2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Multi-Agent Fact-Checker: Adaptive Estimators

Ashwin Verma, Soheil Mohajer, and Behrouz Touri

***Abstract*—We consider a multi-agent dynamics for distributed fact-checking that validates the truth of a statement based on the labels of an ensemble of inexpert agents. Each agent in the system is modeled as a Binary Symmetric Channel (BSC) that incorrectly judges the veracity of each true/false statement with some probability $\pi_i \in (0,1)$ which we refer to as the unreliability parameter of the agent. We introduce a class of adaptive estimators for the unreliability parameters of the agents. For the class of estimators, we provide the necessary conditions for the adaptive estimator to converge to the true unreliability parameters. We show that the estimators for ensembles of two and three agents eventually adhere to a consistent (fixed) update rule. Furthermore, we also show that, surprisingly, the estimator for the unreliability parameters based on the hard-decoded estimate of the statement truths fails to converge to the true unreliability parameters for any number of agents.**

I. Introduction

As online social networks become increasingly effective in disseminating information, the task of distinguishing between true and false information becomes increasingly challenging. This growing efficiency of information dissemination has led to a number of studies on how misinformation spreads through networks [1]–[3], [15], [17]. Conversely, there is growing interest in the development of automated fact-checkers that can perform tasks such as document retrieval, evidence extraction, and claim validation in an automated manner [11], [12], [19].

When there are multiple imperfect *fact checkers*, determining the validity of a source based on their responses becomes a challenge. In such cases, it is important to know the reliability statistics of the fact checkers in question. As a result, a natural question arises: in the presence of multiple imperfect fact checkers, how can we formulate and learn their reliability over time? We provide a model for distributed fact-checking using unreliable or imperfect agents. A key step in our model is to model each imperfect agent as a BSC channel. Given an estimate of the unreliability parameters, a weighted thresholding estimator can be used to identify the validity of the statement [16], [18], [21], where the weights are the log-odds based on the agents' unreliability estimates. We focus on a class of learning rules to estimate the agents' reliability parameters based on a set of desirable properties. Our algorithm has the advantage of requiring minimal memory and having a simplified update rule.

In our problem, we are working with a mixture of product distributions. Determining the parameters of an identifiable mixture has been widely researched [4], [6], [7], [9], [10]. The parameter estimation problem typically involves finding a hypothetical model that produces samples with a distribution that closely resembles the true model. The problem of distributed fact-checking parallels the problem of crowdsourcing labelling tasks popularly studied in the framework of Dawid-Skene model, introduced through empirical studies in [5]. The convergence of Dawid-Skene estimator, which is based on the Expectation-Maximization algorithm, for the offline scenario, that is when the sequence of statements to be verified are available as a batch, has been studied in [8], [22].

This work is a continuation of our earlier efforts on laying framework and studying distributed fake-news detection [20], [21]. In [21], we presented a framework for the multi-agent fact-checker system and identified the class of optimal linear-thresholding estimators for assessing the validity of statements *assuming the full knowledge of agents unreliability*. To learn the agents' unreliability parameters, in [20], we introduced a specific online estimator (which in this paper is referred to as ALL estimator) for the estimation of the unreliability parameters of the agents and studied the convergence property of it for two-agent fact-checker system.

The main contributions of this paper involve (i) moving beyond ALL estimator and proposing a generalized class of online estimators for the agents' unreliability parameters. The estimators are associated with a function, of the agents' opinions and unreliability estimate. The associated function can be interpreted as an estimate for the validity of the statements. We also propose a set of axioms that a natural estimator should satisfy and hence, we call the class of functions satisfying the desired properties as the *natural functions*; (ii) characterizing natural functions for two and three-agent fact-checker systems; (iii) proving that a ALL estimator belongs to the class of natural functions for any n-agent fact-checker system and the hard-estimator does not belong to this class for any $n \geq 2$.

Notation: Let $\mathbb{N}$ denote the set of all natural numbers, $\mathbb{N}_0$ denote $\mathbb{N} \cup \{0\}$, and for any $n \in \mathbb{N}$, define $[n] := \{1, 2, \ldots, n\}$. We denote the set of real numbers by $\mathbb{R}$ and the set of all real-valued n-dimensional vectors by $\mathbb{R}^n$. We use bold letters, such as $\boldsymbol{x}, \boldsymbol{s}$, to denote vectors, and regular letters, such as x, s, to denote scalars in $\mathbb{R}$. We use $\mathbf{1}$ to denote the all ones vector. For a scalar $a \in [0,1]$, we use $\bar{a}$ to denote $1-a$. Throughout this work, all random variables are defined with respect to an underlying probability space $(\Omega, \mathcal{F}, \Pr)$.

II. Problem Formulation

Here, we describe the model for our fact-checker setup as set forth in our earlier works [20], [21] and the problem of interest in this study. Consider a source that streams a

This research is supported by AFOSR under the grant FA9550-23-1-0057.
A. Verma and B. Touri are with the ECE Department of University of California San Diego (email: {a1verma, btouri}@ucsd.edu), and S. Mohajer is with the ECE Department of the University of Minnesota (email: soheil@umn.edu).

sequence of statements. Each statement can be true or false. We use a *hidden* variable $S(t) \in \{+1, -1\}$ to denote the label (true/false) of the statement of discrete-time instance $t \in \mathbb{N}_0$. A *fact-checker* is interested in evaluating the validity of the statements using imperfect (inexpert) agents. We assume that the stream symbols are independently and identically distributed according to the Rademacher distribution, i.e., $\Pr(S(t) = +1) = \Pr(S(t) = -1) = \frac{1}{2}$, for every $t \in \mathbb{N}$.

Model for the fact-checker: A *fact-checker* is an overseer of multiple agents, where each agent is responsible for testing the validity of the statements provided to it. For $n \in \mathbb{N}$, let $[n]$ be the set of agents verifying the validity of the statements. At each time $t \in \mathbb{N}$, the agents observe the shared statement $S(t)$ and output their evaluation regarding the validity of the statement to the fact-checker, by returning their assessment about the statement. In other words, if the agent considers the statement correct it marks the statement as True, otherwise, it marks it as False. However, due to their limited expertise, the agents' assessments may be different from the actual label of the statements. Mathematically, we model agent $i \in [n]$ as a memoryless Binary Symmetric Channel (BSC) with the *error probability* or *crossover probability* $\pi_i \in (0, 1)$, that takes the input $S(t)$ and outputs $R_i(t)$, where for every $s \in \{+1, -1\}$, the distribution of the output is given as

$$\Pr(R_i(t) = -s \,|\, S(t) = s) = \pi_i = 1 - \Pr(R_i(t) = s \,|\, S(t) = s).$$

Therefore, agent $i \in [n]$ observes the input $S(t)$ at time t and outputs the assessment $R_i(t)$, which is independent of the past. Here, π_i represents the unreliability of agent i since the agent misclassifies the statement with probability π_i. We represent the collection of crossover probabilities by $\boldsymbol{\pi}$ and the sequence of all agents' outputs at time t by $\boldsymbol{R}(t)$.

Properties of Output distribution: Let us discuss some properties of the output distribution.

(1) Because of independency of the statement stream $\{S(t)\}$, and since each agent is viewed as a memoryless channel, the random vector process $\{\boldsymbol{R}(t)\}$ is an independent process.

(2) At any time $t \in \mathbb{N}$, given $S(t)$, the outputs $\{R_i(t)\}_{i=1}^n$ are independent of each other. Moreover, for any $t \in \mathbb{N}$ and for every $i \in [n]$, $R_i(t)$ has the Rademacher distribution.

