

Universal Quantile Estimation with Feedback in the Communication-Constrained Setting

Ram Rajagopal¹, Martin J. Wainwright^{1,2} and Pravin Varaiya¹

¹ Department of Electrical Engineering and Computer Science

² Department of Statistics

University of California, Berkeley

{ramr,wainwrig,varaiya}@eecs.berkeley.edu

Abstract—We consider the following problem of decentralized statistical inference: given i.i.d. samples from an unknown distribution, estimate an arbitrary quantile subject to limits on the number of bits exchanged. We analyze a standard fusion-based architecture, in which each of m sensors transmits a single bit to the fusion center, which in turn is permitted to send some number k bits of feedback. Supposing that each of m sensors receives n observations, the mean-squared error of the optimal centralized protocol decays as $O(\frac{1}{nm})$. First, we describe a decentralized protocol based on $k = m$ bits of feedback that is strongly consistent, and achieves the same asymptotic MSE as the centralized optimum. Second, we describe and analyze a decentralized protocol based on only a single bit ($k = 1$) of feedback. For step sizes independent of m , it achieves an asymptotic MSE of order $O(\frac{1}{n\sqrt{m}})$, whereas for step sizes decaying as $m^{-1/2}$, it achieves the same order of MSE—namely, $O(\frac{1}{nm})$ —as the centralized optimum. We discuss the tradeoffs between these different protocols.

I. INTRODUCTION

Whereas classical statistical inference is performed in a centralized manner, many modern scientific and engineering applications (e.g., sensor networks) are inherently *decentralized*: data are distributed throughout a network, and cannot be aggregated due to various forms of communication constraints. In statistical terms, such communication constraints imply that the individual sensors cannot transmit the raw data; rather, they must compress or quantize the data (e.g., from a continuous-valued observation to a binary random variable), and can transmit only this compressed representation back to the fusion center. There is a rich literature in both information theory and statistical signal processing on problems of decentralized statistical inference. A number of researchers, dating back to the seminal paper [9], have studied the problem of hypothesis testing under communication-constraints; see [10] for an overview. A parallel line of work deals with problem of decentralized estimation. Work in signal processing typically formulates it as a quantizer design problem and considers finite sample behavior [1], [4]; in contrast, the information-theoretic approach is asymptotic in nature, based on rate-distortion theory [11]. In much of the literature on decentralized statistical inference, it is assumed that the underlying distributions are known with a specified parametric form (e.g., Gaussian). More recent work has addressed non-parametric and data-driven formulations of these problems, in which the decision-maker

is simply provided samples from the unknown distribution [7], [6].

This paper addresses the problem of estimating an arbitrary quantile of an unknown distribution, for which no unbiased single sample estimator exists. We consider a standard fusion-based architecture, in which each of m sensors is permitted to transmit a single bit to the fusion center, which in turn is permitted to send some number k bits of feedback. For a decentralized protocol with $k = m$ bits of feedback, we prove that the algorithm achieves the order-optimal rate of the best centralized method (i.e., one with access to the full collection of raw data). We also consider a protocol that permits only a single bit of feedback, and establish that it achieves the same rate. This single-bit protocol is advantageous in that, with for a fixed target mean-squared error of the quantile estimate, it yields longer sensor lifetimes than either the centralized or full feedback protocols.

The remainder of the paper is organized as follows. We begin in Section II with background on quantile estimation, and optimal rates in the centralized setting. In Section III, we begin by describing two algorithms for solving the decentralized version. We then state our main theoretical results, and illustrate the agreement between theory and simulation. Section IV contains the proofs of our main results, and we conclude in Section V with a discussion.

II. BACKGROUND

We begin by introducing some background on quantile estimation; see Serfling [8] for further details.

Basic problem: Given a real-valued random variable X , let $F(x) := \mathbb{P}[X \leq x]$ be its distribution function (necessarily right-continuous). For any $0 < \alpha < 1$, the α^{th} -quantile of F is defined as $F^{-1}(\alpha) = \theta(\alpha) := \inf \{x \in \mathbb{R} \mid F(x) \geq \alpha\}$. Moreover, if F is continuous at α , then we have $\alpha = F(\theta(\alpha))$. As a particular example, for $\alpha = 0.5$, the associated quantile is simply the median.

