TIME-DISCRETE STOCHASTIC PROCESSES IN ARBITRARY SETS, WITH APPLICATIONS TO PROCESSES WITH ABSORBING REGIONS AND TO THE PROBLEM OF LOOPS IN MARKOFF CHAINS

by

Prof. Dr D.van Dantzig

Prepublication

National Bureau of Standards, Washington D.C.
Mathematical Centre, Amsterdam
1952

FOREWORD

Time-discrete stochastic processes have many applications in physics, chemistry and elsewhere; for example, in the study of configurations of atoms in large molecules and in connection with the relativity new method of statistical inference known as sequential analysis.

The present report makes available some new results developed by Professor Van Dantzig while a guest worker at the National Bureau of Standards during the summer of 1951. His work was performed under Contract CST-525 between the National Bureau of Standards and the University of North Carolina. It is believed that these results will be of interest both to mathematicians and to workers in applied fields.

J.H.Curtiss
Chief, National Applied
Mathematics Laboratories

A.V.Astin
Acting Director
National Bureau of Standards

TIME-DISCRETE STOCHASTIC PROCESSES IN ARBITRARY SETS, WITH APPLICATIONS TO PROCESSES WITH ABSORBING REGIONS AND TO THE PROBLEM OF LOOPS IN MARKOFF CHAINS

By

D. van Dantzig

Prepublication

ABSTRACT

By means of generalized generating functions (functionals) and generalized matrices, a linear operator $C_{(A)}$ in a space of measurable functions on a given set E (containing e.g. all bounded ones) is introduced, depending on the subset A where absorption takes place. An interpretation, previously given, of the auxiliary variables and functions as probabilities, e.g. of the non-occurrence of some "catastrophe", appears to be useful for this type of problems also. It is proved that $C_{(A)}$ is a projection-operator and that for $A_1 \supset A_2$ $C_{(A_1)}$ and $C_{(A_2)}$ are commutable and that their product equals $C_{(A_2)}$. This identity contains as a special case an identity recently obtained in an Amsterdam Ph.D.-thesis by J.H.B.Kemperman. Moreover A.Wald's fundamental identity is generalized and it is shown that the characteristic function of a distribution can be considered as a special case of an eigenvalue in a generalized sense, where the eigenfunction (in the special case an exponential function) needs only be proportional to its transform by the linear operator outside the absorbing set A. Finally some theorems on loops in the path of a wandering point and some similar results are obtained by derivation of the generating functional with regard to the function on which it depends.

The purely mathematical theory of generalized matrix-multiplication is developed in the Appendix. In particular, sufficient conditions for the associativity are given, in cases also where the matrices are not bounded. Moreover the Appendix contains a generalization of a process by "independent increases", for the case of a set on which (under some conditions) a transitive group operates. I owe several remarks which improved the text to Mr J.J.de Iongh, who assisted me in drawing up the final version of this paper.

1. The Method of Collective Marks.

According to A.Kolmogoroff a probability field is defined as a set Π on which a countably [1] additive set function (defined, $\geqq 0$ and $\leqq 1$ for all subsets Λ of Π belonging to a σ-field σ_Π, which contains Π and each of its elements) is given. Denoting the set function by P the value it takes on a set Λ by P_Λ (instead of $P(\Lambda)$), we have

$$0 \leqq P_\Lambda \leqq 1 = P_\Pi \tag{1.1}$$

$$P_\Lambda = \sum_1^\infty P_{\Lambda_n}, \text{ if } \Lambda = \bigcup_1^\infty \Lambda_n \text{ and } \Lambda_m \cap \Lambda_n = 0 \text{ for } m \neq n \tag{1.2}$$

In many problems of probability theory we have to do with several countably additive set functions, all defined on a single set Π, i.e., P may vary over another set Ω, usually called

---

[1] These countably additive set functions, defined by the condition 1.2 are also sometimes referred to as totally additive absolutely additive, completely additive or σ-additive set functions.

the "set of admissible hypotheses" or "parameter space". Then for any $\Theta \in \Omega$ we have a $\mathcal{P}$, denoted by $\mathcal{P}^\theta$ taking on Λ the value $\mathcal{P}_\Lambda^\theta$. If Π is a countable (i.e. finite or denumerable) set, we have $\mathcal{P}_\Lambda^\theta = \sum_{\lambda \in \Lambda} \mathcal{P}_\lambda^\theta$, and if both Π and Ω are finite sets, $\mathcal{P}_\Lambda^\theta$ is determined by the ordinary rectangular matrix of the numbers $\mathcal{P}_\lambda^\theta$, with $\lambda \in \Pi$. For this reason we call the system of values $\mathcal{P}_\Lambda^\theta$ with $\theta \in \Omega$, $\Lambda \in \sigma_\Pi$, subject to the condition of countable additivity with respect to Λ a (generalized) <u>matrix</u>. Some of the fundamental properties of these matrices will be considered under somewhat more general conditions in an Appendix to this paper (§§ 7 and 8). [2] .

Generalizing an idea of Laplace, according to which a system of probabilities p_n ($n = 0, 1, 2, \ldots$) is represented by its "generating function"

$$\varphi(z) \overset{df}{=} \sum_0^\infty p_n z^n, \tag{1.3}$$

z being an auxiliary variable, it was found to be useful (Cf. D.van Dantzig 1940), to represent a probability field $\mathcal{P}_\Lambda$ by the corresponding functional of an auxiliary function $\mathcal{U}$, defined on Π. Assuming this function to be σ_Π-measurable and the integral 1.4 to exist, and denoting the value, which $\mathcal{U}$ takes in a point $\lambda \in \Pi$, by $\mathcal{U}^\lambda$ (instead of the customary $\mathcal{U}(\lambda)$), the corresponding functional may be written as

$$C(\mathcal{U}) \overset{df}{=} \int \mathcal{P}_{d\lambda}\, \mathcal{U}^\lambda \tag{1.4}$$

---

[2] The notation used in the Appendix ($\mathcal{P}_X^x$ instead of $\mathcal{P}_\Lambda^\theta$) is slightly different from the notation in this paragraph, but conforms to the use in § 2.

If $C(\mathcal{U})$ is known for a sufficiently large class of functions $\mathcal{U}$, it determines P. In the simplest case, when it is known for all bounded σ_{π}-measurable $\mathcal{U}^{\lambda}$, we have

$$P_{\Lambda} = C(\iota_{\Lambda}) \tag{1.5}$$

where ι_{Λ} (ι =iota) denotes the characteristic function of the set Λ, defined by

$$\iota_{\Lambda}^{\lambda} \overset{df}{=} \begin{cases} 1 & \text{if } \lambda \in \Lambda \\ 0 & \text{if not.} \end{cases} \tag{1.6}$$

(Cf. D.van Dantzig, 1935). In the more general case, when P varies over a set Ω, we can also introduce an auxiliary countably additive set function F on a σ-field σ_{Ω} of subsets of Ω, assuming P_{Λ}^{θ} for any fixed Λ to be σ_{Ω}-measurable, and define the corresponding functional

$$C(F,\mathcal{U}) \overset{df}{=} \int F_{d\theta} \int P_{d\lambda}^{\theta} \cdot \mathcal{U}^{\lambda} \tag{1.7}$$

These functionals were introduced in my Amsterdam lectures in 1947 and in a lecture given at Lyon in 1948 (published 1949); and some applications of the method were given there. The auxiliary functions were called "marks", their functional the "collective mark". It will be found useful to use a notation and terminology, introduced on another occasion (1935), viz. to call point functions and countably additive set-functions (subject to certain hardly restrictive conditions specified in the Appendix, functions of the first and the second kind respectively, and to denote systematically the arguments (points) of the former ones by lower case letters written as upper suffixes, and those (sets) of the latter ones by capitals, written as lower suffixes, except that in the case where an integration is performed the limits of "small" or "elementary" sets are represented by the symbol denoting the corresponding variable point, preceded by the letter d, instead of a capital letter.

Previous applications showed that the application of the method was often considerably facilitated, if the auxiliary and collective variables, functions and functionals were interpreted as probabilities, whilst restricting their range temporarily to the real interval $(0,1)$. So in 1.7 we might restrict F_{Θ} $(\Theta \in \mathfrak{S}_{\Omega})$ to $0 \leqq F_{\Theta} \leqq 1 = F_{\Omega}$, and interpret F_{Θ} as the probability that, by means of some random mechanism a $\theta \in \Theta$ were chosen. Also, assuming $0 \leqq \mathcal{U}^{\lambda} \leqq 1$, we might consider an auxiliary event $\mathcal{E}$ (which has nothing to do with the probability problem under consideration) and assume that whenever the latter results in a $\lambda \in \Pi$, a random mechanism, depending on λ, determines with a probability $1 - \mathcal{U}^{\lambda}$ against $\mathcal{U}^{\lambda}$, whether or not $\mathcal{E}$ happens. Then $C(F, \mathcal{U})$ is the total probability that $\mathcal{E}$ does not happen. It is essential, that the random mechanisms remain (at least partly) indetermined, so as to guarantee variability of F and $\mathcal{U}$ over sufficiently large sets of functions. Roughly speaking, the wider the class of functions over which F and $\mathcal{U}$ may vary, the larger the class of problems for which their introduction is useful. The interpretation of an auxiliary F as a probability distribution is due to J.von Neumann (1928) and was extensively used in his and O.Morgenstern's Theory of Games and Economic Behavior, (1947), and in A.Wald's Statistical Decision Functions, (1950). Functions of the first kind, however, whenever they occur in these places, are usually given an economic interpretation as "gains" or "losses". This, doubtless, has the advantage that they may vary over all real numbers, not requiring restriction to the interval $(0,1)$, but the disadvantage, that products and powers of gains or losses have no obvious interpretation, whereas products of probabilities may readily be interpreted as probabilities.

For the same reason C and $\mathcal{U}^x$ have been interpreted as the total and conditional probabilities that an event $\mathcal{E}$ does not (instead of does) occur. We shall call this event $\mathcal{E}$ a "catastrophe". The terminological advantage of taking occurences of non $\mathcal{E}$ instead of $\mathcal{E}$ lies in the fact that a conjunction of several non-occurences of $\mathcal{E}$ can be described as a (total) non-occurence of $\mathcal{E}$, whereas a conjunction of several occurrences of $\mathcal{E}$ cannot be described so simply as an occurrence. Anyhow, the interpretation of the auxiliary quantities is not of primary importance, and moreover, nothing prevents us, so far as no convergence difficulties occur, to apply the probabilistic terminology to negative or complex quantities also, in the same way as it is done with the geometric terminology. (Cf. Bartlett, (1944)).

It is the purpose of the present paper to derive some results concerning stochastic processes by means of this method.

2. Stochastic Processes.

We consider random variables, i.e. functions on Π , the "values" of which are elements $x, y, z, \ldots$ of an arbitrary set E. Random variables (or random events) will be denoted by dropping the argument $\lambda \in \Pi$ and underlining the function symbol, e.g. $\underline{x}$, $\underline{y}$, $\underline{x}_n$, etc. [3] . Again we assume that on the set E and on its direct product sets ($E \times E$, $E \times E \times E$, . . .) σ -fields of subsets σ_E , σ_{E^2} , σ_{E^3}, . . . are given (more general conditions are considered in the appendix § 7). [4] .

We shall define a stochastic process (discrete in time) as a sequence of random variables $\underline{x}_0$, $\underline{x}_1$, $\underline{x}_2$, . . . in E , such that the conditional probability distribution of each of them, the preceding ones being given, exists:

$$P_{(n)}{}^{x_0,\ldots,x_{n-1}}_{X} = P[\underline{x}_n \in X / \underline{x}_0 = x_0, \ldots, \underline{x}_{n-1} = x_{n-1}], \quad (2.1)$$

where x_0, . . . , x_{n-1} are arbitrary elements of E and X is an element of σ_E . These probability distributions will be assumed to be σ_{E^n} -measurable (in x_0, . . . , x_{n-1}) and countably additive in X and to satisfy the relations

$$0 \leqq P_{(n)}{}^{x_0,\ldots,x_{n-1}}_{X} \leqq 1 = P_{(n)}{}^{x_0,\ldots,x_{n-1}}_{E} . \quad (2.2)$$

The common probability distribution of $\underline{x}_1$, . . . , $\underline{x}_n$ is given by

$$P^{(n)}_{X_0,\ldots,X_n} = P[\underline{x}_0 \in X_0, \ldots, \underline{x}_n \in X_n], \quad (2.3)$$

---

[3] Random variables often are denoted by capital letters. This rarely is done consistently, capitals being used for other purposes also, and some random variables being denoted by other symbols. Our system of notation saves a whole alphabet (capital letters) for other purposes.

[4] In order to obtain a definition of the probability distribution on E we have to restrict the random variables to those functions on Π which are $\sigma_E - \sigma_\Pi$ -measurable, i.e. for which the originals of σ_E -sets are always σ_Π -sets.

We restrict ourselves to the case, where it is countably additive, not only in each X_k separately, but also in the X_k together, hence on $\mathfrak{S}_{E^{n+1}}$. Assuming moreover that the multiple integral equals the result of successive integrations, it is then related to 2.1 by the recurrent relation for $n \geqq 1$.

$$P^{(n)}_{X_0,\dots,X_n} = \int_{X_0}\cdots\int_{X_{n-1}} P^{(n-1)}{}_{dx_0,\dots,dx_{n-1}} P_{(n)}{}^{x_0,\dots,x_{n-1}}_{X_n} . \quad (2.4)$$

Descriptively we shall refer to the sequence of random variables as to a "random walk" in E, or a point, "wandering" or "jumping" through or over E, to the elements $x \in E$ as to the "states" or "places" in which it may be, and to $P_{(n)}{}^{x_0,\dots,x_{n-1}}_{X}$ as to the „transition probability" from a point having "passed through" $x_0,\dots,x_{n-1}$ successively to any state in X. The sequence of states $x_0, x_1, \dots$ through which the wandering point passes will be called its path, the sequence $x_0,\dots,x_n$ its n th path-segment, the transition from x_{n-1} to x_n the n th step (or jump).

