[19] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single node case," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, pp. 344–357, 1993.
[20] ——, "A generalized processor sharing approach to flow control: The multiple node case," *IEEE/ACM Trans. Networking*, July 1994.
[21] K. Sriram, "Methodologies for bandwidth allocation, transmission scheduling, and congestion avoidance in broadband ATM networks," *Comput. Net. and ISDN Syst.*, vol. 26, pp. 43–59, 1993.
[22] M. Wernik, O. Aboul-Magd, and H. Gilbert, "Traffic management for B-ISDN services," *IEEE Network*, Sept. 1992.
[23] W. Whitt, "Tail probabilities with statistical multiplexing and effective bandwidths in multi-class queues," *Telecomm. Syst.*, vol. 2, pp. 71–107, 1993.
[24] O. Yaron and M. Sidi, "Generalized processor sharing networks with exponentially bounded burstiness arrivals," in *IEEE INFOCOM Proc.*, 1994, vol. 2, pp. 628–635.
[25] Z. L. Zhang, D. Towsley, and J. Kurose, "Statistical analyisis of generalized processor sharing scheduling discipline," preprint, 1994.

# Lower Bounds on Expected Redundancy for Nonparametric Classes

Bin Yu, *Member, IEEE*

***Abstract*— This correspondence focuses on lower bound results on expected redundancy for universal coding of independent and identically distributed data on [0, 1] from parametric and nonparametric families. After reviewing existing lower bounds, we provide a new proof for minimax lower bounds on expected redundancy over nonparametric density classes. This new proof is based on the calculation of a mutual information quantity, or it utilizes the relationship between redundancy and Shannon capacity. It therefore unifies the minimax redundancy lower bound proofs in the parametric and nonparametric cases.**

## I. Introduction

One important ingredient of Rissanen's stochastic complexity theory is his (almost) pointwise lower bound on expected redundancy for regular parametric models, and a minimax counterpart follows from Clarke and Barron [1] (cf. [8]). A similar lower bound was proved by Rissanen *et al.* [11] and Yu and Speed [13] on expected redundancy for the Lipschitz nonparametric class of densities. This lower bound was shown in two different senses: one extending the parametric pointwise bound to an artificial parameter space with a dimension depending on the sample size [11], and the other in the minimax sense [13].

On the other hand, Rissanen's pointwise lower bound can be viewed in the broader picture of the relationship between *redundancy and Shannon capacity*. The study of this useful relationship can be traced back to Gallager [5], who showed that the Shannon capacity serves as a lower bound on the minimax expected redundancy over a parametric source class. Haussler [6] extended the result to general classes of sources. Merhav and Feder [9] showed that the same Shannon capacity is a lower bound on the expected redundancy also in the pointwise or "almost sure" sense. Thus the Shannon capacity serves as a lower bound on the expected redundancy both in minimax and pointwise senses. It follows that in the parametric case the mutual information corresponding to any prior on the parameter space is a lower bound on redundancy in both senses. Using the expansion of the mutual information from a smooth prior in [1], Rissanen's pointwise lower bound can be rederived through this redundancy–capacity paradigm. In general, however, calculating or lower bounding the capacity or mutual information can be difficult.

The focus of this correspondence is on minimax redundancy lower bounds for nonparametric source classes of independent and identically distributed (i.i.d.) data strings. Our contribution is the calculation of the mutual information corresponding to a uniform prior on a specially selected finite source subclass, therefore providing a minimax lower bound on redundancy. Since the old approach for nonparametric minimax lower bounds in [13] is based on accumulated prediction error, not on capacity or mutual information, our current approach unifies the parametric and nonparametric cases.

