entropy

MDPI

*Article*

# Joint Detection and Communication over Type-Sensitive Networks

**Joni Shaska * and Urbashi Mitra**

Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA 90089, USA; ubli@usc.edu
* Correspondence: shaska@usc.edu

**Abstract:** Due to the difficulty of decentralized inference with conditional dependent observations, and motivated by large-scale heterogeneous networks, we formulate a framework for decentralized detection with coupled observations. Each agent has a state, and the empirical distribution of all agents' states or the type of network dictates the individual agents' behavior. In particular, agents' observations depend on both the underlying hypothesis as well as the empirical distribution of the agents' states. Hence, our framework captures a high degree of coupling, in that an individual agent's behavior depends on both the underlying hypothesis and the behavior of all other agents in the network. Considering this framework, the method of types, and a series of equicontinuity arguments, we derive the error exponent for the case in which all agents are identical and show that this error exponent depends on only a single empirical distribution. The analysis is extended to the multi-class case, and numerical results with state-dependent agent signaling and state-dependent channels highlight the utility of the proposed framework for analysis of highly coupled environments.

**Keywords:** heterogeneous networks; method of types; large-scale networks; information measures



**Citation:** Shaska, J.; Mitra, U. Joint Detection and Communication over Type-Sensitive Networks. *Entropy* **2023**, *25*, 1313. https://doi.org/10.3390/e25091313

Academic Editors: Raúl Alcaraz, Luca Faes, Leandro Pardo and Boris Ryabko

Received: 30 May 2023
Revised: 7 August 2023
Accepted: 24 August 2023
Published: 8 September 2023



## 1. Introduction

Decentralized detection is an important element in a wide range of modern applications, such as the Internet of Things [1], smart grids [2], cognitive radio [3], and millimeter-wave communications [4]. However, many classical results in decentralized detection assume that agents' observations are independent, conditioned on the underlying hypothesis. This assumption fails to hold in many of these recent applications, such as human decision-making [5], sensor networks with correlated observations [6], and quorum sensing in microbial communities [7]. Unfortunately, the problem of decentralized detection with correlated observations is NP-Hard [8], and many of the classical results are not applicable in this case (for examples, see [9–11]). Recent work in decentralized detection has placed greater attention on the case of correlated observations [12–15]. Although recent advancements have been promising, the inherent difficulty of the problem has resulted in approximations and relaxations [13,15]. In this work, we build upon the state-dependent formulation introduced in [16] by allowing agents' observations to depend on both the underlying hypothesis as well as the empirical distribution, or *type*, of their states. The notion of type has a rich history in information theory and statistics, being first introduced by Csiszar [17]. Today, the method of types has been further developed [18] and is used in a variety of fields, such as control [19], machine learning [20], statistics [21], and even DNA storage channels [22].

Conditionally correlated observations can be handled under specific signal models [15] and assumptions [12,16,23–25]. In particular, Ref. [15] studied bandwidth-constrained detection under the Neyman–Pearson criterion and solved a relaxation of the problem. Several works [23–25] have studied the problem under communication constraints, with [23] showing that the network learns the hypothesis exponentially quickly under constrained [23]

and randomized [24] communication. Moreocer, [25] developed a deep learning algorithm for real-time industry constraints. Other works have attempted to decouple agents' observations via algorithms [13] and specific models [12,16,26]. In [12], a hidden variable was introduced that allows the observations of the agents to be independent, conditioned on the hidden variable, and it was proved that threshold-based decisions are optimal under certain model assumptions. Unfortunately, even if a problem of interest falls under this framework, the assumptions are rather strong and fail in a number of applications. In our prior work [16,26], we introduced a *state* variable for each agent and allowed the agents' observations to be independent conditioned on both the hypothesis and the agent's state. We proved similar results to those of [12,27] under much weaker conditions. However, the model proposed in [16,26] grants each agent its own individual state, whereas in [12] agents may share a common hidden variable.

The framework was extended in [16,26]; herein agents' observations, *depend on a common variable*, i.e., the type of the agents' states. In [16], it was assumed that agents know their individual state and that the fusion center knows the states of all agents. We strongly relax this assumption; agents do not know their state, and the fusion center only knows the empirical distribution of the agents' states. Another key difference is that in [16,26] the state variable *is sufficient to decouple agents' observations*, whereas in this work all agent states are necessary to decouple observations, allowing this formulation to handle stronger forms of coupling. The need for the empirical distribution calls for different analysis techniques from those in [12,16,26] via the method of types. We further introduce a communication link between the agents and the centralized decision-maker (called the *fusion center*) which is not present in [12,16,26] .

Many works in decentralized detection include a communication link between the agents and the fusion center; the idea itself is not new [28–30]. However, in prior works the statistical properties of the communication channel were assumed to be independent of the network's behavior. A contribution of our current work is that we *allow the quality of the communication channel to vary with the network's behavior*. This is again accomplished by allowing the channel to vary with the type of the agents' states. The concept of a channel with state-dependent noise has been previously considered in information theory [31], and is in use today [32–34]. However, most of the aforementioned works involving the notion of state have focused primarily on communication over channels with state, and have not examined joint detection and communication. While recent works on estimation exist, they were the context of estimating the channel state to improve communication performance [34–36]. Notably, signal-dependent noise [37] can be accommodated in our proposed model. In particular, these models are relevant to visible light communication [38], magnetic recording [39,40], and imaging applications [41,42].

As an example, we may consider the occurrence of such forms of coupling in microbial systems. Microbial communities synthesize signaling molecules [7]; when sensed in the environment, these can result in individual gene expressions that lead to new collective behaviors through a process called quorum sensing. Specifically, cell $i$ only engages in quorum sensing when the received number of autoinducer molecules from the environment $A_i$ exceeds a certain threshold $\tau_A$. A common model involves assuming that $A_i$ follows a Poisson distribution conditioned on the total number of *synthases* (synthases are enzymes within a cell that are responsible for the production of autoinducer molecules) in the community and the number of receptors in a cell $i$, provided as $R_i$ [43]:

$$\mathbb{P}(A_i = k | R_i, S_1, \ldots; S_n) = \frac{\left(\lambda R_i \sum_{j=1}^{n} S_j\right)^k \exp\left(-\lambda R_i \sum_{j=1}^{n} S_j\right)}{k!}$$

where $S_i$ is the number of synthases present in cell $i$ and $\lambda > 0$ is a normalizing term. Hence, we can think of the number of synthases and receptors in cell $i$ as being the **state** of cell $i$. Then, the observation of cell $i$ depends on *the states of all other cells* through the summed total of synthases across the cells. This example illustrates the need for the current

approach, as the models proposed in [12,16] cannot handle this form of coupling and do not lead to tractable asymptotic results.

In this work, we derive the *error exponent* as the network size grows. Assuming that all priors are known, the optimal asymptotic decay rate of the probability of error is provided by the *Chernoff information* [27,44,45], regardless of whether conditional independence holds. Using the Chernoff information, Ref. [27] proved that identical rules are *asymptotically optimal* for identical agents, while [45] showed that identical binary quantizers are asymptotically optimal in power-constrained networks. The works in [27,45] both relied on conditional independence. A contribution of the present study is to remove the need for conditional independence through the development of a measure that is asymptotically equivalent to the Chernoff information and tractable in our scenario. The primary argument comes from the method of types, which, combined with a series of equicontinuity arguments, shows that asymptotic performance is dominated *by a single distribution*. Surprisingly, this dominating distribution is generally *not the true distribution of the agents' states*.

Using the network type to decouple agents' observations can be extended beyond pure decentralized detection. For instance, consensus algorithms used in blockchain applications often need to deal with faulty or nonconforming nodes [46]. Hence, it is possible to consider whether the node is conforming or not as the state and the total percentage of conforming nodes as the network type. Then, the problems of jointly estimating the network type (the consensus problem) and detecting the underlying hypothesis (the detection problem) can be considered. Much of the structure herein applies to such problems, as observations received by agents depend on the other agents' states. Moreover, the hypothesis and network type are correlated; when more agents are faulty or nonconforming, an attack is more likely to be present.

Our contributions in this paper are as follows:

1. We formulate a framework for distributed inference in which the agents' observations are correlated through both the hypothesis and the empirical distribution (or *type*) of the network state. This formulation captures a high level of coupling between agents.
2. We consider a distributed inference problem with a communication link between the agents and the fusion center, with the additional caveat that the noise over the link is dependent on the agents. Hence, our framework captures joint sensing with correlated observations as well as joint communications with correlated noise.
3. We derive expressions for the error exponent for a single class of agents, then extend our results to the case of heterogeneous groups of identical agents. In particular, assuming that identical agents use a common rule, the optimal error exponent depends only on the ratios of the groups, not on the actual size of the groups themselves. This allows a wide range of problems to be studied in which there are multiple classes of agents that interfere with each other.
4. We present a numerical example for a three-class case to highlight the utility of the proposed expression for the error exponent. In particular, we show how this expression can be used to optimize the ratios of heterogeneous groups in the presence of cross-class interference. This example further illustrates the fact that the true distribution may not dominate the asymptotics. The effect of the channel is observed as well.

*Notation*

Random variables are denoted by capital letters $X$ and specific realizations are denoted as lowercase letters $x$. Random vectors are denoted as boldface capital letters $\mathbf{X}$ and specific realizations are denoted with lowercase boldface letters $\mathbf{x}$. Given a random vector (realization) $\mathbf{X}$ ($\mathbf{x}$), $\mathbf{X}^{\backslash k}$ ($\mathbf{x}^{\backslash k}$) denotes the vector $\mathbf{X}$ ($\mathbf{x}$) with the $k$th element removed. Calligraphic letters $\mathcal{X}$ denote sets. The symbol $\mathbb{P}$ denotes probabilities of events, and $\mathbb{E}_X$ denotes expectations with respect to the random variable $X$.

## 2. Materials and Methods

The details concerning how plots are generated are provided in Section 5, along with a discussion of a specific example.

## 3. Problem Formulation, Definitions, and Assumptions

### *3.1. Problem Setup*

Consider a set of $n$ agents. The global environmental variable $H$ is the binary $H \in \{0,1\}$. Agent $k$ ($k = 1,2,\ldots;n$) receives a *signal* $Y_k \in \mathcal{Y}$, with $\mathcal{Y}$ being the signal space. The probability density of $Y_k$ conditioned on $H = h$ is denoted as $p_h^k(y)$. In addition, each agent takes a *state* $S_k \in \{1,2,\ldots;m\}$, where $m$ is a finite integer. The prior for the state of agent $k$ is $p^k(s)$, and we define the vector $\boldsymbol{p}^k = [p^k(1),\ldots,p^k(m)]^T$. The collection of states $\mathbf{S} = [S_1,\cdots,S_n]^T$ is called the *network state*, with joint density $p(\mathbf{s})$. For a given network state, we denote the empirical distribution (or the *type*) of $\mathbf{S}$ as $\mathbf{Q}_{\mathbf{S}}^n$, that is,

$$Q_{\mathbf{S}}^n(i) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{S_k=i\}}, \qquad i \in \{1,\ldots;m\}, \tag{1}$$

where $\mathbb{1}_{\{S_k=i\}}$ is the indicator that agent $k$ is in state $i$. Let $\mathcal{Q}^n$ denote the set of all empirical distributions corresponding to sequences of length $n$; then, for a given $\mathbf{q}^n \in \mathcal{Q}^n$, $\mathcal{T}(\mathbf{q}^n)$ is the *type-class of* $\mathbf{q}^n$, i.e.,

$$\mathcal{T}(\mathbf{q}^n) = \{\mathbf{s} \in \{1,2,\ldots;m\}^n : \mathbf{Q}_{\mathbf{s}}^n = \mathbf{q}^n\}, \tag{2}$$

where $\{1,2,\ldots;m\}^n$ is the Cartesian product of $\{1,2,\ldots;m\}$ with itself $n$ times. Note that $\mathbf{Q}_{\mathbf{S}}^n$ is a random vector with realization $\mathbf{q}^n$, that is,

$$p(\mathbf{q}^n) = \mathbb{P}(\mathbf{Q}_{\mathbf{S}}^n = \mathbf{q}^n) = \mathbb{P}(\mathcal{T}(\mathbf{q}^n)) = \sum_{\mathbf{s}\in\mathcal{T}(\mathbf{q}^n)} p(\mathbf{s}). \tag{3}$$

The joint probability distribution of $Y_k$ and the network type under hypothesis $H = h$ is provided by $p_h^k(y,\mathbf{q}^n)$. The associated conditional density is denoted as $p_h^k(y|\mathbf{q}^n)$. Let $\mathcal{P}^m$ denote the probability simplex in $\mathbb{R}^m$:

$$\mathcal{P}^m = \{\mathbf{q} \in \mathbb{R}^m : q_i \geq 0, \sum_i q_i = 1\}. \tag{4}$$

For $\mathbf{q} \in \mathcal{P}^m$, the conditional density $p_h^k(y|\mathbf{q})$ is called the *signal model for agent* k. When we write densities conditioned on $\mathbf{q} \in \mathcal{P}^m$, we are assuming that these densities have a functional dependence on $\mathbf{q}$ in order to avoid issues with measurability, as certain types may not be observable regardless of the size of the network. For a simple example, consider $[\frac{1}{e}, 1-\frac{1}{e}]^T \in \mathcal{P}^2$, which is never in $\mathcal{Q}^n$ for any $n$ due to the fact that $\frac{1}{e}$ is irrational. We define $\mathbf{Y} = [Y_1,\cdots,Y_n]^T$, while the joint density of $\mathbf{Y}$ and $\mathbf{Q}^n$ and the density of $\mathbf{Y}$ conditioned on $\mathbf{Q}^n = \mathbf{q}^n$ under $H = h$ are denoted as $p_h(\mathbf{y},\mathbf{q}^n)$ and $p_h(\mathbf{y}|\mathbf{q}^n)$, respectively. The joint density $p_h(\mathbf{y}|\mathbf{q})$ is called the *joint signal model*. For brevity, we call the conditional distribution $p(H = h|\mathbf{q}^n)$ the *hypothesis model*. It is important to note that we *do not assume conditional independence of the agents' observations*, i.e., we can have $p_h(\mathbf{y}) \neq \prod_k p_h^k(y_k)$ for $h \in \{0,1\}$; we do, however, assume that the structure described below holds.

**Assumption 1.** *The joint signal model obeys the following:* $\forall \mathbf{y}$, $\forall \mathbf{q} \in \mathcal{P}^m$, $\forall h \in \{0,1\}$,

$$p_h(\mathbf{y}|\mathbf{q}) = \prod_{k=1}^{n} p_h^k(y_k|\mathbf{q}). \tag{5}$$

Equation (5) states that the signal $Y_k$ of agent $k$ is independent of $\mathbf{Y}^{\setminus k}$ conditioned on *both* $H$ and $\mathbf{Q}^n$.

**Assumption 2.** $\forall \mathbf{y}, \forall \mathbf{q}^n \in \mathcal{Q}^n$, $h \in \{0,1\}$, $p_h(\mathbf{y}, \mathbf{q}^n) > 0$; *the joint densities have the same support under both hypotheses.*

Upon receiving observation $Y_k$, agent $k$ makes a *decision* $U_k \in \{1, 2, \ldots; b\}$ according to a *rule*, which is a (possibly randomized) function from $\mathcal{Y}$ to the decision space $\mathcal{U}$. We denote the possibly randomized rule used by agent $k$ as $\gamma_k$,

$$U_k = \gamma_k(Y_k) \sim p^k(u|y). \tag{6}$$

The collection of rules $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_n]^T$ is called a *strategy*. After agent $k$ has made its decision, it sends $U_k$ to the *fusion center* through a noisy communication link which is allowed to depend on the type $\mathbf{q}^n$. Upon sending $U_k$, the fusion center receives the message $X_k$ with

$$X_k \sim p^k(x|u, \mathbf{q}^n). \tag{7}$$

Given $\mathbf{q} \in \mathcal{P}^m$, the conditional density $p^k(x|u, \mathbf{q})$ is the *channel model for agent* k. We define $\mathbf{X} = [X_1, \cdots, X_n]^T$, and the joint conditional density $p(\mathbf{x}|\mathbf{u}, \mathbf{q})$ is called the *joint channel model*.

