

Fast Distributed Estimation of Empirical Mass Functions over Anonymous Networks

Håkan Terelius*, Damiano Varagnolo*, Carlos Baquero† and Karl Henrik Johansson*

Abstract—The aggregation and estimation of values over networks is fundamental for distributed applications, such as wireless sensor networks. Estimating the average, minimal and maximal values has already been extensively studied in the literature. In this paper, we focus on estimating empirical distributions of values in a network with anonymous agents. In particular, we compare two different estimation strategies in terms of their convergence speed, accuracy and communication costs. The first strategy is deterministic and based on the average consensus protocol, while the second strategy is probabilistic and based on the max consensus protocol.

Index Terms—distributed computation, consensus, data aggregation, order statistics

I. INTRODUCTION

Aggregating data over networks is essential for many distributed systems. However, simple aggregations such as averaging loses a lot of the information contained in the original data set. Aiming to expand the set of available aggregation tools, we propose and characterize algorithms that estimates empirical Probability Mass Functions (PMFs) over networks. Specifically, we consider collaborative anonymous agents that aim to compute estimates of empirical distributions in the shortest possible time.

A. Literature review

The vast literature on the estimation of probability densities / mass functions over networks can be divided in the main classes of parametric and non-parametric approaches.

Parametric approaches generally assume the estimand to have a certain structure before obtaining observations, e.g., to be a sum of Gaussians. Examples are the distributed implementations of the Expectation-Maximization (EM) algorithm [1], [2], [3], [4]. Nonparametric approaches instead do not fix the structure a priori, but rather select it from the observations. This class comprises the various distributed kernel density estimation [5], classification [6] and clustering approaches [7].

*ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden.
Emails: { hakante, damiano, kallej } @kth.se

†HASLab, INESC TEC & Universidade do Minho, Braga, Portugal.
Email: cbm@di.uminho.pt

The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement n°257462 HYCON2 Network of excellence, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, and from Project NORTE-07-0124-FEDER-000058 that is co-financed by the North Portugal Regional Operational Programme (ON.2 - O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF).

Besides the parametric / non-parametric classification, the existing literature can also be characterized on how the information is propagated and aggregated over the network. We notice strategies based on pre-established hierarchical tree routing structures, where the various nodes compute summaries of the empirical distributions in their sub-trees and propagate them towards the root, eventually obtaining the approximated statistics of the whole network in a bottom-up fashion [8], [9], [10], [11]. Other strategies are instead based on gossip communications, and exploit averaging techniques to explicitly compute Cumulative Distribution Functions (CDFs) [12], [13], [14]. Other techniques are based on estimating how many agents are in a certain specific state [15], [16].

B. Statement of contributions

We propose a novel algorithm with the properties: • be *symmetrically distributed*, i.e., without leaders / leader election steps, and with agents executing the same algorithm in parallel. • be *privacy preserving*, i.e., avoiding the possibility of tracing or characterizing a single agent. • exploit *aggregation techniques*, where the size of the exchanged information packets is constant in time. • be *fast*, i.e., s.t. the time for all the agents to share the same estimate is small.

Our proposed strategy is based on max-consensus (see Sec. III). As the max-consensus converges in no more than the steps needed to transmit information between arbitrary nodes in the network, this is one of the fastest possible aggregation mechanism. This emphasis on *fast convergence* is because time can be the crucial factor in many practical situations (e.g., in control applications).

From an algorithmic point of view our strategy departs from [12], [13], [17] by substituting the average consensus schemes with max-consensus. This apparently minor modification actually makes the two estimators completely different, and opens a variety of novel problems. In fact, while the average consensus scheme requires exchanging very few scalars per iteration and where the agents computes the *exact* PMF asymptotically in time, the max consensus scheme converges much faster than the average one, but not to the exact value. Indeed, the statistical performance depends on how many scalars are exchanged per iteration. *We specifically characterize the temporal behavior of the performance of the max-consensus strategy, concluding when it is preferable to the average consensus strategy.*

II. STATEMENT OF THE ESTIMATION PROBLEM

Consider a strongly connected network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of $V = |\mathcal{V}|$ agents communicating through the links \mathcal{E} . Let \mathcal{V}_i denote the set of neighbors of agent i , and $\mathcal{V}_t^{(i)}$ the set of the t -steps neighbors of agent i . We recall that $\mathcal{V}_t^{(i)}$ can be defined for $t = 0$ as $\mathcal{V}_0^{(i)} = \{i\}$ and, for $t \geq 1$, through the recursion

$$\mathcal{V}_t^{(i)} \doteq \bigcup_{(i,j) \in \mathcal{E}} \mathcal{V}_{t-1}^{(j)}. \quad (1)$$

Let every agent $i \in \mathcal{V}$ belong to a discrete state $z^{(i)} \in \mathbb{N}_B \doteq \{0, \dots, B-1\}$ (\mathbb{N}_B being the set of plausible states), e.g., given by sensor measurements. We are then interested in distributively estimating the relative frequencies of the local states $z^{(1)}, \dots, z^{(V)}$. I.e., if $n_b \doteq |\{i \text{ s.t. } z^{(i)} = b\}|$ is the number of agents in state b , then we aim to estimate the PMF

$$p_b \doteq \frac{n_b}{V} \quad b \in \mathbb{N}_B \quad (2)$$

given that the network size V is unknown while the plausible states \mathbb{N}_B are known.

