The Annals of Probability
1977, Vol. 5, No. 6, 966-986

SOURCE CODING THEOREMS FOR STATIONARY, CONTINUOUS-TIME STOCHASTIC PROCESSES

By Michael B. Pursley[1] and Robert M. Gray[2]

University of Illinois and Stanford University

New results are presented on the problem of source coding subject to a fidelity criterion for stationary, measurable, continuous-time stochastic processes. The main result is established without an ergodicity requirement and for quite general reproduction alphabets. This is then specialized to ergodic sources to prove a coding theorem for ergodic continuous-time sources. The approach is to obtain coding theorems for continuous-time sources from the coding theorem for nonergodic discrete-time sources.

1. Introduction. The two basic theorems of information theory—the channel coding theorem and the theorem on source coding subject to a fidelity criterion—have been proved for quite general discrete-time sources and channels. In the case of continuous-time systems, however, quite general channel coding theorems exist [8], but coding theorems for stationary sources have appeared only for the special cases of Gaussian processes with average squared-error distortion measure and block ergodic processes [1]. The principal obstacle to extending the source coding theorem to stationary ergodic continuous-time sources has been the fact that if one segments an ergodic continuous-time random process into functions on intervals of length τ, the resulting processes may not be ergodic. For nonergodic continuous-time sources, the segmented source is nonergodic for all values of τ. The purpose of this paper is to demonstrate that since the segmented source is stationary even if not ergodic, the source coding theorem for stationary nonergodic discrete-time sources [6] can be applied to obtain a coding theorem for stationary continuous-time sources. We then use this result to establish a coding theorem for stationary ergodic continuous-time sources.

The most general source coding problem that we consider is the coding of a continuous-time, stationary, measurable stochastic process with values in a complete separable metric space consisting of a set A and a metric ρ on A. The metric ρ also serves as the measure of distortion between elements of the source alphabet A and elements of the reproduction alphabet $\hat{A} \subset A$.

An example of the type of stationary, ergodic source that we have in mind is a source whose output is modeled as a real-valued, stationary, measurable random process such as the Ornstein–Uhlenbeck process. If, as in this example,

Received February 27, 1975; revised February 1, 1977.

[1] Research supported in part by NSF Grant ENG 75-20864.

[2] Research supported in part by NSF Grant ENG 76-02276.

AMS 1970 *subject classifications.* Primary 94A15; Secondary 60G10, 60G35, 94A05.

Key words and phrases. Source coding subject to a fidelity criterion, information theory, continuous-time sources.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Probability.* www.jstor.org

$A = \hat{A} = \mathscr{R}$, the real line, then the Euclidean distance metric $\rho^{(1)}(u, v) = |u - v|$ is an appropriate distortion measure.

For an example that illustrates the need for considering random processes which take values in a metric space, consider the problem of transmitting time-varying pictorial information over a digital communication system. In this situation we could take the value of the process at a given instant of time to be a real valued function of two variables. The value of the function is intensity or gray-level at a point on the black-and-white two-dimensional image. The alphabet A for such a process could be taken as the space of all nonnegative integrable functions on $[0, 1] \times [0, 1]$ (since the integral of such a function corresponds physically to the total amount of light from the picture received by an observer). It is of interest to determine the minimum possible distortion that can be achieved when this process is transmitted at rate R over a digital communication channel. We will give conditions under which this minimum distortion can be calculated from the probabilistic description of the process in terms of an information-theoretic minimization.

2. Notation and preliminaries. The following notation will be used throughout the paper. The set of all real numbers is denoted by $\mathscr{R}$ and the extended real line is denoted by $\bar{\mathscr{R}}$. $\mathscr{L}$ is the σ-field of Lebesgue measurable subsets of $\mathscr{R}$ and $\mathscr{L}_I$ is the σ-field of Lebesgue measurable subsets of the finite interval I. The Lebesgue measure is denoted by m. The sets A and $\hat{A}$ denote the source and reproduction alphabets which are metric spaces. The source alphabet A is a complete separable metric space under the metric ρ and $\hat{A}$ is a Borel subset of A. The Borel σ-field of subsets of A and of $\hat{A}$ are denoted by $\mathscr{B}_A$ and $\mathscr{B}_{\hat{A}}$, respectively. Let Ω be the function space $A^{\mathscr{R}}$; that is, an element $\omega \in \Omega$ is a function $\omega: \mathscr{R} \to A$. For each $\tau \in \mathscr{R}$ let T^τ be the τ-unit shift transformation on Ω; that is, $[T^\tau\omega](t) = \omega(t + \tau)$ for each $\omega \in \Omega$ and $t \in \mathscr{R}$. Let $(\Omega, \mathscr{F}, \mu)$ be a complete probability space upon which T^τ is a measure-preserving transformation (i.e., $F \in \mathscr{F}$ implies $T^\tau F \in \mathscr{F}$ and $\mu(F) = \mu(T^\tau F)$) for each $\tau \in \mathscr{R}$. In the terminology of ergodic theory, $\{T^\tau \mid \tau \in \mathscr{R}\}$ is a flow. We define the A-valued stochastic process $\{X_t \mid t \in \mathscr{R}\}$ on the probability space $(\Omega, \mathscr{F}, \mu)$ by $X_t(\omega) = \omega(t)$ for each $\omega \in \Omega$ and each $t \in \mathscr{R}$. We assume that X_t is a measurable mapping from $(\Omega, \mathscr{F})$ into $(A, \mathscr{B}_A)$ for each $t \in \mathscr{R}$. Note that according to our definitions, $X_t(T^\tau(\omega)) = X_{t+\tau}(\omega)$. We refer to the flow $\{T^\tau \mid \tau \in \mathscr{R}\}$ and stochastic process $\{X_t \mid t \in \mathscr{R}\}$ on $(\Omega, \mathscr{F}, \mu)$ as a *continuous-time source*. Since the continuous-time source can be specified by giving only the σ-field $\mathscr{F}$ and the probability measure μ, we usually denote the source by $(\Omega, \mathscr{F}, \mu)$. Let $\overline{\mathscr{F} \times \mathscr{L}}$ be the completion of the σ-field $\mathscr{F} \times \mathscr{L}$ with respect to the product measure $\mu \times m$. The restriction of $\omega \in \Omega$ to the finite interval I is denoted by ω^I and A^I is the function space consisting of all mappings $x: I \to A$. The set of all Lebesgue integrable functions $y: I \to \mathscr{R}$ is denoted by $L_1(I)$. The space of all functions $y: I \to \mathscr{R}$ for which y^p is Lebesgue integrable $(p \geqq 1)$ will be denoted by $L_p(I)$. Finally, Z denotes the set of all integers.

The source must satisfy certain additional requirements which will be introduced as needed. The first of these is the following measurability property.

(P.1) The random process $\{X_t \mid t \in \mathscr{R}\}$ is measurable; that is, if $X_t(\omega)$ is viewed as a function of the pair (ω, t), it is a measurable mapping from $(\Omega \times \mathscr{R}, \overline{\mathscr{F} \times \mathscr{L}})$ into $(A, \mathscr{B}_A)$.

Property (P.1) implies that for μ-almost all $\omega \in \Omega$, $X_t(\omega)$ is a Lebesgue measurable function of t. This guarantees that the set Ω_M of all functions $\omega : \mathscr{R} \to A$ which are measurable mappings of $(\mathscr{R}, \mathscr{L})$ into (A, B_A) has μ-measure 1. This will in turn assure that a suitable distortion measure can be defined on the function space A^I. We note that condition (P.1) is not very restrictive in the sense that under quite general conditions, there is a standard extension of a process of the function space type which is measurable and also of the function space type [4].

A set $E \in \mathscr{F}$ is said to be τ-invariant if $\mu(E \triangle T^\tau E) = 0$, where $A \triangle B$ is the symmetric difference $(A \cap B^C) \cup (A^C \cap B)$. A set $E \in \mathscr{F}$ is said to be invariant if it is τ-invariant for each $\tau \in \mathscr{R}$. The flow $\{T^\tau \mid \tau \in \mathscr{R}\}$ is ergodic (or metrically transitive) if every invariant set $F \in \mathscr{F}$ satisfies $\mu(F) \in \{0, 1\}$ and it is τ-ergodic if every τ-invariant set $G \in \mathscr{F}$ satisfies $\mu(G) \in \{0, 1\}$. The flow is said to be totally ergodic (or block ergodic) if it is τ-ergodic for each $\tau \in \mathscr{R}$. In general, an ergodic flow will not be totally ergodic. We will say the source $(\Omega, \mathscr{F}, \mu)$ is ergodic (τ-ergodic, totally ergodic) if the flow $\{T^\tau \mid \tau \in \mathscr{R}\}$ is ergodic (τ-ergodic, totally ergodic).

Given two spaces $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$, a transition probability ([1], [14]) is a mapping $P_2 : \Omega_1 \times \mathscr{F}_2 \to [0, 1]$ for which $P_2(\omega, \cdot)$ is a probability measure on $(\Omega_2, \mathscr{F}_2)$ for each $\omega \in \Omega_1$ and $P_2(\cdot, F)$ is a measurable function on $(\Omega_1, \mathscr{F}_1)$ for each $F \in \mathscr{F}_2$. The term regular conditional probability is often used for such mappings (e.g., [11]).