(3) The joint distribution of the output $\boldsymbol{R}(t)$ is given as

$$\Pr(\boldsymbol{R}(t) = \boldsymbol{r}) = \frac{1}{2}\left(\prod_{i=1}^{n} \pi_i^{\frac{1+r_i}{2}} \bar{\pi}_i^{\frac{1-r_i}{2}} + \prod_{i=1}^{n} \pi_i^{\frac{1-r_i}{2}} \bar{\pi}_i^{\frac{1+r_i}{2}}\right),$$

where $\boldsymbol{r} \in \{+1, -1\}^n$, and $\bar{x} := 1 - x$.

For brevity, for a given unreliability parameters (vector) of the agents $\boldsymbol{x} \in (0, 1)^n$, we define $g_{\boldsymbol{x}} : \{+1, -1\}^n \to (0, 1)$ to be the distribution of the output vector $\boldsymbol{r} \in \{-1, +1\}^n$, i.e., $g_{\boldsymbol{x}}(\boldsymbol{r}) = \Pr(\boldsymbol{R} = \boldsymbol{r}; \boldsymbol{x})$. One goal is to obtain a reliable estimate of the validity of each statement $S(t)$ as a function of $\boldsymbol{R}(t)$ as well as build an online estimate for the unreliability parameters $\boldsymbol{\pi}$ as a function of observations $\boldsymbol{R}(0), \ldots, \boldsymbol{R}(t)$. In next section, we will introduce a class of online estimators with properties that are either necessary or desirable for the adaptive estimation of $\boldsymbol{\pi}$.

III. Natural Estimators

First, let us discuss an online estimator for the unreliability parameters of the agents comprising the fact-checker for any number of agents $n \geq 2$ as introduced in [20]. We have provided convergence guarantees for this algorithm for $n = 2$ agents in [20].

Consider the stream of output observed by the fact-checker, namely, $\{\boldsymbol{R}(t)\}$. At any time $t \in \mathbb{N}_0$, based on earlier output vectors, the fact-checker has an estimate $\boldsymbol{P}(t)$ of the true unreliability parameters $\boldsymbol{\pi}$ and it updates this estimate after observing the new output $\boldsymbol{R}(t+1)$.

Note that if the vector $\boldsymbol{\pi}$ was known, the fact-checker could evaluate the likelihood ratio of source being -1 or $+1$ observing $\boldsymbol{R}(t+1)$ as

$$L^*(t) = \frac{\Pr(\boldsymbol{R}(t+1) | S(t+1) = -1)}{\Pr(\boldsymbol{R}(t+1) | S(t+1) = +1)} = \prod_{i=1}^{n} \left(\frac{\pi_i}{1 - \pi_i}\right)^{R_i(t+1)}$$

to decode $S(t+1)$ by simply announcing

$$\hat{S}(t+1) = 2\mathbb{1}_{\{L^*(t) < 1\}} - 1,$$

where $\mathbb{1}_{\{.\}}$ represents the indicator function. Without the full knowledge of $\boldsymbol{\pi}$ however, a natural approach would be to use its *estimate* $\boldsymbol{P}(t)$, to compute an *approximate* likelihood ratio $L(t)$ of $S(t+1) = -1$ to $S(t+1) = +1$ based on $\boldsymbol{R}(t+1)$. For the received vector $\boldsymbol{R}(t+1)$ and an estimate $\boldsymbol{P}(t)$ of agents' unreliability parameters, we define the likelihood ratio estimate $L(t)$ by

$$L(t) := \prod_{i=1}^{n} \left(\frac{P_i(t)}{1 - P_i(t)}\right)^{R_i(t+1)}. \quad (1)$$

Using $L(t)$, we can estimate $S(t+1)$ by setting

$$\hat{S}(t+1) := 2\mathbb{1}_{\{L(t) < 1\}} - 1 = \begin{cases} -1 & \text{if } L(t) \geq 1 \\ +1 & \text{if } L(t) < 1 \end{cases}. \quad (2)$$

We are ready to discuss the update rule for the unreliability parameters' estimates, given the source symbol estimate $\hat{S}(t+1)$ and the output vector $\boldsymbol{R}(t+1)$. The proposed algorithm/dynamics updates the unreliability parameters as

$$P_i(t+1) = (1 - \eta_t) P_i(t) + \frac{1}{2}\eta_t \left(\frac{L(t) - 1}{L(t) + 1} R_i(t+1) + 1\right), \quad (3)$$

for all $t \in \mathbb{N}_0$ and $i \in [n]$, with some initial condition (guess) $\boldsymbol{P}(0) \in (0, 1)^n$, where $\{\eta_t\}$ is a pre-decided step-size sequence, and $L(t)$ is given in (1).

One popular choice for the step-size sequence is the harmonic sequence $\eta_t = \frac{1}{t+1}$ for all $t \in \mathbb{N}_0$. To grasp the motivation behind the estimator using such a step-size sequence, examine the scenario when the fact-checker knows the source sequence symbols $\{S(t)\}$.

Since, at any time $t \in \mathbb{N}$, the output distribution of the agents given $S(t)$ is independent of each other, the problem of estimating the unreliability parameters of the agents is equivalent to n uncoupled problems of estimating the parameter of Bernoulli distribution from its samples. Estimation of parameters for a Bernoulli distribution from its sample is a well-studied problem and a class of estimators effective to solve it is the *add-constant* estimator [13]. For the current setting, for any $i \in [n]$, the add-β estimator, where $\beta \geq 0$ for parameter π_i at time $t \in \mathbb{N}$ is given by

$$Q_i(t) = \frac{\beta + \sum_{k=1}^{t} \mathbb{1}_{\{R_i(k) \neq S(k)\}}}{t + 2\beta}. \quad (4)$$

The estimator makes use of the empirical frequency of agent i misclassifying the source symbol received and can be expressed recursively as

$$Q_i(t+1) = (1 - \gamma_t) Q_i(t) + \gamma_t \mathbb{1}_{\{R_i(t+1) \neq S(t+1)\}}.$$

Here, $\gamma_t := \frac{1}{t+1+2\beta}$ and $Q_i(0) = 1/2$. The convergence properties of estimator $Q(t)$ for different values of β and various loss functions are studied in [13]. Different values of β lead to well-known estimators, including the empirical estimator ($\beta = 0$), the Krichevsky–Trofimov (KT) estimator ($\beta = \frac{1}{2}$), and Laplace estimator ($\beta = 1$).

To see the connection to our setting, where the source symbol is unknown, consider an extreme case where $L(t) \gg 1$ (which suggests $\hat{S}(t+1) = -1$ high confidence). For $R_i(t+1) = +1$, we get

$$\frac{1}{2}\left(\frac{L(t)-1}{L(t)+1} R_i(t+1) + 1\right) = \frac{L(t)}{L(t)+1} \approx 1,$$

whereas for $R_i(t+1) = -1$ we have $\frac{1}{L(t)+1} \approx 0$. Thus,

$$\frac{1}{2}\left(\frac{L(t)-1}{L(t)+1} R_i(t+1) + 1\right) \approx \mathbb{1}_{\{R_i(t+1) \neq \hat{S}(t+1)\}}.$$

A similar situation holds when $L(t) \approx 0$. Therefore, the update rule (3) with $\eta_t = \frac{1}{t+1}$ can be viewed as an imperfect and adaptive version of the add-β estimator (with $\beta = 0$).