We restrict our focus to distributed algorithms where each agent $i \in \mathcal{V}$ has a local variable $x^{(i)}(t)$ that can be modified at time $t+1$ by accessing the states $x^{(j)}(t)$'s of the neighboring nodes, and performing the aggregation operation

$$x^{(i)}(t+1) = f\left(x^{(i)}(t), x^{(j_1)}(t), x^{(j_2)}(t), \dots\right), \\ j_1, j_2, \dots \in \mathcal{V}_i$$

that preserves the dimension of $x^{(i)}(t)$. Furthermore, at every time t each agent computes a local estimate of the PMF function from the local variable $x^{(i)}(t)$,

$$\hat{p}_b^{(i)}(t) = g\left(x^{(i)}(t)\right)$$

for an appropriate estimation function $g(\cdot)$.

The estimation strategy is thus defined by the initial variables $x^{(i)}(0)$, the update function f and the estimation function g . In order to compare different estimation strategies we consider the Mean Squared Error (MSE) as a performance index J , i.e.,

$$J(\hat{p}_1, \dots, \hat{p}_B) \doteq \mathbb{E} \left[\frac{1}{V \cdot B} \sum_{b \in \mathbb{N}_B, i \in \mathcal{V}} \left(p_b - \hat{p}_b^{(i)} \right)^2 \right] \quad (3)$$

where the expectation is taken over the initial conditions.

Remark 1 For notational simplicity we consider static networks. Nonetheless it is straightforward to handle time-varying topologies by substituting the edges \mathcal{E} with a time-dependent set $\mathcal{E}(t)$, and the neighborhoods $\mathcal{V}^{(i)}$ with the time-dependent counterparts $\mathcal{V}^{(i)}(t)$.

The problem analyzed in this manuscript is to propose and compare different estimation schemes.

III. ESTIMATORS BASED ON CONSENSUS PROTOCOLS

We consider two particular estimators, one based on average consensus strategies (see also [12], [13], [17]), and a novel one based on max consensus strategies and structurally similar to the size estimation techniques in [18], [19].

In the following, we abstract away the message transmission and consider a distributed system where agents communicate by synchronous rounds. At each round, and over each edge, only a constant size message is transmitted, and no messages are lost.

Remark 2 For notational simplicity we consider synchronous communications. Nonetheless this could be relaxed for both estimators, since they can be adapted to operate with gossip asynchronous transmissions.

A. Estimator based on Average consensus

In the average consensus protocol, the local variable is a B -dimensional real vector $x^{(i)}(t) \in \mathbb{R}^B$ containing the estimate of the PMF. At initialization, each node set its local variable based on its own state,

$$x_b^{(i)}(0) = \begin{cases} 1, & \text{if } z^{(i)} = b \\ 0, & \text{otherwise.} \end{cases}$$

Let \mathbf{x}_b denote the vector of all agents' states $x_b^{(i)}$. It is known that if at each time the local variables are updated with an average consensus update like

$$\mathbf{x}_b(t+1) = W \mathbf{x}_b(t) \quad b \in \mathbb{N}_B \quad (4)$$

where W is a doubly-stochastic weight matrix (for example chosen as the Metropolis weights), then, assuming perfect computations¹, every $x_b^{(i)}(t)$ converges to the average of the initial values [21]. Thus

$$x_b^{(i)}(t) \xrightarrow{t \rightarrow \infty} \frac{1}{V} \sum_{j \in \mathcal{V}} x_b^{(j)}(0) = \frac{n_b}{V} = p_b.$$

The PMF estimate is simply

$$\hat{p}_b^{(i)}(t) = x_b^{(i)}(t). \quad (5)$$

To describe the convergence properties of (4), recall that the estimation error can be bounded by an exponential function [22], i.e., by

$$\|p_b - \hat{p}_b(t)\|_2 \leq c e^{-\alpha t}$$

where c and α depend on the initial condition, the network topology and the choice of the weights.

Remark 3 We do not consider more advanced protocols, such as accelerated average consensus [23], or finite-time average consensus [24]. The rationale for this choice is that we want to characterize the simplest averaging algorithm, with the smallest demands from both communication and computational points of view.

¹For simplicity we do not consider quantization effects, e.g., [20].

B. Estimator based on Max consensus

In the max consensus protocol, the local variable is a $B \times M$ -dimensional real matrix $x^{(i)}(t) \in \mathbb{R}^{B \times M}$ whose elements are initially set based on the local state as

$$x_{b,m}^{(i)}(0) \sim \begin{cases} \mathcal{U}[0,1], & \text{if } z^{(i)} = b \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where $\mathcal{U}[0,1]$ is the uniform distribution between 0 and 1. Then at each time t , the local variables are updated with the max consensus update

$$x_{b,m}^{(i)}(t) = \max_{j \in \mathcal{V}_t^{(i)}} \left\{ x_{b,m}^{(j)}(t-1) \right\}, \quad b \in \mathbb{N}_B, m = 1, \dots, M. \quad (7)$$

Notice that the definition of t -steps neighborhood $\mathcal{V}_t^{(i)}$ precisely captures the agents that contributed to the generation of $x_{b,m}^{(i)}(t)$, i.e.,

$$x_{b,m}^{(i)}(t) = \max_{j \in \mathcal{V}_t^{(i)}} \left\{ x_{b,m}^{(j)}(0) \right\}. \quad (8)$$