A stochastic process is a (simple) <u>Markoff process</u>, if the transition-probabilities for $n \geqq 1$ depend on the last state only, through which the wandering point has passed:

$$P_{(n)}{}^{x_0,\dots,x_{n-1}}_{X} = P_{(n)}{}^{x_{n-1}}_{X} \qquad (n \geqq 1) \qquad (2.5)$$

$P_{(n)}$ is called the transition-matrix of the n th transition; $P_{(n)}{}^{x}_{X}$ is the probability that the wandering point, if x is its $(n-1)$ th state, (i.e. $\underline{x}_{n-1} = x$) will jump into an element of X (i.e. $\underline{x}_n \in X$). The probability, under condition $\underline{x}_{n-1} = x$, that $\underline{x}_{n+1} \in X$ is $\int P_{(n)}{}^{x}_{dy} P_{(n+1)}{}^{y}_{X}$, for which we shall write, using the matrix notation discussed in greater detail in the Appendix definition 8,37, $(P_{(n)} P_{(n+1)})^{x}_{X}$. Generally we have for $n \geqq 1$

$$P[\underline{x}_{n-1+k} \in X / \underline{x}_{n-1} = x] = (P_{(n)} \cdots P_{(n-1+k)})_X^x \qquad (2.6)$$

the matrix-multiplication being associative. (Cf. Appendix § 8, lemma 5).

The Markoff process is stationary, if all transition-matrices $P_{(n)}$ are equal to one and the same matrix P :

$$P_{(n)}{}_X^x = P_X^x \qquad (n \geqq 1) \qquad (2.7)$$

The matrix defined by 2.6 then simply is the k th power P^k of P , independent of n . The asymptotic properties of general stationary Markoff processes have been studied in an excellent paper by W.Doblin (1940).

An important special case is the one in which E is the set of all real numbers, and in which the Markoff process is invariant under translation, i.e.

$$P_{(n)}{}_{X+v}^{x+v} = P_{(n)}{}_X^x \qquad (2.8)$$

for all real numbers v . Putting $p_{(n)}X = P_{(n)}{}_X^0$ we have then

$$P_{(n)}{}_X^x = P_{(n)}{}_{X-x}^0 = p_{(n)}X_{-x} = \int p_{(n)}\,dv\, /_X^{x+v} . \qquad (2.9)$$

Such a process is sometimes called a "process by independent increases" (i.i.-process), as the coordinate of the $(n+1)^{st}$ state is the sum of $(n+1)$ independent stochastic variables:

$$\underline{x}_n = \underline{x}_0 + \underline{v}_1 + \cdots + \underline{v}_n , \qquad (2.10)$$

$\underline{v}_k$ being distributed according to $p_{(k)}X$; we might prefer the term "invariant process".

More generally we may consider the case, where E is an r -dimensional Euclidean space. Then, defining an invariant process by 2.8, where now x and v are vectors in r dimensions, 2.9 remains valid. The same is true in the still more general case, where E is an arbitrary Abelian group, additively written, x and v now being group-elements.

The case where E is a non-abelian group can be considered as a specialization of the still more general case, where a group of transformations of E into itself exists, which is transitive over E, and under which P is invariant (Cf. Appendix, § 9).

3. The Collective Matrix of a Stationary Markoff Process in an Absorbing Medium.

We consider a stationary Markoff process in a set E, determined by the transition matrix P_X^x. We assume that the wandering point starts from a given state x, that whenever it is in a state y, there is a probability A^y that the process will not be continued (we say then that the wandering point is "absorbed" in y), in which case there is a probability $1-U^y$ that the catastrophe $\mathcal{E}$ will occur; and a probability $B^y = 1-A^y$ that it will be continued, in which case there is a probability $1-T^y$ that $\mathcal{E}$ will occur.

Instead of the total probability C that $\mathcal{E}$ will not happen we introduce the conditional probability C^x that the wandering point, if starting at x, will be absorbed eventually without a catastrophe having happened.

It is easy to compute C^x. The point being in x, it is either absorbed at once (probability A^x), in which case non-$\mathcal{E}$ has probability $\mathcal{U}^x$, or not, (probability B^x) in which case the probability of non-$\mathcal{E}$ is T^x. But then it jumps to some point y (transition-probability P^x_{dy} if dy represents a "small set" containing y, namely a set on which the variation of the integrand is small), and then the probability of eventual absorption without $\mathcal{E}$ is C^y. Hence

$$C^x = A^x \mathcal{U}^x + B^x T^x \int P^x_{dy} \, C^y . \qquad (3.1)$$

In order to obtain an expression for the solution of this equation we introduce the "collective matrix" C^x_X, which denotes the probability that a point, starting in x, will eventually be absorbed somewhere in X, without a catastrophe having happened. The equation for C^x_X corresponding to 3.1 is now

$$C^x_X = A^x \mathcal{U}^x 1^x_X + B^x T^x \int P^x_{dy} \, C^y_X$$

If we make the assumption that a catastrophe can only happen (with probability $1 - T^x$) in a state where the point is <u>not</u> absorbed, we have to substitute $\mathcal{U}^x = 1$, and the equation for C^x_X is specialized to

$$C^x_X = A^x 1^x_X + B^x T^x \int P^x_{dy} \, C^y_X . \qquad (3.2)$$

The total C^x (for general $\mathcal{U}^x$) can now be expressed in this specialized C^x_X by

$$C^x = C^x(\mathcal{U}) = \int C^x_{dy} \, \mathcal{U}^y . \qquad (3.3)$$

We consider in particular the case where the local absorption-probabilities A^x have only the values 0 or 1. If then A is the set of all x with $A^x = 1$, B its complement, we have

$$A^x = 1_A^x \quad , \quad B^x = 1_B^x \quad . \qquad (3.4)$$

Substitution of 3.4 into 3.3 gives in accordance with the definitions of the diagonal matrices

$$(1_A)_X^x \overset{df}{=} 1_{A \cap X}^x = 1_A^x \cdot 1_X^x \qquad (8.7)$$

$$(1_B)_X^x \overset{df}{=} 1_{B \cap X}^x = 1_B^x \cdot 1_X^x$$

$$(T)_X^x \overset{df}{=} T^x 1_X^x \qquad (8.8)$$

and of matrix multiplication

$$(QP)_X^z \overset{df}{=} \int_E Q_{dy}^z P_X^y \qquad (8.37)$$

considered in greater detail in the Appendix § 8, the equation

$$C_X^x = (1_A)_X^x + (1_B TPC)_X^x$$

or in matrix notation

$$C = 1_A + 1_B TPC \quad . \qquad (3.5)$$

Multiplying both members of 3.5 from the left by 1_A and 1_B and noting that $1_A 1_B = 1_B 1_A = 0$, $1_A 1_A = 1_A$, $1_B 1_B = 1_B$, we obtain the important and often used identities

$$1_A C = 1_A \quad , \quad 1_B C = 1_B TPC \; . \tag{3.6}$$

From 3.5 we obtain by induction

$$C = \sum_0^{N-1} (1_B TP)^n 1_A + (1_B TP)^N C \, . \tag{3.7}$$

Now, assuming first $\| T \| = \sup_{x \in E} T^x < 1$, then by 8.42 of the Appendix

$$\| (1_B TP)^N C \| \leq \| 1_B TP \|^N \| C \| \leq \| T \|^N \tag{3.8}$$

as $\| 1_B \| = 1$ (unless $B = 0$, which trivial case we exclude), $\| P \| = 1$ and $\| C \| \leq 1$, C_X^x being a probability distribution. Hence the second term in the right member of 3.7 tends to zero, and

$$C = \sum_0^{\infty} (1_B TP)^n 1_A \; , \tag{3.9}$$

where the series converges if $\| T \| < 1$. We know, however, only that $\| T \| \leq 1$. If we replace T^x in 3.9 by θT^x , θ being a real number, the series in the right member is convergent for $0 \leq \theta < 1$, hence an analytic function of θ . As all its terms are ≥ 0 and its value, being a probability, remains ≤ 1 , hence bounded for $0 \leq \theta < 1$, the series remains convergent for $\theta = 1$ also, and its value remains ≤ 1 . Hence 3.9 holds not only for $\| T \| < 1$, but also for $\| T \| = 1$. We may then drop (if necessary) the interpretation of the T^x as probabilities, and replace them by arbitrary complex values. Then C becomes a (complex) analytic functional of T , defined (at least) for $\| T \| \leq 1$.

For $\| T \| < 1$ we can write because of 8.50

$$C = (I - I_B TP)^{-1} I_A \tag{3.10}$$

Multiplying both members of 3.9 from the right with I_A we obtain the identity

$$C I_A = C \quad , \tag{3.11}$$

which expresses the fact that absorption occurs in A only, i.e. that

$$C_X^x = C_{A \cap X}^x \quad . \tag{3.12}$$

For $T^x = 1$ (i.e. if no catastrophe occurs at all) 3.9 implies the total probability of a point, starting at x , being absorbed eventually somewhere in X . For the special case where E is the set of natural numbers, so that $P_X^x = \sum_{y \in X} P_y^x$, where

$$P_y^x = \begin{cases} p & \text{if} \quad y = x+1 \\ 1-p & \text{if} \quad y = x-1 \end{cases} \tag{3.13}$$

and $= 0$ in all other cases, and where $B = \{x \in E \,/\, 0 < x < a+b\}$, C_y^x with $T^x = 1$ for $y = x-a$ and $y = x+b$ solves the classical problem of the Ruin of the Gamblers. In fact, 3.9 is essentially equivalent with Abraham de Moivre's classical solution. For $T^x = T =$ constant we obtain the generating function belonging to the problem of Duration of Play. The generalization which was used in Wald's and Barnard's original theory of Sequential Analysis is obtained if in 3.13 the conditions are replaced by $y = x+\beta$ and $y = x-\alpha$ respectively, α and β being natural numbers, whereas $T^x = T$. In particular the generating function used by Barnard corresponds to the case $\alpha = 1$, $\beta > 1$. More general cases were studied by

D.Blackwell and M.A.Girshick, G.Blom, M.A.Girshick and J.H.B.Kemperman. The latter considered the general i.i. stochastic process in a Euclidean n-dimensional space, and studied in greater detail the case where E is the set of all integers and $P_y^x = p_{y-x}$ arbitrary (but, of course, $\geqq 0$ with $\sum_{-\infty}^{+\infty} p_x = 1$). Also the case that a wandering point will come somewhere in a set A_1 without having been in a set A_2 before, is contained in our general formula, by taking $A = A_1 \cup A_2$, $X = A_1$. Some applications of the use of non-constant T^x will be given in section 6.

4. Two Absorbing Regions.

We shall now establish a relation between the collective matrix belonging to an absorbing region A_1 , with the one belonging to a second absorbing region A_2 , contained in A_1 . We may think of two kinds of wandering particles, those of the first kind being absorbed in A_1 , those of the second kind in A_2 only. We shall denote these two collective matrices by $C_{(A_1)}$ and $C_{(A_2)}$ respectively, and we shall prove:

Theorem 1. If $A_1 \supset A_2$, then

$$C_{(A_1)} C_{(A_2)} = C_{(A_2)} = C_{(A_2)} C_{(A_1)} \quad . \tag{4.1}$$

We remark that the theorem applies if $A_1 = A_2$, and then states that C is idempotent (a "projection-operator"). The equation 4.1 suggests an analogy between the matrices $C_{(A)}$ and spectral operators. These matrices however are not additive in the sets A .

Proof: The second equality in 4.1 is trivial. It makes use of the fact only that by 3.11

$$C_{(A_2)} = C_{(A_2)} /_{A_2} , \tag{4.2}$$

and that by 3.5 and 3.6

$$C_{(A_1)} = /_{A_1} + /_{B_1} C_{(A_1)} \tag{4.3}$$

Now the definition of matrix multiplication 8.37 implies for the diagonal matrices $/_X$ and $/_Y$

$$/_X /_Y = /_{X \cap Y} \tag{4.4}$$

hence, as $A_2 \subset A_1$ and therefore $B_1 \subset B_2$, $A_1 \cap B_1 = A_2 \cap B_2 = O$, $A_1 \cup B_1 = A_2 \cup B_2 = E$, $A_2 \cap B_1 = O$, $A_1 \cup B_2 = E$

$$/_{A_1} /_{A_2} = /_{A_2} /_{A_1} = /_{A_2} \tag{4.5}$$

$$/_{B_1} /_{B_2} = /_{B_2} /_{B_1} = /_{B_1} \tag{4.6}$$

$$/_{A_2} /_{B_1} = /_{B_1} /_{A_2} = 0 \tag{4.7}$$

$$/_{A_1} /_{B_2} = /_{B_2} /_{A_1} = /_{A_1} - /_{A_2} = /_{B_2} - /_{B_1} \tag{4.8}$$

From 4.2, 4.3, 4.5, 4.7 we obtain at once

$$C_{(A_2)} C_{(A_1)} = C_{(A_2)} /_{A_2} /_{A_1} + C_{(A_2)} (/_{A_2} /_{B_1}) C_{(A_1)} =$$

$$= C_{(A_2)} /_{A_2} + 0 = C_{(A_2)} .$$

For the first part of the equation 4.1 we shall give two proofs. The second proof, which is purely algebraic like the previous one, will be given later and in a slightly generalized form so that it yields simultaneously another result.

The first proof is based on the probabilistic interpretation of $C_{(A_1)}$ and $C_{(A_2)}$.

$C_{(A_2)}{}^{x}_{X}$ is the probability of the following event $\mathcal{A}$: a point, starting in x, arrives ultimately in $A_2 \cap X$ without a catastrophe having happened and is absorbed there. Now consider the region $D \overset{df}{=} A_1 \cap B_2$, i.e. the part between A_1 and A_2. If $\mathcal{A}$ happens, then the wandering point may either have or have not passed through a point of D. The two cases being exclusive and exhausting, $C_{(A_2)}{}^{x}_{X}$ is the sum of the corresponding probabilities. In the first case the probability is the same as if the larger region A_1 had been the absorbing one, and it being required that the point be absorbed, not only in X, but in $A_2 \cap X$. Hence its probability is $C_{(A_1)}{}^{x}_{A_2 \cap X} = (C_{(A_1)}/A_2)^{x}_{X}$. In the second case there is a point $y \in D$, where the wandering point comes for the <u>first</u> time in D. The probability that it will arrive in any "small set" dy is the probability of being absorbed there if A_1 had been the absorbing region, i.e. $(C_{(A_1)})^{x}_{dy}$. This must be multiplied with the probability that the point goes from y to a point in $X \cap A_2$, A_2 being the absorbing region (always without a catastrophe), i.e., $C_{(A_2)}{}^{y}_{X}$, and integrated over y. Together we obtain

$$C_{(A_2)}{}^{x}_{X} = C_{(A_1)}{}^{x}_{A_2 \cap X} + \int_D C_{(A_1)}{}^{x}_{dy}\; C_{(A_2)}{}^{y}_{X}\,, \tag{4.9}$$

or in matrix-notation

$$C_{(A_2)} = C_{(A_1)}\, /_{A_2} + C_{(A_1)}\, /_D\, C_{(A_2)} . \qquad (4.10)$$

But $/_D = /_{B_2} - /_{B_1}$, $C_{(A_1)} /_{B_1} = C_{(A_1)} /_{A_1} /_{B_1} = 0$, hence

$C_{(A_2)} = C_{(A_1)} (/_{A_2} + /_{B_2} C_{(A_2)}) = C_{(A_1)}\, C_{(A_2)}$ by the identity analogous with 4.3. Hence the theorem has been proved.

