

Introduction to the Correspondence

1 Borel and Hadamard

The two great mathematicians, Jacques Hadamard (1865–1963) and Emile Borel (1871–1956), who were central in French mathematical life for many years, offered mathematical guidance and inspiration to both Maurice Fréchet and Paul Lévy and are frequently referred to in the correspondence. The books of Maz'ja and Shaposhnikova (1998) and Guiraldencq (1999), the article 'Borel' in the *Encyclopedia Universalis* written by Maurice Fréchet as well as the presentation of the life and work of Emile Borel in volume I of Borel (1972) contain extensive descriptions. Moreover, the exchanges between Lévy and Fréchet began just after the end of the First World War in which Borel and Hadamard had been deeply involved—consult Mazliak and Tazzioli (2009).

Jacques Hadamard was born in Versailles in 1865. Heir to the great tradition of 19th century analysts, he threw himself into the investigation of analytic functions, focusing on theorems of analytic continuation and the systematic study of the radius of convergence. In particular he obtained what is now known as the Hadamard three-circle theorem, which affirms that the logarithm of the maximum modulus of an analytic function on a disk centered at O with radius r is a convex function. From his work, especially that on the Riemann ζ function, he drew important conclusions in analytic number theory. In 1896 he announced the remarkable prime number theorem (proven independently by de La Vallée-Poussin) showing that the number of primes less than a positive real number x grows asymptotically as $\frac{x}{\ln x}$.

Under the influence of his illustrious predecessor Poincaré, Hadamard was eager to keep contact with physics, and was a promoter of the systematic study of partial differential equations. He introduced numerous methods for their solution, in particular a general solution for the case of hyperbolic equations.

Hadamard was deeply involved in the scientific life of his time. He began a very active seminar at the Collège de France, which before the Second World War was the obligatory forum for any new results in mathematical analysis in France. In fact, Fréchet and Lévy indicate in their correspondence that this seminar was their natural meeting place. Hadamard was also politically engaged, like Borel as noted below,

but in a different spirit. A committed pacifist largely in sympathy with the goals of the Third International, Hadamard was an ardent partisan of rapprochement with the USSR. He was the guest of honor at the “Week of French science in the USSR” in 1934 (one of the first visits of a French scientist to the Soviet Union). A favorite of the regime, he returned to the Soviet Union several times.

The dramatic developments beginning in 1940 found him an old man of 75, crowned in glory but nonetheless under the threat of racist persecution because of his Jewish origin, despite the loss of two sons for France during the Great War. He felt constrained to accept an invitation to emigrate quickly to the USA in 1942. Returning to France at the Liberation, he lived another twenty years, keeping abreast of the mathematical life of the time. When Hadamard died in 1963 at the age of 98, Lévy was elected to his chair at the Academy of Science.

Emile Borel was born at Saint-Affrique in Aveyron in 1871. Throughout his entire life, he kept in close contact with the city of his birth. After very successful secondary-school studies, he entered the Ecole Normale Supérieure, where he immediately chose to pursue mathematics. He began by working on divergent series, for which he introduced what became known as the Borel sum. This quickly led him to a close examination of the idea of measures of sets (Hawkins 1970). In a series of remarkable works over a dozen years, he introduced the basic concepts of measure theory, which allowed Lebesgue to construct his integral and thereby revolutionize analysis. In related work, from 1905 Borel dedicated himself to the theory of probability. At the start of the 20th century he was its most important innovator, opening the way for its axiomatic formulation in terms of measure theory, which was completed in 1933 by Kolmogorov in his *Grundbegriffe der Wahrscheinlichkeitsrechnung* (*Basic Concepts of Probability Theory*) (КОЛМОГОРОВ 1933).

In addition to his research activities, Borel was very attentive to the role of the mathematician in society. He left numerous writings, both philosophical and popular, seeking to explain his ideas. He was constantly involved in political and social life, especially by way of his wife’s connections. She was the daughter of the mathematician Paul Appell, a literary figure and talented author who wrote under the pseudonym Camille Marbo (who won the Fémina prize in 1913). The First World War drew Borel to a task that soon became, for him, a calling. The Pyrrhic victory of 1918 decimated French youth, in particular students and alumni from the Ecole Normale Supérieure (Borel’s nephew among them) of which Borel felt himself the spiritual father as much as director. He resigned from this position, entered national politics and was elected deputy for Aveyron in 1924 in the so-called *Cartel des Gauches*, a union of moderate leftist parties. His friend Paul Painlevé, President of the Conseil in 1925, named him minister of the Navy. From that time, his life was divided between Paris and Saint-Affrique. He was elected Mayor of the latter several times. In 1921 he became a member of the Academy of Science, and was its President in 1934. He also drew on his political influence in the service of his discipline (Catellier and Mazliak 2012). The opening in 1928 of the Institut Henri Poincaré in Paris, which became one of the most important centers for mathematical research in France, was mainly a result of his efforts. Borel fought constantly, with considerable success, to ensure that numerous mathematicians and physicists visited the IHP, especially representatives of the Soviet school of probability, for whom the difficulty

of travel was ever greater. In fact it is mainly due to him that the French school remained in contact with the most active foreign practitioners of the discipline. During the Occupation, Borel was imprisoned several times because of his political views (Mazliak and Shafer 2011). He was thus a model of the engaged scientist. When he died in 1956 at age 85, he was succeeded at the Academy of Science by Maurice Fréchet, whose ideas were very close to his. There is further reference below to the central role played by Borel in the genesis of modern probability theory (Sect. 2).

2 Probability Theory in France at the Start of the 20th Century

Thanks to the work of Pierre-Simon Laplace and Denis-Siméon Poisson, probability theory was an important subject in French mathematics at the start of the 19th century. Thereafter it fell somewhat out of favor and only returned to a central position a century later. By that time new techniques, especially set theory, enabled the examination of random phenomena from a fresh perspective. This renewed interest was prompted by two exceptional mathematicians, **Henri Poincaré** (1854–1912) and especially **Emile Borel** (1871–1956).

Poincaré was educated at the Ecole Polytechnique in the second half of the 20th century. Marked by important discoveries in Brownian motion, radioactivity, and gas dynamics, this period was especially suited to his progressive realization of the links between the theory of probability and physics. The first paper in which Poincaré used the theory of probability was his famous *Sur le problème des trois corps et les équations de la mécanique*,¹ which appeared in 1890. Here Poincaré announced his recurrence theorem, showing that the initial conditions for which a dynamical system is not recurrent (in the sense that over time it never returns to a position close to the initial one) are a set with probability zero, probability being defined in terms of the set volume. This paved the way for the use of measure theory in probabilistic reasoning, later initiated by Borel and Lebesgue. Questions of the distribution of initial conditions of dynamical systems also play a central role in Poincaré's entry into the debate on the question of irreversibility in the article *Sur la théorie cinétique des gaz*.²

In 1896, the contents of Poincaré's famous course at the Faculté des Sciences de Paris appeared for the first time in print. A second edition was to follow just prior to the author's death in 1912. The course was construed one of mathematical physics, complementing the numerous volumes that Poincaré had written on various physical theories. (His title at the Faculté des Sciences de Paris was Professor of Calculus of Probability and Mathematical Physics.) This series of lessons treated fundamental results in the theory of probability in depth, especially questions pertinent to independence, the law of large numbers, and the error distribution. This

¹(1890) Acta Math. 13:1–270.

²(1894) Rev. Gén. Sci. Pures Appl. 5:513–521. For more on these matters, see Von Plato (1994, Chap. 3) and Mazliak (2014).

work came to replace Joseph Bertrand's classic treatise on probability published by Gauthier-Villars in 1888, and a standard reference work in France.

Emile Borel's personality has been discussed above (see Sect. 1). Borel's route to probability theory was rather different from that of Poincaré and is studied in detail in the paper of Durand and Mazliak (2011). Borel appears to have become interested in probability only when he realized that measure theory, which he had introduced during the final years of the 19th century and which served as basis for the creation of a new integral by Lebesgue, was very useful in studying certain problems connected to randomness. The article of Poincaré on the three-body problem mentioned above is a good example. In that case measure theory provides a structure to express the rarity of initial conditions for non-recurrent trajectories.

Borel's ideas on probability first appeared in 1905 in his *Remarques sur certaines questions de probabilités*.³ The goal of this paper, established at the outset, is to utilize the then-new measure theory and Lebesgue integral to justify the intuitive result of an example that Poincaré had given in his course of 1896, that choosing a rational point 'at random' from the interval $[0, 1]$ has probability zero.

In 1909, number theory furnished Borel with material for a notable article entitled *Les probabilités dénombrables et leurs applications arithmétiques*.⁴ This remarkable work took probability theory into a new era. Borel announced a first version of the strong law of large numbers, stipulating the almost sure convergence of the arithmetic mean of a series of independent random variables, thereby illustrating the interest of the notion of convergence in measure. Borel's most striking result illuminates the structure of the reals: with probability 1, the frequency of 1s in the binary representation of a number between 0 and 1 is equal to $\frac{1}{2}$. This appears to contradict the fact that one can choose "arbitrarily" the sequence of 0s and 1s that comprises the representation of a real.

Despite several spectacular papers, Borel's career in probability never attained the esteem that one might have expected. He was convinced that working on formalizing the theory was only of interest to lesser mathematicians, and hence never pushed as he might to axiomatize probability theory in the framework of measure theory. The program was instead completed, as is well known, by the Russian school (see Sect. 3). In the years following his 1905 article on countable probabilities, Borel mainly published popular articles on mathematics or papers on the philosophy of mathematics that explained his understanding of probabilistic models in applications. His celebrated book *Le Hasard (Chance)* was published in 1914, succeeded by numerous later editions.

The trauma of the Great War without doubt informed a certain disillusionment with mathematics on Borel's part. However it also seems that Borel did not have a very positive opinion of probability theory during these years. In her memoirs, the novelist Camille Marbo, Borel's wife, reported that Borel told her that he had enough of very abstract mathematics and so had *pantoufflait* (taken the comfortable

³(1905) Bull. Soc. Math. Fr. 33:123–128.

⁴(1909) Rend. Circ. Mat. Palermo 27:247–270.

route), working on probability and statistics. Bru et al. (2009) notes a surprising consequence of this thinking: some of Borel's most profound results in the theory of probability are to be found buried in his popular articles or reviews. (Bru et al. (2009) is especially interested in the pamphlet from the collection *Que-Sais-Je* which includes some interesting remarks on the theory of martingales.)

In this overview of the condition of probability theory in France before the appearance of Lévy and his generation, mention should be made of a third figure, **Louis Bachelier** (1870–1946). Bachelier was for a very long time ostracized by French mathematicians, and his prophetic work was only recognized during his lifetime by foreigners, especially members of the Russian school. In 1900 at the Sorbonne Bachelier defended a very original thesis entitled *Théorie de la spéculation* (*The Theory of Speculation*). This work envisaged for the first time a random phenomena (in this case the fluctuations of the stock market) as a mathematical process in continuous time. It in fact corresponded to a model of Brownian motion by Norbert Wiener that was introduced only some twenty years later. Bachelier also laid the groundwork for the study of Markov processes in continuous time. In 1931, in his great paper⁵ (see Sect. 3), Kolmogorov referenced Bachelier's notions on Markovian transitions as one origin of his own views.

It was Bachelier's misfortune to be too original—see Bru (2001) and Courtault and Kabanov (2002). This quality was in fact praised by Poincaré in his report on Bachelier's thesis which was laudatory, despite what has often been claimed. In addition, Bachelier 'came from nowhere,' being neither *normalien* nor *polytechnicien*. As a result he was treated with an indifference bordering on disdain. In applied mathematics today the theory of Brownian motion is widely used, and the role of Louis Bachelier as its forefather in the mathematics of finance is universally acknowledged.

Bachelier's most severe and least sympathetic critic was without doubt Lévy (Bru 1999). In fact, before he assumed his first position as professor in 1928, Lévy had read almost nothing by Bachelier, relying on the partial and incomplete reading of a single page of a 1913 publication by Bachelier, which he assumed to be mathematically incorrect. In this work, Bachelier had repeated, very briefly, certain parts of his thesis (which Lévy had never read), using several unfortunate shortcuts, in particular concerning the question of the speed of a Brownian particle. Lévy said that the "erreur de Bachelier (Bachelier's error)," which Bachelier in fact did not make in his thesis, stopped his reading. This also resulted in a distressing incident. In 1926 Bachelier, at the age of 56, was a candidate for the post left vacant by René Louis Baire in Dijon. Maurice Gevrey, who held the chair of Mechanics, asked Lévy for his opinion of the incriminating passage. Lévy responded that there was an error and the post went to Georges Cerf. Bachelier, who for 25 years had held only temporary positions, very poorly paid (or not paid at all), and had received a few grants, blamed Lévy for this new setback in his unfortunate career. He even circulated an accusatory letter to the academic community. This greatly affected Lévy, who be-

⁵Kolmogoroff, Andrei (1931) Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math. Ann. 104:415–458.

lieved that even without his contribution, Bachelier would never have obtained the Dijon post over a candidate as young and promising as Cerf.

It was only after reading Kolmogorov's article of 1931 that Lévy began to take seriously the content of Bachelier's work. Lévy's letters of September 1943 show how he was ready, when he began the first draft of certain chapters of Lévy (1948), to recognize the accomplishments and priorities of Bachelier on important questions concerning not only Brownian motion but diffusion more generally. Lévy and Bachelier were reconciled at the start of the Second World War. Beginning in 1948, two years after Bachelier's death, Lévy spoke of Brownian motion as 'la fonction déjà étudiée par Wiener et Bachelier (the function already studied by Wiener and Bachelier)'. In his autobiography (Lévy 1970), Lévy revisited his relationship with Bachelier, recognizing the unfairness of his initial judgment.

3 The Russian and Soviet Schools of Probability Before 1940

By the time Lévy had become deeply interested in research on the theory of probability, its center of gravity had already moved east. As of the 1920s, Soviet mathematicians used the results obtained by their immediate predecessors (especially Borel) and steered their efforts toward the investigation of probabilistic structures. In the two ensuing decades, this strategy resulted in fundamental advances by the Soviet school of probability. Because of these successes and many allusions to this school featuring in the correspondence between Lévy and Fréchet, the present section is included. The seemingly arbitrary cut-off date of 1940 is introduced only because after the Second World War, the letters between Lévy and Fréchet contain far fewer references to the Soviet school.

3.1 *Before the 1917 Revolution*

Without question, the founder of the Russian school of probability was Pafnuty Lvovich Chebyshev (1821–1894), a professor in Saint Petersburg (where he was recruited at the urging of Buniakovsky, after failing to obtain a position in Moscow). Because of Chebyshev, Russian research in mathematics, especially in probability, was centered in Saint Petersburg for fifty years. From the beginning of his mathematical career Chebyshev was interested in probability, in particular improving the fundamental limit theorems (the law of large numbers and the central limit theorem) and specifying the assumptions under which they are applicable. The preamble to his thesis, defended in Moscow in 1844, gives an idea of his goal:

Making a belief in all these conclusions via a simple and precise analysis accessible to the majority of students, is a large step to an elementary exposition of the theory of probability.

In all his research in probability, Chebyshev sought to justify the methods used in its applications. Юшкевич (1968) explains how Chebyshev's interest in ballistics

and in mechanics in the 1860s was at the origin of his discovery of the celebrated inequality that bears his name, and the method of moments. Kolmogorov, in a report from 1947 (КОЛМОГОРОВ 1947)—which should be read with a certain filter, in as much as it was written at the height of Soviet pan-Russian nationalism, for which Chebyshev became an emblem—wrote:

[T]he principal is that Chebychev searched continually to obtain precise estimates of the deviation with the [theoretical] limiting laws, that hold for a large but still *finite*⁶ number of trials [...]. In addition, Chebychev was the first to realize (and make full use of) the significance of a “random variable” and the “mathematical expectation” (average value) of a random variable. The sense of these concepts was already known and appeared as an extension of the expressions “event” and “probability,” but random variables and their mathematical expectation take on a more precise meaning via a practical and supple algorithm.

In the 1920s von Mises’ theory of collectives enabled the idea of an algorithmic construction of randomness. According to Kolmogorov this work had its origins in the Petersburg school of probability.

Chebyshev’s students Andrej Andrejevitch Markov (1856–1922) and Aleksandr Mikhailovitch Lyapunov (1857–1918) took up their professor’s torch. Chebyshev recruited Markov to the Imperial Academy of Science in 1886. Seneta (2001) justly observes that Markov, who lived until 1922, had a courageously resistant attitude to the most repressive measures of the tsarist regime. This, naturally, engendered a rather benevolent attitude toward him—or by association the school of Petrograd⁷—on the part of the Bolsheviks. Markov and Lyapunov continued to study the limit theorems, weakening the hypotheses on identical distributions in the central limit theorem. The first systematic development of the theory of characteristic functions, originally begun by Lagrange and Laplace, is from Lyapunov. Markov in turn, searching for a counterexample to the necessity of independence in the law of large numbers, formulated in 1907 a first model of the chains which now bear his name.⁸ There is no doubt that on the eve of the events of 1917, Petrograd was the center of applied mathematics in Russia.

In Moscow, the other capital of the empire, the school of mathematics took on increasing importance and began to rival Saint Petersburg at the start of the 20th century. The notable members of the Moscow group were Dmitri Fedorovitch Egorov (1869–1931) and his student Nikolai Nikolaievitch Luzin (1883–1950). The latter originated a very powerful formalist movement, studying the theory of abstract function spaces along the lines of the work of the French school of Borel, Baire, Lebesgue and Fréchet. It is to this period that the seminal works applying measure

⁶Our emphasis.

⁷Petrograd is the name taken by Saint Petersburg from the beginning of the First World War until 1924 when it became Leningrad.

