Social learning with heterogeneous agents and sequential decision making

Yunlong Wang and Petar M. Djurić

*Department of Electrical and Computer Engineering
Stony Brook University, Stony Brook, New York 11794*

Abstract

In this paper, we consider the problem of social learning in a network of agents where the agents make decisions sequentially by choosing one of two hypotheses on the state of nature. Each agent observes a signal generated according to one of the hypotheses and knows the decisions of all the previous agents in the network. The network contains two types of agents: rational and irrational. A rational agent makes a decision by not only using its private observation but also the decisions of each of the agents which already made decisions. To that end, the agent employs Bayesian theory. An irrational agent makes a decision by ignoring the available information and by randomly choosing the hypothesis. We analyze the asymptotic performance of a system with rational and irrational agents where we study rational agents that use either a deterministic or random decision making policies. We propose a specific random decision making policy that is based on the social belief and the private signals of the agents. We prove that under mild conditions the expected social belief in the true state of nature tends to one if the rational agents use the proposed random policy. In a network with rational agents that use deterministic policy, the conditions for convergence are stricter. We provide simulation results on the studied systems and compare their performance.

Keywords: social learning, Bayesian learning, sequential decision making, irrational agents, random/deterministic decision making

1. Introduction

An important issue in social learning is how agents make decisions and learn from the decisions of other agents. When agents make decisions based on private and imperfect information, it is natural that they also use information in the decisions of other agents made in a similar situation. In this work, we consider agents that make their decisions on one of two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ one at a time in a sequential manner. When an agent makes its own decision, it broadcasts it to all the agents that will subsequently be making decisions. The graph that describes our system is a directed graph shown in Fig. 1. We note that the motivation for studying this network is understanding some basics of social learning with random decision making rather than a particular application of signal processing.

Applications of sequential decision making can be found in many fields including distributed detection in wireless sensor networks. There, the need for low-cost and low-power devices forces every sensor to aggregate all its information into a one bit decision ([1, 2]). In [3], a comprehensive introduction to this problem was provided. Another major application is the understanding of social learning in multi-agent systems. In these systems, the agents make decisions not only from their private observations, but also from the decisions made by others [4]. As indicated in [5], the study of social learning addresses the problems of modeling the interaction of the agents in the network while learning takes place.

In social learning systems, an agent can learn from the decisions of agents that already made them by using either non-Bayesian approaches or the Bayesian methodology. For example, in [6, 7, 8], non-Bayesian social learning methods were studied. In [4, 9, 10, 11, 12, 13, 14, 15], the agents had Bayesian social learning strategies in obtaining information from previous decisions. Although the Bayesian learning methods provide a fast convergence rate to one of the hypotheses, the agents that employ the Bayesian machinery may ignore their private observations and *herd* on the wrong decision. Some work on this issue can be found in [4, 9], where the interest was in studying herding behavior and information cascades. In [11], the authors provided the sufficient conditions for asymptotic learning in Bayesian social learning systems. In [16], the author studied this problem from the perspective of quickest time change detection. Recently, random decision making was addressed in [17]. In [18], the authors considered the effect of the decisions of subsequent agents on the utility of a current agent by us-

ing a Chinese restaurant game model. They proposed an optimal strategy for decision making. This approach was applied to cognitive radio networks [19]. Another extension of this problem was a scenario with noisy links [20], where the decisions were broadcasted through a noisy channel and thereby the decisions could be randomly flipped or erased. An overview of models and techniques for studying social learning can be found in [5, 10]. In this paper, our study is within the framework of Bayesian social learning.

In the analysis of social learning systems, one typical question is the following: Does the probability of error in decision making converge to zero as the number of agents goes to infinity? The answer depends on (a) whether the log-likelihood ratio of private observations is bounded and (b) whether the memories of the agents are finite (the size of the memory of an agent is the number of decisions of other agents available to the agent). In [21, 22, 23], the problem with finite memory was addressed. In [21], a system where each agent could only observe the decision of its immediate previous agent was investigated. It was shown that if the likelihood ratio test is employed for decision making, the error probability does not converge to zero if the ratio is bounded. In [22], social learning with sampling of past decisions was addressed. There, each agent made inference by the Bayesian method and using very few samples from the decision history of the agents. In the case of bounded log-likelihood ratios, a non-Bayesian decision policy was proposed whose probability of decision error converges to zero [23]. In [11, 24], the agents had infinite memory, and the conditions for asymptotic learning were listed.

In the literature, it is common to assume that the agents know the decision making policy of the other agents, or that the policy is identical for every agent (e.g., [4, 16, 20]). In this work, we address the problem of social learning in a system of heterogeneous agents. The agents can be either rational or irrational and they make the decisions sequentially, one at a time. A rational agent makes a decision according to the posterior of the hypotheses (conditioned on the decisions of the previous agents and its private observation). By contrast, an irrational agent simply ignores all its available information and makes a decision by flipping a fair coin. The agents do not know which of them are rational and irrational but they know the percentage of irrational agents in the network.

More related work on social learning with heterogeneous agents can be found in [25], [26] and [27]. In [25], the problem of misinformation spread was addressed, where the social system was modeled by a random network

and the agents used a gossip style method to update their beliefs from the exchanged information. In [26], with a binary voter model, the authors studied the effect of stubborn agents in social systems. There, the stubborn agents are agents that never change their beliefs. In [27], the authors discussed the relationship between the convergence rate of an average-based learning method and the homophily of the multi-type random networks. In all of these papers, however, there is no randomness in the decisions making.

