

# COMPOSITIO MATHEMATICA

PAUL LÉVY

## Processus strictement ou presque strictement markoviens

*Compositio Mathematica*, tome 14 (1959-1960), p. 172-193

[http://www.numdam.org/item?id=CM\\_1959-1960\\_\\_14\\_\\_172\\_0](http://www.numdam.org/item?id=CM_1959-1960__14__172_0)

© Foundation Compositio Mathematica, 1959-1960, tous droits réservés.

L'accès aux archives de la revue « Compositio Mathematica » (<http://www.compositio.nl/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

*Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
http://www.numdam.org/*

# Processus strictement ou presque strictement markoviens

Paul Lévy

## 1. Introduction

1° Nous considérerons en principe des fonctions aléatoires  $X(t)$  du temps  $t$ , dont les valeurs  $x$  sont des nombres réels, ou, dans certains exemples, des points d'une courbe fermée. Le lecteur vérifiera aisément que l'extension de certains résultats au cas où  $x$  est un point d'un espace à  $N$  dimensions ( $N < \infty$ ) ne présente aucune difficulté, à condition de convenir que  $x \leq y$  signifie dans ce cas que toutes les coordonnées de  $y - x$  sont  $\geq 0$ . Pour les définitions et remarques du 1° de cette introduction, on peut généraliser beaucoup plus, et supposer que même la variable  $t$  soit un point d'un espace abstrait quelconque.

Dans de nombreuses questions, il suffit de connaître la loi jointe des  $X(t_n)$  pour n'importe quelle suite finie de nombres  $t_n$ . On la connaît alors pour n'importe quelle suite infinie, même si l'ensemble des  $t_n$  est partout dense sur l'axe des  $t$ . Nous dirons dans ce cas qu'on connaît la *définition faible* de la fonction  $X$ . Dans d'autres cas, on a besoin de définitions *plus complètes*. Mais quand peut-on considérer une définition comme *complète*?

Dans l'axomatique de A. Kolmogorov et J. L. Doob, pour définir une fonction aléatoire, on se donne un ensemble  $\Omega$ , dont chaque élément  $\omega$  correspond à une détermination possible de la fonction, une famille borelienne  $B$  d'ensembles  $E \subset \Omega$ , et une probabilité  $P(E)$  définie dans  $B$ . Si  $\Omega$  n'est pas dénombrable, il est impossible que  $B$  comprenne tous les  $E \subset \Omega$ : une définition, à ce point de vue, n'est jamais complète. Si on veut qu'elle soit au moins aussi complète que la définition faible, il faut que  $B$  comprenne tous les  $E \subset \Omega$  définis par des conditions de la forme  $X(t) < x$ ,  $t$  et  $x$  étant donnés. Pour être considérée comme complète, il faut évidemment que  $B$  ne se réduise pas au plus petit corps vérifiant cette condition. Mais quels ensembles faut-il ajouter? Il semble qu'une réponse directe ne puisse qu'être arbitraire.

Nous éviterons cette difficulté par une méthode constructive,

qui se présentera sous la forme suivante: nous définirons d'abord un ensemble de variables aléatoires auxiliaires  $U_\nu$ ;  $X(t)$  apparaîtra ensuite comme une fonction certaine, bien définie, de  $t$  et des  $U_\nu$ . S'il suffit d'introduire une infinité dénombrable de variables  $U_\nu$ , qui soient scalaires ou du moins bien définies par leurs fonctions de répartition, il n'y a aucune difficulté à les introduire successivement, la loi de chaque  $X_\nu$  pouvant dépendre des choix antérieurs. Dans le cas contraire, pour ne pas avoir à considérer un ensemble bien ordonné non dénombrable, nous supposerons que les  $U_\nu$  comprennent, d'une part une suite de variables  $U'_n$  qu'on peut sans inconvénient supposer enchaînées [ce ne serait d'ailleurs pas une restriction essentielle de les supposer indépendantes; mais cela devient une restriction si on suppose que ce sont des  $X(t_n)$ ], ensuite d'autres variables  $U''_\nu$ , dont les lois peuvent dépendre des  $U'_n$ , mais qui sont indépendantes les unes des autres.

Dans ces conditions, la fonction  $X$  sera considérée comme *complètement définie*.  $B$  correspondra à la famille des ensembles mesurables dans l'espace produit des  $U_\nu$ , compte tenu de la loi de probabilité définie dans cet espace. C'est toujours dans ce sens que le mot *mesurable* sera employé dans la suite.

Précisons que nous ne considérerons jamais deux fonctions comme distinctes si elles peuvent être déduites l'une de l'autre en ajoutant à  $\Omega$  et en retranchant des ensembles de probabilités nulles.

Terminons ces remarques préliminaires en exprimant le regret que l'axiomatique de Kolmogorov, dont il n'est pas question de discuter qu'elle doive être aujourd'hui la base du calcul des probabilités, ait un peu trop fait oublier les méthodes constructives, sans lesquelles la théorie des suites de variables enchaînées n'est pas concevable.

2° En termes peu précis, une fonction aléatoire est dite *markovienne* si à tout instant  $t$ , la valeur actuelle  $X(t)$  de la fonction  $X$  étant connue, l'avenir est indépendant du passé. Cette notion peut être précisée de plusieurs manières. Quand on parle de la loi de probabilité de l'avenir, s'agit-il de la définition faible ou de la définition complète? D'autres distinctions sont possibles, et les plus importantes sont en relation avec le fait que la *probabilité de transition*

$$p(x' - 0; t, t'|x) = \Pr\{X(t') < x' | X(t) = x\} \quad (t' > t),$$

étant une probabilité conditionnelle, n'est pas bien définie quand la condition  $X(t) = x$  est infiniment peu probable. Il y a donc une

*classe de probabilités de transition équivalentes*, et n'importe quelle fonction de cette classe permet de définir le processus markovien considéré, ou plutôt une *classe de processus markoviens équivalents*. Si de plus on se donne, pour une valeur initiale  $t_0$  de  $t$ , la valeur de  $X(t_0)$ , ou au moins la loi de probabilité de cette variable aléatoire, on obtient une *classe de fonctions markoviennes équivalentes*, c'est-à-dire ayant la même définition faible.

Or, s'il s'agit d'un processus stationnaire, où  $p$  est de la forme  $p(x'; t' - t|x)$ , si l'ensemble des racines de  $X(t) = x$  est presque sûrement non vide et fermé à gauche, et si  $T$  désigne la première de ces racines après un instant donné  $\theta$ , il semble d'abord naturel de penser que  $Pr\{X(T+t) < x'\}$  est indépendant de la valeur  $t$  de  $T$  et égal à  $p(x' - 0; t|x)$ . Mais cela n'est évidemment possible que pour une seule fonction  $p$  choisie convenablement dans la classe des fonctions équivalentes, et qui serait la *vraie probabilité de transition*. La question suivante se pose alors naturellement: *y a-t-il, dans le cas le plus général, une fonction p qu'on puisse considérer comme la vraie probabilité de transition, et à l'aide de laquelle on puisse prévoir l'avenir à partir d'un instant aléatoire tel que T* (le sens de ces derniers mots sera précisé plus loin; v. définition 3)?

Des exemples bien connus (v. l'exemple 2 ci-dessous) montrent que, même dans le cas stationnaire, il n'en est pas toujours ainsi. Dans ces exemples, il existe des *valeurs critiques* de  $x$ , qui peuvent dépendre de  $t$ , qui apparaissent comme l'aboutissement de plusieurs trajectoires possibles, et sont telles qu'au moment du passage au point  $x$  les probabilités relatives à l'avenir dépendent du chemin d'arrivée.

Or, si la probabilité  $p$  est une fonction continue de  $x$ , elle se distingue de toutes les fonctions de la même classe. Il est donc naturel de se demander si des conditions de continuité assez peu restrictives ne suffisent pas à la fois pour être assuré qu'il n'y a pas de valeur critique et que  $p$  est la vraie probabilité de transition.

Le résultat principal du présent travail est la démonstration du fait qu'il en est bien ainsi quand  $X(T)$  est un nombre certain [v. le théorème 2, et les formules (8) et (9) qui en précisent la signification]. Il est probable que ce théorème s'étend à tout instant  $T$  vérifiant les conditions de la définition 3, sans la condition restrictive relative à  $X(T)$ . Nous présenterons à ce sujet quelques remarques qui ne constituent pas une démonstration.

