Robust Distributed Maximum Likelihood Estimation with Quantized Data

Xiaojing Shen, Member, IEEE, Pramod K. Varshney, Fellow, IEEE, Yunmin Zhu *†

August 22, 2012

Abstract

In this paper, distributed maximum likelihood estimation (MLE) with quantized data is considered under the assumption that the structure of the joint probability density function (pdf) is known, but it contains unknown deterministic parameters. The parameters may include different vector parameters corresponding to marginal pdfs and parameters that describe dependence of observations across sensors. We first discuss the regularity conditions which should be satisfied by the pdf and vector quantizers such that the MLE with quantized data is asymptotically efficient. Then, the relationship between the asymptotic variance of the MLE and the number of quantization bits is analytically derived. Since the optimal MLE scheme based on quantized data cannot be obtained when the joint pdf of observations is not completely known, a robust distributed MLE scheme is designed for a fixed number of quantization bits. Its asymptotic efficiency is proved under some regularity conditions and the asymptotic variance is derived so that the robustness can be analytically assessed. A numerical example with a bivariate Gaussian pdf is considered. Simulations show that the robustness of the proposed MLE scheme.

keywords: Maximum likelihood estimation; distributed estimation; quantized data; asymptotic variance; Fisher information matrix; wireless sensor networks

1 Introduction

Wireless sensor networks have attracted much attention and have become a fast-growing research area over the past years. Many advances have been made in distributed detection, estimation, tracking and control

*This work was supported in part by U.S. Air Force Office of Scientific Research (AFOSR) under Grants FA9550-10-1-0263 and FA9550-10-1-0458 and in part by the NNSF of China (# 61004138 and 61273074).
†Xiaojing Shen and Pramod K. Varshney (corresponding author, varshney@syr.edu) are with the Department of Electrical Engineering and Computer Science, Syracuse University, NY, 13244, USA. Yunmin Zhu is with Department of Mathematics, Sichuan University, Chengdu, Sichuan 610064, China. Xiaojing Shen (shenxj@scu.edu.cn) is on leave from Department of Mathematics, Sichuan University, Chengdu, Sichuan 610064, China.
Due to limited communication bandwidth and energy, the sensors usually quantize the measurements and send them to a fusion center which makes the final inference for decision, estimation, tracking or control tasks. The focus of this paper is on robust distributed MLE with quantized data. Distributed estimation and quantization problems have been considered in a number of studies. The parameters to be estimated are modeled as random and deterministic in different situations.

For random parameters, there exist various prior studies under the assumption of known joint pdf of parameters and sensor measurements (see [7, 8, 9, 10, 11, 12]). The work in [9] presented necessary conditions for optimal quantizers and optimal estimation. Optimum quantizers are difficult to obtain because coupled nonlinear equations need to be solved. The optimum estimator takes the form of a conditional mean that is usually hard to compute. To simplify computation, an efficient vector quantization algorithm based on the best linear unbiased estimator was proposed in [10, 11]. The convergence of the algorithm in [11] is guaranteed. The authors of [12] proposed a quantization approach where only a training sequence is available.

For deterministic parameters, several universal distributed estimation schemes have been proposed [13, 14, 15] in the presence of unknown, additive sensor noises that are bounded and identically distributed. These universal distributed estimation schemes have a low bandwidth requirement. The work in [16, 17, 18, 19, 20, 21] addressed various design and implementation issues under the assumption of a scalar parameter and using scalar quantizers. An approach consisting of multiple non-identical thresholds is employed in [16, 17]. The authors of [17] studied estimation of a scalar mean location parameter in the presence of zero-mean additive white Gaussian noise. Methods to design quantizers by minimizing the worst case performance in bounded parameter sets were proposed [18]. In [19], the performance limit that a distributed estimation scheme with identical quantizers can achieve was found as well as the set of optimal noise distribution functions and quantizers. In [20, 21], a quantization approach that adaptively adjusts the thresholds from sensor to sensor was proposed. The work of [22, 23, 24] proposed vector quantization design for distributed estimation under the assumption of additive observation noise model. The authors of [22] proposed a hyperplane-based heuristic approach, where the vector quantization problem can be converted to scalar quantization problems. In [24], a class of hyperplane-based vector quantizers was proposed which linearly convert the observation vector into a scalar by using a compression vector and then carry out scalar quantization.

When the structure of pdf is known, in previous works, the MLE with quantized data is extensively used to estimate the deterministic parameters. In this paper, robust distributed MLE with quantized data is considered. Our work differs from previous studies in several aspects. Prior results concentrate on the problem of how to design the quantization schemes for estimating a deterministic parameter where each sensor makes one noisy observation. The observations are usually assumed independent across sensors, and then discuss the relationship between MLE performance and the number of sensors. Here, we focus on the problem of how to design estimation schemes for the unknown parameter vector corresponding to the joint pdf of the observations where the number of sensors is fixed. These observations may be dependent across sensors. The unknown parameters may include different vector parameters corresponding to marginal pdfs and parameters that describe dependence of observations across sensors. Actually, the dependence between sensors is very
important in multisensor fusion systems, for example, see the recent work on distributed Neyman-Pearson
detection fusion, hypothesis testing using heterogeneous data, distributed location estimation with dependent
sensor observations [25, 26, 27]. We also derive the relationship between MLE performance, number of
quantization bits and number of observations. It is worth noting that our work neither requires the knowledge
of observation models, assumptions of Gaussianity and independence of noises across sensors nor requires
scalar quantizers and scalar estimated parameters.

In this paper, we first determine the regularity conditions which should be satisfied by the joint pdf
and quantizers such that the MLE with quantized data is asymptotically efficient. Then, the relationship
between the asymptotic variance of MLE and the number of quantization bits is analytically derived. We
shall prove that the asymptotic variance of MLE with quantized data is monotone decreasing with the number
of quantization bits and has a lower limit, which is equal to the asymptotic variance of MLE with raw
measurements. When the number of quantization bits is given, a robust distributed MLE scheme is designed
by employing $J$ different quantizers. Its asymptotic efficiency is proved under some regularity conditions
and the asymptotic variance is derived to be the inverse of a convex linear combination of Fisher information
matrices based on $J$ different quantizers. Thus, the robustness can be analytically verified. A numerical
example with a bivariate Gaussian pdf with an unknown parameter vector is considered. Simulations show
that the new MLE scheme is robust and much better than that based on the worst quantization scheme from
among the groups of quantizers. Another interesting phenomenon is that the asymptotic variance of the
estimates of parameters of marginal pdfs is almost independent of the dependence between the sensors.

The rest of the paper is organized as follows. Problem formulation is given in Section 2. In Section 3,
the performance analysis of MLE with quantized data is given. The robust MLE scheme is proposed and
the asymptotic results are derived. In Section 4, numerical examples are given and discussed. In Section 5,
conclusions are made.

## 2 Problem formulation

The basic $L$-sensor distributed estimation system is considered (see Figure 1). Each sensor has $k_i$-dimensional
observation population $Y_i$, $i = 1, \ldots, L$. Suppose that the joint observation population $\mathbf{Y} \triangleq (Y_1', \ldots, Y_L')'$
has a given family of joint pdf:

$$\{p(y_1, \ldots, y_L | \theta)\}_{\theta \in \Theta \subseteq \mathbb{R}^k}$$  \hspace{1cm} (1)

where $'$ denotes the transpose and $\theta$ is the unknown $k$-dimensional deterministic parameter vector which may
include marginal parameters and dependence parameters. Here, we do not assume independence across sen-
sors, knowledge of measurement models and Gaussianity of joint pdf. Let $N$ independently and identically
distributed (i.i.d.) sensor observation samples and joint observation samples be

$$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iN}), \; i = 1, \ldots, L;$$  \hspace{1cm} (2)
\[ \vec{Y} = (\vec{Y}_1', \ldots, \vec{Y}_L'). \]  

(3)

Suppose the sensors and the fusion center wish to jointly estimate the unknown parameter vector \( \theta \) based on

the spatially distributed observations. If there is sufficient communication bandwidth and power, the fusion center can obtain asymptotically efficient estimates with the complete observation samples based on MLE procedure under some regularity conditions on the joint pdf.