## II. A Review

In this section we review the existing lower bounds on redundancy in the i.i.d. case. For a given i.i.d. data string $x_1, x_2, \cdots, x_n$ and without knowing the distribution $f$ which generated the data, we would like to compress the data in an efficient way. When $f(x) = f_\theta(x)$ belongs to a smooth $k$-dimensional parametric model class such that the parameter $\theta$ can be estimated at the $n^{-1/2}$ rate, Rissanen [10] showed that we need at least $H(f) + \frac{k}{2}\log n$ bits for the string, asymptotically. That is, for any joint density $q_n$ on $n$-tuples, if we view $-\log q_n(x^n)$ as the code length of an idealized prefix code, then its expected redundancy is

$$E_{f_\theta^n} \log (f_\theta^n / q_n).$$

Rissanen ([10]) showed that

$$\liminf E_{f_\theta^n} \log (f_\theta^n / q_n)/(k \log n/2) \geq 1$$

for all $\theta$ values except for a set which depends on $q$ and has Lebesgue measure zero. With a prefix code achieving this lower bound, Rissanen justified that $\frac{k}{2}\log n$ can be viewed as the coding complexity measure of the model class. For more general classes, Merhav and Feder [9] showed that the Shannon capacity replaces $\frac{k}{2}\frac{\log n}{n}$ as the pointwise or almost sure lower bound on redundancy. As we can derive from [1], $\frac{k}{2}\frac{\log n}{n}$ is naturally the leading term in the capacity in the regular parametric case.

When $f$ is known to be in the smooth nonparametric density class of bounded derivatives (or Lipschitz class) on [0, 1], a complexity rate measure of $n^{1/3}$ was established by Rissanen *et al.* in [11] by embedding the nonparametric class in a parametric class of dimension of order $n^{1/3}/\log n$. This embedding reflects the fact that a smooth nonparametric class is in essence a parametric class whose dimension increases with the sample size.

The other approach to obtain lower bounds on expected redundancy is minimax (cf. [2], [3]). Let $w(\theta)$ be a prior on the parameter space and $q_n$ a joint density on $n$-tuples; then Gallager [5] has shown that

Manuscript received December 4, 1994; revised July 27, 1995. This work was supported by the Army Research Office under Grant DAAH04-94-G-0232 and by the National Science Foundation under Grant DMS-9322817. The material in this correspondence was presented at the IEEE IT/IMS Workshop, Virginia, Oct. 1994.
The author is with the Department of Statistics, University of California, Berkeley, CA 94720-3860 USA.
Publisher Item Identifier S 0018-9448(96)00009-0.

the minimax expected redundancy is bounded from below by the maximum of the mutual information over all priors, i.e., the Shannon channel capacity. That is

$$\begin{aligned}&\min_{q_n}\max_{\theta} E_{f_\theta^n}\log(f_\theta^n/q_n)\\ &=\max_{\theta}\min_{q_n} E_{f_\theta^n}\log(f_\theta^n/q_n)\\ &\geq \max_{w(\theta)}\min_{q_n} E_w E_{f_\theta^n}\log(f_\theta^n/q_n)\\ &=\max_{w} E_w E_{f_\theta^n}\log(f_\theta^n/q_n^w)\\ &=\max_{w} I(w)\end{aligned}$$

where

$$q_n^w(x^n) = \int f_\theta^n(x^n)w(\theta)\,d\theta$$

$$I(w) := I(\theta, X^n), \quad \text{with } \theta \sim w.$$

Conditionally on $\theta$, $X_1,\cdots,X_n \sim_{\text{i.i.d.}} f_\theta$, and $\max_w I(w)$ is the Shannon channel capacity.

When the parameter space is one-dimensional, asymptotic expansions of $I(w)$ are given by Ibragimov and Has'minsky [8] under regularity conditions on $f_\theta$ and $w(\theta)$. For general cases, see Clarke and Barron [1], where they showed that the first term in the expansion of the Bayes redundancy or mutual information $I(w)$ is the Rissanen coding complexity of $\frac{k}{2}\log n$. Hence this complexity measure also serves as the minimax lower bound on expected redundancy.

