

## Some aspects of the probabilistic work

Loïc Chaumont<sup>1</sup>, Laurent Mazliak<sup>2</sup>, and Marc Yor<sup>3</sup>

<sup>1</sup> Laboratoire Angevin de Recherche en Mathématiques (LAREMA), University of Angers, France

<http://math.univ-angers.fr/~chaumont>

[loic.chaumont@univ-angers.fr](mailto:loic.chaumont@univ-angers.fr)

<sup>2</sup> Laboratoire de Probabilités et Modèles Aléatoires, Institut de mathématiques de Jussieu, University Paris VI, France

<http://www.proba.jussieu.fr/users/lma/mazliak.html>

[mazliak@ccr.jussieu.fr](mailto:mazliak@ccr.jussieu.fr)

<sup>3</sup> Laboratoire de Probabilités et Modèles Aléatoires, Institut de mathématiques de Jussieu, University Paris VI, France

[http://www.academie-sciences.fr/Membres/Y/Yor\\_Marc.htm](http://www.academie-sciences.fr/Membres/Y/Yor_Marc.htm)

Translated from the French by Kathleen Qechar

### 3.1 Introduction

Anyone reading about the mathematical works of Kolmogorov must naturally expect to find broad considerations about the axiomatization of probability which Kolmogorov developed at the beginning of the nineteen thirties and which forms the contents of his famous publication *Grundbegriffe der Wahrscheinlichkeitsrechnung* (*Foundations of the theory of probabilities*) [Kol33], published by Springer in 1933. It is certain that among all the works of the Soviet mathematician, this small opusculé of about sixty pages is the most famous part, and is often the only one to which his name is attached for a quite large public but also for some mathematicians. Without wanting in any way to diminish the importance of this work, it is nevertheless quite astonishing that the attention was thus focused on what does not constitute the most original creation of Kolmogorov in the field of probability. The aim of this part, devoted to certain aspects of the probabilistic works of the scientist, is precisely to highlight some of his most remarkable works in that domain. In this imposing monument, a drastic choice was necessary and we chose to focus on the two purely probabilistic directions that Kolmogorov worked on, namely on the one hand the study of the various types of convergence for sums of independent random variables, which enabled him to continue the studies of his Russian predecessors Markov and Lyapounov, and on the other hand literally revolutionary considerations about processes in continuous time, whose

branches extend ahead in time until some discoveries which go back to hardly thirty years. Nevertheless, as we found it difficult, and almost impossible, that a chapter devoted to the probabilistic works of the Soviet mathematician does not refer to the axiomatization of probability, we will begin with a short glance of his main contributions to this topic, inviting the reader to refer to the numerous articles dealing with the question in a more detailed way (see e.g. [vonP94], [SV06]). We will also refer to the essential text of Shiryaev [Shi89] for a more complete chart of Kolmogorov's works. One will find in the article [Maz03] some indications on the life of the mathematician and the status of the discipline in the stalinist USSR.

## 3.2 The axiomatization of probability calculus

### 3.2.1 An abstract framework

As mentioned above, Kolmogorov's publication *Grundbegriffe der Wahrscheinlichkeitsrechnung* [Kol33] is a modest monograph of 60 pages published in 1933 along with several articles devoted to the modern probability theory. The Russian translation of the text is dated 1936, and it was mostly achieved due to some political reasons at the time when an important pressure was put on Soviet scientists so that they publish their works in Russian and in USSR rather than abroad. As for the first English translation, it is dated 1950. This relatively important delay shows that the axiomatization suggested by the Russian scientist wasn't as generally accepted as we usually think it was. Several probabilists, and among the most eminent ones, such as Paul Lévy, will never use the axiomatization of Kolmogorov, which will not prevent them in any way having extraordinary ideas. In fact, outside the USSR, before the 50's, more or less only Cramér's treatise [Cra37] refers to this field. Besides, this author doesn't give any detailed explanation; he only uses Kolmogorov's axiomatization because it is the most practical one among those available at that time (in particular the *theory of collectives* suggested by von Mises). However, from the 50's, it will definitely be adopted by the younger generation. What is attractive in the formal framework proposed by this axiomatization, is the fact that it provides e.g. a global explanation of the multiple paradoxes which had in the past plagued this discipline (like those of Joseph Bertrand, Emile Borel etc.) : each time, the precise definition of the probability space as a description of the considered random experiment allows to suppress the ambiguity. (On this subject, see *infra*, as well as [SV06] and [Szé86]. One can also refer to Itô's comments in the foreword of [Itô86].)

The great force of Kolmogorov's treatise is to voluntarily consider a completely abstract framework, without seeking to establish bridges with the applied aspects of the theory of probability, beyond the case of finite probabilities. In general, the search for such bonds inevitably brings to face delicate philosophical questions and is thus likely to darken mathematical modelling.

While speaking about the questions of application only in the part devoted to the finite probabilities<sup>4</sup>, Kolmogorov is released from this constraint and can avoid the pitfalls that von Mises had not always circumvented. Indeed, the *theory of the collectives* also claimed to establish a discrimination between the experiments for which the application of the probabilities was legitimate, and others. But Kolmogorov, who presents a purely mathematical theory, does not have such an ambition, and thus not such a limitation. Within the abstract framework which he defines, any mathematical work is legitimate, and its validation for applications is a matter for other fields of knowledge. In particular, he allows himself to consider sets not having any topological structure, whereas for finer studies (like the phenomena of convergence), he will be free to work on better spaces through the use of images of probability laws. Besides this fact will place the axiomatization of Kolmogorov in opposition with the Bourbaki topological set-up concerning measure theory. The very general character of his theory will make it possible to the Russian mathematician to use in all its force the measure theory of Borel and Lebesgue, which is still relatively new at that time since its abstract version was mainly developed by Fréchet (quoted in the *Grundbegriffe* as the one who liberated measure theory from geometry) then by the Polish school (Banach, Sierpiński, Kuratowski...) in the 20's.

From the beginning of the twentieth century, Borel had been a promoter of the use of the measure theory and integral of Lebesgue for the treatment of questions of probability. In 1909, he publishes a revolutionary paper where such a method enables him to obtain a first strong version of the law of large numbers and interpretations on the distribution of real numbers. Undoubtedly his moderate consideration for probabilistic mathematics and serious doubts as for the legitimacy of their applications prevented him from fully reaping the crops from the seeds which he had sown.

Kolmogorov introduces the by now classical concept of a probability space in the form of a triplet  $(\Omega, \mathcal{F}, P)$  composed of a set  $\Omega$  provided with  $\sigma$ -algebra (which he calls a *set field*)  $\mathcal{F}$  and a normalized measure (probability)  $P$ . The random variables are simply functions  $X$  with real values defined on  $\Omega$  such that for all  $a \in \mathbb{R}$ ,  $\{\omega \in \Omega, X(\omega) < a\} \in \mathcal{F}$  and their laws are the image measures of the probability  $P$  defined by  $P^{(X)}(A) = P(X^{-1}(A))$ , for all  $A \in \mathcal{B}(\mathbb{R})$ , the Borelian  $\sigma$ -algebra of  $\mathbb{R}$ .

The major contributions of Kolmogorov's work in the clarification of probabilistic concepts are incontestably the construction of a probability measure on an infinite product of spaces, which plays an important part in the theory of stochastic processes, and the formalization of the conditional law via the use of the Lebesgue-Radon-Nikodym theorem (the abstract version of this theorem had been published by Nikodym in 1930). Let us note incidentally

---

<sup>4</sup> Which is not without reminding one about the way in which, in his article of 1931 on Markov processes, he had proposed in a long introduction to defer to other works some reflections about the applicability of his theories

that it wasn't the first time that a probability on a product space was built: the most famous example is given by Wiener [Wie23] who since 1923, by applying techniques that Daniell had developed a few years before to extend Lebesgue's integral to spaces of infinite dimensions, built the probability measure – now called Wiener's measure – associated with Brownian motion (see [RY91]).

### 3.2.2 Construction of the conditional law

Let us present in a few words the construction of the conditional law, while following the text of Kolmogorov but with modernized notations for the sake of clarity of our exposition.

First of all let us recall the definition of the elementary conditional probability of an event  $C$  (i.e. of an element of the  $\sigma$ -algebra  $\mathcal{F}$ ) by an event  $D$  such as  $P(D) > 0$ , by  $P(C \mid D) = P(C \cap D)/P(D)$ . Now, let  $U$  denote a real-valued random variable and  $B$  an event. We try to build a random variable  $\omega \mapsto \pi(U(\omega); B)$ , a Borelian function of  $U$ , so called conditional probability of  $B$  knowing  $U$ , and such that, for all  $A \in \mathcal{B}(\mathbb{R})$  with  $P(U \in A) > 0$ , we have:

$$P(B \mid U \in A) = \int_{\Omega} \pi(U; B) dP(\cdot \mid U \in A).$$

For any  $A \in \mathcal{B}(\mathbb{R})$ , we write  $Q_B(A) = P(B \cap U^{-1}(A))$ . Let us note that if  $P^{(U)}$  is the law of  $U$  defined by  $P^{(U)}(A) = P(U^{-1}(A))$ , then  $P^{(U)}(A) = 0$  implies  $Q_B(A) = 0$  and thus, according to the Lebesgue-Radon-Nikodym theorem, we can find a Borelian function  $f_B$  such that  $\forall A \in \mathcal{B}(\mathbb{R})$ ,  $Q_B(A) = \int_{\mathbb{R}} \mathbb{1}_A f_B dP^{(U)}$ , i.e.

$$P(B \cap U^{-1}(A)) = \int_{\mathbb{R}} \mathbb{1}_A f_B dP^{(U)} = \int_{\Omega} \mathbb{1}_{U \in A} f_B \circ U dP$$

and thus

$$P(B \mid U \in A) = \int_{\Omega} f_B \circ U dP(\cdot \mid U \in A)$$

and we set  $\pi(U; B) = f_B \circ U$ . From this point, Kolmogorov will recover all classical properties of conditional probabilities. He nicely illustrates the strength of his formalism by explaining the paradox of Borel related to the random drawing of a point on a sphere: the interested reader can refer to e.g. [Bil95], p. 462 and [SV06].

