ON THE THEORY OF MARKOFF CHAINS

by

Gustav Elfying

Institute cf Statistics
Mimeo. Series No. 103
May, 1954

Finland 1937

ZUR THEORIE DER MARKOFFSCHEN KETTEN

Von

Gustav Elfving

ACTA SOCIETATIS SCIENTIARUM FENNICAE

Volume 2, No. 8, 1937, Finland

Translated by Edwin L. Cox under the supervision of Fred J. Allred, Assistant Professor of Modern Languages at North Carolina State College of the University of North Carolina, Raleigh, N. C. at the request of the Department of Experimental Statistics, North Carolina State College. (Edited jointly by the Translation Service and the Department of Experimental Statistics.)

May 17, 1954

ON THE THEORY OF MARKOFF CHAINS

1. By a Markoff chain is understood, as you know, a stochastic process with discrete time and a finite[1] number of possible conditions, that is moreover homogeneous with regard to time. The theory of this process was established by Markoff and Poincare. Later, in particulars, it was carried further by Hostinsky, Hadamard and Von Mises[2]. Romanovsky has given in Acta Mathematica, 1933, a systematic and in some measure exhaustive exposition of this theory (7). He depends on the essentials of matrix calculation. The application of this technique is postulated by the fact that the probabilities appearing here satisfy certain systems of linear homogeneous difference equations with constant coefficients.

The Markoff chains - in the narrow sense - stand nearest in the system of probability calculations to those with regard to constant time t and homogeneous stochastic processes with a finite number of possible conditions. Kolmogoroff has shown that the transition probabilities appearing here satisfy certain systems of linear homogeneous differential equations with constant coefficients. If we limit ourselves by a process of this kind to a sequence of equi-distant time points, then we have a Markoff chain; this behaves to the corresponding stationary process almost like the whole number powers of the exponential function. We denote this process in the following, in short, as a stationary Markoff process.

---

1 Compare also; Kolmogoroff (5).

2 For the history and bibliography of the Markoff chain see B. Hostinsky; Les méthodes analytiques du calcul des probabilitiés, Gauthier-Villars, Paris 1931; J. Hadamard and M. Fréchet. Sur les probabilitiés discontinues des enenements en chaîne, Z. angew. Math. Mech. XIII (1933).

The greater part of the work on the Markoff chains is concerned with the end state of the system considered, that is, with the asymptotic representation of the process. The following work will be concerned with two somewhat different questions.

First; The parallelism between the Markoff chains and the stationary Markoff processes poses the following question. If a Markoff chain is given, is there a stationary Markoff process - then whose transition probability coincides nearly for whole number t with each of the chains? Is this process clearly determined? This problem is at first sight, a rather elementary interpolation problem, but the demand that the result should be a stochastic process, brings up a certain characteristic to which we wish to draw attention. Actually it is shown that this problem can have none or one unique or infinitely many solutions.

Secondly; Let there be proposed a law of a Markoff chain or a stationary Markoff process, i.e., that the transition probabilities be known. Now, if one fixes an absolute probability distribution for a certain time point t_o, then the distribution is determined by it in each following time point. But how does this fit with the preceding point of time? Can we continue the process backwards unconditionally or must a certain initial moment be attributed to it? It becomes evident that, in general, the latter is the case.[1]

The content of the problem appears at best with a statistical interpretation. Let us consider a very large set of systems that jointly independent of one another are subject to the considered process. The question then reads: If at the time moment t_o a certain distribution of the system of the possible states is observed, how long at most can the process have continued? Our problem leads us therefore

[1] Moreover compare also Kolmogoroff (4)

to the concept of the age of the distribution with a given chain law. The age is, for example, the analog of the magnitude t in the expression of the normal distribution

$$\frac{1}{\sqrt{t}}\ \phi\left(\frac{x-m}{\sqrt{t}}\right) = \frac{1}{2\pi t}\ e^{-\frac{(x-m)^2}{2t}}$$

if this distribution is explained as the result of a diffusion process homogeneous in t and x with the diffusion speed 1.

In our paper we shall only handle the especially easily handled analytic case, in which the matrix of the process possesses only different characteristic or real values. The essentials of the results obtained below should be independent of this assumption and should be easily extended properly through the stationary considerations, to the case of equal real value.

2. Markoff Chains. We consider a system S which is capable of states E_1, E_2, ..., E_n. We denote by $p_{ik}^{(v)}$ the probability for S, reached in v steps from E_i to E_k; in particular we set $p_{ik}^{(1)} = p_{ik}$. The transition probabilities obviously satisfy the conditions.