In many practical situations, however, to reduce the communication requirement from sensors to the fusion center due to limited communication bandwidth and power, the \( i \)-th sensor quantizes the observation vector to \( r_i \) bits (\( r_i \geq 1 \)) by \( r_i \) measurable indicator quantization functions:

\[ I_i^{(y_i)} : y_i \in \mathbb{R}^{k_i} \rightarrow \{0, 1\}; \ldots; I_i^{(r_i)}(y_i) : y_i \in \mathbb{R}^{k_i} \rightarrow \{0, 1\}, \]  

(4)

for \( i = 1, \ldots, L \). Here, each quantizer \( I_i^{(y_i)} \) may have one or multiple thresholds; and its 0/1 quantization region may be a continuous region or union of discontinuous regions. Moreover, we denote by

\[ I(y|r) \triangleq (I_1^{(y_1)'}, \ldots, I_L^{(y_L)'})' \in \mathbb{R}^r, \quad r = \sum_{i=1}^{L} r_i, \]  

(5)

where

\[ I_i^{(y_i)} \triangleq (I_{i1}^{(y_i)}, \ldots, I_{ir_i}^{(y_i)}), \quad i = 1, \ldots, L, \]  

(6)

and \( r \) is the total number of bits available to transmit observations from the sensors to the fusion center.
Once the \( r \)-bit binary quantized samples \( I_i(Y_{in}), n = 1, \ldots, N \) are generated at sensor \( i, i = 1, \ldots, L \), they are transmitted to the fusion center. The fusion center is then required to estimate the true parameter vector \( \theta^* \) with the quantized data. Usually, the MLE is used to estimate the parameter vector by maximizing the log likelihood function

\[
l(\theta|\vec{Y}) \triangleq \log P(I(\vec{Y}|r)|\theta) = \log \prod_{n=1}^{N} P(I_1(Y_{1n}), \ldots, I_L(Y_{Ln})|\theta) = N \sum_{n=1}^{N} \log \int_{\{I_1(y_1)=I_1(Y_{1n}),\ldots,I_L(y_L)=I_L(Y_{Ln})\}} p(y_1, \ldots, y_L|\theta) dy_1 \ldots dy_L. \tag{7}
\]

In the framework of such a distributed multisensor fusion system, one may question whether this method can estimate the parameters efficiently, since raw measurements \( y_i \) may be compressed to as low as 1-bit \( I_i(y_i), i = 1, \ldots, L \) so that a lot of information can be lost. Indeed, there are examples where MLE with quantized data cannot estimate parameters well, which will be given in Remark 3.3. On the other hand, there are also examples where MLE method with quantized data yields good estimation results (see, e.g., [25, 27]). Thus, in the present paper, we shall concentrate on analytically investigating the following basic problems:

- What conditions should the pdf \( p(y_1, \ldots, y_L|\theta) \) and the quantizers \( I(y|r) \) satisfy to guarantee that MLE with quantized data will be an asymptotically efficient estimator?

- What is the relationship between the asymptotic variance of MLE and the total number of quantization bits \( r \)?

- How should \( I(y|r) \) be designed to derive a robust and asymptotically efficient MLE for a given number of bits \( r \)?

3 Performance Analysis of Maximum Likelihood Estimation with Quantized Data

3.1 Asymptotic efficiency of maximum likelihood estimation with quantized data

By the definition of observation samples and quantizers, we define

\[
\vec{U} \triangleq (\vec{U}_1^t, \ldots, \vec{U}_L^t)',
\tag{8}
\]
\[ \bar{U}_i \triangleq (U_{i1}, \ldots, U_{iN})', \quad i = 1, \ldots, L, \]  
\[ U_{in} \triangleq (U_{in}^1, \ldots, U_{in}^r), \quad n = 1, \ldots, N, \]  
\[ U_{in}^1 \triangleq I_1(Y_{in}), \ldots, U_{in}^r \triangleq I_i(Y_{in}). \]

If we take \( \bar{U} \) as the joint quantized observation samples and denote the quantized observation population by \( U \triangleq I(Y|r) = (I_1(Y_1), \ldots, I_L(Y_L))' \), we know that \( U \) has a discrete/categorical distribution. Based on the pdf of \( Y \) and quantizers \( I(y|r) \), the probability mass function (pmf) of the quantized observation population \( U \) is

\[ f_U(u_1, u_2, \ldots, u_L|\theta) = P_{u_1, u_2, \ldots, u_L} \text{ for } U = (u_1, u_2, \ldots, u_L), \]  
where

\[ (u_1, u_2, \ldots, u_L) \in \mathcal{S}_u = \{(u_1, u_2, \ldots, u_L) \in \mathbb{R}^r : u_i \text{ is a } r_i\text{-dimensional binary row vector}, i = 1, \ldots, L, r = \sum_{i=1}^L r_i\}, \]

\[ P_{u_1, u_2, \ldots, u_L} = \int_{\Xi(u_1, u_2, \ldots, u_L)} p(y_1, y_2, \ldots, y_L|\theta) dy_1 dy_2 \ldots dy_L, \]

\[ \Xi(u_1, u_2, \ldots, u_L) = \{(y_1, y_2, \ldots, y_L) : I_1(y_1) = u_1, I_2(y_2) = u_2, \ldots, I_L(y_L) = u_L\}. \]

Note that \( f_U(u_1, u_2, \ldots, u_L|\theta) \) is determined by \( p(y_1, y_2, \ldots, y_L|\theta) \) and sensor quantizers \( I_1(y_1), \ldots, I_L(y_L) \).

Thus, the quantized observation population \( U \) has a family of joint pmf \( \{f_U(u_1, u_2, \ldots, u_L|\theta)\}_{\theta \in \Theta \subseteq \mathbb{R}^k} \) which yields the following log likelihood function of samples \( \bar{U} \) by (7), (12)-(15):

\[ l(\theta|\bar{U}) \triangleq \log \prod_{n=1}^N f_U(U_{1n}, U_{2n}, \ldots, U_{Ln}|\theta) \]

\[ = \sum_{n=1}^N \log \int_{\{I_1(y_1)=U_{1n}, \ldots, I_L(y_L)=U_{Ln}\}} p(y_1, \ldots, y_L|\theta) dy_1 \ldots dy_L \]

\[ = \sum_{n=1}^N \log \int_{\{I_1(y_1)=I_1(Y_{1n}), \ldots, I_L(y_L)=I_L(Y_{Ln})\}} p(y_1, \ldots, y_L|\theta) dy_1 \ldots dy_L \]

\[ = l(\theta|\bar{Y}). \]

Therefore, MLE with quantized data for the family \( \{p(y_1, \ldots, y_L|\theta)\}_{\theta \in \Theta \subseteq \mathbb{R}^k} \) is equivalent to MLE for the family \( \{f_U(u_1, u_2, \ldots, u_L|\theta)\}_{\theta \in \Theta \subseteq \mathbb{R}^k} \). Therefore, if MLE for the latter is an asymptotically efficient estimator, then MLE for the former is also asymptotically efficient. Thus, we only provide the conditions for
$p(y_1, y_2, \ldots, y_L | \theta)$ and sensor quantizers $I_1(y_1), \ldots, I_L(y_L)$ such that the classical regularity conditions for $f_U(u_1, u_2, \ldots, u_L | \theta)$ are satisfied, then asymptotic efficiency of MLE with quantized data can be guaranteed.