Let $V_t^{(i)} \doteq |\mathcal{V}_t^{(i)}|$,

$$p_b^{(i)}(t) \doteq \frac{|\{i \in \mathcal{V}_t^{(i)} \text{ s.t. } z^{(i)} = b\}|}{V_t^{(i)}}, \quad (9)$$

and $n_b^{(i)}(t) \doteq p_b^{(i)}(t)V_t^{(i)}$. As shown in Sec. IV, the Maximum Likelihood (ML) estimator for $n_b^{(i)}(t)$ given the $x_{b,m}^{(i)}(t)$'s is

$$\hat{n}_b^{(i)} = \left(\frac{1}{M} \sum_{m=1}^M -\ln(x_{b,m}^{(i)}) \right)^{-1}. \quad (10)$$

Now, since

$$p_b^{(i)}(t) = \frac{p_b^{(i)}(t)}{\sum_{\beta \in \mathbb{N}_B} p_\beta^{(i)}(t)} = \frac{n_b^{(i)}(t)}{\sum_{\beta \in \mathbb{N}_B} n_\beta^{(i)}(t)}$$

because of the functional invariance property of ML estimators [25, Thm. 7.2.10, p. 320], the ML estimate of $p_b^{(i)}(t)$ given the $x_{b,m}^{(i)}(t)$'s is

$$\hat{p}_b^{(i)}(t) = \frac{\hat{n}_b^{(i)}(t)}{\sum_{\beta \in \mathbb{N}_B} \hat{n}_\beta^{(i)}(t)}. \quad (11)$$

For $t \geq d$ (d being the network diameter) the max consensus strategy converges globally, and $n_b^{(i)}(t) = n_b$, thus the PMF estimated $p_1^{(i)}(t), \dots, p_B^{(i)}(t)$ converges to an estimate of the global PMF p_1, \dots, p_B .

Remarkably, this estimator provides additional estimates of the distributions of the states in every t -steps neighborhood. Considering a certain agent i , the set of $p_b^{(i)}(0), p_b^{(i)}(1), \dots$ correspond to local views of the neighborhoods empirical distribution that can be used by i to rapidly infer if close neighbors tend to have the similar states.

We notice that estimator (11) has strong similarities with the size estimators proposed in [26], [27], [28].

Nonetheless, as reported in the following section, its statistical properties are essentially different since each vector $[\hat{p}_1^{(i)}(t), \dots, \hat{p}_B^{(i)}(t)]$ has correlated components.

We also notice that appropriate termination rules can be based on estimates of the diameter d of the network, again obtained by exploiting max consensus approaches as in [19], [29].

We finally notice that, under continuity assumptions, the stochastic generation mechanism proposed in (6) is not a design parameter. As soon as we neglect quantization effects, substituting $\mathcal{U}[0,1]$ with another continuous probability distribution leads to estimators with identical statistical performance, see [18].

C. Summary of the differences between the two estimators

The max consensus scheme (11) converges in d steps to an estimate of the true PMF. Given a fixed M , its MSE J (3) will vary up to time $t = d$ and then remain constant. Increasing M , the MSE curves are also expected to get closer and closer to zero, due to the consistency property of ML estimators.

The average consensus scheme (5) requires nodes to exchange less information, and is in general converging asymptotically for $t \rightarrow +\infty$. These comments are graphically represented in fig. 1.

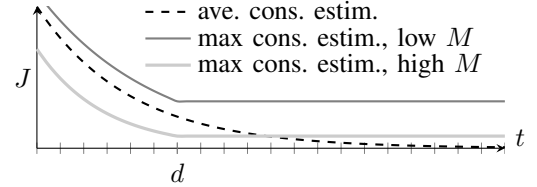


Fig. 1: Graphical representation of the properties from the estimators. By increasing M it is possible to let the max consensus estimator (11) perform better than the average consensus scheme (5) for $t \leq d$.

The aim is to find conditions on M and on the network for which it is possible to state which algorithm is preferable for $t \leq d$, i.e., when time is a concern. To this end, we first need to describe the statistical properties of the max consensus estimator.

IV. STATISTICAL CHARACTERIZATION OF THE MAX CONSENSUS PMF ESTIMATOR

For notational simplicity we consider the stationary state where the max consensus has already been computed, i.e., where $x_{b,m}^{(i)}(t) = x_{b,m} \doteq \max_{i \in \mathcal{V}} \{x_{b,m}^{(i)}(0)\}$. With this assumption the joint PMF $p(\hat{n}_b; n_1, \dots, n_B, M)$ is equal to $p(\hat{n}_b^{(i)}(t); n_1^{(i)}(t), \dots, n_B^{(i)}(t), M)$. To derive these distributions we then consider that if $b \neq \beta$ then $x_{b,m}$ is statistically independent of the parameter n_β . Thus, from simple order-statistics arguments [30],

$$p(x_{b,m}; n_1, \dots, n_B) = p(x_{b,m}; n_b) = n_b (x_{b,m})^{n_b-1}$$

for all m (we omit the dependency on the parameter M for notational brevity). Since the $x_{b,m}$'s are i.i.d. we have

$$\begin{aligned} p(x_{b,1}, \dots, x_{b,M}; n_b) &= \prod_{m=1}^M p(x_{b,m}; n_b) \\ &= n_b^M \prod_{m=1}^M (x_{b,m})^{n_b-1} \end{aligned} \quad (12)$$

