ON THE THEORY OF STOCHASTIC PROCESSES, WITH PARTICULAR REFERENCE TO APPLICATIONS

W. FELLER

CORNELL UNIVERSITY

(ONR. Project for Research in Probability)

1. Introduction

Since Kolmogoroff's famous paper of 1931, "On Analytical Methods in the Theory of Probability," the theory of stochastic processes has been developed and it has been shown that it can successfully be applied to practical problems and used to describe empirical phenomena. However, the theory is new and the most appropriate mathematical techniques have yet to be discovered. It is therefore reasonable to expect that the usefulness of the theory will increase when more pertinent mathematical problems are solved. On the other hand, these new problems are of interest also in pure analysis beyond the theory of stochastic processes. In the past, pure mathematics has always derived great benefits from the interplay with physical theories, and many parts of purest mathematics owe their origin to physical problems. Now we shall see that our theory leads to integrodifferential equations of a type never studied before: they contain as simplest special cases a surprisingly great variety of familiar and unfamiliar functional equations (sec. 7). It seems probable that our methods of deriving solutions and adjoint equations could be utilized also for many functional equations without probabilistic meaning.[1] Another example of a problem of general interest on which we shall touch briefly (sec. 10) is connected with the fact that an empirical phenomenon can often be described in several ways, say by a system of infinitely many ordinary differential equations or by a partial differential equation of the diffusion type. This seems to indicate connections which have yet to be explored.

As for practical usefulness, it should be borne in mind that for a mathematical theory to be applicable it is by no means necessary that it be able to provide accurate models of observed phenomena. Very often in applications the constructive role of mathematical theories is less important than the economy of thought and experimentation resulting from the ease with which qualitatively reasonable working hypotheses can be eliminated by mathematical arguments. Perhaps even more important is the constant interpretation of observations in the light of theory and of theory in the light of observations; in this way mathematical theory can become an indispensable guide not only to a better understanding, but even to a proper formulation of scientific

[1] The integrodifferential equations mentioned in the text contain as a special case, among others, infinite systems of ordinary differential equations, which will be studied in section 4. The theory of these ordinary differential equations has been generalized also to cases which are void of probabilistic meaning [cf. Arley and Borchsenius (1945)].

problems. For example, in geology we are confronted with random processes which have been going on for millions of years, some of them covering the surface of the earth. We observe that certain species go through a period of prosperity and steady increase, only to die out suddenly and without apparent reason. Is it really necessary to introduce new hypotheses for each new observation, to assume cataclysms working one-sidedly against certain species, or to find other explanations? The Volterra-Lotka theory of struggle for existence teaches us that even under constant conditions situations are bound to arise which would appear to the naive observer exactly like many of the cataclysms of geology. Similarly, although it is impossible to give an accurate mathematical theory of evolution, even the simplest mathematical model of a stochastic process, together with observations of age, geographical distribution, and sizes of various genera and species, makes it possible to deduce valuable information concerning the influence on evolution of various factors such as selection, mutation, and the like.[2] In this way undecisive qualitative arguments are supplemented by a more convincing quantitative analysis.

In the mathematical literature stochastic processes are usually treated in a formal and general way which does not clearly show the practical meaning and applicability. On the contrary, practical problems leading to stochastic processes are usually treated by special methods and under various disguises so that the connection with the general theory does not become apparent. It seems therefore desirable to begin (sec. 2) by explaining the very simplest and most intuitive examples of stochastic processes which have been treated in the literature. They do not require new mathematical tools, as they all lead to systems of ordinary differential equations which, though infinite, are of a very simple form. We shall then pass to more general theories, but it should be understood that we shall not in this paper consider the most general type of stochastic processes such as occurs in the analysis of time series. Instead, we shall restrict the considerations to what is now generally called Markov processes, that is to say to processes where all future probability relations are fully determined by the present state, in the same way as in classical mechanics the present state uniquely determines the future development of a system. Furthermore, we shall focus our attention on the so-called discontinuous type of Markov processes where the changes occur in jumps: the system remains for a time unchanged, then undergoes a sudden change into another state. This type of process has found extensive and important applications in the theory of automatic telephone exchanges and in the theory of risk for insurance companies.[3] The continuous type of process is best exemplified by diffusion processes which are continuous in the sense that some change occurs during any time interval, however small, but in small time intervals the changes are small. This type of process is of special importance in physics,

[2] For details, cf. Yule (1924). It is evident that at that time Yule could not use the formalism of stochastic processes (which would have simplified some of his considerations).

[3] Elaborate applications of the theory of stochastic processes to telephone exchanges will be found in C. Palm's book of 1943. The so-called collective theory of risk has been initiated by F. Lundberg and continued by H. Cramér and his collaborators. An account will be found in Segerdahl's book (1939). These by no means simple mathematical theories have found extensive applications in practice.

and two excellent accounts have recently been given.[4] The connection between the two types will be touched upon in the last section of the present paper.

2. The simplest examples

(1) *The Poisson process.*—This is quite familiar to physicists, who often refer to it as "random events" and occasionally call the Poisson distribution after Bateman. Here we consider this example as a point of departure for various generalizations and also to emphasize that the all-important Poisson distribution appears in its own rights and not merely as an approximation to the binomial distribution.

Consider, in order to fix ideas, events occurring in time, such as telephone calls, radioactive disintegrations, impacts of particles (cosmic rays), and the like. Let it be assumed that (i) the probability of an event in any time interval of length dt is, asymptotically, ηdt, where η is a positive constant; (ii) the probability of more than one event in a time interval dt is of smaller order of magnitude than dt, in symbols $o(dt)$; (iii) the numbers of events in non-overlapping time intervals represent independent random variables.

Denote, then, by $P_n(t)$ the probability of having exactly n events in a time interval of length t. We want to compare $P_n(t)$ with $P_n(t + dt)$. Now $n \geqq 1$ events can occur in the interval $(0, t + dt)$ in one of several ways: either n events occur in $(0, t)$ and none in $(t, t + dt)$; or $n-1$ events occur in $(0, t)$ and one in $(t, t + dt)$; or, finally, less than $n - 1$ events occur in $(0, t)$ and more than one in $(t, t + dt)$. Writing down the corresponding probabilities we find

$$P_n(t + dt) = (1 - \eta dt)P_n(t) + \eta dt P_{n-1}(t) + o(dt), \tag{1}$$

and similarly

$$P_0(t + dt) = (1 - \eta dt)P_0(t) + o(dt). \tag{1'}$$

Rearranging these equations and passing to the limit we find easily that our probabilities satisfy the system of differential equations

$$\begin{cases} P'_0(t) = -\eta P_0(t), \\ P'_n(t) = -\eta P_n(t) + \eta P_{n-1}(t), \qquad n \geqq 1. \end{cases} \tag{2}$$

The initial conditions are obviously

$$P_0(0) = 1, \qquad P_n(0) = 0, \qquad n \geqq 1. \tag{3}$$

Fortunately, in this case, the differential equations are of a recursive character and can be solved successively. The required solution is given by the familiar Poisson distribution

$$P_n(t) = e^{-\eta t}\frac{(\eta t)^n}{n!}, \qquad n \geqq 0; t \geqq 0. \tag{4}$$

[4] Cf. Chandrasekhar (1943); Wang and Uhlenbeck (1945).

If $\mathbf{N}_t$ denotes the number of events during $(0, t)$, then the expected value $E(\mathbf{N}_t) = M(t)$ and the variance $E(\mathbf{N}_t^2) - M^2(t) = \sigma^2(t)$ satisfy differential equations which can be deduced directly from (2). For comparison we quote here the familiar fact that

$$M(t) = \eta t, \qquad \sigma^2(t) = \eta t. \tag{5}$$

It is important to notice that the applicability of the Poisson distribution (4) is far wider than would appear on the surface. It describes adequately many phenomena which play in space rather than in time; in such cases the parameter t stands for volume, area, or length, instead of time. With such an interpretation of t it will be seen that our assumptions which led to the differential equations (2) hold for many "events," such as the distribution of stars in space, flaws in a material, raisins in a cake, misprints in a book. In such cases we shall refer to the parameter t as "operational time." We shall see that many stochastic processes do not necessarily play in time, but that the operational time may be anything, such as depth of penetration, force, and the like.

Before passing to other examples we shall introduce the terminology which is most convenient for the general theory, although it sounds somewhat artificial in the special case of the Poisson distribution. Instead of saying that during the time interval $(0,t)$ exactly n events have occurred, we shall say that "the system is in state E_n." Instead of saying that at time t an "event" has occurred, we shall say that "the system has passed from state E_n to E_{n+1}."

