Markov Random Processes Are Neither Bandlimited nor Recoverable From Samples or After Quantization

Daniel Marco, *Member, IEEE*

Abstract—This paper considers basic questions regarding Markov random processes. It shows that continuous-time, continuous-valued, wide-sense stationary, Markov processes that have absolutely continuous second-order distribution and finite second moment are not bandlimited. It also shows that continuous-time, stationary, Markov processes that are continuous-valued or discrete-valued and satisfy additional mild conditions cannot be recovered from uniform sampling. Further it shows that continuous-time, continuous-valued, stationary, Markov processes that have absolutely continuous second-order distributions and are continuous almost surely, cannot be recovered without error after quantization. Finally, it provides necessary and sufficient conditions for stationary, discrete-time, Markov processes to have zero entropy rate, and relates this to information singularity.

Index Terms—Bandlimited, information-singular, Markov, recoverability after quantization, recoverability from samples.

I. INTRODUCTION

THIS paper considers basic questions mostly regarding continuous-time Markov random processes that are either continuous- or discrete-valued. The results are concerned with the nonbandlimitedness and nonrecoverability from sets of samples or after quantization of such processes.

The first result, given in Theorem 1, considers bandlimitedness. Whether or not a random process is bandlimited is an important question in engineering. It is shown that the Markov property (along with some technical conditions) implies nonbandlimitedness. This result is important as it sheds light on the connection between two seemingly unrelated properties of random processes, Markovity and bandlimitedness, by showing that the former implies the latter.

It is well known that bandlimited random processes can be recovered from uniformly spaced samples. An important question is: when is the converse true? The example that precedes Theorem 3 demonstrates the converse is not true in general without further conditions. Theorems 2 and 3 show that it is true in the presence of Markovity. Specifically, it is shown that continuous- and discrete-valued Markov processes (satisfying some mild technical conditions) cannot be recovered with zero average distortion (with respect to a general distortion measure) from any set of samples that has a gap between the sampling

times that occurs with positive frequency (e.g., uniform sampling).

Theorem 4 is motivated by the problem of coding continuous-time and continuous-valued random processes. A common and natural approach to coding such processes is the use of scalar quantization followed by entropy coding. One expects that for a given positive target distortion, as sampling rate increases, coding rate remains bounded. Indeed, for discrete-time processes, dithered scalar quantization plus entropy coding incurs only a small loss of no more than 0.755 bit/sample relative to the rate–distortion function [1]. In [2] it is shown that in continuous-time, coding rate tends to infinity and it is argued that in most cases distortion cannot be made to go to zero (thus making the result of interest). Bar-David [3] showed an example of a bandlimited Gaussian process that cannot be recovered without error after being passed through a binary quantizer with a threshold at the mean. Slepian [4] showed that error-free recovery is not possible for an ergodic Gauss–Markov process (this process is not bandlimited) and a binary quantizer with a threshold at the mean. Theorem 4 generalizes Slepian's result (hence strengthening the result given in [2]) by showing that error-free recovery is not possible for Markov processes in general (adhering to some mild technical conditions) with any quantizer.

Finally, the entropy rate of stationary, discrete-time, Markov processes is considered, and necessary and sufficient conditions for it to be zero are provided. This is related to information singularity.

The remainder of this paper is organized as follows. Section II introduces notation and background. Results are given in Section III. Section IV offers concluding remarks. Finally, some lemmas are left to the Appendix .

II. BACKGROUND AND NOTATION

We represent a continuous-time random process X as the collection of real-valued random variables $\{X_t(\omega), t \in (-\infty, \infty), \omega \in \Omega\}$ defined on the probability space $(\Omega, \mathcal{F}, P)$. Let X_t denote the random variable at time t, and let $X(\omega)$ denote the sample path corresponding to ω. The process X is continuous almost surely if for any time t, $\Pr(\lim_{s \to t} |X_s - X_t| = 0) = 1$ [5, p. 63].

The process X is said to be (*unilateral*) Markov if for any finite $n \geq 1$ and any $t_n < \cdots < t_1 < s < t$, X_t is conditionally independent of $X_{t_1}, \ldots, X_{t_n}$ given X_s. Namely, the future is independent of the past given the present. If X has an absolutely continuous second-order distribution, i.e., a second-order probability density function (pdf), then it has an nth-order pdf for any finite $n \geq 1$.

