"the random quantizer design can be treated as an estimate of the true optimal design . . ."

Statistical Analysis of a Quantizer Design Algorithm

by

Efren F. Abaya, Ph.D.*

Abstract

Suppose that a sequence of probability distribution functions $[F_n]$ converges weakly to a distribution function F. Does the sequence of optimal quantizers for the F_n's converge to an optimal quantizer for F? If so, do the respective distortions converge to the optimal distortion for F? It is shown that uniform integrability of the cost function with respect to the sequence $\{F_n\}$ is sufficient to obtain such convergence for mean-square distortion. These questions are used to motivate a study of the strong consistency properties of optimal quantizer designs based on sampled data.

Introduction

Pulse code modulation (PCM) appears to be the emerging new technology for long haul voice transmission. Its popularity derives in part from the robustness of digital transmission methods to noise and distortion, and in part from the simplicity and flexibility of digital transmission and processing. In PCM, the analog signal is converted at the transmitter to a digital format in a process known as analog-to-digital conversion. The process entails the addition of a controlled amount of distortion, or quantizing noise, to the original signal. In return, noise pick-up in the communication channel can be minimized. This is a trade-off which is fundamental to the design of PCM communication systems. Therefore, the analysis of the quantizing process is crucial to a determination of the overall system performance.

---

*Professor of Electrical Engineering, College of Engineering, University of the Philippines.

Quantization is perhaps the simplest and most popular method for analog-to-digital conversion. Roughly speaking, a quantizer is a device which constructs a discrete-amplitude approximation of a continuous-valued quantity. As such, quantization is philosophically akin to the process of rounding-off numbers. In fact, this viewpoint can be adopted in studying quantizing errors in digital filters, although this paper will not delve into this topic.

Because of its importance, quantization has been the object of many investigations. Lloyd [9] and Max [11] independently pioneered the probabilistic approach to quantizer design which has come to be known as the Lloyd – Max algorithm. This method takes the assumed probability density of the (random) input signal and by successive iterations converges on an optimal quantizer design. The optimality of the result may be limited only to the assumed density. Subsequent workers experimented with different techniques, such as fixed-point methods [8, 12], dynamic programming [5] and Newton – Raphson iterations [13]. Except for reference [8], all of these techniques share the common assumption that the probability distribution of the random inputs is completely specified. In practice, however, this assumption is rarely met because real sources do not conform to any standard statistical description. At most, the input distribution can be modelled as a member of some parametrized class of distributions, where the values of the parameters are unknown. In other situations, even less prior information may be available, so that it is necessary to resort to a nonparametric model.

Under these circumstances, a simple method readily suggests itself. The true input distribution F is not known and cannot be used in the quantizer design algorithms. Nevertheless, it is possible to form an estimate $\hat{F}_n$ based on n observations of the input signal. In the parametric case described above, the unknown parameters can be estimated; in the nonparametric case, ample statistical techniques are available for estimating a distribution function or a probability density. If $\hat{F}_n$ is used in any of the above-mentioned quantizer design algorithms, then it may be hoped for that the resulting quantizer should approximate an optimal quantizer for F. In a recent paper [8], a group of researchers showed experimentally and theoretically that this approach yields reasonable quantizer designs.

Notice that the resulting designs exhibit a random character, since they are based on random data. Alternatively, the distribution $\hat{F}_n$ on which the design is based is random; therefore so is the result. The premise of this paper is that the random quantizer design can be treated as an estimate of the true optimal design. Then it becomes possible to ask about statistical properties of the estimator, such as consistency, unbiasedness, variance, asymptotic distributions, and others. These properties give some indication of the usefulness and suitability of various design procedures.

In this paper we consider the strong consistency property of the above quantizer estimators. Roughly speaking, consistency means that as more data are gathered ($n \to \infty$) the design should "converge" to the true optimal quantizer for the unknown distribution F. Of the different types of convergence, the most useful to the designer is strong or with probability 1 convergence [7]. This ensures that for practically all data sets, the resulting designs tend to the optimal one. The problem studied here is to investigate conditions under which the strong consistency properly holds. This question is made more precise in Section III, using some notations and terminology developed in Section II.