Now suppose that for a fixed level $\alpha^* \in (0, 1)$, we wish to estimate the quantile $\theta^* = \theta(\alpha^*)$. We work in the non-parametric setting, in which the class of possible distributions lacks a particular parameterized form, but rather satisfies only mild conditions. In particular, we assume that the distribution function F is differentiable, so that X has the

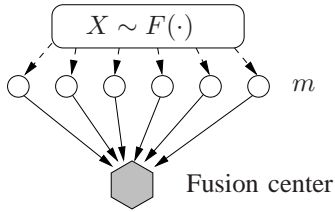


Fig. 1. Sensor network for quantile estimation with m sensors. Each sensor is permitted to transmit a 1-bit message to the fusion center; in turn, the fusion center is permitted to broadcast k bits of feedback.

density function $p_X(x) = F'(x)$ (w.r.t Lebesgue measure), and moreover that $p_X(x) > 0$ for all $x \in \mathbb{R}$.

Estimation rates: In this setting, a standard estimator for θ^* is the *sample quantile* $\xi_N(\alpha^*) := F_N^{-1}(\alpha^*)$ where F_N denotes the empirical distribution function based on i.i.d. samples (X_1, \dots, X_N) . Under the conditions given above, it can be shown [8] that $\xi_N(\alpha^*)$ is strongly consistent for θ^* (i.e., $\xi_N \xrightarrow{a.s.} \theta^*$), and moreover that asymptotic normality holds

$$\sqrt{N}(\xi_N - \theta^*) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha^*(1 - \alpha^*)}{p_X^2(\theta^*)}\right), \quad (1)$$

so that the asymptotic MSE decreases as $O(1/N)$, where N is the total number of samples.

III. DISTRIBUTED QUANTILE ESTIMATION

We consider the standard network architecture illustrated in Figure 1. There are m sensors, each of which has a dedicated two-way link to a fusion center. We assume that each sensor $i \in \{1, \dots, m\}$ collects independent samples $X(i)$ of the random variable $X \in \mathbb{R}$ with distribution function $F(\theta) := \mathbb{P}[X \leq \theta]$. We consider a sequential version of the quantile estimation problem, in which sensor i receives measurements $X_n(i)$ at time steps $n = 0, 1, 2, \dots$, and the fusion center forms an estimate θ_n of the quantile. The key condition—giving rise to the decentralized nature of the problem—is that communication between each sensor and the central processor is constrained, so that the sensor cannot simply relay its measurement $X(i)$ to the central location, but rather must perform local computation, and then transmit a summary statistic to the fusion center. More concretely, we impose the following restrictions on the protocol. First, at each time step $n = 0, 1, 2, \dots$, each sensor $i = 1, \dots, m$ can transmit a single bit $Y_n(i)$ to the fusion center. Second, the fusion center can broadcast k bits back to the sensor nodes at each time step. We analyze two distinct protocols, depending on whether $k = m$ or $k = 1$.

A. Protocol specification

For each protocol, all sensors are initialized with some fixed θ_0 . The algorithms are specified in terms of a constant $K > 0$ and step sizes $\epsilon_n > 0$ that satisfy the conditions

$$\sum_{n=0}^{\infty} \epsilon_n = \infty \quad \text{and} \quad \sum_{n=0}^{\infty} \epsilon_n^2 < \infty. \quad (2)$$

The first condition ensures infinite travel (i.e., that the sequence θ_n can reach θ^* from any starting condition), whereas the second condition (which implies that $\epsilon_n \rightarrow 0$) is required for variance reduction. A standard choice satisfying these conditions—and the one that we assume herein—is $\epsilon_n = 1/n$.

With this set-up, the m -bit scheme consists of the following steps:

Decentralized quantile estimation with m -bit feedback:

- (a) *Local decision:* each sensor computes the binary decision

$$Y_{n+1}(i) \equiv Y_{n+1}(i; \theta_n) := \mathbb{I}(X_{n+1}(i) \leq \theta_n), \quad (3)$$

and transmits it to the fusion center.

