

IMA Workshop, University of Minnesota

**DISTRIBUTED HYPOTHESIS TESTING
IN NETWORKS**

Angelia Nedić

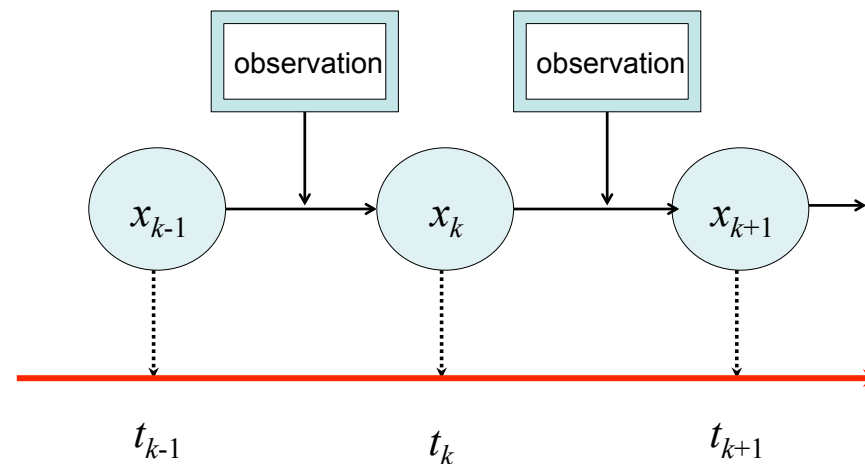
Collaborative work with **Alexander Olshevsky and Cesar Uribe**

Industrial and Enterprise Systems Engineering Department
and Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

October 2, 2015

Learning/Decision Making

Decision Making with Learning can be thought of a process through which we refine our understanding of a "state" of certain phenomena of interest:

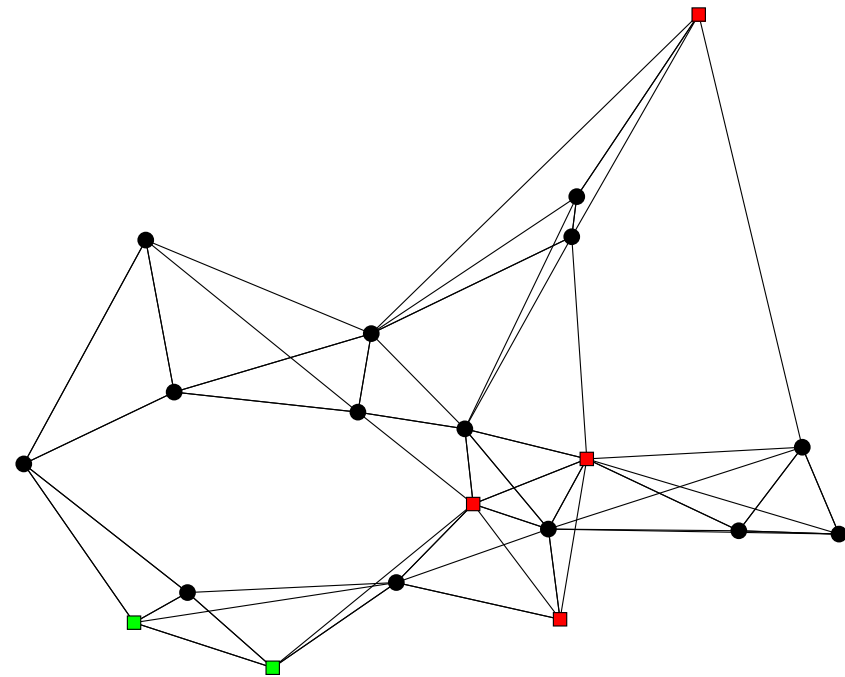


Our focus will be on Bayesian "Decision Process" - Hypothesis Testing.

Decision Process in Networks

The process takes place in a system of networked agents with different learning capabilities.

- A system of interconnected agents observes a "phenomenon"
- Each agent has "learning" capability
- Each agent has its own observations of the phenomenon
- Agents are willing to share their knowledge
- If the agents collaborate (share their knowledge), can they learn better, faster?
- If so, can we quantify how much better, how much faster?



We answer some of those questions for a system of Bayesian "learners"

Related Literature

Spread from social and economic networks to wired/wireless sensor networks

- **Learning and opinion aggregation in social networks:** Gale and Kariv 2003, Epstein et al. 2010, Acemoglu et al. 2011, Jadbabaie et al. 2012, 2013, Mueller-Frank 2013, **Shahrampour et al. 2014, Lalitha et al. 2014, 2015, N. O. Uribe 2014, 2015,** Sarwate and Javidi 2015
- **Detection/estimation in sensor networks:** Olfati-Saber et al. 2006, Saligrama et al. 2006, Rahnema-Rad and Tahbaz-Salehi 2010, Barbarossa et al. 2013, Battisteli and Chisci 2014
- **Consensus literature:** DeGroot 1974, Aumann 1976, Borkar and Varaiya 1982, Tsitisklis 1984, Jadbabaie, Lin, and Morse 2003 ...

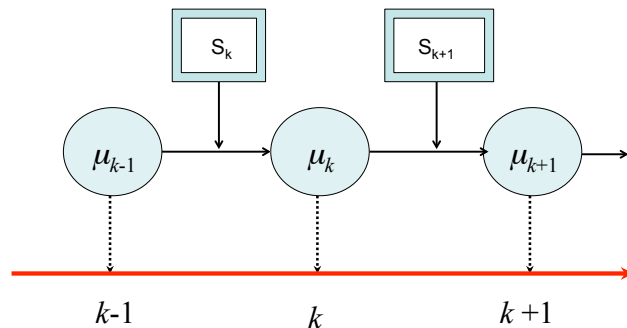
Bayesian Learning: Single Agent

An agent observes a certain phenomenon, and uses the observations to refine its understanding of the state of the phenomenon.

Agent receives observations s_k of the true state, which are random and form an i.i.d. process with a distribution f over a finite set S of the possible states

Agent learning model: a set of probability distributions* $\ell(\cdot|\theta)$ on S parametrized by θ , where θ takes a finite number of possible values: **set of hypotheses** $\Theta = \{\theta_1, \dots, \theta_m\}$.