Cette restriction relative à  $X(T)$  fait que, sous sa forme actuelle, notre théorème ne comprend pas un théorème démontré dès 1945

par J. L. Doob<sup>1)</sup>). Mais il est par ailleurs beaucoup plus général, puisque Doob s'était borné au cas à la fois stationnaire et dénombrable. La méthode de Doob reposait sur la division de l'axe des  $t$  en intervalles très petits. Il semble qu'une combinaison de sa méthode et de la nôtre devrait permettre de traiter simplement un cas beaucoup plus général.

Un important travail de K. L. Chung [3] sur les mêmes problèmes est en cours d'impression. Notre méthode est beaucoup plus simple que celle de Chung.

3° Les n°s 2 et 3 contiennent quelques définitions, et le rappel de notions classiques sur la probabilité de transition. Nous avons renoncé à des définitions proposées antérieurement, et où la notion de probabilité de transition était utilisée d'une manière insuffisamment précise. Les définitions des fonctions markoviennes, et strictement markoviennes, seront les définitions maintenant classiques. Celle des fonctions presque strictement markoviennes sera donnée sous une forme nouvelle.

Au n° 4, nous montrerons comment on peut, dans tous les cas, déduire de n'importe quelle probabilité de transition donnée la définition complète d'une fonction markovienne, et, au n° 5, nous montrerons que la fonction ainsi obtenue est presque strictement markovienne. Il semble inutile d'insister sur le fait que, dans tous les cas, il existe une classe très étendue de fonctions markoviennes ayant la même probabilité de transition. Il s'agit le plus souvent de modifications artificielles d'une ou plusieurs fonctions qui se distinguent des autres par des conditions de continuité.

Le n° 6 contient, indépendamment du théorème fondamental mentionné ci-dessus, quelques remarques se rattachant à la recherche des processus strictement markoviens ayant une probabilité de transition donnée. Nous chercherons à préciser la notion de valeur critique, ce à quoi nous n'avons réussi que moyennant certaines conditions, d'ailleurs assez peu restrictives. Nous indiquerons comment on peut éliminer les valeurs critiques par application d'une idée de D. Ray, qui nous a été communiquée par W. Feller: elle repose sur l'introduction de valeurs fictives, c'est-à-dire d'un ensemble de valeurs n'ayant aucune chance d'être réalisé à un instant donné, mais qui peut l'être tout de même au cours des temps. Cela ne change pas la probabilité de transition. Il est très vraisemblable que cette élimination suffit à

---

<sup>1)</sup> J. L. Doob, [2], théorème 2.1. V. aussi T. Nisida [4], qui par une autre méthode a vérifié dans un cas particulier le théorème général de Doob.

assurer, sinon les conditions de continuité énoncées dans le théorème 2, du moins des conditions plus larges qui permettraient d'aboutir aux mêmes conclusions, et même à leur extension au cas où  $X(T)$  est aléatoire. L'idée de Ray conduirait ainsi à établir l'existence d'une fonction strictement markovienne ayant n'importe quelle probabilité de transition donnée. Nous croyons d'ailleurs savoir que Ray a obtenu ce résultat par une autre méthode.

Le présent travail doit beaucoup à K. L. Chung, et à plusieurs autres savants, qui ont attiré notre attention sur l'insuffisance des considérations heuristiques que nous avions d'abord utilisées. Nous les en remercions. Nous ne nous sommes pas interdit ici ces considérations, mais nous les avons nettement distinguées des raisonnements rigoureux.

## 2. Définitions et exemples

**DÉFINITION 1.** La fonction aléatoire  $X(t)$  sera dite *markovienne* (ou *simplement markovienne*) si, pour n'importe quelle suite de nombres croissants  $t_n$ , la suite des  $X(t_n)$  est une chaîne simple de Markov.

**DÉFINITION 2.** La fonction aléatoire  $X(t)$  sera dite *presque strictement markovienne* s'il existe une fonction aléatoire  $Y(\tau) = Y(\tau; t, x)$  de la variable positive  $\tau$  et des paramètres  $t$  et  $x$  telle que, pour n'importe quelle valeur de  $t$ , la fonction

$$X^*(t') = \begin{cases} X(t') & (t' \leq t) \\ Y[t'-t; t, X(t)] & (t' > t) \end{cases}$$

apparaisse comme une nouvelle définition (complète) de la même fonction aléatoire  $X(t')$ .

**DÉFINITION 3.** Nous appellerons *instant initial conventionnel*<sup>2)</sup> un nombre  $T$ , bien défini pour chaque réalisation de la fonction étudiée  $X(t)$ , tel que, pour savoir si  $T \leq t$ , il suffise de connaître cette fonction jusqu'à l'instant  $t$  inclus, et tel que  $T$  et  $X(T)$  soient des variables aléatoires, c'est-à-dire que les événements  $T < t$  et  $X(T) < x$  aient toujours des probabilités bien définies [exemple, dans le cas où l'ensemble  $\{t : X(t) = x\}$  est presque sûrement fermé à gauche, le premier instant  $t \geq \theta(\theta$  donné) où  $X(t) = x$ ].

**DÉFINITION 4** (d'après E. B. Dynkine et A. A. Jouchkevitch). La fonction  $X(t)$  sera dite strictement markovienne si la propriété par laquelle nous avons défini les fonctions presque strictement

<sup>2)</sup> Nous traduisons l'expression *optional starting time* employée par K. L. Chung.

markoviennes subsiste quand on y remplace  $t$  par  $T$ ,  $T$  étant un instant initial conventionnel quelconque.

Pour qu'une fonction presque strictement markovienne soit strictement markovienne, il suffit que cette propriété soit vérifiée dans le cas où la loi de probabilité de  $T$  est continue.

Il est bien évident que les définitions 1, 2 et 4 ne peuvent être que de plus en plus restrictives. Les exemples qui suivent montrent que chacune est effectivement plus restrictive que la précédente.

**EXEMPLE 1.** Désignons par  $Z(t)$  la fonction aléatoire du mouvement brownien, par  $e$  l'ensemble des points de l'axe des  $t$  pour lesquels une au moins des fonctions  $Z(t-p)$  ( $p$  entier  $\geq 0$ ), admet un maximum ou un minimum, et prenons pour  $X(t)$  la fonction caractéristique de  $e$ .

Cette fonction, étant presque sûrement nulle pour tous les points de n'importe quelle suite  $\{t_n\}$  donnée, est équivalente à la fonction identiquement nulle; elle est donc markovienne.

D'autre part  $X(\theta) = 1$  entraîne  $X(\theta+p) = 1$  pour tout  $p$  entier positif. Donc, à tout instant  $t$ , la connaissance du passé définit un ensemble de points  $t' > t$  pour lesquels  $X(t') = 1$ . La seule donnée de  $X(t)$  ne pouvant pas donner les mêmes renseignements, la fonction  $X(t)$  n'est pas presque strictement markovienne.

**EXEMPLE 2** <sup>3)</sup>.  $X(t)$  est ici la fonction du mouvement brownien sur une courbe fermée, rectifiable, ayant un point double  $a$ , et un seul. Précisons que le point mobile ne peut pas profiter de son passage en  $a$  pour passer d'une branche à l'autre, mais que, si on l'observe au moment d'un passage en  $a$ , on ne sait pas quelle est la branche par laquelle il est arrivé et par laquelle il doit aussi repartir.

Si, à un instant  $t$  donné,  $X(t) = x \neq a$ , son mouvement ultérieur à partir de cette position  $x$  est bien défini par la formule du mouvement brownien. L'éventualité  $X(t) = a$ , étant infiniment peu probable, n'intervient pas dans la définition de  $X^*(t')$ , qui constitue toujours une nouvelle définition de la fonction  $X(t')$ . Le processus est donc presque strictement markovien. Mais il suffit de prendre pour  $T$  l'instant du premier passage en  $a$  (après l'instant  $t_0$  ou après n'importe quel instant donné  $t$ ) pour constater qu'il n'est pas strictement markovien.

### 3. La probabilité de transition

On sait qu'on appelle ainsi la fonction de répartition de

---

<sup>3)</sup> Cet exemple est une variante d'un exemple indiqué à l'auteur par W. Feller.

