2020 59th IEEE Conference on Decision and Control (CDC)
Jeju Island, Republic of Korea, December 14-18, 2020

Event-Triggered Distributed Inference

Aritra Mitra, Saurabh Bagchi and Shreyas Sundaram

Abstract— **We study a setting where each agent in a network receives certain private signals generated by an unknown static state that belongs to a finite set of hypotheses. The agents are tasked with collectively identifying the true state. To solve this problem in a communication-efficient manner, we propose an event-triggered distributed learning algorithm that is based on the principle of diffusing low beliefs on each false hypothesis. Building on this principle, we design a trigger condition under which an agent broadcasts only those components of its belief vector that have adequate innovation, to only those neighbors that require such information. We establish that under standard assumptions, each agent learns the true state exponentially fast almost surely. We also identify sparse communication regimes where the inter-communication intervals grow unbounded, and yet, the asymptotic learning rate of our algorithm remains the same as when agents communicate at every time-step. We then establish, both in theory and via simulations, that our event-triggering strategy has the potential to significantly reduce information flow from uninformative agents to informative agents. Finally, we argue that, as far as only asymptotic learning is concerned, one can allow for arbitrarily sparse communication patterns.**

I. Introduction

We consider a scenario involving a network of agents, where each agent receives a stream of private signals sequentially over time. The observations of every agent are generated by a common underlying distribution, parameterized by an unknown static quantity which we call the *true state of the world*. The task of the agents is to collectively identify this unknown quantity from a finite family of hypotheses, while relying solely on local interactions. The problem described above is known as distributed inference/hypothesis testing, and has been explored using a variety of techniques [1]–[8]. While [1]–[6] proposed consensus based linear and log-linear rules, [7] and [8] propose a min-protocol that leads to an improved asymptotic learning rate over previous approaches.

A much less explored aspect of distributed inference is that of *communication-efficiency* - a theme that is becoming increasingly important as we envision distributed autonomy with low-power sensor devices, and limited-bandwidth wireless communication channels. Motivated by this gap in the literature, we seek to answer the following questions in this paper. (i) When should an agent exchange information with a neighbor? (ii) What piece of information should the agent exchange? To address the questions posed above, we draw on ideas from the theory of event-triggered control. The initial results [9], [10] on this topic were centered around stabilizing dynamical systems by injecting control inputs only when needed, as opposed to periodic inputs. These ideas were then extended to design event-driven control and communication techniques for multi-agent systems, focusing primarily on variations of the basic consensus problem [11].

A. Mitra, S. Bagchi and S. Sundaram are with the School of Electrical and Computer Engineering at Purdue University. Email: {mitra14, sbagchi, sundara2}@purdue.edu. This work was supported in part by NSF CAREER award 1653648, and by the Laboratory Directed Research and Development program at Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Contributions: The main contribution of this paper is the development of a novel event-triggered distributed learning rule that is based on the principle of diffusing low beliefs on each false hypothesis across the network. Building on this principle, we design a trigger condition that enables an agent to decide, using purely local information, whether or not to broadcast its belief[1] on a given hypothesis to a given neighbor. Specifically, based on our event-triggered strategy, an agent broadcasts *only* those components of its belief vector that have adequate "innovation", to *only* those neighbors that are in need of the corresponding pieces of information. In this way, our approach not only reduces the frequency of communication, but also the amount of information transmitted in each communication round.

We establish that our proposed event-triggered learning rule enables each agent to learn the true state exponentially fast under standard assumptions on the observation model and the network. We characterize the learning rate of our algorithm, and identify conditions under which it matches the rate in [8], even when the inter-communication intervals between the agents grow unbounded. In other words, we identify sparse communication regimes where communication-efficiency comes essentially for "free". We further demonstrate, both in theory and in simulations, that our event-triggered scheme can considerably reduce information flow from uninformative agents to informative agents. Finally, we show that if asymptotic learning of the truth is the only consideration, then one can allow for arbitrarily long intervals between successive communications.

II. Model and Problem Formulation

Network Model: We consider a group of agents $\mathcal{V} = \{1, \ldots, n\}$, and model interactions among them via an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. An edge $(i, j) \in \mathcal{E}$ indicates that agent i can directly transmit information to agent j, and vice versa. The set of all neighbors of agent i is defined as

[1] By an agent's "belief vector", we mean a distribution over the set of hypotheses; this vector gets recursively updated over time.

$\mathcal{N}_i = \{j \in \mathcal{V} : (j,i) \in \mathcal{E}\}$. We say that $\mathcal{G}$ is rooted at $\mathcal{C} \subseteq \mathcal{V}$, if for each agent $i \in \mathcal{V} \setminus \mathcal{C}$, there exists a path to it from some agent $j \in \mathcal{C}$. For a connected graph $\mathcal{G}$, we will use $d(i,j)$ to denote the length of the shortest path between i and j.