- (b) *Parameter update:* the fusion center updates its current estimate θ_{n+1} of the quantile parameter as follows:

$$\theta_{n+1} = \theta_n + \epsilon_n K \left(\alpha^* - \frac{\sum_{i=1}^m Y_{n+1}(i)}{m} \right) \quad (4)$$

- (c) *Feedback:* the fusion broadcasts the m received bits $\{Y_{n+1}(1), \dots, Y_{n+1}(m)\}$ back to the sensors.

After the feedback step, each sensor has knowledge of $\{Y_n(1), \dots, Y_n(m)\}$, which (in conjunction with knowledge of m , α^* and ϵ_n) allow it to compute the updated parameter θ_{n+1} . Knowledge of this parameter suffices to compute the local decision (3).

The 1-bit feedback scheme is similar, except that it requires broadcasting only a single bit (Z_{n+1}), and involves an extra step size parameter K_m , which is specified in the statement of Theorem 2.

Decentralized quantile estimation with 1-bit feedback:

- (a) *Local decision:* each sensor computes the binary decision

$$Y_{n+1}(i) = \mathbb{I}(X_{n+1}(i) \leq \theta_n) \quad (5)$$

and transmits it to the fusion center.

- (b) *Aggregate decision and parameter update:* The fusion center computes the aggregate decision

$$Z_{n+1} = \mathbb{I}\left(\frac{\sum_{i=1}^m Y_{n+1}(i)}{m} \leq \alpha^*\right), \quad (6)$$

and uses it update the parameter according to

$$\theta_{n+1} = \theta_n + \epsilon_n K_m (Z_{n+1} - \beta) \quad (7)$$

where the constant β is chosen as

$$\beta = \sum_{i=0}^{\lfloor m\alpha^* \rfloor} \binom{m}{i} (\alpha^*)^i (1 - \alpha^*)^{m-i}. \quad (8)$$

- (c) *Feedback:* The fusion center broadcasts the aggregate decision Z_{n+1} back to the sensor nodes (one bit of feedback).

After the feedback step of this protocol, each sensor has knowledge of the aggregate decision Z_{n+1} , which (in con-

junction with ϵ_n and the constant β) allow it to compute the updated parameter θ_{n+1} . Knowledge of this parameter suffices to compute the local decision (5).

B. Convergence results

We now state our main results on the convergence behavior of these two distributed protocols. In all cases, we assume the step size choice $\epsilon_n = 1/n$.

Theorem 1 (*m-bit feedback*). *For any $\alpha^* \in (0, 1)$, consider a random sequence $\{\theta_n\}$ generated by the m-bit feedback protocol. Then*

(a) *For all initial conditions θ_0 , the sequence θ_n converges almost surely to θ^* such that $\mathbb{P}(X \leq \theta^*) = \alpha^*$.*

(b) *Moreover, if the constant K is chosen to satisfy $p_X(\theta^*) K > \frac{1}{2}$, then*

$$\sqrt{n}(\theta_n - \theta^*) \xrightarrow{d} \mathcal{N}\left(0, \frac{K^2 \alpha^* (1 - \alpha^*)}{[2Kp_X(\theta^*) - 1]} \frac{1}{m}\right), \quad (9)$$

so that the asymptotic MSE is $O(\frac{1}{mn})$.

Remarks: After n steps of this decentralized protocol, a total of $N = nm$ observations have been made, so that our discussion in Section II dictates (see equation (1)) that the optimal asymptotic MSE is $O(\frac{1}{nm})$. Interestingly, then, the m -bit feedback decentralized protocol is order-optimal with respect to the centralized gold standard.

