

# Encoding the Sojourn Times of a Markov Process with a Fidelity Criterion

Rajesh Sundaresan and Sergio Verdú  
Dept. of Electrical Engineering  
Princeton University  
Princeton, NJ 08544.

## Abstract

*The exponential distribution and the Gaussian distribution play a prominent role in queueing theory and communication theory, respectively. The exponential distribution leads to certain information theoretic quantities that are similar to their Gaussian counterparts. In this paper we look at one such similarity.*

*The sojourn times of a Markov process are exponentially distributed, given the knowledge of the state of the process. The problem of encoding the sojourn times is considered for two scenarios.*

*In the first case, the encoder is assumed to have full knowledge of the state of the process. The rate-distortion function according to a fidelity criterion is found and its similarity to its Gaussian counterpart - the water-flooding solution for simultaneous description of independent Gaussian random variables - is discussed.*

*In the second case, the encoder does not have access to the state information, while the decoder has. This is a generalization of the Wyner-Ziv problem ([11]) for stationary and ergodic sources taking values on a continuous alphabet source. We give an upper bound and a lower bound on the rate-distortion function according to a fidelity criterion.*

## 1 Introduction

The exponential distribution leads to certain expressions and properties which are very similar to those in the Gaussian case. To name some: among all mean-constrained positive random variables, the exponentially distributed random variable maximizes the differential entropy; it satisfies a mutual-information saddle-point property analogous to the Gaussian case [9]; the rate-distortion function for encoding the interarrival times of a Poisson process according to a fi-

delity criterion is very similar to that of encoding i.i.d. Gaussian random variables [9].

Consider a homogeneous Poisson process of rate  $\lambda$ . The interarrival times constitute an independent and identically exponentially distributed process with mean  $1/\lambda$ . Suppose we wish to encode these interarrival times with a finite number of bits per second. Encoding the interarrival times is equivalent to encoding a memoryless source with exponential interarrival times. For a chosen fidelity criterion, the rate-distortion function reflects the optimal tradeoff between the encoding rate and the achievable fidelity.

This problem was considered in [8] with a distortion measure equal to the normalized absolute error between the true and the reproduced interarrival times. This fidelity criterion however does not lead to a closed form expression for the rate-distortion function.

An alternative fidelity criterion was given in [9] which indeed led to a closed form expression similar to that of the Gaussian source. The fidelity criterion was that the reproduced interarrival times do not exceed the true value by more than a prescribed amount  $d$ , and that the last arrival be declared only after it occurs. As we will see later, this fidelity criterion is very similar to the absolute-error criterion of [8], except that the reproduced values are constrained to lie on one side of the actual value.

In this paper we consider the problem of encoding a continuous-time Markov process. To encode the Markov process, we encode the sequence of states and the sequence of sojourn times separately. If the jump chain (sequence of states) is encoded at its entropy rate, it can be recovered noiselessly at the decoder. In this paper, we will extend the result of [9] to the problem of encoding the sojourn times of a Markov process, with a fidelity criterion similar to that in [9]. We consider two scenarios, one in which the encoder

for the sojourn time process has access to the state information and the other in which the encoder has no knowledge of the state. The decoder in both cases is assumed to have full knowledge of the state process.

In the next section we present some preliminaries required to state the results. Following this we present the rate-distortion function for encoding the sojourn times when state information is known at the encoder. We compare this solution with the water flooding solution for simultaneous description of independent Gaussian random variables.

In the last section, we look at the rate-distortion function when the encoder does not have access to the state information. This problem is a generalization of the Wyner-Ziv problem for a specific type of a stationary and ergodic sources, taking values on continuous alphabet space. For this case we have not been able to get a closed form expression for the rate-distortion function. We give an upper bound and a lower bound on the rate-distortion function.

## 2 Preliminaries

Let  $I = \{1, 2, \dots\}$  be an index set and  $\mathcal{R}_+$  be the set of positive real numbers  $[0, \infty)$ . Let  $\{V(t) : t \in \mathcal{R}_+\}$  denote a continuous-time, irreducible, Markov process on a finite state space  $\mathcal{S}$ ,  $|\mathcal{S}| < \infty$ . Let  $\{\lambda_{ij} : i, j \in \mathcal{S}, i \neq j\}$  denote the set of transition rates of this Markov process, and  $\{q_i : i \in \mathcal{S}\}$ , where  $\sum_{i \in \mathcal{S}} q_i = 1$ , the unique stationary distribution of the states.