In this paper, we study two cases. In the first, the rational agents for decision making use a deterministic policy whereas in the second, they employ a random policy. In [17], we proposed a random decision making policy and compared it to the deterministic policy. In that comparison, all the agents were rational. We showed that with the deterministic policy, the probability of decision error may fail to converge to zero due to an information cascade. On the contrary, with the random policy, the probability of decision error is guaranteed to converge to zero. In this work, because of the presence of irrational agents, this probability does not converge to zero. However, if an agent is rational, its probability of a decision error converges to zero. Furthermore, if the social belief is defined by the posterior on the true state of nature conditioned on all the known decisions, then we show that it *always* converges to one in probability if the rational agents adopt the random decision making policy. This is not the case when they use the deterministic policy. In summary, the main contributions of this paper are (1) a proof that the presence of irrational agents does not affect the asymptotic probability of the rational agents if they adopt random policy and (2) a proof that in a system with random agents herding does not take place.

The paper is organized as follows. In the next section we describe the models of the sequential system and explain the social learning process. In Section 3, we introduce both the deterministic and the random decision making policies. The analyses of the convergence of the social belief and the probability of decision error are provided in Section 4. Simulation results are presented in Section 5, and concluding remarks in Section 6.

2. Problem formulation

The process of social learning is illustrated in Fig. 1. We consider the decision making problem in networks of agents A_n , $n \in \mathbb{N}^+$, where the agents make decisions and broadcast them sequentially one at a time to all the subsequent agents. Mathematically, each agent A_n receives an independent

Figure 1: Social learning in a sequential decision making system.

private observation y_n that is generated according to one of the following two hypotheses:

$$\begin{aligned} \mathcal{H}_1 : \quad & y_n \sim \phi_1(y_n) \\ \mathcal{H}_0 : \quad & y_n \sim \phi_0(y_n), \end{aligned} \tag{1}$$

where $\phi_k(y_n)$ denotes the distribution of observation y_n under $\mathcal{H}_k$, $\forall k \in \{0, 1\}$. We assume that these two distributions are known to all the agents and that each agent has noninformative prior probabilities on the hypotheses. Thus, we have $p(\mathcal{H}_0) = p(\mathcal{H}_1) = 1/2$. Let $\log \left(\frac{\phi_1(y_n)}{\phi_0(y_n)} \right)$ be the log-likelihood ratio (LLR) of the hypotheses. We discriminate two types of LLRs, bounded and unbounded. An LLR is bounded if there exist two finite real numbers m and M , such that $\forall y_n$, $\log \left(\frac{\phi_1(y_n)}{\phi_0(y_n)} \right) \in [m, M]$. Otherwise, the LLR is unbounded.

In the network, there is a positive probability that an agent A_n behaves irrationally. An irrational agent A_n , makes its decision by drawing from a Bernoulli distribution with parameter $\eta \in (0, 1)$, i.e.,

$$\alpha_n \sim \text{Ber}(\eta), \tag{2}$$

where α_n denotes the decision made by A_n and $\alpha_n \in \mathcal{A} = \{0, 1\}$. Let I_n be an indicator function taking a value of one if the agent A_n is irrational; otherwise, it is zero. Then we set that $p(I_n = 1) = \xi < 1$, $\forall n \in \mathbb{N}^+$. In the rest part of this paper, we assume that both η and ξ are known by all the agents in the network.

Here we remark that the model in (2) is just one possible formulation of irrational behavior. In our work, an irrational agent is an agent that ignores all the available information to it and makes decision randomly according to a certain law. The simplicity of this irrational behavior model notwithstanding, allows us to demonstrate very interesting asymptotical properties of the studied system, which are of primary interest here.

The rational agents, on the other hand, make their decisions by not only using their private observations, but also learning from others' decisions. Let $\alpha_{1:n}$ be the decision sequence from agent A_1 to agent A_n . Then the agent A_n can formulate its private belief in $\mathcal{H}_1$, $\beta_n = p(\mathcal{H}_1|\alpha_{1:n-1}, y_n)$, by using the Bayes' rule given by the following equation, $\forall n \in \mathbb{N}^+$:

$$\beta_n = \frac{1}{1 + \frac{(1 - \pi_{n-1})\phi_0(y_n)}{\pi_{n-1}\phi_1(y_n)}}, \quad (3)$$

where we define π_n to be the social belief in $\mathcal{H}_1$ conditioned on the action sequence until and including agent A_n , i.e.,

$$\pi_n = p(\mathcal{H}_1|\alpha_{1:n}), \forall n \in \mathbb{N}^+, \quad (4)$$

with π_0 being defined by $1/2$. We note that the private belief β_n represents the posterior on $\mathcal{H}_1$ conditioned on all the information available to A_n . Furthermore, the social belief π_{n-1} serves as the prior knowledge of A_n before it receives its private observation y_n .

In this system, once A_n makes its decision α_n , all the following agents should update the social belief by Bayes' rule from π_{n-1} to π_n . Recall that the studied network is composed of heterogenous agents, and therefore when the decision α_n is broadcasted, the following agents do not know if the decision is made by a rational or an irrational agent. Thus, in updating the social belief, this has to be taken into consideration, which is done by marginalizing the hidden random variable I_n . By the definition in (4), we have that the social belief is updated according to

$$\begin{aligned} \pi_n = & \left(\frac{\pi_{n-1}(1 - l_n^{(1)})}{\pi_{n-1}(1 - l_n^{(1)}) + (1 - \pi_{n-1})(1 - l_n^{(0)})} \right)^{1-\alpha_n} \\ & \times \left(\frac{\pi_{n-1}l_n^{(1)}}{\pi_{n-1}l_n^{(1)} + (1 - \pi_{n-1})l_n^{(0)}} \right)^{\alpha_n}, \end{aligned} \quad (5)$$

where depending on the action α_n (zero or one), the updated π_n is either equal to the first or second factor, respectively. In both cases, we have that $\forall k \in \{0, 1\}$,

$$\begin{aligned} l_n^{(k)} &= p(\alpha_n = 1 | \alpha_{1:n-1}, \mathcal{H}_k) \\ &= \sum_{I_n=0}^1 p(\alpha_n = 1 | \alpha_{1:n-1}, \mathcal{H}_k, I_n) p(I_n) \end{aligned} \quad (6)$$

denotes the probability of agent A_n making decision $\alpha_n = 1$ given the decision sequence up to α_{n-1} and the true state of nature being $\mathcal{H}_k$. As in [5], we refer to $l_n^{(k)}$ as *action likelihood*. Due to the possibility that one decision can be made by an irrational agent, the action likelihood is obtained by using the total probability theorem. For $n = 1$, we set $l_1^{(k)} = p(\alpha_1 = 1 | \mathcal{H}_k)$.