(2)*Radioactive disintegrations.*—The terminology is well illustrated by the successive transitions of a radioactive atom. We may say that the atom passes from one state to another, in symbols $E_0 \to E_1 \to \cdots \to E_N$. Of course this time we have only a finite number of possible states. Moreover, in the Poisson case the probabilities of transition were the same for all states. According to accepted theories, if the atom is in state E_i it has a probability asymptotically equal to $\eta_i\, dt$ to disintegrate in the following time interval of length dt; this probability does not depend on the age or history of the atom, but changes from state to state. Clearly this case can be treated exactly as the Poisson case, only that this time the factors η will depend on the state. The corresponding differential equations now take on the form

$$\begin{cases} P'_0(t) = -\eta_0 P_0(t), \\ P'_n(t) = -\eta_n P_n(t) + \eta_{n-1} P_{n-1}(t), \qquad 1 \leqq n \leqq N; \eta_N = 0. \end{cases} \tag{6}$$

The initial conditions are of course the same as before. The system (6) is frequently found in the literature and the explicit solution has been given (and is being given) independently by many authors. If we suppose that no two among the η_i are equal, this solution can be written in the form

$$P_n(t) = (-1)^n \eta_0 \eta_1 \cdots \eta_{n-1} \times$$

$$\sum_{k=0}^{n} e^{-\eta_k t} \{(\eta_k - \eta_0) \cdots (\eta_k - \eta_{k-1})(\eta_k - \eta_{k+1}) \cdots (\eta_k - \eta_n)\}^{-1}. \tag{7}$$

These formulas are rather unwieldy and it is usually simpler to obtain pertinent information directly from the differential equations (6). For example, it is easy to see that for $1 \leqq n \leqq N - 1$ the function $P_n(t)$ first increases monotonically to a unique maximum and then decreases steadily to zero. If t_n denotes the place of this maximum, then $t_1 < t_2 < \cdots < t_{N-1}$. For $t < t_n$ one has $\eta_n P_n(t) < \eta_{n-1} P_{n-1}(t)$, whereas for $t > t_n$ the reverse inequality holds. These facts check and explain experimental observations.

(3) *Growth without absorption.*—The differential equations (6) and their solution (7) retain sense also in case $N = \infty$. This case is met with in practice in connection with all kinds of physical or biological growth processes; the state E_n denotes then the number of "individuals" in the "population." Since only transitions $E_n \to E_{n+1}$ are possible, individuals can only multiply, not die or be absorbed. The factors η_n determine the intensity of multiplication peculiar to the state E_n. The simplest case is that where the individuals multiply independently of one another and each individual has in any time interval of length dt a chance ηdt to split into two. Because of the assumed independence we have then

$$\eta_n = n\eta, \tag{8}$$

and (6) becomes

$$P'_n(t) = -n\eta P_n(t) + (n-1)\eta P_{n-1}(t), \qquad n > 1, \tag{9}$$

usually with the natural initial condition

$$P_1(0) = 1, \qquad P_n(0) = 0, \qquad n > 1. \tag{10}$$

The solution follows either from (7) or by direct integration:

$$P_n(t) = e^{-\eta t}(1 - e^{-\eta t})^{n-1}. \tag{11}$$

This type of process has been used (in an indirect way) by Yule in his mathematical theory of evolution, to which we have referred in the introduction. There the population consists of the species within a genus, or of the various animal or vegetable genera. The multiplication is due to mutation, each species or genus having at any time a (constant) probability of throwing a new species or genus. The theory is used to study relations between the age of genera, the number of composing species, their geographical distribution, and the like. The conclusions permit us to confirm or reject various assumptions concerning evolution. The same type of process has been used by Furry in the theory of cosmic ray showers; Arley (1943) calls (11) the Furry process. The population there consists of electrons or of photons, and the multiplication is actually a complex event (each electron can emit a photon, each photon can be absorbed emitting two electrons; the differential equations then refer only to each second generation). The same equations have been used by Feller (1939) to

study fluctuations in biological growth. In all these cases it would be more natural to take into account the possibility of an individual dying at any time.

The mathematical expectation of the size of the population, that is,

$$M(t) = \sum nP_n(t), \tag{12}$$

can be obtained directly from (9): multiplying by n and adding, one sees that $M'(t) = \eta M(t)$, or

$$M(t) = e^{\eta t}. \tag{13}$$

For the variance

$$\sigma^2(t) = \sum n^2 P_n(t) - M^2(t), \tag{14}$$

one obtains similarly

$$\sigma^2(t) = M^2(t) - M(t), \tag{15}$$

a relation which permits comparison with actual observations [cf. Arley (1943)].

Clearly our process is the statistical counterpart of the deterministic process of growth described by the equation

$$\dot{x}(t) = \eta x(t). \tag{16}$$

In the same way it is possible to translate other differential equations describing a deterministic growth process into an analogous random process. For example, to the differential equation

$$\dot{x}(t) = \eta x^2(t) \tag{17}$$

there corresponds a random process described by the system (6) with $\eta_n = n^2\eta$. Now a population growing according to (17) would increase indefinitely within a finite time interval; the solution

$$x(t) = (c - \eta t)^{-1} \tag{18}$$

"explodes" at $t = c/\eta$. One should therefore assume that a differential equation of type (17) could not naturally describe a growth process. Accordingly, it is of interest to study the corresponding anomaly for the system (6) and to find out under what conditions this system can describe a regular growth. It turns out that every time the series

$$\sum \eta_n^{-1} \tag{19}$$

converges the solutions $P_n(t)$ of (6) have the undesirable property that

$$\sum P_n(t) < 1. \tag{20}$$

This can be interpreted by saying that there is a finite probability that infinitely many jumps will occur during a finite time interval, or that the system "explodes" in a way similar to the singularity of (18). On the contrary,[5] if the series (19) diverges, the sum (20) equals unity and the probability distribution behaves properly in every respect. Similar situations arise in connection with more general processes.

Equations (9) and (16) correspond to a growth where there is no interaction whatsoever between the individuals. The so-called logistic law of growth assumes that the size of the population has an adverse influence on the reproduction rate of the individual, and accordingly one assumes that the intensity functions η_n are of the form

$$\eta_n = n\eta - n^2\zeta, \tag{21}$$

where η and ζ are positive constants. It is generally believed that the corresponding deterministic equation

$$\dot{x}(t) = \eta x(t) - \zeta x^2(t) \tag{22}$$

gives the expected value of the process (6) in the same way as (16) really gives the expected value of the undisturbed growth (9). In reality this is not true and the latter quantity is somewhat smaller than the solution $x(t)$ of (22). The fact that (16) actually gives the expected value for the process (9) must be looked upon as a coincidence (which is very fortunate for the theory of radioactive decay). Similar remarks hold for the more refined problems of the Volterra-Lotka theory.

(4) *Growth with absorption.*—The possibility of an individual dying or of a particle being absorbed is taken into account by assuming that not only transitions $E_n \to E_{n+1}$ are possible, but also the reverse transitions $E_n \to E_{n-1}$. Let us, for example, suppose that in any small time interval of length dt any individual has a chance $\gamma dt + o(dt)$ of being absorbed, and that the intensity γ does not depend on the actual state of the system. An obvious change in the argument leading to the Poisson distribution and to (9) shows that the probabilities $P_n(t)$ now satisfy the system of infinitely many equations

$$\begin{cases} P'_0(t) = \gamma P_1(t), \\ P'_n(t) = -n(\eta + \gamma)P_n(t) + (n-1)\eta P_{n-1}(t) \\ \qquad\qquad + (n+1)\gamma P_{n+1}(t), \qquad n > 0. \end{cases} \tag{23}$$

This system presents a novelty, inasmuch as it is not of recursive type and its integration therefore is not obvious. It can be accomplished by introducing the generating function.

$$P(t,x) = \sum P_n(t)x^n,$$

[5] Feller (1940). Recent results of Doob (1945) reveal the curious fact that, whenever (19) converges and therefore (20) holds, the differential equations have infinitely many solutions which are meaningful as probabilities. They correspond to a ghostlike return of the system into a finite state after an "explosion" has taken place.

for which it is easy to find a partial differential equation of the first order. If at the beginning the population consists of a single individual, that is, if $P_1(0) = 1$, the solution is given[6] by

$$P_0(t) = \gamma\sigma, \qquad P_n(t) = (1 - \eta\sigma)(1 - \gamma\sigma)(\eta\sigma)^{n-1}, \tag{24}$$

where for abbreviation

$$\sigma = \{1 - e^{(\eta-\gamma)t}\}\ \{\gamma - \eta e^{(\eta-\gamma)t}\}^{-1}. \tag{25}$$

In the particular case $\gamma = \eta$, equations (24) have to be replaced by

$$P_0(t) = \eta t\{1 + \eta t\}^{-1}, \qquad P_n(t) = (\eta t)^{n-1}\{1 + \eta t\}^{n+1}. \tag{26}$$

It is interesting to note that the probability of extinction is

$$\lim P_0(t) = \begin{cases} 1, & \text{if } \gamma \geqq \eta, \\ \dfrac{\gamma}{\eta}, & \text{if } \gamma \leqq \eta, \end{cases} \tag{27}$$

whereas for every fixed $n > 0$ the probability $P_n(t) \to 0$. Thus, roughly speaking, for large values of t the population will either have died out or be exceedingly large, whereas there is no probability for moderate sizes. The same phenomenon is known in connection with the problem of survival of family names. For the expectation $M(t)$ and variance $\sigma^2(t)$ [cf. (13) and (15)] we obtain easily

$$M(t) = e^{(\eta-\gamma)t}, \qquad \sigma^2(t) = \{M^2(t) - M(t)\}\frac{\eta + \gamma}{\eta - \gamma}. \tag{28}$$

It is seen that absorption increases the magnitude of the probable random fluctuations.