A partition $\mathscr{P}$ of a measurable space $(\Omega, \mathscr{F})$ is a finite collection of disjoint sets in $\mathscr{F}$ such that the union of all sets in $\mathscr{P}$ is Ω. We define the quantity

$$\mathscr{H}(P, Q \mid \mathscr{P}) = \sum_{B \in \mathscr{P}} P(B) \log [P(B)/Q(B)]$$

where P and Q are two probability measures on the same space $(\Omega, \mathscr{F})$ and $\mathscr{P}$ is a partition of $(\Omega, \mathscr{F})$. As usual, we must set $0 \log (0/u) = 0$ for all $u \geqq 0$ and $u \log (u/0) = +\infty$ for all $u > 0$ in such definitions. The entropy of P relative to Q is then defined by

$$H(P, Q) = \sup_{\mathscr{P}} \mathscr{H}(P, Q/\mathscr{P})$$

where the supremum is over all finite partitions of $(\Omega, \mathscr{F})$. According to Dobrushin's theorem (see [3], Section 2.2 or [17], Theorems 2.1.1 and 2.4.1), if $\mathscr{M}$ is any algebra of subsets of Ω which generates $\mathscr{F}$ then $H(P, Q)$ is equal to the supremum of $\mathscr{H}(P, Q \mid \mathscr{P}')$ over all partitions $\mathscr{P}' \subset \mathscr{M}$. The basic properties of relative entropy (and of mutual information, since it is a special case of relative entropy) can be found in [3], [5], [10], [16], and [17]. An important property that we will need in the sequel is given in the following lemma.

LEMMA 1. *Suppose a probability measure P is defined on $\times_{i=1}^{3}(\Omega_i, \mathscr{F}_i)$ by the iterated integral*

$$P(B) = \int_{\Omega_1}\int_{\Omega_2}\int_{\Omega_3} I_B(\omega_1, \omega_2, \omega_3)P_3(\omega_2, d\omega_3)P_2(\omega_1, d\omega_2)P_1(d\omega_1)$$

where P_1 is a probability measure on $(\Omega_1, \mathscr{F}_1)$; P_2 and P_3 are transition probabilities on $\Omega_1 \times \mathscr{F}_2$ and $\Omega_2 \times \mathscr{F}_3$, respectively; and I_B is the indicator function for the set $B \in \mathscr{F}_1 \times \mathscr{F}_2 \times \mathscr{F}_3$. Assume that $\mathscr{F}_2 \times \mathscr{F}_3$ is generated by a countable algebra $\mathscr{M}$ of subsets of $\Omega_2 \times \Omega_3$. For each $\omega \in \Omega_1$, define p_ω on $\times_{i=2}^{3}(\Omega_i, \mathscr{F}_i)$ by

$$p_\omega(C) = \int_{\Omega_2}\int_{\Omega_3} I_C(\omega_2, \omega_3)P_3(\omega_2, d\omega_3)P_2(\omega, d\omega_2)$$

for each $C \in \mathscr{F}_2 \times \mathscr{F}_3$, and let $\bar{p}_\omega$ be the measure on $\times_{i=2}^{3}(\Omega_i, \mathscr{F}_i)$ for which

$$\bar{p}_\omega(F_2 \times F_3) = p_\omega(F_2 \times \Omega_3)p_\omega(\Omega_2 \times F_3)$$

for $F_i \in \mathscr{F}_i$ $(i = 2, 3)$. Let $\bar{P}$ be the measure on $\times_{i=1}^{3}(\Omega_i, \mathscr{F}_i)$ which satisfies

$$\bar{P}(F_1 \times F_2 \times F_3) = \int_{F_1} \bar{p}_\omega(F_2 \times F_3)P_1(d\omega)$$

for $F_i \in \mathscr{F}_i$ $(i = 1, 2, 3)$. Finally, Q and $\bar{Q}$ are the measures on $\times_{i=2}^{3}(\Omega_i, \mathscr{F}_i)$ which satisfy

$$Q(F_2 \times F_3) = P(\Omega_1 \times F_2 \times F_3)$$

and

$$\bar{Q}(F_2 \times F_3) = Q(F_2 \times \Omega_3)Q(\Omega_2 \times F_3)$$

for each $F_i \in \mathscr{F}_i$ $(i = 2, 3)$. Then

$$H(P, \bar{P}) = \int_{\Omega_1} H(p_\omega, \bar{p}_\omega)P_1(d\omega) \tag{1}$$

$$\leqq H(Q, \bar{Q})\,. \tag{2}$$

PROOF. The proof of (1) will be omitted since it can be obtained in a straightforward manner from well-known results on conditional information (e.g., [17], Chapter 3, especially (3.1.5) on page 30). To prove (2) we use the Markov structure of probability space $(\times_{i=1}^{3}\Omega_i, \times_{i=1}^{3}\mathscr{F}_i, P)$ which is derived from the fact that $P_3(\omega_2, F_3)$ does not depend on ω_1. We can assume that $H(Q, \bar{Q}) < \infty$ in proving (2) since it is trivally true for $H(Q, \bar{Q}) = \infty$. The Markov structure implies $H(Q, \bar{Q}) = H(P, P')$ where P' is defined on $\times_{i=1}^{3}(\Omega_i, \mathscr{F}_i)$ by

$$P'(F_1 \times F_2 \times F_3) = P(F_1 \times F_2 \times \Omega_3)P(\Omega_1 \times \Omega_2 \times F_3)$$

for $F_i \in \mathscr{F}_i$ $(i = 1, 2, 3)$. For the proof of this result see Section 2.8 of [3]. Thus $H(P, P') < \infty$ and so $P' \ll P$. But this implies $H(P, P') \geqq H(P, \bar{P})$ (see [3], Section 2.7 or [17], Theorem 3.6.1) which completes the proof.

3. Properties of the time-average distortion measure. The purpose of source coding is to "compress" a source output process by encoding it with elements from a finite code book with least possible distortion. Let $\hat{B} \subset \hat{A}^{\mathscr{R}}$ be the set of allowable reproducing functions $\xi : \mathscr{R} \to \hat{A}$. We require that each $\xi \in \hat{B}$ is a measurable mapping of $(\mathscr{R}, \mathscr{L})$ into $(\hat{A}, \mathscr{B}_{\hat{A}})$. Define $\hat{B}(I)$ to be the set of all functions which are restrictions of functions in $\hat{B}$ to the interval $I \subset \mathscr{R}$. Later, further restrictions will be placed on $\hat{B}$ to ensure that its members are sufficiently well behaved. For arbitrary $\lambda > 0$, a λ-length code book C_λ is a collection of finitely many elements of $\hat{B}([0, \lambda))$. Included in all such code books is the reference function $a_\lambda : [0, \lambda) \to \hat{A}$ defined by $a_\lambda(t) = a_0$ for each t in the interval $[0, \lambda)$, where a_0 is an arbitrary reference letter in the reproducing alphabet $\hat{A}$. Similarly, we define $a_I : I \to \hat{A}$ by $a_I(t) = a_0$ for each $t \in I$ and $\hat{a} : \mathscr{R} \to \hat{A}$ by $\hat{a}(t) = a_0$ for each t. We require $\hat{a} \in \hat{B}$. This implies that $a_\lambda \in \hat{B}([0, \lambda))$ for all $\lambda > 0$. If $||C_\lambda||$ denotes the total number of elements in C_λ then the rate of this code book is $(\lambda^{-1}) \log ||C_\lambda||$.

For any finite interval I we define a distortion measure $\tilde{\rho}_I$ between functions in Ω_M by letting

$$\tilde{\rho}_I(\omega, \zeta) = [m(I)]^{-1} \int_I \rho[\omega(t), \zeta(t)] m(dt) \tag{3}$$

for all $\omega, \zeta \in \Omega_M$. In this definition we are allowing extended real values for the integral since ρ is in general not bounded. Although there will be functions $\omega \in \Omega$ which are not measurable mappings of $(\mathscr{R}, \mathscr{L})$ into $(A, \mathscr{B}_A)$, the set of all such functions must have μ-measure zero according to (P.1). We can then extend $\tilde{\rho}_I$ to $\Omega \times \Omega$ by defining $\tilde{\rho}_I(\omega, \zeta) = +\infty$ if $(\omega, \zeta) \notin \Omega_M \times \Omega_M$. Note that $\mathscr{B}_{\hat{A}} \subset \mathscr{B}_A$ and any $\xi \in \hat{B}$ is a measurable mapping from $(\mathscr{R}, \mathscr{L})$ into $(A, \mathscr{B}_A)$. Measurability of the integrand of (3) follows from continuity of $\rho : A \times A \to \mathscr{R}$ and measurability of ω and ζ.

Note that (3) can be used to define a distortion measure ρ_I between elements of A^I. If x and y are such elements then set

$$\rho_I(x, y) = \tilde{\rho}_I(\omega_x, \omega_y) \tag{4}$$

where $x(t) = \omega_x(t)$ and $y(t) = \omega_y(t)$ for all $t \in I$ and $\omega_x(t) = \omega_y(t) = a_0$ otherwise. The distortion measure ρ_I is not a metric for A^I since $\rho_I(x, y) = 0$ does not imply $x = y$ and since $\rho_I(x, y)$ need not be finite. However, ρ_I is nonnegative, satisfies the triangle inequality, and is symmetric. We will refer to either ρ_I or $\tilde{\rho}_I$ as the time-average distortion measure. In case $I = [0, \lambda)$ for $\lambda > 0$ we denote ρ_I and $\tilde{\rho}_I$ by ρ_λ and $\tilde{\rho}_\lambda$, respectively.

In order to have a meaningful source coding problem for the process $\{X_t \mid t \in \mathscr{R}\}$ there must exist a reference letter $a_0 \in \hat{A}$ such that

$$\int_\Omega \rho[X_t(\omega), a_0] \mu(d\omega) = \rho^* < \infty . \tag{5}$$

Since $\rho : A \times A \to \mathscr{R}$ is continuous, the integrand in (5) is a measurable mapping of $(\Omega, \mathscr{F})$ into $(\mathscr{R}, \mathscr{L})$ for each $t \in \mathscr{R}$. We have denoted the value of the integral in (5) by ρ^* which does not depend on t since the process is stationary. Note that

$$\rho^* = \int_\Omega \rho[\omega(t), a_0] \mu(d\omega) \quad \forall\, t \in \mathscr{R} \tag{6}$$

and that (5) is valid if and only if the process satisfies the following property.