Before giving the second proof we consider a very special case. Let E be the set of all integers. The integrations can then (and generally if E is countable) be replaced by summations and we know the matrices and functions of the second kind if we know their values for the one element sets X only. Let B_2 be the interval $-a < x < b$, and B_1 the interval $-d < x < b$, where $-a < -d < 0 < b$ (a, b and d are positive integers). Further we take $x = 0$ and X the set with the one element $-\ell$, with $-\ell \leq -a$ (ℓ integer). Let further the process be invariant, so that $P_y^x = P_{y+k}^{x+k} = p_{y-x}$ for all integers x, y and k. Then 4.9 becomes with $y = -j$:

$$C_{(A_2)}{}_{-\ell}^{\,0} = C_{(A_1)}{}_{-\ell}^{\,0} + \sum_{d}^{a-1} C_{(A_1)}{}_{-j}^{\,0}\, C_{(A_2+j)}{}_{-\ell+j}^{\,0} \qquad (4.11)$$

where $\qquad A_2 + j = Ens\{x + j \,/\, x \in A_2\}$.

With a different notation and $T^x =$ constant this relation was found by J.H.B.Kemperman in his Amsterdam Ph.D.-thesis (p. 71).

By means of 4.10 $C_{(A_2)}{}_X^x$ can be found for all x and X if the matrix $C_{(A_1)}$ is completely known, and $C_{(A_2)}{}_X^x$ for $x \in D$ only. By iteration like in section 3, however, we can express $C_{(A_2)}$ completely by $C_{(A_1)}$.

As $C_{(A_1)} 1_D C_{(A_2)} = C_{(A_1)} 1_{B_2} C_{(A_2)} = C_{(A_1)} 1_{B_2} T P C_{(A_2)}$

by 3.6, 4.10 is equivalent with

$$C_{(A_2)} = C_{(A_1)} 1_{A_2} + C_{(A_1)} 1_D T P C_{(A_2)} . \tag{4.12}$$

By means of the same iteration process as used before, this yields

$$C_{(A_2)} = C_{(A_1)} \sum_0^{\infty} (1_D T P C_{(A_1)})^n 1_{A_2} , \tag{4.13}$$

or, by 8.50 if $\|T\| < 1$:

$$C_{(A_2)} = C_{(A_1)} (1 - 1_D T P C_{(A_1)})^{-1} 1_{A_2} \tag{4.14}$$

For the special case $A_1 = E$ hence $B_1 = 0$, $D = B_2$, $C_{(A_1)} = 1$, this leads back to 3.10.

The solution 4.14 is particularly simple if D consists of one point α only. Denoting the middle factor in the last member of 4.14 by $1 + Q$, we have thus

$$(1 - 1_\alpha T P C_{(A_1)})(1 + Q) = 1 \tag{4.15}$$

or

$$Q = 1_\alpha T P C_{(A_1)} + 1_\alpha T P C_{(A_1)} Q , \tag{4.16}$$

showing that Q_X^x vanishes unless $x = \alpha$, in which case

i.e. $$Q_X^\alpha = T^\alpha (P C_{(A_1)})_X^\alpha + T^\alpha (P C_{(A_1)})_\alpha^\alpha Q_X^\alpha ,$$

$$Q^{\alpha}_{X} = T^{\alpha}\{1 - T^{\alpha}(PC_{(A_1)})^{\alpha}_{\alpha}\}^{-1}(PC_{(A_1)})^{\alpha}_{X} . \qquad (4.17)$$

The solution 4.14 then becomes

$$C_{(A_2)}{}^{x}_{X} = C_{(A_1)}{}^{x}_{A_2 \cap X} + C_{(A_1)}{}^{x}_{\alpha} Q^{\alpha}_{A_2 \cap X} . \qquad (4.18)$$

If E is countable, 4.17 and 4.18 give a recursive method for computing successively the $C_{(A_n)}$ if we take $B_0 = 0$ (whence $C_{(A_0)} = 1$), and add successively one element each time to B_n .

We assume, for simplification of the notation, that the elements $x, y, z, \ldots$ of E are the integers $1, 2, 3, \ldots$ themselves, in the order in which they are taken into B , so that $B_n = \{1, \ldots, n\}$, $A_n = \{n+1, n+2, \ldots\}$ and abbreviate $C_{(A_n)}$ by C_n . Moreover we remark that for $y \geqq n$

$$(PC_{n-1})^{n}_{y} = \sum_{1}^{\infty} {}^{z} P^{n}_{z}(C_{n-1})^{z}_{y} = \sum_{1}^{n-1} {}^{z} P^{n}_{z}(C_{n-1})^{z}_{y} + P^{n}_{y}$$

as $(C_{n-1})^{x}_{y} = 0$ if $x \geqq n$ unless $x = y$. The recursive relation then becomes $(C_n)^{x}_{y} = 0$ if $y \leqq n$ and

$$(C_n)^{x}_{y} = (C_{n-1})^{x}_{y} + (C_{n-1})^{x}_{n}\{1 - T^{n}P^{n}_{n} - T^{n}\sum_{1}^{n-1} {}^{z} P^{n}_{z}(C_{n-1})^{z}_{n}\}^{-1} T^{n}\{P^{n}_{y} + \sum_{1}^{n-1} {}^{z} P^{n}_{z}(C_{n-1})^{z}_{y}\} , \qquad (4.19)$$

if $y \geqq n+1$,

where we remind that $(C_{n-1})^{x}_{y}$ and $(C_n)^{x}_{y}$ are functions of the infinite number of independent variables $T^1, T^2, T^3, \ldots$ (the superfixes are not exponents!)

5. Second Proof and Wald's Fundamental Identity.

The second proof of the first equation of 4.1 can be given in several ways, and consists just in verifying 4.1, 4.10 or 4.12 by means of 3.5, 3.9 or 3.10, applied to $C_{(A_1)}$ and $C_{(A_2)}$. We choose the following form. For 3.5 we can write $1-C = 1_B - 1_B TPC$ or

$$1_B(1-TP) = (1-1_B TP)(1-C)\ , \tag{5.1}$$

or if $\|T\|<1$,

$$1-C = (1-1_B TP)^{-1} 1_B (1-TP)\ . \tag{5.2}$$

Hence, if we multiply both members from the right with any bounded matrix D , we find that

$$CD = D \tag{5.3}$$

if D satisfies the identity

$$1_B(1-TP)D = 0\ . \tag{5.4}$$

Now taking like before $C = C_{(A_1)}, B = B_1\ , D = C_{(A_2)}$, 5.4 is satisfied. For, by 3.6 $1_{B_2} C_{(A_2)} = 1_{B_2} TPC_{(A_2)}$ if both members are multiplied from the left with 1_{B_1} , 4.6 proves 5.4. Hence 5.3, i.e. the first part of 4.1 is true.

The identities 5.1, 5.2 lead to other interesting results. As the set X on which D^x_X depends is irrelevant as long as D enters into the equations as a last (right hand) factor only, we can just as well replace it by a function f^x of the first kind. Then formally our previous result 5.3, 5.4 becomes: f^x satisfies the equation

$$\int C^x_{dy} f^y = f^x \tag{5.5}$$

(i.e. f is an eigenfunction of C belonging to the eigenvalue 1), if $1_B^x f^x = 1_B^x T^x \int P_{ay}^x f^y$

i.e. if

$$f^x = T^x (Pf)^x \text{ for } x \in B . \tag{5.6}$$

On the other hand, if 5.5 is satisfied, 5.2 shows that also 5.6 holds. The result that 5.5 is implied by 5.6 is valid if f is bounded, and the series occurring implicitly in C and in the right hand member of 5.2 are uniformly convergent. For this it is sufficient that $\| 1_B T P 1_B \| < 1$, i.e. that $|T^x| < (P_B^x)^{-1}$ for every x in B , or, more generally, that $\| (1_B T P 1_B)^n \| < 1$ for some n.

In the most important application, however, f is not bounded, and then associativity has to be ascertained in another way, say by the conditions of Lemma 4 (Appendix § 8). We have then:

Theorem 2. If a $f_0^x \geq 0$ and a $T_0^x \geq 0$ exist such that $\| P f_0 \|^B < \infty$, that

$$f_0^x \geq T_0^x (P f_0)^x \quad \text{if } x \in B , \tag{5.7}$$

and that

$$\| 1_B T_0 P 1_B \| = \| 1_B T_0 P_B \| = \| T_0 P_B \|^B \leq 1 , \tag{5.8}$$

then for all T^x and f^x with $|f^x| \leq c f_0^x$ and with $|T^x| \leq \theta T_0^x$ for all $x \in B$ with $0 \leq \theta < 1$, c being a constant, the equations 5.5 and 5.6 are equivalent. [5]

Proof: We have by 5.8 for $n \geq 1$: $\| (1_B T_0 P)^n 1_B \| \leq 1$,

$$\| \sum_0^\infty (1_B |T| P)^n \| \leq 1 + \| \sum_0^\infty (1_B |T| P)^n 1_B |T| P \| \leq$$

---

5 Cf. J.H.B.Kemperman, theorem 1 and 2 (p. 13, 14) for the case, where E is the real axis, the process invariant (but not necessarily stationary and A a half-line.

$$\leq 1 + \sum_0^\infty \theta^{n+1} \| 1_B T_o P 1_B \|^n \| 1_B T_o P \| \leq 1 + \theta(1-\theta)^{-1} \| 1_B T_o P \| ,$$

so that C and the right member of 5.2 exist absolutely, and are, moreover, bounded. Hence, with $R \overset{df}{=} \sum_0^\infty (1_B/T/P)^n$, R is bounded and ≥ 0 , and $(R/f/)^x$ exists for all x as it is

$$\leq \left(\sum_0^\infty (1_B/T/P)^n/f/\right)^x \leq cf_o^x + c\theta \sum_0^\infty \theta^n (1_B T_o P 1_B)^n (1_B T_o P) f_o^x \leq$$

$$\leq c_o f_o^x + c\,\theta(1-\theta)^{-1} \| 1_B T_o P \| f_o^x .$$

In the same way $(R/T/P/f/)^x$ exists, hence also $(R/1-TP//f/)^x$, and, similarly $(Cf)^x$. Hence the conditions of Lemma 4 (Appendix § 8) with $E=E'$, R_Y^x instead of F_Y (for any fixed x), $1_B(1-TP)$ instead of P , are satisfied and we have

$$f - Cf = (1-C)f = \{R\,1_B(1-TP)\} f = \{R\; 1_B(1-TP)f\} = 0$$

if 5.6 holds. The transition from 5.5 to 5.6 is similar and even simpler.

We now shall apply this result to the case of an i.i. process. We take f^x to be of the form

$$f^x = e^{\xi x} \qquad \text{and } T^x = T = \text{ constant} \tag{5.9}$$

ξ and T being real or complex numbers.

Then with (cf 2.8)

$$P_X^x = \int p_{dv}\, 1_X^{x+v} \tag{5.10}$$

we have

$$(Pf)^x = \int p_{dv} f^{x+v} = \int p_{dv}\, e^{\xi(x+v)} = \varphi(\xi) f^x \tag{5.11}$$

where

$$\varphi(\xi) \overset{df}{=} \int p_{dv}\, e^{\xi v} \tag{5.12}$$

is the characteristic function of the distribution function. The interchange of the integrations is certainly allowed if the integrals in 5.11 converge absolutely, i.e. if 5.12 converges absolutely, i.e., if $\varphi(Re\xi)$ exists. Hence we have proved

Theorem 3. For each ξ for which $\varphi(Re\xi)$ exists, $f^x = e^{\xi x}$ is an eigenfunction of the i.i. matrix P, satisfying 5.10, belonging to the eigenvalue $\varphi(\xi)$.

Substitution of 5.9, 5.11 into 5.6 shows that the latter equation is satisfied if and only if

$$T\varphi(\xi) = 1 \qquad (5.13)$$

(as soon as B is not empty), and then holds for all $x \in E$ (not only $\in B$).

In order to apply theorem 2 we choose $f_0^x = e^{\xi_0 x}$, ξ_0 being real and such that Pf_0, hence $\varphi(\xi_0)$ exists, and $T_0^x = T_0 =$ const., real, > 0, and $\leq (\|P_B\|^B)^{-1}$. The last condition, hence 5.8, is always satisfied if $T_0 \leq 1$. Condition 5.7 (with equality sign) is satisfied if

$$T_0\,\varphi(\xi_0) \leq 1 \qquad (5.14)$$

Further we take a T and ξ satisfying 5.13 with $|T|\,T_0^{-1} = \theta < 1$, and such that $(\xi_0 - Re\,\xi)x \geq -\ln c$ for all x in B. Then, writing

$$C = \sum_0^\infty T^n C_{(n)}, \qquad (5.15)$$

so that $C_{(n)X}^{\;x}$ is the probability of absorption in X following the n-th step, we have

$$(f^{-1}Cf)^x = e^{-\xi x}\sum_0^\infty T^n \int C_{(n)\,dy}^{\;x}\, e^{\xi y}, \qquad (5.16)$$

and we find that 5.13 entails, for $x = 0$, the left member of 5.16 being equal to 1:

$$\sum_{1}^{\infty} \varphi(\xi)^{-n} \int C_{(n)} \overset{o}{d y} \, e^{\xi y} = 1 . \tag{5.17}$$

The identity 5.17 is known as Wald's Fundamental Identity. We have seen that it is a special case of the identity $Cf = f$, which holds for all f, satisfying $f = TPf$ on B under the conditions of theorem 2.