In Section IV, several propositions and a theorem are developed as tools for investigations on strong consistency. Finally, in Section V, it is shown by two examples how the tools may be applied to study two design procedures.

Optimal Quantizers

An N-level scalar quantizer Q is a mapping from the real line into the real line which assigns to the input x an output Q(x) chosen from a finite set of N distinct points (called output levels) $y_1 < y_2 < \dots y_n$. Let $\{P_1, P_2, \dots, P_n\}$ be an exhaustive partition of the real line. Then $Q(x) = y_i$ if x lies in P_i. We will take the viewpoint that Q (x) is intended to approximate x so that P_i logically becomes an interval and y_i is chosen to be some representative value in that interval. Graphically, Q (x) is then an ascending staircase function with N segments.

Generally, the input X is regarded as a random variable having a known probability distribution function F (x). It is desired to approximate X by the N-valued discrete random variable Q (X). The accuration of the approximation is expressed in terms of an r – th power distortion measure

$$\begin{aligned} D(Q,F) &= E\,[\,|X - Q(X)|^r\,] \\ &= \int |x - Q(x)|^r \, dF(x) \qquad (2.1) \end{aligned}$$

where the integral is taken in the Lebesgue-Stieltjes sense. This is simply an average error measure. Two popular choices for the exponent are r = 2 which gives the mean-square error power and r = 1 which gives the mean-absolute error.

An N-level quantizer is said to be optimum if it achieves the smallest distortion (2.1) among all quantizers having the same number of levels. The existence of this minimum has been shown elsewhere [2].

We may use PCM as an illustration of the use of quantization. The objective of a PCM transmitter is to convert a random analog waveform x(t) representing, for example, a speech utterance, into a sequence of digitally coded numbers. This is achieved by two operations – sampling and quantization. In the process of sampling, only the values of the signal at regularly spaced points in time t_n are retained. All other parts of the waveform are discarded. It can be shown [15, p 400] that if the waveform is band limited and the sampling rate is high enough, then the original waveform can be recovered exactly by the following formula.

$$x(t) = \sum_{n=-\infty}^{\infty} x(t_n) \frac{\sin \Pi f_s (t - t_n)}{\Pi f_s (t - t_n)} \qquad (2.2)$$

In this formula f_s is the sampling rate.

As a result of sampling, the signal is reduced to a sequence of amplitudes at well-defined points in time. Theoretically, this step does not introduce any distortion. Next, the sample values are quantized individually. Letting $x_n = x(t_n)$ be random variables, the resultant signal which is transmitted is the sequence $\{. \ . \ . \ Q(x_{-1}), \ Q(x_0), \ Q(x_2), \ . \ . \ .\}$.

This sequence can be digitally coded at the transmitter. The receiver constructs an approximation to the original waveform by

$$\tilde{x}(t) = \sum_{n=-\infty}^{\infty} Q(x_n) \frac{\sin \Pi f_s (t - t_n)}{\Pi f_s (t - t_n)} \qquad (2.3)$$

This approximation may be viewed as the original signal plus additive quantization noise

$$\tilde{x}(t) = x(t) + n(t)$$

where

$$\tilde{x}(t) = \sum_{n=-\infty}^{\infty} [x_n - Q(x_n)] \frac{\sin \Pi f_s (t - t_n)}{\Pi f_s (t - t_n)}$$

If the waveform x(t) is bandlimited, stationary, ergodic and satisfies a few other technical assumptions, then the mean-square instantaneous noise power is [9]

$$E\{n^2(t)\} = E\{ |x_n - Q(x_n)|^2 \}$$

where the expectation is with respect to the common distribution of the samples. Thus under ideal conditions, the noise perceived at the receiver is precisely the mean-square distortion introduced by quantizing. To maximize the fidelity of transmission, we must minimize the quantizer distortion.