At time k , agent has a belief μ_k (probability distribution on Θ) that the best explains the observations s_1, \dots, s_k collected up to that time. At time $k + 1$, it observes s_{k+1} and updates its belief to μ_{k+1} :



Bayesian update: for all $\theta \in \Theta$,

$$\mu_{k+1}(\theta) = \frac{\mu_k(\theta)\ell(s_{k+1}|\theta)}{\sum_{p=1}^m \mu_k(\theta_p)\ell(s_{k+1}|\theta_p)}$$

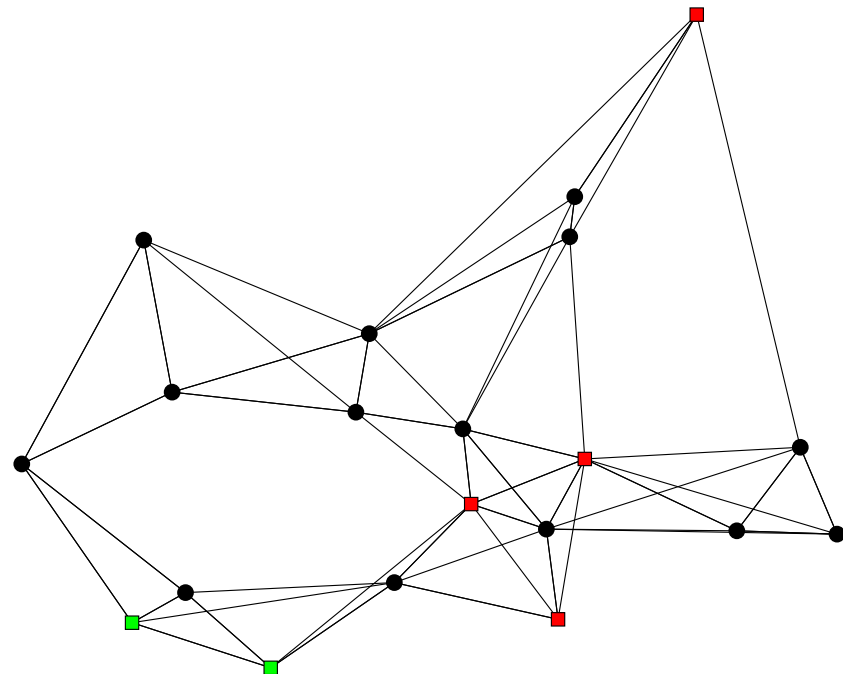
$$\mu_{k+1} = BU(\mu_k, s_{k+1})$$

*Likelihood functions

Learning on Graphs

Learning in a system of networked agents with different learning capabilities.

- A system of interconnected agents observes a "phenomenon"
- Each agent has "learning" capability and maintains a belief about the true state of the phenomenon
- Agents receive noisy private observation of the phenomenon and use them to update their beliefs
- Agents also talk to their neighbors and exchange their beliefs, which can also be used to update the beliefs
- How should agents aggregate their beliefs so that every agent learns the true state?



Agents can do better by collaborating

- Nature chooses a coin which could be biased or not
- Over time, the coin is tossed and the outcomes can be observed (either H or T)
- Assume the toss outcomes are iid
 - Coin model: let $\theta = 1$ if the coin is biased and $\theta = 0$ otherwise
 - The outcome of the k th toss is $s_k \in \{H, T\}$.
- Two agents observe the toss outcomes and want to learn if the coin is biased or not.
- Both agents have some prior belief $\mu_1(0)$ and $\mu_2(0)$ about the biased-ness of the coin (prob. distributions on θ -values).
- Agent learning models captured through likelihoods:
$$\begin{aligned} \ell^1(\cdot \mid \theta = 1) &= [0.5, 0.5], & \ell^1(\cdot \mid \theta = 0) &= [0.55, 0.45] \\ \ell^2(\cdot \mid \theta = 1) &= [0.55, 0.45], & \ell^2(\cdot \mid \theta = 0) &= [0.5, 0.5] \end{aligned}$$
- Independently (by Bayesian updates), none of them can learn on its own
- Collaboratively, both can learn no matter what coin the nature chooses

Known:

- A fully Bayesian approach may not be possible in general since full knowledge of neither the network structure nor other agents likelihood functions might be available; Gale and Kariv 2003.
- Non-Bayesian methods have been shown in the literature to be successful in learning; Epstein, Noor, and Sandroni 2010
- **An aggregation method for local Bayes estimates leading to a global learning in work of Jadbabaie, Molavi, Sandroni, and Tahbaz-Salehi 2012, Jadbabaie, Molavi, and Tahbaz-Salehi 2013, Bandyopadhyay and Chung 2014**

We will start with a slightly more general model akin to that of Jadbabaie et al. 2012

Learning Model

- There are n agents and all have the same set of hypotheses: $\theta \in \Theta \triangleq \{\theta_1, \dots, \theta_m\}$
- At each time k , agent i has a belief μ_k^i - probability distribution on the hypotheses
- Each agent i receives a noisy private observation s_k^i , $s_k^i \in \{s_1^i, \dots, s_{m_i}^i\}$
- We assume that the observations $\{s_k^i\}$ are iid for each agent, and independent across agents
- Each agent i has likelihood functions (probability distributions) capturing the agent learning capability: $\ell^i(\cdot | \theta)$ for all $\theta \in \Theta$
- Agents also exchange beliefs with neighbors and use them when updating their beliefs
- The neighbor structure (static at the moment) is captured by a directed graph: $\mathcal{G} = (V, E)$, where E is the set of directed edges, and
$$(i, j) \in E \text{ if agent } i \text{ receives information from } j$$
- Upon observing the outcome s_{k+1}^i , agent i updates its belief based the prior belief μ_k^i , the outcome, and its likelihood functions

An update has been proposed and analyzed in Jadbabaie et al. 2012 of the following form:

$$\mu_{k+1}^i = A_{ii}BU(\mu_k^i, s_{k+1}^i) + \sum_{j \in \mathcal{N}_i} A_{ij}\mu_k^j,$$

where $\mathcal{N}_i = \{j \in [n] \mid (j, i) \in E\}$, with a row-stochastic matrix A satisfying $A_{ij} > 0$ for $(j, i) \in E$.