Observation Model: Let $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$ denote m possible states of the world, with each state representing a hypothesis. A specific state $\theta^\star \in \Theta$, referred to as the true state of the world, gets realized. Conditional on its realization, at each time-step $t \in \mathbb{N}_+$, every agent $i \in \mathcal{V}$ privately observes a signal $s_{i,t} \in \mathcal{S}_i$, where $\mathcal{S}_i$ denotes the signal space of agent i.[2] The joint observation profile so generated across the network is denoted $s_t = (s_{1,t}, s_{2,t}, \ldots, s_{n,t})$, where $s_t \in \mathcal{S}$, and $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \ldots \mathcal{S}_n$. Specifically, the signal s_t is generated based on a conditional likelihood function $l(\cdot|\theta^\star)$, the i-th marginal of which is denoted $l_i(\cdot|\theta^\star)$, and is available to agent i. The signal structure of each agent $i \in \mathcal{V}$ is thus characterized by a family of parameterized marginals $l_i = \{l_i(w_i|\theta) : \theta \in \Theta, w_i \in \mathcal{S}_i\}$. We make certain standard assumptions [1]–[5]: (i) The signal space of each agent i, namely $\mathcal{S}_i$, is finite. (ii) Each agent i has knowledge of its local likelihood functions $\{l_i(\cdot|\theta_p)\}_{p=1}^m$, and it holds that $l_i(w_i|\theta) > 0, \forall w_i \in \mathcal{S}_i$, and $\forall \theta \in \Theta$. (iii) The observation sequence of each agent is described by an i.i.d. random process over time; however, at any given time-step, the observations of different agents may potentially be correlated. (iv) There exists a fixed true state of the world $\theta^\star \in \Theta$ (unknown to the agents) that generates the observations of all the agents. The probability space for our model is denoted $(\Omega, \mathcal{F}, \mathbb{P}^{\theta^\star})$, where $\Omega \triangleq \{\omega : \omega = (s_1, s_2, \ldots), \forall s_t \in \mathcal{S}, \forall t \in \mathbb{N}_+\}$, $\mathcal{F}$ is the σ-algebra generated by the observation profiles, and $\mathbb{P}^{\theta^\star}$ is the probability measure induced by sample paths in Ω. Specifically, $\mathbb{P}^{\theta^\star} = \prod_{t=1}^{\infty} l(\cdot|\theta^\star)$. We will use the abbreviation a.s. to indicate almost sure occurrence of an event w.r.t. $\mathbb{P}^{\theta^\star}$.

The goal of each agent is to eventually learn the true state $\theta^\star$. However, the key challenge in achieving this objective arises from an *identifiability problem* that each agent might potentially face. To make this precise, define $\Theta_i^{\theta^\star} \triangleq \{\theta \in \Theta : l_i(w_i|\theta) = l_i(w_i|\theta^\star), \forall w_i \in \mathcal{S}_i\}$ as the set of hypotheses that are *observationally equivalent* to $\theta^\star$ from the perspective of agent i. Thus, if $|\Theta_i^{\theta^\star}| > 1$, it will be impossible for agent i to uniquely learn the true state θ^* on its own.

In the next section, we will develop an algorithm that resolves the identifiability problem described above in a communication-efficient manner. Before describing this algorithm, let us first recall the following definition from [7].

Definition 1. *(**Source agents**) An agent i is said to be a source agent for a pair of distinct hypotheses $\theta_p, \theta_q \in \Theta$ if it can distinguish between them, i.e., if $D(l_i(\cdot|\theta_p)||l_i(\cdot|\theta_q)) > 0$, where $D(l_i(\cdot|\theta_p)||l_i(\cdot|\theta_q))$ represents the KL-divergence between the distributions $l_i(\cdot|\theta_p)$ and $l_i(\cdot|\theta_q)$. The set of source agents for pair (θ_p, θ_q) is denoted $\mathcal{S}(\theta_p, \theta_q)$.* □

Throughout the rest of the paper, we will use $K_i(\theta_p, \theta_q)$ as a shorthand for $D(l_i(\cdot|\theta_p)||l_i(\cdot|\theta_q))$.

[2]We use $\mathbb{N}$ and $\mathbb{N}_+$ to represent the set of non-negative integers and positive integers, respectively.

III. An Event-Triggered Distributed Learning Rule

• **Belief-Update Strategy**: In this section, we develop an event-triggered distributed learning rule that enables each agent to eventually learn the truth, despite infrequent information exchanges with its neighbors. Our approach requires each agent i to maintain a local belief vector $\boldsymbol{\pi}_{i,t}$, and an actual belief vector $\boldsymbol{\mu}_{i,t}$, each of which are probability distributions over the hypothesis set Θ. While agent i updates $\boldsymbol{\pi}_{i,t}$ in a Bayesian manner using only its private signals (see eq. (2)), to formally describe how it updates $\boldsymbol{\mu}_{i,t}$, we need to first introduce some notation. Accordingly, let $\mathbb{1}_{ji,t}(\theta) \in \{0,1\}$ be an indicator variable which takes on a value of 1 if and only if agent j broadcasts $\mu_{j,t}(\theta)$ to agent i at time t. Next, we define $\mathcal{N}_{i,t}(\theta) \triangleq \{j \in \mathcal{N}_i | \mathbb{1}_{ji,t}(\theta) = 1\}$ as the subset of agent i's neighbors who broadcast their belief on θ to i at time t. As part of our learning algorithm, each agent i keeps track of the lowest belief on each hypothesis $\theta \in \Theta$ that it has heard up to any given instant t, denoted by $\bar{\mu}_{i,t}(\theta)$. More precisely, $\bar{\mu}_{i,0}(\theta) = \mu_{i,0}(\theta)$, and $\forall t \in \mathbb{N}$,

$$\bar{\mu}_{i,t+1}(\theta) = \min\{\bar{\mu}_{i,t}(\theta), \{\mu_{j,t+1}(\theta)\}_{j \in \{i\} \cup \mathcal{N}_{i,t+1}(\theta)}\}. \quad (1)$$

We are now in position to describe the belief-update rule at each agent: $\boldsymbol{\pi}_{i,t}$ and $\boldsymbol{\mu}_{i,t}$ are initialized with $\pi_{i,0}(\theta) > 0, \mu_{i,0}(\theta) > 0, \forall \theta \in \Theta, \forall i \in \mathcal{V}$ (but otherwise arbitrarily), and subsequently updated as follows $\forall t \in \mathbb{N}$:

$$\pi_{i,t+1}(\theta) = \frac{l_i(s_{i,t+1}|\theta)\pi_{i,t}(\theta)}{\sum_{p=1}^{m} l_i(s_{i,t+1}|\theta_p)\pi_{i,t}(\theta_p)}, \quad (2)$$

$$\mu_{i,t+1}(\theta) = \frac{\min\{\bar{\mu}_{i,t}(\theta), \pi_{i,t+1}(\theta)\}}{\sum_{p=1}^{m} \min\{\bar{\mu}_{i,t}(\theta_p), \pi_{i,t+1}(\theta_p)\}}. \quad (3)$$

