Asymptotically Optimal Quantizers for Detection of I.I.D. Data

GERALD R. BENITZ AND JAMES A. BUCKLEW, MEMBER, IEEE

Abstract —The asymptotic probability of error for quantization in maximum likelihood tests is analyzed. We assume quantizers with large numbers of levels generated from a companding function. A theorem which relates the companding function to the asymptotic probability of error is proven. The companding function is then optimized.

I. INTRODUCTION

A FUNDAMENTAL problem in digital communication systems is the question of how much degradation in performance a receiver suffers given that the incoming data are quantized. A related problem is how to design a quantizer so that this degradation is minimized.

Much research dealing with these questions has been done. Kassam [1] considered the problem of how to design a quantizer to minimize a probability of error criterion. Poor and Thomas [2], [3] consider the problem of quantizer design from an asymptotic relative efficiency type of distortion criterion and then from an Ali–Silvey distance measure. Poor [4], [5] subsequently considers the problem relative to various statistical divergences and also considers some novel estimation theory applications. Bucklew [6] considers the problem of design from a mean-squared error in the likelihood ratio point of view.

None of these papers consider the probability of error as the design distortion measure. The error probability is notoriously intractable, and this fact in our opinion motivated all these other approaches. We take some first steps in the direction of using error probability as the distortion measure. For a large number of samples, the probability of error (in many cases) converges to zero exponentially fast. We propose to take the exponential rate constant as our distortion measure. This is tantamount to using decibels to measure the probability of error for large sample sizes. We develop a theory asymptotic in the number of quantizer output levels for this distortion measure based upon a companding model for the quantizer, and then optimize the asymptotic constant to obtain the optimal companding characteristic.

II. PRELIMINARIES AND OVERVIEW

The primary theoretical tool invoked in this paper is the well-known theorem of Chernoff [7]. Chernoff explored the probability of error in a maximum likelihood binary decision problem and derived an expression which gives the asymptotic efficiency of the test, i.e., the exponential rate at which the probability of error vanishes as the sample size increases.

Suppose that $\{X_i\}_{i=1}^{\infty}$ are independent identically distributed (i.i.d.) random variables. Consider the following detection problem:

$$H_0: X_1 \text{ has distribution } P$$

$$H_1: X_1 \text{ has distribution } Q.$$

Assuming equal prior probabilities, the maximum likelihood detector minimizes the probability of error. Assume that W is a measure on $\mathbf{R}$ which dominates both P and Q, and let dP/dW and dQ/dW denote the Radon–Nikodym derivatives. Define the log likelihood function

$$L(x) \triangleq \log \frac{dP}{dW}(x) - \log \frac{dQ}{dW}(x) \qquad (2.1)$$

and create the test statistic

$$\sum_{j=1}^{n} L(X_j).$$

Define e_n as the total probability of error after observing n samples,

$$e_n \triangleq \frac{1}{2} P\left(\left\{\sum_{j=1}^{n} L(X_j) < 0\right\}\right) + \frac{1}{2} Q\left(\left\{\sum_{j=1}^{n} L(X_j) > 0\right\}\right).$$

Define the alpha entropy H_α,

$$H_\alpha \triangleq \int_R \left[\frac{dP}{dW}(x)\right]^{\alpha} \left[\frac{dQ}{dW}(x)\right]^{1-\alpha} dW(x). \qquad (2.2)$$

Chernoff's theorem says that if $L(X_1)$ is finite W almost-everywhere, then

$$\lim_{n \to \infty} \frac{1}{n} \log e_n = \log H_{\alpha_0}$$

where

$$H_{\alpha_0} = \min_{\alpha \in [0,1]} H_\alpha. \qquad (2.3)$$

Note that $H_\alpha \leq 1$, which can be seen from an application of Hölder's inequality. Also, in Lemma 2, we show that H_α is continuous and differentiable in α.

Chernoff's theorem differs from the traditional Chernoff bound [13] in that the theorem clearly exhibits the exponential tightness of the bound. Exponential tightness means that

$$\exp[n(I - \epsilon)] < e_n < \exp[n(I + \epsilon)]$$

for large n, where $I = \log H_{\alpha_0}$ and $\epsilon > 0$. Thus any design which reduces I must (eventually for large n) reduce the probability of error. Designs which reduce only an upper bound criterion may not always reduce the probability of error.

We choose as our criterion of quantizer performance the value of H_{α_0} in (2.3). We will actually compute the change in H_{α_0} caused by quantization. It is assumed that the source distributions P and Q have Lebesgue densities so that (2.2) exists with $dW = dx$. Quantization of the source can be viewed as a mapping of P and Q into $\hat{P}$ and $\hat{Q}$, respectively, each having finite support. If we choose W to be the counting measure on this support, it is clear that (2.2) exists. It is in this way that we will apply Chernoff's theorem.

Since we can compare the asymptotic performance of quantizers, it is desirable to find the optimal quantizer. We will do so asymptotically in N, the number of quantization levels. In Section V we define a quantization rule which generates a sequence of increasingly fine quantizers indexed by N. As N becomes large the degradation (the change in H_{α_0}) will vanish. The figure of merit of the quantization rule will be the rate at which the degradation vanishes.

Our assumptions and notation are given in Section III. Section IV gives a Taylor series analysis which discerns the change in H_α due to quantization. In Section V we define a quantization rule and relate it to the change in H_α. We satisfy the conditions of Chernoff's theorem in Section VI and then optimize in Section VII. Extensions are given in Sections VIII and IX followed by examples in Section X.

III. Notation

Let $\{X_j\}_{j=1}^\infty$ be an i.i.d. sequence of random variables. The two possible Lebesgue densities for the distribution of X_1 are

$$H_0: \text{density} = p(x)$$
$$H_1: \text{density} = q(x).$$

A large class of densities will be considered in Section VIII, but the core of the analysis concerns compact sets. Consequently, we begin with the following assumptions.

1) A closed interval $A = [a, b]$ ($-\infty < a < b < \infty$) exists such that $\{p > 0\} = \{q > 0\} = A$.
2) Both p and q are three times continuously differentiable (C^3) on (a, b) with the third derivatives bounded.
3) $\text{Min}_A p > 0$, $\text{min}_A q > 0$. (3.1)

The source is mapped into discrete values through a quantizer. For our purposes, a quantizer will be viewed as a partition of the set A. Thus the collection of points

$$a_0, a_1, a_2, \cdots, a_N, \qquad a_j \in R$$

$$(a_i \neq a_j \quad \text{for } i \neq j) \quad (3.2)$$

defines a quantizer; i.e., to quantize X is to indicate which points of (3.2) are greater than X. The numerical value assigned by the quantizer is not important.

Given any $M > 0$, we say that a partition (3.2) is M-allowable for the set A if the following conditions hold:

1) $a = a_0 < a_1 < a_2 < \cdots < a_N = b$;
2) $a_j - a_{j-1} \leq M/N$, $j = 1, 2, \cdots, N$.