A *bilateral* Markov process is one where for any finite $m, n \geq 1$ and any $t_n < \cdots < t_1 < t < \tau_1 < \cdots < \tau_m$, X_t is conditionally independent of $X_{t_2}, \ldots, X_{t_n}, X_{\tau_2}, \ldots, X_{\tau_m}$ given X_{t_1} and X_{τ_1}. When second-order pdfs exist, then, as mentioned, finite-dimensional pdfs exist and a unilateral Markov process is also a bilateral Markov process [6]. For a discrete-valued Markov process this follows directly [6].

We define a scalar quantizer to be a (measurable) function $Q : \mathbb{R} \to \mathbb{Z}$. The quantizer Q induces a partition of $\mathbb{R}$ into cells $\{S_i\}$ such that $Q(x) = i$ when $x \in S_i$. Typically, these cells are intervals. A reconstruction of the process X at time t is a (measurable) function $g_t : A \to \mathbb{R}$, where A is some domain (we consider various reconstructions with various domains). Sometimes $\widehat{X}$ denotes a reconstruction of X. The operator $|\cdot|$ in addition to denoting absolute value also denotes the number of elements in a set or, if the set is an interval, then it denotes its length. If the set contains several intervals, then it denotes the number of these intervals. Its meaning is clear from the context.

Next, define the distortion measure $d(x,y) = \rho(|x - y|)$, where $\rho : [0, \infty) \to [0, \infty)$ is a continuous, nondecreasing (measurable) function for which $\rho(\alpha) > 0$ for $\alpha > 0$ and $\rho(0) = 0$.[1] We comment that the continuity of ρ is used in showing Lemma A1 of the Appendix, which in turn is used in showing Theorems 2–4. These theorems, however, hold for discontinuous ρ as well, since a continuous $\bar\rho$ can be constructed from a discontinuous ρ by setting $\bar\rho(\alpha) = 0$ when $\alpha = 0$, and $\bar\rho(\alpha) = \min\{\alpha, \frac{1}{\alpha} \int_0^\alpha \rho(s)\,ds\}$ when $\alpha > 0$. It is easy to see that $\bar\rho$ is continuous, nondecreasing, $\bar\rho(\alpha) > 0$ for $\alpha > 0$, and $\bar\rho(\alpha) \leq \rho(\alpha)$ for all $\alpha \in [0, \infty]$. The latter implies that if the theorems hold for $\bar\rho$ they also hold for ρ.

The following discussion (found in [7, pp. 231–233]) of discrete-valued Markov processes is relevant to Theorem 3.

Let $\mathcal{S} = \{1, 2, \ldots, L\}$, $1 \leq L \leq \infty$ be the state space of a discrete-time, discrete-valued, Markov chain Y, and let $V = \{v_i\}_{i=1}^L$, a set of distinct real numbers, be its alphabet. The random process X is a continuous-time, discrete-valued, Markov random process if $X_t \in V$ for all t, where we say that X_t is in state i when $X_t = v_i$, such that the amount of time X spends in state i before transitioning to another state is exponentially distributed[2] with rate λ_i (i.e., mean $\frac{1}{\lambda_i}$). The amounts of time the process spends in each state are independent random variables. The state to which the process transitions is determined according to the Markov chain Y, which is called *the underlying* Markov chain. If $\lambda_i = 0$, then state i is *absorbing*, i.e., $X_t = v_i$ implies $X_s = v_i$ for all $s > t$ with probability one. If $\lambda_i = \infty$, then state i is *instantaneous*, since once entered it is instantaneously left. (Note that we allow λ_i to be 0 or ∞, thereby referring to absorbing or instantaneous states, respectively.) The process is *time-homogenous* if its transition probabilities are invariant under shifts of time, which implies that the rates λ_i exist and are constant. For a stationary process X, let $q_i \triangleq \Pr(X_t = v_i)$, which is ordinarily different from the probability that Y is in state i.

---

[1] If $\rho(0) > 0$ is assumed, then Theorems 3 and 4 become trivial.

[2] The exponential distribution follows from the memoryless property, which is implied by the Markovity of the process. For a rigorous proof of this (which assumes that X is separable) see [8, p. 148].

Finally, the state space $\mathcal{S}$ of the Markov chain Y can be partitioned into sets called *communicating classes* such that any two states in the same communicating class have a positive probability to transition from one to the other in a finite number of transitions [7, pp. 167–174].

III. RESULTS

We begin by showing that continuous-time, continuous-valued, Markov random processes are not bandlimited. The idea is to use the fact, implied by Markovity, that when reconstructing a Markov process at any time between two sampling times, the only relevant information for the reconstruction are the two samples at these times, while all other samples provide no constraints on the reconstruction of the process at the given time. Nonbandlimitedness then follows from the inability, due to absolute continuity, to reconstruct the process without error at the given time from the two samples.