Lloyd [9] and Max [11] independently formulated the necessary conditions for optimality of a quantizer. Let the quantization intervals be $P_i = (x_{i-1}, x_i)$, $i = 1, 2, \ldots, N$. For the case of mean-square distortion, the Max-Lloyd necessary conditions are:

$$\int_{x_{i-1}}^{x_i} (x - y_i)\, dF(x) = 0 \qquad 0 = 1, 2, \ldots, N$$

$$x_i = (y_i + y_{i+1})/2 \qquad i = 1, 2, \ldots, (N - 1) \quad (2.4)$$

An optimum mean-square distortion quantizer must satisfy these two conditions simultaneously. The first one says that each output level y_i is a conditional mean for the given quantizing interval. The second one says that the endpoints of each interval lie midway between the two adjacent output levels. If this prescription is followed, then $Q(x) = y_i$ if y_i is the output level closest to x. Thus the second condition is aptly named "nearest neighbor assignment".

An algorithm for constructing optimal quantizers, known as Lloyd's Method I [9], is based on the Max – Lloyd conditions. The given quantities are the number of desired levels N and the distribution F(x) of the input random variable. With these inputs, the algorithm proceeds as follows.

Step 1. Select an arbitrary starting set of output levels

$$y_1 < y_2 < \ldots < y_n.$$

Step 2. (Nearest neighbor assignment). Construct the quantizing intervals so that

$$P_1 = (-\infty, x_1) \qquad P_n = (x_{N-1}, \infty)$$

$$P_i = (x_{i-1}, x_i) \qquad i = 2, \ldots, N-1$$

$$x_i = (y_i + y_{i+1})/2 \qquad i = 1, 2, \ldots, N-1$$

Step 3. (Centroidal assignment). Select a new set of output levels to be the conditional means of the intervals constructed in Step 2. That is,

$$y_i = E\,[\,X \mid X \in P_i\,] \qquad i = 1, 2, \ldots, N$$

Step 4. Repeat Steps 2 and 3 in turn until the output levels converge.

It has been shown that output levels constructed by the above procedure will converge. The set of N limit levels then serves to define a quantizer which simultaneously satisfies the Max-Lloyd conditions (2.4). In most cases, this quantizer is optimal for the distribution F [8].

Lloyd's Method I presupposes that the input distribution F is known. For example, in the case of PCM, F would be the amplitude distribution of speech (assumed independent of time). In many cases, it is impossible to obtain this knowledge. One way around this problem is to take n independent random samples from the distribution F and form the empirical distribution function

$$\hat{F}_n(x) = (\#\text{ of samples} \leq x)/n \qquad (2.5)$$

By the Glivenko-Cantelli Theorem [16]

$$\lim_{n \to \infty} \hat{F}_n(t) = F(t) \qquad \text{almost surely} \qquad (2.6)$$

at all continuity points of F. In other words, $\hat{F}_n$ is a strongly consistent estimator of F. We may construct an optimal quantizer $\hat{Q}_n$ for $\hat{F}_n$ using Lloyd's Method I. In fact, the implementation of the algorithm turns out to be particularly simple for empirical distributions [8]. The quantizer $\hat{Q}_n$ is not optimal for F, but it may be viewed as an estimator of the quantizer Q which is optimal for F. Note that the estimator is a random function. In turn, $\hat{Q}_n$ has output levels $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}$ which are estimators of the corresponding output levels $y_1, y_2, \ldots, y_N$ for Q.

Experiments reported in [8] using speech data show that the $\hat{y}_i$ can be close to y_i for moderately large sample sizes. From a design standpoint, it is desirable to know that this always happens. That is

$$\lim_{n \to \infty} \hat{Q}_n = Q$$

or that

$$\lim_{n \to \infty} \hat{y}_i = y_i \qquad i = 1, 2, \ldots, N.$$

Of the different types of stochastic convergence, the most useful is almost sure convergence [7], and it is the one which this paper uses.

Statement of the Problem

The considerations given at the end of the previous section motivate the formal problem to be stated below. The application to quantizer design is considered later. First we define a term to be used in the problem statement.

Definition. A sequence $\{F_n\}$ of probability distribution functions is said to *converge weakly* to a distribution function F if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

at all continuity points x of F. Weak convergence is denoted by $F_n \to F$

Now suppose that a sequence of distribution functions $\{F_n\}$ converges weakly to a distribution function F. Let Q_n be an optimal N-level quantizer for F_n. Two interesting questions may be posed:

(Q1) Do the distortions $D(Q_n, F_n)$ converge to D(Q, F)?

(Q2) Does the sequence of optimal quantizers $\{Q_n\}$ converge in some sense to a quantizer Q? If so, is Q optimal for F?