Given an M-allowable partition, we define the following (for each $j = 1, 2, \cdots, N$):

$$B_j = (a_{j-1}, a_j)$$
$$\Delta_j = a_j - a_{j-1}$$
$$x_j = (a_j + a_{j-1})/2. \quad (3.3)$$

For notational simplicity we use the following conventions to indicate relative rates of growth. Given sequences $\langle x_n \rangle_{n=1}^\infty$ and $\langle y_n \rangle_{n=1}^\infty$, we write

$$y_n = O(x_n) \text{ if and only if } \limsup_{n \to \infty} \left| \frac{y_n}{x_n} \right| < \infty,$$

$$y_n = o(x_n) \text{ if and only if } \lim_{n \to \infty} \frac{y_n}{x_n} = 0.$$

IV. Expansion of Alpha Entropy

Our goal in this section is to estimate the change in alpha entropy due to quantization. We first analyze the alpha entropy of the source. Pick any $\alpha \in (0,1)$, $M > 0$, and any M-allowable partition. Define

$$f(x) \triangleq p^\alpha(x) q^{1-\alpha}(x).$$

The alpha entropy (2.2) is

$$H_\alpha = \sum_{j=1}^N \int_{x_j - (\Delta_j/2)}^{x_j + (\Delta_j/2)} f(x) \, dx.$$

Note that on any bin B_j, both p and q can be approximated by a second-order Taylor polynomial. The function f also has a Taylor expansion. Consider the derivatives of f on A:

$$f' \triangleq \frac{df}{dx} = \alpha p' p^{\alpha-1} q^{1-\alpha} + (1 - \alpha) q' p^\alpha q^{-\alpha} \quad (4.1)$$

$$f'' \triangleq \frac{df'}{dx} = \alpha p'' p^{\alpha-1} q^{1-\alpha} - \alpha(1 - \alpha)(p')^2 p^{\alpha-2} q^{1-\alpha} \quad (4.2)$$

$$+ 2\alpha(1 - \alpha) p' q' p^{\alpha-1} q^{-\alpha}$$

$$- \alpha(1 - \alpha)(q')^2 p^\alpha q^{-\alpha-1}$$

$$+ (1 - \alpha) q'' p^\alpha q^{-\alpha}.$$

The third derivative of f exists, but all we require is that

f''' be bounded on A. By inspection of (4.2), it is clear that the terms in f''' will involve positive powers of p, p', p'', p''' and negative powers of p, and likewise for q. From (3.1) we have that p, p', p'', p''' must be bounded and likewise for q. Also p^{-1} and q^{-1} are bounded on A. Thus f''' is bounded on A.

We now apply Taylor's theorem [8]. For $t \in B_j$,

$$f(t) = f(x_j) + (t - x_j)f'(x_j) + \frac{1}{2}(t - x_j)^2 f''(x_j)$$
$$+ \frac{1}{6}(t - x_j)^3 f'''(z)$$

for some z between t and x_j. Because f''' is bounded, there is a $K > 0$ such that

$$\left| \int_{B_j} f(t)\, dt - \int_{B_j} f(x_j) + (t - x_j)f'(x_j) \right.$$
$$\left. + \frac{1}{2}(t - x_j)^2 f''(x_j)\, dt \right|$$
$$\leq K \int_{B_j} |t - x_j|^3\, dt = K\Delta_j^4 / 32. \quad (4.3)$$

Consider

$$\sum_{j=1}^N \Delta_j^4 \leq \left(\max_j \Delta_j^3 \right) \sum_{j=1}^N \Delta_j = (a_N - a_0) \max_j \Delta_j^3 = O(N^{-3}),$$

from which it follows that

$$\left| \sum_{j=1}^N \int_{B_j} f(t)\, dt - \int_{B_j} f(x_j) + (t - x_j)f'(x_j) \right.$$
$$\left. + \frac{1}{2}(t - x_j)^2 f''(x_j)\, dt \right|$$
$$\leq \sum_{j=1}^N K\Delta_j^4 / 32 = O(N^{-3}).$$

We complete the integration to conclude that

$$H_\alpha - \sum_{j=1}^N \left[f(x_j)\Delta_j + \frac{1}{24} f''(x_j)\Delta_j^3 \right] = O(N^{-3}). \quad (4.4)$$

The $f'(x_j)$ term vanishes because x_j is the midpoint of B_j. The use of $O(N^{-3})$ will become apparent when we consider sequences of M-allowable partitions.

It is important to observe that the convergence in (4.4) is uniform in some neighborhood of α. To see this, return to the choice of the constant K in (4.3). Because f''' is continuous in α, K may be chosen to bound f''' in some neighborhood of α.

We now analyze the alpha entropy of the quantized source, denoted $\hat{H}_\alpha$. With each quantizer bin associate its probability of occurrence. Define

$$\hat{p}_j \triangleq \int_{B_j} p(x)\, dx \qquad \hat{q}_j \triangleq \int_{B_j} q(x)\, dx.$$

It is then clear that

$$\hat{H}_\alpha = \sum_{j=1}^N \hat{p}_j^\alpha \hat{q}_j^{1-\alpha}.$$

Apply Taylor's theorem and proceed as before, noting that p''' and q''' are bounded,

$$\hat{p}_j = p(x_j)\Delta_j + \frac{1}{24} p''(x_j)\Delta_j^3 + e_j, \qquad e_j = O\left(\Delta_j^4\right)$$

$$\hat{q}_j = q(x_j)\Delta_j + \frac{1}{24} q''(x_j)\Delta_j^3 + d_j, \qquad d_j = O\left(\Delta_j^4\right).$$

These equations define e_j and d_j. Also define

$$\xi_j \triangleq \frac{1}{p(x_j)} \left[\frac{1}{24} p''(x_j) + \frac{e_j}{\Delta_j^3} \right]$$

$$n_j \triangleq \frac{1}{q(x_j)} \left[\frac{1}{24} q''(x_j) + \frac{d_j}{\Delta_j^3} \right]$$

and note that these are bounded. Thus

$$\hat{H}_\alpha = \sum_{j=1}^N p^\alpha(x_j) q^{1-\alpha}(x_j) \left[1 + \xi_j \Delta_j^2\right]^\alpha \left[1 + \eta_j \Delta_j^2\right]^{1-\alpha} \Delta_j. \quad (4.5)$$

To expand (4.5), note that for $\beta \in (0,1)$,

$$(1+t)^\beta = 1 + \beta t + o(t), \qquad |t| < \frac{1}{2}. \quad (4.6)$$

Clearly, $\xi_j \Delta_j^2$ vanishes as $N \to \infty$, and likewise for $\eta_j \Delta_j^2$. Apply (4.6) in (4.5) to obtain

$$\hat{H}_\alpha = \sum_{j=1}^N \left\{ p^\alpha(x_j) q^{1-\alpha}(x_j) \left[1 + \alpha\xi_j\Delta_j^2 + o\left(\Delta_j^2\right)\right] \right.$$
$$\left. \cdot \left[1 + (1-\alpha)\eta_j\Delta_j^2 + o\left(\Delta_j^2\right)\right]\Delta_j \right\}$$

$$\hat{H}_\alpha = \sum_{j=1}^N p^\alpha(x_j) q^{1-\alpha}(x_j)$$
$$\cdot \left[1 + \alpha\xi_j\Delta_j^2 + (1-\alpha)\eta_j\Delta_j^2 + o\left(\Delta_j^2\right)\right]\Delta_j.$$

The latter expression follows from an easily verifiable operation involving o and the fact that all terms are uniformly bounded.