- The update makes sense, since it preserves "probability distributions".
- Nice results were shown (convergence with probability 1 to the true state of "world").

- **What motivates this update rule?**

- **Why not**

$$\mu_{k+1}^i = BU(\hat{\mu}_k^i, s_{k+1}^i) \quad \text{with} \quad \hat{\mu}_k^i = \sum_{j \in \mathcal{N}_i \cup \{i\}} A_{ij}\mu_k^j \quad ?$$

- **Can the update rule be related to some network optimization objective?**

Agent Update: Inference Via Minimization Rule

- *When learning alone*: Standard Bayesian update - agent i updates minimize a cost function which is composed of two terms (Walker 2006)
 - Maximum Likelihood Estimation (MLE) of a state given the observed data and
 - A regularization function KL divergence from the current prior

$$\mu_{k+1}^i = \operatorname{argmin}_{\pi \in \Pi(\Theta)} \left\{ \underbrace{-\mathbb{E}_{\pi} [\log \ell^i(s_{k+1}^i | \cdot)]}_{MLE} + D_{KL}(\pi \| \mu_k^i) \right\}$$

where $\Pi(\Theta)$ is the set of all probability distributions over the set Θ

- *When learning with neighbors*:

$$\mu_{k+1}^i = \operatorname{argmin}_{\pi \in \Pi(\Theta)} \left\{ \underbrace{-\mathbb{E}_{\pi} [\log \ell^i(s_{k+1}^i | \cdot)]}_{MLE} + \sum_{j=1}^n A_{ij} D_{KL}(\pi \| \mu_k^j) \right\}$$

with a row-stochastic matrix A compatible with the graph G : $A_{ij} \geq 0$ when $(j, i) \in E$, and $A_{ii} > 0$.

- There is an explicit expression for μ_{k+1}^i (unique minimizer)

Learning with Neighbors

$$\mu_{k+1}^i(\theta) = \frac{\prod_{j=1}^n [\mu_k^j(\theta)]^{A_{ij}} \ell^i(s_{k+1}^i | \theta)}{\sum_{r=1}^m \prod_{j=1}^n [\mu_k^j(\theta)]^{A_{ij}} \ell^i(s_{k+1}^i | \theta_r)} \quad \text{for all } \theta \in \Theta$$

The minimizer is given by

$$\mu_{k+1}^i(\theta) = \frac{\prod_{j=1}^n [\mu_k^j(\theta)]^{A_{ij}} \ell^i(s_{k+1}^i | \theta)}{\sum_{r=1}^m \prod_{j=1}^n [\mu_k^j(\theta)]^{A_{ij}} \ell^i(s_{k+1}^i | \theta_r)} \quad \text{for all } \theta \in \Theta$$

Proof: Consider Lagrangian function associated with the problem (constraint $\pi' \mathbf{1} = 1$ is penalized)

$$-\sum_{r=1}^m \pi_r \log(y_r) + \sum_{j=1}^n A_{ij} \sum_{r=1}^m \pi_r \log\left(\frac{\pi_r}{z_r^j}\right) + \lambda \left(\sum_{r=1}^m \pi_r - 1\right)$$

Set the gradient to zero

$$-\log(y_r) + \sum_{j=1}^n A_{ij} \log \frac{\pi_r}{z_r^j} + 1 + \lambda = 0 \quad \text{for all } r = 1, \dots, m$$

$$-\log(y_r) + \log(\pi_r) - \log\left(\prod_{j=1}^n (z_r^j)^{A_{ij}}\right) + c = 0$$

$$\log \left(\frac{\pi_r}{y_r \prod_{j=1}^m (z_r^j)^{A_{ij}}} \right) = -c$$

$$\pi_r = y_r \prod_{j=1}^m (z_r^j)^{A_{ij}} C \quad \text{for all } r = 1, \dots, m$$

with $C > 0$, which is just a normalization constant ensuring that $\pi \in \Pi(\theta)$ with $y_r = \ell^i(s_{k+1}^i | \theta)$ and $z_r^j = \mu_k^j(\theta)$ we obtain

$$\mu_{k+1}^i(\theta) = \frac{\prod_{j=1}^n [\mu_k^j(\theta)]^{A_{ij}} \ell^i(s_{k+1}^i | \theta)}{\sum_{r=1}^m \prod_{j=1}^n [\mu_k^j(\theta)]^{A_{ij}} \ell^i(s_{k+1}^i | \theta_r)} \quad \text{for all } \theta \in \Theta$$

- If $\mathcal{N}_i = \emptyset$ (corresponding to $A_{ii} = 1$) the preceding rule gives the standard Bayesian update

$$\mu_{k+1}^i(\theta) = \frac{\mu_k^i(\theta) \ell^i(s_{k+1}^i | \theta)}{\sum_{r=1}^m \mu_k^i(\theta) \ell^i(s_{k+1}^i | \theta_r)} \quad \text{for all } \theta \in \Theta$$

Belief Update

The preceding belief update can be related to opinion aggregation, also referred to as "opinion pooling". A general form of opinion pooling (Gilardoni and Clayton 1993):

$$b_{k+1}^i = BU(\mu_k^i, \ell(s_{k+1}^i | \cdot))$$

Mixing based on g -Opinion pools

$$\mu_{k+1}^i(\theta) = \frac{g^{-1} \left(\sum_{j=1}^n A_{ij} g(b_{k+1}^j(\theta)) \right)}{\sum_{r=1}^m g^{-1} \left(\sum_{j=1}^n A_{ij} g(b_{k+1}^j(\theta_r)) \right)} \quad \text{for all } \theta \in \Theta$$

- It is shown by Gilardoni and Clayton 1993 that a consensus can be reached only for $g(x) = x$ or $g(x) = \log(x)$.