The condition $F_n \to F$ may loosely be interpreted to mean that the distributions F_n become more and more similar to F. The second question then asks if the optimal quantizers for F_n eventually become close to a (hopefully optimal) quantizer for F. We have not yet defined how quantizers converge. This will be done in the next section.

Development

We begin the analysis by defining what it means for a sequence of quantizers to converge. To ease the exposition, we will limit ourselves in this paper to a mean-square distortion criterion (r = 2 in Eq. 2.1). Recall that by the Max-Lloyd necessary conditions (2.4), an optimal quantizer must satisfy the "nearest neighbor assignment" rule. Even if a quantizer Q is not optimal, the distortion may be reduced by applying nearest neighbor assignment (Step 2 of Lloyd's Method I). That is, we retain the output levels of Q and change the quantizing intervals as indicated, thereby getting a modified quantizer Q' having the same output levels but a smaller distortion. In the sequel we will assume that all quantizers have been treated in the way, and so satisfy the nearest neighbor assignment rule. This being so, a quantizer Q is completely specified by its output levels $y_1, y_2, \ldots, y_n$. To evaluate Q(x) is simply a matter of picking the output level closest to x.

With the output level representation, it sounds natural to say that a sequence of N-level quantizers $\{Q_n\}$ converges to a quantizer Q if the output levels of the Q_n's converge respectively to the output levels of Q. This criterion is just slightly deficient because the limit quantizer Q may have less than N levels. Thus we cannot use the product topology to define convergence of quantizers. Instead we turn to the functional representation of a quantizer, which gives the following usable definition.

Definition. A sequence of N-level quantizers $\{Q_n\}$ is said to converge to a quantizer Q if $Q_n(x) \to Q(x)$ at all continuity points of Q.

Notice that this definition allows the limit quantizer Q to have fewer than N levels. Because of nearest neighbor assignment, the definition is consistent with the notion of convergence presented earlier. It follows from the definition that

$$\lim_{n \to \infty} [x - Q_n(x)]^2 = [x - Q(x)]^2 \qquad (4.1)$$

uniformly over compact intervals. This property will prove to be useful later on.

It is possible for the output levels of the Q_n's not to converge to any finite point. In this case we have the following definition.

Definition. A sequence of N-level quantizers $\{Q_n\}$ is said to diverge if $Q_n(x) \to \infty$ for all x.

Question (Q1) asks about the convergence of distortions $D(Q_n,F_n)$. To simplify the situation, suppose that Q_n has a single level y_1, which is the same for all n. Also, let $F_n \to F$. Intuitively, one feels that since the F_n are getting "close" to F, the distortions $D(Q_n, F_n)$ should be near D(Q, F). That is, it would be convenient to be able to say that

$$\lim_{n \to \infty} \int (x - y_1)^2 \, dF_n = \int (x - y_1)^2 \, dF.$$

This happens if and only if the integrand is uniformly integrable with respect to $\{F_n\}$ (7, p 138) . Recall that a function g(x) is uniformly integrable with respect to $\{F_n\}$ if

$$\int |g(x)| \, dF_n(x) < \infty \text{ for all } n$$

and

$$\lim_{a \to \infty} \sup_n \int_{|x|>a} |g_n(x)| \, dF_n(x) = 0.$$

Thus, uniform integrability appears to be just the right condition to obtain convergence of the distortions. This is proven in the following proposition.

Proposition 1. Suppose that $F_n \to F$ and that a sequence of N-level quantizers $\{Q_n\}$ converges to a quantizer Q. If $(x-y)^2$ is uniformly integrable with respect to $\{F_n\}$ for each y, then $D(Q_n, F_n) \to D(Q, F)$.

