9. ON ANALYTICAL METHODS IN PROBABILITY THEORY *

The object of investigation

A physical process (a change of a certain physical system) is called *stochastically determined* if, knowing a state X_0 of the system at a certain moment of time t_0 we also know the probability distribution for all the states X of this system at the moments $t > t_0$.

I systematically consider the simplest cases of stochastically determined processes, and primarily, processes continuous in time (this is what makes the method essentially new: so far, a stochastic process has usually been considered to be a discrete sequence of separate "events").

If the set $\mathfrak{A}$ of different possible states of the system is finite, then a stochastic process can be characterized using ordinary linear differential equations (Chapter II). If a state of the system depends on one or several continuous parameters, then the corresponding analytic apparatus reduces to partial differential equations of parabolic type (Chapter IV) and we obtain various distribution functions, the normal Laplace distribution being the simplest.

INTRODUCTION

1. In order to subject social or natural phenomena to mathematical treatment, these phenomena should first be schematized. The fact is that mathematical analysis can only be applied to studying changes of a certain system if every possible state of this system can be completely determined using known mathematical techniques, for example, by the values of a certain number of parameters. This mathematically defined system is not a reality itself, but a scheme that can be used to describe reality.

Classical mechanics makes use only of the schemes for which the state y of a system at time t is uniquely determined by its state x at any preceding time t_0. Mathematically this can be expressed by the formula

$$y = f(x, t_0, t).$$

If such a unique function f exists, as is always assumed in classical mechanics, then we say that our scheme is a *scheme of a purely deterministic*

---

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

process. These purely deterministic processes also include processes when the state y is not completely determined by giving a state x at a single moment of time t, but also essentially depends on the pattern of variation of this state x prior to t. However, usually it is preferred to avoid such a dependence on the preceding behaviour of the system, and to do this the notion of the state of the system at time t is generalized by introducing new parameters.[1]

Outside the realm of classical mechanics, along with the schemes of purely deterministic processes, one often considers schemes in which the state x of the system at a certain time t_0 only determines a certain probability of a possible event y to occur at a certain subsequent moment $t > t_0$. If for any given t_0, $t > t_0$, and x there exists a certain probability distribution for the states y, we say that our scheme is a *scheme of a stochastically determined process*. In the general case this distribution function can be represented in the form

$$P(t_0, x, t, \mathfrak{E})$$

where $\mathfrak{E}$ denotes a certain set of states y, and P is the probability of the fact that at time t one of the states y belonging to this set will be realized. Here we face a complication: in general, this probability cannot be determined for all sets $\mathfrak{E}$. A rigorous definition of a stochastically determined process which enables one to avoid this complication is given in §1.

As in the case of a purely deterministic process, we could also have considered here schemes in which the probability P essentially depends not only on the state x but also on the past behaviour of the system. Still, this influence of the past behaviour of the process can be bypassed using the same method as in the scheme of a purely deterministic process.

Note also that the possibility of applying a scheme of either a purely deterministic or a stochastically determined process to the study of some real processes is in no way linked with the question whether this process is deterministic or random.

2. In probability theory one usually considers only schemes according to which any changes of the states of a system are only possible at certain moments

[1] A well-known example of this method is to introduce, in addition to positions of points, the components of their velocities when describing a state of a certain mechanical system.

$t_1, t_2, \ldots, t_n, \ldots$ which form a discrete series. As far as I know, Bachelier[2] was the first to make a systematic study of schemes in which the probability $P(t_0, x, t, \mathfrak{E})$ varies continuously with time t. We will return to the cases studied by Bachelier in §16 and in the Conclusion. Here we note only that Bachelier's constructions are by no means mathematically rigorous.

Starting from Chapter II of this paper we mainly consider above-mentioned schemes that are continuous with respect to time. From the mathematical point of view these schemes have an important advantage: they allow one to introduce differential equations for P with respect to time and lead to simple analytic expressions which in the usual theory can be derived only as asymptotic formulas. As for the applications, first the new schemes can be directly applied to real processes, and secondly, from the solutions of differential equations for processes continuous with respect to time new asymptotic formulas for continuous schemes can be derived, as will be shown later in §12.

3. We do not start with the complete system of axioms of probability theory. Let us indicate, however, all the prerequisites we will use in our further discussion. We do not make any special assumptions about the set $\mathfrak{A}$ of possible states x. Mathematically, $\mathfrak{A}$ can be considered as an arbitrary set consisting of arbitrary elements. All assumptions concerning the system $\mathfrak{F}$ of sets and the function $P(t_0, x, t, \mathfrak{E})$ are given in §1. In what follows the theory is developed as a purely mathematical one.

CONTENTS

[2] I. 'Théorie de la spéculation', *Ann. École Norm. Supér.* **17** (1900), 21; II. 'Les probabilités à plusieurs variables', *Ann. École Norm. Supér.* **27** (1910), 339; III. *Calcul des probabilités*, Paris, 1912.

CHAPTER I. GENERALITIES

§1. General scheme of a stochastically determined process

Let $\mathfrak{S}$ be a system that can be in states $x, y, z, \ldots$, and $\mathfrak{F}$ a system of sets $\mathfrak{E}$ formed from the elements $x, y, z, \ldots$. A process of variation of the system $\mathfrak{S}$ is *stochastically determined with respect to* $\mathfrak{F}$ if for any choice of state x, set $\mathfrak{E}$ and moments t_1 and t_2 ($t_1 < t_2$) the probability $P(t_1, x, t_2, \mathfrak{E})$ of the fact that, if x takes place at t_1, then one of the states of $\mathfrak{E}$ takes place at t_2 exists. If $P(t_1, x, t_2, \mathfrak{E})$ is defined only for $t_2 > t_1 \geq t_0$, then we say that the process of variation is stochastically determined for $t \geq t_0$.

Regarding the system $\mathfrak{F}$, we assume that it is first additive (that is, it contains all the differences, as well as finite or countable sums of its elements), and secondly contains the empty set, the set $\mathfrak{A}$ of all possible states $x, y, z, \ldots$ and all the one-element sets. If the set $\mathfrak{A}$ is finite or countable, then clearly $\mathfrak{F}$ consists of all the subsets of $\mathfrak{A}$. In the most important case when $\mathfrak{A}$ is uncountable, the assumption that $\mathfrak{F}$ contains all the subsets of $\mathfrak{A}$ does not hold for any of the schemes known at present.

Of course we assume that

$$P(t_1, x, t_2, \mathfrak{A}) = 1 \tag{1}$$

and for the empty set $\mathfrak{N}$,

$$P(t_1, x, t_2, \mathfrak{N}) = 0.$$

We further assume that $P(t_1, x, t_2, \mathfrak{E})$ is additive as a function of $\mathfrak{E}$, that is, for any decomposition of $\mathfrak{E}$ into a finite or countable number of non-intersecting summands $\mathfrak{E}_n$ the following identity holds

$$\sum_n P(t_1, x, t_2, \mathfrak{E}_n) = P(t_1, x, t_2, \mathfrak{E}). \tag{2}$$

To formulate further assumptions on $P(t_1, x, t_2, \mathfrak{E})$ we need the notion of measurability of a function $f(x)$ with respect to the system $\mathfrak{F}$ and the definition of abstract Stieltjes integral. We given them here in a form suitable for our needs.[3]

A function $f(x)$ is called *measurable with respect to the system* $\mathfrak{F}$ if for any choice of real numbers a and b the set $\mathfrak{E}_\Delta(a < f(x) < b)$ of all x for which $f(x)$ satisfies the inequality in parentheses, belongs to $\mathfrak{F}$. It can easily be shown that if the system $\mathfrak{F}$ is additive and $f(x)$ is measurable with respect to $\mathfrak{F}$, then the set $\mathfrak{E}$ of all x for which $f(x)$ belongs to a given Borel-measurable set is contained in $\mathfrak{F}$.

Now let $f(x)$ be measurable with respect to $\mathfrak{F}$ and bounded, and let $\phi(\mathfrak{E})$ denote a non-negative additive function defined on $\mathfrak{F}$; then, as is known, the sum

$$\sum_m \frac{m}{n}\phi\Big(\frac{m}{n} \le f(x) < \frac{m+1}{n}\Big)$$

tends to a well defined limit as $n \to \infty$. This limit will be called the *integral*

$$\int_{\mathfrak{A}_x} f(x)\phi(d\mathfrak{A}).$$

[3] Concerning these notions, as well as additive sets of systems, etc., see, for example, M. Fréchet, 'Sur l'intégrale d'une fonctionnelle étendu à un ensemble abstrait', *Bull. Soc. Math. France* **43** (1915), 248.

This notation differs from the usual one only in the specification of the variable of integration and the place of the differential inside the parentheses.

In what follows we assume that $P(t_1, x, t_2, \mathfrak{E})$, as a function of the state x, is measurable with respect to the system $\mathfrak{F}$. Finally, $P(t_1, x, t_2, \mathfrak{E})$ must satisfy the fundamental equation

$$P(t_1, x, t_2, \mathfrak{E}) = \int_{\mathfrak{A}_y} P(t_2, y, t_3, \mathfrak{E}) P(t_1, x, t_2, d\mathfrak{A}) \tag{3}$$

for arbitrary t_1, t_2, t_3, $(t_1 < t_2 < t_3)$. If $\mathfrak{A}$ is a finite or countable set of elements $x_1, x_2, \ldots, x_n, \ldots$, then

$$\int_{\mathfrak{A}_y} P(t_2, y, t_3, \mathfrak{E}) P(t_1, x, t_2, d\mathfrak{A}) = \sum_n P(t_2, x_n, t_3, \mathfrak{E}) P(t_1, x, t_2, x_n)$$

and on the right-hand side we have the expression for the total probability $P(t_1, x, t_3, \mathfrak{E})$; therefore (3) is satisfied in this case. In case $\mathfrak{A}$ is uncountable we take (3) as a new axiom.

The above requirements completely define a stochastically determined process: the elements $x, y, z, \ldots$ of an arbitrary set $\mathfrak{A}$ can be considered as characteristics of a state of a certain system, and an arbitrary function $P(t_1, x, t_2, \mathfrak{E})$ satisfying the above requirements as the corresponding probability distribution.

A non-negative function $F(\mathfrak{E})$ defined on $\mathfrak{F}$, additive and such that

$$F(\mathfrak{A}) = 1, \tag{4}$$

will be called a *normal distribution function*. All the requirements imposed on $P(t_1, x, t_2, \mathfrak{E})$ can now be formulated in the following way: $P(t_1, x, t_2, \mathfrak{E})$, as a function of $\mathfrak{E}$, is a normal distribution function; as a function of x it is measurable with respect to the system $\mathfrak{F}$ and satisfies the integral equation (3).

Suppose now that at $t = t_0$ we have a normal distribution function $Q(t_0, \mathfrak{E})$ which gives the probability of the fact that the system $\mathfrak{S}$ at t_0 is in one of the states belonging to $\mathfrak{E}$. The distribution function $Q(t, \mathfrak{E})$ for $t > t_0$ is determined by means of the second fundamental equation

$$Q(t, \mathfrak{E}) = \int_{\mathfrak{A}_x} P(t_0, x, t, \mathfrak{E}) Q(t_0, d\mathfrak{A}). \tag{5}$$

We clearly have

$$Q(t, \mathfrak{A}) = \int_{\mathfrak{A}} Q(t_0, d\mathfrak{A}) = Q(t_0, \mathfrak{A}) = 1, \tag{6}$$

$$\int_{\mathfrak{A}_x} P(t_1, x, t_2, \mathfrak{E}) Q(t_1, d\mathfrak{A}) =$$
$$= \int_{\mathfrak{A}_x} P(t_1, x, t_2, \mathfrak{E}) \int_{\mathfrak{A}'_x} P(t_0, y, t_1, d\mathfrak{A}) Q(t_0, d\mathfrak{A}') =$$
$$= \int_{\mathfrak{A}'_y} \int_{\mathfrak{A}_x} P(t_1, x, t_2, \mathfrak{E}) P(t_0, y, t_1, d\mathfrak{A})) Q(t_0, d\mathfrak{A}') =$$
$$= \int_{\mathfrak{A}'_y} P(t_0, y, t_2, \mathfrak{E}) Q(t_0, d\mathfrak{A}') = Q(t_2, \mathfrak{E}). \tag{7}$$

Formula (5) is considered as the definition of $Q(t, \mathfrak{E})$, not as a new requirement imposed on $\mathfrak{S}$. Note, however, that (5) implies (3) as a particular case.