Our rule is Logarithmic pool: $g(x) = \log(x)$

$$\mu_{k+1}^i(\theta) = \frac{\prod_{j=1}^n [\mu_k^j(\theta)]^{A_{ij}} \ell^i(s_{k+1}^i | \theta)}{\sum_{r=1}^m \prod_{j=1}^n [\mu_k^j(\theta_r)]^{A_{ij}} \ell^i(s_{k+1}^i | \theta_r)} \quad \text{for all } \theta \in \Theta$$

Recent work by Lalitha et. al 2014 also uses $g(x) = \log(x)$, while the rule used in Jadbabaie et al. 2012 is more related to Linear pool, i.e., $g(x) = x$.

Learning Objective

$$\mu_{k+1}^i = \operatorname{argmin}_{\pi \in \Pi(\Theta)} \left\{ \underbrace{-\mathbb{E}_{\pi} [\log \ell^i(s_{k+1}^i | \cdot)]}_{MLE} + D_{KL}(\pi \| \mu_k^i) \right\}$$

Can be interpreted as gradient method for minimizing

$$D_{KL}(f_i \| \ell^i(\cdot | \theta)) \text{ over } \theta \in \Theta,$$

with $D_{KL}(p \| q) = \sum_{i=1}^d p_i \log \left(\frac{p_i}{q_i} \right)$

- **Learning Objective:** The update corresponds to determine a hypothesis $\theta \in \Theta$ that explains best the agent system data

$$\min_{\theta \in \Theta} \sum_{i=1}^n D_{KL}(f_i \| \ell^i(\cdot | \theta))$$

- Let

$$\Theta_i^* = \operatorname{argmin}_{\theta \in \Theta} D_{KL}(f_i \| \ell^i(\cdot | \theta)) \quad \text{and} \quad \Theta^* = \bigcap_{i=1}^n \Theta_i^*$$

- **Consistency** All agents should agree on the hypotheses in Θ^* that best describes the observations for all.
- **Definition** The agents beliefs are said to provide a distributed consistent estimator if with probability 1: for all agents i ,

$$\lim_{k \rightarrow \infty} \mu_k^i(\theta) = 0 \quad \text{for all } \theta \notin \Theta^*$$

Assumptions: Graphs, Likelihoods and Initial Beliefs

Assumption (Graphs) Graph sequence $\{\mathcal{G}_k\}$ and a matrix sequence $\{A_k\}$ are such that:

1. A_k is row-stochastic with $[A_k]_{ij} > 0$ if $(j, i) \in E_k$ and $[A_k]_{ii} > 0$.
2. If $[A_k]_{ij} > 0$ then $[A_k]_{ij} > \eta$ for some positive scalar η .
3. $\{\mathcal{G}_k\}$ is B -strongly connected, i.e., here is an integer $B \geq 1$ such that the graph $\left\{V, \bigcup_{i=kB}^{(k+1)B-1} E_i\right\}$ is strongly connected for all $k \geq 0$.

Assumption (Likelihood Models)

$$\Theta^* \triangleq \bigcap_{i=1}^n \operatorname{argmin}_{\theta \in \Theta} D_{KL}(f_i(\cdot) \parallel \ell^i(\cdot|\theta)) \text{ is nonempty}$$

Assumption (Initial Beliefs) For all agents $i = 1, \dots, n$,

1. The prior beliefs on all $\theta^* \in \Theta^*$ are positive, i.e. $\mu_0^i(\theta^*) > 0$ for all $\theta^* \in \Theta^*$.
2. There exists an $\alpha > 0$ such that $\ell^i(s^i|\theta) > \alpha$ for all outcomes s^i and all $\theta \in \Theta$.

Consistency

Proposition 1 Under these assumptions, the update rule

$$\mu_{k+1}^i = \operatorname{argmin}_{\pi \in \Pi(\Theta)} \left\{ -\mathbb{E}_{\pi} [\log \ell^i(s_{k+1}^i | \cdot)] + \sum_{j=1}^n [A_k]_{ij} D_{KL}(\pi \| \mu_k^j) \right\}$$

or explicitly

$$\mu_{k+1}^i(\theta) = \frac{\prod_{j=1}^n \mu_k^j(\theta)^{[A_k]_{ij}} \ell^i(s_{k+1}^i | \theta)}{\sum_{r=1}^m \prod_{j=1}^n \mu_k^j(\theta)^{[A_k]_{ij}} \ell^i(s_{k+1}^i | \theta_r)}$$

generates sequences $\{\mu_k^i\}$, $i = 1, \dots, m$, of beliefs such that with probability 1, for all agents i :

$$\lim_{k \rightarrow \infty} \mu_k^i(\theta) = 0 \quad \text{for all } \theta \notin \Theta^*$$

Proof: Choose a $\theta^* \in \Theta^*$ and define

$$\varphi_k^i(\theta) = \log \left(\frac{\mu_k^i(\theta)}{\mu_k^i(\theta^*)} \right)$$

From the update rule, we obtain for all $i = 1, \dots, n$

$$\varphi_{k+1}^i(\theta) = \sum_{j=1}^n [A_k]_{ij} \varphi_k^j(\theta) + \log \left(\frac{\ell^i(s_{k+1}^i | \theta)}{\ell^i(s_{k+1}^i | \theta^*)} \right) \quad \text{for all } \theta \in \Theta$$

Stacking φ_{k+1}^i , for $i = 1, \dots, n$ in a vector

$$\varphi_{k+1}(\theta) = A_k \varphi_k(\theta) + \mathcal{L}_k(\theta) \quad \text{for all } \theta \in \Theta$$

after some manipulations akin to "consensus"-type analysis, we find that, almost surely, coordinate-wise we have

$$\lim_{k \rightarrow \infty} \frac{1}{k} \varphi_{k+1}(\theta) \leq -\frac{\delta}{n} \|H(\theta)\|_1 \mathbf{1} \quad \text{for all } \theta \notin \Theta^*$$

where

$$[H(\theta)]_i = D_{KL}(f_i \| \ell^i(\cdot | \theta)) - D_{KL}(f_i \| \ell^i(\cdot | \theta^*)) \quad \text{for } i = 1, \dots, n$$

and δ is a bound for the "influence imbalance for the chain $\{A_k\}$

- $\delta = \inf_{t \geq 0} \min_{i \in [n]} [A_t \cdots A_0 \mathbf{1}_n]_i$
- when the matrices A_k are doubly stochastic $\delta = 1$

