

Sur les liaisons entre les grandeurs aléatoires

Par S. Bernstein, Kharkow

¹⁰ Un des traits caractéristiques les plus frappants de la science moderne est le rôle important qu'elle est obligée de réserver aux schémas de la théorie des probabilités. A première vue cette transformation de la méthode de construction scientifique paraît inconciliable avec le principe déterministe de la science classique, d'après lequel, tout phénomène particulier est équivalent à un certain nombre de grandeurs liées par des relations différentielles, fonctionnelles ou autres, telles que l'influence du reste de l'univers sur le phénomène considéré consiste uniquement à fixer un nombre suffisant de ces grandeurs pour fournir les conditions initiales ou aux limites qui déterminent complètement toutes les autres. Or, cette formule déterministe n'est qu'une déclaration invérifiable, puisqu'elle exclut la possibilité de répéter une expérience dans les mêmes conditions, et la science positive a pratiquement appliqué, en général, une formule causaliste un peu différente qui elle-même n'est qu'approximativement compatible avec la précédente. On remplace le phénomène réel par un schéma abstrait caractérisé par les mêmes grandeurs, et on admet que les données aux limites qui déterminent le phénomène peuvent être fixées plus ou moins arbitrairement indépendamment de l'état du reste de l'univers.

Il est évident que le principe déterministe remplacé ainsi par le principe de causalité, seul pratiquement utilisable, ne saurait être rigoureusement et universellement applicable sans tomber en contradiction avec la formule déterministe complète donnée plus haut; de plus, en construisant des schémas abstraits qui décomposent la réalité en parties indépendantes d'une stabilité parfaite, le principe de causalité impose de lui-même la nécessité d'une nouvelle construction logique qui réunirait dans un système de nature différente l'ensemble des phénomènes qu'il déclare indépendants.

Ainsi, par exemple, le fait que le sens du déplacement d'un point mobile sur un axe donné est indépendant du choix de la direction positive sur cet axe, s'exprime par l'affirmation que les deux signes du déplacement sont également probables. De même, l'équivalence des billets d'une loterie donnée et l'indépendance physique de leurs possesseurs du tirage de la loterie admet comme expression mathématique le fait que chacun d'eux a une probabilité de gagner entièrement déterminée par le nombre de billets qu'il possède. Cette indépendance est-elle absolue? Certainement, non. Mais les schémas abstraits ainsi construits que l'on appelle stochastiques peuvent aussi bien correspondre à la réalité que n'importe quel schéma causaliste de

S. Bernstein: Sur les liaisons entre les grandeurs aléatoires

la science classique qui suppose également son indépendance de tous les autres phénomènes.

D'ailleurs, la théorie mathématique des probabilités fondée sur une base axiomatique simple, exempte de contradictions, conduit, pourvu que le nombre d'observations soit assez grand, à des prévisions pratiquement non moins certaines que celles qui résultent des schémas déterministes, et fournit ainsi une variété infinie de procédés pour vérifier expérimentalement la concordance des schémas stochastiques avec les phénomènes réels qu'ils doivent représenter. Je n'ai pas l'intention de vous entretenir aujourd'hui de cette axiomatique caractérisée par l'affirmation formelle de l'impossibilité d'une réduction parfaite de la réalité à des schémas généraux quels qu'ils soient que j'ai proposée il y a 15 ans. Je remarquerai seulement que mon point de vue qui se dégagera plus nettement de l'exposé qui va suivre, diffère considérablement de la conception empiriste identifiant la probabilité à une fréquence, admise tacitement par la plupart des praticiens et développée systématiquement dans un livre important de M. von Mises. Mais, d'autre part, je tiens à souligner dès à présent que pour les applications effectives de la théorie des probabilités, il est indispensable, à mon avis, de postuler l'équivalence logique de tous les événements ayant la même probabilité mathématique, ce qui équivaut à admettre que la probabilité „*un*“ ne doit représenter que la certitude absolue.

Dans ces conditions, lorsque nous obtenons comme résultat de calcul dans les énoncés des théorèmes les plus importants pour les applications, connus sous le nom de lois des grands nombres, que la probabilité d'un événement A dans une expérience donnée est très petite, inférieure à $\frac{1}{N}$, où N est un nombre entier très grand, cette affirmation signifie simplement qu'il existe au moins N événements incompatibles objectivement équivalents à A (dont la distinction de A ne peut être faite que moyennant une convention indépendante de la marche de l'expérience) parmi lesquels il n'y en a qu'un seul qui se trouve réalisé.

2^o Après ces préliminaires un peu longs nous pouvons définir le schéma stochastique élémentaire, comme une expérience idéalisée, où sous des conditions bien déterminées α , l'apparition d'un événement A n'est ni nécessaire, ni impossible et cependant le lien objectif entre les conditions α et l'événement aléatoire A est caractérisé d'une façon complète et univoque par une grandeur scalaire p , nommée probabilité mathématique de A .

Examinons de plus près ce schéma simple. Souvent, on suppose l'expérience considérée non seulement indépendante de toutes les autres circonstances extérieures, mais on admet de plus que les expériences qui se suivent dans les mêmes conditions α sont mutuellement indépendantes. Cette hypothèse fondamentale de Bernoulli, n'est cependant pas la seule logiquement admissible, et il importe de se rendre compte

Grosse Vorträge

dans ce cas simple de toute la variété des liaisons effectivement possibles entre les phénomènes représentés par un même schéma stochastique stationnaire, c'est-à-dire tel que l'apparition de l'événement A dans chaque expérience admet la même probabilité tant que les résultats des autres expériences restent indéterminés.

Si nous supposons d'abord que la liaison mutuelle entre tous les phénomènes correspondant au schéma stationnaire est *symétrique*, c'est-à-dire indépendante des numéros d'ordre de nos expériences, cette liaison pourra être complètement caractérisée par la fonction $F(n)$ qui exprime la probabilité que l'événement A se reproduira invariablement dans n expériences quelconques choisies arbitrairement. Dans le cas de l'indépendance complète, on aurait $F(n) = p^n$, tandis que dans le cas déterministe, où le résultat d'une seule expérience prédétermine tous les autres, on a $F(n) = p$, quel que soit n . On peut démontrer que dans tous les autres cas, on a $p > F(n) > p^n$, et, de plus, que la fonction $F(n)$ est assujettie, en général, à la seule condition que $(-1)^k \Delta_h^k F(n) \geq 0$, qui revient à dire que la fonction $F(n)$ est complètement monotone, ou encore, qu'il existe toujours une fonction non décroissante $\psi(t)$, telle que l'on a

$$F(n) = \int_0^1 t^n d\psi(t) \quad (1)$$

avec les conditions $1 = \int_0^1 d\psi(t)$, $p = \int_0^1 t d\psi(t)$.

Il en résulte que le cas d'indépendance de Bernoulli est actuellement le seul, où la loi des grands nombres est applicable, c'est-à-dire le seul, où la fréquence $\frac{m}{n}$ admet p pour limite stochastique, car, en vertu de l'égalité (1), la probabilité de l'inégalité

$$p_0 < \frac{m}{n} < p_1$$

a pour limite $\psi(p_1) - \psi(p_0)$, lorsque $n \rightarrow \infty$.

Je signalerai, comme exemple, le cas, où

$$F(n) = p^n \prod_{i=0}^{n-1} \left(\frac{A + i}{A + i p} \right),$$

A étant une constante positive, considéré dans un ordre d'idées un peu différent par Markoff et plus tard par M. Pólya, qui est intéressant, parce qu'il conduit pour

la fréquence $x = \frac{m}{n}$ à la loi de distribution limite

$$C x^{A-1} (1-x)^{B-1}, \text{ où } B = A \frac{1-p}{p},$$

S. Bernstein : Sur les liaisons entre les grandeurs aléatoires

qui, par une transformation linéaire de x , conduit aux types les plus importants des courbes statistiques bien connues de M. Pearson.

3^o Les circonstances sont encore beaucoup plus variées, si l'on rejète l'hypothèse de la symétrie, en admettant que les liaisons entre les expériences dépendent, en général, de leurs numéros d'ordre.