(P.2) For any element $a \in A$,

$$\int_\Omega \rho[X_t(\omega), a] \mu(d\omega) < \infty . \tag{7}$$

If, for example, $A = \mathscr{R}$ and $\rho = \rho^{(1)}$ then (P.2) is equivalent to the requirement that the mean of $\{X_t \mid t \in \mathscr{R}\}$ is finite. If we choose $a_0 = 0$ for this example then we have $\rho^* = E|X_t|$, where E denotes expected value with respect to the probability measure μ. If (P.2) is not true, then the average distortion will be infinite for any code of any (finite) rate. Thus (P.2) is a necessary condition for a meaningful source coding problem. The stochastic process $\{Y_t \mid t \in \mathscr{R}\}$ defined by

$$Y_t(\omega) = \rho[X_t(\omega), a_0] \qquad \forall\, \omega \in \Omega\,, \quad \forall\, t \in \mathscr{R}$$

is a nonnegative real-valued random process which is measurable because of (P.1) and the continuity of the function $\rho(\cdot, a_0): A \to \mathscr{R}$. It follows from Fubini's theorem that $\int_\Omega \rho[X_t(\omega), a_0]\mu(d\omega)$ is a measurable function of t and

$$\int_\Omega \{\int_I \rho[X_t(\omega), a_0]m(dt)\}\mu(d\omega) = \rho^* m(I) < \infty\,. \tag{8}$$

For any finite interval I let $B(I)$ be the set of all measurable mappings x from $(I, \mathscr{L}_I)$ into $(A, \mathscr{B}_A)$ for which $\rho_I(x, a_I) < \infty$. For the special case $A = \mathscr{R}$ and $\rho = \rho^{(1)}$ we have $B(I) = L_1(I)$. For any bounded metric ρ which generates the usual Euclidean topology on $\mathscr{R}$, $B(I)$ is the set of all Lebesgue measurable functions on I.

In order to apply the discrete-time source coding theorem of Gray and Davisson we need a suitable metric space to serve as the alphabet for the discrete-time source which is obtained by segmenting sample functions of the original continuous-time source. The next lemma shows that the set $B(I)$ is a candidate for this alphabet if ρ_I is the distortion measure.

LEMMA 2. *The distortion measure ρ_I is pseudometric for the set $B(I)$. The resulting pseudometric space is complete and separable.*

PROOF. Clearly $\rho_I(y, y) = 0$ for each $y \in B(I)$. The definition of $B(I)$ guarantees that $\rho_I(x, y) < \infty$ for all $x, y \in B(I)$. The triangle inequality for ρ_I follows easily from the triangle inequality for ρ. The proofs of separability and completeness closely parallel the corresponding proofs for $L_p([0, \lambda))$, so we omit the details.

Although ρ_I is not a metric for $B(I)$, we can consider equivalence classes of functions, just as in the case of $L_1(I)$, to obtain an appropriate metric space. Let $\mathscr{B}(I)$ denote the σ-field of Borel subsets of this metric space. In case $I = [0, \lambda)$, we use the notation $(B_\lambda, \rho_\lambda)$ for the metric space and $\mathscr{B}_\lambda$ for its Borel σ-field.

LEMMA 3. *If I is a finite interval and $y \in B(I)$,*

$$\int_I \rho[X_t(\omega), y(t)]m(dt) < \infty \tag{9}$$

for μ-almost all ω. If $S: \Omega \to \bar{\mathscr{R}}$ is defined by letting $S(\omega)$ equal the integral in (9) when $\omega \in \Omega_M$ and letting $S(\omega) = +\infty$ otherwise, then S is an extended-real-valued random variable on $(\Omega, \mathscr{F}, \mu)$ having a finite expectation ES.

PROOF. We first note that the random process $\{V_t \mid t \in \mathscr{R}\}$ defined by

$$V_t(\omega) = \rho[X_t(\omega), y(t)] \qquad \forall\, \omega \in \Omega\,, \quad \forall\, t \in \mathscr{R}$$

is a measurable random process. This follows from (P.1), the definition of $B(I)$, and the continuity of $\rho\colon A \times A \to \mathscr{R}$ with respect to the product topology on $A \times A$. In particular $V_t(\omega)$ is a measurable function of t for all ω in the set Ω_M. Therefore, (9) follows from (8), the definition of $B(I)$, and the triangle inequality. The finiteness of ES then follows from (5), the fact that $y \in B(I)$, and Fubini's theorem.

LEMMA 4. *For each $H \in \mathscr{B}(I)$,*

$$\{\omega \in \Omega_M \mid \omega^I \in H\} \in \mathscr{F}\,.$$

PROOF. It suffices to consider sets H which are open balls in $(B(I), \rho_I)$. Let $\varepsilon > 0$ and $y \in B(I)$ be arbitrary and let $H = \{x \in B(I) \mid \rho_I(x, y) < \varepsilon\}$. The proof is completed by noting that

$$\{\omega \in \Omega_M \mid \omega^I \in H\} = \{\omega \in \Omega_M \mid \textstyle\int_I \rho(\omega(t), y(t))m(dt) < \varepsilon m(I)\}$$

which is $\mathscr{F}$-measurable by Lemma 3.

LEMMA 5. *The set Σ of all sample functions $\omega \in \Omega$ for which $\omega^I \in B(I)$ for each finite interval I satisfies $\Sigma \subset \Omega_M$, $\Sigma \in \mathscr{F}$, and $\mu(\Sigma) = 1$.*

PROOF. That $\Sigma \subset \Omega_M$ follows from the fact that for any open subset G contained in A, $\omega^{-1}(G) = \bigcup_{i=-\infty}^{\infty} \omega_i^{-1}(G)$ where $\omega_i = \omega^{[i\tau,(i+1)\tau)}$. Hence $\omega \in \Sigma$ implies $\omega_i \in B([i\tau, (i+1)\tau))$ for each $i \in Z$ which implies $\omega \in \Omega_M$. For each positive integer j let

$$\Omega_{M,j} = \{\omega \in \Omega_M \mid \textstyle\int_{[-j,j]} \rho[X_t(\omega), a_0]m(dt) < \infty\}$$

and note that $\Sigma = \bigcap_{j=1}^{\infty} \Omega_{M,j}$. Lemma 3 guarantees that $\Omega_{M,j} \in \mathscr{F}$ for each j so $\Sigma \in \mathscr{F}$. According to (8),

$$\textstyle\int_{[-j,j]} \rho[X_t(\omega), a_0]m(dt) < \infty$$

for μ-almost all ω so $\mu(\Omega_{M,j}) = 1$ for each j. Thus $\mu(\Sigma) = 1$.

An important property of the set Σ is that $\omega \in \Sigma$ if and only if $\omega^{[i\tau,(i+1)\tau)} \in B([i\tau, (i+1)\tau))$ for each $i \in Z$. Thus, Σ can be regarded as the set of all doubly infinite sequences $\cdots, \omega_{-1}, \omega_0, \omega_1, \cdots$ where $\omega_i = \omega^{[i\tau,(i+1)\tau)}$ and we can write

$$\Sigma = \times_{i=-\infty}^{\infty} B([i\tau, (i+1)\tau))\,.$$

Since each $x \in B([0, \tau))$ has a unique replica $x_i \in B([i\tau, (i+1)\tau))$ defined by $x_i(t) = x(t - i\tau)$ for $t \in [i\tau, (i+1)\tau)$, then we can also view Σ as the set B_τ^Z of all doubly infinite sequences of elements of B_τ. It is important to note, however, that Σ does not depend on τ since $\omega^{[i\omega,(i+1)\tau)} \in B([i\tau, (i+1)\tau)$ for all i if and only if for each $\lambda > 0$, $\omega^{[i\lambda,(i+1)\lambda)} \in B([i\lambda, (i+1)\lambda))$ for all i.

The only restriction that we have placed on $\hat{B}$ thus far is that the reference function $\hat{a}\colon \mathscr{R} \to \hat{A}$ defined in Section 2 must be an element of $\hat{B}$ and that each

$\xi \in \hat{B}$ must be a measurable mapping from $(\mathscr{R}, \mathscr{L})$ into $(\hat{A}, \mathscr{B}_{\hat{A}})$. We must now add two requirements. The first requirement is as follows.

(P.3) The set $\hat{B}$ is such that

$$\hat{B}([0, \lambda)) \subset B([0, \lambda)) \quad \forall \lambda > 0 . \tag{10}$$

For instance, $\hat{B}$ might be the set of all measurable mappings $\xi : (\mathscr{R}, \mathscr{L}) \to (\hat{A}, \mathscr{B}_{\hat{A}})$ which satisfy

$$\int_I \rho(\xi(t), a_0)m(dt) < \infty$$

for all finite intervals I. For this example, any $\xi \in \hat{B}$ will automatically be a measurable mapping of $(\mathscr{R}, \mathscr{L})$ into $(A, \mathscr{B}_A)$ so that $\hat{B}([0, \lambda)) \subset B([0, \lambda))$; furthermore, if $A = \hat{A}$ then $\hat{B}([0, \lambda)) = B([0, \lambda))$ for all $\lambda > 0$. Since $B([0, \lambda))$ represents functions which must be reproduced at the decoder, a more reasonable choice for $\hat{B}$ would be the set of all measurable mappings ξ for which the restriction ξ^I is a simple function for each finite interval I. If $\hat{A}$ is dense in A, then for this choice of $\hat{B}$, $\hat{B}([0, \lambda))$ will be dense in $B([0, \lambda))$. For any $\hat{B}$ satisfying (10), Lemma 2 guarantees that $\hat{B}([0, \lambda))$ is a separable pseudometric space. The second requirement is that $\hat{B}$ have the following property.

(P.4) For any real numbers λ and τ such that $0 < \tau < \lambda$,

$$\hat{B}([0, \lambda)) = \hat{B}([0, \tau)) \times \hat{B}([\tau, \lambda)) .$$

That is, $y \in \hat{B}([0, \lambda))$ if and only if y can be regarded as a pair of functions (y_1, y_2) such that $y_1 \in \hat{B}([0, \tau))$ and y_2 is a translated version of an element of $\hat{B}([0, \lambda - \tau))$.

