THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 51, No. 3, March, 1972
Printed in U.S.A.

On Maxentropic Discrete Stationary Processes

By D. SLEPIAN

(Manuscript received September 24, 1971)

This paper is concerned with the following mathematical problem. Let **X** *denote a stationary time-discrete random process whose variables,* $\cdots, X_{-1}, X_0, X_1, \cdots$, *take values from the finite set of real numbers* $\{x_1, x_2, \cdots, x_K\}$. *Let* **X** *have mean zero and a given covariance sequence* $\rho_k = EX_jX_{j+k}$, $j, k = 0, \pm 1, \pm 2, \cdots$. *What is the largest entropy that* **X** *can have and what is the probability structure of this most random process of given second moments?*

I. INTRODUCTION

Let **X** denote a stationary time-discrete random process whose variables, $\cdots, X_{-1}, X_0, X_1, \cdots$, take values from the finite set of real numbers $\{x_1, x_2, \cdots, x_K\}$. Let **X** have mean zero and a given covariance sequence $\rho_k = EX_jX_{j+k}$, $j, k = 0, \pm 1, \pm 2, \cdots$. What is the largest entropy that **X** can have and what is the probability structure of this most random process of given second moments?

Our interest in this question arose from the consideration of certain pulse-type communication systems used for the transmission of digital data. In such systems, a customer provides data in the form of an infinite sequence of binary digits that can be represented by a stationary process **Y** whose variables, $\cdots, Y_{-1}, Y_0, Y_1, \cdots$, are independent random variables each taking values zero and one with equal probabilities. An encoder transforms **Y** into a K-level process **X** of the sort described above, whose random variables are then used as amplitudes for successive pulses of a train. The transmitted signal is thus of the form

$$s(t) = \sum_{n=-\infty}^{\infty} X_n g(t - nT + \theta) \tag{1}$$

where $g(t)$ is the pulse shape and $T > 0$ is the pulse repetition period of the system. It is easy to compute that the power density spectrum of the stochastic process (1) is given by

$$\Phi_s(f) = \frac{|G(f)|^2}{T} \Phi_{\mathbf{x}}(fT) \tag{2}$$

where $G(f)$ is the Fourier transform of $g(t)$ and

$$\Phi_{\mathbf{x}}(f) = \sum_{-\infty}^{\infty} \rho_n e^{2\pi i n f} \tag{3}$$

is the spectrum of the discrete-amplitude process **X**. Here it has been assumed that θ is uniformly distributed in $(0, T)$.

Many different encoding schemes for mapping the customer's data stream **Y** onto the pulse amplitude stream **X** have been proposed in the past. Typical are dicode, partial response, pseudo-ternary, run-length-limited codes, etc.. Entry to the literature on this subject can be made through Refs. 1–4. In general, these encoding schemes are employed to give $\Phi_{\mathbf{x}}(f)$, and hence $\Phi_s(f)$, some desirable shape that will be particularly well-suited to the transmission medium, the noise, and the demodulation process. However, such deviations of $\Phi_{\mathbf{x}}(f)$ from a flat shape (Φ = constant) are bought at the price of a decreased information rate for the system as will be seen in an example below. Solution to the problem posed in the opening paragraph would yield the maximum information rate possible with given amplitudes $x_1, x_2, \cdots, x_K$ and given spectrum $\Phi_{\mathbf{x}}(f)$.

To illustrate these matters, consider the simple case of dicode for which the encoding is

$$X_n = Y_n - Y_{n-1}, \quad n = 0, \pm 1, \pm 2, \cdots .$$

Here $K = 3$ and the allowed pulse amplitudes are $x_1 = 1$, $x_2 = 0$, $x_3 = -1$. It is readily computed that for this amplitude-process $\rho_0 = \frac{1}{2}$, $\rho_1 = \rho_{-1} = -\frac{1}{4}$, and $\rho_n = 0$ for $n = \pm 2, \pm 3, \cdots$ and so $\Phi_{\text{dicode}}(f) = \sin^2 \pi f$. This spectrum vanishes like f^2 at zero frequency, a frequently desirable property. But, this 3-level scheme signals at a rate of only one bit of information per pulse whereas a rate of $\log_2 3 = 1.58$ bits per pulse could be had by appropriate mapping of the customer's binary digits onto independent random variables taking the same amplitude values, -1, 0, and 1 each with probability 1/3. This latter encoding would, of course, yield a flat spectrum. Thus dicode achieves a desired spectrum at the cost of about a 1/3 decrease in rate. Can any scheme with the same values and spectrum as dicode attain a rate greater than one bit per pulse? What is the highest rate so achievable?

We have been unable to answer even these seemingly simple specific questions. Quite apart from applications to pulse-amplitude data transmission systems, the general question of finding a maxentropic finite state discrete process of given second moments is of interest in its own right. As we shall see, such a process is a natural finite state analog of the Gaussian process and could serve as a convenient model in many contexts. We have been able to make only slight progress in solving this more general problem.

It is the purpose of this paper to record the progress we have made and the approaches we have followed in pursuing these goals, and to exhibit the difficulties encountered as well. It is hoped that others who may become interested in this problem can thereby avoid some pitfalls and be guided to more successful approaches.

II. REDUCTION TO THE MARKOV CASE

Let **X** be a stationary process $\cdots, X_{-1}, X_0, X_1, \cdots$ where each X takes values from the set of K real numbers $\{x_1, x_2, \cdots, x_K\}$. We denote the probability distribution of n successive Xs by

$$p_n(\epsilon_1, \cdots, \epsilon_n) = \Pr\{X_{i+1} = x_{\epsilon_1}, \cdots, X_{i+n} = x_{\epsilon_n}\}. \tag{4}$$

Here each index $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ takes values from the set $\{1, 2, \cdots, K\}$. We have, of course,

$$\sum_{\epsilon} p_n(\epsilon_1, \epsilon_2, \cdots, \epsilon_n) = 1 \tag{5}$$

and

$$p_n(\epsilon_1, \cdots, \epsilon_n) \geqq 0, \quad \epsilon_1, \epsilon_2, \cdots, \epsilon_n = 1, 2, \cdots, K. \tag{6}$$

The stationarity of **X** implies that the left of (4) is independent of i, and furthermore that

$$\sum_{\alpha=1}^{K} p_n(\epsilon_1, \epsilon_2, \cdots, \epsilon_{n-1}, \alpha) = \sum_{\alpha=1}^{K} p_n(\alpha, \epsilon_1, \epsilon_2, \cdots, \epsilon_{n-1})$$

$$= p_{n-1}(\epsilon_1, \cdots, \epsilon_{n-1}), \qquad \epsilon_1, \epsilon_2, \cdots, \epsilon_{n-1} = 1, 2, \cdots, K. \tag{7}$$

The statements (4) through (6) are to hold for $n = 1, 2, \cdots$ and (7) for $n = 2, 3, \cdots$. Note that if (7) holds for $n = n_o$, this implies the validity of (7) for $n = 2, 3, \cdots, n_o$.

The entropy of **X** is defined by

$$H(\mathbf{X}) = \lim_{n\to\infty} \frac{1}{n} H_n(\mathbf{X}), \tag{8}$$

$$H_n(\mathbf{X}) = -\sum p_n(\epsilon_1, \cdots, \epsilon_n) \log p_n(\epsilon_1, \cdots, \epsilon_n), \tag{9}$$

where the sum is over the K^n allowed values of the ϵ's. We seek to maximize (8) by suitable choice of a hierarchy of distributions $p_n(\epsilon_1, \cdots, \epsilon_n)$, $n = 1, 2, \cdots$ that satisfy (5), (6), and (7) and the constraints

$$EX_i = \sum_{\epsilon_1=1}^{K} x_{\epsilon_1} p_1(\epsilon_1) = 0, \tag{10}$$

$$EX_j X_{j+k} = \sum_{\epsilon} x_{\epsilon_1} x_{\epsilon_{k+1}} p_{k+1}(\epsilon_1, \cdots, \epsilon_{k+1}) = \rho_k, \qquad k = 0, 1, 2, \cdots. \tag{11}$$

Here the ρ_k are given and the sum is over all allowable values of $\epsilon_1, \epsilon_2, \cdots, \epsilon_{k+1}$.

We do not know how to proceed directly with this problem. One approach is to attempt to solve the problem when the constraint (11) is imposed only for $k = 0, 1, 2, \cdots, L$. That is, we seek the process of maximum entropy whose first $L + 1$ covariance elements are prescribed. Let $H^{(L)}(\mathbf{X})$ denote this maximum entropy and let $p_n^{(L)}(\epsilon_1, \cdots, \epsilon_n)$, $n = 1, 2, \cdots$, be the corresponding distribution. We would then investigate the behavior of these quantities as $L \to \infty$. We have, of course, $H^{(L)}(\mathbf{X}) \geqq H(\mathbf{X})$.