Note that a central idea in describing the estimator for the unreliability parameter is to implicitly or explicitly define an estimator for the validity of the statements. Let us define a class of estimators based on functions $\mathcal{B} : \{-1,+1\}^n \times (0,1)^n \to [-1,1]$. The function $\mathcal{B}(\cdot;\cdot)$ represents a soft estimate of the statement truth S. Using the function $\mathcal{B}(\cdot;\cdot)$ we can define the adaptive estimator for $\boldsymbol{\pi}$ as

$$\boldsymbol{P}(t+1) = (1-\eta_t)\boldsymbol{P}(t) + \eta_t \frac{1}{2}\Big(\mathbf{1} - \mathcal{B}(\boldsymbol{R}(t+1); \boldsymbol{P}(t))\boldsymbol{R}(t+1)\Big). \quad (5)$$

Let us look at three examples of the function $\mathcal{B}(\cdot;\cdot)$ representing different potential estimators.

(1) **Approximate Log-Likelihood (ALL) Estimator:** If we use the approximate log-likelihood ratio (1) to get an estimate of the statement validity we get the ALL estimator resulting in the online estimator defined through (3). It can be shown that the $\mathcal{B}$-function for this estimator is given as

$$\mathcal{B}_{\text{ALL}}(\boldsymbol{R};\boldsymbol{x}) := \tanh\left(\sum_{i=1}^{n} \frac{R_i}{2} \log \frac{1-x_i}{x_i}\right) = \frac{1-L(\boldsymbol{R};\boldsymbol{x})}{1+L(\boldsymbol{R};\boldsymbol{x})}, \quad (6)$$

where $L(\boldsymbol{R};\boldsymbol{x}) := \prod_{i=1}^{n} \left(\frac{x_i}{1-x_i}\right)^{R_i}$ is the approximation of the likelihood ratio.

(2) **Hard-Thresholding (HT) Estimator:** Instead of using the approximate likelihood ratio for statement validity we can use the hard estimator which uses the hard estimate of statement validity to compute the empirical frequency of misclassifying the source symbol. In other words, the HT estimator compares the output of each agent with the estimated value for $S(t+1)$ given in (2), if the two values agree, HT Estimator decreases the agent's unreliability parameter down, otherwise, the unreliability parameter will be increased. The $\mathcal{B}$-function for the HT estimator can be expressed as

$$\mathcal{B}_{\text{HT}}(\boldsymbol{R};\boldsymbol{x}) = \text{sign}\left(\sum_{i=1}^{n} R_i \log \frac{1-x_i}{x_i}\right), \quad (7)$$

where $\text{sign}(a) := -\mathbb{1}_{\{a \leq 0\}} + \mathbb{1}_{\{a \geq 0\}}$ for any $a \in \mathbb{R}$.

To corroborate the idea that $\mathcal{B}(\cdot;\cdot)$ is an estimate of the statement validity S, let us introduce the **Oracle Estimator**

$$\mathcal{B}_{\text{oracle}}(\boldsymbol{R}, S; \boldsymbol{x}) = S\mathbf{1}. \quad (8)$$

Using the $\mathcal{B}_{\text{oracle}}$-function results in the add-β estimator defined through (4). Note that the function $\mathcal{B}_{\text{oracle}}$ does not fit in our class of functions $\mathcal{B}(\cdot;\cdot)$ of interest since it takes the truth of the statement S as an argument.

IV. Results

In this section, we state the main results of this paper. Let us start with stating the desirable properties of the function $\mathcal{B}(\cdot;\cdot)$ that must be satisfied in order to have a feasible estimator of unreliability parameters that converges to $\boldsymbol{\pi}$.

A. Natural Estimators: Axioms and Necessary Conditions

First, let us introduce some conditions/axioms that one would expect from a reasonable estimator. Later, we will discuss why such axioms are expected from such an estimator.

Definition 1. *For any $n \in \mathbb{N}$ let us define $\mathcal{C}_n^{nat}$ as the set of all functions $\mathcal{B} : \{-1,+1\}^n \times (0,1)^n \to [-1,1]$ that satisfy*

Assumption (i) *Anti-Symmetry of reliability:*

$$\mathcal{B}(\boldsymbol{R};\boldsymbol{x}) = -\mathcal{B}(\boldsymbol{R};\mathbf{1}-\boldsymbol{x}). \quad (9)$$

Assumption (ii) *Anti-Symmetry of Opinions:*

$$\mathcal{B}(-\boldsymbol{R};\boldsymbol{x}) = -\mathcal{B}(\boldsymbol{R};\boldsymbol{x}). \quad (10)$$

Assumption (iii) *Consistency of Estimators:*

$$\mathbb{E}_{\boldsymbol{R}\sim g_{\boldsymbol{x}}}[\boldsymbol{R} \cdot \mathcal{B}(\boldsymbol{R};\boldsymbol{x})] = \mathbf{1} - 2\boldsymbol{x}. \quad (11)$$

*We refer to $\mathcal{C}_n^{nat}$ as the set of **natural functions** for an n-agent fact-checker system.*

Assumption (i) ensures that the estimates of the statement validity for fact-checker systems with unreliability parameters $\boldsymbol{\pi}$ and $\mathbf{1}-\boldsymbol{\pi}$ take the same absolute value but have different signs. The assumption is justified as the output of the fact-checker system with unreliability parameter vector $\mathbf{1}-\boldsymbol{\pi}$ can be seen as the flipped output of a fact-checker system with unreliability parameter $\boldsymbol{\pi}$. Similarly given a fact-checker system *Assumption (ii)* ensures that the flipping the output of all the agents' opinion flips the sign of the estimate of the statement validity. Finally, regarding *Assumption (iii)*, note that the consistency condition (11) is equivalent to $\boldsymbol{x} = \frac{1}{2}(\mathbf{1} - \mathbb{E}_{\boldsymbol{R}\sim g_{\boldsymbol{x}}}[\boldsymbol{R}\mathcal{B}(\boldsymbol{R};\boldsymbol{x})])$, which is what is expected from the mean-field dynamics of (5), i.e., to have $\boldsymbol{x}$ as its equilibrium point, given that the agents true reliability parameter vector is $\boldsymbol{x}$.

Now we are ready to present our main results. The first result shows that interestingly the only natural estimator for two-agent fact-checker system is the ALL estimator (6).

Proposition 1 (Elements of $\mathcal{C}_2^{\text{nat}}$)**.** *For a two-agent fact-checker system the class of functions $\mathcal{B}(\cdot;\cdot)$ satisfying* Assumption (i)-Assumption (iii) *contains exclusively the function $\mathcal{B}_{ALL}(\boldsymbol{R};\boldsymbol{x})$ as defined in* (6)*,* i.e., $\mathcal{C}_2^{nat} = \{\mathcal{B}_{ALL}(\boldsymbol{R};\boldsymbol{x})\}$.

Remark 1. *In [20] we studied the ALL estimator for a two-agent fact-checker system whose unreliability parameter is $\boldsymbol{\pi}$ and we have shown that the estimates $\{\boldsymbol{P}(t)\}$ converge to the solution set $\mathcal{E}$ of the equation*

$$\mathbb{E}_{\boldsymbol{R}\sim g_{\boldsymbol{\pi}}}[\boldsymbol{R}\mathcal{B}_{ALL}(\boldsymbol{R};\boldsymbol{x})] = \mathbf{1} - 2\boldsymbol{x}. \tag{12}$$

Note the difference in (11) *and* (12) *lies in the distribution over which the expectation is taken. Since $\mathcal{B}_{ALL}(\cdot;\cdot)$ satisfies* Assumption (iii) *we know that $\boldsymbol{\pi} \in \mathcal{E}$. However the set $\mathcal{E}$ is a continuum of points $\boldsymbol{x}$ for which $g_{\boldsymbol{x}}(\boldsymbol{R}) = g_{\boldsymbol{\pi}}(\boldsymbol{R})$ for all $\boldsymbol{R} \in \{-1,+1\}^2$.*

In the following proposition, we identify the functions that satisfy the properties required by a natural estimator for a three-agent fact-checker system.