To derive $p(\hat{n}_b; n_b)$ consider that $z \doteq -\ln((x_{b,m}))$ is an exponential random variable with rate n_b , i.e.,

$$p(z; n_b) = \begin{cases} n_b e^{-n_b z} & \text{if } z \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

From (10), $M\hat{n}_b^{-1}$ is the sum of M i.i.d. exponential random variables with rate n_b , i.e., $M\hat{n}_b^{-1}$ is a Γ variate with shape M and scale $\frac{1}{n_b}$. Thus $M^{-1}\hat{n}_b \sim \text{I-}\Gamma(M, n_b)$, i.e.,

$$\begin{aligned} p(\hat{n}_b; n_b, M) &= \text{I-}\Gamma(M, Mn_b) \\ &= \Gamma(M)^{-1} \frac{1}{\hat{n}_b} \left(\frac{Mn_b}{\hat{n}_b} \right)^M \exp\left(-\frac{Mn_b}{\hat{n}_b}\right) \end{aligned}$$

where M is the shape and Mn_b the scale. For the estimate (11), \hat{p}_b is thus the ratio of correlated sums of inverse-Gamma variates, each with its own scale.

Unfortunately to the best of our knowledge there exists no currently available literature describing the distribution of this kind of variates. The closest manuscripts in fact characterize ratios of the form $\frac{x}{x+y}$ where x and y are independent inverse Γ variates [31]. Moreover both the Gamma and inverse Gamma distributions are not closed, i.e., linear combinations of independent copies of these kind of variates have not the same original distribution, up to location and scale parameters, see [32]. This means that there is no possibility to reduce the fraction (11) to the case described in [31], and characterization of the statistical properties of \hat{p}_b must rely on Monte Carlo (MC) integration methods.

Case $\mathbb{N}_B = \{0, 1\}$

In this case $\hat{p}_b^{(i)}(t)$ becomes a special ratio that is described in [31]. The probability density for \hat{p}_0 is then

$$p_{\hat{p}_0}(x; n_0, n_1, M) = \frac{(x(1-x))^{M-1}}{\left(\frac{n_0}{n_1}\right)^M B(M, M)} \left(1 + \frac{n_1 - n_0}{n_0} x\right)^{-2M} \quad (14)$$

where $B(\cdot, \cdot)$ is the Beta function and $x \in [0, 1]$. Its cumulative distribution is given by (18) where

$${}_2F_1(a, b; c; x) \doteq \sum_{i=0}^{+\infty} \frac{(a)_i (b)_i}{(c)_i \cdot i!} x^i \quad (15)$$

is the Gauss hypergeometric function and

$$(x)_i \doteq x(x+1)\cdots(x+i-1) \quad (16)$$

is the so called Pochhammer symbol (with the convention that $(x)_0 = 1$). From this, it is possible to compute the

moments of \hat{p}_0 (and thus of $\hat{p}_0 - \mathbb{E}[\hat{p}_0]$) using the relation

$$\mathbb{E}[(\hat{p}_0)^k] = \begin{cases} \frac{B(M+k, M)}{B(M, M)} \mathcal{F}(k, M, n_0, n_1) & \text{if } n_0 > n_1 \\ \left(\frac{n_0}{n_1}\right)^k \frac{B(M+k, M)}{B(M, M)} \mathcal{F}(k, M, n_1, n_0) & \text{otherwise.} \end{cases} \quad (17)$$

where

$$\mathcal{F}(k, M, a, b) \doteq {}_2F_1\left(k, M; 2M+k; \frac{a-b}{a}\right)$$

(notice that n_0 and n_1 appear in inverted positions in the two cases in (17)).

When $n_0 = n_1$ the estimators are unbiased for every M , otherwise, as expected, they are only asymptotically unbiased (for $M \rightarrow +\infty$).

In figures 2 and 3 we evaluate the relative bias and MSE of the estimator based on the design parameter M and on the distribution of the states. Notice that the MSE performances follow the typical $O\left(\frac{1}{M}\right)$ property for this kind of estimators.

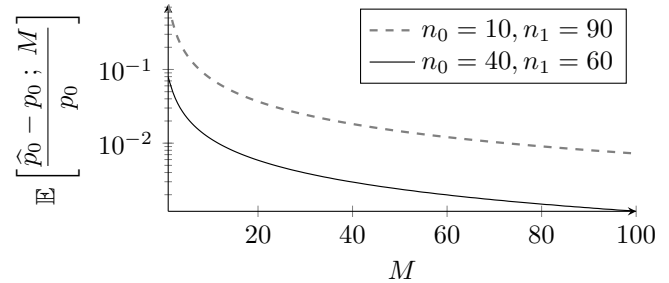


Fig. 2: Dependency of the relative bias $\mathbb{E}\left[\frac{\hat{p}_0 - p_0}{p_0}; M\right]$ on M for various values of n_0 and n_1 . The estimators are unbiased for every M if $n_0 = n_1$.