• **Communication Strategy**: We now focus on specifying *when* an agent broadcasts its belief on a given hypothesis to a neighbor. To this end, we first define a sequence $\mathbb{I} = \{t_1, t_2, t_3, \ldots\} \in \mathbb{N}_+$ of *event-monitoring* time-steps, where $t_1 = 1$, and $t_{k+1} - t_k = g(k), \forall k \in \mathbb{N}_+$. Here, $g : [1, \infty) \to [1, \infty)$ is a continuous, non-decreasing function that takes on integer values at integers. We will henceforth refer to $g(k)$ as the *event-interval* function. At any given time $t \in \mathbb{N}_+$, let $\hat{\mu}_{ij,t}(\theta)$ represent agent i's belief on θ the last time (excluding time t) it transmitted its belief on θ to agent j. Our communication strategy is as follows. At t_1, each agent $i \in \mathcal{V}$ broadcasts its entire belief vector $\boldsymbol{\mu}_{i,t}$ to every neighbor. Subsequently, at each $t_k \in \mathbb{I}, k \geq 2$, i transmits $\mu_{i,t_k}(\theta)$ to $j \in \mathcal{N}_i$ if and only if the following event occurs:

$$\mu_{i,t_k}(\theta) < \gamma(t_k) \min\{\hat{\mu}_{ij,t_k}(\theta), \hat{\mu}_{ji,t_k}(\theta)\}, \quad (4)$$

where $\gamma : \mathbb{N} \to (0,1]$ is a non-increasing function, which we will henceforth call the *threshold* function. If $t \notin \mathbb{I}$, then an agent i does not communicate with its neighbors at time t, i.e., all inter-agent interactions are restricted to time-steps in $\mathbb{I}$, subject to the trigger-condition given by (4). We will describe the functional forms of $g(\cdot)$ and $\gamma(\cdot)$ in Section IV.

• **Summary**: At each time-step $t+1 \in \mathbb{N}_+$, and for each hypothesis $\theta \in \Theta$, the sequence of operations executed by

Fig. 1. The figure shows a network where only agent 1 is informative. In Section III, we design an event-triggered algorithm under which all upstream broadcasts along the path $3 \rightarrow 2 \rightarrow 1$ stop eventually almost surely.

an agent i is summarized as follows. (i) Agent i updates its local and actual beliefs on θ via (2) and (3), respectively. (ii) For each neighbor $j \in \mathcal{N}_i$, it decides whether or not to transmit $\mu_{i,t+1}(\theta)$ to j, and collects $\{\mu_{j,t+1}(\theta)\}_{j\in\mathcal{N}_{i,t+1}(\theta)}$.[3] (iii) It updates $\bar{\mu}_{i,t+1}(\theta)$ via (1) using the (potentially) new information it acquires from its neighbors at time $t+1$.

• **Intuition:** The premise of our belief-update strategy is based on diffusing low beliefs on each false hypothesis. For a given false hypothesis θ, the local Bayesian update (2) will generate a decaying sequence $\pi_{i,t}(\theta)$ for each $i \in \mathcal{S}(\theta^*, \theta)$. Update rules (1) and (3) then help propagate agent i's low belief on θ to the rest of the network.

To build intuition regarding our communication strategy, consider the network in Fig 1. Suppose $\Theta = \{\theta_1, \theta_2\}, \theta^* = \theta_1$, and $\mathcal{S}(\theta_1, \theta_2) = 1$, i.e., agent 1 is the only informative agent. Since our approach is based on eliminating each false hypothesis, it makes sense to broadcast beliefs only if they are low enough. Accordingly, a naive way to enforce sparse communication could be to set a fixed low threshold, and wait until beliefs fall below such a threshold to broadcast. However, it is fairly easy to see that this approach will eventually lead to dense communication. The obvious fix is to introduce an event-condition that is *state-dependent*. For instance, suppose an agent broadcasts its belief on a state θ only if it is sufficiently lower than what it was when it last broadcasted about θ. While an improvement over the "fixed-threshold" strategy, this new scheme has the following demerit: broadcasts are not *agent-specific*. In other words, going back to our example, agent 2 (resp., agent 3) might transmit unsolicited information to agent 1 (resp., agent 2) - information, that agent 1 (resp., agent 2) does not require.

Given the above issues, we ask: Is it possible to devise an event-triggered scheme that eventually stops unnecessary broadcasts from agent 3 to 2, and agent 2 to 1, while preserving essential information flow from agent 1 to 2, and agent 2 to 3? *More generally, we seek a triggering rule that can reduce transmissions from uninformative agents to informative agents.* This leads us to the event condition in Eq. (4). For each $\theta \in \Theta$, an agent i broadcasts $\mu_{i,t}(\theta)$ to a neighbor $j \in \mathcal{N}_i$ only if $\mu_{i,t}(\theta)$ has adequate "innovation" with respect to i's last broadcast about θ to j, *and* j's last broadcast about θ to i. A decreasing threshold function $\gamma(t)$ makes it progressively harder to satisfy the event condition in Eq. (4), demanding more innovation to merit broadcast as time progresses. Checking the event condition only at time-steps in $\mathbb{I}$ saves computations, and provides an additional instrument to control communication-sparsity.[4]

[3] If $t+1 \notin \mathbb{I}$, this step gets bypassed, and $\mathcal{N}_{i,t+1}(\theta) = \emptyset, \forall \theta \in \Theta$.

[4] Without the event condition given by (4), our communication strategy would boil down to a simple time-triggered rule, as in our recent work [12].