§2. The operator $F_1(x, \mathfrak{E}) * F_2(x, \mathfrak{E})$

Let $F_1(x, \mathfrak{E})$ and $F_2(x, \mathfrak{E})$ be two normal distribution functions which, considered as functions of x, are measurable with respect to $\mathfrak{F}$. Set

$$F(x, \mathfrak{E}) = F_1(x, \mathfrak{E}) * F_2(x, \mathfrak{E}) = F_1 * F_2(x, \mathfrak{E}) = \int_{\mathfrak{A}_y} F_2(y, \mathfrak{E}) F_1(x, d\mathfrak{A}); \tag{8}$$

It is easy to see that $F(x, \mathfrak{E})$ satisfies the same conditions of measurability and additivity as $F_1(x, \mathfrak{E})$ and $F_2(x, \mathfrak{E})$ and (4) also holds:

$$F(x, \mathfrak{A}) = \int_{\mathfrak{A}'_y} F_2(y, \mathfrak{A}) F_1(x, d\mathfrak{A}') = \int_{\mathfrak{A}'_y} F_1(x, d\mathfrak{A}') = 1;$$

consequently, $F(x, \mathfrak{E})$ is also a normal distribution function.

Further, the operator $F_1 * F_2$ is associative,

$$F_1 * (F_2 * F_3) = (F_1 * F_2) * F_3, \tag{9}$$

which can easily be seen by the following simple calculation:

$$F_1 * (F_2 * F_3)(x, \mathfrak{E}) = \int_{\mathfrak{A}_y} \int_{\mathfrak{A}'_z} F_3(z, \mathfrak{E}) F_2(y, d\mathfrak{A}') F_1(x, d\mathfrak{A}) =$$
$$= \int_{\mathfrak{A}'_z} F_3(z, \mathfrak{E}) \int_{\mathfrak{A}_y} F_2(y, d\mathfrak{A}') F_1(x, d\mathfrak{A}) = (F_1 * F_2) * F_3(x, \mathfrak{E}).$$

By contrast, $F_1 * F_2$ is not, in general, commutative.

Now we define the unit function $\mu(x, \mathfrak{C})$, which for any normal distribution function $F(x, \mathfrak{C})$ satisfies

$$\mu * F(x, \mathfrak{C}) = F * \mu(x, \mathfrak{C}) = F(x, \mathfrak{C}). \tag{10}$$

To this end it suffices to set $\mu(x, \mathfrak{C}) = 1$ when x belongs to $\mathfrak{C}$ and $\mu(x, \mathfrak{C}) = 0$ otherwise. We then have

$$\mu * F(x, \mathfrak{C}) = \int_{\mathfrak{A}_y} F(y, \mathfrak{C})\mu(x, d\mathfrak{A}) = F(x, \mathfrak{C}),$$

$$F * \mu(x, \mathfrak{C}) = \int_{\mathfrak{A}_y} \mu(y, \mathfrak{C})F(x, d\mathfrak{A}) = \int_{\mathfrak{C}} F(x, d\mathfrak{C}) = F(x, \mathfrak{C}).$$

The probability $P(t_1, x, t_2, \mathfrak{C})$ has been defined so far only for $t_2 > t_1$; now set for any t

$$P(t, x, t, \mathfrak{C}) = \mu(x, \mathfrak{C}). \tag{11}$$

In view of (10), this new definition does not contradict the fundamental equation (3), since (3) can be written as

$$P(t_1, x, t_2, \mathfrak{C}) * P(t_2, x, t_3, \mathfrak{C}) = P(t_1, x, t_3, \mathfrak{C}). \tag{12}$$

§3. Classification of particular cases

If the changes in the state of the system $\mathfrak{S}$ take place only at certain moments which form a discrete series

$$t_0 < t_1 < t_2 < \ldots < t_n < \ldots \to \infty,$$

then obviously

$$P(t', x, t'', \mathfrak{C}) = P(t_m, x, t_n, \mathfrak{C}) \tag{13}$$

for all moments t' and t'' such that

$$t_m \le t' < t_{m+1}, \quad t_n \le t'' < t_{n+1}.$$

Introducing the notation

$$P(t_m, x, t_n, \mathfrak{C}) = P_{mn}(x, \mathfrak{C}), \tag{14}$$

$$P_{n-1,n}(x, \mathfrak{C}) = P_n(x, \mathfrak{C}), \tag{15}$$

we have

$$P_{mn}(x, \mathfrak{E}) = P_{m+1} * P_{m+2} * \ldots * P_n(x, \mathfrak{E}). \tag{16}$$

Hence in this case the process of change of $\mathfrak{S}$ is totally determined by the elementary distribution functions $P_n(x, \mathfrak{E})$.

Now let $P_1(x, \mathfrak{E})$, $P_2(x, \mathfrak{E})$, $\ldots$, $P_n(x, \mathfrak{E})$ be arbitrary normal distribution functions which are assumed to be measurable as functions of x; further, let $t_0 < t_1 < \ldots < t_n < \ldots$ be a certain sequence of moments of time. Defining $P_{mn}(x, \mathfrak{E})$ and $P(t', x, t'', \mathfrak{E})$ by (16), (14) and (13), we also obtain normal distribution functions which satisfy the equations

$$P_{mn}(x, \mathfrak{E}) * P_{np}(x, \mathfrak{E}) = P_{mp}(x, \mathfrak{E}) \quad (m < n < p), \tag{17}$$

and hence the equation

$$P(t', x, t'', \mathfrak{E}) * P(t'', x, t''', \mathfrak{E}) = P(t', x, t''', \mathfrak{E}) \quad (t' < t'' < t''').$$

But this latter equation is none other than the fundamental equation (12) or (13). Thus we see that every sequence of arbitrary normal distribution functions $P_n(x, \mathfrak{E})$, measurable as functions of x, characterizes a certain stochastically determined process.

The schemes with discrete time defined above are those usually considered in probability theory. If all the distribution functions $P_n(x, \mathfrak{E})$ coincide,

$$P_n(x, \mathfrak{E}) = P(x, \mathfrak{E}), \tag{18}$$

we have a homogeneous scheme with discrete time; in this case (16) yields

$$P_{n,n+p}(x, \mathfrak{E}) = \underbrace{P(x, \mathfrak{E}) * P(x, \mathfrak{E}) * \ldots * P(x, \mathfrak{E})}_{p \text{ times}} = [P(x, \mathfrak{E})]^p_* = P^p(x, \mathfrak{E}). \tag{19}$$

As far back as 1900 Bachelier considered stochastic processes continuous in time.[4] There are good grounds for giving schemes with continuous time a central place in probability theory. It seems that most important here are schemes homogeneous in time, in which $P(t, x, t + \tau, \mathfrak{E})$ depends only on the difference $t_2 - t_1$:

$$P(t, x, t + \tau, \mathfrak{E}) * P(\tau_2, x, \mathfrak{E}) = P(\tau, x, \mathfrak{E}). \tag{20}$$

[4] See the first of the papers cited in footnote 2.

The fundamental equation in this case takes the form

$$P(\tau_1, x, \mathfrak{C}) * P(\tau_2, x, \mathfrak{C}) = P(\tau_1 + \tau_2, x, \mathfrak{C}). \tag{21}$$

Another series of particular cases is obtained under special requirements on the set $\mathfrak{A}$ of elementary states x. Here one should distinguish the cases of finite or countable sets $\mathfrak{A}$; in the continuous case the classification is performed with respect to the number of parameters determining the state of the system. The subsequent subdivision of the material in this paper is based on distinguishing such particular cases.

§4. The ergodic principle

Without special assumptions on the set $\mathfrak{A}$ of all possible states x, we can only prove several general theorems, namely those dealing with the ergodic principle. We say that a stochastic process obeys the *ergodic principle* if for any $t^{(0)}, x, y$ and $\mathfrak{C}$

$$\lim_{t\to\infty} [P(t^{(0)}, x, t, \mathfrak{C}) - P(t^{(0)}, y, t, \mathfrak{C})] = 0. \tag{22a}$$

For a scheme with discrete time (22a) is clearly equivalent to the following:

$$\lim_{n\to\infty} [P_{mn}(x, \mathfrak{C}) - P_{mn}(y, \mathfrak{C})] = 0; \tag{22b}$$

and in the latter case the following theorem holds:

Theorem 1. *If for any x, y, and $\mathfrak{C}$*

$$P_n(x, \mathfrak{C}) \geq \lambda_n P_n(y, \mathfrak{C}), \ \lambda_n \geq 0, \tag{23}$$

and the series

$$\sum_{n=1}^{\infty} \lambda_n \tag{24}$$

diverges, then the ergodic principle (22b) *holds and the limit in* (22b) *is uniform with respect to x, y and $\mathfrak{C}$.*

Proof. Let

$$\sup_x P_{kn}(x, \mathfrak{C}) = M_{kn}(\mathfrak{C}), \quad \inf_x P_{kn}(x, \mathfrak{C}) = m_{kn}(\mathfrak{C}).$$

For $i < k$ we clearly have

$$P_{in}(x, \mathfrak{C}) = \int_{\mathfrak{A}_y} P_{kn}(y, \mathfrak{C}) P_{ik}(x, d\mathfrak{A}) \leq M_{kn}(\mathfrak{C}) \int_{\mathfrak{A}_y} P_{ik}(x, d\mathfrak{A}) = M_{kn}(\mathfrak{C}) \tag{25}$$

and, similarly,

$$P_{in}(x, \mathfrak{C}) \geq m_{kn}(\mathfrak{C}). \tag{26}$$

By (23), for any x and y we have

$$\begin{gathered}
P_k(x, \mathfrak{C}) - \lambda_k P_k(y, \mathfrak{C}) \geq 0, \\
P_{k-1,n}(x, \mathfrak{C}) = \int_{\mathfrak{A}_z} P_{kn}(z, \mathfrak{C}) P_k(x, d\mathfrak{A}) = \\
= \int_{\mathfrak{A}_z} P_{kn}(z, \mathfrak{C})[P_k(x, d\mathfrak{A}) - \lambda_k P_k(y, d\mathfrak{A})] + \\
+\lambda_k \int_{\mathfrak{A}_z} P_{kn}(z, \mathfrak{C}) P_k(y, d\mathfrak{A}) \geq \\
\geq m_{kn}(\mathfrak{C}) \int_{\mathfrak{A}_z} [P_k(x, d\mathfrak{A}) - \lambda_k P_k(y, d\mathfrak{A})] + \lambda_k P_{k-1,n}(y, \mathfrak{C}) = \\
= m_{kn}(\mathfrak{C})(1 - \lambda_k) + \lambda_k P_{k-1,n}(y, \mathfrak{C}), \\
P_{k-1,n}(y, \mathfrak{C}) - P_{k-1,n}(x, \mathfrak{C}) \leq (1 - \lambda_k)[P_{k-1,n}(y, \mathfrak{C}) - m_{kn}(\mathfrak{C})];
\end{gathered}$$

hence by (25),

$$P_{k-1,n}(y, \mathfrak{C}) - P_{k-1,n}(x, \mathfrak{C}) \leq (1 - \lambda_k)[M_{kn}(\mathfrak{C}) - m_{kn}(\mathfrak{C})]. \tag{27}$$

Since (27) holds for any x and y, we also have

$$M_{k-1,n}(\mathfrak{C}) - m_{k-1,n}(\mathfrak{C}) \leq (1 - \lambda_k)[M_{kn}(\mathfrak{C}) - m_{kn}(\mathfrak{C})]. \tag{28}$$

Setting $k = m+1, m+2, \ldots, n$ successively in (28) and multiplying all the resulting equalities we find

$$M_{mn}(\mathfrak{C}) - m_{mn}(\mathfrak{C}) \leq \prod_{k=m+1}^{n} (1 - \lambda_k). \tag{29}$$

The right-hand side of (29) tends to zero as $n \to \infty$; this proves the theorem.

For a homogeneous scheme with discontinuous time the following holds:

Theorem 2. *If for any x, y and $\mathfrak{C}$*

$$P(x, \mathfrak{C}) \geq \lambda P(y, \mathfrak{C}) \quad (\lambda > 0), \tag{30}$$

then $P^n(x, \mathfrak{C})$ converges uniformly to a certain distribution function $Q(\mathfrak{C})$.

Proof. We have

$$M_{n,n+p}(\mathfrak{E}) = \sup P^p(x, \mathfrak{E}) = M_p(\mathfrak{E}),$$
$$m_{n,n+p}(\mathfrak{E}) = \inf P^p(x, \mathfrak{E}) = m_p(\mathfrak{E}),$$
$$\lambda_n = \lambda,$$

and, by (29),

$$M_p(\mathfrak{E}) - m_p(\mathfrak{E}) \leq (1-\lambda)^p. \tag{31}$$

But (25) and (26) imply that for $q > p$,

$$p^q(x, \mathfrak{E}) = P_{0q}(x, \mathfrak{E}) \leq M_{q-p,q}(\mathfrak{E}) = M_p(\mathfrak{E}), \tag{32}$$
$$P^q(x, \mathfrak{E}) \geq m_p(\mathfrak{E}); \tag{33}$$

therefore,

$$M_p(\mathfrak{E}) \geq M_q(\mathfrak{E}) \geq m_q(\mathfrak{E}) \geq m_p(\mathfrak{E}). \tag{34}$$

Our theorem now follows immediately from (31) and (34).