$X(t')(t' > t)$  dans l'hypothèse  $X(t) = x$ . Nous la désignerons par  $p(x'; t' | t | x)$ . Considérée comme fonction de  $x'$ , c'est une fonction de répartition. C'est une fonction  $P$ -mesurable de  $x$  [nous entendons par là,  $x$  étant une valeur possible de  $X(t)$ , qu'il s'agit de la mesure liée à la loi de probabilité de  $X(t)$ ]. Enfin elle vérifie l'équation de Chapman-Kolmogorov

$$(3) \quad p(x'; t, t' | x) = \int_{-\infty}^{\infty} p(x'; u, t' | y) d_u p(y; t, u | x) (t \leq u \leq t').$$

À ces conditions près, elle est quelconque.

Il est bien clair que la donnée de cette fonction définit, non pas un processus markovien, mais une classe de processus markoviens équivalents. Complétée par une condition initiale, elle définit une classe de fonctions markoviennes équivalentes.

Rappelons comment, étant donnée une fonction  $p$  et une condition initiale relative à un instant  $t_0$ , on peut définir les valeurs de  $X(t)$  sur un ensemble  $e$ , dénombrable et partout dense dans  $(t_0, \infty)$ . On range ses points en une suite  $\{t_n\}$ . Pour chaque  $n$ , les points  $t_1, t_2, \dots, t_{n-1}$  divisent  $(t_0, \infty)$  en  $n$  intervalles. Désignons par  $(t'_n, t''_n)$  celui qui contient  $t_n$ ;  $t''_n$  peut être fini ou infini, et la définition de  $X(t_n)$ , quand on a préalablement déterminé les  $X(t_\nu)$  d'indices  $\nu < n$ , n'est pas la même dans les deux cas.

Si  $t''_n$  est infini, la seule donnée effectivement utile est celle de  $X(t'_n)$ , et la fonction de répartition conditionnelle de  $X(t_n)$  est  $p[x'; t'_n, t_n | X(t'_n)]$ .

Dans le cas contraire, il faut tenir compte des deux données  $X(t'_n)$  et  $X(t''_n)$ , ce qui oblige à utiliser une formule d'interpolation. La loi à trois variables  $X(t'_n)$ ,  $X(t_n)$  et  $X(t''_n)$  se déduit aisément de la condition initiale et de la probabilité de transition si on range ces variables dans l'ordre indiqué, on sait qu'il est toujours possible de la reconstituer en définissant d'abord  $X(t'_n)$ , puis  $X(t''_n)$ , et enfin  $X(t_n)$ , à l'aide d'une formule qui se déduit aisément de la probabilité de transition, et qui est la *formule d'interpolation*.

Il est clair qu'en définissant ainsi successivement tous les  $X(t_n)$ , on réalisera ainsi, à n'importe quel moment de cette suite d'opérations, la même loi à  $n+1$  variables  $X(t_0), \dots, X(t_n)$  que si on avait rangé ces variables dans l'ordre des  $t_\nu$  croissants<sup>4)</sup>.

Bien entendu enfin, la probabilité de transition, qui est une probabilité conditionnelle, n'est pas bien définie dans tous les cas,

<sup>4)</sup> Une méthode analogue s'applique à n'importe quel processus. Mais en général la loi de probabilité qui définit  $X(t_n)$  dépend de tous les  $X(t'_\nu)$  antérieurement définis.

et l'on ne peut, sauf dans le cas où  $Pr\{X(t) = x\} > 0$ , parler que d'une *classe de fonctions p équivalentes* qui correspond à une classe de processus markoviens équivalents. D'une manière précise, deux déterminations  $p_1$  et  $p_2$  de la fonction  $p$  sont équivalentes si on a  $p_1 = p_2$ , sauf si  $x \in e_t^*$ ,  $e_t^*$  désignant, pour chaque valeur de  $t$ , un ensemble tel que

$$(4) \quad Pr\{X(t) \in e_t^*\} = 0.$$

Il semble inutile d'insister sur le fait que,  $p$  apparaissant un peu comme une dérivée, il arrive souvent qu'une de ses déterminations possibles se présente naturellement, les autres en étant des modifications artificielles. Tel est le cas où une des déterminations possibles de  $p(x'; t, t'|x)$  est une fonction continue de  $x$ . Précisons aussi que l'équivalence de deux fonctions  $p_1$  et  $p_2$  a été définie en fonction de la formule (3); on ne change pas le second membre de cette formule en remplaçant  $p_1(x'; u, t|y)$  par  $p_2(x'; u, t|y)$ . Dans la définition 2, on peut donc, sans rien changer d'essentiel, définir la loi de  $Y(\tau; t, x)$  par n'importe quelle fonction  $p(y; t, t+\tau|x)$  d'une même classe d'équivalence. Cette remarque ne s'étend pas à la définition 4. Si on y a défini  $T$  de manière que  $X(T)$  ait sûrement une valeur donnée  $x$ , et si les deux fonctions  $p_1$  et  $p_2$  sont différentes pour cette valeur  $x$ , il n'est pas indifférent d'utiliser l'une ou l'autre pour définir la loi de probabilité de  $X^*(T+\tau) = Y(\tau; T, x)$ .

#### 4. La définition complète d'une fonction aléatoire markovienne

1° Dans le cas d'un processus séparable<sup>5)</sup>, la construction qui vient d'être indiquée, complétée par la définition de  $\Omega$ , constitue une définition complète. Bien entendu, cette définition de  $\Omega$  n'est pas arbitraire. Dans le cas de la fonction de Poisson, on peut prendre l'ensemble des fonctions continues à gauche, ou au contraire à droite; on ne peut pas prendre l'ensemble des fonctions continues.

---

<sup>5)</sup> Nous disons qu'un processus est *séparable* si on peut déterminer l'ensemble  $\Omega$  et la suite des  $t_n$  de manière que la fonction  $X(\cdot)$  appartienne presque sûrement à  $\Omega$ , et que, si  $f(\cdot) \in \Omega$ ,  $f(t)$  soit une fonction certaine de  $t$  et des  $X(t_n)$ . Nous disons qu'il est *complètement séparable* si on peut prendre pour  $\{t_n\}$  n'importe quel ensemble dénombrable partout dense. Tel est le cas pour une fonction presque sûrement continue à gauche.

Nous regrettons de ne pas avoir réussi à trouver, pour exprimer cette idée, un autre terme que celui employé par Doob dans un autre sens.

Dans le cas général en prenant toujours pour point de départ la construction indiquée au n° 3, l'ensemble  $e$  étant toujours supposé partout dense dans  $(t_0, \infty)$ , nous allons d'abord, pour chaque  $t \in e^c$  (complément de  $e$ ), donner une définition de  $X(t)$  compatible avec les données. Remarquons à cet effet que, l'ensemble  $e \cup t$  étant dénombrable, la loi jointe de  $X(t)$  et des  $X(t_n)$  est bien définie, et peut être obtenue comme il a été dit au n° 3. On peut alors la reconstituer en définissant d'abord la suite des  $X(t_n)$ , et ensuite  $X(t)$ , à l'aide d'une loi  $L = L(t)$  convenablement définie. Naturellement, cette loi n'est pas bien déterminée dans tous les cas. Mais, si  $\omega$  est le paramètre de Doob dont dépendent tous les  $X(t_n)$ , et si deux des déterminations possibles de la loi  $L$  ont des fonctions de répartition  $f_1(x, \omega)$  et  $f_2(x, \omega)$ , l'égalité  $f_1 = f_2$  est presque sûrement réalisée pour chaque  $x$ , donc aussi pour tous les  $x$  rationnels, donc aussi identiquement en  $x$ , pour presque tous les  $\omega$ .