IV. Main Results

In this section, we state the main results of this paper, and then discuss their implications. With the exception of Theorem 1, the proofs of all our results are omitted here due to space constraints, but can be found in [13]. To state our first result, let us define $G(z) \triangleq \int_1^z g(\tau)d\tau, \forall z \in [1,\infty)$. Let $G^{-1}(\cdot)$ represent the inverse of $G(\cdot)$, i.e., $\forall z \in [1,\infty), G^{-1}(G(z)) = z$. Since $g(\cdot)$ is continuous and takes values in $[1,\infty)$ by definition, $G(\cdot)$ is strictly increasing, unbounded, and continuous, with $G(1) = 0$, and hence, $G^{-1}(z)$ is well-defined for all $z \in [0,\infty)$.

Theorem 1. *Suppose the functions $g(\cdot)$ and $\gamma(\cdot)$ satisfy:*

$$\lim_{t\to\infty} \frac{G(G^{-1}(t)-2)}{t} = \alpha \in (0,1]; \quad \lim_{t\to\infty} \frac{\log(1/\gamma(t))}{t} = 0. \tag{5}$$

Furthermore, suppose the following conditions hold. (i) For every pair of hypotheses $\theta_p, \theta_q \in \Theta$, the source set $\mathcal{S}(\theta_p, \theta_q)$ is non-empty. (ii) The communication graph $\mathcal{G}$ is connected. Then, the event-triggered distributed learning rule governed by (1), (2), (3), and (4) guarantees the following.

- ***(Consistency)**: For each agent $i \in \mathcal{V}$, $\mu_{i,t}(\theta^\star) \to 1$ a.s.*
- ***(Exponentially Fast Rejection of False Hypotheses)**: For each agent $i \in \mathcal{V}$, and for each false hypothesis $\theta \in \Theta \setminus \{\theta^\star\}$, the following holds:*

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq \max_{v\in\mathcal{S}(\theta^\star,\theta)} \alpha^{d(v,i)} K_v(\theta^\star, \theta) \ a.s. \tag{6}$$

□

We prove Theorem 1 in Section V. At this point, it is natural to ask: For what classes of functions $g(\cdot)$ does the above result hold? The following result provides an answer.

Corollary 1. *Suppose the conditions in Theorem 1 hold.*

(i) *Suppose $g(x) = x^p, \forall x \in \mathbb{R}_+$, where p is any positive integer. Then, for each $\theta \in \Theta \setminus \{\theta^\star\}$, and $i \in \mathcal{V}$:*

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq \max_{v\in\mathcal{S}(\theta^\star,\theta)} K_v(\theta^\star, \theta) \ a.s. \tag{7}$$

(ii) *Suppose $g(x) = p^x, \forall x \in \mathbb{R}_+$, where p is any positive integer. Then, for each $\theta \in \Theta \setminus \{\theta^\star\}$, and $i \in \mathcal{V}$:*

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq \max_{v\in\mathcal{S}(\theta^\star,\theta)} \frac{K_v(\theta^\star, \theta)}{p^{2d(v,i)}} \ a.s. \tag{8}$$

□

Clearly, the communication pattern between the agents is at least as sparse as the sequence $\mathbb{I}$. Our event-triggering scheme introduces further sparsity, as we next establish.

Proposition 1. *Suppose the conditions in Theorem 1 are met. Then, there exists $\bar{\Omega} \subseteq \Omega$ such that $\mathbb{P}^{\theta^*}(\bar{\Omega}) = 1$, and for each $\omega \in \bar{\Omega}$, $\exists T_1(\omega), T_2(\omega) < \infty$ such that the following hold.*

(i) *At each $t_k \in \mathbb{I}$ such that $t_k > T_1(\omega)$, $\mathbb{1}_{ij,t_k}(\theta^*) \neq 1, \forall i \in \mathcal{V}$ and $\forall j \in \mathcal{N}_i$.*

(ii) *Consider any* $\theta \neq \theta^*$*, and* $i \notin \mathcal{S}(\theta^*, \theta)$*. Then, at each* $t_k > T_2(\omega)$, $\exists j \in \mathcal{N}_i$ *such that* $\mathbb{1}_{ij,t_k}(\theta) \neq 1$.[5]

□

The next result is an immediate application of Prop. 1.

Corollary 2. *Suppose the conditions in Theorem 1 are met. Additionally, suppose* $\mathcal{G}$ *is a tree graph, and for each pair* $\theta_p, \theta_q \in \Theta$, $|\mathcal{S}(\theta_p, \theta_q)| = 1$*. Consider any* $\theta \neq \theta^*$*, and let* $\mathcal{S}(\theta^*, \theta) = v_\theta$*. Then, each agent* $i \in \mathcal{V} \setminus \{v_\theta\}$ *stops broadcasting its belief on* θ *to its parent in the tree rooted at* v_θ *eventually almost surely.* □

A few comments are now in order.

• **On the nature of $g(\cdot)$ and $\gamma(\cdot)$:** Intuitively, if the event-interval function $g(\cdot)$ does not grow too fast, and the threshold function $\gamma(\cdot)$ does not decay too fast, one should expect things to fall in place. Theorem 1 makes this intuition precise by identifying conditions on $g(\cdot)$ and $\gamma(\cdot)$ that lead to exponentially fast learning of the truth. From (5), we note that $\gamma(\cdot)$ can be any sub-exponentially decaying function. Moreover, Corollary 1 reveals that up to integer constraints, $g(\cdot)$ can be any polynomial or exponential function.

• **Design trade-offs:** What is the price paid for sparse communication? To answer the above question, we set as benchmark the scenario studied in our previous work [8], where we did not account for communication efficiency. There, we showed that each false hypothesis θ gets rejected exponentially fast by every agent at the *network-independent* rate: $\max_{v\in\mathcal{V}} K_v(\theta^*, \theta)$.[6] We note from (6) that it is only the event-interval function $g(\cdot)$ that potentially impacts the learning rate, since $\alpha \leq 1$. However, from claim (i) in Corollary 1, we glean that polynomially growing inter-communication intervals between the agents, coupled with our proposed event-triggering strategy, lead to *no loss in the long-term learning rate relative to the benchmark case in [8]*, i.e., communication-efficiency comes essentially for "free" under this regime. With exponentially growing event-interval functions, one still achieves exponentially fast learning, albeit at a *reduced* learning rate that is network-structure dependent (see Eq. 8). Our results thus capture the trade-offs between sparse communication and the learning rate.