Due to the uniform bound on the e_j, d_j, and all the other terms, it follows that $\max_j o(\Delta_j^2) = o(\max_j \Delta_j^2) = o(N^{-2})$. We emphasize that, without the uniform bounds, we would not be able to claim this. It follows that

$$\hat{H}_\alpha - \sum_{j=1}^N p^\alpha(x_j) q^{1-\alpha}(x_j)$$
$$\cdot \left[1 + \alpha \frac{p''(x_j)}{p(x_j)} \frac{\Delta_j^2}{24} + (1-\alpha) \frac{q''(x_j)}{q(x_j)} \frac{\Delta_j^2}{24}\right]\Delta_j = o(N^{-2}). \quad (4.7)$$

The convergence in (4.7) is also uniform in a neighborhood of α. This is clear because errors attributable to e_j and d_j will be continuous functions of α.

It is now possible to compute the change in alpha entropy due to quantization. Combining (4.2), (4.4), and (4.7) gives

$$\hat{H}_\alpha - H_\alpha = \frac{\alpha(1-\alpha)}{24} \sum_{j=1}^{N} p^\alpha(x_j) q^{1-\alpha}(x_j)$$

$$\cdot \left[\frac{p'(x_j)}{p(x_j)} - \frac{q'(x_j)}{q(x_j)} \right]^2 \Delta_j^3 + o(N^{-2}). \quad (4.8)$$

Note that the summation is $O(N^{-2})$. Also note that

$$\frac{p'}{p}(x) - \frac{q'}{q}(x) = \frac{d}{dx} \log \frac{p}{q}(x) \quad (4.9)$$

is the derivative of the log likelihood ratio.

V. THE QUANTIZATION RULE

In this section we construct sequences of M-allowable partitions according to a "companding" rule (see Bennett [9]). Let g be a mapping of the set $A = [a, b]$ into the unit interval. We say that g defines a quantization rule if it satisfies the following conditions:

1) g is a one-one mapping (bijection) between $[a, b]$ and $[0, 1]$;
2) g is differentiable on (a, b) with g' Lipschitz continuous ($|g'(x) - g'(y)| \le M|x - y|$ for some M);
3) For some $\delta > 0$, $g' \ge \delta$ on (a, b). $\quad (5.1)$

The function g is used to create M-allowable partitions as follows. For any integer $N > 1$, define

$$a_j = g^{-1}\left(\frac{j}{N}\right), \qquad j = 0, 1, 2, \cdots, N. \quad (5.2)$$

Note that we are using g to reflect a uniform partition of $[0, 1]$ onto A. The partition of (5.2) is M-allowable; choose $M = 1/\delta$ for δ in 3) of (5.1).

We now estimate the bin length Δ_j as a function of g'. We have that

$$g(a_j) - g(a_{j-1}) = \frac{1}{N}, \qquad j = 1, 2, \cdots, N.$$

Apply the mean-value theorem [8] to g on each bin. For each j there is a $\xi_j \in B_j$ such that

$$\Delta_j = \frac{1}{Ng'(\xi_j)}. \quad (5.3)$$

We wish to substitute (5.3) into (4.8), but we need (4.8) to resemble a Riemann sum. It is necessary to estimate Δ_j^2 in terms of $g'(x_j)$. Consider

$$\Delta_j^2 = \left[\frac{1}{Ng'(x_j)} \right]^2 + \frac{1}{N^2} \left[\left(\frac{1}{g'(\xi_j)} \right)^2 - \left(\frac{1}{g'(x_j)} \right)^2 \right]$$

$$= \left[\frac{1}{Ng'(x_j)} \right]^2 + \frac{1}{N^2} [g'(x_j) - g'(\xi_j)]$$

$$\cdot \left[\frac{1}{g'(x_j)} + \frac{1}{g'(\xi_j)} \right] \frac{1}{g'(x_j)g'(\xi_j)}.$$

From (5.1), g' is bounded below by $\delta > 0$ and is Lipschitz continuous. Thus

$$\Delta_j^2 = \left[\frac{1}{Ng'(x_j)} \right]^2 + \frac{1}{N^2} O(\Delta_j).$$

Let $h(x)$ be continuous on A. Then because h is bounded and the Δ_j vanish uniformly,

$$N^2 \sum_{j=1}^{N} h(x_j)\Delta_j^3 = \sum_{j=1}^{N} \left(\frac{1}{g'(x_j)} \right)^2 h(x_j)\Delta_j + O(N^{-1}). \quad (5.4)$$

Define

$$h(x, \alpha) = \frac{\alpha(1-\alpha)}{24} p^\alpha(x) q^{1-\alpha}(x) \left[\frac{p'(x)}{p(x)} - \frac{q'(x)}{q(x)} \right]^2. \quad (5.5)$$

This function is continuous for $x \in A$ and bounded on $A \times [0, 1]$. Thus (5.4) holds for $h = h(x, \alpha)$ and the convergence is uniform in α.

Lemma 1: For each $\alpha^* \in (0, 1)$, an open neighborhood $G(\alpha^*)$ of α^* exists such that

$$\lim_{N \to \infty} N^2(\hat{H}_\alpha - H_\alpha) = \int_A \left(\frac{1}{g'(x)} \right)^2 h(x, \alpha) \, dx \quad (5.6)$$

uniformly for $\alpha \in G(\alpha^*)$.

Proof: It remains to be shown that the right side of (5.4) converges to the integral in (5.6) uniformly in a neighborhood of α^*. Clearly, equicontinuity of the collection of integrands over $G(\alpha^*)$ would be sufficient. Note that function h has a bounded derivative (by inspection of (5.5)). In fact, a neighborhood of α^* must exist on which $(d/dx)h(x, \cdot)$ is bounded, giving us the desired equicontinuity.

The form of the preceeding limit differs slightly from [4] and [5]. The crux of our work is to prove uniformity over a collection of Ali–Silvey distance measures. This uniformity along with Chernoff's theorem will allow us to give a probability of error meaning to what would otherwise be an Ali–Silvey distance.

VI. MINIMIZATION OF ALPHA ENTROPY

To apply Chernoff's result, it is necessary to evaluate the minimum (in α) of H_α and $\hat{H}_\alpha$. We let α_0 be that value which minimizes H_α; it is assumed that $\alpha_0 \in (0, 1)$. At this point we can use α_0 in (5.6) to see that we have an upper bound to the change in asymptotic efficiency. The purpose of this section is to show that (5.6) provides more than an upper bound.

The shortcoming of (5.6) is that it is not necessary for the minimum of $\hat{H}_\alpha$ to occur at α_0. The following theorem shows that α_0 is, in fact, a good estimate.

Theorem 1: For each N, let $\gamma(N)$ be the value which minimizes $\hat{H}_\alpha$ at each level N. Then

$$\lim_{N \to \infty} N^2 \left(\hat{H}_{\gamma(N)} - H_{\alpha_0} \right) = \int_A \left(\frac{1}{g'(x)} \right)^2 h(x, \alpha_0)\, dx \quad (6.1)$$

with $h(x, \alpha)$ defined in (5.5).

Proof: The first step is to establish the convergence of $\gamma(N)$. We need the fact that H_α has a unique minimum.