(5) *The Arley process for cosmic rays.*—The simplified model (23) does not adequately describe the phenomenon of cosmic ray showers. With (23) the expected number of particles would either steadily increase or steadily decrease, whereas actually the number of electrons first rapidly increases but then decreases, since with increasing depth of penetration the particles lose energy and are more rapidly absorbed. Also, as Arley remarked, the actual fluctuations are much larger than would correspond to (23). The actual probabilities of multiplication and absorption for each particle depend on its energy; to describe this situation adequately, more complicated models of stochastic processes are required, where the state of the system is no longer described by a simple integer. However, as Arley (1943) has shown, satis-

[6] Equations (23) have been considered by Feller (1939), by Arley (1943), and (in an indirect way) by Yule (1924). The solution (24) is due to C. Palm and has been communicated without proof by Arley and Borchsenius (1945).

factory results can be achieved in the following simple manner. We assume that the probability for absorption for each particle increases linearly in time, whereas the probabilities for multiplication remain constant. Then the intensities η_n and γ_n assume the forms $n\eta$ and $n\gamma t$, respectively, and (23) is replaced by

$$P'_n(t) = -n(\eta + \gamma t)P_n(t) + (n-1)\eta P_{n-1}(t) + (n+1)\gamma t P_{n+1}(t). \quad (29)$$

This is the first example for a non-stationary process. For the expected value of the number of particles, one obtains easily

$$M(t) = \exp(\eta t - \frac{\gamma}{2}t^2), \quad (30)$$

with a maximum at $t = \eta/\gamma$. The method of the generating function described above [cf. subsec.(4) above] leads also to the following explicit solution:[7]

$$\begin{cases} P_o(t) = 1 - \exp(\eta t - \frac{\gamma}{2}t)^2, & (31) \\ P_n(t) = \exp(-\eta t + \frac{\gamma}{2}t^2)\, A^{n-1}\{A + \exp([\gamma t - \eta]^2/2\gamma)\}^{-(n+1)}, \end{cases}$$

where

$$A = \eta \int_0^t \exp\left(\frac{[\gamma s - \eta]^2}{2\gamma}\right) ds. \quad (32)$$

(6) *The Pólya process.*—It is well known that Pólya has devised a scheme of urns which in the simplest way illustrates the occurrence of contagious events, in the sense that with Pólya's scheme each favorable event increases the probability of succeeding favorable events. Pólya's probability distribution has proved exceedingly useful and serves as the prototype of contagious distributions. O. Lundberg (1940) has shown that the full power and flexibility of the Pólya distribution appears only if one passes from discrete drawings from an urn to a stochastic process with a continuous time parameter.

In the original scheme, drawings are made from an urn containing Np white and Nq black balls ($p + q = 1$). After each drawing the ball is replaced and, in addition, $N\sigma$ balls of the color last drawn are added to the urn. If X is the number of white balls in n successive drawings, then

$$Pr(X=k) = \binom{n}{k} \frac{p(p+\delta)(p+2\delta)\cdots(p+[k-1]\delta)q(q+\delta)\cdots(q+[n-k-1]\delta)}{1(1+\delta)(1+2\delta)\cdots(1+[n-1]\delta)}, \quad (33)$$

[7] Not published.

and an easy computation shows that

$$E(X) = np, \qquad \sigma^2(X) = npq(1 + n\delta)\,(1 + \delta).^{-1} \tag{34}$$

If we assume that in the first n drawings exactly k white balls have been drawn (that is, that $X = k$), the probability of obtaining a white ball in the next drawing is

$$\frac{p + k\delta}{1 + n\delta} \tag{35}$$

One could pass from these formulas to the limit much in the same way as is customary with the derivation of the Poisson formula from the binomial distribution. One would imagine that within a finite time interval $(0,t)$ a very large number, n, of drawings is made but that the numbers p and δ are very small. In order to obtain a reasonable passage to the limit one puts

$$\lim np = t, \qquad \lim n\delta = td, \tag{36}$$

where d is the new parameter of contagion. We obtain then from (33) the following distribution, depending on the continuous time parameter t and on the parameter of contagion d,

$$\begin{cases} P_0(t) = (1 + dt)^{-1/d}, \\ P_k(t) = \left(\dfrac{t}{1 + td}\right) \dfrac{k(1 + d)(1 + 2d) \cdot \cdot \cdot (1 + [k - 1]d)}{k!} P_0(t). \end{cases} \tag{37}$$

Now it seems simpler and more natural to start directly with this process and to derive its differential equations in the same way as we have obtained the Poisson distribution, without a passage to the limit. Let the "state" denote the number of events during $(0,t)$. Then only passages $E_k \to E_{k+1}$ are possible, and the process is determined by some differential equations of the form (6). Formula (35) suggests putting

$$\eta_k = \frac{1 + kd}{1 + td}. \tag{38}$$

It is readily verified that the probability distribution (37) is really the solution of (6) with (38) and thus represents the continuous limiting case of the Pólya scheme.

It is clear that (38) has been chosen in accordance with the original Pólya scheme but represents only one of the many possible choices for the intensity functions. From a purely abstract point of view there is nothing remarkable in the Pólya scheme, but this does not detract from its value. It has been selected judiciously as the simplest scheme of contagion that lends itself to practical purposes. One of the reasons why it is so practicable is the fact that

with (37) the first and the second moment can be fitted separately to empirical observations. Interesting material illustrating this fact will be found in O. Lundberg's book, where the distribution (37) has been derived and applied to accident and sickness statistics. Arley (1943) used the same distribution as an approximation to theoretical distributions, which are less usable in practice.

3. The nature of contagion

Pólya's original urn scheme (as described in the last section) clearly shows a true contagion, and it has become customary to assume that a kind of actual contagion exists in nature whenever Pólya's or a similar distribution shows an excellent fit to observations. As O. Lundberg has pointed out, there is no justification for this assumption. The argument actually implies that the distribution (37) is "contagious" simply because it satisfies the equations (6) with non-constant coefficients (38), or, in other words, because the transition probabilities depend on previous happenings. In the same sense all the probability distributions derived above should be (and usually are) classified as "contagious." However, the equations for our growth processes have been derived from the fundamental hypothesis that there is no contagion whatsoever, each individual being absolutely independent of the others and of the past. It is interesting to notice that even the Pólya distribution (37) admits of an interpretation which *excludes* any notion of contagion. In fact, it has in this form been anticipated by Greenwood and Yule (1920) in their studies on industrial accident statistics.

Suppose, in order to fix the ideas, that each individual in a large population is liable to accidents: within any short time interval of length dt each individual has a probability asymptotically equal to ηdt to sustain an accident. The parameter η is characteristic for the individual and remains constant in time. The number of accidents of any particular individual during $(0,t)$ is then a random variable distributed according to the Poisson distribution. Now different individuals are characterized by different parameters η, that is to say, they exhibit a variable proneness to accidents. Accordingly, the parameter η becomes itself a random variable, and the probability that an individual taken at random from the entire population should have k accidents during $(0,t)$ is given by

$$P_k(t) = \int_0^\infty e^{-\eta t} \frac{(\eta t)^k}{k!}\, dU(\eta), \tag{39}$$

where $U(\eta)$ denotes the distribution function of the random variable η. Greenwood and Yule assumed in particular that η has a Pearson Type III distribution:

$$U'(\eta) = \frac{c^r}{\Gamma(r)} \eta^{r-1} e^{-c\eta}, \qquad r, c, \eta > 0. \tag{40}$$

In that case, (39) reduces to the Pólya distribution (37).[8]

[8] Cf. O. Lundberg (1940); also Feller (1943).

With the "contagious" interpretation of (37) one is led to believe that for each individual each accident increases the probability of another accident. With the new interpretation this probability remains absolutely constant, an accident in no way influencing the probability of further accidents. It is not now known which of the models of stochastic processes admit of a similar double interpretation and which (if any) necessarily mean true contagion. It seems that none of the distributions which are now used in the literature permits the conclusion that the phenomenon described is contagious.

As an amusing example let us consider football, where the observed number of goals in individual games shows a clearly "contagious" distribution. This could be interpreted as true contagion, where each success increases the probability of further successes; or by assuming that for each team the probabilities remain constant but that the skill (= parameter in a Poisson distribution) varies from team to team. In most cases a combination of the two interpretations would probably come nearest to the truth. Incidentally, the treatment of statistical problems connected with the game of cricket has been suggested by Seal, in his discussion of the paper by Elderton (1945).

4. Markov processes leading to ordinary differential equations

The considerations of the second section can easily be generalized to the case where not only transitions from a state E_k to the neighboring states E_{k+1} and E_{k-1} are possible but also transitions from any E_i to any E_k. For example, in the statistical theory of strength of material one considers[9] bundles of threads under tension. The strength of an individual thread is a random variable. In state E_k the bundle consists of k threads, and it is assumed that each carries the same load. The load plays the role of operational time and will be denoted by t. As t increases, a moment will be reached where the weakest among the k constituent threads will break. Now this does not necessarily mean a transition $E_k \to E_{k-1}$. In fact, if $k - 1$ threads remained, each of them would now carry the load $t/(k - 1)$, and it is possible that the weakest among them is too weak to carry this load. It may happen that among the $k - 1$ remaining threads one is too weak to carry the load $t/(k - 1)$, another too weak to carry $t/(k - 2)$, but that there are $k - 3$ threads able to support the load $t/(k - 3)$. In this case the bundle will pass from state E_k to E_{k-3}, and similarly the passage to any E_j with $j < k$ is theoretically possible. In connection with a discussion of Pareto's law of income distribution we shall find it natural to consider the changes of income of an individual as a random process. If the possible incomes are divided into classes $E_1, E_2, \cdots$, any transition $E_i \to E_j$ becomes imaginable.

