Type-Sensitive Social Learning

Joni Shaska
University of Southern California
shaska@usc.edu

Urbashi Mitra
University of Southern California
ubli@usc.edu

***Abstract*—The problem of distributed hypothesis testing with correlated observations is studied. Specifically, systems in which the behavior is governed by both the underlying hypothesis, as well as an underlying empirical distribution on the network state is considered. Thus, there is significant coupling between the interim decisions of the agents and the signals they transmit. The current model addresses increased coupling relative to prior work. The optimal decay rate for optimal detection is computed; key properties associated with this error rate are derived. The utility of the analysis is shown via the consideration of a multi-class problem wherein agents within each class have specific properties and interact with agents of other classes via signal enhancement or jamming. This multi-class case is studied numerically and it is shown that there is a optimal ratio between class populations that maximizes the decay rate of the error.**

***Index Terms*—decentralized detection, multi-agent networks, state-dependent networks, error exponents**

I. Introduction

Decentralized detection in wireless networks has been persistently studied over the years [1], [2]. Despite its long history, the problem remains of interest as different contexts are considered. Mainly, a large body of recent work has gone into decentralized detection with correlated observations [3]–[5]. Unfortunately, the problem of decentralized detection with correlated observations is NP-Hard [6]. However, many applications involve sensors with correlated observations and coupled signaling. We provide a few motivating examples.

In cyber-security for smart grids [7], the initial state of the grid and the number of nodes that have been attacked affect the signals received at each node and how the nodes react. Thus, there is coupling between the received signals and the network state. Human-decision making is often affected by environmental conditions and the decisions and psychological states of others in a community [8]; abstractions of these couplings can be modeled in our framework. Multi-species microbial communities engage in different collective behaviors (e.g. forming a biofilm, inducing a quorum) as a function of the individual states of the microbes in each community. One species of bacteria can release antibiotics that are detrimental to another species. While our prior work has modeled quorum sensing as a decentralized decision making process [9], the level of coupling considered herein was not present.

In particular, [10] addressed the generalization of decision making wherein each agent observes signals due to a common, unknown hypothesis, but each agent is affected by their individual state. In [10], optimal agent rules are found and the asymptotic performance is studied. However, there are some weaknesses in this previous work. Specifically, analysis is done assuming agents' have a local look at the network state, and that all statistical dependencies are limited to local, as opposed to global, dependencies. Also, it is assumed that the fusion center has total knowledge of the network state, a rather strong assumption.

This work has been funded in part by one or more of the following grants: ONR 00014-15-1-2550, NSF CCF-1817200, ARO W911NF1910269, Cisco Foundation 1980393, DOE DE-SC0021417, Swedish Research Council 2018-04359, NSF CCF-2008927, ONR 503400-78050.

We set out to alleviate some of these issues. In particular, we assume that agents' observations depend not only on the underlying hypothesis, but also the empirical distribution of the network state. Hence, the statistical dependencies of agents' observations are global, as opposed to local in [10]. Moreover, we assume the fusion center knows the empirical distribution of the network state, but not necessarily the overall network state itself. This assumption is significantly weaker than that in [10], but helps facilitate design and analysis while alleviating issues related to computational complexity at the fusion center. This framework allows us to model multi-agent interactions where the agents are drawn from a finite number of distinct classes. Furthermore, we can allow for the agents in different classes to have distinct coupling with agents from another class.

Our contributions are as follows:

1) We introduce the concept of using the underlying empirical distribution of the network state to alleviate statistical dependencies.
2) We analyze the error exponent of the proposed system and show a number of desirable properties. One such property being that the error exponent collapses to a single distribution, alleviating several series design challenges.
3) We formalize the problem of computing optimal population ratios, and use our results to analyze system performance as a function of the agent ratios.

II. Problem Formulation

Consider the setup depicted in Fig. 1. a set of n nodes are oriented in a parallel configuration. Each agent receives an observation consisting of the random variable $Y_k \in \mathcal{Y}$, which we call the *signal*, and $X_k \in \mathcal{S}$, which we call the *state of agent* k, $k = 1, 2, ..., n$. All agents observe the same underlying hypothesis. We assume that each agent is in one of m states: $X_k \in \{0, 1, ..., m-1\}$, $k = 1, 2, ..., n$. The state vector $\boldsymbol{X}$ has a prior probability $q(\boldsymbol{x})$. We denote the type (empirical distribution) of $\boldsymbol{X}$ as $\boldsymbol{Z}_n$. The goal of the fusion center is to assess which of the two possible hypotheses,

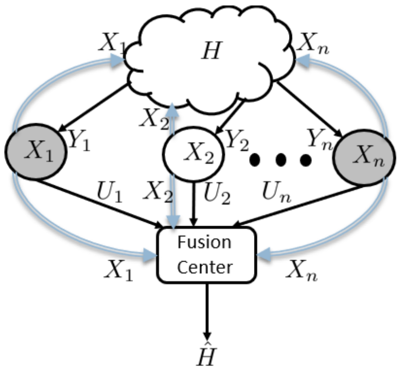

Fig. 1. Problem setup

$H \in \{0, 1\}$, is true, with conditional probability $p(h|\boldsymbol{z}_n)$. All agents observe the same underlying hypothesis. For simplicity, we denote the conditional pmf (pdf) of a random variable X conditioned on $H = h$, $h \in \{0, 1\}$, as $p_h(x)$. Also, the pmf (pdf) of X conditioned on $Y = y$ and $H = h$ is denoted as $p_h(x|y)$. Each agent receives a local observation $Y_k \in \mathcal{Y}$, $k = 1, ..., n$, distributed according to $p_h(y_k|\boldsymbol{z}_n)$. Moreover, we assume independence conditioned on the hypothesis and network type, namely,

$$p_h(\boldsymbol{y}|\boldsymbol{z}_n) = \prod_k p_h(y_k|\boldsymbol{z}_n). \tag{1}$$

