Journal of Statistical Physics, Vol. 46, Nos. 5/6, 1987

Variations on a Theme by Mark Kac

P. W. Kasteleyn[1]

Received December 2, 1986

Mark Kac's theorem on the mean recurrence time in a stationary stochastic process in discrete time with discrete states is taken as the starting point for a series of variations, most of which are formulated in terms of 0–1 processes. Whereas the original theorem deals with the mean recurrence time of a given state under the condition that the state is realized at time 0, this condition is dropped in part of the variations; two others refer to the variance of the recurrence time and two to the Poincaré cycle of a dynamical system. Most variations consist in inequalities and formal identities for the mean first-arrival time and subsequent recurrence times for the given state.

KEY WORDS: Stationary stochastic process; 0–1 process; first-passage time; recurrence time; Poincaré cycle; inequalities.

INTRODUCTORY NOTES

This paper is devoted to a subject to which Mark Kac made a fundamental contribution in his 1947 paper "On the notion of recurrence in discrete stochastic processes."[(1)] In that paper, Kac considered a stationary stochastic process in discrete time with discrete states. He derived a simple, generally valid expression for the mean recurrence time of a given state under the condition that this state is realized at time 0. In the following we shall discuss several variations on this classic theme. Most of these variations are new; some have appeared scattered in the literature, but are presented here for the sake of composition, with proofs adapted to a uniform treatment. In part of the variations the condition on the initial state is dropped; two others refer to the variance rather than the expectation of the recurrence time. Kac also gave in his paper an application of his theorem to the Poincaré cycle of a dynamical system. We shall also briefly touch upon this subject.

[1] Instituut-Lorentz voor Theoretische Natuurkunde, 2311 SB Leiden, The Netherlands.

0022-4715/87/0300-0811$05.00/0

Since in the following, as in Ref. 1, the only events of interest are defined in terms of the occurrence or nonoccurrence of one single state, there is no loss of generality in restricting the discussion to two-state processes with states labeled, say, 0 and 1, where 1 will be the selected state.

Let, therefore, $X = (X_t)_{t=0}^{\infty}$ be a stationary 0–1 process. The probability measure describing the process will be denoted by P, the expectation and variance with respect to P of a random variable Y by $\langle Y \rangle$ and $\mathrm{Var}(Y)$, respectively, the probability of the event $X_0 = x_0$, $X_1 = x_1, \ldots, X_n = x_n$, with $x_j \in \{0, 1\}$, $0 \leqslant j \leqslant n$, by $P(x_0 x_1 \cdots x_n)$ with the following conventions: 0^m will represent a sequence of m (consecutive) 0's, 0^∞ an infinite sequence of 0's, y an arbitrary finite sequence of 0's and 1's, and q the frequently occurring probability $P(1)$; we assume $q > 0$. The stationary process X, or the measure P, is called *cyclic* with *repeating unit* $x_0 x_1 \cdots x_n$ if the only sequences that can occur with positive probability are s-fold repetitions of the unit, $(x_0 x_1 \cdots x_n)^s$, $s \leqslant \infty$, and uninterrupted subsequences of such a sequence.

The analysis concentrates on the consecutive (random) *arrival times* (passage times) T_k for the state 1 (i.e., $T_1 := \min\{t \geqslant 0\colon X_t = 1\}$; $T_k := \min\{t > T_{k-1}\colon X_t = 1\}$, $k \geqslant 2$), more in particular on T_1 and on the differences $n_k := T_{k+1} - T_k$, the *recurrence times* of the state 1; T_1 will be denoted alternatively by n_0.

Use will be made of a few simple identities, which are consequences of the following consistency relations for probabilities:

$$P(0y) + P(1y) = P(y0) + P(y1) = P(y)$$

where, as in what follows, stationarity is essential. The identities are:

(i) $P(0^m 1) = P(10^m)$

which follows from the fact that both sides are equal to $P(0^m) - P(0^{m+1})$;

(ii) $P(y0^\infty) = 0$ if y contains at least one 1

because then

$$P(y0^\infty) \leqslant P(10^\infty) = P(0^\infty) - P(00^\infty) = 0$$

[note that $P(0^\infty)$ need *not* be zero];

(iii) $\lim_{m \to \infty} P(0^m y) = 0$ if y contains at least one 1

because then

$$\lim_{m \to \infty} P(0^m y) \leqslant \lim_{m \to \infty} P(0^m 1) = P(10^\infty) = 0$$

by (i) and (ii);

(iv) $\sum_{m=0}^{\infty} P(0^m 1) + P(0^\infty) = 1$

obtained by iterating the consistency relations; finally, three identities that require a somewhat less trivial proof:

$$\text{(v)}\quad \sum_{m=1}^{\infty} m^r P(y0^{m-1}1) = \sum_{m=0}^{\infty} \Delta(m^r)[P(y0^m) - P(y0^\infty)], \qquad r \geqslant 1$$

$$\text{(vi)}\quad \sum_{m=1}^{\infty} m^r P(10^{m-1}y) = \sum_{m=0}^{\infty} \Delta(m^r)[P(0^m y) - \lim_{N\to\infty} P(0^N y)], \qquad r \geqslant 1$$

$$\text{(vii)}\quad \sum_{m=1}^{\infty} P(10^{m-1}y) = P(y) - \lim_{N\to\infty} P(0^N y)$$

where $\Delta(m^r) := (m+1)^r - m^r$. The identity (v) is derived as follows. For $M \in \mathbb{N}$