Proposition 2 (Elements of $\mathcal{C}_3^{\text{nat}}$)**.** *For a three-agent fact-checker system, the set of natural estimators $\mathcal{C}_3^{nat}$ consists of functions $\mathcal{B}(\cdot;\cdot)$ satisfying*

$$\mathcal{B}(\boldsymbol{R};\boldsymbol{x}) = \mathcal{B}_{ALL}(\boldsymbol{R};\boldsymbol{x}) + \frac{c_{\boldsymbol{x}}R_1R_2R_3}{2g_{\boldsymbol{x}}(\boldsymbol{R})}, \tag{13}$$

where $\mathcal{B}_{ALL}(\boldsymbol{R};\boldsymbol{x})$ is the ALL estimator defined in (6) *and $c_{\boldsymbol{x}}$ is any function of the vector $\boldsymbol{x}$ such that $c_{\boldsymbol{x}} = -c_{\mathbf{1}-\boldsymbol{x}}$.*

Next, we show that for $n \geq 2$ there exists $\boldsymbol{x}$ for which $\mathcal{B}_{\text{HT}}(\cdot;\cdot)$ does not satisfy *Assumption (iii)*.

Proposition 3 (Convergence for Hard-Thresholding Estimator)**.** *The function $\mathcal{B}_{HT}(\cdot;\cdot)$ as defined through* (7)*, based on the hard-thresholding estimator, does not satisfy* Assumption (iii). *In other words, $\mathcal{B}_{HT}(\cdot;\cdot) \notin \mathcal{C}_n^{nat}$ for any $n \geq 2$.*

Recall that the system of equations in *Assumption (iii)* is a necessary condition for the estimates $\{\boldsymbol{P}(t)\}$ to converge to $\boldsymbol{\pi}$. However, for a fact-checker system with unreliability parameter $\boldsymbol{\pi}$, it is also important to identify the solution set $\mathcal{E}$ to the system of equation $\mathbb{E}_{\boldsymbol{R}\sim g_{\boldsymbol{\pi}}}[\boldsymbol{R}\mathcal{B}_{\text{ALL}}(\boldsymbol{R};\boldsymbol{x})] = \mathbf{1} - 2\boldsymbol{x}$. The solution set $\mathcal{E}$ represents the points $\boldsymbol{x} \in (0,1)^n$ that could be the points of convergence for the estimates $\{\boldsymbol{P}(t)\}$. In the following theorem, we identify the set $\mathcal{E}$ for a three-agent fact-checker system to be the set containing the true estimate $\boldsymbol{\pi}$, the 'symmetric' estimate $\mathbf{1} - \boldsymbol{\pi}$ and the degenerate point $\frac{1}{2}\mathbf{1}$.

Theorem 1 (Fixed points of ALL-estimator for three-agent fact-checker)**.** *For a three-agent fact-checker system where the agents have unreliability parameters $\pi_i \in (0,1) \setminus \{\frac{1}{2}\}$ for $i \in [3]$, the set of solutions of the fixed-point equation $\boldsymbol{x} = \frac{1}{2}(\mathbf{1} - \mathbb{E}_{\boldsymbol{R}\sim g_{\boldsymbol{\pi}}}[\boldsymbol{R}\mathcal{B}_{ALL}(\boldsymbol{R};\boldsymbol{x})])$ is $\mathcal{S} := \{\boldsymbol{\pi}, \mathbf{1}-\boldsymbol{\pi}, \frac{1}{2}\mathbf{1}\}$.*

Note that the set $\mathcal{E}$ also represents the set of convergence for the Dawid-Skene estimator [5] and the Theorem 1 is the first result to identify the exact set $\mathcal{E}$. The theorem signifies that for a three-agent system, the only points the Dawid-Skene and its variants would converge to are the relevant points $\boldsymbol{\pi}$, $\mathbf{1} - \boldsymbol{\pi}$ or the degenerate point $\frac{1}{2}\mathbf{1}$.

In the following theorem, we show that for any n-agent fact-checker system the adaptive estimator associated with the ALL estimator satisfies all the desired properties.

Theorem 2. *For $n \geq 2$, $\mathcal{B}_{ALL}(\cdot;\cdot)$, as defined in* (6)*, satisfies* Assumption (i)-Assumption (iii). *In other words, $\mathcal{B}_{ALL} \in \mathcal{C}_n^{nat}$.*

V. Proof of Main Results

In this section, we present the proof of the results discussed in Section IV. First, we establish a notation to impose an ordering on the 2^n distinct possibilities of the output vector $\boldsymbol{R}$.

Definition 2 (Notation)**.** *Consider the binary representation $(b_1, b_2, \ldots, b_n)$ of $N \in \{0\} \cup [2^n - 1]$. Here b_1 represents the most significant bit and b_n the least significant bit. Define the output vector $\boldsymbol{\mathcal{R}}_N$ associated with N as*

$$\boldsymbol{\mathcal{R}}_N := \begin{pmatrix} -1^{b_1} & -1^{b_2} & \ldots & -1^{b_n} \end{pmatrix}^\top.$$

Now we provide the proof for the characterization of elements in $\mathcal{C}_2^{\text{nat}}$.

Proof of Proposition 1: We show that for any fixed $\boldsymbol{x} \in (0,1)^2$, if $\mathcal{B}(\cdot;\cdot) \in \mathcal{C}_2^{\text{nat}}$, the values $\mathcal{B}(\boldsymbol{R};\boldsymbol{x})$ takes for any vector $\boldsymbol{R} \in \{-1,1\}^2$ coincides with that of $\mathcal{B}_{\text{ALL}}(\boldsymbol{R};\boldsymbol{x})$ given in (6). To do this, we utilize *Assumption (iii)*.

So, consider an arbitrary point $\boldsymbol{x} \in (0,1)^2$. To compute $\mathbb{E}[R_1\mathcal{B}(\boldsymbol{R};\boldsymbol{x})]$, note that $g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0) = g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3)$ and $g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1) = g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2)$. Therefore,

$$\begin{aligned}\mathbb{E}[R_1\mathcal{B}(\boldsymbol{R};\boldsymbol{x})] &= g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)\mathcal{B}(\boldsymbol{\mathcal{R}}_0;\boldsymbol{x}) + g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)\mathcal{B}(\boldsymbol{\mathcal{R}}_1;\boldsymbol{x}) \\ &\quad - g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2)\mathcal{B}(\boldsymbol{\mathcal{R}}_3;\boldsymbol{x}) - g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3)\mathcal{B}(\boldsymbol{\mathcal{R}}_2;\boldsymbol{x}) \\ &= g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)\mathcal{B}(\boldsymbol{\mathcal{R}}_0;\boldsymbol{x}) + g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)\mathcal{B}(\boldsymbol{\mathcal{R}}_1;\boldsymbol{x}) \\ &\quad - g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)\mathcal{B}(\boldsymbol{\mathcal{R}}_3;\boldsymbol{x}) - g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)\mathcal{B}(\boldsymbol{\mathcal{R}}_2;\boldsymbol{x}) \\ &= 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)\mathcal{B}(\boldsymbol{\mathcal{R}}_0;\boldsymbol{x}) + 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)\mathcal{B}(\boldsymbol{\mathcal{R}}_1;\boldsymbol{x}),\end{aligned}$$

where the last step follows from the *Assumption (i)*, $\mathcal{B}(-\boldsymbol{R};\boldsymbol{x}) = -\mathcal{B}(\boldsymbol{R};\boldsymbol{x})$. Similarly we have

$$\mathbb{E}[R_2\mathcal{B}(\boldsymbol{R};\boldsymbol{x})] = 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)\mathcal{B}(\boldsymbol{\mathcal{R}}_0;\boldsymbol{x}) - 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)\mathcal{B}(\boldsymbol{\mathcal{R}}_1;\boldsymbol{x}).$$