**Assumption 3.** *The joint channel model obeys the following:* $\forall \mathbf{x}$, $\mathbf{u}$, $\forall \mathbf{q} \in \mathcal{P}^m$,

$$p(\mathbf{x}|\mathbf{u}, \mathbf{q}) = \prod_{k=1}^{n} p^k(x_k|u_k, \mathbf{q}). \tag{8}$$

**Assumption 4.** $\forall \mathbf{x}$, $\mathbf{u}$, $\forall \mathbf{q} \in \mathcal{P}^m$, $p(\mathbf{x}|\mathbf{u}, \mathbf{q}) > 0$.

The fusion center does not know the network state $\mathbf{S}$, however, *we assume that it knows* $\mathbf{Q}_\mathbf{S}^n$. This assumption is not strong, as the empirical distribution $Q_S^n$ can be estimated via consensus methods [47]. Upon receiving messages $\mathbf{X}$ and $\mathbf{Q}_\mathbf{S}^n$, the fusion center makes an inference as to which hypothesis is true, denoted by $\hat{H}$. We seek to minimize the asymptotic decay rate of the probability of the error (as defined in Equation (10)). We assume that the fusion center is using the maximum a posterori (MAP) rule, i.e.,

$$\hat{H} = \begin{cases} 1, & (\mathbf{x}, \mathbf{q}^n) \in \mathcal{A}^\gamma \\ 0, & (\mathbf{x}, \mathbf{q}^n) \in \mathcal{A}^{\gamma^c} \end{cases}, \quad \text{where} \quad \mathcal{A}^\gamma = \left\{ (\mathbf{x}, \mathbf{q}^n) : \frac{p_1(\mathbf{x}|\mathbf{q}^n)}{p_0(\mathbf{x}|\mathbf{q}^n)} \geq \frac{p(H=0|\mathbf{q}^n)}{p(H=1|\mathbf{q}^n)} \right\}, \tag{9}$$

which minimizes the probability of error for a given strategy $\boldsymbol{\gamma}$. The set $\mathcal{A}^\gamma$ depends on the specific strategy $\boldsymbol{\gamma}$ selected; given $\boldsymbol{\gamma}$, it is possible to compute the optimal inference rule as a deterministic function of $\boldsymbol{\gamma}$ using Equation (9). The complete problem setup is summarized in Figure 1.

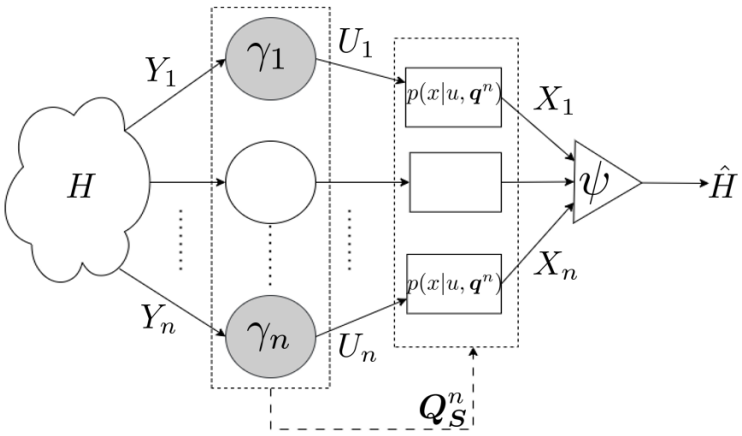


**Figure 1.** A set of $n$ agents receive signals $Y_k$ and states $S_k$. Each agent is characterized by a decision rule $\gamma_k$ and sends a message $X_k$ to the fusion center, which outputs $\hat{H}$. The empirical distribution of the states $\mathbf{Q}_\mathbf{S}^n$ governs the behavior of the signals $Y_k$ as well as the communication channels.

### 3.2. Definitions

We now introduce several definitions and concepts that are used throughout the paper.

**Definition 1.** *Let $\mathbb{P}^{\gamma}(\hat{H} \neq H)$ be the probability of error under strategy $\gamma$. We define the error exponent $\Lambda$ (provided the limit exists) as:*

$$\Lambda = -\lim_{n\to\infty} \inf_{\gamma} \frac{1}{n} \log \mathbb{P}^{\gamma,\psi}(\hat{H} \neq H). \tag{10}$$

The limit $\Lambda$ depends on the strategy $\gamma$. Thus, the strategy $\gamma^*$ that achieves the infimum may be such that the limit does not exist. Moreover, (10) makes no assumption as to how the statistical properties of the agents vary with $n$; in general, it is not possible to say anything about the existence of $\Lambda$. However, in many practical settings, such as homogeneous networks and power-constrained networks, $\Lambda$ exists and has a nice closed-form solution [16,27,45]. The main result of this work is an equivalent characterization of the error exponent defined above, showing that in our scenario the limit does exist. This equivalent expression has several desirable properties, and we can directly optimize the equivalent expression.

**Definition 2.** *The Kullback–Leibler Divergence between two distributions $\mathbf{q}$ and $\mathbf{p}$ is provided as follows:*

$$D(\mathbf{q}||\mathbf{p}) = \sum_{\mathbf{x}} \mathbf{q}(x) \log \frac{\mathbf{q}(x)}{\mathbf{p}(x)}.$$

Here, we are interested in understanding the interactions between different *classes* of agents, where members of a given class are *identical*, defined as follows.

**Definition 3.** *Given a collection of $n$ agents, these agents are identical if the following conditions hold:*

1. $p_h^k(y|\mathbf{q}) = p_h^j(y|\mathbf{q})$ *for all* $k, j \in \{1, 2, \ldots; n\}$, $h \in \{0, 1\}$, $y \in \mathcal{Y}$, $\mathbf{q} \in \mathcal{P}^m$.
2. $p^k(x|u, \mathbf{q}) = p^j(x|u, \mathbf{q})$ *for all* $k, j \in \{1, 2, \ldots; n\}$, $x \in \mathcal{X}$, $u \in \{1, 2, \ldots; b\}$, $\mathbf{q} \in \mathcal{P}^m$.
3. *The agent states $S_k$ are i.i.d. a priori, i.e.,* $p(\mathbf{s}) = \prod_k p^k(s_k) = \prod_k p(s_k)$.

Condition (1) states that, conditioned on the hypothesis $H$ and the network type $\mathbf{Q}_{\mathbf{S}}^n$, the probability distributions on the received signals for all agents are the same. Similarly, Condition (2) states that, conditioned on the network type $\mathbf{Q}_{\mathbf{S}}^n$ and $U_k = u$ for all $k \in \{1, 2, \ldots; n\}$, the probability distributions on the received messages are the same for all agents.

**Definition 4.** *A class is a collection of agents that are all identical.*

### 3.3. Key Assumptions

We first derive the error exponent for the single-class case in Theorem 1, which is then generalized to the case of multiple classes.

**Assumption 5.** *Our key assumptions for Theorem 1 are as follows:*

*(a)* *All agents are identical, as provided in Definition 3. Hence, we remove the notational dependence on k in the sequel.*

*(b)* *The hypothesis model obeys the following:*

$$\lim_{n\to\infty} \frac{1}{n} \log \min_{\mathbf{q}^n} \min\{p(H = 1|\mathbf{q}^n), p((H = 0|\mathbf{q}^n)\} = 0. \tag{11}$$

(c) *The signal model is continuous in* $\mathbf{q}$ *for all agents, that is, if* $\{\boldsymbol{\alpha}_i\}_{i\in\mathbb{Z}}$ *is a sequence in* $\mathcal{P}^m$ *such that* $\lim_{i\to\infty}\boldsymbol{\alpha}_i = \mathbf{q}$*, then* $\forall y$*,*

$$\lim_{i\to\infty} p_h(y|\boldsymbol{\alpha}_i) = p_h(y|\mathbf{q}), \qquad h \in \{0,1\}. \tag{12}$$

(d) *The channel model is continuous in* $\mathbf{q}$ *for all agents. That is, if* $\{\boldsymbol{\alpha}_i\}_{i\in\mathbb{Z}}$ *is a sequence in* $\mathcal{P}^m$ *such that* $\lim_{i\to\infty}\boldsymbol{\alpha}_i = \mathbf{q}$*, then* $\forall x$*,* $\forall u$*,*

$$\lim_{i\to\infty} p(x|u,\boldsymbol{\alpha}_i) = p(x|u,\mathbf{q}). \tag{13}$$

**Remark 1.** *Recall that the fusion center knows the empirical distribution* $\mathbf{q}^n$ *and that the optimal rule is provided by* (9)*. Hence, if* (33) *does not hold, then the threshold* $\frac{p(H=0|\mathbf{q}^n)}{p(H=1|\mathbf{q}^n)}$ *may either grow or decay exponentially quickly, biasing the fusion center to the point that the decisions* $\mathbf{u}$ *become irrelevant. Hence, if the empirical distribution of the state carries too much information about the hypothesis, then the probability of error can be driven to zero exponentially quickly by simply looking at the network state, regardless of the rules used by the agents, leading to the need for Assumption 5.b. Assumptions 5.c and 5.d imply that if two distributions in* $\mathcal{P}^m$ *are close with respect to the standard Euclidean metric, then the resulting signal and channel models should be close for all* $y$ *and* $x$*, respectively.*

## 4. Main Results and Important Corollaries

We first consider the single-class result (Theorem 1). We discuss its implications and outline the needed proof techniques, then turn our attention to the multi-class case, which begins by extending Theorem 1 to Lemma 1 and then stating Theorem 2 and its implications. For the main theorems, we provide proof outlines in this section and the complete proofs in Section 6. The extension of Theorem 1 to Lemma 1 is provided in Appendix A.2.

### *4.1. Single-Class Results*

**Theorem 1.** *Subject to Assumptions 5.a–5.d,*

$$\Lambda = -\lim_{n\to\infty}\inf_{\gamma}\min_{\lambda\in[0,1]}\max_{\boldsymbol{q}\in\mathcal{P}^m} -D(\boldsymbol{q}||\boldsymbol{p}) + \frac{1}{n}\log\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q})^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q})^{\lambda}, \tag{14}$$

*where* $D(\boldsymbol{q}||\boldsymbol{p})$ *is the Kullback–Leibler (KL) divergence between the distribution* $\boldsymbol{q}\in\mathcal{P}^m$ *and the true state distribution* $\boldsymbol{p}$*.*

Theorem 1 provides an alternative asymptotically equivalent expression for the error exponent. In particular, Theorem 1 states that a single distribution dominates the asymptotic performance. Interestingly, the dominating distribution is in general *not the true distribution of the agents' states*, despite the fact that the empirical distribution of the states converges towards the true distribution. We then extend Theorem 1 to multiple classes; if agents with a single class use a common rule, the error exponent for each class depends only on the ratios of numbers of agents between classes.

We underscore why the Chernoff information is challenging to compute for our problem framework:

$$\Lambda = -\lim_{n\to\infty}\inf_{\gamma}\min_{\lambda\in[0,1]}\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x},\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x},\boldsymbol{q}^n)^{\lambda}. \tag{15}$$

As $n$ grows, so does the space of potential strategies $\gamma$, possible messages $\boldsymbol{x}$, and possible types $\mathcal{Q}^n$. Even if we have identical agents using the same rule, the complexity and coupling due to the summation over $\boldsymbol{q}^n$ remains. If agents use the same rule

$$\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n}p_0(\boldsymbol{x},\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x},\boldsymbol{q}^n)^{\lambda}=\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n}p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}p(\boldsymbol{q}^n|H=0)^{1-\lambda}p(\boldsymbol{q}^n|H=1)^{\lambda} \tag{16}$$

$$=\frac{1}{n}\log\sum_{\boldsymbol{q}^n}p(\boldsymbol{q}^n|H=0)^{1-\lambda}p(\boldsymbol{q}^n|H=1)^{\lambda}\int_{\boldsymbol{x}}\prod_k p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda}p_1^k(x_k|\boldsymbol{q}^n)^{\lambda} \tag{17}$$

$$=\frac{1}{n}\log\sum_{\boldsymbol{q}^n}p(\boldsymbol{q}^n|H=0)^{1-\lambda}p(\boldsymbol{q}^n|H=1)^{\lambda}\prod_k\int_x p_0^k(x|\boldsymbol{q}^n)^{1-\lambda}p_1^k(x|\boldsymbol{q}^n)^{\lambda} \tag{18}$$

$$\overset{(a)}{=}\frac{1}{n}\log\sum_{\boldsymbol{q}^n}p(\boldsymbol{q}^n|H=0)^{1-\lambda}p(\boldsymbol{q}^n|H=1)^{\lambda}\left(\int_x p_0^1(x|\boldsymbol{q}^n)^{1-\lambda}p_1^1(x|\boldsymbol{q}^n)^{\lambda}\right)^n, \tag{19}$$

where $(a)$ is due to the fact that agents are identical and use the same rule, then all terms in the product are identical. Note that due to the summation over $\boldsymbol{q}^n$, as previously stated, the complexity of calculating the Chernoff information grows with $n$, leading to the need for Theorem 1.

There are a few key remarks that must be made here about Theorem 1:

1. The maximization occurs over $\mathcal{P}^m$ instead of $\mathcal{Q}^n$; hence, we have directly removed the dependence on $\boldsymbol{q}^n$. Because the expression in Theorem 1 is continuous over the compact set $\mathcal{P}^m$, it always achieves its maximum (versus supremum). This is due to Assumptions 5.c and 5.d.
2. Note that the second term is the classical Chernoff information corresponding to the fixed distributions $p_h(\boldsymbol{x}|\boldsymbol{q})$, $h\in\{0,1\}$ and that the KL divergence term can be thought of as a bias. Hence, we only need to consider the $m$-dimensional probability vector that yields the worst Chernoff information biased by the KL divergence. In a certain sense, $\boldsymbol{q}$ is sufficiently close to the true state distribution $p$, such that its poor performance (under strategy $\gamma$) cannot be ignored even in asymptotically large networks. *Only one distribution in $\mathcal{P}^m$ dominates the asymptotic performance*, as expected, although it may not be the true distribution $p$. An instantiation of this is provided in the numerical results.
3. The maximization for $\boldsymbol{q}$ takes place over all of $\mathcal{P}^m$; however, it is only necessary to search a subset of $\mathcal{P}^m$ to find the maximum, thereby reducing the computational cost. To determine the subset of interest, observe that

$$\min_{\boldsymbol{q}\in\mathcal{P}^m}D(\boldsymbol{q}||\boldsymbol{p})-\frac{1}{n}\log\int_{\boldsymbol{x}}p_0(\boldsymbol{x}|\boldsymbol{q})^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q})^{\lambda}\leq D(\boldsymbol{p}||\boldsymbol{p})-\frac{1}{n}\log\int_{\boldsymbol{x}}p_0(\boldsymbol{x}|\boldsymbol{p})^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{p})^{\lambda} \tag{20}$$

$$=-\frac{1}{n}\log\int_{\boldsymbol{x}}p_0(\boldsymbol{x}|\boldsymbol{p})^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{p})^{\lambda} \tag{21}$$

$$\overset{(a)}{\leq}-\frac{1}{n}\log\sum_{\boldsymbol{u}}p_0(\boldsymbol{u}|\boldsymbol{p})^{1-\lambda}p_1(\boldsymbol{u}|\boldsymbol{p})^{\lambda} \tag{22}$$

$$\overset{(b)}{\leq}-\frac{1}{n}\log\int_{\boldsymbol{y}}p_0(\boldsymbol{y}|\boldsymbol{p})^{1-\lambda}p_1(\boldsymbol{y}|\boldsymbol{p})^{\lambda}, \tag{23}$$

where both $(a)$ and $(b)$ are due to Hölder's inequality. Using the fact that the Chernoff information is non-negative [44], it can be seen that the distribution $\boldsymbol{q}^*$ that achieves the maximum over $\mathcal{P}^m$ must satisfy

$$D(\boldsymbol{q}^*||\boldsymbol{p})\leq-\frac{1}{n}\log\int_{\boldsymbol{y}}p_0(\boldsymbol{y}|\boldsymbol{p})^{1-\lambda}p_1(\boldsymbol{y}|\boldsymbol{p})^{\lambda}=c(\lambda,\boldsymbol{p}). \tag{24}$$

The right-hand side of (24) is the Chernoff information for the signal model under distribution $p$; hence, the maximizing $\boldsymbol{q}^*$ must live in a ball defined by the Kullback–Leibler divergence centered at the distribution $p$ with radius $c(\lambda,\boldsymbol{p})$, thereby reducing the search space for the optimization. In fact, the Chernoff information admits a closed-

form solution for a wide range of distributions, such as members of the exponential family [48]

4. The expression in Theorem 1 admits the following property: for all $\boldsymbol{q} \in \mathcal{P}^m$ and $\lambda \in [0,1]$,

$$\frac{1}{n}\log\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q})^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q})^{\lambda} \overset{(a)}{=} \frac{1}{n}\log\int_{\boldsymbol{x}} \Big[\prod_{k=1}^{n} p_0^k(x_k|\boldsymbol{q})\Big]^{1-\lambda}\Big[\prod_{k=1}^{n} p_1^k(x_k|\boldsymbol{q})\Big]^{\lambda} \tag{25}$$

$$= \frac{1}{n}\log\prod_{k=1}^{n}\int_{x_k} p_0^k(x_k|\boldsymbol{q})^{1-\lambda}p_1^k(x_k|\boldsymbol{q})^{\lambda} = \frac{1}{n}\sum_{k=1}^{n}\log\int_{x_k} p_0^k(x_k|\boldsymbol{q})^{1-\lambda}p_1^k(x_k|\boldsymbol{q})^{\lambda}, \tag{26}$$

where $(a)$ holds due to both Equations (5) and (8). Then, for agents using a common rule, all terms in the sum are equal; thus,

$$\frac{1}{n}\log\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q})^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q})^{\lambda} = \log\int_{x} p_0^1(x|\boldsymbol{q})^{1-\lambda}p_1^1(x|\boldsymbol{q})^{\lambda}, \tag{27}$$

which does not depend on $n$, helping to simplify analysis.