Note that the two examples given in the previous paragraph also satisfy (P.4). However, suppose we take $\hat{B}$ to be all continuous functions $\xi : \mathscr{R} \to A$ for which $\int_I \rho(\xi(t), a_0)m(dt)$ is finite for each finite interval I. In this case we still have $\hat{B}([0, \lambda)) \subset \hat{B}([0, \tau)) \times B([\tau, \lambda))$ but there are functions $y_1 \in \hat{B}([0, \tau))$ and $y_2 \in \hat{B}([\tau, \lambda))$ for which $\lim_{t \to \tau} y_1(t) \neq y_2(\tau)$ so that $y = (y_1, y_2) \notin \hat{B}([0, \tau))$. Since this function $y = (y_1, y_2)$ is defined by $y(t) = y_1(t)$ for $0 \leqq t < \tau$ and $y(t) = y_2(t)$ for $\tau \leqq t < \lambda$, it must be right continuous, however, so that we can enlarge $\hat{B}$ in this example to conform to (P.4).

As in the case of the source alphabet, we can form equivalence classes of elements of the pseudometric space $(\hat{B}([0, \lambda)), \rho_\lambda)$ to obtain the reproducing alphabet space $(\hat{B}_\lambda, \rho_\lambda)$ which is a separable metric space. We let $\hat{\mathscr{B}}_\lambda$ be the Borel σ-field of subsets of the metric space $(\hat{B}_\lambda, \rho_\lambda)$. If $\lambda = n\tau$ for some positive integer n, we identify the measurable space $(\hat{B}_\tau{}^n, \hat{\mathscr{B}}_\tau{}^n)$, which is the n-fold product space formed from $(\hat{B}_\tau, \hat{\mathscr{B}}_\tau)$, with the space $(\hat{B}_\lambda, \hat{\mathscr{B}}_\lambda)$. This follows from (P.4) and the fact that the product topology for $\hat{B}_\tau{}^n$ is generated by the metric ρ_λ which is defined by

$$\rho_\lambda(x, y) = n^{-1} \sum_{i=0}^{n-1} \rho_\tau(x_i, y_i) ,$$

where x_i and y_i are the restrictions of x and y, respectively, to the interval $[i\tau, (i + 1)\tau)$. Note that the definition of $B([0, \lambda))$ implies that $(B_\lambda, \mathscr{B}_\lambda)$ can be identified with the product measurable space $(B_\tau^n, \mathscr{B}_\tau^n)$ in the same manner as for the reproduction alphabet.

4. Coding the continuous-time source. A source word $x \in A^{[0,\lambda)}$ is encoded with the "best" code word in the sense of minimizing the distortion as measured by ρ_λ. If for a given $x \in A^{[0,\lambda)}$ all code words produce infinite distrortion, the reference code word a_λ defined in Section 3 is used. Let $\mathscr{C}_\lambda$ denote the class of all λ-length code books C_λ and define a function $d_\lambda : \Omega \times \mathscr{C}_\lambda \to \mathscr{R}$ by

$$\tilde{d}_\lambda(\omega; C_\lambda) = \min \{\tilde{\rho}_\lambda(\omega, \omega_y) \mid y \in C_\lambda\}$$

for all $\omega \in \Omega$ and all $C_\lambda \in \mathscr{C}_\lambda$ where $\omega_y : \mathscr{R} \to A$ is as defined in Section 3. Note that $\tilde{d}_\lambda(\cdot; C_\lambda)$ is a measurable extended-real-valued function having finite expectation since, by Lemma 3, each $\tilde{\rho}_\lambda(\cdot, \omega_y)$ is such a function.

We can define $d_\lambda : A^{[0,\lambda)} \times \mathscr{C}_\lambda \to \bar{\mathscr{R}}$ by

$$d_\lambda(x; C_\lambda) = \tilde{d}_\lambda(\omega_x; C_\lambda), \quad \forall\, x \in A^{[0,\lambda)}$$

which is analogous to (4). The average distortion for code book C_λ is

$$\bar{\rho}(C_\lambda \mid \mu) = \int_\Omega \tilde{d}_\lambda(\omega; C_\lambda)\mu(d\omega) \tag{11}$$

which is finite by Lemma 3. The minimum average distortion attainable for λ-length codes of rate not greater than the positive number R for the code alphabet $\hat{B}([0, \lambda))$ is

$$\delta_1(R, \mu, \lambda) = \inf \{\bar{\rho}(C_\lambda \mid \mu) \mid C_\lambda \in \mathscr{C}_\lambda(R)\} \tag{12}$$

where $\mathscr{C}_\lambda(R)$ is the set of all λ-length codes which have rate not greater than R and have code words from $\hat{B}([0, \lambda))$. The minimum average distortion for all codes of rate not greater than R is

$$\delta(R, \mu) = \inf \{\delta_1(R, \mu, \lambda) \mid \lambda > 0\}.$$

Note that for any $C_\lambda \in \mathscr{C}_\lambda$

$$\tilde{d}_\lambda(\omega; C_\lambda) \leqq \lambda^{-1} \int_{[0,\lambda)} \rho[X_t(\omega), a_0] m(dt)$$

and therefore by (8)

$$\bar{\rho}(C_\lambda \mid \mu) \leqq \rho^*. \tag{13}$$

This implies that $\delta_1(R, \mu, \lambda) \leqq \rho^*$ for $R > 0$ and $\lambda > 0$. Furthermore, property (P.4) implies that for any $R > 0$, $\lambda > 0$, and $t \in [0, \lambda)$

$$\lambda\delta_1(R, \mu, \lambda) \leqq (\lambda - t)\delta_1(R, \mu, \lambda - t) + t\delta_1(R, \mu, t). \tag{14}$$

It follows from the theory of subadditive functions [7] that

$$\lim_{\lambda\to\infty} \delta_1(R, \mu, \lambda) = \inf \{\delta_1(R, \mu, \lambda) \mid \lambda > 0\}. \tag{15}$$

Note that (12) and (13) imply

$$\delta(R, \mu) \leqq \rho^* \tag{16}$$

for any $R > 0$.

One of the goals of the paper is to express $\delta(R, \mu)$ in terms of an information-theoretic minimization; that is, to provide a coding theorem relating $\delta(R, \mu)$ to the appropriate distortion-rate functions. As in the discrete-time case, there is no hope for evaluating $\delta(R, \mu)$ directly from its definition.

The distortion-rate function $D(R, \mu)$ for the source $(\Omega, \mathscr{F}, \mu)$, reproducing class $\hat{B}$, and distortion measure ρ is defined as follows. First, for $\lambda > 0$ Lemma 4 guarantees that we can define a probability measure μ_λ on $(B_\lambda, \mathscr{B}_\lambda)$ by setting $\mu_\lambda(H) = \mu(\{\omega \in \Omega_M \mid \omega^{[0,\lambda)} \in H\})$ for each $H \in \mathscr{B}_\lambda$. Let Q_λ be the set of all transition probabilities $q: B_\lambda \times \hat{\mathscr{B}}_\lambda \to [0, 1]$. For $q \in Q_\lambda$ let

$$\mathscr{D}_\lambda(\mu, q) = \int_{B_\lambda}\int_{\hat{B}_\lambda} \rho_\lambda(x, y) q(x, dy) \mu_\lambda(dx) .$$

Define $D_1(R, \mu, \lambda)$ by

$$D_1(R, \mu, \lambda) = \inf \{\mathscr{D}_\lambda(\mu, q) \mid q \in Q_\lambda(R, \mu)\}$$

where $Q_\lambda(R, \mu)$ is the set of all transition probabilities for which the information rate $\lambda^{-1}\mathscr{I}_\lambda(\mu, q)$ is not greater than R. The average mutual information $\mathscr{I}_\lambda(\mu, q)$ is defined as the entropy of P_λ relative to π_λ where P_λ and π_λ are probability measures on $\mathscr{B}_\lambda \times \hat{\mathscr{B}}_\lambda$ defined by

$$P_\lambda(E \times F) = \int_E q(x, F) \mu_\lambda(dx)$$

and

$$\pi_\lambda(E \times F) = \mu_\lambda(E) P_\lambda(B_\lambda \times F)$$

for each $E \in \mathscr{B}_\lambda$ and each $F \in \hat{\mathscr{B}}_\lambda$. Clearly, $Q_\lambda(R, \mu)$ is nonempty and $D_1(R, \mu, \lambda) \leqq \rho^*$ for any $R \geqq 0$ since the transition probability $q_0 \in Q_\lambda$ which satisfies $q_0(x, \{a_\lambda\}) = 1$ for each $x \in B_\lambda$ also satisfies $\mathscr{D}_\lambda(\mu, q_0) = \rho^*$ and $\mathscr{I}_\lambda(\mu, q_0) = 0$. Next, define the distortion rate function by

$$D(R, \mu) = \liminf_{\lambda \to \infty} D_1(R, \mu, \lambda) .$$

It can be shown that $D(R, \mu) = \lim_{\lambda \to \infty} D_1(R, \mu, \lambda)$ by first establishing that for $t \in [0, \lambda)$

$$\lambda D_1(R, \mu, \lambda) \leqq (\lambda - t) D_1(R, \mu, \lambda - t) + t D_1(R, \mu, t)$$

which is analogous to (14). The converse source coding theorem [1] guarantees $\delta(R, \mu) \geqq D(R, \mu)$; that is, there is no rate R code which produces distortion less than $D(R, \mu)$.

Equation (15) implies a code with average distortion arbitrarily close to $\delta(R, \mu)$ can be found by first choosing λ sufficiently large and then selecting a good code from $\mathscr{C}_\lambda(R)$. However, this is not a very practical method for finding good codes. In the next section we will approach the problem from a point of view closer to that adopted in practice: the continuous-time source is modeled as a discrete-time source.