On peut très simplement obtenir la loi  $L$  de la manière suivante: on choisit dans  $e$  deux suites  $\{t'_n\}$  et  $\{t''_n\}$ , la première croissante, la seconde décroissante, ayant toutes les deux la limite  $t$ . Observons qu'on peut aisément, de bien des manières, nommer de telles suites. Partant par exemple de  $t'_0 = t_0$ ,  $t''_0 = \infty$ , et procédant par récurrence, on peut prendre pour  $t'_n$  et  $t''_n$  ( $n = 1, 2, \dots$ ) les premiers  $t_v$  intérieurs respectivement aux intervalles  $(t'_{n-1}, t)$  et  $(t, t''_{n-1})$ .  $X(t'_n)$  et  $X(t''_n)$ , qui sont des  $X(t_v)$ , étant connus, la formule d'interpolation donne pour  $X(t)$  une loi conditionnelle  $L_n$ . Alors: *la loi  $L$  est presque sûrement la limite de  $L_n$ , pour  $n$  infini.*

Pour le démontrer, désignons par  $U_n$  une variable aléatoire dont le choix équivaille à l'ensemble des choix de  $X(t'_n)$  et  $X(t''_n)$ . Il résulte du caractère markovien de la fonction  $X$ , d'une part que la loi  $L$  est la loi conditionnelle de  $X(t)$  quand on connaît tous les  $U_n$  [leur connaissance rend en effet successivement inutiles celles des différents  $X(t_v)$ ], d'autre part que  $L_n$  est la loi conditionnelle de  $X(t)$  quand on connaît  $U_1, U_2, \dots, U_n$  (la connaissance de chaque  $U_n$  rend en effet inutiles tous ceux qui le précédent).

Appliquons maintenant un lemme connu<sup>6)</sup>: *si un événement  $A$ , de probabilité bien définie  $\alpha$ , est déterminé par une suite de variables aléatoires  $U_v$ , et si  $\alpha_n$  est sa probabilité évaluée quand on connaît les  $n$  premières variables  $U_v$ , il est presque sûr que la suite des  $\alpha_n$  a une limite, égale à 1 dans le cas où  $A$  est finalement réalisé et à 0 dans le cas contraire.*

6) V. P. Lévy, [3], théorème 47.

Pour appliquer ce lemme, divisons  $[0, 1]$  en  $p$  intervalles égaux  $i_h$  ( $h = 1, 2, \dots, p$ ; il est entendu que chacun des points 0 et  $h/p$  est rattaché à un et un seul des intervalles contigus), désignons par  $m = f(x, \omega)$  et  $m_n = f_n(x, \omega)$  les probabilités attribuées à l'événement  $X(t) < x$  par les lois  $L$  et  $L_n$ , et par  $A_h$  l'événement  $m \in i_h$ , qui est bien un événement déterminé par la suite des  $U_n$ . Le lemme pouvant s'appliquer à la fois à tous les  $A_h$ , dont les probabilités ne peuvent pas être toutes  $< 1/p$ , il est presque sûr que, pour tout  $\varepsilon'$  positif, et pour  $n$  assez grand, il y a un  $A_h$  (et un seul, si  $\varepsilon' < \frac{1}{2}$ ) dont la probabilité conditionnelle quand  $U_n$  est connu est  $> 1 - \varepsilon'$ . Prenant pour  $A$  cet événement  $A_h$ , désignons l'événement contraire par  $A^c$ , et par  $E_n$  une valeur probable conditionnelle évaluée quand on connaît  $U_1, U_2, \dots, U_n$ . On a alors

$$\begin{aligned} m_n &= E_n(m) = \alpha_n E_n(m|A) + (1 - \alpha_n) E_n(m|A^c) \\ &= E_n(m|A) + (1 - \alpha_n) [E_n(m|A^c) - E_n(m|A)] \\ &= h - \theta/p + \theta' \varepsilon' \quad (0 \leq \theta \leq 1, |\theta'| \leq 1). \end{aligned}$$

Il n'est d'ailleurs pas évident a priori que  $h$  soit indépendant de  $n$ . Mais, si  $\beta_n$  est la probabilité d'un changement ultérieur, un tel changement donne à  $A_h$  une probabilité à la fois  $> (1 - \varepsilon')\beta_n$  et  $< \varepsilon'$ . Donc  $(1 - \varepsilon')\beta_n < \varepsilon'$ , ce qui prouve que  $\beta_n$  tend vers zéro. Il est donc presque sûr que, à partir d'un certain moment, la formule précédente s'applique au même événement  $A_h$ , qui est finalement réalisé. Donc, presque sûrement, on a pour tout  $n$  assez grand

$$|m - m_n| \leq 1/p + \varepsilon' = \varepsilon,$$

$\varepsilon$  étant arbitrairement petit (puisque on peut prendre  $p$  arbitrairement grand). Donc, pour chaque  $x$ ,  $m = f(x, \omega)$  est la limite presque sûre de  $m_n$ . Cela est donc vrai aussi pour l'ensemble des  $x$  rationnels; donc, toujours presque sûrement, la loi  $L$  est la limite de  $L_n$ , c.q.f.d.

2° Pour obtenir maintenant la définition complète d'une des fonctions  $X(t)$  ayant la probabilité de transition donnée, il n'y a qu'à convenir que, une fois les  $X(t_n)$  définis, les opérations qui viennent d'être définies sont appliquées indépendamment aux différents  $t \in e^c$ . En d'autres termes, les lois conditionnelles  $L(t)$  sont absolument indépendantes les unes des autres.

Chacune des lois  $L(t)$  étant presque sûrement bien définie, il en est de même pour  $L(T)$ , si  $T$  est une variable aléatoire indépendante des  $X(t_n)$ . Il résulte alors du théorème de Fubini que, presque sûrement, presque tous les  $L(t)$  sont bien définis. Nous entendons par là qu'ils le sont sauf sur un ensemble exceptionnel  $E$ ,

tel que

$$(5) \quad \Pr\{T \in E\} = 0.$$

Il en résulte en particulier que: *l'ensemble  $E$  des points où la loi  $L(t)$  n'est pas bien définie est presque sûrement de mesure nulle.*

Il n'est d'ailleurs pas difficile de se débarrasser de la restriction relative à cet ensemble. Pour  $n$  infini, la plus petite et la plus grande valeur d'accumulation de la suite des  $m_n$  sont deux nombres bien définis, qui ne peuvent que croître avec  $x$ . On n'a qu'à prendre l'un d'eux, ou leur moyenne, pour définir  $m$ . Le seul élément arbitraire qui reste ainsi dans notre définition est le choix de la suite des  $t_n$  (dont l'ordre influe évidemment sur les valeurs des  $t'_n$  et des  $t''_n$ ).

## 5. Propriétés de la fonction définie au n° 4

### 1° THÉORÈME 1. *La fonction définie au n° 4*

- a. *est presque partout indépendante du choix des  $t_n$ ;*
- b. *est presque strictement markovienne;*
- c. *admet la probabilité de transition  $p(x'; t, t'|x)$  initialement donnée.*

COROLLAIRE 1. *Il existe au moins une fonction presque strictement markovienne ayant une probabilité de transition donnée.*

Ce théorème est une conséquence simple des constructions des n°s 3 et 4. Nous avons d'abord déterminé les  $X(t_n)$  de manière que, pour toute suite de nombres croissants  $\theta_\nu \in e$ , la suite des  $X(\theta_\nu)$  soit une chaîne simple de Markov ayant la probabilité de transition donnée. Nous avons ensuite défini chaque  $X(t)$  de manière qu'il puisse être intercalé entre  $X(t'_n)$  et  $X(t''_n)$ , donc aussi, si  $\theta_{\nu-1} < t < \theta_\nu$ , entre  $X(\theta_{\nu-1})$  et  $X(\theta_\nu)$  (puisque, pour  $n$  assez grand,  $\theta_{\nu-1} < t'_n < t''_n < \theta_\nu$ ), sans que ces conditions cessent d'être vérifiées. On peut intercaler de même successivement les termes de n'importe quelle suite de nombres, de sorte que, finalement, ces conditions sont vérifiées pour n'importe quelle suite finie de nombres  $\theta_\nu$ , appartenant ou non à  $e$ . La fonction obtenue a donc bien la définition faible voulue: elle est markovienne, et a la probabilité de transition donnée, ce qui démontre c.

2° Pour démontrer a, donnons-nous deux déterminations  $e_1$  et  $e_2$  de l'ensemble  $e = \cup t_n$ , en partant desquelles nous définirons deux fonctions  $X_1(t)$  et  $X_2(t)$ . D'après ce que nous venons de voir, pour chaque suite finie  $\{\theta_\nu\} \subset (e_1 \cup e_2)$  ( $\theta_{\nu-1} < \theta_\nu$ ), la suite des  $X_i(\theta_\nu)$  ( $i = 1, 2$ ) est une chaîne de Markov ayant la probabilité de

transition donnée. La définition est donc indépendante de  $i$  (si on prend dans les deux cas la même condition initiale).

