ISRAEL JOURNAL OF MATHEMATICS, Vol. 30, No. 3, 1978

GUESSING THE NEXT OUTPUT OF A STATIONARY PROCESS

BY

DONALD S. ORNSTEIN[†]

ABSTRACT

Suppose we start watching a stationary process at time 0. Then the conditional probability of a particular output at time -1, given the outputs at times 0 through k, will converge. In this paper we will show that we can make a guess, depending only on the outputs from 0 through k (and not, of course, on the process) that will converge to the above limit with probability one.

Introduction

In this paper we will consider stationary processes that print out two symbols 0, or 1.[††] (By a stationary process we will mean a shift invariant measure on doubly infinite sequences of 0, 1 or equivalently a measure preserving transformation and a partition into two sets.) We will assume the process is ergodic.[‡]

We will be concerned with the conditional probability that the process will print out a, 0 at time 1 given its past. This makes sense as follows: Let a_i be the printout at time i. Let $G(a_0, a_{-1}, \cdots, a_{-k})$ denote the conditional probability of $a_1 = 0$ given that the printout at times 0 through $-k$ were $(a_0, \cdots, a_{-k})$. The Martingale convergence theorem says that with probability one $G(a_0, \cdots, a_{-k})$ converges as $k \to \infty$.

If we know the infinite past we can find (with probability one) each $G(a_0, \cdots, a_{-k})$ because the ergodic theorem (or law of large numbers) says that the probability of any finite string is equal (with probability one) to the frequency with which it will occur in the past.

[†]Research supported in part by NSF MCS 76–09159

[††] (Our results hold with no added difficulties for stationary processes that print out a finite or countable number of symbols.)

[‡] This is no real restriction because every stationary process decomposes into ergodic ones, i.e. all non ergodic processes are obtained by randomly picking an ergodic process.

Received June 16, 1977

This means that if we are given the infinite past of a stationary process we can, without any further information about the process, determine the probability that the next printout will be a 0.

Thomas Cover pointed out that we can ask a more subtle question: is there a guessing scheme such that given a sequence of 0, 1 of length k there is a guess, $f(a_0, \cdots, a_{-k})$ (f is some number between 0 and 1), with the property that, for any stationary process, if we apply f to the last k outputs and let k go to infinity, then with probability one (the probability is given by the process) $f(a_0, \cdots, a_{-k})$ will converge to the probability that $a_1 = 0$ given the past.

The purpose of this paper is to show that such a scheme exists.

We could rephrase our problem (reversing time) as follows: Suppose we start watching a stationary process at time 0. The conditional probability of a 0 at time -1 given the printout up to time k will converge and we want to make a guess, based only on the printout up to time k, that will converge to the same thing.

The problem is especially interesting in the case of deterministic processes, or processes of 0 entropy (i.e., the conditional probability of a 0 given the past is 0 or 1). In this case it is said that we can predict the future from the past, or the past from the future. It was not commonly realized until Cover pointed it out that that statement could have many meanings. A special case of the result of this paper shows that the above is true in a strong sense, that is: if we start watching a deterministic process at time 0 (we are given no information about the process) we can, at each finite time, make a guess as to the process printed out at time -1 before we started watching and with probability one our guesses will converge to the correct answer.

We will now construct a scheme for guessing the output at -1, given the outputs starting at 0.

We will start with some notation. Let $a_0, a_1, \cdots, a_k$ be a sequence of 0's or 1's ($a_i = 0$ or $a_i = 1$). Let $b_0, \cdots, b_l$, $l < k$ be another such sequence. We will say that $b_0, \cdots, b_l$ *occurs in* $a_0, \cdots, a_k$ *at* a_j if $a_{j+i} = b_i$, $0 \leqq i \leqq l$. *If* $a_{j-1} = 0$ *we will say that the occurrence is preceded by* 0. If we are *given a sequence* $a_0, \cdots, a_1, \cdots$ *define* $g(i,j)$, $i < j$ as follows: Let r denote the number of occurrences of $a_0, \cdots, a_i$ in $a_0, \cdots, a_j$, not counting the occurrence at a_0. Let r_0 denote the number of the above occurrences that are preceded by 0. *Let* $g(i,j) = r_0/r$ (the $g(i,j)$ are "reasonable" guesses for the conditional probability that $a_{-1} = 0$). We will say that $a_0, \cdots, a_k$ *is* (N, K, ε) *acceptable* if there is a subsequence $K = a_{n_0}, < a_{n_1} < \cdots < a_{n_N} = a_k$ such that

(1) $|g(i, n_j) - g(s, n_t)| < \varepsilon$ if $K \leqq i \leqq n_{j-1}$, $K \leqq s \leqq n_{t-1}$ and $1 \leqq j \leqq N$, $1 \leqq t \leqq N$.