A rational agent A_n , after obtaining its private belief by the Bayes' rule in (3), makes a decision according to some policy. We describe the policy as a random map of the private belief β_n to the action space $\mathcal{A} = \{0, 1\}$. Mathematically, the agent makes its decision from a certain distribution conditioned on β_n , i.e.,

$$\alpha_n \sim p(\alpha_n | \beta_n). \quad (7)$$

We note that, both the deterministic and random decision making can formally be described by (7).

Given the decision making policy $p(\alpha_n | \beta_n)$ and the known value of ξ , the agents can obtain the action likelihood by marginalizing out y_n . It is given by

$$l_n^{(k)} = \eta \xi + q_k(1 - \xi), \quad \forall k \in \{0, 1\}, \quad (8)$$

where q_k denotes the probability that a rational agent A_n makes a decision one given that $\mathcal{H}_k$ is true, i.e.,

$$\begin{aligned} q_k &= p(\alpha_n = 1 | \alpha_{1:n-1}, \mathcal{H}_k, I_n = 0) \\ &= \int_{-\infty}^{\infty} p(\alpha_n = 1 | \alpha_{1:n-1}, y_n, I_n = 0) p(y_n | \mathcal{H}_k) dy_n. \end{aligned} \quad (9)$$

In the proposed system, an irrational agent A_n is active only when it is its turn to make a decision, which is made according to (2). On the other hand, a rational agent A_n behaves as follows. While it is following the decisions of

the previous agents, it calculates the action likelihoods $l_t^{(0)}$ and $l_t^{(1)}$ ($t < n$) and updates the social belief π_{t-1} to π_t by (5). When $t = n$, A_n observes y_n , and it calculates its private belief by (3). Finally, it makes its decision according to (7). This sequential learning algorithm of the rational agents belongs to the class of “social learning filters” [10].

Algorithm I: The behavior of a rational agent A_n .

Initialize $\pi_0 = 0.5$.

```

while  $t \leq n$  do
  if  $t < n$  then
     $A_n$  observes the decision made by agent  $A_t$ ;
     $A_n$  calculates the action likelihoods  $l_t^{(0)}$  and  $l_t^{(1)}$  by (8);
     $A_n$  updates the social belief  $\pi_{t-1}$  to  $\pi_t$  by (5);
  else
     $A_n$  calculates its private belief by (3) after observing  $y_n$ ;
     $A_n$  makes its decision according to (7) ;
  end
end

```

3. Decision making policies

In this paper, we address two decision making policies of rational agents and analyze the asymptotic performance of systems with such agents. Namely, we study heterogenous systems composed of rational and irrational agents, where the rational agents follow either deterministic or random decision making policies.

3.1. Deterministic policy

A common deterministic decision making policy of an agent A_n is defined by [4],

$$\alpha_n = \begin{cases} 1, & \text{if } \beta_n > 0.5, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

If we set the reward to one when the agent makes a decision identical to the true hypothesis, and zero otherwise, by the above rule, the expected utility of A_n is maximized [4]. We call this policy deterministic because once the private belief β_n has been formed, the decision α_n is made deterministically. In the literature, such agents are called myopic.

3.2. Random policy

An alternative to the deterministic policy is a random policy. In [17], we defined a random policy where the decision of the agent is drawn from a Bernoulli distribution defined by

$$p(\alpha_n|\beta_n) = \begin{cases} \beta_n, & \text{if } \alpha_n = 1, \\ 1 - \beta_n, & \text{otherwise.} \end{cases} \quad (11)$$

Thus, the decision α_n of A_n is a Bernoulli random variable parameterized by the A_n 's private belief β_n . In social systems, the actions of the agents are sometimes naturally or by design random. Although these agents do not maximize an explicit utility function, we still call them rational.

4. Analysis

In the proposed social learning system, $\pi_n(\alpha_{1:n}) = p(\mathcal{H}_1|\alpha_{1:n})$, that is π_n is a function defined by the action sequence $\alpha_{1:n}$ (there is a deterministic mapping from $\alpha_{1:n}$ to π_n). In the analysis that follows, without loss of generality, we assume that the true hypothesis is $\mathcal{H}_1$. We will also simplify the notation and will express $\pi_n(\alpha_{1:n})$ as π_n . Considering that $\alpha_{1:n}$ is random due to the random observations and the random policy if it is applied, π_n is a discrete random process. Since π_n reflects the performance of the system at n , we want to describe it in statistical terms. To that end we use $\mathbb{E} \pi_n$, which is given by

$$\begin{aligned} \mathbb{E} \pi_n &= \sum_{\pi_n} \pi_n p(\pi_n) \\ &= \sum_{\alpha_{1:n} \in \mathcal{A}^n} \pi_n p(\alpha_{1:n}). \end{aligned} \quad (12)$$

Since the agents use the Bayesian method to learn from the decision sequence, it can be shown that their average performance cannot decrease if more information becomes available to the system. Mathematically, in the following we

present a theorem that states the relationship between the action likelihood and the monotonic property of the $\mathbb{E} \pi_n$:

Theorem 1. *In the proposed sequential decision making system, when the true state of nature is $\mathcal{H}_1$, the expected value of social belief is monotonic, i.e., $\forall t \in \mathbb{N}^+$,*

$$\mathbb{E} \pi_n - \mathbb{E} \pi_{n-1} \geq 0. \quad (13)$$

Proof: By (12), we can write

$$\mathbb{E} \pi_n = \sum_{\alpha_{1:n-1} \in \mathcal{A}^{n-1}} p(\alpha_{1:n-1} | \mathcal{H}_1) \left(\sum_{\alpha_n=0}^1 p(\alpha_n | \alpha_{1:n-1}, \mathcal{H}_1) \pi_n \right), \quad (14)$$

where with the conditioning we have emphasized that the true hypothesis is $\mathcal{H}_1$. Then we have,

$$\mathbb{E} \pi_n - \mathbb{E} \pi_{n-1} = \sum_{\alpha_{1:n-1} \in \mathcal{A}^{n-1}} p(\alpha_{1:n-1} | \mathcal{H}_1) \Delta(\alpha_{1:n}), \quad (15)$$

where $\Delta(\alpha_{1:n})$ is a function of the action sequence $\alpha_{1:n}$, which is given by,

$$\Delta(\alpha_{1:n}) = \sum_{\alpha_n=0}^1 p(\alpha_n | \alpha_{1:n-1}, \mathcal{H}_1) \pi_n - \pi_{n-1}. \quad (16)$$

Noting that π_n is updated from π_{n-1} according to (5) and that $p(\alpha_n | \alpha_{1:n-1}, \mathcal{H}_1)$ can be expressed by the action likelihood in (9), we can rewrite $\Delta(\alpha_{1:n})$ in the following form:

$$\begin{aligned} \Delta(\alpha_{1:n}) &= \frac{\pi_{n-1}(1 - l_n^{(1)})^2}{\pi_{n-1}(1 - l_n^{(1)}) + (1 - \pi_{n-1})(1 - l_n^{(0)})} \\ &\quad + \frac{\pi_{n-1}(l_n^{(1)})^2}{\pi_{n-1}l_n^{(1)} + (1 - \pi_{n-1})l_n^{(0)}} - \pi_{n-1}. \end{aligned} \quad (17)$$

From the above equation, one can derive that

$$\Delta(\alpha_{1:n}) = \frac{\pi_{n-1}}{Z} (1 - \pi_{n-1})^2 (l_n^{(1)} - l_n^{(0)})^2, \quad (18)$$

where Z is the multiplication of the two denominators in (16) and is given by

$$Z = (\pi_{n-1}(1 - l_n^{(1)}) + (1 - \pi_{n-1})(1 - l_n^{(0)}))(\pi_{n-1}l_n^{(1)} + (1 - \pi_{n-1})l_n^{(0)}).$$

Noting that $\pi_n \in [0, 1]$, and $l_n^{(k)} \in [0, 1]$, then it can be shown that $\Delta(\alpha_{1:n}) \geq 0$.

Because $p(\alpha_{1:n-1}|\mathcal{H}_1) \geq 0$, $\forall \alpha_{1:n-1} \in \mathcal{A}^{n-1}$, we have proved that the expected social belief is non-decreasing. $\square$

Considering the symmetric structure of the proposed system, we can make a similar claim about the behavior of $\mathbb{E} \pi_n$ when $\mathcal{H}_0$ is true. Then, the sequence $\mathbb{E} \pi_n$ is monotonically non-increasing. One can prove this by repeating the same process as in the above proof with a few notational changes.

By Theorem 1 and the fact that $\mathbb{E} \pi_n$ is bounded, we can readily show that the expected social belief must converge. We next will analyze the limit of this sequence when the deterministic and random policies are used.

4.1. Deterministic policy

We start our analysis with the definition of information cascade from [4]. It is defined as follows: an information cascade occurs at time $\tilde{n}$ if the public belief stops evolving after agent $A_{\tilde{n}}$, i.e., $\pi_n = \pi_{\tilde{n}}$, $\forall n > \tilde{n}$.

As is shown in [4], if the LLR of the social learning system with deterministic policy is bounded, the probability that information cascade occurs converges to one. Instead of repeating the proof here, we briefly sketch it as follows.

Let $\log \left(\frac{\phi_1(y_n)}{\phi_0(y_n)} \right) \in [m, M]$, $\forall y_n$. Then by (10), we get that if the social belief $\pi_n < u$ or $\pi_n > U$, the agent A_n will make its decision without using its private information even if it is rational, where $u = \frac{1}{1+e^M}$ and $U = \frac{1}{1+e^m}$. Then by (6), we get that $l_n^{(0)} = l_n^{(1)}$, meaning that the social belief does not change after the decision α_n is made, i.e., $\pi_n = \pi_{n-1}$. Moreover, once the social belief leaves the range $[u, U]$, it remains unchanged forever. Hence, information cascade starts from this agent A_n .

It can be shown that for arbitrary $\pi_n \in [u, U]$, there exists a finite $\tau \in \mathbb{N}^+$ such that the probability $p(\pi_{n+\tau} \in [0, u] \cup [U, 1])$ is strictly positive. This is because if all the agents after A_n make decisions as one or as zero, then the social belief will evolve to the range $[U, 1]$ or $[0, u]$ in finite time,

respectively. Noting that the probability for this event is strictly positive, the above statement is true. Then we have that there exists $\tau_0 \in \mathbb{N}^+$ and a $p_0 > 0$ such that $\forall \pi_n \in [u, U]$, $p(\pi_{n+\tau_0} \in ([0, u] \cup [U, 1])) > p_0$, where τ_0 is the upper bound of all the τ s and p_0 is the lower bound of all the $p(\pi_{n+\tau} \in [0, u] \cup [U, 1])$ s.