We next sketch the proof of Theorem 1. We start from the classical Chernoff information and use it to show that

$$\Lambda = -\lim_{n\to\infty}\inf_{\gamma}\min_{\lambda\in[0,1]}\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}p(\boldsymbol{q}^n). \tag{28}$$

To prove the result, we wish to show that

$$\left|\frac{1}{n}\log\frac{\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}p(\boldsymbol{q}^n)}{\max_{\boldsymbol{q}\in\mathcal{P}^m}\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q})^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q})^{\lambda}2^{-nD(\boldsymbol{q}||\boldsymbol{p})}}\right| < \epsilon. \tag{29}$$

*uniformly in* $\lambda$ *and* $\gamma$, that is, we wish to show that for any $\epsilon > 0$ there exists an integer $n_\epsilon$ *that is independent of* $\lambda$ *and* $\gamma$ such that (29) holds for all $n \geq n_\epsilon$. Uniform convergence in $\lambda$ and $\gamma$ enables determination of the minimum and infimum, respectively, yielding

$$\begin{aligned}\inf_{\gamma}\min_{\lambda\in[0,1]}\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}p(\boldsymbol{q}^n)\\ -\inf_{\gamma}\min_{\lambda\in[0,1]}\max_{\boldsymbol{q}\in\mathcal{P}^m} -D(\boldsymbol{q}||\boldsymbol{p}) + \frac{1}{n}\log\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q})^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q})^{\lambda} \to 0,\end{aligned} \tag{30}$$

as $n \to \infty$, which is the desired assertion. Equivalently, it can be shown that

$$(1-\epsilon) < \left[\frac{\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}p(\boldsymbol{q}^n)}{\max_{\boldsymbol{q}\in\mathcal{P}^m}\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q})^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q})^{\lambda}2^{-nD(\boldsymbol{q}||\boldsymbol{p})}}\right]^{\frac{1}{n}} < (1+\epsilon), \tag{31}$$

uniformly in $\lambda$ and $\gamma$.

### *4.2. Multi-Class Results*

We now discuss extending the results in the previous section to the case of multiple classes. Consider a set of $n_c < \infty$ classes. For a given class $c \in \{1,2,\ldots;n_c\}$, let $c_k$ be the number of agents that belong to class $c$. Then, let $Y_{c,k} \in \mathcal{Y}$ and $S_{c,k} \in \{1,2,\ldots;m\}$ be the signal and state, respectively, of the $k$th agent in class $c$. Without loss of generality, assume that the signal space $\mathcal{Y}$ and state space $\{1,2,\ldots;m\}$ are the same for all classes.

Furthermore, for a given network state $\boldsymbol{S}$, let $\boldsymbol{Q}^n_{c,\boldsymbol{S}}$ denote the type of the states of the agents belonging to class $c$, i.e.,

$$\boldsymbol{Q}^n_{c,\boldsymbol{S}}(i) = \frac{1}{c_k}\sum_{k=1}^{c_k} \mathbb{1}_{\{S_{c,k}=i\}}, \qquad i \in \{1,\ldots;m\}. \tag{32}$$

For given realizations of the class types $\boldsymbol{q}^n_1,\ldots;\boldsymbol{q}^n_{n_c}$, the signal model and state prior for class $c$ are denoted as $p^c(y|\boldsymbol{q}^n_1,\ldots;\boldsymbol{q}^n_{n_c})$ and $p^c(s)$, respectively, and with $\boldsymbol{p}^c = [p^c(1),\ldots,p^c(m)]^T$. Recall that per Definition 4, all agents in a given class are identical; thus, the signal models and state priors are the same within the class. Let $U_{c,k} \in \{1,2,\ldots;b\}$ be the decision made by the $k$th agent in class $c$ distributed according to $p^{c,k}(u|y)$, with $\gamma^c_k$ being the rule of the $k$th agent in class $c$ (again assuming that, without loss of generality, the decision space $\{1,2,\ldots;b\}$ is the same for all classes). The message of the $k$th agent in class $c$ received by the fusion center is denoted as $X_{c,k}$ and distributed according to $p^c(x|u,\boldsymbol{q}^n_1,\ldots;\boldsymbol{q}^n_{n_c})$. Again, because agents in the same class are identical, the channel model is the same throughout the class. Moreover, let $\boldsymbol{X}_c = [X_{c,1},\ldots;X_{c,c_k}]^T$ be the vector of received messages from all agents in class $c$. We can then extend Assumption 5 to the case of $c$ classes.

**Assumption 6.** *The following assumptions hold for all classes. Hence, for notational simplicity, when referring to class c we remove the k superscript.*

*(a) The hypothesis model obeys the following.*

$$\lim_{n\to\infty}\frac{1}{n}\log \min_{\boldsymbol{q}^n_1,\ldots,\boldsymbol{q}^n_{n_c}} \min\{p(H=1|\boldsymbol{q}^n_1,\ldots,\boldsymbol{q}^n_{n_c}), p(H=0|\boldsymbol{q}^n_1,\ldots,\boldsymbol{q}^n_{n_c})\} = 0. \tag{33}$$

*(b) The signal model is continuous in $\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c}$ for all classes, that is, if $\{\boldsymbol{\alpha}_{1,i}\}_{i\in\mathbb{Z}},\ldots;\{\boldsymbol{\alpha}_{n_c,i}\}_{i\in\mathbb{Z}}$ are sequences in $\mathcal{P}^m$ such that $\lim_{i\to\infty}\boldsymbol{\alpha}_{j,i} = \boldsymbol{q}_j$, $j = 1,2,\ldots;n_c$, then $\forall y$,*

$$\lim_{i\to\infty} p^c_h(y|\boldsymbol{\alpha}_{1,i},\ldots;\boldsymbol{\alpha}_{n_c,i}) = p^c_h(y|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c}), \qquad h \in \{0,1\}. \tag{34}$$

*(c) The channel model is continuous in $\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c}$ for all classes, that is, if $\{\boldsymbol{\alpha}_{1,i}\}_{i\in\mathbb{Z}},\ldots;\{\boldsymbol{\alpha}_{n_c,i}\}_{i\in\mathbb{Z}}$ are sequences in $\mathcal{P}^m$ such that $\lim_{i\to\infty}\boldsymbol{\alpha}_{j,i} = \boldsymbol{q}_j$, $j = 1,2,\ldots;n_c$, then $\forall x$, $\forall u$,*

$$\lim_{i\to\infty} p^c(x|u,\boldsymbol{\alpha}_{1,i},\ldots;\boldsymbol{\alpha}_{n_c,i}) = p^c(x|u,\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c}) \qquad h \in \{0,1\}. \tag{35}$$

The conditions of Assumption 6 closely resemble those of Assumption 5. Namely, Assumption 6.a retains the assumption that the network type for each class should not carry too much information about the hypothesis, while Assumptions 6.b and 6.c extend the assumption that the signal and channel models are continuous in the univariate case to the multi-dimensional case. Before, the models were continuous only in $\boldsymbol{q}_1$, whereas we now assume that are continuous in $\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c}$.

**Lemma 1.** *Assume that $\forall k$ and $\lim_{n\to\infty}\frac{c_k}{n} > 0$; then, under Assumptions 6.a–6.c,*

$$\Lambda = -\lim_{n\to\infty}\inf_{\gamma}\min_{\lambda\in[0,1]}\max_{\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c}}\frac{1}{n}\sum_{c=1}^{n_c} -c_k D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log\int_{\boldsymbol{x}^c} p_0(\boldsymbol{x}^c|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda} p_1(\boldsymbol{x}^c|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda}. \tag{36}$$

Lemma 1 implies that all agents within a given class $c$ use the same rule $\gamma^c$. When referring to the rule used by all agents in class $c$, we use superscripts to avoid confusion with previously defined notation, where a subscript indicates the rule used by a specific agent. Then, the error exponent takes on a form that allows heterogeneous networks with a high degree of interference to be examined. The details of the extensions of Theorem 1 are provided in Appendix A.2; Lemma 1 leads to the following theorem.

**Theorem 2.** *Let $r_c \in [0,1]$ be the fraction of agents that belong to class $c \in \{1,2,\dots;n_c\}$, i.e., $c_k = \lfloor r_c n \rfloor$, with $\sum_{c=1}^{n_c} r_c = 1$ and where $\lfloor x \rfloor$ denotes the largest integer that is less than or equal to $x$. Moreover, suppose that all $r_c$ are held constant as $n \to \infty$ and that agents in the same class use a common rule. Then, under Assumptions 6.a–6.c,*

$$\Lambda = -\inf_{\gamma^1,\dots;\gamma^{n_c}} \min_{\lambda\in[0,1]} \max_{\boldsymbol{q}_1,\dots;\boldsymbol{q}_{n_c}} \sum_{c=1}^{n_c} r_c \left[ -D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1,\dots;\boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1,\dots;\boldsymbol{q}_{n_c})^{\lambda} \right]. \tag{37}$$

Because identical agents with a common rule may not be optimal, Theorem 2 provides a lower bound on the optimal error exponent. We highlight several important points of Theorem 2 below:

1. Observe that all agents are coupled through the distributions $\boldsymbol{q}_1,\dots;\boldsymbol{q}_{n_c}$, and recall that for a given class $c$, $\boldsymbol{q}_c$ depends on *all agents in class c* through their states $S_{c,k}$. Hence, the distributions $\boldsymbol{q}_1,\dots;\boldsymbol{q}_{n_c}$ collectively depend on all agents in the network, meaning that the received signal, decision, and message for a given agent are dependent on *all agents in the network*. As a result, Theorem 2 captures a very strong form of coupling.
2. Note that the expression in Theorem 2 is not expressed as a limit, does not depend on $n$, and does not depend on the actual size of the classes. Hence, Theorem 2 provides an objective function that can be used to design rules $\gamma^1,\dots;\gamma^{n_c}$ *that do not depend on the size of the network*.
3. Theorem 2 depends only on the ratios of the classes; that is, Theorem 2 provides an explicit objective function to find the optimal ratios for asymptotically large networks. Specifically, to find the optimal ratios we can solve

$$\min_{r_1,\dots;r_{n_c}} \inf_{\gamma^1,\dots;\gamma^{n_c}} \min_{\lambda\in[0,1]} \max_{\boldsymbol{q}_1,\dots;\boldsymbol{q}_{n_c}} \sum_{c=1}^{n_c} r_c \left[ -D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1,\dots;\boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1,\dots;\boldsymbol{q}_{n_c})^{\lambda} \right]. \tag{38}$$

In the next section, we present a numerical example that highlights the utility of the proposed framework.

## 5. Numerical Example

We design an example that highlights the different forms of coupling captured by our framework. Note that the total number of agents is never specified, as it is only the fraction of agents in each class (ratio) that matters. However, considering our asymptotic analysis, the network size must be sufficiently large. Consider a three-class system where all agents take one of two states (1 or 2) with $p^1(S=1) = 0.5$ and $p^2(S=1) = p^3(S=1) = 0.9$, under each hypothesis all classes observe a Gaussian random variable with signal models

$$p_h^1(y|\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3) = \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2}\left(y-\mu(h,\boldsymbol{q}_2)\right)^2 \right), \qquad \begin{bmatrix}\mu(0,\boldsymbol{q}_2)\\ \mu(1,\boldsymbol{q}_2)\end{bmatrix} = \begin{bmatrix}0\\ \alpha r_1 \boldsymbol{q}_2(1)\end{bmatrix} \tag{39}$$

and

$$p_h^c(y|\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3) = \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2}y^2 \right), \qquad c \in \{2,3\}, \tag{40}$$

where $\mu(h,\boldsymbol{q}_2)$ is the mean of the signal model when $H = h \in \{0,1\}$, $\boldsymbol{q}_2$ is the empirical distribution of Class 2, $\alpha$ is a constant that determines the separation between the means of the two hypotheses, and $r_i = i_k / \left(\sum_{c=1}^{3} c_k\right)$ is the ratio for Class $i$.

Important notes about the signal models are as follows:

1. When $H = 1$, the signal model for Class 1 depends only on the number of agents in Class 2 that are in State 1.
2. The signal models for Classes 2 and 3 are constant with respect to the underlying hypothesis as well as the distributions $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, and $\boldsymbol{q}_3$; hence, agents in Class 2 or 3 cannot distinguish between the two hypotheses.

Upon receiving the signal, each agent in class 1 makes a binary decision according to a threshold test, i.e., $u_{1,k} = 1 \iff y_{1,k} \leq \tau$. Observe that because agents in the other two classes cannot distinguish between hypotheses, their decisions do not matter. Note that the agents belonging to class 1 use identical thresholds; while these may not be optimal, they simplify both design and analysis. Each agent then sends its decision over a binary symmetric channel with the following crossover probability:

$$\begin{aligned} p^c(x=1|u=0,\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3) &= p^c(x=0|u=1,\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3) \\ &= \max\left\{|\frac{1}{2} - r_3\boldsymbol{q}_3(1)|,\rho\right\} \qquad c \in \{1,2,3\}, \quad 0 < \rho < \frac{1}{2}. \end{aligned} \tag{41}$$

The parameter $\rho$ governs the minimum achievable crossover probability of the channel. Note that because $|\frac{1}{2} - r_3\boldsymbol{q}_3(1)| \leq \frac{1}{2}$, the crossover probability can never be lower than $\rho$; thus, as $\rho$ increases the channel becomes worse. It can be seen that while Class 2 aids Class 1 in distinguishing between the two hypotheses, Class 3 controls the quality of the channel between the agents and the fusion center. Moreover, if $r_2 = 0$ then agents cannot distinguish the two hypotheses; thus, the error exponent is zero. Similarly, if $r_3 = 0$, the crossover probability for all channels becomes $\frac{1}{2}$; thus, the channel output becomes random and the error exponent becomes zero. This example underscores the impact of cross-class interference on proper optimization of the system. To determine the optimal class ratios, we can solve

$$(r_1^*, r_2^*, r_3^*) = \arg \min_{r_1,r_2,r_2} \max_{\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3} \sum_{c=1}^{3} r_c \left[ -D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log \sum_{x=1}^{2} \sqrt{p_0^c(x|\boldsymbol{q}_1,\boldsymbol{q}_2)p_1^c(x|\boldsymbol{q}_1,\boldsymbol{q}_2)} \right], \tag{42}$$

with $r_1^* + r_2^* + r_3^* = 1$. For computational simplicity, we set $\tau = 15$ and $\lambda = \frac{1}{2}$. These values can be further optimized.