Before stating the analogous result for the 1-bit feedback protocol, we begin by introducing some useful notation. First, we define for any fixed $\theta \in \mathbb{R}$ the random variable $\bar{Y}(\theta) := \frac{1}{m} \sum_{i=1}^m Y(i; \theta) = \frac{1}{m} \sum_{i=1}^m \mathbb{I}(X(i) \leq \theta)$. Note that for each fixed θ , the distribution of $\bar{Y}(\theta)$ is binomial with parameters m and $F(\theta)$. It is convenient to define the function

$$G_m(r, y) := \sum_{i=0}^{\lfloor my \rfloor} \binom{m}{i} r^i (1-r)^{m-i}, \quad (10)$$

with domain $(r, y) \in [0, 1] \times [0, 1]$. With this notation, we have

$$\mathbb{P}(\bar{Y}(\theta) \leq y) = G_m(F(\theta), y).$$

Theorem 2 (*1-bit feedback*). *For any $\alpha^* \in (0, 1)$, consider a random sequence $\{\theta_n\}$ generated by the 1-bit feedback protocol. Then we have:*

(a) *The sequence $\theta_n \xrightarrow{a.s.} \theta^*$ such that $\mathbb{P}(X \leq \theta^*) = \alpha^*$.*

(b) *Suppose that the step size K_m is chosen such that $K_m > \frac{\sqrt{2\pi\alpha^*(1-\alpha^*)}}{2p_X(\theta^*)\sqrt{m}}$, or equivalently such that*

$$\gamma_m(\theta^*) := K_m \left| \frac{\partial G_m}{\partial r}(r; \alpha^*) \Big|_{r=\alpha^*} \right| p_X(\theta^*) > \frac{1}{2}, \quad (11)$$

then

$$\sqrt{n}(\theta_n - \theta^*) \xrightarrow{d} \mathcal{N}\left(0, \frac{K_m^2 G_m(\alpha^*, \theta^*) [1 - G_m(\alpha^*, \theta^*)]}{2\gamma_m(\theta^*) - 1}\right) \quad (12)$$

(c) *If we choose a constant step size $K_m = K$, then as $n \rightarrow \infty$, the asymptotic variance behaves as*

$$\left[\frac{K^2 \sqrt{2\pi\alpha^*(1-\alpha^*)}}{8Kp_X(\theta^*)\sqrt{m} - 4\sqrt{2\pi\alpha^*(1-\alpha^*)}} \right], \quad (13)$$

so that the asymptotic MSE is $O(\frac{1}{n\sqrt{m}})$.

(d) *If we choose a decaying step size $K_m = \frac{K}{\sqrt{m}}$, then*

$$\frac{1}{m} \left[\frac{K^2 \sqrt{2\pi\alpha^*(1-\alpha^*)}}{8Kp_X(\theta^*) - 4\sqrt{2\pi\alpha^*(1-\alpha^*)}} \right], \quad (14)$$

so that the asymptotic MSE is $O(\frac{1}{nm})$.

C. Comparative Analysis

We can compute the relative performance of each proposed algorithm. A simple calculation shows that setting $K = 1/p_X(\theta^*)$ in algorithm m -bf yields the same asymptotic variance as the optimal centralized scheme. In practice, however, the value $p_X(\theta^*)$ is typically not known. For the algorithm 1-bf, making the substitution $\bar{K} = K/\sqrt{2\pi\alpha^*(1-\alpha^*)}$ yields the asymptotic variance

$$\frac{2\pi}{4} \frac{\bar{K}^2 \alpha^* (1 - \alpha^*)}{[2\bar{K}p_X(\theta^*) - 1]} \frac{1}{m}. \quad (15)$$

Since the stability criterion is the same as that for m -bf, the optimal choice is $\bar{K} = 1/p_X(\theta^*)$. So although the rate is the same, the prefactor for the 1-bf algorithm is 57% higher than the optimal centralized scheme. The main advantage of using algorithm 1-bf is that despite the performance loss, the network lifetime scales as $O(m)$ compared to $O(1)$ for m -bf.

D. Simulation example

We now provide some simulation results in order to illustrate the two decentralized protocols, and the agreement between theory and practice. In particular, we consider the quantile estimation problem when the underlying distribution (which, of course, is unknown to the algorithm) is uniform on $[0, 1]$ random. In this case, we have $p_X(x) = 1$ uniformly for all $x \in [0, 1]$, so that taking the constant $K = 1$ ensures that the stability conditions in both Theorem 1 and 2 are satisfied. We simulate the behavior of both algorithms for $\alpha^* = 0.3$ over a range of choices for the network size m . Figure 2(a) illustrates several sample paths of m -bit feedback protocol, showing the convergence to the correct θ^* .

