, 1945.

Joseph L. Doob. Probability in function space. *Bulletin of the American Mathematical Society*, 53:15–30, 1947.

Joseph L. Doob. Asymptotic properties of Markoff transition probabilities. *Transactions of the American Mathematical Society*, 63:393–421, 1948.

Joseph L. Doob. *Stochastic Processes*. Wiley, New York, 1953.

Joseph L. Doob. Wiener's work in probability theory. *Bulletin of the American Mathematical Society*, 72(1, Part II):69–72, 1966. This is a special obituary issue for Wiener, and its pages are numbered separately from the rest of Volume 72.

Joseph L. Doob. William Feller and twentieth century probability. In Lucien M. Le Cam, Jerzy Neyman, and Elizabeth L. Scott, editors, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, pages xv–xx. University of California Press, Berkeley and Los Angeles, 1972.

Joseph L. Doob. Kolmogorov's early work on convergence theory and foundations. *Annals of Probability*, 17:815–821, 1989.

Joseph L. Doob. The development of rigor in mathematical probability, 1900–1950. In Pier (1994a), pages 157–170. Reprinted in *American Mathematical Monthly* 103(7):586–595, 1996.

Karl Dörge. Zu der von R. von Mises gegebenen Begründung der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 32:232–258, 1930.

Karl Dörge. Review of Kolmogorov (1933). *Jahrbuch über die Fortschritte der Mathematik*, 59:1152, 1933.

Louis-Gustave Du Pasquier. *Le calcul des probabilités, son évolution mathématique et philosophique*. Hermann, Paris, 1926.

Evgeny B. Dynkin. Kolmogorov and the theory of Markov processes. *Annals of Probability*, 17:822–832, 1989.

Francis Y. Edgeworth. *Metretike, or the Method of Measuring Probabilities and Utility*. Temple, London, 1887.

Francis Y. Edgeworth. *Writings in Probability, Statistics, and Economics*. Edward Elgar, Cheltenham, United Kingdom, 1996. Three volumes. Edited by Charles R. McCann, Jr.

Albert Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 17:549–560, 1905. English translation in Einstein (1956).

Albert Einstein. Zur Theorie der Brownschen Bewegung. *Annalen der Physik*, 19:371–381, 1906. English translation in Einstein (1956).

Albert Einstein. *Investigations on the Theory of the Brownian Movement*. Dover, New York, 1956. First published by Dutton, New York, in 1926, this is a translation, by A. D. Cowper, of *Untersuchungen über die Theorie der Brownschen Bewegung*. It includes most of Einstein's papers on Brownian motion.
78

Robert Leslie Ellis. On the foundations of the theory of probabilities. *Transactions of the Cambridge Philosophical Society*, 8(1):1–6, 1849. The paper was read on February 14, 1842. Part 1 of Volume 8 was published in 1843 or 1844, but Volume 8 was not completed until 1849.

Georg Faber. Über stetige Funktionen. *Mathematische Annalen*, 69:372–443, 1910.

Gustav Theodor Fechner. *Kollektivmasslehre.* Engelmann, Leipzig, 1897. Edited by G. F. Lipps.

William Feller. Review of Kolmogorov (1933). *Zentralblatt für Mathematik und ihre Grenzegebiete*, 7:216, 1934.

William Feller. Sur les axiomatiques du calcul des probabilités et leurs relations avec les expériences. In Wavre (1938–1939), pages 7–21 of the second fascicle, number 735 *Les fondements du calcul des probabilités*. This celebrated colloquium, chaired by Maurice Fréchet, was held in October 1937 at the University of Geneva. Participants included Cramér, Dœblin, Feller, de Finetti, Heisenberg, Hopf, Lévy, Neyman, Pòlya, Steinhaus, and Wald, and communications were received from Bernstein, Cantelli, Glivenko, Jordan, Kolmogorov, von Mises, and Slutsky. The proceedings were published by Hermann in eight fascicles in their series *Actualités Scientifiques et Industrielles*. The first seven fascicles appeared in 1938 as numbers 734 through 740; the eighth, de Finetti's summary of the colloquium, appeared in 1939 as number 766. See de Finetti (1939).

Arne Fisher. *The Mathematical Theory of Probabilities and Its Application to Frequency Curves and Statistical Methods.* Macmillan, New York, 1915. Second edition 1922.

David Flower, editor. *Kolmogorov in Perspective.* American Mathematical Society and London Mathematical Society, Providence, RI, and London, 2000. Volume 20 of the History of Mathematics Series. Flower is listed as the chair of the editorial board. 70, 94

Robert M. Fortet. Opinions modernes sur les fondements du calcul de probabilités. In Le Lionnais (1948), pages 207–215. This volume was reprinted by Blanchard in 1962 and by Hermann in 1998. An English translation, *Great Currents of Mathematical Thought*, was published by Dover, New York, in 1971.

Robert M. Fortet. *Calcul des probabilités.* Centre National de la Recherche Scientifique, Paris, 1950.

Robert M. Fortet. Faut-il élargir les axiomes du calcul des probabilités? In Bayer (1951), pages 35–47.

Abraham A. Fraenkel. *Lebenskreise.* Deutsche Verlag-Anstalt, Stuttgart, 1967.

Maurice Fréchet. Définition de l'intégrale sur un ensemble abstrait. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 160:839–840, 1915a.

Maurice Fréchet. Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait. *Bulletin de la Société mathématique de France*, 43:248–265, 1915b.

Maurice Fréchet. Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements. *Rendiconti del Reale Istituto Lombardo di Scienze e Lettere*, 63:899–900, 1059–1062, 1930.

Maurice Fréchet. *Généralités sur les probabilités. Variables aléatoires.* Gauthier-Villars, Paris, 1937. This is Book 1 of Fréchet (1937–1938). The second edition (1950) has a slightly different title: *Généralités sur les probabilités. Éléments aléatoires.*

Maurice Fréchet. *Recherches théoriques modernes sur la théorie des probabilités.* Gauthier-Villars, Paris, 1937–1938. This work is listed in the bibliography of the *Grundbegriffe* as in preparation. It consists of two books, Fréchet (1937) and Fréchet (1938a). The two books together constitute Fascicle 3 of Volume 1 of Émile Borel's *Traité du calcul des probabilités et ses applications.*

Maurice Fréchet. *Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles.* Gauthier-Villars, Paris, 1938a. This is Book 2 of Fréchet (1937–1938). Second edition 1952.

Maurice Fréchet. Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités. In Wavre (1938–1939), pages 23–55 of the second fascicle, number 735, *Les fondements du calcul des probabilités.* This celebrated colloquium, chaired by Maurice Fréchet, was held in October 1937 at the University of Geneva. Participants included Cramér, Dœblin, Feller, de Finetti, Heisenberg, Hopf, Lévy, Neyman, Pòlya, Steinhaus, and Wald, and communications were received from Bernstein, Cantelli, Glivenko, Jordan, Kolmogorov, von Mises, and Slutsky. The proceedings were published by Hermann in eight fascicles in their series *Actualités Scientifiques et Industrielles.* The first seven fascicles appeared in 1938 as numbers 734 through 740; the eighth, de Finetti's summary of the colloquium, appeared in 1939 as number 766. See de Finetti (1939).

Maurice Fréchet. Rapport général sur les travaux du Colloque de Calcul des Probabilités. In Bayer (1951), pages 3–21.

Maurice Fréchet. *Les mathématiques et le concret.* Presses Universitaires de France, Paris, 1955. 76

Maurice Fréchet and Maurice Halbwachs. *Le calcul des probabilités à la portée de tous.* Dunod, Paris, 1924.

Hans Freudenthal and Hans-Georg Steiner. Aus der Geschichte der Wahrscheinlichkeitstheorie und der mathematischen Statistik. In Behnke et al. (1966), pages 149–195.

Thornton C. Fry. *Probability and Its Engineering Uses*. Van Nostrand, New York, 1928.

Joseph Gani, editor. *The Making of Statisticians*. Springer, New York, 1982. 83

Paola Gario. Guido Castelnuovo: Documents for a biography. *Historia Mathematica*, 28:48–53, 2001.

Vasily I. Glivenko. *Курс теории вероятностей (Course of Probability Theory)*. GONTI, Moscow, 1939.

Boris V. Gnedenko. *Курс теории вероятностей (Theory of Probability)*. Nauka, Moscow, first edition, 1950. Sixth edition 1988.