⁸Seneta (2001) affirms that this occurred in reaction to an article by Nekrassov, explaining that a proof of the existence of free will emerges from the everyday observation of averaging phenomenon in society (for example, the average number of births remains almost the same from one year to the next) because independence of actions is necessary to apply the law of large numbers. Markov, extremely anti-religious, was eager to disprove this necessity.

theory to the theory of functions belong. This very formal research was not without connections to profound questions in philosophy and religion.⁹

Luzin's mathematical education was influenced by his voyage to Göttingen, home of the formalist school headed by Hilbert. Returning to Moscow, Luzin defended his thesis in 1914. Three years later, with the help of Egorov, he founded a research group on fundamental mathematics (affectionately nicknamed "Luzitania"¹⁰), in which his students participated—including Kolmogorov, after 1922.

There are two other important figures to note in this briefly sketched mathematical landscape. One, Serguei Natanevitch Bernstein (1880–1968), played a major role in the development of modern probability. In 1898 he began a thesis at the Sorbonne on elliptic equations, which he defended in 1904. This put him at the center of activity from which emerged the modern understanding of probability. Bernstein was also an essential point of contact between the French and Russian mathematical communities. In 1911 he began to work on probability in Kharkov, where he had taught since 1907. In 1917 he made an attempt at axiomatizing its foundations (Бернштейн 1917).

Vucinich (1999) notes Bernstein's unyielding attitude after 1917 to the ideological demands of the new revolutionary regime. Bernstein's international reputation, established well before the Soviet era, permitted him to weather the blackest years of the period without great difficulty. Retaliation against him principally took the form of partial concealment, i.e. keeping him obscured behind Khinchin and Kolmogorov (see below).

Finally, as noted by Юшкевич (1968), a very significant figure who should not be overlooked is Evgeni Evgenievitch Slutsky (1880–1948). Slutsky finished his studies in 1911 in Kiev, but not without difficulty, given various political problems. In the years preceding the 1917 revolution he published several articles on mathematical methods for statistics, including a monograph (Слутский 1912), a very modern exposition. In the 1920s he launched himself into pioneering research in stochastic calculus. These works were recognized as of primary importance by Kolmogorov and Khinchin. Slutsky's research remained little known outside the USSR although in the 1930s the great traveler Fréchet, among the most passionate to maintain contact with the USSR, drew the attention of French mathematicians to several of Slutsky's articles.¹¹

⁹Luzin and Egorov had close ties to orthodox theological circles, notably to the extraordinary Pavel Alexandrovitch Florensky (1882–1937), Luzin's close friend, who managed both studies in mathematics at the University of Moscow and religion at the Academy of Theology. In 1912, at the Academy, he defended a thesis searching for, à la V. Soloviev, a spiritual interpretation of science. Luzin also spent some time studying theology. The correspondence between Luzin and Florensky is the object of an interesting paper (Демидов et al. 1989). On Florensky, see the beautiful book Betti (2010), and also Žust (2002) for biographical aspects.

¹⁰Wordplay on the name Luzin, probably an allusion to the Lusitania, sunk in 1916, which was a pretext for the USA to enter the First World War.

¹¹Detailed information about Slutsky is available in Locker (2001), as well as in the 1948 eulogy to him written by Kolmogorov (Колмогоров А.Н. (1948) Е.Е. Слутский. Успехи Математической Наук III(4):142–151).

3.2 *Under the New Regime*

Revolutionary torment, as well as the new, post-1917 ideology, caused Russian mathematicians of the old school to progressively modify their discipline and their research strategies. This reorientation favored the emergence of a new generation naturally more in synch with the ideology of the existing regime. As noted in Vucinich (1999), the 1920s were marked by an attempt to erect a structure able to encompass all mathematics. It is easy to imagine that the very abstract ideas of the Moscow school of Egorov and Luzin were not favorably viewed by the new regime, whereas the new ideologues looked with satisfaction on the growing influence of research groups closer to inspirational Marxist ideas that demanded (Graham 1993) the unification of theory and practice.¹² It is interesting to read in this regard Luzin's own description of the situation of Soviet mathematicians at the end of the 1920s, when he had to defend himself with the "energy of despair" against the raging storm aimed at him in July 1936.¹³ In the notes of the boisterous meetings of the Commission of the Academy of Science which miraculously have been preserved, Luzin cries out:

Can I ignore the opinion of comrade Kolman, who occupies a high position and accuses me of theoretical abstraction and damage arising from my work? No, certainly not. My goal was not to damage my country. And if those who have a clear view of these matters find that a theoretical activity is damaging, well I preferred to do no more damage. But out of me came theorems. What could I do? Retreat into myself? I decided to publish here all the applied research that could help my country. And the parts of my work that have been called damaging. . . , to publish them abroad. Because I had to take account of the situation. Judge for yourselves if continuing to stubbornly publish theoretical work—that not only has no present application, but for which no future application is anticipated—wouldn't have presented an obstacle to our scientific publishing enterprise.¹⁴

Remarkably, it was three mathematicians (Колмогоров 1925; Хинчин 1926 and Бернштейн 1927) whose scientific work was principally in probability who formulated a politically correct and rather anti-idealistic view of the fundamentals and elements of mathematics. Their interpretation rose to the level of dogma during the entire Soviet period, with a practical view of probability theory serving as a prime example. Since the founders of the Russian school had developed the theory of probability in a rather practical context, they were easily able to avoid any accusation of idealism. The affirmation that the discipline was at the interface with the "real" world drew on the authority of Chebyshev, Markov and Lyapunov. As a result of this specifically applied character the theory of probability enjoyed an excep-

¹²See, for instance, the illuminating resume of the XVIIIth Congress of the Party published as an introduction to the 1939 volume of the *Известия Академии Наук СССР*.

¹³A violent campaign orchestrated by various political bodies accused Luzin of submission to the West, on account of his numerous publications in western journals. See Dugac (2003) and Демидов and Левшин (1999).

¹⁴See Демидов and Левшин (1999).

tional level of approval during the Soviet period¹⁵ (in Collective 1948, 51 pages are devoted to this theme). Vavilov's 1946 report (ВАВИЛОВ 1946) considered which fundamental scientific problems the Academy of Science should devote itself to in the next five-year plan. Probability is the first mathematical discipline mentioned, specifically because of its applied aspects. It was grouped with natural sciences and social sciences such as economics.

As in any period of fundamental change, this one was obviously marked by contradictory tendencies. It would be foolhardy to posit an overly happy outlook for the mathematics of randomness. Vucinich (1999) notes that randomness by its nature did not fit a dialectical materialistic interpretation very well, and therefore remained rather suspect in the eyes of those who from the 1920s on took charge of formulating the contours of the new ideology. Vucinich cites, for instance, a contribution to the symposium dedicated to the relation between dialectical materialism and mathematics, organized in 1931 under the auspices of the Communist Academy. This contribution regrets the interest taken by the Leningrad school in the fluctuations of random variables, which are found closer to the anarchic behavior of a capitalist economy than to the planned Soviet economy.

Contacts between Paris and Moscow gradually resumed in the aftermath of the Great War. Leading researchers like Borel and Fréchet sought to maintain relations with their Russian colleagues. The archives of the Institut Henri Poincaré point to the great efforts made by the Institute in this regard, particularly to the tenacity of Borel, who had to deal with an expanding Soviet bureaucracy. Between 1925 and 1939, Bernstein was frequently invited, but could only come to France once, in 1929. Again and again, from 1930 to 1935, Bernstein postponed his arrival, month to month, explaining that his government would not authorize his trip.¹⁶ In 1935 Fréchet mentioned¹⁷ that he had felt the necessity of renewing contacts with Russia as of his installation in Strasbourg in 1919, when relations between the two countries were at a very low point. He managed as best he could to establish an exchange of publications, and mentions in his Kiev speech the personal relationship that he was able to build with Nikolai Mitrofanovitch Krylov. In 1934, Hadamard was one of the first to visit the USSR, followed soon afterwards by Fréchet in 1935, as part of a long journey across eastern Europe.

3.3 *Several Important Steps*

The period from 1925 to 1940 witnessed effervescent mathematical creativity in the USSR. The foundations of the modern theory of probability emerged in the papers

¹⁵Which, most ironically, is also the reason for its lack of esteem by the French “structuralist” school (see Maasshal 1999).

¹⁶The archives of the Institut Henri Poincaré also contain an enormous dossier on the physicist Frenkel, who postponed his visit from 1934 to 1937.

¹⁷Draft of a speech in Kiev, archives of the Paris Academy of Science.

of the three leading figures (Kolmogorov, Khinchin and Bernstein as well as Slutsky, as mentioned) and their impressive list of students.¹⁸

In 1925, in their only joint paper, Khinchin and Kolmogorov obtained the classical criterion for the convergence of a series of independent random variables with mean equal to zero: if $\sum_{n \geq 1} E(X_n^2) < +\infty$, $\sum_n X_n$ converges a.s. To prove this, they made use of the celebrated inequality for martingales discovered by Kolmogorov

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon\right) \leq \frac{E(S_n^2)}{\varepsilon^2}$$

where $S_n = \sum_{k=1}^n X_k$. During the years that followed, a cascade of results on the law of large numbers flowed from this finding, notably the derivation of a general version of the strong law in 1930 by Kolmogorov: if the X_n have zero mean and if $\sum_{n=1}^{\infty} \frac{E(X_n^2)}{n^2} < +\infty$ then $\sigma_n = \frac{S_n}{n} \rightarrow 0$, a.s.

In 1924 Khinchin had obtained a first version of the law of iterated logarithms (for Bernoulli variables) that Kolmogorov was able to generalize in 1929 by showing that under reasonable assumptions $\lim_{n \rightarrow \infty} \frac{S_n}{\sqrt{2B_n \ln \ln B_n}} = 1$ where $B_n = \sum_{k=1}^n E(X_k^2)$. Here again, beyond the result itself, it is notable that the proof rests on a technique (large deviations) that itself found considerable further use.

At the beginning of the 1930s, the Soviet school of probability turned its attention to the study of stochastic processes in continuous time. Between 1930 and 1931, Kolmogorov was abroad, mainly at Göttingen in Germany, then in France, where he met Lévy and Fréchet. He stayed for a month in the latter's house in Sanary-sur-Mer in the Var, discussing Markov processes; then he spent several days in Paris where Lévy invited him to his home (Shiryayev 1989). In 1931 Kolmogorov's remarkable memoir¹⁹ on Markov processes appeared. Kolmogorov gave a systematic treatment of the functional equations satisfied by the semi-group of the transitions. This pioneering opus (directly descended from Bachelier's work, as acknowledged by the Soviet school—see Sect. 2) opened the way for the use of analytic techniques for the definition of stochastic processes. Later in the 1930s, Bernstein, in particular, launched a systematic analysis of diffusion processes. Such studies underwent further very rapid development in the 1940s (due especially to Wolfgang Doeblin, Kiyoshi Itô, William Feller, Robert Fortet, John Doob and, naturally, Lévy—see Sects. 7 and 8 below, and also Mazliak 2007, 2009).

In 1933 Kolmogorov published his celebrated *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Kolmogorov 1933), in which he systematically developed the theoretical foundations of probability theory in the framework of measure theory. As a result his voyage to the West, Kolmogorov undoubtedly became convinced that Borel's idea of putting measure theory at the heart of the mathematics of probability needed adjustment in order to accommodate the theory of processes in con-

¹⁸On Kolmogorov, the reader is invited to consult (Chaumont et al. 2007).

¹⁹Kolmogoroff, Andrei (1931) Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.* 104:415–458.

tinuous time. In addition, the impact on Kolmogorov of the school of Göttingen, where the spirit of Hilbert and his *Grundlagen der Geometrie* (*Foundations of Geometry*) reigned, inspired the Russian to write a small booklet proposing the rigorous foundations of two essential notions: conditional probabilities and the existence of probability distributions on infinite products of measurable spaces (the famous existence theorem of Kolmogorov under the assumption of compatibility of finite dimensional distributions—see Dellacherie and Meyer (1975)). In his autobiography Lévy (1970) claims (such a bitter remark should be considered with care since Lévy was often obsessed by questions of priority) that he could have written the monograph himself, had he better appreciated its importance. Kolmogorov's approach, albeit purely axiomatic, purports to somehow be constructive (a concession, in more or less good grace, to the mathematical orientation of the Soviets). Fréchet, a believer of the objective character of probability (see Fortet (1948)), was no doubt intrigued by the delicate balance that Kolmogorov struck between a frequentist theory based on finite sets representing the results obtained by repeating a random experiment a finite number of times—the only possible interpretation in connection with real events—and the aspects requiring infinity that are a pure mathematical abstraction.²⁰ In fact Fréchet wrote to Kolmogorov on this subject and the latter confirmed to him that he conceived the formal definition of probability in terms of “very large but finite sets,” since infinite sets are only a mathematical idealization. Kolmogorov confirmed that for him “formal axiomatics should be accompanied by an analysis of its significance for reality.”²¹

In the 1930s, Khinchin and Kolmogorov regularly notified Lévy and Fréchet of their work.²² In his book, Lévy (1970) mentions the significant correspondence that he had with Khinchin, notably on the arithmetic of probability distributions. These epistolary discussions culminated at the end of the 1930s with the Lévy-Khinchin formula describing the characteristic function of an indefinitely divisible distribution. It was found by Lévy and then by Khinchin in 1938, using a different method (see Sect. 5). Relations between Lévy and Khinchin deteriorated as a result of an article by Khinchin published at the end of the Stalinist era in a collection dedicated to philosophical problems of modern physics. In it, Khinchin castigated Western ide-

²⁰See Chap. 7 of Von Plato (1994) on this question.

²¹Letter, August 3, 1939 (archives of the Paris Academy of Science). The very late date of this letter, from a moment when contacts between Soviet and western scientists had almost disappeared, is another sign of the benevolent attitude of the regime toward Kolmogorov.

²²For example, during the winter of 1933–1934, Kolmogorov wrote to Fréchet (January 29, 1934, archives of the Paris Academy of Science) that he had obtained the criterion for a process in continuous time in the form $E(|f(x + \Delta) - f(x)|^K) = O(|\Delta|^{1+\varepsilon})$, which he applied to Brownian motion. Oddly, the famous criterion that bears his name was never published by Kolmogorov himself, but appears for the first time (properly attributed to him) in an article by Slutsky in 1937. Incidentally, the letter of Kolmogorov just mentioned is written in an uncharacteristically anxious style. He wrote for instance: “It requires considerable energy to surmount formal difficulties” suggesting that he was asking Borel for some official help so that he could obtain the permission to travel). Kolmogorov apparently encountered unexpected bureaucratic difficulties regarding his trip to France, a sign of the restrictions that the Soviet authorities were imposing on international travel.

alism. Lévy published a virulently critical response in *Bulletin des Sciences Mathématiques*. In his autobiography (Lévy 1970, p. 109), Lévy alluded to the incident and regretted his response, because it did not take account of the violent Stalinist campaign directed against so-called bourgeois science at the end of the 1940s, and thus may have put Khinchin at risk. Kolmogorov resumed contact with Fréchet at the start of de-Stalinization, and in 1958 was invited for a semester to the University of Paris (see Letter 78).

4 Several Common Topics of Interest to Lévy and Fréchet

In the correspondence included here, Lévy is mainly concerned with his own research. However among the letters are many exchanges on topics common to Lévy and Fréchet. Both of them were analysts and probabilists. Fréchet defended his thesis in 1906, five years before Lévy, then immediately joined the school of functional calculus led by his professor and mentor Hadamard. Starting in 1910, he enriched that branch of mathematics with results on abstract sets in *Les ensembles abstraits et le calcul fonctionnel*.²³ In the same year *Leçons sur le calcul des variations* (*Lessons on the Calculus of Variations*) by Hadamard, assembled and edited by Fréchet, was published. Fréchet had, for example, already clarified the idea of compactness, but his treatment was limited to sequential compactness, a limitation that later incurred the wrath of Bourbaki. Fréchet thus opened the way for modern analysis and certain aspects of topology. In the latter, he contributed at the highest level and was considered the equal of Hausdorff by the Polish school, as is mentioned in several editions of the celebrated treatise on topology by Kuratowski.

In 1918 when the Lévy-Fréchet correspondence began, Lévy, like Fréchet, had mostly worked along lines set out by Hadamard. He attacked some interesting theoretical problems in functional calculus and the calculus of variations, and he applied the obtained results to classical questions of potential theory (see Sect. 6). Hadamard had asked Lévy to organize the papers of Gateaux, who died in the war.²⁴ Reading them led Lévy to a number of questions, and he felt himself ready to write an epoch making book about functional calculus. However it took until 1922 before Lévy published *Leçons d'analyse fonctionnelle* (Lévy 1922). In 1918, Fréchet already had the ideas of neighborhood, metric space, and in 1925, after having created the corresponding mathematical vocabulary and defined the appropriate structures, he gave the definition of vector spaces, affine spaces and affine topological spaces. Thus, in a few years time, Fréchet gained a considerable lead over Lévy in these emerging areas of analysis. Hadamard approved this approach in which attention focused, at first, on specific problems involving the nature of variables, and only afterwards on the relation between functionals and variables, and finally on hypotheses concerning

²³(1910) Rend. Circ. Mat. Palermo 30:1–26.

²⁴On Gateaux, see the complete study provided in Mazliak (2011).

sets and transformations between spaces. The French school of analysis, not without considerable grousing, would also follow this path in the aftermath of Fréchet's works.

The correspondence addresses topics in analysis and probability which interested both mathematicians, despite their often very different approaches. These topics appear in the letters in different ways, recurring or sporadically. Since we only have Lévy's letters, Fréchet's contributions seem a bit veiled, despite his considerable mathematical stature. Fully describing the extensive and encyclopedic nature of Fréchet's opus would have required much more than a commentary on his correspondence with Lévy.

