Advances in Applied Mathematics 45 (2010) 36–49

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

Iterative maximum likelihood on networks

Elchanan Mossel [a,b,1], Omer Tamuz [a,*,2]

[a] Weizmann Institute, Rehovot 76100, Israel
[b] University of California, Berkeley, CA 94720-3840, USA

ARTICLE INFO

Article history:
Received 30 April 2009
Accepted 31 August 2009
Available online 3 December 2009

MSC:
62F12
90B18
91D30

Keywords:
Maximum likelihood
Sparse sensing
Learning on networks
Algebraic recursion relation
Bayesian Economics

ABSTRACT

We consider n agents located on the vertices of a connected graph. Each agent v receives a signal $X_v(0) \sim N(\mu, 1)$ where μ is an unknown quantity. A natural iterative way of estimating μ is to perform the following procedure. At iteration $t+1$ let $X_v(t+1)$ be the average of $X_v(t)$ and of $X_w(t)$ among all the neighbors w of v. It is well known that this procedure converges to $X(\infty) = \frac{1}{2}|E|^{-1} \sum d_v X_v$ where d_v is the degree of v.

In this paper we consider a variant of simple iterative averaging, which models "greedy" behavior of the agents. At iteration t, each agent v declares the value of its estimator $X_v(t)$ to all of its neighbors. Then, it updates $X_v(t+1)$ by taking the maximum likelihood (or minimum variance) estimator of μ, given $X_v(t)$ and $X_w(t)$ for all neighbors w of v, and the structure of the graph.

We give an explicit efficient procedure for calculating $X_v(t)$, study the convergence of the process as $t \to \infty$ and show that if the limit exists then $X_v(\infty) = X_w(\infty)$ for all v and w. For graphs that are symmetric under actions of transitive groups, we show that the process is efficient. Finally, we show that the greedy process is in some cases more efficient than simple averaging, while in other cases the converse is true, so that, in this model, "greed" of the individual agents may or may not have an adverse affect on the outcome.

The model discussed here may be viewed as the maximum likelihood version of models studied in Bayesian Economics. The ML variant is more accessible and allows in particular to show the significance of symmetry in the efficiency of estimators using networks of agents.

* Corresponding author.
E-mail addresses: mossel@stat.berkeley.edu (E. Mossel), omertamuz@gmail.com (O. Tamuz).

0196-8858/$ – see front matter
doi:10.1016/j.aam.2009.11.004

1. Introduction

Networks and graphs are often viewed as computational models. In computational complexity several complexity classes are studied in terms of corresponding computation graphs, for example finite-automata, PSPACE and LOG-SPACE. For general background see, e.g., [1]. In parallel computing, networks are used to model the communication network between different computers, while in sparse sensing the connectivity network is of fundamental computation significance (see, e.g., [2] and [10]).

A recent trend emanating from Economics and Game Theory considers networks where different nodes correspond to computational entities with different objectives [5]. Recent models in Bayesian Economics consider models where each player is repeatedly taking actions that are based on a signal he has received that is correlated with the state of the word and past actions of his neighbors [9,3,8].

In this paper we study a simple model where, in each iteration, agents iteratively try to optimally estimate the state of the world, which is a single parameter $\mu \in \mathbb{R}$. It is assumed that originally each agent receives an independent sample from a normal distribution with mean μ. Later at each iteration each agent updates his estimate by taking the maximum likelihood estimator of μ given its current estimator and those of its neighbors, *and given the graph structure*. Note that for normal distributions, the maximum likelihood estimator is identical with the minimum variance unbiased estimator. At the first iteration, the estimator at each node will be the average of the original signal at the node and its neighbors. However, from the second iteration on, the procedure will not proceed by simple averaging due to the correlation between the estimators at adjacent nodes. As we show below, this correlation can be calculated given the structure of the graph and results in dramatic differences from the simple averaging process. Note that under this model, the agents are memoryless and use only the results of the last iteration to calculate those the next.

The results presented here are also applicable for non-normal measurements, when the agents' distributions are such that the minimal variance estimators are still linear combinations of the measurements. In such non-normal cases the estimators will no longer be ML estimators, but will still, using the same calculation, be minimum variance estimators.

The model suggested above raises a few basic questions:

- Is the process above well defined?
- Can the estimators be efficiently calculated? Note that in the Bayesian economic models (such as [8]) there are no efficient algorithms for updating beliefs.
- Does the process converge?

We answer the first two questions positively, and conjecture that the answer to the third is positive as well. Once these questions are addressed we prove a number of results regarding the limit estimators including:

- We show that for connected graphs, as $t \to \infty$, the correlation between the estimators of the different agents goes to one.
- We describe a graph for which the maximum likelihood process converges to an estimator different than the optimal.
- We compare the statistical efficiency of the limiting estimator to the limiting estimator obtained by simple iterative averaging and to the optimal estimator, in different graphs.

Most of the results of this paper were presented in the 2009 Allerton conference [6].

[1] Supported by a Sloan fellowship in Mathematics, by BSF grant 2004105, by NSF Career Award (DMS 054829) by ONR award N00014-07-1-0506 and by ISF grant 1300/08.
[2] Supported by ISF grant 1300/08.