5. The segmented source. In order to represent the continuous-time source for which we have no coding theorem by a discrete-time source for which a coding theorem exists, we segment the sample functions of the continuous-time process into a doubly infinite sequence of time functions as suggested by Berger (1971). The alphabet for the segmented source consists of functions defined on the interval $[0, \tau)$. Given any $\tau > 0$ a sample function $\omega \in \Omega$ can be segmented to give a sequence $\cdots, \omega_{-1}, \omega_0, \omega_1, \cdots$ where ω_i is the restriction of ω to the interval $[i\tau, (i+1)\tau)$. Thus an element $\omega \in \Omega$ can be viewed as a doubly infinite sequence with alphabet $A' = A^{[0,\tau)}$ and the original source can be viewed as a discrete-time source described by the process $\{X_i' \mid i \in Z\}$ defined on $(\Omega, \mathscr{F}, \mu)$ by $X_i'(\omega) = \omega_i$. However, the only coding theorem available for stationary discrete-time sources requires that the underlying measurable space is a product space $\times_{i=-\infty}^{\infty} (A_i, \mathscr{A}_i)$ where $A_i = A_0$ and $\mathscr{A}_i = \mathscr{A}_0$ for $i \in Z$, A_0 is a complete separable metric space, and $\mathscr{A}_0$ is the Borel σ-field of subsets of A_0 ([6] and [13]). Since $(\Omega, \mathscr{F})$ does not have this structure, we cannot apply the source coding theorem directly to the process $\{X_i' \mid i \in Z\}$.

We have already considered an appropriate subset Σ of Ω which does have the necessary structure when the alphabet A_0 is taken to be the metric space B_τ. Thus we can take ρ_τ as the distortion measure for the discrete-time source coding problem that we are designing. Letting $\mathscr{A}_0 = \mathscr{B}_\tau$ we can construct the product measurable space $(\Sigma, \mathscr{S}_\tau) = \times_{i=-\infty}^{\infty} (A_i, \mathscr{A}_i)$ discussed in the preceding paragraph. It is easy to show that if for any $\lambda > 0$ we let $\mathscr{A}_0 = \mathscr{B}_\lambda$ and generate $\mathscr{S}_\lambda$ in the same manner, then $\mathscr{S}_\tau = \mathscr{S}_\lambda$. That is, $\mathscr{S}_\tau$ does not depend on τ. Henceforth, we will denote this σ-field by $\mathscr{S}$.

To be useful, the probability space $(\Sigma, \mathscr{S}, \nu)$ must satisfy $\mathscr{S} \subset \mathscr{F}$, $\mu(\Sigma) = 1$, and $\nu(H) = \mu(H)$ for each $H \in \mathscr{S}$. If the first two of these are satisfied then we can define the measure ν by the third. The second of these is true by Lemma 5 and the first is established by showing that sets of the form $\times_{i=-\infty}^{\infty} G_i$ are in $\mathscr{F}$ where $G_i = B([i\tau, (i+1)\tau))$ for all but finitely many integers i and $G_i \in \mathscr{B}([i\tau, (i+1)\tau))$, $\forall\, i \in Z$. Note that

$$\times_{i=-\infty}^{\infty} G_i = \bigcap_{i \in Z} \{\omega \in \Omega_M \mid \omega^{[i\tau,(i+1)\tau)} \in G_i\} .$$

By Lemma 4 each set in the intersection is $\mathscr{F}$-measurable so $\times_{i=-\infty}^{\infty} G_i \in \mathscr{F}$. This implies that any set in the σ-field $\mathscr{S}$ generated by sets of this form is $\mathscr{F}$-measurable.

For any $\tau > 0$, *the segmented source of duration* τ is the random process $\{X_{\tau,i} \mid i \in Z\}$ on $(\Sigma, \mathscr{S}, \nu)$ defined by $X_{\tau,i}(\omega) = \omega^{[i\tau,(i+1)\tau)}$ for $\omega \in \Sigma$ and $i \in Z$. This source is clearly stationary with respect to the τ-unit shift transformation T^τ but it is not ergodic with respect to T^τ if the original source is not ergodic. Moreover, ergodicity of the original source does not guarantee ergodicity of the segmented source of duration τ. Since $(\Sigma, \mathscr{S}, \nu)$ does not depend on the value of τ used in the construction, it can serve as the underlying probability space for the segmented source of duration τ for any $\tau > 0$. For each positive integer n, let $\mathbf{X}_\tau^n$ be the random vector whose ith component is $X_{\tau,i}$ for $0 \leq i \leq n-1$.

We are now in a position to define the least average distortion $\delta(R, \mu, \tau)$ that can result when the source is encoded with a discrete-time code book. We will view the source as a discrete-time source with alphabet B_τ. Source letters, which are elements of B_τ, are encoded with elements of the reproduction alphabet $\hat{B}_\tau$. The first step is to consider the coding of the segmented source described by the process $\{X_{\tau,i} \mid i \in \Sigma\}$ defined on $(\Sigma, \mathscr{S}, \nu)$.

A (τ, N) block code book $C_{\tau,N}$ is a finite collection of elements of $\hat{B}_\tau^N$. We require that $a_{N\tau} \in C_{\tau,N}$. The code $C_{\tau,N}$ is applied to the segmented source by an encoding rule $U\colon \hat{B}_\tau^N \to C_{\tau,N}$ which satisfies

$$\rho_{\tau,N}(x, U(x)) = \min\{\rho_{\tau,N}(x, y) \mid y \in C_{\tau,N}\}$$

where $\rho_{\tau,N}$ is the distortion measure obtained from the single-letter distortion measure ρ_τ in the usual way; that is,

$$\rho_{\tau,N}(x, y) = N^{-1} \textstyle\sum_{i=1}^N \rho_\tau(x_i, y_i)\,.$$

We can apply a (τ, N) block code to the continuous-time source by extending the encoding rule U to $A^{[0,N\tau)}$. For instance, we can let $U(x) = a_{N\tau}$ whenever $x \in A^{[0,N\tau)}$ but $x \notin \hat{B}_\tau^N$. Since $\hat{B}_\tau^N = \hat{B}_{N\tau}$, then we can view any (τ, N) block code as an $N\tau$-length code for the continuous-time source as discussed in Section 4. In fact there is a one-to-one correspondence between elements of $\mathscr{C}_{N\tau}$ and elements of the class $\mathscr{C}_{\tau,N}$ of all (τ, N) block code books. This correspondence is denoted by $C_{N\tau} \sim C_{\tau,N}$.

The rate of a (τ, N) block code $C_{\tau,N}$ is $(N\tau)^{-1} \log \|C_{\tau,N}\|$. If the logarithm is taken to the base 2, this represents the minimum number of binary digits per unit time required to transmit fixed length binary representations of sequences of code words over a channel which accepts binary digits at a constant rate. Of course $C_{N\tau} \sim C_{\tau,N}$ implies the rate of $C_{N\tau}$ is equal to the rate of $C_{\tau,N}$. Let $\mathscr{C}_{\tau,N}(R)$ be the class of all (τ, N) block codes of rate not greater than R. We say a code is a block code for segments of length τ if it is a (τ, N) block code for some positive integer N. The class $\mathscr{C}(R, \tau)$ of all block codes for segments of length τ which have rate not greater than R is $\bigcup_{N=1}^\infty \mathscr{C}_{\tau,N}(R)$. Note that $(\tau, 1)$ block codes are just the τ-length codes of Sections 3 and 4.

If $d_{\tau,N}(x; C_{\tau,N})$ is the minimum distortion as measured by $\rho_{\tau,N}$ for source output $x \in B_\tau^N$ and code $C_{\tau,N}$, then we can define

$$\delta_N(R, \mu, \tau) = \inf\{\bar{\rho}(C_{\tau,N} \mid \mu) \mid C_{\tau,N} \in \mathscr{C}_{\tau,N}(R)\}$$

where

$$\bar{\rho}(C_{\tau,N} \mid \mu) = \textstyle\int_\Sigma d_{\tau,N}(X_\tau^N(\omega); C_{\tau,N})\mu(d\omega) \tag{17}$$

(since $\mu(\Sigma) = 1$ we can also write (17) as an integral on the space Ω). The quantity $\delta_N(R, \mu, \tau)$ is the greatest lower bound on the distortion that results when the segmented source of duration τ is encoded with a (τ, N) block code of rate R. We then define

$$\delta(R, \mu, \tau) = \inf\{\delta_N(R, \mu, \tau) \mid N = 1, 2, \cdots\}\,.$$

Since $N\delta_N(R, \mu, \tau)$ is a subadditive function of N, $\delta(R, \mu, \tau) = \lim_{N\to\infty} \delta_N(R, \mu, \tau)$.

We note that $\delta_1(R, \mu, N\tau) = \delta_N(R, \mu, \tau)$ so (15) implies that for any $\tau > 0$,

$$\delta(R, \mu, \tau) = \lim_{N\to\infty} \delta_1(R, \mu, N\tau) = \delta(R, \mu) . \tag{18}$$

The implication is that given any $\varepsilon > 0$ we can segment the continuous-time source using any segment length τ and construct a block code of rate R which yields average distortion within ε of the lowest possible distortion $\delta(R, \mu)$ achievable by coding the original continuous-time source. Thus, the segment length τ is not an important parameter in the design of efficient source codes. If this is not the case (e.g., if we drop the requirement (P.4)), we have no guarantee that it is possible to achieve distortion level $\delta(R, \mu) + \varepsilon$ by applying a rate R block code to the segmented source of duration τ.

The quantity $\bar{\rho}(C_{\tau,N} \mid \mu)$ is the average distortion that results when the (τ, N) block code $C_{\tau,N}$ is applied to the source $(\Omega, \mathscr{F}, \mu)$. Note however that $\mathbf{X}_\tau^N$ is actually a measurable mapping from $(\Sigma, \mathscr{S})$ into $(B_\tau^N, \mathscr{B}_\tau^N)$. Thus $\bar{\rho}(C_{\tau,N} \mid \mu)$ is really only a function of ν, the restriction of μ to $(\Sigma, \mathscr{S})$. In particular

$$\bar{\rho}(C_{\tau,N} \mid \mu) = \bar{\rho}'(C_{\tau,N} \mid \nu) = \int_\Sigma d_{\tau,N}(\mathbf{X}_\tau^N(\omega); C_{\tau,N})\nu(d\omega) . \tag{19}$$

The quantity $\bar{\rho}'(C_{\tau,N} \mid \nu)$ is the average distortion that results when the block code $C_{\tau,N}$ is applied to the discrete-time source $(\Sigma, \mathscr{S}, \nu)$. Because of (18) and (19)

$$\delta(R, \nu, \tau) = \delta(R, \mu, \tau) = \delta(R, \mu) \tag{20}$$

for each $\tau > 0$ and hence $\delta(R, \nu) = \delta(R, \mu)$.