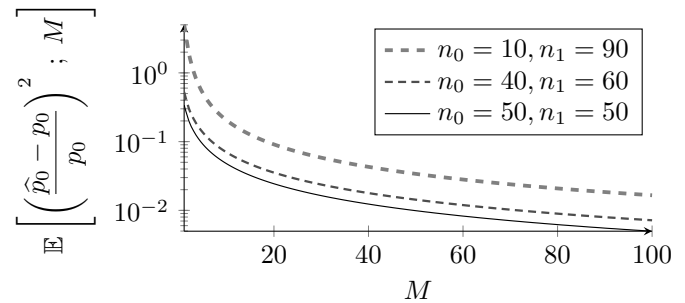


Fig. 3: Dependency of the relative MSE $\mathbb{E}\left[\left(\frac{\hat{p}_0 - p_0}{p_0}\right)^2; M\right]$ on M for various values of n_0 and n_1 .

Remark 4 The performances indicators summarized in figures 2 and 3 are valid for general $\hat{p}_b^{(i)}(t)$'s when associated

$$F_{\hat{p}_0}(x; n_0, n_1, M) = \frac{\left(1 + \frac{n_1}{n_0} \frac{1-x}{x}\right)^{-M}}{MB(M, M)} \cdot {}_2F_1\left(M, 1-M; M+1; \left(1 + \frac{n_1}{n_0} \frac{1-x}{x}\right)^{-1}\right) \quad (18)$$

with the local $n_b^{(i)}(t)$'s. The derivations of this section thus also characterize the behavior of the estimators during the transient.

V. COMPARISONS

Here we compare the performance between the average consensus based estimator (5) and the max consensus based estimator (11) during their transients. Our primary goal is to determine when to choose each algorithm, and how to tune the parameter M for the max consensus estimator.

We consider four different network topologies, i.e., the line topology (fig. 4a), the cyclic topology (fig. 4b), the cyclic grid topology (fig. 4c), and a geometric random topology (fig. 4d), each network consisting of 100 nodes.

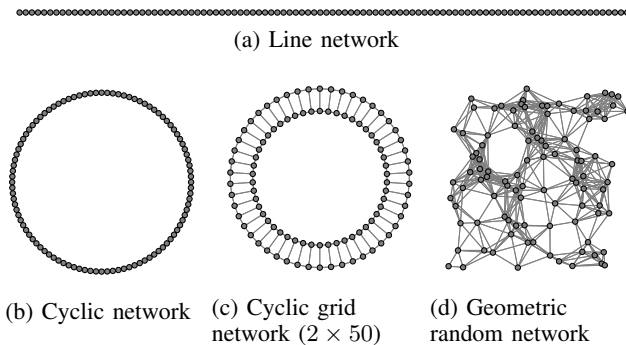


Fig. 4: Network topologies, with 100 nodes.

We evaluate the algorithms with Monte Carlo (MC) simulations, using the MSE (3) as the performance index, where the mean is taken over all agents and all MC runs. For each network the communication protocol proceeds in synchronous rounds, where nodes cyclically repeat the steps described in (4) and (7).

- *First experiment - fig. 5:* we select a random initial state for each MC run, where each agent is in state $z^{(i)} = 0$ or $z^{(i)} = 1$ with equal probability. The figure shows the 95% confidence intervals for both the average consensus based estimator as well as for the max consensus based estimator with $M = 10$, $M = 100$ and $M = 1000$.

As expected, the average consensus based estimator converges *asymptotically* to the true value. The max consensus based estimator converges instead *in a finite time* (after d steps, $d \doteq$ diameter of the network) to an estimate whose MSE decreases with M . In this scenario the choices $M = 100$ and $M = 1000$ yield similar and reasonable precisions that outperform the average consensus in most cases.

We observe that for the max consensus-based scheme a remarkable phenomenon may appear, specially when M is small ($M = 10$), the MSE increases with the number of iterations. This behavior is induced by a combination of facts:

first, the MSE index considered sums the agents' local MSEs. Second, small M 's induce estimates with high statistical variance, i.e., increase the chances that at least one agent will have some $\hat{p}_b^{(i)}(t)$ noticeably overestimated. At time $t = 1$ this overestimation does not greatly influence the overall MSE, since it affects only the erroneous agent, but as time passes, the max consensus spreads the overestimation through the agents.

- *Second experiment - fig. 6:* we now consider a single worst-case initial distribution of the states $z^{(i)}$, where the leftmost half of the agents in fig. 4 are in state 0 and the rightmost half are in state 1. Notice that this is actually not an unreasonable distribution, since for estimation applications in wireless sensor networks the communication topology and the measured environmental quantities might be spatially correlated.

Since there is only one fixed initial state, the average consensus based estimator is deterministic and unique. The figure thus compares the confidence intervals of the max consensus estimators (depending upon the realization of the $x_{b,m}^{(i)}$'s) against the performance of the deterministic average consensus estimate.

The outcome is that the max consensus based estimator (11) can be much faster and more accurate than the average consensus counterpart (5), even for very small M 's (even though a larger M improves the accuracy). The motivation is that if the distribution of the states is not geographically homogeneous then the max consensus is much more efficient at propagating information about certain states through the network.