1.1. Formal definition of the model

We consider a finite, undirected, connected graph $G = (V, E)$, where each vertex has a self-loop so that $\forall v$: $(v, v) \in E$, and a state of the world $\mu \in \mathbb{R}$. We assign each vertex v a normal unbiased estimator $X_v = X_v(0)$ of μ so that $\mathbf{E}[X_v] = \mu$ and $\mathbf{Var}[X_v] = 1$, for all v. These estimators are uncorrelated.

In iteration $t \in \mathbb{N}$ we define $X_v(t+1)$ to be the minimum variance unbiased estimator constructible over the estimators of v and its neighbors $N(v) = \{w \mid (v, w) \in E\}$ at time t

$$X_v(t+1) = \sum_{w \in N(v)} \alpha_w X_w(t), \quad \text{where} \sum_{w \in N(v)} \alpha_w = 1, \quad \text{and} \tag{1}$$

$$\alpha \text{ minimizes } \mathbf{Var}\left[\sum_w \alpha_w X_w(t)\right] \tag{2}$$

(note that α may be positive or negative). The process Y_v is given by simple iterative averaging so $Y_v(0) := X_v$, and

$$Y_v(t+1) = \frac{1}{d_v} \sum_{w \in N(v)} Y_w(t). \tag{3}$$

It is well known that $Y_v(t)$ converges to $Y_v(\infty) = \frac{1}{2}|E|^{-1} \sum d_v X_v$.

Finally, we define $Z(\infty)$ to be the global minimum variance unbiased estimator:

$$Z(\infty) = \frac{1}{|V|} \sum_{v \in V} X_v.$$

Note that in the case of normally distributed X_v's, the minimum variance definitions coincide with those of maximum likelihood.

This scheme can be generalized to the case where the original estimators X_v have a general covariance structure, with the definitions for $X_v(t)$ and $Y_v(t)$ remaining essentially the same, and that of $Z(\infty)$ changing to the form of (4) below.

1.2. Statements of the main results

The process defined by (1) and (2) is well defined. More formally:

Proposition. *(See Proposition 2.1.) For every realization of the random variables $X_v(0)$, $v \in V$ and for all $t \geqslant 1$, $X_v(t)$ is uniquely determined.*

The process can be calculated efficiently:

Proposition. *(See Proposition 2.2.) It is possible to calculate $\{X_v(t) \mid v \in V\}$, given $\{X_v(t-1) \mid v \in V\}$, by performing n operations of finding the point of an n-dimensional affine space (as specified by a generating set of size at most n) with minimal L_2 norm.*

Calculating the latter is a classical convex optimization problem. See, e.g., [4].

For transitive graphs the process always converges to the optimal estimator:

Proposition. *(See Proposition 2.7.) Let G be a transitive graph (defined below). Then $X_v(t)$ converges to $X(\infty) = Z(\infty)$.*

For graphs of large maximal degree, $X_v(t)$ converges to μ:

Proposition. *(See Proposition 3.1.) Let $G_n = (V_n, E_n)$ be a family of graphs where $|V_n| = n$ and $\max_{v\in V_n} d_v \to \infty$. Then*

$$\lim_{n\to\infty} \sup_{v\in V_n} \lim_{t\to\infty} \mathbf{E}\big[\big(X_v(t)-\mu\big)^2\big] = 0.$$

Note by comparison that for any graph,

$$\mathbf{E}\big[\big(Y(\infty)-\mu\big)^2\big] = \frac{1}{4}|E|^{-2}\sum_{v\in V} d_v^2.$$

In particular for a star on n vertices, as $n\to\infty$ it holds that $X(\infty)$ converges to μ but $Y(\infty)$ does not.

Finally, for some graphs, the process converges to a limit different than $Z(\infty)$ and $Y(\infty)$.

Theorem. *(See Theorem in Appendix A.) Let $G = (V, E)$ be the interval of length four where $V = \{a, b, c, d\}$ and $E = \{\{a, b\}, \{b, c\}, \{c, d\}\}$. Then $X_v(t)$ converges to a limit $X(\infty)$, where*

$$X(\infty) = \frac{1}{4}\big[(1-\xi)(X_a + X_d) + (1+\xi)(X_b + X_c)\big],$$

$$\mathbf{Var}\big[X(\infty)\big] = \xi,$$

with $\xi = 2 - \sqrt{3} = \frac{1}{4}(1 + \sqrt{49} - \sqrt{48})$.

Note that for this graph

$$Y(\infty) = \frac{1}{4}\left(\left(1-\frac{1}{5}\right)(X_a + X_d) + \left(1+\frac{1}{5}\right)(X_b + X_c)\right),$$

$$\mathbf{Var}\big[Y(\infty)\big] = 0.26,$$

and

$$Z(\infty) = \frac{1}{4}(X_a + X_b + X_c + X_d),$$

$$\mathbf{Var}\big[Z(\infty)\big] = \frac{1}{4}.$$

1.2.1. Conjectures

Showing some supporting results, we conjecture that the process always converges, and in particular to a state where all agents have the same estimator.

We present a number of additional open problems and conjectures in the conclusion.