Lemma 2: Let P and Q be probability measures on R which are dominated by a measure W. Define

$$p(x) = \frac{dP}{dW}(x), \qquad q(x) = \frac{dQ}{dW}(x).$$

Suppose that $\log(p(x)/q(x))$ is finite W almost everywhere. Then for $\alpha \in (0,1)$ and H_α defined in (2.2);

$$H'_\alpha \triangleq \frac{d}{d\alpha} H_\alpha = \int_R \left(\log \frac{p(x)}{q(x)} \right) p^\alpha(x) q^{1-\alpha}(x)\, dW(x)$$

$$H''_\alpha \triangleq \frac{d}{d\alpha} H'_\alpha = \int_R \left(\log \frac{p(x)}{q(x)} \right)^2 p^\alpha(x) q^{1-\alpha}(x)\, dW(x).$$

The proof of Lemma 2 is in [14]. It is based on a straightforward real variable argument. One can formally get the result by just pulling the derivatives through the integral. Observe that H''_α is positive so that H'_α is strictly increasing in α (unless $p = q$, which is trivial). Thus only one value of α minimizes H_α.

Let γ_0 be a cluster point of the $\gamma(N)$. Then for all $\epsilon > 0$ and $M > 0$, an $N \geq M$ exists such that

$$H_{\gamma(N)} > H_{\gamma_0} - \epsilon,$$

because H is continuous. It is true that $\hat{H} \geq H$ (which can be observed from an application of Hölder's inequality). Thus

$$\hat{H}_{\alpha_0} \geq \hat{H}_{\gamma(N)} \geq H_{\gamma(N)} > H_{\gamma_0} - \epsilon,$$

the first inequality following from the definition of $\gamma(N)$. Suppose $\gamma_0 \neq \alpha_0$. Then $H_{\gamma_0} > H_{\alpha_0}$ (strictly, because α_0 is the unique minimum) which allows us to choose ϵ such that $H_{\gamma_0} - \epsilon > H_{\alpha_0} + \epsilon$, giving

$$\hat{H}_{\alpha_0} > H_{\alpha_0} + \epsilon.$$

Reviewing the discussion, we see that this holds for infinitely many values of N (recall that $\hat{H}$ is a function of N). This however, violates the fact that $\hat{H} \to H$. It follows that $\gamma_0 = \alpha_0$.

Let $G(\alpha_0)$ be the neighborhood defined in Lemma 1. It follows that an $M > 0$ exists such that $\gamma(N) \in G(\alpha_0)$ whenever $N \geq M$. Define

$$\phi(\alpha) = \int_A \left(\frac{1}{g'(x)} \right)^2 h(x, \alpha)\, dx$$

and observe that ϕ is continuous (h is continuous in α and bounded). From the uniformity in Lemma 1 we have that,

for all $\epsilon > 0$, an $M' \geq M$ exists such that

$$\left| N^2 \left(\hat{H}_{\gamma(N)} - H_{\gamma(N)} \right) - \phi(\gamma(N)) \right| < \epsilon$$

whenever $N \geq M'$. Observe that

$$\hat{H}_{\alpha_0} - H_{\alpha_0} \geq \hat{H}_{\gamma(N)} - H_{\alpha_0} \geq \hat{H}_{\gamma(N)} - H_{\gamma(N)}.$$

Multiply all terms by N^2 and apply Lemma 1 to both upper and lower bounds, to get

$$\phi(\alpha_0) + \epsilon \geq N^2 \left(\hat{H}_{\gamma(N)} - H_{\alpha_0} \right) \geq \phi(\gamma(N)) - \epsilon,$$

but from the continuity of ϕ, $\phi(\gamma(N)) \to \phi(\alpha_0)$. Thus the above inequality must imply convergence to $\phi(\alpha_0)$ as $N \to \infty$.

VII. Optimization

As a first attempt to minimize (6.1) one may apply Lagrange techniques to find

$$(g'(x))^3 = \xi h(x)$$

for some ξ. The problem with this solution is that the function h may need to vanish, but our technical conditions say that g' cannot approach zero. To see this, examine (5.5) and (4.9). Note that any zero-slope point of the likelihood function forces $h = 0$.

To overcome this difficulty, we refer to the maximum principle of Pontryagin [10]. We restate our problem as an optimal trajectory problem with constraints. Call $g(t)$ the trajectory and $g'(t)$ the control. The problem is to minimize

$$\int_a^b \left(\frac{1}{g'(t)} \right)^2 h(t)\, dt \quad (7.1)$$

with the constraints that

$$g(t) = \int_a^t g'(x)\, dx, \qquad g(b) = 1,\ g'(t) \geq \delta > 0. \quad (7.2)$$

This is a problem of a time-variant cost with constraints of fixed end time and fixed end point.

Define the Hamiltonian

$$H(g, \xi, g', t) = \left(\frac{1}{g'} \right)^2 h(t) + \xi g' \quad (7.3)$$

where $\xi = \xi(t)$ is an auxiliary function to be determined. From Pontryagin, an optimal trajectory must satisfy the following:

1) $d\xi/dt(t) = -\partial/\partial g H(g, \xi, g', t)|_t$;
2) for each t, $g'(t)$ satisfies

$$H(g(t), \xi(t), g'(t), t) = \min_{\eta \geq \delta} H(g(t), \xi(t), \eta, t);$$

3) $H(g(t), \xi(t), g'(t), t) = H(g(b), \xi(b), g'(b), b) - \int_t^b (1/g'(\tau))^2 dh/dt(\tau)\, d\tau.$ (7.4)

The first condition reveals that ξ is constant. The second condition is satisfied for

$$g'(t) = \max\left[\left(\frac{2h(t)}{\xi}\right)^{1/3}, \delta\right]. \quad (7.5)$$

The third condition is satisfied for h continuous on $[a, b]$ with h' havingt finitely many discontinuities ("piecewise" continuous), and for g' given in (7.5). It remains to choose ξ to satisfy (7.2).

We now claim that this solution is optimal over δ-bounded quantization rules, with h given in (5.5). Consider the following collection of functions on $[a, b]$:

$$C_M = \left\{f: \delta \leq f \leq M, |f(x) - f(y)| \right.$$

$$\left. \leq M|x - y|, \int_a^b f(x) \, dx = 1\right\}.$$

Note that given any quantization rule with $g' \geq \delta$, an M exists such that $g' \in C_M$.

It is easy to verify that C_M is compact under the norm of uniform convergence (the "sup" norm). Thus any continuous functional defined on C_M must attain its minimum at an element of C_M. Lebesgue's dominated convergence theorem [11] assures us that (7.1) is continuous on C_M.

Lemma 3: Let g define a quantization rule on A with $g' \geq \delta > 0$. Suppose $h \geq 0$ is Lipschitz continuous on A with h' piecewise continuous. Define

$$r(x) = \max\left[(\xi h(x))^{1/3}, \delta\right]$$

where ξ satisfies

$$\int_a^b r(x) \, dx = 1.$$

Then $\int_a^x r(t) \, dt$ defines a quantization rule on A and

$$\int_a^b \left(\frac{1}{r(x)}\right)^2 h(x) \, dx \leq \int_a^b \left(\frac{1}{g'(x)}\right)^2 h(x) \, dx.$$

We call $\int_a^x r(t) \, dt$ the δ-optimum quantization rule.

Our main task in the proof is to find an M such that $g' \in C_M$ and $r \in C_M$. Then note that $r(x)$ is the only element of C_M which satisfies the necessary conditions. We know that C_M must contain the minimizing function. Thus it must be $r(x)$.