For the general theory it is convenient to change the notation slightly and to denote by $P_{\nu k}(\tau,t)$ the (conditional) probability of the system to be at time t in state E_k if at the time $\tau < t$ it has been in state E_ν. Except for the last two all examples of the second section were stationary in time so that their transition probabilities $P_{\nu k}(\tau,t)$ would depend only on the difference $t - \tau$. With the present notations the P_n of these examples would be denoted by P_{0k} or P_{1k}, respectively.

[9] Cf. Daniels (1945).

A process of random transitions from one of the states to another is called a Markov process if the transitions in non-overlapping time intervals are statistically independent (so that corresponding probabilities multiply). Consider then an arbitrary moment s, $(\tau < s < t)$. At time s the system is in some state E_j, and a transition $E_\nu \to E_j \to E_k$ has, by assumption, probability $P_{\nu j}(\tau,s)P_{jk}(s,t)$.
Therefore

$$P_{\nu k}(\tau,t) = \sum_j P_{\nu j}(\tau,s)P_{jk}(s,t). \tag{41}$$

This is the fundamental identity which may be taken as the description of Markov processes with a simple sequence of states; it is generally known as the Chapman-Kolmogoroff equation.

Now we can proceed as before and deduce from (41) differential equations for $P_{\nu k}(\tau,t)$. For that purpose we have again to assume that certain elementary probabilities are known: (*i*) If at time t the system is in state E_k it will be assumed that the probability for any change during $(t, t + dt)$ is $p_k(t)dt + o(dt)$. The function $p_k(t)$ will be referred to as intensity for the state E_k. (ii) If at time t the system is in E_k and if a change occurs, the (conditional) probability that this change will take the system into E_n will be denoted by $\Pi_{kn}(t)$. Obviously the conditions

$$p_k(t) \geqq 0, \qquad \Pi_{kn}(t) \geqq 0, \quad \Pi_{kk}(t) = 0, \qquad \sum_n \Pi_{kn}(t) = 1 \tag{42}$$

must be satisfied. Otherwise the $p_k(t)$ and the Π_{kn} can be chosen arbitrarily. The two assumptions can be combined to read

$$\begin{cases} P_{kk}(t, t + dt) = 1 - p_k(t)dt + o(dt), \\ P_{kn}(t, t + dt) = p_k(t)\Pi_{kn}(t)dt + o(dt). \end{cases} \tag{43}$$

Following Kolmogoroff (1931), it is now easy to deduce two systems of differential equations for the $P_{\nu k}(\tau,t)$. In the "forward equations" both ν and τ are fixed and only k and t are the variables. In a formal way they are obtained from (41), putting $s = t - dt$, substituting from (43), and passing to the limit. One obtains the ordinary differential equations

$$\frac{\partial}{\partial t} P_{\nu k}(\tau,t) = - p_k(t)P_{\nu k}(\tau,t) + \sum_j p_j(t)\Pi_{jk}(t)P_{\nu j}(t). \tag{44}$$

The parameters ν and τ appear only in the initial conditions, which are obviously

$$P_{\nu k}(\tau,\tau) = \begin{cases} 1, & \text{if } k = \nu, \\ 0, & \text{if } k \neq \nu. \end{cases} \tag{45}$$

Conversely, the "backward equations" contain k and t only as parameters and refer to ν and τ as variables. They follow from (41) on putting $s = \tau + d\tau$ and passing formally to the limit:

$$\frac{\partial}{\partial \tau} P_{\nu k}(\tau,t) = p_\nu(\tau)\left\{P_{\nu k}(\tau,t) - \sum_j \Pi_{\nu j}(\tau) P_{jk}(\tau,t)\right\}. \tag{46}$$

The initial condition is again (45), with τ replaced by t. The forward system appears more natural from a physical point of view, but the backward system is occasionally preferable for technical reasons. It plays also an important role in connection with questions of reversibility. In the terminology of the theory of differential equations, each of the systems is the adjoint of the other.

As an example, consider the Poisson process. With the present notations its transition probabilities are

$$P_{\nu k}(\tau,t) = \begin{cases} e^{-\eta(t-\tau)} \dfrac{\{\eta(t-\tau)\}^{k-\mu}}{(k-\mu)!}, & \text{if } k \geqq \mu, \\ 0, & \text{if } k < \mu. \end{cases} \tag{47}$$

By definition of the process $p_k(t) = \eta$, $\Pi_{k,k+1}(t) = 1$, whereas all other $\Pi_{kn}(t) = 0$. The two systems satisfied by (47) are

$$\begin{cases} \dfrac{\partial}{\partial t} P_{\nu k}(\tau,t) = -\eta\{P_{\nu k}(\tau,t) - P_{\nu,k-1}(\tau,t)\}, \\ \dfrac{\partial}{\partial \tau} P_{\nu k}(\tau,t) = \eta\{P_{\nu k}(\tau,t) - P_{\nu+1,k}(\tau,t)\}. \end{cases} \tag{48}$$

In the general case the situation is of course less obvious.

The two systems of differential equations could also be interpreted mechanically by considering containers E_i and a liquid flowing from one to the other through channels whose conductivity is regulated by the coefficients of the equations. This interpretation is natural at least in the stationary case where all the coefficients are constants.

It has been shown[10] that each of the two systems of infinitely many ordinary differential equations (44) and (46) has a solution which is automatically a solution of the other and of the fundamental equation (41). In all cases of practical interest this solution is unique.[11] However, there exist particular processes where a phenomenon of "explosion" can occur, completely analogous to that discussed in connection with the example, given in subsection (2) above. Necessary and sufficient conditions for the possibility of such an occurrence have also been found (loc. cit.). Incidentally, it will be seen in section 6

[10] Cf. Feller (1940).

[11] There is in general no uniqueness if the consideration is not restricted to solutions which are meaningful from a probability standpoint. Cf. also footnote 5.

that our differential equations are only a very special case of a couple of integrodifferential equations describing the general discontinuous Markov process.

5. Ergodicity

The classical counterparts of our random processes are the discrete Markov chains, that is to say, random operations such as shuffling of cards, random walk, transferring a molecule from one container to another, and the like. With each operation the system passes from one state to another, but the operations are performed at arbitrary moments. With shuffling cards it seems plausible, and desirable, that after a large number of operations all permutations become essentially equally probable, no matter what the initial arrangement has been. If there are infinitely many possible states, one cannot reasonably expect that all of them will in the limit become equally probable. Instead, we shall say that a Markov process with transition probabilities $P_{\nu k}(\tau,t)$ is regularly erogodic if

$$\lim P_{\nu k}(\tau,t) = P_k \tag{49}$$

exists and is independent of ν and τ; in this case the probability distribution for the states tends to a uniquely determined limiting distribution which is independent of the initial state. We shall apply this notion in particular to stationary processes where the transition probabilities depend only on the difference $t - \tau$. It should be noticed that the constants P_ν do not necessarily add to unity. This fact is illustrated by the limiting distribution (27). In most physical and technical applications, however, $\Sigma P_\nu = 1$, and the P_ν represents the steady state for the system. A non-trivial example connected with a waiting time problem will be found in section 8.

The main headache with the classical theory was to establish conditions under which a Markov chain is ergodic. The trouble is due to the fact that in the discrete case even quite respectable chains are not ergodic but exhibit periodic fluctuations. For example, a Raileigh-Pearson random walk consists of a series of operations each of which carries the moving particle a unit step to the right or to the left. If the state of the system denotes the distance of the particle from its original position, an odd number of operations will carry the system into an odd-numbered state, an even number into an even state. Therefore the system cannot be ergodic. In view of such difficulties it is remarkable that stochastic processes with a continuous time parameter behave much more simply. *Any* Markov process defined by a system of differential equations (44) or (46) with constant coefficients will be regularly ergodic, provided that it is at all possible to reach (in one or more steps) any state from any state.[12] More particularly, only ergodic discrete Markov chains can be interpolated by processes with continuous time parameter.[13]

[12] The latter condition is by no means necessary and is given only for simplicity. However, some restriction is obviously required in order to exclude trivial exceptional cases without practical meaning, such as systems where the possible states split into two groups without

What ergodicity might possibly mean in practice may be illustrated by the much discussed Pareto law of income distribution. According to this law there should exist a universal distribution function, depending on one or two parameters, which (after an appropriate adjustment of the parameters) would describe the income distribution within any economic unit (country). It appears that at present opinions regarding existence of such a mysterious function are, less divided than opinions concerning its precise analytic form. Now, if the transition of an individual from one finite income class to another is considered as a random process which is stationary or almost stationary, the ergodic properties would, to begin with, ensure the existence of a limiting distribution. Moreover, it seems plausible that this stable limiting distribution would to a certain degree be insensitive to the changes in numerical values of the coefficients p_k, Π_{kn} as long as these retain certain qualitative properties. In fact, experience with other limit laws in probability would lead one to expect that a wide class of processes with the same qualitative properties of the elementary transition probabilities Π_{kn} would lead to the same limiting distribution in the same sense as the summation of a wide class of different random variables leads to the normal distribution. If there is any reality to Pareto's law, such considerations may possibly lead to a better understanding.[14] Alternatively, if the hypothesis should prove false, this would not disprove Pareto's claim, but at least it would produce doubts in certain respects.