Moreover, by (16) and (13), the log likelihood function is also equal to

$$l(\theta | \vec{U}) = \sum_{n=1}^{N} \log f_U(U_{1n}, U_{2n}, \ldots, U_{Ln} | \theta)$$

$$= \sum_{j=1}^{2^r} n_j \log f_U(\vec{u}_j | \theta)$$

(19)

where $n_j = \# \{(U_{1n}, U_{2n}, \ldots, U_{Ln}) = \vec{u}_j \in S_u, n = 1, \ldots, N \}$, $\sum_{j=1}^{2^r} n_j = N$; $\# \{ \cdot \}$ is the cardinality of the set. The parameter vector $\theta$ is estimated by maximizing the log likelihood function (19) or equivalently solving the equation:

$$\frac{\partial}{\partial \theta} l(\theta | \vec{U}) = 0.$$  

(20)

Based on the classical asymptotic properties of MLE (see, e.g., [28, 29]), we have the following two lemmas.

**Lemma 3.1.** Assume that $p(y_1, y_2, \ldots, y_L | \theta)$ and sensor quantizers $I_1(y_1), \ldots, I_L(y_L)$ generate the quantized samples and $f_U(u_1, u_2, \ldots, u_L | \theta)$ satisfies the regularity conditions:

(C1) $I_i(Y_{i1}), \ldots, I_i(Y_{iN})$ are i.i.d., $i = 1, \ldots, L$;

(C2) The model parameter is identifiable:

$$\theta \neq \theta' \Leftrightarrow f_U(u_1, u_2, \ldots, u_L | \theta) \neq f_U(u_1, u_2, \ldots, u_L | \theta');$$  

(21)

(C3) $f_U(u_1, u_2, \ldots, u_L | \theta)$ is differentiable with respect to parameter vector $\theta$;

(C4) The parameter space contains an open set of which the true parameter $\theta^*$ is an interior point.

Then, Equation (20) has a solution denoted by $\hat{\theta}$ which is a consistent estimator of $\theta^*$, i.e., $\hat{\theta} \to \theta^*$, a.s.

The next three conditions and (C1)–(C4) are sufficient to guarantee asymptotic normality and efficiency of MLE.

**Lemma 3.2.** Assume that $p(y_1, y_2, \ldots, y_L | \theta)$ and sensor quantizers $I_1(y_1), \ldots, I_L(y_L)$ generate the quantized samples and $f_U(u_1, u_2, \ldots, u_L | \theta)$ satisfies the regularity conditions (C1)–(C4) and

(C5) For every $(u_1, \ldots, u_L) \in S$, $f_U(u_1, u_2, \ldots, u_L | \theta)$ is three times differentiable with respect to parameter vector $\theta$, and $\int f_U(u_1, u_2, \ldots, u_L | \theta) du_1 \ldots du_L$ can be differentiated three times under the integral sign;
(C6) For any \( \theta^* \) in the parameter space, there exists a positive number \( \epsilon \) and a function \( M(u_1, \ldots, u_L) \) such that
\[
\left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_t} \log f_U(u_1, u_2, \ldots, u_L|\theta) \right| \leq M(u_1, \ldots, u_L) \text{ for all } (u_1, \ldots, u_L) \in S, \quad \epsilon > 0 \quad (22)
\]
\[i, j, t = 1, \ldots, k,
\]
\[\theta \in \epsilon \text{ neighborhood of } \theta^*, \text{ with } E_{\theta^*}[M(U_1, \ldots, U_L)] < \infty;\]

(C7) The Fisher information matrix exists and is nonsingular:
\[
I(\theta^*, I(\cdot|\cdot)) = E_{\theta^*}\left[\frac{\partial}{\partial \theta} \log(f_U(U_1, U_2, \ldots, U_L|\theta)) \times \frac{\partial}{\partial \theta} \log(f_U(U_1, U_2, \ldots, U_L|\theta))'\right] > 0. \quad (23)
\]

Then,
\[
\sqrt{N}(\hat{\theta} - \theta^*) \rightarrow N(0, I^{-1}(\theta^*, I(\cdot|\cdot))) \quad (24)
\]
where \( I^{-1}(\theta^*, I(\cdot|\cdot)) \) is the Cramér-Rao lower bound for one quantized sample which depends on the form and number of bits of the quantizer \( I(y|\cdot) \). That is, \( \hat{\theta} \) is a consistent and asymptotically efficient estimator of \( \theta^* \).

Remark 3.3. Note that the identifiability of \( f_U(u_1, u_2, \ldots, u_L|\theta) \) implies that \( p(y_1, y_2, \ldots, y_L|\theta) \) is identifiable. If not, there are \( \theta \neq \theta' \) such that \( p(y_1, y_2, \ldots, y_L|\theta) = p(y_1, y_2, \ldots, y_L|\theta') \) so that, by (12) and (14), \( f_U(u_1, u_2, \ldots, u_L|\theta) = f_U(u_1, u_2, \ldots, u_L|\theta') \) which yields a contradiction. On the contrary, the identifiability \( p(y_1, y_2, \ldots, y_L|\theta) \) does not imply that \( f_U(u_1, u_2, \ldots, u_L|\theta) \) is identifiable. For example, let
\[
(Y_1, Y_2) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix}\right), \quad |\theta| < 1
\]

and
\[
I_1(y_1) = I_2(y_2) = I(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}
\]

We know that the normal pdf \( p(y_1, y_2|\theta) \) is identifiable. However, by (12), the pdf of \( U = (I_1(Y_1), I_2(Y_2)) \) is
\[
f_U(u_1, u_2|\theta) \equiv 1/4, \text{ for all } |\theta| < 1,
\]
which is not identifiable. In this case, we would not be able to distinguish between two parameters even with an infinite amount of data by MLE. For the above example, however, it is easy to make \( f_U(u_1, u_2|\theta) \) identifiable by choosing unsymmetric \( I_1(y_1) \) and \( I_2(y_2) \).
3.2 MLE with multiple bit quantized data and Cramér-Rao lower bound

In this subsection, the relationship between Cramér-Rao lower bound (i.e., asymptotic variance of MLE with quantized data) and the number of bits for quantization is analytically determined. We first derive a useful lemma.

**Lemma 3.4.** If $b = [b_1, b_2, \ldots, b_k]'$, $c = [c_1, c_2, \ldots, c_k]'$, $a = b + c$ and $g = d + e$, $d, e$ are positive constants, then

$$
\frac{aa'}{g} \preceq \frac{bb'}{d} + \frac{cc'}{e}
$$

(25)

where the equality is obtained at $\frac{b}{a} = \frac{c}{e}$.

**Proof.** Since $(cd - be)(cd - be)'$ is a non-negative definite matrix, we have

$$(cd - be)(cd - be)' \succeq 0$$

$\Leftrightarrow$ $cc'd^2 + bb'e^2 \succeq cb'de + bc'de$

$\Leftrightarrow$ $cc'd^2 + bb'e^2 + bb'de + cc'de \succeq cb'de + bc'de + bb'de + cc'de$

$\Leftrightarrow$ $(bb'e + cc'd)(d + e) \succeq (b + c)(b + c)'de$

$\Leftrightarrow$ $\frac{aa'}{d + e} \preceq \frac{bb'}{d} + \frac{cc'}{e}$.

