The Annals of Probability
1973, Vol. 1, No. 1, 43–65

SPECIAL INVITED PAPERS

AN APPLICATION OF ERGODIC THEORY TO PROBABILITY THEORY

BY DONALD S. ORNSTEIN

Stanford University

1. Introduction. The purpose of this paper is to try to introduce some structure in the class of stationary processes, and to take a step towards their classification. The main results of this paper are restatements of or minor modifications of some recent results in ergodic theory ([3], [5]–[10], [16], [17]).

The mathematical model for a stationary process[1]. A stationary process can be thought of as a box that prints out one letter each unit of time, where the probability of printing out a given letter may depend on the letters already printed out but is independent of the time (that is, the mechanism in the box does not change).

EXAMPLE 1. The box contains a roulette wheel. We spin the wheel once each unit of time, and print out the result. (We call such a process an independent process.)

EXAMPLE 2. The box contains a roulette wheel. We look at all possibilities for three consecutive spins of the wheel, and divide these into two classes. Each time we spin the wheel, we look at the last three spins and print out 1 if they fall in the first class, and 2 if they fall in the second class.

EXAMPLE 3. The box contains two coins, one of which is biased so that the probability of heads is not $\frac{1}{2}$. We divide all sequences of heads and tails of length three into two classes. At each unit of time, we look at the sequence of heads and tails which the box has printed out in the last three times. If the sequence lies in the first class, we flip the first coin and print out heads; if it comes up heads and tails, it comes up tails. If the previous three print-outs were in the second class, then we would use the second coin. This is an example of a three-step Markov process; that is, the last three print-outs determine the probability of printing out "heads," but if we know the last three print-outs, the conditional probability of "heads" is unaffected by an additional knowledge of what the process printed out in the past. (Note that Example 2 need not be an n-step Markov process for any n.)

EXAMPLE 4. The box contains a mechanical system such as a gas. At each unit of time, we make a fixed measurement on the system which has only a finite number of possible outcomes, and print out the outcome of the measurement. (If the outcome of the measurement were real-valued, we could divide the line into a finite number of sets, and print out which set the measurement fell into.)

EXAMPLE 5. A teleprinter. This prints out letters where the probability of printing out a given letter depends on what has already been printed. (Many possibilities will have probability 0 because they will not make sense.)

The mathematical model for a process is an invertible, measure preserving transformation T, acting on a space X of total measure 1, *and a partition $\mathcal{P}$ of X into a*

Received April 12, 1972; revised August 24, 1972.

[1] For most of this paper, we will restrict ourselves to discrete-time processes, with only a finite number of possible outputs. The theory will work for continuous time and more general state spaces, but I think it worthwhile studying the simplest case first.