Proof. Recall that $(x - Q_n(x))^2 \to (x - Q(x))^2$ uniformly over compact intervals. Let

$$T_n = X_n - Q_n(X_n))^2 \sim G_n$$

$$T = (X - Q(X))^2 \sim G$$

when X_n and X have the distributions F_n and F, respectively. By a theorem in [4, p. 34] it follows that $G_n \to G$ Thus, if T_n is uniformly integrable with respect to $\{G_n\}$, it follows that $D(Q_n, F_n) \to D(Q, F)$ [4, p. 32]. It is, however, more convenient to have a condition in terms of the F_n 's. Assume that Q has N output vectors $y_1, y_2, \ldots, y_N$. Imagine a hypercube (with sides of unit length) about y_1, and denote the midpoints of the 2k faces by z_1^j, $j = 1, 2, \ldots, 2k$, Then

$$(x - y_1)^2 \leq \sum_{j=1}^{2k} (x - z_1^j)^2$$

Because of nearest neighbor assignment, it follows that

$$[x - Q_n(x)]^2 \leq \sum_{j=1}^{2k} (x - z_1^j)^2$$

Therefore T_n is uniformly integrable with respect to $\{G_n\}$ if $(x-y)^2$ is uniformly integrable with respect to $\{F_n\}$ for all vectors y.

QED

Notice that the Q_n are not necessarily optimal. Also, the limit quantizer Q may have less than N levels. For the uniform integrability condition, it is sufficient to check that x^2 is uniformly integrable with respect to $[F_n]$.

The next result answers part of question (Q2).

Proposition 2. Under the conditions of Prop. 1, if Q_n is an optimal N-level quantizer for F_n, then Q is optimal for F.

Proof. Suppose that Q_n is optimal for F_n. Then $D(Q_n, F_n) \leq D(Q', F_n)$ for any quantizer Q' having N levels or less. By taking limits on both sides of the inequality, it follows that

$$D(Q, F) \leq D(Q', F).$$

Thus the limit quantizer Q is optimal for the distribution F.

QED

We now address the first part of question (Q2), namely: Does a sequence of optimal quantizers converge? In general, the answer is no, because there may be several quantizers which are optimal for the limit distribution. However, all that we really need in applications is a subsequence of $\{Q_n\}$ which converges. This can be isolated by the procedure described in the next proposition.

Proposition 3. Under the conditions of Prop. 2, every subsequence of $[Q_n]$ contains a further subsequence which converges to some quantizer.

Proof. Consider the sequence $\{y_{1,n}\}$ of first output levels of the Q_n's. If this has a finite limit point, then select a convergent subsequence; otherwise, retain the original sequence. To economize on notation, this subsequence will be denoted by $[y_{1,n'}]$. The corresponding quantizers are denoted by $Q_{n'}$. Next, take the sequence $(y_{2,n'})$ of second output levels and select a convergent subsequence $\{y_{2,n''}\}$, if possible; otherwise keep the sequence $(y_{2,n'})$. Do this for all the output levels in turn. This procedure either produces a convergent subsequence, or it determines that none of the sequences $\{y_{i,n}\}$ have finite limit points, i.e., the sequence $\{Q_n\}$ diverges.

We will now show that the latter case is impossible. Suppose that $\{Q_n\}$ diverges. Then

$$\lim_{n \to \infty} [x - Q_n(x)]^2 = \infty \text{ for all } x.$$

Let B denote a closed ball whose surface has F – probability zero, and whose interior has probability greater than 0.9, i.e.

$$\int_B dF > 0.9.$$

Let M be an arbitrary number. Then for n sufficiently large

$$\int_B dF_n > 0.9$$

$$[x - Q_n(x)]^2 > M \qquad \text{for } x \in B.$$

This implies that for large n,

$$\int [x - Q_n(x)]^2 \; dF_n > 0.9 \; M$$

and therefore

$$\lim_{n \to \infty} D(Q_n, F_n) = \infty. \tag{4.2}$$

Mimicking the proof of Prop. 2, we conclude that every quantizer Q with N levels or less has infinite distortion D(Q, F). But we also have, by virtue of uniform integrability, that

$$D(Q, F) \leq = \lim\inf \int [x - Q(x)]^2 \; dF_n < \infty.$$

This contradicts (4.2) which means that there has to be a convergent subsequence of $\{Q_n\}$. Clearly, the same argument can be made for any subsequence of $\{Q_n\}$. This completes the proof of the Proposition.

QED

We are now in a position to answer the two questions posed in Sect. III. This is done in the following theorem.