Important particular cases of Theorem 2 were proved by Gostinskii and Hadamard.[5] As has been shown by Hadamard, in these particular cases $Q(\mathfrak{E})$ satisfies the integral equation

$$Q(\mathfrak{E}) = \int_{\mathfrak{A}_x} P(x, \mathfrak{E}) Q(d\mathfrak{A}). \tag{35}$$

For the most general stochastically determined scheme one has:

Theorem 3. *If for some sequence*

$$t_0 < t_1 < \ldots < t_n < \ldots \to \infty$$

and any x, y and $\mathfrak{E}$,

$$P(t_{n-1}, x, t_n, \mathfrak{E}) \geq \lambda_n P(t_{n-1}, y, t_n, \mathfrak{E}), \quad \lambda_n \geq 0, \tag{36}$$

and if the series $\sum_{n=1}^{\infty} \lambda_n$ diverges, then the ergodic principle (22a) *holds and the convergence in* (22a) *is uniform with respect to $x, y, \mathfrak{E}$.*

Proof. For a given $t^{(0)}$ let

$$\sup_x P(t^{(0)}, x, t, \mathfrak{E}) = M(t, \mathfrak{E}),$$
$$\inf_x P(t^{(0)}, x, t, \mathfrak{E}) = m(t, \mathfrak{E}).$$

---

[5] *C. R. Acad. Sci. Paris* **186** (1928), 59; **189**; 275.

If

$$t^{(0)} \leq t_m \leq t_n \leq t \leq t_{n+1},$$

then, in the same way as in the proof of Theorem 1, we obtain the analogous formula to (29),

$$M(t, \mathfrak{C}) - m(t, \mathfrak{C}) \leq \prod_{k=m+1}^{n} (1 - \lambda_k).$$

Since $n \to \infty$ together with t, it follows that $M(t, \mathfrak{C}) - m(t, \mathfrak{C}) \to 0$ as $t \to \infty$, which proves the theorem.

Finally in case of a scheme homogeneous in time one has the following theorem, analogous to Theorem 2:

Theorem 4. *If there exists σ such that for any $x, y, \mathfrak{C}$,*

$$P(\sigma, x, \mathfrak{C}) \geq \lambda P(\sigma, y, \mathfrak{C}) \quad (\lambda > 0), \tag{37}$$

then $P(\tau, x, \mathfrak{C})$ converges uniformly to a certain distribution function $Q(\mathfrak{C})$ as $\tau \to \infty$.

CHAPTER II. FINITE STATE SYSTEMS

§5. Preliminary remarks

Let us now assume that $\mathfrak{A}$ is formed from a finite number of elements

$$x_1, x_2, \ldots, x_n.$$

In this case set

$$P(t_1, x_i, t_2, x_j) = P_{ij}(t_1, t_2). \tag{38}$$

Since for any set $\mathfrak{C}$ we obviously have

$$P(t_1, x_i, t_2, \mathfrak{C}) = \sum_{x_k \subset \mathfrak{C}} P_{ik}(t_1, t_2), \tag{39}$$

we can confine ourselves to the probabilities $P_{ij}(t_1, t_2)$. The fundamental equation (3) now takes the form

$$\sum_j P_{ij}(t_1, t_2) P_{jk}(t_2, t_3) = P_{ik}(t_1, t_3), \tag{40}$$

whereas (1) can be written as

$$\sum_j P_{ij}(t_1, t_2) = 1. \tag{41}$$

Any non-negative functions $P_{ij}(t_1, t_2)$ satisfying the conditions (40) and (41) determine some stochastically determined process of variation of the systems $\mathfrak{S}$.

In this case the operator is defined as follows:

$$F_{ik} = F_{ik}^{(1)} * F_{ik}^{(2)} = \sum_j F_{ij}^{(1)} F_{jk}^{(2)}, \tag{42}$$

hence the fundamental equation (40) reduces to

$$P_{ik}(t_1, t_2) * P_{ik}(t_2, t_3) = P_{ik}(t_1, t_3). \tag{43}$$

For a scheme with discontinuous time we set

$$P_{pq}(x_i, x_j) = P_{ij}^{(pq)}, \quad P_p(x_i, x_j) = P_{ij}^{(p)}.$$

Then the probabilities $P_{ij}^{(p)}$ satisfy

$$\sum_j P_{ij}^{(p)} = 1, \tag{44}$$

and, conversely, arbitrary non-negative values $P_{ij}^{(p)}$ satisfying (44) can be considered as the corresponding values of the probabilities of a certain stochastically determined process.

The probabilities $P_{ij}^{(pq)}$ can be calculated by the formula

$$P_{ij}^{(pq)} = P_{ij}^{(p+1)} * P_{ij}^{(p+2)} * \ldots * P_{ij}^{(q)}. \tag{45}$$

For a homogeneous scheme with discontinuous time we have

$$P_{ij}^{(p)} = P_{ij}, \quad P_{ij}^{(pq)} = [P_{ij}]_*^{q-p} = P_{ij}^{q-p}.$$

If all the P_{ij} are positive, then obviously the conditions of Theorem 2 (§4) hold, hence P_{ij}^q tend to a certain limit Q_j as $q \to \infty$. The integral equation (35) transforms in our case into the system of equations

$$Q_i = \sum_j Q_j P_{ji} \quad (i = 1, \ldots, n). \tag{46}$$

These results were obtained by Gostinskii and Hadamard.[6]

§6. Differential equations of a continuous stochastic process

By (11) we have

$$P_{ii}(t,t) = 1, \quad P_{ij}(t,t) = 0, \ i \neq j. \tag{47}$$

If the variations of our system $\mathfrak{S}$ are possible at any time t, then it is natural to suppose that

$$\lim_{\Delta\to\infty} P_{ii}(t,t+\Delta) = 1, \quad \lim_{\Delta\to 0} P_{ij}(t,t+\Delta) = 0, \ i \neq j, \tag{47a}$$

that is, for small time intervals the probability of a change in the state of the system is small. This assumption is contained in the hypothesis of the continuity of the functions $P_{ij}(t_1,t_2)$ with respect to t_1 and t_2.

Now assume that the functions $P_{ij}(s,t)$ are continuous and differentiable with respect to t and s for $t \neq s$. We do not require differentiability of these functions at $t = s$. It would be imprudent to assume *a priori* the existence of a derivative at these special points.[7]

For $t > s$ we have

$$\begin{aligned} \frac{\partial P_{ik}(s,t)}{\partial t} &= \lim_{\Delta\to 0} \frac{P_{ik}(s,t+\Delta) - P_{ik}(s,t)}{\Delta} = \\ &= \lim_{\Delta\to 0} \frac{1}{\Delta}\Big[\sum_j P_{ij}(s,t)P_{jk}(t,t+\Delta) - P_{ik}(s,t)\Big] = \\ &= \lim_{\Delta\to 0}\Big[\sum_{j\neq k} P_{ij}(s,t)\frac{P_{ij}(t,t+\Delta)}{\Delta} + P_{ik}(s,t)\frac{P_{kk}(t,t+\Delta)-1}{\Delta}\Big]. \end{aligned} \tag{48}$$

If the determinant

$$\Xi = |P_{ij}(s,t)|$$

is non-zero, then the equations

$$\sum_{j\neq k} P_{ij}(s,t)\frac{P_{jk}(t,t+\Delta)}{\Delta} + P_{ik}(s,t)\frac{P_{kk}(t,t+\Delta)-1}{\Delta} = \alpha_{ik} \quad (i = 1,\ldots,n)$$

can be solved:

$$\frac{P_{kk}(t,t+\Delta)-1}{\Delta} = \frac{\Lambda_{kk}}{\Xi}, \quad \frac{P_{jk}(t,t+\Delta)}{\Delta} = \frac{\Lambda_{jk}}{\Xi}, \ j \neq k. \tag{49}$$

---

[6] See Footnote 5.

[7] Compare with the functions $F(s,x,t,y)$ considered in Chapter 4, which necessarily have points of discontinuity at $t = s$.

Since by (48) α_{ik} tend to the limit values $\partial P_{ik}(s,t)/\partial t$ as $\Delta \to 0$, the values (49) tend to well defined limits[8]

$$\lim \frac{P_{kk}(t,t+\Delta)-1}{\Delta} = A_{kk}(t), \tag{50a}$$

$$\lim \frac{P_{jk}(t,t+\Delta)}{\Delta} = A_{jk}(t), \quad j \neq k. \tag{50b}$$

In fact it is evident from the relation

$$\lim_{s\to t} \Xi = 1, \tag{51}$$

which holds by (47) and the continuity of Ξ, that Ξ may be non-zero under a proper choice of $s < t$.

From (48) and (50) we immediately obtain the *first system of differential equations* for the function $P_{ik}(s,t)$:

$$\frac{\partial P_{ik}(s,t)}{\partial t} = \sum_j A_{jk}(t) P_{ij}(s,t) = P_{ik}(s,t) * A_{ik}(t). \tag{52}$$

In this case, by (47) and (50),

$$A_{jk}(t) = \left[\frac{\partial P_{jk}(t,u)}{\partial u}\right]_{u=t}, \tag{53}$$

$$A_{jk} \geq 0,\ j \neq k, \quad A_{kk} \leq 0, \tag{54}$$

and, by (41) and (50),

$$\sum_k A_{jk} = 0. \tag{55}$$

The equations (52) were established only for $s < t$; however, (47) and (53) show that these equations are valid also for $t = s$.

For $s < t$ we have

$$\frac{\partial P_{ik}(s,t)}{\partial s} = \lim_{\Delta\to 0} \frac{P_{ik}(s+\Delta,t) - P_{ik}(s,t)}{\Delta} =$$

$$= \lim_{\Delta\to 0} \frac{1}{\Delta}\Big[P_{ik}(s+\Delta,t) - \sum_j P_{ij}(s,s+\Delta) P_{jk}(s+\Delta,t)\Big] =$$

$$= -\lim_{\Delta\to 0}\Big[\frac{P_{ii}(s,s+\Delta)-1}{\Delta} P_{ik}(s+\Delta,t) +$$

$$+ \sum_{j\neq i} \frac{P_{ij}(s,s+\Delta)}{\Delta} P_{jk}(s+\Delta,t)\Big] \tag{56}$$

---

[8] We could equally well have taken the opposite approach: to assume *a priori* that the conditions (47a) and (50) hold and to derive from this the continuity and differentiability of the function $P_{ij}(s,t)$ with respect to t.

and, by (50), we obtain the *second system of differential equations*

$$\frac{\partial P_{ik}(s,t)}{\partial s} = -\sum_j A_{ij}(s)P_{jk}(s,t) = -A_{ik}(s) * P_{ik}(s,t). \tag{57}$$

If the functions $A_{ij}(s)$ are continuous, then clearly the equations (57) are also true for $s = t$.

Now assume that at t_0 we know the distribution function

$$Q(t_0, x_k) = Q_k(t_0), \quad \sum_k Q_k(t_0) = 1,$$

of the probabilities that the system $\mathfrak{S}$ at t_0 is in the state x_k. Then equation (5) takes the form

$$Q_k(t) = \sum Q_i(t_0)P_{ik}(t_0,t).$$

By (52) the functions $Q_k(t)$ satisfy

$$\frac{dQ_k(t)}{dt} = \sum_j A_{jk}(t)Q_j(t) \quad (k = 1, \dots, n). \tag{58}$$

If the functions $A_{ik}(t)$ are continuous, then the functions $P_{ik}(s,t)$ form the unique system of solutions of (52) satisfying the initial conditions (47); consequently, the considered stochastic process is totally determined by all the $A_{ik}(t)$. The real meaning of the functions $A_{ik}(t)$ can be illustrated in the following way: for $i \neq k$ $A_{ik}(t)dt$ is the probability of passing from the state x_i to the state x_k during the time from t to $t + dt$, whereas

$$A_{kk}(t) = -\sum_{j\neq k} A_{kj}(t).$$

It can also be shown that if we have any continuous functions $A_{ik}(t)$ satisfying the conditions (54) and (55), then the solutions $P_{ik}(s,t)$ of the differential equations (52) under the initial conditions (47) are non-negative and satisfy the conditions (40) and (41); in other words, they determine a stochastic process.

Indeed, by (52) and (55) we have

$$\frac{\partial}{\partial t}\sum_k P_{ik}(s,t) = \sum\Big[\sum_k A_{jk}(t)\Big]P_{ij}(s,t) = 0, \tag{59}$$

and, by (47),

$$\sum_k P_{ik}(t,t) = 1.$$

Thus (59) implies (41).