**Theorem 3.5.** If we denote the Fisher information matrix with one raw measurement by $\mathcal{I}(\theta)$, then the relationship between the number of bits and the Cramér-Rao lower bound or the asymptotic variance of MLE with quantized data satisfy

$$
\mathcal{I}^{-1}(\theta^*, I(\cdot|\cdot)) \succeq \mathcal{I}^{-1}(\theta^*, I(\cdot|\cdot + 1)),
$$

(26)

and

$$
\lim_{r \to \infty} \mathcal{I}^{-1}(\theta^*, I(\cdot|\cdot)) = \mathcal{I}(\theta)^{-1},
$$

(27)

where $r \to \infty$ means $r_i \to \infty$, $i = 1, \ldots, L$. The inequality in terms of Fisher information Matrices can be obtained from (25) by taking the inverses and flipping the direction of inequality.
Proof. We denote the $r$-bit and $(r+1)$-bit quantized observation population by $U_r$ and $U_{r+1}$ respectively. The pmfs of $U_r$ and $U_{r+1}$ are defined in a similar manner by (12) respectively. Here, they are denoted by $f_{U_r}(u_1, \ldots, u_r|\theta)$ and $f_{U_{r+1}}(u_1, \ldots, u_r, u_{r+1}|\theta)$ respectively, where $u_i = 0/1, i = 1, \ldots, r$. Thus, we have

$$f_{U_r}(u_1, \ldots, u_r|\theta) = f_{U_{r+1}}(u_1, \ldots, u_r, 0|\theta) + f_{U_{r+1}}(u_1, \ldots, u_r, 1|\theta).$$

Moreover,

$$\frac{\partial}{\partial \theta_i} f_{U_r}(u_1, \ldots, u_r|\theta) = \frac{\partial}{\partial \theta_i} f_{U_{r+1}}(u_1, \ldots, u_r, 0|\theta) + \frac{\partial}{\partial \theta_i} f_{U_{r+1}}(u_1, \ldots, u_r, 1|\theta),$$

for $i = 1, \ldots, k$, where $k$ is the dimension of parameter vector $\theta$.

The Fisher information matrices for $U_r$ and $U_{r+1}$ can be derived respectively as follows.

$$\mathcal{I}(\theta^*, I(\cdot|r)) = E \left[ \left( \frac{\partial}{\partial \theta_1} f_{U_r}(U_r), \ldots, \frac{\partial}{\partial \theta_k} f_{U_r}(U_r) \right) \cdot \left( \frac{\partial}{\partial \theta_1} f_{U_r}(U_r), \ldots, \frac{\partial}{\partial \theta_k} f_{U_r}(U_r) \right)' \right]$$

$$= \sum_{u_i=0/1, i=1, \ldots, r} \frac{a_{u_1, \ldots, u_r} \cdot a'_{u_1, \ldots, u_r}}{g_{u_1, \ldots, u_r}}$$

where

$$a_{u_1, \ldots, u_r} \triangleq \left( \frac{\partial}{\partial \theta_1} f_{U_r}(u_1, \ldots, u_r|\theta), \ldots, \frac{\partial}{\partial \theta_k} f_{U_r}(u_1, \ldots, u_r|\theta) \right)'$$

$$g_{u_1, \ldots, u_r} \triangleq f_{U_r}(u_1, \ldots, u_r|\theta);$$

and

$$\mathcal{I}(\theta^*, I(\cdot|r+1)) = E \left[ \left( \frac{\partial}{\partial \theta_1} f_{U_{r+1}}(U_{r+1}), \ldots, \frac{\partial}{\partial \theta_k} f_{U_{r+1}}(U_{r+1}) \right) \cdot \left( \frac{\partial}{\partial \theta_1} f_{U_{r+1}}(U_{r+1}), \ldots, \frac{\partial}{\partial \theta_k} f_{U_{r+1}}(U_{r+1}) \right)' \right]$$

$$= \sum_{u_i=0/1, i=1, \ldots, r} \left( \frac{b_{u_1, \ldots, u_r, 0} \cdot b'_{u_1, \ldots, u_r, 0}}{d_{u_1, \ldots, u_r, 0}} + \frac{c_{u_1, \ldots, u_r, 1} \cdot c'_{u_1, \ldots, u_r, 1}}{e_{u_1, \ldots, u_r, 1}} \right),$$

where

$$b_{u_1, \ldots, u_r, 0} \triangleq \left( \frac{\partial}{\partial \theta_1} f_{U_{r+1}}(u_1, \ldots, u_r, 0|\theta), \ldots, \frac{\partial}{\partial \theta_k} f_{U_{r+1}}(u_1, \ldots, u_r, 0|\theta) \right)'$$
\[
\mathbf{e}_{u_1, \ldots, u_r, 1} \triangleq \left( \frac{\partial}{\partial \theta_1} f_{U_{r+1}}(u_1, \ldots, u_r, 1|\theta), \ldots, \frac{\partial}{\partial \theta_k} f_{U_{r+1}}(u_1, \ldots, u_r, 1|\theta) \right)^{\prime}, \quad (35) \\
\mathbf{d}_{u_1, \ldots, u_r, 0} \triangleq f_{U_{r+1}}(u_1, \ldots, u_r, 0|\theta), \quad \mathbf{e}_{u_1, \ldots, u_r, 1} \triangleq f_{U_{r+1}}(u_1, \ldots, u_r, 1|\theta). \quad (36)
\]

By Lemma 3.4 and Equations (28)-(36), we have
\[
\mathcal{I}(\theta^*, I(\cdot|r)) \leq \mathcal{I}(\theta^*, I(\cdot|r + 1)).
\]
Thus,
\[
\mathcal{I}^{-1}(\theta^*, I(\cdot|r)) \geq \mathcal{I}^{-1}(\theta^*, I(\cdot|r + 1)).
\]

Furthermore, quantization using \( r \) bits means that the measurement space is divided into \( 2^r \) regions. When \( r \) is large enough, for notational simplicity, we consider \( \mathbb{R}^2 \) to be divided into \( 2^r \) regions: \( \Delta_1 = (-\infty, -M_r] \times R, \Delta_2 = (-M_r, +\infty] \times R, \Delta_3 = (-M_r, M_r] \times (-\infty, -M_r], \Delta_4 = (-M_r, M_r] \times (M_r, +\infty], \) and the square \( (-M_r, M_r] \times (-M_r, M_r] \) is further divided into \( 2^{r-4} \) small regions \( \Delta_i, i = 5, 6, \ldots, 2^r \) where \( M_r \) is large enough. By the definition of \( \mathcal{I}(\theta^*, I(\cdot|r)) \) (23), the mean value theorem for integration and existence of \( \mathcal{I}^{-1}(\theta) \), we have
\[
\lim_{r \to \infty} \mathcal{I}(\theta^*, I(\cdot|r)) = \lim_{r \to \infty} \sum_{i=1}^{2^r} \int_{\Delta_i} \frac{1}{p(y|\theta)} dy \left( \frac{\partial}{\partial \theta_1} \int_{\Delta_i} p(y|\theta) dy, \ldots, \frac{\partial}{\partial \theta_k} \int_{\Delta_i} p(y|\theta) dy \right)^{\prime}
\]
\[
= \lim_{r \to \infty} \sum_{i=1}^{2^r} \int_{\Delta_i} \frac{1}{p(y|\theta)} \left( \int_{\Delta_i} \frac{\partial}{\partial \theta_1} p(y|\theta) dy, \ldots, \int_{\Delta_i} \frac{\partial}{\partial \theta_k} p(y|\theta) dy \right)^{\prime}
\]
\[
= \lim_{r \to \infty} \sum_{i=1}^{2^r} \int_{\Delta_i} p(y^i|\theta) \Delta_i \left( \frac{\partial}{\partial \theta_1} p(y^i|\theta) \Delta_i, \ldots, \frac{\partial}{\partial \theta_k} p(y^i|\theta) \Delta_i \right)^{\prime}
\]
\[
= \lim_{r \to \infty} \sum_{i=5}^{2^r} \int_{\Delta_i} p(y^i|\theta) \left( \frac{\partial}{\partial \theta_1} p(y^i|\theta), \ldots, \frac{\partial}{\partial \theta_k} p(y^i|\theta) \right)^{\prime}
\]
\[
= \mathcal{I}(\theta)
\]
For high dimensional case $\mathbb{R}^m, m > 2$, the proof is similar. Therefore, we have the theorem.

