

ON THE NOTION OF RECURRENCE IN DISCRETE STOCHASTIC PROCESSES

M. KAC¹

1. **Introduction.** It is the purpose of this note to discuss "statistical" versions of the "Wiederkehrsatz" and the Poincaré cycle and to relate these versions to the ones encountered in dynamical considerations.² Although the content of the note is entirely elementary and in part known it is hoped that it will help elucidate some of the basic notions of statistical physics.

2. **Recurrence and mean recurrence time in a class of discrete stochastic processes.** Let x_1, x_2, \dots be a sequence of random variables each capable of assuming the values a_1, a_2, \dots . We shall say that the sequence x_1, x_2, \dots is a stationary process if: (a) for each j the probability

$$\text{Prob. } \{x_n = a_j\}$$

is independent of n ; (b) for each set of values $a_{s_1}, a_{s_2}, \dots, a_{s_r}$ the probability

$$\text{Prob. } \{x_{k_1} = a_{s_1}, x_{k_2} = a_{s_2}, \dots, x_{k_r} = a_{s_r}\}$$

depends only on the differences $|k_i - k_j|$.

Let

$$\text{Prob. } \{x_n = a_j\} = W_1(a_j) = W(a_j)$$

and

$$\text{Prob. } \{x_1 = a_{s_1}, x_2 = a_{s_2}, \dots, x_r = a_{s_r}\} = W_r(a_{s_1}, a_{s_2}, \dots, a_{s_r}).$$

We shall use the notation \bar{a}_k to denote any a_j different from a_k . For instance,

$$\text{Prob. } \{x_1 = a_1, x_2 \neq a_2\} = W_2(a_1, \bar{a}_2),$$

Received by the editors March 3, 1947.

¹ John Simon Guggenheim Memorial Fellow.

² That the theory of stationary stochastic processes is mathematically equivalent with an "ergodic" theory (to which one is also led by dynamical considerations) was clearly recognized by Doob in 1934. See J. L. Doob, *Stochastic processes and statistics*, Proc. Nat. Acad. Sci. U.S.A. vol. 20 (1934) pp. 376-379. The analogies discussed in the present paper are but particular cases of Doob's general equivalence principle. However, since the motivations underlying the statistical and the dynamical points of view are of a somewhat different physical character it seemed desirable to treat both cases separately.

where it is understood that

$$W_2(a_1, \bar{a}_2) = \sum_{r \neq 2} W_2(a_1, a_r).$$

We shall assume that $W(a_j) \neq 0$ for all j .

We also define the conditional probabilities $P_k(a_{s_1} | a_{s_2}, \dots, a_{s_k})$ by the usual formula

$$(1) \quad P_k(a_{s_1} | a_{s_2}, \dots, a_{s_k}) = \frac{W_k(a_{s_1}, a_{s_2}, \dots, a_{s_k})}{W(a_{s_1})}.$$

To simplify the notation we shall use a without a subscript to denote a chosen and then fixed a_j .

We now established the following, almost trivial, formula

$$(2) \quad \begin{aligned} W_{k+2}(a, \bar{a}, \dots, \bar{a}, a) &= W_k(\bar{a}, \bar{a}, \dots, \bar{a}) - 2W_{k+1}(\bar{a}, \bar{a}, \dots, \bar{a}) \\ &\quad + W_{k+2}(\bar{a}, \bar{a}, \dots, \bar{a}), \quad k \geq 0, \end{aligned}$$

where it is understood that $W_0 = 1$. In fact, we have

$$\begin{aligned} W_{k+2}(a, \bar{a}, \dots, \bar{a}, a) &= W_{k+1}(a, \bar{a}, \dots, \bar{a}) - W_{k+2}(a, \bar{a}, \dots, \bar{a}), \\ W_{k+1}(a, \bar{a}, \dots, \bar{a}) &= W_k(\bar{a}, \bar{a}, \dots, \bar{a}) - W_{k+1}(\bar{a}, \bar{a}, \dots, \bar{a}), \\ W_{k+2}(a, \bar{a}, \dots, \bar{a}) &= W_{k+1}(\bar{a}, \bar{a}, \dots, \bar{a}) - W_{k+2}(\bar{a}, \bar{a}, \dots, \bar{a}), \end{aligned}$$