• **Sparse communication introduced by event-triggering:** Observe that being able to eliminate each false hypothesis is enough for learning the true state. In other words, agents need not exchange their beliefs on the true state (of course, no agent knows a priori what the true state is). Our event-triggering scheme precisely achieves this, as evidenced by claim (i) of Proposition 1: every agent stops broadcasting its belief on θ^* eventually almost surely. In addition, an important property of our event-triggering strategy is that it reduces information flow from uninformative agents to informative agents. To see this, consider any false hypothesis $\theta \neq \theta^*$, and an agent $i \notin \mathcal{S}(\theta^*, \theta)$. Since $i \notin \mathcal{S}(\theta^*, \theta)$, agent i's local belief $\pi_{i,t}(\theta)$ will stop decaying eventually, making it impossible for agent i to lower its actual belief $\mu_{i,t}(\theta)$ without the influence of its neighbors. Consequently, when acting alone, i will not be able to leverage its own private signals to generate enough "innovation" in $\mu_{i,t}(\theta)$ to broadcast to the neighbor who most recently contributed to lowering $\mu_{i,t}(\theta)$. The intuition here is simple: an uninformative agent cannot outdo the source of its information. This idea is made precise in claim (ii) of Proposition 1. Moreover, Corollary 2 stipulates that when the baseline graph is a tree, then all upstream broadcasts to informative agents stop after a finite period of time.

A. Asymptotic Learning of the Truth

If asymptotic learning of the true state is all one cares about, i.e., if the convergence rate is no longer a consideration, then one can allow for arbitrarily sparse communication patterns, as we shall now demonstrate. We first allow the baseline graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$ to now change over time. To allow for this generality, we set $\mathbb{I} = \mathbb{N}_+$, i.e., the event condition (4) is now monitored at each time-step. Furthermore, we set $\gamma(t) = \gamma \in (0, 1], \forall t \in \mathbb{N}$. At each time-step $t \in \mathbb{N}_+$, and for each $\theta \in \Theta$, an agent $i \in \mathcal{V}$ decides whether or not to broadcast $\mu_{i,t}(\theta)$ to an instantaneous neighbor $j \in \mathcal{N}_i(t)$ by checking the event condition (4). While checking this condition, if agent i has not yet transmitted to (resp., heard from) agent j about θ prior to time t, then it sets $\hat{\mu}_{ij,t}(\theta)$ (resp., $\hat{\mu}_{ji,t}(\theta)$) to 1. Update rules (1), (2), (3) remain the same, with $\mathcal{N}_{i,t}(\theta)$ now interpreted as $\mathcal{N}_{i,t}(\theta) \triangleq \{j \in \mathcal{N}_i(t) | \mathbb{1}_{ji,t}(\theta) = 1\}$. Finally, by an union graph over an interval $[t_1, t_2]$, we will imply the graph with vertex set $\mathcal{V}$, and edge set $\cup_{\tau=t_1}^{t_2} \mathcal{E}(\tau)$.

Theorem 2. *Suppose for every pair of hypotheses* $\theta_p, \theta_q \in \Theta$, $\mathcal{S}(\theta_p, \theta_q)$ *is non-empty. Furthermore, suppose for each* $t \in \mathbb{N}_+$*, the union graph over* $[t, \infty)$ *is rooted at* $\mathcal{S}(\theta_p, \theta_q)$*. Then, the event-triggered distributed learning rule described above guarantees* $\mu_{i,t}(\theta^*) \to 1$ *a.s.* $\forall i \in \mathcal{V}$. □

While a result of the above flavor is well known for the basic consensus setting [14], we are unaware of its analogue for the distributed inference problem. When $\mathcal{G}(t) = \mathcal{G}, \forall t \in \mathbb{N}$, we observe from Theorem 2 that, as long as each agent transmits its belief vector to every neighbor infinitely often, all agents will asymptotically learn the truth, without any other constraints on the *frequency* of agent interactions.

V. Proof of Theorem 1

We start with the following useful result.

Lemma 1. *Suppose the conditions in Theorem 1 hold. Then, there exists a set* $\bar{\Omega} \subseteq \Omega$ *with the following properties. (i)* $\mathbb{P}^{\theta^\star}(\bar{\Omega}) = 1$*. (ii) For each* $\omega \in \bar{\Omega}$*, there exist constants* $\eta(\omega) \in (0, 1)$ *and* $t'(\omega) \in (0, \infty)$ *such that*

$$\pi_{i,t}(\theta^\star) \geq \eta(\omega), \bar{\mu}_{i,t}(\theta^\star) \geq \eta(\omega), \forall t \geq t'(\omega), \forall i \in \mathcal{V}. \quad (9)$$

(iii) Consider a false hypothesis $\theta \neq \theta^*$*, and an agent* $i \in \mathcal{S}(\theta^*, \theta)$*. Then on each sample path* $\omega \in \bar{\Omega}$*, we have:*

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq K_i(\theta^\star, \theta). \quad (10)$$

□

[5] In this claim, j might depend on t_k.

[6] In contrast, for linear [1], [2] and log-linear [3]–[6] rules, the corresponding rate is a convex combination of the relative entropies $K_v(\theta^*, \theta), v \in \mathcal{V}$.

The proofs of claims (ii) and (iii) in the above Lemma essentially follow the same arguments as that of [8, Lemma 2] and [8, Lemma 3], respectively. The following result will be the key ingredient in proving Theorem 1.