(2) If $K \leqq i \leqq n_{j-1}$, $1 \leqq j \leqq N$, then the number of occurrences of $a_0, \cdots, a_i$ in $a_0, \cdots, a_{n_j}$ is $\geqq K$.

We will refer to $a_0, \cdots, a_k$ together with $K = n_0 < \cdots < n_N = k$ *as a partitioned sequence* (the partition tells us which "guesses" $g(i,j)$ to consider).

We can now describe our guessing scheme. Let $a_0, a_1, \cdots$ be the outcomes of our process. Pick a sequence $\varepsilon_i \to 0$. Wait until there is an l_1 such that for some N_1, $a_0, \cdots, a_{l_1}$ is $(N_1, N_1, \varepsilon_1)$ acceptable. Then predict $g(N_1, l_1)$. Keep making the same guess until there is an l_2 such that there is an N_2 and $a_0, \cdots, a_{l_2}$ is $(N_2, N_2, \varepsilon_2)$ etc.

Lemma 1. *Given $\varepsilon > 0$ and M then with probability 1 there will be an L such that if $l > L$ then $a_0, \cdots, a_l$ is (N, N, ε) acceptable for some $N > M$.*

Proof. The Martingale convergence theorem tells us that given $\varepsilon > 0$ there is an M such that if $j > M$ and $i > M$, then the conditional probability that $a_{-1} = 0$, given $a_0, \cdots, a_i$ or $a_0, \cdots, a_j$, differ by $< \frac{1}{2}\varepsilon$. The lemma now follows from the Birkhoff ergodic theorem (law of large numbers).

Define $G(a_0, \cdots, a_k)$ to be the conditional probability that a_{-1} is 0 given that the outputs of the process at times 0 through k are $a_0, \cdots, a_k$. We will say that $a_0, \cdots, a_l$ *is* (N, K, α) *bad* if there exists $K = n_0 < n_1 \cdots < n_N = l$, and $g(i, n_j) - G(a_0, \cdots, a_i) > \alpha$ for all $K \leqq i \leqq n_{j-1}$, $1 \leqq j \leqq N$ (or if all of the above numbers are $< -\alpha$).

Lemma 2. *Either our guessing scheme converges to the right guess or there is an $\alpha > 0$ and increasing sequences l_i, N_i such that $a_0, \cdots, a_{l_i}$ is (N_i, N_i, α) bad.*

Proof. If our guessing scheme does not work there will be an $\alpha > 0$ such that we are wrong by $> 2\alpha$ infinitely often. There is an M (Martingale convergence theorem) such that if $i > M$ and $j > M$ then

$$|G(a_0, \cdots, a_i) - G(a_0, \cdots, a_j)| < \frac{1}{100}\alpha.$$

If $a_0, \cdots, a_l$ is $(N, N, (1/100)\alpha)$ acceptable then all of the $g(i,j)$ that we used are within $(1/100)\alpha$ of each other. Therefore, if $N > M$, $\beta < (1/100)\alpha$ and $a_0, \cdots, a_l$ is (N, N, β) acceptable and gives a guess that is off by $> 2\alpha$ then $a_0, \cdots, a_l$ must be bad.

Lemma 3. *Given $\alpha > 0$ there is a $\beta < 1$ such that if $1/K < (1/100)\alpha$ then the probability that there will be an l and N such that $a_0, \cdots, a_l$ is (N, K, α) bad is $< \beta^N$.*

Lemma 3 implies easily that our guessing scheme will converge with probability one. The proof of Lemma 3 will depend on the next two lemmas.

The badness condition states the existence of a sequence and a partitioning. The next lemma gives us canonical sequences and partitionings. This makes it easier to compute probabilities.