2. General proofs

2.1. The process is well defined

Proposition 2.1. *For every realization of the random variables $X_v(0)$, $v \in V$ and for all $t \geqslant 1$, $X_v(t)$ is uniquely determined.*

Proof. Let $X_v^{(1)}(t) = A$ and $X_v^{(2)}(t) = B$ be minimum variance estimators satisfying (1) and (2), with variance V. Then their average must have variance at least V, since it also is a linear combination of the estimators from which A and B were constructed:

$$V \leqslant \mathbf{Var}\left[\frac{1}{2}(A+B)\right],$$
$$V \leqslant \frac{1}{4}\mathbf{Var}\,A + \frac{1}{4}\mathbf{Var}\,B + \frac{1}{2}\mathbf{Cov}(A,B),$$
$$V \leqslant \frac{1}{2}V + \frac{1}{2}\mathbf{Cov}(A,B),$$
$$V \leqslant \mathbf{Cov}(A,B).$$

Since $\mathbf{Cov}(A,B) \leqslant \sqrt{\mathbf{Var}\,A\,\mathbf{Var}\,B} = V$, then $\mathbf{Cov}(A,B) = V$ and $A = B$. Therefore, there exists a unique minimum variance unbiased estimator and the process is well defined. □

2.2. The algorithm for calculating the estimator

We present an efficient algorithm to calculate $X_v(t)$. Let $E_v(t) = \{X_w(t) \mid w \in N(v)\}$ be the estimators of agent v's neighbors at time t. Let $\mathbf{C}$ be the covariance matrix of $E_v(t)$, so that $C_{wu} = \mathbf{Cov}(X_w(t), X_u(t))$. For each w, let x_w be a realization of $X_w(t)$. Then the log likelihood of $\mu \in \mathbb{R}$ is

$$\log \mathcal{L}(\mu) = -\frac{1}{2}\sum_{wu}(x_w - \mu)C^{-1}_{wu}(x_u - \mu) + \text{const},$$

where $\mathbf{C}^{-1}$ is $\mathbf{C}$'s pseudo-inverse. This expression is maximal for

$$\mu = \frac{\sum_{wu} C^{-1}_{wu} x_w}{\sum_{wu} C^{-1}_{wu}}.$$

Hence, the MLE, and therefore also $X_w(t+1)$, equals

$$X_{ML} = X_v(t+1) = \frac{\sum_{w,u \in N(v)} C^{-1}_{wu} X_w(t)}{\sum_{w,u \in N(v)} C^{-1}_{wu}}. \tag{4}$$

Note that $X_v(t+1)$ is also, among all the unbiased estimators of μ constructible over the estimators in $E_V(t)$, the one with the minimum variance.

Given this last observation, there exists a simple geometric interpretation for (4):

Proposition 2.2. *It is possible to calculate $\{X_v(t+1) \mid v \in V\}$, given $\{X_v(t) \mid v \in V\}$, by performing n operations of finding the point of an n-dimensional affine space (as specified by a generating set of size at most n) with minimal L_2 norm.*

Proof. Consider an n-dimensional vector space $\mathcal{V}$ over $\mathbb{R}$, with an inner product $\langle\cdot,\cdot\rangle$. Let $\mathbf{z}$ be some non-zero vector in $\mathcal{V}$, and let $\mathcal{A} \subset \mathcal{V}$ be the affine space defined by $\mathcal{A} = \{\mathbf{x} \in \mathcal{V} \mid \langle \mathbf{x}, \mathbf{z}\rangle = 1\}$.

(This is a generalization of $\mathcal{V} = \text{span}(\{X_v \mid v \in V\})$, $\langle X, Y\rangle = \mathbf{Cov}(X,Y)$, $\mathbf{z} = \sum_{v \in V} X_v$ and $\mathcal{A}$ being the set of unbiased estimators.)

Given a set of vectors $E = \{\mathbf{x}_k \mid k = 1, \ldots, K \leqslant n\}$, where $\mathbf{x}_k \in \mathcal{A}$, let $C_{kl} = \langle \mathbf{x}_k, \mathbf{x}_l\rangle$. Then $\mathcal{A} \cap \text{span}(E)$ is also an affine space, and the minimum L_2 norm vector in $\mathcal{A} \cap \text{span}(E)$ is

$$\mathbf{x}_{ML} = \frac{\sum_{kl} C^{-1}_{kl} \mathbf{x}_k}{\sum_{kl} C^{-1}_{kl}},$$

where $\mathbf{C}^{-1}$ is the matrix pseudo-inverse of $\mathbf{C}$. This equation is identical to (4). □

Note that if $\mathbf{C}$ is invertible then its pseudo-inverse is equal to its inverse. Otherwise, there are many linear combinations of the vectors in E which are equal to the unique $\mathbf{x}_{ML}$. The rôle of the pseudo-inverse is to facilitate computation: it provides the linear combination with least sum of squares of the coefficients [7].

2.3. Convergence

Denote $V_v(t) := \mathbf{Var}[X_v(t)]$ and $C_{vw}(t) := \mathbf{Cov}(X_v(t), X_w(t))$.