3.3 Robust maximum likelihood estimation with quantized data

When the number of bits $r_i$ for each sensor is given, a natural question is how should quantizers $I(\cdot | r)$ be designed such that the asymptotic variance $\mathcal{I}^{-1}(\theta^*, I(\cdot | r))$ of MLE with quantized data is as small as possible. The true parameter $\theta^*$, however, is not known, i.e., the pdf is not known. To the best of our knowledge, most of the existing work on optimal quantizer depends on the pdf or signal models. When both of them are not known, the optimal quantizer design cannot be derived in general. Actually, in the framework of this paper, we do not assume the Gaussian pdf and knowledge of measurement models. Thus, it is impossible to derive the optimal quantizer. However, we can choose multiple groups of quantizers to fend off against the risk of a poor quantizer. Therefore, we have following robust MLE design scheme.

For notation simplicity, let $r_i = 1, i = 1, \ldots, L$ and $r = L$

1. Choose $J$ groups of different quantizers (see Remark 3.7 for a discussion on the choice of $J$ groups of quantizers)

$$I^{(j)}(y) \triangleq (I^{(j)}_1(y_1), \ldots, I^{(j)}_L(y_L))^\prime \in \mathbb{R}^L, \ j = 1, \ldots, J,$$

where

$$I^{(j)}_i(y_i) : y_i \in \mathbb{R}^{k_i} \rightarrow \{0, 1\}, \ i = 1, \ldots, L.$$  \hfill (37)

2. Observe $N_j$ joint observation samples $\{(Y_{1n_j}, \ldots, Y_{Ln_j})\}_{n_j=1}^{N_j}$ which are quantized by the $j$-th group of quantizers for $j = 1, \ldots, J$. We denote by $N \triangleq \sum_{j=1}^J N_j$. The quantized observation samples $\{(I^{(j)}_1(Y_{1n_j}), \ldots, I^{(j)}_L(Y_{Ln_j}))\}_{n_j=1}^{N_j}$ are denoted by $\{(U^{(j)}_1, \ldots, U^{(j)}_{Ln_j})\}_{n_j=1}^{N_j}$. Moreover, we denote by $\bar{U}^{(j)} \triangleq \{(U^{(j)}_1, \ldots, U^{(j)}_{Ln_j})\}_{n_j=1}^{N_j}$. The corresponding quantized observation population based on $I^{(j)}(y)$ is denoted by $U^{(j)} \text{ whose pmf is}$

$$f^{(j)}_{U}(u_1, u_2, \ldots, u_L | \theta), \ j = 1, \ldots, J,$$

which can be similarly obtained by (12).

3. Estimate the parameter $\theta$ with the $J$ groups of quantized samples by maximizing the log likelihood function:

$$l(\theta|\bar{U}^{(1)}, \ldots, \bar{U}^{(J)}) = \log \prod_{j=1}^{J} \prod_{n_j=1}^{N_j} f^{(j)}_{U}(U^{(j)}_{1n_j}, U^{(j)}_{2n_j}, \ldots, U^{(j)}_{Ln_j} | \theta)$$

$$= \sum_{j=1}^{J} \sum_{n_j=1}^{N_j} \log f^{(j)}_{U}(U^{(j)}_{1n_j}, U^{(j)}_{2n_j}, \ldots, U^{(j)}_{Ln_j} | \theta)$$

12
\[
\sum_{j=1}^{J} \sum_{n_j=1}^{N_j} \log \int_{\left\{ I_1^{(j)}(y_1) = U_{n_j}^{(j)}, \ldots, I_L^{(j)}(y_L) = U_{n_j}^{(j)} \right\}} p(y_1, \ldots, y_L | \theta) dy_1 \ldots dy_L
\]

\[
\sum_{j=1}^{J} \int \{ I_j(y_j) \} = U_j^{(j)}, \ldots, I_j^{(L)}(y_L) = U_j^{(L)} \}
\]

\[
\Delta \sum_{j=1}^{J} l(\theta | U_j^{(j)}),
\]

where \(l(\theta | U_j^{(j)})\) is the log likelihood function of the \(j\)-th group of quantized data \(U_j^{(j)}\). Equivalently, we solve the equation:

\[
\frac{\partial}{\partial \theta} l(\theta | U_1^{(1)}, \ldots, U_J^{(J)}) = 0,
\]

whose solution is called robust MLE and denoted by \(\hat{\theta}_R\).

Obviously, the \(N\) quantized samples are not identically distributed due to the \(J\) different quantizers. One may question whether the new estimator based on the different quantizers is still asymptotically efficient? What is the asymptotic variance of the new estimator? Why is it robust compared to using one group of quantizers? Actually, these questions can be analytically answered by the following Theorem.

**Theorem 3.6.** There are \(J\) groups of different sensor quantizers \(I_j^{(j)}(y)\) defined by \([37], j = 1, \ldots, J\). Assume that \(p(y_1, y_2, \ldots, y_L | \theta)\) and quantizers \(I_j^{(j)}(y)\) generate the quantized samples and the quantized pmf \(\hat{I}_j^{(j)}(u_1, u_2, \ldots, u_L | \theta)\) defined by \([39]\) satisfies the regularity conditions (C1)--(C7). Then,

\[
\sqrt{N}(\hat{\theta}_R - \theta^*) \rightarrow N(0, \mathcal{I}^{-1}(\theta^*; I^{(1)}(\cdot), \ldots, I^{(J)}(\cdot))
\]

where \(N = \sum_{j=1}^{J} N_j, N_j \to \infty, j = 1, \ldots, J,\)

\[
\mathcal{I}^{-1}(\theta^*; I^{(1)}(\cdot), \ldots, I^{(J)}(\cdot)) \triangleq \left( \sum_{j=1}^{J} \frac{N_j}{N} \mathcal{I}(\theta^*; I^{(j)}(\cdot)) \right)^{-1}
\]

\[
= N \left( \sum_{j=1}^{J} N_j \mathcal{I}(\theta^*; I^{(j)}(\cdot)) \right)^{-1}.
\]

\[
\left( \sum_{j=1}^{J} N_j \mathcal{I}(\theta^*; I^{(j)}(\cdot)) \right)^{-1}
\]

is the Cramér-Rao lower bound for \(N\) quantized samples, where \(\mathcal{I}(\theta^*; I^{(j)}(\cdot))\) is the Fisher information matrix for one quantized sample of \(U^{(j)}\). That is, \(\hat{\theta}_R\) is a consistent and asymptotically efficient estimator of \(\theta^*\).