For the Lipschitz class mentioned above, any Bayes redundancy or mutual information still provides a lower bound. However, no prior seems to exist on the whole density class for which the mutual information can be approximated analytically. On the other hand, the expected redundancy is simply the accumulated prediction or estimation error in terms of Kullback–Leibler divergence:

$$E_{f_\theta^n}\log(f_\theta^n/q_n) = \sum_t E_{f_\theta^{t-1}} D(f_\theta(\cdot)\,\|\,q(\cdot\mid X^{t-1})).$$

Techniques such as Assouad's (cf. [4]) have long been developed to obtain minimax lower bounds on density estimation errors in the nonparametric case [12]. By lower-bounding the divergence $D$ by the Hellinger distance and mimicing Assouad's technique (cf. [12]) from the density estimation literature, a minimax rate lower bound of $n^{1/3}$ was established and shown to be the optimal rate by Yu and Speed in [13].

Note that in applying Assouad's technique, one does not calculate the Bayes estimation error over the whole class, but only over a conveniently chosen subclass, and the Bayes estimation error over this subclass provides a lower bound on the minimax estimation error. It turns out that the detour to accumulated prediction or estimation error is not necessary since we can use the subclass directly with the redundancy. Recently, Haussler [7] gave useful general bounds on mutual information. Using one of his lower bounds to the subclass and with a uniform prior, we obtain in the next section minimax lower bounds for general smooth density classes; thus we reconcile the proofs of minimax redundancy lower bounds in parametric and nonparametric cases, because our new proof is based on mutual information rather than the accumulated prediction error. As a corollary to our theorem, we rederive the minimax lower bound rate $n^{1/3}$ given in [13] for the Lipschitz class.

## III. Minimax Lower Bounds for Smooth Density Classes: An Information-Theoretic Proof

In nonparametric density estimation, it is well known that constraints must be imposed on the density class over which a minimax result is sought (cf. Devroye [4]). Smoothness constraints are commonly used since the minimax rates obtained reflect the difficulty of estimation as a function of the smoothness of the class and the sample size—the smoother the class and the larger the sample size, the easier the estimation. The density classes $LIP(s, C)$ defined below are standard smooth classes in nonparametric density estimation. (See [12] and the references cited therein.) Because of the close relationship between estimation and compression or redundancy, these classes are appropriate in our setting as well. In this section, a minimax lower bound on expected redundancy over a class is derived by lower-bounding the mutual information corresponding to a uniform prior over a chosen subclass and using a result in Haussler [7]. Moreover, the minimax rates in these lower bounds are believed to be optimal.

Let the smooth density class $LIP(s, C)$ on [0, 1] contain those $f$'s such that

$$\sup_{0\leq x\leq 1}|f^{(i)}(x)| \leq C, \quad \text{for } i=1,\cdots,k$$

and

$$|f^{(k)}(x) - f^{(k)}(y)| \leq C|x-y|^\nu, \quad \text{for } 0\leq x,\, y\leq 1$$

where $C$ is a fixed constant, $k$ is a positive integer, and $\nu$ is such that $0 < \nu \leq 1$ and $s = k + \nu \geq 1$. Note that taking $k = 0$ and $\nu = 1$ gives the Lipschitz class studied in [13] and [11].

For any finite subclass $\mathcal{F}_n$ of

$$LIP(s, C)^n := \{f^n : f \in LIP(s, C)\}$$

and a prior $\mu$ on $\mathcal{F}_n$ with the corresponding mixture

$$h := \sum_{f^n\in\mathcal{F}_n} f^n \mu(f^n)$$

we have

$$\begin{aligned}&\min_{q_n}\max_{f\in LIP(s,C)} E_{f^n}\log(f^n/q_n)\\ &\geq \min_{q_n}\max_{f\in\mathcal{F}_n} E_{f^n}\log(f^n/q_n)\\ &\geq E_\mu E_{f^n}\log(f^n/h)\\ &= I(\mu).\end{aligned}$$

Thus a minimax lower bound on redundancy can be obtained if we appropriately choose a subclass $\mathcal{F}_n$ and a prior $\mu$ on it, and bound $I(\mu)$ from below.