A partial generalization for arbitrary stochastic processes is considered in the Appendix (§ 10).

6. The Problem of Loops

We consider a stationary process without absorption. As, however, we want to consider properties of initial segments of given length n of the path, we have to admit discontinuation of the process at any moment. We start therefore with the expression 3.1 and substitute

$$B^x = B = const. \quad , \quad A^x = 1 - B \tag{6.1}$$

where B is an auxiliary variable. Moreover, we take $U^x = T^x$ and a fixed initial point $x = \alpha$. We obtain

$$C^{\alpha} = \sum_{0}^{\infty} B^n (1 - B) \big((TP)^n T \big)^{\alpha} \tag{6.2}$$

so that not C^{α} but $C^{\alpha}/(1-B)$ is the generating function with B as variable.

For some applications of the theory of Markoff and other stochastic processes, e.g. in the chemical statistics of longchain molecules (cf.e.g. G.King 1948, 1949, E.W.Montroll, 1950, Ch.M.Tchen, 1951, J.J.Hermans, M.S.Klamkin and R.Ullman, 1952 and the literature quoted there) it may be of some use to have methods by which the occurrence of "loops" in the path, i.e. of returns of the wandering point to a state where it has been before, can be studied. Without going into these applications and without prejudging its workability or its ability to yield non-trivial results in practically important cases, we shall outline sich a method here, which in any case might give a line of attack of this and similar problems, and is apt to be generalized for adaptation to other problems.

Evidently the probability of coincidence of $\underline{x}_k$ and $\underline{x}_\ell$ for $k \neq \ell$ is zero if the $\underline{x}_k$ have continuous distributions. In order to avoid irrelevant complications we shall assume, that these distributions are purely discontinuous, i.e. E is an enumerable set. In some of the applications mentioned E is such a set, namely a point-lattice (e.g. a tetrahedral lattice) in space. According to Appendix 8.1 the transition-matrix $P^x_{\mathcal{X}}$ then is completely determined by the ordinary (infinite) matrix P^x_y $(x \in E, y \in E)$, and the integrals like $\int F_{\alpha x} f^x$ pass into (in general infinite) sums $\sum F_x f^x$.

C^a now is an analytic function of an enumerable infinity of variables T^x , i.e. of a vector in a infinitely dimensional space. The partial derivatives of a function of the T^x with respect to these variables can be considered as the components of another vector, the gradient. We shall denote the differentiation-operator by ∇ with

$$\nabla_x = \frac{\partial}{\partial T^x} \qquad (6.3)$$

The suffix is written as a lower one because 1° in vector-analysis the derivatives with respect to a contravariant vector form a covariant vector, 2° the generalization of 6.3 to non-enumerable E will be found to lead to functions of the second kind.

If

$$C = \sum_0^\infty p_n T^n \tag{6.4}$$

is an ordinary generating function (T^n being here the n-th power of T) of the probabilities $p_n = P[\underline{n} = n]$ of the values taken by a random variable $\underline{n}$, the successive derivatives of C with respect to T give for $T=1$ the successive factorial moments:

$$[C]_{T=1} = 1 \tag{6.5}$$

$$\left[\frac{dC}{dT}\right]_{T=1} = \sum n p_n = \mathcal{E}\underline{n} \tag{6.6}$$

$$\left[\frac{d^2C}{dT^2}\right]_{T=1} = \sum n^{!2} p_n = \mathcal{E}\underline{n}^{!2} \tag{6.7}$$

where $\mathcal{E}$ is the expectation-symbol, whereas

$$x^{!k} \overset{df}{=} x(x-1)\ldots(x-k+1) \tag{6.8}$$

denotes the k-th "factorial power" of x.

By writing out the implicit matrix-summations in 6.2 we get an infinite sequence of terms, each of which is the probability of a definite path of length n, multiplied by $T^{x_0}\ldots T^{x_n}B^n(I-B)$. If the path passes k times through the state x, application of ∇_x and subsequent substitution of $T=1$ gives k times the probability of the path. Hence, if $\underline{n}_x$ denotes the number of times a path passes through x, we have

$$[\nabla_x C^a]_{T=1} = \mathcal{E}\underline{n}_x \tag{6.9}$$

In the same way we find

$$\left[\nabla_y \nabla_x C^{\alpha}\right]_{T=1} = \mathcal{E}\underline{n}_y \underline{n}_x - \mathcal{E}\underline{n}_y \delta_{yx} = \begin{cases} \mathcal{E}\underline{n}_x^{\,/2} & \text{if } y = x \\ \mathcal{E}\underline{n}_y \underline{n}_x & \text{if } y \neq x \end{cases} \tag{6.10}$$

etc.

It follows that

$$\sum_{x \in E} \left[\tfrac{1}{2} \nabla_x \nabla_x C^{\alpha}\right]_{T=1} = \tfrac{1}{2} \mathcal{E} \sum_{x \in E} \underline{n}_x^{\,/2} . \tag{6.11}$$

Now if $\underline{n}_x = 0$ or $\underline{n}_x = 1$, then $\underline{n}_x^{\,/2} = 0$; if $\underline{n}_x = 2$, i.e. if x is a double point of the path, then $\frac{1}{2}\underline{n}_x^{\,/2} = 1$. Hence 6.11 gives the expectation of the number of double-points in a path, a triple-, quadruple-, k-uple point being counted, as usual in algebraic geometry, as 3, 6, $\frac{1}{2}k(k-1)$ double-points. Hence, we have proved

Theorem 6: If $\underline{D}$ denotes the random variable which on each path equals the number of double-points or loops, counted in the aforesaid manner, we have

$$\mathcal{E}\underline{D} = \left[\tfrac{1}{2} \Box\, C^{\alpha}\right]_{T=1} \tag{6.12}$$

where $\Box$ denotes the generalized Laplacian operator

$$\Box \overset{df}{=} \sum_{x \in E} (\nabla_x)^2 = \sum_{x \in E} \left(\frac{\partial}{\partial T^x}\right)^2 . \tag{6.13}$$

Similar relations, of course, hold for the number of triple-points, etc.

In the derivation of 6.12 we used the fact only that C^{α} is a power series in the T^x, but not its special form 6.2. Hence the result holds for arbitrary non-stationary and also non-Markovian processes also. In the special case 6.2 the left members of 6.9, 6.10 and 6.11 are easily computed. In fact, we have

$$\nabla_x((TP)^n T) = \sum_1^n (TP)^m 1_x (PT)^{n-m} \tag{6.14}$$

$$\nabla_y \nabla_x ((TP)^n T) = \sum_{k+\ell+m=n}^{k,\ell,m} (TP)^k 1_y P(TP)^{\ell-1} 1_x (PT)^m + \sum_{k+\ell+m=n}^{k,\ell,m} (TP)^k 1_x P(TP)^{\ell-1} 1_y (PT)^m. \tag{6.15}$$

Hence we obtain

$$\left[\tfrac{1}{2} \square C^\alpha\right]_{T=1} = (1-B) \sum^{k,\ell,m} B^{k+\ell+m+2} \sum_{x \in E} (P^k)_x^\alpha (P^{\ell+1})_x^x (P^{m+1})_E^x \ . \tag{6.16}$$

But $(PT)^x = \int P_{dy}^x T^y = P_E^x = 1$ if $T^x = 1$ for all x . Hence $\sum_1^\infty B^m (P^m)_E^x =$ $= B(1-B)^{-1}$ and 6.12, 6.16 may also be written as

$$\mathcal{E}\underline{D} = B \sum_{x \in E} \Phi_x^\alpha (\Phi_x^x - 1) \tag{6.17}$$

where

$$\Phi_y^x \overset{df}{=} \left(\sum_0^\infty B^n P^n\right)_y^x = \left((1-BP)^{-1}\right)_y^x \ . \tag{6.18}$$

In the special case, where the stationary Markoff process is an invariant one, 6.17 simplifies still further. For in that case Φ_x^x is independent of x , so that the summation is extended over Φ_x^α only and gives $\Phi_E^\alpha = (1-B)^{-1}$. Hence for an invariant process 6.17 becomes

$$\mathcal{E}\underline{D} = B(1-B)^{-1}(\Phi_0^0 - 1) \ . \tag{6.19}$$

Expressed in the conditional expectations $\mathcal{E}_{(n)}\underline{D}$ for given n , the left member of 6.19 is $\sum_0^\infty B^n (1-B) \mathcal{E}_{(n)} \underline{D}$, so that 6.19 shows that $\mathcal{E}_{(n)}\underline{D}$ is the coefficient of B^n in $B(1-B)^{-2}(\Phi_0^0 - 1)$, hence

$$\mathcal{E}_{(n)}\underline{D} = \sum_1^n (n-\ell)(P^\ell)_0^0 \ . \tag{6.20}$$

Here $(P^m)^o_o$ is the probability that a path of length m will be closed, i.e. that its first and its last $((m+1)^{th})$ state coincide (whether it contains double points or not). Like so many simple specializations of general theorems, however, this result is trivial: in a path of length n a loop can have any length ℓ , $1 \leq \ell \leq n$, and, if so, it can begin at the $1^{st}, \ldots, (n-\ell)^{th}$ point; in the case of an invariant process these $n-\ell$ cases have equal probabilities which are equal to $(P^\ell)^o_o$, i.e. the value of the ℓ -fold convolute of p_v (cf. 2.8) for the value $v = 0$.

The method scetched here can be generalized for arbitrary stochastic processes and applied to other problems than the expectation of the mean number of loops.

For an arbitrary functional Φ of a function T^x we define (cf. Hadamard 1910, Fréchet 1912, 1914, Van Dantzig 1935)

$$\nabla_X \Phi(T) = \left(\frac{\partial}{\partial T}\right)_X \Phi(T) = \lim_{\varepsilon \searrow 0} \frac{\Phi(T+\varepsilon 1_X) - \Phi(T)}{\varepsilon} \tag{6.21}$$

if this limit exists. Under certain regularity conditions, which are e.g. satisfied in the case of an analytic Φ , this quantity is a countably additive function of the set X . The dual definition for a functional Ψ of a function V_x of the second kind would be

$$\left(\frac{\partial}{\partial V}\right)^x \Psi(V) = \lim_{\varepsilon \searrow 0} \frac{\Psi(V+\varepsilon 1^x) - \Psi(V)}{\varepsilon} \tag{6.22}$$

yielding for sufficiently regular Ψ a function of the first kind. For a matrix P we can define

$$\left(\frac{\partial}{\partial P}\right)_A^\alpha \Psi(P) = \lim_{\varepsilon \searrow 0} \frac{\Psi(P+\varepsilon 1_A 1^\alpha) - \Psi(P)}{\varepsilon} \tag{6.23}$$

where $({}_{A}{}^{a})^{x}_{X} = {}^{x}_{A}{}^{a}_{X}$ is to be distinguished from ${}^{a}{}_{A} = \int {}^{a}_{dx}{}^{x}_{A} = {}^{a}_{A}$. In this paper we shall have to do with 6.21 only. Applying the operator ∇_X to the general stochastic process

$$C = (1-B)\sum_{0}^{\infty} B^n \iint\ldots\int P^{(n)}_{dx_0\ldots dx_{n-1}} T^{x_0}\ldots T^{x_{n-1}}, \qquad (6.24)$$

from which the special case (6.2) is reobtained by the specialization (2.5) with a $P_{(n)} = P$ independent of n and $P_{(0)X} = {}^{a}_{X}$, we obtain:

$$(\nabla_X C)_{T=1} = \mathcal{E}\, \Omega_X , \qquad (6.25)$$

where the random variable Ω_X for any given X is the number of times that the process passes through a state in X. For $X = E$ $\Omega_E = \Omega$ is the number of steps made until the process is stopped.

In the same way we find, as a generalization of 6.10

$$(\nabla_Y \nabla_X C)_{T=1} = \mathcal{E}(\Omega_X \Omega_Y - \Omega_{X\cap Y}) \qquad (6.26)$$

In the case of continuous distributions of the jumps, the problem of loops becomes trivial: their probability then is zero. We can however consider related problems, e.g. the question, how often a path returns into states within a given distance from one of its previous states, without any pretention that this can be used for the solution of the generalized problem of loops ("almost loops"), which to give a precise form seems itself a rather hard problem.

We shall suppose here that E is a Euclidean space, or, more generally, a metrical space, where any two states x and y have a distance, denoted by r^{xy}. More generally we may take any two-state function f^{xy} (apart from conditions of integrability), and form the operator

$$\Box_{(f)} \overset{df}{=} \iint f^{xy} \nabla_{dx} \nabla_{dy} \qquad (6.27)$$

which passes for the case of a finite set E into the generalized Laplacian

$$\square_{(f)} = \sum^i \sum^j f^{ij} \frac{\partial}{\partial T^i} \frac{\partial}{\partial T^j} , \tag{6.28}$$

where the f^{ij} are constants (independent of the T^i). Under sufficient regularity conditions the order of the derivations may be interchanged, so that 6.27 then vanishes identically for an antisymmetrical f ($f^{xy} = -f^{yx}$). We shall therefore restrict ourselves to symmetrical f ($f^{xy} = f^{yx}$).

Provided the interchange of integrations is allowed, 6.26 gives:

$$\left(\tfrac{1}{2} \square_{(f)} C\right)_{T=1} = \tfrac{1}{2} \mathcal{E} \iint n_{dx}\, n_{dy}\, f^{xy} - \tfrac{1}{2} \mathcal{E} \int n_{dx}\, f^{xx} . \tag{6.29}$$

Here the integrals after the second and the first expectation-sign respectively become for a path of length $n-1$, passing, $n_1, \dots, n_r$ times respectively, through r different points $x_1, \dots, x_r$, (so that $n_X = \sum_1^r {}^i\, n_i\, 1_X^{x_i}$)

$$\int n_{dx}\, f^{xx} = \sum_1^r {}^i\, n_i\, f^{x_i x_i} \tag{6.30}$$

$$\iint n_{dx}\, n_{dy}\, f^{xy} = \sum_1^r {}^i \sum_1^r {}^j\, n_i\, n_j\, f^{x_i x_j} . \tag{6.31}$$

Hence the righthand member of 6.29 is the expectation of

$$S_{(f)} \overset{df}{=} \tfrac{1}{2} \sum\sum_{i \neq j} n_i\, n_j\, f^{x_i x_j} + \sum_i \tfrac{1}{2}\, n_i (n_i - 1)\, f^{x_i x_i} . \tag{6.32}$$

This quantity is the sum of the $f^{x_i x_j}$ ($i, j = 1, 2, \dots, n$) over all pairs $i \neq j$, together with the sum of the $f^{x_i x_i}$ over all multiple points, counted as multiple double points like before.