Lemma 2. *Consider a false hypothesis $\theta \in \Theta \setminus \{\theta^\star\}$ and an agent $v \in \mathcal{S}(\theta^\star, \theta)$. Suppose the conditions stated in Theorem 1 hold. Then, the following is true for each agent $i \in \mathcal{V}$:*

$$\liminf_{t\to\infty} -\frac{\log \mu_{i,t}(\theta)}{t} \geq \alpha^{d(v,i)} K_v(\theta^\star, \theta) \; a.s. \tag{11}$$

□

Proof. Let $\bar{\Omega} \subseteq \Omega$ be the set of sample paths for which assertions (i)-(iii) of Lemma 1 hold. Fix a sample path $\omega \in \bar{\Omega}$, an agent $v \in \mathcal{S}(\theta^\star, \theta)$, and an agent $i \in \mathcal{V}$. When $i = v$, the assertion of Eq. (11) follows directly from Eq. (10) in Lemma 1. In particular, this implies that for a fixed $\epsilon > 0$, $\exists t_v(\omega, \theta, \epsilon)$, such that:

$$\mu_{v,t}(\theta) < e^{-(K_v(\theta^\star,\theta)-\epsilon)t}, \forall t \geq t_v(\omega, \theta, \epsilon). \tag{12}$$

Moreover, since $\omega \in \bar{\Omega}$, Lemma 1 guarantees the existence of a time-step $t'(\omega) < \infty$, and a constant $\eta(\omega) > 0$, such that on ω, $\pi_{i,t}(\theta^\star) \geq \eta(\omega), \bar{\mu}_{i,t}(\theta^\star) \geq \eta(\omega), \forall t \geq t'(\omega), \forall i \in \mathcal{V}$. Let $\bar{t}_v(\omega, \theta, \epsilon) = \max\{t'(\omega), t_v(\omega, \theta, \epsilon)\}$. Let $t_q > \bar{t}_v$ be the first event-monitoring time-step in $\mathbb{I}$ that is larger than $\bar{t}_v$.[7] Now consider any $t_k \in \mathbb{I}$ such that $k \geq q$. In what follows, we will analyze the implications of agent v deciding whether or not to broadcast its belief on θ to a one-hop neighbor $j \in \mathcal{N}_v$ at t_k. To this end, we consider the following two cases.

Case 1: $\mathbb{1}_{vj,t_k}(\theta) = 1$, i.e., v broadcasts $\mu_{v,t_k}(\theta)$ to j at t_k. Thus, since $v \in \mathcal{N}_{j,t_k}(\theta)$, we have $\bar{\mu}_{j,t_k}(\theta) \leq \mu_{v,t_k}(\theta)$ from (1). Let us now observe that $\forall t \geq t_k + 1$:

$$\begin{aligned}\mu_{j,t}(\theta) &\overset{(a)}{\leq} \frac{\bar{\mu}_{j,t-1}(\theta)}{\sum_{p=1}^{m} \min\{\bar{\mu}_{j,t-1}(\theta_p), \pi_{j,t}(\theta_p)\}} \\ &\overset{(b)}{\leq} \frac{\mu_{v,t_k}(\theta)}{\sum_{p=1}^{m} \min\{\bar{\mu}_{j,t-1}(\theta_p), \pi_{j,t}(\theta_p)\}} \overset{(c)}{<} \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_k}}{\eta}.\end{aligned} \tag{13}$$

In the above inequalities, (a) follows directly from (3), (b) follows by noting that the sequence $\{\bar{\mu}_{j,t}(\theta)\}$ is non-increasing based on (1), and (c) follows from (12) and the fact that all beliefs on $\theta^\star$ are bounded below by η for $t \geq \bar{t}_v$.

Case 2: $\mathbb{1}_{vj,t_k}(\theta) \neq 1$, i.e., v does not broadcast $\mu_{v,t_k}(\theta)$ to j at t_k. From the event condition in (4), it must then be that at least one of the following is true: (a) $\mu_{v,t_k}(\theta) \geq \gamma(t_k)\hat{\mu}_{vj,t_k}(\theta)$, and (b) $\mu_{v,t_k}(\theta) \geq \gamma(t_k)\hat{\mu}_{jv,t_k}(\theta)$. Suppose $\mu_{v,t_k}(\theta) \geq \gamma(t_k)\hat{\mu}_{vj,t_k}(\theta)$. From (12), we then have:

$$\hat{\mu}_{vj,t_k}(\theta) \leq \frac{\mu_{v,t_k}(\theta)}{\gamma(t_k)} < \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_k}}{\gamma(t_k)}. \tag{14}$$

In words, the above inequality places an upper bound on the belief of agent v on θ when it last transmitted its belief on θ to agent j, *prior* to time-step t_k; at least one such transmission is guaranteed to take place since all agents broadcast their entire belief vectors to their neighbors at t_1. Noting that $\bar{\mu}_{j,t}(\theta) \leq \hat{\mu}_{vj,t_k}(\theta), \forall t \geq t_k$, using (3), (14), and arguments similar to those for arriving at (13), we obtain:

$$\mu_{j,t}(\theta) < \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_k}}{\eta\gamma(t_k)} \leq \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_k}}{\eta\gamma(t)}, \forall t \geq t_k+1, \tag{15}$$

where the last inequality follows from the fact that $\gamma(\cdot)$ is a non-increasing function of its argument. Now consider the case when $\mu_{v,t_k}(\theta) \geq \gamma(t_k)\hat{\mu}_{jv,t_k}(\theta)$. Following the same reasoning as before, we can arrive at an identical upper-bound on $\hat{\mu}_{jv,t_k}(\theta)$ as in (14). Using the definition of $\hat{\mu}_{jv,t_k}(\theta)$, and the fact that agent j incorporates its own belief on θ in the update rule (1), we have that $\bar{\mu}_{j,t}(\theta) \leq \hat{\mu}_{jv,t_k}(\theta), \forall t \geq t_k$. Using similar arguments as before, observe that the bound in (15) holds for this case too.