Boris V. Gnedenko and Andrei N. Kolmogorov. Теория вероятностей (Probability theory). In *Математика в СССР за тридцать лет, 1917–1947 (Thirty Years of Soviet Mathematics, 1917–1947)*, pages 701–727. Гостехиздат, Moscow and Leningrad, 1948. English translation in Sheynin (1998a), pp. 131–158.

Boris V. Gnedenko and Andrei N. Kolmogorov. *Предельные распределения для сумм независимых случайных величин*. State Publishing House, Moscow, 1949. Translated into English by Kai L. Chung: *Limit Distributions for Sums of Independent Random Variables* by Addison-Wesley, Cambridge, Massachusetts, in 1954, with an appendix by Joseph L. Doob.

Shelby J. Haberman. *Advanced Statistics, Volume I: Description of Populations*. Springer, New York, 1996.

Malachi Haim Hacohen. *Karl Popper: The Formative Years, 1902–1945*. Cambridge University Press, Cambridge, 2000.

Jacques Hadamard. Les principes du calcul des probabilités. *Revue de métaphysique et de morale*, 39:289–293, 1922. A slightly longer version of this note, with the title "Les axiomes du calcul des probabilités", was included in *Oeuvres de Jacques Hadamard*, Tome IV, pp. 2161–2162. Centre National de la Recherche Scientifique, Paris, 1968.

Paul R. Halmos. *Measure Theory*. Von Nostrand, New York, 1950.

Paul R. Halmos. *I want to be a mathematician: An automathography*. Springer, New York, 1985.

Joseph Y. Halpern and Mark R. Tuttle. Knowledge, probability, and adversaries. *Journal of the ACM*, 40(4):917–962, 1993.

Felix Hausdorff. Beiträge zur Wahrscheinlichkeitsrechnung. *Sitzungsberichte der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 53:152–178, 1901.

Felix Hausdorff. *Grundzüge der Mengenlehre.* Von Veit, Leipzig, 1914.

Thomas Hawkins. *Lebesgue's Theory of Integration: Its Origins and Development.* Chelsea, New York, second edition, 1975. First edition 1970, University of Wisconsin Press, Madison. The second edition differs only slightly from the first, but it corrects a consequential error on p. 104. Second edition reprinted in 1979 by Chelsea, New York, and then in 2001 by the American Mathematical Society, Providence, RI.

Georg Helm. Die Wahrscheinlichkeitslehre als Theorie der Kollektivbegriffe. *Annalen der Naturphilosophie*, I:364–384, 1902.

Chris C. Heyde and Eugene Seneta, editors. *Statisticians of the Centuries.* Springer, New York, 2001. 73, 93

David Hilbert. Mathematical problems. *Bulletin of the American Mathematical Society*, 8:437–479, 1902. Hilbert's famous address to the International Congress of Mathematicians in Paris in 1900, in which he listed twenty-three open problems central to mathematics. Translated from the German by Mary W. Newson.

Theophil H. Hildebrandt. On integrals related to and extensions of the Lebesgue integral. *Bulletin of the American Mathematical Society*, 24:113–144, 177–202, 1917.

Thomas Hochkirchen. *Die Axiomatisierung der Wahrscheinlichkeitsrechnung und ihre Kontexte: Von Hilberts sechstem Problem zu Kolmogoroffs Grundbegriffen.* Vandenhoeck & Ruprecht, Göttingen, 1999. 18

Philip Holgate. Independent functions: Probability and analysis in Poland between the wars. *Biometrika*, 84:161–173, 1997.

Eberhard Hopf. On causality, statistics, and probability. *Journal of Mathematics and Physics*, 13:51–102, 1934.

Harold Jeffreys. *Scientific Inference.* Cambridge University Press, Cambridge, 1931. Second edition 1957, third 1973.

Harold Jeffreys. *Theory of Probability.* Oxford University Press, Oxford, 1939. Second edition 1948, third 1961.

Børge Jessen. über eine Lebesguesche Integrationstheorie für Funktionen unendlich vieler Veränderlichen. In *Den Syvende Skandinaviske Mathatiker-kongress I Oslo 19–22 August 1929*, pages 127–138, Oslo, 1930. A. W. Brøggers Boktrykkeri.

Børge Jessen. Some analytical problems relating to probability. *Journal of Mathematics and Physics*, 14:24–27, 1935.

Norman L. Johnson and Samuel Kotz. *Leading Personalities in Statistical Sciences*. Wiley, New York, 1997. 93

Mark Kac. The search for the meaning of independence. In Gani (1982), pages 62–72.

Mark Kac. *Enigmas of Chance, an Autobiography*. Harper and Row, New York, 1985.

Jean-Pierre Kahane. Des séries de Taylor au mouvement brownien, avec un aperçu sur le retour. In Pier (1994a), pages 415–429.

Andreas Kamlah. Probability as a quasi-theoretical concept; J. V. Kries' sophisticated account after a century. *Erkenntnis*, 19:239–251, 1983.

John Maynard Keynes. *A Treatise on Probability*. Macmillan, London, 1921.

Aleksandr Ya. Khinchin. Sur la loi forte des grands nombres. *Comptes rendus des Séances de l'Académie des Sciences*, 186:285–287, 1928.

Aleksandr Ya. Khinchin. Учение Мизеса о вероятностях и принципы физической статистики (Mises's work on probability and the principles of statistical physics). *Успехи физических наук*, 9:141–166, 1929.

Aleksandr Ya. Khinchin. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematischen Annalen*, 109:604–615, 1934.

Aleksandr Ya. Khinchin. On the Mises frequentist theory. *Вопросы философии (Questions of Philosophy)*, 15(1 and 2):91–102 and 77–89, 1961. Published after Khinchin's death by Boris Gnedenko. English translation in Sheynin (1998) 99–137, reproduced with footnotes by Reinhard Siegmund in *Science in Context* **17** 391-422 (2004). We have seen only this English translation, not the original.

Aleksandr Ya. Khinchin and Andrei N. Kolmogorov. Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. *Математический сборник (Sbornik: Mathematics)*, 32:668–677, 1925. Translated into Russian on pp. 7–16 of Kolmogorov (1986) and thence into English on pp. 1–10 of Kolmogorov (1992).

Eberhard Knobloch. Emile Borel's view of probability theory. In Vincent F. Hendricks, Stig Arthur Pedersen, and Klaus Frovin Jørgensen, editors, *Probability Theory: Philosophy, Recent History and Relations to Science*, pages 71–95. Kluwer, Dordrecht, 2001.

Andrei N. Kolmogorov. Une série de Fourier-Lebesgue divergente presque partout. *Fundamenta Mathematicae*, 4:324–328, 1923. Translated into Russian on pp. 8–11 of Kolmogorov (1985) and thence into English on pp. 1–7 of Kolmogorov (1991).

Andrei N. Kolmogorov. La définition axiomatique de l'intégrale. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 180:110–111, 1925a. Translated into Russian on pp. 19–20 of Kolmogorov (1985) and thence into English on pp. 13–14 of Kolmogorov (1991).

Andrei N. Kolmogorov. Sur la possibilité de la définition générale de la dérivée, de l'intégrale et de la sommation des séries divergentes. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 180:362–364, 1925b. Translated into Russian on pp. 39–40 of Kolmogorov (1985) and thence into English on pp. 33–34 of Kolmogorov (1991).

Andrei N. Kolmogorov. Über die Summen durch den Zufall bestimmter unabhängiger Grössen. *Mathematische Annalen*, 99:309–319, 1928. An addendum appears in 1929: Volume 102, pp. 484–488. The article and the addendum are translated into Russian on pp. 20–34 of Kolmogorov (1986) and thence into English on pp. 15–31 of Kolmogorov (1992).

Andrei N. Kolmogorov. Общая теория меры и исчисление вероятностей (The general theory of measure and the calculus of probability). *Сборник работ математического раздела, Коммунистическая академия, Секция естественных и точных наук (Collected Works of the Mathematical Chapter, Communist Academy, Section for Natural and Exact Science)*, 1:8–21, 1929a. The Socialist Academy was founded in Moscow in 1918 and was renamed the Communist Academy in 1923 Vucinich (2000). The date 8 January 1927, which appears at the end of the article in the journal, was omitted when the article was reproduced in the second volume of Kolmogorov's collected works (Kolmogorov (1986), pp. 48–58). The English translation, on pp. 48–59 of Kolmogorov (1992), modernizes the article's terminology somewhat: M becomes a "measure" instead of a "measure specification".