For $t_1 < t_2$ we now assume that

$$P'_{ik}(t_1, t) = P_{ik}(t_1, t), \quad \text{if } t_1 \leq t \leq t_2, \tag{60}$$

$$P'_{ik}(t_1, t) = \sum_j P_{ij}(t_1, t_2) P_{jk}(t_2, t), \quad \text{if } t_2 < t. \tag{61}$$

The functions $P'_{ik}(t_1, t)$ are continuous and satisfy the differential equations (52); consequently, (60) holds for any t and not merely for $t \leq t_2$; but then (61) with $t = t_3$ coincides with (40),

It remains to show that the solutions $P_{ik}(t_1, t)$ are non-negative. For this we assume that for fixed s,

$$\psi(t) = \min P_{ik}(s, t).$$

Choosing appropriate i and k we clearly have

$$D^+\psi(t) = \frac{\partial P_{ik}(s,t)}{\partial t}, \quad P_{ik}(s,t) = \psi(t),$$

and if $\psi(t) \leq 0$, then by (54),

$$A_{kk}(t) P_{ik}(s, t) \geq 0,$$

$$A_{jk}(t) P_{ij}(s,t) \geq A_{jk}(t)\psi(t), \ j \neq k,$$

$$D^+\psi(t) = \frac{\partial P_{ik}(s,t)}{\partial t} = \sum_j A_{jk}(t) P_{ij}(s,t) \geq$$

$$\geq \sum_{j \neq k} A_{jk}(t)\psi(t) = R(t)\psi(t).$$

Since $\psi(s) = 0$, $\psi(t)$ is clearly greater than any negative solution of the equation

$$dy/dt = R(t)y,$$

and therefore it cannot be negative itself.

§7. Examples

In schemes homogeneous in time the coefficients $A_{ik}(t)$ appear to be independent of the time t; in this case the process is completely determined by the n^2 constants A_{ik}. Equations (52) now take the form

$$\frac{dP_{ik}(t)}{dt} = \sum_j A_{ik} P_{ij}(t); \tag{62}$$

and solving these equations is not difficult. If all the A_{ik} are non-zero, then the conditions of Theorem 4 (§4) hold and consequently, $P_{ik}(t)$ tends to a limit Q_k as $t \to \infty$. The quantities Q_k satisfy the equations

$$\sum_k Q_k = 1, \quad \sum_j A_{jk} Q_j = 0 \ (k = 1, \ldots, n).$$

For example, let

$$n = 2, \quad A_{12} = A_{21} = A, \quad A_{11} = A_{22} = -A,$$

that is, the probabilities of transition from the state x_1 to the state x_2 and the reverse transition from x_2 to x_1 are the same. The differential equations (62) in our case give

$$P_{12}(t) = P_{21}(t) = \tfrac{1}{2}(1 - e^{-2At}),$$

$$P_{11}(t) = P_{22}(t) = \tfrac{1}{2}(1 + e^{-2At}).$$

We see that $P_{ik}(t)$ tends to the limit $Q_k = \frac{1}{2}$ as $t \to \infty$.

The following example shows that approaching the limit can be accompanied by oscillations damping with time:

$$n = 3, \quad A_{12} = A_{23} = A_{31} = A,$$

$$A_{21} = A_{32} = A_{13} = 0, \quad A_{11} = A_{22} = A_{33} = -A;$$

$$P_{11}(t) = P_{22}(t) = P_{33}(t) = \frac{2}{3} e^{-3/2At} \cos \alpha t + \frac{1}{3},$$

$$P_{12}(t) = P_{23}(t) = P_{31}(t) = e^{-3/2At}\Big(\frac{1}{\sqrt{3}} \sin \alpha t - \frac{1}{3} \cos \alpha t\Big) + \frac{1}{3},$$

$$P_{21}(t) = P_{32}(t) = P_{13}(t) = -e^{-3/2At}\Big(\frac{1}{\sqrt{3}} \sin \alpha t + \frac{1}{3} \cos \alpha t\Big) + \frac{1}{3},$$

$$\alpha = \frac{\sqrt{3}}{2} A.$$

Similar damping oscillations for schemes with discontinuous time were found by Romanovskii.

Chapter III. COUNTABLE STATE SYSTEMS

§8. Preliminary remarks. Discontinuous schemes

If $\mathfrak{A}$ consists of a countable set of elements

$$x_1, x_2, \ldots, x_n, \ldots,$$

all the notations and results of §5 of Chapter II remain valid. The convergence of the series

$$\sum_k P_{ik}(t_1, t_2) = 1, \quad \sum_k F_{ik} = 1$$

is assumed, and from this we derive convergence of the series (40), (42), (46); by contrast, we do not require that the series

$$\sum_i P_{ik}(t_1, t_2)$$

should converge.

We now make a few remarks on schemes with discontinuous time, in particular homogeneous ones. The conditions of our theorems concerning ergodic principles for schemes with a countable set of states fail in most cases, but nevertheless the principle itself often appears to be satisfied.

Consider, for example, a game studied recently by S.N. Bernshtein: in any separate trial a gambler wins only one rouble with probability A and loses it with probability B ($B > A$, $A + B \leq 1$), the latter, however, provided only that his cash is non-zero; otherwise he does not lose anything.

If we denote by x_n the state in which the cash of our gambler is $n - 1$ roubles, then the conditions of the game can be written as follows:

$$P_{n,n+1} = A, \quad P_{n+1,n} = B \quad (n = 1, 2, 3, \ldots),$$
$$P_{11} = 1 - A, \quad P_{nn} = 1 - A - B \quad (n = 2, 3, 4, \ldots),$$
$$P_{ij} = 0 \quad \text{otherwise.}$$

It can easily be proved that

$$\lim_{p \to \infty} P_{ij}^p = \left(1 - \frac{A}{B}\right)\left(\frac{A}{B}\right)^{j-1} = Q_j, \quad \sum_j Q_j = 1,$$

which implies the ergodic principle in this case.

Note that the fact that the limits

$$\lim_{p \to \infty} P_{ij}^p = \Lambda_j$$

exist, implies the ergodic principle only if

$$\sum_j \Lambda_j = \Lambda = 1.$$

It can be shown that always $\Lambda \leq 1$ and that for $\Lambda < 1$ the ergodic principle fails.

If all the Λ_j exist and are zero, then there arises the question of the asymptotic expression for P_{ij}^p as $p \to \infty$. If such an expression exists independently of i:

$$P_{ij}^p = \lambda_j^p + o(\lambda_j^p),$$

then we say that the *local ergodic principle* holds. This principle seems to be of great significance in the case of a countable set of possible states.

Now let all possible states x be enumerated by the integers $(-\infty < n < +\infty)$. All the notation and formulas of §5 are then true, but now the sums run over all the integers. We consider the case

$$P_{ij} = P_{j-i}^p$$

in more detail. Clearly in this case we have

$$P_{ij}^p = P_{j-i}^p, \quad P_k^{p+1} = \sum_i P_i^p P_{k-i}^n, \quad P_k^{m+n} = \sum_i P_i^m P_{k-i}^n.$$

If the series

$$a = \sum_k k P_k, \quad b^2 = \sum_k k^2 P_k$$

are absolutely convergent, then there arises the question on the conditions of applicability of the generalized Laplace formula

$$P_k^p = \frac{1}{b\sqrt{2\pi p}} \exp\Big[-\frac{(k-pa)^2}{2pb^2}\Big] + o\Big(\frac{1}{\sqrt{p}}\Big). \tag{63}$$

All we know is that it holds in the Bernoulli case, when

$$P_0 = 1 - A, \quad P_1 = A, \tag{64}$$

and the other P_k vanish. Lyapunov's theorem is of no help for our problem, as is clear from the following example:

$$P_{+1} = P_{-1} = \tfrac{1}{2}, \quad P_k = 0, \ k \neq \pm 1,$$

where (63) is inapplicable. In order for (63) to hold, it is necessary[9] that for any integer m there exists k such that

$$k \not\equiv 0 \,(\mathrm{mod}\, m), \quad P_k \neq 0.$$

[9] More details on this question can be found in R. von Mises *Wahrscheinlichkeitsrechnung*, Berlin, 1931, especially the chapter on "local" limit theorems. (Remark by Russian editor.)

Note also that only for $a = 0$ does formula (63) actually give an asymptotic expression for P_k^p for a given k. In this case it follows from (63) that for a given

$$P_k^p = \frac{1}{b\sqrt{2\pi p}} + o\Big(\frac{1}{\sqrt{p}}\Big) \tag{65}$$

and given i and j,

$$P_{ij}^p = \frac{1}{b\sqrt{2\pi p}} + o\Big(\frac{1}{\sqrt{p}}\Big). \tag{66}$$

By (66) we obtain the ergodic principle in the case considered.

We obtain special approximation formulas for P_{ij}^p when the probabilities P_{ii}, that is, the probabilities of the facts that the state of the system does not vary at any particular moment, are very close to one. For example, in the Bernoulli case, for small A the approximate Poisson formula can be used:

$$P_k^p \sim \frac{A^k p^k}{k!} e^{-Ap}. \tag{67}$$

A general method for deriving such formulas can be obtained by using differential equations for processes continuous in time, as is shown in §10 for formula (67).

§9. Differential equations of a process continuous in time

As in §6, we assume that the functions $P_{ij}(s,t)$ are continuous and have derivatives with respect to t and s for $t \neq s$. In the case of a countable set of possible states formulas (48) and (56) still remain valid; but to prove the possibility of changing the order of the sum and the limit in these formulas and thus arrive at the differential equations (52) and (57), we have to introduce new restrictions, namely:

A) the existence of limit values in (50);

B) uniform convergence in (50b) with respect to j for a given k;

C) uniform convergence of the series

$$\sum_{k \neq j} \frac{P_{jk}(t, t+\Delta)}{\Delta} = \frac{1 - P_{jj}(t, t+\Delta)}{\Delta} \tag{68}$$

with respect to Δ (the fact that this series converges follows immediately from (41)).

In §6, for a finite number of states we deduced condition A) from the differentiability of $P_{ij}(s,t)$ for $t \neq s$; by contrast, in the case of a countable set of states this condition does not seem to follow from this property of P_{ij}. With regard to condition B), note that uniform convergence in (50b) with respect to k for a given j follows from the obvious inequality

$$P_{jk}(t,t+\Delta) \leq 1 - P_{jj}(t,t+\Delta).$$

Note, further, that we do not require uniform convergence in (50b) for any j and k, nor do we require uniform convergence in (50a) with respect to k; these requirements would have been inconvenient for applications.

Since the factors $P_{ij}(s,t)$ in (48) form an absolutely convergent series, we can, in view of conditions A) and B), change the order of the signs lim and $\sum$ in this formula and obtain (52). Then the variables $A_{jk}(t)$ clearly satisfy the formulas of the last condition; moreover, since the factors $P_{jk}(s+\Delta,t)$ are uniformly bounded, we can change the order of the sum and limit signs in (56), which suffices for deducing (57).

§10. Uniqueness of solutions and their calculation for a process homogeneous in time

In the present case, (52) takes the form

$$\frac{dP_{ik}(t)}{dt} = \sum_j A_{jk}P_{ij}(t) = P_{ik}(t) * A_{ik}, \tag{69}$$

with constants A_{jk}. We will prove that if the series

$$\begin{gathered} \sum_j |A_{jk}| = B_k^{(1)}, \\ \sum_j B_j^{(1)}|A_{jk}| = B_k^{(2)}, \\ \cdots\cdots\cdots \\ \sum_j B_j^{(n)}|A_{jk}| = B_k^{(n+1)}, \\ \cdots\cdots\cdots \end{gathered} \tag{70}$$

$$\sum_n \frac{B_k^{(n)}}{n!}x^n, \quad k = 1,2,\ldots, \quad |x| \leq \theta\ (>0), \tag{71}$$

converge and the initial conditions

$$P_{ii}(0) = 1, \quad P_{ij}(0) = 0,\ i \neq j, \tag{72}$$

hold, then the equations (69) have the unique system of solutions $P_{ik}(t)$ satisfying the conditions of our problem.

Indeed, since always $P_{ij}(t) \le 1$, (69) and (70) imply

$$|dP_{ik}(t)/dt| \le B_k^{(1)};$$

therefore (69) can be differentiated term by term

$$\frac{d^2 P_{ik}(t)}{dt^2} = \sum A_{jk} \frac{dP_{ij}(t)}{dt} = \frac{d}{dt} P_{ik}(t) * A_{ik}.$$

In a similar way the general relations are obtained:

$$\left| \frac{d^n}{dt^n} P_{ik}(t) \right| \le B_k^{(n)}, \tag{73}$$

$$\frac{d^{n+1}}{dt^{n+1}} P_{ik}(t) = \frac{d^n}{dt^n} P_{ik}(t) * A_{ik}. \tag{74}$$

From (73) and the assumption of convergence of the series (71) it follows that the functions P_{ik} are analytic. Further, by (69) and (74) we find that

$$\frac{d^n}{dt^n} P_{ik}(t) = P_{ik}(t) * [A_{ik}]_*^n; \tag{75}$$

in particular, for $t = 0$ we have, by (72),

$$\frac{d^n}{dt^n} P_{ik}(0) = [A_{ik}]_*^n, \tag{76}$$

which implies that the analytic functions $P_{ik}(t)$ are uniquely determined by the constants A_{ik}. Formulas (76) and (75) serve also for calculating the solutions of the system (69) using Taylor series.