Donnons-nous maintenant un nombre  $t$  n'appartenant ni à  $e_1$ , ni à  $e_2$ . Pour définir  $X_1(t)$  et  $X_2(t)$ , il faut, comme au n° 6, introduire des suites de nombres  $t'_{n,i}$  et  $t''_{n,i}$  ( $i = 1, 2$ ) tendant vers  $t$ .

Nous pouvons ensuite définir une suite d'intervalles  $(t'_v, t''_v)$ , qui soient tantôt des  $(t'_{n,1}, t''_{n,1})$ , et tantôt des  $(t'_{n,2}, t''_{n,2})$ , et dont chacun est intérieur au précédent. Les résultats du n° 4 s'appliquant aussi bien aux  $(t'_v, t''_v)$  qu'aux  $(t'_{n,1}, t''_{n,1})$  et aux  $(t'_{n,2}, t''_{n,2})$ , nous obtenons pour  $X(t)$  [ou  $X_i(t)$ ] trois lois conditionnelles  $L$ ,  $L_1$ ,  $L_2$ , toutes les trois presque sûrement bien définies.  $L$  coïncidant alors à la fois avec  $L_1$  et  $L_2$ ,  $L_1$  et  $L_2$  coïncident.

Faisant maintenant varier  $t$ , le même raisonnement qu'au n° 4, 2° nous montre que  $L_1(t)$  et  $L_2(t)$  sont bien définis et coïncident pour tous les  $t$ , à l'exception d'un ensemble exceptionnel, donc de mesure presque sûrement nulle, qui peut évidemment dépendre des ensembles  $e_1$  et  $e_2$  et même de l'ordre dans lequel on range leurs éléments. Cela n'empêche pas qu'en choisissant d'abord une suite particulière  $X(t_n)$ , on définit une fonction aléatoire qui coïncide presque partout avec chacune des fonctions analogues.

3° Il reste à démontrer la propriété b. Pour cela, il suffit de considérer chaque valeur de  $t$  indépendamment des autres. Nous avons vu qu'il est presque sûr qu'on ne change pas la définition de  $X(t)$  en ajoutant  $t$  à l'ensemble  $e$ . Nous pouvons supposer plus précisément que  $t = t_1$ , c'est-à-dire qu'après s'être donné une condition initiale, la première opération sera la définition de  $X(t)$  pour la valeur considérée. Après cela, les opérations aboutissant à la fonction  $X(t')$  dans  $(t, \infty)$  sont absolument indépendantes de celles relatives à l'intervalle  $(0, t)$ . Seuls  $t$  et  $X(t)$  interviennent, et  $X(t+\tau)$  est bien de la forme  $Y[\tau; t, X(t)]$ , c.q.f.d.

4° REMARQUE. Dans l'énoncé du théorème précédent, nous avons évité de parler de probabilité conditionnelle. Il est entendu qu'on ne peut introduire cette notion qu'en précisant qu'il s'agit d'une des probabilités conditionnelles équivalentes les unes des autres à l'aide desquelles on peut définir le processus. Dans ces conditions, on déduit du théorème précédent que:

**COROLLAIRE 2.** *Pour la fonction  $X(t)$  définie au n° 4, à tout instant  $t$ , on peut déduire la probabilité de n'importe quel événement dépendant de l'avenir d'une probabilité conditionnelle  $\varphi(x)$  relative à l'éventualité  $X(t) = x$  et ayant les propriétés suivantes: b'. Sauf peut-être pour un ensemble exceptionnel  $e_t^*$  [c'est-à-dire vérifiant la*

condition (4)], cette probabilité est indépendante du passé; **d**. Même si  $X(t) \in e_t^*$ , elle ne peut dépendre que du passé immédiat.

La propriété **b'** n'est qu'un autre énoncé de **b**: il est presque sûr qu'on ne change rien en introduisant  $t$  avec le rang 1 dans la suite des  $t_n$ , donc en supposant  $t = t_1$ . On peut donc construire la fonction  $X$  en définissant d'abord  $X(t)$ , et ensuite le passé et l'avenir par des constructions indépendantes les unes des autres.

Cela n'est que presque sûr. Il peut donc y avoir des cas d'exception. Mais il résulte du théorème de Fubini (appliqué en choisissant indépendamment, d'une part une détermination possible de la fonction  $X$ , d'autre part une valeur de  $t$ ) que, presque sûrement, les cas d'exception ne sont réalisés que sur un ensemble de mesure nulle. Quelque petit que soit  $\varepsilon$  positif, on peut donc choisir dans  $(t - \varepsilon, t)$  un  $u$  pour lequel le cas d'exception ne soit pas réalisé (par exemple un des  $t_n$ ). Alors pour l'étude de  $X(t')$  dans  $(u, \infty)$ , et a fortiori pour son étude dans  $(t, \infty)$ , la donnée de  $X(u)$  rend inutile tout renseignement sur un passé plus ancien, ce qui démontre **d**.

## 6. La recherche des processus strictement markoviens ayant une probabilité de transition donnée

Pour cette recherche, nous pouvons maintenant partir de la fonction presque strictement markovienne définie au n° 4. L'exemple 2 montre bien pourquoi une telle fonction peut n'être pas strictement markovienne: il y a un *état critique* (ou *valeur critique* de  $X$ ), où le caractère markovien disparaît. Pour cet état, il y a deux voies d'accès et deux voies de sortie, et la voie de sortie est déterminée par le passé immédiat. Ce passé donne ainsi un renseignement que le présent ne peut pas donner. Il peut arriver aussi qu'il donne des renseignements qui ne fassent que modifier les probabilités relatives à l'avenir. C'est ce que va montrer l'exemple suivant:

**EXEMPLE 3.** Considérons, comme dans l'exemple 2, une courbe  $C$  fermée, ayant un point double  $a$ , et supposons-la orientée. Désignons par  $U(s)$  le point atteint en décrivant sur  $C$ , à partir d'une origine  $0$  et dans le sens positif, un chemin de longueur  $s$ , le point mobile pouvant passer d'une branche à l'autre quand il passe en  $a$ ; à chaque passage, la probabilité d'un tel changement a une valeur  $\alpha$ , qui n'est ni  $0$ , ni  $1$ , ni  $\frac{1}{2}$ . Remplaçons maintenant  $s$  par  $S(t) = S_0(t) + kt$ ,  $k$  étant une constante positive, et  $S_0(t)$  une fonction constamment croissante, dépendant d'un processus additif homogène et à loi de probabilité continue, [par exemple  $\sum_1^\infty N_h(t)/h^2$ ,

les  $N_h(t)$  étant des déterminations indépendantes de la fonction de Poisson]. Posons enfin  $X(t) = U[S(t)]$ .

Compte tenu de l'hypothèse  $k > 0$ , on voit aisément que l'événement  $X(t) = a$ , tout en étant infiniment peu probable pour chaque  $t$  donné, se produit presque sûrement une infinité de fois, sans que jamais ce soit l'instant d'un saut, de sorte qu'il n'y aura jamais d'ambiguité sur la voie d'accès. Le processus est alors presque strictement markovien, mais non strictement markovien, comme on le voit en prenant pour  $T$  l'instant d'un passage en  $a$  (par exemple celui de rang  $n$ , ou le premier après un  $t$  donné).

Il n'y a dans cet exemple qu'une dépendance stochastique de la voie de sortie (en partant du point double) et du passé immédiat. On ne peut donc faire apparaître le caractère critique de ce point qu'en introduisant des probabilités conditionnelles relatives aux instants où le point mobile y passe.

**EXEMPLE 4.** Dans les exemples 2 et 3, il n'y avait qu'une valeur critique. Il est facile d'en déduire des exemples comportant plusieurs valeurs critiques. Mais il n'est peut-être pas inutile de montrer qu'il peut arriver que presque toutes les valeurs de  $x$  soient critiques.