$$\begin{aligned}
\sum_{m=1}^{\infty} m^r P(y0^{m-1}1) &\geqslant \sum_{m=1}^{M} m^r P(y0^{m-1}1) + \lim_{N\to\infty} \sum_{m=M+1}^{N} M^r P(y0^{m-1}1) \\
&= \sum_{m=1}^{M} m^r [P(y0^{m-1}) - P(y0^m)] \\
&\quad + \lim_{N\to\infty} \sum_{m=M+1}^{N} M^r [P(y0^{m-1}) - P(y0^m)] \\
&= \sum_{m=0}^{M-1} \Delta(m^r)\, P(y0^m) - M^r \lim_{N\to\infty} P(y0^N) \\
&= \sum_{m=0}^{M-1} \Delta(m^r)[P(y0^m) - P(y0^\infty)]
\end{aligned}$$

Since the inequality holds for any M, we have

$$\sum_{m=1}^{\infty} m^r P(y0^{m-1}1) \geqslant \sum_{m=0}^{\infty} \Delta(m^r)[P(y0^m) - P(y0^\infty)]$$

If the right-hand side is infinite, so is the left-hand side. On the other hand,

$$\begin{aligned}
&\sum_{m=1}^{\infty} m^r P(y0^{m-1}1) \\
&\quad = \lim_{N\to\infty} \sum_{m=1}^{N} m^r [P(y0^{m-1}) - P(y0^m)] \\
&\quad = \lim_{N\to\infty} \left[\sum_{m=0}^{N-1} \Delta(m^r)\, P(y0^m) - N^r P(y0^N)\right] \\
&\quad \leqslant \sum_{m=0}^{\infty} \Delta(m^r)[P(y0^m) - P(y0^\infty)]
\end{aligned}$$

where the inequality follows from $P(y0^N) \geqslant P(y0^\infty)$. From the two inequalities for $\sum_m m^r P(y0^{m-1}1)$, the identity (v) follows. The "mirror image" of the derivation yields (vi) and a minor adaptation to the case $r = 0$ yields identity (vii).

Since $P(10^\infty) = 0$, the recurrence times n_k are defined for almost all realizations of X with $X_0 = 1$. For $k = 1$ this is the content of Theorem 1 of Ref. 1. Theorem 2 is the one that will serve as the theme for the variations to follow; we present it in an adapted form.

THEME

For a stationary 0–1 process $(X_t)_{t=0}^{\infty}$ with $P(0^\infty) = 0$ the expectation of the first-recurrence time n_1 under the condition $X_0 = 1$ is

$$\langle n_1 \mid X_0 = 1 \rangle = q^{-1} \tag{1}$$

Proof. Applying the identity (vi) with $y = 1$ and $r = 1$ and using (iii) and (iv) and the assumption $P(0^\infty) = 0$, one obtains

$$\begin{aligned}
\langle n_1 \mid X_0 = 1 \rangle &= q^{-1} \sum_{m=1}^{\infty} m P(10^{m-1}1) \\
&= q^{-1} \sum_{m=0}^{\infty} P(0^m 1) \\
&= q^{-1}[1 - P(0^\infty)] = q^{-1}
\end{aligned}$$

This proof, though based on the same consistency relations as Kac's proof, differs from the latter in that no explicit discussion of the asymptotic behavior for $m \to \infty$ of $m[P(0^m) - P(0^{m+1})]$ is required (see Ref. 1; cf. also Ref. 2).

VARIATION I

For an arbitrary stationary 0–1 process

$$\langle n_k \mid X_0 = 1 \rangle = q^{-1}[1 - P(0^\infty)] \qquad \text{for all } k \geqslant 1 \tag{2}$$

For $k=1$ this variation is nearly trivial, the proof being identical to that of the theme, except for the last step, which is not valid if $P(0^\infty) \neq 0$. Kac's theorem was presented in this form (in a broader context) by Blum and Rosenblatt.[2] It was later encountered again in an analysis of random walks on stochastically black–white colored lattices.[3]

The equality of all $\langle n_k \mid X_0 = 1 \rangle$, $k \geqslant 1$ (and also that of all higher moments of the n_k) follows from the fact that under the given condition the n_k, $k \geqslant 1$, are identically distributed. To prove this fact, we apply (vii), with $m = n_1$, $y = 10^{n_2 - 1}10^{n_3 - 1}1 \cdots 10^{n_k - 1}1$ $(k \geqslant 2)$, and (iii):

$$\begin{aligned}
P(n_k > n \mid X_0 = 1) &= q^{-1} \sum_{n_1, \ldots, n_{k-1} = 1}^{\infty} \; \sum_{n_k = n+1}^{\infty} P(10^{n_1 - 1} y) \\
&= q^{-1} \sum_{n_2, \ldots, n_{k-1} = 1}^{\infty} \; \sum_{n_k = n+1}^{\infty} [P(y) - \lim_{N \to \infty} P(0^N y)] \\
&= q^{-1} \sum_{n_1, \ldots, n_{k-2} = 1}^{\infty} \; \sum_{n_{k-1} = n+1}^{\infty} P(10^{n_1 - 1}1 \cdots 10^{n_{k-1} - 1}1) \\
&\qquad\qquad [\text{by relabeling}] \\
&= P(n_{k-1} > n \mid X_0 = 1) \qquad \text{for any } n \geqslant 0
\end{aligned} \tag{3}$$

Iteration of (3) yields the desired property. Note that the n_k need not be mutually independent. An extension of the argument shows, however, that the process $(n_k)_{k=1}^{\infty}$ conditioned on $X_0 = 1$ is stationary. Another relation, to be used in Variation XIV, is also easily derived:

$$P(n_1 > n \mid X_0 = 1) = q^{-1} P(n_0 = n)$$

If the condition $X_0 = 1$ is relaxed to $X \neq 0^\infty$ (i.e., there is at least one t such that $X_t = 1$), the mean recurrence times need not be equal to q^{-1} any more, nor are they independent of k in general. Examples show that the $\langle n_k \rangle$ may vary with k in many different ways. The next variation shows, however, that there are bounds to their behavior. Instead of conditioning on $X \neq 0^\infty$, we assume $P(0^\infty) = 0$. By (ii), the n_k are again well-defined random variables.

