MINIMAX QUANTIZATION FOR DISTRIBUTED MAXIMUM LIKELIHOOD ESTIMATION

Parvathinathan Venkitasubramaniam, Lang Tong, and Ananthram Swami

ABSTRACT

We consider the design of quantizers for the distributed estimation of a deterministic parameter, when the fusion center uses a Maximum-Likelihood estimator. We define a new metric of performance, which is to minimize the maximum ratio between the Fisher Information of the unquantized and quantized observations. Since the estimator is M-L, the criterion is equivalent to the minimizing the maximum asymptotic relative efficiency due to quantization. We propose an algorithm to obtain the quantizer that optimizes the metric and prove its convergence. Through simulations, we illustrate that the quantizer performance is close to the best possible Fisher Information as number of quantization bits increases. Furthermore, under certain conditions, the quantizer structure is found to belong to the class of score-function quantizers, which maximize Fisher Information for a given value of the parameter.

1. INTRODUCTION

1.1. Motivation

Distributed estimation refers to the statistical inference problem where observations from several distributed nodes are collected at a common fusion center which estimates the underlying parameter. Such an estimation system finds numerous applications especially in the context of sensor networks where the estimated parameter is indicative of the phenomenon being sensed. Unlike classical point estimation, here it is impractical to assume that the fusion center has direct access to the observed measurements. The data transmission to the fusion center is severely restricted by communication and energy constraints and nodes are allowed to transmit only a compressed version of the data. In this work, we focus on the optimal quantization of the observations to maximize estimation performance.

Consider a system as shown in Figure 1. Each node measures a random observation X_i independent and identically distributed under a deterministic scalar parameter θ. The nodes then transmit quantized versions of the observations, $\{\gamma_i(X_i)\}$, to the fusion center. We assume that the quantizers are identical, i.e., $\gamma_i(\cdot) = \gamma(\cdot)$. This assumption is motivated by the i.i.d distributions of observations and the possibility of using Type-Based Multiple Access[1] for energy efficient communication to the fusion center.

The fusion center uses the received data $\{\gamma(X_i)\}$ to make an estimate $\hat{\theta}$ of the true parameter. It is well known that the mean squared error (MSE) of any unbiased estimate is lower bounded by the inverse of Fisher Information (FI). Furthermore, when the quantized observations are distributed according to some probability mass function $\mathbf{q}_\theta$, and a Maximum-Likelihood estimator based on q_θ
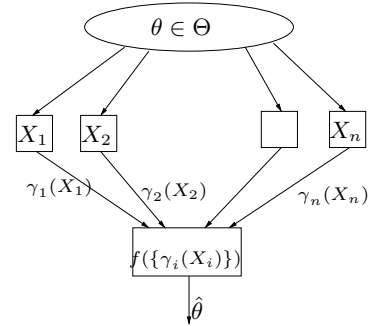
Fig. 1. Distributed Estimation System

is implemented, the MSE of $\hat{\theta}$ asymptotically achieves this bound. The best possible quantizer for an M-L estimator would therefore maximize FI for every value of θ. However, since the FI varies with θ, a single quantizer cannot be optimal for all θ.

We know that, for every θ, the Fisher Information I_θ of the quantized observations is less than the FI J_θ of the unquantized $\{X_i\}$. Therefore, our goal is to design a quantizer that minimizes the maximum ratio between these quantities over θ. In other words, we wish to design $\gamma^* = \arg\min_\gamma \max_\theta \frac{J_\theta}{I_\theta}$. This criterion is also equivalent to minimizing the maximum Asymptotic Relative Efficiency (ARE) between consistent estimators based on quantized and unquantized observations. Furthermore, for source distributions with constant J_θ, such a quantizer requires the least sample size for a given asymptotic error variance amongst all deterministic quantizers.

1.2. Main Contributions

In this work, we define a criterion for performance that minimizes the maximum ARE between M-L estimators based on quantized and unquantized measurements under a deterministic parameter. We propose an iterative algorithm to obtain a quantizer that optimizes the minimax criterion and prove its convergence. We then describe a class of quantizers called *Score Function Quantizers* which can be used to obtain the maximum achievable Fisher Information for a given parameter θ. Through numerical simulations, we compare the performance of the minimax quantizer and optimal Score-Function quantizers. We also provide insights into the possible structure of these quantizers under some conditions on source distribution.