(1) $p_{ik} \geqslant 0, \quad \sum_k p_{ik} = 1,$

On the other hand, we denote by $q_1^{(t)}$, ..., $q_n^{(t)}$ the absolute probabilities of the states E_1, ..., E_n at time t; then it is equally true

(2) $q_i^{(t)} \geqslant 0, \quad \sum_i q_i^{(t)} = 1.$

According to the multiplication and addition laws, there exist between the probabilities introduced before, the relations

(3) $p_{ik}^{(v)} = \sum_j p_{ij}^{(v-1)}\ p_{jk}, \quad q_k^{(t+v)} = \sum_i q_i^{(t)}\ p_{ik}^{(v)}.$

If we set, in turning to matrix notation

$$(4) \quad P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}, \quad Q^{(t)} = (q_1^{(t)} \quad \cdots \quad q_n^{(t)}),$$

then the equations (3) assume the form

$$(3') \qquad P^{(v)} \equiv (p_{ik}^{(v)}) = P^v, \quad Q^{(t+v)} = Q^{(t)}P^v$$

A matrix P with the properties (1) is called <u>stochastic</u> according to Romanovsky.

In the following we will think of matrix P as given once for all. The main task of the theory becomes then naturally the explicit statement of the number $p_{ik}^{(v)}$ as a function of v. This problem has been solved generally in the Romanovsky article (p. 181 and the following) on the basis of a formula of Perron. We shall handle it here only under the assumptions made at the end of the introduction.

Let

$$(5) \quad |\lambda E - p| = \begin{vmatrix} \lambda - p_{11} & -p_{12} & \cdots & -p_{1n} \\ -p_{21} & \lambda - p_{22} & \cdots & -p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -p_{n1} & -p_{n2} & \cdots & \lambda - p_{nn} \end{vmatrix}$$

be the characteristic equation of P; its roots which are, as we know, assumed to be simple are $\lambda_1, \ldots \lambda_n$. We set

$$(5) \qquad \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Frobenius has shown, as is known,[1] that among the roots of a matrix with non-negative elements, there is always a root $\geqslant 0$, that is absolutely greater or equal to each other root. In case of the stochastic matrix P, this root is equal to one (compare (7), page 149). It is consequently true that $\lambda_1 = 1$, $\lambda_i \leqslant 1$ (i=2,..n).

If there exists a certain matrix P, then we can easily specify a narrowed range for the roots. Let the principal elements be identically $p_{11}, \ldots, p_{nn}$, jointly positive, and let p be the smallest of these elements, thus the matrix $P - pE$ is non-negative and its roots $\lambda_i - p$ $(i = 1, \ldots, n)$ are thus absolutely smaller or equal to the greatest positive among them, which obviously is equal to $1 - p$. It is also true that

(6) $$|\lambda_i - p| \leqslant 1 - p$$

that is, the λ_i belong to a circular slice which borders on the interior of the unit circle with $\lambda = 1$.

According to the common theory of the linear equation system for each root λ_i, there are two each of characteristic vectors determined to within a factor that we in matrix notation indicate by

$$U_{i|} = (u_{i1} \ \ldots \ u_{in}), \quad V_{|i} = \begin{pmatrix} v_{1i} \\ \cdot \\ \cdot \\ \cdot \\ v_{ni} \end{pmatrix}$$

they have the properties

(7) $$U_{i|} V_{|j} = 0 \ (i \neq j), \quad U_{i|} V_{|i} \neq 0 \qquad (i = 1, \ldots, n)$$

and

(8) $$U_{i|} P = \lambda_i U_{i|}, \quad PV_{|i} = \lambda_i V_{|i}. \qquad (i = 1, \ldots, n)$$

---

[1] (Proceedings Akad. Wiss. Berlin, 1908, 1909, 1912.)

The elements of $V{\downarrow}_i$ are jointly equal because of (1); we wish to choose them equal to ones. For the residual $V{\downarrow}_i$ we determine some standard and standardize the $U_i{\downarrow}$ accordingly in such a way that

$$U_i{\downarrow}\, V{\downarrow}_i = 1 \qquad (i = 1, \ldots, n) \tag{9}$$

Here are the rotation matrices

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ u_{n1} & u_{n2} & \cdots & u_{nn} \end{bmatrix}, \quad V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix} \tag{10}$$

which are uniquely determined - with a fixed sequence of the λ_i - and by (7), (8) and (9) the equations

$$UV = E, \quad UPV = \Lambda \tag{11}$$

hold. On the other hand, if there is to a given chain P, a matrix Λ of form (5) as well as two matrices U, V with properties (11) then $\lambda_1, \ldots, \lambda_n$ are characteristic roots of P and U, V are the rotation matrices of the chain. It is actually

$$|\lambda E - P| = |U| \cdot |\lambda E - P| \cdot |V| = |\lambda E - UPV| = |\lambda E - \Lambda| = (\lambda - \lambda_1)\ldots(\lambda - \lambda_n);$$

further from (11) $UP = \Lambda U$ and $PV = V\Lambda$ follows, from which the determining equations (8) are given for the single rows or columns of U, V.