VARIATION II

For a stationary 0–1 process with $P(0^\infty) = 0$

$$|\langle n_k \rangle - q^{-1}| \leqslant 1 + 2\langle n_0 \rangle - q^{-1} \tag{4}$$

To prove Eq. (4), we first derive the following identities [cf. Ref. 3, Eqs. (2.17) and (2.18)]

$$\langle n_1^2 \mid X_0 = 1\rangle = q^{-1}(1 + 2\langle n_0\rangle) \tag{5}$$

$$\langle n_1 n_{k+1} \mid X_0 = 1\rangle = q^{-1}\langle n_k\rangle, \qquad k \geqslant 1 \tag{6}$$

Applying (vi), with $y = 1$, $r = 2$, and using (iii) and (iv), one finds

$$\begin{aligned}\langle n_1^2 \mid X_0 = 1\rangle &= q^{-1} \sum_{m=1}^{\infty} m^2 P(10^{m-1}1) \\ &= q^{-1} \sum_{m=0}^{\infty} (1 + 2m)\, P(0^m 1) \\ &= q^{-1}(1 + 2\langle n_0\rangle)\end{aligned}$$

The proof of (6) is similar. We next use the fact that $\mathrm{Var}(n_1 + n_{k+1} \mid X_0 = 1) \geqslant 0$ and $\mathrm{Var}(n_1 - n_{k+1} \mid X_0 = 1) \geqslant 0$. Using (1), (3), (5), and (6), we find

$$\begin{aligned}0 \leqslant{}& \langle n_1^2 \mid X_0 = 1\rangle + \langle n_{k+1}^2 \mid X_0 = 1\rangle \\ &\pm 2\langle n_1 n_{k+1} \mid X_0 = 1\rangle - \langle n_1 \pm n_{k+1} \mid X_0 = 1\rangle^2 \\ ={}& q^{-1}[2(1 + 2\langle n_0\rangle) \pm 2\langle n_k\rangle] - (q^{-1} \pm q^{-1})^2\end{aligned}$$

or

$$-1 - 2\langle n_0\rangle + q^{-1} \leqslant \langle n_k\rangle - q^{-1} \leqslant 1 + 2\langle n_0\rangle - q^{-1} \tag{7}$$

Let us consider the case $k = 1$ and ask whether the lower and upper bounds for $\langle n_1\rangle$ can be attained; for arbitrary k the argument is similar, but lengthy. The left-hand inequality reduces to an equality iff $\mathrm{Var}(n_1 + n_2 \mid X_0 = 1) = 0$, i.e., iff there is $j \in \mathbb{N}$ such that $n_1 + n_2 = j$ a.s. It is easily seen that this implies that P is cyclic with repeating unit $0^{i-1}10^{j-i-1}1$, with $1 \leqslant i \leqslant j - 1$, or a convex combination of such measures with different i and the same j. This implies $q = 2j^{-1}$; therefore equality cannot hold for other values of q.

The right-hand inequality (which was first derived in Ref. 3) reduces to an equality iff $\mathrm{Var}(n_1 - n_2 \mid X_0 = 1) = 0$, i.e., iff there is $j \in \mathbb{Z}$ such that $n_1 - n_2 = j$ almost surely. Since n_1 and n_2 have identical distributions, j must be zero, so that, with probability one, n_1 and n_2 (and hence all n_k, $k \geqslant 1$) are equal. This means that P is cyclic with repeating unit $0^{n-1}1$, $n \geqslant 1$, or a convex combination of such measures with different n. Since any value of q in $(0, 1]$ can always be written as a convex combination of numbers of the form n^{-1}, the upper bound for $\langle n_1\rangle$ can be attained for all q.

Observe that the lower bound in (7) is larger than the trivial lower bound 1 only if $-1-2\langle n_0\rangle+2q^{-1}>1$, i.e., if $\langle n_0\rangle<q^{-1}-1$.

The inequality (4) leads us to ask what can be said about $\langle n_0\rangle$, the mean first-arrival time for the state 1. A first answer to this question is given in the next variation.

VARIATION III

For a stationary 0–1 process with $P(0^\infty)=0$

$$\langle n_0\rangle \geqslant (1-q)/2q \tag{8}$$

The equality sign holds iff $q=j^{-1}$ $(j\in\mathbb{N})$ and P is cyclic with repeating unit $0^{j-1}1$.

This inequality follows from $\mathrm{Var}(n_1 \mid X_0=1)\geqslant 0$ together with (1) and (5) (cf. Ref. 3). The equality sign holds iff there is $j\in\mathbb{N}$ such that $n_1=j$ a.s., which implies that P is cyclic with repeating unit $0^{j-1}1$ and, hence, $q=j^{-1}$.