Lemma 2.3. *All estimators have the same limiting variance*: $\exists \rho_\infty\ \forall v$: $V_v(t) \to \rho_\infty$.

Proof. Since, in every iteration, each agent calculates the minimum variance unbiased estimator over those of his neighbors and its own, then the variance of the estimator it calculates must be lower than that of any other unbiased linear combination:

$$\sum_{w \in N(v)} \alpha_w = 1 \quad \Rightarrow \quad V_v(t+1) \leqslant \mathbf{Var}\Bigg[\sum_{w \in N(v)} \alpha_w X_w(t)\Bigg]. \tag{5}$$

In particular, for each neighbor w of v

$$V_v(t+1) \leqslant V_w(t), \tag{6}$$

and since each vertex is its own neighbor, then

$$V_v(t+1) \leqslant V_v(t). \tag{7}$$

Therefore, since the variance of each agent's estimator is monotonously decreasing (and positive), it must converge to some ρ_v. Now assume $(v, w) \in E$ and $\rho_v < \rho_w$, then, at some iteration t, $V_v(t) < \rho_w \leqslant V_w(t)$. But then, by (6), we have $V_w(t+1) \leqslant V_v(t) < \rho_w$ – a contradiction. Therefore, ρ_v must equal ρ_w, and since the graph is connected, all agents must converge to the same variance, ρ_∞. □

Lemma 2.4. $\forall v, w$: $C_{vw}(t) \to \rho_\infty$.

Proof. The previous lemma is a special case of this one, for $v = w$. Otherwise, for two neighboring agents v and w, for any ϵ, there exists an iteration t where both $\mathbf{Var}[X_v(t)] < \rho_\infty + \epsilon$ and $\mathbf{Var}[X_w(t)] < \rho_\infty + \epsilon$. Then:

$$\begin{aligned}
&\mathbf{Var}\left[\frac{1}{2}\big(X_v(t) + X_w(t)\big)\right] \\
&\quad = \frac{1}{4}\big[V_v(t) + V_w(t) + 2C_{vw}(t)\big] \\
&\quad < \frac{1}{2}\big[\rho_\infty + \epsilon + C_{vw}(t)\big]
\end{aligned}$$

and since (5) implies $V_v(t+1) \leqslant \mathbf{Var}[\frac{1}{2}(X_v(t)+X_w(t))]$, then $\rho_\infty < \frac{1}{2}[\rho_\infty + \epsilon + C_{vw}(t)]$ and $C_{vw}(t) \geqslant \rho_\infty - \epsilon$. Since C_{vw} is also bounded from above: $C_{vw}(t) \leqslant \sqrt{\mathbf{Var}[X_v(t)]\,\mathbf{Var}[X_w(t)]} < \rho_\infty + \epsilon$, we have demonstrated that $C_{vw}(t) \to \rho_\infty$ when v and w are neighbors. This implies that the correlation between neighbors converges to 1, and therefore, since the graph is finite, all correlations converge to 1 and all covariances converge to ρ_∞. □

This last lemma implies that if one agent's estimator converges, then all others' also converge, to the same limit. Even without convergence, however, it implies that all the estimators converge to their average:

$$\lim_{t\to\infty} \mathbf{Var}\left[X_v(t) - \frac{1}{|V|}\sum_{w\in V} X_w(t)\right] = 0. \tag{8}$$

The following lemma will be used to conjecture that all the estimators do converge. It states that an estimator is uncorrelated to the difference between it and any of the estimators which were used to calculate it.

Lemma 2.5. $\forall w \in N(v)$: $\mathbf{Cov}(X_v(t+1), X_v(t+1) - X_w(t)) = 0$.

Proof. We examine the estimators $\hat{X}(\beta) = X_v(t+1)(1-\beta) + X_w(t)\beta$, which are also unbiased estimators of μ, and are linear combinations of the estimators from which $X_v(t+1)$ was constructed. They should all therefore have higher variance than $X_v(t+1)$. Since $\hat{X}(\beta = 0) = X_v(t+1)$, then

$$0 = \left.\frac{\partial\,\mathbf{Var}[\hat{X}]}{\partial\beta}\right|_{\beta=0}.$$

Now:

$$\begin{aligned}
0 &= \left.\frac{\partial\,\mathbf{Var}[\hat{X}]}{\partial\beta}\right|_{\beta=0} \\
&= \left.\frac{\partial[(1-\beta)^2\,\mathbf{Var}[X_v(t+1)]]}{\partial\beta}\right|_{\beta=0} + \left.\frac{\partial[\beta^2\,\mathbf{Var}[X_w(t)]]}{\partial\beta}\right|_{\beta=0} \\
&\quad + \left.\frac{\partial[2\beta(1-\beta)\,\mathbf{Cov}(X_v(t+1), X_w(t))]}{\partial\beta}\right|_{\beta=0} \\
&= \big[-2(1-\beta)\,\mathbf{Var}\big[X_v(t+1)\big]\big]\big|_{\beta=0} + \big[2\beta\,\mathbf{Var}\big[X_w(t)\big]\big]\big|_{\beta=0} \\
&\quad + \big[2(1-2\beta)\,\mathbf{Cov}\big(X_v(t+1), X_w(t)\big)\big]\big|_{\beta=0} + \big[2(1-2\beta)\,\mathbf{Cov}\big(X_v(t+1), X_w(t)\big)\big]\big|_{\beta=0} \\
&= -2\,\mathbf{Var}\big[X_v(t+1)\big] + 2\,\mathbf{Cov}\big(X_v(t+1), X_w(t)\big) \\
&= -2\,\mathbf{Cov}\big(X_v(t+1), X_v(t+1) - X_w(t)\big)
\end{aligned}$$

and so $\mathbf{Cov}(X_v(t+1), X_v(t+1) - X_w(t)) = 0$. □

Note that this implies that $\mathbf{Var}[X_v(t+1)] = \mathbf{Cov}(X_v(t+1), X_w(t))$.