The second term vanishes identically

1° if $f^{xx} = 0$ for all x, e.g. if f^{xy} is the distance r^{xy} of x and y, 2° if $\underline{n}_i$ takes the values 0 and 1 only (except for a probability 0).

In the latter case, which occurs always, if the transition probabilities have continuous distributions, we can omit the $\underline{x}_i$ with $\underline{n}_i = 0$, so that $\underline{n}_i = 1$ for all i and 6.32 becomes

$$\underline{S}_{(f)} = \frac{1}{2} \sum_{i \neq j} \sum f^{\underline{x}_i \underline{x}_j} \tag{6.33}$$

where the $\underline{x}_i$ are the $\underline{n} (= r)$ different points through which the path passes. Hence in this case the operator 6.29 gives the expectation of the sum of the values of f for all pairs of different points of the path.

If in particular f^{xy} is the distance r^{xy} of x and y, this is $\frac{1}{2} \underline{n} (\underline{n} - 1)$ times the mean distance taken over the path of any two states through which the wandering point passes.

If, instead of $f^{xy} = r^{xy}$, we take for f the characteristic function of any relation $R(x,y)$ between two states (e.g. $R(x,y) \overset{df}{=} (r^{xy} \leq a)$, where a is a non-negative number), i.e.

$$f^{xy} = \begin{cases} 1 & \text{if } R(x,y) \text{ holds} \\ 0 & \text{if not,} \end{cases} \tag{6.34}$$

then 6.33 becomes half the number of ordered pairs of different states in a path for which $R(x,y)$ holds, (e.g. which have a distance $\leq a$). In the case of a symmetrical relation this is the number of unordered pairs.

In the stationary Markovian case we again can easily compute the left member of 6.29. Like before we obtain for a symmetrical f^{xy}

$$\tfrac{1}{2}\,\square_{(f)} C = (1-B)\sum_{k,\ell,m} B^{k+\ell+m} \iint ((TP)^k)_{dx} (P(TP)^{\ell-1})^x_{dy}\, f^{xy}\, ((PT)^m)^y$$

For $T^x = 1$ this becomes, as $(PT)^y$ then also equals 1 ,

$$\tfrac{1}{2}\left(\square_{(f)} C\right)_{T=1} = (1-B)\sum_{k,\ell,m} B^{k+\ell+m} \iint (P^k)_{dx} (P^\ell)^x_{dy}\, f^{xy} . \qquad (6.35)$$

For the case of an invariant process, given by 2.8, this simplifies still further, if $f^{xy} = f^{x-y} = f^{y-x}$.
We have then

$$\iint (P^k)_{dx} (P^\ell)^x_{dy}\, f^{xy} = \int p^{(k)}_{dz} \int p^{(\ell)}_{dv} \int 1^z_{dx} \int 1^{x+v}_{dy}\, f^{y-x} = \int p^{(\ell)}_{dv}\, f^v ,$$

($p^{(\ell)}_z$ being the ℓ^{th} convolute of p_z), as $\int 1^z_{dx} = 1$ and $\int p^{(k)}_{dz} = 1$.
Hence the coefficient of $(1-B)B^n$, becomes, as ℓ must be $\geqq 1$,

$$\mathcal{E}_{(n)}\, \tfrac{1}{2} \sum_{i \neq j} f^{x_i, x_j} = \sum_{1}^{n} (n-\ell) \int p^{(\ell)}_{dv}\, f^v . \qquad (6.36)$$

Like before, this relation can easily be proved in an elementary way.

If f^v has the Fourrier transform $\chi(t)$:

$$f^v = \int e^{ivt} \chi(t)\, dt , \qquad (6.37)$$

6.36 becomes because of $\int p^{(\ell)}_{dv} e^{ivt} = (\varphi(t))^\ell$, where

$$\varphi(t) \overset{df}{=} \int p_{dv}\, e^{ivt} ,$$

$$\left.\begin{aligned} \mathcal{E}_{(n)} \tfrac{1}{2} \sum_{(i \neq j)} f^{x_i, x_j} &= \sum_{1}^{n} (n-\ell) \int \varphi(t)^{\ell} \chi(t)\, dt = \\ &= \int \frac{(n-1) - n\varphi(t) + \varphi(t)^{n}}{(1-\varphi(t))^{2}} \varphi(t) \chi(t)\, dt . \end{aligned}\right\} \quad (6.38)$$

For a given p and f , hence φ and χ , 6.38 gives the required expectation in the form of one integral.

APPENDIX

§7. Functions of the First and the Second Kind

According to Kolmogoroff a probability field on a set E is given by a countably additive set function on E, defined and non-negative for all sets belonging to a σ-field σ_E of subsets of E, and taking the value 1 on E.

More generally we shall, instead of a σ-field, consider a δ-field (according to the terminology used in Hahn-Rosenthal), i.e. a system δ_E of sub-sets of E which has the following properties:

i. δ_E is a field (i.e. contains with any two sets X and Y their union ("join") $X \cup Y$, their difference(s) which we denote by $X -- Y$ and $Y -- X$ respectively, hence also their intersection ("meet") $X \cap Y$;

ii. δ_E contains the intersection $\bigcap_1^\infty X_n$ of any sequence $\{X_n\}$ of sets X_n contained in it;

iii. E is the union of a sequence $\{E_n\}$ of sets belonging to δ_E.

It follows by replacing E_n by $\bigcup_1^n E_k$, that we can assume without restriction that

$$E_{n+1} \supset E_n, \quad E_n \in \delta_E \quad \text{for all } n, \tag{7.1}$$

whereas i and ii imply that the union $X = \bigcup_1^\infty X_n$ of a sequence $\{X_n\}$ of sets belonging to δ_E belongs to δ_E if (and only if) all X_n are contained in some $Y \in \delta_E$. For in that case $X \subset Y$ and

$$X = Y -- \bigcap_1^\infty (Y -- X_n)$$

Hence δ_E is a σ-field if and only if $E \in \delta_E$.

The use of a δ-field instead of a σ-field has the advantage that function-"values" $\pm\infty$ (cf. e.g. Hahn-Rosenthal, Halmos, etc) may be avoided.

Generalizing a terminology introduced on a previous occasion (D.van Dantzig, 1935) we define a function of the second kind as a real function F, defined and countably additive on δ_E. We denote the value which F takes on a set X by F_X. Calling a sequence of sets $X_n \in \delta_E$ which are mutually exclusive and which have $X \in \delta_E$ as union a dissection of X, and denoting this by $\{X_n\} \in D(X)$, hence

$$\{X_n\} \in D(X) \overset{df}{\equiv} \cup X_n = X \text{, and } X_m \cap X_n = O \text{ if } m \neq n \tag{7.2}$$

(O being the vacuous set), we then have

$$\{X_n\} \in D(X) \to \Sigma^n F_{X_n} = F_X \tag{7.3}$$

The absolute-value function of F which itself is a function of the second kind, is denoted by $|F|$, and defined by

$$|F|_X \overset{df}{\equiv} \sup_{\{X_n\} \in D(X)} \Sigma^n |F_{X_n}| . \tag{7.4}$$

On the other hand we define a function of the first kind as a real function f, defined for all $x \in E$, and bounded and δ-measurable on each $X \in \delta_E$. Hence, if we denote by f^x (instead of the customary $f(x)$) the value which f takes in x, and by $Ens\{x \in E / A(x)\}$ the set of all $x \in E$ for which the statement $A(x)$ holds, we have le.

$$X \in \delta_E \to \|f\|^X < \infty , \tag{7.5}$$

where

$$\|f\|^X \overset{df}{\equiv} \sup_{x \in X} |f^x| , \tag{7.6}$$

and 2e.

$$Ens\{x \in X / f^x \leq c\} \in \delta_E \tag{7.7}$$

for all real c and $X \in \delta_E$.

The set functions $\|f\|^X$ are in general not countably additive, hence not functions of the second kind.

We remark that $|F|_X$ and $\|f\|^X$ are finite if $X \in \delta_E$ and can be considered as "norms" of F and f on X if X is fixed. In particular, if $|F|_E$ and $\|f\|^E$ are finite, they have the ordinary properties of norms of F and f on E. In the latter case we shall sometimes omit the suffixes E.

More generally we may admit the values of F and f to be complex numbers, and in some cases even to be taken from arbitrary dual Banach spaces.

The integral of f with respect to F exists on each subset $X \in \delta_E$ and will be denoted by

$$\int_X F_{dx} f^x$$

or shortly by $(Ff)_X$. It is, according to well-known theorems, a function of the second kind, satisfying the inequality

$$|(Ff)|_X \leq |F|_X \|f\|^X \tag{7.8}$$

for any $X \in \delta_E$. Also according to known theorems we have:

$$|(Ff) - (\mathcal{G}g)|_X \leq |F - \mathcal{G}|_X \|f\|^X + |\mathcal{G}|_X \|f - g\|^X, \tag{7.9}$$

an inequality which is often applied. In particular, denoting by $var_X f$ the variation of f on X, i.e.

$$var_X f \overset{df}{=} \sup_{x \in X} f^x - \inf_{x \in X} f^x, \tag{7.10}$$

and for any dissection $\{X_n\}$ of X:

$$var_{\{X_n\}} f \overset{df}{=} \sup_n var_{X_n} f \tag{7.11}$$

we have, if $\{X_n\} \in D(X)$, $x_n \in X_n$ for all n :

$$|(Ff)_X - \sum_n F_{X_n} f^{x_n}| \leq |F|_X \, var_{\{X_n\}} f . \tag{7.12}$$

This follows from 7.9 with $G = F$ and $g^x = f^{x_n}$ if $x \in X_n$.

§8. Matrices

Let E and E' be two arbitrary sets, on each of which a δ -field δ_E and $\delta_{E'}$ of subsets is given, satisfying the conditions of § 7.

We consider a function P (to be called a --- generalized --- "matrix", or more explicitly an (E,E') -matrix), determining a real (occasionally a complex) number P_X^y as a function of 1°. an element $y \in E'$ and 2°. a subset $X \in \delta_E$, subject to the conditions:

1°. For fixed $X = A \in \delta_E$ P_A^y determines a function (denoted by P_A) of the first kind on E' ;

2°. For fixed $y = b \in E'$ P_X^b determines a function (denoted by P^b) of the second kind on E .

If in particular E is countably (finite, or enumerably infinite), and if δ_E contains every set consisting of one element, we have

$$P_X^y = \sum_{x \in X} P_x^y \tag{8.1}$$

and if E' also is enumerable, P_X^y is determined by the ordinary rectangular finite or infinite matrix P_x^y .

The norm for fixed y is given by

$$/P^y/_X \overset{df}{=} \sup_{\{X_n\}\in D(X)} \Sigma_n /P^y_{X_n}/. \tag{8.2}$$

It is itself a matrix. Its norm for fixed X is given by

$$/\!/P/\!/^Y_X \overset{df}{=} \sup_{y\in Y} /P^y/_X \tag{8.3}$$

and is not a matrix in the above sense, Y being a set instead of an element of E', and $/\!/P/\!/^Y_X$ moreover not being completely additive in X. The matrix P is called <u>bounded</u> if $/\!/P/\!/^Y_X$ is bounded, i.e. if

$$\sup_{X,Y} /\!/P/\!/^Y_X = /\!/P/\!/^{E'}_E < \infty . \tag{8.4}$$

If $E \notin \delta_E$, then $/\!/P/\!/^{E'}_X$ is not defined for $X = E$. In that case we take the equality 8.4 as a definition of $/\!/P/\!/^{E'}_E$.
Instead of the symbol $/\!/P/\!/^{E'}_E$ we shall also use the symbol $/\!/P/\!/$.

In particular the sets E and E' may coincide. In that case we shall suppose that the fields δ_E and $\delta_{E'}$ also coincide. A special case here is the "unit-matrix", denoted by $/$ (iota), and defined by

$$/^x_X \overset{df}{=} \begin{cases} 1 & \text{if } x \in X \\ 0 & \text{if not} \end{cases} \tag{8.5}$$

For fixed $X = A$ $/_A$ is the characteristic function of the set A; for fixed a $/^a$ might be called the "characteristic function of the second kind of the point a"; it takes the value 1 on each set containing a and 0 on each other one. Evidently $/$ is bounded with $/\!/ / /\!/ = 1$, and

$$/^x_{X_1} /^x_{X_2} = /^x_{X_1 \cap X_2} . \tag{8.6}$$

More generally with every subset $A \in \delta_E$ corresponds a matrix, also denoted by $/_A$, and defined by

$$(1_A)_X^x \overset{df}{=} 1_{A\cap X}^x \, . \tag{8.7}$$

In the case of finite $E = E'$ it corresponds according to 8.1 with a matrix having 1 in all elements of the main diagonal , the suffixes of which belong to A , and 0 in all other places. Evidently 1_A also is bounded and $\|1_A\| = 1$ unless A is empty.

With every function f of the first kind corresponds a matrix, to be denoted by $f1$, and defined by

$$(f1)_X^x \overset{df}{=} f^x 1_X^x = \begin{cases} f^x & \text{if } x \in X, \\ 0 & \text{if not} . \end{cases} \tag{8.8}$$

In the case of an enumerable $E = E'$ the matrix which corresponds to $f1$ according to 8.1 is a diagonal matrix, having the value f^x in the element of the main diagonal which corresponds with x . Evidently $f1$ is bounded if and only if f is, and $\|f1\| = |f|^E$.

