On the Inconsistency of Estimation with Uncertain Models

César A. Uribe[1], James Zachary Hare[2], Lance M. Kaplan[2], and Ali Jadbabaie[3]
[1]Rice University, Houston, TX, USA
[2]DEVCOM Army Research Laboratory, Adelphi, MD, USA
[3]Massachusetts Institute of Technology, Cambridge, MA, USA
[1]cauribe@rice.edu, [2]{james.z.hare.civ, lance.m.kaplan.civ}@army.mil, [3]jadbabai@mit.edu

2024 58th Asilomar Conference on Signals, Systems, and Computers | 979-8-3503-5405-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/IEEECONF60004.2024.10942985

***Abstract*—We study the hypothesis testing problem where an agent seeks to select a hypothesis from a finite set Θ, based on a sequence of i.i.d. observations of a random variable $X_k \sim P^*$ for $k \geq 1$ with unknown distribution P^*. Each hypothesis $\theta^* \in \Theta$ states that $X_k \sim P_\theta$. The objective is to find a hypothesis θ^* such that $\theta = \arg\min_{\theta\in\Theta} D_{KL}(P^*\|P_\theta)$. However, the set of hypotheses $\{P_\theta\}$ is partially known, where only a finite number of observations are available for the random variables $Y_s^\theta \sim P_\theta$ with $\theta \in \Theta$. We show that contrary to classical Bayesian approaches, the obtained estimator will *not* be consistent, and the aggregated log-likelihood ratios will converge in distribution to a Gaussian distribution even when $k \to \infty$. Our result states that estimators with uncertain likelihoods will not concentrate on the true hypothesis. There is a strictly positive probability that the belief in a suboptimal hypothesis is maximal.**

***Index Terms*—Distributed algorithms, compressed communication, algorithm design and analysis, Bayesian update.**

I. Introduction

Detection theory is a classical problem in signal processing that designs algorithms that identify if a change has occurred in the environment. It has many applications, such as situational awareness [1], event detection [2], target detection [3], etc. An approach to designing a detector is to abstract the problem into a hypothesis testing problem that constructs beliefs for each hypothesis that are proportional to a log-likelihood ratio test, where an agent computes their beliefs sequentially as a stream of observations is realized of the true state of the world [4]. However, a key challenge in this problem is the ability to learn the hypothesis that the true state of the world is in highly dynamic and uncertain environments.

Traditionally, the literature assumes that each hypothesis is modeled as a parameterized distribution, where the parameters are known precisely. However, highly dynamic and uncertain environments make the parameters of events (hypotheses) susceptible to changes, or the events may have been unseen previously and are uncertain. An agent might have a limited amount of prior knowledge (training data) to learn the parameters of the likelihood functions.

A frequentist would solve this challenge by estimating the parameters of the distributions using maximum likelihood estimates [5]. Although this approach has been shown to work well as the amount of prior knowledge becomes large, maximum likelihood estimates are prone to bias when the amount of training data is limited, causing the asymptotic properties to break down [6]. Robust hypothesis testing approaches have been studied from a minimax perspective that constructs Least Favorable Distributions (LFD) over the support of the prior knowledge [7]–[9]. However, computing LFDs is computationally expensive. Additionally, the support of the observations may be outside of the support of the training data when limited data is available, making LFDs inaccurate [10]. Other approaches to managing uncertainty include possibility theory [11], probability intervals [12], and belief theory [13].

A natural approach to handling uncertainty in limited training data is constructing surrogate likelihood models called *Uncertain Models* [14], [15]. This Bayesian approach assumes that the parameters are known within a distribution instead of fixed values. Then, the likelihood of the observations is estimated using a posterior predictive distribution that accurately incorporates the parameter uncertainty while reducing biased estimates.

In this work, we study whether learning with beliefs constructed with uncertain models is possible when the amount of prior knowledge is finite. We find that the agent cannot learn the true state of the world with finite prior knowledge. We prove this statement by considering a simple case in which observations are drawn from a Gaussian distribution with an unknown mean and known variance. We show that beliefs with uncertain models are normally distributed with a finite mean and variance. Only when the amount of prior knowledge grows unboundedly the generated estimator is consistent.