In Appendix A we establish

Theorem 1: The K-valued stationary discrete process of largest entropy with mean zero, given values $x_1, \cdots, x_K$, *and given values of* $\rho_0, \rho_1, \cdots, \rho_L$ *is an Lth-order Markov process.*

An Lth-order Markov process is characterized by the fact that

$$\Pr\{X_n = x_{\epsilon_n} \mid X_{n-1} = x_{\epsilon_{n-1}}, \cdots, X_{n-L} = x_{\epsilon_{n-L}}, X_{n-L-1} = x_{\epsilon_{n-L-1}}, \cdots\}$$
$$= \Pr\{X_n = x_{\epsilon_n} \mid X_{n-1} = x_{\epsilon_{n-1}}, \cdots, X_{n-L} = x_{\epsilon_{n-L}}\}$$

for all n and all allowable values of the ϵ's. A stationary Lth-order Markov process can be specified by K^{L+1} transition probabilities

$$q_L(\epsilon_{L+1} \mid \epsilon_1, \cdots, \epsilon_L) \equiv \Pr\{X_{L+1} = x_{\epsilon_{L+1}} \mid X_1 = x_{\epsilon_1}, \cdots, X_L = x_{\epsilon_L}\}$$
$$\epsilon_1, \cdots, \epsilon_{L+1} = 1, 2, \cdots, K$$

and a corresponding Lth-order distribution $p_L(\epsilon_1, \cdots, \epsilon_L)$ that satisfies

$$\sum_{\epsilon_1=1}^{K} q_L(\epsilon_{L+1} \mid \epsilon_1, \cdots, \epsilon_L) p_L(\epsilon_1, \cdots, \epsilon_L) = p_L(\epsilon_2, \cdots, \epsilon_{L+1})$$

$$\epsilon_2, \cdots, \epsilon_{L+1} = 1, 2, \cdots, K. \tag{12}$$

We have, of course

$$q_L(\epsilon_{L+1} \mid \epsilon_1, \cdots, \epsilon_L) \geqq 0, \tag{13}$$

$$\sum_{\alpha=1}^{K} q_L(\alpha \mid \epsilon_1, \cdots, \epsilon_L) = 1, \qquad \epsilon_1, \cdots, \epsilon_{L+1} = 1, 2, \cdots, K. \tag{14}$$

Equations (12) and (14) guarantee that the normalized solutions p_L of (12) have property (7) (with $n = L$). The general term p_n of the probability distribution for such a process is given in terms of p_L by the product rule

$$p_n(\epsilon_1, \cdots, \epsilon_n) = p_L(\epsilon_1, \cdots, \epsilon_L) q_L(\epsilon_{L+1} \mid \epsilon_1, \cdots, \epsilon_L)$$

$$q_L(\epsilon_{L+2} \mid \epsilon_2, \cdots, \epsilon_{L+1}) \cdots q_L(\epsilon_n \mid \epsilon_{n-L}, \epsilon_{n-L+1}, \cdots, \epsilon_{n-1}) \tag{15}$$

for $n > L$. For $n < L$,

$$p_n(\epsilon_1, \cdots, \epsilon_n) = \sum_{\alpha_1=1}^{K} \cdots \sum_{\alpha_{L-n}=1}^{K} p_L(\epsilon_1, \cdots, \epsilon_n, \alpha_1, \cdots, \alpha_{L-n}). \tag{16}$$

It is easy to show that for a stationary Lth-order Markov process the entropy (8) through (9) is given by

$$\begin{aligned} H &= -\sum_{\epsilon} p_L(\epsilon_1, \cdots, \epsilon_L) \sum_{\alpha} q_L(\alpha \mid \epsilon_1, \cdots, \epsilon_L) \log q_L(\alpha \mid \epsilon_1, \cdots, \epsilon_L) \\ &= -\sum p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) \log p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) \\ &\quad + \sum p_L(\epsilon_1, \cdots, \epsilon_L) \log p_L(\epsilon_1, \cdots, \epsilon_L) \\ &= H_{L+1} - H_L . \end{aligned} \tag{17}$$

III. THE DETAILED DISTRIBUTION

Now to find the most random stationary Lth-order Markov process with given $\rho_0, \rho_1, \cdots, \rho_L$, we must maximize $H_{L+1} - H_L$ by proper choice of the K^{L+1} quantities $\rho_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1})$ subject to certain linear constraints of the form

$$\sum_{\epsilon} a_i(\epsilon_1, \cdots, \epsilon_{L+1}) p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) = b_i \qquad i = 1, 2, \cdots, M. \tag{18}$$

We assume this system is of rank $M' \leqq M$.

There are two ways to proceed: (*i*) by the method of Lagrange multipliers which treats the unknown p_{L+1}'s symmetrically; (*ii*) by expressing $H_{L+1} - H_L$ in terms of $K^{L+1} - M'$ independent p_{L+1}'s obtained by solving (18). Both methods lead to unwieldy higher-order algebraic equations with which we have been able to do little in the general case. The form of the solutions is not without some interest, however, and we record it here.

To avoid unnecessary superficial complications, we shall henceforth assume that if x is one of the allowed values $x_1, x_2, \cdots, x_K$, then $-x$ is also in the set of allowed values. This condition will assure that

$$\Pr(X_1 = x_{\epsilon_1}, \cdots, X_n = x_{\epsilon_n}) = \Pr(X_1 = -x_{\epsilon_1}, \cdots, X_n = -x_{\epsilon_n})$$

in the optimal process and that $EX_j = 0, j = 0, \pm 1, \cdots$.

3.1 *Lagrange Multipliers*

Let us define the sample lag sums

$$\begin{aligned}
l_n^{(0)}(\epsilon_1, \cdots, \epsilon_n) &\equiv x_{\epsilon_1}^2 + x_{\epsilon_2}^2 + \cdots + x_{\epsilon_n}^2 \\
l_n^{(1)}(\epsilon_1, \cdots, \epsilon_n) &\equiv x_{\epsilon_1}x_{\epsilon_2} + x_{\epsilon_2}x_{\epsilon_3} + \cdots + x_{\epsilon_{n-1}}x_{\epsilon_n} \\
&\vdots \\
l_n^{(j)}(\epsilon_1, \cdots, \epsilon_n) &\equiv x_{\epsilon_1}x_{\epsilon_{1+j}} + x_{\epsilon_2}x_{\epsilon_{2+j}} + \cdots + x_{\epsilon_{n-j}}x_{\epsilon_n} \\
&\vdots \\
l_n^{(n-1)}(\epsilon_1, \cdots, \epsilon_n) &\equiv x_{\epsilon_1}x_{\epsilon_n} \qquad (19)
\end{aligned}$$

and the function

$$h_n(\epsilon_1, \cdots, \epsilon_n; \lambda_0, \lambda_1, \cdots, \lambda_{n-1}) \equiv \exp \sum_{j=0}^{n-1} \lambda_j l_n^{(j)}(\epsilon_1, \cdots, \epsilon_n). \qquad (20)$$

Then the Lagrange solution can be stated as follows. Solve the homogeneous system of equations

$$\sum_{\epsilon_{L+1}} h_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}; \lambda_0, \cdots, \lambda_L) f(\epsilon_2, \cdots, \epsilon_{L+1}) = \frac{1}{c} f(\epsilon_1, \cdots, \epsilon_L) \qquad \epsilon_1, \epsilon_2, \cdots, \epsilon_L = 1, 2, \cdots, K \qquad (21)$$

for the K^L f's and c. Then the transition probabilities and initial stationary distribution of the maxentropic process are given by

$$q_L(\epsilon_{L+1} \mid \epsilon_1, \cdots, \epsilon_L) = c h_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}; \lambda_0, \cdots, \lambda_L) \frac{f(\epsilon_2, \cdots, \epsilon_{L+1})}{f(\epsilon_1, \cdots, \epsilon_L)}, \qquad (22)$$

$$p_L(\epsilon_1, \cdots, \epsilon_L) = kf(\epsilon_1, \cdots, \epsilon_L)f(\epsilon_L, \cdots, \epsilon_1),$$
$$\epsilon_1, \cdots, \epsilon_{L+1} = 1, 2, \cdots, K. \quad (23)$$

In (23), $k > 0$ must be chosen so that the p_L sum to unity. A derivation of these equations is given in Appendix B.

While equations (21), (22), and (23) are a formal solution to our problem, in practice they are of little value. The solutions p_L and q_L contain the Lagrange multipliers $\lambda_0, \lambda_1, \cdots, \lambda_L$ in a complicated way, and these must be determined to give the prescribed covariance elements $\rho_0, \rho_1, \cdots, \rho_L$. Presumably that eigenvalue c of (21) should be taken which gives maximum entropy and yields $q_L \geqq 0$ in (22). In the small examples we have carried out, p_L and q_L turned out to be independent of the eigenvalue chosen in (21), but we have been unable to prove anything in general about this situation. For particular processes, say the symmetric binary process with $K = 2$, $x_1 = 1$, $x_2 = -1$, for example, equations (21) take a special simple seductive form that suggests the possibility of explicit closed-form solution. We have been unable to find one.

Perhaps the best that can be said for this curious Lagrange solution is that (23) shows clearly that in the maxentropic process $p_L(\epsilon_1, \cdots, \epsilon_L) = p_L(\epsilon_L, \cdots, \epsilon_1)$. It is not hard to see that the product rule (15) and the form of (22) and (23) propagate this property so that for arbitrary n, $p_n(\epsilon_1, \cdots, \epsilon_n) = p_n(\epsilon_n, \cdots, \epsilon_1)$. The maxentropic process treats past and future in a symmetric manner.