Formulas (11) make possible the solution of our iteration problem. Through the involution of the second formula under the examination of the first, we obtain

$$UP^v V = \Lambda^v, \quad P^v = V\Lambda^v U, \tag{12}$$

and so, since $v_{i1} = 1$, $\lambda_1 = 1$, we have

$$p_{ik}^{(v)} = u_{1k} + u_{2k}\, v_{i2}\, \lambda_2^v + \cdots + u_{nk}\, v_{in}\, \lambda_n^v. \tag{13}$$

This equation contains all the known results about the asymptotic behavior of the chain. If all the characteristic roots with the exception of λ_1 are numerically smaller than one, $p_{ik}^{(v)}$ tends - and further more $q_k^{(v)}$ as well - for $v \to \infty$ to the value $p_k = u_{1k}$, independently of i. If, on the other hand, we get negative or complex roots with absolute value one, then $p_{ik}^{(v)}$ becomes asymptotically periodic. The formulas (13) are true only for numerous simple real values. The presentation of Romanovsky (page 221) is incorrect on this point by our reasoning.

3. <u>The Stationary Markoff Process</u>. We consider again a system S, which can be found in the states E_1, ... E_n, but this time, a stationary observation can be attained. Let $P_{ik}(s,t)$ be the conditional probability that if the state E_k at time t be known, then the system will be found in state E_i at time s. We agree with Kolmogoroff ((3), page 428) about a stationary stochastic process with a finite number of states in the case when the matrix $P^{(s,t)}$ of the transition probabilities for $t \to s$ tends to E.

In the following, we confine ourselves to a homogeneous stationary process, i.e. to one in which the $P_{ik}(s,t)$ depend only on $t \to s$; we write in short $P_{ik}(t)$ in place of $P_{ik}(s,s+t)$. These quantities suffice obviously for the probability of the characteristic conditions

(14) $$P_{ik}(t) \geqslant 0, \quad \sum_k P_{ik}(t) = 1$$

and

(15) $$P_{ik}(t_1 + t_2) = \sum_j P_{ij}(t_1)\, P_{jk}(t_2),$$

as well as the stationary conditions

(16) $$P_{ik}(t) \to 0, \quad (i \neq k), \quad P_{ii}(t) \to 1 \qquad (t \to 0).$$

We shall see later that in the homogeneous case the existence of the limits

(17) $$a_{ik} = P'_{ik}(0) = \lim \frac{P_{ik}(t)}{t} \qquad (t \to 0;\ i \neq k)$$

$$a_{ii} = P'_{ii}(0) = \lim \frac{P_{ii}(t) - 1}{t} \qquad (t \to 0)$$

follow from (15) and (16). For the present we shall assume their existence. From (14) and (17) then follow immediately the properties

(18) $$a_{ik} \geqslant 0 \quad (i \neq k), \quad a_{ii} \leqslant 0, \quad \sum_k a_{ik} = 0.$$

The functions $P_{ik}(t)$ satisfy as Kolmogoroff has shown[1], certain simple differential equations. We have actually from (15)

$$\frac{P_{ik}(t+\Delta t) - P_{ik}(t)}{\Delta t} = P_{ik}(t)\,\frac{P_{kk}(\Delta t) - 1}{\Delta t} + \sum_{k} P_{ij}(t)\,\frac{P_{jk}(\Delta t)}{\Delta t}$$

and from this follows, according to (17) for $\Delta t \to 0$

$$(19) \qquad P'_{ik}(t) = \sum_{j} a_{jk} P_{ij}(t) \qquad (k = 1, \ldots, n).$$

Thus we have the statement: The functions $P_{ik}(t)$ satisfy the differential equations (19i) with the initial conditions (16i).

4. Solution of the Differential Equations. Conversely, let a matrix be stated with the properties (18),

$$(20) \qquad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

We ask for the corresponding solutions of the system (19) with the initial conditions (16i) (i = 1, ..., n).

The existence and uniqueness of the solutions are insured by the general existence theorems. Kolmogoroff (p. 431) shows that they possess exactly the properties (14), (15). We shall carry out the known explicit solution immediately below, but again only under the assumption that the matrix A possesses many simple real values.

We turn now entirely to matrix notations. With the notation

---

[1] (3), page 428 Kolmogoroff treated the general function $P_{ik}(s,t)$ and assumed the existence $\frac{\partial}{\partial t} P_{ik}(s,t)$ for $t > s$, not for $t = s$.

$$(21) \qquad P^{(t)} = \begin{bmatrix} P_{11}(t) & \dots & P_{1n}(t) \\ \dots & \dots & \dots \\ P_{n1}(t) & \dots & P_{nn}(t) \end{bmatrix}$$

the differential equations (19) can be condensed

$$(19') \qquad \frac{dP^{(t)}}{dt} = P^{(t)} A ,$$

while the initial conditions (16) take the form

$$(16') \qquad P^{(0)} = E .$$

The equation (19') will undergo a transformation which is entirely analogous to one used in Section 2.