Therefore in order for $\mathcal{B}(\cdot;\cdot)$ to satisfy (11), we need to have

$$\begin{pmatrix} 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0) & 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1) \\ 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0) & -2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1) \end{pmatrix} \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} = \begin{pmatrix} 1 - 2x_1 \\ 1 - 2x_2 \end{pmatrix},$$

where $B_i = \mathcal{B}(\boldsymbol{\mathcal{R}}_i; \boldsymbol{x})$ for $i \in \{0,1\}$. Note that for non-degenerate $\boldsymbol{x}$, i.e., if $x_i \notin \{0,1\}$, the above matrix is invertible. Solving the system of linear equations in B_0, B_1 we get

$$\begin{pmatrix} B_0 \\ B_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{4g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)} & \frac{1}{4g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)} \\ \frac{1}{4g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)} & -\frac{1}{4g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)} \end{pmatrix} \begin{pmatrix} 1-2x_1 \\ 1-2x_2 \end{pmatrix} = \begin{pmatrix} \frac{1-x_1-x_2}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)} \\ \frac{x_2-x_1}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)} \end{pmatrix}.$$

We can simplify B_0 as follow

$$B_0 = \frac{1-x_1-x_2}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)} = \frac{(1-x_1)(1-x_2)-x_1x_2}{(1-x_1)(1-x_2)+x_1x_2} = \mathcal{B}_{\text{ALL}}(\boldsymbol{\mathcal{R}}_0; \boldsymbol{x}).$$

Similarly we have $B_1 = \mathcal{B}_{\text{ALL}}(\boldsymbol{\mathcal{R}}_1; \boldsymbol{x})$. ■

Next, we provide the characterization of elements in $\mathcal{C}_3^{\text{nat}}$.

Proof of Proposition 2: As in the proof of Proposition 1 we show that for any fixed $\boldsymbol{x} \in (0,1)^3$, $\mathcal{B}(\cdot;\cdot) \in \mathcal{C}_3^{\text{nat}}$ iff the value $\mathcal{B}(\boldsymbol{R}; \boldsymbol{x})$ takes for any vector $\boldsymbol{R} \in \{-1,1\}^2$ satisfies (13). Consider an arbitrary point $\boldsymbol{x} \in (0,1)^3$. To compute $\mathbb{E}_{\boldsymbol{R}\sim g_{\boldsymbol{x}}}[R_i \mathcal{B}(\boldsymbol{R}; \boldsymbol{x})]$ note that $g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_i) = g_{\boldsymbol{\mathcal{R}}_{7-i}}$ for any $i \in \{0,1,2,3\}$. Similar to the proof of Proposition 1, we can express the equations in terms of the values of the functions at $\boldsymbol{\mathcal{R}}_i$ for $i \in \{0,1,2,3\}$ through the equation $\mathcal{H}\boldsymbol{B} = \mathbf{1} - 2\boldsymbol{x}$, where

$$\mathcal{H} = \begin{pmatrix} 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0) & 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1) & 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2) & 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3) \\ 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0) & 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1) & -2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2) & -2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3) \\ 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0) & -2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1) & 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2) & -2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3) \end{pmatrix}$$

and $\boldsymbol{B} = \begin{pmatrix} B_0 & B_1 & B_2 & B_3 \end{pmatrix}^\top$. Here $B_i = \mathcal{B}(\boldsymbol{\mathcal{R}}_i; \boldsymbol{x})$ for $i \in \{0,1,2,3\}$. The matrix $\mathcal{H}$ in one of its row echelon form can be expressed as

$$\begin{pmatrix} 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0) & 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1) & 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2) & 2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3) \\ 0 & -2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1) & 0 & -2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3) \\ 0 & 0 & -2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2) & -2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3) \end{pmatrix}.$$

Therefore, we know that $\mathcal{H}$ is a matrix with rank 3 if $g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_i) \neq 0$ for $i \in \{0,1,2,3\}$. By the rank-nullity theorem [14, eq.(4.4.15)] the dimension of the null-space of $\mathcal{H}$ is 1. Therefore, the nullspace of $\mathcal{H}$ is given by $\text{span}(\boldsymbol{z})$, where

$$\boldsymbol{z} = \begin{pmatrix} \frac{1}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)} & -\frac{1}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)} & -\frac{1}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2)} & \frac{1}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3)} \end{pmatrix}^\top.$$

Therefore, the solution set for $\mathcal{H}\boldsymbol{B} = \mathbf{1} - 2\boldsymbol{x}$ is given by $\{\boldsymbol{b} \in \mathbb{R}^4 : \boldsymbol{b} = \boldsymbol{B}_* + c\boldsymbol{z}\}$, where $\boldsymbol{B}_*$ is one solution to the system of linear equation $\mathcal{H}\boldsymbol{B} = \mathbf{1} - 2\boldsymbol{x}$. We can choose

$$\boldsymbol{B}_* = \begin{pmatrix} \frac{(1-x_1)(1-x_2)(1-x_3)-x_1x_2x_3}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)} \\ \frac{(1-x_1)(1-x_2)x_3-x_1x_2(1-x_3)}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)} \\ \frac{(1-x_1)x_2(1-x_3)-x_1(1-x_2)x_3}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2)} \\ \frac{(1-x_1)x_2x_3-x_1(1-x_2)(1-x_3)}{2g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3)} \end{pmatrix}, \tag{14}$$

whose i-th element is in fact $(\boldsymbol{B}_*)_i = \mathcal{B}_{\text{ALL}}(\boldsymbol{\mathcal{R}}_{i-1}; \boldsymbol{x})$. Therefore the functions satisfying Definition 1 take the form

$$\mathcal{B}(\boldsymbol{R}; \boldsymbol{x}) = \mathcal{B}_{\text{ALL}}(\boldsymbol{R}; \boldsymbol{x}) + \frac{c_{\boldsymbol{x}} R_1 R_2 R_3}{2g_{\boldsymbol{x}}(\boldsymbol{R})},$$

where $c_{\boldsymbol{x}}$ is an arbitrary function of $\boldsymbol{x}$. Furthermore, to ensure $\mathcal{B}(\boldsymbol{R}; \boldsymbol{x}) = -\mathcal{B}(\boldsymbol{R}; 1-\boldsymbol{x})$, we need to have

$$\mathcal{B}_{\text{ALL}}(\boldsymbol{R}; \boldsymbol{x}) + \frac{c_{\boldsymbol{x}} R_1 R_2 R_3}{2g_{\boldsymbol{x}}(\boldsymbol{R})} = -\mathcal{B}_{\text{ALL}}(\boldsymbol{R}; 1-\boldsymbol{x}) - \frac{c_{\mathbf{1}-\boldsymbol{x}} R_1 R_2 R_3}{2g_{\mathbf{1}-\boldsymbol{x}}(\boldsymbol{R})}.$$

As $g_{\boldsymbol{x}}(\boldsymbol{R}) = g_{\mathbf{1}-\boldsymbol{x}}(\boldsymbol{R})$ and $\mathcal{B}_{\text{ALL}}(\boldsymbol{R}; \boldsymbol{x}) = -\mathcal{B}_{\text{ALL}}(\boldsymbol{R}; 1-\boldsymbol{x})$, the above equality holds iff $c_{\boldsymbol{x}} = -c_{\mathbf{1}-\boldsymbol{x}}$. ■

Proof of Proposition 3: From Proposition 1, it readily follows that $\mathcal{B}_{\text{HT}}(\cdot;\cdot) \notin \mathcal{C}_2^{\text{nat}}$ for a two-agents fact-checker system. For any $n \geq 2$, we show that there exists $\boldsymbol{x} \in (0,1)^n$ such that $\mathcal{B}_{\text{HT}}(\cdot;\cdot)$ does not satisfy *Assumption (iii)*.