3.2 *The Independent Variable Approach*

We seek to maximize

$$J'' = -\sum_{\epsilon} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) \log p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1})$$
$$+ \sum_{\epsilon} p_L(\epsilon_1, \cdots, \epsilon_L) \log p_L(\epsilon_1, \cdots, \epsilon_L) \quad (24)$$

by choice of the K^{L+1} quantities $p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1})$. Here we define

$$p_L(\epsilon_1, \cdots, \epsilon_L) \equiv \sum_{\alpha} p_{L+1}(\epsilon_1, \cdots, \epsilon_L, \alpha),$$
$$\epsilon_1, \cdots, \epsilon_L = 1, 2, \cdots, K. \quad (25)$$

The p_{L+1}'s must satisfy

$$\sum_{\epsilon} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) = 1, \quad (26)$$

$$\sum_{\alpha} p_{L+1}(\epsilon_1, \cdots, \epsilon_L, \alpha) - \sum_{\alpha} p_{L+1}(\alpha, \epsilon_1, \cdots, \epsilon_L) = 0,$$

$$\epsilon_1 , \cdots , \epsilon_L = 1, 2, \cdots , K, \qquad (27)$$

$$\sum_{\epsilon} x_{\epsilon_1} x_{\epsilon_{k+1}} p_{L+1}(\epsilon_1 , \cdots , \epsilon_{L+1}) = \rho_k , \qquad k = 0, 1, \cdots , L. \qquad (28)$$

These $K^L + L + 2 \equiv M$ equations are of the form (18). We suppose that they can be solved for $M' \leqq M$ of the p_{L+1}'s in terms of the remaining $K^{L+1} - M' \equiv M''$ ones. We denote these M'' independent p_{L+1}'s by the variables $\xi_1 , \xi_2 , \cdots , \xi_{M''}$ and denote the M' dependent p_{L+1}'s by $\eta_1 , \cdots , \eta_{M'}$. Thus we write equations (26), (27), and (28) in the form

$$\eta_i = \alpha_i + \sum_{j=1}^{M''} \beta_{ij}\xi_j , \qquad i = 1, 2, \cdots , M'. \qquad (29)$$

It is convenient also to adopt a single index notation for the p_L of (25) which we now denote by $\zeta_1 , \zeta_2 , \cdots , \zeta_N$, where $N = K^L$. By means of (29) the right of (25) can be expressed in terms of the ξ's. We write

$$\zeta_i = \delta_i + \sum_{j=1}^{M''} \gamma_{ij}\xi_j , \qquad i = 1, 2, \cdots , K^L. \qquad (30)$$

We further note that (26) and (25) imply that

$$\sum_1^{M'} \xi_i + \sum_1^{M''} \eta_i = 1, \qquad \sum_1^{N} \zeta_i = 1.$$

Since these are to hold as identities in the ξ's we must have

$$\sum_1^{M'} \alpha_i = \sum_1^{N} \delta_i = 1, \qquad \sum_{i=1}^{N} \gamma_{ij} = 1 + \sum_{i=1}^{M'} \beta_{ij} = 0,$$

$$j = 1, 2, \cdots , M''. \qquad (31)$$

In this new notation (24) becomes

$$J'' = -\sum_1^{M''} \xi_i \log \xi_i - \sum_1^{M'} \eta_i \log \eta_i + \sum_1^{N} \zeta_i \log \zeta_i$$

where the η's and ζ's are explicit linear functions of the ξ's given by (29) and (30). Maximizing with respect to the ξ's gives

$$\frac{\partial J''}{\partial \xi_j} = 0 = -1 - \log \xi_j - \sum_{i=1}^{M'} (1 + \log \eta_i)\beta_{ij} + \sum_1^{N} (1 + \log \zeta_i)\gamma_{ij} .$$

On taking account of (31), we find finally

$$\xi_j = \frac{\prod^{N} \zeta_i^{\gamma_{ij}}}{\prod^{M'} \eta_i^{\beta_{ij}}}, \qquad j = 1, \cdots, M''. \tag{32}$$

These are M'' equations for the M'' ξ's. There seems to be little that can be said in general about them, and so the trail ends here. (We note only that the form of these equations is appropriate for an iterative numerical solution: a trial set of ξ's used to evaluate the products yields directly a new set of ξ's.)

IV. SOME SIMPLE EXAMPLES

We consider first the binary case and set

$$x_1 = 1, \quad x_2 = -1, \tag{33}$$

In this case we must take $\rho_0 = 1$.

When $L = 1$, we find

$$p(1, 1) = p(2, 2) = \tfrac{1}{4}(1 + \rho_1)$$

$$p(1, 2) = p(2, 1) = \tfrac{1}{4}(1 - \rho_1)$$

$$H = -\tfrac{1}{2}(1 + \rho_1) \log \tfrac{1}{2}(1 + \rho_1) - \tfrac{1}{2}(1 - \rho_1) \log \tfrac{1}{2}(1 - \rho_1)$$

$$\rho_n = \rho_1^n, \qquad n = 0, 1, 2, \cdots,$$

$$\Phi_x(f) = \frac{1 - \rho_1^2}{1 + \rho_1^2 - 2\rho_1 \cos 2\pi f}.$$

When $L = 2$,

$$p(1, 1, 1) = p(2, 2, 2) = \tfrac{1}{8}(1 + 2\rho_1 + \rho_2)$$

$$p(1, 1, 2) = p(2, 2, 1) = \tfrac{1}{8}(1 - \rho_2)$$

$$p(1, 2, 2) = p(2, 1, 1) = \tfrac{1}{8}(1 - \rho_2)$$

$$p(1, 2, 1) = p(2, 1, 2) = \tfrac{1}{8}(1 - 2\rho_1 + \rho_2).$$

Let

$$\alpha_\pm = \frac{1}{2(1 - \rho_1^2)} \left[\rho_1(1 - \rho_2) \pm \sqrt{4\rho_1^4 + \rho_1^2(\rho_2^2 - 6\rho_2 - 3) + 4\rho_2}\right].$$

Then

$$\rho_n = \frac{1}{(1 + \alpha_+\alpha_-)(\alpha_+ - \alpha_-)} \left[(1 - \alpha_-^2)\alpha_+^{n+1} - (1 - \alpha_+^2)\alpha_-^{n+1}\right]$$

$$\Phi_x(f) = \frac{(1 - \rho_1^2)(1 - \rho_2)(1 - 2\rho_1^2 + \rho_2)}{1 - \rho_1^2 + 2\rho_1^4 - 4\rho_1^2\rho_2 + \rho_2^2 + \rho_1^2\rho_2^2 - 2\rho_1(1 - \rho_2)^2 \cos 2\pi f + 2(\rho_1^2 - \rho_2)(1 - \rho_1^2) \cos 4\pi f}.$$

When $L = 3$, we are already in algebraic difficulties. Equations (26), (27), and (28) in the present case permit us to solve for all the p's in terms of $\xi \equiv p(1, 1, 1, 1)$. We find

$$p(1, 1, 1, 2) = p(1, 2, 2, 2) = p(2, 1, 1, 1) = p(2, 2, 2, 1) = \tfrac{1}{8}(1 + 2\rho_1 + \rho_2) - \xi$$

$$p(1, 1, 2, 1) = p(1, 2, 1, 1) = p(2, 1, 2, 2) = p(2, 2, 1, 2) = \tfrac{1}{8}(1 + \rho_1 + \rho_2 + \rho_3) - \xi$$

$$p(1, 1, 2, 2) = p(2, 2, 1, 1) = \tfrac{1}{8}(1 - \rho_1 - 2\rho_2 - \rho_3) + \xi$$

$$p(1, 2, 1, 2) = p(2, 1, 2, 1) = \tfrac{1}{8}(-3\rho_1 - \rho_3) + \xi$$

$$p(1, 2, 2, 1) = p(2, 1, 1, 2) = \tfrac{1}{8}(-2\rho_1 - 2\rho_2) + \xi$$

$$p(2, 2, 2, 2) = \xi.$$

On setting $Z = 8\xi$, equation (32) becomes

$$Z = \frac{[1 + 2\rho_1 + \rho_2 - Z]^4[1 + \rho_1 + \rho_2 + \rho_3 - Z]^4}{Z[-3\rho_1 - \rho_3 + Z]^2[-\rho_1 - 2\rho_2 - \rho_3 + Z]^2[-2\rho_1 - 2\rho_2 + Z]^2}.$$

One can take the square root of both sides of this equation, clear fractions, and expand to obtain a cubic equation in Z. It is not hard to show that there are no roots rational in ρ_1, ρ_2, and ρ_3, so that the simple dependence of p on the ρ's exhibited for the cases $L = 1$ and 2 fails here.