With the above arguments, we get that the probability of absence of information cascade after τ_0 is smaller than $1 - p_0$. From the chain rule of probability theory, it follows that the probability of absence of information cascade will converge to zero exponentially as time goes to infinity.

On the other hand, (6) also implies that if $l_n^{(0)} \neq l_n^{(1)}$, $\pi_n \neq \pi_{n-1}$. Therefore if $l_n^{(0)} \neq l_n^{(1)}$ for arbitrary π_{n-1} , information cascade will not happen.

With the following lemma, we show that when the LLR is unbounded, information cascade cannot occur.

Lemma 1. *With the deterministic policy, if (a) the two distributions ϕ_0 and ϕ_1 are not identical everywhere, (b) $0 < \pi_{n-1} < 1$, and (c) the LLR is unbounded, then $l_n^{(0)} \neq l_n^{(1)}$.*

Proof: By (8), we have

$$l_n^{(1)} - l_n^{(0)} = (1 - \xi)(q_1 - q_0). \quad (19)$$

Then it is sufficient to show $q_1 \neq q_0$, $\forall 0 < \pi_{n-1} < 1$.

By the deterministic policy defined in (10), the decision region for $\alpha_n = 1$ denoted by S_1 can be written as

$$S_1 = \left\{ y_n \left| \frac{\phi_1(y_n)}{\phi_0(y_n)} > \frac{1 - \pi_{n-1}}{\pi_{n-1}} \right. \right\}. \quad (20)$$

Then it follows that $p(\alpha_n = 1 | \alpha_{1:n-1}, y_n, I_n = 0)$ equals to one if $y_n \in S_1$ and zero otherwise. Therefore from (9), if the deterministic policy is used, we have that

$$q_k = \int_{y_n \in S_1} \phi_k(y_n) dy_n. \quad (21)$$

Given that $0 < \pi_{n-1} < 1$ and that the LLR is unbounded, both S_1 and its complimentary set S_0 must be nonempty. Next, we show that $q_1 - q_0 > 0$ by considering the following two cases. If $\pi_{n-1} < 0.5$, we have that

$$q_1 - q_0 = \int_{y_n \in S_1} (\phi_1(y_n) - \phi_0(y_n)) dy_n > 0, \quad (22)$$

because $\forall y_n \in S_1, \phi_1(y_n) > \phi_0(y_n)$. Otherwise, $\pi_{n-1} \geq 0.5$ shows that $\forall y_n \in S_0, \phi_1(y_n) < \phi_0(y_n)$. Then we have

$$\begin{aligned} q_1 - q_0 &= (1 - q_0) - (1 - q_1) \\ &= \int_{y_n \in S_0} (\phi_1(y_n) - \phi_0(y_n)) dy_n > 0. \end{aligned} \quad (23)$$

With (22) and (23), the proof of Lemma 1 is completed. $\square$

We point out that the result in Lemma 1 shows that information cascade cannot occur when the LLR is unbounded in the studied system. In [11], a similar result can be found except that there all the agents are rational.

4.2. Random policy

In contrast to the deterministic policy, the random policy in (11) shows that even if the LLR is bounded, the social belief still can evolve after it becomes greater than U or smaller than u . With the following lemma, we show that the information cascade can be prevented by using the random policy.

Lemma 2. *With the random policy, if the two distributions ϕ_0 and ϕ_1 are not identical everywhere and if $0 < \pi_{n-1} < 1$, it holds that $l_n^{(0)} \neq l_n^{(1)}$.*

Proof: Again, we prove the lemma by showing $q_1 - q_0 \neq 0$ for any π_{n-1} . From (9), we have that

$$q_k = \int_{-\infty}^{\infty} \frac{\pi_{n-1} \phi_1(y_n)}{\pi_{n-1} \phi_1(y_n) + (1 - \pi_{n-1}) \phi_0(y_n)} \phi_k(y_n) dy_n. \quad (24)$$

Noting that $\pi_{n-1} \neq 0$, we have

$$q_1 - q_0 = \int_{-\infty}^{\infty} \frac{\phi_1(y_n)(\phi_1(y_n) - \phi_0(y_n))}{\phi_1(y_n) + \lambda \phi_0(y_n)} dy_n, \quad (25)$$

where $\lambda = \frac{1 - \pi_{n-1}}{\pi_{n-1}}$ represents the ratio of social beliefs in $\mathcal{H}_0$ and $\mathcal{H}_1$.

We now prove that $q_1 \neq q_0$ by contradiction. Define $\mathfrak{I}_1 = q_1 - q_0$ by

$$\mathfrak{I}_1 = \int_{-\infty}^{\infty} \frac{\phi_1(y_n)(\phi_1(y_n) - \phi_0(y_n))}{\phi_1(y_n) + \lambda \phi_0(y_n)} dy_n, \quad (26)$$

and

$$\mathfrak{I}_2 = \int_{-\infty}^{\infty} \frac{\lambda \phi_0(y_n)(\phi_1(y_n) - \phi_0(y_n))}{\phi_1(y_n) + \lambda \phi_0(y_n)} dy_n. \quad (27)$$

From $\int_{-\infty}^{\infty} (\phi_1(y_n) - \phi_0(y_n)) dy_n = 0$, we have that $\mathfrak{I}_1 + \mathfrak{I}_2 = 0$. Assuming that $\mathfrak{I}_1 = 0$, then we must have $\mathfrak{I}_2 = 0$. This, however, is contradicted by the following equation:

$$\mathfrak{I}_1 - \frac{\mathfrak{I}_2}{\lambda} = \int_{-\infty}^{\infty} \frac{(\phi_1(y_n) - \phi_0(y_n))^2}{\phi_1(y_n) + \lambda \phi_0(y_n)} dy_n > 0 \quad (28)$$

because $(\phi_1(y_n) - \phi_0(y_n))^2$ does not equal to zero for all y_n .