To finish this survey of the relevant topics, we present a few details to help introduce the letters.

4.1 *Fréchet's Integral and Lévy's Averages*

In 1898, Borel gave an explicit definition of the measure of a set and measurable sets in his *Leçons sur la théorie des fonctions*. Lebesgue published *Leçons sur la théorie de l'intégration* in 1904. Hawkins' book *Lebesgue's Theory of Integration* (Hawkins 1970) and Alain Michel's book *Naissance de la théorie moderne de l'intégration*²⁵ trace the origin and history of modern integration theory. They also describe Fréchet's integral, and give an overall viewpoint on these developments. *Eléments d'histoire des mathématiques* by Bourbaki also includes the Fréchet integral in its coverage of the history of integration. In the works of Hadamard, Riesz, Fisher, Fréchet and Radon the concept of the integral, due to its linear and continuity properties, gained increasing connections with functional analysis. In 1907 Riesz found the representation of linear functionals on the space of continuous functions (on a given interval) by using Stieltjes integrals. This improved the 1903 result of Hadamard that defined these functionals as limits of integrals. Later, Riesz and Fréchet independently established the theorem representing linear functionals on L^2 by Lebesgue-Stieltjes integrals. In 1915, in several publications, Fréchet made a decisive step by incorporating certain aspects of Radon's work in his theory of abstract spaces. Fréchet exploited partitions on a function space $\mathcal{F}$ in order to define, using superior and inferior Darboux sums, the integral of a functional which he denoted $\int_{\mathcal{F}} G(e) df(e)$. This is a kind of Stieltjes integral for which the function f is an additive function of sets and thus a primitive version of an abstract measure. At the end of the First World War, the American works of the British mathematician Daniell appeared, similar to certain of Fréchet's ideas. Daniell constructed an integral on a set of arbitrary elements that can be points in a space of denumerably infinite dimension but found it impossible to give an example in such a space.²⁶

²⁵(1992) Vrin, Paris.

²⁶On Daniell, see Aldrich (2007).

At almost the same time, Lévy came up with a notion of the average on a unit sphere²⁷ in L^2 . This idea proved to be an important step on the way to integration in infinite dimension. These Lévy averages were first introduced, following Gateaux's suggestions and methodology, in a note.²⁸ In it, Lévy introduced a functional Φ defined on the space of continuous functions on $[0, 1]$. He considered the subspace of functions that he calls simple of order n , meaning constant in each of the n equal intervals of a partition of $[0, 1]$. Restricted to this subspace, Φ appears as a function φ of the n values $z_1, \dots, z_n$ on these intervals. Its average μ_n is given by an integral relation

$$\mu_n \cdot \int_{-\infty}^{+\infty} a_{n-1}(\lambda) d\lambda = \int_{-\infty}^{+\infty} \lambda a_{n-1}(\lambda) d\lambda$$

where $a_{n-1}(\lambda)$ is the $n - 1$ dimensional volume of the intersection of the unit sphere and the hyperplane $z_n = \lambda$. Lévy's idea, which had already occurred to Gateaux, is to let n go to ∞ . In a situation when μ_n goes to a value μ , the expression Φ has the value μ almost everywhere. This results because the measure of the domain in which φ is not between $\mu - \varepsilon$ and $\mu + \varepsilon$ goes to 0 as n goes to ∞ . In the case of averages over spheres, Lévy also notes the necessity of making use of the norm L^2 .

Daniell's integral and Gateaux-Lévy's averages were the source of Wiener's 1923 construction of Brownian motion (see Sect. 7). Wiener's *Differential Space* is the first explicit example of the application of Daniell's integral to a (denumerably) infinite dimensional space. However, in reviewing theories of integration in 1930, Kolmogorov emphasized the importance of Fréchet's integral, which he retained as a source for his *Grundbegriffe* (Kolmogorov 1933).

4.2 Differentials

In extending of the ideas of derivative and differential to infinite dimensional spaces, Hadamard's school of functional calculus focused on certain questions from the calculus of variations (see Sect. 6). This resulted in a fuzzy cohabitation of various definitions corresponding to different situations. Fréchet's general work on analysis helped greatly to clarify these matters.

Before the First World War, Fréchet introduced the differential of a functional U as the continuous linear functional δU satisfying $U(x + \delta x) - U(x) - \delta U(\delta x) = o(r)$ where r is the distance between the functions x and $x + \delta x$. Simultaneously, Gateaux introduced his own "variation" (aka differential). For a functional U , this is the limit $\delta U(\delta x)$ of the ratio $\frac{U(x+\lambda\delta x)-U(x)}{\lambda}$ when λ tends to 0. Gateaux's definition does not require any metric structure on the space of functions x and is not always linear in δx . Lévy, inspired by Gateaux, made considerable use of the Gateaux

²⁷On this important topic, see the many details contained in the paper of Mazliak (2011).

²⁸Lévy, Paul (1919) Sur la notion de moyenne dans le domaine fonctionnel. CRAS (August 25, 1919).

differential in his own work. Starting in 1922 (see Lévy 1922), Lévy compared Gateaux's and Fréchet's differentials and exposed their respective use in the framework of functional calculus. It is clear that in a normed vector space differentiability in Fréchet's sense implies differentiability in Gateaux's sense. Thus the ensuing debate between Lévy and Fréchet focused mainly on the role of linear and metric structures. As a result, these constructs—today part of the foundation of analysis—imposed themselves on Lévy as the natural framework for the calculus of variations, despite his constant concern about economizing structures. In order to reconcile linearity and simplicity, Lévy proposed that U be differentiable if and only if it admits a variation in Gateaux's sense that is also linear. In a 1937 memoir, Fréchet criticized this idea of differentiability in Lévy's sense because it did not necessarily satisfy the composition formula for partial derivatives $\frac{d}{dz}U(\phi_1(z), \dots, \phi_n(z))_{z=z_0} = \sum_{i=1}^n \partial_i U(\phi_1(z_0), \dots, \phi_n(z_0))(\frac{d\phi_i}{dz})_{z_0}$ where $\phi_1, \dots, \phi_n$ are functions of the real variable z .

In the same memoir Fréchet defined his own differential, those of Gateaux and Lévy, and gave a fourth possibility that follows from work by Volterra and Hadamard. The latter had occasionally been used by Lévy to calculate the variation of a functional U on a family of curves parameterized by a real number λ (see Sect. 6). If such a U is defined on set of continuous functions $t \mapsto x(t, \lambda)$ parametrized by λ , U becomes a function of the parameter λ . If one requires that the variation of x at the point t be a first-order approximation to the difference $x(t, \lambda + \delta\lambda) - x(t, \lambda)$, it is natural to make the definition $\delta U = \frac{\partial U}{\partial \lambda}[x(t, \lambda)]\delta\lambda$ with $\delta x = \frac{\partial x}{\partial \lambda}\delta\lambda$. If U is in fact defined on a functional space E , since the preceding formula makes no supposition about the topology of E or the variation of U for general $\delta x \in E$, one must make this formula compatible with the different types of variation of U on E , so that it has the required form when restricted to the parameterized family. Therefore Fréchet, requiring that δU be linear in E , proposed that the formula for the composition of partial derivatives be part of the definition of δU on E , as had been suggested by Hadamard. In the variant known as the Hadamard-Fréchet differential, one demands that $\frac{\partial x}{\partial \lambda}$ belong to the functional space on which U is defined. In 1937, Fréchet described this differential as “the generalized differential of Monsieur Hadamard” and contended that the extension of Hadamard's definition to functionals had up to then neither been studied nor even proposed. Ten years after Lévy's last discussion on this topic in the second edition of his book (Lévy 1937), Manuel Balanzat published a note presented by Fréchet to the Academy of Science²⁹ that extended the Hadamard-Fréchet differential to topological vector spaces. If X and Y are two such spaces, then $f : X \rightarrow Y$ is differentiable at x_0 if there exists a continuous linear function D (the differential of f at x_0) with the following property: for any real function g satisfying $g(\lambda_0) = x_0$ that is differentiable at λ_0 , $G = f \circ g$ is differentiable at λ_0 and $G'(\lambda_0) = D[g'(\lambda_0)]$. Thus the differential made no use of the metric structure. In that note, Balanzat's principal source is Fréchet's memoir of

²⁹Balanat, Manuel (1960) La différentielle d'Hadamard-Fréchet dans les espaces vectoriels topologiques. CRAS (November 28, 1960).

1937, augmented by his work in topology. The note also contains a short discussion of the different ways to axiomatize differential calculus in topological spaces and in certain Fréchet spaces that are not topological in the Bourbaki sense.

4.3 Probability

After the First World War Lévy and Fréchet both became interested in the theory of probability at almost the same time, for pedagogical reasons. They soon realized that there was much to be done (see Sect. 2). Lévy discusses this moment in his autobiography (Lévy 1970), explaining that essentially all required methods and concepts had either to be recovered or created.

Fréchet never overlooked applied research and applications, and his work was quickly noticed by Borel. As opposed to almost all probabilists of that period, Fréchet did not come to probability theory via theoretical physics, set theory, or actuarial studies, but via applications in statistics, economics, and social science.³⁰ In this regard, he followed a German tradition represented in Strasbourg by G.F. Knapp (1842–1926), professor of Statistics and Economics from 1874 to 1918. This focus on applications marked Fréchet’s entire scientific career. His course in the theory of probability at the IECS (*Institut d’Enseignement Commercial de Strasbourg–Strasbourg Institute for Business Studies*), written with Halbwachs (whose contributions are readily identifiable) entitled *Le calcul des probabilités à la portée de tous* was published in 1924. The two authors presented the theory of probability via applications to insurance and statistics, without using high-level mathematical analysis. They wrote in the introduction “We have not attempted to follow the path of our forerunners either in detail or in outline.” One can easily imagine that Borel, involved in politics (he was deputy in 1924 and even briefly became minister in 1925), appreciated this link between social science, statistics, and the theory of probability.³¹ Some of the themes that Fréchet would follow during the next forty years are presented in this book. All his life, Fréchet retained a taste for teaching mathematics applied to social science to students with little scientific background. He returned to this subject in innumerable publications that are collected in his 1955 book *Les mathématiques et le concret*.³²

Lévy’s approach to mathematics was very different. As a result, beginning with his first notes in CRAS (presented by Hadamard) on characteristic functions and the central limit theorem, Lévy found himself in the center of a dispute between the school of the theory of functions and that of functional calculus, represented by Borel and Hadamard, respectively. Laurent Schwartz notes that Hadamard believed Lévy had abandoned mathematics for physics by devoting himself almost

³⁰For details on this important moment in Fréchet’s life, see Havlova et al. (2005).

³¹For Borel’s interest in probability, see Durand and Mazliak (2011).

³²Presses Universitaires de France, Paris (1955).

exclusively to probability, and never pardoned him. Lévy's methods also put him in conflict with Borel (see Lévy 1925). Borel found Lévy's use of rigorous mathematical techniques of little interest, and his analytical reasoning unnecessarily complicated.³³ And yet Lévy's book (Lévy 1925) has a whole chapter on the theory of gases, which must have pleased Borel. In any case, Lévy was very proud of the appreciation that this chapter garnered from physicists. Levi-Civita wrote to Lévy that this was "obviously" the chapter that most interested him. In 1959, Cramer mentioned that Lévy's 1925 book was the first modern treatise on probability

among those books that mark an advance and show the direction in which the science will develop in the future. A central place is given in the book to the general theory of probability distributions, which was novel and audacious, something that the younger generation of today can barely comprehend. . . .³⁴

Although physical considerations were in later years sometimes presented by Lévy as a guide to his thinking or to illustrate a property by analogy (see Sect. 7), Lévy did in fact not go on to study connections between probability and physics. Hadamard's reproach is probably for other reasons; perhaps Hadamard hoped the audacity remarked on by Cramer would be applied by Lévy in more noble areas than probability. But it was more likely the supposed absence of rigor in the language used by Lévy to describe the theory of probability that occasioned the reproach. Lévy had to explain this at length in his own preface to Lévy (1937) (which followed Borel's preface). One sees yet again how Fréchet benefited from the support of his two mentors, while Lévy, not succeeding in capturing the interest of Borel, lost a bit of Hadamard's as well.

Among the first topics in probability discussed in the letters, the law of errors took on considerable importance. The two mathematicians worked on and compared their approaches to this topic from the start of the 1920s, and found themselves at the center of a theoretical debate that interested and impassioned an entire international community. The revival of interest in the question of how to justify practical methods of taking into account errors of measurement had various origins, from the experiments done in physics to ideas about its role in the artillery fire of the war that had just ended. In any case, the scientific dispute between Fréchet and Lévy was cordial, but very animated. It figures extensively in an amazing series of letters at the end of 1925, and reemerges during their unfortunate attempt at election to the Academy of Science in 1951.

Lévy embraced the questions of Gauss' law and the central limit theorem from his first lectures in probability theory at the Ecole Polytechnique in 1920. His lectures included various aspects of the convergence to a Gaussian distribution of which he gave several demonstrations and examples, tailored to his audience.³⁵ In 1921, at a conference of a philosophy meeting, he addressed this question outside the closed

³³See Bru (1999).

³⁴Cramer, Harald (1958) *Eloge de M. Paul Lévy*. In: *Le calcul des probabilités et ses applications* (July 15–20, 1958). Editions du CNRS, Paris. pp. 13–15.

³⁵See Barbut and Mazliak (2008).

circle of lecture halls for the first time. In 1924 and 1925, several publications of Lévy and Fréchet both addressed this matter and cited each other. One may follow most of the debate about the law of errors between Lévy and Fréchet in the letters. One of them was published with Lévy's approval (after proofreading the manuscript) in a 1925 article by Fréchet in the Moscow Mathematical Society (see Letter 14) and a passage from a letter by Fréchet was copied by Lévy at the end of a 1929 article (see Letter 24).

A dialogue arose between Fréchet and the Soviet astronomer M.F. Subbotin of Novocherkassk, who was interested in a justification of the corrections made to measurements in the observation of stars. In 1923 he proposed two axioms that any possible error distribution must satisfy.³⁶ These axioms were inspired by an attempt by Schiaparelli dating from the end of the 19th century, which Subbotin found in the book of Czuber.³⁷ They were formulated as follows.

Axiom I states that the probability (in modern terms, the probability density) of an error depends only on its size via a certain *error function* $\varphi(\varepsilon)$ associated to the particular measurement. In a series of measurements of a quantity, the probability that the error lies between ε and $\varepsilon + d\varepsilon$ is given by $\varphi(\varepsilon) d\varepsilon$. Subbotin also included a technical requirement that “in general” (sic) φ have a continuous first derivative. This caused Fréchet to object.

Axiom II is a homogeneity principle, which dates to Schiaparelli. It requires that the most probable value of the quantity x , of which one has observed measurements $l_1, l_2, \dots, l_n$, that is the value of y which maximizes the product $\varphi(y - l_1) \cdots \varphi(y - l_n)$, be independent of the unit of measure. It was on the interpretation of this axiom that the debate took place.

Lévy's interpretation was, naturally, the most probabilistic. For him, this axiom signifies that if $y = F(l_1, \dots, l_n)$ is the value at which $\varphi(y - l_1) \cdots \varphi(y - l_n)$ attains its maximum, then for all $\alpha > 0$, $\alpha F(l_1, \dots, l_n)$ is the value at which $\varphi(y - \alpha l_1) \cdots \varphi(y - \alpha l_n)$ attains its maximum. There is, as Lévy remarks, a connection with his theory of stable probability distributions (see Sect. 5).

Subbotin's first article had the same interpretation of axiom II as Lévy. However, his study made an implicit use of the fact that φ has a continuous derivative everywhere in order to have a simple maximum condition, which led in turn to an error distribution of the form $\varphi_m(\varepsilon) = \frac{mh}{2\Gamma(m-1)} \cdot e^{-h^m|\varepsilon|^m}$ where $m > 0$ and the constant h is chosen to get a probability density. An article by Fréchet³⁸ then observed the use of this stronger hypothesis, but made an error (with a different interpretation of the axiom) in proposing a counterexample. Surprisingly, this seems to have embroiled Subbotin. In a new note,³⁹ he accepted Fréchet's conclusions and showed that the general form of a function satisfying the two axioms may be written as a convergent product $A \prod_{r \geq 1} [\varphi_{m_r}(u)]^{k_r}$, which can have a non-continuous derivative

³⁶(1923) Математический Сборник 31:296–301.

³⁷Czuber, Emanuel (1891) Theorie des Beobachtungsfehler. Teubner, Leipzig.

³⁸(1924) Математический Сборник 32:5–8.

³⁹(1925) CRAS 180:1716–1719.

at 0. Finally, in another note⁴⁰ Fréchet showed that the general form of admissible functions of errors obtained by Subbotin can be recovered directly as the solution of a problem of functional optimization.

On receiving Fréchet's article (see Letter 9), Lévy made his disagreement known, and their discussion of Axiom II began. Fréchet's understanding of it is at once tortuous and simple: he required that if the measurements $l_1, \dots, l_n$ lead to selecting x as the true value, the measurements $\alpha l_1, \dots, \alpha l_n$, taken with the error distribution φ_1 deduced from φ by multiplication by α (and therefore with probability density $\frac{1}{\alpha}\varphi(\frac{1}{\alpha}x)$) lead to selecting αx . Lévy pointed out to him that this fact holds for any distribution, so that Axiom II becomes vacuous. Subsequently, Fréchet accepted this argument and published an explanation (see Letter 14) in which he only partially gave way, but he did acknowledge publicly that his interpretation of axiom II could not lead to any useful conclusion.