We can expect $I(\mu)$ to give a good lower bound on the minimax redundancy only when the subclass $\mathcal{F}_n$ accurately reflects the complexity of the underlying class, relative to the sample size. Because redundancy and estimation error are intimately related, it is not surprising that the hypercube subclass from density estimation serves us well here (cf. [4] and [12]).

The hypercube subclass is constructed by perturbing the uniform density on [0, 1] over $m$ equally sized subintervals. The perturbation is done on each subinterval by adding a positive or a negative "signal" or properly shifted and scaled perturbation function $g$ (see Fig. 1). Since there are $m$ subintervals and two choices for each interval, the total number of such perturbed functions is $2^m$, which may be regarded as corresponding to an $m$-dimensional hypercube, hence the name. The number of subintervals $m$ will be chosen later according to the sample size $n$ so that the complexity of the hypercube class increases with $n$ and the redundancy rate on this hypercube class approximates that of the full class.

To be more precise, let the perturbation function $g$ be a fixed $(k+1)$-times differentiable function on [0, 1], which is bounded by $c_g$ and for which

$$\int_0^1 g(x)\,dx = 0 \qquad \text{and} \qquad \int_0^1 g^2(x)\,dx = a_g > 0.$$

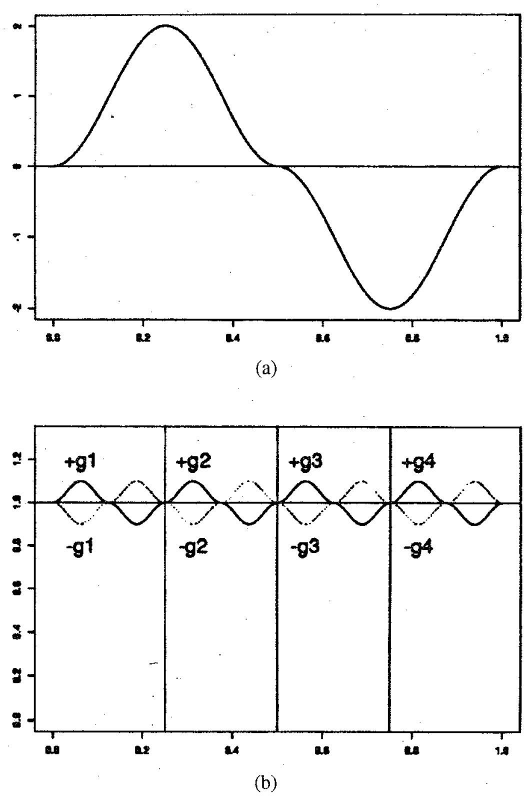


Fig. 1. (a) An example of perturbation function: $g(x) = -\cos(4\pi x) + 1$ on [0, 0.5] and $g(x) = \cos(4\pi x) - 1$ on (0.5, 1]. (b) Shifted and scaled perturbation functions are added to $u(x) \equiv 1$ for $m = 4$ and $c = 0.1$.

Moreover, we require $g(0) = g(1) = 0$, and the right derivatives of $g$ at $x = 0$ and the left derivatives of $g$ at $x = 1$ are zero up to the $k$th order. These requirements ensure that when we piece together shifted and rescaled $g$'s as below, the resulting function is in $LIP(s, C)$.