For the mean first-recurrence time $\langle n_1\rangle$ there is also a lower bound in terms of q:

VARIATION IV

For a stationary 0–1 process with $P(0^\infty)=0$

$$\langle n_1\rangle \geqslant 2-q \tag{9}$$

The equality sign holds iff $P(010)=0$.

Indeed, we have

$$\begin{aligned}\langle n_1\rangle &= P(1)\langle n_1 \mid X_0=1\rangle + P(0)\langle n_1 \mid X_0=0\rangle\\ &= q\cdot q^{-1}+P(0)\langle n_1 \mid X_0=0\rangle\\ &\geqslant 1+P(0)=2-q\end{aligned}$$

since $n_1\geqslant 1$. For the equality sign to hold, $\langle n_1 \mid X_0=0\rangle$ must be equal to 1. This is true iff the first 1 following a sequence of one or more 0's is almost surely followed by another 1, i.e., iff $P(010)=0$.

The inequality (9) holds also for $\langle n_k\rangle$, $k>1$, but the condition for equality is more complicated.

Let us return to n_0. Observe that it is a first-arrival time, not a recurrence time, and therefore cannot be expected to behave similarly to the n_k, $k \geqslant 1$. In order to derive a lower bound for $\langle n_0 \rangle$ in terms of q which, like the one found for $\langle n_1 \rangle$, can be attained for *any* value of q, we will need the following identity.

VARIATION V

For a stationary 0–1 process with $P(0^\infty)=0$

$$\langle n_0 \rangle = \sum_{m=1}^{\infty} P(0^m) \tag{10}$$

The identity holds both if the sum on the right-hand side is finite and if it is infinite.

The proof is based on (v), with $y=0$, $r=1$:

$$\langle n_0 \rangle = \sum_{m=1(0)}^{\infty} mP(0^m 1) = \sum_{m=0}^{\infty} [P(0^{m+1}) - P(0^\infty)] = \sum_{m=1}^{\infty} P(0^m)$$

Equation (10) is essentially equivalent to a classical identity (derived along other lines in, e.g., Ref. 12, Vol. I, Section XI, 1). The announced lower bound for $\langle n_0 \rangle$ is given in the next variation.

VARIATION VI

For a stationary 0–1 process with $P(0^\infty)=0$

$$\langle n_0 \rangle \geqslant j - \tfrac{1}{2} j(j+1)\, q \tag{11}$$

where j is the largest integer $\leqslant q^{-1}$. The equality sign holds iff

$$P(10^{j-1}1) = (j+1)\, q - 1, \qquad P(10^j 1) = 1 - jq$$
$$P(10^m 1) = 0 \qquad \text{for} \quad m \neq j-1, j$$

To prove this, observe that

$$P(0^m) = P(0^{m-1}) - P(0^{m-1}1) \geqslant P(0^{m-1}) - P(1)$$

and hence, by iteration, $P(0^m) \geqslant 1 - mq$. Therefore, by Eq. (10),

$$\langle n_0 \rangle \geqslant \sum_{m=1}^{M} P(0^m) \geqslant \sum_{m=1}^{M} (1 - mq) \tag{12}$$

for any M. If $j \leqslant q^{-1} < j+1$, the terms in the second sum are nonnegative iff $m \leqslant j$, so that the largest lower bound for $\langle n_0 \rangle$ of the type (12) is obtained by choosing $M = j$, which yields (11).

For the equality sign to hold, it is necessary and sufficient that $P(0^m)=0$ for $m>j$, $P(0^m)=1-mq$ for $m\leqslant j$. Since

$$P(10^m1)=P(0^m1)-P(0^{m+1}1)=P(0^m)-2P(0^{m+1})+P(0^{m+2})$$

this condition on the $P(0^m)$ is equivalent to the one stated in the variation, which is formulated in terms of the $P(10^m1)$. The latter is more transparent: it says that a 1 can only be followed by $0^{j-1}1$ or by 0^j1, with the prescribed probabilities. It is evident that a measure satisfying this requirement exists.

Note that inequality (11) is stronger than (8) because $(1-jq)[1-(j+1)q]\leqslant 0$; it coincides with (8) iff $q=j^{-1}$. Note also that the lower bound given by (11) is piecewise linear and continuous in q.

The inequality (11) can be generalized to one in terms of the probabilities of sequences of length $\leqslant l$, $l\geqslant 1$:

VARIATION VII

For a stationary 0–1 process with $P(0^\infty)=0$ and for $l\geqslant 1$

$$\langle n_0\rangle \geqslant \sum_{m=1}^{l-1} P(0^m)-\binom{j}{2}P(0^{l-1})+\binom{j+1}{2}P(0^l) \tag{13}$$

where j is the largest integer $\leqslant P(0^{l-1})/P(0^{l-1}1)$. Equality obtains iff

$$P(10^{j+l-2}1)=jP(0^{l-1})-(j+1)P(0^l)$$

$$P(10^{j+l-1}1)=jP(0^l)-(j-1)P(0^{l-1})$$

$$P(10^m1)=0 \qquad \text{for} \quad l-1\leqslant m\leqslant j+l-3 \qquad \text{and} \qquad m\geqslant j+l$$

The proof is a generalization of that of the previous variation. From $P(0^m)\geqslant P(0^{m-1})-P(0^{l-1}1)$, valid for $1\leqslant l\leqslant m$, it follows by iteration that

$$P(0^m)\geqslant P(0^{l-1})-(m-l+1)P(0^{l-1}1)$$

and, hence,

$$\langle n_0\rangle \geqslant \sum_{m=1}^{l-1} P(0^m)+\sum_{m=l}^{M}[P(0^{l-1})-(m-l+1)P(0^{l-1}1)] \tag{14}$$

for any M. If $j\leqslant P(0^{l-1})/P(0^{l-1}1)<j+1$, the terms in brackets are nonnegative iff $m\leqslant j+l-1$, so that the largest lower bound for $\langle n_0\rangle$ of the type (14) is obtained by choosing $M=j+l-1$, which yields (13).