Andrei N. Kolmogorov. Über das Gesetz der iterierten Logarithmus. *Mathematische Annalen*, 101:126–135, 1929b. Translated into Russian on pp. 34–44 of Kolmogorov (1986) and thence into English on pp. 32–42 of Kolmogorov (1992).

Andrei N. Kolmogorov. Untersuchungen über den Integralbegriff. *Mathematische Annalen*, 103:654–696, 1930a. Translated into Russian on pp. 96–136 of Kolmogorov (1985) and into English on pp. 100–143 of Kolmogorov (1991).

Andrei N. Kolmogorov. Sur la loi forte des grands nombres. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 191:910–912, 1930b. Translated into Russian on pp. 44–47 of Kolmogorov (1986) and thence into English on pp. 60–61 of Kolmogorov (1992).

Andrei N. Kolmogorov. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104:415–458, 1931. Translated into Russian on pp. 60–105 of Kolmogorov (1985) and thence into English on pp. 62–108 of Kolmogorov (1992).

Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin, 1933. A Russian translation, by Grigory M. Bavli, appeared under the title Основные понятия теории вероятностей (Nauka, Moscow) in 1936, with a second edition, slightly expanded by Kolmogorov with the assistance of Albert N. Shiryaev, in 1974, and a third edition (ФАЗИС, Moscow) in 1998. An English translation by Nathan Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956.

Andrei N. Kolmogorov. Zufällige Bewegungen (Zur Theorie der Brownschen Bewegung). *Annals of Mathematics*, 35:116–117, January 1934a. Translated into Russian on pp. 168–170 of Kolmogorov (1986) and thence into English on pp. 176–178 of Kolmogorov (1992).

Andrei N. Kolmogorov. Review of Lévy (1934). *Zentralblatt für Mathematik und ihre Grenzegebiete*, 8:367, 1934b.

Andrei N. Kolmogorov. О некоторых новых течениях в теории вероятностей (On some modern currents in the theory of probability). In *Труды 2-го Всесоюзного математического съезда, Ленинград, 24–30 июня 1934 г. (Proceedings of the 2nd All-Union Mathematical Congress, Leningrad, 24–30 June 1934)*, volume 1 (Plenary sessions and review talks), pages 349–358, Leningrad and Moscow, 1935. Издательство АН СССР. English translation in Sheynin (2000), pp. 165–173.

Andrei N. Kolmogorov. Letter to Maurice Fréchet. Fonds Fréchet, Archives de l'Académie des Sciences, Paris, 1939.

Andrei N. Kolmogorov. Евгений Евгениевич Слуцкий; Некролог (Obituary for Evgeny Evgenievich Slutsky). *Успехи математических наук (Russian Mathematical Surveys)*, III(4):142–151, 1948a. English translation in Sheynin (1998a), pp. 77–88, reprinted in *Mathematical Scientist* 27:67–74, 2002.

Andrei N. Kolmogorov. The main problems of theoretical statistics (abstract). In *Второе всесоюзное совещание по математической статистике (Second National Conference on Mathematical Statistics)*, pages 216–220, Tashkent, 1948b. English translation in Sheynin (1998b), pp. 219–224.

Andrei N. Kolmogorov. Вероятность (Probability). In *Большая Советская Энциклопедия (Great Soviet Encyclopedia)*, volume 7, pages 508–510. Soviet Encyclopedia Publishing House, Moscow, second edition, 1951. The same article appears on p. 544 of Vol. 4 of the third edition of the encyclopedia, published in 1971, and in a subsidiary work, the Математическая энциклопедия, published in 1987. English translations of both encyclopedias exist.

Andrei N. Kolmogorov. Summary, in Russian, of his address to a conference on statistics in Moscow in 1954. In Anonymous (1954), pages 46–47. English translation in Sheynin (1998b), pp. 225–226.

Andrei N. Kolmogorov. Теория вероятностей (Probability theory). In Aleksandrov et al. (1956), pages Chapter XI; 33–71 of Part 4 in the 1963 English edition; 252–284 of Volume 2 in the 1965 English edition. The Russian edition had three volumes. The English translation, *Mathematics, Its Content, Methods, and Meaning*, was first published in 1962 and 1963 in six volumes by the American Mathematical Society, Providence, RI, and then republished in 1965 in three volumes by the MIT Press, Cambridge, MA. Reprinted by Dover, New York, 1999.

Andrei N. Kolmogorov. On tables of random numbers. *Sankhya, The Indian Journal of Statistics. Series A*, A 25:369–376, 1963. Translated into Russian on pp. 204–213 of Kolmogorov (1987) and back into English on pp. 176–183 of Kolmogorov (1993).

Andrei N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:1–7, 1965. Translated into Russian on pp. 213–223 of Kolmogorov (1987) and back into English as "Three approaches to the definition of the notion of amount of information" on pp. 184–193 of Kolmogorov (1993).

Andrei N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory*, IT-14:662–664, 1968. Translated into Russian on pp. 232–237 of Kolmogorov (1987) and back into English as "To the logical foundations of the theory of information and probability theory" on pp. 203–207 of Kolmogorov (1993).

Andrei N. Kolmogorov. *Избранные труды. Математика и механика.* Nauka, Moscow, 1985. 83, 84, 86

Andrei N. Kolmogorov. *Избранные труды. Теория вероятностей и математическая статистика.* Nauka, Moscow, 1986. 83, 84, 85, 86

Andrei N. Kolmogorov. *Избранные труды. Теория информации и теория алгоритмов.* Nauka, Moscow, 1987. 86

Andrei N. Kolmogorov. *Selected Works of A. N. Kolmogorov. Volume I: Mathematics and Mechanics.* Nauka, Moscow, 1991. Translation by V. M. Volosov of Kolmogorov (1985). 83, 84

Andrei N. Kolmogorov. *Selected Works of A. N. Kolmogorov. Volume II: Probability Theory and Mathematical Statistics.* Kluwer, Dordrecht, 1992. Translation by G. Lindquist of Kolmogorov (1986). 83, 84, 85

Andrei N. Kolmogorov. *Selected Works of A. N. Kolmogorov. Volume III: Information Theory and the Theory of Algorithms.* Kluwer, Dordrecht, 1993. Translation by A. B. Sossinsky of Kolmogorov (1987). 86

Andrei N. Kolmogorov and Sergei V. Fomin. *Elements of the Theory of Functions and Functional Analysis.* Dover, New York, 1999. 68

Bernard O. Koopman. The axioms and algebra of intuitive probability. *Annals of Mathematics*, 41:269–292, 1940.

Samuel Kotz. Statistics in the USSR. *Survey*, 57:132–141, 1965.

Ulrich Krengel. Wahrscheinlichkeitstheorie. In Gerd Fischer, Friedrich Hirzebruch, Winfried Scharlau, and Willi Törnig, editors, *Ein Jahrhundert Mathematik, 1890–1990. Festschrift zum Jubiläum der Deutsche Mathematiker-Vereingigung*, pages 457–489. Vieweg, Braunschweig and Wiesbaden, 1990.

Sylvestre-François Lacroix. *Traité élémentaire du calcul des probabilités*. Bachelier, Paris, 1822. First edition 1816.

Rudolf Laemmel. *Untersuchungen über die Ermittlung von Wahrscheinlichkeiten*. PhD thesis, Universität Zürich, 1904. Excerpts reprinted on pp. 367–377 of Schneider (1988).

Pierre Simon Laplace. *Théorie analytique des probabilités*. Courcier, Paris, 1812. Second edition 1814, third 1820. The third edition was reprinted in Volume 7 of Laplace's *Œuvres complètes*.

Pierre Simon Laplace. *Essai philosophique sur les probabilités*. Courcier, Paris, 1814. The essay first appeared as part of the second edition of the *Théorie analytique*. The fifth and definitive edition of the essay appeared in 1825. A modern edition of the essay, edited by Bernard Bru, was published by Christian Bourgois, Paris, in 1986. The most recent English translation, by Andrew I. Dale, *Philosophical Essay on Probabilities*, was published by Springer, New York, in 1994.