After receiving signal Y_k, each agent determines a local decision, $U_k = \gamma_k(Y_k) \in \mathcal{U} = \{0, 1, ..., b-1\}$. These local decisions are then sent to the fusion center along with the system state for the final decision. The fusion center output is given by $U_0 = \psi(\boldsymbol{U}, \boldsymbol{Z}_n) \in \{0, 1\}$. Let the set Γ be the set of all decision rules and Ψ be the set of all fusion rules. We call a collection of decision rules $\gamma_k \in \Gamma$, $k = 1, 2, ..., n$, together with a fusion rule $\psi \in \Psi$ a *strategy* denoted by $\boldsymbol{\psi} \in \Gamma^n \times \Psi$, where Γ^n is the Cartesian product of Γ with itself n times.

It is often the case that one wishes to design the system so as to minimize the probability of error. That is, one wishes to solve the optimization problem given by

$$\inf_{\boldsymbol{\psi}} \mathbb{P}(U_0 \neq H). \tag{2}$$

To minimize the probability of error, the optimal rule at the fusion center is given by the maximum a posteriori (MAP) rule [1], regardless of the rules used at the individual agents. Hence, we only concern ourselves with the optimization over the agents' rules $\boldsymbol{\gamma}$. Moreover, we focus our attention to asymptotically large networks, and so we instead choose to optimize the asymptotic error rate, or *error exponent* of the system. The error exponent is given as

$$-\lim_n \frac{1}{n} \log \mathbb{P}(U_0 \neq H). \tag{3}$$

Following proof techniques as in [10], [11], one can show that the error exponent is in fact equal to

$$-\lim \min_{s\in[0,1]} \frac{1}{n} \log \sum_{\boldsymbol{u},\boldsymbol{z}} p_0(\boldsymbol{u}|\boldsymbol{z}_n)^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z}_n)^s q(\boldsymbol{z}_n). \tag{4}$$

As a result, we elect to restrict our attention to the following problem,

$$\inf_{\boldsymbol{\gamma}} \min_{s\in[0,1]} \frac{1}{n} \log \sum_{\boldsymbol{u},\boldsymbol{z}_n} p_0(\boldsymbol{u}|\boldsymbol{z}_n)^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z}_n)^s q(\boldsymbol{z}_n). \tag{5}$$

Notice that $\boldsymbol{z}_n$ is a function of *all* states X_k for $k = 1, 2, ..., n$. Hence, the received observation Y_k of agent k is a function of *the behavior of all agents in the network*. That is, the given signal model accounts for *global network interference* across the agents.

III. Asymptotic Analysis

Before we begin our analysis, we introduce a few definitions and concepts that we employ throughout our analysis. Let $\mathcal{P}^m$ denote the probability simplex in $\mathbb{R}^m$. That is,

$$\mathcal{P}^m = \{\boldsymbol{z} \in \mathbb{R}^m : z_i \geq 0, \sum_i z_i = 1\}. \tag{6}$$

As stated in the problem formulation, the signal of agent k is distributed according to $p_h(y_k|\boldsymbol{z})$.[1] Hence, the densities of the agents' are indexed by elements in $\mathcal{P}^m$. We would like these densities to have certain properties over $\mathcal{P}^m$, upon which we now elaborate.

Let $\mathcal{Y}$ be an outcome space. Let $\mathcal{F}$ be a family of probability densities with respect to some underlying measure μ, denoted by $f_{\boldsymbol{z}}$. Then, we make the following assumption.

Assumption 1. *$\forall \epsilon > 0$, $\exists \delta > 0$ such that if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are any two points in $\mathcal{P}^m$ that satisfy $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta$, then*

$$\int |f_{\boldsymbol{\alpha}} - f_{\boldsymbol{\beta}}| d\mu < \epsilon \tag{7}$$

We now give a convenient condition to check whether a given family $\mathcal{F}$ is satisfies Assumption 1.

Lemma 1. *If $f_{\boldsymbol{\alpha}} \longrightarrow f_{\boldsymbol{\beta}}$ a.e. whenever $\boldsymbol{\alpha} \longrightarrow \boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathcal{P}^m$, then $\mathcal{F}$ satisfies Assumption 1.*

For the remainder of this correspondence, we make the following assumptions.

Assumption 2. *We assume agents are identical, that is, for all $n \in \mathbb{N}$:*

1) Agent states are i.i.d.
2) All agents have the same signal model $p_h(y|\boldsymbol{z}_n)$.
3)

$$\inf_{\boldsymbol{\gamma}} \min_{s\in[0,1]} \inf_{\boldsymbol{z}\in\mathcal{P}^m} \sum_u p_0(u|\boldsymbol{z})^{1-s} p_1(u|\boldsymbol{z})^s 2^{-D(\boldsymbol{z}||q)} > 0. \tag{8}$$

Moreover, while the optimization of s takes place over the interval $[0, 1]$, we restrict ourselves to the interval $[\epsilon_0, 1-\epsilon_0]$, for any $\epsilon_0 > 0$. The reason for this is purely technical, and do not change any of the statements regarding optimality. We are now ready to state the main theorem.