Dans ces conditions la loi des grands nombres sera respectée dans des cas très étendus; il est aisé de montrer, par exemple, que pour qu'il en soit ainsi, c'est-à-dire pour que la fréquence $\frac{m}{n}$ ait la probabilité p , comme limite stochastique, il suffit que le coefficient de corrélation entre deux expériences dont les numéros d'ordre sont assez éloignés tende uniformément vers 0. Cette condition et d'autres analogues qui ne font intervenir que les liaisons entre les expériences considérées deux à deux sont cependant insuffisantes pour assurer la validité du théorème de Laplace. Mais il est digne d'intérêt que, même en admettant une certaine périodicité des liaisons qui entraîne une régularité très considérable dans la succession effective des apparitions de l'évènement A , telle par exemple, que dans chaque suite de 5 expériences, l'évènement A , ayant la probabilité $\frac{1}{2}$, se produise, au moins une et au plus quatre fois, on puisse constater néanmoins que la fréquence $\frac{m}{n}$ satisfait à la loi limite de Laplace avec la même dispersion dite normale qui caractérise le cas classique de Bernoulli, de sorte qu'aucune analyse statistique macroscopique ne permettrait de distinguer ces deux cas, et il serait nécessaire d'isoler de petits groupes d'expériences pour déceler la presque périodicité microscopique indiquée de la suite considérée.

Les exemples analogues que nous rencontrerons plus loin, où l'on ne trouve aucune trace des particularités des liaisons élémentaires dans les effets macroscopiques, ont comme base essentielle des propositions de la nature suivante :

Pour que la loi de Laplace-Gauss ($[R]$ désigne la probabilité pour que les relations R aient lieu, et le symbole \lim veut dire : limite pour n infini)

$$\lim \left[t_0 \sqrt{B_n} < m - np < t_1 \sqrt{B_n} \right] = \frac{1}{\sqrt{2\pi}} \int_{t_0}^{t_1} e^{-\frac{t^2}{2}} dt$$

soit applicable à une suite d'évènements A_n de probabilité constante p , où la dispersion $E(m - np)^2 = B_n > M n^l (l > \frac{2}{3})$, il suffit que des liaisons intenses quelconques ne soient possibles qu'entre des expériences voisines, telles que la différence i entre leurs numéros d'ordre ne dépasse pas un nombre R ($i \leq R$), nommé rayon d'activité, où $\frac{R^2}{B_n} \rightarrow 0$, pourvu qu'en dehors du rayon d'activité (pour $i > R$), les expériences

Grosse Vorträge

soient presque indépendantes: c'est-à-dire que la variation de la probabilité de A_n en puisse dépasser $\frac{B_n}{n^2}$, quels que soient les résultats d'un nombre quelconque d'expériences extérieures au rayon d'activité de l'événement A_n . Cette circonstance se présente, en particulier, dans le cas des chaînes de Markoff, dont nous parlerons plus loin. La propriété que $\frac{R^2}{B_n} \rightarrow 0$ est essentielle, car il est aisé de construire des exemples, où le théorème tomberait en défaut, dès que le rayon d'activité devient plus grand (si $\lim \frac{R^2}{B_n} > K > 0$).

Je n'ai pas besoin d'insister que l'hypothèse de la presque indépendance introduite ici est, malgré sa grande généralité, bien plus restrictive que celle des coefficients de corrélation tendant uniformément vers 0 qui, comme nous l'avons vu, suffit pour la validité de la loi des grands nombres.

On a des résultats analogues, si le schéma de chaque expérience est plus compliqué et contient un nombre quelconque de grandeurs, dont les valeurs dépendent des différentes issues possibles de l'expérience. Ainsi, par exemple un schéma stochastique parfait contenant n quantités aléatoires continues x_1, x_2, \dots, x_n sera complètement caractérisé par la surface de distribution des probabilités ou de corrélation $F(x_1, x_2, \dots, x_n)$ à n dimensions qui, des conditions extérieures déterminées α étant réalisées, possède la même réalité objective, comme si les conditions α l'avaient construite matériellement. La signification la plus simple de cette surface, consiste, d'après la loi des grands nombres, en ce qu'elle représente approximativement les fréquences correspondantes des ensembles de valeurs (x_1, x_2, \dots, x_n) , si l'expérience était susceptible d'un nombre suffisamment grand de répétitions. Cependant, sauf des cas de symétrie ou isotropie semblables à ceux qui ont été indiqués au début, les schémas stochastiques parfaits n'apparaissent le plus souvent que comme limites de suites très nombreuses d'expériences représentées par des schémas stochastiques, en général, imparfaits ou incomplètement connus, soit stationnaires, soit dynamiquement variables, suivant une loi suffisamment déterminée.

4^o C'est l'étude de la formation de ces schémas limites parfaits qui va nous occuper à présent.

Ainsi, supposons d'abord que l'on considère une succession de quantités aléatoires x_i , ayant la même probabilité d'être positives ou négatives, de sorte que $E x_i = 0$, et qui admettent toutes la même dispersion b^2 qui d'ailleurs peut nous être inconnue. La loi des probabilités de x est également inconnue et, peut être, même n'y a-t-il pas de sens d'en parler, comme d'un fait physique déterminé; supposons seulement qu'il soit suffisamment improbable que x devienne trop grand par comparaison à son écart étalon b , ou pour préciser, introduisons une hypothèse déterminée: que

S. Bernstein: Sur les liaisons entre les grandeurs aléatoires

$\frac{E|x|^3}{b^3}$ reste borné, que l'on pourrait élargir de bien des façons. Ces circonstances se présentent, par exemple, pour le déplacement horizontal x pendant un intervalle de temps Δt d'une particule en mouvement brownien. Si l'on admet de plus que ces déplacements x_i sont indépendants, on a, et c'est là un cas particulier typique du théorème de Liapounoff, que, N croissant indéfiniment, la probabilité de l'inégalité

$$S_0 < \sum_1^N x_i < S_1$$

a pour limite

$$\frac{1}{\sqrt{2\pi b^2 N}} \int_{s_0}^{s_1} e^{-\frac{s^2}{2b^2 N}} ds.$$

En posant $\frac{t}{\Delta t} = N$, on trouve donc pour le déplacement horizontal total $S(t)$ de la particule pendant un temps t , la loi de probabilité limite

$$\sqrt{\frac{\Delta t}{2\pi b^2 t}} e^{-\frac{s^2 \Delta t}{2b^2 t}} = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{s^2}{4D t}}, \text{ où } D = \frac{l^2}{2\Delta t}.$$

On obtient ainsi la formule connue du mouvement brownien comme conséquence du théorème de Liapounoff, et la constante de diffusion D pouvant être déterminée expérimentalement, on en déduit la valeur de b^2 .

Seulement il faut observer que cette formule ne saurait être rigoureusement exacte, puis qu'il n'est pas permis de faire tendre Δt vers zéro. En effet, la vitesse moyenne pendant le temps Δt de la particule est finie $\frac{x_i}{\Delta t} < c$, où c est la vitesse de la lumière; donc $b^2 < c^2 \Delta t^2$, et, par conséquent, on doit avoir $\Delta t > \frac{2D}{c^2}$.

D'ailleurs, l'hypothèse de l'indépendance des déplacements voisins x_i paraît a priori inadmissible pour des intervalles Δt suffisamment petits, et il est naturel de se demander, si on ne pourrait pas donner une interprétation microscopique plus satisfaisante de la loi macroscopique de diffusion en utilisant les généralisations du théorème de Liapounoff, analogues à celle que nous avons indiquée plus haut pour le théorème de Laplace.

Il serait trop long de détailler ici ces généralisations, applicables à un nombre quelconque de variables aléatoires qui servent, en particulier, de fondement à la théorie de la corrélation normale; l'idée essentielle de ces résultats se dégager assez clairement, je l'espère, de l'énoncé suivant.