### 3.2.3 The 0-1 law (or *the all or nothing law*)

As mentioned previously, in 1933, measure theory isn't yet very commonly accepted, in any case not in its abstract form, and when Fréchet will discover the monograph of the Russian mathematician, he will be disconcerted by the

very abstract form taken by certain arguments, like that of the law known as the 0-1 law which Kolmogorov placed in the appendix of his work. This law was stated independently, in particular by Lévy in 1934 (when he didn't yet know the *Grundbegriffe*), and it is interesting to compare the two approaches of this result, which we will do by way of illustration of the strongly synthesizing character of the axiomatic proposed by Kolmogorov. For an easy reading, we will use today's vocabulary and notations, keeping only the spirit of the two proofs.

**Theorem 1.** *Let  $(X_n)_{n \geq 1}$  denote a sequence of independent real-valued random variables.*

*We introduce  $\mathcal{G}_n = \sigma(X_n, X_{n+1}, \dots)$  (the  $\sigma$ -algebra generated<sup>5</sup> by  $X_n, X_{n+1}, \dots$ ), and  $\mathcal{G} = \bigcap_{n \geq 1} \mathcal{G}_n$  (the “tail  $\sigma$ -algebra”).*

*Then, any element of  $\mathcal{G}$  is of probability 0 or 1.*

*Kolmogorov's proof:* It is the most common proof taught today. Let  $A$  an element of  $\mathcal{G}$ . Let us suppose that  $P(A) > 0$  and note  $P_A$  the conditional probability knowing  $A$ . According to the independence hypothesis made on the  $X_k$ 's, for all  $B \in \mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$ ,  $B$  is independent from the elements of  $\mathcal{G}_{n+1}$  and thus from  $A$ , and we have:

$$P_A(B) = \frac{P(A \cap B)}{P(A)} = P(B).$$

Therefore, the probabilities  $P_A$  and  $P$  coincide on all  $\mathcal{F}_n$ 's, and hence on Boole's algebra  $\bigcup_{n \geq 1} \mathcal{F}_n$  and thus, according to the monotone class theorem,

on the  $\sigma$ -algebra thus generated, i.e.  $\mathcal{F} = \sigma(X_1, X_2, \dots, X_n, \dots)$ . Since in particular  $A \in \mathcal{F}$ , we have  $P_A(A) = P(A)$ , i.e.  $P(A) = 1$ .  $\square$

*Lévy's proof:* In fact, Lévy contents himself to prove the result when variables  $X_n$  follow a uniform law on  $[0,1]$ . In this case, the obtention of a realization of the sequence  $(X_n)_{n \geq 1}$  can be conceived as the one of a point in a cube of size 1 with an infinity of dimensions, the law of probability being given by Lebesgue's measure. Lévy's argument is then based on an observation which he affirms being obvious and which is equivalent in fact to a monotone class result: he points out (we employ the modern formalism) that for any event  $A$  of the  $\sigma$ -algebra  $\mathcal{F} = \sigma(X_1, X_2, \dots, X_n, \dots)$  (which may therefore be written as follows  $[(X_1, X_2, \dots, X_n, \dots) \in B]$ , where  $B$  is a measurable set of  $\mathbb{R}^{\mathbb{N}}$ ), and for all  $\varepsilon > 0$ , we can find  $n > 0$  and  $D_n \in \mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$  such that<sup>6</sup>  $P(D_n \Delta A) < \varepsilon$ . In fact, his explanation is to say that measurable sets inside the infinitely dimensional cube are obtained by “*M. Lebesgue's constructions*” from the “intervals” of the cube, which are the sets of the

<sup>5</sup> i.e. the smallest  $\sigma$ -algebra which allows all of them to be measurable

<sup>6</sup>  $\Delta$  indicates the symmetric difference:  $D_n \Delta A = (D_n \cup A) \setminus (D_n \cap A)$

type  $]a_0, b_0[ \times ]a_1, b_1[ \times \cdots \times ]a_n, b_n[ \times ]0, 1[ \times ]0, 1[ \dots$ , in the same way as the Borel sets of  $\mathbb{R}$  are built from the real-valued open intervals; and we know that for any element  $A \in \mathcal{B}([0, 1])$ , there exists a finite collection of intervals  $]\alpha_0, \beta_0[, \dots, ]\alpha_m, \beta_m[$  such that  $\lambda(A \Delta \bigcup_{k=0}^m ]\alpha_k, \beta_k[) < \varepsilon$  (where  $\lambda$  is Lebesgue's measure on  $\mathbb{R}$ ).

Let now  $E$  an element of  $\mathcal{G}$ . Let us note that the independence of the  $(X_n)_{n \geq 1}$ 's allows us to write that  $\forall n, P(E \mid \mathcal{F}_n) = P(E)$ . Let  $N$  and  $D_N \in \mathcal{F}_N$  such that  $P(E \Delta D_N) < \varepsilon$  (and so, in particular,  $P(D_N) > P(E) - \varepsilon$ ). Then, we have:

$$\varepsilon > P(D_N \cap E^c) = P(D_N)P(E^c \mid D_N).$$

However,  $P(E^c \mid \mathcal{F}_N) = 1 - P(E \mid \mathcal{F}_N) = 1 - P(E) = P(E^c)$ , thus  $P(E^c \mid D_N) = P(E^c)$ , hence

$$\varepsilon > P(D_N)P(E^c) > (P(E) - \varepsilon)(1 - P(E)) > P(E)(1 - P(E)) - \varepsilon$$

and so  $P(E)(1 - P(E)) < 2\varepsilon$ . This is true for all  $\varepsilon > 0$ , thus  $P(E)(1 - P(E)) = 0$ .  $\square$

### 3.3 Limit theorems and series of independent random variables

The direction in which Kolmogorov will develop his first works in Probability, undoubtedly guided by his elder Khinchin<sup>7</sup>, may be found to be in continuity with the former studies which had specified the conditions of validity of the limit theorems (in particular the law of large numbers) for sums of random variables throughout the nineteenth century. The first paper, which goes back to 1925 [KK25] and which is the only article jointly written by Kolmogorov and Khinchin, is remarkable in the way that it introduces a number of techniques which will be at the base of some later developments of the theory of probability, in particular in the study of the results of convergence for martingales. This first work relates to the convergence of series of independent random variables. The main result is stated as follows (in modern terms):

**Theorem 2.** *Let  $(X_n)_{n \geq 1}$  a sequence of centered (i.e. of 0 expectation) independent real-valued random variables. Let us suppose that  $\sum_{n \geq 1} E(X_n^2) < +\infty$ .*

*Then  $\sum_n X_n$  converges almost surely (a.s.), i.e. with probability 1.*

The proof suggested by Kolmogorov is based on a famous inequality which is named after him today:

---

<sup>7</sup> One also spells: Khintchine

**Lemma 1 (Kolmogorov's inequality).** *Let  $S_n = X_1 + \cdots + X_n$ . Then*

$$P(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon) \leq \frac{E(S_n^2)}{\varepsilon^2}. \quad (3.1)$$

*Proof.* We write

$$\{\max_{1 \leq k \leq n} |S_k| \geq \varepsilon\} = \bigcup_{p=1}^n A_p$$

where

$$A_p = \{|S_1| < \varepsilon, |S_2| < \varepsilon, \dots, |S_{p-1}| < \varepsilon, |S_p| \geq \varepsilon\}.$$

Note that the  $A_p$ 's form a partition of the entire probability space and that for all  $1 \leq p \leq n$ ,  $S_n - S_p$  is independent from  $S_p \mathbb{1}_{A_p}$  and has 0 expectation. Therefore, for  $1 \leq p \leq n$ ,

$$E(S_n^2 \mathbb{1}_{A_p}) = E((S_n - S_p)^2 \mathbb{1}_{A_p}) + E(S_p^2 \mathbb{1}_{A_p}) \geq \varepsilon^2 P(A_p)$$

and summing with respect to  $p$ ,  $E(S_n^2) \geq \varepsilon^2 P(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon)$ .  $\square$

It is only a few years later, in a note for the *Comptes Rendus de l'Académie des Sciences* (CRAS) of Paris in 1930 [Kol30], that Kolmogorov will obtain from the previous result his most famous consequence, that is to say this nowadays classical version of the strong law of large numbers:

**Corollary 1.** *Let us suppose that the variables  $X_n$  are independent and centered. We write  $E(X_n^2) = b_n$  and we suppose that  $\sum_{n=1}^{\infty} \frac{b_n}{n^2} < +\infty$ . Then*

$$\sigma_n = \frac{S_n}{n} \rightarrow 0, \text{ a.s.}$$

*Proof.* First of all, let us note that for all fixed  $N > 0$ ,

$$\overline{\lim}_n |\sigma_n| = \overline{\lim}_n \frac{|S_n - S_{2^N}|}{n},$$

where  $\overline{\lim}_n$  indicates the superior limit for  $n \rightarrow +\infty$ . Moreover, it is obvious that:

$$\begin{aligned} \{\overline{\lim}_n \left| \frac{S_n - S_{2^N}}{n} \right| > \varepsilon\} &\subset \bigcup_{n \geq 2^N} \left\{ \left| \frac{S_n - S_{2^N}}{n} \right| > \varepsilon \right\} \\ &\subset \bigcup_{m \geq N} \left\{ \max_{2^m \leq k \leq 2^{m+1}} \left| \frac{S_k - S_{2^N}}{k} \right| > \varepsilon \right\} \\ &\subset \bigcup_{m \geq N} \left\{ \max_{2^m \leq k \leq 2^{m+1}} \left| \frac{S_k - S_{2^N}}{2^m} \right| > \varepsilon \right\}. \end{aligned}$$