On sait que W. Feller et McKean ont défini une fonction aléatoire  $X_1(t)$  ayant les caractères suivants: elle est presque sûrement continue, constamment variable sans être jamais monotone, mais se distingue de la fonction  $X_2(t)$  du mouvement brownien parce que, pour chaque  $t$  donné, il y a une probabilité unité que  $X_1(t)$  appartienne à un ensemble dénombrable partout dense, que nous supposerons être l'ensemble des nombres rationnels. Chaque détermination de  $X_1(t)$  peut se déduire d'une détermination de  $X_2(t)$  par un changement sur l'échelle des  $t$ . La succession des maxima et minima est la même pour les deux fonctions.

Partons alors de la fonction  $X_2(t)$ , et supposons que, dans chacun des intervalles où elle varie entre deux nombres entiers, nous fassions un tirage au sort à pile ou face pour savoir si nous conserverons l'échelle des  $t$ , ou si nous la changerons de manière à en faire une fonction  $X_1(t)$ . Prenons pour  $X(t)$  la fonction ainsi obtenue.

Toutes les valeurs de  $x$  sont ainsi presque sûrement réalisées, tantôt dans des intervalles où  $X(t)$  est de la forme  $X_1(t+\text{const.})$ , tantôt dans des intervalles où  $X(t)$  est de la forme  $X_2(t+\text{const.})$ . Cela n'empêche pas la fonction  $X(t)$  d'être presque strictement markovienne. En effet il est presque sûr, à tout instant  $t$ , que, ou bien  $X(t)$  a une valeur rationnelle non entière et varie comme

$X_1(t)$  au voisinage de l'instant  $t$ , ou bien a une valeur irrationnelle et varie comme  $X_2(t)$  au voisinage de cet instant. On reconstitue donc la fonction  $X(t+\tau) = Y[\tau; t, X(t)]$  en la définissant, suivant que  $X(t)$  est rationnel ou irrationnel, comme s'il s'agissait de  $X_1(t+\tau)$  ou de  $X_2(t+\tau)$ , et cela jusqu'à ce qu'elle prenne une valeur entière. Le cas où  $X(t)$  est entier est infiniment peu probable. Donc  $X(t)$  est bien une fonction presque strictement markovienne. Les cas négligés se produisant tout de même pour des valeurs exceptionnelles de  $t$ , on voit que toutes les valeurs non entières de  $x = X(t)$  sont des valeurs critiques. Mais, pour ces valeurs, un des deux chemins qui y passent est infiniment peu probable. Pour une fonction presque strictement markovienne, il n'en saurait être autrement que pour des valeurs exceptionnelles de  $x$ .

2° Si on cherche à définir d'une manière générale ce qu'il convient d'appeler *valeur critique* d'une fonction presque strictement markovienne pour l'instant  $t$ , on est conduit d'abord à dire: c'est une valeur pour laquelle il y a plusieurs voies d'accès possibles (au moins deux) et plusieurs voies de sortie possibles (au moins deux), le choix de la voie de sortie dépendant, au moins stochastiquement, de la voie d'entrée (7). Mais une telle définition repose sur la notion de probabilité conditionnelle, et n'a aucun sens pour un point du plan des  $tx$  tel que  $Pr\{X(t) = x\} = 0$ . On ne peut que considérer un ensemble de points qui ait une probabilité positive d'être coupé par la courbe  $x = X(t)$ , choisir l'instant initial conventionnel  $T$  de manière que le point  $T, X(T)$  appartienne à cette intersection, et se demander si le mouvement après l'instant  $t$  peut être retrouvé à l'aide d'une probabilité de transition indépendante du passé, et aussi si cette probabilité est bien  $p(x'; t, t'|x)$ .

L'idée intuitive de point critique, qui nous a guidé d'abord dans nos recherches, est donc insuffisante pour une théorie générale. Mais il y a des cas où on peut la sauver, et c'est de ces cas que nous allons nous occuper.

Le premier, que nous ne ferons que mentionner, est celui des processus additifs, pour lesquels la loi de probabilité de  $X(t') - X(t)$  est donnée inconditionnellement. Il n'est dans ce cas pas

7) Précisons qu'en parlant de la voie d'accès, nous ne parlons pas seulement de l'ensemble ordonné (continu ou discontinu) des valeurs de  $X(u)$  prises avant l'instant  $t$ , mais de la rapidité de leur succession. En d'autres termes, il s'agit de la courbe décrite dans le plan des  $tx$  par le point  $u, X(u)$ , et la partie de cette courbe qui précède immédiatement le point d'arrivée importe seule.

question de point critique; un tel point n'apparaît dans les exemples 2 et 3 que parce que l'enroulement sur une courbe fermée ayant un point double réunit en un même point de la courbe des points distincts de la droite. Cette remarque s'étend au cas plus général où, à l'instant  $t$ ,  $X(t')$  est, pour chaque  $t' > t$ , une fonction certaine de  $t$ ,  $t'$ ,  $X(t)$ , et d'une variable aléatoire définie inconditionnellement.

3° Le second cas que nous voulons signaler est celui où l'ensemble  $E_x$  des racines de  $X(t) = x$  a une probabilité positive  $\alpha$  de n'être pas vide, et où il est possible de définir un instant initial conventionnel  $T$  de manière que  $X(T) = x$  (par exemple la première réalisation de cette équation après un instant donné)<sup>8)</sup>. Nous désignerons par  $F(t)$  la fonction de répartition de  $T$ . Donc  $0 < F(\infty) = \alpha \leq 1$ .

Si  $t' > t$ , on a

$$(6) \quad \Pr\{T < t, X(t') < x'\} = \int_{t_0}^{t-0} P(x' - 0; u, t'|x) dF(u),$$

$P$  étant une probabilité conditionnelle définie par cette formule même. On peut convenir que, pour  $u = t$ , ce soit la dérivée du premier membre par rapport à  $F(t)$ . C'est une fonction sommable de  $F(t)$ , et, sur l'axe des  $t$ , elle peut n'être pas bien définie en certains points ou même dans certains intervalles. Mais, ce qui est important, c'est que la valeur  $x$  peut être considérée individuellement.

On peut alors avoir une définition précise des valeurs critiques: *La valeur  $x$  est dite critique si la probabilité (6) n'est pas indépendante du passé*, c'est-à-dire si la connaissance d'un événement dépendant du passé qui ait une probabilité positive d'être réalisé, peut la modifier. Ainsi, dans l'exemple 4, que  $x$  soit rationnel ou irrationnel, si  $T$  est la première racine de  $X(t) = x$ , les deux cas distingués dans la définition de cet exemple sont également probables. Tandis que, suivant que  $x$  est rationnel ou irrationnel,  $p(x'; t, t'|x)$  est égal à l'une ou l'autre de deux fonctions continues de  $x$ ,  $P(x'; t, t'|x)$  est leur moyenne. La connaissance de la voie d'accès à la valeur  $x$  rend infiniment probable l'un ou l'autre des deux cas possibles, et, si  $T$  est la première racine de  $X(t) = x$  au moins égale à un nombre donné, il n'y a qu'une probabilité  $\frac{1}{2}$  que

---

<sup>8)</sup> Ces conditions sont naturellement beaucoup plus restrictives dans le cas d'un espace à plusieurs dimensions que dans le cas linéaire. Toutefois, même dans ce dernier cas, il existe des processus discontinus pour lesquels elles ne sont réalisées pour aucun  $x$  (voir le 9° ci-dessous).

ce soit celui qui, pour une valeur donnée de  $t$ , était presque sûr<sup>9)</sup>. Naturellement la probabilité (5) n'est pas la même dans les deux cas.

4° Considérons maintenant le cas d'une fonction markovienne stationnaire, vérifiant de plus la condition considérée au 3° ci-dessus. Elle est déduite d'une probabilité de transition de la forme  $p(x'; \tau|x)$  ( $\tau = t' - r$ ), et on pourrait penser que  $P$  a la même forme. Cela n'est pas vrai s'il y a des points critiques. En effet les différents itinéraires pouvant conduire de la valeur initiale à la valeur  $x$  peuvent être plus ou moins longs à parcourir, de sorte que la donnée de  $T$  peut modifier les probabilités des différentes voies d'accès possibles.