The derivation of the condition for equality is analogous to the one given in the proof of the previous variation.

The lower bound for $\langle n_0 \rangle$ given in (13) [which reduces to (11) for $l=1$] is nondecreasing in l, since the lower bound for $P(0^m)$ is nondecreasing in l. Indeed, subtracting the bounds for $l+1$ and for l gives, after a little algebra,

$$\begin{aligned}&[P(0^l)-(m-l)\,P(0^l1)]-[P(0^{l-1})-(m-l+1)\,P(0^{l-1}1)]\\&\quad=(m-l)\,P(10^{l-1}1)\geqslant 0\end{aligned}$$

The lower bound at "level l" represents the lowest value that $\langle n_0 \rangle$ can take if P is allowed to vary in the set of measures with fixed values of the l-point probabilities $P(x_0 x_1 \cdots x_{l-1})$. This makes the monotonicity in l of the bound even obvious.

Identity (10) guarantees that the bounds converge to the exact value of $\langle n_0 \rangle$ as $l \to \infty$.

The question of whether there is not only a lower, but also an upper bound for $\langle n_0 \rangle$ within the set of measures with fixed values of the $P(x_0 x_1 \cdots x_{l-1})$ for given l must be answered in the negative: at each level l one can construct measures with arbitrarily large (even infinite) values of $\langle n_0 \rangle$. The reason is that fixing $P(0^m)$ for $1 \leqslant m \leqslant l$ does not give sufficient information about the way in which $P(0^m)$ behaves for $m > l$. In particular, it does not exclude that $\sum_m P(0^m)$ diverges; examples where $\langle n_0 \rangle = \infty$ are easily constructed.

For more restricted classes of 0–1 processes, however, upper bounds at a given level do exist. For 1-dependent processes, e.g. [characterized by the property $P(y0y') + P(y1y') = P(y)\,P(y')$ for any y and y'] an upper bound is readily found. From $P(y0y') \leqslant P(y)\,P(y')$ with $y = 0^m$, $y' = 0$ one derives by iteration $P(0^{2s-1}) \leqslant P(0)^s$, $P(0^{2s}) \leqslant P(0)^s$ for $s \geqslant 1$, and hence, by Eq. (10),

$$\langle n_0 \rangle \leqslant 2P(0)/P(1) = 2(1-q)\,q^{-1}$$

A more detailed analysis shows, however, that for all q there is a sharper upper bound, which can be attained for $q \leqslant \frac{1}{2}$.[4]

Let us now consider the mean recurrence time $\langle n_1 \rangle$ again. The analogue for $\langle n_1 \rangle$ of the identity (10) is given in the following variation:

VARIATION VIII

For a stationary 0–1 process with $P(0^\infty) = 0$

$$\langle n_1 \rangle = 1 + P(0) + \sum_{m,n=1}^{\infty} P(0^m 1 0^n) \tag{15}$$

The identity holds both if the sum on the right-hand side is finite and if it is infinite.

The proof is again based on (v), this time with $y = 0^m 1$, $r = 1$:

$$\begin{aligned}\langle n_1 \rangle &= P(1)\langle n_1 \mid X_0 = 1\rangle + P(0)\langle n_1 \mid X_0 = 0\rangle \\ &= 1 + \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} nP(0^m 1 0^{n-1} 1) \\ &= 1 + \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} [P(0^m 1 0^n) - P(0^m 1 0^\infty)] \\ &= 1 + \sum_{m=1}^{\infty} P(0^m 1) + \sum_{m,n=1}^{\infty} P(0^m 1 0^n)\end{aligned}$$

from which Eq. (15) follows because $\sum_{m=1}^{\infty} P(0^m 1) = P(0)$.

The next, somewhat baroque, variation constitutes a sharpening of the inequality (9) for $\langle n_1 \rangle$ in terms of the probabilities of sequences of length $\leqslant l$.

VARIATION IX

For a stationary 0–1 process with $P(0^\infty) = 0$ and for $l \geqslant 1$

$$\begin{aligned}\langle n_1 \rangle \geqslant{} & 1 + P(0) + \sum_{\substack{m,n \geqslant 1 \\ m+n \leqslant l-1}} P(0^m 1 0^n) \\ & + \max_{J \subset I_l} \sum_{(m,n) \in J} (a_{m,n;l} - a_{m-1,n-1;l-1})\end{aligned} \tag{16}$$

where

$$I_l := \{(m, n) \colon 1 \leqslant m, n \leqslant l-1;\ l \leqslant m+n \leqslant 2l-2\}$$

$$a_{m,n;l} := \sum_{\substack{i,j \geqslant 0 \\ i+j = m+n+1-l}} P(0^{m-i} 1 0^{n-j})$$