The definition of the distortion-rate function $D(R, \nu, \tau)$ for the segmented source of duration τ is analogous to Berger's definition of the rate-distortion function for discrete-time stationary source with abstract alphabet [2]. We first define the nth-order distortion-rate function $D_n(R, \nu, \tau)$. The random vector $\mathbf{X}_\tau^n$ induces a probability measure ν^n on $(B_\tau^n, \mathscr{B}_\tau^n)$ which is defined by

$$\nu^n(E) = \nu\{\omega \in \Sigma \mid \mathbf{X}_\tau^n(\omega) \in E\} \quad \forall\, E \in \mathscr{B}_\tau^n .$$

For any transition probability $q: B_\tau^n \times \hat{\mathscr{B}}_\tau^n \to [0, 1]$, the average mutual information $\mathscr{I}_{\tau,n}(\nu, q)$ is defined as the entropy of p relative to π where p and π are the measures on $\mathscr{B}_\tau^n \times \hat{\mathscr{B}}_\tau^n$ defined by

$$p(E \times F) = \int_E q(x, F)\nu^n(dx) \tag{21}$$

and

$$\pi(E \times F) = \nu^n(E)p(B_\tau^n \times F) \tag{22}$$

for each $E \in \mathscr{B}_\tau^n$ and each $F \in \hat{\mathscr{B}}_\tau^n$.

Let $Q_{\tau,n}(R, \nu)$ be the set of all transition probabilities q for which the information rate $(n\tau)^{-1}\mathscr{I}_{\tau,n}(\nu, q)$ is not greater than the positive number R. Define

$$\mathscr{D}_{\tau,n}(\nu, q) = \int_{B_\tau^n} \int_{\hat{B}_\tau^n} \rho_{\tau,n}(x, y)q(x, dy)\nu^n(dx) .$$

The nth order distortion-rate function is given by

$$D_n(R, \nu, \tau) = \inf\{\mathscr{D}_{\tau,n}(\nu, q) \mid q \in Q_{\tau,n}(R, \nu)\}\,.$$

Finally, the distortion-rate function for the segmented source of duration τ is

$$D(R, \nu, \tau) = \lim_{n\to\infty} D_n(R, \nu, \tau)\,.$$

The distortion-rate function for $(\Sigma, \mathscr{S}, \nu)$ is then $D(R, \nu) = \inf\{D(R, \nu, \tau) \mid \tau > 0\}$. Since a transition probability $q \in Q_{\tau,n}(R, \nu)$ can be viewed as a function on $B_\tau^n \times \hat{\mathscr{B}}_\tau^n$ or $B_\lambda \times \hat{\mathscr{B}}_\lambda$ where $\lambda = n\tau$, it is clear from the definitions that $\mathscr{I}_{\tau,n}(\nu, q) = \mathscr{I}_\lambda(\mu, q)$ and $\mathscr{D}_{\tau,n}(\nu, q) = \mathscr{D}_\lambda(\mu, q)$ so that $D(R, \nu, \tau) = D(R, \mu)$ for each $\tau > 0$ and thus $D(R, \nu) = D(R, \mu)$.

To summarize, we have constructed a source with underlying probability space $(\Sigma, \mathscr{S}, \nu)$ which has the structure needed for the discrete-time source coding theorem of Gray and Davisson. This source satisfies $\delta(R, \mu) = \delta(R, \nu, \tau) = \delta(R, \nu)$ and $D(R, \mu) = D(R, \nu, \tau) = D(R, \nu)$ regardless of the value of τ that was used to construct the segmented source. If the original continuous-time source is totally ergodic, then the segmented source of duration τ consisting of the process $\{X_{\tau,i} \mid i \in Z\}$ on the probability space $(\Sigma, \mathscr{S}, \nu)$ is ergodic with respect to the shift transformation T^τ. In this case the abstract alphabet discrete-time source coding theorem of Berger (1968, 1971) can be applied to establish $\delta(R, \nu, \tau) = D(R, \nu, \tau)$ and therefore not only is it true that $\delta(R, \mu) = D(R, \mu)$ but also $\delta(R, \mu, \tau) = D(R, \mu)$ for each $\tau > 0$.

Because of the fact that we do not require the original source to be ergodic (let alone block ergodic), the segmented source cannot be assumed to be ergodic. Thus, the more general result of Gray and Davission is required. Since the alphabet space (B_τ, ρ_τ) is a complete separable metric space, the nonergodic source $(\Sigma, \mathscr{S}, \nu)$ can be decomposed into its ergodic (or metrically transitive) components [18], [19]. Let $\mathscr{T}_\tau$ be the σ-field of τ-invariant sets in $\mathscr{S}$. Since $\mathscr{S}$ is the σ-field of Borel subsets of the complete separable metric space Σ, then there exists a family $\{\nu_\omega^\tau \mid \omega \in \Sigma\}$ of τ-ergodic probability measures on $(\Sigma, \mathscr{S})$ for which $\nu_\omega^\tau(S)$ is a $\mathscr{T}_\tau$-measurable function of ω for each fixed $S \in \mathscr{S}$ and

$$\nu(S \cap H) = \int_H \nu_\omega^\tau(S)\nu(d\omega)$$

for each $S \in \mathscr{S}$ and each $H \in \mathscr{T}_\tau$. Furthermore, for each nonnegative measurable function $f\colon \Sigma \to \mathscr{R}$,

$$\int_\Sigma f\, d\nu = \int_\Sigma \{\int_\Sigma f\, d\nu_\omega^\tau\}\nu(d\omega)\,. \tag{23}$$

Note that we require $\nu_\omega^\tau(S)$ to be $\mathscr{T}_\tau$-measurable and not just $\mathscr{S}$-measurable as in [15]. Thus, $\nu_\omega^\tau(S)$ is a version of the conditional probability of S given $\mathscr{T}_\tau$. The results of Gray and Davisson (1974) as generalized in [13] provide a coding theorem for the nonergodic source in terms of the distortion-rate functions $D(R, \nu_\omega^\tau)$ for the ergodic sources of this decomposition. We need only the following special case.

THEOREM 1. *Given a discrete-time stationary source consisting of the random process* $\{X_{\tau,i} \mid i \in Z\}$ *on the probability space* $(\Sigma, \mathscr{S}, \nu)$ *with complete separable metric space alphabet* (B_τ, ρ_τ) *and reproduction alphabet* $\hat{B}_\tau \subset B_\tau$ *for which there exists a letter* $a_\tau \in B_\tau$ *such that*

$$\int_\Sigma \rho_\tau[X_{\tau,0}(\omega), a_\tau]\nu(d\omega) < \infty$$

then

$$\delta(R, \nu, \tau) = \int_\Sigma D(R, \nu_\omega{}^\tau)\nu(d\omega) .$$

6. Continuous-time source coding theorems. In this section we will present several new coding results for continuous-time random processes. We first summarize the results of the previous sections by stating a source coding theorem for stationary sources. We then specialize this result to stationary ergodic sources.

Stationary sources. Our basic coding theorem follows from Theorem 1 and (20). As in the case of Theorem 1, this result can be applied for any $\tau > 0$.

THEOREM 2. *For a stationary continuous-time source satisfying* (P.1) *and* (P.2) *with a reproduction alphabet satisfying* (P. 3) *and* (P.4),

$$\delta(R, \mu) = \int_\Sigma D(R, \nu_\omega{}^\tau)\mu(d\omega) , \tag{24}$$

where $\Sigma = \{\omega \in \Omega \mid \omega^{[i\tau,(i+1)\tau)} \in B([i\tau, (i+1)\tau)), \forall i \in Z\}$ *and* $\{\nu_\omega{}^\tau \mid \omega \in \Sigma\}$ *is the family of* τ*-ergodic probability measures obtained from the ergodic decomposition of the segmented source of duration* τ.

Equation (24) can be written in various equivalent forms. Since $\nu_\omega{}^\tau$ is τ-ergodic, we can replace $D(R, \nu_\omega{}^\tau)$ by $\delta(R, \nu_\omega{}^\tau, \tau)$ in the integrand and we can also replace $\delta(R, \mu)$ by either $\delta(R, \mu, \lambda)$ or $\delta(R, \nu, \lambda)$ for any $\lambda > 0$.

Theorem 2 provides a formula for evaluating the optimum performance $\delta(R, \mu)$ for the stationary source. Since an uncountable number of distortion-rate functions must be evaluated in order to calculate the right-hand side of (24), it is important to have conditions under which the simpler relationship $\delta(R, \mu) = D(R, \mu)$ is valid. A stronger relationship which has considerably greater practical significance is $\delta(R, \mu, \tau) = D(R, \mu)$ for each $\tau > 0$. In order to establish such relationships we must place further restrictions on the source.

Ergodic sources. Before proceeding to the source coding theorem for ergodic sources, we will show that the coding theorem for totally ergodic sources follows immediately from Theorem 2. Suppose the totally ergodic source satisfies all of the conditions of Theorem 2. Total ergodicity implies $\mathscr{T}_\tau$ is a trivial σ-field for each $\tau > 0$ (i.e., any set in $\mathscr{T}_\tau$ has ν-measure equal to 0 or 1). Therefore, for each set $S \in \mathscr{S}$ there is a set $\Sigma_S \in \mathscr{S}$ such that $\nu(\Sigma_S) = 1$ and

$$\nu_\omega{}^\tau(S) = E\{I_S \mid \mathscr{T}_\tau\}(\omega) = E\{I_S\} = \nu(S) \quad \forall\, \omega \in \Sigma_S$$

where I_S is the indicator function for S. Since $\mathscr{S}$ is countably generated this implies that the measures $\nu_\omega{}^\tau$ and ν are identical for ν-almost all ω. Theorem 2 and the results of Section 5 then imply $\delta(R, \mu) = D(R, \mu)$ and in fact $\delta(R, \mu, \tau) = D(R, \mu)$ for every $\tau > 0$.