For comparison to our theory, we measure the empirical variance by first computing the average $\hat{\epsilon}_n^l = \sqrt{n}(\theta_n - \theta^*)$ over $L = 20$ runs. The normalization by \sqrt{n} is used to isolate the effect of m . We estimate the variance by running algorithm for $n = 2000$ steps, and computing the empirical variance of $\hat{\epsilon}_n^l$ for time steps $n = 1800$ through to $n = 2000$. Figure 2(b) shows these empirically computed variances, and a comparison to the theoretical predictions of Theorems 1 and 2 for constant step size; note the excellent agreement between theory and practice. Panel (c) shows the comparison between the m -bf algorithm, and the 1-bf algorithm with decaying $1/\sqrt{m}$ step size. Here the asymptotic MSE of both algorithms decays like $1/m$ for m up to roughly 500; after this point, our fixed choice of n is insufficient to reveal the asymptotic behavior.

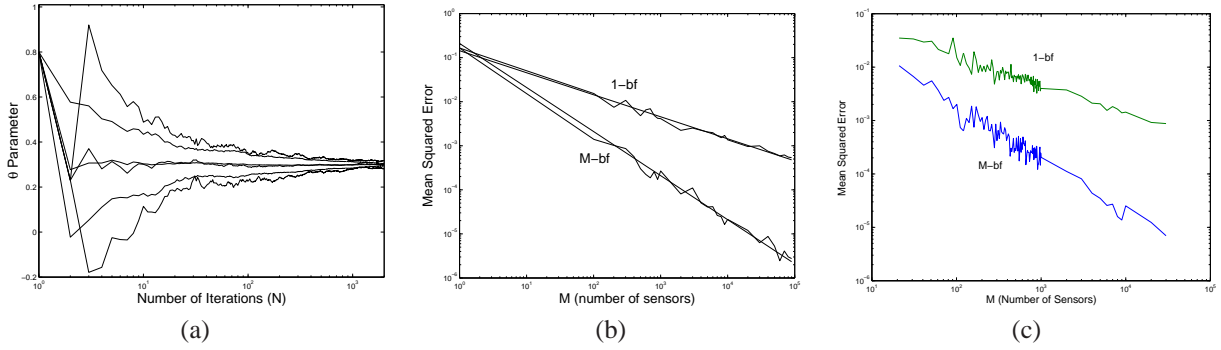


Fig. 2. Convergence of θ_n to θ^* with $m = 11$ nodes, and quantile level $\alpha^* = 0.3$. (b) Log-log plots of the variance against m for both algorithms (m -bf and 1-bf) with constant step sizes, and theoretically-predicted rate. (c) Log-log plots of the variance against m for m -bf and 1-bf algorithms with constant step size (c) Log-log plots of m -bf with constant step size versus 1-bf algorithm with decaying step size.

IV. ANALYSIS

Our proofs of Theorem 1 and 2 exploit results from the stochastic approximation literature [5], [2]. In particular, both types of parameter updates (4) and (7) can be written in the general form

$$\theta_{n+1} = \theta_n + \epsilon_n H(\theta_n, Y_{n+1}) \quad (16)$$

where $Y_{n+1} = (Y_{n+1}(1), \dots, Y_{n+1}(m))$.

Note that the step size choice $\epsilon_n = 1/n$ satisfies the conditions in equation (2). Moreover, the sequence (θ_n, Y_{n+1}) is Markov, since θ_n and Y_{n+1} depend on the past only via θ_{n-1} and Y_n .

In addition to these assumptions, convergence requires an additional attractiveness condition. For each fixed $\theta \in \mathbb{R}$, let $\mu_\theta(\cdot)$ denote the distribution of Y conditioned on θ . A key quantity in the analysis of stochastic approximation algorithms is the averaged function

$$h(\theta) := \int H(\theta, y) \mu_\theta(dy) = \mathbb{E}[H(\theta, Y) | \theta]. \quad (17)$$

We assume (as is true for our cases) that this expectation exists. Now the differential equation method dictates that under suitable conditions, the asymptotic behavior of the update (16) is determined essentially by the behavior of the ODE $\frac{d\theta}{dt} = h(\theta(t))$.