Soient deux suites, pour fixer les idées, de grandeurs aléatoires: $x_1, x_2, \dots, x_n, \dots, y_1, y_2, \dots, y_n$. Soit

Grosse Vorträge

$$E \sum_1^n x_i = E \sum_1^n y_i = 0, \quad E \left(\sum_1^n x_i \right)^2 = A_n^2, \quad E \left(\sum_1^n y_i \right)^2 = C_n^2,$$

$$E \left[\left(\sum_1^n x_i \right) \left(\sum_1^n y_i \right) \right] = R_n A_n C_n.$$

En admettant que x_i, x_K, y_i, y_K ne peuvent être liées d'une façon sensible que tant que $\frac{|i-k|}{A_n} L \rightarrow 0, \frac{|i-k|}{C_n} M \rightarrow 0$, où L et M sont les maxima respectifs de $|x_i|$ et $|y_i|$, et qu'en dehors de ce voisinage les quantités x et y deviennent presque indépendantes dans un sens analogue à celui, que nous avons défini antérieurement, la probabilité des inégalités simultanées

$$t_0 A_n < \sum_1^n x_i < t_1 A_n$$

$$t_0' C_n < \sum_1^n y_i < t_1' C_n$$

a pour limite

$$\frac{1}{2\pi\sqrt{1-R_n^2}} \int_{t_0'}^{t_1'} \int_{t_0}^{t_1} e^{-\frac{t^2 + t'^2 - 2R_n t t'}{2(1-R_n^2)}} dt dt',$$

pour $n \rightarrow \infty$.

Nous voyons ainsi que dans des cas très étendus, l'itération d'expériences mutuellement liées, représentées par des schémas stochastiques extrêmement variés, qui peuvent être imparfaits, pourvu qu'elles deviennent presque indépendantes, à l'extérieur d'un domaine d'activité suffisamment petit, conduit à des schémas stochastiques parfaits bien déterminés dans lesquels la corrélation normale et la loi de Gauss, en particulier, apparaissent comme des lois de la nature d'un caractère aussi universel que le principe de l'inertie en mécanique.

C'est donc un mérite incontestable de Galton et de M. Pearson d'avoir fondé et développé l'étude de la corrélation normale et d'avoir prévu leur importance pour les applications; mais ce n'est que grâce à des propositions analogues à celle qui vient d'être énoncée que cette prévision reçoit une justification mathématique satisfaisante et que nous pouvons nous rendre compte, par exemple, de la raison profonde de l'existence approximative de la corrélation normale entre les grandeurs de divers organes chez les êtres vivants, ainsi que de l'applicabilité approchée aux propriétés qui dépendent de plusieurs gènes de la loi de régression héréditaire de Galton. D'ailleurs, cette loi d'hérédité de Galton qui suppose que le coefficient de corrélation entre les générations successives diminue en progression géométrique, fournit le premier exemple de chaînes stochastiques dont l'étude mathématique a été fondée par Markoff sur une base toute différente.

S. Bernstein : Sur les liaisons entre les grandeurs aléatoires

En revenant au mouvement brownien nous pourrions déduire de la proposition citée appliquée à une seule grandeur une variété infinie de schémas microscopiques qui conduirait à la même formule de diffusion. Les expériences macroscopiques ordinaires ne permettraient pas de faire le choix entre ces différentes interprétations, mais, s'il était possible d'effectuer une sorte de filtration des déplacements microscopiques, correspondant à des durées de temps assez courtes pour que la loi de diffusion soit encore inapplicable, on serait amené probablement à remplacer l'hypothèse de l'indépendance complète par une autre plus conforme à la réalité.

Ainsi, par exemple, pour faire une hypothèse mathématique précise aussi simple que possible, on pourrait supposer que

$$x_{i+1} = \left(1 - \frac{1}{A}\right) x_i + a_{i+1}, \quad (2)$$

où a_{i+1} est une quantité aléatoire indépendante de x_i avec $E a_i = 0$, $E a_i^2 = \beta^2$, $\frac{E |a_i|^3}{\beta^3}$ borné, $A \geq 1$ étant une constante que nous supposerons ensuite très grande,

pour tenir compte de l'inertie de la particule pendant des intervalles de temps très petits. Un calcul facile montre alors que $S_N = \sum_1^N x_i$ satisfait pour $N = \frac{t}{\Delta t}$ très grand

à la formule de diffusion avec $D = \frac{A^2 \beta^2}{2 \Delta t}$, et, si on admet de plus que $A \Delta t \rightarrow 0$,

la différence $S_{N_1} - S_{N_0}$ (où $N_1 > N_0$) satisfera également à la loi de Gauss avec la dispersion $\frac{A^2 \beta^2}{\Delta t} (N_1 - N_0) = 2 D (t_1 - t_0)$, comme dans le cas de l'indépendance complète des déplacements x_i ; donc, toutes les conséquences macroscopiques seront les mêmes. L'augmentation de A permet de diminuer la dispersion élémentaire

$b^2 = \frac{A \beta^2}{2 - \frac{1}{A}} \sim \frac{D}{A} \Delta t$ du déplacement x_i , et par conséquent, d'envisager des inter-

valles de temps Δt plus petits sans introduire des vitesses inadmissibles; mais la condition $A \Delta t \rightarrow 0$ (nécessaire pour la concordance des effets macroscopiques) exclut la possibilité de faire b^2 de l'ordre de Δt^2 ; le passage rigoureux à la limite $\Delta t \rightarrow 0$ reste donc toujours mécaniquement illégitime, amenant des vitesses infinies.

J'ai insisté un peu longuement sur la discussion de ce cas typique simple pour mettre en évidence que l'emploi des relations différentielles dont nous parlerons tout à l'heure, introduit par les physiciens MM. Einstein, Smoluchowsky, Fokker et Planck pour l'étude des liaisons stochastiques qui joue un rôle important dans la théorie moderne des probabilités, doit être considéré surtout, comme une puissante méthode technique, qui ne devrait pas nous laisser perdre de vue le caractère essen-

Grosse Vorträge

tiellement discontinu du mécanisme microscopique de la formation des schémas stochastiques qui échappe à la représentation différentielle.

5° Ainsi, sous les réserves indiquées au sujet de sa signification mécanique, on peut définir a priori une variable aléatoire $S(t) = \lim_{n \rightarrow \infty} S_{N=nt}$ dépendant de t qui satisfera rigoureusement à la loi limite

$$\frac{1}{\sqrt{4\pi Dt}} e^{-\frac{s^2}{4Dt}},$$

en jouissant, conformément aux résultats précédents, de la propriété que les valeurs de $S(t)$ à des moments différents t_0 et $t_1 > t_0$ se trouvent en corrélation normale, ayant pour coefficient de corrélation $R(t_0, t_1) = \sqrt{\frac{t_1}{t_0}}$, ce qui signifie que $S_0 = S(t_0)$ étant fixé, la loi conditionnelle des probabilités de $S(t_1)$ ou loi de passage est

$$w(s_0, t_0, s_1, t_1) = \frac{1}{\sqrt{4\pi D(t_1 - t_0)}} e^{-\frac{(s_1 - s_0)^2}{4D(t_1 - t_0)}}, \quad (3)$$

quelles que soient les valeurs de $S(t)$ aux moments $t < t_0$. Lorsque cette dernière circonstance se présente, c'est-à-dire, si la probabilité du passage de $S_0(t_0)$ à $S_1(t_1)$ est indépendante des états de $S(t)$ antérieurs à t_0 , dès que la valeur $S_0(t_0)$ est fixée, on dit que les valeurs de $S(t)$ forment une chaîne. Lorsque, en plus, comme ici, la corrélation est normale nous dirons que la chaîne est normale. Dans ce dernier cas, on a $R(t_0, t_1) = \frac{\varphi(t_0)}{\varphi(t_1)}$, où $\varphi(t)$ est une fonction arbitraire croissante.

Si l'espérance mathématique $E S(t) = f(t)$ est une fonction continue ainsi que $R(t_0, t)$ et que de plus $R(t, t) = 1$, comme cela a lieu actuellement, nous dirons que la variable aléatoire $S(t)$ est une fonction *quasi continue* ou stochastiquement continue de t . Je crois qu'il est préférable de ne pas employer ici le terme ordinaire de continuité. En effet, il est facile de montrer, par exemple, dans le cas actuel que la probabilité des inégalités simultanées $|S - S_0| < |x - x_0|^{\frac{1}{2} - \alpha}$ pour toutes les valeurs de $t_0 < T$, tend vers 1, quel que soit $\alpha > 0$, lorsque $t - t_0$ tend vers 0; mais, d'autre part, il est aisé de voir que, si la variable aléatoire $S(t)$ est liée en chaîne, il n'est pas possible de fixer un module de continuité $\omega(\Delta t)$ tel que l'inégalité

$$\Delta S < \omega(\Delta t)$$

soit certaine.