Assume that a transition occurs at time  $t = 0^-$ . Let  $\{Y_k : k \in I\}$  denote the jump chain formed by this Markov process. That is,  $Y_1 = V(0)$ ,  $Y_2$  is the next state the process  $\{V(t) : t \in \mathcal{R}_+\}$  moves to after time  $t = 0$ ,  $Y_3$  is the next state after that, and so on.  $Y_k$  also takes values on  $\mathcal{S}$ . This jump chain is a discrete-time Markov chain whose transition probabilities are given by

$$\begin{aligned} p_{ji} &= \frac{\lambda_{ji}}{\lambda_j}, \quad j, i \in \mathcal{S}, j \neq i, \\ p_{jj} &= 0, \quad j \in \mathcal{S} \end{aligned}$$

where  $\lambda_j = \sum_{i \in \mathcal{S} - \{j\}} \lambda_{ji}$ ,  $j \in \mathcal{S}$ . We also have that

$$B^{-1} = \sum_{j \in \mathcal{S}} q_j \lambda_j < \infty \quad (1)$$

and that

$$\pi_j = B q_j \lambda_j \quad (2)$$

is an equilibrium distribution for the jump chain ([7], Exercise 1.1.5). We note here that the jump chain in

general might not be an aperiodic Markov chain. But the above distribution is indeed the unique stationary distribution.

For convenience, we will restrict ourselves to only those processes that lead to jump chains which are irreducible and aperiodic Markov chains. This condition precludes, for example, two-state Markov processes, and Markov processes whose associated undirected graphs are trees. This is because in the jump chain jumps to the same state are not allowed, and hence to return to a state, an even number of jumps are required. This will ensure that the distribution of the states will eventually converge to the equilibrium distribution regardless of its initial distribution. Further, we need this condition in the proof of the result in Section 4.

Let  $\{X_k : k \in I\}$  denote the sequence of sojourn times, where  $X_k$  is the sojourn time in state  $Y_k$  at (discrete) time  $k$ .  $X_k$  takes values in  $\mathcal{R}_+$ . The process  $\{X_k : k \in I\}$  has memory. But given the state  $Y_k = j$ ,  $X_k$  is independent of all other random variables and is exponentially distributed with parameter  $\lambda_j$  ([6]). We can think of the process  $\{X_k : k \in I\}$  as obtained by a memoryless random transformation (a memoryless channel) characterized by  $P_{X|Y}(x|j) = \lambda_j e^{-\lambda_j x}$ , from the input process  $\{Y_k : k \in I\}$ . Our results of Section 4 can be shown to hold for some special semi-Markov processes. These processes have to satisfy condition  $G$  defined below.

**Definition 1** *The joint process  $\{(Y_k, X_k) : k \in I\}$  is said to satisfy condition  $G$  if  $\{Y_k : k \in I\}$  is an irreducible, aperiodic, finite-state Markov chain, and if  $\{X_k : k \in I\}$  is a conditionally independent real-valued process obtained from  $\{Y_k : k \in I\}$ .*

In particular, these processes are hidden Markov processes whose embedded Markov chain  $\{Y_k : k \in I\}$  is ergodic.

Let  $T$  denote the (left) shift transformation on the  $\{Y_k : k \in I\}$  process.  $T$  is said to be ergodic with respect to the probability law of the  $\{Y_k : k \in I\}$  process if every shift invariant event has probability 0 or 1. With a slight abuse of notation we will let  $T$  denote the left shift transformation on the  $\{X_k : k \in I\}$  process and the joint process  $\{(Y_k, X_k) : k \in I\}$  as well.

For the finite state Markov chain  $\{Y_k : k \in I\}$ ,  $T$  is ergodic since the chain is irreducible ([3]). Using

this, the stationarity of the Markov chain and ([2], Theorem 7.2.1), it can easily be shown that the joint process  $\{(Y_k, X_k) : k \in I\}$  is a stationary and ergodic process, if it satisfies the condition  $G$ .