We next consider the case $K = 3$ and choose

$$x_1 = 1, \quad x_2 = 0, \quad x_3 = -1.$$

Here with $L = 1$ we already meet with higher-degree algebraic equations. The constraints permit solution of all p's in terms of $p(1, 1) = \xi$. We find

$$p(1, 2) = p(2, 1) = p(2, 3) = p(3, 2) = \tfrac{1}{2}(\rho_0 + \rho_1) - 2\xi$$

$$p(3, 1) = p(1, 3) = -\tfrac{1}{2}\rho_1 + \xi$$

$$p(2, 2) = 1 - 2\rho_0 - \rho_1 + 4\xi$$

$$p(3, 3) = \xi.$$

On setting $Z = 4\xi$, equation (32) becomes

$$Z = \pm\frac{(\rho_0 + \rho_1 - Z)^4}{(-2\rho_1 + Z)(1 - 2\rho_0 - \rho_1 + Z)^2} \tag{34}$$

which is a cubic in Z. One finds

$$\rho_n = \rho_0\left(\frac{\rho_1}{\rho_0}\right)^n, \qquad n = 0, 1, 2, \cdots \tag{35}$$

quite independent of the value chosen for ξ. The spectrum is given by

$$\Phi_x(f) = \frac{\rho_0(\rho_0^2 - \rho_1^2)}{\rho_0^2 + \rho_1^2 - 2\rho_0\rho_1 \cos 2\pi f}. \tag{36}$$

Using the dicode values $\rho_0 = \frac{1}{2}$, $\rho_1 = -\frac{1}{4}$, one finds from (34) that $\xi = 0.0103$. The entropy of the resulting simple Markov process is found to be 1.299 bits, which is greater than the one-bit rate of dicode. While the first two terms of the covariance sequence agree with the dicode values, the higher terms are given by (35) and the spectrum, as given by (36), does not vanish for $f = 0$.

The case $K = 3$, $L = 2$ begins to reveal the complexity of the general case. We denote each of the 27 quantities $p(i, j, k)$ by x with a subscript ranging from 1 to 27. The association is made by listing the p's in order, interpreting (ijk) as a three-digit number. Thus $x_1 = p(1, 1, 1)$, $x_2 = p(1, 1, 2)$, $x_3 = p(1, 1, 3)$, $\cdots$, $x_{27} = p(3, 3, 3)$. Equations (26), (27), and (28) can be solved to express all the x's in terms of five of them. Equations (29) are

$$\begin{aligned}
\eta_1 &= -\tfrac{1}{4} + \rho_1 - \tfrac{1}{4}\rho_2 - \xi_1 + 3\xi_2 + 4\xi_3 + \tfrac{3}{2}\xi_4 + \tfrac{1}{4}\xi_5 \\
\eta_2 &= \tfrac{1}{4} - \tfrac{1}{2}\rho_1 + \tfrac{1}{4}\rho_2 + \xi_1 - 2\xi_2 - 3\xi_3 - \tfrac{3}{2}\xi_4 - \tfrac{1}{4}\xi_5 \\
\eta_3 &= \tfrac{1}{8} + \tfrac{1}{4}\rho_0 - \tfrac{1}{8}\rho_2 - \tfrac{1}{2}\xi_1 + \tfrac{1}{4}\xi_4 + \tfrac{1}{8}\xi_5 \\
\eta_4 &= \tfrac{1}{2}\rho_0 - \rho_1 + \tfrac{1}{2}\rho_2 + \xi_1 - 4\xi_2 - 4\xi_3 - \xi_4 \\
\eta_5 &= \tfrac{1}{4} - \tfrac{1}{2}\rho_0 + \tfrac{1}{2}\rho_1 - \tfrac{1}{4}\rho_2 - \xi_1 + 2\xi_2 + 2\xi_3 + \tfrac{1}{2}\xi_4 - \tfrac{1}{4}\xi_5 \\
\eta_1 &= x_1 = x_{27} \\
\eta_2 &= x_2 = x_{10} = x_{18} = x_{26} \\
\eta_3 &= x_3 = x_9 = x_{19} = x_{25} \\
\eta_4 &= x_4 = x_{24} \\
\eta_5 &= x_5 = x_{13} = x_{15} = x_{23} \\
\xi_1 &= x_6 = x_{22} \\
\xi_2 &= x_7 = x_{21} \\
\xi_3 &= x_8 = x_{12} = x_{16} = x_{20} \\
\xi_4 &= x_{11} = x_{17} \\
\xi_5 &= x_{14}\,.
\end{aligned} \tag{37}$$

Equations (30) become

$$\begin{aligned}
\zeta_1 &= -\tfrac{1}{8} + \tfrac{1}{4}\rho_0 + \tfrac{1}{2}\rho_1 - \tfrac{1}{8}\rho_2 - \tfrac{1}{2}\xi_1 + \xi_2 + \xi_3 + \tfrac{1}{4}\xi_4 + \tfrac{1}{8}\xi_5 \\
\zeta_2 &= \tfrac{1}{4} - \tfrac{1}{2}\rho_1 + \tfrac{1}{4}\rho_2 + \xi_1 - 2\xi_2 - 2\xi_3 - \tfrac{1}{2}\xi_4 - \tfrac{1}{4}\xi_5 \\
\zeta_3 &= -\tfrac{1}{8} + \tfrac{1}{4}\rho_0 - \tfrac{1}{8}\rho_2 - \tfrac{1}{2}\xi_1 + \xi_2 + \xi_3 + \tfrac{1}{4}\xi_4 + \tfrac{1}{8}\xi_5 \\
\zeta_4 &= \tfrac{1}{2} - \rho_0 + \rho_1 - \tfrac{1}{2}\rho_2 - 2\xi_1 + 4\xi_2 + 4\xi_3 + \xi_4 + \tfrac{1}{2}\xi_5
\end{aligned} \tag{38}$$

where

$$\begin{aligned}
\zeta_1 &= p(1, 1) = p(3, 3) \\
\zeta_2 &= p(1, 2) = p(3, 2) = p(2, 1) = p(2, 3) \\
\zeta_3 &= p(1, 3) = p(3, 1) \\
\zeta_4 &= p(2, 2).
\end{aligned}$$

For the case at hand, equations (32) become

$$\begin{aligned}
\xi_1^2 &= \frac{\zeta_1^{-1}\zeta_2^{4}\zeta_3^{-1}\zeta_4^{-2}}{\eta_1^{-2}\eta_2^{4}\eta_3^{-2}\eta_4^{2}\eta_5^{-4}}, &\qquad \xi_2^2 &= \frac{\zeta_1^{2}\zeta_2^{-8}\zeta_3^{2}\zeta_4^{4}}{\eta_1^{6}\eta_2^{-8}\eta_4^{-8}\eta_5^{8}} \\
\xi_3^4 &= \frac{\zeta_1^{2}\zeta_2^{-8}\zeta_3^{2}\zeta_4^{4}}{\eta_1^{8}\eta_2^{-12}\eta_4^{-8}\eta_5^{8}}, &\qquad \xi_4^2 &= \frac{\zeta_1^{\frac{1}{2}}\zeta_2^{-2}\zeta_3^{\frac{1}{2}}\zeta_4^{1}}{\eta_1^{3}\eta_2^{-6}\eta_3\eta_4^{-2}\eta_5^{2}} \\
\xi_5 &= \frac{\zeta_1^{\frac{1}{4}}\zeta_2^{-1}\zeta_3^{\frac{1}{4}}\zeta_4^{\frac{1}{2}}}{\eta_1^{\frac{1}{2}}\eta_2^{-1}\eta_3^{\frac{1}{2}}\eta_5^{-1}}.
\end{aligned} \tag{39}$$

The right members of these equations can be written in terms of the ξ's by using (37) and (38). Equations (39) can then be written as five multinomial equations in the five ξ's. In principle, by using Sylvester's method,[5] the ξ's could be systematically eliminated to yield a single high-order polynomial equation for ξ_1. The other ξ's can be similarly determined. To carry this out in practice would be a formidable task.

V. THE COVARIANCE PROBLEM

We have seen that the maxentropic discrete stationary process with given values $x_1, x_2, \cdots, x_K$ and given truncated covariance sequence $\rho_0, \rho_1, \cdots, \rho_L$ is an Lth order Markov process. In Section III a formal solution was given to the problem of determining the complete probability structure of this process. This structure in turn determines the remaining elements $\rho_{L+1}, \rho_{L+2}, \cdots$ of the covariance sequence. It is shown in Appendix C that for a K-valued Lth order Markov process the covariance sequence can always be written in the form

$$\rho_n = \sum_{j=1}^{K^L} A_j \theta_j^n, \qquad n = 0, 1, 2, \cdots . \tag{40}$$

Thus, only a restricted class of covariance sequences, those expressible as a finite sum of exponentials, can be obtained by our procedure. The dicode covariance, $\rho_0 = \frac{1}{2}$, $\rho_1 = -\frac{1}{4}$, $\rho_n = 0$, $n = 2, 3, \cdots$, is excluded, for example.