In Figure 2a, we compute the optimal error exponent as a function of the channel quality $\rho$ for various values of $\alpha$. Note that the class ratios are optimized for each data point. Recall that as $\rho$ increases, so does the interference, causing the channel to worsen. The importance of the channel on the overall system performance can be clearly seen. As $\rho$ increases, the minimum achievable crossover probability increases and the best-case quality of the channel decreases; hence, the optimal error exponent decreases along with the quality of the channel. In fact, when $\rho = 0.4$, the optimal error exponent is 0.0136, an entire order of magnitude less than when $\rho = 0.1$. The impact of the signal mean for Class 1 is determined by $\alpha$. Not surprisingly, as the mean increases, the error exponent increases as well; however, we begin to see diminishing returns as we move from $\alpha = 100$ to $\alpha = 150$.

In Figure 2b, the optimal ratio between the three classes is determined as a function of channel quality when $\alpha = 150$. Figure 2b reveals the impact of cross-class interactions. Recall that each class serves a different purpose; Class 1 is the only class that can distinguish between hypotheses, Class 2 controls the sensing capabilities of Class 1, and Class 3 controls the channel quality for Class 1. Hence, the performance of the system relies on the interactions between the three classes. In particular, as $\rho$ increases Class 3 becomes less important to the overall system, as the quality of the channel degrades. This can be seen in Figure 2b by the decreasing $r_3^*$ and the fact that Class 1 becomes more important to the system, hence the increasing $r_1^*$.

Finally, we examine the optimizing distribution for computing the error exponents when $\alpha = 100$. As previously noted, the true class distributions of the states ($p^c$) do not necessarily dominate asymptotic performance. This can be seen in Figure 2c, which shows that the optimal types are sometimes different from the true distributions. Recall that under $S = 1$ we have $p^1 = 0.5$ and $p^2 = p^3 = 039$; thus, in this three-class example, it is only when $\rho = 0.06$ that we see the optimizing distribution aligning with the true distribution. We underscore that the network type converges to the true state distribution. Recall that we assume the signal and channel models to be continuous; hence, as the network types converge to the true distributions, the performances of all other distributions in a

neighborhood around the true distributions are relatively close. Then, it may be beneficial to design the rule $\gamma$ to optimize detection for a distribution *close to the true distributions*, as the performance difference is small. This trade-off is captured by our result, where the closeness to $p^c$ is captured by the KL divergence and the asymptotic detection performance is captured by the Chernoff information term. Hence, the dominating distribution is the one that offers the best trade-off.

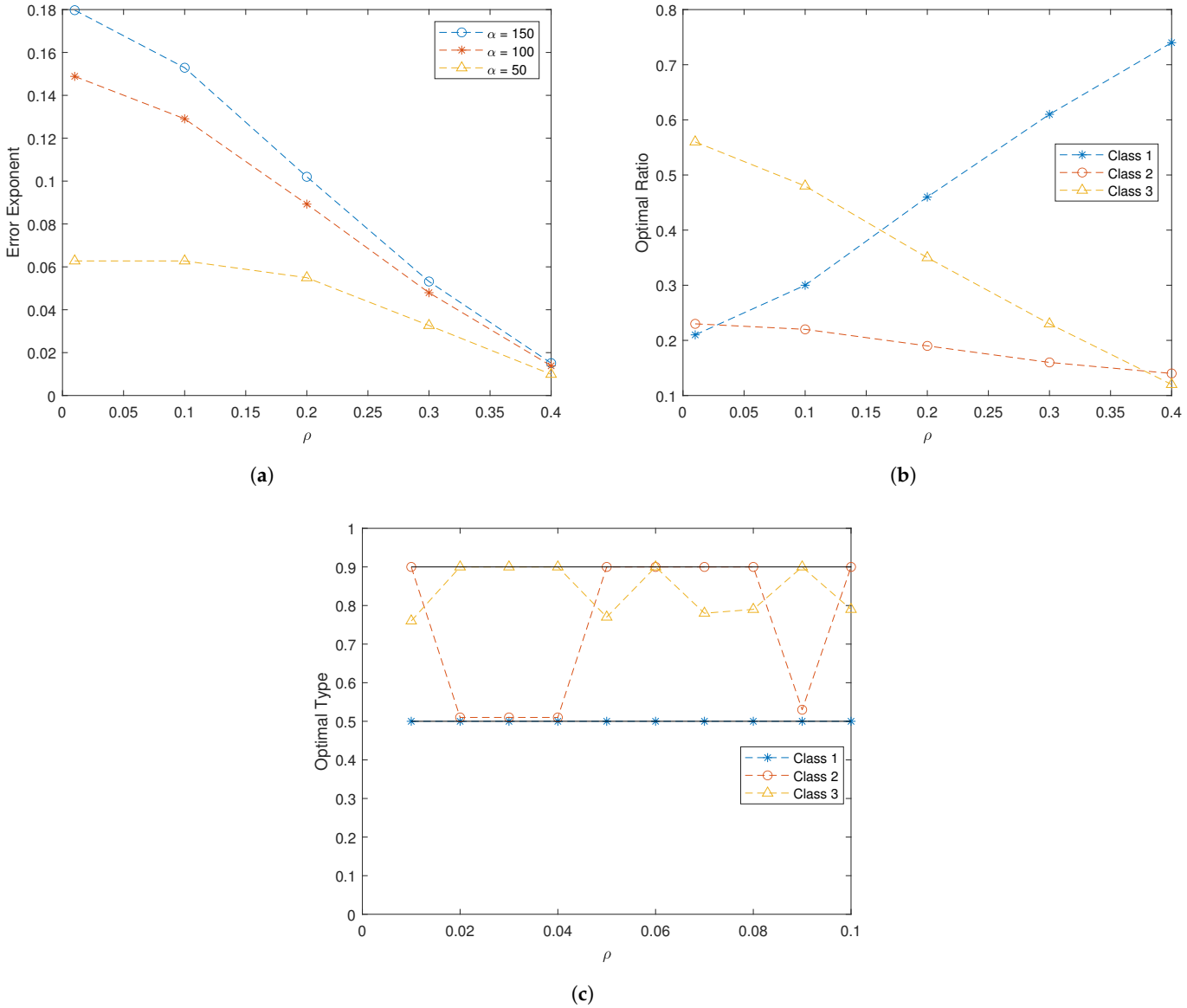


**Figure 2.** Three-class example with coupled signaling and state-dependent channels: (**a**) the optimal error exponent as a function of $\rho$, highlighting the importance of the channel on the overall system; (**b**) the optimal class ratios for $\alpha = 150$ (as $\rho$ increases, Class 3 becomes less important to the overall system); and (**c**) the dominating distributions $\alpha = 100$, which may be different from the true distributions.

## 6. Proofs

### *6.1. Proof of Theorem 1*

Before we begin the proof, we must introduce a number of important definitions and lemmas. There are two sets of lemmas. The first set of lemmas is a series of well-known mathematical facts. Because these are not our contributions but are necessary for the proof of Theorem 1, we omit the proofs, though we provide appropriate citations as

necessary. The second set of lemmas is a series of results that, while necessary, are not major contributions of this work; these proofs are provided in Appendix A.1.

#### 6.1.1. Definitions

**Definition 5.** *A family of functions $\mathcal{F}$ defined on a common domain is **equicontinuous** at a point $x_o$ if for any $\epsilon > 0$ there exists a $\delta > 0$ (possibly a function of $\epsilon$ and $x_o$) such that whenever $|x - x_o| < \delta$ we have $|f(x) - f(x_o)| < \epsilon$ for all $f \in \mathcal{F}$.*

Observe that while the $\delta$ above may depend on $\epsilon$ and the specific point $x_o$, it is not allowed to depend on the specific function $f$, i.e., the chosen $\delta$ must work for all functions in $\mathcal{F}$. The next definition removes the dependence on $x_o$.

**Definition 6.** *A family of functions $\mathcal{F}$ is **uniformly equicontinuous** if for any $\epsilon > 0$ there exists a $\delta > 0$ (possibly a function of $\epsilon$) such that whenever $|x - y| < \delta$ we have $|f(x) - f(y)| < \epsilon$ for all $f \in \mathcal{F}$.*

The above definition states that the same $\delta$ must work for all functions $f \in \mathcal{F}$ at all points in the domain.

**Definition 7.** *Given a family of Lebesgue measurable functions $\mathcal{F}$ with $\int_x |f(x)| < \infty$ for all $f(x) \in \mathcal{F}$, the integrals $\int_x f(x)$ are **uniformly absolutely continuous** if $\forall \epsilon > 0$ and $\exists \delta > 0$ such that for all Lebesgue measurable sets $\mathcal{A}$ with $\nu(\mathcal{A}) < \delta$*

$$\int_{\mathcal{A}} |f(x)| < \epsilon, \tag{43}$$

*for all $f \in \mathcal{F}$, where $\nu$ denotes the Lebesgue measure. Of course, these definitions can be extended to any general measure space; however, we focus on the Lebesque measure here for simplicity and to avoid endlessly defining notation. For a thorough discussion of abstract measure spaces, see [49].*

Again, it is important to distinguish that the same $\delta$ must work for all functions $f \in \mathcal{F}$ for a given $\epsilon$.

**Definition 8.** *Assume that we have a family of measurable functions $\mathcal{F}$ with $\int_x |f(x)| < \infty$ for all $f \in \mathcal{F}$. Moreover, define $\mathcal{I}_a = [-a, a]$. Then, the integrals $\int_x f(x)$ are said to be **uniformly absolutely convergent** if*

$$\lim_{a \to \infty} \int_{\mathcal{I}_a} |f(x)| = \int_x |f(x)|, \tag{44}$$

*uniformly in $\mathcal{F}$.*

This is a powerful property, stating that for a given $\epsilon > 0$ there is a large enough $a$ that all functions in $\mathcal{F}$ satisfy

$$\left| \int_{\mathcal{I}_a} |f(x)| - \int_x |f(x)| \right| < \epsilon. \tag{45}$$

#### 6.1.2. Key Lemmas

The following lemmas are needed to prove Theorem 1. However, because most are simply known mathematical facts (except Lemma 3, the proof of which is provided in Appendix A.1), we omit the proofs.

**Lemma 2.** *Let $\mathcal{F}$ be an equicontinuous and pointwise-bounded family of functions defined on a common domain $\mathcal{D}$. If $\mathcal{D}$ is compact, then $\mathcal{F}$ is uniformly equicontinuous on $\mathcal{D}$.*

Observe that $\mathcal{P}^m$ is compact due to it being closed and bounded; because all of our functions (signal models, channel models, etc.) are defined on this space, Lemma 2 allows us to simplify the proof.

**Lemma 3.** *Let $\mathcal{F}$ and $\mathcal{G}$ be families of equicontinuous strictly positive functions defined on a common domain $\mathcal{D}$; furthermore, assume that for each point $x \in \mathcal{D}$ we have $\inf_{f\in\mathcal{F}} f(x) > 0$, $\inf_{g\in\mathcal{G}} g(x) > 0$, $\sup_{f\in\mathcal{F}} f(x) < \infty$, and $\sup_{g\in\mathcal{G}} g(x) < \infty$. Then, the family $\{f(x)^\lambda g(x)^{1-\lambda}\}_{f,g,\lambda}$ for $f \in \mathcal{F}$, $g \in \mathcal{G}$, and $\lambda \in [0,1]$ is equicontinuous on $\mathcal{D}$.*

The next lemma is taken from [49], Theorem 21.

**Lemma 4.** *Let $\{f_i\}$ be a sequence of real measurable functions with $\int_x |f_i(x)| < \infty$. Assume that the integrals $\int_x f_i(x)$ are uniformly absolutely continuous and uniformly absolutely convergent. Moreover, assume that $f_i \to f$ almost everywhere (a.e.); then, $\int_x f(x) < \infty$ and*

$$\lim_{i\to\infty} \int_x |f_i(x) - f(x)| = 0. \tag{46}$$

Lemma 4 provides a nice immediate result. In particular, suppose we have a function of two variables $f(x,y)$ with $\int_x |f(x,y)| < \infty$ for all $y$ and with $\int_x |f(x,y)|$ uniformly absolutely continuous and uniformly absolutely convergent with respect to $y$. In this case, Lemma 4 states that the integral $\int_x f(x,y)$ is *continuous in* $y$. To see this, observe that if $\{y_i\}$ is a sequence with $y_i \to y$, then, per the triangle inequality,

$$\lim_{i\to\infty} \left| \int_x f(x,y_i) - \int_x f(x,y) \right| \le \lim_{i\to\infty} \int_x |f(x,y_i) - f(x,y)| = 0. \tag{47}$$

### *6.2. Intermediate Lemmas*

We next present several intermediate results. The proofs of all these results can be found in Appendix A.1. Moreover, recalling that we assume all agents to be identical, we consequently omit the $k$ superscript in the following lemmas as well as in the proof.

**Lemma 5.** *Subject to Assumptions 5.a–5.d, the following two statements hold:*

*(a)* *There exists a non-negative function $g(x)$ such that $\int_x g(x) < \infty$ and $\forall x, h \in \{0,1\}$, $\forall \gamma$, $\forall \boldsymbol{q} \in \mathcal{P}^m$, and*

$$\sum_u \int_y p(x|u,\boldsymbol{q}) p(u|y) p_h(y|\boldsymbol{q}) = p_h(x|\boldsymbol{q}) \le g(x). \tag{48}$$

*(b)* *We have*

$$\inf_\gamma \min_{\lambda\in[0,1]} \min_{\boldsymbol{q}\in\mathcal{P}^m} \int_x p_0(x|\boldsymbol{q})^{1-\lambda} p_1(x|\boldsymbol{q})^\lambda > 0. \tag{49}$$

**Lemma 6.** *For all $\epsilon > 0$, there exists a $\delta > 0$ (which depends only on $\epsilon$ and $h$) such that whenever $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are two distributions in $\mathcal{P}^m$ with $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta$, then $\int_y |p_h(y|\boldsymbol{\alpha}) - p_h(y|\boldsymbol{\beta})| < \epsilon$ for all $h \in \{0,1\}$.*

**Lemma 7.** *For a fixed $x \in \mathcal{X}$ and $h \in \{0,1\}$, the family $\{p_h(x|\boldsymbol{q}) 2^{-D(q\|p)}\}_\gamma$ which is indexed by $\gamma$ is uniformly equicontinuous on $\mathcal{P}^m$.*

**Lemma 8.** *For a fixed $x \in \mathcal{X}$, the family $\{p_0(x|\boldsymbol{q})^{1-\lambda} p_1(x|\boldsymbol{q})^\lambda 2^{-D(q\|p)}\}_{\gamma,\lambda}$ which is indexed by $\gamma$ and $\lambda \in [0,1]$ is uniformly equicontinuous on $\mathcal{P}^m$.*

**Lemma 9.** *For any $\epsilon > 0$, there exists a $\delta > 0$ (which depends only on $\epsilon$) such that whenever $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are two distributions in $\mathcal{P}^m$ with $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta$, then*

$$\int_x |p_0(x|\boldsymbol{\alpha})^{1-\lambda} p_1(x|\boldsymbol{\alpha})^{\lambda} 2^{-D(\boldsymbol{\alpha}||\boldsymbol{p})} - p_0(x|\boldsymbol{\beta})^{1-\lambda} p_1(x|\boldsymbol{\beta})^{\lambda} 2^{-D(\boldsymbol{\beta}||\boldsymbol{p})}| < \epsilon,$$

*for all* $\gamma$ *and* $\lambda \in [0,1]$.

An immediate consequence of Lemma 9 follows.

**Lemma 10.** *For any* $\epsilon > 0$*, there exists a* $\delta > 0$ *(which depends only on* $\epsilon$*) such that, whenever* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$ *are two distributions in* $\mathcal{P}^m$ *with* $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta$*, we have*

$$\left| \frac{\int_x p_0(x|\boldsymbol{\alpha})^{1-\lambda} p_1(x|\boldsymbol{\alpha})^{\lambda} 2^{-D(\boldsymbol{\alpha}||\boldsymbol{p})}}{\int_x p_0(x|\boldsymbol{\beta})^{1-\lambda} p_1(x|\boldsymbol{\beta})^{\lambda} 2^{-D(\boldsymbol{\beta}||\boldsymbol{p})}} - 1 \right| < \epsilon, \tag{50}$$

*for all* $\gamma$ *and* $\lambda \in [0,1]$.