Moreover, without loss of generality, we assume that the first group of quantizers corresponds to the worst asymptotic variance

\[
\mathcal{I}^{-1}(\theta^*; I^{(1)}(\cdot)) \geq \mathcal{I}^{-1}(\theta^*; I^{(j)}(\cdot)), j = 2, \ldots, J,
\]

13
\[ I^{-1}(\theta^*; I^{(1)}(\cdot), \ldots, I^{(J)}(\cdot)) \leq I^{-1}(\hat{\theta}_R; I^{(1)}(\cdot)). \tag{48} \]

That is, \( \hat{\theta}_R \) is a robust estimator of \( \theta^* \).

**Proof.** From Lemma 3.1, the regularity of \( p(y_1, y_2, \ldots, y_L | \theta) \) and quantizers \( I^{(j)}(y) \) ensures that the quantized samples and the corresponding pmf \( f_U^{(j)}(u_1, u_2, \ldots, u_L | \theta) \) defined by (39) satisfy the regularity conditions (C1)–(C4), it is easy to prove that Equation (43) has a solution denoted by \( \hat{\theta}_R \) which is a consistent estimator of \( \theta^* \), i.e., \( \hat{\theta}_R \to \theta^* \), a.s. The proof is similar to that of Lemma 3.1 (see, e.g., [28, 29]). However, \( N \) quantized samples are independent but not identically distributed due to the use of different quantizers. Thus, to prove the asymptotic normality, we will use Lyapunov central limit theorem by checking the Lyapunov condition. Denote by (C5) and (51),

First, we expand the first derivative of the log likelihood function (40) around the true value \( \theta^* \),

\[
\frac{\partial l(\theta)}{\partial \theta} \bigg|_{\hat{\theta}_R} = \frac{\partial l(\theta)}{\partial \theta} \bigg|_{\theta^*} + \frac{\partial^2 l(\theta)}{\partial \theta^2} \bigg|_{\theta^*} (\hat{\theta}_R - \theta^*) + \cdots \tag{49}
\]

where higher-order terms can be ignored under regularity conditions (see e.g. [28]). Substitute \( \hat{\theta}_R \) for \( \theta \) and by (43), we have

\[
0 = \frac{\partial l(\theta)}{\partial \theta} \bigg|_{\hat{\theta}_R} = \frac{\partial l(\theta)}{\partial \theta} \bigg|_{\theta^*} + \frac{\partial^2 l(\theta)}{\partial \theta^2} \bigg|_{\theta^*} (\hat{\theta}_R - \theta^*).
\]

Thus,

\[
\sqrt{N}(\hat{\theta} - \theta^*) = - \left[ \frac{1}{N} \frac{\partial^2 l(\theta)}{\partial \theta^2} \bigg|_{\theta^*} \right]^{-1} \frac{1}{\sqrt{N}} \frac{\partial l(\theta)}{\partial \theta} \bigg|_{\theta^*} \bigg|_{\theta^*} \tag{50}
\]

Then, we check the Lyapunov condition. Denote by \( S_N^2 \triangleq \sum_{j=1}^{J} N_j I(\theta^*, I^{(j)}) \), \( (S_N^2)^2 \triangleq \sum_{j=1}^{J} N_j \tau^2 I(\theta^*, I^{(j)}) \tau \) for an arbitrary \( \tau \), and

\[
\mathcal{M}_j \triangleq E \left[ \tau^2 \frac{\partial}{\partial \theta} f_U^{(j)}(U_{1n_j}, U_{2n_j}, \ldots, U_{Ln_j} | \theta) \right] \tag{51}
\]

which exists, since condition (C3) is satisfied and \( U^{(j)} \) is a categorical distribution. Moreover, by condition (C5) and (51),

\[
\lim_{N \to \infty} \frac{1}{(S_N^2)^3} \sum_{j=1}^{J} \sum_{n_j=1}^{N_j} E \left[ \tau^2 \frac{\partial}{\partial \theta} \log f_U^{(j)}(U_{1n_j}, U_{2n_j}, \ldots, U_{Ln_j} | \theta) \right]
\]

14
\[ - E \left[ \tau' \frac{\partial}{\partial \theta} \log f_U^{(j)}(U_{1n_j}, U_{2n_j}, \ldots, U_{Ln_j} | \theta) \right]^3 \right] = \lim_{N \to \infty} \frac{1}{(S_N^*)^3} \sum_{j=1}^{J} \sum_{n_j=1}^{N_j} E \left[ \left| \tau' \frac{\partial}{\partial \theta} \log f_U^{(j)}(U_{1n_j}, U_{2n_j}, \ldots, U_{Ln_j} | \theta) - 0 \right|^3 \right] \]

\[ \leq \lim_{N \to \infty} \frac{1}{(S_N^*)^3} \sum_{j=1}^{J} \sum_{n_j=1}^{N_j} \mathcal{M}_j \]

\[ \leq \lim_{N \to \infty} \frac{1}{(S_N^*)^3} N \max \{ \mathcal{M}_1, \ldots, \mathcal{M}_J \} \]

\[ \leq \lim_{N \to \infty} \frac{N \max \{ \mathcal{M}_1, \ldots, \mathcal{M}_J \}}{(N \min \{ \tau' \mathcal{I}(\theta^*, I^{(1)}(\cdot)), \ldots, \tau' \mathcal{I}(\theta^*, I^{(J)}(\cdot)) \})^{\frac{3}{2}}} \]

\[ \leq \lim_{N \to \infty} \frac{1}{\sqrt{N}} \min \{ \tau' \mathcal{I}(\theta^*, I^{(1)}(\cdot)), \ldots, \tau' \mathcal{I}(\theta^*, I^{(J)}(\cdot)) \} = 0. \]

That is, the Lyapunov condition is satisfied. Thus, by Lyapunov central limit theorem (see, e.g., [29]), for all \( \tau \),

\[ \frac{1}{\sqrt{N}} \tau' \frac{\partial l(\theta | \vec{U}^{(1)}, \ldots, \vec{U}^{(J)})}{\partial \theta} \bigg|_{\theta^*} \to N\left(0, \frac{(S_N^*)^2}{N}\right). \]

Moreover, by the Cramér-Wold device (see, e.g., [29]), we have

\[ \frac{1}{\sqrt{N}} \frac{\partial l(\theta | \vec{U}^{(1)}, \ldots, \vec{U}^{(J)})}{\partial \theta} \bigg|_{\theta^*} \to N\left(0, \frac{S_N^2}{N}\right). \] (52)

By application of the weak law of large number, we have

\[ \frac{1}{N_j} \frac{\partial^2 l(\theta | \vec{U}^{(j)})}{\partial \theta^2} \bigg|_{\theta^*} \to -\mathcal{I}(\theta^*; I^{(j)}(\cdot)), j = 1, \ldots, J, \] (53)

where \( l(\theta | \vec{U}^{(j)}) \) is defined in (42).