La turbulence de la variable aléatoire dépend essentiellement du coefficient de corrélation $R(t_0, t_1)$; si l'on veut, par exemple, que la variable quasi continue $S(t)$ possède une dérivée stochastique (ce qui, dans le cas d'une chaîne, ne saurait se

S. Bernstein : Sur les liaisons entre les grandeurs aléatoires

présenter qu'en des points isolés), il faut que l'on ait $\left[\frac{\partial R(t_0, t)}{\partial t} \right]_{t=t_0} = 0$. Ainsi, si dans l'équation (2) nous rejetons la condition $\Delta \Delta t = \frac{A}{n} \rightarrow 0$, en supposant au contraire, $\frac{1}{A} = 0$, de sorte qu'elle prend la forme

$$\Delta_2 S_{i-1} = \alpha_{i+1},$$

on voit que, n croissant indéfiniment, les valeurs de $S(t)$ aux différents moments sont encore en corrélation normale, mais la variable $S(t)$ ne forme plus une chaîne, le coefficient de corrélation ayant pour valeur

$$R(t_0, t_1) = \sqrt{\frac{t_0}{t_1} \left[\frac{3}{2} - \frac{1}{2} \frac{t_0}{t_1} \right]},$$

$$\left[R(t, t) = 1, \frac{\partial}{\partial t} R(t, t) = 0 \right];$$

on peut montrer que, $t_2 - t_0$ étant pris suffisamment petit, la probabilité de l'inégalité

$$\left[\frac{S(t_1) - S(t_0)}{t_1 - t_0} - \frac{S(t_2) - S(t_0)}{t_2 - t_0} \right] < \varepsilon,$$

où ε est un nombre positif donné et $t_0 < t_1 < t_2$, devient aussi voisine de 1 qu'on le veut.

Sans insister sur les propriétés des variables aléatoires, quasi continues analogues à celle là, j'observerai que, $S(t_0)$ et $S(t_1)$ étant fixées, la loi des probabilités de $S(t_2)$ dépendra actuellement des valeurs de $S(t)$ antérieures à t_0 , mais cette dépendance sera d'autant plus faible que t_1 , sera voisin de t_0 ; l'introduction d'une variable auxiliaire $S'(t)$ (dont l'observation simultanée avec $S(t)$ pourrait ne pas être réalisable) permettrait de former une chaîne double. En tout cas la loi de $S(t_n)$ ne saurait jamais être indépendante des valeurs de $S(t)$ pour $t < t_0$, lorsque $S(t_0), S(t_1) \dots S(t_{n-1})$ sont données en des points arbitraires $t_0 < t_1 < \dots < t_{n-1} < t_n$, sans que cette loi soit complètement déterminée par la valeur unique $S(t_{n-1})$, c'est-à-dire sans représenter une chaîne, d'après la définition précédente.

Un des problèmes typiques les plus importants qui se présentent au sujet des variables aléatoires stochastiquement continues est, par exemple, celui de la détermination de la probabilité $P_{a,b}$ que $a < S(t_1) < b$, si, outre la condition $S(0) = 0$ on exige que $S(t)$ satisfasse aux inégalités $F_1(t) < S(t) < F_2(t)$ pour toute valeur positive de $t < t_1$, où $F_1(t), F_2(t)$ sont deux fonctions données de t .

L'étude directe de ces problèmes par les anciennes méthodes de la théorie des probabilités est, en général, assez pénible, se ramenant à la recherche de la limite de certaines intégrales multiples d'ordre infiniment croissants qui s'expriment ex-

Grosse Vorträge

plicitement au moyen de la surface de corrélation, connue, par hypothèse, entre les valeurs successives de la variable aléatoire considérée.

Au contraire, l'utilisation des équations linéaires aux dérivées partielles s'est montrée très féconde pour la résolution des problèmes de ce genre qui se rencontrent souvent en physique.

Je résumerai, à titre d'exemple, la solution du problème indiqué pour le cas fondamental, correspondant à la loi de diffusion, d'après M. Kolmogoroff, qui a le plus contribué au développement mathématique de la méthode.

Il est bien connu et facile de vérifier que la probabilité de passage $w(s_0, t_0, s_1, t_1)$ donnée par la formule (3) satisfait, comme fonction de s_1 et t_1 à l'équation

$$D \frac{\partial^2 w}{\partial s_1^2} = \frac{\partial w}{\partial t_1},$$

et considérée, comme fonction du lieu s_0 et du moment t_0 de départ, elle satisfait à l'équation conjuguée

$$D \frac{\partial^2 w}{\partial s_0^2} = - \frac{\partial w}{\partial t_0} \quad (4)$$

Il est à peu près évident que, si l'on pose $P_{a,b} = P_{a,b}(0, 0)$, la probabilité $P_{a,b}(s_0, t_0)$ considérée comme fonction de (s_0, t_0) , satisfera également à la dernière équation pour toutes les valeurs de (s_0, t_0) de la région ω limitée par les droites $t = 0$ et $t = t_1$, et par les courbes $s = F_1(t)$ et $s = F_2(t)$. La fonction $P_{a,b}(s_0, t_0)$ pourra ainsi être déterminée, comme la solution de l'équation (4) satisfaisant aux conditions aux limites $P_{a,b}(s, t_1) = 1$, si $a < s < b$, et $P_{a,b}(s, t_1) = 0$, si $s < a$ ou si $s > b$, ainsi que $P_{a,b}[F_1(t), t] = P_{a,b}[F_2(t), t] = 0$.

Toutes les fois, où la loi de probabilités de passage de la chaîne satisfera à une équation analogue du type parabolique, la même méthode sera applicable. Mais le cas le plus simple et aussi, sans doute, le plus important est celui, où la loi de passage est une loi de Gauss; il se ramène alors formellement par un changement de variables évident au cas qui vient d'être examiné.

On reconnaît, de plus, facilement que si la dispersion de cette loi est une fonction de $t_1 - t_0$ seulement, elle est nécessairement de la forme

$$\frac{1}{\sqrt{4\pi D(t_1 - t_0)}} e^{-\frac{[s_1 - s_0 - f(t_1) + f(t_0)]^2}{2D(t_1 - t_0)}}$$

et correspond ainsi au cas, où la variable S est, comme précédemment, la limite d'une somme S_N de quantités ΔS_i de dispersion constante, pouvant, en outre, posséder des espérances mathématiques dépendant de t .

Un autre cas intéressant est celui, où c'est le coefficient de corrélation $R(t_0, t_1)$ qui est fonction de $t_1 - t_0$ seulement. Puisque dans le cas d'une chaîne normale on a

S. Bernstein : Sur les liaisons entre les grandeurs aléatoires

toujours, comme nous l'avons vu, $R(t_0, t_1) = \frac{\varphi(t_0)}{\varphi(t_1)}$, il est aisé de montrer que l'on doit avoir actuellement $R(t_0, t_1) = \rho^{t_1-t_0}$; le coefficient de corrélation décroîtra ainsi en progression géométrique avec le temps.

D'ailleurs, la variable aléatoire $S(t)$ liée en chaîne normale la plus générale peut être obtenue comme limite de la somme S_N de quantités aléatoires dépendantes $\Delta S_i = S_{i+1} - S_i$; définies par les relations

$$\Delta S_i = c_i S_i \Delta t + a_i, \quad (5)$$

où c_i est une fonction donnée de i (ou de t), et a_i des quantités indépendantes de S_i telles que $E a_i \propto a_i \Delta t$, $E a_i^2 \propto b_i \Delta t$, $\frac{E |a_i|^3}{\Delta t^{\frac{3}{2}}} \rightarrow 0$, a_i et b_i étant des fonctions finies de t .