In 1927, the discussion moved to another point. From Fréchet's two articles⁴¹ arose the possibility of replacing the hypothesis of the additivity of errors by consideration of the maximal error. Lévy then wrote a final article on the topic.⁴² He criticized Fréchet's hypothesis, justifying once more the for him indisputably additive character of errors and Gauss' law. By now, it had been generally accepted that according to the specific "concrete" situation, one could follow either Lévy or Fréchet's interpretations. This long debate had some unfortunate consequences for the two men. It gave rise to unpleasant remarks about Lévy's supposed errors and the propensity of Fréchet to philosophize rather than do mathematics, remarks that reappeared during their candidacy for the Academy of Science.

By contrast to Lévy, axiomatizing the foundations of the theory of probability and random processes was of immediate interest to Fréchet. He soon realized that the semi-formalized probabilistic notions had much in common with the concepts of general mathematical analysis. Laplace had observed that Bernoulli's theorem arose from a double approximation, the frequency being very likely to be very close to the probability. This very particular convergence corresponds completely to the convergence in measure in the theory of functions. This immediately suggested that random quantities are naturally governed by several modes of convergence, just as functions of a real variable. These modes are not equivalent, which implies that there are several possible versions of the law of large numbers. Clearly, Fréchet was struck by this. He made it the subject of his first course on probabilities at the Institut Henri Poincaré in 1928–1929, and in 1930 published a beautiful article that undoubtedly influenced Kolmogorov, who visited him during the summer of 1930 (see Sect. 3). This point gives an interesting hint about the personalities and thoughts of Fréchet and Kolmogorov. During the second half of the academic year 1928–1929, Fréchet

⁴⁰Fréchet, Maurice (1925) Sur la loi des erreurs d'observation. CRAS 181:204–205.

⁴¹Fréchet, Maurice (1928) Sur la loi de probabilités de l'écart maximum. Ann. Soc. Pol. Math. 6:93–122 and Fréchet, Maurice (1928) Sur l'hypothèse de l'additivité des erreurs partielles. Bull. Soc. Math. Fr. 52:203–216.

⁴²Lévy, Paul (1929) Sur quelques travaux relatifs à la théorie des erreurs. Bull. Soc. Math. Fr. 53:11–32.

lectured in Paris on the theory of random variables in a manner very close to his course at the IECS in Strasbourg. He defined

a random variable as a number $X(E)$ of which the value is fixed by the result E of a trial belonging to a certain category C of trials. . . . One sees that X is a functional with a numerical value $X(E)$, with variable E which belongs to a domain C . In the following, whenever we consider several random variables $X(E), Y(E), \dots$, at once, we will assume that they are all defined over the same category C of trials, so that the symbol C can be omitted.⁴³

Fréchet noted that by making the probability play the role of a linear measure, the types of convergence of numerical functions of real variables, namely convergence almost everywhere, convergence in measure, and convergence in the average of order r , could be extended to sequences of random variables. Therefore one may extend the usual properties of numerical functions of a real variable to random variables, notably the usual connections between the various modes of convergence, the possible metrization, the questions of compactness (in Fréchet's sense), etc. It was, incidentally, Fréchet who introduced various designations such as *convergence in probability* by analogy with the *convergence in measure* used by analysts. Even though C was clearly an abstract set, it would be an exaggeration to think that Fréchet presented his students at the IECS and the IHP with an explicit set-theoretic axiomatics of the theory of probability. In fact, he followed the 1843 treatise *Exposition of the Theory of Randomness* by Cournot very closely, interpreting it in his own way and using a more general analysis to develop it. However, for Fréchet as for Borel, probability defined on a category of trials is a physical (or psychological) reality, while the linear measure of Borel is a concept belonging to the realm of mathematics. Fréchet did not seem to find it useful to clearly affirm in his books or courses that both could be related to the general definition of an additive function on abstract sets. For him, the theory of probability and the theory of linear measure do not belong to the same world, but this does not prevent them from being linked by the same mathematical structures and finding from this connection a basis for the richness of their possible applications. Thus he put himself in a purely French tradition, reaching from Poincaré to Borel, that declined to see real mathematics in probability theory.⁴⁴ Kolmogorov belonged to another generation and a different culture. The frontiers between analysis and probability must have appeared strangely Byzantine to him. Nonetheless, he always considered Fréchet a mentor.

With a few reservations, Lévy would little by little adopt Fréchet's understanding of probability spaces, i.e. via categories of trials. In Lévy (1937) he adopted the language employed by his colleague almost totally, and referred to it explicitly, using Borel's *fields*, the ancestor of modern σ -algebras with a then somewhat fuzzy definition; he later improved this definition and called the category of trials a *closed field*. In the same publication, he also returned to his previous work, dating from 1924, addressing the question of the definition of a probability distribution on an

⁴³See Fréchet and Halbwachs (1924, Chap. V).

⁴⁴See Mazliak (2014).

abstract set. His notable idea was to distribute the positive mass over the elements of a successively refined partition. This allowed him to construct a probability measure on a set that has the cardinality of the continuum, thus implicitly using the axiom of choice. He connected this approach to the integrals of Fréchet and Daniell, but never succeeded in relating this theory of partitions to the Fréchet integral. Lévy always defended a subjective interpretation of probability, and never gave a complete and general model of probability in the now-standard Doob form $(\Omega, \mathcal{A}, P)$, the only exception being the special case of $[0, 1]$ equipped with Borel sets and the Lebesgue measure. Quite the contrary, Lévy's practice permits one to leave Ω and the σ -algebra out and only speak of events for which there is no doubt of having a well determined probability because of their definition or because their probability is fixed via asymptotic conditions dependent on a chain of simpler events of well defined probability (for example, the zero-one law). This did not stop Lévy from being very precise and complete on questions of measurability (see Sect. 7). Lévy especially objected to the space Ω and its technicalities for stochastic processes.⁴⁵ However this did not hinder him from considering questions as profound as the separability of processes or their modifications. For Lévy, a stochastic process X is a function that depends on chance at every instant, and the miracle of the choice of an element ω determining the trajectory $X(\cdot, \omega)$ does not correspond to any concrete situation, and certainly not to the philosophy of randomness that he wished to defend.

"In principle," wrote Lévy, "a stochastic process is a phenomenon for which chance intervenes at every instant t of its evolution. For Doob, a stochastic process is simply a random function $X(t)$ of a variable t which one can imagine to be time. Whatever the set E of possible realizations of $X(t)$, a symbolic value ω may be associated to each element of E . This formulation permits Doob to consider $X(t)$ as a well-defined function $x_t(\omega)$ of t and a random variable ω , the choice of which encodes all successive influences of chance. It is a notation that is often convenient, even though it gives, as if born in one instant, that which for me is essentially a perpetual future. It should be remarked that if the set E has only the cardinality of the continuum, one can represent ω as a scalar variable, or more precisely as a variable chosen between 0 and 1 with uniform probability. . . ."⁴⁶

5 Lévy's Stable Probability Distributions

One of Paul Lévy's principal contributions to the theory of probability and its applications is his theory of stable probability distributions. This theory appears often, implicitly or explicitly, in his correspondence with Maurice Fréchet (especially in Letters 9 and 21).

⁴⁵In the complementary notes to the second edition of Lévy (1937) Lévy mentions that Fréchet concurred.

⁴⁶Lévy (1937), note II in the second edition of 1954.

Lévy designated a family of probability distributions on $\mathbb{R}$ (or the corresponding random variables) as a *distribution type*⁴⁷ if that family is invariant under translation and changes of scale; in other words, if an affine transformation $X \mapsto aX + b$ ($a > 0$) of the random variable X gives a variable governed by the same family of laws (see Letter 9).

Classic examples of distribution types are normal (Gaussian) distributions $\mathcal{N}(m, \sigma)$, exponential distributions $P(X > x) = e^{-(x-x_0)/h}$, $x \geq x_0$ and the Pareto distributions $P(X > x) = (\frac{x_0+c}{x+c})^\alpha$, $\alpha > 0$, $x \geq x_0 > -c$.

A distribution type is called *stable* if for any two independent random variables X_1 and X_2 belonging to the distribution type and a_1 and a_2 two positive constants, there exists a positive real $a(a_1, a_2)$ such that the variable X defined by

$$aX = a_1X_1 + a_2X_2 \quad (1)$$

belongs to the same distribution type. More generally, if random variables $X_1, X_2, \dots, X_n$ belong to a stable distribution type, and $a_1, a_2, \dots, a_n$ are given positive reals, one can find $a(a_1, \dots, a_n)$ such that X defined by

$$a(a_1, \dots, a_n)X = \sum_{k=1}^n a_k X_k \quad (2)$$

with $a_i > 0 \forall i$ and $a(a_1, \dots, a_n) > 0$, belongs to the same distribution type.

Stability may of course be defined by operations other than the addition of independent random variables: for example stability with respect to the maximum (as for the Fréchet, Gumbel and Weibull distributions) or for positive variables by the product, as is the case for the log-normal distribution. Here we always assume stability with respect to addition which is the only case considered by Lévy.

The importance of stable distributions for the theory of probability and its applications derives mainly from the following: if a normalized sum of n independent random variables with the same distribution⁴⁸ has a limiting distribution as n goes to infinity, that limit is obviously stable under addition.

The best-known stable distribution is the normal distribution, and until at least 1919 it was the only one known, with the exception of the Cauchy distribution. This is the probability density $\frac{1}{\pi} \frac{a}{a^2 + (x-x_0)^2}$ ($a > 0$) named by Lévy. Poisson, and later Cauchy, demonstrated that it is stable.

Lévy began his research on stable distributions in 1919, but his definitive characterization of them did not appear until a joint note with A.Y. Khinchin soberly entitled *Sur les lois stables*.⁴⁹ Up to a translation, for such a distribution the distribution function F is specified by its characteristic function $\varphi(z) = \int_{-\infty}^{\infty} e^{izt} dF(t)$,

⁴⁷French: *type de loi*.

⁴⁸Or an average, or more generally a positive linear combination.

⁴⁹CRAS (February 3, 1936).

namely

$$\varphi(z) = \exp \left[- \left(a + i \frac{z}{|z|} b \right) |z|^\alpha \right] \quad (3)$$

with $0 < \alpha \leq 2$, $a > 0$ and

$$\left| b \cos \left(\frac{\pi}{2} \alpha \right) \right| \leq a \sin \left(\frac{\pi}{2} \alpha \right). \quad (4)$$

In (3), the parameters a and b take into account the units and possible asymmetry of the random variable, respectively, so that a stable distribution type is defined for each possible value of α . Here, L_α denotes the stable distribution type with parameter α .

The note just referenced demonstrates inequality (4). There Lévy treats the case $0 < \alpha < 1$ and Khinchin the case $1 < \alpha < 2$. For $\alpha = 2$, equation (3) defines the normal distribution, for $\alpha = 1$ and $b = 0$, the Cauchy distribution. For all other values of α , although the distribution L_α is well defined by (3) and by an inverse Fourier transform, an explicit expression for the distribution function is not known. However approximate expressions are available, a point to be considered later.

At the Ecole Polytechnique in 1919, Lévy lectured several times on the theory of probabilities, as a substitute for Humbert. He covered the central limit theorem, mentioned the name *stable limit* of a sequence of distributions, and employed the Fourier transform.⁵⁰ In his autobiography (Lévy 1970, p. 77), he wrote

I would probably never have dreamt of this problem without a remark by Captain Lhoste,⁵¹ who heard my lectures in 1919.... He told me (using different language, of course) that Gauss' law is the only stable distribution. I did not dare reply. But the same evening....

What he discovered that same evening was that all distributions with the characteristic function

$$\varphi(z) = \exp(-a|z|^\alpha), \quad a > 0, 0 < \alpha \leq 2 \quad (5)$$

are stable, and when the distribution is symmetric, only these. He did not publish this result until 1922.⁵² By the academic year 1920–1921 Lévy held the chair of Analysis at the Ecole Polytechnique. The part of his course dedicated to the theory of probability was by then much more substantial than his lectures of 1919. In the course, he showed that the only stable distributions with finite variance are normal distributions. The proof is based on the functional equation (7) below, verified by the characteristic functions of the distributions. In the 1922 note, Lévy gave the definition (1) of stable distributions; he demonstrated again that the Cauchy distribution is

⁵⁰See pp. 495–513 of Humbert's lecture notes of his course in analysis at the Ecole Polytechnique (first division, 1919); the note 319; and also Barbut and Mazliak (2008).

⁵¹One can read about Lhoste in Hadjadji Seddik-Ameur, Nacira (2003) Les tests de normalité de Lhoste. Math. Sci. Hum. 162:19–43.

⁵²Lévy, Paul (1922) Sur le rôle de la loi de Gauss dans la théorie des erreurs. CRAS 174:855–857.

stable, and gave the explicit relation between the coefficients of the two expressions (1) and (2) above

$$a^\alpha = \sum_k a_k^\alpha. \quad (6)$$

In the case $\alpha = 2$, i.e. Gauss' law, (6) becomes the familiar expression for the variance of a sum of independent random variables. Finally, in the same note Lévy introduced the expression *domaine d'attraction* (*domain of attraction*) of a stable distribution, and stated that this domain includes all distributions for which the expectation $E(|X|^p)$ is finite when $p < \alpha$ and infinite when $p > \alpha$. In 1923 three notes to the CRAS of Paris appeared; one on April 23, which proved the existence of a stable distribution for every $\alpha \in (0, 2]$; one on May 7, which gave the expression (3) above (but not the inequality (4)) and proved the existence of asymmetric stable distributions; and finally one on May 14 which gave an example of a distribution L_α obtained as a limiting distribution.

In 1923 Lévy also wrote the important article *Théorie des erreurs. La loi de Gauss et les lois exceptionnelles*.⁵³ The *exceptional distributions* are, of course, the cases L_α with $\alpha < 2$. Lévy drew attention to the fact that taking the average of an increasing number of random *errors* with stable distributions having $\alpha \leq 1$, far from increasing the precision of a measurement, on the contrary makes the imprecision of the measurement tend to infinity, because for such distributions not only is the variance infinite (as in fact holds for all $\alpha < 2$) but also the expected value. This article contains Lévy's proof, celebrated for its elegance and conciseness, of the formula (3) or (5) for the characteristic function of a stable distribution, starting from the functional equation

$$\varphi(az) = \varphi(a_1z)\varphi(a_2z). \quad (7)$$

Because of the definition (1), the characteristic function of a stable distribution must satisfy this condition.

Maurice Fréchet used Lévy's proof in his article *Fréchet, Maurice. 1927. Sur la loi de probabilité de l'écart maximum*.⁵⁴ Fréchet applied Lévy's technique not to the characteristic function but to the distribution function itself to establish the form of the distributions stable with respect to the maximum of n independent random variables. This appears to have been the origin of an argument between Fréchet and Lévy (see Letter 21 and especially Lévy's article *Sur quelques travaux relatifs à la théorie des erreurs*).⁵⁵

Aside from the book of Lévy (1925) and the article of 1929, Lévy did not return to stable distributions until 1936 in his *Note commune* co-authored with Khinchin, in which he completed the one-dimensional case, and another note⁵⁶ which extends

⁵³(1924) Bull. Soc. Math. Fr. 52:49–85.

⁵⁴Ann. Soc. Pol. Math. 5:93–116.

⁵⁵(1929) Bull. Soc. Math. Fr. 53:1–21.

⁵⁶Intégrales à éléments aléatoires indépendants et lois stables à n variables. CRAS (March 17, 1936).

the idea to vector random variables in n dimensions. Then Lévy (1937) appeared. In his autobiography (Lévy 1970, p. 78), Lévy wrote: ‘I believe that this time I can say that in my book of 1937, the theory took its definitive form’. In fact this was only a stopping point. The definitive theory was developed later in the work of Doebelin, Lévy, Feller and the Soviet school. Doebelin (1938) and Gnedenko (1939), especially, discovered the characterization of the domains of attraction of the stable distributions.

In his work (Lévy 1937, § 56 and 57, pp. 198–204) Lévy identifies two properties that relate stable distributions to those discovered empirically by Pareto for the statistical distributions of income and inheritances.⁵⁷ The simplest Paretian form for a random variable X is

$$P(X > x) = \left(\frac{x_0}{x}\right)^\alpha, \quad x \geq x_0 > 0, \alpha > 0. \quad (8)$$

The first property is that the probability distribution function L_α ($0 < \alpha < 2$) is asymptotically Paretian in the sense that for x large⁵⁸

$$P(X > x) \equiv \frac{K}{x^\alpha}. \quad (9)$$

In other words, for large values of the variable X , Pareto’s Law with exponent α is a good approximation to the stable distribution L_α .

One can also obtain the distribution L_α as follows:

- (1) Choose n independent random variables $X_1, X_2, \dots, X_n$ from a *Paretian urn*, i.e. a distribution of form (8) with exponent α and,
- (2) if $1 < \alpha < 2$, define

$$Y_n = \frac{1}{n^{1/\alpha}} \sum_{k=1}^n (X_k - m), \quad (10)$$

where $m = x_0 \frac{\alpha}{\alpha-1}$ is the expectation of the Paretian distribution (8)

- (3) or, if $0 < \alpha \leq 1$, define

$$Z_n = \frac{1}{n^{1/\alpha}} \sum_{k=1}^n X_k. \quad (11)$$

(The expectation of the Paretian distribution is infinite for this range of α .)

As n goes to infinity, the probability distribution of Y_n for the larger α values and that of Z_n for the smaller approach the stable distribution L_α . This second property is the analog of the central limit theorem for non-Gaussian stable probability distributions. In order to assure convergence, it is necessary to replace the divisor $\sqrt{n}$

⁵⁷Pareto, Vilfredo. *Cours d’Économie Politique* Professé a l’Université de Lausanne. Volume I (1896), Volume II (1897). F. Rouge, Lausanne.