This proof rests on the compactness of C_M. The details involve application of the Ascoli–Arzelà theorem [12] of functional analysis, and Lebesgue's dominated convergence theorem. A rigorous proof is omitted (but can be found in [14]).

Suppose that we could allow $\delta = 0$ in Lemma 3. Define

$$g_0'(x) = h^{1/3}(x)\left[\int_a^b h^{1/3}(t) \, dt\right]^{-1},$$

which is possibly the $\delta = 0$ minimization of (7.1). Let $\langle \delta_k \rangle_{k=1}^\infty$ be a positive sequence vanishing monotonically. For each k, let g_k be the δ_k-optimum quantization rule.

Define

$$E_k = \int_a^b \left(\frac{1}{g_k'(x)}\right)^2 h(x) \, dx, \quad k = 0, 1, 2, \cdots. \quad (7.6)$$

Note that, for $k \geq 1$, the E_k cannot increase (i.e., g_{k+1} can only improve upon g_k). It is to be shown that $E_k \to E_0$.

There exist $\langle \xi_k \rangle_{k=1}^\infty$ such that

$$g_k'(x) = \max\left[\xi_k h^{1/3}(x), \delta_k\right],$$

and it is clear that the ξ_k converge so that $g_k' \to g_0'$. From Lebesgue's dominated convergence theorem, $g_k \to g_0$.

Consider the integrand in (7.6) and note that it converges to $h/(g_0')^2$. Also, $h/(g_k')^2$ is bounded for all k. By Lebesgue's dominated convergence theorem,

$$\lim_{k \to \infty} E_k = E_0 = \left[\int_a^b h^{1/3}(x) \, dx\right]^3. \quad (7.7)$$

Theorem 2: Let p and q satisfy (3.1). Let $\alpha_0 \in (0, 1)$ minimize H_α. Define $h(x) = h(x, \alpha_0)$, with $h(x, \alpha)$ defined in (5.5). For this function h, let E_0 assume its value in (7.7).

Recall the maximum likelihood detector and let e_n be defined as in Section II. Similarly, let $\hat{e}_n(N)$ be the detected error probability of an N-level quantized version of the source. Then for any $\epsilon > 0$, there exists a quantization rule such that

$$\lim_{N \to \infty} \lim_{n \to \infty} \frac{N^2}{n} \log \frac{\hat{e}_n(N)}{e_n} < \frac{E_0}{H_{\alpha_0}} + \epsilon. \quad (7.8)$$

Conversely, there is no quantization rule (as defined in Section V) such that

$$\lim_{N \to \infty} \lim_{n \to \infty} \frac{N^2}{n} \log \frac{\hat{e}_n(N)}{e_n} < \frac{E_0}{H_{\alpha_0}}. \quad (7.9)$$

Proof: The existence of a quantization rule is guaranteed by (7.7). The converse follows from Lemma 3 and the fact that E_0 is a lower bound for the E_k of (7.6). To see (7.8) consider

$$\lim_{N \to \infty} \lim_{n \to \infty} \frac{N^2}{n} \log \frac{\hat{e}_n(N)}{e_n} = \lim_{N \to \infty} N^2 \log \frac{\hat{H}_{\gamma(N)}}{H_{\alpha_0}},$$

which follows from Chernoff's result. However,

$$\log \frac{\hat{H}}{H} = \log\left[1 + \frac{\hat{H} - H}{H}\right]$$

$$= \frac{\hat{H} - H}{H} + o\left(\frac{\hat{H} - H}{H}\right).$$

Thus

$$\log \frac{\hat{H}_{\gamma(N)}}{H_{\alpha_0}} = \frac{\hat{H}_{\gamma(N)} - H_{\alpha_0}}{H_{\alpha_0}} + o(N^{-2})$$

because $\hat{H}_{\gamma(N)} - H_{\alpha_0} = O(N^{-2})$. (7.9) follows from this and Theorem 1.

VIII. EXTENSION TO UNBOUNDED DENSITIES

The problem encountered in this extension is that the log likelihood function may no longer be bounded. If it is bounded, then the random variables can be mapped into the unit interval, with the resultant densities satisfying (3.1). This is an easy exercise left to the reader.

We consider one-tailed densities. The two-tailed case is an obvious extension. The method is to define a class of densities and prove the existence of a companding function, which allows a result similar to Lemma 1. There are two classes of densities sharing the following properties:

1) $\{ p > 0 \} = \{ q > 0 \} = [0, \infty)$;
2) $\inf_{(0, M)} p > 0$ and $\inf_{(0, M)} q > 0$ for all $M > 0$;
3) both p and q are C^3 on $(0, \infty)$ with bounded third derivatives;
4) either $p \geq q$ or $p \leq q$ as $x \to \infty$.   (8.1)

Class 1 (Polynomial Rates): There exist $a > 1/\alpha_0$ and $b > 1/(1 - \alpha_0)$ such that

1) $0 < \liminf_{x \to \infty} x^a p(x)$, $\limsup_{x \to \infty} x^a p(x) < \infty$,
2) $\limsup_{x \to \infty} x^{a+2} |p''(x)| < \infty$,
3) $\liminf_{x \to \infty} x^b q(x) > 0$,
4) $\limsup_{x \to \infty} x^{b+2} |q''(x)| < \infty$.

Also $K > 0$ and $\epsilon > 0$ exist such that

5) $\left| \dfrac{d^2}{dx^2} p^\alpha(x) q^{1-\alpha}(x) \right| \leq K x^{-3-\epsilon}$, $x \in (0, \infty)$,

for all α in some neighborhood of α_0.

Class 2 (Exponential Rates): There exist a, b, and c such that for, all $\epsilon > 0$,

1) $\lim_{x \to \infty} e^{(a+\epsilon)x^c} p(x) = \infty$, $\lim_{x \to \infty} e^{(a-\epsilon)x^c} p(x) = 0$,
2) $\lim_{x \to \infty} e^{(a-\epsilon)x^c} p''(x) = 0$,
3) $\lim_{x \to \infty} e^{(b+\epsilon)x^c} q(x) = \infty$,
4) $\lim_{x \to \infty} e^{(b-\epsilon)x^c} q''(x) = 0$.

Also $K > 0$ and $\epsilon > 0$ exist such that

5) $\left| \dfrac{d^2}{dx^2} p^\alpha(x) q^{1-\alpha}(x) \right| \leq K e^{-\epsilon x^c}$, $x \in (0, \infty)$,

for all α in some neighborhood of α_0.

These classes are not restrictive and are usually easy to verify by inspection. We define a quantization rule g:

1) $g: [0, \infty) \to [0, 1)$ is one-to-one;
2) g' is Lipschitz continuous, except possibly at one point;
3) $\inf_{(0, M)} g' > 0$ for all $M > 0$.   (8.2)

Define a_j as in (5.2) with $a_N = \infty$.

Lemma 4: Let p and q be class 1 densities. If there is an $M > 0$ such that $g(x) = 1 - x^{-1/m}$ for $x > M$, then an m and $G(\alpha_0)$ exist such that

$$\lim_{N \to \infty} N^2 (\hat{H}_\alpha - H_\alpha) = \int_0^\infty \left(\frac{1}{g'(x)} \right)^2 h(x, \alpha) \, dx \quad (8.3)$$

uniformly for $\alpha \in G(\alpha_0)$. The function h is defined in (5.6).