This raises an important pertinent question that we have sidestepped thus far: what are the possible covariance sequences for a discrete stationary process taking values $x_1, x_2, \cdots, x_K$? When the restriction on allowed values of the process is removed, one has the elegant Bochner theorem[6] that characterizes the covariance sequences as Fourier cosine-series coefficients of non-negative finite measures, that is, as non-negative definite sequences. No comparable description is available for the proper subset of these non-negative definite sequences that comprises the covariance sequences of processes restricted to the values $x_1, x_2, \cdots, x_K$. The situation has been discussed by B. McMillan[7] and L. A. Shepp.[8]

More germaine to our discussion, and less ambitious, is the question: "What sequences of $L + 1$ numbers, $\rho_0, \rho_1, \cdots, \rho_L$, can be the first $L + 1$ terms of the covariance sequence of a discrete stationary process taking values $x_1, \cdots, x_K$?" If we consider such a sequence as a point in $\mathcal{E}_{L+1}$, Euclidean space of $L + 1$ dimensions, then the region $\mathcal{R}$ of admissible points is a convex one bounded by fewer than $2K^L$ hyperplanes. This is shown in Appendix D. Such a region can be characterized as the convex hull of its extreme points, or vertices (finite in number), and a convenient description of the region is a list of these vertices. An alternate economical description is a list of the hyperplane boundaries of $\mathcal{R}$. We have been unable to sort out the combinatorics involved, even in the simple case $K = 2$, $x_1 = 1$, $x_2 = -1$, to provide such lists for arbitrary values of L.

It is to be expected that the formalisms of Section III will have solutions $p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1})$ that are probability distributions if and only if $\rho_0, \rho_1, \cdots, \rho_L$ is contained in $\mathcal{R}$. In the few cases that we have been able to carry out in detail, this is indeed the case. For example, from Section IV, we see that the solution presented is valid only if

$$1 + 2\rho_1 + \rho_2 \geqq 0,$$

$$1 - 2\rho_1 + \rho_2 \geqq 0,$$

$$1 - \rho_2 \geqq 0.$$

These inequalities do indeed describe the region of admissible values of ρ_1 and ρ_2 for a process having values +1 and −1.

VI. THE UNIFILAR MARKOV MESSAGE SOURCE

The dicode process is not an Lth order Markov process for any L. It can, however, be described simply in terms of a two-state Markov chain. Consider the chain with states S_1 and S_2 and transition probabilities $\frac{1}{2}$ as shown in Fig. 1. Along each transition path in the figure is an associated number enclosed in a box. When the chain passes along a path from one state to another, the associated number is "emitted." The sequence of emitted numbers is the dicode process.

The foregoing is an example of a class of discrete processes which we call unifilar Markov message sources. An ergodic Markov chain with states $S_1, S_2, \cdots, S_N$ is given along with the transition probabilities $p_{ij} = \Pr\{\text{next state is } S_j \mid \text{last state is } S_i\}$. Associated with each pair of states S_i, S_j for which $p_{ij} > 0$ is a number $X(S_i, S_j)$ that is emitted when the chain passes from S_i to S_j. The word unifilar refers to the fact that we demand that whenever $S_j \neq S_k$, then $X(S_i, S_j) \neq X(S_i, S_k)$, $i = 1, 2, \cdots, N$. If this condition is met and the initial state of the chain is known, the sequence of emitted letters determines the sequence of states followed by the chain and a simple formula is available for the entropy of the emitted X process, namely

$$H = -\sum_{i,j} p(S_i)p_{ij} \log p_{ij} . \tag{41}$$

(See Ref. 9, p. 68.) Here $p(S_i)$, the probability that the chain be in state S_i, is the stationary measure satisfying

$$\sum_i p(S_i)p_{ij} = p(S_j), \qquad j = 1, 2, \cdots, N.$$

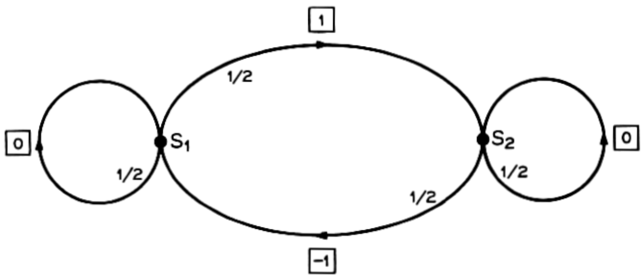

Fig. 1—Diagram of a two-state Markov chain with states S_1 and S_2 and transition probabilities 1/2.

It is easy to write an expression for the covariance sequence of a unifilar Markov message source. When $n \geqq 2$, we have

$$\rho_n = \sum_{\substack{i,j\\k,l}} X(S_i, S_j)p(S_i)p_{ij}p_{jk}^{(n-1)}p_{kl}X(S_k, S_l) \tag{42}$$

where $p_{jk}^{(n)}$ is the probability that the chain be in state S_k after n transitions given that it started from S_j. We have also

$$\rho_1 = \sum_{i,j,k} X(S_i, S_j)p(S_i)p_{ij}p_{jk}X(S_j, S_k), \tag{43}$$

$$\rho_0 = \sum_{i,j} X(S_i, S_j)^2 p(S_i)p_{ij}. \tag{44}$$

Since $p_{ij}^{(n)}$ has an expression analogous to (75), in Appendix C, equation (42) can be written

$$\rho_n = \sum_1^N B_j \varphi_j^n, \qquad n = 2, 3, \cdots. \tag{45}$$

Comparison with (40) shows that the covariance sequences achievable are of almost the same form as for the Lth-order Markov processes. For the unifilar Markov message source, deviation from the sum of exponential form may occur for ρ_0 and ρ_1.

To find the unifilar Markov message source of largest entropy with N states and given truncated covariance sequence appears to be a most difficult problem. We have not found a unifilar Markov source with values $0, \pm 1$, with the dicode covariance sequence and an entropy greater than unity.

VII. THE n-VARIATE GAUSSIAN ANALOGUE

Closely related to the problem we have been discussing is the following question. Let $X_1, X_2, \cdots, X_n$ be n random variables each taking values from the set $x_1, x_2, \cdots, x_K$. What distribution for the Xs, $p_n(\epsilon_1, \cdots, \epsilon_n)$ say, has maximum entropy and given second moments $EX_iX_j = \rho_{ij}$? Using Lagrange multipliers, one finds at once that

$$p_n(\epsilon_1, \cdots, \epsilon_n) = c \exp\left\{-\frac{1}{2}\sum_1^n \sigma_{ij}x_{\epsilon_i}x_{\epsilon_j}\right\}. \tag{46}$$

Here

$$\frac{1}{c} \equiv S = \sum_\epsilon \exp\left\{-\frac{1}{2}\sum_{i,j=1}^n \sigma_{ij}x_{\epsilon_i}x_{\epsilon_j}\right\} \tag{47}$$

and the σ's must be determined so that

$$\rho_{ij} = -\frac{2}{S}\frac{\partial S}{\partial \sigma_{ij}}$$

$$= c \sum_{\epsilon} x_{\epsilon_i} x_{\epsilon_j} \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{n} \sigma_{ij} x_{\epsilon_i} x_{\epsilon_j}\right\}, \qquad i, j = 1, \cdots, n. \tag{48}$$

The entropy of (46) can be written

$$H = -\sum_{\epsilon} p_n(\epsilon_1, \cdots, \epsilon_n) \log p_n(\epsilon_1, \cdots, \epsilon_n)$$

$$= \log S + \tfrac{1}{2} \sum \rho_{ij}\sigma_{ij} . \tag{49}$$

The analogy of (46) with the n-variate Gaussian density is striking. Let $Y_1, Y_2, \cdots, Y_n$ be n real-valued random variables having probability density $\hat{p}_n(y_1, y_2, \cdots, y_n)$. Let $EY_iY_j = \rho_{ij}$. Under these constraints, the density having largest entropy is the Gaussian density

$$\hat{p}_n(y_1, \cdots, y_n) = \hat{c} \exp\{-\tfrac{1}{2} \sum \hat{\sigma}_{ij} y_i y_j\} \tag{50}$$

Here

$$\frac{1}{\hat{c}} \equiv \hat{S} = \int_{-\infty}^{\infty} dy_1 \cdots \int_{-\infty}^{\infty} dy_n \exp\{-\tfrac{1}{2} \sum \hat{\sigma}_{ij} y_i y_j\} \tag{51}$$

and the $\hat{\sigma}$'s are related to the ρ's by

$$\rho_{ij} = -\frac{2}{\hat{S}}\frac{\partial \hat{S}}{\partial \hat{\sigma}_{ij}}, \qquad i, j = 1, 2, \cdots, n. \tag{52}$$

The entropy of (50) can be written

$$\hat{H} = -\int_{-\infty}^{\infty} dy_1 \cdots \int_{-\infty}^{\infty} dy_n \hat{p}_n(y_1, \cdots, y_n) \log \hat{p}_n(y_1, \cdots, y_n)$$

$$= \log \hat{S} + \tfrac{1}{2} \sum \rho_{ij}\hat{\sigma}_{ij} . \tag{53}$$

Note the complete parallel between (46) through (49) and (50) through (53).

In the case of the Gaussian density, the integral (51) can be performed to yield the simple expression

$$\hat{S} = \frac{(2\pi)^{n/2}}{|\hat{\sigma}|^{\frac{1}{2}}} \tag{54}$$

where $|\hat{\sigma}|$ is the determinant of the matrix with elements $\hat{\sigma}_{ij}$. Equation (52) then gives at once the well-known results $\rho_{ij} = \hat{\sigma}^{-1}_{ij}$ or $\hat{\sigma} = \rho^{-1}$, where we use obvious matrix notation. Surprisingly, the σ_{ij} are rational in the ρ_{ij} in spite of the more complicated nature of the dependence of $\hat{S}$ on the $\hat{\sigma}_{ij}$, as given by (54).