Therefore, for all  $\varepsilon > 0$  and all  $N > 0$ :

$$\begin{aligned}
P\left(\overline{\lim}_n |\sigma_n| > \varepsilon\right) &= P\left(\overline{\lim}_n \frac{|S_n - S_{2^N}|}{n} > \varepsilon\right) \\
&\leq \sum_{m=N}^{\infty} P\left(\max_{2^m \leq k \leq 2^{m+1}} |S_k - S_{2^N}| > \varepsilon 2^m\right) \leq \frac{1}{\varepsilon^2} \sum_{m=N}^{\infty} \frac{1}{2^{2m}} \sum_{k=2^N+1}^{2^{m+1}} b_k \\
&= \frac{1}{\varepsilon^2} \sum_{i=N}^{\infty} \left(\sum_{m \geq i} \frac{1}{2^{2m}}\right) \sum_{k=2^i+1}^{2^{i+1}} b_k \leq \frac{16}{3} \frac{1}{\varepsilon^2} \sum_{n \geq 2^N} \frac{b_n}{n^2}.
\end{aligned}$$

By hypothesis, the last term can be made as small as we like, and so  $\overline{\lim}_n \sigma_n = 0$  a.s.  $\square$

*Remark 1.* The independence of the  $X_n$  variables occurs twice in this proof: first when we apply Kolmogorov's inequality, and secondly when we write

$$E\left((S_{2^{m+1}} - S_{2^N})^2\right) = \sum_{k=2^N+1}^{2^{m+1}} b_k$$

(using the additivity of the variance for non-correlated variables).

As mentioned above, the result of Lemma 1 and its proof can directly be applied to the case of discrete martingales<sup>8</sup>:

**Corollary 2.** *Let  $(M_n)_{n \geq 1}$  denote a square integrable martingale such that  $E(M_n) = 0$ . Then*

$$P\left(\max_{1 \leq k \leq n} |M_k| \geq \varepsilon\right) \leq \frac{E(M_n^2)}{\varepsilon^2}. \quad (3.2)$$

It is easily proven that (3.2) may be strengthened a little, in the form of the important *Doob inequality*: if  $(M_n)_{n \geq 1}$  is a square integrable martingale such that  $E(M_n) = 0$ , we have

$$E\left[\max_{1 \leq k \leq n} (M_k)^2\right] \leq 4E[M_n^2].$$

As we know, the theory of martingales has, after Doob's works, invaded the scene of contemporary probability theory. In order to illustrate the strength of the inequality (3.2) and the notion of martingale, let us prove the following result and its corollary.

---

<sup>8</sup> Let us recall that a discrete martingale is a sequence  $(M_n)_{n \geq 1}$  of integrable random variables such as, for all  $n$ , the expectation of  $M_{n+1}$ , knowing all the previous values, is equal to the last one:  $E\left(M_{n+1} \middle| M_1, \dots, M_n\right) = M_n$ . (The martingale is called square integrable, or etc. if each  $M_n$  satisfies this property.)



**Proposition 1.** *Let  $(M_n)$  be a  $L^2$  martingale (i.e. a square integrable martingale) such that  $\sup_n E(M_n^2) < +\infty$ . Then  $M_n$  converges, at the same time in  $L^2$  and a.s., towards a random variable  $M$ .*

*Proof.* It is easy to show the convergence in  $L^2$  by proving that  $(M_n)_{n \geq 1}$  is a Cauchy sequence in  $L^2$ . The interested reader can refer to any elementary course on martingales. Let us now deduce the a.s. convergence.

Let  $\varepsilon > 0$ . For each  $p > 0$ , let us write  $V_p = \sup_{n \geq p} |M_n - M_p|$ . As  $(M_n - M_p)_{n \geq p}$  is a square integrable martingale, we have according to (3.2) for all  $N$

$$P\left(\max_{p \leq k \leq N} |M_k - M_p| \geq \varepsilon\right) \leq \frac{E((M_N - M_p)^2)}{\varepsilon^2}. \quad (3.3)$$

As mentioned above,  $(M_k)_{k \geq 0}$  is a Cauchy sequence in  $L^2$  and so, for each given  $m > 0$ , we can choose  $p_m$  such that  $\forall N \geq p_m$ ,  $E((M_N - M_{p_m})^2) \leq \varepsilon^2/2^m$ . Therefore, passing to the limit in (3.3) when  $N \rightarrow +\infty$ :  $P(V_{p_m} \geq \varepsilon) \leq 1/2^m$ . According to the Borel-Cantelli Lemma, a.s. there exists  $m > 0$  such that  $V_{p_m} < \varepsilon$  i.e.  $\forall n \geq p_m$ ,  $|M_n - M_{p_m}| < \varepsilon$ , which is equivalent to saying that a.s.  $(M_n)$  is a Cauchy sequence in  $\mathbb{R}$ .  $\square$

**Corollary 3.** *Let  $(Z_n)_{n \geq 1}$  a sequence of independent random variables with the same Bernoulli law  $P(Z = 1) = P(Z = -1) = \frac{1}{2}$ . We consider the random walk on  $\mathbb{Z}$ :  $S_0 = 0, S_n = Z_1 + \dots + Z_n$ . Let  $a > 0$  an integer and  $\tau = \inf\{n \geq 0, S_n = a\}$  the first passage time of  $a$ . Then the Laplace transform of the law of  $\tau$  is given for all  $\theta \geq 0$  by*

$$E[(\cosh \theta)^{-\tau}] = e^{-\theta a}.$$

*Proof.* We only indicate the broad outline of the proof, leaving the details to the interested reader (see e.g. [BMP01]). Let  $X_n^\theta = \frac{e^{\theta S_n}}{(\cosh \theta)^n}$ . We verify that it is a martingale and that<sup>9</sup>  $(X_{n \wedge \tau}^\theta)_{n \geq 1}$  is a  $L^2$  martingale, which converges a.s. and in  $L^2$  towards the variable  $W^\theta = \frac{e^{\theta a}}{(\cosh \theta)^\tau} \mathbb{I}_{\tau < +\infty}$ . Passing to the limit as  $\theta \rightarrow 0$ , which is made possible by dominated convergence, we obtain that  $P(\tau < +\infty) = 1$ , and thus the desired result.  $\square$

In 1924, Khinchin [Khi24] had proven a result which brought a radical precision to the law of large numbers, the *law of the iterated logarithm*. The generalization of Khinchin's result by Kolmogorov in 1929 ([Kol29]) was one of his greatest achievements.

**Theorem 3.** *Let  $(X_n)_{n \geq 1}$  a sequence of independent real-valued random variables. Let us suppose that for all  $n$ ,  $E(X_n) = 0$  and  $b_n = E(X_n^2) < +\infty$ . We*

---

<sup>9</sup>  $n \wedge \tau$  means  $\min\{n, \tau\}$ , the smallest number of  $n$  and  $\tau$

let  $B_n = \sum_{k=1}^n b_k$  (that is  $B_n = E(S_n^2)$ ). If  $B_n \rightarrow +\infty$  and  $|X_n| \leq M_n = o(\sqrt{\frac{B_n}{\ln \ln B_n}})$ , we have a.s.

$$\overline{\lim} \frac{S_n}{\sqrt{2B_n \ln \ln B_n}} = 1.$$

Today, Kolmogorov's proof still remains very much up to date, as it introduces techniques, in particular of large deviations, which became fundamental in the study of many limiting phenomena in probability theory. We will only show the less technical part of the result, leaving the reader consult one of the innumerable texts which present the complete proof (e.g. [Bil95]).

Let us note, for  $\varepsilon > 0$ ,  $\phi^\varepsilon(n) = (1 + \varepsilon)\sqrt{2B_n \ln \ln B_n}$ ; we shall prove that  $P(S_n \geq \phi^\varepsilon(n), \text{ infinitely often}) = 0$ . According to Borel-Cantelli's Lemma, it suffices to prove that for a well-chosen subsequence  $n_k \uparrow +\infty$ , we have

$$\sum_{k=1}^{\infty} P(\max_{n \leq n_k} S_n \geq \phi^\varepsilon(n_{k-1})) < \infty. \quad (3.4)$$

As mentioned above, we will obtain the result thanks to the following lemma which gives some large deviations estimates for the sequence  $(S_n)$ .

**Lemma 2.** *Let  $x \geq 0$ .*

- (i) *If  $x \leq B_n/M_n$  then  $P(S_n > x) < e^{-(\frac{x^2}{2B_n})(1 - \frac{xM_n}{2B_n})}$*
- (ii) *If  $x \geq B_n/M_n$ , then  $P(S_n > x) < e^{-\frac{x}{4M_n}}$*
- (iii)  *$P(\max_{1 \leq k \leq n} S_k \geq x) \leq 2P(S_n > x - \sqrt{2B_n})$ .*

*Proof.*

Let us fix  $n$  and in order to simplify the writing, this index  $n$  will be omitted in the next lines.

Let  $a > 0$  such that  $aM \leq 1$ . Then,

$$\begin{aligned} E(e^{aX_k}) &= 1 + \sum_{r \geq 2} E\left(\frac{a^r X_k^r}{r!}\right) \leq 1 + \frac{a^2 b_k}{2} \sum_{r \geq 2} 2 \frac{a^{r-2} M^{r-2}}{r!} \\ &\leq 1 + \frac{a^2 b_k}{2} \left(1 + \frac{aM}{2}\right) < \exp\left[\frac{a^2 b_k}{2} \left(1 + \frac{aM}{2}\right)\right], \end{aligned}$$

and thus: 
$$E(e^{aS}) = \prod_{k=1}^n E(e^{aX_k}) < \exp\left[\frac{a^2 B}{2} \left(1 + \frac{aM}{2}\right)\right].$$

As  $P(S > x) < E(\frac{e^{aS}}{e^{ax}})$  (for all  $a > 0$ ), we obtain the inequality  $P(S > x) < \exp[-ax + \frac{a^2 B}{2}(1 + \frac{aM}{2})]$  from which we easily deduce the points (i) and (ii) by taking successively  $a = x/B$  and  $a = 1/M$ .