1.3. Related Work

There has been extensive literature on quantization for distributed estimation. When the parameter is random, Lam and Reibman[2] developed an iterative quantizer that maximizes the Bayesian Fisher Information. Gubner[3] and Zhang and Li[4] developed quantizers that minimize the MSE of a random parameter for certain restricted

classes of estimators. For a deterministic parameter, Luo[5] optimized the number of quantization bits for a uniform quantizer to minimize MSE. When the observation is a deterministic parameter in additive noise, Ribiero and Giannakis[6] showed that the Fisher Information is maximized when the threshold of a $1-$bit quantizer is equal to the parameter θ. In an earlier work[7], we had shown the optimality of *Score-Function Quantizers* in maximizing Fisher Information for a particular θ. Data compression for distributed estimation has also been considered from an information theoretic perspective [8], where nodes encode long sequences of observations.

Our paper is organized as follows. The basic system model and the minimax criterion are discussed in Section 2. The iterative algorithm and the convergence are explained in Section 3. Score-Function Quantizers and optimality are briefly discussed in Section 3.1. Numerical results and quantizer structure are presented in Section 4. Conclusions and future extensions are given in Section 5.

2. MINIMAX CRITERION

Consider the setup as shown in Figure 1. Let $\{P_\theta; \theta \in \Theta \subseteq \mathbb{R}\}$ represent the family of distributions for the random observation $X_i \in \mathcal{X}$ indexed by the deterministic scalar parameter θ, where P_θ denotes a probability measure on the space $\mathcal{X}$ with $\sigma-$field G. A *deterministic quantizer* can be defined as a $G-$measurable mapping $\gamma : \mathcal{X} \mapsto \{1, \cdots, D\}$. In this work, we shall restrict our study to the set of deterministic quantizers, represented by Γ. Since we assume nodes use identical quantizers, the variables $\gamma(X_i)$ are distributed i.i.d according to a probability mass function $\mathbf{q}_\theta$ specified by

$$
\begin{aligned}
\mathbf{q}_\theta &= \{q_\theta(1), \cdots, q_\theta(D)\}, \\
q_\theta(k) &= P_\theta(\gamma(X_i) = k).
\end{aligned}
$$

Let $\hat{\theta}$ be an unbiased estimate of θ based on the quantized observations $\gamma(X_i)$. If the family of probability mass functions $\{\mathbf{q}_\theta; \theta \in \Theta\}$ satisfies certain regularity conditions (see [9],pp. 169), then

$$
\mathrm{Var}_\theta[\hat{\theta}] \geq \frac{1}{nI_\theta}, \tag{1}
$$

$$
\text{where } I_\theta = \sum_{i=1}^{D} q_\theta(i) \left(\frac{d}{d\theta} \log q_\theta(i) \right)^2. \tag{2}
$$

The quantity $1/nI_\theta$ is the Cramer-Rao Bound (CRB) for the p.m.f $\mathbf{q}_\theta$. Furthermore, if $\hat{\theta}$ is the M-L estimate of θ based on $\mathbf{q}_\theta$, then $\hat{\theta} \sim \mathcal{N}(\theta, 1/nI_\theta)$, as $n \to \infty$ under more regularity conditions.

A quantizer that maximizes I_θ over all Γ is optimal for that θ when coupled with the M-L estimator. However, as mentioned earlier, it is not possible to design one quantizer that maximizes I_θ for all θ. Hence, we consider the minimax criterion:

$$
\gamma^* = \arg\min_\gamma \max_\theta \frac{J_\theta}{I_\theta}, \tag{3}
$$

$$
\text{where } J_\theta = \int_{\mathcal{X}} p_\theta(x) \left(\frac{d \log p_\theta(x)}{d\theta} \right)^2. \tag{4}
$$

Let the subscripts q and uq denote variables for the quantized and unquantized estimators respectively. Using (1),

$$
\gamma^* = \arg\min_\gamma \max_\theta \frac{\mathrm{CRB}_q}{\mathrm{CRB}_{uq}} \tag{5}
$$

Since we use a M-L estimator, (5) implies that for a fixed sample size, the quantizer γ^* would make the maximum ratio of the asymptotic error variances as close to 1 as possible.