### 3 Encoding the Sojourn times of a Markov Process

Consider a Poisson process with rate  $\lambda$  arrivals per second. To encode this process, we encode the interarrival times of this process. Equivalently, we encode the outputs of a memoryless source, each of which is exponentially distributed with mean  $1/\lambda$  seconds.

By an  $(n, M, \epsilon)$ -code, we mean a set  $\{(\hat{x}^n)_m\}_{m=1}^M$  and a mapping  $f : \mathcal{R}_+^n \rightarrow \{(\hat{x}^n)_m\}_{m=1}^M$  such that the following criteria are met.

1. For every possible realization of the sequence of interarrival times,  $x^n$ ,

$$\hat{x}_i \leq x_i, \quad \text{for } i = 1, \dots, n, \quad (3)$$

$$2. \quad P \left( \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i) > d \right) = \epsilon. \quad (4)$$

This fidelity criterion is similar to that of [8], except that the reproduced values are constrained not to exceed the actual value.

Let  $W_n = \sum_{i=1}^n X_i$  be the time of the last arrival. A reasonable definition of the rate of the code in (nats per second) is  $(\log M)/W_n$  which behaves as  $\lambda(\log M)/n$  as  $n \rightarrow \infty$ . Therefore we define the rate of the code as follows.

**Definition 2**  $R$  is an achievable rate (in nats/sec) for distortion parametrized by  $d$  if for every  $\delta > 0$ , there exists a sequence of  $(n, M, \epsilon_n)$ -codes with

$$\frac{\log M}{n} \leq \frac{R}{\lambda} + \delta, \quad \text{for all sufficiently large } n, \quad (5)$$

$$\text{and} \quad \lim_{n \rightarrow \infty} \epsilon_n = 0. \quad (6)$$

**Definition 3** The rate-distortion function  $R(d)$  is the smallest achievable encoding rate (in nats/sec) for distortion  $d$ .

An  $(n, M, \epsilon)$ -code with respect to the criteria (3) and (4) is equivalent to an  $(n, M, \epsilon)$ -code with respect to the following two criteria:

1. For every possible realization of the sequence of interarrival times,

$$\hat{x}_i \leq x_i + d, \quad \text{for } i = 1, \dots, n, \quad (7)$$

$$2. \quad P \left( \sum_{i=1}^n \hat{X}_i < \sum_{i=1}^n X_i \right) = \epsilon. \quad (8)$$

That is, we require that the reproduced interarrival times not exceed the true value by more than  $d$ , and that the last arrival be declared only after it occurs with high probability.

**Theorem 1** :[9]. For the Poisson process with rate  $\lambda$ , the rate-distortion function according to the fidelity criterion described by (3) and (4) (or equivalently (7) and (8)) is given by

$$\lambda^{-1} R(d) = \begin{cases} \log \left( \frac{1}{\lambda d} \right), & \text{if } d \leq \frac{1}{\lambda}, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

This result was used to find the rate-distortion function of a continuous-time, irreducible, finite-state Markov process  $\{V(t) : t \in \mathcal{R}_+\}$ . Let this Markov process give rise to the joint state and sojourn time process given by  $\{(Y_k, X_k) : k \in I\}$  that satisfies condition  $G$ . Then this joint process is both stationary and ergodic. In order to encode the continuous-time Markov process, we first encode the sequence of states so that it can be recovered at the decoder noiselessly. This takes a number of nats per second which equals  $\lambda$  times the entropy rate of the jump chain  $\{Y_k : k \in I\}$ . The problem now is to encode the sojourn times, assuming that the decoder knows the true state sequence.

Now, by an  $(n, M, \epsilon)$ -code, we again mean a set of  $M$  codewords  $\{(\hat{x}^n)_m\}_{m=1}^M$  and a mapping  $f : \mathcal{S}^n \times \mathcal{R}_+^n \rightarrow \{(\hat{x}^n)_m\}_{m=1}^M$  such that the following criteria are met.