We will say a continuous-time source $(\Omega, \mathscr{F}, \mu)$ is continuous in the sense of Pinsker if

$$\lim_{\tau\to 0} \mu(E \triangle T^\tau E) = 0 , \quad \forall E \in \mathscr{F} .$$

A source which is weakly mixing and continuous in the sense of Pinsker is known to be totally ergodic [1]. Thus, the above comments constitute a proof of the coding theorem for weakly mixing sources which are continuous in the sense of Pinsker. Our main result for ergodic sources is the following theorem.

THEOREM 3. *Suppose a stationary, ergodic, continuous-time source* $(\Omega, \mathscr{F}, \mu)$ *satisfies* (P.1) *and* (P.2) *and is continuous in the sense of Pinsker. Suppose the reproduction alphabet satisfies* (P.3) *and* (P.4). *Then* $\delta(R, \mu, \tau) = D(R, \mu)$ *for all* $\tau > 0$.

PROOF. Let τ and ε be arbitrary positive numbers and define $\lambda_k = 2^{-k}$ for each nonnegative integer k. For each k, let $\{\nu_{\omega,k} \mid \omega \in \Sigma\}$ be the family of λ_k-ergodic probability measures. Choose a positive integer n and a transition probability $q \in Q_{\tau,n}(R, \nu)$ such that

$$\mathscr{D}_{\tau,n}(\nu, q) \leqq D(R, \mu) + \varepsilon . \tag{25}$$

Let $\nu^n_{\omega,k}$ be the measure induced on $\mathscr{B}_\tau^n$ by $\nu_{\omega,k}$. For each nonnegative integer k define

$$R_k(\omega) = (n\tau)^{-1}\mathscr{I}_{\tau,n}(\nu_{\omega,k}, q) , \quad \forall \omega \in \Sigma$$

and note that

$$\mathscr{D}_{\tau,n}(\nu_{\omega,k}, q) \geqq D(R_k(\omega), \nu_{\omega,k}) , \quad \forall \omega \in \Sigma . \tag{26}$$

It follows from (25) that $\mathscr{D}_{\tau,n}(\nu, q) \leqq \rho^* + \varepsilon$. Since

$$\begin{aligned}\mathscr{D}_{\tau,n}(\nu, q) &= \textstyle\int_{B_\tau^n} \int_{\hat{B}_\tau^n} \rho_{\tau,n}(x, y) q(x, dy) \nu^n(dx) \\ &= \textstyle\int_\Sigma \int_{\hat{B}_\tau^n} \rho_{\tau,n}(\mathbf{X}_\tau{}^n(\xi), y) q(\mathbf{X}_\tau{}^n(\xi), dy) \nu(d\xi)\end{aligned}$$

and

$$\mathscr{D}_{\tau,n}(\nu_{\omega,k}, q) = \textstyle\int_\Sigma \int_{\hat{B}_\tau^n} \rho_{\tau,n}(\mathbf{X}_\tau{}^n(\xi), y) q(\mathbf{X}_\tau{}^n(\xi), dy) \nu_{\omega,k}(d\xi) ,$$

then it follows from (23) that

$$\mathscr{D}_{\tau,n}(\nu, q) = \textstyle\int_\Sigma \mathscr{D}_{\tau,n}(\nu_{\omega,k}, q) \nu(d\omega) . \tag{27}$$

We will later show that there exists a $\delta > 0$, a positive integer k, and a set $\Sigma_0 \in \mathscr{S}$ such that $\nu(\Sigma_0{}^C) \leqq \delta$;

$$R_k(\omega) < R + \varepsilon , \quad \forall \omega \in \Sigma_0 ; \tag{28}$$

and

$$\textstyle\int_{\Sigma_0^C} \int_\Sigma \rho_{\tau,n}(\mathbf{X}_\tau{}^n(\xi), a_{n\tau}) \nu_{\omega,k}(d\xi) \nu(d\omega) < \varepsilon . \tag{29}$$

Assuming that (28) is true, then it is clear that

$$D(R_k(\omega), \nu_{\omega,k}) \geqq D(R + \varepsilon, \nu_{\omega,k}) , \quad \forall \omega \in \Sigma_0 . \tag{30}$$

Combining (26) and (30) we have

$$\mathscr{D}_{\tau,n}(\nu_{\omega,k}, q) \geqq D(R+\varepsilon, \nu_{\omega,k}), \quad \forall\, \omega \in \Sigma_0,$$

and then from (27)

$$\mathscr{D}_{\tau,n}(\nu, q) \geqq \int_{\Sigma_0} D(R+\varepsilon, \nu_{\omega,k})\nu(d\omega). \tag{31}$$

Next we notice that

$$\int_{\Sigma_0^C} D(R+\varepsilon, \nu_{\omega,k})\nu(d\omega) \leqq \int_{\Sigma_0^C} \int_\Sigma \rho_{\tau,n}(\mathbf{X}_\tau^n(\xi), a_{n\tau})\nu_{\omega,k}(d\xi)\nu(d\omega)$$

so that (29) and (31) give

$$\mathscr{D}_{\tau,n}(\nu, q) \geqq \int_\Sigma D(R+\varepsilon, \nu_{\omega,k})\nu(d\omega) - \varepsilon. \tag{32}$$

Applying Theorem 2 and (25) and (32) we have

$$D(R, \mu) \geqq \delta(R+\varepsilon, \mu) - 2\varepsilon. \tag{33}$$

Since ε was arbitrary and $\delta(\cdot, \mu)$ is a convex function, (33) implies

$$D(R, \mu) \geqq \delta(R, \mu)$$

which, together with (20) and the converse source coding theorem, guarantees $\delta(R, \mu, \tau) = D(R, \mu)$. Thus, the proof is complete if we establish (28) and (29). Clearly, these two results follow from the two lemmas which are stated and proved below.

LEMMA 6. *For any $\varepsilon > 0$ there exists a δ such that $\nu(S) < \delta$ implies*

$$\int_S \int_\Sigma \rho_{\tau,n}(\mathbf{X}_\tau^n(\xi), a_{n\tau})\nu_{\omega,k}(d\xi)\nu(d\omega) < \varepsilon$$

for all nonnegative integers k.

PROOF. Suppose the random variable Y is defined by

$$Y(\xi) = \rho_{\tau,n}(\mathbf{X}_\tau^n(\xi), a_{n\tau}), \quad \forall\, \xi \in \Sigma \tag{34}$$

and, for each k, X_k is defined by

$$X_k(\omega) = \int_\Sigma Y\, d\nu_{\omega,k}, \quad \forall\, \omega \in \Sigma. \tag{35}$$

Then, the random variables X_k are uniformly integrable and therefore the integrals $\int_\Sigma X_k\, d\nu$ are uniformly bounded and uniformly continuous in the sense that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\sup\{\int_S X_k\, d\nu \mid k = 0, 1, 2, \cdots\} \leqq \varepsilon \tag{36}$$

whenever $S \in \mathscr{S}$ and $\nu(S) \leqq \delta$. Substituting in (36) using (34) and (35) completes the proof of the lemma.

LEMMA 7. *For any $\varepsilon > 0$ and $\delta > 0$ there exists a set $\Sigma_0 \in \mathscr{S}$ and an integer k such that $\nu(\Sigma_0^C) \leqq \delta$ and*

$$\mathscr{I}_{\tau,n}(\nu_{\omega,k}, q) \leqq \mathscr{I}_{\tau,n}(\nu, q) + \varepsilon, \quad \forall\, \omega \in \Sigma_0.$$

PROOF. Since convergence in mean implies almost uniform convergence of a subsequence, it suffices to show

$$(37)\qquad \lim_{k\to\infty} \int_\Sigma |\mathscr{I}_{\tau,n}(\nu_{\omega,k}, q) - \mathscr{I}_{\tau,n}(\nu, q)|\nu(d\omega) = 0\,,$$

which follows from

$$(38)\qquad \liminf_{k\to\infty} \mathscr{I}_{\tau,n}(\nu_{\omega,k}, q) \geqq \mathscr{I}_{\tau,n}(\nu, q)$$

for ν-almost all ω and

$$(39)\qquad \int_\Sigma \mathscr{I}_{\tau,n}(\nu_{\omega,k}, q)\nu(d\omega) \leqq \mathscr{I}_{\tau,n}(\nu, q)\,.$$

We first establish (38) for ν-almost ω. Let $\mathscr{T}^* = \bigcap_{k=0}^\infty \mathscr{T}_{\lambda_k}$. Clearly for each k, $\mathscr{T}_{\lambda_k} \subset \mathscr{T}_{m\lambda_k}$ each positive integer m. Thus,

$$\mathscr{T}^* = \bigcap_{k=0}^\infty \bigcap_{m=1}^\infty \mathscr{T}_{m\lambda_k}\,.$$

Let $\mathscr{T}$ denote the class of invariant sets in $\mathscr{S}$. Clearly $\mathscr{T} \subset \mathscr{T}^*$. To show that in fact $\mathscr{T} = \mathscr{T}^*$, let λ be an arbitrary positive number, let $E \in \mathscr{T}^*$, and let (τ_i) be a sequence of positive numbers such that each τ_i is of the form $m\lambda_k$ and $\lim_{i\to\infty} \tau_i = \lambda$. For each i, stationarity of μ implies

$$(40)\qquad \mu(E \triangle T^\lambda E) \leqq \mu(E \triangle T^{\tau_i} E) + \mu(E \triangle T^{(\lambda-\tau_i)} E)\,.$$