Conjecture 2.6 *(Convergence).* $\exists X(\infty)\ \forall v : X_v(t) \to X(\infty)$.

The following observation supports this conjecture:

$$\begin{aligned}
&\mathbf{Var}\big[X_v(t+1)-X_v(t)\big]\\
&\quad=\mathbf{Cov}\big(X_v(t+1)-X_v(t),X_v(t+1)-X_v(t)\big)\\
&\quad=\mathbf{Cov}\big(X_v(t+1),X_v(t+1)-X_v(t)\big)-\mathbf{Cov}\big(X_v(t),X_v(t+1)-X_v(t)\big).
\end{aligned}$$

Using Lemma 2.5

$$\begin{aligned}
&=-\mathbf{Cov}\big(X_v(t),X_v(t+1)-X_v(t)\big)\\
&=V_v(t)-\mathbf{Cov}\big(X_v(t+1),X_v(t)\big),
\end{aligned}$$

and using it again:

$$=V_v(t)-V_v(t+1).$$

This implies that if t_0 is such that $V_v(t_0)=\rho_\infty+\epsilon$ and therefore $\sum_{t=t_0}^{\infty}V_v(t)-V_v(t+1)=\epsilon$, then

$$\sum_{t=t_0}^{\infty}\mathbf{Var}\big[X_v(t+1)-X_v(t)\big]=\epsilon.$$

2.4. *Efficiency for transitive graphs*

Vertex transitive graphs (henceforth referred to as transitive graphs), are graphs where all vertices are essentially equivalent, or "equally important". Alternatively, one may say that the graph "looks the same" from all vertices. Formally, $G=(V,E)$ is transitive iff, for every pair of vertices $v,w\in V$ there exists a function $f:V\to V$ which is a graph automorphism (i.e. f is a bijection and $(a,b)\in E\Leftrightarrow(f(a),f(b))\in E$) and maps v to w.

Proposition 2.7. *When G is transitive then the process converges and $X(\infty)=Z(\infty)$.*

Proof. By the symmetry of the graph, the average of the agents' estimators cannot give more weight to one agent's original estimator than to another:

$$\frac{1}{|V|}\sum_v X_v(t)=\frac{1}{|V|}\sum_v X_v=Z(\infty),$$

and hence the average of the agents' estimators is constant and in particular converges. By Lemma 2.4, (8), if the average converges then each of the estimators converges to the same limit:

$$\forall v\lim_{t\to\infty}X_v(t)=\lim_{t\to\infty}\frac{1}{|V|}\sum_v X_v(t)=Z(\infty)=X(\infty).\qquad\square$$

Note that for regular graphs (i.e. graphs where all vertices have the same degree), which are a superset of transitive graphs, $Y(\infty)=Z(\infty)$.

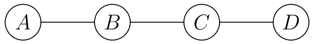

Fig. 1. Interval of length four.

3. Analytic examples

Complete analytical analysis of these iterations for general graphs seems difficult, since (4) is quadratic. In fact, we found only two simple examples amenable to complete analysis: the star, a graph with a central node connected to all other nodes, and the interval of length four, a graph of four linearly ordered nodes.

In the former, we show that the minimum variance scheme is efficient, so that $X(\infty) = Z(\infty)$. In the latter, we show that it isn't, but that the "price of anarchy" is low.

3.1. High degree graphs and the star

We consider a graph of n vertices, of which u is the central node and is connected to all others, and no additional edges exist.

The averaging estimator $Y(\infty)$ gives weight $\frac{n}{3n-2}$ to X_u and $\frac{2}{3n-2}$ to the rest. Its variance is $\frac{n^2+4n-4}{(3n-2)^2}$, which is asymptotically $\frac{1}{9}$.

On the other hand, $X_u(1) = Z(\infty)$, since node u, neighboring all nodes of the graph, immediately finds the global minimum variance estimator. In the next iteration, all nodes w set $X_w(2) = X_u(1)$, and the process essentially halts, since all nodes have the same estimator, $X_w(2) = X(\infty) = Z(\infty)$, with $\mathbf{Var}[X(\infty)] = \frac{1}{n}$.