1. For every possible realization of the sequence of  $n$  interarrival times  $x^n$ , and state sequence  $y^n$ ,

$$\hat{x}_i \leq x_i, \quad \text{for } i = 1, \dots, n, \quad (10)$$

2.

$$P \left( \frac{1}{N_j} \sum_{i=1}^n (X_i - \hat{X}_i) 1_{\{Y_i=j\}} > d_j, \text{ for some } j \in \mathcal{S} \right) = \epsilon, \quad (11)$$

where  $N_j = \sum_{i=1}^n 1_{\{Y_i=j\}}$  is the number of occurrences of the state  $j$  within the first  $n$  jumps. The criterion in (11) indicates a tolerance of  $d_j$  to encode the sojourn times when in state  $j$ .

Analogous to Definitions 2 and 3, we define the achievable rate (in nats/sec) for distortion parametrized by  $(d_1, \dots, d_{|\mathcal{S}|})$  and the rate-distortion

function  $R_{X|Y}^i(d_1, \dots, d_{|S|})$ . Now  $B^{-1}$  in (1), the average number of transitions per second, plays the role of  $\lambda$  in the definitions. Since both the encoder and decoder know the true sequence of states, the problem can be viewed as the decoupled encoding of  $|S|$  different sources, each of which is a random process of independent exponentially distributed random variables with mean  $\lambda_j^{-1}$  seconds.

**Theorem 2** :[9]. *If  $\{(Y_k, X_k) : k \in I\}$  satisfies the condition G, then*

$$B \cdot R_{X|Y}^i(d_1, \dots, d_{|S|}) = \sum_{i \in S} \pi_i \left[ \log \left( \frac{1}{\lambda_i d_i} \right) \right]^+, \quad (12)$$

where  $[\cdot]$  stands for the positive function of the argument.

Using (2) we can rewrite (12) as

$$R_{X|Y}^i(d_1, \dots, d_{|S|}) = \sum_{i \in S} q_i \lambda_i \left[ \log \left( \frac{1}{\lambda_i d_i} \right) \right]^+.$$

The criterion in (11) defines a tolerance level for the sojourn times from *each* of the states. Instead of this criterion, suppose that the fidelity criterion does not distinguish between sojourn times from the various states. In particular, we require instead of (11) that

$$P \left( \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i) > d \right) = \varepsilon. \quad (13)$$

Note that the above criterion in conjunction with (10) is the same as criteria (3) and (4), and is thus equivalent to criteria (7) and (8).

We define codes, achievable rates for distortion parametrized by  $d$ , and the rate-distortion function  $R_{X|Y}(d)$  analogously to Definitions 2 and 3. Again,  $B^{-1}$  in (1), the average number of transitions per second, plays the role of  $\lambda$  in the definitions.

Loosely speaking, we need to allocate our resources appropriately and arrive at allowable distortions for encoding the sojourn times in each state so that criterion (13) can be met in an efficient way. The following theorem gives  $R_{X|Y}(d)$  according to this fidelity criterion and also the efficient allocation of resources.

**Theorem 3** *If  $\{(Y_k, X_k) : k \in I\}$  satisfies the condition G, then*

$$B \cdot R_{X|Y}(d) = \sum_{i \in S} \pi_i \left[ \log \left( \frac{1}{\lambda_i d_i} \right) \right]^+,$$

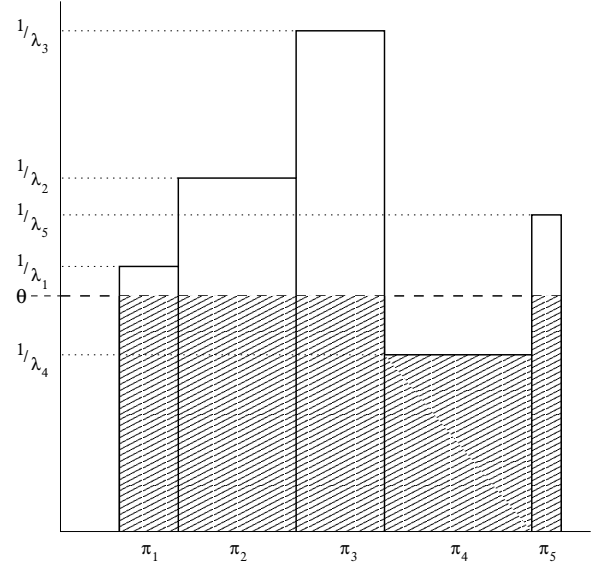


Figure 1: Water-flooding solution.