Theorem 1: A continuous-time, wide-sense-stationary, Markov random process with an absolutely continuous second-order distribution and finite second moment is not bandlimited.

Proof: If a wide-sense-stationary random process X has finite second moment and is bandlimited, then the sampling theorem for such a process [9] shows that it can be reconstructed in mean square at any time t from a set of sufficiently dense uniformly spaced samples. Namely, there exists $\tau > 0$ such that for any t, $\lim_{N \to \infty} E(X_t - \widetilde{X}_{N,t}^\tau)^2 = 0$, where $\widetilde{X}_{N,t}^\tau$ is an appropriately chosen function of t and the samples of X at times $\{-N\tau, \ldots, N\tau\}$ (e.g., $\widetilde{X}_{N,t}^\tau$ can be chosen to be a sum of sinc functions). To show that the given Markov process, denoted by X, is not bandlimited we consider the contrapositive. Namely, we show that for any sampling interval $\tau > 0$ and any reconstruction procedure there exists a time t for which mean-square convergence fails. This implies that X is not bandlimited.

Let $\tau > 0$ be given. Set $\widehat{X}_{N,t}^\tau = E[X_t | X_{-N\tau}, \ldots, X_{N\tau}]$ and observe that it is the best mean-square estimate of X_t given the samples at times $\{-N\tau, \ldots, N\tau\}$. Setting $t = \frac{\tau}{2}$ (any $t \in (0, \tau)$ will do) we have

$$\liminf_{N \to \infty} E(X_t - \widehat{X}_{N,t}^\tau)^2 \overset{(a)}{=} E(X_t - E(X_t \mid X_0, X_\tau))^2 \overset{(b)}{>} 0$$

where (a) is due to X being a bilateral Markov process since it has a second-order pdf, and (b) follows since X_0, X_t, and X_τ have an absolutely continuous joint distribution, which implies that $\Pr(X_t = E(X_s | X_0, X_\tau)) = 0$. $\square$

Before proceeding we make two definitions relevant to Theorems 2 and 3.

Definition 1: A set of sampling times $\{t_k\}_{k=-\infty}^\infty$ is *nondense* if $t_{k+1} > t_k$, $\lim_{k \to \infty} t_k = \infty$, $\lim_{k \to -\infty} t_k = -\infty$, and for some $\delta > 0$, $\liminf_{T \to \infty} \frac{|I_T^\delta|}{T} > 0$, where

$$I_T^\delta = \bigcup \{l_k : |l_k| \geq \delta, \, t_{k+1} \leq T, \, t_k \geq -T\}$$

and $l_k = (t_k, t_{k+1})$. We also write $I_T^\delta = \cup_{i=1}^{N_T^\delta} l_{k_i}$, where $N_T^\delta = |I_T^\delta|$.

Definition 2: A process X is *nonrecoverable* (with respect to $d(\cdot,\cdot)$) from the set of samples at times $\{t_k\}_{k=-\infty}^{\infty}$, if there exists $\varepsilon > 0$ such that any reconstruction of X from these samples incurs distortion greater than ε. Specifically

$$D(X, g(\{X_{t_k}\}_{k=-\infty}^{\infty}))$$
$$\triangleq \liminf_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} E[d(X_t, g_t(\{X_{t_k}\}_{k=-\infty}^{\infty}))]\, dt > \varepsilon$$

where g is any reconstruction of X from the samples at times $\{t_k\}_{k=-\infty}^{\infty}$ and $g_t(\cdot)$ is the random variable that represents its value at time t.

Theorem 2: A continuous-time, stationary, Markov random process with absolutely continuous second-order distribution is nonrecoverable (with respect to $d(\cdot,\cdot)$) from samples taken at a nondense set of times.

Proof: Let X denote the random process. Let $\delta > 0$ be such that $\liminf_{T \to \infty} \frac{|I_T^\delta|}{T} = \varepsilon_1 > 0$. It follows from Lemma A1 that there exists a set of time-invariant reconstructions $\tilde{g}_t(X_{t_{k_i}}, X_{t_{k_i+1}})$, $i \in \{1, \ldots, N_T\}$ of X at time t, each of which uses $X_{t_{k_i}}$ and $X_{t_{k_i+1}}$, respectively, such that

$$D_T(X, g(\{X_{t_k}\}_{k=-\infty}^{\infty})) \triangleq \frac{\int_{-T}^{T} E[d(X_t, g_t(\{X_{t_k}\}_{k=-\infty}^{\infty}))]\, dt}{2T}$$
$$\geq \frac{1}{2T} \int_{I_T^\delta} E[d(X_t, g_t(\{X_{t_k}\}_{k=-\infty}^{\infty}))]\, dt$$
$$\geq \frac{1}{2T} \sum_{i=1}^{N_T^\delta} \int_{t_{k_i}}^{t_{k_i+1}} E[d(X_t, \tilde{g}_t(X_{t_{k_i}}, X_{t_{k_i+1}}))]\, dt \qquad (1)$$