For example, if

$$A_{i,i+1} = A, \quad A_{ii} = -A,$$
$$A_{ij} = 0 \text{ otherwise},$$

then we easily obtain

$$P_{mn}(t) = \frac{(At)^{n-m}}{(n-m)!} e^{-At}, \quad n \ge m,$$
$$P_{mn}(t) = 0, \quad m > n,$$

that is, the formula of the Poisson distribution: for $k = n - m$, $p = t$ the resulting formula coincides with (67).

If the ergodic principle holds and $P_{ik}(t) \to Q_k$ as $t \to \infty$, then obviously the constants Q_k satisfy the equations

$$\sum_k Q_k = 1, \quad \sum_i A_{ik} Q_i = 0 \quad (k = 1, 2, \ldots). \tag{77}$$

If, for example

$$\begin{aligned} A_{i,i+1} &= A, \quad A_{i+1,i} = B, \quad B > A, \\ A_{11} &= -A, \quad A_{ii} = -(A+B), \quad i > 1, \\ A_{ij} &= 0 \quad \text{otherwise}, \end{aligned}$$

then we easily obtain from (77)

$$Q_n = (1 - A/B)(A/B)^{n-1}.$$

As a second example, we set

$$\begin{aligned} A_{i,i+1} &= A, \quad A_{i+1,i} = iB, \\ A_{ii} &= -A - (i-1)B, \\ A_{ij} &= 0 \quad \text{otherwise}, \end{aligned}$$

so that from (77) we have

$$Q_{n+1} = \frac{1}{n!}\left(\frac{A}{B}\right)^n e^{-A/B},$$

which again is Poisson's formula.

CHAPTER IV. CONTINUOUS STATE SYSTEMS, THE CASE OF ONE PARAMETER

§11. Preliminary remarks

Suppose now that the state of the system considered is determined by the values of a certain real parameter x; in this case we denote by x both the state of the system and the value of the parameter corresponding to this state. If $\mathfrak{E}_y$ is the set of all states x for which $x \leq y$, then we set

$$P(t_1, x, t_2, \mathfrak{E}_y) = F(t_1, x, t_2, y).$$

As a function of y, $F(t_1, x, t_2, y)$ is monotone and right continuous and satisfies the boundary conditions

$$F(t_1, x, t_2, -\infty) = 0, \quad F(t_1, x, t_2, +\infty) = 1. \tag{78}$$

For the function $F(t_1, x, t_2, y)$ the fundamental equation (3) transforms into:

$$F(t_1, x, t_3, z) = \int_{-\infty}^{\infty} F(t_2, y, t_3, z) dF(t_1, x, t_2, y). \tag{79}$$

Thus we have to use integral distribution functions of random variables and ordinary Stieltjes integrals.

Integral (79) exists according to Lebesgue[10] if $F(t_2, y, t_3, z)$ is Borel-measurable with respect to y. In what follows we assume that the system $\mathfrak{F}$ (see §1) coincides with the system of all Borel sets, which implies Borel-measurability of $F(t_1, x, t_2, y)$ as a function of x. In this case, as is known, the additive set function $P(t_1, x, t_2, \mathfrak{E})$, for all Borel sets $\mathfrak{E}$, is uniquely determined by the corresponding function $F(t_1, x, t_2, y)$.

A function $F(y)$, monotone and right continuous, such that

$$F(-\infty) = 0, \quad F(+\infty) = 1$$

is called a *normal distribution function.* If $F_1(x, y)$ and $F_2(x, y)$, as functions of x, are Borel measurable and, as functions of y, are normal distribution functions, then the same is true for the function

$$F(x, y) = F_1(x, y) \oplus F_2(x, y) = \int_{-\infty}^{\infty} F_2(z, y) dF_1(x, z). \tag{80}$$

This operator $\oplus$, like $*$, obeys the associative law; using this law, the fundamental equation (79) can be expressed as

$$F(t_1, x, t_3, y) = F(t_1, x, t_2, y) \oplus F(t_2, x, t_3, y). \tag{81}$$

with $F_1(x, y) = V_1(y - x)$, $\quad F_2(x, y) = V_2(y - x)$. Then, as can easily be shown,

$$F_1(x, y) \oplus F_2(x, y) = V(y - x) = V_1(y - x) \odot V_2(y - x), \tag{82}$$

[10] H. Lebesgue, *Leçons sur l'intégration et la recherche des fonctions primitives*, Gauthier-Villars, Paris, 1928.

where

$$V(x) = V_1(x) \odot V_2(x) = \int_{-\infty}^{\infty} V_2(x-z)dV_1(z). \tag{83}$$

The associative law also holds for the operator $\odot$ while for normal distribution functions the commutative law holds as well; if $V_1(x)$ and $V_2(x)$ are considered as distribution functions of two independent random variables X_1 and X_2, then $V_1(x) \odot V_2(x)$, as is known, is the distribution function of the sum[11] $X = X_1 + X_2$.

If $F(t_1, x, t_2, y)$ is absolutely continuous as a function of y, then we have

$$F(t_1, x, t_2, y) = \int_{-\infty}^{y} f(t_1, x, t_2, y)dy. \tag{84}$$

In this case the non-negative function $f(t_1, x, t_2, y)$ is Borel measurable with respect to x and y and satisfies

$$\int_{-\infty}^{\infty} f(t_1, x, t_2, y) = 1, \tag{85}$$

$$f(t_1, x, t_3, z) = \int_{-\infty}^{\infty} f(t_1, x, t_2, y)f(t_2, y, t_3, z)dy. \tag{86}$$

Conversely, if (85), (86) hold for $f(t_1, x, t_2, y)$, then the function $F(t_1, x, t_2, y)$ defined by (84) satisfies (78) and (79): hence, such a function determines the scheme of a stochastic process. This function $f(t_1, x, t_2, y)$ will be called the *differential distribution function* for the random variable y.

Note also that the following mixed formulas hold:

$$F(t_1, x, t_3, z) = \int_{-\infty}^{\infty} F(t_2, y, t_3, z)f(t_1, x, t_2, y)dy, \tag{87}$$

$$f(t_1, x, t_3, z) = \int_{-\infty}^{\infty} f(t_2, y, t_3, z)dF(t_1, x, t_2, y). \tag{88}$$

When the scheme is discontinuous in time, the functions

$$F_{mn}(x, y) = F(t_m, x, t_n, y), \quad F_n(x, y) = F_{n-1,n}(x, y),$$

are considered; they satisfy the equations

$$F_{m,n+1}(x, y) = F_{mn}(x, y) \oplus F_{n+1}(x, y), \tag{89}$$

$$F_{kn}(x, y) = F_{km}(x, y) \oplus F_{mn}(x, y) \quad (k < m < n). \tag{90}$$

[11] See P. Lévy, *Calcul des probabilités*, Paris, 1927, p.187.

If

$$F_{mn}(x,y) = \int_{-\infty}^{y} f_{mn}(x,y)dy, \quad f_n(x,y) = f_{n-1,n}(x,y),$$

then, in addition, we have

$$f_{m,n+1}(x,z) = \int_{-\infty}^{\infty} f_{mn}(x,y)f_{n+1}(y,z)dy, \tag{91}$$

$$f_{kn}(x,z) = \int_{-\infty}^{\infty} f_{km}(x,y)f_{mn}(y,z)dy \quad (k < m < n). \tag{92}$$

§12. Lindeberg's method. Passage from discontinuous to continuous schemes

As we noted in §3, probability theory usually deals only with schemes that are discontinuous in time. For these schemes, the main problem is to find approximate expressions for the distributions $F_{mn}(x,y)$ for large $n-m$, or what is essentially the same, to construct asymptotic formulas for $F_{mn}(x,y)$ as $n \to \infty$. The Laplace-Lyapunov theorem is the most important result achieved in this direction. Now we will consider in more detail the proof of this theorem given by Lindeberg[12] with the purpose of outlining his main idea in as general a form as possible and thus obtaining a general method for constructing asymptotic expressions for $F_{mn}(x,y)$.

Let

$$F_n(x,y) = V_n(y-x),$$

$$a_n(x) = \int_{-\infty}^{\infty} (y-x)dF_n(x,y) = \int_{-\infty}^{\infty} ydV_n(y) = 0,$$

$$b_n^2(x) = \int_{-\infty}^{\infty} (y-x)^2 dF_n(x,y) = \int_{-\infty}^{\infty} y^2 dV_n(y) = b_n^2,$$

$$B_{mn}^2 = b_{m+1}^2 + b_{m+2}^2 + \ldots + b_n^2.$$

The Laplace-Lyapunov theorem states that under certain additional assumptions, for constant m and as $n \to \infty$ we have

$$F_{mn}(x,y) = \Phi\left(\frac{y-x}{B_{mn}}\right) + o(1)$$

[12] *Math. Z.* **15** (1922), 211.

uniformly with respect to x and y, where

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2} dz.$$

Along with the stochastic process with discontinuous time determined by the functions $F_n(x, y)$, we will consider another one, with continuous time; we suppose that it is characterized by the function

$$\overline{F}(t', x, t'', y) = \Phi\Big(\frac{y - x}{\sqrt{t'' - t'}}\Big).$$

Further, let

$$t_0 = 0, \quad t_n = B_{0n}^2,$$
$$\overline{F}_{mn}(x, y) = \overline{F}(t_m, x, t_n, y), \quad \overline{F}_n(x, y) = \overline{F}_{n-1,n}(x, y).$$

Clearly we have

$$\overline{F}_n(x, y) = \Phi\Big(\frac{y - x}{b_n}\Big),$$
$$\bar{a}_n(x) = \int_{-\infty}^{\infty} (y - x) d\overline{F}_n(x, y) = 0,$$
$$\bar{b}_n^2(x) = \int_{-\infty}^{\infty} (y - x)^2 d\overline{F}_n(x, y) = b_n^2.$$

The first and second moments $\bar{a}_n(x)$ and $\bar{b}_n^2(x)$ of the distribution $\overline{F}_n(x, y)$ coincide with the corresponding moments $a_n(x)$ and $b_n^2(x)$ of the distribution $F_n(x, y)$. From this Lindeberg deduced that

$$F_{mn}(x, y) - \overline{F}_{mn}(x, y) \to 0 \quad \text{as } n \to \infty,$$

after which the Laplace-Lyapunov theorem follows directly from the obvious identity

$$\overline{F}_{mn}(x, y) = \Phi\Big(\frac{y - x}{B_{mn}}\Big).$$

In the general case of arbitrary functions $F_n(x, y)$ we can only apply Lindeberg's method if we know a function $\overline{F}(t', x, t'', y)$ characterizing a continuous stochastic process, and which, for a certain sequence of instants of time

$$t_0 < t_1 < t_2 < \ldots < t_n < \ldots$$

gives the moments $\bar{a}_n(x)$, $\bar{b}_n^2(x)$ which coincide with $a_n(x)$, $b_n^2(x)$, or are close to them. A general method for constructing such functions $\overline{F}$ is obtained by using differential equations of continuous processes, considered in the sections below. To pass from $\overline{F}$ to F we can use the following:

Transition Theorem. *Let the functions $F_n(x,y)$ and $\overline{F}_n(x,y)$ determine two stochastic processes with discontinuous time. If*

$$\int_{-\infty}^{\infty}(y-x)dF_n(x,y) = a_n(x) \quad \int_{-\infty}^{\infty}(y-x)d\overline{F}_n(x,y) = \bar{a}_n(x), \tag{93}$$

$$\int_{-\infty}^{\infty}(y-x)^2dF_n(x,y) = b_n^2(x), \quad \int_{-\infty}^{\infty}(y-x)^2d\overline{F}_n(x,y) = \bar{b}_n^2(x), \tag{94}$$

$$\int_{-\infty}^{\infty}|y-x|^3dF_n(x,y) = c_n(x), \quad \int_{-\infty}^{\infty}|y-x|^3d\overline{F}_n(x,y) = \bar{c}_n(x), \tag{95}$$

$$|a_n(x)-\bar{a}_n(x)| \leq p_n, \quad |b_n(x)-\bar{b}_n^2(x)| \leq q_n, \quad c_n(x) \leq r_n, \quad \bar{c}_n(x) \leq \bar{r}_n, \tag{96}$$

and if there exists a function $R(x)$ such that

$$\begin{aligned} R(x) &= 0, \quad \text{for } x \leq 0, \\ 0 \leq R(x) &\leq 1, \quad \text{for } 0 < x < l, \\ R(x) &= 1, \quad \text{for } l \leq x, \end{aligned} \tag{97}$$

and for

$$U_{kn}(x,z) = \int_{-\infty}^{\infty} R(z-y)d\overline{F}_{kn}(x,y) \tag{98}$$

the inequalities

$$\begin{aligned} &\left|\frac{\partial}{\partial x}U_{kn}(x,z)\right| \leq K_n^{(1)}, \quad \left|\frac{\partial^2}{\partial x^2}U_{kn}(x,z)\right| \leq K_n^{(2)}, \\ &\left|\frac{\partial^3}{\partial x^3}U_{kn}(x,z)\right| \leq K_n^{(3)}, \quad (k=0,1,\ldots,n), \end{aligned} \tag{99}$$

hold, then the relation

$$\overline{F}_{0n}(x,y-l) - \epsilon_n \leq F_{0n}(x,y) \leq \overline{F}_{0n}(x,y+l) + \epsilon_n, \tag{100}$$

holds, where

$$\epsilon_n = K_n^{(1)}\sum_{k=1}^{n}p_k + \frac{1}{2}K_n^{(2)}\sum_{k=1}^{n}q_k + \frac{1}{6}K_n^{(3)}\sum_{k=1}^{n}(r_k+\bar{r}_k).$$

In applying this theorem to the case when the moments $a(x), b(x), c(x)$ are unbounded as x increases, it is often possible to eliminate this unboundedness by introducing a new properly chosen variable $x' = \phi(x)$.