Lemma 5: Let p and q be class 2 densities. If there is an $M > 0$ such that $g(x) = 1 - e^{-\xi x^c}$ for $x > M$, then ξ and $G(\alpha_0)$ exist such that (8.3) holds uniformly for $\alpha \in G(\alpha_0)$.

The proof of Lemma 4 is provided in the Appendix. Lemma 5 is proved in [14].

Theorem 3: Under the conditions of Lemma 4 or Lemma 5

$$\lim_{N \to \infty} N^2 (\hat{H}_{\gamma(N)} - H_{\alpha_0}) = \int_0^\infty \left(\frac{1}{g'(x)} \right)^2 h(x, \alpha_0) \, dx,$$

with $\gamma(N)$ defined in Theorem 1.

The proof of this theorem is given in Section VI. The only point to prove is that $\phi(\alpha) = \int_0^\infty (1/g'(x))^2 h(x, \alpha) \, dx$ is continuous. From the proof of Lemmas 4 and 5, $\phi(\alpha) < \infty$ for $\alpha \in G(\alpha_0)$. We have that $h(x, \alpha)$ is continuous in α, in fact monotone in α for large x due to (8.1). Thus ϕ must be continuous.

It remains to optimize the function g. Choose any function g which satisfies Theorem 3. For each $m \in Z^+$ and $k \in Z^+$, let $g_{m,k}$ satisfy the following (if possible):

1) $g_{m,k} = g(x)$ for $x \geq m$;
2) $g_{m,k}$ minimizes $\int_0^m (1/g'_{m,k})^2 h$ under the constraints $\int_0^m g'_{m,k} = g(m)$ and $g'_{m,k} \geq 1/k$ on $(0, m)$, $(h(x) = h(x, \alpha_0))$.   (8.4)

For $m > 1$, $g_{m,k}$ exists as $k \to \infty$. Condition 2) is guaranteed by Lemma 3 (with slight modifications).

From Section VII we know that for each m,

$$\int_0^m \left(\frac{1}{g'_{m,k}} \right)^2 h \xrightarrow[k \to \infty]{} \left(\frac{1}{g(m)} \right)^2 \left(\int_0^m h^{1/3} \right)^3. \quad (8.5)$$

Also for each $m > 1$ there is a k such that

$$\int_0^m \left(\frac{1}{g'} \right)^2 h \geq \int_0^m \left(\frac{1}{g'_{m,k}} \right)^2 h \geq \left(\frac{1}{g(m)} \right)^2 \left(\int_0^m h^{1/3} \right)^3. \quad (8.6)$$

From (8.6) we see that $h^{1/3}$ is integrable. Because the right side of (8.5) converges as $m \to \infty$, a companding function clearly can be chosen to perform arbitrarily close to $(\int_0^\infty h^{1/3})^3$.

Theorem 4: Let p and q be either class 1 or class 2 densities. Define $E_0 = (\int_0^\infty h^{1/3})^3$. Then, for all $\epsilon > 0$, a quantization rule exists such that (7.8) holds. Conversely, no quantization rule (as defined in this section) exists such that (7.9) holds.

The proof is in the preceding discussion along with the proof of Theorem 2.

IX. Uniform Quantizers

We wish to establish a result similar to Theorem 3 for a uniform quantizer with infinitely many bins. Define the partition as follows:

1) $a_j = j\Delta$, $j = 0,1,2,\cdots$,

2) $B_j = (a_{j-1}, a_j)$, $j = 1,2,3,\cdots$.   (9.1)

We shall consider the performance as $\Delta \to 0$.

Lemma 6: Let p and q be either class 1 or class 2 densities. Then a neighborhood $G(\alpha_0)$ of α_0 exists such that

$$\lim_{\Delta \to 0} \frac{1}{\Delta^2}(\hat{H}_\alpha - H_\alpha) = \int_0^\infty h(x,\alpha)\,dx$$

uniformly for $\alpha \in G(\alpha_0)$. The function h is defined in (5.5).

The proof of the lemma is in [14].[1] As in Section VIII, we establish the continuity of $\int_0^\infty h(x,\alpha)\,dx$ in α. This allows the following result.

Theorem 5: Let p and q be either class 1 or class 2 densities. Then

$$\lim_{\Delta \to 0} \frac{1}{\Delta^2}(\hat{H}_{\gamma(N)} - H_{\alpha_0}) = \int_0^\infty h(x,\alpha_0)\,dx.$$

X. Examples and Conclusion

Optimal and suboptimal quantizers exist for a large class of decision problems. In practice, it is Lemma 3 which provides the compressor design. Observe that the constant δ is arbitrary. With ever smaller choices of δ, the optimum compressor approaches the function

$$g_0' \propto \left[p^{\alpha_0} q^{1-\alpha_0} \left(\frac{p'}{p} - \frac{q'}{q} \right)^2 \right]^{1/3}.$$

As a first example, consider Gaussian densities

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x-s)^2 \right]$$

$$q(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x+s)^2 \right].$$

Then $\alpha_0 = 1/2$ by symmetry. The log likelihood function is a line and its slope (4.9) is constant. Thus $g_0' \propto p^{1/6}q^{1/6}$, which is a Gaussian density with mean zero and variance 3. It is interesting that g_0', the density of quantizer bins, does not depend on s. Table I provides a comparison of two eight-level quantizers, one chosen according to the function g_0 and the other chosen to uniformly divide $[-s-2\sigma, s+2\sigma]$, where $\sigma^2 = 1$ is the variance of p and q. The figure of merit, rate, is the asymptotic limit of $-(1/n)\log e_n$ (see Section II). Although the change in rate is small, any increase in rate indicates a significant decrease in error probability for large sample sizes. (The

[1] The limit, without the uniformity result, is already known [4].

probability of error for n samples should be of the form $k \cdot \exp[n \cdot I]$, where k is a slowly varying constant and I is a positive constant. Clearly, small variations in I cause tremendous variations in the error for large n.)

A second example is the Laplacian density. Let

$$p(x) = \frac{1}{2}\exp[-|x-s|] \qquad q(x) = \frac{1}{2}\exp[-|x+s|].$$

The fact that the derivatives have discontinuities is not a problem. The proofs can be extended to piecewise-C^3 densities. Again $\alpha_0 = 1/2$. The log likelihood function is a line for $|x| \le s$ and is constant elsewhere. Thus $g_0' = 0$ for $|x| > s$ and $g_0' \propto p^{1/6}q^{1/6}$ for $|x| \le s$, i.e., g_0' is constant. The result is that optimal quantization is uniform for $|x| \le s$. No partitioning is necessary for $|x| > s$, because the likelihood function is constant.

The following examples illustrate the effect of density shape and likelihood ratio on the resultant quantizer. We consider densities $p(x)$ on the unit interval tested against $q(x) = 1$. In this case $g_0' \propto [p^{\alpha_0}(p'/p)^2]^{1/3}$. Note that p'/p is the slope of the log likelihood function. Table II presents eight-level quantizers derived from g_0. Table III presents the asymptotic rates for the quantizers of Table II along with rates for eight-level uniform quantizers. Data are also given for corresponding four-level quantizers (the partitions are the second, fourth, and sixth points in Table II).