Non-Asymptotic Learning Rate

Proposition 2 Let Assumptions 1-3 hold. Also, let $\rho \in (0, 1)$ and consider the update rule

$$\mu_{k+1}^i = \operatorname{argmin}_{\pi \in \Pi(\Theta)} \left\{ -\mathbb{E}_{\pi} [\log \ell_i(s_{k+1} | \cdot)] + \sum_{j=1}^n [A_k]_{ij} D_{KL}(\pi \| \mu_k^j) \right\}$$

Then, the following property is true: for any $\theta \notin \Theta^*$, there is an integer $\mathbf{N}(\rho)$ such that, with probability $1 - \rho$, for all $k \geq \mathbf{N}(\rho)$ there holds

$$\mu_k^i(\theta) \leq \exp\left(-\frac{k}{2}\gamma_2 + \gamma_1\right) \quad \forall i = 1, \dots, n,$$

where

$$\mathbf{N}(\rho) \triangleq \left\lceil \frac{8(\log(\alpha))^2 \log\left(\frac{1}{\rho}\right)}{\gamma_2^2} + 1 \right\rceil,$$

α is a lower bound on the likelihoods ℓ_i (see Assum. 3),

$$\gamma_1 \triangleq \max_{\substack{\theta \in \Theta \setminus \Theta^* \\ \theta^* \in \Theta^*}} \left\{ \max_{1 \leq i \leq n} \log \frac{\mu_0^i(\theta)}{\mu_0^i(\theta^*)} + \frac{C}{1-\lambda} \|\mathbf{H}(\theta)\|_1 \right\} \quad \gamma_2 \triangleq \frac{\delta}{n} \min_{\theta \in \Theta \setminus \Theta^*} \|\mathbf{H}(\theta)\|_1.$$

$$[\mathbf{H}(\theta)]_i = D_{KL}(f_i \| \ell_i(\cdot | \theta)) - D_{KL}(f_i \| \ell_i(\cdot | \theta^*))$$

The constants C , δ and λ are related to the graphs and satisfy the following relations:

□ For general B -connected graph sequences $\{\mathcal{G}_k\}$,

$$C = 2, \quad \lambda \leq (1 - \eta^{nB})^{\frac{1}{B}}, \quad \delta \geq \eta^{nB}.$$

□ If every matrix A_k is doubly stochastic,

$$C = \sqrt{2}, \quad \lambda = \left(1 - \frac{\eta}{4n^2}\right)^{\frac{1}{B}}, \quad \delta = 1.$$

□ If each \mathcal{G}_k is an undirected graph and each A_k is the lazy Metropolis matrix, i.e. the stochastic matrix which satisfies

$$[A_k]_{ij} = \frac{1}{2 \max(d(i), d(j))} \quad \text{for all } \{i, j\} \in \mathcal{G}_k,$$

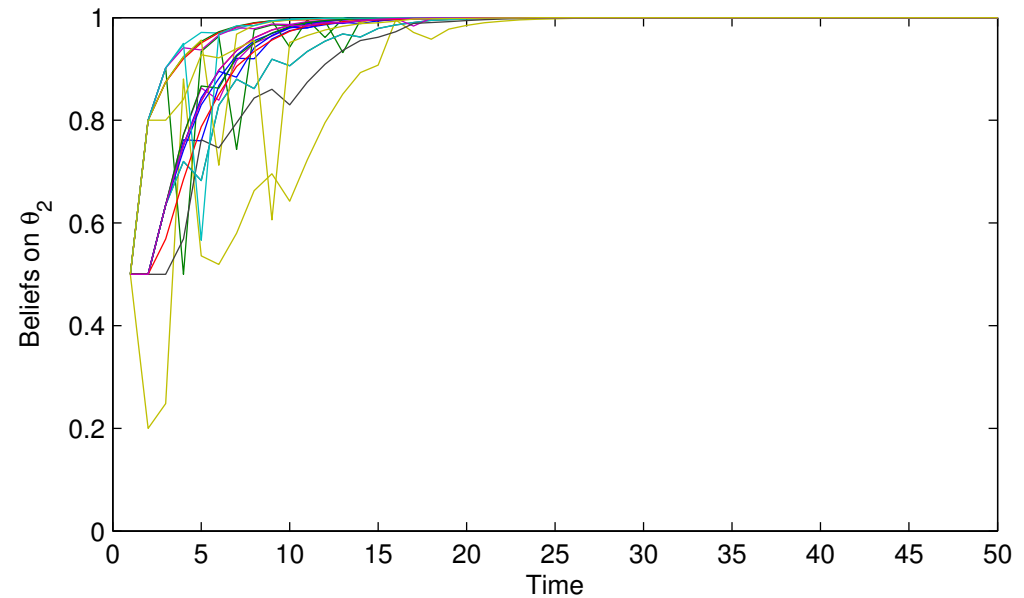
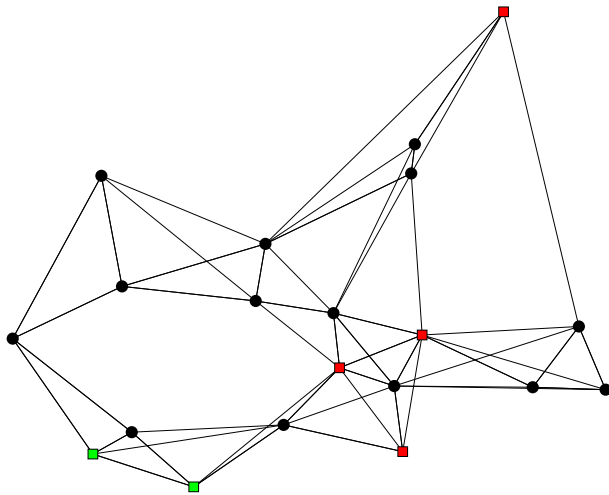
then

$$C = \sqrt{2}, \quad \lambda = 1 - \frac{1}{\Theta(n^2)}, \quad \delta = 1.$$

Simulation result

$n = 20$, $\Theta = \{\theta_1, \theta_2, \theta_3\}$, $\mathcal{S}^i = \{0, 1\}$ for all agents

