# A. N. Kolmogorov

Quelques aspects de l'œuvre probabiliste*

Loïc CHAUMONT, Laurent MAZLIAK et Marc YOR
Laboratoire de Probabilités et Modèles Aléatoires
Universités Paris VI et VII
4, Place Jussieu
75252 Paris Cedex 05, France

November 5, 2003

*Résumé* : La partie la plus célèbre de l'œuvre mathématique d'Andrei Nikolaevitch Kolmogorov (1903-1987) concerne le calcul des probabilités. Nous présentons ici quelques-uns des résultats les plus emblématiques du mathématicien soviétique dans ce domaine.

*Mots Clés* : fondements des probabilités, théorèmes limites, lois indéfiniment divisibles, processus stochastiques

*Codes AMS* : 01A60, 60Axx, 60Fxx, 60G05.

## 1 Introduction

Qui consulte un ouvrage sur l'œuvre mathématique de Kolmogorov doit naturellement s'attendre à trouver de larges considérations au sujet de l'axiomatisation des probabilités que Kolmogorov a mise au point au début des années 1930 et qui forme le contenu de sa célèbre publication *Grundbegriffe der Wahrscheinlichkeitsrechnung* (*Fondements de la théorie des probabilités*) [26], parue chez Springer en 1933. Il est certain que de toute la production du mathématicien soviétique, ce petit opuscule d'une soixantaine de pages est la part la plus célèbre, et souvent l'une des seules à laquelle son nom soit attaché pour un public assez vaste mais aussi pour certains mathématiciens. Sans vouloir aucunement diminuer l'importance de ce travail, il reste néanmoins étonnant que l'attention se soit ainsi focalisée sur ce qui ne constitue pas la création la plus originale de Kolmogorov dans le domaine des probabilités. L'objet de cette partie, consacrée à certains aspects de l'œuvre probabiliste du savant, est justement de mettre en avant quelques-uns de ses travaux les plus remarquables. Dans ce monument imposant, un choix drastique a été nécessaire et

*Ce texte constitue un chapitre du livre « L'héritage de Kolmogorov en mathématiques » à paraître aux éditions Belin, collection « Échelles », 2003.

nous avons choisi de nous concentrer sur les deux directions purement probabilistes que Kolmogorov a le plus travaillées, à savoir d'une part l'étude des situations limites pour les sommes de variables aléatoires indépendantes, qui lui permit de prolonger les études de ses prédécesseurs russes Markov et Lyapounov, et d'autre part des considérations littéralement révolutionnaires sur les processus en temps continu, dont les ramifications se prolongent très en avant dans le temps jusqu'à des découvertes qui datent d'à peine trente ans. Comme il nous a néanmoins semblé difficile, si ce n'est impossible, qu'un chapitre consacré au travail probabiliste du mathématicien soviétique ne fasse pas référence à l'axiomatisation des probabilités, nous commencerons par un bref survol des principaux apports de celle-ci, renvoyant le lecteur aux nombreux articles traitant de la question de façon détaillée (voir par exemple [40], [33]). On se reportera par ailleurs au texte indispensable de Shiryaev [34] pour un tableau plus complet des travaux de Kolmogorov. On trouvera dans l'article [30] des indications sur la vie du mathématicien et le statut de la discipline dans l'URSS stalinienne.

# 2 L'axiomatisation du calcul des probabilités

## 2.1 Un cadre abstrait

Comme mentionné précédemment, la publication de Kolmogorov *Grundbegriffe der Wahrscheinlichkeitsrechnung* [26] est une modeste monographie de 60 pages parue en 1933 dans un ensemble de textes consacrés à la théorie moderne des probabilités. La traduction russe du texte date de 1936, et fut en grande partie réalisée en hâte pour des raisons politiques au moment où une pression importante s'exerçait sur les scientifiques soviétiques pour qu'ils publient leurs travaux en russe et en URSS plutôt qu'à l'étranger. La première traduction anglaise date, elle, de 1950. Ce délai relativement important montre que l'axiomatisation proposée par le savant russe ne s'est pas imposée d'une manière aussi foudroyante qu'on le pense habituellement. De nombreux probabilistes, et parmi les plus éminents, comme Paul Lévy, ne se serviront jamais de l'axiomatisation de Kolmogorov, ce qui ne les empêchera en rien d'avoir des idées extraordinaires. En fait, hors de l'URSS, avant les années 1950, il n'est plus ou moins que le traité de Cramér [10] qui fasse référence à ce cadre. L'auteur n'en donne d'ailleurs pas de justification profonde ; c'est uniquement le fait que l'axiomatisation de Kolmogorov soit la plus pratique parmi celles disponibles à cette époque (en particulier la *théorie des collectifs* proposée par von Mises) qui la lui fait utiliser. Cependant, à partir des années 1950, elle sera adoptée définitivement par la jeune génération. Ce qui est séduisant dans le cadre formel proposé par cette axiomatisation, c'est qu'elle fournit par exemple une explication globale des multiples paradoxes qui avaient été énoncés dans le passé de la discipline (comme ceux de Joseph Bertrand, Émile Borel etc.) : à chaque fois, la donnée précise de l'espace de probabilités comme description de l'expérience aléatoire considérée permet de lever l'ambiguïté. (Sur ce sujet, voir *infra*, ainsi que [33] et [39]. On pourra aussi consulter les commentaires d'Itô dans la préface de [16].)

La grande force du traité de Kolmogorov est de se placer volontairement dans un cadre totalement abstrait, sans chercher à établir des ponts avec les aspects appliqués de la théorie des probabilités, au-delà de la situation des probabilités finies. La recherche de tels liens dans le cas général amène en effet forcément à affronter des questions philosophiques délicates et risque ainsi d'obscurcir la modélisation mathématique. En ne parlant des questions d'application que dans la partie consacrée aux probabilités finies[1], Kolmogorov se libère de cette contrainte et peut éviter les écueils que n'avait pas toujours contournés von Mises. En effet, la *théorie des collectifs* prétendait aussi établir une discrimination entre les expériences pour lesquelles l'application des probabilités était légitime, et les autres. Mais Kolmogorov, qui présente une théorie purement mathématique, n'a pas une telle ambition, et donc pas une telle limitation. À l'intérieur du cadre abstrait qu'il énonce, tout travail mathématique est légitime, sa validation pour les applications relevant d'autres domaines. En particulier, il se permet de considérer des ensembles n'ayant aucune structure topologique quitte, pour des études plus fines (comme les phénomènes de convergence), à travailler sur de meilleurs espaces à travers l'utilisation de lois images. Ce fait mettra d'ailleurs l'axiomatisation de Kolmogorov à l'opposé des conceptions de Bourbaki sur la théorie de la mesure. Le caractère très général de sa théorie va permettre au mathématicien russe d'utiliser dans toute sa force la théorie de la mesure de Borel et de Lebesgue, encore relativement neuve à cette époque puisque sa version abstraite a été en grande partie mise au point par Fréchet (cité dans les *Grundbegriffe* comme celui qui a libéré la théorie de la mesure de la géométrie) puis par l'école polonaise (Banach, Sierpiński, Kuratowski...) dans les années 1920.

Dès les premières années du vingtième siècle, Borel avait été un promoteur de l'utilisation de la théorie de la mesure et de l'intégrale de Lebesgue pour le traitement de questions de probabilité. En 1909, il publie un papier révolutionnaire où une telle méthode lui permet d'obtenir une première version forte de la loi des grands nombres et des interprétations sur la répartition des nombres réels. Sans doute sa considération modérée pour les mathématiques probabilistes et de sérieux doutes quant à la légitimité de leurs applications l'empêchèrent de récolter jusqu'au bout ce qu'il avait semé.

Kolmogorov introduit la notion désormais classique d'espace de probabilités sous forme d'un triplet $(\Omega, \mathcal{F}, P)$ composé d'un ensemble $\Omega$ muni d'une tribu $\mathcal{F}$ et d'une mesure normalisée (probabilité) $P$. Les variables aléatoires sont simplement des fonctions $X$ à valeurs réelles définies sur $\Omega$ telles que pour tout $a \in \mathbb{R}$, $\{\omega \in \Omega, X(\omega) < a\} \in \mathcal{F}$ et leur loi est la mesure image de la probabilité $P$ définie par $P^{(X)}(A) = P(X^{-1}(A))$, pour tout $A \in \mathcal{B}(\mathbb{R})$, la tribu borélienne de $\mathbb{R}$.