Steffen L. Lauritzen. *Thiele. Pioneer in Statistics*. Oxford University Press, Oxford, 2002.

François Le Lionnais, editor. *Les grands courants de la pensée mathématique*. Cahiers du Sud, Paris, 1948. This volume was reprinted by Blanchard in 1962 and by Hermann in 1998. An English translation, *Great Currents of Mathematical Thought*, was published by Dover, New York, in 1971. 79

Henri Lebesgue. Sur une généralisation de l'intégrale définie. *Comptes rendus des Séances de l'Académie des Sciences*, 132:1025–1028, 1901.

Henri Lebesgue. *Leçons sur l'intégration et la recherche des fonctions primitives*. Gauthier-Villars, Paris, 1904. Second edition 1928.

Paul Lévy. *Calcul des probabilités*. Gauthier-Villars, Paris, 1925. 9

Paul Lévy. Généralisation de l'espace différentiel de N. Wiener. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 198:786–788, 1934.

Paul Lévy. *Théorie de l'addition des variables aléatoires*. Gauthier-Villars, Paris, 1937. Second edition 1954. 9

Paul Lévy. Les fondements du calcul des probabilités. *Dialectica*, 3(9/10): 55–64, 1949.

Paul Lévy. Les fondements du calcul des probabilités. *Revue de métaphysique et de morale*, 59(2):164–179, 1954. Reprinted in Lévy (1973–1980), Volume VI, pp. 95–110.

Paul Lévy. Un paradoxe de la théorie des ensembles aléatoire. *Comptes rendus des Séances de l'Académie des Sciences*, 248:181–184, 1959. Reprinted in Lévy (1973–1980), Volume VI, pp. 67–69.

Paul Lévy. *Quelques aspects de la pensée d'un mathématicien.* Blanchard, Paris, 1970.

Paul Lévy. *Œuvres de Paul Lévy.* Gauthier-Villars, Paris, 1973–1980. In six volumes. Edited by Daniel Dugué. 88

Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer, New York, second edition, 1997.

Jean-Baptiste-Joseph Liagre. *Calcul des probabilités et théorie des erreurs avec des applications aux sciences d'observation en général et à la géodésie en particulier.* Muquardt, Brussels, second edition, 1879. First edition 1852. Second edition prepared with the assistance of Camille Peny.

Michel Loève. *Probability theory: foundations, random sequences.* Van Nostrand, Princeton, NJ, 1955. Second edition 1960.

Antoni Łomnicki. Nouveaux fondements du calcul des probabilités (Définition de la probabilité fondée sur la théorie des ensembles). *Fundamenta Mathematicae*, 4:34–71, 1923.

Zbigniew Łomnicki and Stanisław Ulam. Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités. I. Variables indépendantes. *Fundamenta Mathematicae*, 23:237–278, 1934.

George G. Lorentz. Mathematics and politics in the Soviet Union from 1928 to 1953. *Journal of Approximation Theory*, 116:169–223, 2002.

Jeff Loveland. Buffon, the certainty of sunrise, and the probabilistic reductio ad absurdum. *Archive for History of Exact Sciences*, 55:465–477, 2001.

Jan Łukasiewicz. *Die logischen Grundlagen der Warhscheinlichkeitsrechnung.* Wspolka Wydawnicza Polska, Kraków, 1913.

Hugh MacColl. The calculus of equivalent statements (fourth paper). *Proceedings of the London Mathematical Society*, 11:113–121, 1880.

Hugh MacColl. The calculus of equivalent statements (sixth paper). *Proceedings of the London Mathematical Society*, 28:555–579, 1897.

Saunders MacLane. Mathematics at Göttingen under the Nazis. *Notices of the American Mathematical Society*, 42:1134–1138, 1995.

Leonid E. Maistrov. *Probability Theory: A Historical Sketch.* Academic Press, New York, 1974. Translated and edited by Samuel Kotz.

Andrei A. Markov. *Исчисление вероятностей (Calculus of Probability).* Типография Императорской Академии Наук, Saint Petersburg, 1900. Second edition 1908, fourth 1924. 89

Andrei A. Markov. *Wahrscheinlichkeitsrechnung.* Teubner, Leipzig, 1912. Translation of second edition of Markov (1900). Free on-line at http://historical.library.cornell.edu.

Thierry Martin. *Probabilités et critique philosophique selon Cournot.* Vrin, Paris, 1996.

Thierry Martin. *Bibliographie cournotienne.* Annales littéraires de l'Université Franche-Comté, Besançon, 1998.

Thierry Martin. Probabilité et certitude. In Thierry Martin, editor, *Probabilités subjectives et rationalité de l'action*, pages 119–134. CNRS Éditions, Paris, 2003.

Per Martin-Löf. The literature on von Mises' Kollektivs revisited. *Theoria*, 35: 12–37, 1969.

Per Martin-Löf. Statistics from the point of view of statistical mechanics. Notes by Ole Jørsboe. Matematisk Institut. Aarhus University. 36 pp, December 1966–February 1967.

Per Martin-Löf. Statistiska modeller. Notes by Rold Sundberg. Institutet för försäkringsmatematik ock mathematisk statistik. Stockholm University. 158 pp, September 1969–May 1970.

Pesi R. Masani. *Norbert Wiener, 1894–1964.* Birkhäuser, Basel, 1990.

Laurent Mazliak. Andrei Nikolaievitch Kolmogorov (1902–1987). Un aperçu de l'homme et de l'œuvre probabiliste. Prépublication PMA-785, http://www.probab.jussieu.fr. Université Paris VI, 2003.

Paolo Medolaghi. La logica matematica ed il calcolo delle probabilitá. *Bollettino dell'Associazione Italiana per l'Incremento delle Scienze Attuariali*, 18: 20–39, 1907.

Alexius Meinong. *Über Möglichkeit und Wahrscheinlichkeit: Beiträge zur Gegenstandstheorie und Erkenntnistheorie.* Barth, Leipzig, 1915.

Philip Mirowski, editor. *Edgeworth on Chance, Economic Hazard, and Statistics.* Rowman & Littlefield, Lanham, MD, 1994.

Edward C. Molina. Arne Fisher, 1887–1944. *Journal of the American Statistical Association*, 39:251–252, 1944.

Jacob Mordukh. О связанных испытаниях, отвечающих условию стохастической коммутативности (On connected trials satisfying the condition of stochastic commutativity). *Труды русских ученых за границей (Work of Russian Scientists Abroad, published in Berlin)*, 2:102–125, 1923. English translation in Sheynin (2000), pp. 209–223. We have seen only this English translation, not the original.

Ernest Nagel. *Principles of the Theory of Probability*. University of Chicago Press, Chicago, 1939. Volume I, Number 6 of the International Encyclopedia of Unified Science, edited by Otto Neurath.

Jerzy Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society, Series A*, 236:333–380, 1937. 56

Jerzy Neyman. L'estimation statistique, traitée comme un problème classique de probabilités. In Wavre (1938–1939), pages 25–57 of the sixth fascicle, number 739, *Conceptions diverses*. This celebrated colloquium, chaired by Maurice Fréchet, was held in October 1937 at the University of Geneva. Participants included Cramér, Dœblin, Feller, de Finetti, Heisenberg, Hopf, Lévy, Neyman, Pòlya, Steinhaus, and Wald, and communications were received from Bernstein, Cantelli, Glivenko, Jordan, Kolmogorov, von Mises, and Slutsky. The proceedings were published by Hermann in eight fascicles in their series *Actualités Scientifiques et Industrielles*. The first seven fascicles appeared in 1938 as numbers 734 through 740; the eighth, de Finetti's summary of the colloquium, appeared in 1939 as number 766. See de Finetti (1939).

Jerzy Neyman. Foundation of the general theory of statistical estimation. In Bayer (1951), pages 83–95.

Otton Nikodym. Sur une généralisation des intégrales de M. J. Radon. *Fundamenta Mathematicae*, 15:131–179, 1930.

Kh. O. Ondar, editor. *The Correspondence Between A. A. Markov and A. A. Chuprov on the Theory of Probability and Mathematical Statistics*. Springer, New York, 1981. Translated from the Russian by Charles M. and Margaret Stein.

Octav Onicescu. Le livre de G. Castelnuovo "Calcolo della Probabilità e Applicazioni" comme aboutissant de la suite des grands livres sur les probabilités. In *Simposio internazionale di geometria algebrica (Roma, 30 settembre–5 ottobre 1965)*, pages xxxvii–liii, Rome, 1967. Edizioni Cremonese.