As for the point (iii), let us note  $U = \max_{1 \leq k \leq n} S_k$ , and that  $(U \geq x)$  is the union of the events  $E_k = (S_1 < U, \dots, S_{k-1} < U, S_k = U \geq x)$  for  $1 \leq k \leq n$ .

Thus, we have

$$\begin{aligned} P(S > x - \sqrt{2B}) &\geq \sum_{k=1}^n P(E_k \cap (S > x - \sqrt{2B})) \geq \sum_{k=1}^n P(E_k \cap (S > U - \sqrt{2B})) \\ &= \sum_{k=1}^n P(E_k) P(S > U - \sqrt{2B} \mid E_k) = \sum_{k=1}^n P(E_k) P(S - S_k > -\sqrt{2B} \mid E_k). \end{aligned}$$

But  $S - S_k$  is independent of  $E_k$ , and therefore this last expression is also

$$= \sum_{k=1}^n P(E_k) P\left(\sum_{i=k+1}^n X_i > -\sqrt{2B}\right) \geq \sum_{k=1}^n P(E_k) P\left(\left(\sum_{i=k+1}^n X_i\right)^2 < 2B\right).$$

Now,

$$1 - P\left(\left(\sum_{i=k+1}^n X_i\right)^2 < 2B\right) = P\left(\left(\sum_{i=k+1}^n X_i\right)^2 \geq 2B\right) \leq \frac{1}{2B} \sum_{i=k+1}^n b_i \leq \frac{1}{2},$$

and hence

$$P(S > x - \sqrt{2B}) \geq \frac{1}{2} \sum_{k=1}^n P(E_k) = \frac{1}{2} P(U \geq x). \quad \square$$

From Lemma 2, we deduce (3.4) for some well-chosen subsequence  $(n_k)$ . Indeed, let us choose these integers such as for all  $k$ ,  $B_{n_{k-1}} \leq (1 + \tau)^k \leq B_{n_k}$ . From Lemma 2 (i)-(ii), we obtain, by using the hypotheses, for all  $\mu > 0$ , and  $k$  large enough (such that  $M_{n_k} < \frac{\mu}{1+\varepsilon} \sqrt{2B_{n_k}/\ln \ln B_{n_k}}$ ):

$$P(S_{n_k} > \phi^\varepsilon(n_k)) < [k \ln(1 + \tau)]^{-(1+\varepsilon)^2(1-\mu)}.$$

Then, choosing  $\mu$  such that  $(1 + \varepsilon)^2(1 - \mu) > 1$ , we have  $\sum_{k=1}^\infty P(S_{n_k} > \phi^\varepsilon(n_k)) < \infty$ . Thus, we conclude by applying Lemma 2 (iii) and the fact that  $\frac{\sqrt{2B_{n_k}}}{\phi^\varepsilon(n_k)} \rightarrow 0$ .  $\square$

### 3.4 Processes in continuous time

In the beginning of the years 1930, a great number of probabilistic works of the Soviet school are related to the study of the stochastic processes in continuous time, meeting thus in particular the needs in physics or aiming at describing some “social phenomena”. The axiomatization due to Kolmogorov, which we commented on above, brought an essential element to the establishment of this theory. The theorem of construction of probability measures on a space of infinite dimension shows that the law of a stochastic process in continuous time is given in a unique way starting from the family of the finite-dimensional marginal laws of the process in question.

### 3.4.1 Chapman-Kolmogorov's equation

The first family of processes to which Kolmogorov, as a good heir to the Russian school of probability, turns naturally to, is that of the Markov processes, i.e. those which satisfy the property (known as the Markov property) of independence of the future with respect to the past conditionally to the knowledge of the present. The article *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung* [Kol31] sets definitely the analytical bases of the theory of Markov processes.

In this fundamental article published in 1931, the entire study of the process is focused around the function:

$$P(s, x, t, A)$$

which represents the probability such that at time  $t$  the random phenomenon is in one of the states of the set  $A$ , if it is in the state  $x$  at time  $s$ , prior to  $t$  ( $0 \leq s < t$ ). As the necessary measurability assumptions are supposed to be satisfied, this function must verify the integral equation

$$P(s, x, t, A) = \int_E P(s, x, u, dy) P(u, y, t, A), \quad \text{for all } u \in ]s, t[, \quad (3.5)$$

where  $E$  stands for the set of all the possible states of the process. Equation (3.5), commonly called today Chapman-Kolmogorov's equation (Chapman had indeed noted it in a report [Cha28] on Brownian Motion in 1928), is the analytic translation of Markov's property and the measures  $P(s, x, t, dy)$  represent the transition probabilities of the process: if we denote by  $(X_t)_{t \geq 0}$  this process, then, for all measurable  $A$  (for the measure  $dy$ ):

$$P(s, x, t, A) = P(X_t \in A \mid X_s = x).$$

However, as Kolmogorov's article is purely analytical, as we can easily see from its title, it doesn't mention any pathwise realizations of the random process.

Equation (3.5) can't be entirely solved in an explicit way in the too global framework in which it is posed. Thus, Kolmogorov seeks conditions of regularity on the probabilities  $P(s, x, t, dy)$  which would make it possible to obtain a more accessible form. Eager to use the new techniques of analysis related to Lebesgue's integral, he naturally focuses on the case where  $P(s, x, t, dy)$  is absolutely continuous, with density  $f(s, x, t, y) \geq 0$ , according to Lebesgue's measure.

Equation (3.5), which is satisfied by the transition probabilities  $P(s, x, t, dy)$ , translates for their densities as:

$$\begin{aligned} \int_{-\infty}^{\infty} f(s, x, t, z) dz &= 1 \\ \int_{-\infty}^{\infty} f(s, x, u, z) f(u, z, t, y) dz &= f(s, x, t, y), \end{aligned}$$

for all  $u \in ]s, t[$  and all  $y \in \mathbb{R}$ . Therefore, to obtain local conditions starting from (3.5), the natural idea is to realize a Taylor development of  $f$ , which needs regularity conditions on  $f$  and assumptions on the moments.

Kolmogorov asks that for all  $s, t, y$ ,  $f(s, x, t, y)$  admits third order derivatives on  $x$  and on  $y$  which are uniformly bounded on  $s$  and  $t$ , on any set of the type  $\{s, t : s - t > k\}$ ,  $k > 0$ . Besides, under the following assumptions for the moments:

$$\text{for all } t \geq 0, \lim_{\Delta \rightarrow 0} \int_{-\infty}^{\infty} |y - x|^i f(t, x, t + \Delta, y) dy = 0, \quad i = 1, 2, 3, \quad (3.6)$$

$$\lim_{\Delta \rightarrow 0} \frac{\int_{-\infty}^{\infty} |y - x|^3 f(t, x, t + \Delta, y) dy}{\int_{-\infty}^{\infty} |y - x|^2 f(t, x, t + \Delta, y) dy} = 0, \quad (3.7)$$

he shows the existence of the limits

$$A(s, x) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (y - x) f(s, x, s + \Delta, y) dy, \quad (3.8)$$

$$B^2(s, x) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (y - x)^2 f(s, x, s + \Delta, y) dy, \quad (3.9)$$

which he calls respectively the infinitesimal mean and the infinitesimal variance of the process and which will be known in the future, in the diffusions case, as the drift coefficient and diffusion coefficient. Thus, from the (3.5), the existence of the limits (3.8) and (3.9) and under the differentiability assumption of  $f$  mentioned previously, Kolmogorov obtains the two following partial differential equations:

$$\frac{\partial}{\partial s} f(s, x, t, y) = -A(s, x) \frac{\partial}{\partial x} f(s, x, t, y) - B^2(s, x) \frac{\partial^2}{\partial x^2} f(s, x, t, y), \quad (3.10)$$

$$\frac{\partial}{\partial t} f(s, x, t, y) = -\frac{\partial}{\partial y} [A(t, y) f(s, x, t, y)] + \frac{\partial^2}{\partial y^2} [B^2(t, y) f(s, x, t, y)]. \quad (3.11)$$

The importance of these equations is such that one can consider them as being at the origin of the modern theory of stochastic processes. Let us give e.g. the main arguments of the proof of the first equation, which the author calls *first fundamental differential equation* and which is now known as the *backward equation* (the second being the *forward equation*).