Consider $\boldsymbol{x}^* \in (0,1)^n$ such that

$$\log \frac{1-x_1^*}{x_1^*} > \sum_{i=2}^{n} \left| \log \frac{1-x_i^*}{x_i^*} \right|. \tag{15}$$

Then, for any $\boldsymbol{R} \in \{-1,+1\}^n$, we have $\mathcal{B}_{\text{HT}}(\boldsymbol{R}; \boldsymbol{x}^*) = R_1$. Therefore $\mathbb{E}[R_1 \mathcal{B}_{\text{HT}}(\boldsymbol{R}; \boldsymbol{x})] = \mathbb{E}[R_1^2] = 1$. However $1 - 2x_1^* < 1$. So, $\mathcal{B}_{\text{HT}}(\boldsymbol{R}; \boldsymbol{x}^*)$ does not satisfy (11), at least for vectors $\boldsymbol{x}$ satisfying (15). ■

In order to prove Theorem 1 for $a,b,c \in (0,1)$, we define a function $h(a,b,c) = abc + \bar{a}\bar{b}\bar{c}$. For convenience, with an abuse of notation, we also define $h(a,b) = ab + \bar{a}\bar{b}$.

Proof of Theorem 1: Using the fact that $x_i \notin \{0,1\}$ for $i \in [3]$, we can perform algebraic manipulations and express the fixed-point equation $\boldsymbol{x} = \frac{1}{2}\left(\mathbf{1} - \mathbb{E}_{\boldsymbol{R}\sim g_{\boldsymbol{\pi}}}[\boldsymbol{R}\mathcal{B}_{\text{ALL}}(\boldsymbol{R}; \boldsymbol{x})]\right)$ as $\mathcal{X}\boldsymbol{u} = 0$, where

$$\mathcal{X} = \begin{pmatrix} x_2+x_3-1 & 1-x_2-x_3 & x_3-x_2 & x_2-x_3 \\ x_3+x_1-1 & x_3-x_1 & 1-x_1-x_3 & x_1-x_3 \\ x_1+x_2-1 & x_2-x_1 & x_1-x_3 & 1-x_1-x_2 \end{pmatrix}$$

and $\boldsymbol{u} = (u_0, u_1, u_2, u_3)^\top$, with $u_0 = \frac{g_{\boldsymbol{\pi}}(\boldsymbol{\mathcal{R}}_0)}{g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0)}$, $u_1 = \frac{g_{\boldsymbol{\pi}}(\boldsymbol{\mathcal{R}}_3)}{g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_3)}$, $u_2 = \frac{g_{\boldsymbol{\pi}}(\boldsymbol{\mathcal{R}}_2)}{g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_2)}$, and $u_3 = \frac{g_{\boldsymbol{\pi}}(\boldsymbol{\mathcal{R}}_1)}{g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_1)}$.

Summing equations in $\mathcal{X}\boldsymbol{u} = \mathbf{0}$ and multiplying by $\frac{1}{2}$, we get

$$\left(\frac{3}{2} - (x_1+x_2+x_3)\right) u_0 = \left(\frac{1}{2} - x_1\right) u_1 + \left(\frac{1}{2} - x_2\right) u_2 + \left(\frac{1}{2} - x_3\right) u_3. \tag{16}$$

Case 1: Consider the case where $x_1 + x_2 + x_3 \neq \frac{3}{2}$. For $i \in [3]$ define $w_i = \frac{\frac{1}{2}-x_i}{\sum_{j=1}^{3} \frac{1}{2}-x_j}$. Then we have

$$u_0 = w_1 u_1 + w_2 u_2 + w_3 u_3, \tag{17}$$

where $\sum_{i=1}^{3} w_i = 1$. Replacing u_0 from (17) in $\mathcal{X}\boldsymbol{u} = 0$, we get

$$\sum\nolimits_{j=1}^{3} w_j u_j - u_i = \frac{w_{i\oplus 2} - w_{i\oplus 1}}{w_{i\oplus 2} + w_{i\oplus 1}} (u_{i\oplus 2} - u_{i\oplus 1}), \tag{18}$$

for $i \in [3]$. Note that, here, $\oplus$ represents summation modulo 3. We can rewrite the above system as

$$u_i = \gamma_i u_{i\oplus 1} + (1-\gamma_i) u_{i\oplus 2}, \qquad i \in [3], \tag{19}$$

where coefficients γ_i are given by

$$\gamma_i = \frac{w_{i\oplus 1}(w_{i\oplus 1} + w_{i\oplus 2}) + (w_{i\oplus 2} - w_{i\oplus 1})}{(1-w_i)^2}, \quad i \in [3]. \tag{20}$$

The system of equation (19) is equivalent to

$$(1 - \gamma_i(1-\gamma_{i\oplus 1}))(u_i - u_{i\oplus 2}) = 0, \qquad i \in [3]. \tag{21}$$

We note that, for any $i \in [3]$, $(1 - \gamma_i(1-\gamma_{i\oplus 1})) = 1$ holds iff $w_1 w_2 w_3 = 0$. Thus, the system of equations in (21) holds only if either $w_1 w_2 w_3 = 0$ (Case 1-1) or $u_1 = u_2 = u_3$ (Case 1-2) .

Case 1-1: Note that $w_1w_2w_3 = 0$ implies that $x_i = \frac{1}{2}$ for some $i \in [3]$. Let us consider the case with $w_1 = 0$, or equivalently $x_1 = \frac{1}{2}$. The system of equations $\mathcal{X}\boldsymbol{u} = \boldsymbol{0}$ can then be simplified to

$$\begin{aligned}
(x_2 + x_3 - 1)\frac{h(\pi_1,\pi_2,\pi_3) - h(\bar{\pi}_1,\pi_2,\pi_3)}{h(\pi_1,\pi_2,\bar{\pi}_3) - h(\pi_1,\bar{\pi}_2,\pi_3)} &= \frac{h(x_2,x_3)}{h(x_2,\bar{x}_3)}(x_3 - x_2),\\
\left(x_3 - \frac{1}{2}\right)(u_0 + u_1 - u_2 - u_3) &= 0, \qquad (22)\\
\left(x_2 - \frac{1}{2}\right)(u_0 + u_1 - u_2 - u_3) &= 0.
\end{aligned}$$

It is clear that $\boldsymbol{x} = \frac{1}{2}\boldsymbol{1}$ is a feasible solution for (22). In the following, we prove that (22) has no other solution.