We write the roots of the matrix A in the form

$$(22) \qquad \chi_1, \ \chi_2, \ \dots , \chi_n; \ K = \begin{bmatrix} \chi_1 & 0 & \dots & 0 \\ 0 & \chi_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \chi_n \end{bmatrix} .$$

Because of $(18)_3$, one of these, for example, the first $\chi_1 = 0$. Out of the characteristic roots we construct exactly as in Section 2 normal real vectors and from them rotation matrices U, V with

$$(23) \qquad UV = E, \ UAV = K .$$

Also here we can take $v_{11} = \dots v_{n1} = 1$. If we pre-multiply (19') by U and post-multiply by V and set besides

$$(24) \qquad R^{(t)} = UP^{(t)}V ,$$

then it follows, since the elements $r_{ik}^{(t)}$ from $R^{(t)}$ are put together linearly from those of $P^{(t)}$,

$$(25) \qquad \frac{dR}{dt} = RK$$

or writing in full

(25') $$\frac{dr_{ik}^{(t)}}{dt} = \chi_k r_{ik}^{(t)}$$

Because of (16'), $R^{(0)} = E$, and (25') gives for solutions

(26) $$r_{kk}^{(t)} = e^{\chi_k t}, \quad r_{ik}^{(t)} = 0, \quad (i \neq k),$$

in short

(26') $$R^{(t)} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & e^{\chi_2 t} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & e^{\chi_n t} \end{bmatrix}.$$

Finally with the help of (24) we turn to the original variables and obtain

(27) $$P^{(t)} = VR^{(t)}U$$

or in detail completely similar to equation (13)

(27') $$P_{ik}^{(t)} = u_{1k} + u_{2k} v_{i2} e^{\chi_2 t} + \dots + u_{nk} v_{in} e^{\chi_n t}.$$

As will be shown in Section 7, if the real parts of $\chi_2, \dots, \chi_n$ are jointly negative, then it is true that

$$P_{ik}^{(t)} \rightarrow u_{1k} \qquad (t \rightarrow +\infty).$$

Under the application of the above aids introduced, let us demonstrate only the formulated proposition. With a homogeneous process there follows the existence of the derivative $\frac{dP_{ik}^{(t)}}{dt}$

of the fundamental equation

(15') $$P^{(t_1 + t_2)} = P^{(t_1)} P^{(t_2)}$$

and the fixed condition

(16') $$P^{(t)} \longrightarrow E \qquad (t \rightarrow 0).$$

For this purpose, we first make the following remark. If the matrix P has roots $\lambda_1, \ldots, \lambda_n$ and the rotation matrices U, V, then P^v has the roots $\lambda_1^v, \ldots, \lambda_n^v$ and the same rotation matrices as P. From $UPV = \Lambda$, it follows identically through powers that $UP^v V = \Lambda^v$ and from this on the basis of the reflection made in Section 2 (p. 6) (follows) the correctness of our remark.

According to the assumption (16'), we now take a time interval (0,h) inside which the roots of $P^{(t)}$ lie together in the right half plane. It does not hurt for the purpose of easier writing to take h = 1. Let the rotation matrices of $P^{(1)}$ be U, V, the roots $\lambda_1, \ldots, \lambda_n$; further set $\chi_i = \log \lambda_i$, where the branch of the logarithm is fixed by the restriction $|\Im(x_i)| \leqslant \frac{\pi}{2}$.

On the basis of the remarks made above, the matrix $P^{(1/2)}$ has for example the rotation matrices U, V and the roots $\wedge\sqrt{\lambda_i} = e^{\chi_i/2} \ldots (i = 1, \ldots, n)$; the values $e^{\chi_i/2}$ do not come into consideration since they lie in the left half plane. In the same way we show generally that the matrix $P^{(1/2v)}$ has the rotation matrices U, V and the roots $e^{\chi_i/2v}$. If we define $R^{(t)}$ in consequence of (26') and $P_*^{(t)}$ by

$$P_*^{(t)} = VR^{(t)}U,$$

then $P^{(t)}$ and $P_*^{(t)}$ agree with one another for t = 1, 1/2, 1/4, ..., it follows also because of (15') for all t values expressible through finite double divisions; finally, because of the continuity of $P^{(t)}$, it follows from (15') and (16') for all t. But the elements of $P_*^{(t)}$ are evidently differentiable with respect to t, therefore those of $P^{(t)}$ are also, which was to be proved.