*Proof of equation (3.10).* If we apply Taylor-Lagrange's formula to the 3<sup>rd</sup> order in the variable  $z$  on the function  $f(s + \Delta, z, t, y)$  and at the points  $x$  and  $z$ , we obtain for  $s < s + \Delta < t$ ,

$$\begin{aligned} f(s, x, t, y) &= \int_{-\infty}^{\infty} f(s, x, s + \Delta, z) f(s + \Delta, z, t, y) dz \\ &= \int_{-\infty}^{\infty} f(s, x, s + \Delta, z) [f(s + \Delta, x, t, y) \end{aligned}$$

$$\begin{aligned}
& + \frac{\partial}{\partial x} f(s + \Delta, x, t, y)(z - x) + \frac{\partial^2}{\partial x^2} f(s + \Delta, x, t, y) \frac{(z - x)^2}{2} \\
& + \frac{\partial^3}{\partial x^3} f(s + \Delta, x, t, y) \frac{(z - x)^3}{6} ] dz,
\end{aligned}$$

with  $\alpha = x + c(z - x)$ , in the case where  $c$  is such that  $0 < c < 1$ . Thus, if we apply the notations

$$\begin{aligned}
a(s, x, \Delta) &= \int_{-\infty}^{\infty} (y - x) f(s, x, s + \Delta, y) dy, \\
b^2(s, x, \Delta) &= \int_{-\infty}^{\infty} (y - x)^2 f(s, x, s + \Delta, y) dy, \\
c(s, x, \Delta) &= \int_{-\infty}^{\infty} |y - x|^3 f(s, x, s + \Delta, y) dy,
\end{aligned}$$

we can write, under the boundedness assumption on the third order derivative, that for a value  $C$  independent from  $\Delta$  and for  $\theta$  such that  $|\theta| < C$ ,

$$\begin{aligned}
f(s, x, t, y) &= f(s + \Delta, x, t, y) + \frac{\partial}{\partial x} f(s + \Delta, x, t, y) a(s, x, \Delta) \\
&+ \frac{\partial^2}{\partial x^2} f(s + \Delta, x, t, y) \frac{b^2(s, x, \Delta)}{2} + \theta \frac{c(s, x, \Delta)}{6},
\end{aligned}$$

which brings us immediately to the finite difference formula

$$\begin{aligned}
\frac{f(s + \Delta, x, t, y) - f(s, x, t, y)}{\Delta} &= -\frac{\partial}{\partial x} f(s + \Delta, x, t, y) \frac{a(s, x, \Delta)}{\Delta} \\
&- \frac{\partial^2}{\partial x^2} f(s + \Delta, x, t, y) \frac{b^2(s, x, \Delta)}{2\Delta} - \theta \frac{c(s, x, \Delta)}{6\Delta}.
\end{aligned}$$

To conclude, let us note that under the above assumptions, the ratio  $c(s, x, \Delta)/\Delta$  tends towards 0 when  $\Delta$  tends towards 0.  $\square$

As we already pointed out, the study of “random movements” whose law is governed by the (3.5) had already been outlined by Chapman in 1928 ([Cha28]) in a context of theoretical physics. The name of “Chapman-Kolmogorov” equation should not let one believe however that it was the only occasion, before the article of 1931, when this equation appeared. Kolmogorov himself, in this article, mentions a particular case studied by Louis Bachelier in 1900 ([Bac00]). He underlines, in a section that he devotes to the work of Bachelier, that the equation (3.11) had been written in his work of 1900 without however having been proven, in the case where the process is homogeneous in space, i.e. when the densities  $f(s, x, t, y)$  only depend on  $s, t$  and on the difference  $y - x$ . The equation appears also in the works of Marian Smoluchowski about the Brownian motion during the years 1910.

Some continuations of the article *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung* [Kol31] appeared shortly after its publication on

behalf of other probabilists like Bernstein, to whom Kolmogorov's equations inspired his theory of the stochastic differential equations in 1932. However, this theory is based on the discrete model and only allows to obtain weak solutions in the continuous case. Another important work was that of Wolfgang Doeblin, carried out in 1940. Doeblin sent it whilst at war (where he died) to the *Académie des Sciences de Paris*, in a sealed envelope which wasn't opened until 2000 ([Doe40])<sup>10</sup>. In this manuscript, Doeblin, very much ahead of his time, considers the pathwise aspects of the stochastic processes. More exactly, he establishes links between the strictly analytical point of view of Kolmogorov and that of Lévy who concentrates primarily on the paths construction and the fine properties of the processes, and especially of the Brownian motion, by purely probabilistic methods which often left his contemporaries perplexed. Doeblin builds equations very close to Itô's stochastic differential equations established some ten or fifteen years later and whose solutions are Brownian motions with a modified temporal variable: if the law of  $(X_t)$  satisfies the (3.5), then

$$X_t = x + \beta_{H(t)} + \int_0^t A(s, X_s) ds,$$

where  $\beta$  is a real-valued Brownian motion and  $H$  the time change

$$H(t) = \int_0^t B^2(s, X_s) ds.$$

This pathwise vision of processes offers then quite more than the analytical *forward* and *backward* (3.10) and (3.11). It allows Doeblin to establish results on the regularity of trajectories, the comparison of solutions, the properties of iterated logarithm, the functional central limit theorems and especially a preliminary version of the formula of change of variable that Itô will obtain a few years later ([Itô44]) and which will inaugurate the era of stochastic calculus itself.

To establish his formula, Doeblin considers a function  $\varphi(t, x)$  of class  $\mathcal{C}^{1,2}$  (i.e. of class  $\mathcal{C}^1$  with respect to  $t$  and  $\mathcal{C}^2$  with respect to  $x$ ) and increasing with respect to  $x$ , which ensures quite easily that the law of the process  $Y_t = \varphi(t, X_t)$  is a solution of Kolmogorov's equation as soon as the law of  $(X_t)$  is also a solution. Thus, he proves that  $(Y_t)$  satisfies

$$Y_t = \varphi(0, x) + \gamma_{\overline{H}(t)} + \int_0^t \overline{A}(s, X_s) ds,$$

where  $\gamma$  is a real-valued Brownian motion and

$$\begin{aligned} \overline{H}_t &= \int_0^t \overline{B}^2(s, X_s) ds, \quad \overline{B}(s, x) = \left( \frac{\partial}{\partial x} \varphi(s, x) \right) B(s, x) \\ \overline{A}(s, x) &= \frac{\partial \varphi(s, x)}{\partial x} A(s, x) + \frac{\partial}{\partial s} \varphi(s, x) + \frac{1}{2} \left( \frac{\partial^2}{\partial x^2} \varphi(s, x) \right) B^2(s, x). \end{aligned}$$

---

<sup>10</sup> The readers who may be less interested in the technical aspects can content themselves by reading the article [BY03]

It was necessary to await the construction of Itô's stochastic integral during and after the second world war to see the solutions of the (3.10) and (3.11) under a new aspect. These new equations are satisfied<sup>11</sup> by the process itself and no longer only by its transition probabilities. If  $\beta$  stands for a real-valued Brownian motion and if  $A(t, x)$  and  $B(t, x)$  are the functions defined as presented at the beginning of this section, then the transition probabilities of the process  $X$  solution of the stochastic differential equation

$$dX_t = A(t, X_t) dt + B(t, X_t) d\beta_t, \quad (3.12)$$

satisfy the (3.5), (3.10) and (3.11). An essential tool for the obtention of this result was Itô's formula mentioned above. In its most common current form, this fundamental formula is stated as follows: if  $\varphi(t, x)$  is a function of class  $\mathcal{C}^{1,2}$  then

$$\begin{aligned} d\varphi(t, X_t) = & B(t, X_t) \frac{\partial}{\partial x} \varphi(t, X_t) d\beta_t \\ & + \left( \frac{\partial}{\partial t} \varphi(t, X_t) + A(t, X_t) \frac{\partial}{\partial x} \varphi(t, X_t) + \frac{1}{2} B^2(t, X_t) \frac{\partial^2}{\partial x^2} \varphi(t, X_t) \right) dt, \end{aligned} \quad (3.13)$$

where  $X$  solves the stochastic differential equation (3.12). These are the bases of the stochastic calculus which was going to encounter several developments during all the second half of the twentieth century.

From the years 1950 onwards, Doob's martingale theory [Doo90], developed afterwards by P.A. Meyer and his school in Strasbourg, was going to make it possible to weaken the conditions imposed until then on the functions  $A(t, x)$  and  $B(t, x)$  to ensure the construction of stochastic processes. An essential remark in this direction was the observation that under natural assumptions of local boundedness and lipschitzianity, the (3.12) admits a single solution in law, in the sense that if  $\beta'$  is another Brownian motion (eventually defined on another probability space), a solution  $X'$  of:

$$dX'_t = A(t, X'_t) dt + B(t, X'_t) d\beta'_t$$

follows the same law as  $X$ . This made it possible in the years 1970 to define the concept of weak solution for the (3.12) which isn't related to the specific choice of a particular Brownian motion any longer. A famous theorem of Yamada and Watanabe [YW71] asserts that pathwise uniqueness (the one which corresponds to an equation directed by a fixed Brownian motion) implies the unicity in law of the weak solutions. A powerful formulation was proposed by Stroock and Varadhan [SV79] in terms of *martingale problems*. The associated

---

<sup>11</sup> In the very particular way of stochastic differential equations which necessitate the concept of Itô's stochastic integral



generator to the Markovian process  $(X_t)$  solution of (3.12) is the operator  $L$  on  $\mathcal{C}^{1,2}$  defined by:

$$Lf(t, x) = \frac{1}{2}B^2(t, x)\frac{\partial^2}{\partial x^2}f(t, x) + A(t, x)\frac{\partial}{\partial x}f(t, x) + \frac{\partial}{\partial t}f(t, x).$$

Itô's formula allows to express this definition by saying that for all  $f \in \mathcal{C}^{1,2}$ ,

$$(M_t) = \left( f(t, X_t) - \int_0^t Lf(s, X_s)ds \right)_{t \geq 0} \quad (3.14)$$

is a local martingale<sup>12</sup>.

Then, it becomes natural to define a solution for (3.12) as follows. Let  $\mathcal{C} = \mathcal{C}(\mathbb{R}^+, \mathbb{R})$  the set of continuous functions from  $\mathbb{R}^+$  to  $\mathbb{R}$ . We define the *canonical projections* on  $\mathcal{C}$  by  $X_t(\omega) = \omega(t)$  for  $\omega \in \mathcal{C}$  and the *canonical filtration* by  $\mathcal{C}_t = \sigma(X_s, s \leq t)$ ,  $t \geq 0$ . Thus, a solution of (3.12) is a probability  $P$  on  $(\mathcal{C}, (\mathcal{C}_t))$  such that under  $P$  the processes defined by (3.14) are local martingales.

The most interesting aspect of the work of Stroock and Varadhan is that under very weak conditions (approximately, the continuity of the functions  $A$  and  $B$ ), they showed that the preceding martingales problem admits a solution  $P$ . Under this probability, the canonical process satisfies the Markovian properties which were at the origin of Kolmogorov's studies. The reader interested by these subjects can consult with interest the important treatise of Jacod [Jac79].