Section presents the overall problem and analyzes the asymptotic analysis of learning or hypothesis testing with uncertain models. Then, Section III validates the explicit characterization of the asymptotic inconsistency of the estimators. Section IV provides conclusions and future work.

II. Inconsistency of Estimation

Assume we have access to realizations of a sequence of i.i.d. random variables $\{X_k\}$ for $k \geq 1$, such that $X_k \sim P^*$ for an unknown distribution P^*. The objective is to select

This research was sponsored by the DARPA Lagrange, Vannevar Bush Fellowship, and OSD LUCI programs.

a hypothesis from a finite set $\Theta = \{\theta_1, \cdots, \theta_d\}$, about the distribution P^*. The hypothesis $\theta \in \Theta$ states that $X_k \sim P_\theta$.

A classical Bayesian approach would suggest updating a set of beliefs $\mu_t(\theta) = P(\theta \mid \{X_k\}_{k=1}^t)$ for each hypothesis by a sequential updating of posteriors of the form

$$\mu_{t+1}(\theta) = \prod_{k=1}^{t+1} P_\theta(X_k)\mu_0(\theta) = P_\theta(X_{t+1})\mu_t(\theta), \quad (1)$$

where $P_\theta(\cdot)$ is the likelihood function of the random variables $\{X_t\}$ conditioned on the hypothesis θ. Moreover, the Bernstein-von Mises theorem [16, Chapter 10.2] shows that under weak regularity assumptions, the beliefs concentrate around $\theta^* = \arg\min_{\theta\in\Theta} D_{KL}(P^*\|P_\theta)$, i.e., $\lim_{t\to\infty} \mu_t(\theta^*) = 1$ almost surely. Therefore, we obtain an asymptotically consistent estimator [17].

Now, under the uncertain model setup, we assume we *do not* have access to the hypotheses $\{P_\theta\}$ for $\theta \in \Theta$. Our partial knowledge of the hypotheses set is built from a set of realizations of a *finite* sequence of random variables $\{Y_s^\theta\}_{s=1}^M$ such that $Y_s^\theta \sim P_\theta$ with $0 < M < \infty$. The realizations of $\{Y_s^\theta\}_{s=1}^M$ are the only information available to the agent to build up an understanding of Θ. We refer to this as the *prior information*.

The main difficulty in learning with uncertain models is that an agent cannot follow a Bayesian approach as in (1) because the likelihood functions $P_\theta(\cdot)$ are not available. Instead, the agent has the realizations of $\{Y_s^\theta\}_{s=1}^M$ from which one can define a surrogate likelihood function as the posterior predictive distribution:

$$\begin{aligned} &P(X_1, \cdots, X_{k+1} \mid \{Y_s^\theta\}_{s=1}^M) \\ &= \int_{m\in\mathbb{R}} P(X_1, \cdots, X_{k+1} \mid m) P(\theta \mid \{Y_s^\theta\}_{s=1}^M) dm \\ &= \prod_{t=0}^{k} P(X_{t+1} \mid \{X_i\}_{i=1}^t, \{Y_s^\theta\}_{s=1}^M). \end{aligned} \quad (2)$$

Here, we show that an estimator of the form

$$\mu_{k+1}(\theta) \propto P(X_{k+1} \mid \{X_i\}_{i=1}^k, \{Y_s^\theta\}_{s=1}^M)\mu_k(\theta), \quad (3)$$

is inconsistent.

A. Problem Setup

For simplicity of exposition, we will consider the setup where $\Theta = \{\theta_1, \theta_2\}$, $X_k \sim \mathcal{N}(\cdot; 0, 1)$, $Y_k^{\theta_1} \sim \mathcal{N}(\cdot; \theta_1, 1)$ and $Y_k^{\theta_2} \sim \mathcal{N}(\cdot; \theta_2, 1)$. Moreover, without loss of generality, we assume that $D_{KL}(P^*\|P_{\theta_1}) < D_{KL}(P^*\|P_{\theta_2})$. Here, we denote $\mathcal{N}(\cdot; \mu, \tau)$ as the Normal distribution function with mean μ and precision τ, i.e.,

$$\mathcal{N}(x; \mu, \tau) = \sqrt{\tau/(2\pi)} \exp\left(-\tau(x-\mu^2)/2\right). \quad (4)$$