6. <u>The Connection between the Chain and the Stationary Processes.</u> If the differential law of a stationary process exists, that is to say, the matrix A is given, and one limits the observations to an equidistant series of timepoints,

perhaps t = 0, 1, ... , we have a Markoff chain for which the matrix P follows from (27) for t = 1. This proposition shows, that P has the characteristic roots

(28) $\lambda_1 = 1, \quad \lambda_2 = e^{\chi_2}, \; . \; . \; . \; , \quad \lambda_n = e^{\chi_n}$

and the rotation matrices U, V.

We turn now to the oposite problem; if P is given, can we give such a matrix A with the properties (18) so that the transition matrix $P^{(t)}$ determined by A agrees with P for t = 1? In short, can the given chain be interpolated stochastically - significantly?

From the statements above it follows immediately that the roots of the desired matrix A must have the values $\chi_\nu = \log \lambda_\nu$ while its rotation matrices must agree with those of P. If we set (the fixing of the branch of the logarithm may first of all be left uncertain)

(29) $$L = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & \log \lambda_2 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & \log \lambda_n \end{pmatrix},$$

then there is

(30) $$A = VLU$$

or writing the desired matrix in full

(31) $$a_{ik} = u_{2k}\, v_{i2} \log \lambda_2 + \; . \; . \; . \; + u_{nk}\, v_{in} \log \lambda_n \quad (i, k = 1, \; . \; . \; . \; , n).$$

Let it be noted that complex roots λ - which always appear in conjugate pairs - bring no difficulties with them since in (29) we merely take the conjugate branch for the logarithm of the conjugate root. Then λ_α, λ_β are two conjugate roots, then in (31) $u_{\alpha k}$, $u_{\beta k}$ are also conjugate and also $v_{i\alpha}$, $v_{i\beta}$, and the terms λ_α, λ_β come together to give a real term. On the other hand, a negative brings in a complex term in (31).

Further if the last condition (18) is definitely fulfilled, then because $UV = E$, we have

$$\sum_k u_{jk} = \sum_k u_{jk} v_{kl} = 0 \qquad (j \neq 1)$$

and from this it follows by summation of (31) over k, that $\sum_k a_{ik} = 0$.

But if the matrix (30) is also real and our interpolation problem solved formally, then it is not yet certain that the conditions

(32) $$a_{ik} \geqslant 0 \quad (1 \neq k), \; a_{ii} \leqslant 0$$

are sufficient. Actually examples can be given which permit none or one or a finite number of stochastically - significant solutions. That infinitely many solutions can never appear we will show in the next section. Since we anticipate this result, we can put together our results up to this point in the following way;

The Markoff chain determined by P can be interpolated stochastically - significantly, and indeed by means of the "matrix of differential transition probabilities" (30) in case the elements of this matrix are non-negative. The problem can have either none or one or a finite number of solutions.

7. Interpolation Criteria. It is of a certain interest to set up criteria that allow us to decide the interpolatibility of a Markoff chain without calculating the entire matrix (30).

To this end we make first a comment on the position of the roots χ_i of a matrix A with the properties (18); if k is the greatest of the numbers $-a_{ii}$ ($i = 1, \ldots, n$) then the matrix $\frac{A}{K} + E$ is stochastic and its roots $\frac{\chi_i}{k} + 1$ are hence, according to Frobenius (page 5) absolutely smaller or equal to one. It is thus true

(33) $$|\chi_i + k| \leqslant k \qquad (i = 1, \ldots, n)$$

that is, the roots χ_i lie inside a certain circle which touches the imaginary axis on the left in the origin.

Let now P be a given stochastic matrix, and D its determinant. If then A is a matrix with properties (18) that solves the interpolation problem belonging to P, then we can indicate an upper limit for the radius of the above mentioned circle. In the characteristic equation of A is namely the coefficient of χ^{n-1} equal to $a_{11} + \cdots + a_{nn}$, thus we have

(34) $$a_{11} + \cdots + a_{nn} = \chi_1 + \cdots + \chi_n = \log (\lambda_1 \cdots \lambda_n) = \log D.$$

Since the a_{ii} together are ≤ 0, we have in $|\log D| = \log \frac{1}{D}$, the desired limit. The roots of A belong thus to the circular slice

(35) $$\chi - \log D \leq \log D$$

and indeed because of $\chi_1 + \cdots + \chi_n = \log D$, in the right half.

We conclude first from this that only a finite number of branches of the logarithm log λ_i can give stochastically - significant matrices A. As was asserted in Section 6, the interpolation problem is thus only solvable in a finite number of ways.

Further we obtain the criterion; in order that P be interpolable stochastically - significantly, it is necessary that $D > 0$ and that the principal values of $\log \lambda_1, \ldots, \log \lambda_n$ lie within the circles (35). In order that P be interpolable in a μ - fold way it is necessary that $|\log D| \geq (\mu - 1)\pi$.