and formula (2) follows at once if one also notices that

$$W_1(a) = 1 - W_1(\bar{a}) = W_0 - W_1(\bar{a}).$$

THEOREM 1. *For each a*

$$(3) \quad \sum_{k=0}^{\infty} P_{k+2}(a | \bar{a}, \dots, \bar{a}, a) = 1.$$

In view of the definition (1) of conditional probabilities it is sufficient to prove that

$$(4) \quad \sum_{k=0}^{\infty} W_{k+2}(a, \bar{a}, \dots, \bar{a}, a) = W(a).$$

Using (2) we obtain

$$(5) \quad \begin{aligned} \sum_{k=0}^n W_{k+2}(a, \bar{a}, \dots, \bar{a}, a) &= 1 - W(\bar{a}) - (w_{n+1} - w_{n+2}) \\ &= W(a) - (w_{n+1} - w_{n+2}) \end{aligned}$$

where from now on we shall use the abbreviation

$$w_s = W_s(\bar{a}, \bar{a}, \dots, \bar{a}).$$

Since $w_{n+1} \geq w_{n+2}$ and the w 's are non-negative we deduce that $\lim_{n \rightarrow \infty} w_n$ exists and hence

$$\lim_{n \rightarrow \infty} (w_{n+1} - w_{n+2}) = 0.$$

This together with (5) implies (4) and hence (3). The interpretation of (3) is almost immediate.

In fact, $P_{k+2}(a | \bar{a}, \dots, \bar{a}, a)$ is the probability that if the "state" a was observed at the beginning, then it will be observed again, for the first time, after $k+1$ steps. Consequently,

$$\sum_{k=0}^{\infty} P_{k+2}(a | \bar{a}, \dots, \bar{a}, a)$$

is the probability that if a was observed at the beginning it will eventually be observed again (the events of observing a for the first time after 1, 2, 3, \dots steps are obviously mutually exclusive).

Theorem 1 can thus be restated as follows: *with probability 1 each state is bound to recur.*

This is the "statistical" version of the "Wiederkehrrsatz" and it will be seen in the next section that it implies the classical "Wiederkehrrsatz."

If we think of the random variable x_n as being associated with the observation of the system at time $n\tau$ we can define the mean recurrence time of the state a as

$$(6) \quad \theta_a = \sum_{k=0}^{\infty} (k+1)\tau P_{k+2}(a | \bar{a}, \dots, \bar{a}, a).$$

This is the statistical analogue of Poincaré's cycle. In this connection we have the following theorem.

THEOREM 2. *If*

$$(7) \quad \lim_{n \rightarrow \infty} W_n(\bar{a}, \bar{a}, \dots, \bar{a}) = 0$$

then

$$(8) \quad \theta_a = \frac{\tau}{W(a)}.$$

It will be evident from the proof that (7) is also necessary for the validity of (8).

It is sufficient to prove that

$$(9) \quad \sum_{k=0}^{\infty} (k+1)W_{k+2}(a, \bar{a}, \dots, \bar{a}, a) = 1.$$

Using (2) we have

$$\sum_{k=0}^{n-1} (k+1)W_{k+2}(a, \bar{a}, \dots, \bar{a}, a) = 1 - \{(n+1)w_n - nw_{n+1}\}.$$

Now,

$$(n+1)w_n - nw_{n+1} = n(w_n - w_{n+1}) + w_n$$

and since we have assumed that $w_n \rightarrow 0$ it is sufficient to prove that

$$\lim_{n \rightarrow \infty} n(w_n - w_{n+1}) = 0.$$

Since the partial sums of (9) form a nondecreasing sequence it follows that $n(w_n - w_{n+1}) + w_n$ is a nonincreasing sequence. Since furthermore its terms are non-negative we deduce that

$$\lim_{n \rightarrow \infty} \{n(w_n - w_{n+1}) + w_n\}$$

exists.

Moreover, $\lim_{n \rightarrow \infty} w_n$ exists (this is true regardless of the assumption (7)) and hence $\lim_{n \rightarrow \infty} n(w_n - w_{n+1})$ exists.

Now, the series of non-negative terms

$$\sum_{n=1}^{\infty} (w_n - w_{n+1})$$

converges (since $\lim_{n \rightarrow \infty} w_n$ exists) and hence (since $\lim_{n \rightarrow \infty} n(w_n - w_{n+1})$ exists)

$$\lim_{n \rightarrow \infty} n(w_n - w_{n+1}) = 0.$$

This completes the proof of (8).