Si au contraire il n'y a pas de points critiques, la fonction  $P$  est de la forme  $P(x'; \tau|x)$  et indépendante du passé. Ce dernier point résulte de l'hypothèse qu'il n'y ait pas de point critique, du moins s'il s'agit de la fonction définie au 4° pour laquelle le passé immédiat peut seul avoir de l'influence. Alors la valeur  $t$  de  $T$  est aussi sans influence: puisque le processus est stationnaire, l'heure qu'il est ne joue aucun rôle, et l'influence indirecte provenant de renseignements a posteriori sur le passé n'intervient pas non plus, puisqu'il n'y a pas de point critique.

Cette forme de la fonction  $P$  était presque évidente. Ce qui est peut-être moins évident, mais a priori très vraisemblable, c'est que  $P = p$ . Nous le démontrerons, en faisant seulement sur la continuité de  $p$  des hypothèses très peu restrictives. Mais nous allons d'abord montrer avec D. Ray comment on peut obtenir une fonction presque strictement markovienne, sans valeur critique, et ayant la fonction  $p$  donnée comme probabilité de transition.

5° A cet effet, considérons une variable auxiliaire  $\rho$  qui contient tout ce qu'il faut connaître du passé à l'instant où  $X(t) = x$  pour définir la loi de probabilité de l'avenir. Les voies d'accès possibles seront ainsi réparties en classes d'équivalence, correspondant chacune à une valeur de  $\rho$ ; deux voies qui coïncident dans un petit intervalle  $(t - \eta, t)$  appartiennent naturellement à la même classe. L'idée de Ray consiste à substituer à  $X(t)$  la fonction  $X^*(t)$  égale au produit cartésien de  $X(t)$  et  $\rho$ . Si la valeur  $x$  de  $X(t)$  n'est pas critique,  $\rho$  n'a qu'une valeur possible et  $X^*(t)$  ne se distingue pas de  $X(t)$ . Mais s'il y a des valeurs critiques, il n'en est pas de même, et la définition de  $X^*(t)$  implique une extension de

<sup>9)</sup> Il faut remarquer que le point  $T$  de  $E_x$  que nous considérons ici est isolé à gauche, donc exceptionnel. Le résultat, un peu paradoxal, obtenu ne subsiste pas si on prend pour  $T$  l'instant où la mesure de  $(t_0, t) \cap E_x$  atteint une valeur donnée; alors  $P = p$ .

l'espace des valeurs possibles. L'axe des  $x$  apparaît comme une section du plan des  $x\rho$ , et  $X(t)$  comme la projection de  $X^*(t)$ .

Cette fonction  $X^*(t)$  remplit les conditions voulues. En effet l'ensemble  $e$  des valeurs critiques est un ensemble de valeurs fictives, c'est-à-dire qu'à tout instant  $t$  la probabilité que  $X(t) \in e$  est nulle. La probabilité  $p(x'; t|x)$  n'étant changée que sur un tel ensemble, cela ne change pas la classe d'équivalence à laquelle appartient cette fonction. Donc  $X^*(t)$  a bien la probabilité de transition donnée.

D'autre part la donnée de  $\hat{X}^*(t)$ , impliquant à n'importe quel instant celle de  $X(t)$  et de  $\rho$ , contient tout ce qu'il faut pour prévoir l'avenir. Les points critiques ont disparu.

Dans ce qui suit, nous supposerons cette élimination des points critiques préalablement effectuée. Donc  $X^*(t) = X(t)$ .

6° Revenant au cas général, et demandons-nous s'il n'est pas possible de définir, pour chaque processus markovien, une fonction  $p(x'; t, t'|x)$  qui apparaisse comme la *vraie probabilité de transition*. L'idée d'exiger qu'elle soit continue en  $x$ , du moins sur l'ensemble des valeurs de  $X(t)$  qui sont possibles à l'instant  $t$ , est bien naturelle. En effet cette condition de continuité, si elle peut être vérifiée, la distingue de toutes les autres fonctions équivalentes. Mais on remarque qu'elle ne saurait être remplie en un point critique  $x$ , puisque  $p[x'; u, t'|X(u)]$  devrait avoir, quand  $u \uparrow t$ ,  $X(t) \rightarrow x$ , une limite qui dépend de la voie d'accès. De simples conditions de continuité suffisent donc à écarter les circonstances les plus gênantes. Nous allons préciser cette remarque en montrant que, sans parler de points critiques, des conditions de continuité, constituant au moins des conditions suffisantes pour qu'il n'y ait pas de tels points, suffisent aussi pour affirmer que la fonction  $P$  définie par la formule (6) est identique à  $p$ .

Nous désignerons par  $C$  l'ensemble des deux conditions suivantes:  $p(x'; t, t'|x)$  est une fonction de  $t$  continue à gauche, et, pour chaque valeur de  $t$ , il est presque sûr que la loi  $L[t, t'|X(t)]$  dont la fonction de répartition est  $p[x'; t, t'|X(t)]$  est une fonction de  $t$  continue à gauche.

**THÉORÈME 2.** *Si la condition  $C$  est vérifiée, et si  $T$  est un instant initial conventionnel tel que  $X(T)$  ait une valeur certaine  $x$ , on a*

$$P(x'; t, t'|x) = p(x'; t, t'|x) \quad (t' > t).$$

Bien entendu, cette formule n'a de sens que si la fonction  $F(u)$  varie au voisinage du point  $t$ . Nous supposerons d'une manière plus précise qu'elle varie à gauche de ce point, c'est-à-dire que,

pour tout  $\eta$  positif, on ait

$$Pr(H_\eta) = Pr\{t-\eta < T \leq t\} > 0,$$

et  $P$  est par définition la limite de  $Pr\{X(t') < x'|H_n\}$  quand  $\eta \downarrow 0$ . Il s'agit de montrer que cette limite existe et est égale à  $p$ . Or, si  $G_\eta(\xi)$  est la fonction de répartition conditionnelle de  $X(t)$  dans l'hypothèse  $H_\eta$ , on a

$$Pr\{X(t') < x'|H_\eta\} = \int_{-\infty}^{+\infty} p(x'-0; t, t'|\xi) dG_\eta(\xi),$$

de sorte qu'il s'agit de montrer que,  $\eta$  tendant vers zéro, on a

$$(7) \quad \int_{-\infty}^{+\infty} [p(x'; t, t'|x) - p(x'; t, t'|\xi)] dG_\eta(\xi) \rightarrow 0.$$

Or, dans l'hypothèse  $H_\eta$ ,  $X(\theta) = x$  a une racine  $T$  dans  $(t-\eta, t)$ . Si donc  $\eta$  est assez petit, il résulte de la condition  $C$  que,  $\varepsilon$  et  $\varepsilon'$  étant arbitrairement petits, et si  $x'$  n'est pas pour la loi  $L(t, t'|x)$  une valeur à probabilité positive, on a

$$\begin{aligned} |p(x'; T, t'|x) - p(x'; t, t'|x)| &< \varepsilon, \\ Pr\{|p(x'; t, t'|X(t)) - p(x; T, t'|x)| < \varepsilon\} &> 1 - \varepsilon'. \end{aligned}$$

Donc, dans l'intégrale (7), la fonction intégrée, où  $\xi$  représente la valeur de  $X(t)$ , est  $< 2\varepsilon$ , sauf sur un ensemble sur lequel la variation de  $G_\eta(\xi)$  est  $< \varepsilon'$ , et où elle est évidemment  $\leq 1$ . Donc cette intégrale est inférieure au nombre arbitrairement petit  $2\varepsilon + \varepsilon'$ , c.q.f.d.

Il résulte de la formule (6) et du théorème 2 que:

$$(8) \quad F(t') Pr\{X(t') < x'\} = \int_{t_0}^{t'} p(x'-0; t, t'|x) dF(t) \quad (t' > t),$$

$$(9) \quad F(\infty) Pr\{X(T+\tau) < x'\} = \int_{t_0}^{\infty} p(x'-0; t, t+\tau|x) dF(t) \quad (\tau > 0).$$

7° *Remarques et applications.* La condition  $C$  a pour effet d'écartier les points critiques, qui peuvent évidemment exister même dans le cas d'un mouvement rectiligne; on peut en effet établir une correspondance biunivoque et continue entre la courbe considérée dans l'exemple 2 et un intervalle ouvert sur l'axe des  $x$ .