In general, in graphs of large maximal degree, $X_v(t)$ converges to μ:

Proposition 3.1. *Let $G_n = (V_n, E_n)$ be a family of graphs where $|V_n| = n$ and $\max_{v\in V_n} d_v \to \infty$. Then*

$$\lim_{n\to\infty} \sup_{v\in V_n} \lim_{t\to\infty} \mathbf{E}\big[\big(X_v(t) - \mu\big)^2\big] = 0.$$

Proof. Since all estimators at all iterations have mean μ, then $\mathbf{E}[(X_v(t) - \mu)^2] = \mathbf{Var}[X_v(t)]$. By Lemma 2.3, the limiting variances of all the agents in a graph G_n are equal to some ρ_n, and therefore

$$\lim_{n\to\infty} \sup_{v\in V_n} \lim_{t\to\infty} \mathbf{E}\big[\big(X_v(t) - \mu\big)^2\big] = 0 \leftrightarrow \lim_{n\to\infty} \rho_n = 0.$$

The condition $\max_{v\in V_n} d_v \to \infty$ implies that given $\epsilon > 0$, there exists a high enough N, so that in any G_n with $n > N$ there exists a node w_n with degree d_{w_n} larger than $1/\epsilon$. Then $\mathbf{Var}[X_{w_n}(1)] < \epsilon$, since agent w_n would, on the first iteration, average the estimators of all its neighbors, resulting in a new estimator of variance $1/d_{w_n}$. Since variance never increases in the iterative process (Lemma 2.3), then $\rho_n < \epsilon$ for n larger then some N. Since this is true for arbitrary ϵ, ρ_n goes to zero as n goes to infinity. $\square$

3.2. Interval of length four

We analyze the case of $G = (V, E)$ where $V = \{a, b, c, d\}$ and $E = \{\{a, b\}, \{b, c\}, \{c, d\}\}$ (see Fig. 1).

In Appendix A, we prove that the process converges with

$$\mathbf{Var}\big[X(\infty)\big]$$
$$= 2 - \sqrt{3}$$

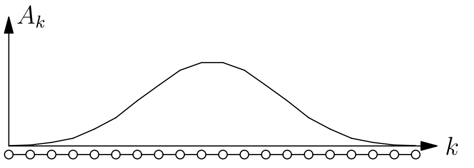

Fig. 2. Simulation of interval of length 20.

$$= \frac{1}{4}(1 + \sqrt{49} - \sqrt{48})$$
$$> \frac{1}{4} = \mathbf{Var}\big[Z(\infty)\big],$$

thus proving that a case exists where $X(\infty) \neq Z(\infty)$. We also, for this case, derive an asymptotic convergence rate of $2 - \sqrt{3}$.

The averaging estimator $Y(\infty)$ is:

$$Y(\infty) = 0.2X_a + 0.3X_b + 0.3X_c + 0.2X_d,$$

with

$$\mathbf{Var}\big[Y(\infty)\big] = 0.26.$$

This is slightly lower than $\mathbf{Var}[X(\infty)]$, which equals about 0.268.

The averaging process, being linear, converges to it first eigenvector, with eigenvalue one. Its second eigenvalue determines its convergence rate: the process will converge exponentially, with rate equal to the second eigenvalue.

In this case the convergence rate for the averaging process is $\frac{1}{4} + \sqrt{33}/12 \approx 0.73$, which is significantly slower than the minimum variance process's rate of $2 - \sqrt{3} \approx 0.268$.

4. Numerical examples and conjectures

Numerical simulations on intervals of lengths larger than four suggest a surprising result.

4.1. Intervals of arbitrary length

Numerical simulations suggest that $X(\infty)$, for intervals of length n, approaches a normal distribution around the center of the interval, with variance proportional to n (see Fig. 2):

Conjecture 4.1. *For interval graphs of length* $2n$*, index the agents by* $k \in \{0, \dots, 2n-1\}$*. Then*

$$X(\infty) = \sum_k A_k X_k$$

where A_k *approaches a normal distribution in the sense that*

$$\lim_{n\to\infty} \sum_k \big(A_k - C_n e^{-(k-n+1/2)^2/\nu(n)}\big)^2 = 0 \quad \text{with } \nu(n) \in \Theta(n),\ C_n \in \mathbb{R}.$$

This implies that while $\lim_{n\to\infty} \mathbf{E}[(X(\infty) - \mu)^2] = 0$, $X(\infty)$ quickly becomes less efficient when compared to $Z(\infty)$:

Conjecture 4.2. $\mathbf{Var}[X(\infty)] \propto \sqrt{n}\,\mathbf{Var}[Z(\infty)]$.

Note that $\mathbf{Var}[Y(\infty)]$, on the other hand, approaches $\mathbf{Var}[Z(\infty)]$ as n increases, for intervals of length n.

4.2. Agents with memory

A model which is perhaps more natural than the memoryless model is the model in which the agents remember all their own values from the previous iterations.

Proposition 4.3. *If the agents have memory, then the process converges to* $Z(\infty)$.