Les apports majeurs de l'ouvrage de Kolmogorov dans la clarification de notions probabilistes sont incontestablement la construction d'une mesure de probabilité sur un produit infini d'espaces, qui joue un rôle important dans la théorie des processus stochastiques, et la formalisation de la loi conditionnelle via l'utilisation du théorème de Lebesgue-Radon-Nikodym, dont la version abstraite avait été publiée par Nikodym en 1930. Notons au passage que ce n'était pas la première fois qu'était construite une probabilité sur un espace produit : l'exemple le plus célèbre est donné par le mémoire de Wiener [41] qui dès 1923, en

[1]Ce qui n'est pas sans rappeler la façon dont, dans son article de 1931 sur les processus de Markov, il avait proposé dans une longue introduction de reporter à d'autres travaux les réflexions sur l'applicabilité de ses théories.

appliquant des techniques que Daniell avait développées quelques années auparavant pour étendre l'intégrale de Lebesgue à des espaces de dimension infinie, construisait la mesure associée au mouvement brownien (voir [31]).

## 2.2 Construction de la loi conditionnelle

Présentons en quelques mots la construction de la loi conditionnelle, en suivant le texte de Kolmogorov mais avec des notations modernisées pour la clarté de l'exposé.
On rappelle tout d'abord la définition de la probabilité conditionnelle élémentaire d'un événement $C$ (c'est-à-dire d'un élément de la tribu $\mathcal{F}$) par un événement $D$ tel que $P(D) > 0$, par $P(C \mid D) = \frac{P(C \cap D)}{P(D)}$. Soit maintenant $U$ une variable aléatoire réelle et $B$ un événement. On cherche à construire une variable aléatoire $\omega \mapsto \pi(U(\omega); B)$, fonction borélienne de $U$, appelée probabilité conditionnelle de $B$ sachant $U$, et telle que, pour tout $A \in \mathcal{B}(\mathbb{R})$ tel que $P(U \in A) > 0$, on ait :

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

Pour $A \in \mathcal{B}(\mathbb{R})$, on pose $Q_B(A) = P(B \cap U^{-1}(A))$. Notons que si $P^{(U)}$ est la loi de $U$ définie par $P^{(U)}(A) = P(U^{-1}(A))$, alors $P^{(U)}(A) = 0$ implique $Q_B(A) = 0$ et donc, par le théorème de Lebesgue-Radon-Nikodym, on peut trouver une fonction borélienne $f_B$ telle que $\forall A \in \mathcal{B}(\mathbb{R}), Q_B(A) = \displaystyle\int_{\mathbb{R}} \mathbb{1}_A f_B dP^{(U)}$, i.e.

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

et donc

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

et on pose $\pi(U; B) = f_B \circ U$. De là, Kolmogorov va redémontrer toutes les propriétés classiques des probabilités conditionnelles. Il illustre joliment la puissance de son formalisme en levant le paradoxe de Borel concernant le tirage aléatoire d'un point sur une sphère : le lecteur intéressé pourra par exemple consulter [6], p. 462 et [33].

## 2.3 La loi du 0-1 (ou *loi du tout ou rien*)

Comme nous l'avons mentionné précédemment, en 1933, la théorie de la mesure n'est pas encore parfaitement entrée dans les mœurs, en tout cas pas sous sa forme abstraite, et Fréchet, quand il va découvrir la monographie du mathématicien russe, va être décontenancé par la forme très abstraite que prennent certaines démonstrations, comme celle de la loi dite du 0-1 que Kolmogorov a placée en annexe de son ouvrage. Cette loi a été énoncée indépendamment, en particulier par Lévy en 1934 (date à laquelle il ne connaissait pas encore les *Grundbegriffe*), et il est intéressant de comparer les deux approches de ce résultat, ce que nous allons faire à titre d'illustration du caractère fortement synthétisant de l'axiomatique proposée par Kolmogorov. Pour le confort de la lecture, nous utiliserons le vocabulaire et les notations d'aujourd'hui, ne gardant que l'esprit des deux preuves.

**Théorème 2.1** *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires réelles indépendantes. On pose* $\mathcal{G}_n = \sigma(X_n, X_{n+1}, \dots)$ *(la tribu engendrée*[2] *par* $X_n, X_{n+1}, \dots$*), et* $\mathcal{G} = \bigcap_{n\geq 1} \mathcal{G}_n$ *(la* « tribu de queue »*).*
*Alors, tout élément de* $\mathcal{G}$ *est de probabilité 0 ou 1.*

*Preuve de Kolmogorov* : Il s'agit de la preuve classiquement enseignée aujourd'hui. Soit $A$ un élément de $\mathcal{G}$. Supposons que $P(A) > 0$ et notons $P_A$ la probabilité conditionnelle sachant $A$. Par l'hypothèse d'indépendance des $X_k$, quel que soit $B \in \mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$, $B$ est indépendant des éléments de $\mathcal{G}_{n+1}$ et donc de $A$, et on a :

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

De ce fait, les probabilités $P_A$ et $P$ coïncident sur toutes les $\mathcal{F}_n$, et donc sur l'algèbre de Boole $\bigcup_{n\geq 1} \mathcal{F}_n$ et donc, par le théorème de classe monotone, sur la tribu qu'elle engendre, c'est-à-dire $\mathcal{F} = \sigma(X_1, X_2, \dots, X_n, \dots)$. Comme en particulier $A \in \mathcal{F}$, on a $P_A(A) = P(A)$, c'est-à-dire $P(A) = 1$. $\square$

*Preuve de Lévy* : Lévy se contente en fait de démontrer le résultat dans le cas où les variables $X_n$ suivent une loi uniforme sur [0,1]. Dans ce cas, le tirage d'une réalisation de la suite $(X_n)_{n\geq 1}$ peut être conçu comme celui d'un point dans un cube de côté 1 à une infinité de dimensions, la loi de probabilité étant donnée par la mesure de Lebesgue. L'argument de Lévy repose alors sur une observation qu'il affirme comme évidente et qui est équivalente en fait à un résultat de classe monotone : il fait remarquer (nous employons le formalisme moderne) que pour tout événement $A$ de la tribu $\mathcal{F} = \sigma(X_1, X_2, \dots, X_n, \dots)$ (qui s'écrit donc sous la forme $[(X_1, X_2, \dots, X_n, \dots) \in B]$, où $B$ est un ensemble mesurable de $\mathbb{R}^{\mathbb{N}}$), et pour tout $\varepsilon > 0$, on peut trouver $n > 0$ et $D_n \in \mathcal{F}_n = \sigma(X_1, X_2, \dots, X_n)$ tels que[3] $P(D_n \Delta A) < \varepsilon$. En fait, sa justification est de dire que les ensembles mesurables dans le cube infini-dimensionnel sont obtenus par « *les constructions de M. Lebesgue* » à partir des « intervalles » du cube, qui sont les ensembles du type $]a_0, b_0[\times]a_1, b_1[\times \cdots \times]a_n, b_n[\times[0,1] \times [0,1] \dots$, de la même façon que les boréliens de $\mathbb{R}$ sont construits à partir des intervalles ouverts réels ; et l'on sait que pour n'importe quel élément $A \in \mathcal{B}([0,1])$, il existe une collection finie d'intervalles $]\alpha_0, \beta_0[, \dots, ]\alpha_m, \beta_m[$ telle que $\lambda(A \Delta \cup_{k=0}^{m} ]\alpha_k, \beta_k[) < \varepsilon$ (où $\lambda$ est la mesure de Lebesgue dans $\mathbb{R}$).
Soit maintenant $E$ un élément de $\mathcal{G}$. Notons que l'indépendance des $(X_n)_{n\geq 1}$ permet d'écrire que $\forall n$, $P(E \mid \mathcal{F}_n) = P(E)$. Soient $N$ et $D_N \in \mathcal{F}_N$ tels que $P(E \Delta D_N) < \varepsilon$ (et donc, en particulier, $P(D_N) > P(E) - \varepsilon$). On a alors :

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

[2] C'est-à-dire la plus petite tribu qui les rende toutes mesurables.
[3] $\Delta$ désigne la différence symétrique : $D_n \Delta A = (D_n \cup A) \setminus (D_n \cap A)$.