Divide [0, 1] into $m$ disjoint intervals $I_j$ of size $1/m$. For $j = 1, 2, \cdots, m$, let the shifted and scaled $g$ be the signal added to the $j$th subinterval

$$g_j(x) = cm^{-s} g[m(x - (j-1)/m)]$$

with $c$ small enough so that $|g_j| < 1$. Let the collection of such perturbed uniform densities be

$$\mathcal{M}_m = \left\{ f_\tau = 1 + \sum_{j=1}^{m} (2\tau_j - 1) \times g_j(x) I_{I_j} : \tau = (\tau_1, \cdots, \tau_m) \in \{0, 1\}^m \right\}$$

and define the *hypercube class*

$$\mathcal{F}_n = \mathcal{M}_m^n = \{f_\tau^n : f_\tau \in \mathcal{M}_m\}.$$

For any two density functions $u$ and $v$ on [0, 1], define their Hellinger distance $H$ through

$$H^2(u, v) = \int (\sqrt{v} - \sqrt{w})^2.$$

Let us take the uniform prior $\mu_m$ on $\mathcal{F}_n$, or equivalently, the product distribution of i.i.d. symmetric Bernoulli trials for $\tau = (\tau_1, \cdots, \tau_m)$. The first inequality of Haussler [7, Theorem 1] can now be stated in our notation as

$$I(\mu_m) \geq -E_{\tau^*} \log E_\tau e^{-n/4 \cdot H^2(f_\tau, f_{\tau^*})}$$

where $\tau$ and $\tau^*$ are independent product distributions of i.i.d. symmetric Bernoulli trials, that is, $\tau_j$ (and $\tau_j^*$) are i.i.d. with $P(\tau_j = 0) = P(\tau_j = 1) = 1/2$.

Let

$$C_m = \int_{I_1} (\sqrt{1 + g_1} - \sqrt{1 - g_1})^2 \equiv \int_{I_j} (\sqrt{1 + g_j} - \sqrt{1 - g_j})^2, \quad \text{for } 1 \leq j \leq m.$$

It is easy to see that

$$H^2(f_\tau, f_{\tau^*}) = \sum_{j=1}^{m} \int_{I_j} (\sqrt{1 + (2\tau_j - 1) g_j} - \sqrt{1 + (2\tau_j^* - 1) g_j})^2 = \sum_{j=1}^{m} |\tau_j - \tau_j^*| C_m.$$

*Lemma:* For positive constant $B_g = a_g c^2$

$$C_m \geq B_g m^{-2s-1}.$$

*Proof:* It is easy to see that

$$(\sqrt{1 + g_1} + \sqrt{1 - g_1})^2 \leq 4$$

and

$$\int_0^{1/m} g_1^2 \, dx = \int_0^{1/m} c^2 m^{-2s} g^2(mx) \, dx = c^2 m^{-2s-1} \int_0^1 g^2(x) \, dx = a_g c^2 m^{-2s-1}.$$

$$\begin{aligned} C_m &= \int_0^{1/m} (\sqrt{1 + g_1} - \sqrt{1 - g_1})^2 \, dx \\ &= \int_0^{1/m} (1 + g_1 - (1 - g_1))^2 / (\sqrt{1 + g_1} + \sqrt{1 - g_1})^2 \, dx \\ &= 4 \int_0^{1/m} g_1^2 / (\sqrt{1 + g_1} + \sqrt{1 - g_1})^2 \, dx \\ &\geq 4 \int_0^{1/m} g_1^2 \, dx / 4 \\ &= a_g c^2 m^{-2s-1}. \end{aligned}$$

□

Now fix $\tau^*$ and let

$$k = \sum_j^m \tau_j^*.$$

Without loss of generality, assume $\tau_j^* = 1$ for $1 \leq j \leq k$ and $\tau_j^* = 0$ for $k + 1 \leq j \leq m$. Then

$$W := \sum_{j=1}^{k} (1 - \tau_j) + \sum_{j=k+1}^{m} \tau_j$$

is binomial $(m, 1/2)$.

$$E_\tau e^{-n/4 \cdot H^2(f_\tau, f_\tau^*)} = E_\tau e^{-n/4 \cdot C_m \left( \sum_{j=1}^{k} (1 - \tau_j) + \sum_{j=k+1}^{m} \tau_j \right)} = E e^{-n/4 \cdot C_m W}.$$