Proof. Since each vertex v always remembers $X_v = X_v(0)$, then X_v is always part of the set over which $X_v(t)$ was constructed. Then, by Lemma 2.5:

$$\mathbf{Cov}\big(X_v(t), X_v\big) = \mathbf{Var}\big[X_v(t)\big], \tag{9}$$

and by Lemma 2.4, $\forall v, w \in V$:

$$\begin{aligned}&\lim_{t\to\infty} \mathbf{Cov}\big(X_w(t), X_v\big)\\ &\quad= \lim_{t\to\infty} \mathbf{Cov}\big(X_v(t), X_v\big)\\ &\quad= \lim_{t\to\infty} \mathbf{Var}\big[X_v(t)\big] = \rho_\infty.\end{aligned}$$

This means that for any agent w, the covariance of its limit estimator with each of the original estimators X_v is identical, and so it must be their average: $X(\infty) = Z(\infty)$. □

This proof relied only on the agents' memory of their original estimators. Since they also gain more estimators over the iterations, and seemingly expand the space that they span, we conjecture that:

Conjecture 4.4. $\forall v \in V\colon X_v(t) = X(\infty)$, *for* $t \geqslant |V|$.

5. Conclusion

An number of interesting open problems can be raised with respect to this model, some of which we conjecture about above:

- Does it always converge? We conjecture above that this is indeed the case.
- For what graphs does it converge to the optimal estimator $Z(\infty)$?
- Otherwise, what is the "price of anarchy", $\mathbf{Var}[X(\infty)]/\mathbf{Var}[Z(\infty)]$? Is it bounded? We conjecture above that it isn't.
- What is the convergence rate?

Appendix A. Analysis of interval of length four

Theorem. *Let* $G = (V, E)$ *be the interval of length four where* $V = \{a, b, c, d\}$ *and* $E = \{\{a, b\}, \{b, c\}, \{c, d\}\}$. *Then* $X_v(t)$ *converges to a limit* $X(\infty)$, *where*

$$X(\infty) = \frac{1}{4}\big[(1-\xi)(X_a + X_d) + (1+\xi)(X_b + X_c)\big],$$
$$\mathbf{Var}\big[X(\infty)\big] = \xi,$$

with $\xi = 2 - \sqrt{3} = \frac{1}{4}(1 + \sqrt{49} - \sqrt{48})$.

Proof. We define $M_{vw}(t) = \mathbf{Cov}(X_v, X_w(t))$, so that each column of $\mathbf{M}$ is the coordinates of an agent's estimator at time t, viewed as a vector in the space spanned by $\{X_a, X_b, X_c, X_d\}$. We define $Z(\infty)$-subtracted M as $\tilde{M}_{vw}(t) = \mathbf{Cov}(X_v - Z(\infty), X_w(t) - Z(\infty))$, where $Z(\infty) = \frac{1}{4}(X_a + X_b + X_c + X_d)$, and likewise define the $Z(\infty)$-subtracted covariance matrix $\tilde{C}_{vw}(t) = \mathbf{Cov}(X_v(t) - Z(\infty), X_w(t) - Z(\infty))$.

We now shift to an alternative orthonormal basis $B = \{b_1, b_2, b_3, b_4\}$:

$$b_1 = \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{pmatrix}, \qquad b_2 = \begin{pmatrix} -1/\sqrt{2} \\ 0 \\ 0 \\ 1/\sqrt{2} \end{pmatrix},$$
$$b_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}, \qquad b_4 = \begin{pmatrix} -1/2 \\ 1/2 \\ 1/2 \\ -1/2 \end{pmatrix}.$$

The vector $b_1 (= 2Z(\infty))$ was chosen because its coordinate is one half in every unbiased estimator, and we know that if $X(\infty)$ exists then it is unbiased. b_2 and b_3 are anti-symmetric to inversion of the interval, a transformation which should leave $X(\infty)$ invariant by the symmetry of the graph. Therefore we expect their coordinates in $X(\infty)$ to vanish. We have no freedom, then, in choosing the last vector, and expect $X(\infty)$ to equal $\frac{1}{2}b_1$ plus some constant ξ times $\frac{1}{2}b_4$:

$$X(\infty) = \frac{1}{4}\big[(1-\xi)(X_a + X_d) + (1+\xi)(X_b + X_c)\big].$$

We omit here the calculation of the first two iterations of the process: twice, each agent declares its estimator to its neighbors and recalculates a new ML estimator based on theirs. After two iterations, under this basis (B), the $Z(\infty)$-subtracted coordinates matrix of the estimators, $\tilde{\mathbf{M}}_B(2)$, is of the form:

$$\tilde{\mathbf{M}}_B(2) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ x & 0 & 0 & -x \\ 0 & z & -z & 0 \\ y & w & w & y \end{pmatrix},$$

with

$$yw = z^2 + w^2, \tag{10}$$

as the investigative reader may verify. Note that relation (10), which can be reproduced by following through with the calculation of the first two iterations, can also be derived from the fact that agents a and d merely copy b and c's (respectively) estimators at each iteration (and Lemma 2.5).

Application of another iteration yields a matrix of the same form:

$$\tilde{\mathbf{M}}_B(3) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{-xz}{x^2+(y-w)^2}z & \frac{xz}{x^2+(y-w)^2}z & 0 \\ z & 0 & 0 & -z \\ w & \frac{x^2}{x^2+(y-w)^2}w & \frac{x^2}{x^2+(y-w)^2}w & w \end{pmatrix}, \tag{11}$$

with the relation of (10) preserved.