Another example illustrating the importance of ergodic properties is supplied by statistical mechanics. It is usually proved (by more or less controversial arguments) that the Gauss-Maxwell distribution for the velocities of gas molecules is the only one having certain properties which are desirable to the physicist. In itself this does not prove that any gas obeys the Gauss-Maxwell law. What is really required is the knowledge that, whatever the initial velocity distribution (which, in fact, is arbitrary), the gas will tend to a state in which the velocities obey the Gauss-Maxwell law. In classical statistical mechanics the molecules move in a perfectly deterministic manmer, and "statistics" enters into the considerations only in connection with the initial state. It is well known that the ergodic theory of such systems has led to very deep and beautiful mathematical results. Seen against this back-

transitions from one group to the other. This case would actually represent two independent physical systems which have arbitrarily been combined into one.

Even if the process is not regularly ergodic, the limits (49) will exist, but they will depend on ν. If the system of differential equations is finite, the existence of these limits can be established in an elementary way: the solutions $P_{\nu k}(\tau,t)$ are then linear combinations of exponential terms of the form $\exp(a_i t_j)$, and it is not difficult to see that the real part of any non-vanishing a_i is negative. For infinite systems the ergodicity can be established either directly from the explicit form of the solutions or (according to a remark of Doob) as a simple consequence of abstract ergodic theorems.

[13] For finite chains, cf. Elfving (1937).

[14] The thoughts expressed in the text are perhaps related to, and certainly influenced by, a recent paper by Bernardelli (1944). Starting from deterministic considerations, Bernardelli also arrives at a system of differential equations which may be interpreted as describing a stochastic process. The present author is unable to follow all the mathematical arguments in Bernardelli's paper. In particular, with a division into finitely many classes, it would seem that all income classes should be equally probable, which would contradict the original thesis. This is not the first instance where an oversimplification completely hides the actual mechanism and contradicts all facts and theories; unfortunately, such a contradiction is not always manifest on the surface.

ground, the comparative triviality of the ergodic theory for random processes becomes the more remarkable. From the point of view of the physicist it may well pay to introduce random from the beginning at a more essential place.

6. Examples of general discontinuous Markov processes

Up to now we have considered only the case where the possible states form a simple sequence $E_1, E_2, \cdots$ In general, this will of course not be the situation. Even in the Pareto case the classification of possible incomes into classes E_i is an artifice which makes the theoretical considerations in many respects more complicated. In fact, a classification into a finite number of classes obscures the essential point.[15] It is therefore safer and at the same time more convenient to characterize the possible states by a real number $x \geqq 0$. Similar situations arise frequently. For example, in the collective theory of risk for insurance companies as developed by F. Lundberg and H. Cramér[16] the transitions of the system are connected with the occurrence of a claim; if a claim occurs, it may be for any amount. Again the actual state is expressed by a real number, positive or negative. Particles of ore are subject to splitting, and such a change can occur at any moment. This is another instance where it is more convenient not to restrict the consideration to a simple sequence of possible states. The more general method has been used by Kolmogoroff (1941) to derive the so-called logarithmico-normal law for the distribution of particle sizes.

A simple example where the solution can be obtained without recourse to the general theory is supplied by the theory of transport of stones by rivers, which has been treated by Pólya.[17] It has applications in engineering. A stone will lie still on the bottom of the river for a relatively long time. For some reason or other it will sooner or later be set into motion and transported a stretch $\mathbf{X}$ farther down the river. The traveling time is so short that the change in position can be treated as instantaneous. We are then concerned with a random process where the state of the system (= stone) is given by the distance x from its position, say at $t = 0$. If one wishes to take into consideration the fact that the river is not homogeneous, the integrodifferential equation to be developed later is required. In the first approximation, however, we may, with Pólya, treat the river as homogeneous. Then the probability of a transport taking place in any time interval of length dt will remain the same (independent of the actual position of the stone); let this intensity be denoted by η. Moreover, the distance $\mathbf{X}_n$ traveled t the nth step is a random variable and under the hypothesis of homogeniety all $\mathbf{X}_n$ have the same distribution function, say $U(x)$. The total distance traveled in the first n steps is $\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n$, and, since the $\mathbf{X}_k$ are mutually independent, the distribution function $U_n(x)$ of this sum is given by the well known formulas

$$U_1(x) = U(x), \qquad U_n(x) = \int_0^x U_{n-1}(x - y)dU(y). \tag{50}$$

[15] Cf. the preceding footnote.
[16] Cf. Segerdahl (1939).
[17] Pólya (1937, 1938); H. A. Einstein (1937). The methods given in the text are different.

With the present assumptions, the number **N** of transports during a time t is a random variable subject to the Poisson distribution. Accordingly, the distribution function for the state (that is, the total distance traveled) can be written down directly. It is the compound Poisson distribution

$$\sum_{n=0}^{\infty} e^{-\eta t} \frac{(\eta t)^n}{n!} U_n(x). \tag{51}$$

The same distribution has been extensively used by O. Lundberg and plays an important role in the collective theory of risk. It has also been given in Khintchine's well-known booklet published in 1933 (p.23). For many purposes it is simpler to pass from (51) to the corresponding equation for the characteristic functions. According to the central limit theorem the $U_n(x)$ tend, with increasing n, to a normal distribution. For larger values of t the expression (51) can therefore be approximated in an exceedingly simple way and the final result depends essentially only on three parameters: η, and the first two moments of the distribution function $U(x)$.

The case where the possible states of the system are determined by two, three, or more real numbers, that is to say, by a point in the plane or in a Euclidean space, presents no essential novelty. However, even more general cases present themselves. In connection with the theory of automatic telephone exchanges one is led to consider units in which certain conversations may be going on; the state will be described by stating the duration of each of the conversations and the number of people in the waiting line: the state is therefore given by a system of real numbers plus an integer. In Arley's theory of cosmic rays one is interested in the number not only of electrons but also of photons; the state is then characterized by two integers. If energy is to be taken into account, the situation becomes more complicated.

It is true that all these examples can, more or less artificially, be reduced to the case where the state is described by a point in a Euclidean space or even by a point on the real axis. However, there is no analytic gain in this procedure and we shall therefore write the equations in their general form. The reader can always interpret the space to mean the real axis.

7. The integrodifferential equations of Markov processes

We shall denote a possible state by a simple symbol such as x or ξ, which thus may stand for real numbers, points in space, or anything else. The set E of all possible states is, in accordance with physical terminology, called the "phase space." Instead of the transition probabilities $P_{\nu k}(\tau,t)$ we are now led to consider transition probabilities of the form $P(\tau,\xi;\ t,\Gamma)$ where $\tau < t$ are, as before, the time parameters, ξ is an arbitrary state, and Γ an arbitrary set of states (which may also consist of a single state). This $P(\tau,\xi;\ t,\Gamma)$ denotes the (conditional) probability that at time t the system will be in some state included in Γ, if at time τ it has been in state ξ. As an example, consider the case where E is the real axis. The transition from a point ξ into a *point* x will in

general have probability zero; however, in many cases a *density* $p(\tau,\xi;\, t,x)$ for such transitions will exist. If then Γ is an interval $a < x < b$, we shall have

$$P(\tau,\xi;\, t,\Gamma) = \int_a^b p(\tau,\xi;\, t,x)dx. \tag{52}$$

On the other hand, the case where the possible states form a simple sequence $E_1, E_2, \cdots$ can also be interpreted on the real axis by letting the integer $x = k$ stand for the state E_k. Then with our previous notations

$$P(\tau,\xi;\, t,k) = P_{\xi k}\,(\tau,t), \tag{53}$$

whereas transitions into any set (interval) not containing an integer are impossible (have probability zero). Equation (53) has an actual meaning only when ξ is an integer, since the state can never assume other values.

The theory of Markov processes hinges on the general Chapman-Kolmogoroff equation analogous to (41) and expressing the fundamental assumption that the changes of the system in any time interval (s,t) are statistically independent of the changes in the previous interval (τ,s). It reads now

$$P(\tau,\xi;\, t,\Gamma) = \int_E P(\tau,\xi;\, s,dE_y)\, P(s,y;\, t,\Gamma). \tag{54}$$

We can now proceed as before and introduce elementary probabilities analogous to the $p_k(t)$ and $\Pi_{rk}(t)$ of section 4. Accordingly, we shall assume:

1. If the system is at time t in state x, there is a probability $p(t,x)dt + o(dt)$ that *any* change of state occurs during $(t,\, t + dt)$.