Recall that the moment generating function of a Binomial$(N, p)$ variable $W$ is

$$E e^{-tW} = (pe^{-t} + 1 - p)^N.$$

For our $W$, $N$ is $m$ and $p$ is 1/2. Therefore, for $t_m = n/4 \cdot C_m$

$$E e^{-n/4 \cdot C_m W} = \left( \frac{1}{2} e^{-t_m} + 1 - \frac{1}{2} \right)^m = 2^{-m} (1 + e^{-t_m})^m.$$

Because the last expression is independent of $k$ or $\tau^*$, its negative logarithm is a lower bound on $I(\mu_m)$, that is

$$\begin{aligned} I(\mu_m) &\geq -E_{\tau^*} \log E_\tau e^{-n/4 \cdot H^2(f_\tau, f_{\tau^*})} \\ &= -E_{\tau^*} \log E e^{-n/4 \cdot C_m W} \\ &= -\log (2^{-m}(1+e^{-t_m})^m) \\ &= m \log 2 - m \log (1+e^{-t_m}) \\ &= m \log 2 - m \log \left(1+e^{-\frac{n}{4}C_m}\right) \\ &\geq m \log 2 - m \log \left(1+e^{-\frac{n}{4}B_g m^{-2s-1}}\right). \end{aligned}$$

Choosing $m = An^{1/(2s+1)}$ $(A > 0)$ to maximize the rate in the above lower bound, we get the following theorem.

*Theorem:*

$$\min_{q_n} \max_{f \in LIP(s,C)} E_{f^n} \log (f^n/q_n) \geq I(\mu_m) \geq A_{g,s} n^{1/(2s+1)}$$

where

$$\begin{aligned} A_{g,s} &= A(\log 2 - \log (1+e^{-B_g/(4A^{2s+1})})) \\ &= A(\log 2 - \log (1+e^{-a_g c^2/(4A^{2s+1})})) > 0. \end{aligned}$$

Taking $k = 0$, $\nu = 1$ therefore $s = 1$ in the theorem, we obtain the optimal rate lower bound in [13], as shown in the corollary below.

*Corollary:*

$$\min_{q_n} \max_{f \in LIP(1,C)} E_{f^n} \log (f^n/q_n) \geq O(n^{1/3}).$$

*Remarks:*

1) In general, we can consider the $LIP(s, C)$ classes on $[0, 1]^d$ $(d \geq 1)$. Minimax lower bounds on redundancy of rates $O(n^{d/(2s+d)})$ can be obtained. These rates are believed to be optimal in the sense that universal codes can be constructed to achieve these rates. In the case of $LIP(1, C)$ the rate $n^{1/3}$ has been shown to be optimal in [13].
2) The proof for the minimax lower bound $\frac{k}{2} \log n$ in the parametric case follows from the asymptotic expansion of $I(\mu)$ in [1] or [8] for smooth priors. Superficially, this approach has a continuous flavor since $\mu$ needs to have nice smoothness properties on the whole parameter space, whereas the proof in the nonparametric case as we just saw has a discrete flavor because of the hypercube subclass over which $I(\mu_m)$ is estimated. Heuristically, however, the continuous prior can be made discrete. Under regularity conditions, we believe that $I(\mu)$ should give the same lower bound $\frac{k}{2} \log n$ even for a discrete uniform prior $\mu$ on a grid subset of the parametric space, as long as the grid size is of the order or smaller than $n^{-1/2}$. Note that the nearest neighbors on the hypercube for the optimal choice $m = n^{1/(2s+1)}$ also have Hellinger distances of order $n^{-1/2}$, the rate at which $n$ i.i.d. data points can possibly distinguish two probability densities. In other words, what seems essential to both the parametric and the nonparametric case is to find a subclass of densities whose closest elements are $n^{-1/2}$ apart from each other in terms of Hellinger distance.