Before we proceed, we state three useful facts about Normal distributions:

- $$\begin{aligned} &\mathcal{N}(\{x_i\}_{i=1}^n; \mu, \tau)\mathcal{N}(\mu; \mu_0, \tau_0) \\ &\qquad \propto \mathcal{N}(\mu; \frac{\tau_0\mu_0 + \tau\sum_{i=1}^n x_i}{\tau_0 + n\tau}, \tau_0 + n\tau) \end{aligned} \quad (5a)$$
- $$\begin{aligned} &\int_{\mu\in\mathbb{R}} \mathcal{N}(x; \mu, \tau)\mathcal{N}(\mu; \mu_0, \tau_0) d\mu \\ &\qquad = \mathcal{N}(x; \mu_0, 1/(\tau_0^{-1} + \tau^{-1})) \end{aligned} \quad (5b)$$
- $$\begin{aligned} &D_{KL}(\mathcal{N}(\cdot; \mu_1, \tau_1)\|\mathcal{N}(\cdot; \mu_2, \tau_2)) \\ &\qquad = \frac{1}{2}\log\frac{\tau_1}{\tau_2} + \frac{1/\tau_1 + (\mu_1-\mu_2)^2}{2/\tau_2} - \frac{1}{2}. \end{aligned} \quad (5c)$$

The next proposition states that the surrogate likelihood functions built from the posterior predictive distributions are also Normal functions.

Proposition 1. *Let $Y_s^\theta \sim \mathcal{N}(\cdot; \theta, 1)$ for $s = 1, \cdots, M$, $\theta \in \Theta = \{\theta_1, \theta_2\}$ and $X_k \sim \mathcal{N}(\cdot; 0, 1)$ for $k \geq 1$. Then, for all $x \in \mathbb{R}$, it holds that*

$$P(x \mid \{X_i\}_{i=1}^k, \{Y_s^{\theta_j}\}_{s=1}^M) = \mathcal{N}(x; m_k^\theta, \tau_k^\theta),$$

where

$$m_k^\theta = \frac{\sum_{i=1}^k X_i + \sum_{s=1}^M Y_s^\theta}{k+M} \quad \text{and} \quad \tau_k^\theta = \frac{M+k}{M+k+1}.$$

Proof. Following the definition of the posterior predictive distribution, it holds that.

$$\begin{aligned} &P(x \mid \{X_i\}_{i=1}^k, \{Y_s^\theta\}_{s=1}^M) \\ &= \int_{m\in\mathbb{R}} P(x \mid m) P(m \mid \{X_i\}_{i=1}^k, \{Y_s^\theta\}_{s=1}^M) dm \\ &\propto \int_{m\in\mathbb{R}} \mathcal{N}(x; m, 1) P(\{X_i\}_{i=1}^k \mid m) P(m \mid \{Y_s^\theta\}_{s=1}^M) dm \\ &\overset{(5a)}{=} \int_{m\in\mathbb{R}} \mathcal{N}(x; m, 1)\mathcal{N}(\{X_i\}_{i=1}^k \mid m, 1)\mathcal{N}(m; \frac{1}{M}\sum_{s=1}^M Y_s^\theta, M) dm \\ &\overset{(5a)}{=} \int_{m\in\mathbb{R}} \mathcal{N}(x; m, 1)\mathcal{N}(m; \frac{\sum_{i=1}^k X_i + \sum_{s=1}^M Y_s^\theta}{k+M}, M+k) dm \\ &\overset{(5b)}{=} \mathcal{N}(x; \frac{\sum_{i=1}^k X_i + \sum_{s=1}^M Y_s^\theta}{k+M}, \frac{M+k}{M+k+1}), \end{aligned}$$

and the desired result follows. □

B. Inconsistency Analysis

Following Proposition 1, we are now ready to analyze the asymptotic properties of (3).