Let $\boldsymbol{x} \neq \frac{1}{2}\boldsymbol{1}$, and without loss of generality, $x_2 \neq \frac{1}{2}$. Hence, we should have $u_0 + u_1 = u_2 + u_3$. However, we have $u_0 + u_1 = 2\frac{h(\pi_2,\pi_3)}{h(x_2,x_3)}$ and $u_2 + u_3 = 2\frac{h(\pi_2,\bar{\pi}_3)}{h(x_2,\bar{x}_3)}$. Therefore, $u_0 + u_1 = u_2 + u_3$ holds if and only if

$$h(\pi_2, \pi_3) = h(x_2, x_3). \qquad (23)$$

Plugging (23) in (22), we arrive at $(x_2 + x_3 - 1) = \tilde{c}(x_3 - x_2)$, or equivalently,

$$x_2 = \frac{x_3(\tilde{c} - 1) + 1}{1 + \tilde{c}}, \qquad (24)$$

where

$$\begin{aligned}
\tilde{c} &= \frac{h(\pi_2,\pi_3)}{h(\pi_2,\bar{\pi}_3)} \frac{(h(\pi_1,\pi_2,\bar{\pi}_3) - h(\pi_1,\bar{\pi}_2,\pi_3))}{h(\pi_1,\pi_2,\pi_3) - h(\bar{\pi}_1,\pi_2,\pi_3)}\\
&= \frac{h(\pi_2,\pi_3)}{h(\pi_2,\bar{\pi}_3)}\left(\frac{\pi_2 - \pi_3}{\pi_2 + \pi_3 - 1}\right) = \frac{h(\pi_2,\pi_3)}{h(\pi_2,\bar{\pi}_3)}\left(\frac{\pi_2 - \pi_3}{\pi_2 - \bar{\pi}_3}\right).
\end{aligned}$$

Plugging (24) into (23), we get

$$\begin{aligned}
0 &= h(x_2, x_3) - h(\pi_2, \pi_3) = 2x_2x_3 - x_2 - x_3 + 1 - h(\pi_2, \pi_3)\\
&= \frac{\tilde{c} - 1}{2(\tilde{c} + 1)}(2x_3 - 1)^2 + \frac{1}{2} - h(\pi_2, \pi_3). \qquad (25)
\end{aligned}$$

We know $h(\pi_2, \pi_3) = \frac{1}{2}(2\pi_2 - 1)(2\pi_3 - 1) + \frac{1}{2} = 2\tilde{\pi}_2\tilde{\pi}_3 + \frac{1}{2}$, where $\tilde{\pi}_i = \frac{1}{2} - \pi_i \in (-\frac{1}{2}, \frac{1}{2})$ for $i \in [3]$. Moreover, we have

$$\begin{aligned}
\frac{\tilde{c} - 1}{\tilde{c} + 1} &= \frac{h(\pi_2,\pi_3)(\pi_2 - \pi_3) - (1 - h(\pi_2,\pi_3))(\pi_2 + \pi_3 - 1)}{h(\pi_2,\pi_3)(\pi_2 - \pi_3) + (1 - h(\pi_2,\pi_3))(\pi_2 + \pi_3 - 1)}\\
&= \frac{-4\tilde{\pi}_2^2\tilde{\pi}_3 + \tilde{\pi}_3}{4\tilde{\pi}_2\tilde{\pi}_3^2 - \tilde{\pi}_2} = -\frac{\tilde{\pi}_3(4\tilde{\pi}_2^2 - 1)}{\tilde{\pi}_2(4\tilde{\pi}_3^2 - 1)}. \qquad (26)
\end{aligned}$$

Using this in (25), we arrive at

$$\begin{aligned}
0 &= -\frac{1}{2}\frac{\tilde{\pi}_3(4\tilde{\pi}_2^2 - 1)}{\tilde{\pi}_2(4\tilde{\pi}_3^2 - 1)}(2x_3 - 1)^2 - 2\tilde{\pi}_2\tilde{\pi}_3\\
&= -\frac{\tilde{\pi}_3}{2\tilde{\pi}_2}\left(\frac{4\tilde{\pi}_2^2 - 1}{4\tilde{\pi}_3^2 - 1}(2x_3 - 1)^2 + 4\tilde{\pi}_2^2\right).
\end{aligned}$$

This last equation holds if and only if $\tilde{\pi}_3 = 0$. Plugging this in (26) implies $\tilde{c} = 1$, which together with (24) leads to $x_2 = \frac{1}{2}$, which is a contradiction. Hence, the only solution for Case 1-1 is $\boldsymbol{x} \neq \frac{1}{2}\boldsymbol{1}$.

Case 1-2: Next, we study the case of $u_1 = u_2 = u_3$, which together with (16) leads to $u_0 = u_1 = u_2 = u_3 = K$ for some $K \in \mathbb{R}$. Equivalently, we get $g_{\boldsymbol{\pi}}(\boldsymbol{\mathcal{R}}_i) = Kg_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_i)$ for $i \in [3] \cup \{0\}$. Summing up the equations over i, we get $K = 1$, since $g_{\boldsymbol{\pi}}$ and $g_{\boldsymbol{x}}$ are probability mass functions. Therefore we get $g_{\boldsymbol{\pi}}(\boldsymbol{\mathcal{R}}_i) = g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_i)$ for $i \in [3] \cup \{0\}$.

From the definition of the function h we have

$$\begin{aligned}
h(x_1, x_2, x_3) - h(x_1, x_2, \bar{x}_3) &= (1 - 2x_3)(1 - x_1 - x_2),\\
h(x_1, x_2, x_3) - h(x_1, \bar{x}_2, x_3) &= (1 - 2x_2)(1 - x_3 - x_1), \qquad (27)\\
h(x_1, x_2, x_3) - h(\bar{x}_1, x_2, x_3) &= (1 - 2x_1)(1 - x_2 - x_3).
\end{aligned}$$

Using (27) and $g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_0) - g_{\boldsymbol{x}}(\boldsymbol{\mathcal{R}}_i) = g_{\boldsymbol{\pi}}(\boldsymbol{\mathcal{R}}_0) - g_{\boldsymbol{\pi}}(\boldsymbol{\mathcal{R}}_i)$ for $i \in [3]$ we get

$$\frac{\tilde{\pi}_1(\tilde{\pi}_2 + \tilde{\pi}_3)}{\tilde{x}_1(\tilde{x}_2 + \tilde{x}_3)} = \frac{\tilde{\pi}_2(\tilde{\pi}_3 + \tilde{\pi}_1)}{\tilde{x}_2(\tilde{x}_3 + \tilde{x}_1)} = \frac{\tilde{\pi}_3(\tilde{\pi}_1 + \tilde{\pi}_2)}{\tilde{x}_3(\tilde{x}_1 + \tilde{x}_2)}, \qquad (28)$$

where $\tilde{x}_i = \frac{1}{2} - x_i$ and $\tilde{\pi}_i = \frac{1}{2} - \pi_i$ for $i \in [3]$. Further simplifying we get the following set of equations

$$\tilde{x}_1\tilde{x}_2 = \tilde{\pi}_1\tilde{\pi}_2,\ \tilde{x}_2\tilde{x}_3 = \tilde{\pi}_2\tilde{\pi}_3,\ \tilde{x}_3\tilde{x}_1 = \tilde{\pi}_3\tilde{\pi}_1,$$

whose solution is $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) = \pm(\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\pi}_3)$. Equivalently, the solution for $u_0 = u_1 = u_2 = u_3$ is $\boldsymbol{x} = \boldsymbol{\pi}$ or $\boldsymbol{1} - \boldsymbol{\pi}$.