TABLE I
Gaussian Case

	Partition							Rate
$s = 1.0$								
Nonuniform	−1.99	−1.17	−0.55	0	0.55	1.17	1.99	0.482
Uniform	−3.0	−2.0	−1.0	0	1.0	2.0	3.0	0.462
$s = 0.1$								
Nonuniform	−1.99	−1.17	−0.55	0	0.55	1.17	1.99	0.00482
Uniform	−2.1	−1.4	−0.7	0	0.7	1.4	2.1	0.00479

TABLE II
Quantizers on [0,1]. $q = 1$

$p(x)$	Partition						
$29(x^{40} + 0.01)$	0.883	0.912	0.932	0.949	0.0964	0.977	0.989
$38(e^{-40x} + 0.001)$	0.015	0.030	0.048	0.068	0.092	0.121	0.160
$40e^{-40x}$	0.036	0.077	0.126	0.184	0.258	0.361	0.526
$7.1e^{-40x^2}$	0.151	0.236	0.313	0.389	0.471	0.567	0.698
$2.8e^{-40x^4}$	0.361	0.461	0.537	0.604	0.669	0.738	0.825
$\cos(2\pi x) + 1.001$	0.249	0.367	0.448	0.500	0.552	0.633	0.751
Uniform	0.125	0.250	0.375	0.500	0.625	0.750	0.875

TABLE III
Asymptotic Rates

		$N = 8$		$N = 4$	
$p(x)$	α_0	Nonuniform	Uniform	Nonuniform	Uniform
$29(x^{40} + 0.01)$	0.56	0.317	0.250	0.312	0.167
$38(e^{-40x} + 0.001)$	0.49	0.752	0.639	0.732	0.458
$40e^{-40x}$	0.27	1.351	1.224	1.295	0.964
$7.1e^{-40x^2}$	0.25	0.763	0.747	0.732	0.672
$2.8e^{-40x^4}$	0.25	0.410	0.398	0.394	0.364
$\cos(2\pi x) + 1.001$	0.41	0.097	0.090	0.071	0.061

The densities chosen are highly skewed to illustrate the interplay of the factors comprising g_0'. The first density is practically constant, except near $x = 1$, causing the partition to be skewed. The dramatic increase in rate over that of a uniform quantizer is of considerable significance. Even a four-level quantizer of this design performs quite well. The second and third densities produce partitions skewed toward the origin due to the concentration of $p(x)$. The difference in the partitions is due to the additive constant, which causes the likelihood ratio to be nearly constant for $x > 1/4$. A comparison of the third, fourth, and fifth densities further illustrates the effect of likelihood ratio on the resultant partitions. It is important to note that, although the change in rate may seem small, it necessarily indicates a substantial decrease in error probability for large sample sizes.

APPENDIX
PROOF OF LEMMA 4

Define for each $j = 1, \cdots, N$,

$$d_j = \left(\int_{B_j} p \right)^{\alpha} \left(\int_{B_j} q \right)^{1-\alpha} - \int_{B_j} p^{\alpha} q^{1-\alpha} \qquad (A.1)$$

with B_j defined in (3.3). It follows that

$$\hat{H}_{\alpha} - H_{\alpha} = \sum_{j=1}^{N} d_j. \qquad (A.2)$$

From Section V, we know that, for each $w > 0$,

$$\lim_{N \to \infty} N^2 \sum_{j: \, a_j \le w} d_j = \int_0^w \left(\frac{1}{g'(x)} \right)^2 h(x, \alpha) \, dx \qquad (A.3)$$

uniformly in some neighborhood of α_0. Observe the agreement with Lemma 3 for bounded densities.

Recall that we have allowed for a discontinuity in g'. This is to allow us more freedom in the construction of g. It will not affect the limit in (A.3), because the bin which straddles the discontinuity will have a value of d_j, which vanishes as N^{-3}.

We have to show that a neighborhood $G(\alpha_0)$ of α_0 exists such that the following holds. For all $\epsilon > 0$, a $W > 0$ and $M > 0$ exist such that

$$N^2 \sum_{j: \, a_j > w} d_j < \epsilon \qquad (A.4)$$

whenever $w \ge W$, $N \ge M$, and $\alpha \in G(\alpha_0)$. Given (A.3) and (A.4), we can say that there is a neighborhood $G(\alpha_0)$ of α_0 such that the following holds. For all $\epsilon > 0$, a $W > 0$ exists such that, for each $w \ge W$, there exists an $M > 0$ such that

$$\left| N^2 \left[\hat{H}_{\alpha} - H_{\alpha} \right] - \int_0^w \left(\frac{1}{g'(x)} \right)^2 h(x, \alpha) \, dx \right| < \epsilon, \qquad (A.5)$$

whenever $N \ge M$ and $\alpha \in G(\alpha_0)$.

The first implication is that $\int_0^{\infty} (1/g')^2 h < \infty$. To see this, assume it to be false. Then pick $w_2 > w_1 \ge W$ such that $\int_{w_1}^{w_2} (1/g')^2 h > 2\epsilon$. Note that there must be a value of N which satisfies (A.5) for both w_1 and w_2. This lowercase cannot be true.

With this observation it follows that the lemma is true. That is, it is now possible to pick $w = \infty$ in (A.5).

To prove (A.4), choose $\beta \in (0,1)$ and rewrite (A.2) as

$$\hat{H}_{\alpha} - H_{\alpha} = \sum_{j=1}^{[(1-\beta)N]} d_j + \sum_{j=[(1-\beta)N]+1}^{N-1} d_j + d_N \qquad (A.6)$$

where $[\cdot]$ denotes the greatest integer function. The constant β is related to the constant w in (A.3) by $(1 - \beta) = g(w)$. Thus we will examine the behavior as $\beta \to 0$.

The first term of (A.6) is covered in (A.3). Consider the second term. In a way similar to Section IV, expand the terms on each bin. For example, for each j there is a ξ_j such that

$$\int_{B_j} p(x) \, dx = p(x_j) \Delta_j + \frac{1}{24} p''(\xi_j) \Delta_j^3.$$

This is possible because p'' is continuous. Likewise, η_j and ψ_j exist such that

$$d_j = \left[p(x_j) + \frac{1}{24} p''(\xi_j) \Delta_j^2 \right]^{\alpha} \left[q(x_j) + \frac{1}{24} q''(\eta_j) \Delta_j^2 \right]^{1-\alpha} \Delta_j$$
$$- p^{\alpha}(x_j) q^{1-\alpha}(x_j) \Delta_j - \frac{1}{24} \left(\frac{d^2}{dx^2} p^{\alpha} q^{1-\alpha} \right) \bigg|_{\psi_j} \Delta_j^3.$$

Using the fact that $(1 + x)^{\alpha} \le 1 + \alpha x$, $x > -1$, it follows that

$$d_j \le p^{\alpha}(x_j) q^{1-\alpha}(x_j) \left[1 + \frac{\alpha}{24} \frac{p''(\xi_j)}{p(x_j)} \Delta_j^2 \right]$$
$$\cdot \left[1 + \frac{1-\alpha}{24} \frac{q''(\eta_j)}{q(x_j)} \Delta_j^2 \right] \Delta_j$$
$$- p^{\alpha}(x_j) q^{1-\alpha}(x_j) \Delta_j - \frac{1}{24} \left(\frac{d^2}{dx^2} p^{\alpha} q^{1-\alpha} \right) \bigg|_{\psi_j} \Delta_j^3.$$