Proof of the Transition Theorem. By (98) we have

$$U_{k-1,n}(x,y) = \overline{F}_{k-1,n}(x,y) \oplus R(y-x) =$$
$$= \overline{F}_k(x,y) \oplus \overline{F}_{k+1}(x,y) \oplus \ldots \oplus \overline{F}_n(x,y) \oplus R(y-x) =$$
$$= \overline{F}_k(x,y) \oplus U_{kn}(x,y) \tag{101}$$

and by (93)–(95), (99),

$$U_{k-1,n}(x,y) = \int_{-\infty}^{\infty} U_{kn}(z,y)d\overline{F}_k(x,z) =$$
$$= \int_{-\infty}^{\infty} \Big[U_{kn}(x,y) + \frac{\partial}{\partial x}U_{kn}(x,y)\frac{z-x}{1}+$$
$$+\frac{\partial^2}{\partial x^2}U_{kn}(x,y)\frac{(z-x)^2}{2} + \frac{\partial^3}{\partial x^3}U_{kn}(\xi,y)\frac{(z-x)^3}{6}\Big]d\overline{F}_k(x,z) =$$
$$= U_{kn}(x,y) + \frac{\partial}{\partial x}U_{kn}(x,y)\bar{a}_k(x)+$$
$$+\frac{\partial^2}{\partial x^2}U_{kn}(x,y)\frac{\bar{b}_k^2(x)}{2} + \bar{\theta}K_n^{(3)}\frac{\bar{c}_k(x)}{6}, \quad |\bar{\theta}| \leq 1. \tag{102}$$

Setting

$$V_{k-1,n}(x,y) = F_k(x,y) \oplus U_{kn}(x,y), \tag{103}$$

we obtain a formula similar to (102),

$$V_{k-1,n}(x,y) = U_{kn}(x,y) + \frac{\partial}{\partial x}U_{kn}(x,y)a_k(x)+$$
$$+\frac{\partial^2}{\partial x^2}U_{kn}(x,y)\frac{b_k^2(x)}{2} + \theta K_n^{(3)}\frac{c_k(x)}{6}, \quad |\theta| \leq 1. \tag{104}$$

From (102) and (104) and using (96) and (99) it follows that

$$|U_{k-1,n}(x,y) - V_{k-1,n}(x,y)| \leq K_n^{(1)}p_k + \frac{1}{2}K_n^{(2)}q_k + \frac{1}{6}K_n^{(3)}(r_k + \bar{r}_k). \tag{105}$$

Now let

$$W_{kn}(x,y) = F_{0k}(x,y) \oplus U_{kn}(x,y) =$$
$$= F_1(x,y) \oplus F_2(x,y) \oplus \ldots \oplus F_k(x,y) \oplus U_{kn}(x,y) =$$
$$= F_{0,k-1}(x,y) \oplus V_{k-1,n}(x,y). \tag{106}$$

Then by (105) we have

$$|W_{kn}(x,y) - W_{k-1,n}(x,y)| =$$
$$= |F_{0,k-1}(x,y) \oplus V_{k-1,n}(x,y) - F_{0,k-1}(x,y) \oplus U_{k-1,n}(x,y)| \leq$$
$$\leq \int_{-\infty}^{\infty} |V_{k-1,n}(z,y) - U_{k-1,n}(z,y)| dF_{0,k-1}(x,z) \leq$$
$$\leq \sup |V_{k-1,n}(z,y) - U_{k-1,n}(z,y)| \leq K_n^{(1)} p_k + \frac{1}{2} K_n^{(2)} q_k + \frac{1}{6} K_n^{(3)} (r_k + \bar{r}_k). \tag{107}$$

$$|W_{nn}(x,y) - W_{0n}(x,y)| \leq$$
$$\leq K_n^{(1)} \sum_{k=1}^{n} p_k + \frac{1}{2} K_n^{(2)} \sum_{k=1}^{n} q_k + \frac{1}{6} K_n^{(3)} \sum_{k=1}^{n} (r_k + \bar{r}_k) = \epsilon_n.$$

But

$$W_{nn}(x,y) = F_{0n}(x,y) \oplus R(y-x) = \int_{-\infty}^{\infty} R(y-z) dF_{0n}(x,z)$$

and

$$W_{0n}(x,y) = \overline{F}_{0n}(x,y) \oplus R(y-x) = \int_{-\infty}^{\infty} R(y-z) d\overline{F}_{0n}(x,z).$$

Taking into account (97) we obtain

$$\begin{aligned} W_{nn}(x,y) &\leq \int_{-\infty}^{y} dF_{0n}(x,z) = F_{0n}(x,y), \\ W_{nn}(x,y+l) &\geq \int_{-\infty}^{y} dF_{0n}(x,z) = F_{0n}(x,y), \\ W_{0n}(x,y) &\geq \int_{-\infty}^{y-l} d\overline{F}_{0n}(x,z) = \overline{F}_{0n}(x,y-l), \\ W_{0n}(x,y+l) &\leq \int_{-\infty}^{y+l} d\overline{F}_{0n}(x,z) = \overline{F}_{0n}(x,y+l), \end{aligned} \tag{108}$$

Formula (100) now follows immediately from (107) and (108). The details of the proof can be found in Lindeberg's paper referred to above.

§13. The first differential equation for processes continuous in time

If the state of our system $\mathfrak{S}$ can be changed at any moment t, then it is natural to assume that significant changes of the parameter x during small time intervals will occur very seldom, or, more exactly, that for any positive ϵ,

$$P(t,x,t+\Delta,\ |y-x|>\epsilon)\to 0 \quad \text{as } \Delta\to 0. \tag{109}$$

In most cases we may assume that the stricter condition

$$m^{(p)}(t,x,\Delta)=\int_{-\infty}^{+\infty}|y-x|^p dF(t,x,t+\Delta,y)\to 0 \quad \text{as } \Delta\to 0 \tag{110}$$

holds, at least for the first three moments $m^{(1)}, m^{(2)}$ and $m^{(3)}$. A general study of the possibilities that arise under these assumptions is of great interest; some remarks to this end will be given below in §19.

In the following sections we also assume that the following important condition holds:

$$\frac{m^{(3)}(t,x,\Delta)}{m^{(2)}(t,x,\Delta)}\to 0 \quad \text{as } \Delta\to 0. \tag{111}$$

This condition will certainly hold if in the definition of $m^{(3)}(t,x,\Delta)$ via (110) only infinitesimally small differences $y-x$ play a significant role for infinitesimally small Δ or, more precisely, if

$$\frac{\int_{x-\epsilon}^{x+\epsilon}|y-x|^3 dF(t,x,t+\Delta,y)}{\int_{-\infty}^{+\infty}|y-x|^3 dF(t,x,t+\Delta,y)}\to 1 \quad \text{as } \Delta\to 0. \tag{112}$$

Strictly speaking, only in this case is our process continuous in time. Formula (111) also implies that

$$\frac{m^{(2)}(t,x,\Delta)}{m^{(1)}(t,x,\Delta)}\to 0 \quad \text{as } \Delta\to 0.$$

Finally, we will also assume that for $s\neq t$ all the partial derivatives of the function $F(s,x,t,y)$ up to the fourth order exist, and that these derivatives for constant t,y are uniformly bounded with respect to s and x for $t-s>k>0$. From (78) and (110) we conclude that for $s=t$ the function $F(s,x,t,y)$ is, by contrast, discontinuous. The function

$$f(s,x,t,y)=\frac{\partial}{\partial y}F(s,x,t,y), \tag{113}$$

clearly satisfies (84)–(86) and, at given t, y, has for $t - s > k > 0$ derivatives up to the third order that are uniformly bounded with respect to s and t. All further calculations are made for this differential distribution function $f(s, x, t, y)$.

Set

$$a(t, x, \Delta) = \int_{-\infty}^{\infty} (y - x) f(t, x, t + \Delta, y) dy, \tag{114}$$

$$b^2(t, x, \Delta) = \int_{-\infty}^{\infty} (y - x)^2 f(t, x, t + \Delta, y) dy = m^{(2)}(t, x, \Delta), \tag{115}$$

$$c(t, x, \Delta) = \int_{-\infty}^{\infty} |y - x|^3 f(t, x, t + \Delta, y) dy = m^{(3)}(t, x, \Delta). \tag{116}$$

By (85) and (86) we have

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

$$= \int_{-\infty}^{\infty} f(s, x, s + \Delta, z) \Big[f(s + \Delta, x, t, y) +$$

$$+ \frac{\partial}{\partial x} f(s + \Delta, x, t, y)(z - x) +$$

$$+ \frac{\partial^2}{\partial x^2} f(s + \Delta, x, t, y) \frac{(z - x)^2}{2} +$$

$$+ \frac{\partial^3}{\partial \xi^3} f(s + \Delta, \xi, t, y) \frac{(z - x)^3}{6} \Big] dz =$$

$$= f(s + \Delta, x, t, y) + \frac{\partial}{\partial x} f(s + \Delta, x, t, y) a(s, x, \Delta) +$$

$$+ \frac{\partial^2}{\partial x^2} f(s + \Delta, x, t, y) \frac{b^2(s, x, \Delta)}{2} + \theta \frac{c(s, x, \Delta)}{6}, \quad |\theta| < C, \tag{117}$$

where for $s + \Delta < \tau < t$, C can be chosen independently of Δ. From (117) we immediately obtain

$$\frac{f(s + \Delta, x, t, y) - f(s, x, t, y)}{\Delta} =$$

$$= - \frac{\partial}{\partial x} f(s + \Delta, x, t, y) \frac{a(s, x, \Delta)}{\Delta} -$$

$$- \frac{\partial^2}{\partial x^2} f(s + \Delta, x, t, y) \frac{b^2(s, x, \Delta)}{2\Delta} - \theta \frac{c(s, x, \Delta)}{6\Delta}. \tag{118}$$

First we prove that if for given x and s the determinant

$$D(s,x,t',y',t'',y'') = \begin{vmatrix} \frac{\partial}{\partial x} f(s,x,t',y') & \frac{\partial}{\partial x} f(s,x,t'',y'') \\ \frac{\partial^2}{\partial x^2} f(s,x,t',y') & \frac{\partial^2}{\partial x^2} f(s,x,t'',y'') \end{vmatrix} \tag{119}$$

does not vanish identically for any t', y', t'', y'', then the ratios

$$a(s,x,\Delta)/\Delta \text{ and } b^2(s,x,\Delta)/2\Delta$$

tend to well defined limits $A(s,x)$ and $B^2(s,x)$ as $\Delta \to 0$.