Lorsque $c_i = c \neq 0$ et $b_i = b$ sont des constantes, on trouve justement le coefficient de corrélation $R(t_0, t_1) = \rho^{t_1-t_0} = e^{c(t_1-t_0)}$, tandis que pour $c \rightarrow 0$ on retombe dans le cas précédent correspondant au mouvement brownien.

La considération d'équations analogues linéaires aux différences finies à plusieurs variables conduit pareillement à des chaînes multiples dont la loi limite est encore la corrélation normale. Dans ce cas important, le passage à la limite peut être fait rigoureusement et explicitement par les méthodes classiques. Il n'en sera plus de même, si l'équation (5) est remplacée par une équation non-linéaire.

Cette fois, la loi limite des probabilités ne sera plus une loi de Gauss, et on n'a même pas encore donné, autant que je sache, de démonstration rigoureuse suffisamment générale de son existence. Pour obtenir la loi limite des probabilités $p(s, t)$ on admet, au contraire, a priori que cette loi existe, et, de plus, qu'elle est représentée par une fonction possédant des dérivées jusqu'à un certain ordre; dans ces conditions, on démontre que celle-ci satisfait à une certaine équation aux dérivées partielles l'équation de Fokker et Planck.

Vu l'importance de cette méthode, je résumerai brièvement la déduction générale de l'équation de Fokker-Planck que je présenterai sous une forme un peu différente de celle qu'on lui donne habituellement, en introduisant, comme fonction inconnue, la probabilité intégrale $P(s_1, t_1) = [s < s_1]$ au moment t_1 au lieu de la densité des probabilités $p(s_1, t_1) = \frac{\partial P(s_1, t_1)}{\partial s_1}$, ce qui permet de ne pas supposer cette dernière deux fois dérivable.

La signe $E_s(u)$ désignant l'espérance mathématique de u , lorsque S est donné, admettons seulement que

$$\lim_{\Delta t \rightarrow 0} E_s \left(\frac{\Delta s}{\Delta t} \right) = f(s, t), \quad \lim_{\Delta t \rightarrow 0} E_s \left(\frac{\Delta s^2}{\Delta t} \right) = B^2(s, t), \quad \lim_{\Delta t \rightarrow 0} E_s \left(\frac{|\Delta s|^3}{\Delta t} \right) = 0,$$

Grosse Vorträge

où les fonctions f et B sont continues, ainsi que $\frac{\partial B}{\partial s}$; supposons, en outre, que $P(s, t)$ est continue et bornée avec ses dérivées des deux premiers ordres. Introduisons ensuite une fonction arbitraire $F(s)$ trois fois dérivable, s'annulant à l'infini avec ses dérivées et telle que les espérances mathématiques correspondantes aient un sens. Dans ces conditions, on a

$$\lim E \left[\frac{F(s + \Delta s) - F(s)}{\Delta t} \right] = \lim E \left[F'(s) \frac{\Delta s}{\Delta t} + \frac{1}{2} F''(s) \frac{\Delta s^2}{\Delta t} \right].$$

Or, le premier membre est égal à

$$\int_{-\infty}^{\infty} F(s) \frac{\partial^2 P(s, t)}{\partial s \partial t} ds = - \int_{-\infty}^{\infty} F'(s) \frac{\partial P(s, t)}{\partial t} ds$$

et le second membre est

$$\int_{-\infty}^{\infty} \left[F'(s) f(s, t) + \frac{1}{2} F''(s) B^2(s, t) \right] \frac{\partial P(s, t)}{\partial s} ds =$$

$$\int_{-\infty}^{\infty} F'(s) \left[f(s, t) \frac{\partial P(s, t)}{\partial s} - \frac{1}{2} \frac{\partial}{\partial s} B^2(s, t) \frac{\partial P(s, t)}{\partial s} \right] ds;$$

donc, le caractère arbitraire de $F'(s)$ amène, comme conséquence nécessaire, à l'identité

$$\frac{\partial P(s, t)}{\partial t} = -f(s, t) \frac{\partial P(s, t)}{\partial s} + \frac{1}{2} \frac{\partial}{\partial s} \left[B^2(s, t) \frac{\partial P(s, t)}{\partial s} \right].$$

L'équation de Fokker-Planck se déduit de celle là par différentiation par rapport à s , si l'on admet, en plus, l'existence et la continuité de $\frac{\partial f}{\partial s}$ et $\frac{\partial^2 B}{\partial s^2}$:

$$\frac{\partial p}{\partial t} = - \frac{\partial}{\partial s} (fp) + \frac{1}{2} \frac{\partial^2}{\partial s^2} (B^2 p).$$

La probabilité de passage $p(s_0, t_0, s, t)$ de la valeur s_0 au moment t_0 à la valeur s au moment t satisfait à cette équation; considérée, comme fonction s_0, t_0 , elle satisfait à une équation analogue conjuguée de la précédente. Le cas de plusieurs variables se traite d'une façon semblable.

Observons, d'ailleurs, que, d'après la définition même de la chaîne, la fonction $p(s_0, t_0, s, t)$ satisfait à la relation intégrale

$$p(s_0, t_0, s_2, t_2) = \int_{-\infty}^{\infty} p(s_0, t_0, s_1, t_1) p(s_1, t_1, s_2, t_2) ds, \quad (t_0 < t_1 < t_2)$$

introduite par M. Chapman et considérée antérieurement par Smoluchowsky dans

S. Bernstein: Sur les liaisons entre les grandeurs aléatoires

le cas particulier, où la probabilité de passage est de la forme $p(s_0, s, t - t_0)$. Dans un mémoire tout récent que je n'ai pas eu le temps d'examiner en détail, M. Hostinský étudie directement cette équation intégrale sans introduire les hypothèses spéciales qui précèdent, et la solution sous forme de série qu'il en donne, ne satisfait pas nécessairement à l'équation aux dérivées partielles de Fokker-Planck; il serait intéressant, comme l'indique M. Hostinský, de discuter d'une façon complète les conditions nécessaires et suffisantes pour que $p(s_0, t_0, s, t)$ satisfasse à l'équation de Planck.

6° L'équation de M. Chapman pour le cas, où $t_0 < t_1 < t_2$ sont des moments successifs entre les-quels on ne peut interposer, tout au plus, qu'un nombre limité de termes, représente la définition des chaînes discrètes, dont les propriétés essentielles ont été découvertes par Markoff dans une série de travaux remarquables (antérieurs aux recherches des physiciens nommés plus haut) consacrés au cas, où la variable aléatoire $S(t_i) = x_i$ ne prend qu'un nombre limité de valeurs: a_1, a_2, \dots, a_n .

Grace aux développements qu'ont reçu les idées de Markoff dans les travaux récents de MM. Romanowsky, Hostinský, Fréchet et v. Mises, la théorie des chaînes discrètes de Markoff est devenu un des chapitres les plus parfaits du calcul classique des probabilités, susceptible de nombreuses applications et de diverses généralisations.

D'après Markoff, une suite de quantités x_i forme une chaîne simple, si on a une succession d'expériences stochastiques E_i admettant h issues possibles, correspondant, respectivement, aux h égalités $x_i = a_k^{(i)}$ ($k = 1, 2, \dots, h$), et telles que les probabilités $p_{k,e}^{(i+1)}$ des égalités $x_{i+1} = a_e^{(i+1)}$ sont bien déterminées, quels que soient les résultats des expériences antérieures, dès qu'il est établi que l'expérience E_i a amené l'égalité $x_i = a_k^{(i)}$, de sorte que le schéma de l'expérience E_{i+1} , devient parfait, lorsque le résultat de l'expérience E_i est fixé. Dans le cas des suites quelconques la probabilité de l'égalité $x_{i+1} = a_e^{(i+1)}$ pourrait dépendre des expériences antérieures; on peut donc définir avec Markoff, comme première généralisation de la chaîne simple, une chaîne du second ordre par la propriété que la probabilité de l'égalité $x_{i+1} = a_e^{(i+1)}$ ne dépend, en plus, que du résultat de l'expérience E_{i-1} .