Almost sure convergence: Suppose that the following *attractiveness condition*

$$h(\theta) [\theta - \theta^*] < 0 \quad \text{for all } \theta \neq \theta^* \quad (18)$$

is satisfied. If, in addition, the variance $R(\theta) := \text{Var}[H(\theta; Y) | \theta]$ is bounded, then we are guaranteed that $\theta_n \xrightarrow{a.s.} \theta^*$ (see §5.1 in Benveniste et al. [2]).

Asymptotic normality: In our updates, the random variables Y_n take the form $Y_n = g(X_n, \theta_n)$ where the X_n are i.i.d. Suppose that the following stability condition is satisfied:

$$\gamma(\theta^*) := -\frac{dh}{d\theta}(\theta^*) > \frac{1}{2}. \quad (19)$$

Then we have

$$\sqrt{n} (\theta_n - \theta^*) \xrightarrow{d} \mathcal{N}\left(0, \frac{R(\theta^*)}{2\gamma(\theta^*) - 1}\right) \quad (20)$$

See §3.1.2 in Benveniste et al. [2] for further details.

A. Proof of Theorem 1

(a) The m -bit feedback algorithm is a special case of the general update (16), with $\epsilon_n = \frac{1}{n}$ and $H(\theta_n, Y_{n+1}) = K [\alpha^* - \frac{1}{m} \sum_{i=1}^m Y_{n+1}(i; \theta_n)]$. Computing the averaged function, we have

$$\begin{aligned} h(\theta) &= K \mathbb{E} \left[\alpha^* - \frac{1}{m} \sum_{i=1}^m Y_{n+1}(i; \theta_n) \mid \theta_n \right] \\ &= K (\alpha^* - F(\theta_n)), \end{aligned}$$

where $F(\theta_n) = \mathbb{P}(X \leq \theta_n)$. We then observe that θ^* satisfies the attractiveness condition (18), since

$$[\theta - \theta^*] h(\theta_n) = K [\theta - \theta^*] [\alpha^* - F(\theta_n)] < 0$$

for all $\theta \neq \theta^*$, by the monotonicity of the cumulative distribution function. Finally, we compute the conditional variance of H as follows:

$$\begin{aligned} R(\theta_n) &= K^2 \text{Var} \left[\alpha^* - \frac{\sum_{i=1}^m Y_{n+1}(i)}{m} \mid \theta_n \right] \\ &= \frac{K^2}{m} F(\theta_n) [1 - F(\theta_n)] \leq \frac{K^2}{4m}, \end{aligned} \quad (21)$$

using the fact that H is a sum of Bernoulli variables. Thus, we can conclude that $\theta_n \rightarrow \theta^*$ almost surely.

(b) Note that $\gamma(\theta^*) = -\frac{dh}{d\theta}(\theta^*) = K p_X(\theta^*) > \frac{1}{2}$, so that the stability condition (19) holds. Applying the asymptotic normality result (20) with the variance $R(\theta^*) = \frac{K^2}{m} \alpha^* (1 - \alpha^*)$ (computed from equation (21)) yields the claim. ■

B. Proof of Theorem 2

This argument involves additional analysis, due to the aggregate decision (6) taken by the fusion center. First, the aggregate decision Z_{n+1} is a Bernoulli random variable; we begin by computing its parameter. Each transmitted bit $Y_{n+1}(i)$ is $\text{Ber}(F(\theta_n))$, where we recall the notation $F(\theta) := \mathbb{P}(X \leq \theta)$. Using the definition (10), we have the equivalences

$$\mathbb{P}(Z_{n+1} = 1) = G_m(F(\theta_n), \alpha^*) \quad (22a)$$

$$\beta = G_m(\alpha^*, \alpha^*) = G_m(F(\theta^*), \alpha^*). \quad (22b)$$

The following result is elementary (proof omitted due to space constraints):

Lemma 1. For fixed $x \in [0, 1]$, the function $f(r) := G_m(r, x)$ is non-negative, differentiable and monotonically decreasing.