R. E. A. C. Paley, Norbert Wiener, and Antoni Zygmund. Notes on random functions. *Mathematische Zeitschrift*, 37:647–668, 1933.

A. Pallez. Normes de la statistique, du calcul des probabilités et des erreurs de mesures. *Journal de la Société de Statistique de Paris*, pages 125–133, 1949.

Egon S. Pearson. *'Student': A Statistical Biography of William Sealy Gosset.* Oxford University Press, Oxford, 1990. Based on writings by E. S. Pearson. Edited and augmented by R. L. Plackett, with the assistance of G. A. Barnard.

Charles Saunders Peirce. On an improvement in Boole's calculus of logic. *Proceedings of the American Academy of Arts and Sciences*, 7:250–261, 1867.

Charles Saunders Peirce. Illustrations of the logic of science. Fourth paper— The probability of induction. *Popular Science Monthly*, 12:705–718, 1878.

Jean-Paul Pier, editor. *Development of Mathematics 1900–1950.* Birkhäuser, Basel, 1994a. 78, 83, 91

Jean-Paul Pier. Intégration et mesure 1900–1950. In *Development of Mathematics 1900–1950* Pier (1994a), pages 517–564.

Henri Poincaré. Sur le problème des trois corps et les équations de la dynamique. *Acta Mathematica*, 13:1–271, 1890.

Henri Poincaré. *Calcul des probabilités. Leçons professées pendant le deuxième semestre 1893–1894.* Gauthier-Villars, Paris, 1896. Free on-line at http://historical.library.cornell.edu. 91

Henri Poincaré. *Calcul des probabilités.* Gauthier-Villars, Paris, 1912. Second edition of Poincaré (1896).

Siméon-Denis Poisson. *Recherches sur la probabilité des judgments en matière criminelle et en matière civile, précédés des règles générale du calcul des probabilités.* Bachelier, Paris, 1837.

Karl R. Popper. *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft.* Springer, Vienna, 1935. An English translation, *The Logic of Scientific Discovery*, with extensive new appendices, was published by Hutchinson, London, in 1959.

Karl R. Popper. A set of independent axioms for probability. *Mind*, 47:275–277, 1938.

Karl R. Popper. *Realism and the Aim of Science.* Hutchinson, London, 1983.

Theodore Porter. *The Rise of Statistical Thinking, 1820–1900.* Princeton University Press, Princeton, NJ, 1986.

Pierre Prevost and Simon Lhuilier. Mémoire sur l'art d'estimer la probabilité des causes par les effets. *Mémoires de l'Académie royale des Sciences et Belle-Lettres, Classe de Mathématique, Berlin*, pages 3–25, 1799. Volume for 1796.

Hans Rademacher. Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen. *Mathematische Annalen*, 87:112–138, 1922.

Johann Radon. Theorie und Anwendungen der absolut additiven Mengenfunktionen. *Sitzungsberichte der kaiserlichen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse*, 122_{IIa}:1295–1438, 1913. Reprinted in his *Collected Works*, 1:45–188. Birkhäuser, Basel, 1987.

Frank P. Ramsey. *The Foundations of Mathematics and Other Logical Essays*. Routledge, London, 1931.

Eugenio Regazzini. Teoria e calcolo delle probabilità. In Angelo Guerraggio, editor, *La Matematica Italiana tra le Due Guerre Mondiali*, pages 339–386, Bologna, 1987a. Pitagora. Reprinted with minor changes as pp. 569–621 of *La Matematica Italiana dopo l'Unità: Gli anni tra le due guerre mondiali*, edited by Simonetta Di Sieno, Angelo Guerraggio, and Pietro Nastasi, Marcos y Marcos, Milan, 1998.

Eugenio Regazzini. Probability theory in Italy between the two world wars. A brief historical review. *Metron*, 45(3–4):5–42, 1987b.

Hans Reichenbach. *Der Begriff der Wahrscheinlichkeit für die mathematische Darstellung der Wirklichkeit*. Barth, Leipzig, 1916.

Hans Reichenbach. Axiomatik der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 34:568–619, 1932.

Hans Richter. Zur Begründung der Wahrscheinlichkeitsrechnung. *Dialectica*, 8:48–77, 1954.

Hans Richter. *Wahrscheinlichkeitstheorie*. Springer, Berlin, 1956.

Henry L. Rietz. Review of Kolmogorov (1933). *Bulletin of the American Mathematical Society*, 40:522–523, 1934.

L. C. G. Rogers and David Williams. *Diffusions, Markov Processes, and Martingales. Volume 1. Foundations*. Cambridge University Press, Cambridge, second, reprinted edition, 2000.

Stanisław Saks. *Théorie de l'intégrale*. Monografie Matematyczne, Warsaw, 1933.

Stanisław Saks. *Theory of the Integral*. Stechert, New York, 1937.

Ivo Schneider, editor. *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933: Einführungen und Texte*. Wissenschaftliche Buchgesellschaft, Darmstadt, 1988. 73, 87

Laurent Schwartz. *Radon Measures on Arbitrary Spaces and Cylindrical Measures*. Oxford, New York, 1973.

Laurent Schwartz. *Un Mathématician aux prises avec le siècle.* Jacob, Paris, 1997. A translation, *A mathematician grappling with his century*, was published by Birkhäuser, Basel, in 2001.

Irving Ezra Segal. Norbert Wiener. November 26, 1894–March 18, 1964. *Biographical Memoires, National Academy of Sciences of the United States of America*, 61:388–436, 1992.

Sanford L. Segal. *Mathematicians under the Nazis.* Princeton University Press, Princeton, NJ, 2003.

Eugene Seneta. On the history of the law of large numbers and Boole's inequality. *Historia Mathematica*, 19:24–39, 1992.

Eugene Seneta. Boltzmann, Ludwig Edward. In *Leading Personalities in Statistical Sciences* Johnson and Kotz (1997), pages 353–354.

Eugene Seneta. Aleksander Aleksandrovich Chuprov (or Tschuprow). In Heyde and Seneta (2001), pages 303–307.

Eugene Seneta. Mathematics, religion and Marxism in the Soviet Union in the 1930s. *Historia Mathematica*, 71:319–334, 2003.

Glenn Shafer. Conditional probability. *International Statistical Review*, 53: 261–277, 1985.

Glenn Shafer. *The Art of Causal Conjecture.* The MIT Press, Cambridge, MA, 1996.

Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game.* Wiley, New York, 2001.

Oscar Sheynin. *Aleksandr A. Chuprov: Life, Work, Correspondence. The making of mathematical statistics.* Vandenhoeck & Ruprecht, Göttingen, 1996.

Oscar Sheynin, editor. *From Markov to Kolmogorov. Russian papers on probability and statistics. Containing essays of S. N. Bernstein, A. A. Chuprov, B. V. Gnedenko, A. Ya. Khinchin, A. N. Kolmogorov, A. M. Liapunov, A. A. Markov and V. V. Paevsky.* Hänsel-Hohenhausen, Egelsbach, Germany, 1998a. Translations from Russian into English by the editor. Deutsche Hochschulschriften No. 2514. In microfiche.

Oscar Sheynin, editor. *From Davidov to Romanovsky. More Russian papers on probability and statistics. S. N. Bernstein, B. V. Gnedenko, A. N. Kolmogorov, A. M. Liapunov, P. A. Nekrasov, A. A. Markov, Kh.O. Ondar, T. A. Sarymsakov, N. V. Smirnov.* Hänsel-Hohenhausen, Egelsbach, Germany, 1998b. Translations from Russian into English by the editor. Deutsche Hochschulschriften No. 2579. In microfiche.

Oscar Sheynin. *Russian Papers on the History of Probability and Statistics.* Hänsel-Hohenhausen, Egelsbach, Germany, 1999a. This volume contains articles by Oscar Sheynin, originally published in Russian and now translated into English by the author. Deutsche Hochschulschriften No. 2621. In microfiche.

Oscar Sheynin, editor. *From Bortkiewicz to Kolmogorov. Yet more Russian papers on probability and statistics.* Hänsel-Hohenhausen, Egelsbach, Germany, 1999b. Translations from Russian into English by the editor. Deutsche Hochschulschriften No. 2656. In microfiche.