⁵⁸See the note 6 to the Preface by Kai Lai Chung.

of the sum of independent random variables by $n^{1/\alpha}$, which diverges more rapidly. When $\alpha = 1$ the limit is the Cauchy distribution, and here the arithmetic mean appears; for $\alpha = \frac{1}{2}$, the divisor is n^2 .

Since these two properties are quite general, the argument is a satisfying explanation of why distributions that are (at least approximately) asymptotically Paretian appear so frequently: for example, the distribution of the frequency of words in a sufficiently long text, the populations of cities, the length of rivers, etc. Fréchet (see for example Fréchet 1955), and later Benoît Mandelbrot and other authors, give multiple further examples in the physical sciences (physical geography, the physics of “sandpiles”), in the social sciences (demographics and economics), and even in the somewhat controversial subject of mathematical finance. It is appropriate and satisfying that these distributions are known as *Pareto-Lévy probability distributions*, which honors both their discoverer, Pareto, and their principal theoretician, Lévy.

6 Lévy and Potential Theory: From Analysis to Probability

Before he focused on probability, Paul Lévy was an analyst, a part of his career that is somewhat overlooked today. Lévy’s expertise in potential theory was considerable, and resulted in studies in mathematical analysis as well as in probability theory by way of Brownian motion. Several of Lévy’s letters to Fréchet mention these matters, which bestow a certain continuity on Lévy’s work.

There is extensive documentation of the intense and complex history of classical potential theory before 1945, which marked the earliest appearance of axiomatic and probabilistic theories.⁵⁹ The solution of the Dirichlet problem, the goal of generations of mathematical physicists and mathematicians, was the central problem of the classical theory. Up to the Second World War, the focus of attention alternated between the conditions on the domain, the boundary itself, and the boundary conditions. Research drew on all the central developments of functional analysis and contributed to the reanimation of entire branches of mathematics. The history of the probabilistic theory of potential has also been extensively investigated.⁶⁰

A brief overview of this immense subject will open the way to a deeper understanding of the spirit of Lévy’s work and certain of his letters to Fréchet. The work of George Green, made accessible by the young Sir William Thomson (the future Lord Kelvin) in 1845, is founded on three integral formulas which Green applied to the theory of electricity. These formulas also apply to gravity and heat, evidencing a formal mathematical analogy between these different physical phenomena.

⁵⁹For details, see Dieudonné (1978), Bouleau (1986), Locker (2001), Taton (1961), Brelot (1964), Choquet (1990), Chung (1995) and the very complete article by De la Pradelle (1988).

⁶⁰See the historical notices in the work of Dellacherie and Meyer (1975), the text by Bouleau (1986) and numerous remarks and notes by Doob (1953, 1980).

The three formulas in question are:

$$\iiint_V v \Delta u \, dv + \iiint_V \nabla u \nabla v \, dv = \iint_S v \frac{\partial u}{\partial v} \, ds \quad (\text{GI})$$

$$\iiint_V (v \Delta u - u \Delta v) \, dv = \iint_S \left(v \frac{\partial u}{\partial v} - u \frac{\partial v}{\partial v} \right) \, ds \quad (\text{GII})$$

$$u(P) = -\frac{1}{4\pi} \iiint_V \frac{1}{r} \Delta u \, dv + \frac{1}{4\pi} \iint_S \frac{\partial u}{\partial v} \frac{1}{r} \, ds - \frac{1}{4\pi} \iint_S u \frac{\partial}{\partial v} \frac{1}{r} \, ds \quad (\text{GIII})$$

In these formulas the normal derivative $\frac{\partial}{\partial v}$ is oriented inward, the volume integrals are over V and the surface integrals are over the surface S which bounds V with normal derivatives evaluated on S . The distance from the point P inside V to the point M being integrated over is r . On the right hand side of the third identity (GIII), the first term is the potential caused by a volume density $\frac{\Delta u}{4\pi}$, the second term is caused by a surface density $\frac{1}{4\pi} \frac{\partial u}{\partial v}$ and the third term is the potential due to a double layer potential⁶¹ of moment $\frac{1}{4\pi} u$, which Lévy put to use in a very clever way in 1910.

The Green formulas are a succinct presentation of the relations between functions and potentials. They reveal the central feature of potential theory, namely that a sufficiently regular function on a sufficiently regular domain is the sum of three potentials. If the function is in addition harmonic, it is the sum of the potential of a single layer and the potential of a double layer.⁶² Equation (GIII) gives a harmonic function u at every point in the volume V in terms of its normal derivative on the surface; the Neumann problem⁶³ in fact consists of determining a harmonic function in a domain in terms of the values of its normal derivative on the boundary. By manipulating (GII) one is led to introduce the function $H(P, M)$ such that, for fixed P in V , the Green function $G(P, M) = \frac{1}{PM} - H(P, M)$ ⁶⁴ is a harmonic function of M in V which vanishes on the boundary. One arrives at the equation

$$4\pi u(P) = \iint_S u(M) \frac{\partial G(P, M)}{\partial v_M} \, ds(M) \quad (*)$$

⁶¹See footnote 62 below.

⁶²The double layer potential consists of a surface distribution of dipole moments all oriented toward the same side of a double-sided surface. The double layer potential may also be viewed as the result of two surface charge distributions when the surfaces are infinitesimally close. The usage of the word “distribution” here follows that of physics of the period, that is as a density of electric charges. During the first quarter of the 20th century this was a continuous function. Only later, in the work of Schwartz, did this concept as well as that of a *unit-doublet* find definitions not requiring a limit by means of distribution theory.

⁶³This refers to the Berlin academic Carl Neumann (1832–1925), who was editor of *Mathematische Annalen*.

⁶⁴In two dimensions the potential is logarithmic and the singularity which must be removed is $\ln PM = \ln r$, rather than $1/PM = 1/r$.

which determines u by its values on the boundary.⁶⁵ This same equation was employed by Lévy during the Second World War to discover new relations between Brownian motion and the harmonic measure.

The “harmonic Dirichlet problem”⁶⁶ seeks to determine a harmonic function u in a domain by its values on the boundary. The Green function solves this problem if the domain and its boundary satisfy certain appropriate conditions. Green used ideas from electrostatics to derive equation (*), which also gives the temperature $u(P)$ at a point P in an isotropic homogeneous body when the boundary is held at a temperature $u(M)$. This visualization played an important role in potential theory during Lévy’s time and was very significant for Lévy’s intuition about diffusion and Brownian motion.

The general Dirichlet problem is to determine a function u that satisfies $\Delta u = \rho$ in a domain Ω for a given function ρ by specifying its values on the boundary. Under sufficient conditions on the domain, the solution is obtained by adding to the solution of the corresponding harmonic Dirichlet problem the unique function v which vanishes on the boundary and satisfies $\Delta v = \rho$ in the interior of the open domain. In three dimensions, one has $v(P) = \frac{1}{4\pi} \iiint_{\Omega} \rho(M) G(M, P) dM$. The existence of these solutions for domains of which the shape (i.e. the geometry and topology) is that of a real conductor is evident from the results of physical experiments and common sense. The work of mathematical physicists gave this observation a rigorous basis.

Dirichlet’s principle shows that the solution of the Dirichlet problem is characterized as the minimum of an energy integral. This guarantees the existence of a solution and serves as a method of calculating it. However, the extent of its applicability came into question with the development of a mathematical potential theory encompassing domains of more general shape. This occurred, notably, in Weierstrass’ criticisms of the work of Riemann⁶⁷ who had used the Dirichlet principle in his theorem on holomorphic transformations and conformal representations. As for the method of Green, not only does the issue of the conditions necessary for the existence of the Green function in a general domain arise, but finding an explicit Green function generally entails intractable computations.⁶⁸

⁶⁵Independently of Green, Gauss’ mean-value theorem for harmonic functions also shows that the potential is determined by its values on the boundary. This result allows one to derive the Poisson representation formula.

⁶⁶This designation is used deliberately, since it was introduced by Lévy and employed in 1926 by Bouligand in the volume of the *Mémoire des Sciences Mathématiques* he devoted to these questions.

⁶⁷Especially for the question of the theorem on the conformal representation in the complex plane which is related to the questions considered here. It is well known that in two dimensions, potential theory has a close relation to the theory of holomorphic functions.

⁶⁸The Green function is itself the solution of a Dirichlet problem. Except for domains with a boundary consisting of curves or surfaces of a simple geometry, for which the Green function is known from the work of Poisson, Liouville, Cauchy and their successors, very few explicit examples are available. However numerical methods may be used to determine Green functions to a high degree of accuracy.

One of Lévy's first articles as a professional mathematician⁶⁹ concerns a question in potential theory. It announces his thesis defense of 'Integro-differential equations defining functions of line' (Lévy 1911) of the following year, when he also published another paper.⁷⁰ Lévy's thesis committee was composed of Hadamard, Picard and Poincaré, three eminent scientists who had worked on potential theory. Picard was interested in ways of solving the differential and integro-differential equations of mathematical physics and published extensively on potential theory between 1900 and 1904. He gained distinction by introducing the famous 'Picard Principle' of successive approximations in his lectures at the Sorbonne. Poincaré introduced his 'balayage method' in 1887 (see Bouleau 1986). In his authoritative article of 1887, he put forward a solution for 'a conductor of arbitrary shape' after having criticized the solutions of Riemann, Schwarz and Harnack.⁷¹ Despite the efforts of Hilbert, the systematic utilization of Dirichlet's principle fell into question.⁷²

In 1906 Hadamard furnished new tools for potential theory utilizing his functional calculus. He discovered variational equations for the Green function of a domain whose boundary undergoes continuous deformations. In such circumstances, the Green function ceases to be a symmetric function of two points and becomes a function of the domain and of its boundary (and hence a function of line in Volterra's sense). This framework admits a continuous deformation of the domain and the corresponding Green function may be characterized.

As Hadamard's student Lévy was seemingly destined to pursue the French approach to potential theory, which for a time favored the treatment of Dirichlet's problem by functional analysis, before returning to Poincaré's method. In the work of Hadamard and then Lévy, Green function techniques were brought to center stage, together with the equations, including functional derivatives, that they imply and the generalizations that they entail. In fact, the longest part of Lévy's thesis is devoted to these generalizations.

Lévy began his five chapter opus with a very dense and relatively long introduction, constituting a third of the total 130 pages. In this introduction, Lévy presents what he calls "a summary of the works of Hadamard and Volterra on the functions of line," an expression taken from the works of Volterra and his Italian school, later rebaptized functionals by Hadamard. The two sole examples given by Lévy are:

⁶⁹Lévy, Paul (1909) Sur les valeurs de la fonction de Green dans le voisinage du contour. *Bull. Soc. Math. Fr.* 34:186–190.

⁷⁰Lévy, Paul (1911) Une généralisation de la méthode de Fredholm pour la résolution du problème de Dirichlet. *J. Ec. Polytech.* 15:197–210. Fredholm, the Swedish creator of the theory of integral equations that bears his name, developed this method in 1900.

⁷¹He called Riemann's method (which uses Dirichlet's principle) 'simple and elegant but not rigorous,' and the other methods are quite insufficient in Poincaré's opinion for solving Dirichlet's problem.

⁷²Hilbert declared at the congress of 1900: 'Every problem in the calculus of variations has a solution provided that one makes the appropriate hypotheses on the boundary conditions and, if necessary sufficiently extends the concept of solution'. However in 1901, in order to give a proper proof of Riemann's theorem, he had to substitute the calculus of variations for very sophisticated theorems on equicontinuous families of harmonic functions.

“the coefficient of mutual induction of two boundary curves” and “the Green function which is a function of two points of a plane, the form of which depends on a closed contour C . This is thus a function of two points and of the contour C . It is of a type of analogous functions that play an important role in mathematical physics.”

By far the most original part of Lévy’s thesis was Chap. I, which covers higher-order functional derivatives. With these new derivatives (especially those of order two), Lévy became a creative mathematician and set out the tools that would later permit him to generalize the Laplacian and the Dirichlet equation to functional calculus.

In his second chapter, Lévy studied equations of the form

$$\delta\Phi_A^B = \int_C f(\Phi_A^B, \Phi_M^A, \Phi_M^B; A, B, M) dS(M).$$

Chapter III was dedicated to the integro-differential equation satisfied by the Green functions. It examines the consequences of the Hadamard variational equation for the Green function

$$\delta G(A, B) = \frac{1}{(n-2)S_n} \oint_{\Sigma} \frac{\partial G(A, M)}{\partial \nu_M} \cdot \frac{\partial G(M, B)}{\partial \nu_M} (\nu, \delta M) d\sigma(M)$$

in dimension $n > 2$.

Chapter IV is concerned with the potential of nested charged slabs, which it treats by successive approximation, an application of Picard’s method.

Finally, in Chap. V, which is more an elegant exercise than a fundamental contribution, Lévy addresses the solution of a functional version of the equation $\Delta \Delta u = 0$.

Aside from the appealing theoretical questions he solved in his thesis (Hadamard told him⁷³ “since you have taken up this subject, I leave it to you”), Lévy remained true to the French tradition of mathematical physics. He used the most recent tools of functional analysis, which seemed to him to be the most appropriate, and did not neglect the physical interpretation of his results. Such interpretations served him as a guide to various topics, including questions of the relation between potential theory and probability in Brownian motion (cf. Sect. 7). Between 1912 and 1919 Lévy published several articles touching on potential theory, gave a presentation of this research in 1922 (Lévy 1922), and in 1935 published a note on tensorial forms of Green functional equations.⁷⁴

Lévy’s first letters to Fréchet are about functional analysis and included a list of the types of functions that interested him. The function spaces that are today canonical in analysis can be read between the lines. The main topics are various features of Green functions and the Laplace and Dirichlet equations. Above all, Lévy rails against Volterra’s formulation of the Laplace equation in terms of “functions of line,” as may be read in the scathing Letter 5. It is important to emphasize the extent

⁷³Lévy (1970, p. 42).

⁷⁴Lévy, Paul (1935) Sur une forme tensorielle des équations fonctionnelles des fonctions de Green et de Neumann. CRAS 200:1723–1725.

to which Lévy presented himself as an analyst who followed Hadamard and differentiated himself from Volterra, because the theoretical links between Volterra's "functions of line" and the functionals of Hadamard are in fact very close. Laurent Schwartz wrote in 1988 that the title of Lévy's thesis "corresponded to current ideas that followed from the work of Volterra."⁷⁵ Once he had completed his thesis, Lévy possessed the mathematical tools to modify and extend certain facets of Volterra's calculations relevant to potential theory.⁷⁶ Not only in his thesis, but in his entire opus of analytical research (which often consisted of reworking certain problems so that they fell under the conditions set out in his thesis), Lévy kept to generalizations of functional derivatives, especially the Laplacian. He thus inserted himself into the framework built by Hadamard, becoming his mentor's theoretical and technical successor. It was only after the First World War that Hadamard prompted Lévy to investigate Fréchet's techniques and asked him to assemble the posthumous edition of Gateaux's work.⁷⁷ In his 1922 book (Lévy 1922) and his letters to Fréchet, Lévy used Volterra's notation for functional variations and derivatives, but clearly departed from Volterra's conclusions. Volterra used the equation $\delta U = \int_0^1 \phi(t) \delta x(t) dt$ for the variation of a functional, and Lévy specified that in cases where that equation applies "we say that U is differentiable in the sense of M. Volterra. . . M. Volterra represents it by $\phi(t) = U'[x(t)]$." Lévy also gave cases to which Volterra's formula does not apply.⁷⁸ He took care to distinguish in the first-order terms between the variation or differential in the sense of Fréchet and the variation or differential in the sense of Gateaux. He defined the corresponding second variation and paid particular attention to second variations of the form $\delta^2 U = \int_0^1 f(t) \delta x(t) dt + \int_0^1 \int_0^1 g(t, t_1) \delta x(t) \delta x_1(t_1) dt dt_1$ where the kernels f and g generalize the square ($\frac{\partial^2}{\partial x^2}$) and rectangle ($\frac{\partial^2}{\partial x \partial y}$) second partial derivatives. Lévy thus highlighted the absence of a term in Volterra's formula for the second variation and gave, as a generalization of the Laplace equation, the equation $\int_0^1 f(t) dt = 0$ in place of Volterra's expression $\int_0^1 g(t, t) dt = 0$, which Lévy considered to be meaningless (Lévy 1922, p. 86). In 1970 (Lévy 1970), he revisited this question, which had always seemed to him to be one of his principal contributions to the calculus of variations and potential theory, maintaining that as of 1901 Hadamard "had drawn attention to the importance of the second term" which "plays such a large role in Gateaux's work and mine."

⁷⁵ Astérisque (1988) n° 157–158.

⁷⁶ See the article Fichera, Gaetano (1994) Vito Volterra and the birth of functional analysis. In Pier (1994), pp. 171–183.

⁷⁷ For this fascinating story, see the detailed study of Mazliak (2011).

⁷⁸ For the classic functionals of mechanics of the form $\int_0^1 f(t, x, \dot{x}) dt$, the variation cannot be put in this form, except in certain special cases. For linear functionals, which were characterized in 1903 by Hadamard and transformed in 1907 by Riesz into a representation as Stieltjes integrals $\int_0^1 x(t) dF(t)$, the variation can again not be put in Volterra's form, because of the possibility of points of discontinuity of F . Also, Volterra's form is not possible when $U(x) = x(\tau)$ (i.e. when F is the Dirac measure at the point τ) for which $\delta U = (\delta x)(\tau)$.