Or, $P(E^c \mid \mathcal{F}_N) = 1 - P(E \mid \mathcal{F}_N) = 1 - P(E) = P(E^c)$, donc $P(E^c \mid D_N) = P(E^c)$, d'où

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

et donc $P(E)(1 - P(E)) < 2\varepsilon$. C'est vrai pour tout $\varepsilon > 0$, donc $P(E)(1 - P(E)) = 0$. $\square$

# 3 Théorèmes limites et séries

La direction dans laquelle Kolmogorov va aborder ses premiers travaux de probabilités, sans doute guidé par son aîné Khinchin[4], s'inscrit dans la continuité d'études qui tout au long du dix-neuvième siècle avaient précisé les conditions de validité des théorèmes limites (en particulier la loi des grands nombres) pour des sommes de variables aléatoires. Le premier papier, qui date de 1925 [19] et qui est le seul article cosigné par Kolmogorov et Khinchin, est remarquable en cela qu'il introduit un certain nombre des techniques qui seront à la base de certains développements ultérieurs de la théorie des probabilités, en particulier dans l'étude des résultats de convergence des martingales. Ce premier travail concerne la convergence des séries de variables aléatoires indépendantes. Le résultat principal s'énonce ainsi (en termes modernes) :

**Théorème 3.1** *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes réelles centrées (i.e. d'espérance nulle). On suppose que $\sum_{n\geq 1} E(X_n^2) < +\infty$. Alors $\sum_n X_n$ converge presque sûrement (p.s.), c'est-à-dire avec probabilité 1.*

La preuve proposée par Kolmogorov repose sur une célèbre inégalité qui porte aujourd'hui son nom :

**Lemme 3.2 (Inégalité de Kolmogorov)** *Posons $S_n = X_1 + \cdots + X_n$. Alors*

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

*Démonstration.* On écrit

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

où

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

Noter que les $A_p$ forment une partition de tout l'espace de probabilités et que pour $1 \leq p \leq n$, $S_n - S_p$ est indépendante de $S_p \mathbb{1}_{A_p}$ et d'espérance nulle. De ce fait, pour $1 \leq p \leq n$,

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

et sommant sur $p$, $E(S_n^2) \geq \varepsilon^2 P(\max_{1\leq k\leq n} \mid S_k \mid \geq \varepsilon)$. $\square$

[4]On orthographie aussi Khintchine.

Ce n'est que quelques années plus tard, dans une note aux *Comptes Rendus de l'Académie des Sciences* (CRAS) de Paris de 1930 [22], que Kolmogorov tirera du résultat précédent sa conséquence la plus célèbre, à savoir cette désormais classique version de la loi forte des grands nombres :

**Corollaire 3.3** *Supposons que les variables $X_n$ soient indépendantes et centrées. On pose $E(X_n^2) = b_n$ et on suppose que $\sum_{n=1}^{\infty} \frac{b_n}{n^2} < +\infty$. Alors $\sigma_n = \frac{S_n}{n} \to 0$, p.s.*

*Démonstration.* Notons tout d'abord que pour tout $N > 0$ fixé,

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

où $\overline{\lim}_n$ désigne la limite supérieure pour $n \to +\infty$. Par ailleurs, il est évident que :

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

Donc, pour tout $\varepsilon > 0$ et pour tout $N > 0$ :

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

Par hypothèse, le dernier terme peut être rendu aussi petit que l'on veut, et donc $\overline{\lim}_n \sigma_n = 0$ p.s. $\square$

**Remarque 3.4** *L'indépendance des variables $X_n$ intervient à deux reprises dans cette démonstration : quand on applique l'inégalité de Kolmogorov, et quand on écrit $E\left((S_{2^{m+1}} - S_{2^N})^2\right) = \sum_{k=2^N+1}^{2^{m+1}} b_k$ (additivité de la variance pour des variables non corrélées).*

Comme mentionné plus haut, le résultat du lemme 3.2 et sa preuve se transposent directement au cas des martingales discrètes[5] :

[5]Rappelons qu'une martingale discrète est une suite $(M_n)_{n \geq 1}$ de variables aléatoires intégrables telle que, pour tout $n$, l'espérance de $M_{n+1}$, connaissant toutes les valeurs antérieures, soit égale à la dernière d'entre elles : $E\left(M_{n+1} \middle| M_1, \ldots, M_n\right) = M_n$. (La martingale est dite de carré intégrable, ou etc., si chacune des $M_n$ l'est.)

**Corollaire 3.5** *Soit $(M_n)_{n\geq 1}$ une martingale de carré intégrable telle que $E(M_n) = 0$. Alors*

$$P(\max_{1\leq k\leq n} \mid M_k \mid\geq \varepsilon) \leq \frac{E(M_n^2)}{\varepsilon^2}. \tag{2}$$

On démontre aisément que (2) se renforce quelque peu, sous la forme de l'importante *inégalité de Doob* : si $(M_n)_{n\geq 1}$ est une martingale de carré intégrable telle que $E(M_n) = 0$, on a

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

Comme on le sait, la théorie des martingales a depuis les travaux de Doob envahi la scène des probabilités contemporaines. À titre d'illustration de la puissance de l'inégalité (2) et de la notion de martingale, montrons le résultat suivant (qui généralise le théorème 3.1) et son corollaire.

**Proposition 3.6** *Soit $(M_n)$ une martingale $L^2$ (c'est-à-dire une martingale de carré intégrable) telle que $\sup_n E(M_n^2) < +\infty$. Alors $M_n$ converge, à la fois dans $L^2$ et p.s., vers une variable aléatoire $M$.*

*Démonstration.* Il est facile de montrer la convergence dans $L^2$ en prouvant que $(M_n)_{n\geq 1}$ est une suite de Cauchy dans $L^2$. Le lecteur intéressé pourra se reporter à n'importe quel cours élémentaire sur les martingales. Déduisons-en ici la convergence p.s.
Soit $\varepsilon > 0$. Pour chaque $p > 0$, posons $V_p = \sup_{n\geq p} \mid M_n - M_p \mid$. Comme $(M_n - M_p)_{n\geq p}$ est une martingale de carré intégrable, on a par (2) pour tout $N$

$$P(\max_{p\leq k\leq N} \mid M_k - M_p \mid\geq \varepsilon) \leq \frac{E((M_N - M_p)^2)}{\varepsilon^2}. \tag{3}$$

Comme signalé ci-dessus, $(M_k)_{k\geq 0}$ est une suite de Cauchy dans $L^2$ et donc, pour chaque $m > 0$ donné, on peut choisir $p_m$ tel que $\forall N \geq p_m$, $E((M_N - M_{p_m})^2) \leq \frac{\varepsilon^2}{2^m}$. De ce fait, passant à la limite dans (3) quand $N \to +\infty$ : $P(V_{p_m} \geq \varepsilon) \leq \frac{1}{2^m}$. Par le lemme de Borel-Cantelli, p.s. il existe $m > 0$ tel que $V_{p_m} < \varepsilon$ i.e. $\forall n \geq p_m, \mid M_n - M_{p_m} \mid< \varepsilon$, ce qui revient à dire que p.s. $(M_n)$ est une suite de Cauchy dans $\mathbb{R}$. $\square$

**Corollaire 3.7** *Soit $(Z_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes de même loi de Bernoulli $P(Z = 1) = P(Z = -1) = \frac{1}{2}$. On considère la marche aléatoire sur $\mathbb{Z}$ : $S_0 = 0, S_n = Z_1 + \cdots + Z_n$. Soient $a > 0$ un entier et $\tau = \inf\{n \geq 0, S_n = a\}$ le premier temps de passage en $a$. Alors la transformée de Laplace de la loi de $\tau$ est donnée, pour tout $\theta \geq 0$ par*

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

*Démonstration.* Nous n'indiquons que les grandes lignes de la preuve, laissant les détails au lecteur intéressé (voir par exemple [2]). Soit $X_n^\theta = \dfrac{e^{\theta S_n}}{(\cosh\theta)^n}$. On vérifie qu'il s'agit d'une martingale et que[6] $(X^\theta_{n\wedge\tau})_{n\geq 1}$ est une martingale $L^2$, qui converge p.s. et dans $L^2$ vers la variable $W^\theta = \frac{e^{\theta a}}{(\cosh\theta)^\tau}\mathbb{1}_{\tau<+\infty}$. Par passage à la limite $\theta \to 0$, possible par convergence dominée, on obtient que $P(\tau < +\infty) = 1$, puis le résultat cherché. $\square$

En 1924, Khinchin [18] avait démontré un résultat qui apportait une précision radicale à la loi des grands nombres, la *loi du logarithme itéré*. La généralisation du résultat de Khinchin par Kolmogorov en 1929 ([21]) fut une de ses plus grandes réalisations.

**Théorème 3.8** *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles indépendantes. On suppose que pour tout $n$, $E(X_n) = 0$ et $b_n = E(X_n^2) < +\infty$. On pose $B_n = \sum_{k=1}^n b_k$. ($B_n = E(S_n^2)$.) Si $B_n \to +\infty$ et $\mid X_n \mid \leq M_n = o(\sqrt{\frac{B_n}{\ln\ln B_n}})$, on a p.s.*

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

La démonstration de Kolmogorov reste encore aujourd'hui très actuelle, en cela qu'elle introduit des techniques, en particulier de grandes déviations, qui sont devenues fondamentales dans l'étude de nombreux phénomènes limites des probabilités. Nous ne démontrerons que la partie la moins technique du résultat, laissant le lecteur consulter l'un des innombrables textes qui présentent la démonstration complète (par exemple [6]).
Posant, pour $\varepsilon > 0$, $\phi^\varepsilon(n) = (1+\varepsilon)\sqrt{2B_n \ln\ln B_n}$, nous allons montrer que $P(S_n \geq \phi^\varepsilon(n)$ infiniment souvent$) = 0$. D'après le lemme de Borel-Cantelli, il suffit de montrer que pour une suite d'entiers $n_k \uparrow +\infty$ bien choisie, on a

$$\sum_{k=1}^\infty P(\max_{n\leq n_k} S_n \geq \phi^\varepsilon(n_{k-1})) < \infty. \tag{4}$$

Comme mentionné plus haut, le résultat va alors découler du lemme suivant qui donne des estimations de grandes déviations pour la suite $(S_n)$.