## Acknowledgment

The author wishes to thank Dr. J. Rissanen, an anonymous referee, the Associate Editor, and J. Huang for their very helpful comments. Special thanks are due to Dr. N. Hengartner for spotting an inexcusable error in an earlier version of the correspondence and for pointing to the Haussler paper.


## References


[1] B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of Bayes methods," *IEEE Trans. Inform. Theory,* vol. 36, pp. 453–471, May 1990.
[2] I. Csiszár, "Information theoretical methods in statistics," class notes, University of Maryland, College Park, Spring 1990.
[3] L. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory,* vol. IT-19, pp. 783–795, Nov. 1973.
[4] L. Devroye, *A Course in Density Estimation.* Basel, Switzerland: Birkhäuser-Verlag, 1987.
[5] R. G. Gallager, "Source coding with side information and universal coding," unpublished manuscript, Sept. 1979.
[6] D. Haussler, "A general minimax result for relative entropy," preprint.
[7] ——, "General bounds on the mutual information between a parameter and n conditionally independent observations," preprint.
[8] I. A. Ibragimov and R. Z. Has'minsky, "On the information in a sample about a parameter," in *Proc. 2nd Int. Symp. on Information Theory,* B. N. Petrov and F. Csáki, Eds. Budapest, Hungary: Akademiai Kiado, 1973.
[9] N. Merhav and M. Feder, "A strong version of the redundancy-capacity theorem of universal coding," *IEEE Trans. Inform. Theory,* vol. 41, pp. 714–722, May 1995.
[10] J. Rissanen, "Stochastic complexity and modeling," *Annals Statist.,* vol. 14, pp. 1080–1100, 1986.
[11] J. Rissanen, T. Speed, and B. Yu, "Density estimation by stochastic complexity," *IEEE Trans. Inform. Theory,* vol. 38, pp. 315–323, May 1993.
[12] B. Yu, "Assouad, Fano, and Le Cam," to appear in *Festschrift in Honor of L. Le Cam on His 70th Birthday,* 1995.
[13] B. Yu and T. Speed, "Data compression and histograms," *Probab. Theory Related Fields,* vol. 92, pp. 195–229, 1992.


# Error Exponents for Successive Refinement by Partitioning


Angelos Kanlis and Prakash Narayan



***Abstract*—Given a discrete memoryless source (DMS) with probability mass function $P$, we seek first an asymptotically optimal description of the source with distortion not exceeding $\Delta_1$, followed by an asymptotically optimal refined description with distortion not exceeding $\Delta_2 < \Delta_1$. The rate-distortion function for successive refinement by partitioning, denoted $R(P, \Delta_1, \Delta_2)$, is the overall optimal rate of these descriptions obtained via a two-step coding process. We determine the error exponents for this two-step coding process, namely, the negative normalized asymptotic log likelihoods of the event that the distortion in either step exceeds its prespecified acceptable value, and of the conditional event that the distortion in the second step exceeds the prespecified value given the rate and distortion of the code for the first step. We show that even when the rate-distortion functions for one- and two-step coding coincide, the error exponent in the former case may exceed those in the latter.**



***Index Terms*—Covering Lemma, error exponent, rate-distortion function, successive refinement by partitioning.**

Manuscript received May 17, 1994; revised August 16, 1995. This work of A. Kanlis was supported by the Bodossaki Foundation, Greece, and the Institute for Systems Research at the University of Maryland, College Park, MD, under NSF Grant OIR-85-00108. The work of P. Narayan was supported by the Institute for Systems Research at the University of Maryland, College Park, MD, under NSF Grant OIR-85-00108. The material in this paper was proesented at the 1995 IEEE International Symposium on Information Theory, Whistler, BC, Canada, September 17–22, 1995.



The authors are with the Electrical Engineering Department, Institute for Systems Research, University of Maryland, College Park, MD 20742 USA.



Publisher Item Identifier S 0018-9448(96)00034-X.