But $E \in \mathscr{T}^*$ implies $\mu(E \triangle T^{\tau_i} E) = 0$ for each i. Thus, (40) and continuity in the sense of Pinsker implies

$$\mu(E \triangle T^\lambda E) \leqq \lim_{i\to\infty} \mu(E \triangle T^{(\lambda-\tau_i)} E) = 0$$

and therefore $E \in \mathscr{T}_\lambda$. But since λ was arbitrary this implies $E \in \mathscr{T}$. Therefore, $\mathscr{T} = \bigcap_{k=0}^\infty \mathscr{T}_{\lambda_k}$. Notice also that $(\mathscr{T}_{\lambda_k})$ is a contracting sequence of sub-σ-fields of $\mathscr{S}$. Therefore, if $Y: \Sigma \to \bar{\mathscr{R}}$ is any random variable with finite expectation EY, then the sequence (X_k) of random variables defined by

$$X_k(\omega) = E\{Y \mid \mathscr{T}_{\lambda_k}\}(\omega) = \int_\Sigma Y\, d\nu_{\omega,k}$$

is a martingale relative to the sequence $(\mathscr{T}_{\lambda_k})$ and the sequence (X_k) converges to $E\{Y \mid \mathscr{T}\}$ ν-almost everywhere. But the source is ergodic so $\mathscr{T}$ is a trivial σ-field and $E\{Y \mid \mathscr{T}\} = EY$ ν-almost everywhere. Let E' and F' be arbitrary sets in $\mathscr{B}_\tau^n$ and $\hat{\mathscr{B}}_\tau^n$, respectively, and define the random variable Y by $Y(\omega) = q(\mathbf{X}_\tau^n(\omega), F')$ if $\mathbf{X}_\tau^n(\omega) \in E'$ and $Y(\omega) = 0$ otherwise. If we define the random variables (X_k) by

$$X_k(\omega) = \int_\Sigma Y\, d\nu_{\omega,k}$$

then we see that for ν-almost all ω

$$\begin{aligned}\lim_{k\to\infty} p_{\omega,k}(E' \times F') &= \lim_{k\to\infty} \int_\Sigma Y\, d\nu_{\omega,k} \\ (41)\qquad &= \int_\Sigma Y\, d\nu = p(E' \times F')\,,\end{aligned}$$

where, for each $\omega \in \Sigma$ and each k, $p_{\omega,k}$ is the probability measure induced on $\mathscr{B}_\tau^n \times \hat{\mathscr{B}}_\tau^n$ by $\nu_{\omega,k}$ and q. That is, for each $E \in \mathscr{B}_\tau^n$ and each $F \in \hat{\mathscr{B}}_\tau^n$,

$$(42)\qquad p_{\omega,k}(E \times F) \doteq \int_E q(x, F)\nu_{\omega,k}^n(dx)\,.$$

For each $\omega \in \Sigma$, $p_{\omega,k}$ can be extended in a unique way to $\mathscr{B}_\tau^n \times \hat{\mathscr{B}}_\tau^n$. The measure p is defined in (21). If $\mathscr{M}$ is a countable algebra of sets which generates $\mathscr{B}_\tau^n \times \hat{\mathscr{B}}_\tau^n$ then it follows that there exists a set $\Sigma' \in \mathscr{S}$ with $\nu(\Sigma') = 1$ such that for each $\omega \in \Sigma'$

$$\lim_{k\to\infty} p_{\omega,k}(M) = p(M)\,, \quad \forall\, M \in \mathscr{M}. \tag{43}$$

The proof of (38) is now easily completed by applying (43) and Dobrushin's theorem to the definition of mutual information. This step is very similar to the proof of lower semi-continuity of mutual information and relative entropy (see [17], pages 13 and 20), so the details are omitted.

Finally, we see that (39) is a direct consequence of Lemma 1 if we define P on the σ-field $\mathscr{S} \times \mathscr{B}_\tau^n \times \hat{\mathscr{B}}_\tau^n$ by

$$P(B) = \int_\Omega \int_{B_\tau^n} \int_{\hat{B}_\tau^n} I_B(\omega, x, y) q(x, dy) \nu_{\omega,k}(dx) \nu(d\omega)$$

and then let $\bar{P}$ be the measure on $\mathscr{S} \times \mathscr{B}_\tau^n \times \hat{\mathscr{B}}_\tau^n$ which satisfies

$$\bar{P}(G \times E \times F) = \int_G \bar{p}_{\omega,k}(E \times F)\nu(d\omega)$$

where $\bar{p}_{\omega,k}$ is the measure which satisfies

$$\bar{p}_{\omega,k}(E \times F) = p_{\omega,k}(E \times \hat{B}_\tau^n) p_{\omega,k}(B_\tau^n \times F)\,.$$

If Q and $\bar{Q}$ are defined as in Lemma 1, then it follows from (23) that

$$Q(E \times F) = \int_E q(x, F)\nu^n(dx)$$

and

$$\bar{Q}(E \times F) = \nu^n(E) \int_{B_\tau^n} q(x, F)\nu^n(dx)\,.$$

Therefore, $H(Q, \bar{Q}) = \mathscr{I}_{\tau,n}(\nu, q)$ and $H(p_\omega, \bar{p}_\omega) = \mathscr{I}_{\tau,n}(\nu_{\omega,k}, q)$. This completes the proof of Lemma 7.

7. Discussion and conclusions. The main results of this paper are Theorems 2 and 3 which relate optimal code performance to the distortion-rate function of information theory. These results apply to a wide class of stationary sources of physical interest including nonergodic sources (Theorem 2) and ergodic sources which are continuous in the sense of Pinsker (Theorem 3). We should mention that the measurability requirement on the process is not unrelated to the continuity requirement. If a stationary source is continuous in the sense of Pinsker, the source output process is continuous in probability. If the source alphabet is $\bar{\mathscr{R}}$ or any compact metric space, this guarantees there is a measurable standard extension of the process.

Theorems 2 and 3, as stated and proved here, require that the distortion measure be a metric. However, the only place this is used in our proof is where we invoke Theorem 1 (the Gray–Davisson theorem). Thus, Theorems 2 and 3 are also valid for any bounded, measurable distortion measure; the proof is exactly as given here except we use Theorem 3 of Kieffer (1975) in place of Theorem 1. Furthermore, as pointed out in [6], it appears that Theorem 1 can easily be extended to allow the distortion measure to be a nonnegative, continuous, nondecreasing, function of a metric.

Finally we should mention that source coding theorems can be obtained for information-stable continuous-time sources with certain uniform integrability conditions on the distortion measure. In unpublished lecture notes (in Hungarian), Csiszar has proved a continuous-time source coding theorem via this method. This approach was used earlier by Dobrushin (1959) for discrete-time sources. Marton (1972) has established information stability for stationary ergodic continuous-time processes which are continuous in the sense of Pinsker and have separable metric space alphabets. Her approach was based on the observation that the set of all τ for which the ergodic process is not τ-ergodic is a countable set.

Acknowledgment. We wish to thank the referee for several valuable suggestions which led to a considerable strengthening of Theorem 3. We are also grateful to A. D. Wyner and B. E. Hajek for helpful discussions related to this paper.

REFERENCES

[1] BERGER, T. (1968). Rate distortion theory for sources with abstract alphabets and memory. *Information and Control* **13** 254–273.

[2] BERGER, T. (1971). *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice-Hall, Englewood Cliffs, N.J.

[3] DOBRUSHIN, R. L. (1959). A general formulation of the fundamental Shannon theorem in information theory. *Uspehi Mat. Nauk.* **14** No. 6 3–104. (English translation in *Amer. Math. Soc. Transl.* (Series 2) **33** 323–438.)

[4] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[5] GELFAND, I. M. and YAGLOM, A. M. (1957). Calculation of the amount of information about a random function contained in another such function. *Uspehi Mat. Nauk.* **12** No. 1 3–52. (English translation in *Amer. Mat. Soc. Transl.* (Series 2) **12** 199–246.)

[6] GRAY, R. M. and DAVISSON, L. D. (1974). Source coding theorems without the ergodic assumption. *IEEE Trans. Information Theory* **20** 502–516.

[7] HILLE, E. and PHILLIPS, R. S. (1957). *Functional Analysis and Semi-Groups.* Amer. Math. Soc., Providence, R.I.

[8] KADOTA, T. T. and WYNER, A. D. (1972). Coding theorems for stationary, asymptotically memoryless, continuous-time channels. *Ann. Math. Statist.* **43** 1603–1611.

[9] KIEFFER, J. C. (1975). On the optimum average distortion attainable by fixed-rate coding of a nonergodic source. *IEEE Trans. Information Theory* **21** 190–193.

[10] KOLMOGOROV, A. N. (1956). On the Shannon theory of information transmission in the case of continuous signals *IEEE Trans. Information Theory* **2** 102–108.

[11] LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, N.Y.

[12] MARTON, K. (1972). Information and information stability of ergodic sources. *Problemy Peredachi Informatsii* **8** No. 3 3–8. (English translation in *Problems of Information Transmission* **8** 179–183.)

[13] NEUHOFF, D. L., GRAY, R. M. and DAVISSON, L. D. (1975). Fixed rate universal block source coding with a fidelity criterion. *IEEE Trans. Information Theory* **21** 511–528.

[14] NEVEU, J. (1964). *Mathematical Foundations of the Calculus of Probability* (translated by A. Feinstein). Holden-Day, San Francisco.

[15] OXTOBY, J. C. (1952). Ergodic sets. *Bull. Amer. Math. Soc.* **58** 116–136.

[16] PEREZ, A. (1957). Notions generalisées d'incertitude, d'entropie et d'information du point de vue de la théorie de martingales. *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes* 183–208.

[17] PINSKER, M. S. (1964). *Information and Information Stability of Random Variables and Processes* (translated by A. Feinstein, ed.). Holden-Day, San Francisco.

[18] ROHLIN, V. A. (1949). Selected topics from the metric theory of dynamical systems. *Uspehi Mat. Nauk.* **4** No. 2 57–128. (English translation in *Amer. Math. Soc. Translations* (Series 2) **49** 171–240.)

[19] ROZANOV, YU. A. (1967). *Stationary Random Processes* (translated by A. Feinstein). Holden-Day, San Francisco.

COORDINATED SCIENCE LABORATORY
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801

ELECTRICAL ENGINEERING DEPARTMENT
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305