[1] When we write $\boldsymbol{z}_n$, we are referring to a specific type class for a sequence of length n When we drop the n subscript and write only $\boldsymbol{z}$, we are referring to an element in $\mathcal{P}^m$.

Theorem 1. *Fix any $\epsilon_0 > 0$. For any $s \in [\epsilon_0, 1-\epsilon_0]$. Then, subject to assumptions 1 and 2, we have*

$$\lim_n \inf_{\gamma} \min_{s\in[\epsilon,1-\epsilon]} \frac{1}{n} \log \sum_{\boldsymbol{u},\boldsymbol{z}_n} p_0(\boldsymbol{u}|\boldsymbol{z}_n)^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z}_n)^s q(\boldsymbol{z}_n) = \lim_n \inf_{\gamma} \inf_{s\in[\epsilon,1-\epsilon]} \max_{\boldsymbol{z}\in\mathcal{P}^m} \frac{1}{n} \log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z})^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z})^s 2^{-nD(\boldsymbol{z}||q)} \quad (9)$$

where $D(\boldsymbol{z}||q)$ is the Kl-divergence between the mass function $\boldsymbol{z} \in \mathcal{P}^m$ and the true state mass function q.

We provide an interpretation of the above theorem. Define $\boldsymbol{z}^*(n)$ as

$$\boldsymbol{z}^*(n) = \arg\max_{\boldsymbol{z}\in\mathcal{P}^m} \sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z})^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z})^s 2^{-nD(\boldsymbol{z}||q)} \quad (10)$$

$$= \arg\max_{\boldsymbol{z}\in\mathcal{P}^m} \frac{1}{n} \log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z})^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z})^s 2^{-nD(\boldsymbol{z}||q)} \quad (11)$$

$$= \arg\max_{\boldsymbol{z}\in\mathcal{P}^m} -D(\boldsymbol{z}||q) + \frac{1}{n} \log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z})^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z})^s. \quad (12)$$

Notice that second term is the classical Chernoff information corresponding to the fixed distributions $p_h(\boldsymbol{u}|\boldsymbol{z})$, $h = 0, 1$, and the KL-divergence term can be thought of as a bias. Hence, $\boldsymbol{z}^*(n)$ denotes the m-dimensional probability vector that yields the worst Chernoff information, biased by the KL-divergence. In some sense, $\boldsymbol{z}$ is sufficiently close to q so that its poor performance cannot be ignored even in asymptotically large networks. Moreover, an important note of the above theorem is that *only one distribution in $\mathcal{P}^m$ dominates the asymptotic performance*. This equivalent exponent admits a number of desirable properties that we discuss in the sequel, such as identical rules being optimal for identical agents, as well as characterizing a region that contains $\boldsymbol{z}^*(n)$.

We are now ready to give the important points of the proof of Theorem 1. We omit full proofs for brevity and space constraints.

The first step in proving the main result is to notice the following. Assume $\mathcal{F}_h$ is a family of probability densities $p_h(y|\boldsymbol{z})$ that satisfy Assumption 1. Then, for any $\epsilon > 0$, there exists a $\delta > 0$, such that for all $\gamma \in \Gamma$ and all $u \in \mathcal{U}$, if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are any two points in $\mathcal{P}^m$ that satisfy $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta$, then

$$|p_h(u|\boldsymbol{\alpha}) - p_h(u|\boldsymbol{\beta})| < \epsilon \quad (13)$$

That is, we are claiming that for any $\epsilon > 0$, *the same δ works for all γ and messages u*. Hence, the "error" incurred from assuming $\boldsymbol{\alpha}$ when $\boldsymbol{\beta}$ is true is dependent *only on the distance between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and not on the actual strategy being used*. This statement will prove immensely useful later on.

We now concern ourselves with the continuity of the function

$$p_0(u|\boldsymbol{z})^{1-s} p_1(u|\boldsymbol{z})^s 2^{-D(\boldsymbol{z}||q)} \quad (14)$$

over $\mathcal{P}^m$. Notice that this expression is for a single agent. The uniform continuity of $p_0(u|\boldsymbol{z})$ and $p_1(u|\boldsymbol{z})$ follows from (13). Moreover, notice that $\mathcal{P}^m$ is compact, and that $D(\boldsymbol{z}||q)$ is also uniformly continuous on $\mathcal{P}^m$ (assuming finite support, i.e., $m < \infty$). Hence, using a series of composition and product arguments (the composition and product of uniformly continuous functions on a compact set is uniformly continuous), one can show that (14) is indeed uniformly continuous on $\mathbf{P}^m$ for any $s \in [0, 1]$. This gives us the following.