We now return to the general case, where E and E' may differ. Then let P be a matrix like before, and f a function of the first kind on E . Then for any $X \in \delta_E$ and for any fixed $y \in E'$, writing $Pf1$ for $P(f1)$,

$$(Pf1)_X^y = \int_X P_{dx}^y f^x \tag{8.9}$$

exists and is countably additive in X for $X \in \delta_E$ and measurable and bounded in y on each $Y \in \delta_{E'}$, as

$$\|Pf1\|_X^Y \le \|P\|_X^Y \|f\|^X \tag{8.10}$$

so that $Pf1$ is a matrix. In particular $(Pf)^y = (Pf1)_E^y$, if it exists for all y , is a function of the first kind on E' , and is bounded if P and f are.

On the other hand, let F be a function of the second kind on E'. Then

Lemma 1. For any $X \in \delta_E$, $Y \in \delta_{E'}$

$$(FP)_{Y,X} \overset{df}{=} \int_Y F_{dy} P^y_X \tag{8.11}$$

exists and is a function of the second kind on δ_E and on $\delta_{E'}$.

Proof: The countable additivity with respect to Y on $\delta_{E'}$ for constant X follows from that for $\int_Y F_{dy} f^y$. For constant Y $(FP)_{Y,X}$ is trivally additive in X. For the countable additivity on δ_E it is therefore sufficient to prove that $\lim_{n\to\infty} (FP)_{Y,X_n} = 0$ if X_n is a decreasing sequence of sets $X_n \in \delta_E$ with empty intersection.

We choose an $\varepsilon > 0$ and define

$$B_n = Ens\{y \in Y / |P^y|_{X_n} \geqq \varepsilon\}.$$

Then, as $|P^y|$ is countably additive and $\geqq 0$ for each $y \in E'$, and as $X_{n+1} \subset X_n$, we have also $B_{n+1} \subset B_n$. Moreover $\bigcap_1^\infty B_n = 0$, for if there existed an $y \in \bigcap_1^\infty B_n$, then for this y $|P^y|_{X_n} \geqq \varepsilon$ for all n, contradicting the countable additivity of $|P^y|$, as $\bigcap_1^\infty X_n = 0$.

In the same way for all sufficiently large n, $|F|_{B_n} \leqq \varepsilon$. Now, if C_n is the complement Y--B_n of B_n with respect to Y then

$$(FP)_{Y,X_n} = (FP)_{B_n,X_n} + (FP)_{C_n,X_n}.$$

But

$$|(FP)_{B_n,X_n}| \leqq \int_{B_n} |F|_{dy} |P^y|_{X_n} \leqq |F|_{B_n} \|P\|^{B_n}_{X_n} \leqq \varepsilon C_1$$

with $C_1 = \|P\|^{B_n}_{X_n}$. And

$$|(FP)_{C_n,X_n}| \leqq \int_{C_n} |F|_{dy} |P^y|_{X_n} \leqq |F|_{C_n} \|P\|^{C_n}_{X_n} \leqq C_2\, \varepsilon$$

with $C_2 = |F|_Y$, as $C_n \subset Y$, and $\|P\|^{C_n}_{X_n} = \sup_{y \in C_n} |P^y|_{X_n} \leqq \varepsilon$ by definition of B_n and C_n. Hence $|(FP)_{Y,X_n}| \leqq (C_1 + C_2)\varepsilon$ for all sufficiently large

n , $\varepsilon > 0$ being given, i.e.

$$\lim_{n\to\infty} (FP)_{Y, X_n} = 0 , \quad QED .$$

It follows easily that for all X, Y

$$|(FP)_{Y,X}| \leq |F|_Y \, \|P\|_X^Y \tag{8.12}$$

If, in particular, F and P are bounded, then also FP is.

By means of the inequalities 8.10 and 8.12 we have

Lemma 2. A bounded matrix P transforms the bounded functions f of the first kind on E into such ones on E' and the bounded functions of the second kind on E' into such ones on E with

$$\|(Pf)\| \leq \|P\| \, \|f\| \tag{8.13}$$

$$|(FP)| \leq |F| \, \|P\| \tag{8.14}$$

(where the suffixes E and E' have been omitted).

Lemma 3. If f and F are functions of the first and second kind on E and E' respectively, and P is a matrix, then if $X \in \delta_E$ and $Y \in \delta_{E'}$,

$$\int_X \left\{ \int_Y F_{dy} P_{dx}^y \right\} f^x = \int_Y F_{dy} \int_X P_{dx}^y f^x , \tag{8.15}$$

i.e. the order of the integrations may be reversed.

Proof: If we put for arbitrary $U \in \delta_E$, $V \in \delta_{E'}$,

$$F'_V \overset{df}{=} F_{Y \cap V} \, , \; f'^x \overset{df}{=} f^x 1_U^x \; , \; P'^y_U \overset{df}{=} 1_Y^y P^y_{X \cap U} \tag{8.16}$$

then F' , P' and f' are bounded and 8.15 is equivalent with

$$\int_E \left\{ \int_{E'} F'_{dy} P'^y_{dx} \right\} f'^x = \int_{E'} F'_{dy} \int_E P'^y_{dx} f'^x . \tag{8.17}$$

Hence it is sufficient, to prove the theorem for bounded functions and matrices with the integrations extended over the whole sets E and E' (which we shall then omit). We omit the primes and we may write 8.17 in the abbreviated form

$$(FP)f = F(Pf). \tag{8.18}$$

We put then

$$Pf \overset{df}{=} g\,, \quad \text{i.e.} \quad g^y \overset{df}{=} \int_E P^y_{dx} f^x\,, \tag{8.19}$$

$$FP \overset{df}{=} \mathcal{G}\,, \quad \text{i.e.} \quad \mathcal{G}_X \overset{df}{=} \int_{E'} F_{dy} P^y_X\,, \tag{8.20}$$

and we have to prove

$$\mathcal{G}f = Fg\,, \quad \text{i.e.} \quad \int_E \mathcal{G}_{dx} f^x = \int_{E'} F_{dy} g^y. \tag{8.21}$$

Let $\varepsilon > 0$ be given. As f is bounded, we can find a <u>finite</u> dissection $\{X_n\} \in D(E)$ with $n \leq N$ and $var_{\{X_n\}} f \leq \varepsilon$.
(Divide e.g. the finite interval $(-|f|^E, +|f|^E)$ into N exclusive subintervals J_n , and take $X_n \overset{df}{=} Ens\{x \in E / f^x \in J_n\}$). Hence choosing the $x_n \in X_n$ arbitrarily, by 7.12

$$\Big| \int \mathcal{G}_{dx} f^x - \Sigma^n \mathcal{G}X_n f^{x_n} \Big| \leq |\mathcal{G}|\, \varepsilon. \tag{8.22}$$

As each of the N functions P_{X_n} is bounded, we can find a <u>finite</u> dissection of E' (obtained e.g. by intersecting the N dissections belonging to the P_{X_n} separately) $\{Y_m\} \in D(E')$ $(m \leq M)$ with $var_{\{Y_m\}} P_{X_n} \leq \varepsilon/N$ for all $n \leq N$. Hence, choosing the $y_m \in Y_m$ arbitrarily, again by 7.12:

$$\Big| \mathcal{G}_{X_n} - \Sigma^m F_{Y_m} P^{y_m}_{X_n} \Big| \leq |F|\, \varepsilon/N \tag{8.23}$$

and with 8.22 and $|\mathcal{G}| \leq |F| \, \|P\|$

$$\Big| \int \mathcal{G}_{dx} f^x - \Sigma^n \Sigma^m F_{Y_m} P^{y_m}_{X_n} f^{x_n} \Big| \leq C_1 \varepsilon \tag{8.24}$$

with $$C_1 \overset{df}{=} |F| \big(\|P\| + \|f\| \big).$$

On the other hand

$$\left| \int P^y_{dx} f^x - \sum^n P^y_{X_n} f^{x_n} \right| \leqq \|P\| \varepsilon \tag{8.25}$$

for every $y \in E'$, in particular for $y = y_m$.

Hence

$$\left| \int F_{dy} g^y - \sum^n \left(\int F_{dy} P^y_{X_n} \right) f^{x_n} \right| \leqq |F| \|P\| \varepsilon. \tag{8.26}$$

But

$$\left| \int F_{dy} P^y_{X_n} - \sum^m F_{Y_m} P^{y_m}_{X_n} \right| \leqq |F| \varepsilon/N \tag{8.27}$$

as $\mathrm{var}_{\{Y_m\}} P_{X_n} \leqq \varepsilon/N$. Hence, as $n \leqq N$ 8.25 - 8.27 give:

$$\left| \int F_{dy} g^y - \sum^n \sum^m F_{Y_m} P^{y_m}_{X_n} f^{x_n} \right| \leqq C_1 \varepsilon . \tag{8.28}$$

From 8.24 and 8.28 we obtain

$$\left| \int G_{dx} f^x - \int F_{dy} g^y \right| \leqq 2 C_1 \varepsilon . \tag{8.29}$$

Hence, as this holds for every $\varepsilon > 0$, 8.21, whence 8.17 and 8.15, QED.

We need a sufficient condition for 8.15 to hold over the whole sets E and E' where F, P and f may be unbounded and prove therefore along well-known lines:

Lemma 4. If

$$H_X = \int_{E'} |F|_{dy} |P^y|_X \tag{8.30}$$

$$h^y = \int_E |P^y|_{dx} |f^x| \tag{8.31}$$

exist for all X and y respectively, and if either

$$\int_E H_{dx} |f^x| < \infty , \tag{8.32}$$

or

$$\int_{E'} |F|_{dy} h^y < \infty , \tag{8.33}$$

then the other one of these two inequalities holds also, and

$$\int_E \left\{ \int_{E'} F_{dy} P^y_{dx} \right\} f^x = \int_{E'} F_{dy} \int_E P^y_{dx} f^x . \tag{8.34}$$

Proof: It is sufficient to prove the theorem for non-negative F, P and f, in which case $H = G$ and $h = g$. According to lemma 3, 8.15 holds for $X = E_k$, $Y = E'_\ell$ with arbitrary k and ℓ (cf. 7.1). Now, if 8.32 is satisfied, the left member of 8.34 is

$$\int_E G_{dx} f^x \geqq \int_{E_k} G_{dx} f^x \geqq \int_{E_k} \left\{ \int_{E'_\ell} F_{dy} P^y_{dx} \right\} f^x =$$

$$= \int_{E'_\ell} F_{dy} \int_{E_k} P^y_{dx} f^x .$$

As the latter expression is non-decreasing if $k \to \infty$, $\ell \to \infty$, and bounded, it has a limit, which is the second member of 8.34, so that its existence has been proved, i.e. 8.32 and $\int_E G_{dx} f^x \geqq \int_{E'} F_{dy} g^y$. In the same way it is proved that $\int_{E'} F_{dy} g^y \geqq \int_E G_{dx} f^x$, whence 8.34. Analogously in the second case, i.e. if 8.33 is assumed.

Lemma 5. (Cf. R.G.Cooke (1950) page 29). If, E_1, E_2, E_3, E_4 are arbitrary sets with corresponding δ-fields δ_{E_1}, δ_{E_2}, δ_{E_3}, δ_{E_4}, if t, x, y, z and T, X, Y, Z denote arbitrary elements and subsets of E_1, E_2, E_3, E_4, and δ_{E_1}, δ_{E_2}, δ_{E_3}, δ_{E_4} respectively, and if M, P and Q are (E_1, E_2)-, (E_2, E_3)-, and (E_3, E_4)-matrices respectively, then

$$\int_X \left\{ \int_Y Q^z_{dy} P^y_{dx} \right\} M^x_T = \int_Y Q^z_{dy} \int_X P^y_{dx} M^x_T \tag{8.35}$$

If, in particular, M, P and Q are bounded, or if

$$\int_E |P^y|_{dx} |M^x|_T = K^y_T , \quad \int_{E'} |Q^x|_{dy} |P^y|_X = L^z_X ,$$
$$\int_E L^x_{dx} |M^x|_T < \infty \quad \text{or} \quad \int_{E'} |Q^z|_{dy} K^y_T < \infty , \tag{8.36}$$

then the matrix-product, defined by

$$(QP)^z_X = \int_{E_3} Q^z_{dy} \, P^y_X \tag{8.37}$$

$$(PM)^y_T = \int_{E_2} P^y_{dx} \, M^x_T \tag{8.38}$$

is associative:

$$(QP)M = Q(PM) \tag{8.39}$$

i.e.

$$\int_{E_2} (QP)^z_{dx} \, M^x_T = \int_{E_3} Q^z_{dy} \, (PM)^y_T \; . \tag{8.40}$$

Proof: Follows immediately from lemmas 3 and 4 with, for constant z and T , $F_Y = Q^z_Y$ and $f^x = M^x_T$.

From this lemma 5 it follows that the ordinary matrix-calculus, with Lebesgue-Stieltjes-Radon integration instead of summation can be applied as soon as all matrices and functions of either kind concerned are bounded, or, more generally, if they satisfy 8.36.

For simplicity we shall further omit (unless special notice is given) the first case, i.e. restrict ourselves to bounded matrices and function of either kind, unless the contrary is stated explicitly, and assume moreover that the sets E , E' , E_1 , E_2 , etc. all coincide. In this case the increasing sequences E_k also can be omitted, as we may take $E_k = E$ for all k .

If P , hence, is a "square" bounded matrix, i.e. $E = E'$, and

$$\|P\| = \|P\|_E^E = \sup_{x\in E} \sup_{\{X_n\}\in D(E)} \Sigma_n |P^x_{X_n}| < \infty , \tag{8.41}$$

then powers of P and polynomials in P can be formed, and as long as the coefficients are constants ("scalars") these are commutable. From 8.13, 8.14, 8.37, 8.41 follows easily that

$$\|PQ\| \leq \|P\| \|Q\| . \tag{8.42}$$

We note for later use the following identities:

$$F1 = F , \text{ i.e. } \int F_{dx} 1^x_X = F_X ; \tag{8.43}$$

$$1f = f , \text{ i.e. } \int 1^x_{dy} f^y = f^x ; \tag{8.44}$$

$$1P = P1 = P , \text{ i.e. } \int 1^z_{dy} P^y_X = \int P^x_{dy} 1^y_X = P^x_X . \tag{8.45}$$

Further we remark that any two diagonal matrices are commutable:

$$\int (f1)^x_{dy} (g1)^y_X = \int (g1)^x_{dy} (f1)^y_X = f^x g^x 1^x_X . \tag{8.46}$$

We can, however, not replace the symbol $f1$ for the diagonal matrix $f^x 1^x_X$ by $1f$, which denotes the function of the first kind 8.44. If $f1$ is followed by another factor (matrix or function of the first kind) we may omit the 1, e.g.