**Lemme 3.9** *Soit $x \geq 0$.*
*(i) Si $x \leq B_n/M_n$ alors $P(S_n > x) < e^{-(\frac{x^2}{2B_n})(1-\frac{xM_n}{2B_n})}$*
*(ii) Si $x \geq B_n/M_n$, alors $P(S_n > x) < e^{-\frac{x}{4M_n}}$*
*(iii) $P(\max_{1\leq k\leq n} S_k \geq x) \leq 2P(S_n > x - \sqrt{2B_n})$.*

*Démonstration.*
On fixe $n$ et pour simplifier l'écriture, l'indice $n$ sera omis dans la suite.
Soit $a > 0$ tel que $aM \leq 1$. Alors,

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

[6] $n\wedge\tau$ signifie $\min\{n,\tau\}$, le plus petit des nombres $n$ et $\tau$.

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

Puisque $P(S > x) < E(\frac{e^{aS}}{e^{ax}})$ (pour tout $a > 0$), on obtient l'inégalité $P(S > x) < \exp[-ax + \frac{a^2B}{2}(1+\frac{aM}{2})]$ d'où l'on déduit aisément les points (i) et (ii) en prenant successivement $a = \frac{x}{B}$ et $a = \frac{1}{M}$.
Pour le point (iii), posons $U = \max_{1\leq k\leq n} S_k$, et notons que $(U \geq x)$ est la réunion des événements $E_k = (S_1 < U, \ldots, S_{k-1} < U, S_k = U \geq x)$ pour $1 \leq k \leq n$.
On a alors

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

Mais $S - S_k$ est indépendante de $E_k$, et donc cette dernière expression est aussi

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

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

d'où finalement

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

Du lemme 3.9, on déduit (4) pour des $n_k$ bien choisis. En effet, choisissons-les tels que pour tout $k$, $B_{n_{k-1}} \leq (1+\tau)^k \leq B_{n_k}$. Par le lemme 3.9 (i)-(ii), on obtient, en utilisant les hypothèses, pour tout $\mu > 0$, et $k$ assez grand (tel que $M_{n_k} < \frac{\mu}{1+\varepsilon}\sqrt{2B_{n_k}/\ln\ln B_{n_k}}$) :

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

Choisissant alors $\mu$ tel que $(1+\varepsilon)^2(1-\mu) > 1$, on a $\sum_{k=1}^{\infty} P(S_{n_k} > \phi^{\varepsilon}(n_k)) < \infty$. On conclut alors en appliquant le lemme 3.9 (iii) et le fait que $\frac{\sqrt{2B_{n_k}}}{\phi^{\varepsilon}(n_k)} \to 0$. $\Box$

# 4 Processus en temps continu

Au début des années 1930, un grand nombre des travaux probabilistes de l'école soviétique s'orientent vers l'étude des processus stochastiques en temps continu, répondant ainsi en particulier aux besoins de la physique ou visant à décrire certains « phénomènes sociaux ». L'axiomatisation due à Kolmogorov, commentée plus haut dans cet exposé, apporta un élément essentiel à l'établissement de cette théorie. Le théorème de construction d'une mesure de probabilité sur un espace de dimension infinie montre que la loi d'un processus stochastique en temps continu est déterminée de manière unique à partir de la famille des lois marginales fini-dimensionnelles du processus en question.

## 4.1 L'équation de Chapman-Kolmogorov

Le premier ensemble de processus vers lequel Kolmogorov, en bon héritier de l'école russe de probabilités, se tourne naturellement, est celui des processus de Markov, c'est-à-dire ceux qui satisfont la propriété (dite de Markov) d'indépendance du futur vis-à-vis du passé conditionnellement à la connaissance du présent. L'article *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung* [23] fixe de manière définitive les fondements analytiques de la théorie des processus de Markov.

Dans cet article fondamental paru en 1931, toute l'étude du processus se concentre autour de la fonction :

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

qui représente la probabilité pour qu'au temps $t$ le phénomène aléatoire soit dans l'un des états de l'ensemble $A$, s'il se trouve dans l'état $x$ au temps $s$, antérieur à $t$ ($0 \leq s < t$). Les hypothèses de mesurabilité nécessaires étant supposées satisfaites, cette fonction doit vérifier l'équation intégrale

$$P(s, x, t, A) = \int_E P(s, x, u, dy)\, P(u, y, t, A)\,, \quad \text{pour tout } u \in ]s, t[, \tag{5}$$

où $E$ est l'ensemble des états possibles du processus. L'équation (5), communément appelée aujourd'hui équation de Chapman-Kolmogorov (Chapman en effet l'avait posée dans un mémoire [8] sur le mouvement brownien en 1928), est la traduction analytique de la propriété de Markov et les mesures $P(s, x, t, dy)$ représentent les probabilités de transition du processus : si on note $(X_t)_{t\geq 0}$ ce processus, alors, pour tout $A$ mesurable (pour la mesure $dy$) :

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

Toutefois, l'article de Kolmogorov étant purement analytique, comme son titre l'indique, celui-ci ne fait pas état des réalisations trajectorielles du processus aléatoire.

L'équation (5) ne peut être entièrement résolue de manière explicite dans l'état trop global où elle est posée. Kolmogorov cherche donc des conditions de régularité sur les probabilités $P(s, x, t, dy)$ qui permettraient de lui donner une forme plus abordable. Désireux d'utiliser les nouvelles techniques d'analyse liées à l'intégrale de Lebesgue, il se place naturellement dans l'hypothèse où $P(s, x, t, dy)$ est absolument continue, de densité $f(s, x, t, y) \geq 0$, par rapport à la mesure de Lebesgue.

L'équation (5), vérifiée par les probabilités de transition $P(s,x,t,dy)$, se traduit sur leurs densités par :

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

pour tout $u \in ]s,t[$ et tout $y \in \mathbb{R}$. De ce fait, pour obtenir des conditions locales à partir de (5), l'idée naturelle est de faire un développement de Taylor de $f$, ce qui nécessite des conditions de régularité sur $f$ et des hypothèses sur les moments.
Kolmogorov demande que pour tous $s,t,y$, $f(s,x,t,y)$ admette des dérivées du troisième ordre en $x$ et en $y$ qui sont uniformément bornées en $s$ et $t$, sur tout ensemble du type $\{s,t\ : s-t>k\}$, $k>0$. Par ailleurs, sous les hypothèses suivantes sur les moments :

$$
\text{pour tout } t \geq 0, \quad \lim_{\Delta \to 0} \int_{-\infty}^{\infty} |y-x|^i\, f(t,x,t+\Delta,y)\,dy = 0\,, \quad i=1,2,3, \tag{6}
$$

$$
\lim_{\Delta \to 0} \frac{\int_{-\infty}^{\infty} |y-x|^3\, f(t,x,t+\Delta,y)\,dy}{\int_{-\infty}^{\infty} |y-x|^2\, f(t,x,t+\Delta,y)\,dy} = 0\,, \tag{7}
$$

il montre l'existence des limites

$$
A(s,x) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (y-x) f(s,x,s+\Delta,y)\,dy\,, \tag{8}
$$

$$
B^2(s,x) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (y-x)^2 f(s,x,s+\Delta,y)\,dy\,, \tag{9}
$$

qu'il appelle respectivement la moyenne infinitésimale et la variance infinitésimale du processus et qui seront plus tard connues, dans le cas des diffusions, sous le nom de coefficient de dérive et de coefficient de diffusion. De l'équation (5), de l'existence des limites (8) et (9) et sous l'hypothèse de différentiabilité de $f$ mentionnée précédemment, Kolmogorov tire alors les deux équations différentielles suivantes :

$$
\frac{\partial}{\partial s} f(s,x,t,y) = -A(s,x)\frac{\partial}{\partial x} f(s,x,t,y) - B^2(s,x)\frac{\partial^2}{\partial x^2} f(s,x,t,y)\,, \tag{10}
$$

$$
\frac{\partial}{\partial t} f(s,x,t,y) = -\frac{\partial}{\partial y}[A(t,y)f(s,x,t,y)] + \frac{\partial^2}{\partial y^2}[B^2(t,y)f(s,x,t,y)]\,. \tag{11}
$$

L'importance de ces équations est telle que l'on peut considérer qu'elles sont à l'origine de la théorie moderne des processus stochastiques. Donnons à titre d'exemple les principaux arguments de la démonstration de la première d'entre elles, que l'auteur nomme *première équation différentielle fondamentale* et qui est appelée aujourd'hui *équation backward* (la seconde étant l'*équation forward*).