L'étude des chaînes d'ordre supérieur ainsi que des chaînes multiples contenant plusieurs variables peut être ramenée formellement à celle des chaînes simples par l'introduction, pour le cas du second ordre, pas exemple, de la nouvelle variable $z_i = \lambda x_i + \mu x_{i-1}$, où λ et μ sont deux paramètres arbitraires, qui prend h^2 valeurs $\lambda a_k^{(i)} + \mu a_e^{(i-1)}$. Ainsi dans l'aperçu général qui va suivre nous pouvons nous borner aux chaînes simples sans nous arrêter sur quelques particularités intéressantes des chaînes d'ordre supérieur.

Un problème fondamental qui se pose est la question de l'existence d'une loi limite des probabilités des égalités $x_n = a_e^{(n)}$ indépendante de la loi des probabilités des égalités initiales $x_1 = a_e^{(1)}$. Son étude relève de la théorie des équations linéaires

Grosse Vorträge

aux différences finies, et pour simplifier, nous allons nous borner au cas important, où les probabilités de passage $p_{k,e}^{(i+1)} = p_{k,e} \geq 0$, ainsi que les valeurs $a_e^{(i)}$ ne dépendent pas de i . Dans ces conditions, $P_e^{(n)}$ désignant la probabilité de l'égalité $x_n = a_e$ à la n -ième expérience, on a manifestement

$$P_e^{(n+1)} = \sum_{i=1}^h P_i^{(n)} p_{i,e}, \quad (6)$$

avec $\sum_{e=1}^h p_{i,e} = 1$ qui exprime que les valeurs a_e sont les seules possibles. Il en résulte que le système de h équations homogènes à h inconnues P_i

$$P_e = \sum_{i=1}^h P_i p_{i,e} \quad (7)$$

admet nécessairement un système de solutions, différentes de 0, tel que $\sum_{i=1}^h P_i = 1$, et on démontre par des considérations élémentaires qu'il y en a toujours au moins un tel que $P_i \geq 0$.

Donc, quelles que soient les probabilités de passage $p_{i,e} \geq 0$, un état stationnaire de la chaîne est toujours possible, et il sera réalisé, si $P_i^{(1)} = P_i$ dans la première expérience. C'est d'ailleurs évident dans le cas simple très important, où $p_{i,k} = p_{k,i}$, c'est-à-dire, lorsque la probabilité du passage de a_i à a_k est la même que celle du passage inverse; la solution stationnaire toujours possible sera alors la distribution uniforme

$$P_i = \frac{1}{h}.$$

Pour que le système de solutions de (7), correspondant à l'état stationnaire soit unique, il est nécessaire et suffisant que l'unité qui est toujours une racine de l'équation caractéristique

$$\begin{vmatrix} p_{11} - \lambda & p_{21} & \dots & p_{h1} \\ p_{12} & \dots & \dots & p_{h2} \\ \dots & \dots & \dots & \dots \\ p_{1h} & \dots & \dots & p_{hh} - \lambda \end{vmatrix} = 0$$

soit une racine simple de cette équation. D'après M. v. Mises, cette condition est équivalente au fait que la matrice $\| p_{i,k} \|$ est non décomposable, c'est-à-dire que, quels que soient i et k , le passage de a_i à a_k dans un sens au moins soit toujours possible après un nombre fini assez grand d'expériences: pour éviter ce cas d'exception il suffirait donc, qu'on n'ait pas simultanément $p_{i,k} = 0$ et $p_{k,i} = 0$.

Le cas particulier important, où tous les $p_{i,k} > 0$, avait déjà été complètement étudié par Markoff qui a montré que dans ces conditions on a non seulement un

S. Bernstein: Sur les liaisons entre les grandeurs aléatoires

système stationnaire unique P_i , mais que l'on a aussi $\lim_{n \rightarrow \infty} P_i^{(n)} = P_i$, quelle que soit la distribution des probabilités initiales, de sorte que la fréquence limite P_i des égalités $x_n = a_i$ est a priori complètement déterminée par la matrice des probabilités de passage $p_{ik} > 0$.

Cependant, en général, l'existence d'un système stationnaire unique n'implique pas qu'il servira de limite à tous les systèmes dynamiques, ce qui est évident, par exemple, pour la matrice

$$\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$$

Pour que tous les systèmes dynamiques tendent vers le système stationnaire unique, il est nécessaire et suffisant que l'équation caractéristique ne possède pas d'autre racine que 1 de module un . Dans le cas contraire, c'est-à-dire, lorsqu'il existera une puissance m de la matrice fondamentale qui sera décomposable, comme cela a lieu, par exemple, pour la matrice

$$\begin{vmatrix} 0 & \alpha & 0 & \beta \\ \beta & 0 & \alpha & 0 \\ 0 & \beta & 0 & \alpha \\ \alpha & 0 & \beta & 0 \end{vmatrix}$$

le système dynamique au lieu d'avoir nécessairement pour limite l'état stationnaire, tendra, en général, à se reproduire périodiquement avec une période égale à m (qui est 2 dans notre exemple).

Cependant, même dans ce dernier cas la fréquence moyenne des égalités $x_n = a_e$ dans l'ensemble des expériences tendra, quelles que soient les conditions initiales, vers la même limite P_e qui correspond à l'état stationnaire. Seulement le théorème limite généralisé de Liapounoff ne sera plus, en général, applicable à la somme $\sum_1^n x_i$.

Je n'insisterai pas ici sur les conditions suffisantes plus ou moins larges pour la validité de ce dernier théorème, en me bornant à observer que dans l'hypothèse de Markoff $p_{ik}^{(n)} > \alpha > 0$ et, même, lorsque les nombres $p_{ik}^{(n)}$ tendent vers 0 assez lentement, il est une conséquence de la proposition générale mentionnée antérieurement.

⁷⁰ L'étude du cas, où x_e peut prendre une infinité de valeur, se traite par un passage à la limite, la théorie des équations intégrales de Fredholm fournissant un appareil mathématique parfaitement adapté à cette étude dans le cas ou moins, où l'ensemble des valeurs possibles est continu et borné. En supposant, pour fixer les idées, que l'on a $a \leq x_e \leq b$, l'équation (6) se trouve remplacée par

Grosse Vorträge

$$P_{n+1}(x_{n+1}) = \int_a^b P_n(x_n) f(x_n, x_{n+1}) dx_n,$$

où la probabilité $f(x_n, x_{n+1}) \geq 0$ de passage de x_n à x_{n+1} est assujettie à la condition $\int_a^b f(x_n, x_{n+1}) dx_{n+1} = 1$; la distribution stationnaire est déterminée par l'équation

$$P(x) = \int_a^b P(y) f(y, x) dy, \quad \left[\int_a^b P(x) dx = 1 \right]$$

et on démontre facilement que dans les conditions de Markoff: $f(y, x) > 0$, elle est unique, et de plus, que $P_n(x) \rightarrow P(x)$, quelle que soit la distribution initiale $P_1(x_1)$.

En particulier, il est évident que l'on aura $P(x) = \frac{1}{b-a}$, si $f(y, x)$ est symétrique.

Ce résultat général, s'appliquant à un domaine fini à un nombre quelconque de dimensions, justifie dans des cas très étendus l'hypothèse de la distribution uniforme des probabilités.

Le cas, où les limites (a, b) deviennent infinies présente une exception qui mérite d'être signalée.

Il est aisé de montrer, en effet, que la fonction positive et continue $f(y, x)$ étant symétrique, on a, quelles que soient les valeurs données α et β et quelque petit que soit ε

$$\int_\alpha^\beta P_n(x) dx < \varepsilon,$$

pourvu que n soit assez grand. Ainsi, la variable aléatoire finirait par disparaître de tout domaine fini donné, quelle que soit sa distribution initiale.

D'après cette remarque, si l'on admet que la variable aléatoire reste indéfiniment observable, il faut ou bien qu'elle soit bornée a priori, ou bien que la loi de passage ne soit pas symétrique: il est nécessaire alors que la loi de passage présente une dissymétrie telle, que la diminution de la grandeur considérée devienne plus probable que son augmentation, dès qu'elle reçoit des valeurs suffisamment grandes. Ainsi il suffirait, pour que $P_n(x)$ tende vers une limite déterminée, que l'on ait

$$\int_{-\infty}^{\infty} x^2 f(y, x) dx < a + \rho y^2, \quad \text{où } \rho < 1;$$

au contraire, si $\rho \geq 1$ cette limite pourra ne pas exister.