In the discrete case, matters are not so simple. For example, when $n = 3$, $K = 2$ and $x_1 = 1$, $x_2 = -1$,

$$S = 2e^{-\frac{1}{2}(\sigma_{11}+\sigma_{22}+\sigma_{33})}[e^{-\frac{1}{2}(\sigma_{12}+\sigma_{13}+\sigma_{23})} + e^{+\frac{1}{2}(\sigma_{12}-\sigma_{13}-\sigma_{23})} + e^{-\frac{1}{2}(-\sigma_{12}+\sigma_{13}-\sigma_{23})} + e^{-\frac{1}{2}(-\sigma_{12}-\sigma_{13}+\sigma_{23})}].$$

One finds

$$\sigma_{12} = -\tfrac{1}{8} \log \frac{\beta_1\beta_2}{\beta_3\beta_4}$$

$$\sigma_{13} = -\tfrac{1}{8} \log \frac{\beta_1\beta_3}{\beta_2\beta_4}$$

$$\sigma_{23} = -\tfrac{1}{8} \log \frac{\beta_1\beta_4}{\beta_2\beta_3}$$

$$c = \tfrac{1}{8}[\beta_1\beta_2\beta_3\beta_4]^{\frac{1}{4}}$$

where

$$\beta_1 = 1 + \rho_{12} + \rho_{13} + \rho_{23}$$

$$\beta_2 = 1 + \rho_{12} - \rho_{13} - \rho_{23}$$

$$\beta_3 = 1 - \rho_{12} + \rho_{13} - \rho_{23}$$

$$\beta_4 = 1 - \rho_{12} - \rho_{13} + \rho_{23} .$$

Thus σ_{12}, σ_{13}, and σ_{23} are not rational in the ρ's. (The quantities σ_{11}, σ_{22}, and σ_{33} can be chosen to be zero in this binary case.) The probabilities themselves, however, turn out to be linear in the ρ's. One has $p_3(1, 1, 1) = p_3(2, 2, 2) = \frac{1}{8}\beta_1$, $p(1, 1, 2) = p(2, 2, 1) = \frac{1}{8}\beta_2$, $p(1, 2, 1) = p(2, 1, 2) = \frac{1}{8}\beta_3$, $p(1, 2, 2) = p(2, 1, 1) = \frac{1}{8}\beta_4$. When $n > 2$, the p's become algebraic in the ρ's.

VIII. CONCLUDING REMARKS

In addition to the methods discussed here, I have pursued several other attacks on the problem at hand. All approaches seem to end in unmanageable algebraic complexities. Perhaps it is the nature of the problem; perhaps the answer can't be stated in simple terms. The mathematician, whose pleasure it is to make order out of chaos, will likely disagree. He will feel that surely so basic a construct as we consider here must be simple at heart and that we have only failed to find the appropriate language to make it so appear. In analogy, to the uninitiated, the relationship found in the last section between ρ and $\hat{\sigma}$, namely $\hat{\sigma} = p^{-1}$, must surely at first have appeared formidable. The matrix language of Cayley brings us apparent order here. Can we find the right point of view in which to describe the discrete maxentropic process?

IX. ACKNOWLEDGMENT

I am happy to acknowledge many most pleasant and profitable talks with L. A. Shepp on all phases of the work reported here.

APPENDIX A

We are concerned with maximizing (8) subject to (5), (6), (7), (10), and the $L + 1$ constraints

$$\sum_{\epsilon_1=1}^{K} \cdots \sum_{\epsilon_{k+1}=1}^{K} x_{\epsilon_1} x_{\epsilon_{k+1}} p_{k+1}(\epsilon_1, \cdots, \epsilon_{k+1}) = \rho_k, \qquad k = 0, 1, \cdots, L. \tag{55}$$

We proceed by maximizing H_n of (9), subject to these same constraints, for each $n > L + 1$.

Observe now that (55) and (10) can be stated solely in terms of p_{L+1}:

$$\sum_{\epsilon_1=1}^{K} \cdots \sum_{\epsilon_{L+1}=1}^{K} x_{\epsilon_1} x_{\epsilon_{k+1}} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) = \rho_k, \qquad k = 0, 1, \cdots, L \tag{56}$$

$$\sum_{\epsilon_1=1}^{K} \cdots \sum_{\epsilon_{L+1}=1}^{K} x_{\epsilon_1} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) = 0. \tag{57}$$

Thus the maximization can be carried out by first maximizing H_n subject to (5), (6), and (7) *given* the K^{L+1} quantities $p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1})$, $\epsilon_1, \cdots, \epsilon_{L+1} = 1, 2, \cdots, K$, then maximizing further over these quantities subject to the additional constraints (56) and (57). For this first maximization problem, the constraints are (5), (6), (7), and

$$\begin{aligned} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) &= \sum_{\alpha} p_n(\epsilon_1, \cdots, \epsilon_{L+1}, \alpha_1, \cdots, \alpha_{n-L-1}) \\ &= \sum_{\alpha} p_n(\alpha_1, \epsilon_1, \cdots, \epsilon_{L+1}, \alpha_2, \cdots, \alpha_{n-L-1}) \\ &= \sum_{\alpha} p_n(\alpha_1, \alpha_2, \epsilon_1, \cdots, \epsilon_{L+1}, \alpha_3, \cdots, \alpha_{n-L-1}) \\ &\;\;\vdots \\ &= \sum_{\alpha} p_n(\alpha_1, \cdots, \alpha_{n-L-1}, \epsilon_1, \cdots, \epsilon_{L+1}) \end{aligned}$$

$$\epsilon_1, \cdots, \epsilon_{L+1} = 1, 2, \cdots, K \tag{58}$$

where we regard the p_{L+1} as given. These quantities, of course, must themselves satisfy (5), (6), and (7) with $n = L + 1$.

Introducing Lagrange multipliers we seek the maximum of

$$J = -\sum_{\varepsilon} p_n(\epsilon_1, \cdots, \epsilon_n) \log p_n(\epsilon_1, \cdots, \epsilon_n) + \lambda \sum_{\varepsilon} p_n(\epsilon_1, \cdots, \epsilon_n)$$
$$+ \sum_{j=0}^{n-L-1} \sum_{\alpha, \varepsilon} \nu^{(j)}_{\epsilon_1 \cdots \epsilon_{L+1}}$$
$$p_n(\alpha_1, \cdots, \alpha_j, \epsilon_1, \cdots, \epsilon_{L+1}, \alpha_{j+1}, \cdots, \alpha_{n-L-1}) \tag{59}$$

where in the last sum for $j = 0$ the first argument of p_n is ϵ_1 and for $j = n - L - 1$ the last argument of p_n is ϵ_{L+1}. In (59) we have omitted terms corresponding to the constraint (7). It turns out that this constraint will be met automatically. Differentiation of (59) with respect to $p_n(\epsilon_1, \cdots, \epsilon_n)$ yields

$$-1 - \log p_n(\epsilon_1, \cdots, \epsilon_n)$$
$$+ \lambda + \nu^{(0)}_{\epsilon_1 \cdots \epsilon_{L+1}} + \nu^{(1)}_{\epsilon_2 \cdots \epsilon_{L+2}} + \cdots + \nu^{(n-L-1)}_{\epsilon_{n-L} \cdots \epsilon_n} = 0$$

or

$$p_n(\epsilon_1, \cdots, \epsilon_n) = c \exp \{\nu^{(0)}_{\epsilon_1 \cdots \epsilon_{L+1}} + \nu^{(1)}_{\epsilon_2 \cdots \epsilon_{L+2}} + \cdots + \nu^{(n-L-1)}_{\epsilon_{n-L} \cdots \epsilon_n}\} \tag{60}$$

where c is independent of the ϵ's. Equation (58) and (5) serve in principle to determine c and the $K^{L+1}(n - L)$ Lagrange multipliers $\nu^{(j)}_{\epsilon_1 \cdots \epsilon_{L+1}}$.