$$\text{where } d_i = \begin{cases} \frac{1}{\lambda_i}, & \text{if } \theta > \frac{1}{\lambda_i}, \\ \theta, & \text{otherwise,} \end{cases}$$

and  $\theta$  is the water-level given by the solution of

$$\sum_{i \in S} \pi_i d_i = d. \quad (14)$$

**Remarks:** 1. Using (2) we can rewrite  $R_{X|Y}(d)$  as

$$R_{X|Y}(d) = \sum_{i \in S} q_i \lambda_i \left[ \log \left( \frac{1}{\lambda_i d_i} \right) \right]^+.$$

2.  $R_{X|Y}(d)$  in Theorem 3 is a convex function of  $d$  for  $d > 0$ . Further, for  $0 < d < d_{max} = \sum_{i \in S} \pi_i / \lambda_i$ , the derivative of  $R_{X|Y}(d)$  with respect to  $d$  is continuous. These two properties are easy to verify.

3. The pair  $(0, \sum_{i \in S} \pi_i / \lambda_i)$  is an achievable rate-distortion pair. That is, when we are allowed a distortion of  $d = \sum_{i \in S} \pi_i / \lambda_i$ , we can achieve reliable compression with rates as close to 0 as we wish.

4. Theorem 3 says that the solution is one where we do not encode the fast states at all (states  $i$  with  $\theta > 1/\lambda_i$ ). All slow states (states  $i$  with  $\theta \leq 1/\lambda_i$ ) are assigned the same distortion level  $\theta$ . The solution is reminiscent of the rate-distortion function for simultaneous description of independent Gaussian random variables ([5], Theorem 13.3.3). The main difference here is that only one state is active at a time. So the distortions for each state are weighted appropriately by their probability so as to find their contribution to the total distortion.

## 4 The Wyner-Ziv Problem

In this section we study the case when the encoder does not have access to the state information, while the decoder has full access to this information. The Slepian-Wolf problem deals with noiseless separate encoding of a correlated, finite alphabet source (each vector  $(Y_k, X_k)$  is drawn in an i.i.d. fashion). Wyner and Ziv [11] dealt with the problem of encoding one of the sources noiselessly, and studied the corresponding rate-distortion problem of encoding the other process. Wyner [10] extended the result to sources taking values on arbitrary alphabet spaces with a slight restriction on the choice of the distortion function. His result is that if  $\{Y_k : k \in I\}$  is the side-information process and if  $\{X_k : k \in I\}$  is to be encoded according to a single letter, additive, expected distortion criterion, then the rate-distortion function is given by (if  $I(X; Y) < \infty$  and for certain smooth distortion functions) :

$$R(d) = \inf (I(X; Z) - I(Y; Z))$$

where the infimum is taken over all joint distributions  $P_{Y,X,Z}$  such that the following two conditions hold:

1.  $Y \rightarrow X \rightarrow Z$  constitutes a Markov chain,
2. There exists a measurable function  $f : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{R}_+$  such that  $E\{\rho(X, f(Y, Z))\} \leq d$ ,

where  $Z$  takes values in an alphabet space  $\mathcal{A}$  (standard measurable space). It is remarked in [10] that this rate-distortion function is in most cases strictly greater than  $R_{X|Y}(d)$ , the rate-distortion function when the encoder has full knowledge of the side-information process. Equality is known to be achieved in some special cases, for example, when  $\{(Y_k, X_k) : k \in I\}$  is jointly Gaussian. In other words, side information does not pay off in terms of rate for the Gaussian case, but we might require a more complex encoder. Our aim is to see if we can bound this Wyner-Ziv loss for the problem we are dealing with, namely, encoding the sojourn times of a Markov process.