Hence we conclude that for arbitrary distributions ϕ_0 and ϕ_1 , the above lemma holds. $\square$

Lemma 2 indicates the difference between the deterministic and random policies. When the random policy is adopted, the social belief will not stop evolving regardless of the LLR.

4.3. Asymptotic learning

With the above lemmas, we can now get the limits of the expected social belief. Also we show when the probability for a rational agent making an error decision converges to zero. With the next theorem, we first show the limit of the expected social belief.

Theorem 2. *Suppose that the agents implement the deterministic policy in (10) with an unbounded LLR or the random decision policy in (11). Let $\mathcal{H}_1$ be the true state of nature, then we have*

$$\lim_{n \rightarrow \infty} \mathbb{E} \pi_n = 1. \quad (29)$$

If $\mathcal{H}_0$ is true,

$$\lim_{n \rightarrow \infty} \mathbb{E} \pi_n = 0. \quad (30)$$

Proof: Here we only prove (29). The proof of (30) is similar. First by theorem one, noting that π_n is bounded, we have that the limit of (15) must be zero, i.e.,

$$\lim_{n \rightarrow \infty} (\mathbb{E} \pi_n - \mathbb{E} \pi_{n-1}) = 0, \quad (31)$$

which implies that $\forall \alpha_{1:n-1} \in \mathcal{A}^{n-1}$, $\lim_{n \rightarrow \infty} p(\alpha_{1:n-1} | \mathcal{H}_1) \Delta(\alpha_{1:n-1}) = 0$. By Lemmas 1 and 2, we have that $(l_n^{(1)} - l_n^{(0)})^2 = 0$ only when $\pi_n = 0$ or $\pi_n = 1$. Then the equation (18) shows that $\Delta(\alpha_{1:n-1}) = 0$ if and only if $\pi_n = 0$ or $\pi_n = 1$. Therefore, for arbitrary $\alpha_{1:n-1}$, it must hold that either $\lim_{n \rightarrow \infty} p(\alpha_{1:n-1} | \mathcal{H}_1)(1 - \pi_{n-1}) = 0$ or $\lim_{n \rightarrow \infty} p(\alpha_{1:n-1} | \mathcal{H}_1)\pi_{n-1} = 0$.

By (5), π_{n-1} converges to zero only when every agent herds on $\mathcal{H}_0$ after a certain agent. Since an information cascade cannot happen and given $\mathcal{H}_1$ is true, we have that the probability of $\pi_{n-1} = 0$ converges to zero. Namely, for any decision sequence $\alpha_{1:n-1}$ for which π_{n-1} converges to zero, we must have

$$\lim_{n \rightarrow \infty} p(\alpha_{1:n-1} | \mathcal{H}_1) = 0. \quad (32)$$

By summarizing the two cases, for an arbitrary decision sequence $\alpha_{1:n-1} \in \mathcal{A}^{n-1}$, we write

$$\lim_{n \rightarrow \infty} p(\alpha_{1:n-1} | \mathcal{H}_1)(1 - \pi_{n-1}) = 0. \quad (33)$$

Hence, we have

$$\lim_{n \rightarrow \infty} \sum_{\alpha_{1:n-1} \in \mathcal{A}^{n-1}} p(\alpha_{1:n-1} | \mathcal{H}_1)(1 - \pi_{n-1}) = 0. \quad (34)$$

Then by (12), we show that the limit of $\mathbb{E} \pi_n$ is one, given by

$$\lim_{n \rightarrow \infty} \mathbb{E} \pi_n = 1. \quad (35)$$

□

Considering that the rational agents are using Bayes' rule to formulate their private beliefs, it can be shown that given $\mathcal{H}_1$ is true, the expected private belief is no smaller than the expected social belief. Then the expected private belief for rational agents also converges to one. Symmetrically, under $\mathcal{H}_0$, the expected private belief will converge to zero. In summary, we have that

$$\lim_{n \rightarrow \infty} \mathbb{E}(\beta_n | I_n = 0) = s, \quad \forall s \in \{0, 1\}, \quad (36)$$

where $\mathcal{H}_s$ is the true state of nature.

Considering the decision policies in (10) and (11), the result in (36) suggests that the probability of decision error for a rational agent can converge to

zero when the deterministic policy with unbounded LLR or the random policy (independently of the condition on boundedness of the LLR) are applied. Given $\mathcal{H}_s$ is the true state of nature, we get that

$$\lim_{n \rightarrow \infty} p(\alpha_n = s | I_n = 0) = 1, \forall s \in \{0, 1\}. \quad (37)$$

Therefore, as n goes to infinity, the probability for one agent to make a right decision is given by

$$\lim_{n \rightarrow \infty} p(\alpha_n = s) = \begin{cases} 1 - \xi\eta, & \text{if } \mathcal{H}_s = \mathcal{H}_0, \\ 1 - \xi(1 - \eta), & \text{if } \mathcal{H}_s = \mathcal{H}_1. \end{cases} \quad (38)$$

where, we recall, ξ is the probability that an agent in the network is irrational. Note that in the above derivation, ξ can take any value between zero and one. We remark that asymptotic learning can be achieved by random policy with any $0 < \xi < 1$. However, as we show in the next section, the convergence speed decreases as ξ grows.

5. Simulations

In this section, we present simulation results on the evolution of social beliefs in sequential decision making systems along with some numerical comparisons between different decision policies.

In the first experiment, we applied the methods for a binomial data model given by

$$\begin{aligned} \mathcal{H}_1 : \quad & y_n \sim B(m, p_1), \\ \mathcal{H}_0 : \quad & y_n \sim B(m, p_0), \end{aligned} \quad (39)$$

where $B(m, p)$ represents a binomial distribution parameterized by m and p . It can be shown that the LLR of this model is bounded.