To establish almost sure convergence, we use a similar approach as in the previous theorem. We begin by computing the function h as follows

$$\begin{aligned} h(\theta) &= K_m \mathbb{E}[Z_{n+1} - \beta \mid \theta] \\ &= K_m [G_m(F(\theta), \alpha^*) - G_m(F(\theta^*), \alpha^*)], \end{aligned}$$

using the equivalences (22). We now establish the attractiveness condition (18). In particular, for any θ such that $F(\theta) \neq F(\theta^*)$, we calculate that $h(\theta) [\theta - \theta^*]$ is given by

$$K_m [G_m(F(\theta_n), \alpha^*) - G_m(F(\theta^*), \alpha^*)] [\theta_n - \theta^*] < 0,$$

where the inequality follows from the fact that $G_m(r, x)$ is monotonically decreasing in r for each fixed $x \in [0, 1]$ (using Lemma 1), and that the function F is monotonically increasing. Finally, computing the variance $R(\theta) := \text{Var}[H(\theta, Y) \mid \theta]$, we have

$$R(\theta) = K_m^2 G_m(F(\theta), \alpha^*) [1 - G_m(F(\theta), \alpha^*)] \leq \frac{K_m^2}{4}$$

since (conditioned on θ), the decision Z_{n+1} is Bernoulli with parameter $G_m(F(\theta); \alpha^*)$. Thus, we can conclude that $\theta_n \rightarrow \theta^*$ almost surely.

(b) To show asymptotic normality, we need to verify the stability condition. By chain rule, we have $\frac{h}{d\theta}(\theta^*) = K_m \frac{\partial G_m}{\partial r}(r, \alpha^*) \Big|_{r=F(\theta^*)} p_X(\theta^*)$. From Lemma 1, we have $\frac{\partial G_m}{\partial r}(F(\theta), \alpha^*) < 0$, so that the stability condition holds as long as $\gamma_m(\theta^*) > \frac{1}{2}$ (where γ_m is defined in the statement). Thus, asymptotic normality holds.

In order to compute the asymptotic variance, we need to investigate the behavior of $R(\theta^*)$ and $\gamma(\theta^*)$ as $m \rightarrow +\infty$. First examining $R(\theta^*)$, the central limit theorem guarantees that $G_m(F(\theta^*), y) \rightarrow \Phi\left(\sqrt{m} \frac{y - \alpha^*}{\alpha^*(1 - \alpha^*)}\right)$. Consequently, we have

$$R(\theta^*) = K_m^2 G_m(F(\theta^*), \alpha^*) [1 - G_m(F(\theta^*), \alpha^*)] \rightarrow \frac{K_m^2}{4}.$$

We now turn to the behavior of $\gamma(\theta^*)$:

Lemma 2. As $m \rightarrow +\infty$, we have

$$\frac{\partial G_m(r, \alpha^*)}{\partial r} \Big|_{r=F(\theta^*)} \rightarrow -\sqrt{\frac{m}{2\pi\alpha^*(1-\alpha^*)}}.$$

Proof: We compute that $G'_m(r) = \frac{\partial G_m(r, \alpha^*)}{\partial r}$ is equal to

$$\sum_{i=0}^{\lfloor m\alpha^* \rfloor} \binom{m}{i} [i r^{i-1} (1-r)^{m-i} - (m-i) r^i (1-r)^{m-i-1}].$$

With a bit of algebra, we obtain that

$$G'_m(r) = \frac{\mathbb{E}[X \mathbb{I}(X \leq \alpha^* m)] - \mathbb{E}[X] \mathbb{E}[\mathbb{I}(X \leq \alpha^* m)]}{r(1-r)}$$

where X is binomial with parameters (m, α^*) , and mean $\mathbb{E}[X] = \alpha^* m$. Applying the central limit theorem

yields that $\frac{1}{\sqrt{m}} X \xrightarrow{d} \mathcal{N}(\alpha^* \sqrt{m}, \alpha^*(1 - \alpha^*))$, whence $\mathbb{E}[\mathbb{I}(X \leq \alpha^* m)] \rightarrow \frac{1}{2}$. Furthermore, we can write

$$\mathbb{E}[X \mathbb{I}(X \leq \alpha^* m)] = \sqrt{m} \mathbb{E} \left[\frac{X}{\sqrt{m}} \mathbb{I} \left(\frac{X}{\sqrt{m}} \leq \alpha^* \sqrt{m} \right) \right].$$