Note now that from (60) we find that

$$\Pr \{X_n = x_{\epsilon_n} \mid X_1 = x_{\epsilon_1}, \cdots, X_{n-1} = x_{\epsilon_{n-1}}\} = \frac{p_n(\epsilon_1, \cdots, \epsilon_n)}{\sum_{\alpha=1}^{K} p_n(\epsilon_1, \cdots, \epsilon_{n-1}, \alpha)}$$
$$= \exp (\nu^{(n-L-1)}_{\epsilon_{n-L} \cdots \epsilon_n}) \Big/ \sum_{\alpha=1}^{K} \exp (\nu^{(n-L-1)}_{\epsilon_{n-L} \cdots \epsilon_{n-1} \alpha}) \equiv f_n(\epsilon_{n-L}, \epsilon_{n-L+1}, \cdots, \epsilon_n),$$

since of the ν's in (60) only $\nu^{(n-L-1)}$ involves ϵ_n. Similarly, for each m satisfying $L + 1 \leqq m \leqq n$ we find from (60) that

$$\Pr \{X_m = x_{\epsilon_m} \mid X_1 = x_{\epsilon_1}, \cdots, X_{m-1} = x_{\epsilon_{m-1}}\}$$
$$= \frac{\Pr \{X_1 = x_{\epsilon_1}, \cdots, X_m = x_{\epsilon_m}\}}{\Pr \{X_1 = x_{\epsilon_1}, \cdots, X_{m-1} = x_{\epsilon_{m-1}}\}}$$
$$= f_m(\epsilon_{m-L}, \epsilon_{m-L+1}, \cdots, \epsilon_m), \qquad L + 1 \leqq m \leqq n, \tag{61}$$

that is, this conditional probability depends only on $\epsilon_{m-L}, \epsilon_{m-L+1}, \cdots, \epsilon_m$. Writing (61) in another form,

$$\Pr\{X_1 = x_{\epsilon_1}, \cdots, X_m = x_{\epsilon_m}\} = f_m(\epsilon_{m-L}, \cdots, \epsilon_m) \Pr\{X_1 = x_{\epsilon_1}, \cdots, X_{m-1} = x_{\epsilon_{m-1}}\}$$

and then summing on $\epsilon_1, \epsilon_2, \cdots, \epsilon_{m-L-1}$, we find that

$$\Pr\{X_{m-L} = x_{\epsilon_{m-L}}, \cdots, X_m = x_{\epsilon_m}\} = f_m(\epsilon_{m-L}, \cdots, \epsilon_m) \cdot \Pr\{X_{m-L} = x_{\epsilon_{m-L}}, \cdots, X_{m-1} = x_{\epsilon_{m-1}}\}. \tag{62}$$

We have then, substituting the value of f_m from (62) into (61),

$$\begin{aligned} \Pr\{X_m = x_{\epsilon_m} \mid X_1 = x_{\epsilon_1}, \cdots, X_{m-1} = x_{\epsilon_{m-1}}\} &= \frac{\Pr\{X_{m-L} = x_{\epsilon_{m-L}}, \cdots, X_m = x_{\epsilon_m}\}}{\Pr\{X_{m-L} = x_{\epsilon_{m-L}}, \cdots, X_{m-1} = x_{\epsilon_{m-1}}\}} \\ &= \Pr\{X_m = x_{\epsilon_m} \mid X_{m-1} = x_{\epsilon_{m-1}}, \cdots, X_{m-L} = x_{\epsilon_{m-L}}\}, \\ &\qquad L + 1 \leqq m \leqq n. \end{aligned} \tag{63}$$

Let us now define

$$p_L(\epsilon_1, \cdots, \epsilon_L) \equiv \sum_{\alpha=1}^{K} p_{L+1}(\epsilon_1, \cdots, \epsilon_L, \alpha) \tag{64}$$

$$q_L(\epsilon_{L+1} \mid \epsilon_1, \cdots, \epsilon_L) \equiv \frac{p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1})}{p_L(\epsilon_1, \cdots, \epsilon_L)}. \tag{65}$$

Repeated application of (63) then shows that

$$\begin{aligned} \Pr\{X_1 = x_{\epsilon_1}, \cdots, X_m = x_{\epsilon_m}\} &= p_L(\epsilon_1, \cdots, \epsilon_L) q(\epsilon_{L+1} \mid \epsilon_1, \cdots, \epsilon_L) \\ &\cdot q(\epsilon_{L+2} \mid \epsilon_2, \cdots, \epsilon_{L+1}) \cdots q(\epsilon_m \mid \epsilon_{m-L}, \epsilon_{m-L+1}, \cdots, \epsilon_{m-1}), \\ &\qquad m = L + 1, L + 2, \cdots, n. \end{aligned}$$

This expression is independent of n. Thus, among stationary processes with a given $(L + 1)$st-order distribution $p_{L+1}(\epsilon_1, \cdots, \epsilon_n)$, the Lth-order Markov process generated by the initial distribution (64) and the transition mechanism (65) has maximum entropy H_n for every $n \geqq L + 1$. Q.E.D.

APPENDIX B

We seek to maximize $H_{L+1} - H_L$ by choice of the $p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1})$ subject to the stationarity constraints

$$\sum_{\alpha} p_{L+1}(\epsilon_1, \cdots, \epsilon_L, \alpha) = \sum_{\alpha} p_{L+1}(\alpha, \epsilon_1, \cdots, \epsilon_L) \qquad \epsilon_1, \cdots, \epsilon_L = 1, \cdots, K, \tag{66}$$

the distribution constraint

$$\sum_{\epsilon} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) = 1, \tag{67}$$

and the covariance restrictions

$$\sum_{\epsilon} l_{L+1}^{(j)}(\epsilon_1, \cdots, \epsilon_{L+1}) p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) = (L + 1 - j)\rho_j \qquad j = 0, 1, \cdots, L \tag{68}$$

where the $l_{L+1}^{(j)}$ are defined in (19). The constraints (68) treat the variables in a more symmetric manner than do the constraints

$$\sum_{\epsilon} x_{\epsilon_1} x_{\epsilon_{k+1}} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) = \rho_k, \qquad k = 0, 1, \cdots, L. \tag{69}$$

It is easy to show that (66), (67), and (68) are equivalent to (66), (67), and (69).

Introducing Lagrange multipliers, we must maximize

$$\begin{aligned} J' = &- \sum_{\epsilon} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) \log p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) \\ &+ \sum_{\epsilon} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) \log [\sum_{\alpha} p_{L+1}(\epsilon_1, \cdots, \epsilon_L, \alpha)] \\ &+ \sum_{\epsilon} \nu_{\epsilon_1 \cdots \epsilon_L} [\sum_{\alpha} p_{L+1}(\epsilon_1, \cdots, \epsilon_L, \alpha) - \sum_{\alpha} p_{L+1}(\alpha, \epsilon_1, \cdots, \epsilon_L)] \\ &+ \mu \sum_{\epsilon} p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) \\ &+ \sum_{j=0}^{L} \lambda_j \sum_{\epsilon} l_{L+1}^{(j)}(\epsilon_1, \cdots, \epsilon_{L+1}) p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}). \end{aligned}$$

Differentiation with respect to $p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1})$ gives the necessary condition

$$\begin{aligned} -1 - \log p_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) + 1 + \log \sum_{\alpha} p_{L+1}(\epsilon_1, \cdots, \epsilon_L, \alpha) \\ + \nu_{\epsilon_1 \cdots \epsilon_L} - \nu_{\epsilon_2 \cdots \epsilon_{L+1}} + \mu + \sum_{j=0}^{L} \lambda_j l_{L+1}^{(j)}(\epsilon_1, \cdots, \epsilon_{L+1}) = 0. \end{aligned} \tag{70}$$

With the notation $c = e^{\mu}$, $f(\epsilon_1, \cdots, \epsilon_L) = e^{-\nu_{\epsilon_1 \cdots \epsilon_L}}$ and the definition (20), (70) becomes (22).

Inserting (22) in (14) yields (21). Again using (22) for q_L in (12) gives

$$\sum_{\epsilon_1} h_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}) \frac{p_L(\epsilon_1, \cdots, \epsilon_L)}{f(\epsilon_1, \cdots, \epsilon_L)} = \frac{1}{c} \frac{p_L(\epsilon_2, \cdots, \epsilon_{L+1})}{f(\epsilon_2, \cdots, \epsilon_{L+1})}$$
$$\epsilon_2, \cdots, \epsilon_{L+1} = 1, 2, \cdots, K,$$

where for simplicity we suppress the λ dependence of h. Relabeling variables, this can be written

$$\sum_{\epsilon_{L+1}} h_{L+1}(\epsilon_{L+1}, \cdots, \epsilon_1) \frac{p_L(\epsilon_{L+1}, \cdots, \epsilon_2)}{f(\epsilon_{L+1}, \cdots, \epsilon_2)} = \frac{1}{c} \frac{p_L(\epsilon_L, \cdots, \epsilon_1)}{f(\epsilon_L, \cdots, \epsilon_1)}$$
$$\epsilon_1, \cdots, \epsilon_L = 1, 2, \cdots, K. \quad (71)$$

But from the definition (20) and (19), $h_{L+1}(\epsilon_{L+1}, \cdots, \epsilon_1) = h_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1})$. Equation (71) is then

$$\sum_{\epsilon_{L+1}} h_{L+1}(\epsilon_1, \cdots, \epsilon_{L+1}; \lambda_0, \cdots, \lambda_L) \cdot \frac{p_L(\epsilon_{L+1}, \cdots, \epsilon_2)}{f(\epsilon_{L+1}, \cdots, \epsilon_2)} = \frac{1}{c} \frac{p_L(\epsilon_L, \cdots, \epsilon_1)}{f(\epsilon_L, \cdots, \epsilon_1)}.$$

Comparison with (21) now shows that we must have

$$\frac{p_L(\epsilon_L, \cdots, \epsilon_1)}{f(\epsilon_L, \cdots, \epsilon_1)} = kf(\epsilon_1, \cdots, \epsilon_L) \quad (72)$$

if the eigenvalue $1/c$ is not degenerate, which is the general case. A change of notation reduces (72) to (23).