Lemma 2. *For any $\epsilon > 0$, there exists a $\delta > 0$, independent of γ, such that if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ satisfy $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta$, then*

$$\left| \frac{\sum_u p_0(u|\boldsymbol{\alpha})^{1-s} p_1(u|\boldsymbol{\alpha})^s 2^{-D(\alpha||q)}}{\sum_u p_0(u|\boldsymbol{\beta})^{1-s} p_1(u|\boldsymbol{\beta})^s 2^{-D(\beta||q)}} - 1 \right| < \epsilon. \quad (15)$$

Recall that all agents are identical. Hence, $p_h(u|\boldsymbol{z})$ varies only according to γ. Since δ does not depend on γ, all agents satisfy (15) with the same δ. That is, even if agents are using different rules, the difference in performance in cannot be too large among the agents. Moreover, δ does not depend on any point in $\mathcal{P}^m$. It is worth pointing out this fact, since $\boldsymbol{z}^*(n)$ may change with n. Hence, regardless of how $\boldsymbol{z}^*(n)$ changes, all agents will still satisfy (15) with the same δ. Furthermore, for notational simplicity, we write $\boldsymbol{z}^*(n)$ as $\boldsymbol{z}^*$, as we hope the dependency on n is clear. Finally, we have the following.

Lemma 3. *For any $s \in [0, 1]$,*

$$\lim_n \frac{1}{n} \log \sum_{\boldsymbol{u},\boldsymbol{z}_n} p_0(\boldsymbol{u}|\boldsymbol{z}_n)^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z}_n)^s q(\boldsymbol{z}_n) = \lim_n \frac{1}{n} \log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z}^*(n))^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z}^*(n))^s 2^{-nD(\boldsymbol{z}^*(n)||q)}. \quad (16)$$

where

$$\boldsymbol{z}^*(n) = \arg\max_{\boldsymbol{z}\in\mathcal{P}^m} \sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z})^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z})^s 2^{-nD(\boldsymbol{z}||q)}. \quad (17)$$

We briefly discuss the major points of the proof for Lemma 3. By the preceding lemma, for any $\epsilon > 0$, $\exists\delta > 0$, independent of γ, such that whenever $\|\boldsymbol{z}_n - \boldsymbol{z}^*(n)\|_2 < \delta$,

$$\left| \frac{\sum_u p_0(u|\boldsymbol{z}_n)^{1-s} p_1(u|\boldsymbol{z}_n)^s 2^{-D(\boldsymbol{z}_n||q)}}{\sum_u p_0(u|\boldsymbol{z}^*)^{1-s} p_1(u|\boldsymbol{z}^*)^s 2^{-D(\boldsymbol{z}^*||q)}} - 1 \right| < \epsilon \quad (18)$$

for all agents. Moreover, notice that

$$\frac{\sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z}_n)^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z}_n)^s 2^{-nD(\boldsymbol{z}_n||q)}}{\sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z}^*)^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z}^*)^s 2^{-nD(\boldsymbol{z}^*||q)}} \quad (19)$$

$$= \frac{\sum_{\boldsymbol{u}} \left[\prod_k p_0(u_k|\boldsymbol{z}_n)\right]^{1-s} \left[\prod_k p_1(u_k|\boldsymbol{z}_n)\right]^s \prod_k 2^{-D(\boldsymbol{z}_n||q)}}{\sum_{\boldsymbol{u}} \left[\prod_k p_0(u_k|\boldsymbol{z}^*)\right]^{1-s} \left[\prod_k p_1(u_k|\boldsymbol{z}^*)\right]^s \prod_k 2^{-D(\boldsymbol{z}^*||q)}} \quad (20)$$

$$= \frac{\sum_{\boldsymbol{u}} \prod_k p_0(u_k|\boldsymbol{z}_n)^{1-s} p_1(u_k|\boldsymbol{z}_n)^s 2^{-D(\boldsymbol{z}_n||q)}}{\sum_{\boldsymbol{u}} \prod_k p_0(u_k|\boldsymbol{z}^*)^{1-s} p_1(u_k|\boldsymbol{z}^*)^s 2^{-D(\boldsymbol{z}^*||q)}} \quad (21)$$

$$= \prod_k \frac{\sum_{u_k} p_0(u_k|\boldsymbol{z}_n)^{1-s} p_1(u_k|\boldsymbol{z}_n)^s 2^{-D(\boldsymbol{z}_n||q)}}{\sum_{u_k} p_0(u_k|\boldsymbol{z}^*)^{1-s} p_1(u_k|\boldsymbol{z}^*)^s 2^{-D(\boldsymbol{z}^*||q)}}. \quad (22)$$

Then, define the set

$$\mathcal{T}_\delta^n = \{\boldsymbol{z}_n \in \mathcal{P}_n : \|\boldsymbol{z}_n - \boldsymbol{z}^*\|_2 < \delta\} \quad (23)$$

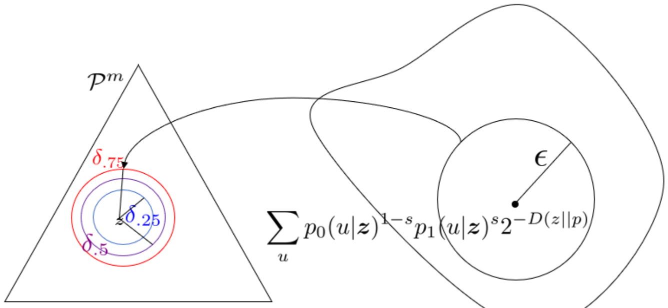

Fig. 2. As the parameter s changes, so to does the needed δ. We can alleviate this issue by simply selecting the smallest such δ over all $s \in [0,1]$. We must however, guarantee that this smallest δ is non-zero.