3. The classical formulations. Let Ω be a set in which a Lebesgue measure has been established, and assume for the sake of simplicity that the measure of Ω is 1. Let ω denote an element of Ω and let T be a one-to-one measure-preserving transformation of Ω into itself. Let furthermore A be a measurable subset of Ω . The classical "Wiederkehrsatz" can now be stated as follows:

THEOREM 1'. *For almost every ω in A there exists an $n \geq 1$ such that $T^n \omega$ is again in A ,*

Let $f(\omega)$ be the characteristic function of the set A and consider the sequence of measurable functions

$$f(\omega), f(T\omega), f(T^2\omega), \dots$$

These functions can be considered as random variables which form a stationary stochastic process. Each random variable can assume only values 1 and 0 with respective probabilities $|A|$ and $1 - |A|$.

Note that $W_{k+2}(1, 0, \dots, 0, 1)$ is nothing but the measure of the set of those ω in A which return to A for the first time after $k+1$ steps (that is, $T^{k+1}\omega$ is the first iteration to fall into A).

Formula (4) thus asserts that the measure of the set of those ω which eventually return to A is equal to the measure of A , and this is equivalent to the statement of Theorem 1'. It should be emphasized that if we were to rewrite the proof of Theorem 1 in the terminology of Theorem 1' the resulting proof would not differ essentially from the orthodox one.³

We are led to a more interesting situation in trying to reformulate Theorem 2. The interest lies in the fact that condition (7) is essentially equivalent to the condition that T is metrically transitive.

For $\omega \in A$ let $n(\omega)$ denote the first $n \geq 1$ for which $T^n\omega \in A$. Thus $n(\omega)$ is the Poincaré cycle of ω relative to the set A . We have now the following theorem.

THEOREM 2'. *If T is metrically transitive and $|A| > 0$ then,*

$$(10) \quad \int_A n(\omega) d\mu = 1.$$

If we define the *average Poincaré cycle* relative to A as

$$\bar{n}_A = \frac{1}{|A|} \int_A n(\omega) d\mu$$

we can restate (10) in the form

$$\bar{n}_A = \frac{1}{|A|}$$

which is more reminiscent of (8).

It is clear that all one has to prove is that metric transitivity of T implies condition (7) which in the present terminology can be written as

³ See C. Carathéodory, *Über den Wiederkehrrsatz von Poincaré*, Berliner Berichte (1919) pp. 580–584.

$$\lim_{n \rightarrow \infty} \int_{\Omega} \prod_{k=0}^n (1 - f(T^k \omega)) d\mu = 0.$$

Let B denote the set whose characteristic function is

$$\prod_{k=0}^{\infty} (1 - f(T^k \omega)).$$

We must prove that $|B| = 0$.

Consider the sets $T^n B$. Since the characteristic function of $T^n B$ is

$$\prod_{k=n}^{\infty} (1 - f(T^k \omega))$$

we see that the sets $T^n B$ form a nondecreasing sequence of sets all of which include B . Moreover, $|T^n B| = |B|$.

Let the set C be the limit of the sets $T^n B$, that is,

$$C = \lim_{n \rightarrow \infty} T^n B.$$

Clearly $TC = C$ and hence, by metric transitivity, $|C|$ is either 0 or 1. Since $|C| = |B|$ we have that either $|B| = 0$ or $|B| = 1$. If $|B| = 1$ then noticing that B is contained in the complement of A we would have $|A| = 0$, contrary to the assumption. Hence $|B| = 0$ and the proof of Theorem 2' is completed.

An even shorter proof of the fact that $|B| = 0$ can be given by appealing to Birkhoff's ergodic theorem. In fact, this theorem asserts (T being metrically transitive) that for almost all ω we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(T^k \omega) = |A| > 0.$$

If $\omega \in B$ then $f(T^k \omega) = 0$ and hence for all $\omega \in B$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n f(T^k \omega) = 0.$$

Thus $|B| = 0$.