By 1924, analysis and potential theory had made significant progress. Lebesgue became aware of the very advanced results of the new American school of potential theory. The Americans had maintained tight relations with the German school. Oliver Dimon Kellogg (whose book *Foundations of Potential Theory* is still in print by Dover and has become a classic on American campuses) wrote his thesis in German, guided by Hilbert, and then went to the University of Missouri, where together with William Osgood (president of the AMS from 1905 to 1906) he helped develop the discipline in the USA. Griffith Conrad Evans (president of the AMS from 1924 to 1926) was a student of Osgood and an eminent specialist in potential theory. By 1924, he had numerous publications in this area. At about the same time the great German specialist Arthur Korn, who met Poincaré in 1891 and won the Valliant prize of the Paris Academy of Science in 1907, had established good relations on the international scene, especially with the Americans. The French school, by contrast seemed quite diffident and inward looking. Thus Lebesgue regarded the American work as so important that he followed up on Wiener's note of March 24, 1924 in the CRAS⁷⁹ with an 'Observation' of his own and with the presentation of a note by Georges Bouligand, of which Lebesgue is probably the source. Lebesgue wrote:

Last January, an article by G. Bouligand gave me the occasion to present several results obtained in the course of research on insolvable examples of the Dirichlet problem, research that began in 1913.⁸⁰ Wiener spoke at length about my article, but what he did not say, and I would like to mention, is that during recent years the Dirichlet problem has been, unbeknownst to me, the object of fruitful research by various American researchers, especially Wiener. Their theorems often go beyond mine, and they have so well addressed the various questions that concerned me that had I been aware of their results I would have undoubtedly found it pointless to follow up on my results of 1913.... The work of the American scientists was not known to Bouligand either, however one should add that the important note by Bouligand in 1919⁸¹ seems to have been overlooked by the Americans. Now that these researchers are aware of each other, one may hope that they will become yet more fruitful.

Lévy turned most of his attention to probability in 1920, but did continue for some time to work on analysis. In addition Wiener, who originated the mathematical theory of Brownian motion in 1923, participated assertively in potential theory and the French, American and German schools converged on this problem. Wiener responded to the articles of Oskar Perron from the USA, discovering one of the final steps that resolved the Dirichlet problem in the case of a continuous boundary.⁸² He also took important steps toward the solution of the problem with a discontinu-

⁷⁹Wiener, Norbert (1924) Une condition nécessaire et suffisante de possibilité pour le problème de Dirichlet. CRAS 178:1050–1053.

⁸⁰Lebesgue discovered a famous counter-example of insolvability of the Dirichlet problem (called Lebesgue's spine). This demonstrates the close relation between the existence of a solution and the topology of the boundary. He was preceded in this type of example by the Polish mathematician Zaremba.

⁸¹Lebesgue is referring to the two notes in CRAS by Bouligand on November 3 and 17, 1919.

⁸²Wiener, Norbert (1925) Note on a paper of O. Perron. J. Math. Phys. 4:21–32.

ous boundary. Perron and Wiener are at the origin of the PWB (Perron, Wiener and Brelot) methods, which play an important role in potential theory.

In 1926 Georges Bouligand took up the task of writing a survey of potential theory for *Mémoires des Sciences Mathématiques* (Bouligand 1926). He cited Lévy extensively—for his thesis work and his sophisticated introduction of the double layer potential—as well as the other members of the French school. The German contributions are prominently referenced, and several pages are devoted to the Americans, through an analysis of the results of Wiener and Phillips. Lévy's work, however, soon became outdated and the Green function approach of Hadamard did not achieve its desired goals. Wiener's theory of Brownian motion and his remarkable results in classical potential theory (including the characterization of sets with vanishing electrostatic capacity, important for probability) included all Wiener needed to begin probabilistic potential theory. It is a bit surprising he did not do so. Instead the Japanese Shizuo Kakutani began this work in 1944 (Bouleau 1986), and the theorem which carries his name is recognized today as the starting point of the probabilistic potential theory developed after the Second World War. It should be recognized, however, that Kakutani's theorem had already been proven by Lévy in 1943, as can be witnessed in the riveting letters of that year (Letters 44–54).

The concept of harmonic measure, initiated in 1935 by the Nevanlinna (1936) played an important role in the finalization of classical potential theory and generated some beautiful results on holomorphic functions. In the 1930s the French school—and before long the entire international community concerned with potential theory—was dominated by the imposing figure of Marcel Brelot. Just before the Second World War, by introducing the idea of resolutive boundary conditions, Brelot expanded the Dirichlet problem to functions integrable with respect to the harmonic measure and gave the problem, in the words of Choquet, a 'definitive form.' At the end of the war, using integrability over the ramified harmonic measure,⁸³ Brelot gave the necessary and sufficient conditions for solvability. This emphasizes the importance of the result published independently by Lévy in 1943 and Kakutani at the end of 1944. The connections between Brownian motion and potential theory were already known, and then were extended to random walks, leading to meetings between probabilists and potential theory specialists (Courant, Friedrichs and Lewy in 1928, Petrovski, Khinchin and Kolmogorov during the 1930s, etc.). This marked an important step; the tool which allowed the solution of the central problem of classical potential theory also ushered in the theory of probability.

Lévy was only able to publish his result after the Liberation of France, whereas Kakutani's work did not become known until the Japanese capitulation in 1945. The result is that under suitable assumptions, for a domain D with boundary F and a subset F' of F , the Dirichlet problem $\Delta U = 0$ in D with boundary condition $U = 1$ on F' and 0 on $F \setminus F'$ has a unique solution $P(A, F')$ for each point $A \in D$ which is the harmonic measure of F' with respect to F . For a domain with a Green

⁸³CRAS (November 26, 1945).

function G , one also has $P(A, F') = \oint_{F'} \frac{\partial G(A, M)}{\partial \nu} d\sigma(M)$.⁸⁴ Today Kakutani's theorem may be stated as follows: the conditional expectation value $g(x) = E_A(f(B_\tau))$, where τ is the earliest time at which a Brownian variable B starting at x reaches the boundary F on a region of the plane is a harmonic function of x which solves the Dirichlet problem $\Delta u = 0$ in D with boundary condition f . This formulation is completely equivalent to the theorem stated by Lévy and then by Kakutani. Their explicit statements differ very little:

Lévy (1948): "The probability that a Brownian motion beginning at a point A inside a domain D attains the boundary F for the first time on F' is the harmonic measure of F' with respect to F viewed from the point A ."

Kakutani (1944b): "Let D be a domain in the Gaussian plane R_2 , and let E be a closed set on the boundary $\text{Bd}(D)$ of D . Then, under certain assumptions on D and E , the probability that the Brownian motion starting from a point $\xi \in D$ will enter into E without entering into $\text{Bd}(D) - E$ before it, is equal to the harmonic measure in the sense of Nevanlinna of E with respect to the domain D and the point."

Lévy's result holds in dimension n , so it might seem more general than Kakutani's, which is limited to dimension two. However a careful analysis (see Locker 2001) of Kakutani (1944a) shows that Kakutani already knew at the end of 1944 that his theorem held for arbitrary dimension. Although Kakutani's proof is written with complex variables, which shows his preoccupation with two dimensions and includes several crucial steps of a probabilistic nature that were not yet proven in 1944, it nevertheless immediately generalizes to arbitrary dimension. This analysis also shows that the assumptions of Lévy and Kakutani on the type of domains and their boundaries are essentially identical. The assumptions are explicit for Kakutani and implicit for Lévy. However, for Lévy they are made clear by the theorems from analysis that he used which require that the domain D must be simply connected and its boundary composed of Jordan curves.

There is a technical difference between Lévy's and Kakutani's results. Lévy's approach to the harmonic measure is via Green functions (as one would expect, given his extensive study of Green functions after the First World War and the way in which he preferred to be guided by analogy with the theory of heat conduction), whereas Kakutani completely avoided their use. There is also a difference in the consequences for Lévy and Kakutani. The theorem did not lead Lévy beyond considerations of Brownian motion and its trajectories, although he was able (Lévy 1948)

⁸⁴This version of the theorem is sufficient to understand the Lévy-Fréchet correspondence. Modern definitions of the harmonic measure are more subtle and their connection with the Dirichlet problem more general. This more recent work was originated by Wiener and then extended by Brelot. The harmonic measure appears as a Radon measure H . If f is super-harmonic in a given open set and F is the set of super-harmonic functions g such that $\liminf(x) \geq f(y)$ (for x an interior point and y on the boundary), one may consider $\overline{H}(f) = \inf F$ and $\underline{H}(f) = -\overline{H}(-f)$. When $\underline{H}(f) = \overline{H}(f) = H(f)$ the boundary condition f is resolute, meaning that $H(f)$ is a generalized solution of the Dirichlet problem. Wiener had already shown that $f \mapsto H(f)(x)$ is a Radon measure H_x . This Radon measure is nothing but the functional representation of the measure. In 1945 Marcel Brelot considered the compactified projective space of a real vector space of arbitrary finite dimension, and redefined the ramified harmonic measure by an extension of the previous method to the space, making use of Daniell's integral.

to connect it with some very significant work of the Soviet school from the 1930s. This link allowed an expansion of the theorem's probabilistic significance by connecting it to the law of the iterated logarithm, via the work of Bernstein, Khinchin, Kolmogorov and Petrovski. Kolmogorov was able to deduce this law from Petrovski (1933, 1935) by considering parabolic equations, but he never published a proof. It was Lévy (1948) who, having heard of it directly from Kolmogorov, made this result known, some years after Ville (1939), Doeblin (1940) and Fortet (1943) had alluded to it. By contrast, in his December 1944 article, Kakutani presented his theorem as the basis for five additional theorems, among them one that established that the sets of capacity zero already characterized by Wiener are composed of precisely those points which Brownian motion almost surely does not reach. These five theorems became part of a program to represent harmonic functions with Brownian motion. Kakutani thus originated a strategy that ultimately led to the fusion of important areas of probability theory and potential theory.

It is nonetheless true that Lévy found the theorem before Kakutani and that Lévy, who gave three different proofs of it, differs from Kakutani in at least the two dimensional case, which makes use of the conformal invariance of planar Brownian motion (see Sects. 7 and 8). Today, this proof is taken as the starting point for the application of the theory of conformal martingales to holomorphic functions. Its extraordinary content, only briefly alluded to in the letters, is as follows: For a circle in the plane with center A , the harmonic measure of an arc F' of the circle viewed from A reduces to the measure of the angle at the center defined by the arc, divided by 2π . Due to the homogeneity of Brownian motion starting at A , this is also the probability that the motion leaves the disk for the first time via the arc. A conformal transformation maps the general case into this simple situation.

To conclude this brief overview of an enormous subject, potential theory, it is necessary to touch on Lévy's research on Markov processes, begun in 1950. This is mentioned in the last letters of the correspondence and also by Kai Lai Chung in the Preface to this volume. The rediscovery of Doeblin by the Americans in 1950 and the work of Doob, Chung and also Hunt led to a renewal of interest in the theory of Markov processes, which in turn ushered in a merging of whole areas of potential theory and probability (see Bouleau 1986). One might add that Lévy, whose theory used "fictitious states" (the set of which constitutes the "Chung boundary set"), divided his classification into seven types of instantaneous states (the states E for which the sojourn time is zero (see Lévy 1970)). In this way he generalized the idea of a Markov process by replacing the time with surfaces in Euclidean space of dimension n , showing that the corresponding generalization of Brownian motion is no longer Markovian in the usual sense. Thus, beginning in 1955, Lévy was plunged into the difficult task of separating Markovian from strictly Markovian processes. His last academic publication (at age 84!) is entitled *Un nouveau point de vue pour l'étude des processus Markoviens*.⁸⁵ In it, using the example of Brownian motion on a curve in the form of a figure eight, i.e. a curve with a double point, Lévy attempted

⁸⁵CRAS (May 25, 1970).

to relate Markov processes to strictly Markov processes (Lévy 1970) by introducing Riemann surfaces above the trajectories. C. Dellacherie and P.-A. Meyer (1975)⁸⁶ put Lévy among the ‘ancestors of the study of trajectories.’ They add that Lévy had considered ‘irregular Markov chains [and] that his work remained misunderstood for a long time.’

7 Lévy and Brownian Motion

In 1827 the Scottish botanist Robert Brown looked through his microscope at small particles found in pollen grains suspended in water and noticed their unusual, haphazard motion. For the next sixty years explanations of this phenomenon were primarily animist in nature before a physical reason emerged: In 1887 Father Ignace Carbonnelle attributed Brownian motion to thermal agitation; a year later Georges Gouy performed experiments varying several parameters and concluded that “this motion is the subtle result of thermal agitation.”⁸⁷

At the end of the 19th century and dawn of the twentieth, certain features of Brownian motion were deemed relevant to basic question of physics, namely: are matter and energy continuous or discontinuous and is the validity of the Carnot principle an absolute law or a statement of high probability? Although the second half of the 19th century—thanks to Maxwell and Boltzmann—had seen probabilities penetrate deep into physics, notably in statistical mechanics and the dynamics of gases, nevertheless the meaning and significance of probability for physics had been obscured by philosophical issues. Now it became an important tool for the study of certain physical phenomena for which a particle description seemed necessary.

The earliest mathematical model relevant to Brownian motion appeared in Louis Bachelier’s thesis,⁸⁸ defended in 1900 (Bachelier 1900). The thesis includes a model of the fluctuations $X(t)$ of the stock market as independent increments $dX(t) = X(t + \delta t) - X(t)$ following a Gaussian law with standard deviation $\sqrt{\delta t}$. Such stochastic processes are now called Wiener processes, Wiener-Lévy processes, or linear Brownian motion. The equation $\frac{\partial f}{\partial t} = \frac{\partial^2 f}{\partial x^2}$ satisfied by the density $f(x, t)$ of the process is also the heat equation for a homogeneous substance in steady state. Bachelier included it under the name “radiation of probability,” but it was a dozen years before he attempted to connect his work with the topic of this section, Brownian motion. In the meantime physicists, notably Albert Einstein, had found an explanation for the phenomenon and begun a mathematical description of it. Beginning in 1902 Einstein developed a new molecular kinetic theory of heat. Thus Brownian motion was incorporated in the new physics of the day and merged with the theory of probability.⁸⁹

⁸⁶Chapters XII–XVI, p. 340, revised edition of 1987.

⁸⁷An analysis of Gouy’s work and its importance for physics appeared in 1941 (Picard 1941).

⁸⁸See the discussion in Sect. 2.

⁸⁹At this time Einstein also revolutionized mechanics with his theory of special relativity and reintroduced the particle picture of light, i.e. photons and the photoelectric effect.

In a 1905 *Annalen der Physik* article (Einstein 1905) Einstein described the mathematical properties of Brownian motion implied by his new molecular kinetic theory. The physical existence of Brownian motion validated his mathematical theory. In 1906, again in *Annalen der Physik*, Einstein showed how one could determine molecular dimensions, and in a third contribution gave a consistent theory of Brownian motion, including gravitational effects. From the diffusion equation $\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2}$, where f is the number of particles per unit volume at the point with coordinate x and D the diffusion constant, Einstein showed $f(x, t)$ to be proportional to $\frac{1}{\sqrt{4D\pi}} \frac{e^{-x^2/(4Dt)}}{\sqrt{t}}$. Up to a multiplicative factor, this is just a Gaussian with standard deviation $\sqrt{2Dt}$. Experiments, especially those of Jean Perrin (who learned of Einstein's results from Langevin), could then measure molecular masses and Avogadro's number by using Einstein's theory. Einstein was in fact not the only one to deal with Brownian motion. In the same year (1906) and the same journal (*Annalen der Physik*), Marian von Smoluchowski began a series of very mathematical articles on physical Brownian motion (see Von Plato 1994 for details).

Now it would seem that the Brownian increments, being proportional to $\sqrt{\delta t}$, rule out any consideration of particle speed, and in fact in his book *Les atomes* (Perrin 1912) Jean Perrin declared that

the changes to the trajectory are so numerous and so rapid that they are impossible to follow. The apparent average speed varies wildly in magnitude and direction, without tending to a limit as the observation time decreases. This is a case in which it is natural to think of those functions without derivatives imagined by mathematicians.

In the original edition of *Les atomes*, Perrin further remarked that the trajectories are nonetheless continuous “because we cannot imagine that the particle goes from one position to another without cutting any plane which separates the two positions.” Other mathematicians and physicists constructed models of physical Brownian motion, or similar motions. Paul Langevin (1908), Adrian Fokker (1914) and Max Planck (1917) constructed models with continuous trajectories for which one can find a distribution of speeds. The process formalized by Leonard Ornstein and George Eugene Uhlenbeck (1930) of the Lorentz school, known today as the Ornstein-Uhlenbeck process, is sometimes considered as an alternative to Brownian motion. It can be derived from Brownian motion X by the analytic formula $U(t) = e^{-t/2} X(e^t)$.⁹⁰

Wiener as of 1922 and as of 1934 Lévy each gave a mathematically consistent theory of Brownian motion as a continuous process with continuous but non-differentiable trajectories. Wiener described general relations between his constructions and statistical mechanics. Lévy made frequent use of the analogy between Brownian motion and the theory of heat as a guide or aide in his research (see, e.g., Letter 45). Aside from the forays of Bachelier twenty years earlier, this work signaled the emergence of Brownian motion as a subject for mathematical research. Wiener, with his Hilbertian approach, was more analyst than probabilist. Lévy, by

⁹⁰For more on these questions as well as a general history of Brownian motion, see Kahane (1998) and Pier (1994).

contrast, made Brownian motion a central theme in the development of the theory of stochastic processes in continuous time. In Wiener's *Differential Space* (Wiener 1923), his seminal article of 1923, the continuity of the trajectories is assumed *a priori*. Lévy began, in 1934, a reformulation of Brownian motion which led him to an effective procedure for constructing trajectories that are almost surely continuous.