We remark that  $\{(Y_k, X_k) : k \in I\}$  in our problem is a process with memory, and hence our results are not implied by those of [11] or [10]. In Section 3 we dealt with a fidelity criterion which required that the probability that the distortion criteria are not met vanish as  $n \rightarrow \infty$ . Using the same fidelity criterion for this problem we have only been able to show that there exists an infinite set of  $n$  for which we can find codes which have the desired rate and vanishing probabilities for  $n$  belonging to this set.

In what follows, we deal with the single-letter expected distortion criterion. We first make precise the problem statement. We follow the notation and definitions of [10].

Let  $\{(Y_k, X_k) : k \in I\}$  be a process that satisfies condition  $G$  described in Definition 1. Let the distortion function be defined as

$$\rho(x, \hat{x}) = |x - \hat{x}|, \quad (15)$$

where  $\hat{x}$  is a reproduction of  $x$ . Also define

$$\rho_n(x^n, \hat{x}^n) \triangleq \frac{1}{n} \sum_{i=1}^n \rho(x_i, \hat{x}_i). \quad (16)$$

An  $(n, M, \Delta)$ -code is a pair of measurable mappings  $f_E$  and  $f_D$ , called the encoder and the decoder given by

$$\begin{aligned} f_E &: \mathcal{R}_+^n \rightarrow \mathcal{J}_M = \{1, 2, \dots, M\}, \\ f_D &: \mathcal{S}^n \times \mathcal{J}_M \rightarrow \mathcal{R}_+^n, \text{ and} \\ \hat{X}^n &= f_D(Y^n, f_E(X^n)) \end{aligned}$$

$$\text{and} \quad E \frac{1}{n} \sum_{k=1}^n \rho(X_k, \hat{X}_k) = \Delta. \quad (17)$$

In this section, we take the unit of rate to be nats per transition. A pair  $(R, d)$ ,  $R \geq 0, d \geq 0$  is said to be achievable if for an arbitrary  $\varepsilon > 0$ , there exists an  $(n, M, \Delta)$ -code with

$$\frac{\log M}{n} \leq R + \varepsilon, \quad \text{and} \quad \Delta \leq d + \varepsilon. \quad (18)$$

Define  $\mathcal{H}$  to be the set of all achievable  $(R, d)$  pairs. Also define

$$R(d) = \min_{(R,d) \in \mathcal{H}} R, \text{ for } d \geq 0. \quad (19)$$

It follows directly from the definition of  $\mathcal{H}$  that if  $\lim_{k \rightarrow \infty} R_k = R$  and if  $(R_k, d) \in \mathcal{H}$ , then  $(R, d) \in \mathcal{H}$ . This justifies the use of minimum instead of infimum in (19). Further it is shown in [10] that  $R(0) = \lim_{d \rightarrow 0} R^*(d)$ .

Let  $Z$  take values in the standard measurable space  $(\mathcal{A}, \mathcal{B}_{\mathcal{A}})$ . Let  $\mathcal{M}_n$  be the class of all random variables  $Z$  such that  $Y^n \rightarrow X^n \rightarrow Z$  form a Markov chain. Let  $\mathcal{M}_n(d)$  be the class of all random variables in  $\mathcal{M}_n$  such that there exists a measurable function  $g_n : \mathcal{S}^n \times \mathcal{A} \rightarrow \mathcal{R}_+^n$  with  $E\rho_n(X^n, g_n(Y^n, Z)) \leq d$ .

Finally we define the following:

$$\begin{aligned} R_1^*(d) &\triangleq \inf_{Z \in \mathcal{M}_1(d)} (I(X; Z) - I(Y; Z)), \\ R_n^*(d) &\triangleq \frac{1}{n} \inf_{Z \in \mathcal{M}_n(d)} (I(X^n; Z) - I(Y^n; Z)), \text{ and} \\ R^*(d) &\triangleq \lim_{n \rightarrow \infty} R_n^*(d). \end{aligned}$$

**Remarks:** 1. First note that since  $Y^n$  takes values on a finite alphabet space, we have  $I(Y^n; X^n) < \infty$ , and thus by the data processing lemma  $I(Y^n; Z) < \infty$ . Therefore, the first two definitions are unambiguous.

2.  $R_n^*(d)$  is a convex function of  $d$ . (See Appendix B of [10] for a proof).

