ON RANDOM SAMPLING FROM A STOCHASTIC PROCESS[1]

By J. R. Blum and Judah Rosenblatt

University of New Mexico and Sandia Corporation

1. The problem. Let $\{X_n, n = 1, 2, \cdots\}$ be a stochastic process which is stationary and ergodic. Then it follows from the individual ergodic theorem that we may estimate the entire probability structure of the process by an observation $\{x_1, x_2, \cdots\}$ on the process.

Assume from now on that the random variables of the process are two-valued, i.e. $P\{X_n = 0\} = 1 - p = 1 - P\{X_n = 1\}$, where $0 < p < 1$. This is an unessential restriction which serves to simplify the ideas involved.

Now suppose that there are physical reasons which prohibit us from observing each of the successive random variables X_n. If we are then forced to observe a subsequence $\{X_{k_n}, n = 1, 2, \cdots\}$, we may ask whether it is still possible to estimate the probability structure of the original process from observing $\{X_{k_1}, X_{k_2}, \cdots\}$. In general the answer to this question is in the negative. For example, if k is an integer, $k > 1$, and $k_n = kn$ for $n = 1, 2, \cdots$; then while the process $\{X_{k_n}\}$ is still stationary and may also be still ergodic, it may be impossible to estimate the joint distribution of X_1 and X_2. Moreover in general the process $\{X_{k_n}\}$ may not be ergodic, or even stationary.

In this paper we shall consider what can be done with random sampling, that is when we assume that $\{k_n\}$ is a sequence of random variables. To formalize this notion we shall assume that in addition to the $\{X_n\}$ process we have at our disposal a sequence of random variables $\{Y_n, n = 1, 2, \cdots\}$ where the $\{Y_n\}$ process is independent of the $\{X_n\}$ process and consists of positive integer-valued random variables. We shall assume throughout that the $\{Y_n\}$ process is a stationary, ergodic process. In terms of the bivariate $\{X_n, Y_n\}$ process we can define a new bivariate process $\{Y_n, Z_n\}$ where the $\{Y_n\}$ process is as above and $Z_n = X_{N(n)}$, where $N(n) = \sum_{j=1}^{n} Y_j$, $n = 1, 2, \cdots$, and we assume that it is the $\{Y_n, Z_n\}$ process which is being observed. As we shall see below, it is easy to prove that the $\{Y_n, Z_n\}$ process is stationary. Under certain assumptions it will be shown that this process is also ergodic. Assume for the moment that this has already been done. Define

$$\begin{aligned} f(y, z_1, z_2) &= 0, \qquad &&\text{if } y \neq 1 \\ &= z_1 z_2, &&\text{if } y = 1. \end{aligned}$$

Then

Received January 9, 1964.

[1] Research supported by the National Science Foundation, Grant GP-1816, and in part by the United States Atomic Energy Commission. Reproduction in whole or in part is permitted for any purposes of the U. S. Government.