*Démonstration de l'équation* (10).

En appliquant la formule de Taylor-Lagrange à l'ordre 3 en la variable $z$ à la fonction $f(s+\Delta, z, t, y)$ et aux points $x$ et $z$, on obtient pour $s < s+\Delta < t$,

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

avec $\alpha = x + c(z-x)$, pour une valeur $c$ telle que $0 < c < 1$. Ainsi, en adoptant les notations

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

on peut écrire, sous l'hypothèse de bornitude de la dérivée de troisième ordre, que pour une valeur $C$ indépendante de $\Delta$ et pour un $\theta$ tel que $|\theta| < C$,

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

ce qui conduit immédiatement à la formule de différence finie

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

Il reste à noter pour conclure que sous les hypothèses faites, le rapport $c(s,x,\Delta)/\Delta$ tend vers 0 lorsque $\Delta$ tend vers 0. □

Comme nous l'avons déjà fait remarquer, l'étude de « mouvements aléatoires » dont la loi est régie par l'équation (5) avait déjà été esquissée par Chapman en 1928 ([8]) dans un contexte de physique théorique. Le nom d'équation de « Chapman-Kolmogorov » ne doit pas laisser croire toutefois que ce fut la seule occasion, avant l'article de 1931, où cette équation apparut. Kolmogorov lui-même, dans cet article, mentionne un cas particulier étudié par Louis Bachelier dès 1900 ([1]). Il souligne, dans une section qu'il consacre au travail de Bachelier, que l'équation (11) avait été écrite dans son travail de 1900 sans toutefois y être démontrée, dans le cas où le processus est homogène en espace, c'est-à-dire lorsque les densités $f(s,x,t,y)$ ne dépendent que de $s$, $t$ et de la différence $y-x$.

Les suites de l'article *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung* [23] ne tardèrent pas à voir le jour de la part d'autres probabilistes comme Bernstein, auquel les équations de Kolmogorov inspirèrent sa théorie des équations différentielles stochastiques en 1932. Cependant, celle-ci s'appuie sur le modèle discret et ne permet d'obtenir que des solutions faibles dans le cas continu. Un autre travail important fut celui de Wolfgang Doeblin, effectué en 1940. Doeblin l'envoya du Front (où il laissa la vie) à l'Académie des Sciences de Paris, sous la forme d'un pli cacheté qui ne fut ouvert qu'en 2000 ([12])[7]. Dans ce manuscrit, Doeblin, très en avance sur son temps, considère l'aspect trajectoriel du processus stochastique. Plus exactement, il fait le lien entre le point de vue strictement analytique de Kolmogorov et celui de Lévy qui se concentre essentiellement sur la construction trajectorielle et les propriétés fines des processus, et spécialement du mouvement brownien, par des méthodes purement probabilistes qui laissèrent souvent perplexes ses contemporains. Doeblin construit des équations très proches des équations différentielles stochastiques d'Itô établies vingt ans plus tard et dont les solutions sont des mouvements browniens avec une variable temporelle modifée : si la loi de $(X_t)$ satisfait à l'équation (5), alors

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

où $\beta$ est un mouvement brownien réel et $H$ le changement de temps $H(t) = \int_0^t B^2(s, X_s)\, ds$. Cette vision trajectorielle des processus offre alors bien plus que les équations analytiques *forward* et *backward* (10) et (11). Elle permet à Doeblin d'établir des résultats de régularité des trajectoires, de comparaison des solutions, des propriétés de logarithme itéré, des théorèmes limites centraux fonctionnels et surtout une version préliminaire de la formule de changement de variable qu'Itô obtiendra quelques années plus tard ([15]) et qui inaugurera l'ère du calcul stochastique proprement dit.

Pour établir sa formule, Doeblin considère une fonction $\varphi(t, x)$ de classe $\mathcal{C}^{1,2}$ (c'est-à-dire de classe $\mathcal{C}^1$ par rapport à $t$ et $\mathcal{C}^2$ par rapport à $x$) et croissante par rapport à $x$, ce qui assure de manière très simple que la loi du processus $Y_t = \varphi(t, X_t)$ est solution de l'équation de Kolmogorov dès lors que la loi de $(X_t)$ est elle-même solution. Il montre alors que $(Y_t)$ satisfait

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

où $\gamma$ est un mouvement brownien réel et

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

Il fallut attendre la construction de l'intégrale stochastique d'Itô pendant et après la guerre pour voir les solutions des équations (10) et (11) sous un aspect nouveau. Il s'agissait

[7]Les lecteurs moins intéressés par les aspects techniques pourront se contenter de lire l'article [7].

d'établir des équations satisfaites[8] par le processus lui-même et non plus seulement par ses probabilités de transition. Si $\beta$ désigne un mouvement brownien réel et si $A(t,x)$ et $B(t,x)$ sont des fonctions définies comme au début de cette section, alors les probabilités de transition du processus $X$ solution de l'équation différentielle stochastique

$$dX_t = A(t, X_t)\, dt + B(t, X_t)\, d\beta_t\,, \tag{12}$$

satisfont les équations (5), (10) et (11). Un outil indispensable à l'obtention de ce résultat fut la formule d'Itô citée plus haut. Sous sa forme actuelle la plus courante, celle-ci s'énonce comme suit : si $\varphi(t,x)$ est une fonction de classe $\mathcal{C}^{1,2}$ alors

$$\begin{aligned} d\varphi(t, X_t) =& B(t, X_t)\frac{\partial}{\partial x}\varphi(t, X_t)\, d\beta_t \\ &+ \left(\frac{\partial}{\partial t}\varphi(t, X_t) + A(t, X_t)\frac{\partial}{\partial x}\varphi(t, X_t) + \frac{1}{2}B^2(t, X_t)\frac{\partial^2}{\partial x^2}\varphi(t, X_t)\right)\, dt\,, \end{aligned} \tag{13}$$

où $X$ est solution de l'équation différentielle stochastique (12). Ce sont là les fondements du calcul stochastique qui allait connaître une multitude de développements pendant toute la seconde moitié du vingtième siècle.

À partir des années 1950, la théorie des martingales de Doob [13], étendue par la suite par Meyer et son école, allait permettre d'affaiblir les conditions jusqu'alors imposées aux fonctions $A(t,x)$ et $B(t,x)$ pour assurer la construction de processus stochastiques. Une remarque essentielle dans cette direction fut l'observation que sous des hypothèses naturelles de bornitude et de lipschitzianité locales, l'équation (12) a une solution unique en loi, dans le sens où si $\beta'$ est un autre mouvement brownien (éventuellement défini sur un autre espace de probabilités), une solution $X'$ de :

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

suit la même loi que $X$. Ceci permit dans les années 1970 de définir une notion de solution faible pour l'équation (12) qui ne fait plus intervenir le choix spécifique d'un mouvement brownien particulier. Un célèbre théorème de Yamada et Watanabe [42] affirme que l'unicité trajectorielle (celle qui correspond à une équation dirigée par un mouvement brownien fixé) entraîne l'unicité en loi des solutions faibles. Une formulation puissante fut proposée par Stroock et Varadhan [37] sous la forme des *problèmes de martingales.* Le générateur associé au processus markovien $(X_t)$ solution de (12) est l'opérateur $L$ sur $\mathcal{C}^{1,2}$ défini par :

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

La formule d'Itô permet d'exprimer cette définition en disant que pour toute $f \in \mathcal{C}^{1,2}$,

$$(M_t) = \left(f(t, X_t) - \int_0^t Lf(s, X_s)ds\right)_{t\geq 0} \tag{14}$$

[8]Dans le sens très particulier des équations différentielles stochastiques qui fait un appel central à la notion d'intégrale stochastique d'Itô.

est une martingale locale[9].
On est alors naturellement conduit à définir une solution de (12) de la façon suivante. Soit $\mathcal{C} = \mathcal{C}(\mathbb{R}^+, \mathbb{R})$ l'ensemble des fonctions continues de $\mathbb{R}^+$ dans $\mathbb{R}$. On définit les *projections canoniques* sur $\mathcal{C}$ par $X_t(\omega) = \omega(t)$ pour $\omega \in \mathcal{C}$ et la *filtration canonique* par $\mathcal{C}_t = \sigma(X_s, s \leq t)$, $t \geq 0$. Une solution de (12) est alors une probabilité $P$ sur $(\mathcal{C}, (\mathcal{C}_t))$ telle que sous $P$ le processus défini par (14) soit une martingale locale.
L'aspect le plus intéressant du travail de Stroock et Varadhan est que sous des conditions très faibles (en gros, la continuité des fonctions $A$ et $B$), le problème de martingales précédent admet une solution $P$. Lorsqu'il y a unicité de la probabilité solution, sous cette probabilité, le processus canonique satisfait les propriétés markoviennes qui furent à l'origine des études de Kolmogorov. Le lecteur intéressé pourra sur ces sujets consulter avec intérêt l'important traité de Jacod [17].