where $\mathcal{P}_n$ is the set of all possible type classes for sequences of length n. Thus, for the types $\boldsymbol{z}_n \in \mathcal{T}_\delta^n$, we have

$$(1-\epsilon)^n \leq \frac{\sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z}_n)^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z}_n)^s 2^{-nD(\boldsymbol{z}_n||q)}}{\sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z}^*)^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z}^*)^s 2^{-nD(\boldsymbol{z}^*||q)}} \leq (1+\epsilon)^n. \tag{24}$$

However, in order for this statement to hold, we must have that $\mathcal{T}_\delta^n$ be non-empty. Notice that by the density of the rationals in $\mathbb{R}$, for any $\delta > 0$, and $\boldsymbol{z} \in \mathbb{R}^m$, $\exists \boldsymbol{q} \in \mathbb{Q}^m$, such that $\|\boldsymbol{q} - \boldsymbol{z}\|_2 < \delta$. Moreover, the elements of $\mathcal{P}_n$ quantize $\mathcal{P}^m$, and that this quantization becomes finer with increasing n. Hence, for some sufficiently large $n(\delta)$, the set $\mathcal{T}_\delta^n$ is non-empty for all $n \geq n(\delta)$. The proof then relies on standard arguments from the method of types [11].

While the result above serves as a good preliminary result, there is a weakness. Namely, for a given ϵ, δ can be chosen independently of the strategy and $\boldsymbol{z}^*$, but *is dependent on* s. Recall that in the proof of the lemma, for a given ϵ, we select the δ so as to guarantee a desired sense of "closeness" in the performance of all types that are within the δ-ball centered around $\boldsymbol{z}^*$. However, we must have this δ-ball ($\mathcal{T}_\delta^n$) be non-empty which is true for all n greater that some $n(\delta) \in \mathbb{N}$. However, notice that for smaller δ, a larger $n(\delta)$ is required. Since δ depends on s, we see that *speed of convergence depends on* s. This argument is illustrated in Fig. 2.

As a result, we cannot take the infimum over s. If we can select this δ to be independent of the strategy, $\boldsymbol{z}^*$, *and* s, then the same $n(\delta)$ can be used for all s, and we may take all necessary infimums. This can be done by having the family of functions corresponding to equation (14) for $s \in [0,1]$ be *uniformly equicontinuous*[2] on $\mathcal{P}^m$. Unfortunately, for $s \in [0,1]$, we cannot guarantee that this family of functions is equicontinuous. However, if we restrict ourselves to $s \in [\epsilon_0, 1-\epsilon_0]$ *for any* $\epsilon_0 > 0$, then the family $\{p_0(u|\boldsymbol{z})^{1-s} p_1(u|\boldsymbol{z})^s : s \in [\epsilon_0, 1-\epsilon_0]\}$ *is* uniformly equicontinuous on $\mathcal{P}^m$. Hence, taking all appropriate infimums in 3 gives us Theorem 1.

[2]A family of functions $\mathcal{F}$ from $\mathcal{X}$ to $\mathbb{R}$ is said to be *equicontinuous* if $\forall x \in \mathcal{X}, \forall \epsilon > 0, \exists \delta > 0$ such that if $|x-y| < \delta$, $|f(x) - f(y)| < \epsilon, \forall f \in \mathcal{F}$. $\mathcal{F}$ is said to be *uniformly equicontinuous* if $\forall \epsilon > 0$, $\exists \delta > 0$ such that if $|x-y| < \delta$, $x, y \in \mathcal{X}$, then $|f(x) - f(y)| < \epsilon$, $\forall f \in \mathcal{F}$.

While it may seem necessary to search all of $\mathcal{P}^m$ for $\boldsymbol{z}^*$, it is not hard to show that

$$D(\boldsymbol{z}^*||q) \leq -\min_{s\in[0,1]} \frac{1}{n} \log \int_{\boldsymbol{y}} p_0(\boldsymbol{y}|p)^{1-s} p_1(\boldsymbol{y}|p)^s = C^*(p) \tag{25}$$

where $C^*(p)$ is the Chernoff information of the signal model at p. Hence, $\boldsymbol{z}^*$ must live in the KL-ball centered at p with radius $C^*(p)$. We provide the following interpretation. $\boldsymbol{z}^*$ must belong to the points in $\mathcal{P}^m$ such that their divergence from p gives us less information, and subsequently a higher error rate, than had we simply assumed p was the observed type and tested accordingly. This confirms the intuition that only the observed types that are close to the true network distribution p need to be considered.

We highlight a corollary of our results.

Corollary 1.1. *For all* $n \in \mathbb{N}$,

$$\begin{aligned}\inf_{\gamma\in\Gamma^n} \inf_{s\in[\epsilon,1-\epsilon]} \max_{\boldsymbol{z}\in\mathcal{P}^m} \frac{1}{n} \log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z})^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z})^s 2^{-nD(\boldsymbol{z}||q)} \\ = \inf_{\gamma\in\Gamma} \inf_{s\in[\epsilon,1-\epsilon]} \max_{\boldsymbol{z}\in\mathcal{P}^m} \log \sum_{u} p_0(u|\boldsymbol{z})^{1-s} p_1(u|\boldsymbol{z})^s 2^{-nD(\boldsymbol{z}||q)}\end{aligned} \tag{26}$$

Proof. Notice that for any $\boldsymbol{\gamma} \in \Gamma^n$, and $s \in [\epsilon, 1-\epsilon]$. Then, for any $\boldsymbol{z} \in \mathcal{P}^m$, we have that

$$\frac{1}{n} \log \sum_{\boldsymbol{u}} p_0(\boldsymbol{u}|\boldsymbol{z})^{1-s} p_1(\boldsymbol{u}|\boldsymbol{z})^s 2^{-nD(\boldsymbol{z}||q)} \tag{27}$$

$$= \frac{1}{n} \sum_k \log \sum_{u_k} p_0(u_k|\boldsymbol{z})^{1-s} p_1(u_k|\boldsymbol{z})^s 2^{-D(\boldsymbol{z}||q)} \tag{28}$$

With this property, the proof can be completed using techniques such as those in [10], [12]. □

The above corollary states that, asymptotically, there is no loss of optimality in having all agents use the same rule. This considerably simplifies design.