Thus, let t', y', t'', y'' be chosen so that (119) is non-zero; in this case, for any sufficiently small Δ we also have

$$D(s+\Delta,x,t',y',t'',y'') \neq 0,$$

so that the equations

$$\begin{aligned} \lambda(\Delta)\frac{\partial}{\partial x} f(s+\Delta,x,t',y') + \mu(\Delta)\frac{\partial}{\partial x} f(s+\Delta,x,t'',y'') &= 0, \\ \lambda(\Delta)\frac{\partial^2}{\partial x^2} f(s+\Delta,x,t',y') + \mu(\Delta)\frac{\partial^2}{\partial x^2} f(s+\Delta,x,t'',y'') &= 1 \end{aligned} \tag{120}$$

have a unique solution. In this case $\lambda(\Delta)$ and $\mu(\Delta)$ tend to $\lambda(0)$ and $\mu(0)$ as $\Delta \to 0$. Further, by (118) we obtain

$$\begin{aligned} \lambda(\Delta)\frac{f(s+\Delta,x,t',y') - f(s,x,t',y')}{\Delta} &+ \\ +\mu(\Delta)\frac{f(s+\Delta,x,t'',y'') - f(s,x,t'',y'')}{\Delta} &= \\ = -\frac{b^2(s,x,\Delta)}{2\Delta} - (\theta' + \theta'')\frac{c(s,x,\Delta)}{6\Delta}&. \end{aligned} \tag{121}$$

The left-hand side of formula (121) tends to

$$\Omega = \lambda(0)\frac{\partial}{\partial s} f(s,x,t',y') + \mu(0)\frac{\partial}{\partial s} f(s,x,t'',y'')$$

as $\Delta \to 0$, whereas on the right-hand side the second term is infinitesimally small as compared with the first one, by (111); therefore this term tends to the limit

$$B^2(s,x) = \lim_{\Delta \to 0} \frac{b^2(s,x,\Delta)}{2\Delta} = -\Omega. \tag{122}$$

It follows immediately from (122) and (111) that

$$c(s,x,\Delta)/\Delta \to 0 \quad \text{as } \Delta \to 0. \tag{123}$$

By (122) and (123), formula (118) for $\Delta = 0$ becomes:

$$\lim_{\Delta\to 0}\left[\frac{\partial}{\partial x}f(s,x,t,y)\frac{a(s,x,\Delta)}{\Delta}\right] = -\frac{\partial}{\partial s}f(s,x,t,y)-$$

$$-\frac{\partial^2}{\partial x^2}f(s,x,t,y)B^2(s,x).$$

Since $\partial f(s,x,t,y)/\partial x$ does not vanish identically for any t and y, the following limit also exists:

$$A(s,x) = \lim_{\Delta\to 0}\frac{a(s,x,\Delta)}{\Delta} =$$

$$= \frac{-\partial f(s,x,t,y)/\partial s - B^2(s,x)\partial^2 f(s,x,t,y)/\partial x^2}{\partial f(s,x,t,y)/\partial x}. \tag{124}$$

Passing to the limit in (118), (122), (123) and (124) we obtain the *first fundamental differential equation*

$$\frac{\partial}{\partial s}f(s,x,t,y) = -A(s,x)\frac{\partial}{\partial x}f(s,x,t,y) - B^2(s,x)\frac{\partial^2}{\partial x^2}f(s,x,t,y). \tag{125}$$

When the determinant $D(s,x,t',y',t'',y'')$ vanishes for any t',y',t'',y'', then the limits $A(s,x)$ and $B^2(s,x)$ do not in general exist, as is clear from the following example:

$$f(s,x,t,y) = \frac{3y^2}{2\sqrt{\pi(t-s)}}e^{-(y^3-x^3)^2/4(t-s)}. \tag{126}$$

Here, for $x = 0$ we have

$$b^2(s,x,\Delta)/2\Delta \to +\infty \quad \text{as } \Delta \to 0.$$

It can be shown, however, that these singular points (s,x) form a nowhere dense set in the (s,x)-plane.

The practical significance of these very important quantities $A(s,x)$ and $B(s,x)$ is as follows: $A(s,x)$ is the mean rate of variation of the parameter x over an infinitesimally small time interval and $B(s,x)$ is the differential variance of the process. The variance of the difference $y - x$ for the time interval Δ is

$$b(s,x,\Delta) = B(s,x)\sqrt{2\Delta} + o(\sqrt{\Delta}) = O(\sqrt{\Delta}); \tag{127}$$

and the expectation of this difference is

$$a(s,x,\Delta) = A(s,x)\Delta + o(\Delta) = O(\Delta). \tag{128}$$

It is worth mentioning that the expectation $m^{(1)}(t,x,\Delta)$ of $|y-x|$, like the variance $b(s,x,\Delta)$, is a quantity of order $\sqrt{\Delta}$.

As will be shown in the following sections, the functions $A(s,x)$ and $B(s,x)$ in some cases uniquely determine our stochastic scheme.

§14. The second differential equation

In this section we retain all the requirements imposed on $f(s,x,t,y)$ in §13 and, in addition, assume that $f(s,x,t,y)$ has continuous derivatives up to the fourth order. Then, from (120) it clearly follows that if the determinant (119) is non-zero, then $\lambda(0)$ and $\mu(0)$ have continuous derivatives with respect to s and x up to the second order; by (122) and (124) the same is also true for $B^2(s,x)$ and $A(s,x)$.

Now, assume that for a certain t we are given an interval $a \le y \le b$ such that at each point of this interval the determinant $D(t,y,u',z',u'',z'')$ does not vanish identically for any u',z',u'',z''. Next, let $R(y)$ be a function that is non-zero only on the interval $a < y < b$, is non-negative and has bounded derivatives up to the third order. In this case we have

$$\int_a^b \frac{\partial}{\partial t} f(s,x,t,y)R(y)dy = \frac{\partial}{\partial t}\int_a^b f(s,x,t,y)R(y)dy =$$

$$= \lim_{\Delta\to 0}\frac{1}{\Delta}\int_{-\infty}^{\infty}[f(s,x,t+\Delta,y) - f(s,x,t,y)]R(y)dy =$$

$$= \lim_{\Delta\to 0}\frac{1}{\Delta}\Big\{\int_{-\infty}^{\infty} R(y)\int_{-\infty}^{\infty} f(s,x,t,z)f(t,z,t+\Delta,y)dz\,dy -$$

$$- \int_{-\infty}^{\infty} f(s,x,t,y)R(y)dy\Big\} = \lim_{\Delta\to 0}\frac{1}{\Delta}\times$$

$$\times\Big\{\int_{-\infty}^{\infty} f(s,x,t,z)\int_{-\infty}^{\infty} f(t,z,t+\Delta,y)\Big[R(z) + R'(z)(y-z) +$$

$$+R''(z)\frac{(y-z)^2}{2} + R'''(\xi)\frac{(y-z)^3}{6}\Big]dy\,dz -$$

$$- \int_{-\infty}^{\infty} f(s,x,t,z)R(z)dz\Big\} = \lim_{\Delta\to 0}\frac{1}{\Delta}\int_{-\infty}^{\infty} f(s,,x,t,z)\times$$

$$\times\Big[R'(z)a(t,z,\Delta)+R''(z)\frac{b^2(t,z,\Delta)}{2}+\theta\frac{c(t,z,\Delta)}{6}\Big]dz=$$

$$=\int_{-\infty}^{\infty} f(s,x,t,z)[R'(z)A(t,z)+R''(z)B^2(t,z)]dz=$$

$$=\int_a^b f(s,x,t,y)[R'(y)A(t,y)+R''(y)B^2(t,y)]dy, \tag{129}$$

$$|\theta|\le \sup|R'''(\xi)|.$$

The passage to the limit with respect to Δ in deriving these formulas is justified by the fact that $a(t,z,\Delta)/\Delta$, $b^2(t,z,\Delta)/2\Delta$ and $c(t,z,\Delta)/\Delta$ tend uniformly to $A(t,z)$, $B^2(t,z)$ and 0 respectively, and the integral of the factor $f(s,x,t,z)$ with respect to z is finite.

Integrating by parts we obtain

$$\int_a^b f(s,x,t,y)R'(y)A(t,y)dy=-\int_a^b \frac{\partial}{\partial y}[f(s,x,t,y)A(t,y)]R(y)dy. \tag{130}$$

In exactly the same way, integrating by parts twice, we obtain

$$\int_a^b f(s,x,t,y)R''(y)B^2(t,y)dy=\int_a^b \frac{\partial^2}{\partial y^2}[f(s,x,t,y)B^2(t,y)]R(y)dy, \tag{131}$$

since $R(a)=R(b)=R'(a)=R'(b)=0$. Formulas (129)–(131) immediately imply that

$$\int_a^b \frac{\partial}{\partial t}f(s,x,t,y)R(y)dy=\int_a^b\Big\{-\frac{\partial}{\partial y}[A(t,y)f(s,x,t,y)]+ +\frac{\partial^2}{\partial y^2}[B^2(t,y)f(s,x,t,y)]\Big\}R(y)dy. \tag{132}$$

However, since the function $R(y)$ can be chosen arbitrarily only if the above conditions are fulfilled, we easily see that for points (t,y) at which the determinant $D(t,y,u',z',u'',z'')$ does not vanish identically the *second fundamental differential equation* also holds:

$$\frac{\partial}{\partial t}f(s,x,t,y)=-\frac{\partial}{\partial y}[A(t,y)f(s,x,t,y)]+\frac{\partial^2}{\partial y^2}[B^2(t,y)f(s,x,t,y)]. \tag{133}$$

This second equation could also have been obtained without using the first one, using the methods described in §13 directly; then, however, new and more stringent restrictions (that we omit here) would have to be imposed on $f(s,x,t,y)$. In that case we start from the formula, similar to (118),

$$\frac{1}{\Delta}[f(s,x,t,y)-f(s,x,t-\Delta,y)]=$$

$$=f(s,x,t-\Delta,y)\frac{1}{\Delta}\Big[\int_{-\infty}^{\infty}f(t-\Delta,z,t,y)dz-1\Big]+$$

$$+\frac{\partial}{\partial y}f(s,x,t-\Delta,y)\frac{1}{\Delta}\int_{-\infty}^{\infty}f(t-\Delta,z,t,y)(z-y)dz+$$

$$\frac{\partial^2}{\partial y^2}f(s,x,t-\Delta,y)\frac{1}{2\Delta}\int_{-\infty}^{\infty}f(t-\Delta,z,t,y)(z-y)^2dz+$$

$$+\frac{1}{6\Delta}\int_{-\infty}^{\infty}f(t-\Delta,z,t,y)|z-y|^3dz. \tag{134}$$

Then we prove that

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\int_{-\infty}^{\infty}f(t-\Delta,z,t,y)|z-y|^3dz=0$$

and that the limits

$$\lim_{\Delta\to 0}\frac{1}{2\Delta}\int_{-\infty}^{\infty}f(t-\Delta,z,t,y)|z-y|^2dz=\overline{B}^2(t,y), \tag{135}$$

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\int_{-\infty}^{\infty}f(t-\Delta,z,t,y)(z-y)dz=\overline{A}(t,y), \tag{136}$$

$$\lim_{\Delta\to 0}\frac{1}{\Delta}\Big(\int_{-\infty}^{\infty}f(t-\Delta,z,t,y)dz-1\Big)=\overline{N}(t,y), \tag{137}$$

exist. Thus we would have obtained our second equation in the following form

$$\frac{\partial}{\partial t}f(s,x,t,y)=\overline{N}(t,y)f(s,x,t,y)+$$

$$+\overline{A}(t,y)\frac{\partial}{\partial y}f(s,x,t,y)+\overline{B}^2(t,y)\frac{\partial^2}{\partial y^2}f(s,x,t,y). \tag{138}$$

To show the identity of this equation with the one derived before, we would have to prove that

$$\overline{B}^2(t,y)=B^2(t,y), \tag{139}$$

$$\overline{A}(t,y)=-A(t,y)+\frac{\partial}{\partial y}B^2(t,y), \tag{140}$$

$$\overline{N}(t,y)=\frac{\partial}{\partial y}A(t,y)+\frac{\partial^2}{\partial y^2}B^2(t,y). \tag{141}$$

§15. Statement of the uniqueness and existence problem of solutions of the second differential equation

In order to define the function $f(s, x, t, y)$ uniquely by the differential equations (125) and (133) we have to set up some initial conditions. For the second equation (133) the following approach can be adopted: according to (85), the function $f(s, x, t, y)$ satisfies the condition

$$\int_{-\infty}^{\infty} f(s, x, t, y) dy = 1 \tag{142}$$

for every $t > s$ and, in view of (110), we also have

$$\int_{-\infty}^{\infty} (y - x)^2 f(s, x, t, y) dy \to 0, \quad \text{as } t \to s. \tag{143}$$

The main question regarding the uniqueness of solutions is as follows: under what conditions can we assert that for given s and x there exists a unique non-negative function $f(s, x, t, y)$ of the variables t, y defined for all y and $t > s$ and satisfying (133) and the conditions (142), (143) ? In certain important particular cases such conditions may be described: these are, for example, all the cases considered in the following two sections.

Now given the functions $A(t, y)$ and $B^2(t, y)$, the question is whether there exists a non-negative function $f(s, x, t, y)$ such that, on the one hand, it satisfies (85) and (86) (as was indicated in §11, these requirements are needed for $f(s, x, t, y)$ to determine a stochastic system), and on the other hand, after passing to the limit via formulas (122) and (124), it gives these functions $A(t, y)$ and $B^2(t, y)$.

To solve this problem, we can, for example, first determine some non-negative solution of our second differential equation (133) satisfying the conditions (142), (143) and then check if it is indeed a solution to our problem. In doing this, the following two general questions arise:

1) Under what conditions does there exist such a solution of equation (133)?

2) Under what conditions does this solution really satisfy (85) and (86)?

There are good grounds for assuming that these conditions are of a sufficiently general character.