where the last inequality derives from Lemma A1 and the fact that X is also a bilateral Markov process. We proceed by lower-bounding for every i the last integral on the right-hand side of the last equation. Specifically, letting $\tau_i = |l_{k_i}| - \delta$ (recall that $|l_{k_i}| \geq \delta$), by Lemma A1 there exist reconstructions of X at time t, $\hat{g}_t(X_0, X_\delta, X_{\delta+\tau_i})$, and $\bar{g}_t(X_0, X_\delta)$ that use $\{X_0, X_\delta, X_{\delta+\tau_i}\}$ and $\{X_0, X_\delta\}$, respectively, such that

$$\int_{t_{k_i}}^{t_{k_i+1}} E\left[d(X_t, \tilde{g}_t(X_{t_{k_i}}, X_{t_{k_i+1}}))\right] dt$$
$$\overset{(a)}{=} \int_{0}^{|l_{k_i}|} E\left[d(X_t, \tilde{g}_t(X_0, X_{|l_{k_i}|}))\right] dt$$
$$\overset{(b)}{\geq} \int_{0}^{\delta+\tau_i} E\left[d(X_t, \hat{g}_t(X_0, X_\delta, X_{\delta+\tau_i}))\right] dt$$
$$\overset{(c)}{\geq} \int_{0}^{\delta} E\left[d(X_t, \bar{g}_t(X_0, X_\delta))\right] dt, \qquad (2)$$

where (a) is due to the stationarity of X and time invariance of $\tilde{g}$, (b) derives from the fact that $\hat{g}$ has a larger parameter input set, and (c) follows since X is a bilateral Markov process. Combining (1) and (2) we get

$$D_T(X, g(\{X_{t_k}\}_{k=-\infty}^{\infty})) \geq \frac{N_T^\delta}{2T} \int_{0}^{\delta} E\left[d(X_t, \bar{g}_t(X_0, X_\delta))\right] dt$$
$$\triangleq \frac{N_T^\delta}{2T} \varepsilon_2 \overset{(a)}{>} 0 \qquad (3)$$

where (a) follows since $E\left[d(X_t, \bar{g}_t(X_0, X_\delta))\right] > 0$ for any $t \in (0, \delta)$, which, as in the proof of Theorem 1, is due to X_0, X_t, and X_δ being jointly absolutely continuous.

Finally, using (3) we obtain

$$D(X, g(\{X_{t_k}\}_{k=-\infty}^{\infty})) = \liminf_{T \to \infty} D_T(X, g(\{X_{t_k}\}_{k=-\infty}^{\infty}))$$
$$\geq \liminf_{T \to \infty} \frac{N_T^\delta}{2T} \varepsilon_2 = \frac{\varepsilon_1 \varepsilon_2}{2} \triangleq \varepsilon > 0$$

as was needed to be shown. $\square$

The next theorem shows the same nonrecoverability result of Theorem 2 for discrete-valued Markov random processes. In both cases, the processes considered are not bandlimited. These results are significant since there exist nonbandlimited processes that can be recovered from equally spaced samples. For example, consider the following stationary random process Y constructed from a random variable U that is uniformly distributed in $[0, 1]$, and from an independent and identically distributed (i.i.d.) sequence of random variables A_k each of which is either 1 or -1 with equal probability.

$$Y_t = A_k, \quad t \in [k-1+U, k+U].$$

With probability one, any sampling interval $\tau < 1$ is an irrational multiple of U, and since the process holds fixed for at least two samples, one can determine U to arbitrary accuracy. Hence, the process is recoverable from samples taken every τ seconds. Notice that this processes is not degenerate, in the sense that it regularly produces innovations and the entropy rate of any sampled version of the process is positive.

Theorem 3: Let X be a continuous-time, discrete-valued, stationary, time-homogenous, Markov random process whose underlying Markov chain has a communicating class, which has at least two noninstantaneous states, and occurs with positive probability. The process X is nonrecoverable (with respect to $d(\cdot,\cdot)$) from samples taken at a nondense set of times.

Proof: Let C be a communicating class that occurs with positive probability, P, and has at least two noninstantaneous states. Such a class exists by assumption. Let $\{1, 2, \ldots, L\}, 2 \leq L \leq \infty$ denote the states of C. Let $X_t = v_i$ when state $i \in C$ occurs at time t, where $v_i \neq v_j$ if $i \neq j$. Let $S(X)$ denote the state of X, i.e., if $X_t = v_i$, then $S(X_t) = i$.