The derivation starts from the fact that for $m, n \geqslant 1$

$$\begin{aligned}P(0^m 1 0^n) &= P(0^{m-1} 1 0^n) - P(1 0^{m-1} 1 0^{n-1}) + P(1 0^{m-1} 1 0^{n-1} 1) \\ &\geqslant P(0^{m-1} 1 0^n) - P(1 0^{m-1} 1 0^{n-1}) \\ &= P(0^{m-1} 1 0^n) + P(0^m 1 0^{n-1}) - P(0^{m-1} 1 0^{n-1})\end{aligned}$$

and hence, by iteration, for $\max\{m, n\} + 1 \leqslant l \leqslant m+n$

$$P(0^m 1 0^n) \geqslant a_{m,n;l} - a_{m-1,n-1;l-1} \tag{17}$$

with $a_{m,n;l}$ as defined above. Choosing a "level" l and $J \subset I_l$, applying (17) to all terms in (15) with $(m, n) \in J$, omitting all other terms in (15) with $m+n \geqslant l$, and maximizing with respect to J yields (16).

In contrast with the situation encountered in the proof of Variation VII, the condition for the right-hand side of (17) to be nonnegative is not of a form that admits an immediate answer to the question of how to characterize the pairs (m, n) for which this lower bound to $P(0^m 1 0^n)$ should be included in (16) in order to obtain the best lower bound to $\langle n_1 \rangle$ for a given level l and given l-point probabilities. Therefore, the problem of finding the maximizing J in (16) will not be addressed.

For $\langle n_1 \rangle$, as for $\langle n_0 \rangle$, there is no upper bound within the set of measures with fixed values of the $P(x_0 x_1 \cdots x_{l-1})$ for given l; the only general upper bound known is the one given in Eq. (7).

The next variation deals with the *variance* of n_1 under the condition $X_0 = 1$.

VARIATION X

For a stationary 0–1 process with $P(0^\infty) = 0$

$$\operatorname{Var}(n_1 \mid X_0 = 1) \geqslant a(1-a) \tag{18}$$

where a is the nonintegral part of q^{-1}.

Equation (18) follows simply from (1), (5), and (11). The necessary and sufficient condition for equality is the same as in Variation VI. Similarly, a lower bound for $\operatorname{Var}(n_1 \mid X_0 = 1)$ at an arbitrary level l is obtained from (13). There is, of course, also an analogue of the identity (10):

VARIATION XI

For a stationary 0–1 process with $P(0^\infty) = 0$

$$\operatorname{Var}(n_1 \mid X_0 = 1) = q^{-1} - q^{-2} + 2q^{-1} \sum_{m=1}^{\infty} P(0^m) \tag{19}$$

This follows directly from (1), (5), and (10). Equation (19) was first derived by Blum and Rosenblatt,[2] who also gave a formal expression for $\langle n_1^r \mid X_0 = 1 \rangle$ for general $r \geqslant 2$ (cf. also Ref. 5).

The bounds on $\langle n_0 \rangle$, $\langle n_1 \rangle$, and $\operatorname{Var}(n_1 \mid X_0 = 1)$ derived so far can, of course, be sharpened if more information on the 0–1 process is available, e.g., if the process is known to be a Markov process, a renewal process, or a 1-dependent process. The next variation gives an example of such sharper bounds. It deals with processes of which the probability measure is log convex with respect to the ordering of 0–1 sequences defined as follows: for two 0–1 sequences y, y' of the same length we put $y > y'$ iff one or more entries of y are larger than the corresponding entries of y' and the other entries (if any) of y and y' are equal (e.g., $1 > 0$, $11 > 01 > 00$, $11 > 10 > 00$). The measure P is called *log convex* if $P(y \vee y')\, P(y \wedge y') \geqslant P(y)\, P(y')$ for all y, y' of the same length, where $y \vee y'$ ($y \wedge y'$) is the least upper bound (greatest lower bound) of y and y'.

VARIATION XII

For a stationary 0–1 process with a log convex probability measure and $P(0^\infty) = 0$

$$\langle n_0 \rangle \geqslant q^{-1} - 1 \tag{20}$$

$$\langle n_1 \rangle \geqslant q^{-1} \tag{21}$$

In both equations the equality sign holds iff the process is Bernoulli [i.e., iff $P(x_0 x_1 \cdots x_n) = P(x_0)\, P(x_1) \cdots P(x_n)$ for all n].

To prove (20) and (21), observe that the definition of log convexity implies

$$\begin{aligned} &P(0y0)\, P(y) - P(0y)\, P(y0) \\ &\quad = P(0y0)[P(0y0) + P(0y1) + P(1y0) + P(1y1)] \\ &\qquad - [P(0y0) + P(0y1)][P(0y0) + P(1y0)] \\ &\quad = P(0y0)\, P(1y1) - P(0y1)\, P(1y0) \geqslant 0 \end{aligned}$$

Choosing $y = 0^{m-2}$ $(m \geqslant 2)$ and $y = 0^{m-1} 10^{n-1}$ $(m, n \geqslant 1)$, one finds, respectively,

$$P(0^m) \geqslant P(0^{m-1})^2 / P(0^{m-2})$$

$$P(0^m 10^n) \geqslant P(0^m 10^{n-1})\, P(0^{m-1} 10^n) / P(0^{m-1} 10^{n-1})$$

Iteration yields

$$\begin{aligned} P(0^m) &\geqslant P(0^{m-1})\, P(0) \\ &\geqslant \cdots \geqslant P(0)^m \end{aligned}$$

$$\begin{aligned} P(0^m 10^n) &\geqslant P(0^m 10^{n-1})\, P(10^n) / P(10^{n-1}) \\ &\geqslant \cdots \geqslant P(0^m 1)\, P(10^n) / P(1) \end{aligned}$$

which, inserted into (10) and (15), yields (20) and (21). The condition for equality follows trivially.