In particular the above criterion teaches us that the matrix P of an interpolable chain can have no vanishing roots, no roots with absolute value one (outside of $\lambda_1 = 1$) and finally, no simple negative roots. Then for a root belonging to negative λ, χ = long $|\lambda| + \pi_i$ there would correspond necessarily a conjugate root $\bar{\chi} = \log |\lambda| - \pi_i$ and λ will then be a double root.

8. <u>An Example</u>. We consider that particular Markoff chain, whose law is given by the cyclic matrix,

$$(36) \qquad P = \begin{bmatrix} p & p_1 & p_2 & \cdots & p_{n-1} \\ p_{n-1} & p & p_1 & \cdots & p_{n-2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ p_1 & p_2 & p_3 & & p \end{bmatrix}$$

Let η be a possible value of $\sqrt[n]{1}$, thus there is, as we can easily confirm, a characteristic root of P

$$(37) \qquad \lambda = p = p_1\eta + p_2\eta^2 + \cdots + p_{n-1}\eta^{n-1};$$

(that is) generally

$$\lambda_v = p + \sum_{\mu=1}^{n-1} p_\mu e^{\mu\frac{2\pi i}{n}(v-1)} \qquad (v = 1, \ldots, n).$$

The real vectors belonging to the roots (37) are determined from the equations (8). These are evidently satisfied with

$$v_1 = 1, v_2 = \eta, \ldots, v_n = \eta^{n-1},$$

$$u_1 = \eta^n = 1, u_2 = \eta^{n-1}, \ldots, u_n = \eta$$

There can therefore with $\varepsilon = e^{\frac{2\pi i}{n}}$, be chosen

$$v_{ik} = \varepsilon^{(i-1)(k-1)}, \quad u_{ik} = \frac{1}{n}\varepsilon^{(i-1)(1-k)}$$

and the coefficients in (13) and (27') become

$$(38) \qquad v_{ij}\, u_{jk} = \frac{1}{n}\varepsilon^{(i-k)(j-1)}.$$

Hence we learn that the interpolation matrix of P is likewise cyclic.

To obtain a single example of the result of the previous section, we consider the case n = 3. We have

$$\lambda_2 = p - \frac{p_1 + p_2}{2} + \frac{p_1 - p_2}{2}\sqrt{3}\, i, \qquad \lambda_3 = \bar{\lambda}_2.$$

We set for the sake of brevity

$$\lambda_2 = \rho e^{i\theta}, \quad \lambda_3 = \rho e^{-i\theta} \qquad (|\theta| \leqslant \pi),$$

and find under the application of the formulas (27') and (38) that if the elements of A be denoted in the same way as those from Γ,

$$a = \frac{2}{3} \log \rho ,$$

$$a_1 = -\frac{1}{3} \log \rho - \frac{\sqrt{3}}{3} \theta ,$$

$$a_2 = -\frac{1}{3} \log \rho + \frac{\sqrt{3}}{3} \theta .$$

Here a is evidently negative. In order that a_1, a_2 be non-negative it is necessary and sufficient that (we suppose for example $\theta \geqslant 0$)

$$\theta \leqslant \frac{1}{\sqrt{3}} \log \frac{1}{\rho} .$$

We obtain two significant solutions of the interpolation problem as soon indeed as the argument value $\theta - 2\pi$ provides a_1, a_2 non-negative, that is if

$$2\pi - \theta \leqslant \frac{1}{\sqrt{3}} \log \frac{1}{\rho} ,$$

one obtains three solutions if

$$\theta + 2\pi \leqslant \frac{1}{\sqrt{3}} \log \frac{1}{\rho}$$

and so on.

9. The Age of a Distribution. We turn now to the second of the problems proposed in the introduction. Its treatment proceeds in a wholly parallel way for chains and for stationary processes. We avail ourselves in the following in both cases of the notation $p_{ik}^{(t)}$ for the transition matrix.

Let there be an absolute probability distribution prescribed, for example, for the moment t = 0,

$$(39) \qquad Q^{(0)} = (q_1^{(0)}, \ldots, q_n^{(0)}); \; q_i^{(0)} \geqslant 0; \; \sum_i q_i^{(0)} = 1.$$

We ask first; can we determine for a given earlier time point - t ($t > 0$) such a distribution $Q^{(-t)} = (q_1^{(-t)}, \dots, q_n^{(-t)})$ with the properties

(40 a) $$q_i^{(-t)} \geqslant 0$$

and

(40 b) $$\sum_i q_i^{(-t)} = 1$$

that $Q^{(0)}$ results from our process from $Q^{(-t)}$, that is

(41) $$q_k^{(0)} = q_1^{(-t)} p_{1k}^{(t)} + \dots + q_n^{(-t)} p_{nk}^{(t)}$$

in short; is there a stochastically - significant solution of the matrix equation

(41') $$Q^{(-t)} P^{(t)} = Q^{(0)} \ ?$$

In formal algebra the question is quite simple to answer. Only the inversion of the matrix $P^{(t)}$ is involved.