## 4.2 Processus à accroissements indépendants et stationnaires

Parmi les processus dont la loi vérifie l'équation de Chapman-Kolmogorov (5), il en est une famille très importante que Lévy commence à étudier dès le début des années 1930 : c'est celle des processus dont les fonctions $f(s, x, t, y)$ sont homogènes en temps et en espace, i.e. ne dépendent que des différences $t - s$ et $y - x$. Autrement dit, il s'agit des *processus à accroissements indépendants et stationnaires* dont Kolmogorov tente de caractériser la loi dans un article paru en deux parties en 1932 : *Sulla forma generale di un processo stocastico omogeneo* [24] et *Ancora sulla forma generale di un processo omogeneo* [25]. Il considère simplement une « fonction du temps aléatoire » $X(\lambda)$, où $\lambda \geq 0$ représente la variable de temps, telle que pour tous $\lambda_1$ et $\lambda_2$, $(\lambda_2 \geq \lambda_1)$ la différence $X(\lambda_2) - X(\lambda_1)$ est indépendante de $(X(\lambda), \lambda \leq \lambda_1)$ et dont la loi ne dépend que de $\lambda_2 - \lambda_1$, i.e. en posant $\Delta = \lambda_2 - \lambda_1$,

$$\Phi_\Delta(x) = P(X(\lambda_2) - X(\lambda_1) < x) = P(X(\lambda_2 - \lambda_1) < x) .$$

Puis il remarque que la relation

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

est un cas particulier de l'équation (5). On peut toutefois noter que ces processus n'entrent pas à cette époque dans le cadre des processus de Markov, objet de l'étude précédemment décrite. Cette formalisation n'interviendra en fait que dans les années 1950.
Le but de l'article de 1932 est, selon Kolmogorov lui-même, de généraliser certains résultats dûs à Bruno de Finetti [14] au cas où les lois données par les fonctions de répartition $\Phi_\Delta$ admettent des moments du second ordre, $\int x^2\, d\Phi_\Delta(x) < \infty$. Nous utiliserons les notations :

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

[9] C'est-à-dire qu'il existe une suite de temps d'arrêt $(T_n)$ presque sûrement finis et croissant vers $+\infty$, telle que pour tout $n$, le processus $(M_{t \wedge T_n})$ soit une martingale.

Ainsi Kolmogorov obtient-il un cas particulier de la célèbre formule de Lévy-Khinchin : si $\psi_\Delta(t) = \int e^{itx}\, d\Phi_\Delta(x)$ alors $\psi_\Delta(t) = [\psi_1(t)]^\Delta$ et

$$\log \psi_1(t) = itm_1 - \frac{\sigma_0^2}{2} t^2 + \int_{-\infty}^{\infty} \left( e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2}\, G(dx)\,, \tag{15}$$

où $G$ est une mesure finie telle que $G(\{0\}) = 0$ et où $m_1 \in \mathbb{R}$, $\sigma_0^2 \geq 0$. Cette formule est communément attribuée à Lévy et Khinchin qui en obtinrent en 1934 ([28]) et 1937 ([20]) la version finale, bien que, comme nous le constatons, de Finetti et Kolmogorov l'eussent déjà établie dans des cas particuliers. Plus précisément, le résultat énoncé par Kolmogorov est le suivant.

**Théorème 4.1** *Lorsque la loi donnée par la fonction de répartition $\Phi_\Delta$ possède un moment d'ordre 2, on a*

$$\log \psi_1(t) = itm_1 - \frac{\sigma_0^2}{2} t^2 + \int_{-\infty}^{\infty} \pi(x,t)\, dF(x)\,, \tag{16}$$

*où $m_1 \in \mathbb{R}$, $\sigma_0^2 \geq 0$, $\pi(x,t) = (e^{itx} - 1 - itx)/x^2$ et où la mesure $dF(x)$ est définie par une fonction $F$, croissante et bornée.*

Les formules ci-dessus ne concernent que les lois unidimensionnelles de la fonction aléatoire $(X(\lambda))$ ; par conséquent, il serait plus juste de parler de la caractérisation des *lois indéfiniment divisibles*[10] plutôt que d'un résultat sur les processus à accroissements indépendants et stationnaires. Notons toutefois qu'à cette époque, cette terminologie n'avait pas cours.

*Démonstration de Kolmogorov* : Tout d'abord, on vérifie que $\psi_\Delta$ est continue par rapport à $\Delta$. En effet, pour $\Delta \leq 1/n$, on a $\sigma_\Delta^2 = \sigma_{1/n}^2 - \sigma_{1/n-\Delta}^2 \leq \sigma_{1/n}^2 = (1/n)\sigma_1^2$, et donc $\sigma_\Delta^2 \to 0$ lorsque $\Delta \to 0$. Par conséquent, $\psi_\Delta(t) \to 1$ lorsque $\Delta \to 0$. On conclut grâce à l'égalité $\psi_{\Delta_1+\Delta_2}(t) = \psi_{\Delta_1}(t)\psi_{\Delta_2}(t)$. La continuité de $\psi_\Delta$ en $\Delta$ permet de montrer que l'égalité

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

vraie pour $\Delta$ rationnel, est aussi vérifiée pour tout $\Delta$ réel. L'auteur en déduit (en invoquant la preuve de de Finetti) que

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

D'autre part, on a

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

[10]Une loi de probabilité $P$ est dite *indéfiniment divisible* si, pour tout $n$, elle est la $n$-ième puissance de convolution d'une loi $\mu_n$. Cela équivaut à dire que $P$ est la loi d'une somme de $n$ variables aléatoires réelles indépendantes et de même loi $\mu_n$.

Posons $F_\Delta(x) = \frac{1}{\Delta}\int_{-\infty}^{x} y^2\, d\Phi_\Delta(x)$, de sorte que

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

Un argument classique permet de justifier que pour toute suite $\Delta_n$ décroissant vers 0, il existe une sous-suite $\Delta_{n_k}$ telle que la suite $F_{\Delta_{n_k}}(x)$ converge lorsque $k$ tend vers $+\infty$, vers la fonction $F(x)$ en tout point $x$ où cette dernière est continue. Remarquons alors que $F$ est une fonction croissante telle que :

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

et compte tenu de ce qu'à $t$ fixé, $\pi(x,t) \to 0$ lorsque $x \to \pm\infty$, on obtient

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

ce qui entraîne que

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

Mais puisque $\sigma_1^2$ est fini, on a $\log\psi_1(t) = itm_1 - \frac{\sigma_1^2}{2}t^2 + o(t^2)$ $(t \to 0)$, et d'après le calcul précédent

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

Ceci implique en particulier que $F(\infty) - F(-\infty) = \sigma_1^2$. Enfin on déduit de ce qui précède que $F(-\infty) = 0$ et $F(\infty) = \sigma_1^2$ ; ainsi $F$ est-elle entièrement déterminée.
Montrons maintenant que pour toute fonction $F$ croissante (et continue à gauche), avec pour valeurs aux extrémités $F(-\infty) = 0$ et $F(+\infty) = \sigma_1^2 < +\infty$, la fonction $\psi_1$ donnée par $\log\psi_1(t) = itm_1 + \int_{-\infty}^{\infty}\pi(x,t)\, dF(x)$ est la fonction caractéristique d'une loi indéfiniment divisible. Considérons pour cela une fonction étagée $T$ ayant pour sauts :

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

en nombre fini $x_1, x_2, \ldots, x_n$. On suppose de plus que $T$ ne saute pas au point 0. On note $\sigma_1^2 = \omega_1 + \omega_2 + \cdots + \omega_n$ la somme de ces sauts. Posons aussi $\log\overline{\psi}_1(t) = itm_1 + \int_{-\infty}^{\infty}\pi(x,t)\, dT(x)$, $\eta = m_1 - \sum_k p_k$, $p_k = \frac{\omega_k}{x_k^2}$ et $\chi_k(t) = \exp(itx_k)$. Alors on vérifie facilement les égalités

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

On en déduit que

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

Remarquons que $\chi_k(t)$ et $\exp(it\eta\Delta)$ sont des fonctions caractéristiques et par conséquent $\psi_\Delta(t)$, en tant que produit de fonctions caractéristiques, est elle-même une fonction caractéristique. Pour conclure, remarquons que l'on peut trouver une suite de fonctions étagées $T_n$ telle que $\int_{-\infty}^{\infty} \pi(x,t)\, dT_n(x)$ converge vers $\int_{-\infty}^{\infty} \pi(x,t)\, dF(x)$ uniformément sur tout intervalle borné. Les fonctions caractéristiques correspondantes $\overline{\psi}_\Delta^{(n)}(t)$ convergent donc vers $\psi_\Delta(t)$. Ceci entraîne que $\psi_\Delta$ est une fonction caractéristique, d'une part, et que cette fonction vérifie $\psi_\Delta(t) = (\psi_1(t))^\Delta$, d'autre part. $\square$