Un exemple instructif à cet égard est fourni par la loi de passage

$$f(y, x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\rho y)^2}{2\sigma^2}}$$

S. Bernstein: Sur les liaisons entre les grandeurs aléatoires

qui, pour $\rho < 1$, conduit à la distribution limite

$$P(x) = \frac{1}{\sigma} \sqrt{\frac{1-\rho^2}{2\pi}} e^{-\frac{1-\rho^2}{2\sigma^2} x^2},$$

tandis que pour $\rho \geq 1$ on obtient une raréfaction infinie.

En revenant au cas, où le domaine de la variable est borné nous devons nous arrêter spécialement sur l'application qu'on a faite de la théorie des chaînes de Markoff pour suppléer à l'hypothèse ergodique dans la théorie cinétique des gaz. L'état A d'une agglomération de molécules possédant une quantité donnée d'énergie, renfermées dans une enceinte, est représenté habituellement, comme on sait, par la position A d'un point assujéti à rester dans une partie déterminée de volume v_0 de l'espace à $6n$ dimensions, où n est le nombre de molécules. Admettons que la probabilité de passage p_{ik} d'un état A_i au moment t à l'état A_k au moment $t + \Delta t$ soit la même que celle du passage de A_k à A_i ($p_{ik} = p_{ki} > 0$), pourvu que l'intervalle de temps Δt ne soit pas trop petit. On peut alors affirmer, d'après ce qui précède, sans faire aucune hypothèse sur le mécanisme du phénomène que, quel que soit l'état initial, au bout d'un temps t assez grand par rapport à l'intervalle Δt , tous les états A_i tendent à devenir également probables, ou, ce qui revient au même, la probabilité de rencontrer le point représentatif A dans une région donnée de volume v tend vers le rapport des volumes $\frac{v}{v_0}$. De plus, en vertu de la loi des grands nombres qui est applicable aux chaînes de Markoff, si le système est abandonné à lui-même pendant un temps T_0 suffisamment grand, la probabilité du séjour du point A dans un volume v pendant une durée de temps $T = \frac{v}{v_0} T_0$ différera de 1 aussi peu qu'on veut.

Ce théorème ergodique, fondamental pour la théorie cinétique des gaz qui, comme on le voit, est une conséquence immédiate de l'interprétation stochastique, si plausible, donnée plus haut, n'a pu être déduit pendant longtemps de l'étude directe du système d'équations différentielles qui déterminent la trajectoire du point représentatif, et il était permis de douter de sa compatibilité avec l'ancienne représentation mécanique du mouvement des molécules. Ce n'est que récemment que M. Birkhoff est parvenu à démontrer un théorème équivalent en faisant des hypothèses d'une nature très générale sur les équations différentielles considérées.

Cette concordance remarquable semble prouver, d'une part, que les schémas causalistes ordinaires de la mécanique classique restent suffisants pour expliquer les lois macroscopiques de la théorie des gaz et, d'autre part, que les interprétations stochastique et déterministe d'un phénomène peuvent conduire à des conséquences limites équivalentes. Actuellement, la différence entre les deux interprétations consiste essentiellement en ce que la mécanique opère avec des intervalles de temps

Grosse Vorträge

infiniment petits, tandis que l'emploi des chaînes de Markoff exclut la possibilité de faire tendre Δt vers 0; si ce passage à la limite était légitime, le régime stationnaire se trouverait réalisé momentanément.

Mais il y a lieu de noter une autre propriété fondamentale des chaînes non stationnaires qui les distingue des trajectoires dynamiques: c'est leur irréversibilité.

En effet, l'égalité $p_{ik} = p_{ki}$ exprime seulement que la probabilité d'aller de A_i à A_k est la même que celle d'aller de A_k à A_i au passage de l'état antérieur à l'état ultérieur; donc, si p'_{ik} désigne la probabilité que le point arrivé à A_k soit parti de A_i , on a les égalités

$$P_k^{(n+1)} p'_{ik} = P_i^{(n)} p_{ik}, \quad P_i^{(n+1)} p'_{ki} = P_k^{(n)} p_{ki}.$$

Par conséquent, tant que le régime stationnaire n'est pas établi, on n'aura pas, en général, $p'_{ik} = p'_{ki} = p_{ik}$, et, de plus, tandis que p_{ik} dans une chaîne est, par définition, fixé d'avance, indépendamment de l'état initial, les probabilités p'_{ik} sont, au contraire, a priori indéterminées. D'ailleurs, le fait que la différence entre les extréma de $P_e^{(n+1)}$ est toujours plus petite que la différence correspondante des $P_e^{(n)}$ prouve assez clairement que dans une chaîne de Markoff la tendance vers la distribution uniforme des probabilités ne saurait avoir lieu dans les deux sens.

8° Ainsi, tandis que les équations de la mécanique classique déterminent d'une façon analogue l'avenir et le passé, il n'en est pas de même pour les chaînes. Par conséquent, si l'on veut reconstituer cette symétrie entre le passé et le futur (ce dont, personnellement, je ne vois pas la nécessité) il faut renoncer à l'emploi des chaînes du type de Markoff et les remplacer par des schémas d'une nature différente. A cet effet, considérons une succession de grandeurs aléatoires x_i , telles que la loi des probabilités x_i ne soit déterminée, $f(x_{i-h}, x_{i+k}, x_i)$, que lorsqu'on connaît une valeur antérieure x_{i-h} et une valeur postérieure x_{i+k} , de sorte que la connaissance supplémentaire des valeurs d'indice inférieur à $i-h$ ou supérieur à $i+k$ ne modifie pas la loi des probabilités de x_i . Dans ces conditions, si l'on se donne arbitrairement les valeurs initiale et finale de x_0 et x_n , par exemple, toutes les grandeurs intermédiaires x_i ($i=1, 2, \dots, n-1$) deviendront stochastiquement parfaites – nous dirons alors que ces grandeurs forment une chaîne réciproque. (Dans une chaîne ordinaire les grandeurs x_i deviennent stochastiquement parfaites dès que x_0 est donnée, de sorte que la loi de x_n ne peut plus alors être choisie arbitrairement.)

Supposons, pour traiter un cas précis assez général, qu'on a nécessairement $x_{i-h} \leq x_i \leq x_{i+k}$, et que la fonction f est de la forme

$$f = \frac{C_{h,k}}{x_{i+k} - x_{i-h}} f_{h,k} \left(\frac{x_i - x_{i-h}}{x_{i+k} - x_{i-h}} \right),$$

où $C_{h,k}$ est une constante qui se déterminera par la condition de normalité

S. Bernstein: Sur les liaisons entre les grandeurs aléatoires

$$C_{h,k} \int_0^1 f_{h,k}(z) dz = 1.$$

On démontre alors que la loi des probabilités doit être de la forme

$$f_{h,k}(z) = z^{\lambda h-1} (1-z)^{\lambda k-1},$$

où λ est un paramètre positif arbitraire, et se réduit ainsi à une courbe de M. Pearson. On a évidemment

$$E x_i = \frac{k x_{i-h} + h x_{i+k}}{k+h}$$

et la dispersion

$$\sigma^2 x_i = \frac{hk}{\lambda(h+k)+1} \left(\frac{x_{i+k} - x_{i-h}}{h+k} \right)^2.$$

Si λ est fixe et n très grand, les courbes de Pearson tendront vers des courbes normales avec des dispersions tendant vers 0, et pour $n = \infty$, les points x_i se disposeraient rigoureusement en ligne droites. Si λ tend vers 0 de telle sorte, que $\lambda n = a$ reste constant, la chaîne réciproque limite définira une variable aléatoire monotone quasi-continue, pour laquelle $f_{h,k}(z)$ se mettra sous la forme

$$f_{h,k}(z) = z^{a(t-t_0)-1} (1-z)^{a(t_1-t)-1},$$

où on pose $\frac{i}{n} = t, \frac{i-h}{n} = t_0, \frac{i+k}{n} = t_1$; ce qui nous donne un schéma général du mécanisme stochastique de la formation et de la transformation continue des courbes de M. Pearson.