This operation is generally impossible because the determinant $|P^{(t)}| = 0$, that is, whenever the matrix has a null root. This is, as we know, only possible with chains. We put this case aside. In all other cases $(P^{(t)})^{-1}$ is determined in an unique way. Under the application of the formula found earlier for positive t

(27) $$P^{(t)} = VR^{(t)}U$$

$(P^{(t)})^{-1}$ can be easily expressed explicitly; it becomes

$$(P^{(t)})^{-1} = U^{-1} (R^{(t)})^{-1} V^{-1} = V(R^{(t)})^{-1}U,$$

and hence evidently $(R^{(t)})^{-1} = R^{(-t)}$

(42) $$(P^{(t)})^{-1} = VR^{(-t)} U = P^{(-t)} ,$$

whereby $P^{(t)}$ for $t < 0$ is defined by (27). Generally, one can by virtue of (27) calculate with $P^{(t)}$ as exactly as with powers.

The solution of the system (41') will be hence

(43) $$Q^{(-t)} = Q^{(0)} P^{(-t)},$$

which written in full, becomes

(43') $$q_i^{(-t)} = q_1^{(0)} p_{1i}^{(-t)} + \ldots\ldots + q_n^{(0)} p_{ni}^{(-t)} \qquad (i = 1,.., n)$$

The solution automatically satisfies the condition (40 b). This follows from (41) through summation over k.

But we do not say that (43) satisfies the condition (40 a) or describes an actual stochastically significant distribution.

We show easily that if this holds for an earlier, it holds also for a later time point. It may be expressly in the state $t_1 < t_2 < 0$; then it follows

$$Q^{(t_1)} P^{(t_1)} = Q^{(t_1)} P^{(t_2-t_1)} P^{(-t_2)} = Q^{(0)},$$

and thus moreover $Q = Q^{(t_2)}$ is a single solution of the equation $QP^{(-t_2)} = Q^{(0)}$, therefore

(44) $$Q^{(t_2)} = Q^{(t_1)} P^{(t_2-t_1)},$$

and the elements of $Q^{(t_2)}$ must then be $\geqslant 0$.

Hence, we conclude that there is given a well determined period $-\tau$ $(0 \leq \tau \leq \infty)$, which has the following properties; for $t \geqslant -\tau$, $Q^{(t)}$ is a distribution, for $t < -\tau$ (it is) not. We shall indicate by τ the age of the distribution $Q^{(0)}$. From what has been said, it follows directly that $\tau + t$ is the age of $Q^{(t)}$. In the case of a stationary process, τ is evidently the smallest not directly divided zero point of any of the probabilities $q_i^{(-t)}$ $(t \geqslant 0)$ defined.

10. Finally we wish to give necessary and sufficient conditions for $\tau = \infty$. For the purpose, we call attention to the idea of a stationary distribution. A distribution $Q = (q_1 \ \ldots \ q_n)$ is called stationary which satisfies the equation

$QP^{(t)} = Q$ for all t. Besides Q is in the case of a chain, a solution of the system

$$p_{1i}\,q_1 + \cdots + p_{ni}\,q_n = q_i \qquad (i = 1, \ldots , n)$$

while this distribution in the case of a stationary process satisfies the equations

$$a_{1i}\,q_1 + \cdots + a_{ni}\,q_n = 0 \qquad (i = 1, \ldots , n).$$

The solutions of this system are noted in Sections 2 and 4 for the roots λ_1 or χ_1 pertaining to the real vector $U_{1|}$. Their components have the same symbols as one can show easily. Moreover on account of $\sum_i u_{1i} = \sum_i u_{1i}\,v_{i1} = 1$, they are already regulated in such a way that they are positive and have the sum one. There is thus for a single valued stationary distribution

(45) $$q_1 = u_{11},\ q_2 = u_{12},\ \ldots ,\ q_n = u_{1n}.$$

We need distinguish only two cases. Either (1°) there are jointly $|\lambda_v| < 1$ $(v > 2)$ or (2°) there is besides $\lambda_1 = 1$ also other roots with absolute value one. The second case can, as we know, happen only with chains, not with stationary processes. The matrix P will then be called imprimitive.

(1°) In this case, it appears directly from equations (13) or (27'), that

(46) $$p_{ik}^{(t)} \to q_k \qquad (t \to \infty)$$

is true, independent of i. We conclude from this that the age of the distribution $Q^{(0)}$ can be infinite when and only when $Q^{(0)} = Q$. If certainly $-\tau < -t < 0$, then it is true that

(47) $$q_k^{(0)} = q_1^{(-t)}\,p_{1k}^{(t)} + \cdots + q_n^{(-t)}\,p_{nk}^{(t)}.$$

If now $\tau = \infty$, then t can be chosen sufficiently large that the right side comes near the chosen value $q_k = \lim p_{ik}^{(t)}$; it is also necessary that $q_k^{(0)} = q_k$. Obviously this condition is also sufficient in the case $Q^{(0)} = Q$, then also (41') is always satisfied with $Q^{(-t)} = Q$.