Graph is static and undirected: A is lazy metropolis



Allowing Conflicting Models

Relaxing the assumption that

$$\bigcap_{i=1}^n \operatorname{argmin}_{\theta \in \Theta} D_{KL}(f_i(\cdot) \parallel \ell^i(\cdot | \theta)) \text{ is nonempty}$$

Note that

$$\Theta^* = \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1}^n D_{KL}(f_i \parallel \ell^i(\cdot | \theta))$$

is non-empty and the agents' beliefs will vanish for all $\theta \notin \Theta^*$ as long as the matrices A_k satisfy some additional properties. Three different settings:

1. Time-varying undirected graphs: A_k is doubly-stochastic with $[A_k]_{ij} > 0$ if $\{i, j\} \in E_k$.

2. Time-varying directed graphs $[A_k]_{ij} = \begin{cases} \frac{1}{d_k^j} & \text{if } j \in N_k^i \\ 0 & \text{if otherwise} \end{cases}$

d_k^i is the out degree of node i at time k , N_k^i is the set of in neighbors of node i .

3. Acceleration in static graphs $\bar{A}_{ij} = \begin{cases} \frac{1}{\max\{d^i, d^j\}} & \text{if } \{i, j\} \in E, \\ 0 & \text{if } \{i, j\} \notin E, \end{cases}$

d^i degree of the node i and $\bar{A} = \frac{1}{2}I + \frac{1}{2}A$,

Learning Rules

Time-varying undirected graphs

$$\mu_{k+1}^i(\theta) = \frac{1}{Z_{k+1}^i} \prod_{j=1}^n \mu_k^j(\theta)^{[A_k]_{ij}} l^i(s_{k+1}^i | \theta), \quad (1)$$

Z_{k+1}^i is a normalization factor, i.e.,

$$Z_{k+1}^i = \sum_{p=1}^m \prod_{j=1}^n \mu_k^j(\theta_p)^{[A_k]_{ij}} l^i(s_{k+1}^i | \theta_p).$$

Time-varying directed graphs

$$\mu_{k+1}^i(\theta) = \frac{1}{Z_{k+1}^i} \left(\prod_{i=1}^n \mu_k^j(\theta)^{[A_k]_{ij} y_k^j} l_i(s_{k+1}^i | \theta) \right)^{\frac{1}{y_{k+1}^i}} \quad (2)$$

$$Z_{k+1}^i = \sum_{p=1}^m \left(\prod_{i=1}^n \mu_k^j(\theta_p)^{[A_k]_{ij} y_k^j} l_i(s_{k+1}^i | \theta_p) \right)^{\frac{1}{y_{k+1}^i}}$$

$$y_{k+1}^i = \sum_{i=1}^n [A_k]_{ij} y_k^j$$

Acceleration in static graphs based on a recent paper by Olshevsky 2014

$$\mu_{k+1}^i(\theta) = \frac{1 \prod_{j=1}^n \mu_k^j(\theta)^{(1+\sigma)A_{ij}} l^i(s_{k+1}^i|\theta)}{Z_{k+1}^i \prod_{j=1}^n \left(\mu_{k-1}^j(\theta) l^j(s_k^j|\theta) \right)^{\sigma A_{ij}}} \quad (3)$$

$$Z_{k+1}^i = \sum_{p=1}^m \frac{\prod_{j=1}^n \mu_k^j(\theta_p)^{(1+\sigma)A_{ij}} l^i(s_{k+1}^i|\theta_p)}{\prod_{j=1}^n \left(\mu_{k-1}^j(\theta_p) l^j(s_k^j|\theta_p) \right)^{\sigma A_{ij}}} \quad (4)$$

Acceleration in Static Graphs

Theorem 1 *Let assumptions for Case 3 hold and let $\rho \in (0, 1)$. Furthermore let $U \geq n$ and let $\sigma = 1 - 2/(9U + 1)$. Then, the update rule of Eq. (3) with this σ , initial condition $\mu_{-1}^i(\theta) = \mu_0^i(\theta)$ and β_{-1}^i fixed to zero, has the following property: there is an integer $\mathbf{N}(\rho)$ such that, with probability $1 - \rho$, for all $k \geq \mathbf{N}(\rho)$ and for all $\theta_v \notin \Theta^*$, there holds*

$$\mu_k^i(\theta_v) \leq \exp\left(-\frac{k}{2}\gamma_2 + \gamma_1\right) \quad \text{for all } i = 1, \dots, n,$$

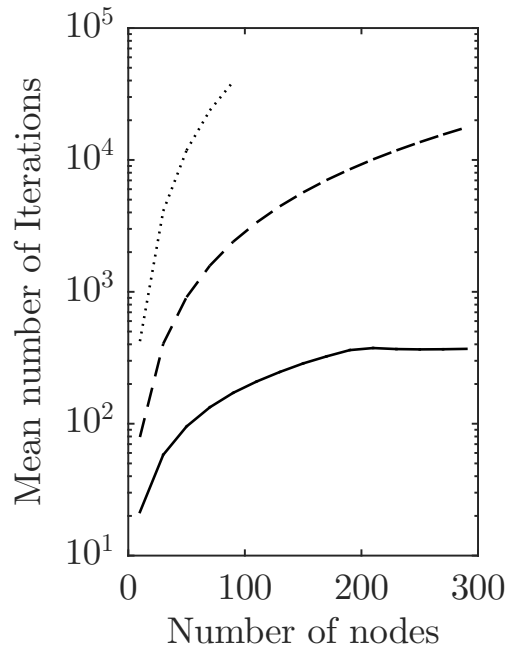
where

$$\mathbf{N}(\rho) \triangleq \left\lceil \frac{72 (\log(\alpha))^2 n \log\left(\frac{1}{\rho}\right)}{\gamma_2^2} \right\rceil,$$