By Slustkey Theorem and (53), we have

\[ \frac{1}{N} \frac{\partial^2 l(\theta | \vec{U}^{(1)}, \ldots, \vec{U}^{(J)})}{\partial \theta^2} \bigg|_{\theta^*} \to -\sum_{j=1}^{J} \frac{N_j}{N} \mathcal{I}(\theta^*; I^{(j)}(\cdot)) = -\frac{S_N^2}{N}. \] (54)

Based on (50), (52), (54) and Slustkey Theorem, we have

\[ \sqrt{N}(\hat{\theta}_R - \theta^*) \to N\left(0, \mathcal{I}^{-1}(\theta^*; I^{(1)}(\cdot), \ldots, I^{(J)}(\cdot))\right) \] (55)

where \( \mathcal{I}^{-1}(\theta^*; I^{(1)}(\cdot), \ldots, I^{(J)}(\cdot)) \) defined by (45).
Furthermore, by \((45)\) \((47)\) and the condition \((C7)\) \((23)\),

\[
\mathcal{I}^{-1}(\theta^*; I^{(1)}(\cdot), \ldots, I^{(J)}(\cdot)) = \left( \sum_{j=1}^{J} \frac{N_j}{N} \mathcal{I}(\theta^*; I^{(j)}(\cdot)) \right)^{-1} \leq \left( \sum_{j=1}^{J} \frac{N_j}{N} \mathcal{I}(\theta^*; I^{(1)}(\cdot)) \right)^{-1} = \mathcal{I}^{-1}(\theta^*; I^{(1)}(\cdot)).
\]

Therefore, we have the theorem.

**Remark 3.7.** In the first step of the robust MLE scheme, how do we choose \(J\) different quantizers?

1. If some prior information that may be due to previous experience or feedback information that can be obtained such as the fact that thresholds should be in a bounded region, the \(J\) groups of different quantizers can be uniformly chosen in that region. If there is no prior information available, then some real-time training samples are required. A basic heuristic criterion is that do not let all samples to be in the 0 region or the 1 region. For example, for \(i\)-th sensor, we can choose \(J\) different quantizers by observing \(N_{si}\) training samples. First, find the \(p\)-quantile and \((1-p)\)-quantile of \(N_{si}\) training samples, and then uniformly choose \(J\) different thresholds between \(p\)-quantile and \((1-p)\)-quantile (assume that \(p < 1 - p\)). If \(N_{si}\) is large, one can choose \(p\) close to \(\frac{1}{2}\), and vice versa.

2. How to determine \(J\)? Actually, there is a tradeoff between robustness, optimality and memory requirements. Theorem 3.6 shows that the asymptotic variance is the inverse of a convex linear combination of Fisher information matrices based on \(J\) different groups of quantizers. Thus, a larger \(J\) means more robustness compared with the worst case. However, it may also mean worse performance, since the best quantizer can be averaged by other worse ones. In addition, note that the above procedure requires the \(N_{si}\) training samples to be stored at the \(i\)-th sensor. Thus, \(N_{si}\) cannot be too large for sensors with limited memory so that the median of the samples may be randomly biased with respect to the median of the population. If \(N_{si}\) is large, one can choose one quantizer \((J = 1)\) with one threshold—the median number, it may be the “best” quantizer. However, if \(N_{si}\) is small which means a possibly large bias so that it may result in a very poor performance. Thus, it is necessary to choose a larger \(J\). Based on experiments that we have run, \(J\) ranging between 2 and 10 is likely to yield good results.

\[^1\text{Here, } p\text{-quantile } (0 \leq p \leq 1) \text{ is point such that the cumulative probability of samples will be less than } p. \text{ Thus, } \frac{1}{2}\text{-quantile is median.}\]
4 Numerical Examples

Let us consider a two-sensor Gaussian example: \((Y_1, Y_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)\) with joint pdf as follows

\[
p(y_1, y_2|\theta) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_1\sigma_2} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1-\mu_1}{\sigma_1}\right)^2 + 2\rho \left(\frac{y_1-\mu_1}{\sigma_1}\right)\left(\frac{y_2-\mu_2}{\sigma_2}\right) + \left(\frac{y_2-\mu_2}{\sigma_2}\right)^2\right]\right),
\]

where the parameter vector to be estimated \(\theta \triangleq [\theta_0, \theta_1, \theta_2] = [\rho, \mu_1, \mu_2]\). We will assume \(\mu_1 = 5\), \(\mu_2 = 7\), \(\sigma_1 = 6\), \(\sigma_2 = 8\) and obtain results for several different values of \(\rho = [0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]\). We will compare the robust MLE with MLE based on a single quantizer. Assume that the prior information is that thresholds are in \([0, 15]\). We uniformly choose the following four groups of different quantizers.

\[
I^{(1)}(y) = (I^{(1)}_1(y_1), I^{(1)}_2(y_2)) = (I[y_1 - 15], I[y_2 - 15]),
\]

\[
I^{(2)}(y) = (I^{(2)}_1(y_1), I^{(2)}_2(y_2)) = (I[y_1 - 10], I[y_2 - 10]),
\]

\[
I^{(3)}(y) = (I^{(3)}_1(y_1), I^{(3)}_2(y_2)) = (I[y_1 - 5], I[y_2 - 5]),
\]

\[
I^{(4)}(y) = (I^{(4)}_1(y_1), I^{(4)}_2(y_2)) = (I[y_1 - 0], I[y_2 - 0]),
\]

where

\[
I[x - c] = \begin{cases} 
1, & \text{if } x \geq c, \\
0, & \text{if } x < c. 
\end{cases}
\]

For robust MLE, we let \(N_1 = N_2 = N_3 = N_4 = N/4\) where \(N\) is number of samples for MLE with fixed quantizer \(I^{(j)}(y), j = 1, 2, 3, 4\) respectively.

In Figs 2-4, theoretical CRLBs with different number of measurements are plotted for parameters \(\theta = [0.5, 5, 7]\) respectively. The number of samples \(N = [800, 900, 1000, 1100, 1200]\) are considered respectively. The corresponding MSEs based on 2000 Monte Carlo runs are also plotted in Figs 5-7 respectively. In Fig. 8, theoretical CRLBs with different values of \(\rho\) and \(N = 1200\) are plotted for parameters \(\theta_0 = [0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]\), \(\theta_1 = 5\) and \(\theta_2 = 7\) respectively. The corresponding MSEs based on 2000 Monte Carlo runs are also plotted for \(\theta_0, \theta_1, \theta_2\) in Fig. 9 respectively.

From Figs 2-9, we have the following observations:

1. Both theoretical CRLBs and MSEs based on 2000 Monte Carlo runs for robust MLE are much smaller than those of the MLE based on the single quantizer that is worst in the group. This phenomenon is consistent with the analytical result in Theorem 3.6. Robust MLE is a kind of conservative estimate,
but it can avoid large errors in the worst case. The advantage of robustness (MSE of the worst MLE minus MSE of Robust MLE) is much larger than the loss due to conservative estimation to enhance robustness (MSE of Robust MLE minus MSE of the best one), especially in Figs 2, 3, 5 and 6.

2. From Figs. 8–9, an interesting phenomenon is observed in that the theoretical CRLBs of marginal parameters are almost equal for different correlation parameter values between the two sensors. It means that the estimation of marginal parameters is not related to the correlation of sensors for the numerical example. Thus, when one only concentrates on estimating a marginal parameter, the correlation between sensors need not be considered.

3. From Figs. 8–9, both theoretical CRLBs and MSEs based on 2000 Monte Carlo runs of $\theta_0$ are less than those of $\theta_1$ and $\theta_2$. The reason may be that the value of $\theta_0$ is less than those of $\theta_1$ and $\theta_2$.