2. If a change occurs, the probability that it will take the system into a state included in the set Γ is given by the distribution function $\Pi(t,x,\Gamma)$. This function must, of course, satisfy the conditions of respectability, such as being non-negative, $\Pi(t,x,E) = 1$, etc. Analytically the two conditions mean simply that, whenever x is not included in Γ, then

$$P(t,x;\, t + dt,\Gamma) = p(t,x)\,\Pi\,(t,x,\Gamma)dt + o(dt), \tag{54'}$$

whereas

$$P(t,x;\, t + dt,x) = 1 - p(t,x)dt + o(dt). \tag{54''}$$

Under some weak regularity conditions on the functions $p(t,x)$ and $\Pi(t,x,\Gamma)$ it is possible to show [Feller (1940)] that the transition probabilities satisfy two integrodifferential equations which now take the place of (44) and (46): The *forward equation*

$$\frac{\partial}{\partial t}P(\tau,\xi;\, t,\Gamma) \tag{55}$$

$$= -\int_\Gamma p(t,y)P(\tau,\xi;\, t,dE_y) + \int_E p(t,y)\Pi(t,y,\Gamma)P(\tau,\xi;\, t,dE_y),$$

and the *backward equation*

$$\frac{\partial}{\partial \tau} P(\tau,\xi; t,\Gamma) \tag{56}$$

$$= p(\tau,\xi)\{P(\tau,\xi; t,\Gamma) - \int_E P(\tau,y; t,\Gamma)\Pi(\tau,\xi,dE_y)\}.$$

Again the common boundary condition is given by the obvious requirement that, as $t - \tau$ approaches zero, $P(\tau,\xi; t,\Gamma)$ tends to one or zero according as Γ does or does not include ξ. With (55) τ and ξ are parameters occurring only in the initial conditions, and conversely (56) represents an equation in the variables τ and ξ only.

As an illustrative example let us return to Pólya's problem of the transport of stones which led to the solution (51). The states were there given by numbers $x \geqq 0$, and by assumption for any $x > 0$, $t > 0$

$$p(t,x) = \eta = \text{const.} \tag{57}$$

For the set Γ we shall naturally take an interval, $\Gamma = (0,x)$. Then, by definition,

$$P(t,\xi,\Gamma) = \begin{cases} U(x - \xi), & \text{if } \xi \leqq x, \\ 0 & \text{if } x \leqq \xi. \end{cases} \tag{58}$$

Obviously the transition probabilities depend only on the differences $t - \tau$ and $x - \xi$, and we can in this case write $P(\tau,\xi; t,\Gamma) = P(t - \tau; x - \xi)$. Equations (55) and 56) take on the form

$$\frac{\partial}{\partial t} P(t - \tau; x - \xi) = -\eta P(t - \tau; x - \xi) + \eta \int_\xi^x U(x - y) d_y P(t - \tau; y - \xi), \tag{55'}$$

and

$$\frac{\partial}{\partial \tau} P(t - \tau; x - \xi) = \eta P(t - \tau; x - \xi) - \eta \int_\xi^x P(t - \tau; x - y) dU(y - \xi). \tag{56'}$$

Introducing more natural variables of integration we see that these two equations actually represent the same equation, which fortunately is of the simple "renewal" type. Its integration presents no difficulty, and one arrives also in this way at the solution (51), where t has naturally to be replaced by $t - \tau$ and x by $x - \xi$.

Returning to the general case (55) and (56) we shall mention only that the existence and uniqueness of solutions has been established[18] and the situation is in every respect similar to that described in connection with the differential equations (44) and (46).

[18] Feller (1940); under somewhat weaker hypotheses, also Feller (1936), where a derivation of the equations is given.

The integrodifferential equations (55) and (56) seem to be of a type not studied before. They present some general mathematical interest inasmuch as they contain as special cases many functional equations of unexpected types. To begin with, the systems of infinitely many differential equations (44) and (46) are special cases which one obtains by taking for E the real axis and supposing that all probability mass is concentrated in the points $x = 1, 2, \cdots$ (or, even more naturally, letting E be the abstract space consisting of the denumerably many points E_k). An equation of quite different character is obtained by again taking for E the real axis but putting

$$\Pi(t,x,\Gamma) = \begin{cases} 1, & \text{if } x - q \text{ is contained in } \Gamma, \\ 0, & \text{otherwise;} \end{cases} \tag{59}$$

here q may be a constant or an arbitrary function. Equation (56) then reads

$$\frac{\partial}{\partial \tau} P(\tau,\xi\, t,\Gamma) = p(\tau,\xi)\{P(\tau,\xi;t,\Gamma) - P(\tau,\xi - q;t,\Gamma)\}. \tag{60}$$

Now t and Γ occur herein only as parameters, and the equation is really of the type of a difference-differential equation

$$\frac{\partial}{\partial \tau} u(\tau,\xi) = p(\tau,\xi)\,\{u(\tau,\xi) - u(\tau,\xi - q)\}, \tag{61}$$

where p and q are arbitrarily prescribed. By virtue of our theory, equation (61) has an adjoint which is obtained by substituting from (59) into (55); the two equations are, in a certain sense, equivalent, although they are very different in appearance. Taking for E a finite or infinite system of lines one can similarly obtain finite or infinite systems of equations analogous to (61):

$$u_i(\tau,\xi) = \sum p_k(\tau,\xi)u_k(\tau,\xi - q_k). \tag{62}$$

If we select for $\Pi(\tau,\xi,\Gamma)$ a function with more than one jump it is possible to replace the right-hand side in (61) by arbitrary linear combinations $u(\tau,\xi - q_n)$, and the like. It is evident that our existence and uniqueness theorems apply to these equations only to the extent that the latter represent stochastic processes, which imply restrictions such as positivity, bounded variation, and the like.

8. Waiting time and telephone problems

Although we have here been considering only Markov processes, it must continually be borne in mind that they represent only a special case of the general stochastic process. As in classical mechanics the present state completely determines the future of the system, so in a Markov process the present state determines all future probability relations. The past history of

the system has an influence only in so far as it has produced the present state ("influence globale," in Pólya's terminology). Accordingly, a Markov process is completely determined when its transition probabilities are known. Now if a man arrives at a counter and observes three people in the waiting line, this knowledge of the present state, in itself, does not permit him to compute the probability distribution of his waiting time. The latter depends in an essential manner on *how* the present state has developed, namely, on the time elapsed since the moment when the customer actually being served got access to the counter. Thus, at least if "state of a system" is defined so that it can really be observed, the problem of waiting times does not lead to a Markov process. (For possible redefinitions of the notion of state, cf. the next section.)

Fortunately, there exists an artifice which is often justified in practice and which reduces waiting time and similar problems to a Markov process.The time $\mathbf{T}$ which it takes to serve a customer is a random variable and will be called the "holding time." Let then

$$F(t) = Pr(\mathbf{T} \leqq t) \tag{63}$$

be the distribution function of the holding time. If it is known that the present customer has been at the counter for exactly t time units, the probability that he will leave within the next dt time units is, up to terms of higher order, equal to $dF(t)/[1 - F(t)]$, provided that the derivative $F'(t)$ exists. Consider now the particular case of exponential holding times, that is,

$$F(t) = 1 - e^{-ct}, \qquad c > 0. \tag{64}$$

In this case the probability just considered is independent of t and equals $c\,dt$: thus, with (64), if the counter is at any time occupied, the probability that the holding time will terminate within the next time interval of length dt is $c\,dt + o(dt)$ and independent of what has happened before. It is exactly the problem we have with telephone conversations of certain people where time plays no role and the probability of a continuation is independent of how long the conversation has already been going on.

The importance of exponential holding times is emphasized in the well-known book of T. C. Fry (1928), where several stochastic processes pertaining to the theory of telephone exchanges are discussed.[19] Many practical problems can at present be solved only under this simplifying assumption. Fortunately, this assumption is in many cases less artificial and more justified than would appear at first glance. For example, if many trunk lines serve the same customers, a waiting line will form only if all these lines are busy. In that event one is not so much interested in when a given line will become free, but rather when any line will become free. The waiting time is in this case given by the smallest among many random variables, and it follows from known limit theorems that under fairly general conditions the corresponding distribution function is of type (64).

[19] Fry generally discusses only the steady state. Many other problems will be found in Palm's book.