The final lemma provides us with a starting point for the proof.

**Lemma 11.**

$$\Lambda = -\lim_{n\to\infty} \inf_{\gamma} \min_{\lambda\in[0,1]} \frac{1}{n} \log \int_{\boldsymbol{x}} \sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda} p(\boldsymbol{q}^n).$$

Hence, rather than starting directly with the Chernoff information, we start from the expression in Lemma 11. We are now ready to begin the proof.

**Proof of Theorem 1.** Define

$$\boldsymbol{q}^* = \arg\max_{\boldsymbol{q}\in\mathcal{P}^m} -D(\boldsymbol{q}||\boldsymbol{p}) + \frac{1}{n} \log \int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q})^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q})^{\lambda}. \tag{51}$$

and note that $\boldsymbol{q}^*$ depends on $n$, $\gamma$, and $\lambda$; then, for any $0 < \epsilon < 1$, per Lemma 10, $\exists\delta > 0$, *which depends only on* $\epsilon$, such that whenever $\|\boldsymbol{q} - \boldsymbol{q}^*\|_2 < \delta$,

$$\left| \frac{\int_x p_0^k(x|\boldsymbol{q})^{1-\lambda} p_1^k(x|\boldsymbol{q})^{\lambda} 2^{-D(\boldsymbol{q}||\boldsymbol{p})}}{\int_x p_0^k(x|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}} - 1 \right| < 1 - \sqrt{1-\epsilon}, \tag{52}$$

for all $\gamma$ and $\lambda \in [0,1]$. Because the agents are identical, they differ only by the rules they use; hence, the same $\delta$ works for all agents. For this $\delta$, define

$$\mathcal{T}_\delta^n = \{\boldsymbol{q}^n \in \mathcal{Q}^n : \|\boldsymbol{q}^n - \boldsymbol{q}^*\|_2 < \delta\}, \tag{53}$$

that is, $\mathcal{T}_\delta^n$ is the set of all types that are less than $\delta$ away from $\boldsymbol{q}^*$ based on the Euclidean distance. There are two important points to make here regarding $\mathcal{T}_\delta^n$:

1. Because both $\mathcal{Q}^n$ and $\boldsymbol{q}^*$ depend on $n$, $\mathcal{T}_\delta^n$ does as well; however, because $\delta$ depends only on $\epsilon$, any type in $\mathcal{T}_\delta^n$ satisfies Equation (52) regardless of $n$ or $\boldsymbol{q}^*$.
2. Observe that for any $\boldsymbol{q} \in \mathcal{P}^m$ there exists a type $\boldsymbol{q}^n$ such that $\|\boldsymbol{q} - \boldsymbol{q}^n\|_2 < \frac{1}{n}$. Hence, $\exists n_o$ such that for all $n \geq n_o$ and for any $\boldsymbol{q} \in \mathcal{P}^m$, $\exists \boldsymbol{q}^n$ such that $\|\boldsymbol{q} - \boldsymbol{q}^n\|_2 < \delta$. That is, $\mathcal{T}_\delta^n$ is non-empty for all $n \geq n_o$. Because $n_o$ depends only on $\delta$ and $\delta$ depends only on $\epsilon$, $n_o$ depends only on $\epsilon$, and the same $n_o$ *works for all agents and all* $\lambda \in [0,1]$.

The following argument holds for any $n \geq n_o$. We begin by observing that

$$\frac{\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}}{\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^*)^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q}^*)^{\lambda} 2^{-nD(\boldsymbol{q}^*||\boldsymbol{p})}} = \frac{\int_{\boldsymbol{x}} \left[\prod_k p_0^k(x_k|\boldsymbol{q}^n)\right]^{1-\lambda} \left[\prod_k p_1^k(x_k|\boldsymbol{q}^n)\right]^{\lambda}}{\int_{\boldsymbol{x}} \left[\prod_k p_0^k(x_k|\boldsymbol{q}^*)\right]^{1-\lambda} \left[\prod_k p_1^k(x_k|\boldsymbol{q}^*)\right]^{\lambda} 2^{-nD(\boldsymbol{q}^*||\boldsymbol{p})}} \tag{54}$$

$$= \frac{\int_{\boldsymbol{x}} \prod_k p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^n)^{\lambda}}{\int_{\boldsymbol{x}} \prod_k p_0^k(x_k|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}} \tag{55}$$

$$= \prod_k \frac{\int_x p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^n)^{\lambda}}{\int_x p_0^k(x_k|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}}. \tag{56}$$

Then, we have the following:

$$\sum_{\boldsymbol{q}^n} \frac{\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}}{\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^*)^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q}^*)^{\lambda} 2^{-nD(\boldsymbol{q}^*||\boldsymbol{p})}} p(\boldsymbol{q}^n) = \sum_{\boldsymbol{q}^n} \prod_k \frac{\int_x p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^n)^{\lambda}}{\int_x p_0^k(x_k|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}} p(\boldsymbol{q}^n) \tag{57}$$

$$\overset{(a)}{\leq} \sum_{\boldsymbol{q}^n} \prod_k \frac{\int_x p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^n)^{\lambda}}{\int_x p_0^k(x_k|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}} 2^{-nD(\boldsymbol{q}^n||\boldsymbol{p})} \tag{58}$$

$$= \sum_{\boldsymbol{q}^n} \prod_k \frac{\int_x p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^n)^{\lambda} 2^{-D(\boldsymbol{q}^n||\boldsymbol{p})}}{\int_x p_0^k(x_k|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}} \tag{59}$$

$$\overset{(b)}{\leq} \sum_{\boldsymbol{q}^n} 1 \overset{(c)}{\leq} (n+1)^m, \tag{60}$$

where $(a)$ holds, as $p(\boldsymbol{q}^n) \leq 2^{-nD(\boldsymbol{q}^n||\boldsymbol{p})}$ [17,50], where $(b)$ is due to the definition of $\boldsymbol{q}^*$ and $(c)$ holds because for any $n$ the number of types is upper-bounded by $(n+1)^m$ ([50], Theorem 11.1.1). Then, taking the $n$-th root yields the upper bound

$$\left[\frac{\sum_{\boldsymbol{q}^n} \int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda} p(\boldsymbol{q}^n)}{\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^*)^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q}^*)^{\lambda} 2^{-nD(\boldsymbol{q}^*||\boldsymbol{p})}}\right]^{\frac{1}{n}} \leq (n+1)^{\frac{m}{n}}. \tag{61}$$

Observe that $(n+1)^{\frac{m}{n}} \to 1$; thus, $\exists n_1$ such that $\forall n \geq n_1$, $(n+1)^{\frac{m}{n}} \leq 1+\epsilon$. Turning our attention to the lower bound,

$$\sum_{\boldsymbol{q}^n} \prod_k \frac{\int_x p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^n)^{\lambda}}{\int_x p_0^k(x_k|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}} p(\boldsymbol{q}^n) \tag{62}$$

$$\geq \sum_{\boldsymbol{q}^n \in \mathcal{T}_o^n} \prod_k \frac{\int_x p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^n)^{\lambda}}{\int_x p_0^k(x_k|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}} p(\boldsymbol{q}^n) \tag{63}$$

$$\overset{(a)}{\geq} \frac{1}{(n+1)^m} \sum_{\boldsymbol{q}^n \in \mathcal{T}_o^n} \prod_k \frac{\int_x p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^n)^{\lambda}}{\int_x p_0^k(x_k|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}} 2^{-nD(\boldsymbol{q}^n||\boldsymbol{p})} \tag{64}$$

$$= \frac{1}{(n+1)^m} \sum_{\boldsymbol{q}^n \in \mathcal{T}_o^n} \prod_k \frac{\int_x p_0^k(x_k|\boldsymbol{q}^n)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^n)^{\lambda} 2^{-D(\boldsymbol{q}^n||\boldsymbol{p})}}{\int_x p_0^k(x_k|\boldsymbol{q}^*)^{1-\lambda} p_1^k(x_k|\boldsymbol{q}^*)^{\lambda} 2^{-D(\boldsymbol{q}^*||\boldsymbol{p})}} \tag{65}$$

$$\overset{(b)}{\geq} \frac{1}{(n+1)^m} \sum_{\boldsymbol{q}^n \in \mathcal{T}_o^n} (1-\epsilon)^n \overset{(c)}{\geq} \frac{1}{(n+1)^m} (1-1+\sqrt{1-\epsilon})^n, \tag{66}$$

where $(a)$ holds because $p(\boldsymbol{q}^n) \geq \frac{1}{(n+1)^{|\mathcal{S}|}} 2^{-nD(\boldsymbol{q}^n||\boldsymbol{p})}$ [17,50], $(b)$ is due to the definition of $\mathcal{T}_o^n$, and $(c)$ holds because $\mathcal{T}_o^n$ is non-empty for $n \geq n_o$. Taking the $n$-th root provides

$$\left[\frac{\sum_{\boldsymbol{q}^n} \int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda} p(\boldsymbol{q}^n)}{\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^*)^{1-\lambda} p_1(\boldsymbol{x}|\boldsymbol{q}^*)^{\lambda} 2^{-nD(\boldsymbol{q}^*||\boldsymbol{p})}}\right]^{\frac{1}{n}} \geq (n+1)^{\frac{-m}{n}} (1-1+\sqrt{1-\epsilon}). \tag{67}$$

Observe that $(n+1)^{\frac{-m}{n}} \to 1$; thus, $\exists n_2$ such that $\forall n \geq n_2$, $(n+1)^{\frac{-m}{n}}(1-1+\sqrt{1-\epsilon}) \geq (1-1+\sqrt{1-\epsilon})(1-1+\sqrt{1-\epsilon}) = 1-\epsilon$. Then, we can take $n_\epsilon = \max\{n_o, n_1, n_2\}$, meaning that for all $n \geq n_\epsilon$ we have

$$1-\epsilon \leq \left[\frac{\sum_{\boldsymbol{q}^n}\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}p(\boldsymbol{q}^n)}{\int_{\boldsymbol{x}} p_0(\boldsymbol{x}|\boldsymbol{q}^*)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^*)^{\lambda}2^{-nD(\boldsymbol{q}^*||\boldsymbol{p})}}\right]^{\frac{1}{n}} \leq 1+\epsilon. \tag{68}$$

Because none of $n_o$, $n_1$, or $n_2$ depend on $\boldsymbol{q}^*$, $\gamma$, or $\lambda$, it is the case that $n_\epsilon$ does not depend on $\boldsymbol{q}^*$, $\gamma$, or $\lambda$; hence, we have uniform convergence, which completes the proof. □

*6.3. Proof of Theorem 2*

Because we assume that agents of the same class use the same rule, if we focus on class $c$ we have

$$p_h(\boldsymbol{x}^c|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c}) = \prod_{k=1}^{c_k} p^{c,k}(x_{c,k}|\boldsymbol{q}_1^n,\ldots;\boldsymbol{q}_{n_c}^n), \tag{69}$$

which is a consequence of Equations (5) and (8). Then, we have

$$\begin{aligned} &-c_k D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log\int_{\boldsymbol{x}} p_0(\boldsymbol{x}^c|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda}p_1(\boldsymbol{x}^c|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda} \\ &= \sum_{k=1}^{c_k} -D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log\int_x p_0^{c,k}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda}p_1^{c,k}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda}, \end{aligned} \tag{70}$$

with

$$p_h^{c,k}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c}) = \sum_{u=1}^{b} p^c(x|u,\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})\int_y p^{c,k}(u|y)p_h^c(y|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c}), \qquad h\in\{0,1\}. \tag{71}$$

If all agents in Class $c$ use rule $\gamma^c$, then every term in the sum of Equation (70) is equal. Hence,

$$\begin{aligned} &\frac{1}{n}\sum_{c=1}^{n_c} -c_k D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log\int_{\boldsymbol{x}} p_0(\boldsymbol{x}^c|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda}p_1(\boldsymbol{x}^c|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda} \\ &= \sum_{c=1}^{n_c}\frac{c_k}{n}\left[-D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log\int_x p_0^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda}p_1^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda}\right]. \end{aligned} \tag{72}$$

We now turn our attention to the difference

$$\begin{aligned} &\sum_{c=1}^{n_c} r_c\left[-D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log\int_x p_0^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda}p_1^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda}\right] \\ &\quad -\sum_{c=1}^{n_c}\frac{c_k}{n}\left[-D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log\int_x p_0^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda}p_1^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda}\right], \end{aligned} \tag{73}$$

which is equivalent to

$$\sum_{c=1}^{n_c}\frac{r_c n - \lfloor r_c n\rfloor}{n}\left[-D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log\int_x p_0^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda}p_1^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda}\right]. \tag{74}$$

Observe that

$$-D(\boldsymbol{q}_c||\boldsymbol{p}^c) + \log\int_x p_0^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda}p_1^{c,1}(x|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda}\Big] \leq 0, \tag{75}$$

for all classes, which is a consequence of the non-negativity of the KL divergence [50] and the non-positivity of the Chernoff information [44]. Combining this with the fact that

$\frac{r_c n - \lfloor r_c n \rfloor}{n} \geq 0$, we see that (74) is upper-bounded by zero. For a lower bound, observe that $\frac{r_c n - \lfloor r_c n \rfloor}{n} \leq \frac{1}{n}$, which yields the result that Equation (74) is lower-bounded by

$$\sum_{c=1}^{n_c} \frac{1}{n} \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big] \tag{76}$$

$$\geq \sum_{c=1}^{n_c} \frac{1}{n} \Big[ - \max_{\boldsymbol{q}} D(\boldsymbol{q}||\boldsymbol{p}^c) + \inf_{\gamma} \min_{\lambda \in [0,1]} \min_{\boldsymbol{q}_1, \cdots; \boldsymbol{q}_{n_c}} \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big]. \tag{77}$$

The KL divergence (for finite alphabets) is bounded, and repeating the proof of Lemma 5 for the multi-class case using Assumptions 6.b and 6.c guarantees that the logarithm terms are finite. Hence, Equation (77) goes to zero as $n \to \infty$. Moreover, note that this lower bound is independent of the strategies $\gamma$ and $\lambda$ and the distributions $\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c}$. This means that Equation (74) converges uniformly in $\gamma$ and $\lambda$ and the distributions $\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c}$, which allows us to take the infimum, minimum, and maximum, respectively. To see this, observe that the upper bound provides

$$\begin{aligned} &\max_{\boldsymbol{q}_1, \cdots; \boldsymbol{q}_{n_c}} \sum_{c=1}^{n_c} \frac{c_k}{n} \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big] \\ &\geq \sum_{c=1}^{n_c} \frac{c_k}{n} \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big] \\ &\geq \sum_{c=1}^{n_c} r_c \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big]. \end{aligned} \tag{78}$$

As this is true for all $\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c}$, we have

$$\begin{aligned} &\max_{\boldsymbol{q}_1, \cdots; \boldsymbol{q}_{n_c}} \sum_{c=1}^{n_c} \frac{c_k}{n} \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big] \\ &\geq \max_{\boldsymbol{q}_1, \cdots; \boldsymbol{q}_{n_c}} \sum_{c=1}^{n_c} r_c \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big]. \end{aligned} \tag{79}$$

The same argument can be repeated to obtain

$$\begin{aligned} &\max_{\boldsymbol{q}_1, \cdots; \boldsymbol{q}_{n_c}} \sum_{c=1}^{n_c} r_c \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big] \\ &\geq \max_{\boldsymbol{q}_1, \cdots; \boldsymbol{q}_{n_c}} \sum_{c=1}^{n_c} \frac{c_k}{n} \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big] \\ &+ \sum_{c=1}^{n_c} \frac{1}{n} \Big[ - \max_{\boldsymbol{q}} D(\boldsymbol{q}||\boldsymbol{p}^c) + \inf_{\gamma} \min_{\lambda \in [0,1]} \min_{\boldsymbol{q}_1, \cdots; \boldsymbol{q}_{n_c}} \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big]. \end{aligned} \tag{80}$$

Hence, the difference

$$\begin{aligned} &\max_{\boldsymbol{q}_1, \cdots; \boldsymbol{q}_{n_c}} \sum_{c=1}^{n_c} r_c \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big] \\ &- \max_{\boldsymbol{q}_1, \cdots; \boldsymbol{q}_{n_c}} \sum_{c=1}^{n_c} \frac{c_k}{n} \Big[ - D(\boldsymbol{q}_c || \boldsymbol{p}^c) + \log \int_x p_0^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{1-\lambda} p_1^{c,1}(x|\boldsymbol{q}_1, \ldots; \boldsymbol{q}_{n_c})^{\lambda} \Big], \end{aligned} \tag{81}$$

goes to zero as $n \to \infty$. Repeating the same argument with the minimum over $\lambda$ followed by the infimum over $\gamma$ completes the proof.