Let $\delta > 0$ be such that $\liminf_{T \to \infty} \frac{|I_T^\delta|}{T} = \varepsilon_1 > 0$. Using similar steps to those used in the proof of Theorem 2 to derive (3), it can be shown that

$$D_T(X, g(\{X_{t_k}\}_{k=-\infty}^{\infty}))$$
$$\geq \frac{N_T^\delta}{2T} \int_{0}^{\delta} E\left[d(X_t, \bar{g}_t(X_0, X_\delta)) \mid S(X) \in C\right] P\, dt \quad (4)$$

where $\bar{g}_t(X_0, X_\delta)$ is a reconstruction of X at time t that uses $\{X_0, X_\delta\}$.

Next, we show that there exists $\varepsilon_2 > 0$ that lower-bounds the integral in (4). Let $i, j \in C$ be two different states that are not instantaneous and let $p_{ij}(s) = \Pr(X_s = v_j \mid X_0 = v_i)$. Using

the Markovity of X and assuming, for now, that X is separable, it is not hard to see that

$$p_1 \triangleq \Pr\left(X_t = v_i, t \in \left[\frac{\delta}{4}, \frac{\delta}{2}\right] \mid X_0 = v_i, X_\delta = v_j\right)$$

$$\geq \frac{p_{ii}(\frac{\delta}{4})p_{ij}(\frac{\delta}{2})}{p_{ij}(\delta)} \int_{\frac{\delta}{4}}^\infty \lambda_i e^{-\lambda_i x} dx$$

$$\overset{(a)}{=} \frac{p_{ii}(\frac{\delta}{4})p_{ij}(\frac{\delta}{2})}{p_{ij}(\delta)} e^{-\lambda_i \frac{\delta}{4}} \overset{(b)}{>} 0$$

where (a) and (b) follow since i and j are neither instantaneous[3] nor absorbing (since $|C| > 1$), hence $0 < \lambda_i < \infty$, and (b) also derives from C having no absorbing states, which implies $p_{ij}(s) > 0$ for all $s > 0$. Similarly

$$p_2 \triangleq \Pr\left(X_t = v_j, t \in \left[\frac{\delta}{4}, \frac{\delta}{2}\right] \mid X_0 = v_i, X_\delta = v_j\right)$$

$$\geq \frac{p_{ij}(\frac{\delta}{4})p_{jj}(\frac{\delta}{2})}{p_{ij}(\delta)} e^{-\lambda_j \frac{\delta}{4}} > 0.$$

We now have

$$\int_0^\delta E\left[d(X_t, \bar{g}_t(X_0, X_\delta)) \mid S(X) \in C\right] P\, dt$$

$$\geq \int_{\frac{\delta}{4}}^{\frac{\delta}{2}} E\left[d(X_t, \bar{g}_t(X_0, X_\delta)) \mid X_0 = v_i, X_\delta = v_j, S(X) \in C\right]$$

$$\times P \frac{p_{ij}(\delta) q_i}{P} dt$$

$$\geq p_{ij}(\delta) q_i \int_{\frac{\delta}{4}}^{\frac{\delta}{2}} \inf_{k \in C} \{\rho(|v_i - v_k|)p_1 + \rho(|v_j - v_k|)p_2\}\, dt$$

$$\triangleq \varepsilon_2 > 0$$

where the last inequality follows since $p_1 > 0$, $p_2 > 0$, $v_i \neq v_j$, $\rho(\alpha) > 0$ when $\alpha > 0$, and ρ being nondecreasing.[4] Combining the last equation with (4) we obtain

$$D(X, g(\{X_{t_k}\}_{k=-\infty}^\infty)) = \liminf_{T \to \infty} D_T(X, g(\{X_{t_k}\}_{k=-\infty}^\infty))$$

$$\geq \liminf_{T \to \infty} \frac{N_T^\delta}{2T} \varepsilon_2 = \frac{\varepsilon_1 \varepsilon_2}{2} \triangleq \varepsilon > 0$$

which shows the theorem for the case that X is separable. Finally, if X is not separable, then there exists an equivalent separable random process Y, i.e., $\Pr(Y_t = X_t) = 1$ for all t[10, pp. 88–89], for which the theorem does hold. This implies that the theorem holds for X as well. $\qquad \square$

The next theorem shows that when the Markov process considered in Theorem 1 is stationary and continuous almost surely, it cannot be recovered without error from its quantized version. We comment that stationary, separable Gauss–Markov processes are examples of processes that have absolutely continuous second-order distribution (as follows from the Gauss–Markov property) and are continuous almost surely (in fact, they are almost surely sample path continuous as can be verified using the Kolmogorov condition for sample continuity [5, p. 57]), [11, p. 519].