Let us set $a = \alpha^*(1 - \alpha^*)$ to simplify notation. Since $\alpha^* \sqrt{m}$ is a continuity point of the normal distribution, we have (see §7.2, [3]) that

$$\mathbb{E} \left[\frac{X}{\sqrt{m}} \mathbb{I} \left(\frac{X}{\sqrt{m}} \leq \alpha^* \sqrt{m} \right) \right] \rightarrow \int_{-\infty}^{\alpha^* \sqrt{m}} x \frac{\exp\left(-\frac{(x - \alpha^* \sqrt{m})^2}{2a}\right)}{\sqrt{2\pi a}} dx$$

Making the change of variable $y = x - \alpha^* \sqrt{m}$ and evaluating the integral, we obtain $\frac{\alpha^* \sqrt{m}}{2} - \sqrt{\frac{a}{2\pi}}$. Finally, some algebra establishes that $\frac{\partial G_m(r, \alpha^*)}{\partial r} \Big|_{r=\alpha^*}$ converges to $-\sqrt{\frac{m}{2\pi\alpha^*(1-\alpha^*)}}$ as claimed. ■

Returning now to the proof of the theorem, we use Lemma 2 and put the pieces together to obtain that $\frac{R(\theta^*)}{2K_m \left| \frac{\partial G_m(r, \theta^*)}{\partial r} \Big|_{r=\alpha^*} p_X(\theta^*) - 1} \right.}$ converges to

$$\frac{K_m^2/4}{2K_m \sqrt{m} p_X(\theta^*) - 1} = \frac{1}{m} \left[\frac{K^2 \sqrt{2\pi\alpha^*(1-\alpha^*)}}{8K p_X(\theta^*) - 4\sqrt{2\pi\alpha^*(1-\alpha^*)}} \right],$$

with $K > \frac{\sqrt{2\pi\alpha^*(1-\alpha^*)}}{2p_X(\theta^*)}$ for stability, thus completing the proof of the theorem. ■

V. DISCUSSION

This paper treated the problem of decentralized quantile estimation under communication constraints, and proposed two different approaches. We showed that an m -bit feedback algorithm achieves the same asymptotic variance as the centralized estimator. We also analyzed the asymptotic behavior of 1-bit feedback schemes. Among the extensions that we are currently investigating involve noisy channels, and protocols based on $1 < L < m$ bits of feedback for reducing the constant in the rate.

REFERENCES

- [1] E. Ayanoglu. On optimal quantization of noisy sources. *IEEE Trans. Info. Theory*, 36(6):1450–1452, 1990.
- [2] A. Benveniste, M. Metivier, and P. Priouret. *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, New York, NY, 1990.
- [3] G.R. Grimmett and D.R. Stirzaker. *Probability and random processes*. Oxford Science Publications, Clarendon Press, Oxford, 1992.
- [4] J. A. Gubner. Decentralized estimation and quantization. *IEEE Trans. Info. Theory*, 39(4):1456–1459, 1993.
- [5] H. J. Kushner and G. G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, New York, NY, 1997.
- [6] Z.Q. Luo. Universal decentralized estimation in a bandwidth-constrained sensor network. *IEEE Trans. Info. Theory*, 51(6):2210–2219, 2005.
- [7] X. Nguyen, M. J. Wainwright, and M. I. Jordan. Nonparametric decentralized detection using kernel methods. *IEEE Trans. Signal Processing*, 53(11):4053–4066, November 2005.
- [8] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley Series in Probability and Statistics. Wiley, 1980.
- [9] R. R. Tenney and N. R. Jr. Sandell. Detection with distributed sensors. *IEEE Trans. Aero. Electron. Sys.*, 17:501–510, 1981.
- [10] J. N. Tsitsiklis. Decentralized detection. In *Advances in Statistical Signal Processing*, pages 297–344. JAI Press, 1993.
- [11] Z. Zhang and T. Berger. Estimation via compressed information. *IEEE Trans. Info. Theory*, 34(2):198–211, 1988.