Kolmogorov complète cette description de la loi des processus à accroissements indépendants en observant que si $P_1(x) = \int_x^\infty y^{-2}\, dF(y)$ et $P_2(x) = \int_{-\infty}^x y^{-2}\, dF(y)$ alors $P_1(x)\, d\lambda$ (respectivement $P_2(x)\, d\lambda$) est la probabilité pour que le processus $X(\lambda)$ ait un saut positif (respectivement négatif), de hauteur supérieure (respectivement inférieure) ou égale à $x$, pendant l'intervalle de temps $d\lambda$. Les mesures $P_1$ et $P_2$ sont respectivement les restrictions à $\mathbb{R}_+$ et à $\mathbb{R}_-$ de ce que l'on appellera par la suite la mesure de Lévy du processus $X(\lambda)$. Faisant suite aux travaux de Lévy, un certain nombre d'auteurs dans les années 1960 à 1980 tels que Skorokhod, Zolotarev, Blumenthal, Getoor, Ray, Taylor, Fristedt, Bingham, Pitman, Jacod,... ont étudié les propriétés fines des trajectoires des processus à accroissements indépendants et stationnaires (appelés depuis processus de Lévy). Puis au début des années 1990, les ouvrages de synthèse de Bertoin [5] et de Sato [32] ont suscité, parmi la communauté probabiliste, un regain d'intérêt pour l'étude des processus de Lévy. Celui-ci s'est encore renforcé depuis la découverte de nouveaux terrains d'applications, comme les mathématiques financières, où les modèles faisant intervenir les processus de Lévy permettent de compenser les défauts du modèle dit de Black-Scholes du mouvement brownien géométrique.

## 4.3 Critères de continuité et de compacité relative

Il est remarquable que Kolmogorov ait aussi développé un certain nombre d'outils pour l'étude des propriétés trajectorielles des processus. Parmi ceux-ci, un très efficace critère garantissant la continuité trajectorielle des processus porte son nom. Il fut trouvé par le mathématicien soviétique en 1934 et exposé la même année au Séminaire de l'Université de Moscou. Il ne fit cependant l'objet d'aucune de ses publications et ce fut Slutsky qui l'énonça et en fournit le premier une démonstration publiée dans le *Giornale dell'Istituto Italiano dei Attuari* en 1937 [36], en en attribuant la paternité à Kolmogorov. Vingt ans plus tard, Kolmogorov fit encore remarquer à Chentsov une extension de ce critère pour des processus discontinus que ce dernier publia en 1956 [9] et qui permet de conclure que le processus en question n'a pas de discontinuité du second ordre[11].

**Théorème 4.2** *Si une famille de variables aléatoires réelles* $(X_t,\ 0 \le t \le 1)$ *est telle qu'il existe trois constantes strictement positives* $\gamma$, $c$ *et* $\varepsilon$ *telles que :*

$$E(|X_t - X_s|^\gamma) \le c|t-s|^{1+\varepsilon}\,, \tag{4.3.1}$$

[11] Dans un ordre d'idées voisin, soulignons encore l'apport de Kolmogorov à la topologie de Skorokhod pour l'espace des trajectoires continues à droite et limitées à gauche [27].

*alors il existe une modification de $X$ qui est presque sûrement continue.*

Rappelons qu'une modification d'un processus $X$ est un processus $\tilde{X}$ tel que pour tout $t$ fixé, $\tilde{X}_t = X_t$, presque sûrement. La preuve suivante s'inspire de celle de [31], p. 26 ; voir aussi [11], vol. V, chap. XXIII, p. 332.

*Démonstration* : Soit $D_m$ l'ensemble des réels de [0,1] de la forme $2^{-m}i$, où $i = 0, 1, \ldots, 2^m$ et $D = \cup_m D_m$ l'ensemble des nombres diadiques de [0,1]. Soit $\Delta_m$ l'ensemble des couples $(s,t)$ de $D_m^2$ tels que $|t-s| = 2^{-m}$. On pose

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

Puisque pour tous $s,t \in \Delta_i$, $E(|X_s - X_t|^\gamma) \leq c\,(2^{-i})^{1+\varepsilon}$ et $\#\Delta_i = 2^i$, on a pour tout $i$,

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

Soient maintenant $s,t \in D$, $s \neq t$ et $m$ l'entier tel que $2^{-m-1} < |t-s| \leq 2^{-m}$ ; alors il existe des suites finies $s_m, s_{m+1}, \ldots, s_p = s$ et $t_m, t_{m+1}, \ldots, t_k = t$ telles que pour tout $i = m, m+1, \ldots, p$ et $j = m, m+1, \ldots, k$, on ait $(s_i, s_{i+1}) \in \Delta_{i+1}$, $(t_j, t_{j+1}) \in \Delta_{j+1}$, $(s_m, t_m) \in \Delta_m$ et

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

On en déduit les inégalités

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

Soit $\alpha \in [0, \varepsilon/\gamma[$. Définissons la variable aléatoire $M_\alpha = \sup\{|X_t - X_s|/|t-s|^\alpha \ : s,t \in D,\ s \neq t\}$. Alors on déduit de ce qui précède que :

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

Dans le cas où $\gamma \geq 1$, cette inégalité entraîne que

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

et lorsque $\gamma < 1$, la même inégalité s'applique à $E(M_\alpha^\gamma)$. En particulier, la variable $M_\alpha$ est finie et l'on a, pour tous $s, t \in D$, $|X_t(\omega) - X_s(\omega)| < K(\omega)|t-s|^\alpha$, où $K(\omega)$ est une constante indépendante de $s$ et de $t$. On en déduit que $X$ est uniformément continu sur $D$ (et même höldérien d'ordre $\alpha$). Le processus $X_{|_D}$ (i.e. $X$ restreint à $D$) se prolonge donc de manière unique en un processus continu sur $[0,1]$ :

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

qui est naturellement également höldérien d'ordre $\alpha$. Enfin, d'après l'hypothèse, pour tout $t \in [0,1]$, $\lim_{s \to t} X_s = X_t$ dans $L^1$, donc $\tilde{X}_t = X_t$, p.s. pour tout $t \in [0,1]$. $\square$

Comme on le voit, on a en fait montré un résultat plus fort : *sous les hypothèses précédentes, il existe une modification de $X$ dont les trajectoires sont presque sûrement höldériennes d'ordre $\alpha$ pour tout $\alpha \in [0, \varepsilon/\gamma[$.* En particulier, le mouvement brownien est (presque sûrement) localement höldérien d'ordre $\alpha$ pour tout $\alpha < 1/2$.
Le théorème 4.2 admet certaines extensions comme celle de Chentsov citée plus haut. Il existe aussi un critère du même genre pour des processus à plusieurs variables ($X_t$, $t \in [0,1]^d$) ; l'hypothèse 4.3.1 s'écrit dans ce cas sous la forme :

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

La preuve de ce résultat est très proche de celle que nous venons de donner et s'applique également à des processus multivariés à valeurs dans des espaces de Banach quelconques.
En 1970, Garsia, Rodemich et Rumsey ont proposé une autre généralisation de ce critère de continuité qui repose sur un raisonnement entièrement déterministe (voir [37], p. 47 ou encore [11], vol. V, chap. XXIII, p. 336). Énonçons brièvement leur résultat :
Soit $f$ une fonction borélienne définie sur une boule $B$ de $\mathbb{R}^d$, satisfaisant à la condition intégrale :

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

où $\Psi$ et $\varphi$ sont deux fonctions continues sur $\mathbb{R}_+$, nulles en 0, strictement croissantes et non bornées, telles que les intégrales

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

soient convergentes pour tout $x > 0$. Alors la fonction $f$ admet une version continue $\tilde{f}$ (i.e. $f(u) = \tilde{f}(u)$, p.s. au sens de la mesure de Lebesgue) qui possède sur la même boule $B$ le module de continuité

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

où la constante $k_d$ ne dépend que de la dimension $d$.
Il s'est avéré par la suite que ce résultat est un cas particulier des théorèmes de plongement de Sobolev. Le lecteur intéressé pourra consulter [11], vol. V, chap. XXIII, p. 334. La forme explicite ainsi obtenue de cette version raffinée du critère de Kolmogorov a pu

être appliquée à des résultats de continuité des temps locaux de processus de Lévy, et plus généralement des semi-martingales, voir [3] et [4]. Cette forme est à rapprocher de la technique des mesures majorantes de Fernique, Marcus, Talagrand,...