Indépendamment de cela elle apparaît comme très peu restrictive. Il suffit évidemment pour qu'elle soit vérifiée que la fonction  $p(x'; t, t'|x)$  soit une fonction continue du point  $tx$ , et que la fonction  $X(t)$  soit, pour chaque valeur donnée de  $t$ , presque sûrement continue à gauche. C'est le cas notamment des processus additifs sans discontinuités fixes, et, dans le cas dénombrable, de tous les processus sans états instantanés.

L'extension du théorème 2 à certains cas où il y a des discontinuités fixes n'est d'ailleurs pas difficile, et il nous paraît très

vraisemblable que ce théorème peut s'étendre à tous les processus sans valeurs critiques et pour lesquels  $p(x'; t, t'|x)$  est une fonction mesurable de  $t$ . Rapproché de la méthode de Ray pour l'élimination des valeurs critiques, cela permettrait de conclure qu'à toute probabilité de transition qui est une fonction mesurable de  $t$  correspond au moins un processus auquel on peut appliquer le théorème 2.

8° Pour savoir si on peut trouver aussi un processus strictement markovien ayant la probabilité de transition donnée, il faut examiner le cas où  $X(T)$  n'est pas un nombre certain. Remarquons d'abord que le fait que la même probabilité de transition peut convenir à la fois pour une intégration le long d'une parallèle à l'axe des  $t$  et pour une intégration le long d'une parallèle à l'axe des  $x$  suggère l'idée qu'elle doit aussi convenir dans des cas beaucoup plus généraux.

Si on cherche à étendre à ces cas la démonstration du théorème 2, la seule difficulté provient de ce que ce qu'on a besoin de savoir si  $Pr\{X(T) < x|T = t\}$  peut être défini comme limite de  $Pr\{X(T) < x|H_\eta\}$ . La grande variété des définitions possibles de  $T$  rend une démonstration générale difficile. Il est d'ailleurs évident que, si on prend pour  $T$  la première racine de  $X(t) = f(t)$ , on a  $X(T) = f(T)$ , et, si la fonction donnée  $f(t)$  est discontinue,  $X(T)$ , dans l'hypothèse  $T = t$ , se réduit à une fonction discontinue de  $t$ . Mais cela ne fait que compliquer la démonstration, et tous les cas que nous avons examinés nous confirment dans l'idée que: *si la condition C est vérifiée, la fonction p est bien une vraie probabilité de transition, qu'on peut toujours appliquer à l'étude de la fonction X(t) après l'instant τ*. Nous avons même l'impression qu'il devrait y avoir une démonstration relativement simple.

9° Quoi qu'il en soit, un problème intéressant est d'étudier la loi de probabilité dont dépend  $X(T)$  dans l'hypothèse  $T = t$ ; c'est en effet de cette valeur qu'il faut repartir. Nous allons traiter ce problème dans un cas simple où  $X(t)$  est une fonction additive, à accroissements stationnaires, et somme ou somme compensée de sauts mobiles positifs, c'est-à-dire que le logarithme de sa fonction caractéristique est d'une des formes

$$(10) \quad t \int_0^\infty (e^{izu} - 1) d\mathbf{n}(u),$$

$$(11) \quad t \int_0^\infty (e^{izu} - 1 - izu) d\mathbf{n}(u),$$

où  $\mathbf{n}(u)$  est une fonction non décroissante, vérifiant dans chacun

des deux cas la condition bien connue qui est nécessaire pour la convergence de l'intégrale. L'hypothèse de l'homogénéité dans le temps n'est pas essentielle, mais et écarte les discontinuités fixes et simplifie les notations.

Dans les deux cas, nous désignerons par  $X_1(t)$  et  $X_2(t)$  les deux termes indépendants l'un de l'autre, dont la somme est  $X(t)$ , que l'on obtient en intégrant par rapport à  $u$ , d'une part de 0 à  $h+0$ , d'autre part de  $h+0$  à  $\infty$ .

Faisons d'abord une remarque très simple, qui semble n'avoir pas encore été faite, sur le cas où  $T$  est la première racine de  $X(t) = x$ . S'il s'agit de la fonction définie par l'intégrale (10), l'ensemble des valeurs prises par  $X(t)$  est de mesure nulle, et, pour presque toutes les valeurs de  $x$ , l'ensemble des racines de  $X(t) = x$  est presque sûrement vide. Dans certains cas, comme celui de Poisson, il existe des valeurs de  $x$  pour lesquelles cet ensemble n'est pas vide. Mais si la loi dont dépend  $X(t)$  est continue, ces valeurs exceptionnelles n'existent pas, de sorte qu'on n'a aucune possibilité d'appliquer le théorème 2.

Il n'en est pas de même dans le cas de l'intégrale (11). Dans ce cas, il est presque sûr que  $X(t) - x$  change de signe une infinité de fois. Si  $X(0) = 0, x > 0$ , la valeur  $x$  sera en général d'abord dépassée par un saut, sans que  $X(t) - x$  s'annule. Mais de toute façon  $X(t) - x$  doit ensuite devenir négatif, et, comme il n'y a pas de sauts négatifs, cela implique que  $X(t)$  prenne exactement la valeur  $x$ ,  $T$  existe presque sûrement et on peut appliquer le théorème 2. Cette conclusion s'applique d'ailleurs aussi bien si

$$\int_0^\infty u d\mu(u) < \infty,$$

cas où  $X(t)$  est à variation bornée, et de la forme  $X^*(t) - \lambda t$ , le premier terme étant la somme des sauts, que dans le cas où le terme  $-izu$  est nécessaire pour la convergence de l'intégrale (11).

10° Prenons maintenant pour  $T$  l'instant du premier saut au moins égal à un nombre donné  $h > 0$ . Nous pouvons supposer  $\mu(\infty) = 0$  et la fonction  $X(t)$  continue à droite. Alors, pourvu que  $\mu(h)$  soit négatif,  $T$  existe presque sûrement, on a

$$Pr\{T < t\} = 1 - e^{-\lambda t} \quad [\lambda = -\mu(h)],$$

et,  $U$  désignant la hauteur (nécessairement  $\geq h$ ) du saut réalisé à l'instant  $T$ ,  $X(T)$  est de la forme  $X_1(T) + U$  ou  $X_1(T) + U - k_h T$ , avec

$$k_h = \int_{h-0}^\infty u d\mu(u) < \infty,$$

suivant que le processus étudié correspond à l'une des formes (10) et (11) du logarithme de la fonction caractéristique. Les lois de  $U$  et de  $X_1(t)$  se déduisant immédiatement de cette fonction, celle de  $X(T)$  s'obtient par une combinaison de formules classiques. Une extension simple du théorème 2 montre ensuite qu'en partant de  $X(T)$ , on peut étudier  $X(T+\tau)$  à l'aide de la probabilité de transition. Dans le présent exemple,  $X(T+\tau)-X(T)$  dépend donc de la même loi que  $X(\tau)$ . Mais le raisonnement indiqué peut s'appliquer à des cas plus étendus.

L'étude du cas où  $T$  est le premier instant où  $X(t)$  est  $\geq x > 0$  est moins simple. Signalons seulement que, dans le cas des lois stables, c'est-à-dire dans le cas où  $d\pi(u) = du/u^{\alpha+1}$  ( $\alpha > 0, < 2, \neq 1$ ), il est évident, par raison d'homogénéité, que  $X(T+0)$  est de la forme  $U_1 x$ , la loi de  $U_1$  ne dépendant pas de  $x$ ;  $X(T-0)$  est de la même forme, mais avec un autre coefficient  $U_0$ . Bien entendu,  $U_0$  et la hauteur  $(U_1 - U_0)x$  du saut réalisé à l'instant  $T$  ne sont pas indépendants, puisqu'il faut que  $U_1 \geq 1$ .

#### BIBLIOGRAPHIE

##### K. L. CHUNG

- [1] On a basic property of Markov chains. En cours d'impression.

##### J. L. DOOB

- [2] Markoff chains. Denumerable case. Trans. of the Amer. math. Soc., 58, 1945, p. 455—473.

##### P. LÉVY

- [3] Théorie de l'addition des variables aléatoires. 1<sup>re</sup> éd. 1937, 2<sup>ième</sup> éd. 1954.

##### T. NISIDA

- [4] On the inverse function of Poisson process. Mathematica japonicae. II, 1952, p. 135—142.

(Oblatum: 30-6-58).