On doit à M. Schrödinger un procédé important de formation de chaînes réciproques qu'il a proposé dans son Mémoire „Über die Umkehrung der Naturgesetze“.

Admettons que deux grandeurs x_0 et x_1 soient en corrélation imparfaite, telle que l'on sache seulement que la densité des probabilités de x_1 est $f(x_0, x_1)$, lorsque x_0 est donné. Pour fixer les idées, posons avec M. Schrödinger

$$f(x_0, x_1) = \frac{1}{\sqrt{4\pi D(t_1-t_0)}} e^{-\frac{(x_1-x_0)^2}{4D(t_1-t_0)}}, \tag{8}$$

en rappelant seulement, comme nous l'avons vu, qu'il est, en tout cas, impossible que la probabilité de x_0 , pour x_1 donné, soit représentée par la même fonction. La surface de corrélation la plus générale, correspondant à cette loi de passage sera

$$F(x_0, x_1) = P(x_0) f(x_0, x_1),$$

où $P(x_0)$ est la distribution des probabilités de x_0 qui peut être donnée arbitrairement; la loi de distribution de x_1 sera alors déterminée par la formule

$$P_1(x_1) = \int_{-\infty}^{\infty} P(x_0) f(x_0, x_1) dx_0 \tag{9}$$

Grosse Vorträge

Ainsi sans contradiction avec la loi de passage $f(x_0, x_1)$ on ne pourra pas, en effectuant un grand nombre d'observations, obtenir des fréquences $w_0(x_0)$ et $w_1(x_1)$ sensiblement éloignées de la relation (9). Mais rien n'empêche de collectionner avec assez de patience des paires de valeurs (x_0, x_1) très improbables et de tirer de la distribution primitive un tableau de corrélation arbitraire entièrement nouveau. Cependant, si ce n'est pas la coïncidence de (x_0, x_1) dans une même paire ou chez un même individu qui nous intéresse, mais seulement les fréquences $w_0(x_0)$ et $w_1(x_1)$ de x_0 et x_1 considérées isolément, une telle sélection sera relativement bien plus facile à réaliser, et on a un problème mathématique parfaitement déterminé, si l'on demande quel est le tableau de corrélation, correspondant aux fréquences choisies $w_0(x_0), w_1(x_1)$, le plus probable que l'on formera ainsi. En supposant que le nombre de paires collectionnées soit très grand, on trouve, d'après M. Schrödinger, que la nouvelle surface de corrélation aura pour limite

$$F(x_0, x_1) = f(x_0, x_1) \psi(x_0) \varphi(x_1)$$

où l'on a, pour déterminer $\psi(x_0)$ et $\varphi(x_1)$, les équations intégrales

$$\begin{aligned} w_0(x_0) &= \psi(x_0) \int_{-\infty}^{\infty} f(x_0, x_1) \varphi(x_1) dx_1 \\ w_1(x_1) &= \varphi(x_1) \int_{-\infty}^{\infty} f(x_0, x_1) \psi(x_0) dx_0. \end{aligned} \tag{10}$$

On peut démontrer que dans le cas, où $f(x_0, x_1)$ a la forme (8) considérée par M. Schrödinger, le système (10) admet une solution, quelles que soient les fonctions continues $w_0(x)$ et $w_1(x)$.

Cela étant, si x_0, x, x_1 formaient primitivement une chaîne simple déterminée par les lois de passage $f_0(x_0, x), f_1(x, x_1)$, la loi des probabilités de x deviendra

$$w(x) = \int_{-\infty}^{\infty} f_0(x_0, x) \psi(x_0) dx_0 \int_{-\infty}^{\infty} f_1(x, x_1) \varphi(x_1) dx_1 = \Psi(x) \cdot \Phi(x), \tag{11}$$

et, puisqu'elle ne sera déterminée que, si $w_0(x_0)$ et $w_1(x_1)$ sont données toutes les deux, la chaîne primitive se trouvera transformée en une chaîne réciproque.

$$\begin{aligned} \text{Dans le cas considéré par M. Schrödinger } f_0(x_0, x) &= \frac{e^{-\frac{(x-x_0)^2}{4\pi D(t-t_0)}}}{\sqrt{4\pi D(t-t_0)}}, \\ f_1(x, x_1) &= \frac{e^{-\frac{(x_1-x)^2}{4\pi D(t_1-t)}}}{\sqrt{4\pi D(t_1-t)}}, \end{aligned}$$

de sorte que $\Psi(x)$ et $\Phi(x)$ satisfont respectivement aux équations conjuguées

S. Bernstein: Sur les liaisons entre les grandeurs aléatoires

$$D \frac{\partial^2 \Psi}{\partial x^2} = \frac{\partial \Psi}{\partial t}, \quad D \frac{\partial^2 \Phi}{\partial x^2} = -\frac{\partial \Phi}{\partial t},$$

et se réduisent respectivement à $\psi(x_0)$ pour $t = t_0$ et à $\varphi(x_1)$ pour $t = t_1$.

Si l'on assujetti les distributions initiales et finales $w_0(x_0)$ et $w_1(x_1)$ à certaines conditions supplémentaires qu'il faut introduire pour que les deux intégrales $\int_{-\infty}^{\infty} \psi(x_0) dx_0$ et $\int_{-\infty}^{\infty} \varphi(x_1) dx_1$, aient un sens [ainsi, par exemple, en supposant $w_0(x_0) = \frac{e^{-\frac{x_0^2}{4\hbar^2}}}{2\hbar\sqrt{\pi}}$, $w_1(x_1) = \frac{e^{-\frac{x_1^2}{4\hbar_1^2}}}{2\hbar_1\sqrt{\pi}}$, il faudrait que $|\hbar_1^2 - \hbar^2| < D(t_1 - t_0)$], la chaîne réciproque

de M. Schrödinger est susceptible de l'interprétation simple suivante. Considérons deux variables aléatoires indépendantes x' et x'' que nous appellerons conjuguées, formant deux chaînes ordinaires de sens inverse, déterminées, respectivement, par les lois de passage $f_0(x_0, x)$, $f_1(x_1, x)$, et admettons que la grandeur x que nous observons ne soit définie que lorsqu'on a simultanément $x' = a$, $x'' = a$ et qu'elle reçoive alors la même valeur a . Dans ces conditions, il résulte immédiatement du théorème de la multiplication des probabilités, que $w_0(x_0)$ et $w_1(x_1)$ seront représentées à un facteur constant près par les intégrales (10), $\psi(x_0')$ et $\varphi(x_1'')$ étant les probabilités initiales des variables conjuguées x' et x'' , et la probabilité cherchée $w(x)$ (au même facteur près) sera donnée par la formule (11).

Remarquons que les chaînes réciproques, quel que soit leur mode de formation, étant stochastiquement déterminées par les états initial et final, sont naturellement incapables de fournir des schémas stochastiques parfaits des états antérieurs à l'état initial et postérieurs à l'état final; ainsi, elles sont assimilables aux trajectoires dynamiques considérées au point de vue des principes variationnels, mais tandis qu'une trajectoire mécanique ainsi définie, peut, en général, être prolongée d'une façon déterminée dans les deux sens, les chaînes réciproques ne jouissent pas de cette propriété: pour leur validité universelle il serait donc nécessaire que le temps soit fini. Quoiqu'il en soit, la forme, sous laquelle nous avons présenté ici les chaînes réciproques qui est sensiblement différente de la conception de M. Schrödinger, est logiquement indépendante du temps, et par conséquent, peut également être appliquée, en particulier, à des problèmes statiques de nature stochastique analogue.

Je suis loin d'avoir épuisé l'étude générale des liaisons entre les quantités aléatoires; mais vous voyez par ce qui précède l'immense variété des moyens dont dispose la théorie des probabilités pour l'interprétation mathématique des phénomènes de la nature, et vous m'excuserez, je l'espère, si je n'ai pas réussi à vous exposer d'une façon suffisamment complète toutes les principales idées modernes relatives à ce domaine si vaste.