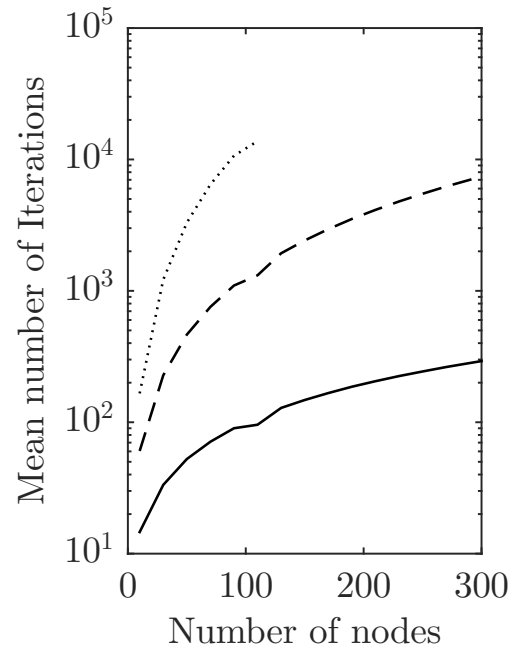
$$\gamma_1 \triangleq \max_{\theta_v \notin \Theta^*} \max_{\theta_w \in \hat{\Theta}^*} \left\{ \frac{\sqrt{2}}{1-\lambda} \sum_{i=1}^n |C_{\mathbf{q}}^i(\theta_v) - C_{\mathbf{q}}^i(\theta_w)| \right\}$$

$$\gamma_2 = \frac{1}{n} \min_{\theta_v \notin \Theta^*} (C_{\mathbf{q}}^* - C_{\mathbf{q}}(\theta_v)),$$

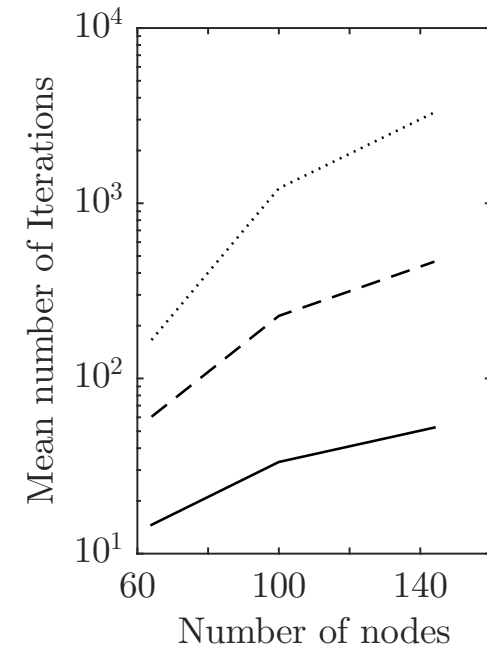
with α from the Assumption on likelihoods and $\lambda = 1 - \frac{1}{18U}$.



(a) Path Graph



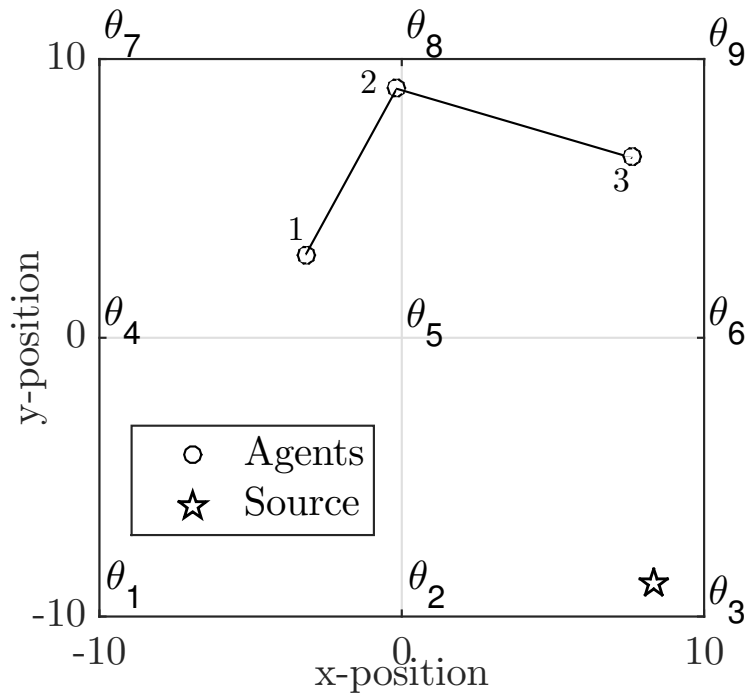
(b) Circle Graph



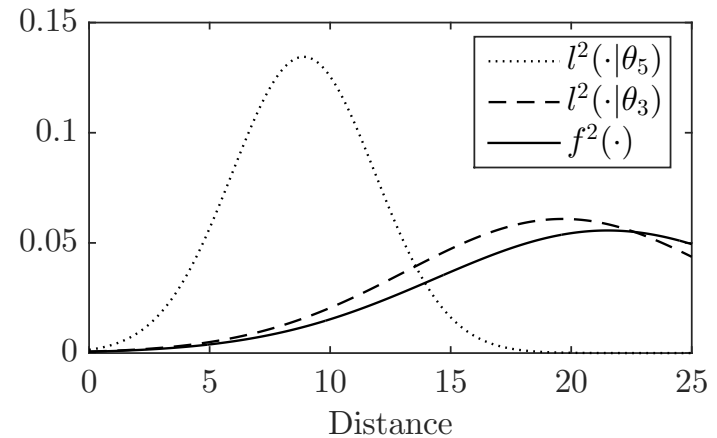
(c) Grid Graph

Figure 1: Empirical mean over 50 Monte Carlo runs of the number of iterations required for $\mu_k^i(\theta) < \epsilon$ for all agents on $\theta \notin \Theta^*$. All agents but one have all their hypotheses to be observationally equivalent. Dotted line for the algorithm proposed in [Jadbabaie 2012], Dashed line for the procedure described in Eq. (1) and solid line for the procedure described in Eq. (3)

Distributed Source Localization



(a) Network of Agents



(b) Hypothesis Distributions

Figure 2: Figure (a) shows a group of 3 agents in a grid of 3×3 hypotheses. Each hypothesis corresponds to a possible location of the source. For example, hypothesis θ_2 locates the source at the $(-10, 0)$ point in the plane. Figure (b) shows the likelihood functions for θ_2 and θ_5 and distribution of observations f^2 for agent 2.