Lemma 1. *Let $X_k \sim P^* = \mathcal{N}(\cdot; 0, 1)$ for $k \geq 1$, $Y_s^\theta \sim P_\theta = \mathcal{N}(\cdot; \theta, 1)$ for $1 \leq s \leq M$ for $M < \infty$ and $\theta \in \Theta = \{\theta_1, \theta_2\}$. Moreover, assume $D_{KL}(P^*\|P_{\theta_1}) < D_{KL}(P^*\|P_{\theta_2})$ and uniform prior beliefs, i.e., $\mu_0(\theta) = 1/2$. Then, for any $k \geq 1$, the sequence of believes $\{\mu_k(\theta)\}$ generated by (3) have the following property:*

$$\lim_{k\to\infty} \mathbb{E}\left[\log\frac{\mu_{k+1}(\theta_2)}{\mu_{k+1}(\theta_1)} \mid \mathcal{F}_k\right] \overset{d}{=} W_M \sim \mathcal{N}\left(\hat{m}_M, \frac{1}{\hat{\sigma}_M^2}\right),$$

where $\mathcal{F}_k$ is the filtration generated by $\{X_i\}_{i=1}^k$ and $\{Y_s^{\theta_1,\theta_2}\}_{s=1}^M$, and

$$\hat{m}_M = \frac{1}{2}M\left(\left(\frac{1}{M}\sum_{s=1}^M Y_s^{\theta_1}\right)^2 - \left(\frac{1}{M}\sum_{s=1}^M Y_s^{\theta_2}\right)^2\right)$$

$$\hat{\sigma}_M^2 = \left(\frac{1}{M}\sum_{s=1}^{M} Y_s^{\theta_1} - \frac{1}{M}\sum_{s=1}^{M} Y_s^{\theta_2}\right)^2 M^2 \left(\frac{\pi^2}{6} - \sum_{s=1}^{M}\frac{1}{s^2}\right).$$

Proof. Initially, it follows from (2) that

$$\log\frac{\mu_{k+1}(\theta_2)}{\mu_{k+1}(\theta_1)} = \log\frac{\prod_{t=1}^{k+1} P(X_t \mid \{X_i\}_{i=1}^{t-1}, \{Y_s^{\theta_2}\}_{s=1}^{M})}{\prod_{t=1}^{k+1} P(X_t \mid \{X_i\}_{i=1}^{t-1}, \{Y_s^{\theta_1}\}_{s=1}^{M})}$$
$$= \sum_{t=0}^{k} \log\frac{P(X_{t+1} \mid \{X_i\}_{i=1}^{t}, \{Y_s^{\theta_2}\}_{s=1}^{M})}{P(X_{t+1} \mid \{X_i\}_{i=1}^{t}, \{Y_s^{\theta_1}\}_{s=1}^{M})}. \quad (6)$$

Our proof technique analyzes each of the terms in (6). Let us start with analyzing only one of the terms for an arbitrary time $t \geq 0$. Note that following the same argument as [18, Lemma 6], we have that

$$\mathbb{E}\left[\log\frac{P(X_{t+1} \mid \{X_i\}_{i=1}^{t}, \{Y_s^{\theta_2}\}_{s=1}^{M})}{P(X_{t+1} \mid \{X_i\}_{i=1}^{t}, \{Y_s^{\theta_1}\}_{s=1}^{M})} \mid \mathcal{F}_k\right]$$
$$= D_{KL}(P^* \| \mathcal{N}(\cdot; m_t^{\theta_2}, \tau_t^{\theta_2}) - D_{KL}(P^* \| \mathcal{N}(\cdot; m_t^{\theta_1}, \tau_t^{\theta_1}),$$

where m_t^θ and τ_t^θ are as in Proposition 1.