## 7. Conclusions

In this paper, we have introduced a new framework for decentralized inference that captures a high degree of coupling between the agents. Under our framework, the

empirical distribution of the network state induces a global coupling across agents. We find an asymptotically equivalent expression to the Chernoff information and unveil a number of interesting properties, such as the fact that the true state distribution does not always dominate asymptotic performance. For the multi-class case, we characterize how ratios of classes of agents affect performance. We further allow for a lossy communication link between the agents and the fusion center and investigate the effects of the channel on overall performance. Our work extends prior work on distributed detection, and is able to break the requirement of conditionally independent observations when correlation is present. In future work, we will remove the fusion center from the system and require agents to directly communicate with each other, as in a purely decentralized ad hoc system. In addition, we will consider the introduction of actions by the agents which can affect observations by other agents to enable the consideration of active hypothesis testing in a distributed setting.

**Author Contributions:** J.S. and U.M. made serious contributions to the work and have had thorough discussions together on the problem formulation, proof techniques, and technical challenges. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work has been funded in part by one or more of the following grants: NSF CCF-1817200, ARO W911NF1910269, DOE DE-SC0021417, Swedish Research Council 2018-04359, NSF CCF-2008927, NSF CCF-2200221, ONR 503400-78050, ONR N00014-15-1-2550, USC + Amazon Center on Secure and Trusted Machine Learning.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors have no conflicts of interest to report.

## Appendix A

*Appendix A.1. Proofs of Lemmas for Theorem 1*

**Proof of Lemma 3.** Fix a point $x_o \in \mathcal{D}$. For $\epsilon > 0$, let $\delta > 0$ (which can depend on $\epsilon$ and $x_o$) be such that $|f(x_o) - f(x)| < \frac{\min\{\inf_{f\in\mathcal{F}} f(x_o),1\}\epsilon}{2}$ for all $f \in \mathcal{F}$ whenever $|x_o - x| < \delta$. Assume w.l.o.g. that $f(x_o) \geq f(x)$; then,

$$f(x_o)^\lambda - f(x)^\lambda = f(x_o)^\lambda - f(x)^\lambda \frac{f(x_o)^{1-\lambda} + f(x)^{1-\lambda}}{f(x_o)^{1-\lambda} + f(x)^{1-\lambda}} \tag{A1}$$

$$= \frac{f(x_o) - f(x) + f(x_o)^\lambda f(x)^{1-\lambda} - f(x_o)^{1-\lambda} f(x)^\lambda}{f(x_o)^{1-\lambda} + f(x)^{1-\lambda}} \tag{A2}$$

$$\overset{(a)}{\leq} \frac{2(f(x_o) - f(x))}{f(x_o)^{1-\lambda}} \tag{A3}$$

$$\overset{(b)}{\leq} \frac{2(f(x_o) - f(x))}{\min\{f(x_o), 1\}} \tag{A4}$$

$$\leq \frac{2(f(x_o) - f(x))}{\min\{\inf_{f\in\mathcal{F}} f(x_o), 1\}} < \epsilon, \tag{A5}$$

where $(a)$ follows from

$$(f(x_o)^\lambda + f(x)^\lambda)(f(x)^{1-\lambda} - f(x_o)^{1-\lambda}) \leq 0, \tag{A6}$$

and $(b)$ follows from $\min_{\lambda\in[0,1]} f(x_o)^{1-\lambda} = f(x_o)$ if $f(x_o) < 1$ and $\min_{\lambda\in[0,1]} f(x_o)^{1-\lambda} = 1$ if $f(x_o) \geq 1$. Hence, we have just shown that for every $\epsilon > 0$, $\exists \delta_{\mathcal{F}}(x_o) > 0$ (independent of $\lambda$ and $f$) such that $\forall \lambda \in [0,1]$ and $\forall f \in \mathcal{F}$, we have $|f(x_o)^\lambda - f(x)| < \epsilon$ whenever $|x_o - x| < \delta_f(x_o)$. A similar argument holds for $g^{1-\lambda}$.

Then, for any $\epsilon > 0$ let $\delta_{\mathcal{F}}(x_o)$ and $\delta_{\mathcal{G}}(x_o)$ be such that $|f(x_o)^\lambda - f(x)^\lambda| < \frac{\epsilon}{2(\max\{\sup_{g\in\mathcal{G}} g(x_o),1\}+\epsilon)}$ and $|g(x_o)^{1-\lambda} - g(x)^{1-\lambda}| < \frac{\epsilon}{2(\max\{\sup_{f\in\mathcal{G}} f(x_o),1\})}$ for all $\lambda$ $f$ and $g$ whenever $|x_o - x| < \delta_{\mathcal{F}}(x_o)$ and $|x_o - x| < \delta_{\mathcal{G}}(x_o)$, respectively. Take $\delta(x_o) = \min\{\delta_f(x_o), \delta_g(x_o)\}$; then, if $|x_o - x| < \delta(x_o)$ for all $\lambda$ we have

$$|f(x_o)^\lambda g(x_o)^{1-\lambda} - f(x)^\lambda g(x)^{1-\lambda}| = |f(x_o)^\lambda g(x_o)^{1-\lambda} - f(x_o)^\lambda g(x)^{1-\lambda} + f(x_o)^\lambda g(x)^{1-\lambda} - f(x)^\lambda g(x)^{1-\lambda}| \tag{A7}$$

$$\overset{(a)}{\leq} f(x_o)^\lambda |g(x_o)^{1-\lambda} - g(x)^{1-\lambda}| + g(x)^{1-\lambda}|f(x_o)^\lambda - f(x)^\lambda| \tag{A8}$$

$$\overset{(b)}{\leq} f(x_o)^\lambda |g(x_o)^{1-\lambda} - g(x)^{1-\lambda}| + (g(x_o)^{1-\lambda} + \epsilon)|f(x_o)^\lambda - f(x)^\lambda| \tag{A9}$$

$$\leq f(x_o)^\lambda \frac{\epsilon}{2(\max\{\sup_{f\in\mathcal{F}} f(x_o),1\})} + (g(x_o)^{1-\lambda} + \epsilon)\frac{\epsilon}{2(\max\{\sup_{g\in\mathcal{G}} g(x_o),1\}+\epsilon)} \tag{A10}$$

$$\overset{(c)}{\leq} \max\{f(x_o),1\}\frac{\epsilon}{2(\max\{\sup_{f\in\mathcal{F}} f(x_o),1\})} + (\max\{g(x_o),1\}+\epsilon)\frac{\epsilon}{2(\max\{\sup_{g\in\mathcal{G}} g(x_o),1\}+\epsilon)} \tag{A11}$$

$$\leq \max\{\sup_{f\in\mathcal{F}} f(x_o),1\}\frac{\epsilon}{2(\max\{\sup_{f\in\mathcal{F}} f(x_o),1\})} + (\max\{\sup_{g\in\mathcal{G}} g(x_o),1\}+\epsilon)\frac{\epsilon}{2(\max\{\sup_{g\in\mathcal{G}} g(x_o),1\}+\epsilon)} \tag{A12}$$

$$= \epsilon, \tag{A13}$$

where $(a)$ is due to the triangle inequality, $(b)$ is due to $g(x)^{1-\lambda} = g(x)^{1-\lambda} - g(x_0)^{1-\lambda} + g(x_o)^{1-\lambda} < \epsilon + g(x_o)^{1-\lambda}$, and $(c)$ is due to the fact that $\max_{\lambda\in[0,1]} f(x_o)^{1-\lambda} = 1$ if $f(x_o) < 1$ and $\max_{\lambda\in[0,1]} f(x_o)^{1-\lambda} = f(x_o)$ if $f(x_o) \geq 1$. □

**Proof of Lemma 5.** The function $g(x) = \sum_u \sup_{\boldsymbol{q}} p(x|u,\boldsymbol{q})$ will always satisfy part (a). To see this, first observe that

$$\sum_u \int_y p(x|u,\boldsymbol{q})p(u|y)p_h(y|\boldsymbol{q}) \leq \sum_u p(x|u,\boldsymbol{q}) \leq \sum_u \sup_{\boldsymbol{q}} p(x|u,\boldsymbol{q}). \tag{A14}$$

for all $h$, $\gamma$, and $\boldsymbol{q}$. Moreover, for each $x$ and $u$, per Assumption 5.d the channel model is continuous in $\boldsymbol{q}$, and because $\mathcal{P}^m$ is compact, $p(x|u,\boldsymbol{q})$ attains its maximum, meaning that $\sup_{\boldsymbol{q}} p(x|u,\boldsymbol{q}) = \max_{\boldsymbol{q}} p(x|u,\boldsymbol{q})$ is a valid density; thus, $\int_x g(x) = \int_x \sum_u \max_{\boldsymbol{q}} p(x|u,\boldsymbol{q}) = b < \infty$. For $b$), repeated use of Hölder's inequality provides

$$\begin{aligned}\int_x p_0(x;\boldsymbol{q})^{1-\lambda} p_1(x;\boldsymbol{q})^\lambda &\geq \int_y p_0(y|\boldsymbol{q})^{1-\lambda} p_1(y|\boldsymbol{q})^\lambda \\ &\overset{(a)}{\geq} \int_y \min\{p_0(y|\boldsymbol{q}), p_1(y|\boldsymbol{q})\} \geq \int_y \inf_{\boldsymbol{q}\in\mathcal{P}^m} \min\{p_0(y|\boldsymbol{q}), p_1(y|\boldsymbol{q})\},\end{aligned} \tag{A15}$$

where $(a)$ holds, as for any two numbers $a$ and $b$ we have $a^{1-\lambda}b^\lambda \geq \min\{a,b\}$ for any $\lambda \in [0,1]$. Per Assumption 5.c, the signal model is continuous in $\boldsymbol{q}$ for $h$ as well as all $y$; thus, $\min\{p_0(y|\boldsymbol{q}), p_1(y|\boldsymbol{q})\}$ is continuous and attains its minimum. Moreover, we assume that $p_h(y|\boldsymbol{q})$ is always strictly positive for any $h$, $y$, and $\boldsymbol{q}$, meaning that

$$\int_x p_0(x;\boldsymbol{q})^{1-\lambda} p_1(x;\boldsymbol{q})^\lambda \geq \int_y \min_{\boldsymbol{q}\in\mathcal{P}^m} \min\{p_0(y|\boldsymbol{q}), p_1(y|\boldsymbol{q})\} > 0. \tag{A16}$$

Because this lower bound holds for all $\gamma$, $\lambda$, and $\boldsymbol{q}$, part $b$) holds. □

**Proof of Lemma 6.** Fix $\boldsymbol{q} \in \mathcal{P}^m$ and let $\{\boldsymbol{\alpha}_i\}$ be any sequence that converges to $\boldsymbol{q}$. Per Assumption 5.c, $p_h(y|\boldsymbol{\alpha}_i) \to p_h(y|\boldsymbol{q})$ for each $y$. Moreover, $\int_y p_h(y|\boldsymbol{\alpha}_i) = \int_y p_h(y|\boldsymbol{q}) = 1$ for all $i$; thus, per Scheffé's lemma [51],

$$\int_y |p_h(y|\boldsymbol{\alpha}_i) - p_h(y|\boldsymbol{q})| \to 0. \tag{A17}$$

The remainder of the proof proceeds via proof by contradiction. Suppose that $\exists \epsilon > 0$ and that there exist two sequences $\{\boldsymbol{\alpha}_i\}$ and $\{\boldsymbol{\beta}_i\}$ such that $\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i \to 0$ and that for all $i$,

$$\int_y |p_h(y|\boldsymbol{\alpha}_i) - p_h(y|\boldsymbol{\beta}_i)| > \epsilon. \tag{A18}$$

Because $\mathcal{P}^m$ is bounded, per the Bolzano–Weierstrass theorem [52], $\{\boldsymbol{\alpha}_i\}$ and $\{\boldsymbol{\beta}_i\}$ have convergent subsequences $\{\boldsymbol{\alpha}_{i_k}\}$ and $\{\boldsymbol{\beta}_{i_k}\}$ that converge to some point $\boldsymbol{\theta}$. Because $\mathcal{P}^m$ is closed, $\boldsymbol{\theta}$ must be in $\mathcal{P}^m$. Then, Equation (A17) provides

$$\int_y |p_h(y|\boldsymbol{\alpha}_{i_k}) - p_h(y|\boldsymbol{\beta}_{i_k})| \leq \int_y |p_h(y|\boldsymbol{\alpha}_{i_k}) - p_h(y|\boldsymbol{\theta})| + \int_y |p_h(y|\boldsymbol{\beta}_{i_k}) - p_h(y|\boldsymbol{\theta})| \to 0, \tag{A19}$$

which contradicts Equation (A18). □

**Proof of Lemma 7.** Recall that Assumption 5.d states that $p(x|u,\boldsymbol{q})$ is continuous in $\boldsymbol{q}$ for any $x$ and $u \in \{1,2,\ldots;b\}$; thus, $\max_u p(x|u,\boldsymbol{q})$ is continuous in $\boldsymbol{q}$ for a fixed $x$. Moreover, because $\mathcal{P}^m$ is compact, $\max_u p(x|u,\boldsymbol{q})$ achieves its maximum on $\mathcal{P}^m$; thus, $0 < \max_{\boldsymbol{q}\in\mathcal{Z}} \max_u p(x|u,\boldsymbol{q}) < \infty$. Then, per Lemma 6, for any $\epsilon > 0$ $\exists \delta_0 > 0$ such that whenever $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta_0$ we have $\int_y |p_h(y|\boldsymbol{\alpha}) - p_h(y|\boldsymbol{\beta})| < \frac{\epsilon}{2b \max_{\boldsymbol{q}\in\mathcal{Z}} \max_u p(x|u,\boldsymbol{q})}$. Then, for all $u$ and $\gamma$ we have

$$\begin{aligned} |p_h(u|\boldsymbol{\alpha}) - p_h(u|\boldsymbol{\beta})| &= \left| \int_y p(u|y)p_h(y|\boldsymbol{\alpha}) - \int_y p(u|y)p_h(y|\boldsymbol{\beta}) \right| \\ \overset{(a)}{\leq} \int_y p(u|y)|p_h(y|\boldsymbol{\alpha}) - p_h(y|\boldsymbol{\beta})| &\overset{(b)}{\leq} \int_y |p_h(y|\boldsymbol{\alpha}) - p_h(y|\boldsymbol{\beta})| < \frac{\epsilon}{2b \max_{\boldsymbol{q}\in\mathcal{Z}} \max_u p(x|u,\boldsymbol{q})}, \end{aligned} \tag{A20}$$

where $(a)$ holds due to the triangle inequality and $(b)$ holds because $p(u|y) \leq 1$ for all $u$ and $\gamma$. Moreover, because $p(x|u,\boldsymbol{q})$ is uniformly continuous in $\boldsymbol{q}$ for each $u \in \{1,2,\ldots;b\}$, $\exists \delta_u > 0$ (which depends only on $\epsilon$, $x$, and $u$) such that whenever $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta_u$, $|p(x|u,\boldsymbol{\alpha}) - p(x|u,\boldsymbol{\beta})| < \frac{\epsilon}{2b}$. For a fixed $x$, take $\delta = \min\{\delta_0, \delta_1, \ldots; \delta_b\}$; thus, $\delta$ depends only on $\epsilon$, $h$, and $x$, and does not depend on $\gamma$ or $u$. Then, whenever $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta$,