---

[3] States i and j are required to not be instantaneous so as to allow p_1 and p_2 (that follows) to be positive. It is clear that if i and j are instantaneous, then this should make reconstruction even harder and so the theorem would hold in this case as well.

[4] Instead of requiring ρ to be nondecreasing it suffices to require that for some $m > 0$ and $s > 0$, $\rho(x) \geq m$ for all $x \geq s$.

Theorem 4: Let X be a continuous-time, stationary, Markov random process that is continuous almost surely and has absolutely continuous second-order distribution. Let Q be an arbitrary scalar quantizer that has a cell that contains an interval, and X is in this interval with positive probability. Let X_Q denote the quantized version of X, i.e., $X_{Q,t} = Q(X_t)$. Then the distortions (with respect to $d(\cdot, \cdot)$) induced by all reconstruction schemes are bounded away from zero. Specifically

$$\inf_g D(X, g(X_Q))$$

$$\triangleq \inf_g \liminf_{T \to \infty} \frac{1}{2T} \int_{-T}^T E[d(X_t, g_t(X_Q))]\, dt > 0$$

where g is any reconstruction of X from $\{X_{Q,t}\}_{t=-\infty}^\infty$ (i.e., g has as input the entire quantized process) and $g_t(X_Q)$ is the random variable that represents its value at time t.

Proof: Lemma A1 of the Appendix shows that there exists a time-invariant best estimator g^*. This and the stationarity of X imply that for any time s

$$\inf_g D(X, g(X_Q)) \geq \liminf_{T \to \infty} \frac{1}{2T} \int_{-T}^T E[d(X_t, g_t^*(X_Q))]\, dt$$

$$= E[d(X_s, g_s^*(X_Q))]. \qquad (5)$$

Before proceeding to show that the last term above is positive, consider an interval $W = (a, b)$ that occurs with positive probability and is contained by a quantization cell of Q. We show that there exists $\tau > 0$ such that $\Pr(X_t \in W, \text{ for all } 0 \leq t \leq \tau) > 0$ or, more precisely, that there exists $G_\tau \subseteq \Omega$ such that $P(G_\tau) > 0$ and $X_t(\omega) \in W$ for all $0 \leq t \leq \tau$ and all $\omega \in G_\tau$.

Since $\Pr(X_0 \in W) > 0$ and since X_0 has an absolutely continuous distribution, it follows that there exists $W' = (a + \nu, b - \nu)$ for some $\nu > 0$ such that $q \triangleq \Pr(X_0 \in W') > 0$. Further, since X is continuous almost surely, $\Pr(X \text{ is continuous at } 0) = 1$. Thus, let

$$\bar{\Omega} = \{\omega \in \Omega : X(\omega) \text{ is continuous at } 0 \text{ and } X_0(\omega) \in W'\}.$$

We have that $P(\bar{\Omega}) = q$. Set

$$G_\delta = \{\omega \in \bar{\Omega} : X_t(\omega) \in W \text{ for all } 0 \leq t \leq \delta\}.$$

We notice that the continuity of $X(\omega), \omega \in \bar{\Omega}$ at 0 implies that for every $\omega \in \bar{\Omega}$ there exists $\delta > 0$ such that $\omega \in G_\delta$. We observe further that $P(G_0) = P(\bar{\Omega}) = q > 0$ and that if $\delta_1 > \delta_2$, then $G_{\delta_1} \subseteq G_{\delta_2}$, i.e., the G sets are nested. Consequently, $\lim_{n \to \infty} G_{\frac{1}{n}} = G_0$, and by continuity of probability there exists $N > 0$ such that $P(G_{\frac{1}{n}}) > \frac{P(G_0)}{2} = \frac{q}{2} > 0$ for all $n \geq N$. Setting $\tau = \frac{1}{N}$ we obtain $P(G_\tau) > \frac{q}{2} > 0$.