APPENDIX C

We have considered stationary Lth-order Markov processes $\cdots X_{-1}, X_0, X_1, \cdots$ whose variables take values $x_1, x_2, \cdots, x_k$. The probability structure of such a process can be generated from transition probabilities $q_L(\epsilon_{L+1} \mid \epsilon_1, \cdots, \epsilon_L)$ via the mechanism of equations (12) through (16). Such a process can also be regarded as a function on the states of an ordinary Markov chain. The chain has K^L states, each one labeled by an Ltuple of integers $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_L)$. The conditional probability that the chain pass in one move from state $\boldsymbol{\alpha}$ to state $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\beta} \mid \boldsymbol{\alpha}) = q_L(\beta_L \mid \alpha_1, \alpha_2, \cdots, \alpha_L)\delta_{\alpha_2\beta_1}\delta_{\alpha_3\beta_2} \cdots \delta_{\alpha_L\beta_{L-1}} \quad (73)$$

where δ_{ij} is the Kronecker symbol. The chain thus has a very special nature. One can pass only to a state whose initial $L - 1$ labels agree with the last $L - 1$ labels of the state just left. The states correspond to successive Ltuples, $X_j, X_{j+1}, \cdots, X_{j+L-1}$, in the Lth-order Markov process. We regain that process from the chain by defining on the states of the chain the numerical valued function

$$X(\boldsymbol{\alpha}) = x_{\alpha_1} . \tag{74}$$

From well-known results in the theory of finite Markov chains (see for example Ref. 10, Chapt. 16, Sec. 1), we see that the probability of going from state $\boldsymbol{\alpha}$ to state $\boldsymbol{\beta}$ in exactly n moves can be written in the form

$$p^{(n)}(\boldsymbol{\beta} \mid \boldsymbol{\alpha}) = \sum_{j=1}^{K^L} u_{\boldsymbol{\alpha}}^{(j)} v_{\boldsymbol{\beta}}^{(j)} \lambda_j^n \qquad n = 1, 2, \cdots$$

$$\alpha_1, \alpha_2, \cdots, \alpha_L, \beta_1, \cdots, \beta_L = 1, 2, \cdots, K. \tag{75}$$

Here the u's and v's are left and right eigenvectors of $p(\boldsymbol{\beta} \mid \boldsymbol{\alpha})$,

$$\sum_{\boldsymbol{\beta}} u_{\boldsymbol{\beta}}^{(j)} p(\boldsymbol{\beta} \mid \boldsymbol{\alpha}) = \theta_j u_{\boldsymbol{\alpha}}^{(j)} \qquad \sum_{\boldsymbol{\beta}} p(\boldsymbol{\alpha} \mid \boldsymbol{\beta}) v_{\boldsymbol{\beta}}^{(j)} = \theta_j v_{\boldsymbol{\alpha}}^{(j)}$$

$$j = 1, 2, \cdots, K^L, \quad \alpha_1, \alpha_2, \cdots, \alpha_L = 1, 2, \cdots, K,$$

normalized so that

$$\sum_{j=1}^{K^L} u_{\boldsymbol{\alpha}}^{(j)} v_{\boldsymbol{\beta}}^{(j)} = \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}} .$$

Note that this gives (75) the special value

$$p^{(0)}(\beta \mid \alpha) = \delta_{\boldsymbol{\alpha}\boldsymbol{\beta}} .$$

In terms of this Markov chain, the covariance of the X process can be written

$$\rho_n = EX_jX_{j+n} = \sum_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} x_{\alpha_1} x_{\beta_1} p(\boldsymbol{\alpha}) p^{(n)}(\boldsymbol{\beta} \mid \boldsymbol{\alpha}) \qquad n = 0, 1, 2, \cdots \tag{76}$$

where $p(\boldsymbol{\alpha})$ is the stationary distribution for the chain, i.e.,

$$\sum_{\boldsymbol{\alpha}} p(\boldsymbol{\alpha}) p(\boldsymbol{\beta} \mid \boldsymbol{\alpha}) = p(\boldsymbol{\beta}).$$

Using (75) in (76) we have the desired result

$$\rho_n = \sum_{j=1}^{K^L} A_j \theta^n, \qquad n = 0, 1, 2, \cdots$$

where

$$A_j = \sum_{\alpha} x_{\alpha_1} p(\boldsymbol{\alpha}) u_{\alpha}^{(j)} \sum_{\beta} x_{\beta_1} v_{\beta}^{(j)} .$$

APPENDIX D

For any discrete stationary process taking values $x_1 , x_2 , \cdots , x_k$, the truncated covariance sequence can be formed from the $(L + 1)$st-order distribution $p_{L+1}(\epsilon_1 , \cdots , \epsilon_{L+1})$. Thus

$$\rho_j = \sum_{\epsilon} x_{\epsilon_1} x_{\epsilon_{j+1}} p_{L+1}(\epsilon_1 , \cdots , \epsilon_{L+1}) \qquad j = 0, 1, \cdots , L. \tag{77}$$

Such a distribution satisfies the stationarity conditions

$$\sum_{\alpha} p_{L+1}(\alpha, \epsilon_2 , \cdots , \epsilon_{L+1}) = \sum_{\alpha} p_{L+1}(\epsilon_2 , \cdots , \epsilon_{L+1} , \alpha)$$

$$\epsilon_2 , \epsilon_3 , \cdots , \epsilon_{L+1} = 1, 2, \cdots , K, \tag{78}$$

the constraint

$$\sum_{\epsilon} p_{L+1}(\epsilon_1 , \cdots , \epsilon_{L+1}) = 1, \tag{79}$$

and the inequalities

$$p_{L+1}(\epsilon_1 , \cdots , \epsilon_{L+1}) \geqq 0 \qquad \epsilon_1 , \epsilon_2 , \cdots , \epsilon_{L+1} = 1, 2, \cdots , K. \tag{80}$$

Conversely, from K^{L+1} quantities $p_{L+1}(\epsilon_1 , \cdots , \epsilon_{L+1})$ satisfying (78) through (80) we can construct a discrete stationary (Lth-order Markov) process having values $x_1 , \cdots , x_K$ and truncated covariance given by (77). To do so, define

$$p_L(\epsilon_1 , \cdots , \epsilon_L) \equiv \sum_{\epsilon_{L+1}} p_{L+1}(\epsilon_1 , \cdots , \epsilon_{L+1})$$

$$\epsilon_1 , \epsilon_2 , \cdots , \epsilon_L = 1, 2, \cdots , K.$$

Let

$$q_L(\epsilon_{L+1} \mid \epsilon_1 , \cdots , \epsilon_L) = \frac{p_{L+1}(\epsilon_1 , \cdots , \epsilon_{L+1})}{p_L(\epsilon_1 , \cdots , \epsilon_L)}.$$

It is easy to verify that (12), (13), and (14) are satisfied, so that the measure described by (15) and (16) defines the desired process.

Thus equations (77) through (80) serve to define parametrically the region $\mathcal{R}$ of admissible truncated covariance sequences. Consider an $(L + 1)$st-order density $p_{L+1}(\epsilon_1 , \cdots , \epsilon_{L+1})$ as a point in a Euclidean space $\mathcal{E}_{K^{L+1}}$ of dimension K^{L+1}. Equations (78), (79), and (80) define a convex region $\mathcal{V}$ in this space that is bounded by no more than $K^L + 1 + K^{L+1} \leqq 2K^{L+1}$ hyperplanes ($K \geqq 2$). Equations (77) provide a linear mapping of $\mathcal{E}_{KL+1}$ into $\mathcal{E}_{L+1}$ and in particular the image of $\mathcal{V}$ is $\mathcal{R}$. The hyperplane boundaries of $\mathcal{V}$ map into hyperplanes in $\mathcal{E}_{L+1}$ that include all the hyperplane boundaries of $\mathcal{R}$. Q.E.D.

REFERENCES

1. Kobayashi, H., "Coding Schemes for Reduction of Intersymbol Interference in Data Transmission Systems," IBM J. Res. & Dev., *14*, No. 4 (July 1970), pp. 343–353.
2. Croisier, A., "Introduction to Pseudoternary Transmission Codes," IBM J. Res. & Dev., *14*, No. 4 (July 1970), pp. 354–367.
3. Kobayashi, H., and Tang, D. T., "Application of Partial-Response Channel Coding to Magnetic Recording Systems," IBM J. Res. & Dev., *14*, No. 4 (July 1970), pp. 368–375.
4. Franaszek, P. A., "Sequence-State Methods for Run-Length-Limited Coding," IBM J. Res. & Dev., *14*, No. 4 (July 1970), pp. 376–383.
5. van de Waerden, B. L., *Modern Algebra*, New York: Frederick Ungar Pub. Co., 1953, Vol. 1, pp. 84–85.
6. Loève, M., *Probability Theory*, New York: D. Van Nostrand Co., 1960, p. 207.
7. McMillan, B., "History of a Problem," J. Soc. Ind. and Appl. Math., *3*, (1955), pp. 114–128.
8. Shepp, L. A., "Covariances of Unit Processes," Unpublished Conference Notes, Working Conference on Stochastic Processes, IDA, Santa Barbara, Calif., 1969.
9. Gallager, R., G. *Information Theory and Reliable Communication*, New York: John Wiley & Son, 1968.
10. Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, New York: John Wiley & Son, 1957.