An important interpretation of the criterion would be in terms of the Asymptotic Relative Efficiency (ARE)[9]. ARE between two maximum likelihood estimates can be defined as the ratio of sample sizes required by the two estimators to achieve the same asymptotic error variance. From (1), we see that for a given asymptotic variance, the criterion can be written as

$$
\gamma^* = \arg\min_\gamma \max_\theta \frac{n_q}{n_{uq}}
$$

For some source distributions p_θ, the Fisher Information of unquantized variable, J_θ can be independent of θ. For example, if $X_i = \theta + N_i$, where N_i has infinite support, it can be proven analytically that J_θ is a constant. Under such circumstances, the minimax criterion simplifies to:

$$
\gamma^* = \arg\min_\gamma \max_\theta n_q
$$

In other words, the minimax quantizer requires the least sample size amongst all deterministic quantizers.

3. ITERATIVE ALGORITHM

In this section, we propose the algorithm for iteratively obtaining the quantizer that achieves the minimax criterion. The basic idea behind the algorithm is this: We start with an arbitrary initial guess of the quantizer γ. We then iteratively improve the quantizer by finding the optimal partition for each $x \in \mathcal{X}$ given the others. The algorithm converges because the performance metric improves with every iteration.

Let $S_\theta(x), S_\theta(i)$ represent the score-function of the unquantized and quantized observations respectively. i.e.,

$$
S_\theta(x) = \frac{d \log p_\theta(x)}{d\theta}, \ x \in \mathcal{X} \tag{6}
$$

$$
S_\theta(i) = \frac{d \log q_\theta(i)}{d\theta}, \ i \in \{1, \cdots, D\} \tag{7}
$$

The score-function is a measure of sensitivity of the likelihood function and the definition is subject to conditions similar to those for CRB. Let the partitions in $\mathcal{X}$ in iteration j be represented by $T^{(j)}$. If the quantizer at the iteration j is represented by $\gamma^{(j)}$, then

$$
T^{(j)}(i) = \{x : \gamma^{(j)}(x) = i\}
$$

The formal statement of the algorithm is as follows:

1. *Initialization:* Divide $\mathcal{X}$ into D arbitrary non-overlapping partitions represented by $\{T^{(0)}(i)\}$. Evaluate the score-function and Fisher Information as

$$
S_\theta^{(0)}(i) = \frac{d}{d\theta} \log q_\theta^{(0)}(i) = \frac{d}{d\theta} \log \mathrm{Pr}\{x \in T^{(0)}(i)\}
$$

$$
I_\theta^{(0)} = \sum_{i=1}^{D} \frac{1}{q_\theta^{(0)}(i)} \left(\frac{dq_\theta^{(0)}(i)}{d\theta} \right)^2
$$

2. *Iteration $n + 1$:* Let $\theta^* = \arg\max_\theta \frac{J_\theta}{I_\theta^{(n)}}$.

   For every $x \in T^{(n)}(j)$, let

$$
k^* = \arg\min_k \{M_n^k(x), \ k = 1, .., D\} \tag{8}
$$

$$
M_n^k(x) = \begin{cases} \max_\theta \frac{p_\theta(x)}{J_\theta} (S_\theta(x) - S_\theta^{(n)}(k))^2, \ k \neq j \\ \frac{p_{\theta^*}(x)}{J_{\theta^*}} (S_{\theta^*}(x) - S_{\theta^*}^{(n)}(k))^2, \ k = j \end{cases} \tag{9}
$$

   Then $x \in T^{(n+1)}(k)$.

III - 653

3. *Termination:* Choose ϵ close to zero. At each iteration, evaluate $I_\theta^{(n)}$. If $\max_\theta |I_\theta^{(n)} - I_\theta^{(n-1)}| < \epsilon$, then the algorithm terminates. In other words, when the change in Fisher Information is very close to zero, we terminate the algorithm.

Theorem 1 *At the nth iteration, let $R^{(n)} = \max_\theta J_\theta / I_\theta^{(n)}$. Then $R^{(n+1)} \leq R^{(n)}, \forall n$. In the absence of local minima, the algorithm converges to $\min \max_\theta J_\theta / I_\theta$.*

Proof: Refer Appendix

Since, the algorithm iteratively assigns x to the D bins, it is computationally intensive. However, for a given parameter set, it is possible to run the algorithm offline and provide only the bin information to the nodes to perform the quantization. Although, the quantizer does not have a particular structure, depending on the source distribution, it may be possible to express the quantizer as thresholds on a function of x. This is discussed in detail in Section 4.3.