3. Since  $Y^n \rightarrow X^n \rightarrow Z$  form a Markov chain, we have that  $I(X^n; Z) - I(Y^n; Z) = I(X^n; Z|Y^n)$ , (see (2.12) of [10]).

4. It can be shown that the limit of  $R_n^*(d)$  exists if  $R_1(d)$  is finite.

5.  $R^*(d)$  is a monotone decreasing, convex function of  $d$ , over the interval where it is finite.

Using a straightforward extension of the techniques in [10], the ergodic theorem, and the results of [4] and [1], we can prove the following theorem. In fact, this result can be shown to hold for all processes satisfying condition G, if the distortion function satisfies conditions (2.10) and (2.11) of [10].

**Theorem 4** For  $d \geq 0$ , if  $R_1^*(d) < \infty$ , then  $R(d) = R^*(d)$ .

From now on we take  $R(d)$  and  $R^*(d)$  to be synonymous. As in [8] we have not been able to find closed form expressions for  $R(d)$ . But we do have some bounds on this function. Based on the result of the previous section, we can show the following upper bound on  $R(d)$ .

**Theorem 5**

$$R(d) \leq \text{conv} \left( \sum_{i \in S} \pi_i \log \left( 1 + \frac{1}{\lambda_i} \left( \frac{1}{\theta} - \lambda_{\min} \right) \right) \right),$$

where  $\theta$  is the water-level as defined in Theorem 3,  $\lambda_{\min}$  is the minimum of the  $\lambda_i$ 's, and  $\text{conv}(f)$  denotes the convex hull of the function  $f$  (the greatest convex function upper bounded by  $f$ ).

By noting that we can do no better than the case when side information is made available at the encoder, we can show the following lower bound on the rate-distortion function  $R(d)$ .

**Theorem 6**

$$R(d) \geq \left[ \sum_{i \in S} \pi_i \log \left( \frac{1}{\lambda_i d_i} \right) - \log 2 \right]^+,$$

where  $d_i = \begin{cases} \frac{1}{\lambda_i}, & \text{if } \theta > \frac{1}{\lambda_i}, \\ \theta, & \text{otherwise,} \end{cases}$

and  $\theta$  is the water-level given by the solution of

$$\sum_{i \in S} \pi_i d_i = d.$$

For small distortion  $d$  the difference between the upper and the lower bounds is very close to 1 bit or less.

## References

- [1] A.R.Barron, "The Strong Ergodic Theorem for Densities: Generalized Shannon-McMillan-Breiman Theorem", *The Annals of Probability*, pp. 1292-1303, 1985.
- [2] T.Berger, *Rate distortion Theory: A Mathematical Basis for Data Compression*, Englewood-Cliffs, NJ, Prentice-Hall, 1971.
- [3] P.Billingsley, *Ergodic theory and Information*. Huntington, NY, Robert.E.Krieger, 1978.
- [4] T.M.Cover, "A proof of the Data Compression Theorem of Slepian and Wolf for Ergodic Sources", *IEEE-IT Trans.*, pp. 226-228, Mar. 1975.
- [5] T.M.Cover and J.A.Thomas, *Elements of Information Theory*, New York : Wiley, 1991.
- [6] G.R.Grimmett and D.R.Stirzaker, *Probability and Random Processes*. New York : Oxford University Press, 1992.
- [7] F.P.Kelly, *Reversibility and Stochastic Networks*. New York : Wiley, 1979.
- [8] I. Rubin, "Information rates of Poisson Processes", *IEEE-IT Trans.*, vol. IT-20, pp. 200-210, Mar. 1974.
- [9] S. Verdú, "The Exponential Distribution in Information Theory," Invited Paper, *Problemy Peredachi Informatsii*, vol. 32, no.1, pp.100-111, Jan.-Mar. 1996 (In Russian). English version in *Problems of Information Transmission*, vol. 32, no. 1, pp. 86-95, Jan.-Mar. 1996.
- [10] A.D.Wyner, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder-II: General Sources", *Information and Control*, pp-60-80, 1978.
- [11] A.D.Wyner and J.Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", *IEEE-IT Trans.*, vol. IT-22, pp. 1-10, Jan. 1976.