We verified the analytical result of the expected social belief by Monte Carlo methods, which were conducted with 2,000 trials. In each trial, we set the number of agents to be $N = 1,000$, and they observed data generated from the data model listed in (39) and parameterized by $m = 6$, $p_0 = 0.6$, and $p_1 = 0.7$. We tested the system with $\eta = 0.5$ and with two values of ξ , $\xi = 0.1$ and $\xi = 0.3$. The private signals of the agents were generated according to $\mathcal{H}_1$.

Figure 2: The convergence of social beliefs with two policies, where the agents get observations from a binomial data model.

We show the results of the deterministic and random policies in Fig. 2. On the abscissa, we plotted the agent index and on the ordinate the estimate of the expected social belief, which was given by the average social belief from all the 2,000 trials. From the figure, we see that the expected social beliefs kept increasing when the agents used the random policy, whereas, as expected, it leveled off when they employed the deterministic policy. We also observe that a larger ξ resulted in a lower convergence speed of the expected social belief.

Another performance metric of interest is the proportion of agents that made the right decision. The results of its investigation are plotted in Fig. 3. As shown in the figure, with the random policy, the probability for agents to make correct decision converged to 0.95 and 0.85 when $\xi = 0.1$ and $\xi = 0.3$, respectively. Recall that this probability converged to 0.85 and 0.95, respectively, instead of 1 because both the rational and irrational agents are counted in the calculation. Also, the figure indicates that the probability of decision error for the rational agents with random policy converged to zero. By contrast, if the deterministic policy was used, it can be seen that the right decision probability stopped increasing because of information cascade. Again, we observe that ξ can decrease the speed of learning.

As shown in Section 4, when the deterministic policy is adopted, herding occurs in finite time with probability one. In Fig. 4, for each agent A_n , we plotted the proportion of trials with herding already started before A_n .

Figure 3: The proportion of agents making the right decision with a binomial data model.

In other words, we plotted the empirical cumulative distribution function (c.d.f.) of the starting time of herding when the deterministic policy is used. There, the two curves represent the performance of the studied system with $\xi = 0.1$ and $\xi = 0.3$, respectively. It can be seen that the herding starts earlier when there is a larger proportion of rational agents.

In order to test the random policy with different values of η , we conducted a second experiment. The data model was identical to that of the first experiment in (39). We let the agents implement the random policy with $\eta = 0.9$, $\eta = 0.5$, and $\eta = 0.1$, respectively. With each of the settings, the random policy was implemented 2,000 times. On the left side of Fig. 5, we plotted the evolution of the expected social belief. It can be seen that the evolution of the expected social belief is insensitive to the value of η . This is because the decisions of the irrational agents provide no information to the other agents, irrespective of the decision rule they use. Thus, we can see that η has little effect on the decision making of the rational agents. On the right side of Fig. 5, we plotted how the proportion of agents making the correct decision evolves for different values of η . There we observe that the converged values are as predicted by (38), which are 0.97, 0.85, and 0.73, respectively.

In the third experiment, we let the data be generated by a Gaussian data

Figure 4: The empirical cumulative distribution function for the herding starting time of the deterministic policy.

model given by

$$\begin{aligned}\mathcal{H}_1 : \quad & y_n \sim \mathcal{N}(\theta, \sigma_w^2), \\ \mathcal{H}_0 : \quad & y_n \sim \mathcal{N}(0, \sigma_w^2),\end{aligned}\tag{40}$$

where y_n was a Gaussian random variable, and the parameters θ and σ_w were known by the agents. Compared with the binomial data model, we have proved that the expected social belief will converge with both policies since the LLR is now unbounded.

We conducted the second experiment by repeating the same process of experiment one. The sequential decision making was carried out by 1,500 agents in 2,000 trials. The parameters were $\theta = 1$ and $\sigma_w = 1$. Again, we let the private observations be generated according to $\mathcal{H}_1$.

In Fig. 6, we plotted the evolution of social belief with the Gaussian model. We observe that the expected social beliefs increase for both policies. Unlike with the binomial data model, information cascade did not occur.

In Fig. 7, we plotted the proportion of agents making correct decision. As shown in the figure, the agents applying both policies had a trend of making the right decision asymptotically. This shows that asymptotic learning can be achieved with both policies. As opposed to the result from experiment one, we observed that the probability for correct decision kept increasing at all times. As we have proved in Theorem 2, because of the unbounded

Figure 5: The asymptotical performance of the random policy with different values of η .

LLR, even with the deterministic policy, the proportion of agents making the correct decision will converge to 0.95 when $\xi = 0.1$ and to 0.85 when $\xi = 0.3$. In Fig. 7, it is also shown that the irrational agents can slow down the learning in the system. Namely, there was a significant decrease of convergence speed when ξ was changed from $\xi = 0.1$ to $\xi = 0.3$. We also observe that the deterministic policy has a faster speed in the short run, while the random policy converges faster in the long run.

6. Conclusions

In this paper we presented a sequential learning system constituted by rational and irrational agents. The rational agents made their decisions by using the Bayesian methodology whereas the irrational agents made decisions arbitrarily. The rational agents can exploit the information from previous decisions by Bayesian learning. We applied two policies to the system and showed their asymptotical properties by investigating the convergence of the expected social belief in the system. In analyzing the deterministic policy, we proved that information cascade takes place with probability one if the data model entails a bounded log-likelihood ratio. By contrast, with the random policy, information cascade does not occur and the probability of decision error of the rational agents converges to zero. We demonstrated the performance of the proposed method by Monte Carlo simulations. We showed that the expected social belief converged to one with the random

Figure 6: The convergence of expected social belief with the Gaussian model.

policy even with a bounded LLR, whereas the expected social belief stopped evolving with the deterministic policy.