Oscar Sheynin, editor. *From Daniel Bernoulli to Urlanis. Still more Russian papers on probability and statistics.* Hänsel-Hohenhausen, Egelsbach, Germany, 2000. Translations from Russian into English by the editor. Deutsche Hochschulschriften No. 2696. In microfiche.

Oscar Sheynin. Fechner as statistician. *British Journal of Mathematical and Statistical Psychology*, 57:53–72, 2004.

Albert N. Shiryaev. Kolmogorov: Life and creative activities. *Annals of Probability*, 17:866–944, 1989.

Albert N. Shiryaev, editor. *Колмогоров в воспоминаниях (Kolmogorov in Remembrance).* Nauka, Moscow, 1993. 70

Albert N. Shiryaev. Andrei Nikolaevich Kolmogorov (April 25, 1903–October 20, 1987). A biographical sketch of his life and creative paths. In Flower (2000), pages 1–87. This article originally appeared in Russian in *Kolmogorov in Remembrance*, edited by Albert N. Shiryaev, Nauka, Moscow, 1993.

Albert N. Shiryaev. Жизнь и творчество. Биографический очерк (Life and creative work. Biographical essay). In *Колмогоров: Юбилейное издание в 3-х книгах (Kolmogorov: Jubilee Publication in Three Volumes)* Shiryaev (2003b), pages 17–209.

Albert N. Shiryaev, editor. *Колмогоров: Юбилейное издание в 3-х книгах (Kolmogorov: Jubilee Publication in Three Volumes).* ФИЗМАТЛИТ (Physical-Mathematical Literature), Moscow, 2003b. 94

Reinhard Siegmund-Schultze. Mathematicians forced to philosophize: An introduction to Khinchin's paper on von Mises' theory of probability. *Science in Context*, 17(3):373–390, 2004.

Wacław Sierpiński. Sur une définition axiomatique des ensembles mesurables (L). *Bulletin International de l'Academie des Sciences de Cracovie A*, pages 173–178, 1918. Reprinted on pp. 256–260 of Sierpiński's *Oeuvres choisies*, Volume II, PWN (Polish Scientific Publishers), Warsaw, 1975.

Evgeny Slutsky. К вопросу о логических основах теории вероятностей (On the question of the logical foundation of the theory of probability). *Вестник статистики (Bulletin of Statistics)*, 12:13–21, 1922.

Evgeny Slutsky. Über stochastische Asymptoten und Grenzwerte. *Metron*, 5: 3–89, 1925.

Hugo Steinhaus. Les probabilités dénombrables et leur rapport à la théorie de la mesure. *Fundamenta Mathematicae*, 4:286–310, 1923.

Hugo Steinhaus. Über die Wahrscheinlichkeit dafür, daß der Konvergenzkreis einer Potenzreihe ihre natürliche Grenze ist. *Mathematische Zeitschrift*, 31: 408–416, 1930a. Received by the editors 5 August 1929.

Hugo Steinhaus. Sur la probabilité de la convergence de séries. Première communication. *Studia Mathematica*, 2:21–39, 1930b. Received by the editors 24 October 1929.

Stephen M. Stigler. Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. *Journal of the American Statistical Association*, 68: 872–879, 1973.

Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge, MA, 1986.

Marshall Stone. Mathematics and the future of science. *Bulletin of the American Mathematical Society*, 63:61–76, 1957.

Angus E. Taylor. A study of Maurice Fréchet: I. His early work on point set theory and the theory of functionals. *Archive for History of Exact Sciences*, 27:233–295, 1982.

Angus E. Taylor. A study of Maurice Fréchet: II. Mainly about his work on general topology, 1909–1928. *Archive for History of Exact Sciences*, 34: 279–380, 1985.

Angus E. Taylor. A study of Maurice Fréchet: III. Fréchet as analyst, 1909–1930. *Archive for History of Exact Sciences*, 37:25–76, 1987.

Erhard Tornier. Grundlagen der Wahrscheinlichkeitsrechnung. *Acta Mathematica*, 60:239–380, 1933.

Stanisław Ulam. Zur Masstheorie in der allgemeinen Mengenlehre. *Fundamenta Mathematicae*, 16:140–150, 1930.

Stanisław Ulam. Zum Massbegriffe in Produkträumen. In *Verhandlung des Internationalen Mathematiker-Kongress Zürich*, volume II, pages 118–119, 1932.

Stanisław Ulam. What is measure? *American Mathematical Monthly*, 50: 597–602, 1943.

Friedrich Maria Urban. *Grundlagen der Wahrscheinlichkeitsrechnung und der Theorie der Beobachtungsfehler*. Teubner, Leipzig, 1923.

James V. Uspensky. *Introduction to Mathematical Probability*. McGraw-Hill, New York, 1937.

David Van Dantzig. Sur l'analyse logique des relations entre le calcul des probabilités et ses applications. In Bayer (1951), pages 49–66.

E. B. Van Vleck. The rôle of the point-set thoery in geometry and physics. Presidential address delivered before the American Mathematical Society, January 1, 1915. *Bulletin of the American Mathematical Society*, 21:321–341, 1915.

John Venn. *The Logic of Chance*. Macmillan, London and New York, third edition, 1888.

Jean-André Ville. Sur la notion de collectif. *Comptes rendus des Séances de l'Académie des Sciences*, 203:26–27, 1936.

Jean-André Ville. *Étude critique de la notion de collectif*. Gauthier-Villars, Paris, 1939. This differs from Ville's dissertation, which was defended in March 1939, only in that a 17-page introductory chapter replaces the dissertation's one-page introduction.

Jean-André Ville. Letter to Pierre Crépel, dated February 2, 1985, 1985.

Ladislaus von Bortkiewicz. Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik. In *Encyklopädie der mathematischen Wissenschaften, Bd. I, Teil 2*, pages 821–851. Teubner, Leipzig, 1901.

Ladislaus von Bortkiewicz. *Die Iterationen. Ein Beitrag zur Wahrscheinlichkeitstheorie*. Springer, Berlin, 1917. Free on-line at http://historical.library.cornell.edu.

Guy von Hirsch. Sur un aspect paradoxal de la théorie des probabilités. *Dialetica*, 8:125–144, 1954.

Richard von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919.

Richard von Mises. *Wahrscheinlichkeitsrechnung, Statistik und Wahrheit*. Springer, Vienna, 1928. Second edition 1936, third 1951. A posthumous fourth edition, edited by his widow Hilda Geiringer, appeared in 1972. English editions, under the title *Probability, Statistics and Truth*, appeared in 1939 and 1957.

Richard von Mises. *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. Deuticke, Leipzig and Vienna, 1931.

John von Neumann. Zur Operatorenmethode in der klassischen Mechanik. *Annals of Mathematics*, 33(2):587–642, 1932.

Jan von Plato. *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective*. Cambridge University Press, Cambridge, 1994.

Marian von Smoluchowski. Zarys kinetycznej teoriji ruchów Browna i roztworów metnych. *Rozprawy i Sprawozdania z Posiedzeń Wydziału Matematyczno-Przyrodniczego Akademii Umiejetności*, 3:257–282, 1906. A German translation appeared in the *Annalen der Physik*, 21, pp. 756–780 in 1906.

Vladimir Vovk and Glenn Shafer. Kolmogorov's contributions to the foundations of probability. *Problems of Information Transmission*, 39(1):21–31, 2003. Prepublication version at http://www.probabilityandfinance.com.

Alexander Vucinich. Mathematics and dialectics in the Soviet Union: The pre-Stalin period. *Historia Mathematica*, 26:107–124, 1999.

Alexander Vucinich. Soviet mathematics and dialectics in the Stalin era. *Historia Mathematica*, 27:54–76, 2000. 84

Alexander Vucinich. Soviet mathematics and dialectics in the post-Stalin era: New horizons. *Historia Mathematica*, 29:13–39, 2002.

Abraham Wald. Sur la notion de collectif dans le calcul des probabilités. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, 202: 180–183, 1936.

Abraham Wald. Die Widerspruchfreiheit des Kollectivbegriffes der Wahrscheinlichkeitsrechnung. *Ergebnisse eines Mathematischen Kolloquiums*, 8:38–72, 1937. This journal, or series of publications, reported results from Karl Menger's famous Vienna Colloquium. Participants included von Neumann, Morgenstern, and Wald. The eighth volume was the last in the series, because the colloquium ended with the Nazi invasion of Austria in 1938. In 1939, Menger started a second series in English, *Reports of a mathematical colloquium*, at the University of Notre Dame.