3.1. Score Function Quantizers

In this section, we discuss a class of quantizers known as *Score Function Quantizers* which can be used to maximize Fisher Information I_θ for a given θ. In [10], the class of Likelihood ratio quantizers was shown to optimize performance metrics in distributed detection like Chernoff Information (Bayesian detection) and Kullback-Leibler distance (Neyman-Pearson Detection). The equivalent of the likelihood ratio in parameter estimation is the score function, which has been defined in (6).

We define the *threshold set T* as the set of all vectors $t = (t_1, ..., t_{D-1}) \in \mathbb{R}^{D-1}$, satisfying $-\infty \leq t_1 \leq \cdots \leq t_{D-1} \leq \infty$. For any $t \in T$, the associated intervals $I_1, ..., I_D$ are defined by $L_1 = [0, t_1], L_2 = [t_1, t_2], \cdots, L_D = [t_{D-1}, \infty]$.

Definition 1 *: A quantizer $\gamma \in \Gamma$ is a monotone Score-function quantizer (SFQ) with threshold vector $t \in T$, if $\gamma(x) = d \iff S_\theta(x) \in L_d, \forall x$. We say that a quantizer is a SFQ is there exists a permutation mapping $\pi : \{1, \cdots, D\} \mapsto \{1, \cdots, D\}$ such that $\pi \circ \gamma$ is a monotone SFQ ($\circ$ is the composition operator).*

It has been shown in [7] that the Fisher Information I_θ for a given θ is maximized by a score-function quantizer. Furthermore, the optimal SFQ that maximizes I_θ can be obtained by using the Lloyd-Max iterative algorithm. It is to be noted that since the score-function depends on θ, this can be used to maximize the FI only at that value of θ and hence is not a practical quantizer. It however serves as a bound for the Fisher Information for any $D-$bit quantizer at that value of θ. The Fisher Information function obtained by evaluating the best SFQ at each θ in the parameter set is a useful benchmark for performance comparison. In the proceeding section, we shall compare the performance of the minimax quantizer to the maximum achievable Fisher Information obtained using SFQs.

4. NUMERICAL RESULTS

4.1. Parameter in AWGN

Here, the parameter is assumed to be an element of the bounded set $[\theta_{\min}, \theta_{\max}] \subset \mathbb{R}$. The source is distributed as $X_i \sim \mathcal{N}(0, \sigma^2)$. It is easy to see that the Fisher Information of unquantized X_i is a constant $J_\theta = \sigma^2$. Therefore, optimizing the minimax criterion corresponds to minimizing the sample size for a given asymptotic error variance.
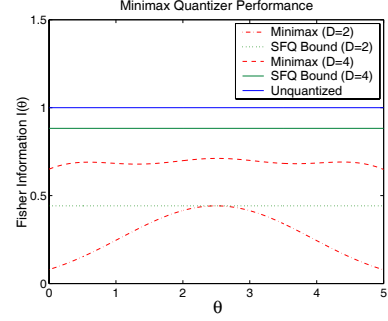

Fig. 2. Parameter in AWGN: $\theta \in [0, 5]$, $\sigma^2 = 1$

Figure 2 plots the Fisher information of the minimax quantizer versus θ and compares it to the bound obtained by evaluating the FI of the score-function quantizer at every θ. As can be seen from the figure, the minimax quantizer for $D = 2$ bits achieves the bound at $\theta = \frac{\theta_{\min} + \theta_{\max}}{2}$. As the number of quantization bits increase, the performance of the quantizer gets closer to the SFQ bound.