### 3.4.2 Processes with independent and stationary increments

Among the processes whose laws verify Chapman-Kolmogorov's equation (3.5), there is a very important family that Lévy started to study at the beginning of the years 1930: those where functions  $f(s, x, t, y)$  are homogeneous in time and space, i.e. depend only on the differences  $t - s$  and  $y - x$ . In other terms, we are talking about the *processes with independent and stationary increments* for which Kolmogorov attempts to characterize the law in an article edited in two parts in 1932: *Sulla forma generale di un processo stocastico omogeneo* [Kol32a] and *Ancora sulla forma generale di un processo omogeneo* [Kol32b]. He simply considers a "random time function"  $X(\lambda)$ , where  $\lambda \geq 0$  represents the time variable, such that for all  $\lambda_1$  and  $\lambda_2$ , ( $\lambda_2 \geq \lambda_1$ ) the difference  $X(\lambda_2) - X(\lambda_1)$  is independent from  $(X(\lambda), \lambda \leq \lambda_1)$  and the law of which only depends on  $\lambda_2 - \lambda_1$ , i.e. if we note  $\Delta = \lambda_2 - \lambda_1$ ,

$$\Phi_\Delta(x) = P(X(\lambda_2) - X(\lambda_1) < x) = P(X(\lambda_2 - \lambda_1) < x), \text{ for all } x \in \mathbb{R}.$$

---

<sup>12</sup> I.e. there exists a sequence of a.s. finite stopping times  $(T_n)$ , increasing towards  $+\infty$ , such that for all  $n$ , the process  $(M_{t \wedge T_n})$  is a martingale

Then, he discovers that the relation

$$\Phi_{\Delta_1+\Delta_2}(x) = \int_{-\infty}^{\infty} \Phi_{\Delta_1}(x-y) d\Phi_{\Delta_2}(y),$$

is a particular case of the (3.5). Nevertheless, we can notice that at that time, these processes aren't related to Markov processes, which are the subject of the study mentioned above. In fact, this formalization will only appear in the years 1950.

The aim of the article of 1932 is, according to Kolmogorov himself, to generalize some results given by Bruno de Finetti [Fin30] in the case where the laws given by the repartition functions  $\Phi_{\Delta}$  admit second order moments,  $\int x^2 d\Phi_{\Delta}(x) < \infty$ . We will use the following notations:

$$m_{\Delta} = \int x d\Phi_{\Delta}(x), \quad \text{and} \quad \sigma_{\Delta}^2 = \int (x - m_{\Delta})^2 d\Phi_{\Delta}(x).$$

Thus Kolmogorov obtains a particular case of the famous Lévy-Khinchin's formula: if  $\psi_{\Delta}(t) = \int e^{itx} d\Phi_{\Delta}(x)$  then  $\psi_{\Delta}(t) = [\psi_1(t)]^{\Delta}$  and

$$\log \psi_1(t) = itm_1 - \frac{\sigma_0^2}{2}t^2 + \int_{-\infty}^{\infty} \left( e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2} G(dx), \quad (3.15)$$

where  $G$  is a finite measure such that  $G(\{0\}) = 0$  and where  $m_1 \in \mathbb{R}$ ,  $\sigma_0^2 \geq 0$ . This formula is commonly attributed to Lévy and Khinchin who obtained its final version in 1934 ([Lév34]) and 1937 ([Khi37]), although, as we have seen, de Finetti and Kolmogorov had already established it in particular cases. More precisely, the result given by Kolmogorov is the following.

**Theorem 4.** *When the law given by the repartition function  $\Phi_{\Delta}$  has a second order moment, we have*

$$\log \psi_1(t) = itm_1 - \frac{\sigma_0^2}{2}t^2 + \int_{-\infty}^{\infty} \pi(x, t) dF(x), \quad (3.16)$$

where  $m_1 \in \mathbb{R}$ ,  $\sigma_0^2 \geq 0$ ,  $\pi(x, t) = (e^{itx} - 1 - itx)/x^2$  and where the measure  $dF(x)$  is defined by an increasing and bounded function  $F$ .

The previous formulae only concern unidimensional laws of the random function  $(X(\lambda))$ ; consequently, it would be better to talk about the characterization of *indefinitely divisible laws*<sup>13</sup> rather than of a result on processes with independent and stationary increments. Nevertheless, let us note that at that time, this terminology didn't exist.

---

<sup>13</sup> A law of probability  $P$  is said to be *indefinitely divisible* if, for all  $n$ , it is the  $n$ -th convolution power of a law  $\mu_n$ . That amounts to saying that  $P$  is the law of the sum of  $n$  real-valued independent random variables with the same law  $\mu_n$

*Kolmogorov's proof:* First of all, we verify that  $\psi_\Delta$  is continuous with respect to  $\Delta$ . Indeed, for  $\Delta \leq 1/n$ , we have  $\sigma_\Delta^2 = \sigma_{1/n}^2 - \sigma_{1/n-\Delta}^2 \leq \sigma_{1/n}^2 = (1/n)\sigma_1^2$ , and so  $\sigma_\Delta^2 \rightarrow 0$  when  $\Delta \rightarrow 0$ . Consequently,  $\psi_\Delta(t) \rightarrow 1$  when  $\Delta \rightarrow 0$ . We conclude thanks to the equality  $\psi_{\Delta_1+\Delta_2}(t) = \psi_{\Delta_1}(t)\psi_{\Delta_2}(t)$ . The continuity of  $\psi_\Delta$  in  $\Delta$  allows to show that the equality

$$\psi_\Delta(t) = [\psi_1(t)]^\Delta,$$

which is true for all rationals  $\Delta$ , is also verified for all reals  $\Delta$ . The author deduces then (using de Finetti's proof) that

$$\log \psi_1(t) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} [\psi_\Delta(t) - 1].$$

Moreover, we have

$$\frac{1}{\Delta} [\psi_\Delta(t) - 1] = itm_1 + \frac{1}{\Delta} \int_{-\infty}^{\infty} (e^{itx} - 1 - itx) d\Phi_\Delta(x).$$

Let us note  $F_\Delta(x) = \frac{1}{\Delta} \int_{-\infty}^x y^2 d\Phi_\Delta(y)$ , so that

$$\frac{1}{\Delta} \int_{-\infty}^{\infty} (e^{itx} - 1 - itx) d\Phi_\Delta(x) = \int_{-\infty}^{\infty} \pi(x, t) dF_\Delta(x).$$

A classical argument allows to justify that for any sequence  $\Delta_n$  decreasing towards 0, there exists a subsequence  $\Delta_{n_k}$  such that the sequence  $F_{\Delta_{n_k}}(x)$  converges when  $k$  tends towards  $+\infty$ , towards the function  $F(x)$  in all points  $x$  where the latter is continuous. Let us remark that  $F$  is an increasing function such that:

$$0 = \lim_{\Delta \downarrow 0} F_\Delta(-\infty) \leq F(-\infty) \leq F(\infty) \leq \lim_{\Delta \downarrow 0} F_\Delta(\infty) = \sigma_1^2$$

and taking into account that at fixed  $t$ ,  $\pi(x, t) \rightarrow 0$  when  $x \rightarrow \pm\infty$ , we obtain

$$\lim_{k \rightarrow \infty} \int_{-\infty}^{\infty} \pi(x, t) dF_{\Delta_{n_k}}(x) = \int_{-\infty}^{\infty} \pi(x, t) dF(x),$$

which implies that

$$\log \psi_1(t) = itm_1 + \int_{-\infty}^{\infty} \pi(x, t) dF(x).$$

But as  $\sigma_1^2$  is finite, we have  $\log \psi_1(t) = itm_1 - \frac{\sigma_1^2}{2}t^2 + o(t^2)$  ( $t \rightarrow 0$ ), and according to the previous computation

$$\log \psi_1(t) = itm_1 - \frac{t^2}{2}(F(\infty) - F(-\infty)) + o(t^2) \quad (t \rightarrow 0).$$

This implies in particular that  $F(\infty) - F(-\infty) = \sigma_1^2$ . Finally we deduce from the above that  $F(-\infty) = 0$  and  $F(\infty) = \sigma_1^2$ ; thus  $F$  is entirely determined.

Now, let us prove that for all increasing (and left continuous) functions  $F$ , with values between  $F(-\infty) = 0$  and  $F(+\infty) = \sigma_1^2 < +\infty$ , the function  $\psi_1$  given by  $\log \psi_1(t) = itm_1 + \int_{-\infty}^{\infty} \pi(x, t) dF(x)$  is the characteristic function of an indefinitely divisible law. For this purpose, let us consider a step function  $T$  with steps:

$$\omega_k = T(x_k+) - T(x_k),$$

for a finite number of reals  $x_1, x_2, \dots, x_n$ . We also suppose that  $T$  doesn't jump at point 0. We note  $\sigma_1^2 = \omega_1 + \omega_2 + \dots + \omega_n$  the sum of these jumps. Let us also note  $\log \bar{\psi}_1(t) = itm_1 + \int_{-\infty}^{\infty} \pi(x, t) dT(x)$ ,  $\eta = m_1 - \sum_k p_k$ ,  $p_k = \omega_k/x_k^2$  and  $\chi_k(t) = \exp(itx_k)$ . Then we can easily verify the identities

$$\begin{aligned} \int_{-\infty}^{\infty} \pi(x, t) dT(x) &= \sum_k p_k (e^{itx_k} - 1 - itx_k), \\ \log \bar{\psi}_1(t) &= it\eta + \sum_k p_k (\chi_k(t) - 1). \end{aligned}$$

Thus, we deduce that

$$\bar{\psi}_\Delta(t) = (\bar{\psi}_1(t))^\Delta = \exp(it\eta\Delta) + \sum \Delta p_k (\chi_k(t) - 1).$$

Let us note that  $\chi_k(t)$  and  $\exp(it\eta\Delta)$  are characteristic functions and consequently  $\psi_\Delta(t)$ , as a product of characteristic functions, is a characteristic function itself. To conclude, let us note that we can find a sequence of step functions  $T_n$  such that  $\int_{-\infty}^{\infty} \pi(x, t) dT_n(x)$  converges towards  $\int_{-\infty}^{\infty} \pi(x, t) dF(x)$  uniformly on any bounded interval. Thus, the corresponding characteristic functions  $\bar{\psi}_\Delta^{(n)}(t)$  converge towards  $\psi_\Delta(t)$ . This implies that  $\psi_\Delta$  is a characteristic function, on one hand, and that this function verifies  $\psi_\Delta(t) = (\psi_1(t))^\Delta$ , on the other hand.  $\square$

Kolmogorov completes this description of the law of processes with independent increments by observing that if  $P_1(x) = \int_x^\infty y^{-2} dF(y)$  et  $P_2(x) = \int_{-\infty}^x y^{-2} dF(y)$  then  $P_1(x) d\lambda$  (respectively  $P_2(x) d\lambda$ ) is the probability such that the process  $X(\lambda)$  has a positive jump (respectively negative), of height bigger than (respectively smaller than) or equal to  $x$ , during the time increment  $d\lambda$ . The measures  $P_1$  and  $P_2$  are respectively the restrictions to  $\mathbb{R}_+$  and  $\mathbb{R}_-$  of what will be called afterwards the Lévy measure of the process  $X(\lambda)$ .