We now proceed with the proof of the main statement. Setting $s = \frac{\tau}{2}$ (any $s \in (0, \tau)$ will do), we have that there exist reconstructions of X at time t, $\tilde{g}_t(X_Q, X_0, X_\tau)$, $\hat{g}_t(\{X_{Q,t}\}_{t=0}^\tau, X_0, X_\tau)$, and $\bar{g}_t(X_0, X_\tau)$ that use $\{X_Q, X_0, X_\tau\}$, $\{\{X_{Q,t}\}_{t=0}^\tau, X_0, X_\tau\}$, and $\{X_0, X_\tau\}$, respectively, such that

$$E[d(X_s, g_s^*(X_Q))] \overset{(a)}{\geq} E[d(X_s, \tilde{g}_s(X_Q, X_0, X_\tau))]$$

$$\overset{(b)}{=} E[d(X_s, \hat{g}_s(\{X_{Q,t}\}_{t=0}^\tau, X_0, X_\tau))]$$

$$\overset{(c)}{\geq} E[d(X_s, \bar{g}_s(X_0, X_\tau)) \mid G_\tau] P(G_\tau)$$

where (a) follows since $\tilde{g}$ has a larger parameter input set, (b) derives from X being a bilateral Markov process[5] since it has a second-order pdf, and (c) follows since $\{X_{Q,t}\}_{t=0}^{\tau}$ is constant given G_τ. Lemma A1 implies that there exists a best estimator $\bar{g}^*$ that minimizes $E[d(X_s, \bar{g}_s(X_0, X_\tau)) \,|\, G_\tau] P(G_\tau)$. This, (5), and the last equation imply that all that remains to show is that

$$E\left[\rho(|X_s - \bar{g}_s^*(X_0, X_\tau)|) \,\big|\, G_\tau\right] P(G_\tau) > 0.$$

As in the proof of Theorem 1, the expectation above is positive since X_0, X_s, and X_τ are jointly absolutely continuous (a fact that remains true also when conditioning on G_τ, since $P(G_\tau) > 0$) and $\rho(\alpha) > 0$ when $\alpha > 0$. $\square$

We conclude this section with a brief discussion tieing Markov processes, entropy rate, and information singularity. Consider the following definition.

Definition 3: A random process X has a *degenerate* two-dimensional distribution if for any time t, any number τ, and almost all v, $\Pr(X_{t+\tau} = u \,|\, X_t = v) = 1$ for some u that can depend on v.

It is not hard to show that this condition is necessary and sufficient for stationary, discrete-time, Markov processes to have zero entropy rate (Markovity is essential since the condition is only on the two-dimensional distribution). It then follows directly via Hajek [13] (Theorem 2 therein) that a stationary, discrete-time, Markov random process with finite second moment is information-singular, in the sense of [13] and [14], if and only if it has a degenerate two-dimensional distribution.

The notion of information singularity can be extended in a natural way to continuous-time processes [15]. For such extension it can be shown that the Markov process mentioned (except for being continuous time and also measurable), is not information-singular if it has a nondegenerate two-dimensional distribution.[6]

IV. CONCLUSION

It was shown that continuous-time Markov processes are not bandlimited and that they cannot be recovered without error from a uniform set of samples or after quantization. Further, necessary and sufficient conditions were provided for discrete-time Markov processes to have zero entropy rate. This was related to information singularity.

APPENDIX

Lemma A1: Given a set of times T, let $f(X_{t \in T-a})$ be a deterministic function of the stationary random process X at times $t \in T - a$, which is defined for all a, where $T - a$ denotes shift by a of the set T. For any time s, there exists a time-invariant

best estimator of X at time s, g_s^*, that uses $f(X_{t \in T})$ and attains the infimum $\inf_{g_s} E[d(X_s, g_s(f(X_{t \in T}))]$.

Proof: We first show that there exists a best estimator. Recalling that $d(x, y) = \rho(|x - y|)$, and that ρ is a nondecreasing function, Lemma A2 below implies that for any observed value of $f(X_{t \in T})$ (or $f(X_T)$ for short), the infimum $\inf_a E[\rho(|X_s - a|) \,|\, f(X_T) = f(x_T)]$ can be taken over a compact set. Since ρ is continuous, it is easy to see that $E[\rho(|X_s - \alpha|) \,|\, f(X_T) = f(x_T)]$ is continuous in α. Hence, the infimum $\inf_\alpha E[\rho(|X_s - \alpha|) \,|\, f(X_T) = f(x_T)]$ is a minimum, which implies that the infimum $\inf_{g_s} E[\rho(|X_s - g_s(f(X_T))]$ is a minimum.

This shows that for any s there exists a best estimator of X_s, g_s^*, that attains the infimum $\inf_{g_s} E[d(X_s, g_s(f(X_T))]$. The stationarity of X implies that for any s and a, $g_{s-a}^*(f(X_{T-a})) = g_s^*(f(X_T))$, i.e., g^* is time-invariant. $\square$

Lemma A2: If $\rho : [0, \infty) \to [0, \infty)$ is a nondecreasing function and X is a random variable, then $\inf_y E\rho(|X - y|) = \inf_{y \in [a,b]} E\rho(|X - y|)$ for some finite interval $[a, b]$.