As an example, let us consider the simplest problem in telephone traffic [C. Palm (1943)]: Infinitely many trunk lines are available and calls arrive at a constant intensity η [the number of calls being distributed according to the Poisson law (4)]. Every new call is directed to a free trunk line, and the holding time (length of the ensuing conversation) is distributed according to (64). Required is the probability $P_{ik}(t)$ that k trunk lines will be busy at time t if initially i lines were busy. If at any time exactly n lines are busy, the probability of any change in state during the next time interval of length dt is $(nc + \eta)dt + o(dt)$; here the term nc accounts for the probability that a busy line is set free, the term η for new calls. The differential equations of our problem are therefore.

$$\begin{cases} P'_{i0}(t) = -\eta P_{i0}(t) + cP_{i1}(t), \\ P'_{ik}(t) = -(\eta + kc)P_{ik}(t) + \eta P_{i,k-1}(t) + (k+1)cP_{i,k+1}(t). \end{cases} \tag{65}$$

Alternatively, we could use the "backward equations"

$$\begin{cases} P'_{0k}(t) = -\eta P_{0k}(t) + cP_{1k}(t), \\ P'_{ik}(t) = -(\eta + ic)P_{ik}(t) + \eta P_{i-1,k}(t) + (i+1)cP_{i+1,k}(t). \end{cases} \tag{66}$$

The initial conditions are, in either case,

$$P_{ik}(0) = \begin{cases} 1, & \text{if } k = i, \\ 0, & \text{if } k \neq i. \end{cases}$$

The solution is

$$P_{ik}(t) = \exp\left[-\frac{\eta}{c}(1 - e^{-ct})\right]\{1 - e^{ct}\}^i \cdot \sum_{r=0}^{\min\{i,k\}} \frac{1}{(k-r)!}\binom{i}{r}\left\{\frac{\eta}{c}(1 - e^{-ct})\right\}^{k-r}\left\{e^{ct} - 1\right\}^{-r}. \tag{67}$$

In particular,

$$P_{0k}(t) = \exp\left[-\frac{\eta}{c}(1 - e^{-ct})\right] \cdot \frac{1}{k!}\left\{\frac{\eta}{c}(1 - e^{-ct})\right\}^k, \tag{68}$$

which is simply a Poisson distribution with parameter $\eta(1 - e^{-ct})/c$. Palm has also treated the more general case of intensities varying in time with ensuing periodic oscillations for the required probabilities.

If we assume exponential holding times, the problem of waiting lines with one or more counters can be treated in a similar way, but the corresponding differential equations are less easy to handle. However, at least the ergodic steady-state limit is easy to obtain.

9. Non-Markov processes

The definition of a Markov process depends on what is meant by the state of a system, and the question may well be asked whether it is not always possible to redefine the notion of state in such a way that the process becomes a Markov process. In theory this is possible. For example, in the waiting time problem of the beginning of the last section one might define the state so as to include not only the number of people in the waiting line but also the moment when the customer being served got access to the counter; the state would then be characterized by an integer and a real number. This would be in accordance with the procedure of mechanics where the primitive notion of state would include only the positions of particles, not their velocities; the latter have been included only for the convenience of having the present state determine the future. On the other hand, the new definition would have the disadvantage of the state not being directly observable. More important is that in most cases our integrodifferential equations would no longer hold because the intensity function $p(t,x)$ shows too strong discontinuities. The new equations can usually be written down, but they are so complicated that it is doubtful whether anything is gained. It is interesting to illustrate the situation by means of the simplest and, perhaps, most important process which is not of the Markov type.

We shall consider a process which is best known from the so-called "renewal theory," although exactly the same type and the same integral equation appear in many other applications, some of them even more important than the renewal theory. In order to explain the process and its practical meaning in the very simplest case, let us consider a certain unit (such as a machine or bulb) which, when installed, has a certain probability density $\phi(t)$ to last for a time t; in other words, the time of service of our unit is a random variable with the elementary distribution function $\phi(t)$. When the lifetime of the unit expires, it is immediately replaced by a similar unit. Required is the probability density $u(t)$ for the necessity of a replacement at time t. Now such a replacement can occur only in one of two mutually exclusive ways. Either the original unit is being replaced, or it has been replaced at some time $s < t$ and one of its successors is to be replaced at time $t - s$ after the installation of the first successor. Therefore

$$u(t) = \phi(t) + \int_0^t u(t - s)\, \phi(s) ds. \tag{69}$$

This is the well-known integral equation of renewal theory,[20] whose integration presents no particular difficulties.

In order to treat the same process as a Markov process we would define the state of the system at time t to be given by the moment $x < t$ of installation of the unit which actually serves at time t. The transition probabilities could be expressed by means of one symbol, but it is more convenient to use

[20] Cf., for example, Feller (1941).

two. If at time τ the state is ξ, then at a later time $t > \tau$ the state may still be ξ: the probability for this event we shall denote by $p(\tau,\xi,t)$. If the state has changed, the new state will be given by a number x with $\tau < x < t$. There exists a corresponding probability density which will be denoted by $p(\tau,\xi; t,x)$. These two functions together give all transition probabilities, since at time t any state with $x < \xi$, $\xi < x < \tau$, or $x > t$ is impossible. Now it is seen that these transition probabilities have a bad discontinuity at the particular point $x = t$, which is not even fixed in the phase space. The result is that the forward equation (55) alone does not describe the process. It yields in the present case

$$\frac{\partial}{\partial t} p(\tau,\xi,t) = -\frac{\phi(t-\xi)}{1-\Phi(t-\xi)} p(\tau,\xi,t), \tag{70}$$

and

$$\frac{\partial}{\partial t} p(\tau,\xi;t,x) = -\frac{\phi(t-\xi)}{1-\Phi(t-\xi)} p(\tau,\xi;t,x), \tag{71}$$

where $\Phi(x) = \int_0^x \phi(y)dy$. From the singularity at $x = t$ we obtain another equation, namely,

$$p(\tau,\xi;t,x) = \frac{\phi(t-\xi)}{1-\Phi(t-\xi)} p(\tau,\xi,t) + \int_\tau^t \frac{\phi(t-z)}{1-\Phi(t-z)} p(\tau,\xi;t,z)dz. \tag{72}$$

It can again be shown that these three equations together completely determine the transition probabilities. On the other hand, these transition probabilities can also be written down in terms of $u(t)$, and the equation (69) has certain advantages over the system (70) to (72).

It may be of interest to remark that many counting mechanisms and other simple apparatus change an incoming Markov process into an outgoing non-Markov process of the renewal type, which is described by (69). For example consider a trunk line or other unit in a telephone exchange. An incoming call is served by this line if the latter is free; otherwise it is directed to a second line. Suppose that the incoming calls are Markovian, or distributed according to the Poisson law. It is easily seen that the outgoing traffic of the first line, which is the incoming traffic of the second line, is no longer Markovian: instead, the time between consecutive calls will be regulated by an equation of type (69). For each consecutive line one integral equation of that type has to be solved. For a detailed analysis we refer the reader to Palm's book. Similar remarks apply to Geiger-Müller counters, where, owing to the resolving time of the counter, not all events are actually counted.

10. The connection with diffusion processes

In this paper we have been considering only the "purely discontinuous" type of Markov process: in a small time interval there is an overwhelming probability that the state will remain unchanged; however, *if* it changes, the change may be radical. The other extreme, that of a "purely continuous" process, is

represented by diffusion and by Brownian motion; there it is certain that *some* change will occur in *any* time interval, however small; only, here it is certain that the changes during small time intervals will be also small. Consider, for simplicity, a one-dimensional diffusion process. It is most convenient to consider densities of the transition probabilities and to let $u(\tau,\xi;t,x)$ denote the probability density of finding the particle at time t at the place x if it is known that at a previous time τ it has been at ξ. The Chapman-Kolmogoroff equation expressing that the changes in position during time intervals (τ,s) and (s,t) are independent now reads

$$u(\tau,\xi;t,x) = \int_{-\infty}^{+\infty} u(\tau,\xi;s,y)u(s,y;t,x)dy, \tag{73}$$

where of course $\tau < s < t$. The forward and backward equations of the discontinuous processes have a counterpart in the so-called Fokker-Planck equations for $u(\tau,\xi;t,x)$, which are familiar to the physicists and have served as a model to Kolmogoroff in laying the foundations of the general theory. In order to derive these equations we have again to introduce natural assumptions concerning the limits of the transition probabilities for very short time intervals. To begin with, we shall assume that the mathematical expectation of the displacement of the particle during a time interval of length dt is of the order of magnitude of dt; more precisely, we shall assume that if the particle is at time t at the place x, this mathematical expectation satisfies the relation

$$\lim_{dt\to 0}\frac{1}{dt}\int_{-\infty}^{+\infty} u(t,x;t+dt,y)\,(y-x)dy = b(t,x), \tag{74}$$

where the function $b(t,x)$ characterizes the direction and intensity of diffusion at place x. In the case of a symmetric diffusion, $b(t,x)$ naturally vanishes. The second and last assumption is that the variance of the displacements satisfies a relation analogous to (74):

$$\lim_{dt\to 0}\frac{1}{dt}\int_{-\infty}^{+\infty} u(t,x;t+dt,y)\,(y-x)^2dy = a(t,x). \tag{75}$$

Under these conditions[21] $u(\tau,\xi;t,x)$ satisfies the "forward equation"

$$u_t(\tau,\xi;t,x) = \tfrac{1}{2}\,[a(t,x)u(\tau,\xi;t,x)]_{xx} + [b(t,x)u(\tau,\xi;t,x]_x, \tag{76}$$

and the "backward equation"

$$u_\tau(\tau,\xi;t,x) = \tfrac{1}{2}\,a(\tau,\xi)u_{\xi\xi}(\tau,\xi;t,x) + b(\tau,\xi)u_\xi(\tau,\xi;t,x); \tag{77}$$

[21] Feller (1936). There is also shown that each of the differential equations (76) and (77) has a unique solution which is automatically a solution of the other equation and of the Chapman-Kolmogoroff equation, and which satisfies all prescribed conditions, such as (74), (75), etc.

here subscripts denote differentiation. Again (76) is an equation in the variables t and x where τ and ξ enter only as parameters occurring in the initial condition; conversely, (77) is an equation in the variables τ and ξ. The initial condition expresses the simple fact that as the difference $t - \tau$ approaches zero the probability of any finite displacement tends to zero: in other words, for every fixed $\epsilon > 0$,

$$\int_{\xi-\epsilon}^{\xi+\epsilon} u(\tau,\xi;t,y)dy \to 1. \tag{78}$$

The most familiar case is that of a homogeneous diffusion where $b(t,x) = 0$ and $a(t,x)$ is a constant which can be chosen as 2. Then

$$u(\tau,\xi;t,x) = \frac{1}{2\{\pi(t-\tau)\}^{1/2}} \exp\left\{-\frac{(x-\xi)^2}{4(t-\tau)}\right\} \tag{79}$$

is the familiar Gaussian distribution.