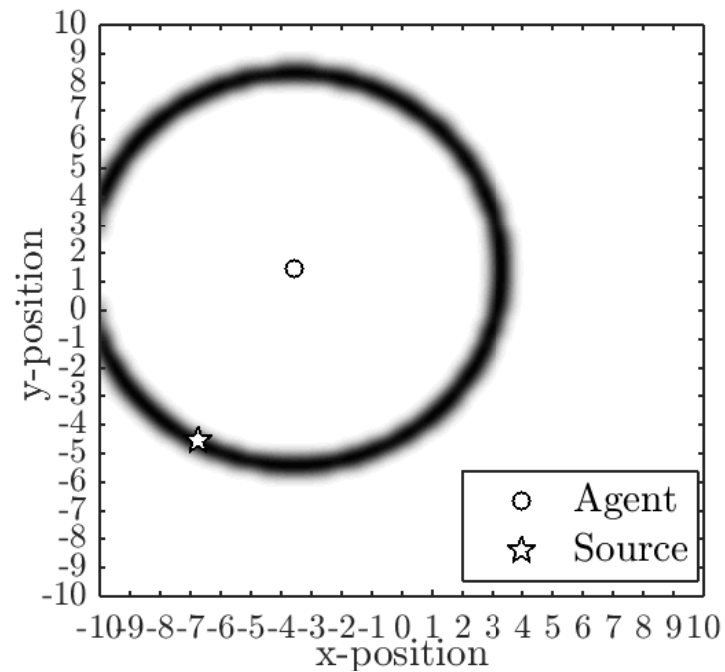
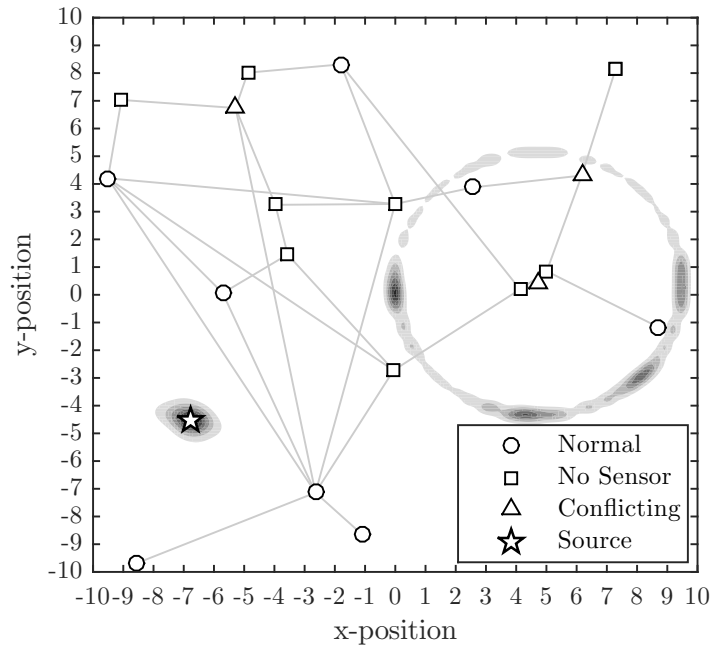
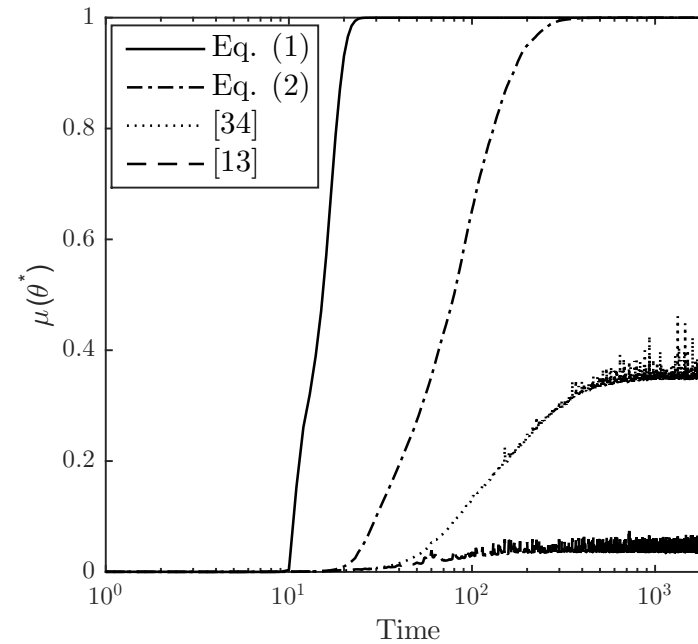


Figure 3: Belief distribution of one agent over the hypotheses grid. Darker shades of gray indicates higher beliefs on the corresponding hypothesis.



(a) Network of Agents



(b) Beliefs on the optimal hypothesis

Figure 4: Figure (a) shows a network of heterogeneous agents. \triangle indicates agents whose observations have been modified such that the optimal hypothesis is the $(0, 0)$ point in the grid. \square indicates agents for whom all hypothesis are observationally equivalent (i.e. no data is measured). \circ indicates regular agents with correct observation models and informative hypothesis. Figure (b) shows the belief evolution on the optimal hypothesis θ^* for different belief update protocols.

Closely Related

- Shahrampour, Rakhlin, and Jadbabaie 2014
- Lalitha, Javidi, and Sarwate 2014, 2015
- N. O. Uribe 2014, 2015