Combining the analyses of cases 1 and 2, referring to (13) and (15), and noting that $\gamma(t) \in (0,1], \forall t \in \mathbb{N}$, we conclude that the bound in (15) holds for each $t_k \in \mathbb{I}$ such that $t_k > \bar{t}_v$. Now since $t_{k+1} - t_k = g(k)$, for any $\tau \in \mathbb{N}_+$ we have:

$$t_{q+\tau} = t_q + \sum_{z=q}^{q+\tau-1} g(z). \tag{16}$$

Next, noting that $g(\cdot)$ is non-decreasing, observe that:

$$t_q + \int_q^{q+\tau} g(z-1)dz \leq t_{q+\tau} \leq t_q + \int_q^{q+\tau} g(z)dz. \tag{17}$$

The above yields: $l(q,\tau) \triangleq t_q + G(q+\tau-1) - G(q-1) \leq t_{q+\tau} \leq t_q + G(q+\tau) - G(q) \triangleq u(q,\tau)$. Fix any time-step $t > u(q,1)$, let $\tau(t)$ be the largest index such that $u(q,\tau(t)) < t$, and $\bar{\tau}(t)$ be the largest index such that $t_{q+\bar{\tau}(t)} < t$. Observe:

$$\bar{t}_v < t_q < t_{q+1} \leq t_{q+\tau(t)} \leq t_{q+\bar{\tau}(t)} < t. \tag{18}$$

Using the above inequality, the fact that $l(q,\tau(t)) \leq t_{q+\tau(t)}$, and referring to (15), we obtain:

$$\mu_{j,t}(\theta) < \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)t_{q+\bar{\tau}(t)}}}{\eta\gamma(t)} \leq \frac{e^{-(K_v(\theta^\star,\theta)-\epsilon)l(q,\tau(t))}}{\eta\gamma(t)}. \tag{19}$$

From the definitions of $\tau(t)$ and $u(q,\tau)$, we have $q + \tau(t) = \lceil G^{-1}(t - t_q + G(q)) \rceil - 1$. This yields:

$$\begin{aligned} l(q,\tau(t)) &= t_q + G(\lceil G^{-1}(t - t_q + G(q)) \rceil - 2) - G(q-1) \\ &\geq t_q + G(G^{-1}(t - t_q + G(q)) - 2) - G(q-1). \end{aligned} \tag{20}$$

From (19) and (20), we obtain the following $\forall t > u(q,1)$:

$$-\frac{\log \mu_{j,t}(\theta)}{t} > \frac{\tilde{G}(t)}{t}(K_v(\theta^\star,\theta) - \epsilon) - \frac{\log c}{t} - \frac{\log(1/\gamma(t))}{t}, \tag{21}$$

where $\tilde{G}(t) = G(G^{-1}(t - t_q + G(q)) - 2)$, and $c = e^{-(K_v(\theta^\star,\theta)-\epsilon)(t_q - G(q-1))}/\eta$. Now taking the limit inferior on both sides of (21) and using (5) yields:

$$\liminf_{t\to\infty} -\frac{\log \mu_{j,t}(\theta)}{t} \geq \alpha(K_v(\theta^\star,\theta) - \epsilon). \tag{22}$$

Finally, since the above inequality holds for any sample path $\omega \in \bar{\Omega}$, and an arbitrarily small ϵ, it follows that the assertion

[7] We will henceforth suppress the dependence of various quantities on ω, θ, and ϵ for brevity.

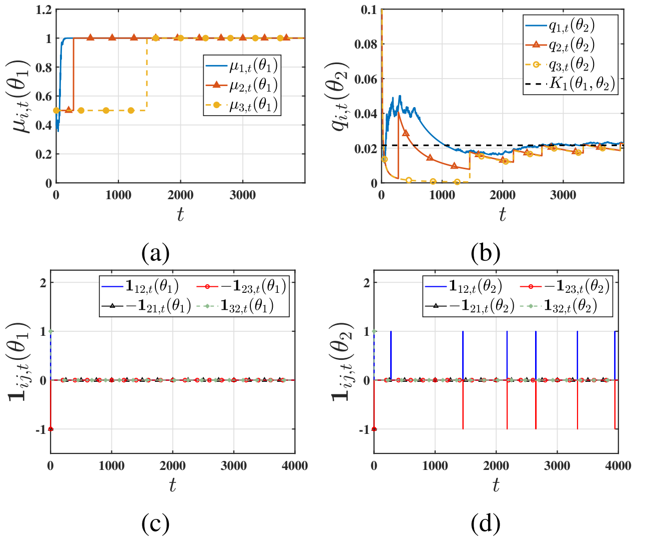

Fig. 2. Plots pertaining to the simulation example in Sec. VI. Fig. 2(a) plots the belief evolutions on the true state θ_1. Fig. 2(b) plots the rate at which each agent rejects the false hypothesis θ_2, namely $q_{i,t}(\theta_2) = -\log(\mu_{i,t}(\theta_2))/t$. Fig.'s 2(c) and 2(d) demonstrate the sparse communication patterns generated by our event-triggering scheme.

in (11) is true for every one-hop neighbor j of agent v. Now consider any agent i such that $d(v,i) = 2$. Clearly, there must exist some $j \in \mathcal{N}_v$ such that $i \in \mathcal{N}_j$. Following identical arguments as before, it is easy to see that $\mu_{i,t}(\theta)$ decays exponentially at a rate that is at least α times the rate at which $\mu_{j,t}(\theta)$ decays to zero. From (22), the latter rate is at least $\alpha K_v(\theta^*, \theta)$, and hence, the former is at least $\alpha^2 K_v(\theta^*, \theta)$. This establishes the claim of the lemma for all two-hop neighbors of agent v. Noting that $\mathcal{G}$ is connected, the proof can be easily completed by induction. □