It is both of theoretical interest and of practical importance that many empirical phenomena can be described as well by a discontinuous as by a continuous model. For example, a biological population changes size only in jumps, individuals being born or dying. Nevertheless, if the population is large, its size can in the usual manner be treated as a continuous function. It is then only a matter of analytical convenience whether the process is to be treated as a diffusion-type random process or handled by one of the models considered in this paper. In many cases the diffusion equation lends itself better for an analytical treatment [Feller (1939)].

From these considerations it follows that, for practical purposes, it is possible to replace the differential equations of diffusion (76) and (77) by systems of ordinary differential equations. In fact, if the real axis is divided into small intervals of length δ, we may in the first approximation neglect the precise position of the particle and consider only the interval in which it is contained. We have then a simple sequence of states E_k, where E_k stands for the fact that the particle is contained in the interval $k\delta < x < (k+1)\delta$. We may then, by the method of section 4, consider the transition probabilities $P_{\nu k}(\tau,t)$ which will satisfy a system of infinitely many ordinary differential equations, say (46). One would then expect that as $\delta \to 0$ the solutions $P_{\nu k}(\tau,t)$ will converge toward a solution $u(\tau,\xi;t,x)$ of an equation of type (77). In other words, passing to the limit $\delta \to 0$ so that

$$\nu\delta \to \xi, \qquad k\delta \to x, \tag{80}$$

one would expect that

$$\delta^{-1}P_{\nu k}(\tau,t) \to u(\tau,\xi;t,x). \tag{81}$$

Let us consider this passage to a limit in a purely formal and heuristic way. As the length δ of the individual intervals decreases, the probability of passing

from one into another increases, and we have to assume that the intensities of the discontinuous process (46) satisfy $p_\nu(t)\delta \to 1$. If we wish to come from (46) to (77) it is necessary to assume that the mathematical expectation and the variance of the displacement in the discontinuous process tend to the corresponding quantities for the process (46), or, in other words, we have to assume that

$$\Pi_{\nu k}(k-\nu)\delta \to b(\tau,\xi), \tag{82}$$

$$\Pi_{\nu k}(k-\nu)^2\delta^2 \to a(\tau,\xi). \tag{83}$$

Finally, since in small time intervals small changes are preponderant, we are led to assume that

$$\Pi_{\nu k}\,|k-\nu|^3\delta^3 \to 0, \tag{84}$$

a relation which, under certain regularity conditions, would be implied by (83).

It is now easy to say what becomes of the individual terms in (46). Putting $j = y\delta$ we obtain from (81)

$$\delta^{-1}P_{jk}(\tau,t) \sim u(\tau,y;t,x) \tag{85}$$

$$= u(\tau,\xi;t,x) + (y-\xi)u_\xi(\tau,\xi;t,x) + \frac{(y-\xi)^2}{2}u_{\xi\xi}(\tau,\xi;t,x) + \cdots$$

$$\sim u(\tau,\xi;t,x) + (j-i)\sigma u_\xi(\tau,\xi;t,x) + \frac{(j-i)^2\delta^2}{2}u_{\xi\xi}(\tau,\xi;t,x).$$

Substituting from (85) into (46) and remembering (82) and (83), we see that (46) passes formally into (77).

It appears therefore that every continuous process can be considered as a limiting case of discontinuous processes, and that solutions of partial differential equations of form (76) and (77) can be approximated by solutions of systems of infinitely many ordinary differential equations of type (44) and (46).

REFERENCES

ARLEY, N.
1934. *"On the Theory of Stochastic Processes and Their Application to the Theory of Cosmic Radiation."* 240 pp. Copenhagen.

ARLEY, N. and V. BORCHSENIUS
1945. "On the theory of infinite systems of differential equations and their application to the theory of stochastic processes and the perturbation theory of quantum mechanics," *Acta Mathematica*, vol. 76, pp. 261-322.

BERNARDELLI, H.
1944. "The stability of income distributions," *Sankhya*, vol. 6, pp. 351-362.

CHANDRASEKHAR, S.
1943. "Stochastic problems in physics and astronomy," *Reviews of Modern Physics*, vol. 15, pp. 1-89.

DANIELS, H. E.
1944. "The statistical theory of the strength of bundles of threads I," *Proc. Roy. Soc. London*, vol. 183-A, pp. 405-435.

DOOB, J. L.
1945. "Markoff chains—Denumerable case," *Trans. Amer. Math. Soc.*, vol. 58, pp. 455-473.

DUBROVSKI, W.
1938. "Eine Verallgemeinerung der Theorie der rein unstetigen stochastischen Prozesse von W. Feller," *C. R. (Doklady) Académie Sciences URSS*, vol. 19, pp. 439-445.

EINSTEIN, H. A.
1937. Der Geschiebetrieb als Wahrscheinlichkeitsproblem. Mitteilungen derVersuchsanstalt für Wasserbau der Eidgenössischen Technischen Hochschule in Zürich, Thesis, 111 pp.

ELDERTON, W.
1945. "Cricket scores and some skew correlation distributions" (discussion by Mr. Seal), *Jour. Roy. Stat. Soc.*, vol. 108, pp. 34-37.

ELFVING, G.
1937. "Zur Theorie der Markoffschen Ketten," *Acta Societatis Scientiarum Fennicae (Helsingfors) N. S. A.*, vol. 2, no. 8, 17 pp.

FELLER, W.
1936. "Zur Theorie der stochastischen Prozesse," *Mathematische Annalen*, vol. 113, pp. 113-160.
1939. "Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung," *Acta Biotheoretica* (Leiden), vol. 5, pp. 11-40.
1940. "On the integrodifferential equations of purely discontinuous Markoff processes," *Trans. Amer. Math. Soc.*, vol. 48, pp. 488-515.
1941. "On the integral equation of renewal theory," *Annals of Math. Stat.*, vol. 12, pp. 243-267.
1943. "On a general class of 'contagious distributions'," *ibid.*, vol. 14, pp. 389-399.

FRY, T. C.
1928. *"Probability and Its Engineering Uses."* 476 pp. New York, D. Van Nostrand.

GREENWOOD, M., and G. UDNY YULE
1920. "An inquiry into the nature of frequency distribution representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents," *Jour. Roy. Stat. Soc.*, vol. 83, pp. 255-279.

KHINTCHINE, A.
1939. "Asymptotische Gesetze der Wahrscheinlichkeitsrechnung." Ergebnisse der Mathematik, vol. 2, Berlin, J. Springer, 77 pp.

KOLMOGOROFF, A.

1931*a*. "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung," *Math. Annalen*, vol. 104, pp. 415-458.

1931*b*. "Sur le problème d'attente," *Bull. Académie Sciences URSS*, pp. 101-106.

1941. "Über das logarithmisch normale Verteilungsgesetz der Dimension der Teilchen bei Zerstuckelung," *C. R. (Doklady) Académie Sciences URSS*, vol. 31, pp. 99-101.

LUNDBERG, O.

1940. *"On Random Processes and Their Application to Sickness and Accident Statistics."* 172 pp. Uppsala.

PALM, C.

1943. "Intensitätsschwankungen im Fernsprechverkehr," *Ericcson Technics (Stockholm)*, vol. 44, pp. 1-189.

POLLACZEK, F.

1946. "Sur l'application de la théorie des fonctions au calcul de certaines probabilités continues utilisécs dans la théorie des réseaux téléphoniques," Annales Inst. H. Poincaré, vol. 10, pp. 1-55.

1947. "Sur un problème du calcul des probabilités qui se rapporte à la téléphonie." Journ. Mathém. Purcs. Appl. Vol. 25, pp. 307-334.

PÓLYA, G.

1930. "Sur quelques points de la théorie des probabilités," *Annales Institut Henri Poincaré*, vol. 1, pp. 117-161.

1937. "Zur Kinematik der Geschiebebewegung." Mitteilungen der Versuchsanstalt fur Wasserbau der Eidgenössischen Technischen Hochschule, Zürich, pp. 1-21.

1938. "Sur la promenade au hasard dans un réseau de rues." Actualités Scientifiques et Industrielles, no. 734, pp. 25-44.

SEGERDAHL, C. O.

1939. *"On Homogeneous Random Processes and Collective Risk Theory."* 132 pp. Uppsala.

WANG, MING CHEN, and G. E. UHLENBECK

1945. "On the theory of the Brownian Motion II," *Reviews of Modern Physics*, vol.17, pp. 323-342.

YULE, G. UDNY

1924. "A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S.," *Philo. Trans. Roy. Soc. London*, Series B, vol. 213, pp. 21-87.