Comme on peut facilement l'imaginer, les applications du théorème 4.2 sont multiples et variées. L'une des plus importantes est sans doute le critère de compacité relative. Rappelons qu'une suite de mesures de probabilité est dite (faiblement) relativement compacte si de toute sous-suite, on peut extraire une nouvelle sous-suite qui converge faiblement. Selon le théorème de Prohorov, *une suite de mesures de probabilité* $(P_n)$ *sur un espace séparable et complet est relativement compacte si et seulement si elle est tendue*, c'est-à-dire si pour tout $\varepsilon > 0$, il existe un compact $K$ tel que $P_n(K) \geq 1 - \varepsilon$, pour tout $n$. Ce théorème présente un grand intérêt lorsqu'en particulier, l'espace de probabilités en question est l'ensemble des trajectoires continues sur la demi-droite positive. Notons alors $\mathcal{C}$ l'espace des fonctions continues sur la demi-droite positive et à valeurs dans $\mathbb{R}^d$, que l'on munit de la topologie de la convergence uniforme et de la tribu borélienne. Pour montrer qu'une suite de mesures $(P_n)$ sur $\mathcal{C}$ converge faiblement vers une mesure $P$, on sait qu'il suffit de vérifier qu'elle est relativement compacte et que ses marginales fini-dimensionnelles convergent vers les marginales correspondantes de $P$. Comme $\mathcal{C}$ est séparable et complet, la compacité relative est équivalente à la tension. La convergence des marginales fini-dimensionnelles se vérifie en général facilement. La compacité relative est souvent problématique. Un critère très utile est obtenu comme une conséquence directe du résultat de Kolmogorov énoncé au théorème 4.2. Nous renvoyons à [6] pour une démonstration.

**Corollaire 4.3** *Si une suite de processus continus* $(X_n(t),\, t \geq 0)$[12] *satisfait :*

*(i) la suite de v.a.* $(X_n(0))$ *est tendue,*

*(ii) il existe trois constantes strictement positives* $\gamma$, $c$ *et* $\varepsilon$ *telles que pour tout* $n$ *:*

$$E(|X_n(t) - X_n(s)|^\gamma) \leq c|t - s|^{d+\varepsilon}\,, \tag{4.3.2}$$

*alors la suite constituée par les lois des processus* $(X_n)$ *est relativement compacte.*

[12]Ceci équivaut à considérer une suite de variables aléatoires $(X_n)$ à valeurs dans $\mathcal{C}$.

# Références

Tous les articles de A. N. Kolmogorov cités dans la liste suivante se trouvent dans [35].


[1] L. Bachelier : Théorie de la spéculation, *Ann. Sci. École Norm. Sup.* **17**, 21-86 (1900).

[2] P. Baldi, L. Mazliak et P. Priouret : *Martingales et Chaînes de Markov*, Hermann, 2001.

[3] M.T. Barlow : Continuity of local times for Lévy processes. *Z. Wahrsch. Verw. Gebiete* **69**, 23-35 (1985).

[4] M.T. Barlow et M. Yor : Semimartingale inequalities via the Garsia-Rodemich-Rumsey lemma and application to local times. *J. Funct. Analysis.* **49**, 198-229 (1982).

[5] J. Bertoin : *Lévy Processes*, Cambridge University Press, Cambridge, 1996.

[6] P. Billingsley : *Probability and Measure*, Wiley, 3-ième édition, 1995.

[7] B. Bru et M. Yor : Wolfgang Doeblin, l'équation de Kolmogoroff. *La Recherche*, Juin 2003.

[8] S. Chapman : 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).

[9] N. N. Chentsov : Weak convergence of stochastic processes whose trajectories have non discontinuities of second kind and the « heuristic » approach to the Kolmogorov-Smirnov tests, Teop. Bep, **1**, 140-144 (1956).

[10] H. Cramér : *Random Variables and Probability distributions*, Cambridge, 1937.

[11] C. Dellacherie et P.A. Meyer : *Probabilités et Potentiel*, Hermann, 1975-92.

[12] W. Doeblin : Sur l'équation de Kolmogoroff (pli cacheté), édité et commenté par B. Bru et M. Yor, *C.R. Acad. Sci. Paris Sér. I Math.* **331**, 1033-1187 (2000).

[13] J. L. Doob : *Stochastic processes.* Wiley Classics Library. John Wiley & Sons, Inc., New York, 1990.

[14] B. de Finetti : Le funzioni caratteristiche di legge istantanea, *Rend. Acc. Lincei* **12**, 278-282 (1930).

[15] K. Itô : Stochastic Integrals, *Proc. Imp. Acad. Tokyo* **20** (1944), 519-524.

[16] K. Itô : *Selected papers*, édité par D.W. Stroock et S.R.S.Varadhan, Springer, 1986.

[17] J. Jacod : *Calcul stochastique et problèmes de martingales.* Lecture Notes in Mathematics, 714. Springer, Berlin, 1979.

[18] A.Y. Khinchin : Über einen Satz der Wahrscheinlichkeitrechnung, *Fund. Mat.* **6**, 9-20 (1924).

[19] A.Y. Khinchin et A.N. Kolmogorov : Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden, *Mat. Sb.,* **32**, 668-677 (1925).

[20] A.Y. Khinchin : Déduction nouvelle d'une formule de M. Paul Lévy, *Bull. Univ. Moscou Math-Mec* **1**, 1-5 (1937).

[21] A.N. Kolmogorov : Über das Gesetz des iterierten Logarithmus, *Math. Ann.* **101**, 126-135 (1929).

[22] A.N. Kolmogorov : Sur la loi forte des grands nombres, *CRAS* Paris, **191**, 910-912 (1930).

[23] A. N. Kolmogorov : Über die analytischen Methoden in der Wahrscheinlichkeitrechnung, *Math. Ann.* **104**, 415-458 (1931).

[24] A. N. Kolmogorov : Sulla forma generale di un processo stocastico omogeneo, *Atti Accad. Naz. Lincei Rend.* **15**, 805-808 (1932).

[25] A. N. Kolmogorov : Ancora sulla forma generale di un processo stocastico omogeneo, *Atti Accad. Naz. Lincei Rend.* **15**, 866-869 (1932).

[26] A. N. Kolmogorov : *Grundbegriffe der Wahrscheinlichkeitrechnung*, Springer, 1933.

[27] A. N. Kolmogorov : On the Skorohod convergence. *Teor. Veroyatnost. i Primenen.* **1**, 239-247 (1956).

[28] P. Lévy : Sur les intégrales dont les éléments sont des variables aléatoires indépendantes, *Ann. Sc. Norm. Pisa* **3**, 337-366 (1934).

[29] P. Lévy : *Théorie de l'Addition des Variables aléatoires*, Gauthier-Villars, 1937.

[30] L. Mazliak : Andrei Nikolaevitch KOLMOGOROV (1903-1987). Un aperçu de l'homme et de l'œuvre probabiliste. Cahiers du CAMS-EHESS, Paris, (2003). Disponible sur http://www.proba.jussieu.fr/users/lma/recherche.html

[31] D. Revuz et M. Yor : *Brownian motion and stochastic calculus*, Springer, 1991 ; 3-ième édition, 1999.

[32] K.I. Sato : *Lévy processes and infinitely divisible distributions.* Translated from the 1990 Japanese original. Cambridge Studies in Advanced Mathematics, 68. University Press, Cambridge, 1999.

[33] G. Shaefer et V. Vovk : The origins of Kolmogorov's Grundbegriffe, preprint, 2003. Disponible sur http://www.probabilityandfinance.com/

[34] A. N. Shiryaev : Kolmogorov : Life and creative activities, *Ann. Prob.* **17**, 866-944 (1989).

[35] A. N. Shiryaev : *Selected works of A. N. Kolmogorov, Vol. II (*traduit par G. Lindquist), Kluwer Academic Publications, 1992.

[36] E.B. Slutsky : Qualche proposizione relativa alla teoria delle funzioni aleatorie, *Gio. Ist. It. Att.* **8**, 183-199 (1937).

[37] D. W. Stroock et S.R.S. Varadhan : *Multidimensional diffusion processes*, Springer, 1979.

[38] D. W. Stroock : *Markov processes from Itô's perspective*, Princeton University Press, 2003.

[39] G. J. Székely : *Paradoxes in probability theory and mathematical statistics.* Mathematics and its Applications (East European Series) **15**. D. Reidel Publishing Co., Dordrecht, 1986.

[40] J. Van Plato : *Creating Modern Probability*, Cambridge Studies, 1994.

[41] N. Wiener : Differential space, *Jour. Math. and Phys.* **58**, 131-174 (1923).

[42] T. Yamada and S. Watanabe : On the uniqueness of solutions of stochastic differential equations II, *J. Math. Kyoto Univ.*, **11**, 553-563, (1971).