§16. Bachelier's case

We now assume that $f(s,x,t,y)$ is a function of the difference $y-x$, depending arbitrarily on s and t, that is, that our process is homogeneous with respect to the parameter

$$f(s,x,t,y) = v(s,t,y-x). \tag{144}$$

In this case, clearly $A(s,t)$ and $B(s,t)$ depend only on s, so that the differential equations (125) and (133) are now expressed as

$$\frac{\partial f}{\partial s} = -A(s)\frac{\partial f}{\partial x} - B^2(s)\frac{\partial^2 f}{\partial x^2}, \tag{145}$$

$$\frac{\partial f}{\partial t} = -A(t)\frac{\partial f}{\partial y} + B^2(t)\frac{\partial^2 f}{\partial y^2}. \tag{146}$$

For the function $v(s,t,z)$, we obtain from (145) and (146):

$$\frac{\partial v}{\partial s} = A(s)\frac{\partial v}{\partial z} - B^2(s)\frac{\partial^2 v}{\partial z^2}, \tag{147}$$

$$\frac{\partial v}{\partial t} = -A(t)\frac{\partial v}{\partial z} + B^2(t)\frac{\partial^2 v}{\partial z^2}. \tag{148}$$

Equation (148) was found by Bachelier,[13] but strictly speaking, was not proved.

If we have $A(t) = 0$ and $B(t) = 1$ identically, then (133) (respectively (146)) turns into the heat equation

$$\partial f/\partial t = \partial^2 f/\partial y^2, \tag{149}$$

for which the only non-negative solution satisfying (142), (143) is given, as is well known, by Laplace's formula

$$f(s,x,t,y) = \frac{1}{\sqrt{\pi(t-s)}}e^{-(y-x)^2/4(t-s)}. \tag{150}$$

In general we assume that

$$x' = x - \int_a^s A(u)du, \quad y' = y - \int_a^t A(u)du,$$
$$s' = \int_a^s B^2(u)du, \quad t' = \int_a^t B^2(u)du.$$

[13] See papers Nos. 1 and 3 in footnote 2.

Then (146) turns into

$$\partial f/\partial t' = \partial^2 f/\partial y'^2,$$

and the conditions (142), (143) retain the same form in the new variables s', x', t', y' as in the variables s, x, t, y. Hence in the general case, the function

$$f(s,x,t,y) = \frac{1}{\sqrt{\pi(t'-s')}} e^{-(y'-x')^2/4(t'-s')} = \frac{1}{\sqrt{\pi\beta}} e^{-(y-\alpha)^2/4\beta}$$

$$\left(\beta = \int_s^t B^2(u)du, \quad \alpha = x + \int_s^t A(u)du\right) \tag{151}$$

is the only solution of equation (146) satisfying our conditions.

§17. A method of transforming distribution functions

Let

$$s' = \phi(s), \quad t' = \phi(t), \quad x' = \psi(s,x), \quad y' = \psi(t,y)$$

$$f(s,x,t,y) = (\partial\psi(t,y)/\partial y) f'(s',x',t',y'), \tag{152}$$

and assume that $\phi(t)$ is a continuous, nowhere decreasing function, whereas $\psi(t,y)$ is arbitrary with respect to t and has a continuous positive derivative with respect to y. If $f(s,x,t,y)$ satisfies (85) and (86), then the same is also true, as can easily be demonstrated, for the function f' with respect to the new variables s', x', t', y'; in other words, our transformation gives a new function $f'(s',x',t',y')$ which, like $f(s,x,t,y)$, determines a stochastic scheme.

If $\phi(t)$ and $\psi(t,y)$ have appropriate derivatives, then under transition to the new variables (125) and (133) turn into

$$\frac{\partial f'}{\partial s'} = -A'(s',x')\frac{\partial f'}{\partial x'} - B'^2(s',x')\frac{\partial^2 f'}{\partial x'^2}, \tag{153}$$

$$\frac{\partial f'}{\partial t'} = -\frac{\partial}{\partial y'}[A'f'] + \frac{\partial^2}{\partial y'^2}[B'^2 f'], \tag{154}$$

where we have set

$$A'(t',y') = \frac{(\partial^2\psi(t,y)/\partial y^2)B^2(t,y) + (\partial\psi(t,y)/\partial y)A(t,y) + \partial\psi(t,y)/\partial t}{\partial\phi(t)/\partial t},$$

$$B'^2(t',y') = \frac{(\partial\psi(t,y)/\partial y)^2 B^2(t,y)}{\partial\phi(t)/\partial t}. \tag{155}$$

With the help of the above transformation the solutions of (133) can be obtained for many new types of coefficients $A(t,y)$ and $B^2(t,y)$. For example, let

$$A(t,y) = a(t)y + b(t), \quad B^2(t,y) = c(t); \tag{156}$$

we set

$$\begin{aligned} \phi(t) &= \int c(t)e^{-2\int a(t)dt}dt, \\ \psi(t,y) &= ye^{-\int a(t)dt} - \int b(t)e^{-\int a(t)dt}dt \end{aligned} \tag{157}$$

and obtain in the new variables s', x', t', y', f' the simplest heat equation:

$$\partial f'/\partial t' = \partial^2 f'/\partial y'^2. \tag{158}$$

In this case the initial conditions (142) and (143) remain valid for $f'(s', x', t', y')$ as well; therefore the formula

$$f' = \frac{1}{\sqrt{\pi(t'-s')}}e^{-(y'-x')^2/4(t'-s')} \tag{159}$$

together with (157) and (152) gives the unique solution $f(s,x,t,y)$ of (133) with coefficients of the form (156) satisfying our conditions. It is easy to see that in this case the function $f(s,x,t,y)$ is of the form

$$\frac{1}{\sqrt{\pi\beta}}e^{-(y-\alpha)^2/4\beta}, \tag{160}$$

where α and β depend only on s, x and t, but not on y.

It is an important problem to find all possible types of coefficients $A(t,y)$ and $B^2(t,y)$ such that for any s, t, x we always obtain a function of the form (160), that is, the Laplace distribution function.

As a second example consider the case

$$A(t,y) = a(t)(y-c), \quad B^2(t,y) = b(t)(y-c)^2. \tag{161}$$

This time, setting

$$\phi(t) = \int b(t)dt, \quad \psi(t) = \ln(y-c) + \int [b(t) - a(t)]dt \tag{162}$$

we again obtain for $f'(s', x', t', y')$ equation (158) for which the solution (159) is already known. Note that here it suffices to consider only the values $x > c$,

$y > c$, since as x or y varies from c to $+\infty$ the variable x' (hence y') runs through all the values from $-\infty$ to $+\infty$. Certain complications arising in connection with this when transferring the conditions (142) and (143) to f' can easily be eliminated.

In particular, for

$$A(t, y) = 0, \quad B^2(t, y) = y^2 \tag{163}$$

we have the formula

$$f(s, x, t, y) = \frac{1}{y\sqrt{\pi(t-s)}} \exp\Big\{-\frac{(\ln y + t - \ln x - s)^2}{4(t-s)}\Big\}. \tag{164}$$

For applications the most important is the case when $A(t, y)$ and $B^2(t, y)$ depend only on y, but do not depend on the time t. The next step in this direction would be to solve our problem for coefficients of the form

$$A(y) = ay + b, \quad B^2(y) = cy^2 + dy + e. \tag{165}$$

§18. Stationary distribution functions

If at time t_0 the differential function of a probability distribution $g(t_0, y)$ is known, then, as for the general formula (5), the distribution function $g(t, y)$ is determined for any $t > t_0$ by the formula

$$g(t, y) = \int_{-\infty}^{\infty} g(t_0, x) f(t_0, x, t, y) dx. \tag{166}$$

Clearly, $g(t, y)$ satisfies the equation

$$\frac{\partial g}{\partial t} = -\frac{\partial}{\partial y}[A(t, y)g] + \frac{\partial^2}{\partial y^2}[B^2(t, y)g]. \tag{167}$$

We now assume that the coefficients $A(t, y)$ and $B^2(t, y)$ depend only on y (the process is homogeneous in time) and study the functions $g(t, y)$ which in this case do not change with time. It is clear that for such functions we have

$$-Ag + (B^2 g)' = C. \tag{168}$$

If we assume that g and g' tend to 0 so rapidly as $y \to \pm\infty$ that the entire left-hand side of (168) tends to 0, then clearly $C = 0$ and we have

$$g'/g = [A - (B^2)']/B^2. \tag{169}$$

Moreover the function $g(y)$ must also satisfy the condition

$$\int_{-\infty}^{\infty} g\,dy = 1. \tag{170}$$

In most cases it appears possible to prove that, if there exists a stationary solution $g(x)$, then $f(s,x,t,y)$ tends to $g(y)$ as $t \to \infty$ and for arbitrary constants s and x; thus, $g(y)$ appears to be not only a stationary, but also the limiting solution.

If the coefficients A and B^2 are of the form (165) , then (169) turns into the Pearson equation

$$\frac{g'}{g} = \frac{y-p}{q_0 + q_1 y + q_2 y^2}, \tag{171}$$

with

$$p = \frac{d-b}{a-2c}, \quad q_0 = \frac{e}{a-2c}, \quad q_1 = \frac{d}{a-2c}, \quad q_2 = \frac{e}{a-2c}. \tag{172}$$

Hence we can construct stochastic schemes for which any of the functions of the Pearson distribution is a stationary solution.

§19. Other possibilities

The theory presented in §§13–18 is essentially determined by the assumption (111). If we get rid of this assumption, then even when the condition (110) is retained, a number of new possibilities appear. For example, consider the scheme determined by the distribution function

$$F(s,x,t,y) = e^{-a(t-s)}\sigma(y-x) + (1 - e^{-a(t-s)}) \int_{-\infty}^{y} u(z)dz, \tag{173}$$

where $\sigma(z) = 0$ for $z < 0$ and $\sigma(z) = 1$ for $z \geq 0$, and $u(z)$ is a continuous non-negative function for which

$$\int_{-\infty}^{\infty} u(z)dz = 1$$

and the moments

$$\int_{-\infty}^{\infty} u(z)|z|^i dz \quad (i = 1,2,3)$$

are finite. It can easily be shown that the function $F(s,x,t,y)$ satisfies (78) and (79), as well as (110).

This scheme can be interpreted as follows: during an infinitely small time interval $(t, t+dt)$ the parameter y either remains constant with probability

$1 - a\,dt$, or takes a value $y', z < y' < z + dz$ with probability $au(z)dt\,dz$. Thus a jump is possible in any time interval, and the distribution function of the values of the parameter after the jump does not depend on the values of this parameter prior to the jump.

This scheme could also be generalized in the following way: imagine that, during an infinitely small time interval $(t, t + dt)$ the parameter y retains its former value with probability $1 - a(t, y)dt$ and turns into y', $z < y' < z + dz$ with probability $u(t, y, z)dt\,dz$. Clearly we assume that

$$\int_{-\infty}^{\infty} u(t, y, z)dz = a(t, y). \tag{174}$$

In this case for $g(t, y)$ the integro-differential equation

$$\frac{\partial}{\partial t} g(t, y) = -a(t, y)g(t, y) + \int_{-\infty}^{\infty} g(t, z)u(t, z, y)dz \tag{175}$$

should hold.

If we wish to consider not only jumps but also continuous changes in y, then it is natural to expect that $g(t, y)$ satisfies

$$\frac{\partial}{\partial t} g(t, y) = -a(t, y)g(t, y) + \int_{-\infty}^{\infty} g(t, z)u(t, z, y)dz - \\ - \frac{\partial}{\partial y}[A(t, y)g(t, y)] + \frac{\partial^2}{\partial y^2}[B^2(t, y)g(t, y)], \tag{176}$$

provided (174) holds and the coefficients $A(t, y)$ and $B^2(t, y)$ are as indicated in §13.

CONCLUSION

If the state of the system under consideration is determined by n real parameters $x_1, x_2, \ldots, x_n$, then under certain conditions similar to those of §13 we have the following differential equations for the differential distribution function $f(s, x_1, \ldots, x_n, t, y_1, \ldots, y_n)$:

$$\frac{\partial f}{\partial s} = -\sum_{i=1}^{n} A_i(s, x_1, \ldots, x_n)\frac{\partial f}{\partial x_i} - \sum_{i=1}^{n}\sum_{j=1}^{n} B_{ij}(s, x_1, \ldots, x_n)\frac{\partial^2 f}{\partial x_i \partial x_j}, \tag{177}$$

$$\frac{\partial f}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial y_i}[A_i(t, y_1, \ldots, y_n)f] + \\ + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial y_i \partial y_j}[B_{ij}(t, y_1, \ldots, y_n)f]. \tag{178}$$

For the case when $A_i(t, y_1, \ldots, y_n)$ and $B_{ij}(t, y_1, \ldots, y_n)$ depend only on t, these equations were discovered and solved by Bachelier.[14] In this case the solutions satisfying the conditions of our problem have the form

$$f = P \exp\left\{-\frac{1}{Q} \sum p_{ij}(y_i - x_i - q_i)(y_j - x_j - q_j)\right\}, \tag{179}$$

with P, Q, p_{ij} and q_i depending only on s and t.

It is also possible to consider mixed schemes, where the state of the system is determined by parameters some of which are discrete and others continuous.

Moscow, 26 July 1930

[14] See item II in footnote 2.