Following the works of Lévy, a number of authors during the years 1960 to 1980 such as Skorokhod, Zolotarev, Blumenthal, Gettoor, Ray, Taylor, Fristedt, Bingham, Pitman, Jacod, ... studied the fine properties of the trajectories of the processes with independent and stationary increments (now called Lévy processes). Then at the beginning of the years 1990, the synthesis works of Bertoin [Ber96] and Sato [Sat99] caused, among the probabilistic community, a renewed interest for the study of Lévy processes. This was again reinforced

with the discovery of new grounds of applications, like financial mathematics, where the models using Lévy processes allow to compensate for the defects of the model known as the Black-Scholes model involving the geometrical Brownian motion.

### 3.4.3 Continuity and relative compactness criteria

It is remarkable that Kolmogorov also developed a certain number of tools for the study of the pathwise properties of random processes. Among those, a very effective criterion guaranteeing the pathwise continuity of the processes bears his name. It was found by the Soviet mathematician in 1934 and presented the same year during the Seminar at the University of Moscow. It was not the subject however of any of his publications and it was Slutsky who stated it and provided the first proof published in the *Giornale dell'Istituto Italiano dei Attuari* in 1937 [Slu37], allotting its paternity to Kolmogorov. Twenty years after this, Kolmogorov pointed out again to Chentsov an extension of this criterion for discontinuous processes which he published in 1956 [Che56] and which allows to conclude that the process in question does not have these discontinuities of the second order<sup>14</sup>.

**Theorem 5.** *If for a family of real-valued random variables  $(X_t, 0 \leq t \leq 1)$  there exist three strictly positive constants  $\gamma, c$  and  $\varepsilon$  such that:*

$$E(|X_t - X_s|^\gamma) \leq c|t - s|^{1+\varepsilon}, \quad (3.4.3.1)$$

*then there exists a modification of  $X$  which is a.s. continuous.*

Let us recall that a modification of a process  $X$  is a process  $\tilde{X}$  such that for all fixed  $t$ ,  $\tilde{X}_t = X_t$ , a.s. The following proof takes as a starting point the proof of [RY91], p. 26; see also [DM75], Vol. V, Chap. XXIII, p. 332.

*Proof:* Let  $D_m$  the set of reals on  $[0,1]$  of the form  $2^{-m}i$ , where  $i = 0, 1, \dots, 2^m$  and  $D = \cup_m D_m$  the set of dyadic numbers on  $[0,1]$ . Let  $\Delta_m$  the set of pairs  $(s, t)$  of  $D_m^2$  such that  $|t - s| = 2^{-m}$ . We note

$$Y_i = \sup_{(s,t) \in \Delta_i} |X_s - X_t|.$$

Since for all  $s, t \in \Delta_i$ ,  $E(|X_s - X_t|^\gamma) \leq c(2^{-i})^{1+\varepsilon}$  and  $\#\Delta_i = 2^i$ , we have for all  $i$ ,

$$E(Y_i^\gamma) \leq \sum_{(s,t) \in \Delta_i} E(|X_s - X_t|^\gamma) \leq c2^i(2^{-i})^{1+\varepsilon} \leq c2^{-i\varepsilon}.$$

<sup>14</sup> In a nearby order of ideas, let us also underline the contribution of Kolmogorov to Skorokhod's topology for the space of right continuous and left limited trajectories [Kol56]

Now, let  $s, t \in D$ ,  $s \neq t$  and  $m$  the integer such that  $2^{-m-1} < |t - s| \leq 2^{-m}$ ; thus there exist finite sequences  $s_m, s_{m+1}, \dots, s_p = s$  and  $t_m, t_{m+1}, \dots, t_k = t$  such that for all  $i = m, m+1, \dots, p$  and  $j = m, m+1, \dots, k$ , we have  $(s_i, s_{i+1}) \in \Delta_{i+1}$ ,  $(t_j, t_{j+1}) \in \Delta_{j+1}$ ,  $(s_m, t_m) \in \Delta_m$  and

$$X_s - X_t = \sum_{i=m}^{p-1} X_{s_{i+1}} - X_{s_i} + X_{s_m} - X_{t_m} + \sum_{j=m}^{k-1} X_{t_{j+1}} - X_{t_j}.$$

Thus, we deduce the inequalities

$$|X_s - X_t| \leq Y_m + 2 \sum_{i=m+1}^{\infty} Y_i \leq 2 \sum_{i=m}^{\infty} Y_i.$$

Let  $\alpha \in [0, \varepsilon/\gamma[$ . Let us define the random variable

$$M_\alpha = \sup\{|X_t - X_s|/|t - s|^\alpha : s, t \in D, s \neq t\}.$$

Then, we deduce from the above that:

$$\begin{aligned} M_\alpha &\leq \sup_{m \in \mathbb{N}} \{2^{(m+1)\alpha} \sup_{2^{-m-1} < |t-s| \leq 2^{-m}} |X_t - X_s| : s, t \in D, s \neq t\} \\ &\leq \sup_{m \in \mathbb{N}} \{2 \cdot 2^{(m+1)\alpha} (\sum_{i=m}^{\infty} Y_i)\} \leq 2^{\alpha+1} \sum_{i=0}^{\infty} 2^{i\alpha} Y_i. \end{aligned}$$

In the case where  $\gamma \geq 1$ , this inequality implies that

$$E(M_\alpha^\gamma)^{1/\gamma} \leq 2^{\alpha+1} \sum_{i=0}^{\infty} 2^{i\alpha} E(Y_i^\gamma)^{1/\gamma} \leq 2^{\alpha+1} \sum_{i=0}^{\infty} 2^{i\alpha} 2^{-i\varepsilon/\gamma} < \infty$$

and when  $\gamma < 1$ , the same inequality applies to  $E(M_\alpha^\gamma)$ . In particular, the variable  $M_\alpha$  is finite and we have, for all  $s, t \in D$ ,  $|X_t(\omega) - X_s(\omega)| < K(\omega)|t - s|^\alpha$ , where  $K(\omega)$  is a constant which does not depend on  $s$  and  $t$ . Then, we deduce that  $X$  is uniformly continuous on  $D$  (and even hölderian of order  $\alpha$ ). The process  $X|_D$  (i.e.  $X$  restricted to  $D$ ) extends then in a unique manner to a continuous process on  $[0, 1]$ :

$$\tilde{X}_t(\omega) = \lim_{s \rightarrow t, s \in D} X_s(\omega), \quad t \in [0, 1]$$

which is naturally also hölderian of order  $\alpha$ . Finally, according to the assumption, for all  $t \in [0, 1]$ ,  $\lim_{s \rightarrow t} X_s = X_t$  in  $L^1$ , thus  $\tilde{X}_t = X_t$ , a.s. for all  $t \in [0, 1]$ .  $\square$

As we can see, we have in fact proven a stronger result: *under the previous assumptions, there exists a modification of  $X$  with a.s. hölderian trajectories of order  $\alpha$  for all  $\alpha \in [0, \varepsilon/\gamma[$ . In particular, the Brownian motion is (a.s.) locally hölderian of order  $\alpha$  for all  $\alpha < 1/2$ .*

Theorem 5 admits some extensions such as the one of Chentsov mentioned previously. There also exists a criterion of the same type for processes with several indexes  $(X_t, t \in [0, 1]^d)$ ; then, in this case, we write the assumption (3.4.3.1) as follows:

$$E(|X_t - X_s|^\gamma) \leq c|t - s|^{d+\varepsilon}.$$

The proof of this result is very close to the one we just gave and also applies to processes with values in general Banach spaces.

In 1970, Garsia, Rodemich and Rumsey suggested another generalization of this continuity criterion which is based on an entirely deterministic argument (see [SV79], p. 47 or [DM75], Vol. V, Chap. XXIII, p. 336). Let us briefly expose their result:

Let  $f$  denote a Borelian function defined on a ball  $B$  of  $\mathbb{R}^d$ , satisfying the integral condition:

$$\int_{B \times B} \Psi \left( \frac{|f(v) - f(u)|}{\varphi(|v - u|)} \right) dudv \stackrel{(def)}{=} A < \infty,$$

where  $\Psi$  and  $\varphi$  are two continuous functions on  $\mathbb{R}_+$ , null in 0, strictly increasing and unbounded, such that the integrals

$$h(x, t) = \int_0^t \Psi^{-1}(x/s^{2d}) d\varphi(s)$$

converge for all  $x > 0$ . Then the function  $f$  admits a continuous version  $\tilde{f}$  (i.e.  $f(u) = \tilde{f}(u)$ , a.e. in the sense of Lebesgue's measure) which possesses on the same ball  $B$  the modulus of continuity:

$$|\tilde{f}(v) - \tilde{f}(u)| \leq 8h(k_d A, 2|v - u|),$$

where the constant  $k_d$  only depends on the dimension  $d$ .