Abraham Wald. Die Widerspruchfreiheit des Kollectivbegriffes. In Wavre (1938–1939), pages 79–99 of the second fascicle, number 735, *Les fondements du calcul des probabilités*. This celebrated colloquium, chaired by Maurice Fréchet, was held in October 1937 at the University of Geneva. Participants included Cramér, Dœblin, Feller, de Finetti, Heisenberg, Hopf, Lévy, Neyman, Pòlya, Steinhaus, and Wald, and communications were received from Bernstein, Cantelli, Glivenko, Jordan, Kolmogorov, von Mises, and Slutsky. The proceedings were published by Hermann in eight fascicles in their series *Actualités Scientifiques et Industrielles*. The first seven fascicles appeared in 1938 as numbers 734 through 740; the eighth, de Finetti's summary of the colloquium, appeared in 1939 as number 766. See de Finetti (1939).

Rolin Wavre. *Colloque consacré à la théorie des probabilités.* Hermann, Paris, 1938–1939. This celebrated colloquium, chaired by Maurice Fréchet, was held in October 1937 at the University of Geneva. Participants included Cramér, Dœblin, Feller, de Finetti, Heisenberg, Hopf, Lévy, Neyman, Pòlya, Steinhaus, and Wald, and communications were received from Bernstein, Cantelli, Glivenko, Jordan, Kolmogorov, von Mises, and Slutsky. The proceedings were published by Hermann in eight fascicles in their series *Actualités Scientifiques et Industrielles.* The first seven fascicles appeared in 1938 as numbers 734 through 740; the eighth, de Finetti's summary of the colloquium, appeared in 1939 as number 766. See de Finetti (1939). 76, 79, 80, 90, 97

André Weil. Calcul de probabilités, méthode axiomatique, intégration. *Revue scientifique*, 78:201–208, 1940.

Harald Westergaard. *Die Grundzüge der Theorie der Statistik.* Fischer, Jena, 1890. Second edition, with H. C. Nybølle as co-author, 1928.

Peter Whittle. *Probability via Expectation.* Springer, New York, fourth edition, 2000.

William A. Whitworth. *Choice and Chance.* Cambridge University Press, Cambridge, third edition, 1878. Fourth edition 1886, fifth 1901.

Norbert Wiener. The mean of a functional of arbitrary elements. *Annals of Mathematics*, 22:66–72, December 1920. 98

Norbert Wiener. The average of an analytical functional. *Proceedings of the National Academy of Science*, 7:253–260, 1921a. 98

Norbert Wiener. The average of an analytical functional and the Brownian movement. *Proceedings of the National Academy of Science*, 7:294–298, 1921b. 98

Norbert Wiener. Differential-space. *Journal of Mathematics and Physics*, 2: 131–174, 1923. 98

Norbert Wiener. The average value of a functional. *Proceedings of the London Mathematical Society*, 22:454–467, 1924. 98

Norbert Wiener. Random functions. *Journal of Mathematics and Physics*, 14: 17–23, 1935.

Norbert Wiener. *I am a Mathematician. The Later Life of a Prodigy.* Doubleday, Garden City, NY, 1956.

Norbert Wiener. *Collected Works.* MIT Press, Cambridge, MA, 1976–1985. Four volumes. Edited by Pesi Masani. Volume 1 includes Wiener's early papers on Brownian motion (Wiener 1920 1921ab 1923 1924), with a commentary by Kiyosi Itô.

Anders Wiman. Über eine Wahrscheinlichkeitsaufgabe bei Kettenbruchent-wicklungen. *Öfversigt af Kongliga Vetenskaps-Akademiens Förhandlingar. Femtiondesjunde Årgången*, 57(7):829–841, 1900.

Anders Wiman. *Bemerkung über eine von Gyldén aufgeworfene Wahrschein-lichkeitsfrage*. Håkan Ohlssons boktrykeri, Lund, 1901.

William Henry Young. On a new method in the theory of integration. *Procee-dings of the London Mathematical Society*, 9:15–30, 1911.

William Henry Young. On integration with respect to a function of bounded variation. *Proceedings of the London Mathematical Society*, 13:109–150, 1915.

Smilka Zdravkovska and Peter L. Duren, editors. *Golden Years of Moscow Mat-hematics*. American Mathematical Society and London Mathematical Society, Providence, RI, and London, 1993. Volume 6 of the History of Mathematics Series. 70

Joseph D. Zund. George David Birkhoff and John von Neumann: A question of priority and the ergodic theorems, 1931–1932. *Historia Mathematica*, 29: 138–156, 2002.

Lifespans

George Biddell Airy (1801–1892)
Aleksandr Danilovich Aleksandrov (1912–1999) (Александр Данилович Алек-
    сандров)
Pavel Sergeevich Aleksandrov (1896–1982) (Павел Сергеевич Александров)
Erik Sparre Andersen (1919–2003)
Oskar Johann Victor Anderson (1887–1960) (Оскар Николаевич Андерсон)
Vladimir Igorevich Arnol'd (1937–2010) (Владимир Игоревич Арнольд)
Louis Jean-Baptiste Alphonse Bachelier (1870–1946)
Stefan Banach (1892–1945)
Marc Barbut (1928–2011)
Maya Bar-Hillel
I. Alfred Barnett (1894–1975)
Jack Barone
Maurice Stephenson Bartlett (1910–2002)
Grigory Minkelevich Bavli (1908–1941) (Григорий Минкелевич Бавли)
Raymond Bayer (1898–1959)
Margherita Benzi (born 1957)
Jacob Bernoulli (1654–1705)
Claude Bernard (1813–1878)
Felix Bernstein (1878–1956)
Sergei Natanovich Bernstein (1880–1968) (Сергей Натанович Бернштейн)
Joseph Louis François Bertrand (1822–1900)
Nic H. Bingham (born 1945)
George David Birkhoff (1884–1944)

David Blackwell (1919–2010)

Alain Blum

Georg Bohlmann (1869–1928)

Ludwig Eduard Boltzmann (1844–1906)

George Boole (1815–1864)

Émile Félix-Edouard-Justin Borel (1871–1956)

Ladislaus von Bortkiewicz (1868–1931) (Владислав Иосифович Борткевич)

Marcel Brissaud

Ugo Broggi (1880–1965)

Robert Brown (1773–1858)

Bernard Bru (born 1942)

Ernst Heinrich Bruns (1848–1919)

Stephen George Brush

George-Louis Leclerc de Buffon (1707–1788)

Francesco Paolo Cantelli (1875–1966)

Constantin Carathéodory (1873–1950)

Rudolf Carnap (1891–1970)

Nicolas Léonard Sadi Carnot (1796–1832)

Guido Castelnuovo (1865–1952)

Carl Wilhelm Ludvig Charlier (1862–1934)

Kai Lai Chung (1917–2009)

Aleksandr Aleksandrovich Chuprov (1874–1926) (Александр Александрович
    Чупров)

Alonzo Church (1903–1995)

Cifarelli, Donato Michele

Auguste Comte (1798–1857)

Julian Lowell Coolidge (1873–1954)

Arthur H. Copeland Sr. (1898–1970)

Antoine-Augustin Cournot (1801–1877)

Jean-Michel Courtault

Thomas Merrill Cover (1938–2012)

Richard T. Cox (1898–1991)

Harald Cramér (1893–1985)

Pierre Crépel (born 1947)

Emanuel Czuber (1851–1925)

Jean-le-Rond D'Alembert (1717–1783)

Percy John Daniell (1889–1946)

Lorraine Daston (born 1951)

Bruno de Finetti (1906–1985)

Claude Dellacherie (born 1943)

Sergei S. Demidov (born 1942) (Сергей Демидов)

Abraham De Moivre (1667–1754)

Augustus De Morgan (1806–1871)

Jean Alexandre Eugène Dieudonné (1906–1992)

Wolfgang Doeblin (1915–1940)

Joseph L. Doob (1910–2004)

Karl Dörge (1899–1977)

Louis-Gustave Du Pasquier (1876–1957)

Peter Larkin Duren (born 1935)

Evgeny Borisovich Dynkin (1924–2014) (Евгений Борисович Дынкин)

Francis Ysidro Edgeworth (1845–1926)

Albert Einstein (1879–1955)

Robert Leslie Ellis (1817–1859)

Agner Krarup Erlang (1878–1929)

Georg Faber (1877–1966)

Ruma Falk

Gustav Theodor Fechner (1801–1887)

Willy Feller (1906–1970) (William after his immigration to the U.S.)