4.2. Gaussian Fading

In this example, the source is a faded version of the parameter, $X = H\theta$, $H \sim \mathcal{N}(0, \sigma^2)$. The parameter is again assumed to be an element of an interval $[\theta_{\min}, \theta_{\max}]$. In this case, the Fisher Information of the unquantized variable is given by $\frac{2}{\sigma^2 \theta^2}$. It is no longer independent of the parameter.
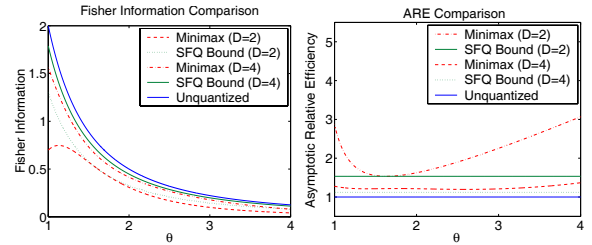

Fig. 3. Parameter in Gaussian Fading: $\theta \in [1, 4]$, $\sigma^2 = 1$

Figure 3 plots the Fisher Information and Asymptotic Relative Efficiency (ARE) of the minimax quantizer for the Gaussian faded distribution for $D = 2, 4$ bits. Similar to the previous example, the minimax quantizer for $D = 2$ achieves the SFQ bound at a particular value of θ. It is also interesting to note that the quantizer ARE at $\theta_{\min}$ and $\theta_{\max}$ are close to each other.

The observations that the minimax quantizer for $D = 2$ achieves the SFQ bound for a given value of θ suggest that those quantizers are score-function quantizers. We shall now look at the partitions in $\mathcal{X}$ space for the examples and draw some insights about the possible structure of the minimax quantizer. As will be seen in the following section, although the quantizers for $D = 4$ are not optimal for any particular θ, they also belong to the class of SFQs.

4.3. Partitions in Observation space

In this section, we illustrate the nature of the quantization bins in the $\mathcal{X}$ domain for the examples and provide some insights on the general nature of the partitions. Figure 4 shows the partitions in the $\mathcal{X}$ space for the two examples, when $D = 4$.
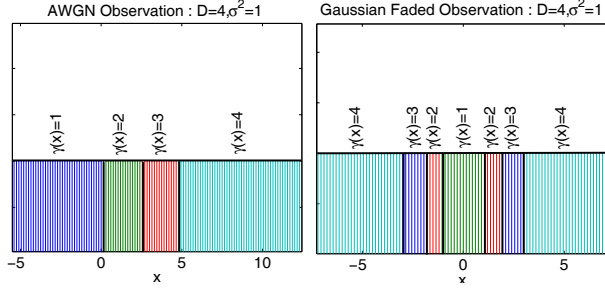
III - 654

Fig. 4. Partitions in $\mathcal{X}$ domain

As can be seen from the figures, the partitions for the AWGN example are the outcome of thresholds on the observation x. In other words, each bin corresponds to an interval $x \in [x_1, x_2]$. The partitions for the fading example are symmetric around zero and hence can be written in the form $x^2 \in [y_1, y_2]$. This may be attributed to the monotone likelihood behaviour of the score-function in these examples. The score-function for the AWGN and Gaussian Fading examples are given by:

$$S_\theta(x) = x - \theta \qquad \text{(AWGN)} \qquad (10)$$

$$S_\theta(x) = \frac{x^2}{\theta^3} - \frac{1}{\theta} \qquad \text{(Fading)} \qquad (11)$$

As can be seen from the above equations, for any value of θ, the score-function is a monotone increasing function of x and x^2 for the AWGN and Gaussian Fading examples respectively. Therefore, any score-function quantizer corresponds to a set of thresholds on x and x^2 respectively. For a general distribution, if the score-function is a monotone increasing function of the sufficient statistic, then all score-function quantizers can be expressed as thresholds on the sufficient statistic.

By using a technique similar to [10], it can be shown that any extreme point in the set of quantizer p.m.fs is achieved by a score-function quantizer. Therefore, if a minimax quantizer corresponds to an extreme point, it would be equivalently expressible as thresholds on the sufficient statistic. Although, we do not have an analytical proof, our numerical simulations suggest that this is true.

5. CONCLUSION AND FUTURE EXTENSIONS

In this work, we proposed a metric for the design of quantizers in deterministic parameter estimation. We developed an iterative algorithm to optimize the minimax metric and showed the convergence. For certain distributions like additive noise observations, the quantizer requires the least sample size amongst all deterministic quantizers. Although the algorithm is computationally intensive, it is possible to implement it offline and pass on the partition information alone to the nodes. Possible future extensions of this problem include extending the idea to vector parameters and generalizing the metric to include noisy channels. When the channel is not ideal, the quantizer design would depend on channel characteristics as well.