Now perform some algebraic manipulations to see that

$$N^2 \sum_{j=[(1-\beta)N]+1}^{N-1} d_j \le N^2 \sum_j \left\{ \left| \frac{p''(\xi_j)}{p^{1-\alpha}(x_j)} \right| \Delta_j^3 + \left| \frac{q''(\eta_j)}{q^{\alpha}(x_j)} \right| \Delta_j^3 \right.$$
$$\left. + \left| \frac{p''(\xi_j)}{p^{1-\alpha}(x_j)} \frac{q''(\eta_j)}{q^{\alpha}(x_j)} \right| \Delta_j^5 + \left| \frac{d^2}{dx^2} p^{\alpha} q^{1-\alpha} \right|_{\psi_j} \Delta_j^3 \right\}. \qquad (A.7)$$

The discussion up to this point is applicable to the proofs of Lemmas 4–6. We now specialize to the case of Lemma 4. It is given that $g(x) = 1 - x^{-1/m}$ on the tail of the line. Thus

$$a_j = g^{-1} \left(\frac{j}{N} \right) = \left(1 - \frac{j}{N} \right)^{-m}.$$

For simplicity, let $i = N - j$. Then redefine

$$a_i = \left(\frac{N}{i} \right)^m, \qquad i = 1, 2, \cdots, [\beta N],$$

$$\Delta_i = a_{i-1} - a_i \le kN^m \frac{1}{i^{m+1}}, \qquad i = 2, 3, \cdots, [\beta N].$$

The constant k depends on m. Using the class 1 density assumptions, we analyze (A.7) as

$$N^2 \sum_{i=2}^{[\beta N]} \left| \frac{p''(\xi_i)}{p^{1-\alpha}(x_i)} \right| \Delta_i^3 \le k_1 N^2 \sum_{i=2}^{[\beta N]} (\xi_i)^{-a-2}(x_i)^{(1-\alpha)a} \Delta_i^3$$

$$\le k_1 N^2 \sum_{i=2}^{[\beta N]} (a_i)^{-a-2}(a_{i-1})^{(1-\alpha)a} \Delta_i^3$$

$$\le k_2 N^2 \sum_{i=2}^{[\beta N]} \left(\frac{i}{N} \right)^{m(a+2)}$$

$$\cdot \left(\frac{N}{i-1} \right)^{(1-\alpha)ma} \frac{N^{3m}}{i^{3m+3}}$$

$$\le k_3 N^{[-\alpha am + m + 2]} \sum_{i=2}^{[\beta N]} i^{[\alpha am - m - 3]}.$$

Choose m such that $\alpha am - m > 2$ in a neighborhood of α_0 (recall that $a > 1/\alpha_0$). Then

$$N^2 \sum_{i=2}^{[\beta N]} \left| \frac{p''(\xi_i)}{p^{1-\alpha}(x_i)} \right| \Delta_i^3 \le k_4 N^{[-\alpha am + m + 2]} (\beta N)^{[\alpha am - m - 2]}$$

$$= k_4 \beta^{[\alpha am - m - 2]} \xrightarrow[\beta \to 0]{} 0.$$

The constants k_1 through k_4 are chosen after the choice of m. Note that the convergence is uniform in a neighborhood of α_0. The q''/q^α term will vanish as $\beta \to 0$ by the same argument (pick m such that $(1-\alpha)bm - m > 2$).
Continuing,

$$N^2 \sum_{i=2}^{[\beta N]} \left| \frac{p''(\xi_i)}{p^{1-\alpha}(x_i)} \frac{q''(\eta_i)}{q^\alpha(x_i)} \right| \Delta_i^5$$

$$\le k_1 N^2 \sum_{i=2}^{[\beta N]} \left(\frac{i}{N} \right)^{[m(a+2)-(1-\alpha)am]} \left(\frac{i}{N} \right)^{[m(b+2)-\alpha bm]} \frac{N^{5m}}{i^{5m+5}}$$

$$= k_1 N^{[m(1-\alpha a + (1-\alpha)b) + 2]} \sum_{i=2}^{[\beta N]} i^{[m(\alpha a + (1-\alpha)b) - 5]}$$

$$\le k_2 N^{-2} \beta^{[m(\alpha a + (1-\alpha)b) - 4]}.$$

Choose m such that $[m(\alpha a + (1-\alpha)b) - 4] > 0$ so that the above will vanish as $\beta \to 0$. The constants k_1 and k_2 depend on m (these constants are not necessarily the same as in the previous discussion). The convergence is again uniform in a neighborhood of α_0.

Define $f(x) = p^\alpha(x)g^{1-\alpha}(x)$, and then

$$N^2 \sum_{i=1}^{[\beta N]} f''(\psi_i) \Delta_i^3 \le K_1 N^2 \sum_{i=2}^{[\beta N]} \left(\frac{i}{N} \right)^{(3+\epsilon)m} \frac{N^{3m}}{i^{3m+3}}$$

$$\le K_2 N^{2-m\epsilon} \sum_{i=2}^{[\beta N]} i^{\epsilon m - 3}$$

$$\le K_3 \beta^{\epsilon m - 2}.$$

Choose m such that $\epsilon m > 2$.
The final term of (A.6) is d_N. Consider the following:

$$N^2 d_N \le N^2 \left[\int_{N^m}^\infty p(x) \, dx \right]^\alpha$$

$$\le k_1 N^2 \left[\int_{N^m}^\infty x^{-a} \, dx \right]^\alpha$$

$$\le k_2 N^2 N^{-\alpha m(a-1)}.$$

Choose m such that $\alpha m(a-1) > 2$ in a neighborhood of α_0.
There are finitely many conditions which m must satisfy, and thus all the preceding constants may be chosen. Also, the intersection of the neighborhoods of α_0 yields a neighborhood of α_0. Thus (A.4) is true.

REFERENCES

[1] S. A. Kassam, "Optimum quantization for signal detection," *IEEE Trans. Commun.*, vol. COM-25, pp. 479–484, May 1977.

[2] H. V. Poor and J. B. Thomas, "Optimum quantization for local decisions based on independent samples," *J. Franklin Inst.*, pp. 549–561, June 1977.

[3] ——, "Applications of Ali–Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Trans. Commun.*, vol. COM-25, pp. 893–900, Sept. 1977.

[4] H. V. Poor, "A companding approximation for the statistical divergence of quantized data," in *Proc. 22nd IEEE Conf. Decision and Control*, San Antonio, TX, Dec. 1983.

[5] ——, "Fine quantization in signal detection and estimation," *IEEE Trans. Inform. Theory*, vol. 34, pp. 960–972, Sept. 1988, Part I.

[6] J. A. Bucklew, "Multidimensional digitization of data followed by a mapping," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 107–110, Jan. 1984.

[7] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, vol. 23, pp. 493–507, 1952.

[8] R. C. Buck, *Advanced Calculus*. New York: McGraw-Hill, 1978.

[9] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, July 1948.

[10] L. S. Pontryagin *et al.*, *The Mathematical Theory of Optical Processes*. New York: Wiley, 1962.

[11] W. Rudin, *Real and Complex Analysis*. New York: McGraw-Hill, 1974.

[12] K. Yosida, *Functional Analysis*. New York: Springer-Verlag, 1980.

[13] J. Wozencraft and I. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.

[14] G. Benitz, doctoral dissertation, Dept. Elec. and Comput. Eng., Univ. of Wisconsin, Madison, 1987.