It should be noted that if T is not metrically transitive then a set A can be found such that $|A| > 0$ and

$$\int_A n(\omega) d\mu < 1.$$

In fact, let D be such that $0 < |D| < 1$ and $TD = D$. Put $A = \Omega - D$ and note that in this case

$$\lim_{n \rightarrow \infty} \int_{\Omega} \prod_{k=0}^n (1 - f(T^k \omega)) d\mu = |D|.$$

This implies that

$$\int_A n(\omega) d\mu = 1 - |D| < 1.$$

4. **Smoluchowski's definition of the mean recurrence time.** In his important researches concerning the question of irreversibility in thermodynamics, Smoluchowski⁴ used a different definition of the mean recurrence time. He defines the mean recurrence time of the state a by means of the formula

$$(11) \quad \theta_a^* = \tau \frac{1 - W(a)}{W_2(\bar{a}, a)}.$$

In order to justify this definition he considered an actual sequence of observations which consists of a 's and \bar{a} 's. He then defines the *observed* mean recurrence time as τ times the ratio of the number of \bar{a} 's to the number of unbroken blocks of \bar{a} 's. For instance, from the sequence of fifteen observations

$$a\bar{a}\bar{a}\bar{a}\bar{a}a\bar{a}\bar{a}\bar{a}\bar{a}a$$

he would find the observed mean recurrence time to be $7\tau/4$. Denoting by $k_n(\bar{a})$ the number of \bar{a} 's in a sequence of n observations and by $l_n(\bar{a})$ the number of unbroken blocks of \bar{a} 's in that sequence we see that the observed mean recurrence time is

$$\tau \frac{k_n(\bar{a})}{l_n(\bar{a})}.$$

Let furthermore $k_n(\bar{a}, a)$ be the number of pairs of adjacent observations of the form (\bar{a}, a) . Then,

$$l_n(\bar{a}) = k_n(\bar{a}, a)$$

if the sequence ends with a , and

⁴ M. v. Smoluchowski, *Drei Vorträge über Diffusion, Brownsche Molekularbewegungen und Koagulation von Kolloidteilchen*, Physikalische Zeitschrift vol. 17 (1916) pp. 557–571 and 585–599, in particular pp. 564–568.

The author was also fortunate to have had access to unpublished notes of Professor G. E. Uhlenbeck, where many aspects of Smoluchowski's work were thoroughly analyzed. He is also indebted to Professor Uhlenbeck for many stimulating discussions.

$$l_n(\bar{a}) = k_n(\bar{a}, a) + 1$$

if the sequence ends with \bar{a} .

If, in the limit as $n \rightarrow \infty$, we replace (following Smoluchowski) the frequencies

$$\frac{k_n(\bar{a})}{n} \quad \text{and} \quad \frac{k_n(\bar{a}, a)}{n}$$

by the corresponding probabilities

$$W(\bar{a}) = 1 - W(a) \quad \text{and} \quad W_2(\bar{a}, a) = W_2(a, \bar{a})$$

we can replace (again in the limit as $n \rightarrow \infty$) the ratio $k_n(\bar{a})/l_n(\bar{a})$ by

$$\frac{1 - W(a)}{W_2(\bar{a}, a)}.$$

Smoluchowski considered the process of replacing limits of frequencies by probabilities as self-evident. Actually he tacitly assumed the validity of certain laws of large numbers. Fortunately, Smoluchowski applied his definition mainly to the process of fluctuations of concentration for which the necessary laws of large numbers can easily be established.

Smoluchowski's definition can be also introduced in an "a priori" fashion by setting

$$(12) \quad \theta_a^* = \sum_{k=0}^{\infty} (k+1) \tau P_{k+3}(a, \bar{a} | \bar{a}, \dots, \bar{a}, a).$$

It can then be shown using (2) and (7) that

$$(13) \quad \theta_a^* = \tau \frac{1 - W(a)}{W_2(a, \bar{a})} = \tau \frac{1 - W(a)}{W_2(\bar{a}, a)}.$$

It is well to point out that in accepting this definition we agree not to count as a recurrence the event when a state is followed by itself.

By giving an "a priori" definition we have not eliminated the necessity for laws of large numbers. In fact, laws of large numbers are indispensable whenever we want to check predictions based on "a priori" results through calculations performed on actual sequences of observations.

The advantage of Smoluchowski's definition can be best appreciated if we consider the situation in which the random variables x_n are derived from a continuous stochastic process $x(t)$. In fact, in most applications

$$x_n = x(n\tau).$$

In many cases the limit, as τ approaches 0, of

$$\frac{\tau}{W_2(a, \bar{a})}$$

exists and is different from 0 (note that now $W_2(a, \bar{a})$ depends on τ). Thus it is often possible to define the mean recurrence time in the limit as observations are taken continuously.

If we use the definition of §2 we always get in the limit as $\tau \rightarrow 0$ the trivial answer 0. The reason for this is quite obvious, inasmuch as for small τ the probability of a state being followed by itself is very close to 1 and consequently too much weight is attached to what after all is only a fake recurrence.

CORNELL UNIVERSITY