The application of this quantizer is not restricted to distributed estimation. For example, when estimating frequency of a signal, it may not be possible to handle a high resolution for each sample and so the minimax quantizer can be used before processing so that the final estimation performance is optimal.

6. APPENDIX

Proof of Theorem 1

Consider the minimax criterion:

$$
\begin{aligned}
\arg\min_\gamma \max_\theta \frac{J_\theta}{I_\theta} &= \arg\max_\gamma \min_\theta \frac{I_\theta}{J_\theta} \\
&= \arg\min_\gamma \max_\theta 1 - \frac{I_\theta}{J_\theta} \\
&= \arg\min_\gamma \max_\theta \frac{J_\theta - I_\theta}{J_\theta}
\end{aligned}
$$

It is easily shown that

$$\frac{J_\theta - I_\theta}{J_\theta} = \sum_i \int_{T_i} \frac{p_\theta(x)}{J_\theta} \left[S_\theta(x) - S_\theta(i) \right]^2 dx$$

where $T_i = \{x : \gamma(x) = i\}$. Using (8) and (9), we get

$$
\begin{aligned}
\max_\theta \frac{J_\theta - I_\theta^{(n)}}{J_\theta} &= \max_\theta \sum_i \int_{T_i^{(n)}} \frac{p_\theta(x)}{J_\theta} \left[S_\theta(x) - S_\theta^{(n)}(i) \right]^2 dx \\
&\leq \max_\theta \sum_i \int_{T_i^{(n)}} \frac{p_\theta(x)}{J_\theta} \left[S_\theta(x) - S_\theta^{(n-1)}(i) \right]^2 dx \\
&\leq \sum_i \int_{T_i^{(n-1)}} \frac{p_{\theta^*}(x)}{J_{\theta^*}} \left[S_{\theta^*}(x) - S_{\theta^*}^{(n-1)}(i) \right]^2 dx \\
&= \max_\theta \frac{J_\theta - I_\theta^{(n-1)}}{J_\theta}
\end{aligned}
$$

We know that $I_\theta \leq J_\theta$ for every value of θ. Therefore the metric is lower bounded by 0. Since, the algorithm reduces the metric at every iteration, in the absence of local minima, the inequalities are strict and hence the algorithm converges to the optimal quantizer in finite steps.

7. REFERENCES

[1] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *to appear in IEEE Trans. Signal Processing*, Feb. 2006.

[2] W. M. Lam and A. R. Reibman, "Design of Quantizers for Decentralized Estimation Systems," *IEEE Trasactions Comm.*, vol. 41, no. 11, pp. 1602–1605, Nov 1993.

[3] J. A. Gubner, "Distributed Estimation and Quantization," *IEEE Trans. Info. Thy.*, vol. 39, no. 4, pp. 1456–1459, July 1993.

[4] K. Zhang and X. R. Li, "Optimal Sensor Data Quantization for Best Linear Unbiased Estimation Fusion," in *Proc. 43rd IEEE Conf. on Decision and Control*, Dec 2004, vol. 3, pp. 2656–2661.

[5] Z. Q. Luo, "Universal Decentralized Estimation in a Bandwidth Constrained Sensor Network," *IEEE Trans. Info. Thy.*, vol. 51, no. 6, pp. 2210–2219, June 2005.

[6] A. Ribiero and G. B. Giannakis, "Non-Parametric Distributed Quantization-Estimation Using Wireless Sensor Networks," in *Proc. 2005 ICASSP Conf.*, Philadelphia, PA, March 2005.

[7] P. Venkitasubramaniam, G. Mergen, L. Tong, and A. Swami, "Quantization for Distributed Estimation in Large Scale Sensor Networks," in *2005 ICISIP*, Dec 2005, pp. 122–127.

[8] Te Sun Han and S. Amari, "Statistical Inference Under Multiterminal Data Compression," *IEEE Trans. Info. Theory*, vol. 44, no. 6, pp. 2300–2325, Oct. 1998.

[9] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1994.

[10] John N. Tsitsiklis, "Extremal Properties of Likelihood Ratio Quantizers," *IEEE Trans. Comm.*, vol. 41, no. 4, pp. 550–558, April 1993.