Arne Fisher (1887–1944)

Ronald Aylmer Fisher (1890–1962)

Sergei Vasil'evich Fomin (1917–1975) (Сергей Васильевич Фомин)

Robert M. Fortet (1912–1998)

Jean Baptiste Joseph Fourier (1768–1830)

Abraham Adolf Fraenkel (1891–1965)

Réné-Maurice Fréchet (1878–1973)

John Ernst Freund (1921–2004)

Hans Freudenthal (1905–1990)

Thornton Carl Fry (1892–1991)

Peter Gács (born 1947)

Joseph Mark Gani (1924–2016)

Réné Gâteaux (1889–1914)

Hans Geiger (1882–1945)

Valery Ivanovich Glivenko (1897–1940) (Валерий Иванович Гливенко)

Boris Vladimirovich Gnedenko (1912–1995) (Борис Владимирович Гнеденко)

Vasily Leonidovich Goncharov (1896–1955) (Василий Леонидович Гончаров)

William Sealy Gossett (1876–1937)

Jorgen Pederson Gram (1850–1916)

Robert Molten Gray (born 1943)

Shelby Joel Haberman (born 1947)

Ian Hacking (born 1936)

Malachi Haim Hacohen

Jacques Hadamard (1865–1963)

Maurice Halbwachs (1877–1945)

Anders Hald (1913–2007)

Paul Richard Halmos (1916–2006)

Joseph Y. Halpern

Godfrey H. Hardy (1877–1947)

Felix Hausdorff (1868–1942)

Thomas Hawkins (born 1938)

Georg Helm (1851–1923)

Christopher Charles Heyde (1939–2008)

David Hilbert (1862–1943)

Theophil Henry Hildebrandt (1888–1980)

Eberhard Hopf (1902–1983)

Philip Holgate (1934–1993)

Harold Hotelling (1895–1973)

Kiyosi Itô (1915–2008)

Harold Jeffreys (1891–1989)

Børge Jessen (1907–1993)

Norman Lloyd Johnson (1917–2004)

François Jongmans

Camille Jordan (1838–1922)

Youri Kabanov (born 1948) (Юрий Кабанов)

Mark Kac (1914–1984)

Jean-Pierre Kahane (1926–2017)

Immanuel Kant (1724–1804)

John Maynard Keynes (1883–1946)

Aleksandr Yakovlevich Khinchin (1894–1959) (Александр Яковлевич Хинчин)

Eberhard Knobloch (born 1943)

Andrei Nikolaevich Kolmogorov (1903–1987) (Андрей Николаевич Колмого-
ров)

Bernard O. Koopman (1900–1981)

Samuel Borisovich Kotz (1930–2010)

Ulrich Krengel (born 1937)

Sylvestre-François Lacroix (1765–1843)

Rudolf Laemmel (1879–1972)

Pierre Simon de Laplace (1749–1827)

Steffen Lauritzen (born 1947)

Mikhail Alekseevich Lavrent'ev (1900–1980) (Михаил Алексеевич Лаврентьев)

Henri Lebesgue (1875–1941)

Mikhail Aleksandrovich Leontovich (1903–1981) (Михаил Александрович
Леонтович)

Boris Venediktovich Levshin (Борис Венедиктович Левшин)

Paul Pierre Lévy (1886–1971)

Simon Antoine Jean Lhuilier (1750–1840)

Ming Li (born 1955)

Jean-Baptiste-Joseph Liagre (1815–1891)

G. Lindquist

Michel Loève (1907–1979)

Bernard Locker (born 1947)

Antoni Marjan Łomnicki (1881–1941)

Zbigniew Łomnicki (1904–1994)

Jerzy Łoś (1920–1998)

Jeff A. Loveland

Jan Łukasiewicz (1878–1956)

Ernest Filip Oskar Lundberg (1876–1965)

Nikolai Nikolaevich Luzin (1883–1950) (Николай Николаевич Лузин)

Leonid Efimovich Maistrov (born 1920) (Леонид Ефимович Майстров)

Hugh MacColl (1837–1909)
Andrei Andreevich Markov (1856–1922) (Андрей Андреевич Марков)
Thierry Martin (born 1950)
Per Martin-Löf (born 1942)
Pesi R. Masani
Laurent Mazliak (born 1968)
Paolo Medolaghi (1873–1950)
Alexius Meinong (1853–1920)
Karl Menger (1902–1985)
Martine Mespoulet
Paul-André Meyer (1934–2003)
Philip Mirowski (born 1951)
Edward Charles Molina (1877–1964)
Jacob Mordukh (born 1895)
Ernest Nagel (1901–1985)
Jerzy Neyman (1894–1981)
Otton Nikodym (1889–1974)
Albert Novikoff
Octav Onicescu (1892–1983)
Kheimerool O. Ondar (born 1936) (Хеймероол Опанович Ондар)
Egon Sharpe Pearson (1895–1980)
Karl Pearson (1857–1936)
Charles Saunders Peirce (1839–1914)
Ivan V. Pesin (Иван Песин)
Jean-Paul Pier
Jules Henri Poincaré (1854–1912)
Siméon-Denis Poisson (1781–1840)
Karl R. Popper (1902–1994)
Pierre Prevost (1751–1839)
Johann Radon (1887–1956)
Hans Rademacher (1892–1969)
Frank Plumpton Ramsey (1903–1930)
Eugenio Regazzini (born 1946)
Hans Reichenbach (1891–1953)
Alfréd Rényi (1921–1970)
Hans Richter (born 1912)
Henry Lewis Rietz (1875–1943)
Leonard Christopher Gordon Rogers (born 1954)
Bertrand Russell(1872–1970)
Ernest Rutherford (1871–1937)
Stanisław Saks (1847–1942)
Leonard Jimmie Savage (1917–1971)
Ivo Schneider (born 1938)
Laurent Schwartz (1915–2002)
Irving Ezra Segal (1918–1998)
Stanford L. Segal (born 1937)

Eugene Seneta (born 1941)

Oscar Sheynin (born 1925) (Оскар Борисович Шейнин)

Albert Nikolaevich Shiryaev (born 1934) (Альберт Николаевич Ширяев)

Wacław Sierpiński (1882–1969)

Evgeny Slutsky (1880–1948) (Евгений Слуцкий)

Aleksei Bronislavovich Sossinsky (born 1937) (Алексей Брониславович Сосинский)

Charles M. Stein (1920–2016)

Margaret Stein

Hans-Georg Steiner (1928–2004)

Hugo Dyonizy Steinhaus (1887–1972)

Stephen Mack Stigler (born 1941)

Erling Sverdrup (1917–1994)

Angus E. Taylor (1911–1999)

Thorvald Nicolai Thiele (1838–1910)

Erhard Tornier (1894–1982)

Mark R. Tuttle

Stanisław Ulam (1909–1984)

Friedrich Maria Urban (1878–1964)

James Victor Uspensky (1883–1947)

David Van Dantzig (1900–1959)

Edward Burr Van Vleck (1863–1943)

John Venn (1834–1923)

Jean-André Ville (1910–1988)

Paul Vitányi (born 1944)

V. M. Volosov (В. М. Волосов)

Johannes von Kries (1853–1928)

Richard Martin Edler von Mises (1883–1953)

John von Neumann (1903–1957)

Raymond Edward Alan Christopher Paley (1907–1933)

Jan von Plato (born 1951)

Marian von Smoluchowski (1872–1917)

Vito Volterra (1860–1940)

Alexander Vucinich (1914–2002)

Abraham Wald (1902–1950)

Rolin Wavre (1896–1949)

Harald Ludvig Westergaard (1853–1936)

Peter Whittle (born 1927)

William Allen Whitworth (1840–1905)

Norbert Wiener (1894–1964)

David Williams (born 1938)

Anders Wiman (1865–1959)

William Henry Young (1863–1942)

Smilka Zdravkovska

Joseph David Zund (born 1939)

Antoni Zygmund (1900–1992)