Proof: Let $\inf_y E\rho(|X - y|) = \varepsilon$. If $\varepsilon = \infty$, then the interval $[a, b] = [0, 1]$ can be chosen. Suppose $\varepsilon < \infty$ and let $M > \varepsilon$ be such that $\rho(s) \geq M$ for some $s > 0$ (and, consequently, $\rho(x) \geq M$ for all $x \geq s$ since ρ is nondecreasing). If such M does not exist, then choosing $[a, b] = [0, 1]$ suffices. Let $\delta < M - \varepsilon$ be positive and set $m = \frac{\varepsilon + \delta}{M} < 1$. Choose the interval $[T_a, T_b]$ so that $\Pr(X \in [T_a, T_b]) \geq m$. Set $b = T_b + s$ and $a = T_a - s$. For any $x \in [T_a, T_b]$ and $y \in (-\infty, a] \cup [b, \infty)$ we have $|x - y| \geq s$. Thus, for all $y \in (-\infty, a] \cup [b, \infty)$

$$E\rho(|X - y|) \geq E\left[\rho(|X - y|) \,|\, X \in [T_a, T_b]\right] \Pr(X \in [T_a, T_b])$$
$$\geq M \cdot m = \varepsilon + \delta$$

which implies that the infimum need only be taken over the finite interval $[a, b]$. $\square$

---

[5]Given X_0 and X_τ, X_s is independent of the uncountable collection $\{X_t\}_{t<0, t>\tau}$. While typically such independence is stated with respect to a countable collection, it holds in the uncountable case also as seen in [11, pp. 562–565],[12, pp. 393–395], where this is given for unilateral processes but holds for bilateral processes as well.

[6]Theorem 9 in [15] has an incomplete proof of this as it makes a small error by considering the condition "not constant with probability one" on the distribution of the process, instead of the condition "nondegenerate two-dimensional distribution."

REFERENCES

[1] J. Ziv, "On universal quantization," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 344–347, May 1985.
[2] D. Marco and D. L. Neuhoff, "Entropy of highly correlated quantized data," *IEEE Trans. Inf. Theory*, to be published.
[3] I. Bar-David, "Sample functions of a Gaussian process cannot be recovered from their zero crossings," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 86–87, Jan. 1975.
[4] D. Slepian, "Estimation of the Gauss-Markov process from observation of its sign," *Stoch. Proc. Appl.*, vol. 14, pp. 249–265, 1983.
[5] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*. New York: Springer-Verlag, 1985.
[6] K. Abend, T. J. Harley, and L. N. Kanal, "Classification of binary random patterns," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 4, pp. 538–544, Oct. 1965.
[7] S. Ross, *Stochastic Processes*, 2nd ed. New York: Wiley, 1981.
[8] K. L. Chung, *Markov Chains with Stationary Transition Probabilities*. Berlin, Germany: Springer-Verlag, 1967.
[9] A. V. Balakrishnan, "A note on the Shannon sampling principle for continuous signals," *IRE Trans. Inf. Theory*, vol. IT-3, no. 3, pp. 143–146, Jun. 1957.
[10] J. Neveu, *Mathematical Foundations of the Calculus of Probability*. San Francisco, CA: Holden-Day, 1965.
[11] M. Loève, *Probability Theory*, 3rd ed. New York: Van Nostrand, 1963.

[12] H. Bauer, *Probability Theory and Elements of Measure Theory*, 2nd ed.   New York: Academic, 1981.

[13] B. Hajek, "Information-singularity and recoverability of random processes," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 3, pp. 422–429, May 1982.

[14] T. Berger, "Information-singular random processes," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 5, pp. 502–511, Sep. 1975.

[15] D. Marco, "Markov random processes are not recoverable after quantization and mostly not recoverable from samples," in *IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 2886–2890.

Daniel Marco (S'02–M'04) received the B.Sc. degree in computer engineering from the Technion–Israel Institute of Technology, Haifa, Israel, in 1999, and the M.S. degree in electrical engineering, the M.S. degree in mathematics, and the Ph.D. degree in electrical engineering, all from the University of Michigan, Ann Arbor, in 2001, 2003, and 2004, respectively.

From 2004 to 2006, he has been with the California Institute of Technology, Pasadena, as a Postdoctoral Scholar. His research interests lie primarily in information theory, high- and low-resolution quantization theory, rate–distortion theory, coding with side information, and sensor networks.