Equation (21) can also be derived in another way, namely by applying the FKG inequality,[6] generalized to countable sets. Consider the unique extension of the process X to a process on $\mathbb{Z}$, $\bar{X} := (X_t)_{t\in\mathbb{Z}}$, and define $\bar{n}_0 := -\max\{t<0\colon X_t = 1\}$ (not to be confused with n_0, which equals zero if $X_0 = 1$), $\bar{n}_k := n_k$ $(k \geqslant 1)$. Then, in an obvious notation,

$$\langle n_1 n_2 \mid X_0 = 1\rangle_X = \langle \bar{n}_1 \bar{n}_2 \mid X_0 = 1\rangle_{\bar{X}} = \langle \bar{n}_0 \bar{n}_1 \mid X_0 = 1\rangle_{\bar{X}}$$

by the stationarity of the process $(\bar{n}_k)_{k\in\mathbb{Z}}$ under the condition $X_0 = 1$, which is easily established. Since both $\bar{n}_0$ and $\bar{n}_1$ are nonincreasing with respect to the ordering of sequences, they are positively correlated. Hence,

$$\langle n_1 n_2 \mid X_0 = 1\rangle \geqslant \langle n_1 \mid X_0 = 1\rangle \langle n_2 \mid X_0 = 1\rangle = q^{-2}$$

from which (21) follows in virtue of (6).

Statements of a different kind about the recurrence times n_k can be made if the process X is not only stationary, but *ergodic*, i.e., if P is not a convex combination $\lambda P_1 + (1-\lambda) P_2$, with $0<\lambda<1$, of two other stationary probability distributions. They are contained in the next variation (the first part of which is mentioned in Ref. 7).

VARIATION XIII

For an ergodic 0–1 process

$$\lim_{k\to\infty} k^{-1} \sum_{j=1}^{k} n_j = q^{-1} \qquad \text{with probability 1} \tag{22}$$

$$\lim_{k\to\infty} k^{-1} \sum_{j=1}^{k} \langle n_j\rangle = q^{-1} \qquad \text{if} \quad \langle n_0\rangle < \infty \tag{23}$$

Equation (22) is derived by applying Birkhoff's ergodic theorem to the random variable X_0:

$$\lim_{t\to\infty} (t+1)^{-1} \sum_{n=0}^{t} X_n = \langle X_0\rangle = q \quad \text{a.s.}$$

Taking the subsequence formed by the time averages up to those t-values for which $X_t = 1$, i.e., the times $t = T_k$, $k \geqslant 1$ (which are defined with probability 1), one obtains

$$\lim_{k\to\infty} \frac{k}{T_k} = q \quad \text{a.s.}$$

which is equivalent to (22) because $\sum_{j=1}^{k} n_j = T_{k+1} - T_1$.

By Eq. (7), all $\langle n_k \rangle$ are uniformly bounded if $\langle n_0 \rangle < \infty$. Equation (23) then follows from (22) by the dominated convergence theorem.

Equation (23) may be said to express the asymptotic behavior of $\langle n_k \rangle$ in the limit $k \to \infty$, not in the strict sense, but in the weaker sense of Cesàro. A much stronger statement, namely on the asymptotic behavior in the strict sense of the probability distribution for n_k, can be made if P is not only ergodic, but if its extension $\bar{P}$ describing the process $\bar{X} = (X_t)_{t \in \mathbb{Z}}$ is an ergodic (=extremal) *Gibbs state.*

VARIATION XIV

For a stationary 0–1 process with a probability measure P of which the extension $\bar{P}$ is an ergodic Gibbs state

$$\lim_{k \to \infty} P(n_k > n) = P(n_1 > n \mid X_0 = 1) = q^{-1} P(n_0 = n) \qquad \text{for all} \quad n \geqslant 0 \tag{24}$$

$$\lim_{k \to \infty} \langle n_k \rangle = q^{-1} \qquad \text{if} \quad \langle n_0 \rangle < \infty \tag{25}$$

Equation (24) is a corollary of a theorem on random walks on a stochastically black and white colored lattice that has quite recently been established by den Hollander.[8] This theorem gives conditions under which, loosely speaking, the arrangement of black and white points "perceived" by the walker in an arbitrary finite region of fixed form around him at the time of his kth visit to a black point will in the limit $k \to \infty$ become asymptotically independent of the color arrangement in any finite neighborhood of his starting point. For a precise formulation and the proof of the theorem see den Hollander's paper. The application to 0–1 processes is obtained by choosing a one-dimensional lattice and a deterministic random walk [$p(1) = 1$]. Equation (25) follows from (24) (see Ref. 9, p. 192).

The final two variations, the first of which is due to Kac himself (see also Ref. 10), are in a different key. They refer to a dynamical system $(\Omega, \mathscr{B}, \mu, T)$, i.e., a probability space $(\Omega, \mathscr{B}, \mu)$ together with an automorphism T of this space (i.e., an invertible measure-preserving mapping of Ω onto itself).