Wiener had already produced three articles on Brownian motion before writing *Differential Space*. In the references to all these articles, one finds Lévy, with his *Leçons d'analyse fonctionnelle (Lessons in Functional Analysis)* (Lévy 1922), in company with Gateaux, Daniell and Einstein.⁹¹

In Wiener's *Differential Space*, the basic constituents of the theory are the measures on the space of continuous functions associated to the increments of the function f which represents a trajectory of the Brownian motion. The differences (or increments) $f(t) - f(t')$ between two positions of the particle are thus the bricks of Wiener's mathematical edifice. Wiener constructs a measure on the space of continuous trajectories that assigns a probability to all "reasonable" events constructed with the Brownian motion. He does this in two different ways. The first is inspired by the work of Lévy and Gateaux, and the second by Daniell's integral, which was the most general and abstract extension of the idea of an integral then available. Wiener's measure μ can be represented by its definition on certain functions F that only depend on a finite number of increments of functions

$$\begin{aligned} \mu(F) = & \pi^{-n/2} \prod_{k=1}^n (t_k - t_{k-1})^{-1/2} \\ & \times \int_{\mathbb{R}^n} du \exp \left[- \sum_{k=1}^n \frac{u_k^2}{t_k - t_{k-1}} \right] F(u_1, \dots, u_1 + \dots + u_n). \end{aligned}$$

Daniell's theorem then allows an extension to all μ -integrable functionals and the Lévy and Gateaux averages (see Mazliak 2011 and Sects. 6 and 8) give an equivalent extension. Lévy mentions in his autobiography (Lévy 1970) how close he came in his *Lessons in Functional Analysis* (Lévy 1922) to Wiener's measure, but at that time (1922) Lévy's methods were familiar only to a very few specialists, and Daniell's integral appeared so sophisticated that Wiener felt obliged to project his new measure on the better understood Lebesgue measure. Wiener was able to take account of the observable properties of Brownian motion and its expected mathematical behavior: independence of the increments of position $X(t) - X(s)$, which follow a Gaussian law with standard deviation $\sqrt{t - s}$, and non-differentiable trajectories. Wiener also identified other more subtle mathematical properties such as the almost sure Hölder continuity with exponent $\frac{1}{2} - \varepsilon$ of the trajectories in a given interval.

By 1923 Wiener had constructed the first mathematical model of Brownian motion. In a certain sense this was the first model of a stochastic process in continuous time, ten years before Kolmogorov's *Grundbegriffe* (Sect. 3). A new mathematical

⁹¹For details on this transition from Gateaux and Lévy's works to Wiener's, see Mazliak (2011).

dimension was given to Brownian motion in the 1930s by Wiener's work with Paley and Zygmund. This introduced a second mathematical formulation making use of the Paley-Wiener series

$$X(t) + iY(t) = \frac{\xi' + i\eta'}{\sqrt{2\pi}} + \sum_1^{\infty} \frac{(\xi_n + i\eta_n)(\cos nt - 1)}{\sqrt{n\pi}} + \sum_1^{\infty} \frac{\xi'_n + i\eta'_n}{\sqrt{n\pi}} \sin nt$$

which represents the motion as a trigonometric series with random independent Gaussian coefficients. This series had been introduced by Wiener in a paper in 1924, with the collaboration of Lévy. Wiener had visited Lévy at Pougues-les-Eaux during the summer of 1922. Lévy describes this pivotal encounter in Lévy (1970), relating how Wiener became familiar with the theory of countable probabilities on which Lévy, following Borel, focused. As a result, Wiener was able to offer a new presentation of *Differential Space* (Wiener 1923) in the *Bulletin de la Société mathématique de France*.⁹² On the first page of that article, written in French, Wiener “thanks M. Paul Lévy for his help in editing the French version of this article.” This is followed by the representation of the motion as a trigonometric series and a proof of its equivalence to the representation in *Differential Space* (Wiener 1923). Thus, Lévy was involved alongside Wiener in the mathematical history of Brownian motion, at the time of Wiener's first publications. In 1934 Lévy returned to work on Brownian motion again and began to combine it with his theory of additive processes. Then the motion appears naturally via the decomposition of such processes into three parts, with Wiener's function being the continuous part. In 1934 Lévy published two notes and a supplement on these matters.⁹³ He further refined these articles in 1937 (Lévy 1937).

Lévy began his reformulation of Brownian motion with the idea that it could be a paradigm for random functions. For Lévy, a random function was indeed a function, but one that is created by means of an intervention of randomness at each instant. To account for this development through the existence of a stochastic process, it is necessary to specify an effective procedure for construction of the process. This is quite different from Wiener's approach, which *a priori* takes the space of trajectories to be a space of continuous functions and constructs a measure that incorporates the required properties, even if he uses the theory of countable probabilities for that aim. As culminating point of the reformulation that he began in 1934, Lévy proposed the “constructive definition of Brownian motion,” which consisted of constructive procedures using successively finer approximations. At each step the approximation employs a finite number of random points, for which the joint laws satisfy the properties of independent increments. In the limit, the required properties of Wiener's function remain satisfied and continuity, assumed *a priori* by Wiener, is assured. In

⁹²Wiener, Norbert (1924) Un problème de probabilités dénombrables. Bull. Soc. Math. Fr. 52:569–578.

⁹³Lévy, Paul (1934) Les généralisations de l'espace différentiel de N. Wiener. CRAS (February 26); Lévy, Paul (1934) Sur les espaces *V* et *W*. CRAS (March 26); and Lévy, Paul (1934) Complément à l'étude des espaces *V* et *W*. CRAS (May 7).

articles in 1940 and 1948 (Lévy 1940 and Lévy 1948, respectively), the interpolation formula is the only constructive method that Lévy retains. This interpolative method is related to the first works by Bernstein on the Kolmogorov problem in the 1930s, i.e. the construction of all Markov processes with continuous trajectories (see Sect. 8).

Lévy's beautiful constructive definition of Brownian motion reveals many properties in arbitrary dimension that he would discover later. Lévy worked with subdivisions associated with the dyadic numbers in $[0, 1]$. If $X(t)$ is the function of linear Brownian motion, and $t_0 < t_1 < t_2$ are three values of t in the interval under consideration, given $X(t_0)$ there are two equivalent stochastic methods to determine $X(t_1)$ and $X(t_2)$. Either one determines the two successive increments $X' = X(t_1) - X(t_0)$ and $X'' = X(t_2) - X(t_1)$ independently, or first finds the total increment, and then interpolates, that is determines $X(t_1)$ according to the conditional law for that variable when $X(t_0)$ and $X(t_2)$ are known. This conditional law is a Gaussian of mean $m_1 = \frac{(t_2-t_1)X(t_0)+(t_1-t_0)X(t_2)}{t_2-t_0}$ and variance $\sigma_1^2 = \frac{(t_2-t_1)(t_1-t_0)}{t_2-t_0}$. Thus one can construct a conditional linear interpolation by writing

$$X(t_1) = [X(t_1) - E(X(t_1)/X(t_0), X(t_2))] + E(X(t_1)/X(t_0), X(t_2))$$

and exploiting the independence of the two terms in the sum, the first having variance σ_1^2 and the second having the mean m_1 given above.

It is clear that if the two successive time intervals defined by the three values of t above have the same length, then the difference between $X(t_1)$ and its conditional expectation given $X(t_0)$ and $X(t_2)$ is a Gaussian variable of standard deviation $\sqrt{2}$ times smaller than the conditional expectation given $X(t_0)$ only (i.e. a simple extrapolation). This suggests good properties of convergence: given $X(t_0)$, one may determine $X(t_1)$, then, by successive interpolations, $X(\frac{1}{2})$, $X(\frac{1}{4})$ and $X(\frac{3}{4})$... and continue dyadically. Let $(X_n(t))_{n \geq 1}$ be the sequence of continuous functions obtained by this series of interpolations, and δ_n the maximum of $|X_n(t) - X_{n+1}(t)|$. Due to the Gaussian character of this difference, one finds $P(\delta_n > (\frac{1}{\sqrt{2}})^{n+2}y) \leq \frac{2^{n+1}}{y\sqrt{2\pi}}e^{-y^2/2}$. Letting $y = c\sqrt{2n \ln 2}$ with $c > 1$, one sees that this probability is a term in a convergent series. Lévy concluded that

according to Cantelli's lemma (or Borel's, when the δ_n are independent),⁹⁴ there exists almost surely a finite number N such that, for $n > N$, one has $\delta_n < c \frac{1}{(\sqrt{2})^{n+1}} \sqrt{n \text{Log } 2}$, which establishes the almost sure uniform convergence to an evidently continuous limit $X(t)$ on $[0, 1]$. It is easy to check that $X(t)$ obtained this way verifies all the conditions of the theoretical definition...

Lévy's proof is an exemplary moment in the history of the theory of probability. It is the most direct and most probabilistic way to both the construction of Brownian motion and the continuity of the trajectories, i.e. to standard Brownian motion. It also embodies Lévy's understanding of randomness and it combines Lévy's entire

⁹⁴The correspondence makes clear that when using the Borel-Cantelli lemma, Lévy always emphasized the advantage of Borel's result over Cantelli's in the independent case.

and unvarying notion of stochastic processes: the influence of randomness at every instant, and its incorporation in a concrete, constructive algorithm. This proof is extensively discussed in the treatise of Blanc-Lapierre and Fortet (1953) intended for engineers and physicists, where it sets the stage for introducing the most useful properties of Brownian motion. It is also the origin of practical algorithms. Lévy's construction is no doubt the only one possible in a work combining mathematical theory with the applications, calculations, and models of engineers and physicists. Itô, who read this proof in Lévy (1948), wrote in 1965 in the first pages of Itô (1965) that he only retained two formulations of standard Brownian motion, Wiener's and Lévy's, and that each had its advantages. In 1940 Lévy not only put his formulation at the start of Lévy (1940), but drew many consequences for Brownian motion from it. He certainly experienced a sense of victory, but later in life it was mixed with remorse: in 1970 he wrote of his regret "to have left to Wiener the discovery of the function $X(t)$," adding "I think, incidentally, that my manner of defining $X(t)$ for Brownian motion is clearer than Wiener's."

Following his insightful work on linear Brownian motion and Brownian motion in a plane (which, because of its singular nature, requires special treatment), Lévy continued his research, investigating Brownian motion with several parameters and Brownian motion with values in a Hilbert space. He also considered Brownian motion in various geometries (on a Riemann sphere, using the work of Francis Perrin and even to some extent in hyperbolic space). Loève (1973) notes the surprise with which, after the war, Lévy discovered the deterministic nature of Brownian motion in a Hilbert space with a parameter; Lévy mentions this to Fréchet in the Letter 58.

At the Second Berkeley Symposium on Probability and Statistics in 1950, shortly before his 65th birthday, Lévy announced to applause his discovery of the Lévy area law (Sect. 8). This victory came after fifteen years of effort by the then incontestable master of Brownian motion. From 1934 to 1950 practically all the advances in the subject were due to Lévy: characterizations, detailed studies of trajectories, and analysis of the geometrical properties of the trajectories. In addition Lévy made innumerable interesting remarks, later amplified. It is clear why Brownian motion imbued his correspondence with Fréchet during these years.

8 Lévy and Stochastic Integrals

The stochastic integral is a major theme in the correspondence of Lévy and Fréchet, especially during the Second World War. The following brief history of stochastic calculus from 1928 until the end of the 1950s is limited to those developments explicitly or implicitly referred to by Lévy in his letters.

Between 1934 and 1950 Lévy invented, investigated and used various types of stochastic integrals. In 1950 he showed that many processes could be represented in the form of an integral with respect to Brownian motion, and then studied these representations. In his 1964 Notice sur les travaux scientifiques de Paul Lévy in support of his election to the Academy of Science, Lévy attached the greatest importance to a certain type of integral that in 1939 had allowed him to define the

area determined by an arc of Brownian motion and its chord (Sect. 7). It was this kind of integral that he wished had been denoted *stochastic*, and in Lévy (1970) he regrets that this did not come to pass. The usage established after the Second World War is that an integral $\int f dg$ is called stochastic if at least one of the two functions f or g is random, and if it satisfies certain regularity properties. The construction, study and interpretation of these integrals is closely tied to the need for a stochastic calculus suitable for applications (especially in statistics and physics). However purely mathematical considerations arising from both the extension of functional analysis to random functions and the needs of probability theory itself played an essential role in the creation of various versions of stochastic calculus. For Lévy, as can be seen in his exchanges with Fréchet, an integral $\int f dg$ should only be called stochastic if it can be defined on the individual trajectories of Brownian motion as the stochastic limit of Stieltjès sums estimated at random times, a point of view that Lévy expressed in his correspondence. Thus, Lévy's stochastic integrals appear to the modern reader as doubly stochastic, and it should come as no surprise that the integral $\int f dg$ which defines the Lévy area has a meaning on the trajectories. There is no paradox; the Lévy integral is not the Itô integral!

We begin our account in 1928, when Slutsky constructed a whole class of stochastic processes for the first time, those that today are called second order processes. He clarified the possible types of convergence of probabilities and elucidated different notions of continuity, differentiability, and stochastic integrability. In fact the terms stochastic derivative and stochastic integral have their origin in Slutsky's work.⁹⁵ For second order processes, it is natural that convergence in quadratic mean plays a privileged role. Slutsky's stochastic integral is thus of the form $\int x(t) dt$ for $x(t)$ continuous in quadratic mean. Integrals of this type, defined by convergence in quadratic mean of Riemann or Riemann-Stieltjès sums, have been called Riemann-Stieltjès stochastic integrals since the 1950s (Blanc-Lapierre and Fortet 1953), with $\int x(t) dt$ defined as the limit in quadratic mean of Riemann sums. In 1937 Slutsky gave more general sufficient conditions on the trajectories that allow a definition of the integral.

Wiener is given credit for a first approach to stochastic integration with respect to Brownian trajectories—even though he never used the term—for square integrable non-random functions. Doob and Itô traced the origin of this integral to *Differential Space*, Wiener's seminal article of 1923 laying out his original theory of Brownian motion (Sect. 7). However its explicit introduction came in 1933 and 1934 in Wiener's *Notes on Random Functions*⁹⁶ and Chap. IX of his book *Fourier Transforms in the Complex Domain*,⁹⁷ and represents the major break between ordinary

⁹⁵In his important note Slutsky, Eugène (1928) Sur les fonctionnelles éventuelles continues et dérivables dans le sens stochastique. CRAS 187:878–880, Slutsky defined and constructed “regular random functions,” i.e. the processes called second order today. He showed that a quadratic mean existed as well as an integral which he wrote as $(S) \int_a^b x_t dt$ and called the *stochastic integral*.

⁹⁶Wiener, Norbert (1933). Math. Z. 37:647–668.

⁹⁷Wiener, Norbert (1934) Colloq. Publ.-Am. Math. Soc. XIX. Most of this text was written with Raymond Paley who died in a skiing accident in 1933.

and stochastic integration. If neither f nor g are of bounded variation—and in 1923 Wiener had shown that Brownian trajectories are not of bounded variation—the integral $\int f dg$ has no possible interpretation in Lebesgue-Stieltjes theory.⁹⁸ Wiener dealt with this problem, giving in the process two successive but equivalent definitions of stochastic integration. The 1933 version begins with the class of functions f of bounded variation and defines its integral with respect to g via integration by parts, then extends the definition to the class of continuous functions. The 1934 version, in the style of Wienerian folklore, uses the Paley-Wiener series (Sect. 7). The formal primitives of that trigonometric series have good convergence properties and permit one, via integration by parts, to obtain integrals of trigonometric polynomials. An extension to L^2 using Hilbert space techniques then follows. These techniques are common in mathematical analysis today, but at the time were rarely used in probability theory, which was not yet completely embedded in the framework of measure theory. Wiener had therefore to make use of the theory of Rademacher functions.

Wiener showed that the limiting object, defined either way, exists on almost all trajectories. In both of these versions, $\int f dg$ (with $f \in L^2$ and g Brownian motion) is nothing but a random function according to Wiener's theory, satisfying the requirements of Einstein and Smoluchowsky (Sect. 7). Its construction is thus strongly influenced by statistics and physics and can be understood as an isometry between the space L^2 and the Gaussian space of Brownian increments.

As of the years 1942 to 1944 Itô originated stochastic calculus with respect to Brownian motion. This work became the basis of theories of integration with respect to martingales. Itô's 1944 article *Stochastic Integral* remains the most well-known of the period. Its only references are to Doob, Lévy and Wiener. In a subsequent letter to Lévy, Itô explained that his goal had been to give an integral sense to the differential expression $dX(t) = A(t, X(t))dt + B(t, X(t))dg(t)$ (where g is a Brownian motion). Itô and his colleague Kakutani, following Doob, put themselves squarely in Kolmogorov's camp, contrary to Lévy (Sect. 3). Itô's formula

$$\int_0^t F'(g(\tau, \omega)) dg(\tau, \omega) = F(g(t, \omega)) - F(g(0, \omega)) - \frac{1}{2} \int_0^t F''(g(\tau, \omega)) d\tau$$

figures explicitly in *Stochastic Integral*; Itô considered it as expressing the essential difference between ordinary and stochastic integration.

The construction of Itô's stochastic integral can be described as follows. Riemann sums define the integral of simple processes defined on subdivisions of $[0, 1]$ that before the time t ($0 < t < 1$) are independent of the Brownian increments after t . Closure of these simple processes under the norm $L^2([0, 1] \times \Omega)$ gives the integral of processes $L^2([0, 1] \times \Omega)$, independent of the time t of the Brownian increments after t . By weakening the topology, Itô was finally able to extend the integral to the class of processes that are sufficiently close in probability to the preceding class. Itô's integral clearly makes no sense on individual trajectories; Lévy subsequently

⁹⁸See Sect. I.8, p. 43 ff, *Naïve Stochastic Integration is Impossible* in Protter (1990).

called it a global procedure by contrast to his integral, which may be defined on trajectories.