VI. CONCLUDING REMARKS AND FUTURE DIRECTIONS

The two distributed estimators of empirical PMFs over networks considered here, one based on max consensus and one based on average consensus, have several intrinsic differences. With the average consensus, agents exchange messages containing only 1 scalar, while with the max consensus they exchange messages containing M scalars (M being a design parameter). For the average, convergence is asymptotic in time, while for the max, convergence is in finite time. With the average, the final estimate is equal to the true value, while with the max the final estimate has a statistical precision that is directly related to M .

The results indicate that there is no uniformly better algorithm: while in certain situations the average consensus strategy is the most reasonable approach, in some others it is outperformed by the max consensus. The rationale is based on how the states of the peers are distributed across the network and how fast the consensus strategies mix the information. If the states are geographically clustered (close nodes have similar states), then the max consensus scheme

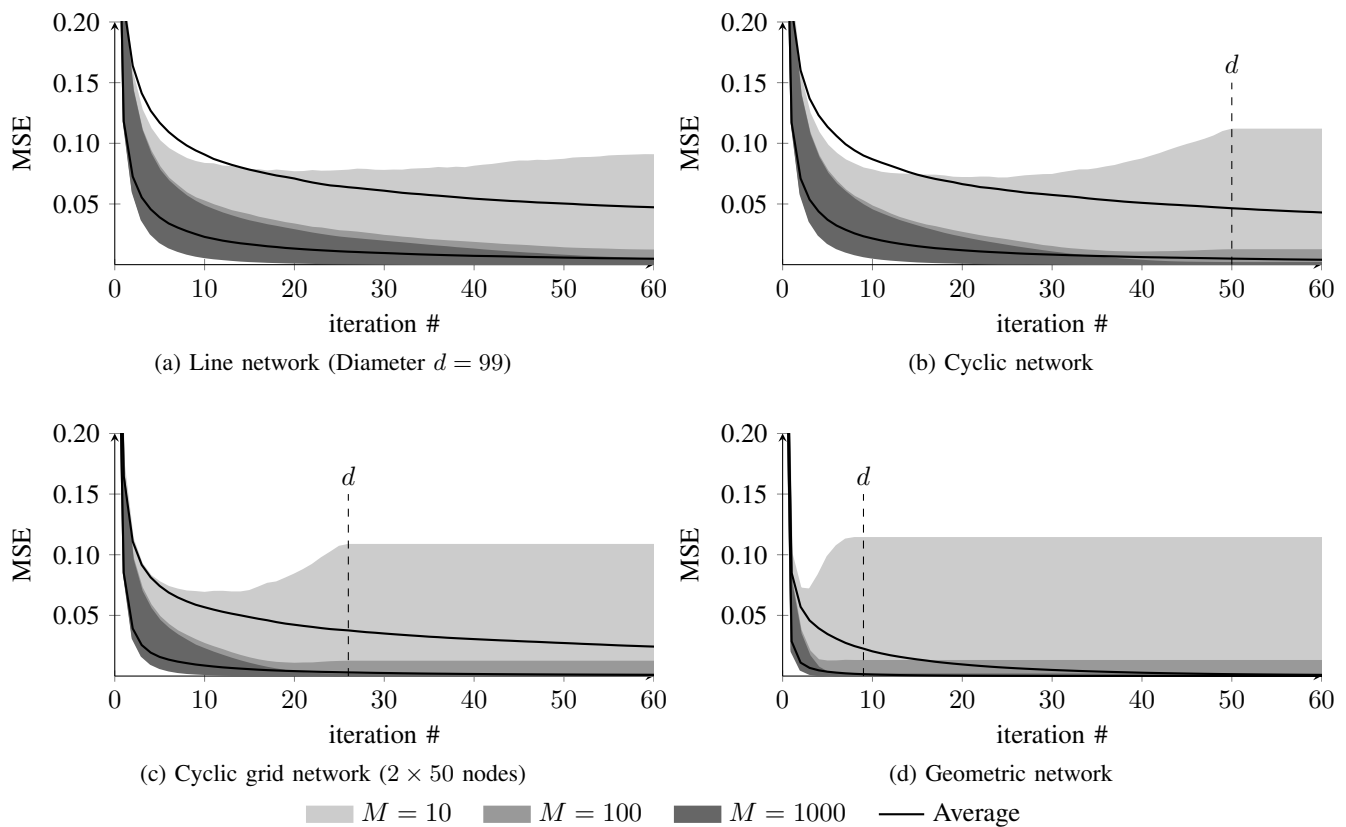


Fig. 5: Comparison of max-consensus based estimator against the average consensus based estimator. Each network consists of 100 nodes, and the network diameter d is marked in the figures. The shaded regions mark the 95% confidence interval for the max-consensus estimator, while the two solid lines mark the upper and lower end of the 95% confidence interval for the average-consensus estimator.

is generally preferable, because it is faster in spreading information of the existence of other states across the network.

This work opens a variety of future research directions. The first one is a more precise characterization of when each strategy performs better than the other, and how to tune M . The next one is how to exploit the estimation to perform fast detection of changes in the aggregated network state. Another important direction is to associate the state with local topological properties, e.g., by setting it equal to the number of neighbors, and estimate the most likely shape of the network.