5 Conclusion

In this paper, an approach for robust distributed MLE with quantized data has been proposed under the assumption that the structure of the joint pdf is known, but it contains unknown deterministic parameters. First, we discussed regularity conditions which should be satisfied by the pdf and quantizers such that the MLE with quantized data is asymptotically efficient. Then, we analytically derived that the asymptotic variance of MLE with quantized data is monotone decreasing with the number of quantization bits and has a lower limit, which is equal to the asymptotic variance of MLE with raw measurements. When the number of quantization bits is given, a robust distributed MLE scheme was designed by employing $J$ different quantizers. Its asymptotic efficiency was proved under some regularity conditions and the asymptotic variance was derived to be the inverse of a convex linear combination of Fisher information matrices based on $J$ different quantizers. Thus, the robustness was analytically shown. A numerical example with a joint Gaussian pdf was considered. Simulations show that the new MLE scheme is robust and much better than that based on the worst quantization scheme from among the groups of quantizers. Another interesting phenomenon is that the asymptotic variance of marginal parameters is almost not related to the correlation between two sensors.

The future work will involve the application of Robust MLE with quantized data to distributed location estimation, distributed detection fusion and hypothesis testing using heterogeneous data.

Acknowledgment

We would like to thank Kush Varshney for his valuable comments on earlier versions of this paper.
References

[1] P. K. Varshney, *Distributed Detection and Data Fusion*. New York: Springer-Verlag, 1997.

[2] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York: Wiley, 2001.

[3] C.-Y. Chong and S. P. Kumar, “Sensor networks: Evolution, opportunities, and challenges,” *Proceedings of the IEEE*, vol. 91, pp. 1247–1256, August 2003.

[4] B. Chen, L. Tong, and P. K. Varshney, “Channel-aware distributed detection in wireless sensor networks,” *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 16–26, 2006.

[5] Y. Zhu, J. Zhou, X. Shen, E. Song, and Y. Luo, *Networked Multisensor Decision and Estimation Fusion: Based on Advanced Mathematical Methods*. CRC Press, 2012.

[6] V. V. Veeravalli and P. K. Varshney, “Distributed inference in wireless sensor networks,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 370, pp. 100–117, January 2012.

[7] E. Ayanoglu, “On optimal quantization of noisy sources,” *IEEE Transactions on Information Theory*, vol. 36, pp. 1450–1452, November 1990.

[8] J. Gubner, “Distributed estimation and quantization,” *IEEE Transactions on Information Theory*, vol. 39, pp. 1456–1459, 1993.

[9] W. Lam and A. Reibman, “Design of quantizers for decentralized estimation systems,” *IEEE Transactions on Communications*, vol. 41, pp. 1602–1605, August 1993.

[10] K. Zhang and X. R. Li, “Optimal sensor data quantization for best linear unbiased estimation fusion,” in *The 43rd IEEE Conference on Decision and Control*, vol. 3, pp. 2656–2661, December 2004.

[11] X. Shen, Y. Zhu, and Z. You, “An efficient sensor quantization algorithm for decentralized estimation fusion,” *Automatica*, vol. 47, pp. 1053–1059, May 2011.

[12] V. Megalooikonomou and Y. Yesha, “Quantizer design for distributed estimation with communication constraints and unknown observation statistics,” *IEEE Transactions on Communications*, vol. 48, pp. 181–184, February 2000.

[13] Z.-Q. Luo, “Universal decentralized estimation in a bandwidth constrained sensor network,” *IEEE Transactions on Information Theory*, vol. 51, pp. 2210–2219, June 2005.

[14] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, and G. B. Giannakis, “Distributed compression-estimation using wireless sensor networks,” *IEEE Signal Processing Magazine*, vol. 23, pp. 27–41, July 2006.
[15] J.-J. Xiao, Z.-Q. Luo, and G. B. Giannakis, “Performance bounds for the rate-constrained universal decentralized estimators,” *IEEE Signal Processing Letters*, vol. 14, no. 1, pp. 47–50, 2007.

[16] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, “Sequential signal encoding from noisy measurements using quantizers with dynamic bias control,” *IEEE Transactions on Information Theory*, vol. 47, pp. 978–1002, March 2001.

[17] A. Ribeiro and G. B. Giannakis, “Bandwidth-constrained distributed estimation for wireless sensor networks–Part I: Gaussian case,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 1131–1143, March 2005.

[18] P. Venkitasubramaniam, L. Tong, and A. Swami, “Quantization for maximin ARE in distributed estimation,” *IEEE Transactions on Signal Processing*, vol. 55, pp. 3596–3605, July 2007.

[19] H. Chen and P. K. Varshney, “Performance limit for distributed estimation systems with identical one-bit quantizers,” *IEEE Transactions on Signal Processing*, vol. 58, pp. 466–471, January 2010.

[20] H. B. Li and J. Fang, “Distributed adaptive quantization and estimation for wireless sensor networks,” *IEEE Signal Processing Letters*, vol. 14, pp. 669–672, 2007.

[21] J. Fang and H. Li, “Distributed adaptive quantization for wireless sensor networks: From delta modulation to maximum likelihood,” *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5246–5257, 2008.

[22] A. Ribeiro and G. B. Giannakis, “Bandwidth-constrained distributed estimation for wireless sensor networks–Part II: Unknown probability density function,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 2784–2796, July 2006.

[23] T. Zhao and A. Nehorai, “Information-driven distributed maximum likelihood estimation based on gauss-newton method in wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 55, pp. 4669–4682, September 2007.

[24] J. Fang and H. B. Li, “Hyperplane-based vector quantization for distributed estimation in wireless sensor networks,” *IEEE Transactions on Information Theory*, vol. 55, pp. 5682–5699, 2009.

[25] A. Sundaresan, P. K. Varshney, and N. S. V. Rao, “Copula-based fusion of correlated decisions,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, pp. 454–471, January 2011.

[26] S. G. Iyengar, P. K. Varshney, and T. Damarla, “A parametric copula-based framework for hypothesis testing using heterogeneous data,” *IEEE Transactions on Signal Processing*, vol. 59, pp. 2308–2319, May 2011.

[27] A. Sundaresan and P. K. Varshney, “Location estimation of a random signal source based on correlated sensor observations,” *IEEE Transactions on Signal Processing*, vol. 59, pp. 787–799, February 2011.
Figure 2: Theoretical CRLBs of MLE of $\theta_0$ while using raw measurements, different quantizers and the robust MLE of $\theta_0$ for different number of measurements
Figure 3: Theoretical CRLBs of MLE of $\theta_1$ while using raw measurements, different quantizers and the robust MLE of $\theta_1$ for different number of measurements.

Figure 4: Theoretical CRLBs of MLE of $\theta_2$ while using raw measurements, different quantizers and the robust MLE of $\theta_2$ for different number of measurements.
Figure 5: MSEs of MLE of $\theta_0$ based on 2000 Monte Carlo runs while using raw measurements, different quantizers and the robust MLE of $\theta_0$ for different number of measurements.

Figure 6: MSEs of MLE of $\theta_1$ based on 2000 Monte Carlo runs while using raw measurements, different quantizers and the robust MLE of $\theta_1$ for different number of measurements.
Figure 7: MSEs of MLE of $\theta_2$ based on 2000 Monte Carlo runs while using raw measurements, different quantizers and the robust MLE of $\theta_2$ for different number of measurements.

Figure 8: Theoretical CRLBs of robust MLE of $\theta_0$, $\theta_1$ and $\theta_2$ for different correlation parameter $\rho$s.
Figure 9: MSEs of robust MLE of $\theta_0$, $\theta_1$ and $\theta_2$ based on 2000 Monte Carlo runs for different correlation parameter $\rho$s.