$$|p_h(x|\boldsymbol{\alpha}) - p_h(x|\boldsymbol{\beta})| = \left| \sum_{u=1}^{b} p(x|u,\boldsymbol{\alpha})p_h(u|\boldsymbol{\alpha}) - \sum_{u=1}^{b} p(x|u,\boldsymbol{\beta})p_h(u|\boldsymbol{\beta}) \right| \tag{A21}$$

$$\overset{(a)}{\leq} \sum_{u=1}^{b} |p(x|u,\boldsymbol{\alpha})p_h(u|\boldsymbol{\alpha}) - p(x|u,\boldsymbol{\beta})p_h(u|\boldsymbol{\beta})| \tag{A22}$$

$$= \sum_{u=1}^{b} |p(x|u,\boldsymbol{\alpha})p_h(u|\boldsymbol{\alpha}) - p(x|u,\boldsymbol{\alpha})p_h(u|\boldsymbol{\beta}) + p(x|u,\boldsymbol{\alpha})p_h(u|\boldsymbol{\beta}) - p(x|u,\boldsymbol{\beta})p_h(u|\boldsymbol{\beta})| \tag{A23}$$

$$\overset{(b)}{\leq} \sum_{u=1}^{b} p(x|u,\boldsymbol{\alpha})|p_h(u|\boldsymbol{\alpha}) - p_h(u|\boldsymbol{\beta})| + p_h(u|\boldsymbol{\beta})|p(x|u,\boldsymbol{\alpha}) - p(x|u,\boldsymbol{\beta})| \tag{A24}$$

$$\overset{(c)}{\leq} \sum_{u=1}^{b} |p_h(u|\boldsymbol{\alpha}) - p_h(u|\boldsymbol{\beta})| \max_{\boldsymbol{q}\in\mathcal{Z}} \max_u p(x|u,\boldsymbol{q}) + |p(x|u,\boldsymbol{\alpha}) - p(x|u,\boldsymbol{\beta})| \tag{A25}$$

$$\leq \sum_{u=1}^{b} \frac{\epsilon}{2b \max_{\boldsymbol{q}\in\mathcal{Z}} \max_u p(x|u,\boldsymbol{q})} \max_{\boldsymbol{q}\in\mathcal{Z}} \max_u p(x|u,\boldsymbol{q}) + \frac{\epsilon}{2b} = \epsilon, \tag{A26}$$

where $(a)$ and $(b)$ are due to the triangle inequality and $(c)$ is due to $p_h(u|\boldsymbol{\beta}) < 1$. Now, because $D(\boldsymbol{q}||\boldsymbol{p})$ is continuous in $\boldsymbol{q}$ (for finite alphabets) and $\mathcal{P}^m$ is compact, $D(\boldsymbol{q}||\boldsymbol{p})$ is uniformly continuous. Clearly, $D(\boldsymbol{q}||\boldsymbol{p})$ does not depend on $\gamma$; hence, we may repeat the exact same argument with $p_h(x|\boldsymbol{q})2^{-D(\boldsymbol{q}||\boldsymbol{p})}$, providing the desired result. □

**Proof of Lemma 8.** In order to use Lemma 3, we need to show that for each $\boldsymbol{q} \in \mathcal{P}^m$ and $h \in \{0,1\}$ it is the case that $\inf_\gamma p_h(x|\boldsymbol{q})2^{-D(\boldsymbol{q}||\boldsymbol{p})} > 0$ and $\sup_\gamma p_h(x|\boldsymbol{q})2^{-D(\boldsymbol{q}||\boldsymbol{p})} < \infty$.

Because $0 < 2^{-D(q||p)} \leq 1$ (due to $D(\boldsymbol{q}||\boldsymbol{p})$ being bounded for finite alphabets), it suffices to show that $\inf_\gamma p_h(x|\boldsymbol{q}) > 0$ and $\sup_\gamma p_h(x|\boldsymbol{q}) < \infty$. For any $\gamma$, observe that

$$p_h(x|\boldsymbol{q}) = \sum_u p(x|u,\boldsymbol{q}) p_h(u|\boldsymbol{q}) \geq \min_u p(x|u,\boldsymbol{q}) \sum_u p_h(u|\boldsymbol{q}) = \min_u p(x|u,\boldsymbol{q}) \overset{(a)}{>} 0, \tag{A27}$$

where $(a)$ is due to Assumption 4. Taking the infimum over $\gamma$ provides us with $\inf_\gamma p_h(x|\boldsymbol{q}) \geq \min_u p(x|u,\boldsymbol{q}) > 0$. A similar argument shows that $\sup_\gamma p_h(x|\boldsymbol{q}) < \infty$. Then, Lemma 3 yields the desired assertion. □

**Proof of Lemma 9.** We proceed via proof by contradiction. Suppose that $\exists \epsilon > 0$ and that we have sequences $\{\boldsymbol{\alpha}_i\}$, $\{\boldsymbol{\beta}_i\}$, and $\{\lambda_i\}$ such that $\boldsymbol{\alpha}_i - \boldsymbol{\beta}_i \to 0$ and

$$\int_x |p_0(x|\boldsymbol{\alpha}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\alpha}_i)^{\lambda_i} 2^{-D(\boldsymbol{\alpha}_i||\boldsymbol{p})} - p_0(x|\boldsymbol{\beta}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\beta}_i)^{\lambda_i} 2^{-D(\boldsymbol{\beta}_i||\boldsymbol{p})}| \geq \epsilon, \tag{A28}$$

for all $i$. Then, let $f_i(x) = |p_0(x|\boldsymbol{\alpha}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\alpha}_i)^{\lambda_i} 2^{-D(\boldsymbol{\alpha}_i||\boldsymbol{p})} - p_0(x|\boldsymbol{\beta}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\beta}_i)^{\lambda_i} 2^{-D(\boldsymbol{\beta}_i||\boldsymbol{p})}|$. Per Lemma 8, $f_i \to 0$ a.e.; then, in order to use Lemma 4 we must show the integrals $\int_x f_i(x)$ are uniformly absolutely continuous, uniformly absolutely convergent, and $\int_x |f_i(x)| < \infty$ for all $i$. Recall that per Lemma 5 there exists a function $g(x)$ such that for all $\boldsymbol{q}$ and $h$ we have $p_h(x|\boldsymbol{q}) \leq g(x)$ and $\int_x g(x) < \infty$. Then,

$$\begin{aligned}
\int_x f_i(x) &= \int_x |p_0(x|\boldsymbol{\alpha}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\alpha}_i)^{\lambda_i} 2^{-D(\boldsymbol{\alpha}_i||\boldsymbol{p})} - p_0(x|\boldsymbol{\beta}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\beta}_i)^{\lambda_i} 2^{-D(\boldsymbol{\beta}_i||\boldsymbol{p})}| && \text{(A29)}\\
&\overset{(a)}{\leq} \int_x |p_0(x|\boldsymbol{\alpha}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\alpha}_i)^{\lambda_i} 2^{-D(\boldsymbol{\alpha}_i||\boldsymbol{p})}| + \int_x |p_0(x|\boldsymbol{\beta}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\beta}_i)^{\lambda_i} 2^{-D(\boldsymbol{\beta}_i||\boldsymbol{p})}| && \text{(A30)}\\
&\leq \int_x |p_0(x|\boldsymbol{\alpha}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\alpha}_i)^{\lambda_i}| + \int_x |p_0(x|\boldsymbol{\beta}_i)^{1-\lambda_i} p_1(x|\boldsymbol{\beta}_i)^{\lambda_i}| && \text{(A31)}\\
&\overset{(b)}{\leq} \int_x \max\{p_0(x|\boldsymbol{\alpha}_i), p_1(x|\boldsymbol{\alpha}_i)\} + \int_x \max\{p_0(x|\boldsymbol{\beta}_i), p_1(x|\boldsymbol{\beta}_i)\} && \text{(A32)}\\
&\leq \int_x g(x) + \int_x g(x) = 2\int_x g(x) < \infty, && \text{(A33)}
\end{aligned}$$

where $(a)$ is due to the triangle inequality, $(b)$ holds due to $a^\lambda b^{1-\lambda} \leq \max\{a,b\}$ for any two real numbers $a, b$, and $\lambda \in [0,1]$. Furthermore, due to the absolute continuity of the Lebesgue integral, $\exists \delta > 0$ for any $\epsilon > 0$ such that for any measurable set $\mathcal{A}$ with Lebesgue measure $\nu(\mathcal{A}) < \delta$ we have $\int_{\mathcal{A}} g(x) < \frac{\epsilon}{2}$. Then, we have

$$\int_{\mathcal{A}} f_i(x) \leq 2\int_{\mathcal{A}} g(x) < \epsilon \tag{A34}$$

for all $i$, meaning that the integrals $\int_x f_i(x)$ are uniformly absolutely continuous. Moreover, because $\int_x g(x) < \infty$, defining $\mathcal{I}_a = [-a, a]$, we have $\lim_{a\to\infty} \int_{\mathcal{I}_a} g(x) = \int_x g(x)$ from the Dominated Convergence theorem [52]. Let $\bar{\mathcal{I}}_a = (-\infty, -a) \cup (a, \infty)$; then, for all $\epsilon > 0$ we have $\exists a_o$ such that for all $a \geq a_o$ we have $\int_{\bar{\mathcal{I}}_a} g(x) < \frac{\epsilon}{2}$, which provides

$$0 \leq \int_x f_i(x) - \int_{\mathcal{I}_a} f_i(x) = \int_{\bar{\mathcal{I}}_a} f_i(x) \leq 2\int_{\bar{\mathcal{I}}_a} g(x) < \epsilon, \tag{A35}$$

for all $i$, making the integrals uniformly absolutely convergent. Then, per Lemma 4 we have $\int_x f_i(x) \to 0$, which contradicts Equation (A28). □

**Proof of Lemma 10.** For any $\epsilon > 0$, per Lemma 9 we have $\exists \delta > 0$, *which does not depend on* $\gamma$, $\lambda$, *or* $\boldsymbol{q}$, such that whenever $\|\boldsymbol{\alpha} - \boldsymbol{\beta}\|_2 < \delta$ we have

$$\begin{aligned}\int_x |p_0(x|\boldsymbol{\alpha})^{1-\lambda} p_1(x|\boldsymbol{\alpha})^{\lambda} 2^{-D(\boldsymbol{\alpha}||\boldsymbol{p})} - p_0(x|\boldsymbol{\beta})^{1-\lambda} p_1(x|\boldsymbol{\beta})^{\lambda} 2^{-D(\boldsymbol{\beta}||\boldsymbol{p})}| \\ < \epsilon \inf_{\gamma} \min_{\lambda\in[0,1]} \min_{\boldsymbol{q}\in\mathcal{P}^m} 2^{-D(\boldsymbol{q}||\boldsymbol{p})} \int_x p_0(x|\boldsymbol{q})^{1-\lambda} p_1(x|\boldsymbol{q})^{\lambda},\end{aligned} \tag{A36}$$

where the right side is strictly positive due to Lemma 5. This leads us to

$$\left| \frac{\int_x p_0(x|\boldsymbol{\alpha})^{1-\lambda} p_1(x|\boldsymbol{\alpha})^{\lambda} 2^{-D(\boldsymbol{\alpha}||\boldsymbol{p})}}{\int_x p_0(x|\boldsymbol{\beta})^{1-\lambda} p_1(x|\boldsymbol{\beta})^{\lambda} 2^{-D(\boldsymbol{\beta}||\boldsymbol{p})}} - 1 \right| \tag{A37}$$

$$= \frac{1}{\int_x p_0(x|\boldsymbol{\beta})^{1-\lambda} p_1(x|\boldsymbol{\beta})^{\lambda} 2^{-D(\boldsymbol{\beta}||\boldsymbol{p})}} \left| \int_x p_0(x|\boldsymbol{\alpha})^{1-\lambda} p_1(x|\boldsymbol{\alpha})^{\lambda} 2^{-D(\boldsymbol{\alpha}||\boldsymbol{p})} - p_0(x|\boldsymbol{\beta})^{1-\lambda} p_1(x|\boldsymbol{\beta})^{\lambda} 2^{-D(\boldsymbol{\beta}||\boldsymbol{p})} \right| \tag{A38}$$

$$\leq \frac{1}{\int_x p_0(x|\boldsymbol{\beta})^{1-\lambda} p_1(x|\boldsymbol{\beta})^{\lambda} 2^{-D(\boldsymbol{\beta}||\boldsymbol{p})}} \int_x |p_0(x|\boldsymbol{\alpha})^{1-\lambda} p_1(x|\boldsymbol{\alpha})^{\lambda} 2^{-D(\boldsymbol{\alpha}||\boldsymbol{p})} - p_0(x|\boldsymbol{\beta})^{1-\lambda} p_1(x|\boldsymbol{\beta})^{\lambda} 2^{-D(\boldsymbol{\beta}||\boldsymbol{p})}| \tag{A39}$$

$$\leq \frac{\epsilon \inf_{\gamma} \min_{\lambda\in[0,1]} \min_{\boldsymbol{q}\in\mathcal{P}^m} 2^{-D(\boldsymbol{q}||\boldsymbol{p})} \int_x p_0(x|\boldsymbol{q})^{1-\lambda} p_1(x|\boldsymbol{q})^{\lambda}}{\int_x p_0(x|\boldsymbol{\beta})^{1-\lambda} p_1(x|\boldsymbol{\beta})^{\lambda} 2^{-D(\boldsymbol{\beta}||\boldsymbol{p})}} \tag{A40}$$

$$\leq \frac{\epsilon \inf_{\gamma} \min_{\lambda\in[0,1]} \min_{\boldsymbol{q}\in\mathcal{P}^m} 2^{-D(\boldsymbol{q}||\boldsymbol{p})} \int_x p_0(x|\boldsymbol{q})^{1-\lambda} p_1(x|\boldsymbol{q})^{\lambda}}{\inf_{\gamma} \min_{\lambda\in[0,1]} \min_{\boldsymbol{q}\in\mathcal{P}^m} 2^{-D(\boldsymbol{q}||\boldsymbol{p})} \int_x p_0(x|\boldsymbol{q})^{1-\lambda} p_1(x|\boldsymbol{q})^{\lambda}} = \epsilon. \tag{A41}$$

□

**Proof of Lemma 11.** As long as the fusion center implements the MAP rule, the error exponent is characterized by the Chernoff information [27,44]. Moreover, assuming that the fusion center knows the network type, the Chernoff information becomes

$$\min_{\lambda\in[0,1]} \frac{1}{n} \log \int_{\boldsymbol{x}} \sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x}, \boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{x}, \boldsymbol{q}^n)^{\lambda}. \tag{A42}$$

Then, for any strategy $\gamma$ and $\lambda \in [0,1]$,

$$\frac{1}{n} \log \int_{\boldsymbol{x}} \sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x}, \boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{x}, \boldsymbol{q}^n)^{\lambda} \tag{A43}$$

$$= \frac{1}{n} \log \int_{\boldsymbol{x}} \sum_{\boldsymbol{q}^n} \left(p_0(\boldsymbol{x}|\boldsymbol{q}^n) \frac{p(H=0|\boldsymbol{q}^n)}{\mathbb{P}(H=0)}\right)^{1-\lambda} \left(p_1(\boldsymbol{x}|\boldsymbol{q}^n) \frac{p(H=1|\boldsymbol{q}^n)}{\mathbb{P}(H=1)}\right)^{\lambda} p(\boldsymbol{q}^n). \tag{A44}$$

Now, we have

$$\frac{1}{n} \log \int_{\boldsymbol{x}} \sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x}, \boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{x}, \boldsymbol{q}^n)^{\lambda} \tag{A45}$$

$$\leq -\frac{1}{n} \log \mathbb{P}(H=0) - \frac{1}{n} \log \mathbb{P}(H=1) + \frac{1}{n} \log \int_{\boldsymbol{x}} \sum_{\boldsymbol{q}^n} p_0(\boldsymbol{u}|\boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{u}|\boldsymbol{q}^z)^{\lambda} p(\boldsymbol{q}^n). \tag{A46}$$