REFERENCES

- [1] H. Jiang and S. Jin, "Scalable and Robust Aggregation Techniques for Extracting Statistical Information in Sensor Networks," in *IEEE International Conference on Distributed Computing Systems*, 2006.
- [2] R. Nowak, "Distributed EM algorithms for density estimation and clustering in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2245–2253, Aug. 2003.
- [3] P. A. Forero, A. Cano, and G. B. Giannakis, "Consensus-based distributed expectation-maximization algorithm for density estimation and classification using wireless sensor networks," in *Acoustics, Speech and Signal Processing*, Las Vegas, Nevada, 2008.
- [4] N. Vlassis, Y. Sfakianakis, and W. Kowalczyk, "Gossip-based greedy Gaussian mixture learning," *Advances in Informatics*, vol. 3746, pp. 349–359, 2005.
- [5] Y. Hu, J.-g. Lou, H. Chen, and J. Li, "Distributed density estimation using non-parametric statistics," in *Distributed Computing Systems*, no. 49, Toronto, Canada, 2007.
- [6] M. Klusch, S. Lodi, and G. Moro, "Distributed clustering based on sampling local density estimates," in *International joint conference on Artificial intelligence*, 2003.
- [7] X. Nguyen, M. Wainwright, and M. Jordan, "Nonparametric decentralized detection using kernel methods," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4053–4066, Nov. 2005.
- [8] M. B. Greenwald and S. Khanna, "Power-conserving computation of order-statistics over sensor networks," in *ACM Symposium on Principles of Database Systems*. New York, USA: ACM Press, 2004, pp. 275–285.
- [9] N. Shrivastava, C. Buragohain, D. Agrawal, and S. Suri, "Medians and beyond: new aggregation techniques for sensor networks," in *International Conference on Embedded Networked Sensor Systems*, 2004, pp. 239–249.
- [10] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "TAG: a Tiny AGgregation service for ad-hoc sensor networks," *ACM SIGOPS Operating Systems Review*, vol. 36, no. SI, pp. 131–146, Dec. 2002.
- [11] S. Motegi, K. Yoshihara, and H. Horiuchi, "DAG based in-network aggregation for sensor network monitoring," in *International Symposium on Applications and the Internet*, 2006, pp. 292–299.
- [12] M. Borges, P. Jesus, C. Baquero, and P. S. Almeida, "Spectra: Robust Estimation of Distribution Functions in Networks," *Distributed Applications and Interoperable Systems*, vol. 7272, pp. 96 – 103, 2012.
- [13] J. Sacha, J. Napper, C. Stratan, and G. Pierre, "Adam2: Reliable distribution estimation in decentralised environments," in *Distributed Computing Systems*, Genova, Italy, 2010.
- [14] M. Haridasan and R. V. Renesse, "Gossip-based distribution