$f1P = fP$ with $(fP)^x_X = f^x P^x_X$, and also $(fg)^x = (f1g)^x = f^x g^x$.

The integrals over subsets of E can be expressed by means of the partial unit matrices 1_A (cf. 8.7):

$$\int_X F_{dx} f^x = \int_E F_{dy} \int_E 1^y_{dx \cap X} f^x = F1_X f . \tag{8.47}$$

The partial unit matrices also serve the purpose of truncating functions:

$$(1_A f)^x = \begin{cases} f^x & \text{if } x \in A \\ 0 & \text{if not} \end{cases} \tag{8.48}$$

$$(F1_A)_X = F_{A \cap X} = \begin{cases} F_X & \text{if } X \subset A \\ 0 & \text{if } X \subset E\text{--}A \end{cases} \tag{8.49}$$

and analogously $1_A P 1_B$.

A sufficient condition for the convergence of $\sum_1^\infty (P_n)_X^x$ (where the P_n are bounded matrices) for all x and X is the convergence of $\sum_1^\infty \|P_n\|$.

It follows in particular that if $\|P\| < 1$, whence by 8.32 $\|P^n\| \leq \|P\|^n < 1$, P^n denoting the n-th matrix-power, $\sum_0^\infty P^n$ converges and is the (unique) left and right-handed inverse of $1 - P$:

$$(1 - P)^{-1} = \sum_0^\infty P^n , \qquad \text{if } \|P\| < 1 . \tag{8.50}$$

More generally, if a matrix M can be brought in the form

$$M = R + Q \tag{8.51}$$

where R^{-1} exists (we use this expression only if $RX = XR = 1$ has one and only one solution $X = R^{-1}$), whereas either $\|QR^{-1}\| < 1$ or $\|R^{-1}Q\| < 1$, then M^{-1} exists and

$$M^{-1} = R^{-1} \sum_0^\infty (-QR^{-1})^n = \sum_0^\infty (-R^{-1}Q)^n R^{-1} . \tag{8.52}$$

The proof that both series in 8.52 are identical, converge if one of the inequality conditions is satisfied, satisfy the equations for the inverse, and are their only solution , are trivial.

A special and well-known case [6] occurs if $R = f1$ is a diagonal matrix, 8.51 and 8.52 become

$$M = f1 + Q \,, \tag{8.53}$$

$$M^{-1} = \sum_{0}^{\infty} (-f^{-1}Q)^n f^{-1} 1 , \tag{8.54}$$

provided $f^x \neq 0$ for all x and e.g. $\| f^{-1}Q \| < 1$, i.e.

$$|Q^x| = \sup_{\{X_n\} \in D(E)} \Sigma |Q^x_{X_n}| < |f^x| \tag{8.55}$$

for all x .

§ 9. Markoff processes invariant under a transitive group.

As a generalization of the invariant processes studied in § 2, we consider a Markoff process on a set E such that a group $\mathcal{G}$ of transformations τ of E in itself exists, which is transitive over E . An important special case of this is the one in which E is a sphere in any number of dimensions, $\mathcal{G}$ being the group of all rotations of E in itself. We consider then a matrix P which is invariant under all transformations of $\mathcal{G}$:

$$P^x_X = P^{\tau x}_{\tau X} \tag{9.1}$$

for all $\tau \in \mathcal{G}$, where

----

[6] Cf. R.G.Cooke (1950, pag. 31, (2.4, II)) and for the case of finite matrices Olga Taussky (1949), and the older literature mentioned there.

$$\tau X \overset{df}{=} Ens\{\tau x / x \in X\} . \tag{9.2}$$

We choose arbitrarily a point $a \in E$ and define H as the subgroup consisting of all $\tau \in \mathcal{G}$ which leave a invariant

$$H \overset{df}{=} Ens\{\tau \in \mathcal{G} / \tau a = a\} . \tag{9.3}$$

Putting for any $\tau \in \mathcal{G}$

$$F_\tau \overset{df}{=} H\tau H = Ens\{\eta_1 \tau \eta_2 / \eta_1 \in H, \eta_2 \in H\} . \tag{9.4}$$

The sets F_τ are mutually exclusive and have $\mathcal{G}$ as their union. Evidently $\tau \in F_\tau$; $\tau \in H \rightleftarrows F_\tau = H$; $F_\tau \cap F_{\tau'} \neq 0 \rightleftarrows \tau' \in F_\tau \rightleftarrows \tau \in F_{\tau'} \rightleftarrows F_{\tau'} = F_\tau$. Choosing arbitrarily in every F_τ a unique representative γ_τ , let Δ be the set of all these representatives.

Now to each $x \in E$ exists a $\tau \in \mathcal{G}$ with $\tau a = x$, hence a $\gamma \in \Delta$ and $\eta, \eta' \in H$ with $\tau = \eta\gamma\eta'$, hence with $x = \eta\gamma\eta' a = \eta\gamma a$ as $\eta' a = a$ because of $\eta' \in H$. If also $x = \eta_1 \gamma_1 a$ then $\eta'' = \gamma^{-1}\eta^{-1}\eta_1\gamma_1 \in H$, hence $\gamma_1 = \eta_1^{-1}\eta\gamma\eta'' \in H\gamma H$, hence $\gamma_1 = \gamma$ as the representatives are unique so that γ is uniquely determined by $x = \eta\gamma a$ (in general η is not). Then 9.1 entails, putting $p_Y \overset{df}{=} P_Y^a$:

$$P_X^x = P_X^{\eta\gamma a} = p_Y \quad with \quad Y = \gamma^{-1}\eta^{-1}X . \tag{9.5}$$

Here p_Y also is countably additive, and ≥ 0 with $p_E = 1$, and, also by 9.1

$$p_Y = p_{\eta Y} \qquad \text{for all} \qquad \eta \in H . \tag{9.6}$$

For any $\Gamma \subset \Delta$ we put

$$q_\Gamma \overset{df}{=} p_{H\Gamma a} , \tag{9.7}$$

so that q_Γ is countably additive and ≥ 0 with $q_\Delta = 1$.

Then for a given Y and varying $\Gamma \subset \Delta$ $p_{Y \cap H\Gamma a}$ is completely additive in Γ, ≥ 0 , and $\leq q_\Gamma$, and can therefore be written in the form

$$p_{Y \cap H\Gamma a} = \int_\Gamma q_{d\gamma}\, \ell_Y^\gamma \; . \tag{9.8}$$

Here ℓ_Y^γ can be chosen to be completely additive in Y and to vanish unless $\gamma a \in HY$. In fact ℓ_Y^γ is the conditional probability that $\underline{x} \in Y$ under the condition that $\underline{x} \in H\gamma a$.

For every fixed Y the functions ℓ_Y^γ are defined as measurable functions on Δ except for a set of q -measure o .

We introduce now:

Assumption A: The definition of the quantities ℓ_Y^γ can be extended in such a way that they are defined for all Y and all γ as univalent functions of their arguments, for fixed Y measurable in γ and for fixed γ countably additive in Y .

According to Doob (1948) a sufficient condition for assumption A to be fulfilled is: E is a Borel-set in a Euclidean space and the σ -field over E consists of all Borel-subsets of E .

The equality 9.6 entails

$$\ell_{\eta Y}^\gamma = \ell_Y^\gamma \qquad \text{for all} \quad \eta \in H. \tag{9.9}$$

Taking in 9.9 especially $Y = K\Gamma a$ with $K \subset H$, $\Gamma \subset \Delta$, $\gamma \in \Gamma$, $\ell_{K\Gamma a}^\gamma$ for variable K is countably additive, ≥ 0 , bounded, and invariant under all $\eta \in H$, hence it has the properties of a Haar-measure.

We now introduce:

Assumption B: The Haar-measure on H is uniquely determined except for a proportionality factor.

According to A.Haar (1933) (cf. also J.von Neumann (1934, 1936), L.H.Loomis (1945)) a sufficient condition for assumption B to be fulfilled is: H is a locally compact topological group.

Assumptions A and B together are certainly satisfied e.g. if E is a Euclidean sphere of any number of dimensions and $\mathcal{G}$ the group of all rotations of E into itself. Hence, if μ_K is the Haar-measure, normalized to $\mu_H = 1$, and m_Γ^γ the proportionality-factor,

$$\ell_{K\Gamma a}^{\gamma} = m_\Gamma^\gamma \mu_K = m^\gamma 1_\Gamma^\gamma \mu_K \tag{9.10}$$

as $m_\Gamma^\gamma = 0$ unless $\gamma \in \Gamma$. Moreover $\ell_{H\Gamma a}^{\gamma} = 1_\Gamma^\gamma$, whence $m^\gamma = 1$.

Extension of $\ell_{K\Gamma a}^{\gamma}$ to ℓ_Y^γ with arbitrary Y gives

$$\ell_Y^\gamma = \int \mu_{d\eta}\, 1_Y^{\eta\gamma a} \tag{9.11}$$

and substitution into 9.8 with $\Gamma = \Delta$ gives

$$p_Y = \int_\Delta q_{d\gamma} \int_H \mu_{d\eta}\, 1_Y^{\eta\gamma a} \tag{9.12}$$

and finally by 9.5

$$p_X^x = \int_\Delta q_{d\gamma} \int_H \mu_{d\eta}\, 1_X^{\tau_x \eta\gamma a}, \tag{9.13}$$

where τ_x is <u>any</u> $\tau \in \mathcal{G}$ with $\tau_x a = x$. On the other hand it is clear that 9.13 satisfies our conditions.

If in particular $\mathcal{G}$ is <u>simply</u> transitive over E, i.e., if H consists of the unit element only, $\Delta = \mathcal{G}$ and 9.7 and 9.13 simplify to $q_\Gamma = p_{\Gamma a}$ and

$$p_X^x = \int_{\mathcal{G}} q_{d\gamma}\, 1_X^{\tau_x \gamma a} \tag{9.14}$$

If, more in particular, E and $\mathcal{G}$ are identical, we can take $a = 1$, $\tau_x = x$, and we get $q_\Gamma = p_\Gamma$ and

$$p_X^x = \int_E p_{dy}\, 1_X^{xy} \tag{9.15}$$

If $\mathcal{G}$ is commutative and additively written, this leads back to 2.8.

§ 10. Wald's fundamental identity for arbitrary stochastic processes.

Exactly the same argument, which leads to Wald's fundamental identity (5.17), holds if E is an r-dimensional Euclidean space, x and y are real vectors and ξ is a complex vector in r dimensions, and ξx is the scalar product. It also holds in the more general case, where E is an additive Abelian group and $\xi = \xi_1 + i\xi_2$, where ξ_1 and ξ_2 are homomorphisms on the set of real numbers, i.e. for all x and y in E, $\xi_i x$ is a real number such that $\xi_i(x+y) = \xi_i x + \xi_i y$ ($i = 1,2$).

We can resume our result in:

Theorem 4: If E is an (additive) Abelian group, if ξ_0, ξ_1 and ξ_2 are homomorphisms in the additive group of real numbers; if $\xi = \xi_1 + \xi_2$ and $\varphi(\xi_0) < \infty$ and $\inf_{x \in B} (\xi_0 - \xi_1)x > -\infty$ then the identity

$$\sum_1^\infty \varphi(\xi)^{-n} \int C_{(n)} \overset{x}{dy}\, e^{\xi y} = 1 \tag{5.17}$$

is satisfied.

A partial generalization of this theorem to arbitrary stochastic processes, determined by the transition probabilities $P_{(n)}{}^{x_0,\ldots,x_{n-1}}_{X}$ satisfying 2.1-2.4 is derived as follows.

We make the additional assumption, that when the wandering point arrives in the state x_{n-1} $(n \geq 1)$ after having passed if $n \geq 2$ through $x_0,\ldots,x_{n-2}$, there is a probability $A_{(n)}^{x_0,\ldots,x_{n-1}}$ that the point will be absorbed hence a probability

$$B_{(n)}^{x_0,\dots,x_{n-1}} = 1 - A_{(n)}^{x_0,\dots,x_{n-1}}$$

that it will not be absorbed.

Moreover we introduce the auxiliary quantities $\mathcal{U}_{(n)}^{x_0,\dots,x_{n-1}}$ and $T_{(n)}^{x_0,\dots,x_{n-1}}$ interpreted as the conditional probabilities that the catastrophe $\mathcal{E}$ does not happen under the condition that the wandering pointhas passed through $x_0,\dots,x_{n-1}$ and has (for $\mathcal{U}$) or has not (for T) been absorbed in x_{n-1} . Finally we define $C_{(n)}^{x_0,\dots,x_{n-1}}$ as the total conditional probability that $\mathcal{E}$ will not occur under' condition that the point has passed through $x_0,\dots,x_{n-1}$.

In order to compute $C_{(n)}^{x_0,\dots,x_{n-1}}$ for $n \geq 1$, we remark that there are two complementary cases: either the point is absorbed in x_{n-1} (probability $A_{(n)}^{x_0,\dots,x_{n-1}}$), in which case non-$\mathcal{E}$ has probability $\mathcal{U}_{(n)}^{x_0,\dots,x_{n-1}}$, or it is not absorbed (probability $B_{(n)}^{x_0,\dots,x_{n-1}}$) and $\mathcal{E}$ does not occur (probability $T_{(n)}^{x_0,\dots,x_{n-1}}$) and it jumps into some "small" set dy (probability $P_{(n)}^{x_0,\dots,x_{n-1}}dy$), and it lives happily ever after (i.e. $\mathcal{E}$ does not occur further on; probability $C_{(n+1)}^{x_0,\dots,x_{n-1},y}$). Hence we have

$$C_{(n)}^{x_0,\dots,x_{n-1}} = A_{(n)}^{x_0,\dots,x_{n-1}}\mathcal{U}_{(n)}^{x_0,\dots,x_{n-1}} + B_{(n)}^{x_0,\dots,x_{n-1}}T_{(n)}^{x_0,\dots,x_{n-1}}\int P_{(n)}^{x_0,\dots,x_{n-1}}{}_{dy}\, C_{(n+1)}^{x_0,\dots,x_{n-1},y} \quad (10.1)$$

For $n=0$ we have a similar equation, with the only difference that the upper suffixes except y fail.