While the assumption that all agents must be identical can seem limiting, it is worth noting that one can relax the assumption that all agents are identical to one in which there is a finite number of agent classes, and all classes satisfy the conditions given in Assumption 1.

IV. A SPECIFIC SIGNAL MODEL

Herein, we show the utility of the proposed model by employing a specifid signal model which allows further specification of our error rates. We shall consider a signal model motivated by microbial interactions; however, other signal models are easily adapted. As such, the signals received by agents are Poisson random variables [13].

Specifically, denote each class of agents by $g = 1, 2, .., G$, where G is the total number of classes in the network, and n_g is the total number of members of class g. For simplicity, assume each class may take one of two states, 0 or 1. Similarly to above, let $\boldsymbol{h}_g$ be the *class interference vector* for class g. That is, $\boldsymbol{h}_g[i] = \alpha_{g,i}$, where $\alpha_{g,i}$ captures the interference induced on g by class i. Then, for members of class g, the signal model is given as

$$y \sim Poiss(\sum_{i=1}^{G} r_g \boldsymbol{h}_g[i]\boldsymbol{z}_n[i] + \lambda_g^h) \tag{29}$$

where $\boldsymbol{z}_n[i] = \frac{1}{n_i}\sum_{k=1}^{n_i} x_k$ (the ratio of agents that belong to class i that are in state 1), r_g is the ratio of agents that belong to class g, and λ_g^h is the base rate for class g under hypothesis h. Then, in order to use our previous results, we must show for each class g,

$$\inf_{\boldsymbol{\gamma}} \inf_{s\in[0,1]} \min_{\boldsymbol{z}\in\mathcal{P}^{|\mathcal{S}_g|}} \sum_u p_0(u|\boldsymbol{z})^{1-s} p_1(u|\boldsymbol{z})^s 2^{-D(\boldsymbol{z}||q)} > 0. \tag{30}$$

Since $D(\boldsymbol{z}||\boldsymbol{p})$ is bounded on $\mathcal{S}_g$, we get that $2^{-D(\boldsymbol{z}||q)} > 0$ for all $\boldsymbol{z} \in \mathcal{S}_g$. Hence, it is sufficient to show that

$$\inf_{\boldsymbol{\gamma}} \inf_{s\in[0,1]} \min_{\boldsymbol{z}\in\mathcal{P}^{|\mathcal{S}_g|}} \sum_u p_0(u|\boldsymbol{z})^{1-s} p_1(u|\boldsymbol{z})^s > 0. \tag{31}$$

To this end, the following result is quite useful.

Lemma 4. *For all* $\gamma \in \Gamma$, $s \in [0,1]$, *and* $\boldsymbol{z} \in \mathcal{P}^{|\mathcal{S}_g|}$,

$$\sum_u p_0(u|\boldsymbol{z})^{1-s} p_1(u|\boldsymbol{z})^s \geq \int_y \min\{p_0(y|\boldsymbol{z}), p_1(y|\boldsymbol{z})\} \tag{32}$$

Observe that this lower bound depends on neither γ nor s. Hence, it is sufficient to show that

$$\inf_{\boldsymbol{z}\in\mathcal{P}^{|\mathcal{S}_g|}} \int_y \min\{p_0(y|\boldsymbol{z}), p_1(y|\boldsymbol{z})\} > 0. \tag{33}$$

We turn to our chosen model, of which we obtain

$$\sum_u p_0(u|\boldsymbol{z})^{1-s} p_1(u|\boldsymbol{z})^s \geq \exp\left\{-\sum_{i=1}^{G} r_g \boldsymbol{h}_g[i] - \max\{\lambda_g^0, \lambda_g^1\}\right\} \tag{34}$$

for all $\boldsymbol{z}$. Moreover, note that $\sum_{i=1}^{G} \boldsymbol{h}_g[i]\boldsymbol{z}[i] \leq \sum_{i=1}^{G} \boldsymbol{h}_g[i] < \infty$ (it is assumed that $0 \leq \boldsymbol{h}_g[i] < \infty$ for all i and g. Hence,

$$\begin{aligned} &\inf_{\boldsymbol{\gamma}} \inf_{s\in[0,1]} \min_{\boldsymbol{z}\in\mathcal{P}^{|\mathcal{S}_g|}} \sum_u p_0(u|\boldsymbol{z})^{1-s} p_1(u|\boldsymbol{z})^s \\ &\geq \exp\left\{ -\sum_{i=1}^{G} r_g \boldsymbol{h}_g[i] - \max\{\lambda_g^0, \lambda_g^1\}\right\} > 0. \end{aligned} \tag{35}$$