Theorem 1. Assume that $F_n \to F$, and that Q_n is an optimum N-level quantizer for F_n. If $(x\text{-}y)^2$ is uniformly integrable with respect to $\{F_n\}$ for each y, then every convergent subsequence of $\{Q_n\}$ converges to an optimal quantizer for F. Such a subsequence always exists. Moreover, $\{D(Q_n, F_n)\}$ converges to the optimal distortion (mean square error) for quantizing F with N levels.

Proof. The first two assertions are merely re-statements of Propositions 2 and 3. To show the last assertion, note that by Propositions 1 and 3, every subsequence of $\{D(Q_n, F_n)\}$ contains a further subsequence which converges. The limiting distortion is the same in every case (Prop. 2), and equals the optimal distortion for quantizing F. Therefore the entire sequence $\{D(Q_n, F_n)\}$ converges.

QED

Applications

The conditions hypothesized in the previous sections may arise when an N-level scalar quantizer is to be designed for an unknown univariate distribution F. One approach to this problem is to take n independent samples from F and form the empirical distribution

$$\hat{F}_n(x) = (\# \text{ of samples} \leq x) \; / \; n. \tag{5.1}$$

According to the Glivenko – Cantelli Theorem [16] $\hat{F}_n$ converges uniformly to F with probability 1 as $n \to \infty$. By the Strong Law of Large Numbers [10, p. 239]

$$\lim_{n \to \infty} \int_{|x| > a} (x - y)^2 \, d\hat{F}_n(x) = \int_{|x| > a} (x - y)^2 \, dF(x)$$

with probability 1 provided that the second moment of F is finite. Thus $(x-y)^2$ is uniformly integrable with respect to the sequence of empirical distributions $\{\hat{F}_n\}$. As discussed in the introduction, we may choose a suitably large number of observations n as the basis for a quantizer design. Then the optimal quantizer Q_n for $\hat{F}_n$ serves as an estimate of the desired optimal quantizer for the unknown distribution F. According to Theorem 1, this estimate is strongly consistent with probability 1, that is, Q_n converges to Q as $n \to \infty$. Moreover the distortion $D(Q_n, \hat{F}_n)$ computed for the estimate is also a strongly consistent estimate of the optimal distortion for F.

The analysis can be extended to vector quantizers. Let $(X^1, X^2, \ldots, X^k)$ denote the scalar components of a k-dimensional vector X. Similarly the marginal distributions of a probability distribution F on R^k will be denoted by $F^1, F^2, \ldots, F^k$. Let

$$I = \mathop{\times}_{i=1}^{k} [-a, a].$$

We have

$$\int_{I^c} (x - y)^2 \, dF(x) = \sum_{i=1}^{k} \int_{I^c} (x^i - y^i)^2 \, dF(x)$$

$$\leq \sum_{i=1}^{k} \int_{|x^i| > a} (x^i - y^i)^2 \, dF^i(x^i).$$

Each of the terms on the right – hand side involve scalar quantities of the type previously discussed. Therefore multi-variate empirical distribution functions also result in uniformly integrable sequences (with probability 1) provided that all second moments are finite. The succeeding developments readily generalize to several dimensions, including the Glivenko – Cantelli Theorem [6]. Thus the comments made above for scalar quantizers apply to vector quantizers also. Numerical experiments verifying these conclusions are presented in [8].

As a second example, consider that the probability density f to be quantized is univariate normal with unknown mean μ and variance $\sigma^2 > 0$.

Let $n(x)$ denote the standard normal density

$$n(x) = \frac{1}{\sqrt{2\Pi}} e^{-x^2/2}$$

Then we may write

$$f(x) = \frac{1}{\sigma} n\left(\frac{x-\mu}{\sigma}\right) \tag{5.2}$$

Let $X_1, X_2, \ldots$ be independent, identically distributed samples from f. Strongly consistent estimates of the unknown parameters μ and σ^2 are, respectively, the sample mean and variance

$$\hat{\mu}_n = n^{-1} \sum_{i=1}^{n} X_i$$

$$\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2 \tag{5.3}$$

A reasonable estimate for f is a normal density with these parameters, i.e.,

$$\hat{f}_n(x) = \frac{1}{\hat{\sigma}_n} n\left(\frac{x - \hat{\mu}_n}{\hat{\sigma}_n}\right) \tag{5.4}$$