Proof. (**Theorem 1**) Fix a $\theta \in \Theta \setminus \{\theta^\star\}$. Based on condition (i) of the Theorem, $\mathcal{S}(\theta^\star, \theta)$ is non-empty, and based on condition (ii), there exists a path from each agent $v \in \mathcal{S}(\theta^\star, \theta)$ to every agent $i \in \mathcal{V} \setminus \{v\}$; Eq. (6) then follows from Lemma 2. By definition of a source set, $K_v(\theta^\star, \theta) > 0, \forall v \in \mathcal{S}(\theta^\star, \theta)$; Eq. (6) then implies $\lim_{t\to\infty} \mu_{i,t}(\theta) = 0$ a.s., $\forall i \in \mathcal{V}$. □

VI. A Simulation Example

In this section, we illustrate our theoretical findings via a simple simulation example. To do so, we consider the network in Fig. 1. Suppose $\Theta = \{\theta_1, \theta_2\}, \theta^* = \theta_1$, and let the signal space for each agent be $\{0, 1\}$. The likelihood models are as follows: $l_1(0|\theta_1) = 0.7, l_1(0|\theta_2) = 0.6$, and $l_i(0|\theta_1) = l_i(0|\theta_2) = 0.5, \forall i \in \{2, 3\}$. Clearly, agent 1 is the only informative agent. To isolate the impact of our event-triggering strategy, we set $g(k) = 1, \forall k \in \mathbb{N}_+$, i.e., the event condition in Eq. (4) is monitored at every time-step. We set the threshold function as $\gamma(k) = 1/k^2$. The performance of our algorithm is depicted in Fig. 2. We make the following observations. (i) From Fig. 2(a), we note that all agents eventually learn the truth. (ii) From Fig. 2(b), we note that the asymptotic rate of rejection of the false hypothesis θ_2, namely $q_{i,t}(\theta_2) = -\log(\mu_{i,t}(\theta_2))/t$, agrees with the theoretical bound in Thm. 1. (iii) From Fig. 2(c), we note that after the first time-step, all agents stop broadcasting about the true state θ_1, complying with claim (i) of Prop. 1. (iv) From Fig. 2(d), we note that broadcasts about θ_2 along the path $3 \to 2 \to 1$ stop after the first time-step, in accordance with claim (ii) of Prop. 1, and Corr. 2. We also observe that in the first 4000 time-steps, agent 1 (resp., agent 2) broadcasts its belief on θ_2 to agent 2 (resp., agent 3) only 7 times (resp., 6 times). Despite such drastic reduction in the communication frequency, all agents still learn the truth at the same learning rate as with the baseline algorithm in [8].

VII. Conclusion

We introduced a new event-triggered distributed learning rule and established that it leads to exponentially fast learning of the true state. In particular, we identified sparse communication regimes where the inter-communication intervals between the agents grow unbounded, with no loss in the long-term learning rate. We then demonstrated, both in theory and in simulations, that our event-triggered scheme has the ability to drastically reduce information flow from uninformative agents to informative agents in the network. As future work, we plan to investigate the complementary direction of compressing information. Exploring connections to Federated Learning is also of interest [15].

References

[1] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, "Non-Bayesian social learning," *Games and Economic Behavior*, vol. 76, no. 1, pp. 210–225, 2012.

[2] A. Jadbabaie, P. Molavi, and A. Tahbaz-Salehi, "Information heterogeneity and the speed of learning in social networks," *Columbia Bus. Sch. Res. Paper*, pp. 13–28, 2013.

[3] S. Shahrampour, A. Rakhlin, and A. Jadbabaie, "Distributed detection: Finite-time analysis and impact of network topology," *IEEE Trans. on Autom. Control*, vol. 61, no. 11, pp. 3256–3268, 2016.

[4] A. Nedić, A. Olshevsky, and C. A. Uribe, "Fast convergence rates for distributed Non-Bayesian learning," *IEEE Trans. on Autom. Control*, vol. 62, no. 11, pp. 5538–5553, 2017.

[5] A. Lalitha, T. Javidi, and A. Sarwate, "Social learning and distributed hypothesis testing," *IEEE Trans. on Info. Theory*, vol. 64, no. 9, 2018.

[6] C. A. Uribe, J. Z. Hare, L. Kaplan, and A. Jadbabaie, "Non-Bayesian social learning with uncertain models over time-varying directed graphs," *arXiv preprint arXiv:1909.04255*, 2019.

[7] A. Mitra, J. A. Richards, and S. Sundaram, "A new approach for distributed hypothesis testing with extensions to Byzantine-resilience," in *Proc. of the American Control Conference*, 2019, pp. 261–266.

[8] ——, "A new approach to distributed hypothesis testing and non-Bayesian learning: Improved learning rate and Byzantine-resilience," *arXiv:1907.03588*, 2019.

[9] P. Tabuada, "Event-triggered real-time scheduling of stabilizing control tasks," *IEEE Trans. on Autom. Control*, vol. 52, no. 9, pp. 1680–1685, 2007.

[10] W. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *Proc. of the Conf. on Decision and Control*. IEEE, 2012, pp. 3270–3285.

[11] C. Nowzari, E. Garcia, and J. Cortés, "Event-triggered communication and control of networked systems for multi-agent consensus," *Automatica*, vol. 105, pp. 1–27, 2019.

[12] A. Mitra, J. A. Richards, and S. Sundaram, "A communication-efficient algorithm for exponentially fast non-Bayesian learning in networks," in *Proc. of the Conf. on Decision and Control*, 2019, pp. 8347–8352.

[13] A. Mitra, S. Bagchi, and S. Sundaram, "Event-triggered distributed inference," *arXiv preprint arXiv:2004.01302*, 2020.

[14] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Trans. on Autom. control*, vol. 50, no. 2, pp. 169–182, 2005.

[15] A. Lalitha, O. C. Kilinc, T. Javidi, and F. Koushanfar, "Peer-to-peer Federated Learning on graphs," *arXiv:1901.11173*, 2019.