- [1] P. F. Swaszek, On the performance of serial networks in distributed detection, *IEEE Transactions on Aerospace and Electronic Systems* 29 (1) (1993) 254–260.
- [2] J. D. Papastavrou, M. Athans, Distributed detection by a large team of sensors in tandem, *IEEE Transactions on Aerospace and Electronic Systems*, 28 (3) (1992) 639–653.
- [3] P. K. Varshney, *Distributed detection and data fusion*, Springer-Verlag New York, Inc., 1996.
- [4] C. Chamley, *Rational Herds: Economic Models of Social Learning*, Cambridge University Press, 2004.
- [5] V. Krishnamurthy, H. V. Poor, A tutorial on interactive sensing in social networks, *IEEE Transactions on Computational Social Systems*, (2014) 3–21.
- [6] S. Bikhchandani, D. Hirshleifer, I. Welch, A theory of fads, fashion, custom, and cultural change as informational cascades, *Journal of political Economy* (1992) 992–1026.

Figure 7: The proportion of agents making the right decision with the Gaussian model.

- [7] A. Jadbabaie, P. Molavi, A. Sandroni, A. Tahbaz-Salehi, Non-Bayesian social learning, *Games and Economic Behavior* 76 (1) (2012) 210–225.
- [8] B. Golub, M. O. Jackson, Naive learning in social networks and the wisdom of crowds, *American Economic Journal: Microeconomics* 2 (1) (2010) 112–149.
- [9] A. V. Banerjee, A simple model of herd behavior, *The Quarterly Journal of Economics* 107 (3) (1992) 797–817.
- [10] V. Krishnamurthy, H. V. Poor, Social learning and Bayesian games in multiagent signal processing: How do local and global decision makers interact?, *IEEE Signal Processing Magazine* 30 (3) (2013) 43–57.
- [11] D. Acemoglu, M. A. Dahleh, I. Lobel, A. Ozdaglar, Bayesian learning in social networks, *The Review of Economic Studies* 78 (4) (2011) 1201–1236.
- [12] D. Gale, S. Kariv, Bayesian learning in social networks, *Games and Economic Behavior* 45 (2) (2003) 329–346.
- [13] P. M. Djurić, Y. Wang, Distributed Bayesian learning in multiagent systems: Improving our understanding of its capabilities and limitations, *IEEE Signal Processing Magazine* 29 (2) (2012) 65–76.

- [14] Y. Wang, P. M. Djurić, A gossip method for optimal consensus on a binary state from binary actions, *IEEE Journal of Selected Topics in Signal Processing* 7 (2) (2013) 274–283.
- [15] C. Chamley, A. Scaglione, L. Li, Models for the diffusion of beliefs in social networks: An overview, *IEEE Signal Processing Magazine* 30 (3) (2013) 16–29.
- [16] V. Krishnamurthy, Bayesian sequential detection with phase-distributed change time and nonlinear penalty pomdp lattice programming approach, *IEEE Transactions on Information Theory* 57 (10) (2011) 7096–7124.
- [17] Y. Wang, P. M. Djurić, Sequential bayesian learning in linear networks with random decision making, in: *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, IEEE, 2014, pp. 6404–6408.
- [18] C. Wang, Y. Chen, K. J. R. Liu, Sequential Chinese restaurant game, *IEEE Transactions on Signal Processing* 60 (3) (2012) 571–584.
- [19] C. Jiang, Y. Chen, Y. Yang, C. Wang, K. J. R. Liu, Dynamic chinese restaurant game: Theory and application to cognitive radio networks, *IEEE Transactions on Wireless Communications* 13 (4) (2014) 1960–1973.
- [20] Z. Zhang, E. Chong, A. Pezeshki, W. Moran, Hypothesis testing in feedforward networks with broadcast failures, *IEEE Journal of Selected Topics in Signal Processing* 7 (5) (2013) 797–810.
- [21] T. M. Cover, Hypothesis testing with finite statistics, *The Annals of Mathematical Statistics* (1969) 828–835.
- [22] L. Smith, P. Sørensen, Rational social learning by random sampling, Available at SSRN 1138095.
- [23] K. Drakopoulos, A. Ozdaglar, J. N. Tsitsiklis, On learning with finite memory, *IEEE Transactions on Information Theory* 59 (10) (2013) 6859–6872.

- [24] E. Mossel, A. Sly, O. Tamuz, Asymptotic learning on Bayesian social networks, *Probability Theory and Related Fields* (2012) 1–31.
- [25] D. Acemoglu, A. Ozdaglar, A. ParandehGheibi, Spread of (mis) information in social networks, *Games and Economic Behavior* 70 (2) (2010) 194–227.
- [26] E. Yildiz, D. Acemoglu, A. Ozdaglar, A. Saberi, A. Scaglione, Discrete opinion dynamics with stubborn agents, SSRN eLibrary.
- [27] B. Golub, M. O. Jackson, How homophily affects the speed of learning and best-response dynamics, *The Quarterly Journal of Economics* 127 (3) (2012) 1287–1338.

7. Acknowledgment

This work has been supported by the NSF under Award CCF-1320626.

8. Vitae

Yunlong Wang received his B.E. Degree in electrical engineering from Xian JIAOTONG University, Xian, China, in 2009. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook. His main research interests are Bayesian learning over networks, distributed estimation and their applications.

Petar M. Djurić received his B.S. and M.S. degrees in electrical engineering from the University of Belgrade and his Ph.D. degree in electrical engineering from the University of Rhode Island. Currently, he is a professor in the Department of Electrical and Computer Engineering at Stony Brook University. He was an associate editor of several signal processing journals including IEEE Transactions on Signal Processing. In 2007, he received the IEEE Signal Processing Magazine Best Paper Award and, during 2008 – 2009, he was Distinguished Lecturer of the IEEE Signal Processing Society. In 2012, he received the European Association for Signal Processing Technical Achievement Award. He is a Fellow of the IEEE.