(2^{0}) Now consider a Markoff chain with imprimitive matrix P. From such a matrix the following can be proved (Romanovsky (7) p. 261, compare also von Mises (6) p. 545). If there is such a whole number r that it is divisible by n, let n = rd, then P^r is totally decomposed, that is, P^r is of the form

$$P^r = \begin{bmatrix} M_1 & 0 & \dots & 0 \\ 0 & M_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & M_r \end{bmatrix}$$

with appropriate enumeration of the conditions, where the matrices $M_1, \dots, M_r$ are primitive quadratic partition matrices of order d.

We limit ourselves to a look at the time point sequence, 0, -r, -2r, ... The absolute distribution in one of these time points, for example $t = -vr$, is determined from

$$Q^{(-vr)} P^{vr} = Q^{(0)}.$$

This system obviously falls among the r independent systems with d equations always. We look at the first, for example; if we denote the element from $M_1^{(v)}$ by $m_{ik}^{(v)}$, this system maintains the form

$$m_{ik}^{(v)} q_1^{(-vr)} + \dots + m_{dk}^{(v)} q_d^{(-vr)} = q_k^{(0)} \; (k=1, \dots, d). \tag{48}$$

We reduce it to a form completely analogous to (47) in which we divide by the sum $s = q_1^{(0)} + \dots + q_d^{(0)}$ and set

$$\frac{q_i^{(t)}}{s} = q_i'(t) .$$

We thus obtain the system

(48') $$m_{1k}^{(v)} q_1'^{(-vr)} + \dots + m_{dk}^{(v)} q_d'^{(-vr)} = q_k'^{(0)} \quad (k = 1, \dots, d)$$

We see, exactly as in (1) that this system permits only stochastically-significant solutions for any large v, whenever the distribution $(q_1'^{(0)}, \dots, q_d'^{(0)})$ with that from the matrix M_1 belongs to a stationary distribution. If the solution of the systems $UP = U$ is $U_1 = (u_{11}, \dots, u_{1n})$ as earlier (Through regulating the relation $u_{11} + \dots + u_{1n} = 1$), then this vector satisfies also the system $Up^{vr} = U$; and also if the independent system in r is broken down, then, besides $(u_{11}, \dots, u_{1d})$ is the solution of the systems $UM_1^v = U$. The components of $(q_1^{(0)}, \dots, q_d^{(0)})$ must also be proportional to the numbers $u_{11}, \dots, u_{1d}$. If this is the case, then (48') is real and solvable with $q_i'^{(-vr)} = q_i'^{(0)}$ for all v.

These reflections can be carried through for each of the index groups

(49) $$1, \dots, d; \; d + 1, \dots, 2d; \; \dots; \; (r - 1) d + 1, \dots, n.$$

The distribution $Q^{(0)}$ is then and only then arbitrarily far "backwards continuable", whenever the $q_i^{(0)}$ inside of each group (49) is proportional to the corresponding u_{1i}.

We sum up.

Let there be a law P applied to a Markoff process; we assume that P has only single valued non-vanishing characteristic roots.

Then there comes from each given distribution $Q^{(0)}$, a number τ $(0 \leq \tau \leq \infty)$, the age of the distribution with the following properties: to each $t \leq \tau$, there is a distribution $Q^{(-t)}$ that subject to the given process, produces after a lapse of time t, the distribution $Q^{(0)}$; for $t > \tau$, there is no such distribution.

In order tha $\tau = \infty$, it is necessary and sufficient:

(1) with stationary processes and chains with primitive matrices, that $Q^{(0)}$ correspond with the stationary distribution of the processes.

(2) with chains with imprimitive matrices, that the $q^{(0)}$ be proportional inside each of the cyclic condition groups with the corresponding probabilities of the stationary distributions.

Literature

(1) J. Hadamard and M. Frechet; Sur les probabilitiés discontinues des évènements en chaîne, Z. angew. Math. Mech. XIII (1933).

(2) B. Hostinsky; Méthodes générales du calcul des probabilitiés Gauthier-Villars, Paris, 1931.

(3) A. Kolmogoroff; Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Math. Ann. 104. (1931).

(4) A. Kolmogoroff; Zur Umkehrbarkeit der Naturgesetze. Math. Ann. 104, (1936).

(5) A. Kolmogoroff; Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen, Mat. Sbornik 1936.

(6) R. v. Mises; Wahrscheinlichkeitrechnung, etc. Leipzig and Vienna, 1931.

(7) V. Romanovsky; Recherches sur les chaînes de Markoff. Acta Math. 66, 1936.