Finally, it remains to be shown that this class of mass functions satisfies Assumption 1. Recall that thanks to Lemma 1, it suffices to show that $f_{\boldsymbol{z}} \longrightarrow f_{\boldsymbol{\theta}}$ whenever $\boldsymbol{z} \longrightarrow \boldsymbol{\theta}$. Suppose $\boldsymbol{z} \longrightarrow \boldsymbol{\theta}$. We then have that

$$\lambda_{g,\boldsymbol{z}}^h = \sum_{i=1}^{G} \boldsymbol{h}_g[i]\boldsymbol{z}[i] + \lambda_g^h \longrightarrow \sum_{i=1}^{G} \boldsymbol{h}_g[i]\boldsymbol{\theta}[i] + \lambda_g^h = \lambda_{g,\boldsymbol{\theta}}^h. \tag{36}$$

For any y, by continuity of x^y and e^{-x}, we have

$$\frac{(\lambda_{g,\boldsymbol{z}}^h)^y}{y!} e^{-\lambda_{g,\boldsymbol{z}}^h} \longrightarrow \frac{(\lambda_{g,\boldsymbol{\theta}}^h)^y}{y!} e^{-\lambda_{g,\boldsymbol{\theta}}^h}. \tag{37}$$

Hence, this family of signal models satisfies Assumption 1. Since this signal model obeys all assumptions, and members of the same class can use the same rule we can restrict our attention to

$$\frac{1}{n}\sum_{g=1}^{G} n_g \log \sum_{u_g} p_0(u_g|\boldsymbol{z})^{1-s} p_1(u_g|\boldsymbol{z})^s 2^{-n_g D(\boldsymbol{z}||q_g)}. \tag{38}$$

Moreover, since members of the same class may use the same rule, we get that

$$\boldsymbol{z}^*(n) = \arg\max_{\boldsymbol{z}\in\mathcal{P}^m} \sum_g n_g \log \sum_{u_g} p_0(u_g|\boldsymbol{z})^{1-s} p_1(u_g|\boldsymbol{z})^s 2^{-D(\boldsymbol{z}||q_g)} \tag{39}$$

$$= \arg\max_{\boldsymbol{z}\in\mathcal{P}^m} \sum_g \frac{n_g}{n} \log \sum_{u_g} p_0(u_g|\boldsymbol{z})^{1-s} p_1(u_g|\boldsymbol{z})^s 2^{-D(\boldsymbol{z}||q_g)}. \tag{40}$$

Now, suppose *fix the quantity* $\frac{n_g}{n} = r_g$. That is, for any n, we wish to keep the ratios of the populations constant. Recall that we allow n to be arbitrarily large, and so we can always get arbitrarily close to any r_g. Moreover, since we assume all exponents are finite under any rule, the difference in performance between two systems that differ only by at most a fixed number of agents tends towards zero in the limit of the network size n. Hence, without loss of asymptotic optimality, we have that

$$\boldsymbol{z}^*(n) = \arg\max_{\boldsymbol{z}\in\mathcal{P}^m} \sum_g r_g \log \sum_{u_g} p_0(u_g|\boldsymbol{z})^{1-s} p_1(u_g|\boldsymbol{z})^s 2^{-D(\boldsymbol{z}||p_q)}. \tag{41}$$

Recall that our signal model given in equation (29) depends only on the *ratio of agents from class g that are in state 1, and not on the actual number itself.* Hence, $p_h(u_g|\boldsymbol{z})$ does not depend on n for any h or g. This, combined with the fact that r_g is held fixed, tells us that $\boldsymbol{z}^*(n)$ *can be chosen independently of* n. That is, the optimal error exponent becomes

$$\begin{aligned} \inf_{\gamma_1,\ldots,\gamma_g} \min_{s\in[0,1]} \sum_g r_g \log \sum_{u_g} p_0(u_g|\boldsymbol{z}^*)^{1-s} \\ p_1(u_g|\boldsymbol{z}^*)^s 2^{-D(\boldsymbol{z}^*||p_q)}. \end{aligned} \tag{42}$$

This exponent is for a fixed set of ratios r_g. The utility of this expression is that it is a function of *only the population ratios*. That is, the size of the network or colony n, does not appear anywhere in the expression and so one can study the impact of population ratios on the overall network/colony performance, without worrying about the size of the colony itself (of course, provided that one is interested in sufficiently large colonies, but in microbial applications, this is often the case). Indeed, if one has control of the population ratios, that one can find the optimal population ratios for asymptotic performance by solving

$$\begin{aligned} \inf_{\boldsymbol{r}} \inf_{\gamma_1,\ldots,\gamma_g} \min_{s\in[0,1]} \sum_g r_g \log \sum_{u_g} p_0(u_g|\boldsymbol{z}^*)^{1-s} \\ p_1(u_g|\boldsymbol{z}^*)^s 2^{-D(\boldsymbol{z}^*||p_q)} \end{aligned} \tag{43}$$

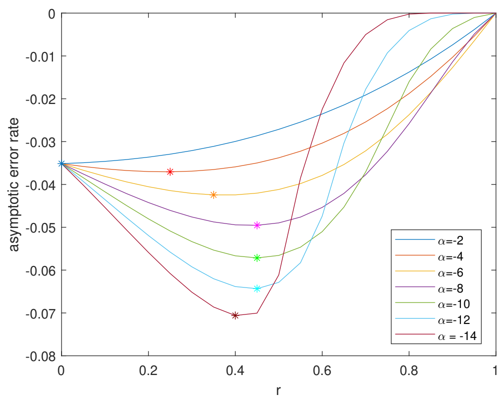

Fig. 3. System performance as a function of the ratio of class 1 for different interference vectors. The optimal ratio of class 1 for a given α curve is indicated by a star.

with $\boldsymbol{r} = [r_1, r_2, ..., r_G]$.