Case 2: $x_1 + x_2 + x_3 = \frac{3}{2}$**.** Then, with $\tilde{x}_i = \frac{1}{2} - x_i$ for $i \in [3]$, the case condition is equivalent to $\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 = 0$. Using this fact in (16), we get $\tilde{x}_1u_1 + \tilde{x}_2u_2 + \tilde{x}_3u_3 = 0$. Thus, the equations in $\mathcal{X}\boldsymbol{u} = \boldsymbol{0}$ can be simplified to

$$\tilde{x}_i\left(4u_i - \sum_{j=0}^{3} u_j\right) = 0, \qquad i \in [3]. \qquad (29)$$

The system in (29) can be satisfied only if one of the following two scenarios holds: (i) if $\tilde{x}_i = 0$ or equivalently, $x_i = \frac{1}{2}$ for some $i \in [3]$. This case has been discussed under Case 1-1, and it is shown that $\boldsymbol{x} = \frac{1}{2}\boldsymbol{1}$ is the only solution; (ii) alternatively, if $\tilde{x}_1\tilde{x}_2\tilde{x}_3 \neq 0$, then for every $i \in [3]$, we should have

$$u_i = \frac{\sum_{j=0}^{3} u_j}{4}. \qquad (30)$$

This set of equations leads to $u_0 = u_1 = u_2 = u_3$, which is studied under Case 1-2. It is shown that $\boldsymbol{x} = \boldsymbol{\pi}$ and $\boldsymbol{1} - \boldsymbol{\pi}$ are the only solutions for Case 1-2. This concludes the proof. ■

Proof of Theorem 2: For any $m \in \mathbb{N}$, for $\boldsymbol{x} \in (0,1)^m$ and $\boldsymbol{R} \in \{-1,+1\}^m$ define $\Pi(\boldsymbol{R};\boldsymbol{x})$ as

$$\Pi(\boldsymbol{R};\boldsymbol{x}) := \prod_{i=1}^{m}\left(x_i\mathbb{1}_{\{R_i=1\}} + (1 - x_i)\mathbb{1}_{\{R_i=-1\}}\right).$$

Note that summing over all possible realizations of $\boldsymbol{R}$ we get

$$\sum_{\boldsymbol{R}\in\{-1,+1\}^m} \Pi(\boldsymbol{R};\boldsymbol{x}) = \prod_{i=1}^{m}(x_i + (1 - x_i)) = 1.$$

For any $\boldsymbol{x} \in (0,1)^n$ and any $\boldsymbol{R} \in \{-1,+1\}^n$ we know that

$$\begin{aligned}
R_i g_{\boldsymbol{x}}(\boldsymbol{R})\mathcal{B}_{\text{ALL}}(\boldsymbol{R};\boldsymbol{x}) &= \frac{1}{2}\left(\Pi(-R_i\boldsymbol{R};\boldsymbol{x}) - \Pi(R_i\boldsymbol{R};\boldsymbol{x})\right)\\
&= \frac{1}{2}\left((1 - x_i)\Pi(-\boldsymbol{R}_{-i};\boldsymbol{x}_{-i}) - x_i\Pi(\boldsymbol{R}_{-i};\boldsymbol{x}_{-i})\right),
\end{aligned}$$

where $\boldsymbol{x}_{-i} \in (0,1)^{n-1}$ and $\boldsymbol{R}_{-i} \in \{-1,+1\}^{n-1}$ are obtained by removing the ith element in $\boldsymbol{x}$ and $\boldsymbol{R}$, respectively.

Therefore, for any $i \in [n]$ we have

$$\mathbb{E}[R_i \mathcal{B}_{\text{ALL}}(\boldsymbol{R};\boldsymbol{x})] = \sum_{\boldsymbol{R}\in\{-1,+1\}^n} g_{\boldsymbol{x}}(\boldsymbol{R}) R_i \mathcal{B}_{\text{ALL}}(\boldsymbol{R};\boldsymbol{x})$$
$$= \frac{1}{2} \sum_{\boldsymbol{R}\in\{-1,+1\}^n} (1-x_i)\Pi(-\boldsymbol{R}_{-i};\boldsymbol{x}_{-i}) - x_i \Pi(\boldsymbol{R}_{-i};\boldsymbol{x}_{-i})$$
$$= (1-x_i) - x_i = 1 - 2x_i,$$

which concludes the proof. ■

REFERENCES

[1] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi. Spread of (mis) information in social networks. *Games and Economic Behavior*, 70(2):194–227, 2010.

[2] D. Acemoglu, A. Ozdaglar, and J. Siderius. Misinformation: Strategic sharing, homophily, and endogenous echo chambers. Technical report, National Bureau of Economic Research, 2021.

[3] C. Budak, D. Agrawal, and A. El Abbadi. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th international conference on World wide web*, pages 665–674, 2011.

[4] S. Chen and A. Moitra. Beyond the low-degree algorithm: mixtures of subcubes and their applications. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 869–880, 2019.

[5] A. P. Dawid and A. M. Skene. Maximum likelihood estimation of observer error-rates using the em algorithm. *J. R. Stat. Soc.: Series C (Applied Statistics)*, 28(1):20–28, 1979.

[6] J. Feldman, R. O'Donnell, and R. A. Servedio. Learning mixtures of product distributions over discrete domains. *SIAM Journal on Computing*, 37(5):1536–1564, 2008.

[7] Y. Freund and Y. Mansour. Estimating a mixture of two product distributions. In *Proceedings of the twelfth annual conference on Computational learning theory*, pages 53–62, 1999.

[8] C. Gao, Y. Lu, and D. Zhou. Exact exponent in optimal rates for crowdsourcing. In *International Conference on Machine Learning*, pages 603–611. PMLR, 2016.

[9] S. Gordon, B. H. Mazaheri, Y. Rabani, and L. Schulman. Source identification for mixtures of product distributions. In *Conference on Learning Theory*, pages 2193–2216. PMLR, 2021.

[10] S. L. Gordon and L. J. Schulman. Hadamard extensions and the identification of mixtures of product distributions. *IEEE Transactions on Information Theory*, 2022.

[11] A. Hanselowski and I. Gurevych. A framework for automated fact-checking for real-time validation of emerging claims on the web. In *NIPS 2017 Workshop on Prioritising Online Content*, pages 1–3, 2017.

[12] A. Hanselowski, C. Stab, C. Schulz, Z. Li, and I. Gurevych. A richly annotated corpus for different tasks in automated fact-checking. *arXiv preprint arXiv:1911.01214*, 2019.

[13] S. Kamath, A. Orlitsky, D. Pichapati, and A. T. Suresh. On learning distributions from their samples. In *Conference on Learning Theory*, pages 1066–1100. PMLR, 2015.

[14] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*, volume 71. SIAM, 2000.

[15] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz. Containment of misinformation spread in online social networks. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 213–222, 2012.

[16] S. Nitzan and J. Paroush. The characterization of decisive weighted majority rules. *Economics Letters*, 7(2):119–124, 1981.

[17] Y. Papanastasiou. Fake news propagation and detection: A sequential model. *Management Science*, 66(5):1826–1846, 2020.

[18] L. Shapley and B. Grofman. Optimizing group judgmental accuracy in the presence of interdependencies. *Public Choice*, 43(3):329–343, 1984.

[19] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal. Fever: a large-scale dataset for fact extraction and verification. *arXiv preprint arXiv:1803.05355*, 2018.

[20] A. Verma, S. Mohajer, and B. Touri. Distributed fact checking:learning unreliability. In *2024 American Control Conference (ACC)*, 2024.

[21] A. Verma, A. Sharbafchi, B. Touri, and S. Mohajer. Distributed fact checking. In *2023 IEEE Int. Symp. on Inform. Theory (ISIT)*, 2023.

[22] Y. Zhang, X. Chen, D. Zhou, and M. I. Jordan. Spectral methods meet em: A provably optimal algorithm for crowdsourcing. *The Journal of Machine Learning Research*, 17(1):3537–3580, 2016.