VARIATION XV[1]

Let $(\Omega, \mathscr{B}, \mu, T)$ be a dynamical system, $A \subset \Omega$ a set of positive measure, and, for $\omega \in A$, $n_{1A}(\omega) = \min\{t \geqslant 1 \colon T^t \omega \in A\}$ the Poincaré cycle of ω. If T is ergodic, then

$$\langle n_{1A} \mid A \rangle := \frac{\int_A d\mu \, n_{1A}(\omega)}{\int_A d\mu} = \mu(A)^{-1} \tag{26}$$

The proof is essentially Kac's. Associate with $(\Omega, \mathscr{B}, \mu, T)$ a 0–1 process $(X_t)_{t=0}^{\infty}$ by putting $X_t(\omega) = 1$ iff $T^t\omega \in A$. Since T is measure-preserving, the process is stationary. Further, $q = P(1) = \mu(A)$ and $P(0^\infty) = \mu(B)$, where

$$B \equiv B_A := \{\omega \in \Omega\colon T^t\omega \notin A \text{ for all } t \geqslant 0\}$$

To show that $\mu(B) = 0$, consider the sets

$$T^{-n}B = \{\omega\colon T^t\omega \notin A \text{ for all } t \geqslant n\}$$

which form a nondecreasing sequence with $\mu(T^{-n}B) = \mu(B)$. Let $C := \bigcup_{n=0}^{\infty} T^{-n}B$. Clearly, $TC = C$ and hence, by the ergodicity of T, $\mu(C) = 0$ or 1. Now

$$\mu(C) = \mu(B) \leqslant \mu(\Omega \backslash A) = 1 - \mu(A) < 1$$

Hence $\mu(B) = 0$. Application of (1) gives (26).

VARIATION XVI

Let $(\Omega, \mathscr{B}, \mu, T)$ be a dynamical system, with T an arbitrary automorphism. Then, with A, B_A, and n_{1A} as defined before,

$$\langle n_{1A} \mid A \rangle = [1 - \mu(B_A)]/\mu(A) \tag{27}$$

Equation (27) is a direct application of Variation I. The translation of the other variations to dynamical systems is equally straightforward.

CODA

For an *arbitrary* stationary 0–1 process [i.e., one where $P(0^\infty)$ need not be zero] the inequalities and identities appearing in Variations II–XII remain valid if the probability measure $P(\cdot)$ is replaced by the conditional probability $P(\cdot \mid X \neq 0^\infty)$ and the unconditional expectations $\langle n_k \rangle$, $k \geqslant 0$, by $\langle n_k \mid X \neq 0^\infty \rangle$.

This statement is easily verified by inspection of the proofs of the variations. Note that

$$P(0^m \mid X \neq 0^\infty) = \frac{P(0^m) - P(0^\infty)}{1 - P(0^\infty)}$$

$$P(y \mid X \neq 0^\infty) = \frac{P(y)}{1 - P(0^\infty)} \qquad \text{if} \quad y \text{ contains a } 1$$

so that q is to be replaced by $q/[1 - P(0^\infty)]$. The factor $1 - P(0^\infty)$ is, of course, the probability that at least one 1 will appear; if this happens, then with probability one infinitely many 1's will appear, so that all n_k are proper random variables. Variation I contains another example of a generalization to arbitrary stationary P.

In connection with the question of whether or not $P(0^\infty) = 0$, the following remark may be in place. It is clear that a sufficient (though not necessary) condition for $P(0^\infty) = 0$ is that the process X be ergodic. Indeed, the event $X = 0^\infty$ is invariant under the *time shift* $\mathscr{T}$, defined by $(\mathscr{T}X)_t = X_{t+1}$, and for an ergodic process all sets of events that are invariant under $\mathscr{T}$ have probability zero or one; the possibility $P(0^\infty) = 1$ is excluded by the assumption $q > 0$.

An example of an ergodic 0–1 process is found in the class of processes the analysis of which gave rise to the work reported on above and which has been alluded to already several times. Let there be given an infinite d-dimensional lattice L, a translation-invariant probability distribution on the set of all black and white colorings of the sites of L, and a random walk on L starting at the origin at time zero and moving independently of the coloring. Then the process $(X_t)_{t=0}^\infty$, defined by $X_t = 0$ (1) if the lattice site occupied by the walker at time t is white (black), is a stationary 0–1 process. It can be shown (a hint of a proof is given in Ref. 11) that X is ergodic if the coloring probability distribution is ergodic with respect to lattice translations ("extremal translation invariant" in the terminology of Ref. 11) and the random walk aperiodic ("spanning").

REFERENCES

1. M. Kac, *Bull. Am. Math. Soc.* **53**:1002 (1947); cf. also M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience, London, 1959), Chapter III, Sections 4–6.
2. J. R. Blum and J. I. Rosenblatt, *J. Math. Sci. (Delhi)* **2**:1 (1967).
3. W. Th. F. den Hollander and P. W. Kasteleyn, *Physica* **117A**:179 (1983).
4. A. Gandolfi, private communication.
5. J. Wolfowitz, *Proc. Am. Math. Soc.* **18**:613 (1967).
6. C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, *Commun. Math. Phys.* **22**:89 (1971).
7. L. Breiman, *Probability* (Addison-Wesley, Reading, Massachusetts, 1968), Chapter 6.
8. W. Th. F. den Hollander, *Mixing Properties for Random Walk in Random Scenery*, to be published.
9. M. Keane and W. Th. F. den Hollander, *Physica* **138A**:183 (1986).
10. J. Wolfowitz, *Bull. Am. Math. Soc.* **55**:394 (1949).
11. P. W. Kasteleyn, *Bull. ISI* **45**:27.I (1985).
12. W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I (Wiley, New York, 1968).