It was then proven that this result is in fact a particular case of Sobolev's embedding theorems. The reader who may be interested in this topic can refer to [DM75], Vol. V, Chap. XXIII, p. 334. It was possible to apply the explicit form of this elaborated version of Kolmogorov's criterion to obtain fine continuity results for the local times of Lévy processes, and more generally of semi-martingales, see [Bar85] and [BY81]. One can compare this form of Kolmogorov's criterion with the *majorizing measures* technique of Fernique, Marcus, Talagrand,...

As one can easily imagine, the applications of Theorem 5 are numerous and diverse. One of the most important is doubtlessly the relative compactness criterion. Let us recall that a sequence of probability measures is said to be (weakly) relatively compact if one can extract from any subsequence, a new subsequence which converges weakly.

According to Prohorov's theorem, a sequence of probability measures  $(P_n)$  on a separable and complete space is relatively compact if and only if it is tight, i.e. if for any  $\varepsilon > 0$ , there exists a compact  $K$  such that  $P_n(K) \geq 1 - \varepsilon$ , for

all  $n$ . This theorem presents a great interest, in particular when the relevant probability space is the set of continuous paths on the positive half-line. Then, let us denote by  $\mathcal{C}$  the space of continuous functions on the positive half-line and with values in  $\mathbb{R}^d$ , which is fitted with the uniform convergence topology and the Borelian  $\sigma$ -algebra. In order to prove that a sequence of measures  $(P_n)$  on  $\mathcal{C}$  converges weakly towards a measure  $P$ , we know that it suffices to verify that it is relatively compact and that its finite dimensional marginals converge towards the corresponding marginals of  $P$ . As  $\mathcal{C}$  is separable and complete, relative compactness is equivalent to the tension property. In general, the convergence of the finite dimensional marginals can easily be stated. Relative compactness is often a problem. A very useful criterion is obtained as a direct consequence of Kolmogorov's theorem exposed in Theorem 5. One can refer to [Bil95] for a proof.

**Corollary 4.** *If a sequence of continuous processes  $(X_n(t), t \geq 0)^{15}$  satisfies*

- (i) *The sequence of r.v.'s  $(X_n(0))$  is tight,*
- (ii) *There exist three strictly positive constants  $\gamma, c$  and  $\varepsilon$  such that for any  $n$ :*

$$E(|X_n(t) - X_n(s)|^\gamma) \leq c|t - s|^{d+\varepsilon}, \quad (3.4.3.2)$$

*then, the sequence of laws of the processes  $(X_n)$  is relatively compact.*

## References

All the articles written by A. N. Kolmogorov listed below can be found in [KolII].

- [Bac00] Bachelier, L.: Théorie de la spéculation. Ann. Sci. École Norm. Sup., **17**, 21–86 (1900)
- [Bar85] Barlow, M.T.: Continuity of local times for Lévy processes. Z. Wahrsch. Verw. Gebiete, **69**, 23–35 (1985)
- [Ber96] Bertoin, J.: Lévy Processes. Cambridge University Press, Cambridge (1996)
- [Bil95] Billingsley, P.: Probability and Measure. Wiley, 3rd edition (1995)
- [BMP01] Baldi, P., Mazliak, L., Priouret, P.: Martingales et Chaînes de Markov. Hermann, paris (2001)
- [BY81] Barlow, M.T., Yor, M.: Semimartingale inequalities via the Garsia-Rodemich-Rumsey Lemma and application to local times. J. Funct. Analysis, **49**, 198–229 (1982)
- [BY03] Bru, B., Yor, M.: Wolfgang Doeblin, l'équation de Kolmogoroff. *La Recherche* (june 2003)
- [Cha28] Chapman, S.: On the brownian displacement and thermal diffusion of grains suspended in a non-uniform fluid. Proc. Roy. Soc. London, Ser. A, **119**, 34–54 (1928)

---

<sup>15</sup> This is equivalent to consider a sequence of random variables  $(X_n)$  with values in  $\mathcal{C}$



- [Che56] Chentsov, N.N.: Weak convergence of stochastic processes whose trajectories have no discontinuities of second kind and the “heuristic” approach to the Kolmogorov-Smirnov tests. *Teor. Ver.*, **1**, 140–144 (1956)
- [Cra37] Cramér, H.: *Random Variables and Probability distributions*. Cambridge (1937)
- [DM75] Dellacherie, C., Meyer, P.A.: *Probabilités et Potentiel*, Hermann, Paris (1975, 1992)
- [Doe40] Doeblin, W.: Sur l’équation de Kolmogoroff (sealed envelope, 1940), edited and commented by B. Bru and M. Yor. *C.R. Acad. Sci. Paris Sér. I Math.*, **331**, 1033–1187 (2000)
- [Doo90] Doob, J.L.: *Stochastic processes*. Wiley Classics Library. John Wiley & Sons, Inc., New York (1990)
- [Fin30] de Finetti, B.: Le funzioni caratteristiche di legge istantanea. *Rend. Acc. Lincei*, **12**, 278–282 (1930)
- [GS80] Guikhman, I., Skorokhod, A.: *Introduction à la théorie des processus aléatoires*. (Fr. trans. D. Embarek). Éditions MIR, Moscow (1980)
- [Itô44] Itô, K.: *Stochastic Integrals*. *Proc. Imp. Acad. Tokyo*, **20**, 519–524 (1944)
- [Itô86] Itô, K.: *Selected papers*, edited by Stroock, D.W. and Varadhan, S.R.S. Springer (1986)
- [Jac79] Jacod, J.: *Calcul stochastique et problèmes de martingales*. *Lecture Notes in Mathematics*, 714. Springer, Berlin (1979)
- [Khi24] Khinchin, A.Y.: Über einen Satz der Wahrscheinlichkeitrechnung. *Fund. Mat.*, **6**, 9–20 (1924)
- [Khi37] Khinchin, A.Y.: Dédution nouvelle d’une formule de M. Paul Lévy. *Bull. Univ. Moscow Math-Mec.*, **1**, 1–5 (1937)
- [KK25] Khinchin, A.Y., Kolmogorov, A.N.: Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden. *Mat. Sb.*, **32**, 668–677 (1925)
- [Kol29] Kolmogorov, A.N.: Über das Gesetz des iterierten Logarithmus. *Math. Ann.*, **101**, 126–135 (1929)
- [Kol30] Kolmogorov, A.N.: Sur la loi forte des grands nombres. *CRAS Paris*, **191**, 910–912 (1930)
- [Kol31] Kolmogorov, A.N.: Über die analytischen Methoden in der Wahrscheinlichkeitrechnung. *Math. Ann.*, **104**, 415–458 (1931)
- [Kol32a] Kolmogorov, A.N.: Sulla forma generale di un processo stocastico omogeneo. *Atti Accad. Naz. Lincei Rend.*, **15**, 805–808 (1932)
- [Kol32b] Kolmogorov, A.N.: Ancora sulla forma generale di un processo stocastico omogeneo, *Atti Accad. Naz. Lincei Rend.*, **15**, 866–869 (1932).
- [Kol33] Kolmogorov, A.N.: *Grundbegriffe der Wahrscheinlichkeitrechnung*. Springer (1933)
- [Kol56] Kolmogorov, A.N.: On the Skorohod convergence. *Teor. Veroyatnost. i Primenen.*, **1**, 239–247 (1956)
- [KolII] Shiryaev, A.N. (ed): *Selected works of A.N. Kolmogorov*, Vol. II (translated by G. Lindquist). Kluwer Academic Publications (1992)
- [Lév34] Lévy, P.: Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. *Ann. Sc. Norm. Pisa*, **3**, 337–366 (1934)
- [Lév37] Lévy, P.: *Théorie de l’Addition des Variables aléatoires*. Gauthier-Villars, Paris (1937)
- [Maz03] Mazliak, L.: Andrei Nikolaevitch Kolmogorov (1903–1987). Un aperçu de l’homme et de l’œuvre probabiliste. *Cahiers du CAMS-EHESS*,

- Paris (2003). Available at <http://www.proba.jussieu.fr/users/lma/recherche.html>
- [RY91] Revuz, D., Yor, M.: Continuous martingales and Brownian motion. Springer (1991). 3rd edition, Springer (1999)
- [Sat99] Sato, K.I.: Lévy processes and infinitely divisible distributions. Translated from the 1990 Japanese original. Cambridge Studies in Advanced Mathematics, 68. University Press, Cambridge (1999)
- [Shi89] Shiryaev, A.N.: Kolmogorov: Life and creative activities. Ann. Prob., **17**, 866–944 (1989)
- [Slu37] Slutsky, E.B.: Qualche proposizione relativa alla teoria delle funzioni aleatorie. Gio. Ist. It. Att., **8**, 183–199 (1937)
- [Str03] Stroock, D.W.: Markov processes from Itô’s perspective. Princeton University Press (2003)
- [SV06] Shafer, G., Vovk, V.: The origins of Kolmogorov’s Grundbegriffe. Statistical Science, **21**:1, 70–98 (2006)
- [SV79] Stroock, D.W., Varadhan, S.R.S.: Multidimensional diffusion processes. Springer (1979)
- [Szé86] Székely, G.J.: Paradoxes in probability theory and mathematical statistics. Mathematics and its Applications (East European Series), **15**. D. Reidel Publishing Co., Dordrecht (1986)
- [vonP94] von Plato, J.: Creating Modern Probability. Its mathematics, physics and philosophy in historical perspective. Cambridge University Press, Cambridge (1994)
- [Wie23] Wiener, N.: Differential space. Jour. Math. and Phys., **58**, 131–174 (1923)
- [YW71] Yamada, T., Watanabe, S.: On the uniqueness of solutions of stochastic differential equations II. J. Math. Kyoto Univ., **11**, 553–563 (1971)