With the resumption of relations between Japanese scholars and the international community after the war, Doob made note of the extraordinary versatility of Itô's integral. Doob's remark that $g^2(t) - t$ is a martingale together with the introduction of classical conditional expectation values in the theory of martingales opened the way for the modern theory of stochastic integration. Lévy was at the origin of the theory of martingales⁹⁹ and recognized their appearance in Brownian motion. It is therefore puzzling that he did not even try to use them to attack stochastic integration theory (see Dellacherie 1980), and sometimes found himself among those resisting this approach.

The first contact between Lévy and Itô was an exchange of letters in April 1954. Decades later, after receiving a prize from the Kyoto Academy in 1998, Itô described the difficulties that confronted him and the extent of his isolation during the long years when no one showed interest in his research. He greatly admired Lévy, who motivated his work and inspired his espousal of probability at the start of the war. When an article by Lévy arrived in Japan¹⁰⁰ Itô was in awe. However he was completely ignorant of Lévy's integrals, the essential account having been published in an American journal in 1940.¹⁰¹ He only found out about them with the appearance of Lévy (1948) which gave a rather abbreviated account of Lévy's work on stochastic integrals during the war and was never completed, Lévy being absorbed by questions about Markov processes (Sect. 4). In 1934, as Lévy began his reformulation of Brownian motion (Sect. 7), the first stochastic integrals appeared, which in 1934 Lévy named 'integrals whose elements are independent random variables.'¹⁰² Lévy's book (Lévy 1937) brought notice of his integrals to a much broader circle than his habitual audience. It contained equations of the form $X(T) = \int_0^T \xi_t \sqrt{dt}$ and $Y(T) = \int_0^T \xi_t \sqrt{d\sigma^2(t)}$, where the positive function σ^2 is non-decreasing in t and the ξ_t are independent Gaussians. These were extraordinary, due to the presence of $\sqrt{dt}$ ¹⁰³ and became somewhat notorious, even being attributed to some eccentricity of Lévy's (see, especially, the remarks of Schwartz¹⁰⁴). In fact, they are

⁹⁹On this topic see Mazliak (2009).

¹⁰⁰According to certain accounts, including Itô's notices of the American Mathematical Society ((1998) 45:1455), Lévy's papers arrived via Romania.

¹⁰¹Lévy's American article, in which his stochastic integral is defined, is included in the bibliography of an article by Kakutani (Kakutani, Shinzo (1944) On Brownian motion in n -space. Proc. Imp. Acad. (Tokyo) (November 20)). Lévy's results apparently arrived in Japan from the USA before Pearl Harbor.

¹⁰²Lévy, Paul (1934) Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. Ann. Sc. Norm. Super. Pisa 3:337–366.

¹⁰³The notation $\sqrt{dt}$ had in fact been introduced by Lévy in various publications and appeared in Lévy (1937) where it appears in a note to the chapter entitled *Les intégrales à éléments aléatoires indépendants*.

¹⁰⁴Number 157–158 of *Astérisque* (1988).

already implicitly present in the elements of a Wiener space W . Decomposing additive processes Lévy showed that the continuous component is Wiener Brownian motion. Eliminating components of Poisson type and the various jumps, a Brownian component may be recovered by a procedure closely related to Lévy's 1934 study of random series with independent terms or with terms forming a Markov chain. Sums of weighted increments $\Delta X = \xi_t \sqrt{\Delta t}$ on nested subdivisions with step $\sqrt{\Delta t}$, when properly reordered into the terms of a series, almost surely approach a limit $X(T)$ independent of the choice of the subdivisions. A generalization leads immediately to $Y(T)$, a Gaussian process with independent increments of variance σ^2 , which Lévy called a generalized Brownian motion. The time change $\tau = \sigma^2(t)$ then connects $Y(T)$ to a standard Wiener-Lévy process. Lévy had followed Bernstein's work on diffusion during the 1930s. In particular he knew very well the connections between the equation $dZ(t) = A(t, Z(t))dt + B(t, Z(t))\xi_t \sqrt{dt}$ satisfied by a process Z and the diffusion equation satisfied by its probability density $\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 (B^2 u)}{\partial x^2} - \frac{\partial (Au)}{\partial u}$. In Lévy (1948), enriched by his deliberations during the war, he again took up his work on diffusion and stochastic differential equations. His principal sources were two memoirs by Bernstein¹⁰⁵ and certain of Kolmogorov's results.

Lévy's approach to the method of Cauchy-Bernstein for solving stochastic differential equations required that the convergence of the Cauchy approximations not only occur in distribution, as for Bernstein, but also with respect to spatial topologies important for probability calculus, especially almost sure convergence. The connections between stochastic differential equations, diffusion equations, and stochastic integrals are not so easy to detect in Lévy (1948) because these topics arose in various places, as they were needed. Skorokhod's 1961 book¹⁰⁶ recognizes Lévy as the first to elucidate the methods of Bernstein for differential equations and to be able to study the trajectories of the solutions. In this context an integral of a non-random function $B(t)$ with respect to a Brownian motion $X(T) = \int_0^T B(t)\xi_t \sqrt{dt}$ may be recast as a solution of the stochastic partial differential equation $dX(t) = B(t)\xi_t \sqrt{dt}$. Further, the equation $\frac{\partial u}{\partial t} = \frac{1}{2} B^2(t) \frac{\partial^2 u}{\partial x^2}$ shows that this X is an additive process that becomes Brownian motion¹⁰⁷ on redefining the time via $\tau(t) = \int_0^t B^2(s) ds$.

At the end of the 1930s Lévy had little use for Wiener's integral, which remained poorly understood by probabilists and considered by many analysts as useful only for constructing counter-examples in the theory of trigonometric series. As December 1938 drew to a close Lévy immersed himself in a study of the geometry of the planar Brownian curve. Between 1939 and 1941 he produced six articles on what he called in Lévy (1940) "stochastic integrals of an entirely new type." Lévy's view

¹⁰⁵Bernstein, Serge (1934) *Principes de la théorie des équations différentielles stochastiques*. Труды Физического Математического Института Стеклова 5:95–124 and Bernstein, Serge (1938) *Equations différentielles stochastiques*. Hermann, Paris.

¹⁰⁶Скороход Анатолий В. (1961) *Исследования по теории случайных процессов*. Киевский Университет, Киев. English translation (1965): *Studies in the theory of random processes*. Addison-Wesley.

¹⁰⁷Lévy (1948, pp. 71–72).

of Brownian motion was built around scale invariance and stochastic self-similarity. Starting at the beginning of 1939 Lévy used these fractal properties and stochastic measures to establish that the area measure of the curve is almost surely zero. He believed thus that the area defined by the arc of the curve and its cord should exist in some sense, similar to the area defined by those extraordinary curves without tangents that had fascinated him since his youth. He constructed entire classes of fractals, whether stochastic or not, that served as models for Brownian motion and reproduced some of its properties. He calculated the characteristic functions of the corresponding Lévy areas. Some of his examples have entered fractal folklore (Lévy flights, Lévy dust, Lévy dragon, etc.)—see Mandelbrot (1975). The bends and loops imposed by randomness on the trajectories of Brownian motion down to the smallest distances imply that defining the area by the usual Green-Riemann formula $A(t) = \frac{1}{2} \int_0^t \{X dY - Y dX\} = \int_0^t X dY$ where X and Y are the coordinates of the two dimensional Brownian motion makes no sense in the usual theory of integration. To circumvent this problem, Lévy introduced another independent randomness in the time scale with a probability P' . The corresponding Stieltjes sums $S_n = \sum \frac{f(t_{i+1}) + f(t_i)}{2} [g(t_{i+1}) - g(t_i)]$ are calculated on nested random subdivisions of the integration interval $[a, b]$ with a random step tending P' -almost surely to zero. Between 1939 and 1940, in several stages and with great finesse, Lévy proved the following assertions:

- (a) Given a series $\{t_n\}$ that is dense in the integration interval $[a, b]$ and with $t_0 = a$, the series of Stieltjes sums $\{S_n\}$ of the form given above converges almost surely in quadratic mean to a limit S independent of the series $\{t_n\}$.
- (b) Random independent choices of the points t_n , each following a uniform distribution on $[a, b]$ will lead almost surely to a dense sequence in $[a, b]$.
- (c) In the computations, one may exchange the probability P' used for the time scale and the probability P for the trajectories by application of Fubini's theorem. This theorem had always been emphasized by Lévy as a fundamental tool for the study of random processes.

The formal integral that gives the Lévy area is thus defined on P -almost every trajectory with a P' -almost sure limit. In Lévy's articles from 1938 to 1941 stochastic integrals of an entirely new type are defined by a series of areas of polygons with vertices chosen randomly on non-random curves. It is clear that this construction corresponds to the constructions leading to limiting objects which are among the curves without tangents of which each part is similar to the whole, i.e. that possess self-similar properties. This early vision of fractals may be found in a remarkable article of 1941¹⁰⁸ in which Lévy compares his integrals with the features that one can observe with more and more powerful microscopes.

During the Second World War and up to 1954, Lévy worked at extending his method of stochastic integration. He defined stochastic integration for non-random functions by applying the 0–1 law to the superior and inferior limits of random

¹⁰⁸Lévy, Paul (1941) *Intégrales stochastiques*. Bull. Soc. Math. Fr.-Sud-Est 67–74.

Stieltjès sums regarded as asymptotic events. In Lévy's parlance (which he even used in his notes to the CRAS), a function is *stoch-int* if the two limits exist and are equal with probability one. If f is Lebesgue integrable, it is also *stoch-int* in Lévy's sense. In addition Lévy compared the range of application of his stochastic integrals of the form $\int f dg$, with f and g continuous and non-random, with that of Young's integral and of various non-stochastic extensions of Stieltjès integrals. Lévy also differentiated various modes of convergence of the Stieltjès sums defining his integrals, with the terminology *stoch-int a.s.*, *stoch-int Pr* (for convergence in probability), etc. The Lévy area thus appears on almost every Brownian trajectory as a Stieltjès integral that is stochastic in Lévy's sense. Encouraged by this result, Lévy expected that integrals $\int f dg$, where both f and g are random processes, would be stochastically well-defined in his sense on almost every trajectory.

Lévy's work on Gaussian processes, which he presented in 1950 in Berkeley, used an integral (that he was careful not to call stochastic) of the form $\int_0^t F(t, u) f_u \sqrt{du}$, later rewritten as $\int F(t, u) dY(u)$ where $Y(u)$ encodes Brownian motion. This integral is defined as the limit of a Riemann sum on partitions of the integration interval, the sum being a Laplacian variable well-defined by its covariance matrix. The limit is thus Gaussian and its covariance is obtained from the kernel F . In 1950 Lévy also gave an interpretation of the Lévy area using a generalization of Wiener's integral. Later developments in stochastic calculus showed, however, that it was Itô who had uncovered the Brownian properties which best allowed an extension of stochastic integration by using martingales.

In 1954, at a time when he would have no further occasion to work on stochastic integration, Lévy gave his personal view of its history and summarized his own theory, which he never ceased to defend:

Our stochastic integrals have nothing in common with what various authors, since Slutsky, refer to under that name. Ours is a definition that applies to individual functions. We use it, in particular, for each instance of a random function. Such a definition, to our knowledge, cannot be found before 1940 except in Paley-Wiener. . . . However, these authors start with the representation of a complex function $g(t) = Z(t)$ by the Fourier-Wiener formula. For the general theory, it is interesting to start instead with the intrinsic definitions of $f(t)$ and $g(t)$ and Riemann-Stieltjès sums. Since 1944, a theory that uses Riemann sums has been developed by K. Itô,¹⁰⁹ who replaces our sum $\sum \frac{f(t_{i+1}) + f(t_i)}{2} [g(t_{i+1}) - g(t_i)]$ by $\sum f(t_i) [g(t_{i+1}) - g(t_i)]$ and does not attempt to find a definition applicable to individual functions.

Thus, although Lévy's name is nowadays not immediately associated with stochastic integration, there is abundant evidence of his deep and protracted involvement in the development of that field.

¹⁰⁹In his seminal article (Itô, Kyoshi (1944) Stochastic integrals. Proc. Imp. Acad. (Tokyo) 20), Itô integrates certain simple processes with respect to Brownian motion. These Riemann sums were the object of an exchange of letters between Itô and Lévy in April 1954. Itô's construction, especially useful in the space L^2 , was later employed for the development of a theory of stochastic integration with respect to martingales. See Dellacherie (1980).

Paris - 29 décembre 1918
9 rue Chernoviz

Monsieur et cher Collègue,

Ayant eu votre adresse par M. Hadamard, je me permets de vous demander s'il vous serait possible de me faire parvenir un exemplaire de votre mémoire sur une extension de la méthode de Jacobi Hamilton. Il me serait utile pour la préparation du cours Peccot, que j'espère faire cette année.

Les *Annales de Mathématique* sont peu répandues en France, et étant mobilisé, il ne m'est pas facile d'aller travailler à la bibliothèque de la Sorbonne, qui ne prête pas les périodiques. Pous m'excuserez de m'adresser à vous dans ces conditions.

Progez, je vous prie, à mes sentiments dévoués

Paul Lévy

ANNUAIRE
D'ARCHIVAGE
MATHÉMATIQUES

Fig. 1 First letter from Lévy to Fréchet (29 December 1918). (Photo: © Académie des Sciences, Paris)

Paris - 38 Av. Th. Gautier
19 - 10 - 65

Mon cher collègue et confrère,

La Commission du tricentenaire s'est réunie hier après le Comité secret (jusqu'à 18 h. 15 !). Il a été décidé que chaque Section devrait prendre la responsabilité du travail fait en son nom.

Montel, Garnier et moi avons tenu une réunion préliminaire. Denjoy est dans le Midi. Je prévient Julia, qui voudra peut-être nous réunir le mercredi 27 après la séance (renvoyée au mercredi à cause de la séance de l'Institut du lundi). Je crois bien faire de vous prévenir de cette éventualité.

J'ai dû exposer longuement à Julia les critiques auxquelles mon travail a donné lieu. Il est trop long, et je n'aurais pas dû être les vivants ! Excusez-moi de ne pas vous donner plus de détails. Il est trop tard pour tout recommencer.

Bien cordialement

P. Lévy

ACADEMIE DES SCIENCES
ARCHIVES
1910-1965

Fig. 2 Letter from Lévy to Fréchet (19 October 1965). (Photo: © Académie des Sciences, Paris)

Fig. 3 Maurice Fréchet in 1910. (Photo: © Private collection F. Lederer)



Fig. 4 Suzanne and Maurice Fréchet and their children Hélène and Henri in 1915. (Photo: © Private collection F. Lederer)



Fig. 5 Paul Lévy in 1913.
(Photo: © Private collection
D. Piron)



Fig. 6 Paul Lévy in 1933.
(Photo: © Private collection
D. Piron)



Fig. 7 Suzanne and Paul Lévy in 1952. (Photo: © Private collection D. Piron)



Fig. 8 Maurice Fréchet in 1942. (Photo: © Private collection F. Lederer)



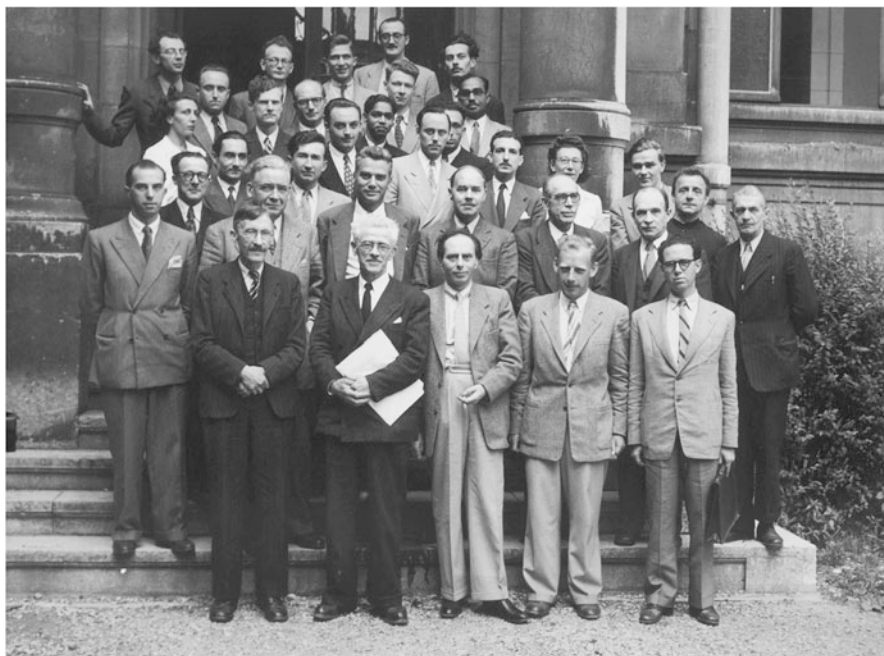


Fig. 9 Colloque International sur le Calcul des Probabilités, Lyon 1948. First row: Paul Lévy and Maurice Fréchet. On the picture one can find among others J. Doob, R. Fortet, D. Van Dantzig, E. Mourier, J. Kampé de Fériet, A. Blanc-Lapierre. . . . (Photo: © Private collection F. Lederer)

Paul Lévy and Maurice Fréchet

50 Years of Correspondence in 107 Letters

Barbut, M.; Locker, B.; Mazliak, L.

2014, XVII, 213 p. 14 illus., 2 illus. in color., Hardcover

ISBN: 978-1-4471-5618-5