In order to simplify the formulae we shall omit the lower suffixes (n) and replace the sequence of upper suffixes $x_0,\dots,x_{n-1}$ by a single symbol π (= path). Then 10.1 becomes:

$$C^{\pi} = A^{\pi}\mathcal{U}^{\pi} + B^{\pi}T^{\pi}\int P^{\pi}{}_{dy}\, C^{\pi y} \quad (10.2)$$

By successive substitution we obtain a formal development f[illegible] C^{π} , π being given (in particular also if π is "empty"):

$$C^{\pi} = A^{\pi}U^{\pi} + B^{\pi}T^{\pi}\int P^{\pi}_{dy_1} A^{\pi y_1}U^{\pi y_1} + B^{\pi}T^{\pi}\int P^{\pi}_{dy_1} B^{\pi y_1}T^{\pi y_1}\int P^{\pi y_1}_{dy_2} A^{\pi y_1 y_2}U^{\pi y_1 y_2} + \ldots =$$

$$= \sum_0^{\infty} k\, B^{\pi}T^{\pi}\int P^{\pi}_{dy_1} B^{\pi y_1}T^{\pi y_1}\int \ldots \int P^{\pi y_1 \cdots y_{k-1}}_{dy_k} A^{\pi y_1 \cdots y_{k-1}} U^{\pi y_1 \cdots y_{k-1}} . \tag{10.3}$$

In order that this development be convergent and a solution of 10.2 it is necessary and sufficient, that

$$R_k^{\pi} = B^{\pi}T^{\pi}\int P^{\pi}_{dy_1} B^{\pi y_1}T^{\pi y_1}\int \ldots \int P^{\pi y_1 \cdots y_{k-1}}_{dy_k} C^{\pi y_1 \cdots y_k} \tag{10.4}$$

tends to zero for $k \to \infty$.

In order to generalize theorem 2 we consider solutions of the system of equations 10.1 in the unknowns $C_{(n)}^{x_0, \ldots, x_{n-1}}$ where the U and the T need not be probabilities, but may be arbitrary real or complex sequences of (not necessarily bounded) functions.

Theorem 5: If A. for all n and all $n = x_0, \ldots, x_{n-1} \in E^n$ real numbers U_0^{π}, T_0^{π} exist, such that

A.1 $T_0^{\pi} \geqq 0$ for all n and all π ,

A.2 $U_0^{\pi} \geqq 0$ for all n and all π ,

A.3 $T_0^{\pi}\int P^{\pi}_{dy} U_0^{\pi y} \leqq U_0^{\pi}$ for all n and all π with $B^{\pi} \neq 0$;

If B. the functions T^{π}, U^{π} are defined for all n and all $\pi \in E^n$ if θ and c are real numbers, and if

B.1 $0 \leqq \theta < 1$,

B.2 $|T^{\pi}| \leqq \theta T_0^{\pi}$ for all n and all π ,

B.3 $|U^{\pi}| \leqq c U_0^{\pi}$ for all n and all π ,

then the equations 10.2 have solutions C, given by 10.3, which satisfy the equations

$$C^{\pi} = U^{\pi} \tag{10.5}$$

for all n and all π, if and only if

$$\mathcal{U}^{\pi} = T^{\pi} \int P^{\pi}_{dy}\, \mathcal{U}^{\pi y} \tag{10.6}$$

for all n and all π with $B^{\pi} \neq 0$.

Proof: Writing

$$Q^{\pi}_{X} = B^{\pi} T^{\pi} P^{\pi}_{X}, \quad Q_{o}{}^{\pi}_{X} = B^{\pi} T_{o}{}^{\pi} P^{\pi}_{X}, \tag{10.7}$$

A.3 becomes

$$\int Q_{o}{}^{\pi}_{dy}\, \mathcal{U}_{o}{}^{\pi y} \leqq B^{\pi} \mathcal{U}_{o}{}^{\pi} \tag{.10.8}$$

from which the existence of the left member follows. The terms of the sum 10.3 with T_o, $\mathcal{U}_o$ instead of T, $\mathcal{U}$ are $\geqq 0$. The $(k+1)$st term of this sum is $\leqq$ the same expression with the factor $A^{\pi y_1 \cdots y_k}$ (which is $\leqq 1$) omitted, hence by 10.8 with $\pi y_1 \ldots y_{k-1}$ instead of π and y_k instead of y $\leqq$ the k-th term with $A^{\pi y_1 \cdots y_{k-1}}$ replaced by $B^{\pi y_1 \cdots y_{k-1}}$. Added to the k-th term itself the sum is $\leqq$ the k-th term with the factor $A^{\pi y_1 \cdots y_{k-1}}$ omitted. Continuing in this way we find that the sum of the first $k+1$ terms is $\leqq \mathcal{U}_o^{\pi}$, so that the series is bounded, hence convergent. Denoting its sum by C_o^{π} we have then $0 \leqq C_o^{\pi} \leqq \mathcal{U}_o^{\pi}$ and also $0 \leqq R^{\pi}_{o,k} \leqq \mathcal{U}_o^{\pi}$, where $R^{\pi}_{o,k}$ is the expression 10.4 with T and C replaced by T_o and C_o. From B.1-B.3 it follows then that the integrals in 10.3 as well as the whole series, exist absolutely and that $|R^{\pi}_{k}| \leqq \theta^{k} R^{\pi}_{o,k} \leqq \theta^{k} \mathcal{U}_o^{\pi}$, whence $\lim_{k \to \infty} R^{\pi}_{k} = 0$, so that 10.3 satisfies 10.2. Now, if a solution of 10.2 satisfies 10.5, it is trivial (by 5.17) that also 10.6 holds. If, on the other hand, 10.6 holds together with 10.3, we have

$$\mathcal{U}^{\pi} - C^{\pi} = \lim_{n\to\infty}\left\{B^{\pi}\mathcal{U}^{\pi} - \sum_{1}^{n} k \int Q^{\pi}_{dy_1}\int\cdots\int Q^{\pi y_1\cdots y_{k-1}}_{dy_k} A^{\pi y_1\cdots y_k}\mathcal{U}^{\pi y_1\cdots y_k}\right\} =$$

$$= \lim_{n\to\infty}\left[\sum_{1}^{n-1}\int Q^{\pi}_{dy_1}\int\cdots\int Q^{\pi y_1\cdots y_{k-1}}_{dy_k}\left\{B^{\pi y_1\cdots y_k}\mathcal{U}^{\pi y_1\cdots y_k} - \int Q^{\pi y_1\cdots y_k}_{dy_{k+1}}\mathcal{U}^{\pi y_1\cdots y_{k+1}}\right\} + \right.$$

$$\left. + \int Q^{\pi}_{dy_1}\int\cdots\int Q^{\pi y_1\cdots y_{n-1}}_{dy_n} B^{\pi y_1\cdots y_n}\mathcal{U}^{\pi y_1\cdots y_n}\right]. \tag{10.9}$$

The expressions between the curved brackets vanish because of 10.6, 10.7 (with $\pi y_1,\ldots,y_{k-1}$ instead of π and y_k instead of y), hence they remain zero after repeated integrations, and the first **sum** between the square brackets in the right member of 10.9 vanishes. The last term between the square brackets is absolutely $\leq \theta^k \mathcal{U}^{\pi}_0$; hence it tends to zero, which proves 10.5 and the theorem.

In the proof of theorem 5 lemma 5 (Appendix § 8) has not been used. It is therefore to be expected that theorem 5 will not completely imply theorem 4. In fact, in the special case of theorem 2 we have

$$\left.\begin{aligned} P^{x_0,\ldots,x_{n-1}}_{(n)\,X} = P^{x_{n-1}}_{X},\ A^{x_0,\ldots,x_{n-1}}_{(n)} = A^{x_{n-1}},\ B^{x_0,\ldots,x_{n-1}}_{(n)} = B^{x_{n-1}}, \\ \mathcal{U}^{x_0,\ldots,x_{n-1}}_{(n)} = f^{x_{n-1}},\ T^{x_0,\ldots,x_{n-1}}_{(n)} = T^{x_{n-1}}. \end{aligned}\right\} \tag{10.10}$$

Then the conditions of theorem 2 except 5.8 are satisfied, but 10.6 passes into

$$C^{x_0,\ldots,x_{n-1}}_{(n)} = f^{x_{n-1}} \tag{10.11}$$

so that it remains to be proved that

$$C^{x_0,\ldots,x_{n-1}}_{(n)} = \int C^{x_{n-1}}_{dy} f^{y} \tag{10.12}$$

where C^{x}_{X} is given by 5.10, if 5.8 is satisfied. We shall not go into the question, under which precise conditions this latter relation is satisfied. In any case a generalization of lemma 2 with $f^{x,y}$ instead of f^{y} is needed.

References.

N.Arley, 1943, On the theory of stochastic processes and their application to the theory of cosmic radiation. Copenhagen, Thesis University of Copenhagen.

G.A.Barnard, 1946, Sequential tests in industrial statistics. Journal Royal Statistical Soc. Supplement 8, 1-21.

M.S.Bartlett, 1944, Negative probability, Proc. Cambr. Phil. Soc. 41, 71-73.

G.Birkhoff, 1948, Lattice theory (2nd edition), Am. Math. Soc. Colloquium Publications, Vol. 25.

D.Blackwell and M.A.Girshick, 1946, On functions of sequences of independent chance vectors with applications to the problem of 'random walk' in k dimensions, Annals Math. Statistics 17, 310-317.

G.Blom, 1949, A generalization of Wald's fundamental identity, Annals Math. Statistics 20, 439-444.

R.G.Cooke, 1950, Infinite matrices and sequence spaces, London, Macmillan and Co.

D.van Dantzig, 1935, La nôtion de derivée d'une fonctionnelle, C.R.Ac.Sc. Paris, 201, 1008-1010.

idem, 1936, Ricci-calculus and functional analysis, Proc. Kon. Ned. Ak. v. Wet., 39, 785-794.

idem, 1941, Mathematische en empiristische grondslagen der waarschijnlijkheidsrekening, Ned. Tijdschrift voor Natuurkunde 8, 70-93.

idem, 1947, Mathematische Statistiek, Hoofdstuk 2, Collectieve kenmerken, voortbrengende en karakteristieke functies, Mathematisch Centrum, Amsterdam, stencilled.

D.van Dantzig, 1949, Sur la méthode des fonctions génératrices, Colloques internationaux du centre national de la recherche scientifique, 13, 29-45.

W.Do(e)blin, 1940, Éléments d'une théorie générale des chaînes simples constantes de Markoff, Annales sc. Éc. Norm. Sup. (3), 57, 61-111.

J.L.Doob, 1938, Stochastic processes with an integral-valued parameter, Transactions Am. Math. Soc. 44, 87-151.

idem, 1945, Markoff chains - denumerable case, Transactions Am. Math. Soc. 58, 455-473.

idem, 1948, Asymptotic properties of Markoff transition probabilities, Transactions Am. Math. Soc. 63, 393-421.

W.Feller, 1949, On the theory of stochastic process, with particular reference to applications, Proceedings Berkeley Symposion on Mathematical Statistics and Probability, 403-432.

idem, 1950, An introduction to probability theory and its applications, Vol. I, New York, John Wiley and Sons.

M.Fréchet, 1913, Sur la notion de différentielle dans le calcul des fonctionnelles, C.R. Congrès des soc. savantes (1912), 45-58.

idem, 1914, Sur la notion de différentielle d'une fonction de ligne, Transactions Am. Math. Soc. 15, 135-161.

idem, 1950, Généralités sur les probabilités. Éléments aléatoires - Traité du calcul des probabilités et de ses applications. Tome I (Les principes de la théorie des probabilités), Fascicule III (Recherches théoriques modernes sur le calcul des probabilités), Livre I. Paris, Gauthier Villars.

M.A.Girshick, 1946, Contributions to the theory of sequential analysis, part II, Annals Math. Statistics 17, 282-291.

A.Haar, 1933, Der Massbegriff in der Theorie der kontinuierlichen Gruppen, Annals of Math. (2), 34, 147-169.

J.Hadamard, 1910, Leçons sur le calcul des variations, Tome I, Paris, A.Hermann.

H.Hahn and A.Rosenthal, 1948, Set functions, Albuquerque, Univ. of New Mexico Press.

P.R.Halmos, 1950, Measure theory, New York, D.van Nostrand Company.

J.J.Hermans, M.S.Klamkin, R.Ullman, 1952, The excluded volume of polymer chains. To appear in: Journal of Chem. Physics.

J.H.B.Kemperman, 1950, The general one-dimensional random walk with absorbing barriers, with applications to sequential analysis, (Ph.D-Thesis, Amsterdam), 's-Gravenhage, Excelsiors Foto-Offset.

G.W.King, 1948, Dependence of physical properties of high polymers on molecular structure; statistical investigations with the aid of punched carth methods, Office of Naval Research, Technical Report no. 1.

idem, 1949, New methods in the statistical mechanics of high polymers, Office of Naval Research, Technical Report no. 2.

L.H.Loomis, 1945, Abstract congruence and the uniqueness of Haar measure, Annals of Math. (2) 46, 348-355.

E.W.Montroll, 1950, Markoff chains and excluded volume effect in polymer chains, Journal of Chem. Physics 18, 734-743.

J.von Neumann, 1928, Zur Theorie der Gesellschaftsspiele, Math. Annalen, 100, 295-320.

idem, 1934, Zum Haarschen Mass und topologischen Gruppen, Compositio Math., 1, 106-114.

idem, 1936, The uniqueness of Haar measure, Math. Sbornik 1, 721-734.

J.von Neumann and O.Morgenstein, 1944, Theory of games and economic behavior, Princeton, N.J., Princeton University Press, 2nd ed., 1947.

Olga Taussky, 1949, A recurring theorem on determinants, Am. Math. Monthly, 56, 672-676.

Ch.M.Tchen, 1951, Stochastic processes and dispersion of configurations of linked events, Journal of Research, National Bureau of Standards, 46, 480-488.

A,Wald, 1950, Statistical decision functions, New York, John Wiley and Sons.

W.Wasow, 1951, On the duration of random walks, Annals of Math. Statistics 22, 199-216.

National Bureau of Standards, Washington D.C.
Mathematical Centre, Amsterdam
1952.