Clearly, this estimator is strongly consistent, so that $\hat{f}_n$ converges weakly to f. Moreover,

$$\int (x-y)^2 \hat{f}_n(x)\, dx = \hat{\sigma}_n^2 + (\mu_n - y)^2$$

Since the estimates are strongly consistent, it follows that

$$\lim_{n\to\infty} \int (x-y)^2 \hat{f}_n(x)\, dx = \int (x-y)^2 f(x)\, dx$$

with probability 1. This is equivalent to having $(x-y)^2$ uniformly integrable with respect to $\{\hat{f}_n\}$. Therefore the optimal quantizer Q_n designed for density $\hat{f}_n$ is a strongly consistent estimate of the optimal quantizer for the unknown f.

The family of normal densities used in this example can be replaced with any class of almost everywhere continuous densities parametrized by location and/or scale parameters. For example, the Laplace, gamma and Rayleigh densities have been occasionally used to model the distribution of speech amplitudes.

Many other estimation methods for probability densities can be used in quantizer design. The strong consistency properties of designs based on these methods will be reported in another paper.

Conclusion

We have introduced the concept of convergence of quantizer sequences, and used it to investigate strong consistency properties of quantizer designs based on sampled data. Several propositions and a theorem were developed as tools in the study. Uniform integrability of certain functions with respect to the estimated distributions is shown to be a sufficient criterion for strong consistency. Finally, these propositions were applied to two procedures for designing quantizers; one involves empirical distributions and the other uses density estimates.

For clarity of exposition, the discussion was limited to mean-square distortion for vector quantizers. However, the methods described are capable of much greater generalization. Other treatments similar to the one described in the pages may be found in references [1] and [3].

References

1. Efren F. Abaya, *A Theoretical Study of Optimal Vector Quantizers,* Ph.D. dissertation, The University of Texas at Austin, 1982.

2. Efren F. Abaya and Gary L. Wise, "On the existence of optimal quantizers," *IEEE Trans. Inform. Th.,* Vol. IT – 28, No. 6, pp. 937-940, Nov. 1982.

3. Efren F. Abaya and Gary L. Wise, "Convergence of vector quantizers with applications to optimal quantization," to appear in *SIAM Journ. Appl. Math.*

4. Patrick Billingsley, *Convergence of Probability Measures,* N. Y.: John Wiley and Sons, Inc., 1968.

5. James D. Bruce, "On the optimum quantization of stationary signals," *IEEE International Convention Record Pt. I,* 1964, pp. 118-24.

6. Luc P. Devroye, "A uniform bound for the deviation of empirical distribution functions," *Journal of Multivariate Analysis*, Vol. 7, No. 4, pp. 594-57, Dec. 1977.

7. R. G. Laha and V. K. Rohatgi, *Probability Theory*, N. Y.: John Wiley and Sons, 1979.

8. Yoseph Linde, Andres Buzo and Robert M. Gray, "And algorithm for vector quantizer design," *IEEE Trans. Commun. Tech.*, Vol. Com-28, No. 1, pp. 84-95, Jan. 1980.

9. S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Th.*, Vol. IT-28, No. 2, pp. 129-137, Mar. 1982.

10. Michel Loeve, *Probability Theory*, N.J.: D. Van Nostrand Company, Inc., 1963.

11. Joel Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Th.*, Vol. IT-6, No. 1, pp. 7-12, Mar. 1960.

12. A. N. Netravali and R. Saigal, *"Optimum quantizer design using a fixed point algorithm"* *BSTJ*, Vol. 55, No. 9, pp. 1423-35, Nov. 1976.

13. W. A. Pearlman and G. H. Senge, "Optimal quantization of the Rayleigh probability distribution," *IEEE Trans. Commun.*, Vol. Com - 27, No. 1, pp. 101-112, Jan. 1979.

14. H. L. Royden, *Real Analysis*, 2nd ed., N.Y.: Macmillan Publishing Co., Inc., 1968.

15. John B. Thomas, *An Introduction to Statistical Communication Theory*, N.Y.: John Wiley and Sons, Inc., 1969.

16. H. G. Tucker, *A Graduate Course in Probability*, N.Y.: Academic Press, 1967.