LEMMA 4. *Assume* $a_0, \cdots, a_k$ *is* (N, K, α) *bad. Let* n_i, $i \leqq N$ *be the smallest integer such that* $a_0, \cdots, a_{n_i}$ *is* (i, K, α) *bad. Then* $a_0, \cdots, a_{n_N}$ *is* (N, K, α) *bad with respect to the partition* $K = n_0 < n_1 < \cdots < n_N = l$.

PROOF. We must check that

(a) $g(i, n_j) - G(a_0, \cdots, a_i) > \alpha$ and

(b) $a_0, \cdots, a_i$ occurs $> K$ times in $a_0, \cdots, a_{n_j}$, for all pairs n_j, $1 \leqq j \leqq N$ and $i \leqq n_{j-1}$. Consider one n_j. Because $a_0, \cdots, a_{n_j}$ is (j, K, α) bad there exists $K = \bar{n}_0 < \bar{n}_1 < \bar{n}_j = n_j$.

(a) and (b) must hold for n_j and all i less than $\bar{n}_{j-1}$. We will therefore be finished if we show $\bar{n}_{j-1} \geqq n_{j-1}$. However, $a_0, \cdots, a_{\bar{n}_{j-1}}$ is $(j-1, K, \alpha)$ bad and therefore $\bar{n}_{j-1} \geqq n_{j-1}$.

LEMMA 5. *Given* $\alpha > 0$ *there exists a* $\beta < 1$ *with the following property. Fix a sequence* $b_0, \cdots, b_n$. *Let* $p(b_0, \cdots, b_n)$ *denote the conditional probability of printing a sequence* $a_0, \cdots, a_l$, $l > n$ *given that* $a_i = b_i$, $0 \leqq i \leqq n$ *and such that* (1) $g(n, l) - G(a_0, \cdots, a_n) > \alpha$ *and* (2) $a_0, \cdots, a_n$ *occurs more than* $100/\alpha$ *times in* $a_0, \cdots, a_l$. *Then* $p(b_0, \cdots, b_n) < \beta$.

PROOF. Using the ergodic theorem we can pick an infinite sequence $c_0, c_1, \cdots$ with the property that for any finite sequence $d_1, \cdots, d_j$, the frequency of occurrences of $d_1, \cdots, d_j$ will equal the probability that the ith output is d_i, $1 \leqq i \leqq j$. We will therefore compute probabilities by computing frequencies of occurrences in c_0, c_1.

It will be convenient to prove our lemma with l restricted to be $<$ some M; and note that for all M we get the same β.

Look at all of the occurrences of $b_0, \cdots, b_n$. We will call an occurrence special if it can be extended to the occurrence of a sequence $a_0, \cdots, a_l$ satisfying (1), (2), and $l < M$. Becasse of (1) and (2) the occurrences of $b_0, \cdots, b_n$ in the occurrence of $a_0, \cdots, a_l$ (including the first) will form a collection of consecutive occurrences in which the percentage of occurrences preceded by 0 is too high. (We need (2) because we are including the first occurrence.) Call such a collection a special collection. We now take the next special occurrence of $b_0, \cdots, b_n$ that occurs to the right of the previous special collection and form the special collection starting with it. (This will be disjoint from our previous special collection.) If we proceed in this way, we will get a set, S, of disjoint, special collections, which include all of the special occurrences. Since the percentage of occurrences preceded by 0, in S, is greater than the percentage of occurrences preceded by 0 in the whole string, it follows that a certain percentage of occurrences will not be in S and will therefore not be special. (Note that this percentage can be taken to be independent of M.)

PROOF OF LEMMA 3. We will show that the probability of printing out an $(N+1, K, \alpha)$ bad sequence is $<\beta$ times the probability of an (N, K, α) bad sequence. Partition the infinite sequences, that have an (N, K, α) bad initial segment, according to their canonical (N, K, α) bad sequence $a_0, \cdots, a_{n_N}$ defined in Lemma 4. Call these classes C_i. In each C_i the probability (by Lemma 5) of extending $a_0, \cdots, a_{n_N}$ to some $a_0, \cdots, a_l$, where $a_0, \cdots, a_{n_N}$ occurs more than $100/\alpha$ times in $a_0, \cdots, a_l$ and $|G(a_0, \cdots, a_{n_N}) - g(n_N, l)| > \alpha$, is less than β. But by Lemma 4 if $a_0, \cdots, a_{n_N}$ extended to an $(N+1, K+\alpha)$ bad sequence it would have an extension of the above form.

DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305 USA