We explore the affect of two populations on the asymptotic performance with respect to the signal model given in (29). Specifically, $\lambda_g^0 = 1$ for both $g = 1, 2$, $\lambda_1^1 = 3$, $\lambda_2^1 = 5$. The systems differ in how the classes interact with each other, which is captured by the class interference vectors $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$. Specifically, $\boldsymbol{h}_1 = [0, 0]$, and $\boldsymbol{h}_2 = [\alpha, 2]$. That is, only class 2 experiences interference. Also, $q_1(X = 1) = \frac{1}{2}$ and $q_2(X = 1) = \frac{9}{10}$. This specific models captures a case wherein classes of agents send signals that are detrimental to the other. Furthermore, the model enables the consideration of a wide range of phenomena. In particular, our analytical results will be dependent on the ratios of the populations in each class. Moreover, notice that $\lambda_1^0 = \lambda_1^1$, and so class 1 can never distinguish between the two hypotheses. However, although class 1 is useless for the process of inference, it serves the purpose of mitigating the harmful effects of class 2 on itself. In our future work, we will specialize our analysis to specific multi-species microbial models where signal jamming and interference are common [14].

In Fig. 3, we plot the function

$$rD(\boldsymbol{z}^*[1]||q_1) + (1-r)\big[D(\boldsymbol{z}^*[2]||q_2) - \log \sum_{u_2} \sqrt{p_0(u_2|\boldsymbol{z}^*)p_1(u_2|\boldsymbol{z}^*)}\big] \quad (44)$$

as a function of the ratio of population one, r, for various α. A few interesting phenomena are illustrated by Fig. 3 . First, the optimal ratio r depends the strength of coupling captured by α. Notice that for $\alpha = 0$, it is optimal to have *no members of population 1*. Indeed, when $\alpha = 0$, notice that class 1 does not influence class 2 at all. Hence, the trait that made class 1 useful in design (the ability to assist class 2) is no longer present, rendering it useless. Furthermore, even though the parameter of the Poisson distribution for class 2 depends linearly on α, the received signal Y, and thus the decisions made by the agents, exhibit a non-linear dependency. This is shown in Fig. 3 by the fact that the optimal ratio of class 1 is non-monotonic in α.

V. Conclusions

In this work, we have studied the problem of decentralized detection over sensor type-sensitive networks. We have shown that the asymptotic error rate is dominated by a single distribution. This framework and result allows us to model complex interactions among agents as well as study the effect of ratios on network or colony performance. Our future work involves extending this framework to incorporate communication links between the agents and the fusion center, as well as analyzing complex intraspecies interactions in microbial colonies.

References

[1] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*. JAI Press, 1993, pp. 297–344.

[2] I. Y. Hoballah and P. K. Varshney, "Distributed bayesian signal detection," *IEEE Transactions on Information Theory*, vol. 35, no. 5, pp. 995–1000, Sep. 1989.

[3] H. Chen, B. Chen, and P. K. Varshney, "A new framework for distributed detection with conditionally dependent observations," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1409–1419, Mar. 2012.

[4] V. Aalo and R. Viswanathou, "On distributed detection with correlated sensors: two examples," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 3, pp. 414–421, 1989.

[5] O. A. Hanna, X. Li, C. Fragouli, and S. Diggavi, "Can we break the dependency in distributed detection?" in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 2720–2725.

[6] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, May. 1985.

[7] A. Tajer, S. Kar, H. V. Poor, and S. Cui, "Distributed joint cyber attack detection and state recovery in smart grids," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Oct. 2011, pp. 202–207.

[8] B. Geng, S. Brahma, T. Wimalajeewa, P. K. Varshney, and M. Rangaswamy, "Prospect theoretic utility based human decision making in multi-agent systems," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1091–1104, 2020.

[9] M. M. Vasconcelos, U. Mitra, O. Câmara, M. S. Gangan, and J. Boedicker, "A continuous-time decision-making model for bacterial growth via quorum sensing: theory and evidence," in *Proc. ACM International Conference on Nanoscale Computing and Communication*, Sep. 2019, pp. 1–6.

[10] J. Shaska and U. Mitra, "Decentralized decision making in multi-agent networks: the state-dependent case," in *2021 IEEE Global Communcations Conference*, Dec 2021.

[11] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley-Interscience, 2006.

[12] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals and Systems*, vol. 1, no. 2, pp. 167–182, Jun. 1988.

[13] M. C. Gursoy, M. Nasiri-Kenari, and U. Mitra, "Towards high data-rate diffusive molecular communications: A review on performance enhancement strategies," *Digital Signal Processing*, vol. 124, p. 103161, 2022, signal Processing Aspects of Molecular Communications. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1051200421002001

[14] G. L. Lozano, J. I. Bravo, M. F. G. Diago, H. B. Park, A. Hurley, S. B. Peterson, E. V. Stabb, J. M. Crawford, N. A. Broderick, and J. Handelsman, "Introducing thor, a model microbiome for genetic dissection of community behavior," *mBio*, vol. 10, no. 2, pp. e02 846–18, 2019.