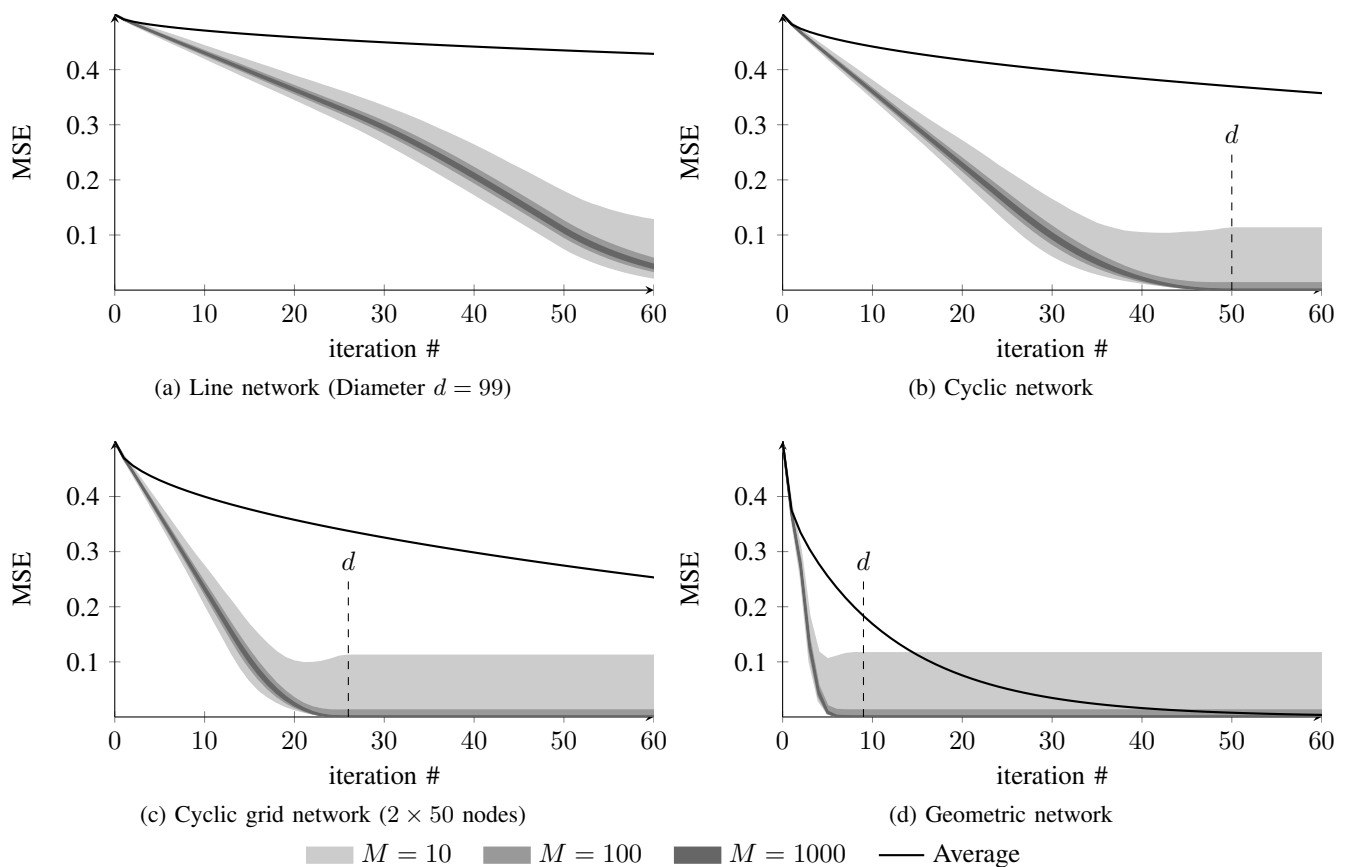


Fig. 6: Comparison of max-consensus based estimator against the average consensus based estimator for a single worst-case initial condition. Each network consists of 100 nodes, and the initial state is determined by the agents spatial configuration. The shaded regions mark the 95% confidence interval for the max-consensus estimator, while the solid line mark the deterministic estimation for the average-consensus estimator.

estimation in peer-to-peer networks,” in *International Conference on Peer-to-Peer Systems*, 2008.

[15] S. Cheng, J. Li, Q. Ren, and L. Yu, “Bernoulli Sampling Based (e, d)-Approximate Aggregation in Large-Scale Sensor Networks,” in *IEEE INFOCOM*, Mar. 2010, pp. 1181–1189.

[16] L. Massoulié, E. L. Merrer, A.-M. Kermarrec, and A. Ganesh, “Peer counting and sampling in overlay networks: random walk methods,” in *ACM symposium on Principles of distributed computing*, 2006, pp. 123–132.

[17] P. C. d. O. Jesus, “Robust Distributed Data Aggregation,” Ph.D. Thesis, Universidade do Minho, 2011.

[18] D. Varagnolo, G. Pillonetto, and L. Schenato, “Distributed statistical estimation of the number of nodes in Sensor Networks,” in *IEEE Conference on Decision and Control*, Atlanta, USA, Dec. 2010, pp. 1498–1503.

[19] J. C. S. Cardoso, C. Baquero, and P. S. Almeida, “Probabilistic Estimation of Network Size and Diameter,” in *Fourth Latin-American Symposium on Dependable Computing*, João Pessoa, Brasil, Sept. 2009, pp. 33–40.

[20] R. Carli, F. Fagnani, P. Frasca, and S. Zampieri, “Gossip consensus algorithms via quantized communication,” *Automatica*, vol. 46, no. 1, pp. 70–80, 2010.

[21] F. Fagnani and S. Zampieri, “Randomized consensus algorithms over large scale networks,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 634–649, May 2008.

[22] L. Xiao and S. Boyd, “Fast linear iterations for distributed averaging,” in *IEEE Conference on Decision and Control*, 2003, pp. 4997–5002.

[23] T. Aysal, B. Oreshkin, and M. Coates, “Accelerated Distributed Average Consensus via Localized Node State Prediction,” *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1563–1576, Apr. 2009.

[24] Y. Yuan, G.-B. Stan, M. Barahona, L. Shi, and J. Goncalves, “Decentralised minimal-time consensus,” in *IEEE Conference on Decision and Control and European Control Conference*, Dec. 2011, pp. 4282–4289.

[25] G. Casella and R. L. Berger, *Statistical inference*, 2nd ed. Duxbury Thomson Learning, 2001.

[26] C. Baquero, P. S. S. Almeida, R. Menezes, and P. Jesus, “Extrema Propagation: Fast Distributed Estimation of Sums and Network Sizes,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 4, pp. 668 – 675, Apr. 2012.

[27] H. Terelius, D. Varagnolo, and K. H. Johansson, “Distributed size estimation of dynamic anonymous networks,” in *IEEE Conference on Decision and Control*, 2012, pp. 5221–5227.

[28] J. Cichon, J. Lemiesz, W. Szpankowski, and M. Zawada, “Two-Phase Cardinality Estimation Protocols for Sensor Networks with Provable Precision,” in *IEEE Wireless Communications and Networking Conference*, Paris, France, Apr. 2012.

[29] F. Garin, D. Varagnolo, and K. H. Johansson, “Distributed estimation of diameter, radius and eccentricities in anonymous networks,” in *3rd IFAC Workshop on Distributed Estimation and Control in Networked Systems*, 2012, pp. 13–18.

[30] H. A. A. David and H. N. N. Nagaraja, *Order Statistics*. Wiley series in Probability and Statistics, 2003.

[31] M. Ali, M. Pal, and J. Woo, “On the Ratio of Inverted Gamma Variates,” *Austrian Journal of Statistics*, vol. 36, no. 2, pp. 153–159, 2007.

[32] V. Witkovsky, “Computing the distribution of a linear combination of inverted gamma variables,” *Kybernetika*, vol. 37, no. 1, pp. 79–90, 2001.