Since the result is a matrix of essentially the same form, equivalent equations apply for consecutive iterations, and we may denote as x_t, y_t, w_t and z_t the corresponding matrix entries at time t. So for even t:

$$\tilde{\mathbf{M}}_B(t) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ x_t & 0 & 0 & -x_t \\ 0 & z_t & -z_t & 0 \\ y_t & w_t & w_t & y_t \end{pmatrix},$$

and for odd t

$$\tilde{\mathbf{M}}_B(t) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & z_t & -z_t & 0 \\ x_t & 0 & 0 & x_t \\ y_t & w_t & w_t & y_t \end{pmatrix}.$$

Since (11) implies that $y_t = w_{t-1}$ and $x_t = z_{t-1}$, then if w_t and z_t converge then the process converges and $X(\infty)$ exists.

The maximum likelihood calculation performed by the agents at each iteration (11) yields

$$w_{t+1} = \frac{z_{t-1}^2}{z_{t-1}^2 + (w_{t-1} - w_t)^2} w_t \tag{12}$$

and

$$z_{t+1} = \frac{z_{t-1} z_t}{z_{t-1}^2 + (w_{t-1} - w_t)^2} z_t. \tag{13}$$

Dividing (12) by (13), we discover that:

$$\frac{w_{t+1}}{z_{t+1}} = \frac{z_{t-1}}{z_t} \frac{w_t}{z_t},$$

and therefore, by repeated application:

$$w_t = \frac{w_2 z_2}{z_3} \frac{z_t}{z_{t-1}} = \frac{1}{2} \frac{z_t}{z_{t-1}}.$$

(10) can alternatively be written as: $w_{t-1} w_t = w_t^2 + z_t^2$. Then:

$$w_{t-1} - w_t = z_t^2 / w_t = 2 z_t z_{t-1},$$

and we can write (13) as:

$$z_{t+1} = \frac{z_{t-1} z_t}{z_{t-1}^2 + 4 z_t^2 z_{t-1}^2} z_t$$

or

$$\frac{z_t}{z_{t+1}} = \frac{z_{t-1}}{z_t}\left(1 + 4z_t^2\right). \tag{14}$$

To solve this recursion we make the following guess:

$$\frac{z_t}{z_{t+1}} = 2 + \sqrt{3 - 4z_t^2}, \tag{15}$$

which is a solution of the following quadratic equation in z_t/z_{t+1}:

$$\frac{z_t^2}{z_{t+1}^2} - 4\frac{z_t}{z_{t+1}} + 1 + 4z_t^2 = 0.$$

This is equivalent to the following relation:

$$\mathbf{Var}\left[X_b(t)\right] = w_t^2 + z_t^2 + \frac{1}{4} = 2w_t,$$

upon which we serendipitously stumbled during our examination of this problem.

This guess satisfies (14), as some manipulation of the two equations will show. Since z_2 and z_3 satisfy (15), then the rest of the z's must, too.

Since (14) implies $z_t \to 0$, we can conclude from $\frac{z_{t-1}}{z_t} = 2 + \sqrt{3 - 4z_{t-1}^2}$ that

$$\lim_{t\to\infty} \frac{z_{t+1}}{z_t} = 2 - \sqrt{3} := \xi.$$

This is the process's asymptotic convergence rate. Since $w_t = \frac{1}{2}\frac{z_t}{z_{t-1}}$, then $w_t \to \frac{1}{2}\xi$, and

$$X(\infty) = \frac{1}{4}\left[(1-\xi)(X_a + X_d) + (1+\xi)(X_b + X_c)\right],$$

with

$$\mathbf{Var}\left[X(\infty)\right] = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\xi\right)^2 = \xi. \qquad \square$$

References

[1] S. Aaronson, G. Kuperberg, C. Granade, The complexity zoo, at http://qwiki.stanford.edu/wiki/Complexity_Zoo.
[2] H.M. Ammari, S.K. Das, Integrated coverage and connectivity in wireless sensor networks: A two-dimensional percolation problem, IEEE Trans. Comput. 57 (10) (2008) 1423–1434.
[3] A.V. Banerjee, A simple model of herd behavior, Quart. J. Econ. 107 (3) (1992) 797–817.
[4] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[5] M. Jackson, The economics of social networks, in: R. Blundell, W. Newey, T. Persson (Eds.), Advances in Economics and Econometrics, vol. I, Theory and Applications: Ninth World Congress of the Econometric Society, Cambridge University Press, 2006, pp. 1–56.
[6] E. Mossel, O. Tamuz, Iterative maximum likelihood on networks, in: Proc. of Allerton, 2009, in press.
[7] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes in C, Cambridge University Press, 1992.
[8] L. Smith, P. Sorensen, Pathological outcomes of observational learning, Econometrica 68 (2) (2000) 371–398.
[9] D.H. Sushil Bikhchandani, I. Welch, Learning from the behavior of others: Conformity, fads, and informational cascades, J. Econ. Perspect. 12 (3) (2008) 151–170.
[10] H. Zhang, J.C. Hou, Asymptotic critical total power for k-connectivity of wireless networks, IEEE/ACM Trans. Netw. 16 (2) (2008) 347–358.