Because this argument holds for any $\gamma$ and $\lambda \in [0,1]$, we have

$$\inf_{\gamma} \min_{\lambda\in[0,1]} \frac{1}{n} \log \int_{\boldsymbol{x}} \sum_{\boldsymbol{q}^n} p_0(\boldsymbol{x}, \boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{x}, \boldsymbol{q}^n)^{\lambda} - \inf_{\gamma} \min_{\lambda\in[0,1]} \frac{1}{n} \log \int_{\boldsymbol{x}} \sum_{\boldsymbol{q}^n} p_0(\boldsymbol{u}|\boldsymbol{q}^n)^{1-\lambda} p_1(\boldsymbol{u}|\boldsymbol{q}^z)^{\lambda} p(\boldsymbol{q}^n) \tag{A47}$$

$$\leq -\frac{1}{n} \log \mathbb{P}(H=0) - \frac{1}{n} \log \mathbb{P}(H=1) \to 0. \tag{A48}$$

Turning our attention to the lower bound,

$$\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n}\big(p_0(\boldsymbol{x}|\boldsymbol{q}^n)\frac{p(H=0|\boldsymbol{q}^n)}{\mathbb{P}(H=0)}\big)^{1-\lambda}\big(p_1(\boldsymbol{x}|\boldsymbol{q}^n)\frac{p(H=1|\boldsymbol{q}^n)}{\mathbb{P}(H=1)}\big)^{\lambda}p(\boldsymbol{q}^n) \tag{A49}$$

$$\geq\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n}\big(p_0(\boldsymbol{x}|\boldsymbol{q}^n)p(H=0|\boldsymbol{q}^n)\big)^{1-\lambda}\big(p_1(\boldsymbol{x}|\boldsymbol{q}^n)p(H=1|\boldsymbol{q}^n)\big)^{\lambda}p(\boldsymbol{q}^n) \tag{A50}$$

$$\geq\frac{1}{n}\log\min_{\boldsymbol{q}^n}\{p(H=0|\boldsymbol{q}^n),p(H=1|\boldsymbol{q}^n)\}\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n}p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}p(\boldsymbol{q}^n) \tag{A51}$$

$$=\frac{1}{n}\log\min_{\boldsymbol{q}^n}\{p(H=0|\boldsymbol{q}^n),p(H=1|\boldsymbol{q}^n)\}+\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n}p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}p(\boldsymbol{q}^n) \tag{A52}$$

$$\geq\frac{1}{n}\log\min_{\boldsymbol{q}^n}\{p(H=0|\boldsymbol{q}^n),p(H=1|\boldsymbol{q}^n)\}+\inf_{\gamma}\min_{\lambda\in[0,1]}\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n}p_0(\boldsymbol{x}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x}|\boldsymbol{q}^n)^{\lambda}p(\boldsymbol{q}^n). \tag{A53}$$

Because this argument holds for any $\gamma$ and $\lambda\in[0,1]$, we have

$$\inf_{\gamma}\min_{\lambda\in[0,1]}\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n}p_0(\boldsymbol{x},\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{x},\boldsymbol{q}^n)^{\lambda}-\inf_{\gamma}\min_{\lambda\in[0,1]}\frac{1}{n}\log\int_{\boldsymbol{x}}\sum_{\boldsymbol{q}^n}p_0(\boldsymbol{u}|\boldsymbol{q}^n)^{1-\lambda}p_1(\boldsymbol{u}|\boldsymbol{q}^z)^{\lambda}p(\boldsymbol{q}^n) \tag{A54}$$

$$\geq\frac{1}{n}\log\min_{\boldsymbol{q}^n}\{p(H=0|\boldsymbol{q}^n),p(H=1|\boldsymbol{q}^n)\}\to 0, \tag{A55}$$

per Assumption 5.b. This completes the proof. □

*Appendix A.2. Extension of Theorem 1*

We begin by defining

$$(\boldsymbol{q}_1^*,\ldots;\boldsymbol{q}_{n_c}^*)=\arg\max_{\boldsymbol{q}_1,\cdots;\boldsymbol{q}_{n_c}}\frac{1}{n}\sum_{c=1}^{n_c}-c_kD(\boldsymbol{q}_c||\boldsymbol{p}^c)+\log\int_{\boldsymbol{x}}p_0(\boldsymbol{x}^c|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{1-\lambda}p_1(\boldsymbol{x}^c|\boldsymbol{q}_1,\ldots;\boldsymbol{q}_{n_c})^{\lambda}. \tag{A56}$$

Then, for any $0<\epsilon<1$, in order for the argument in the proof of Theorem 1 to hold, there must exist an $n_{c,o}$ for each class such that the set

$$\mathcal{T}_{c,\delta}^n=\{\boldsymbol{q}_c^n\in\mathcal{Z}^n:\|\boldsymbol{q}_c^n-\boldsymbol{q}_c^*\|_2<\delta\} \tag{A57}$$

is non-empty for all $n\geq n_{c,o}$. Recall that we assume $\lim_{n\to\infty}\frac{c_k}{n}>0$ for all classes, meaning that $n_{c,o}$ are guaranteed to exist. This means that $\cap_c\mathcal{T}_{c,\delta}^n$ is non-empty for all $n\geq\max_c\{n_{c,o}\}$. Then, repeating the exact same argument as before for each class, we can show that there exists $n_\delta$ such that for all $n\geq n_\delta$ we have

$$\begin{aligned}1-\epsilon\leq(1-1+\sqrt{1-\epsilon})\prod_{c=1}^{n_c}(c_k+1)^{\frac{-m}{n}}\leq\\ \left[\sum_{\boldsymbol{q}_1^n,\cdots;\boldsymbol{q}_{n_c}^n}\prod_{c=1}^{n_c}\frac{\int_{\boldsymbol{x}^c}p_0(\boldsymbol{x}^c|\boldsymbol{q}_1^n,\ldots;\boldsymbol{q}_{n_c}^n)^{1-\lambda}p_1(\boldsymbol{x}^c|\boldsymbol{q}_1^n,\ldots;\boldsymbol{q}_{n_c}^n)^{\lambda}p(\boldsymbol{q}_c^n)}{\int_{\boldsymbol{x}^c}p_0(\boldsymbol{x}^c|\boldsymbol{q}^*)^{1-\lambda}p_1(\boldsymbol{x}^c|\boldsymbol{q}^*)^{\lambda}2^{-nD(\boldsymbol{q}^*||\boldsymbol{p})}}\right]^{\frac{1}{n}}\leq\prod_{c=1}^{n_c}(c_k+1)^{\frac{m}{n}}\leq 1+\epsilon.\end{aligned} \tag{A58}$$

## References


1. Shanthamallu, U.S.; Spanias, A.; Tepedelenlioglu, C.; Stanley, M. A brief survey of machine learning methods and their sensor and IoT applications. In Proceedings of the 2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA), Larnaca, Cyprus, 27–30 August 2017; pp. 1–8.
2. Tajer, A.; Kar, S.; Poor, H.V.; Cui, S. Distributed joint cyber attack detection and state recovery in smart grids. In Proceedings of the 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, Belgium, 17–20 October 2011; pp. 202–207.
3. Patel, A.; Ram, H.; Jagannatham, A.K.; Varshney, P.K. Robust cooperative spectrum sensing for MIMO cognitive radio networks under csi uncertainty. *IEEE Trans. Signal Process.* **2018**, *66*, 18–33.

4. Chawla, A.; Singh, R.K.; Patel, A.; Jagannatham, A.K.; Hanzo, L. Distributed detection for centralized and decentralized millimeter wave massive MIMO sensor networks. *IEEE Trans. Veh. Technol.* **2021**, *70*, 7665–7680.
5. Geng, B.; Cheng, X.; Brahma, S.; Kellen, D.; Varshney, P.K. Collaborative human decision making with heterogeneous agents. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 469–479.
6. Gupta, S.S.; Mehta, N.B. Ordered transmissions schemes for detection in spatially correlated wireless sensor networks. *IEEE Trans. Commun.* **2021**, *69*, 1565–1577.
7. Gangan, M.S.; Vasconcelos, M.M.; Mitra, U.; Câmara, O.; Boedicker, J.Q. Intertemporal trade-off between population growth rate and carrying capacity during public good production. *iScience* **2022**, *25*, 104117.
8. Tsitsiklis, J.; Athans, M. On the complexity of decentralized decision making and detection problems. *IEEE Trans. Autom. Control* **1985**, *30*, 440–446.
9. Aalo, V.; Viswanathou, R. On distributed detection with correlated sensors: Two examples. *IEEE Trans. Aerosp. Electron. Syst.* **1989**, *25*, 414–421.
10. Willett, P.; Swaszek, P.F.; Blum, R.S. The good, bad and ugly: Distributed detection of a known signal in dependent Gaussian noise. *IEEE Trans. Signal Process.* **2000**, *48*, 3266–3279.
11. Gül, G. Minimax robust decentralized hypothesis testing for parallel sensor networks. *IEEE Trans. Inf. Theory* **2020**, *67*, 538–548.
12. Chen, H.; Chen, B.; Varshney, P.K. A new framework for distributed detection with conditionally dependent observations. *IEEE Trans. Signal Process.* **2012**, *60*, 1409–1419.
13. Hanna, O.A.; Li, X.; Fragouli, C.; Diggavi, S. Can we break the dependency in distributed detection? In Proceedings of the 2022 IEEE International Symposium on Information Theory (ISIT), Espoo, Finland, 26 June–1 July 2022; pp. 2720–2725.
14. Kasasbeh, H.; Cao, L.; Viswanathan, R. Soft-decision-based distributed detection with correlated sensing channels. *IEEE Trans. Aerosp. Electron. Syst.* **2019**, *55*, 1435–1449.
15. Maleki, N.; Vosoughi, A. On bandwidth constrained distributed detection of a known signal in correlated Gaussian noise. *IEEE Trans. Veh. Technol.* **2020**, *69*, 11428–114440.
16. Shaska, J.; Mitra, U. State-dependent decentralized detection. *IEEE Trans. Inf. Theory* 2023, *submitted*.
17. Csiszar, I. The method of types [information theory]. *IEEE Trans. Inf. Theory* **1998**, *44*, 2505–2523.
18. Raginsky, M. Empirical processes, typical sequences, and coordinated actions in standard borel spaces. *IEEE Trans. Inf. Theory* **2013**, *59*, 1288–1301.
19. Schuurmans, M.; Patrinos, P. A general framework for learning-based distributionally robust mpc of markov jump systems. *IEEE Trans. Autom. Control* **2023**, *68*, 2950–2965.
20. Haghifam, M.; Tan, V.Y.; Khisti, A. Sequential classification with empirically observed statistics. *IEEE Trans. Inf. Theory* **2021**, *67*, 3095–3113.
21. Guo, F.R.; Richardson, T.S. Chernoff-type concentration of empirical probabilities in relative entropy. *IEEE Trans. Inf. Theory* **2021**, *67*, 549–558.
22. Weinberger, N.; Merhav, N. The dna storage channel: Capacity and error probability bounds. *IEEE Trans. Inf. Theory* **2022**, *68*, 5657–5700.
23. Lalitha, A.; Javidi, T.; Sarwate, A.D. Social learning and distributed hypothesis testing. *IEEE Trans. Inf. Theory* **2018**, *64*, 6161–6179.
24. Inan, Y.; Kayaalp, M.; Telatar, E.; Sayed, A.H. Social learning under randomized collaborations. In Proceedings of the 2022 IEEE International Symposium on Information Theory (ISIT), Espoo, Finland, 26 June–1 July 2022; pp. 115–120.
25. Goetz, C.; Humm, B. Decentralized real-time anomaly detection in cyber-physical production systems under industry constraints. *Sensors* **2023**, *23*, 4207.
26. Shaska, J.; Mitra, U. Decentralized decision making in multi-agent networks: The state-dependent case. In Proceedings of the 2021 IEEE Global Communcations Conference, Madrid, Spain, 7–11 December 2021.
27. Tsitsiklis, J.N. Decentralized detection by a large number of sensors. *Math. Control. Signals Syst.* **1988**, *1*, 167–182.
28. Chen, B.; Tong, L.; Varshney, P.K. Channel-aware distributed detection in wireless sensor networks. *IEEE Signal Process. Mag.* **2006**, *23*, 16–26.
29. Duman, T.; Salehi, M. Decentralized detection over multiple-access channels. *IEEE Trans. Aerosp. Electron. Syst.* **1998**, *34*, 469–476.
30. Liu, B.; Chen, B. Channel-optimized quantizers for decentralized detection in sensor networks. *IEEE Trans. Inf. Theory* **2006**, *52*, 3349–3358.
31. Gelfand, S.I.; Pinsker, M.S. Coding for channel with random parameters. *Probl. Control Inf. Theory* **1980**, *9*, 19–31.
32. Choudhuri, C.; Kim, Y.H.; Mitra, U. Causal state communication. *IEEE Trans. Inf. Theory* **2013**, *59*, 3709–3719.
33. Miretti, L.; Kobayashi, M.; Gesbert, D.; De Kerret, P. Cooperative multiple-access channels with distributed state information. *IEEE Trans. Inf. Theory* **2021**, *67*, 5185–5199.
34. Zhang, W.; Vedantam, S.; Mitra, U. Joint transmission and state estimation: A constrained channel coding approach. *IEEE Trans. Inf. Theory* **2011**, *57*, 7084–7095.
35. Kobayashi, M.; Caire, G.; Kramer, G. Joint state sensing and communication: Optimal tradeoff for a memoryless case. In Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 111–115.
36. Bross, S.I.; Lapidoth, A. The rate-and-state capacity with feedback. *IEEE Trans. Inf. Theory* **2018**, *64*, 1893–1918.
37. Moon, J.; Park, J. Pattern-dependent noise prediction in signal-dependent noise. *IEEE J. Sel. Areas Commun.* **2001**, *19*, 730–743.

38. Tsiatmas, A.; Baggen, C.P.; Willems, F.M.; Linnartz, J.P.M.; Bergmans, J.W. An illumination perspective on visible light communications. *IEEE Commun. Mag.* **2014**, *52*, 64–71.
39. Kavcic, A.; Moura, J.M. Correlation-sensitive adaptive sequence detection. *IEEE Trans. Magn.* **1998**, *34*, 763–771.
40. Hareedy, A.; Amiri, B.; Galbraith, R.; Dolecek, L. Non-binary ldpc codes for magnetic recording channels: Error floor analysis and optimized code design. *IEEE Trans. Commun.* **2016**, *64*, 3194–3207.
41. Kuan, D.T.; Sawchuk, A.A.; Strand, T.C.; Chavel, P. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Trans. Pattern Anal. Mach. Intell.* **1985**, *2*, 165–177.
42. JMeola, J.; Eismann, M.T.; Moses, R.L.; Ash, J.N. Modeling and estimation of signal-dependent noise in hyperspectral imagery. *Appl. Opt.* **2011**, *50*, 3829–3846.
43. Michelusi, N.; Boedicker, J.; El-Naggar, M.Y.; Mitra, U. Queuing models for abstracting interactions in bacterial communities. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 584–599.
44. Shannon, C.E.; Gallager, R.G.; Berlekamp, E.R. Lower bounds to error probability for coding on discrete memoryless channels. I. *Inf. Control* **1967**, *10*, 65–103.
45. Chamberland, J.-F.; Veeravalli, V. Asymptotic results for decentralized detection in power constrained wireless sensor networks. *IEEE J. Sel. Areas Commun.* **2004**, *22*, 1007–1015.
46. Yin, D.; Chen, Y.; Kannan, R.; Bartlett, P. Byzantine-robust distributed learning: Towards optimal statistical rates. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 5650–5659.
47. Li, S.; Zhao, S.; Yang, P.; Andriotis, P.; Xu, L.; Sun, Q. Distributed consensus algorithm for events detection in cyber-physical systems. *IEEE Internet Things J.* **2019**, *6*, 2299–2308.
48. Nielson, F. Revisiting Chernoff information with likelihood ratio exponential families. *Entropy* **2022**, *24*, 1400.
49. Graves, L.M. *The Theory of Functions of Real Variables*; Courier Corporation: North Chelmsford, MA, USA, 2012.
50. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley-Interscience: Hoboken, NJ, USA, 2006.
51. Williams, D. *Probability with Martingales*; Cambridge University Press: Cambridge, MA, USA, 1991.
52. Bartle, R.G.; Sherbert, D.R. *Introduction to Real Analysis*; Wiley: New York, NY, USA, 2000; Volume 2.