It follows from (5c) that

$$D_{KL}(P^* \| \mathcal{N}(\cdot \mid m_t^\theta, \tau_t^\theta)$$
$$= -\frac{1}{2}\log\frac{M+t}{M+t+1} + \frac{1}{2}\frac{M+t}{M+t+1}(1 + (m_t^\theta)^2) - \frac{1}{2}.$$

Therefore, we obtain

$$D_{KL}(P^* \| \mathcal{N}(\cdot; m_t^{\theta_1}, \tau_t^{\theta_1}) - D_{KL}(P^* \| \mathcal{N}(\cdot; m_t^{\theta_2}, \tau_t^{\theta_2})$$
$$= \frac{1}{2}\frac{M+t}{M+t+1}(m_t^{\theta_1})^2 - \frac{1}{2}\frac{M+t}{M+t+1}(m_t^{\theta_2})^2$$
$$= \frac{1}{2}\frac{M+t}{M+t+1}\times$$
$$\times\left(\left(\frac{\sum_{i=1}^{t} X_i + \sum_{s=1}^{M} Y_s^{\theta_1}}{t+M}\right)^2 - \left(\frac{\sum_{i=1}^{t} X_i + \sum_{s=1}^{M} Y_s^{\theta_2}}{t+M}\right)^2\right).$$

Moreover using the fact that $(a+b_1)^2 - (a+b_2)^2 = 2a(b_1 - b_2) + (b_1^2 - b_2^2)$, we have

$$D_{KL}(P^* \| \mathcal{N}(\cdot; m_t^{\theta_1}, \tau_t^{\theta_1})) - D_{KL}(P^* \| \mathcal{N}(\cdot; \mid m_t^{\theta_2}, \tau_t^{\theta_2})$$
$$= \frac{1}{2}\frac{M+t}{M+t+1}\left(2\frac{\sum_{i=1}^{t} X_i}{t+M} \times \frac{\sum_{s=1}^{M} Y_s^{\theta_1} - \sum_{s=1}^{M} Y_s^{\theta_2}}{t+M} + \right.$$
$$\left. + \left(\frac{\sum_{s=1}^{M} Y_s^{\theta_1}}{t+M}\right)^2 - \left(\frac{\sum_{s=1}^{M} Y_s^{\theta_2}}{t+M}\right)^2\right). \quad (7)$$

Let's focus on the second term on the right-hand of 7.

$$\frac{1}{2}\frac{M+t}{M+t+1}\left(\left(\frac{\sum_{s=1}^{M} Y_s^{\theta_1}}{t+M}\right)^2 - \left(\frac{\sum_{s=1}^{M} Y_s^{\theta_2}}{t+M}\right)^2\right)$$
$$= \frac{1}{2}\left(\left(\sum_{s=1}^{M} Y_s^{\theta_1}\right)^2 - \left(\sum_{s=1}^{M} Y_s^{\theta_2}\right)^2\right)\frac{1}{(M+t+1)(M+t)}.$$

Also, note that $\lim_{k\to\infty}\sum_{t=0}^{k} 1/(M+t+1)(M+t) = 1/M$.

Now, let us analyze the behavior of the first term on the right-hand of 7 as

$$\frac{M+t}{M+t+1}\left(\frac{\sum_{i=1}^{t} X_i}{M+t}\right)\left(\frac{\sum_{s=1}^{M} Y_s^{\theta_1} - \sum_{s=1}^{M} Y_s^{\theta_2}}{M+t}\right)$$
$$= \left(\sum_{s=1}^{M} Y_s^{\theta_1} - \sum_{s=1}^{M} Y_s^{\theta_2}\right)\frac{1}{(M+t+1)(M+t)}\sum_{i=1}^{t} X_i$$

We are interested in understanding the following limit, where we can ignore the term $t = 0$, as this is the case when the first observation X_1 is made:

$$\lim_{k\to\infty}\sum_{t=1}^{k}\frac{1}{(M+t+1)(M+t)}\sum_{i=1}^{t} X_i = \lim_{k\to\infty}\sum_{t=1}^{k}\frac{X_t}{M+t}.$$

Note that individually, each $X_t \sim \mathcal{N}(\cdot; 0, 1)$. Thus, it holds that

$$\sum_{t=1}^{k}\frac{X_t}{M+t} \sim \mathcal{N}\left(0, 1/\sum_{t=1}^{k}\frac{1}{(M+t)^2}\right).$$

Additionally, we can use the fact that

$$\lim_{k\to\infty}\sum_{t=1}^{k}\frac{1}{(M+t)^2} = \frac{\pi^2}{6} - \sum_{s=1}^{M}\frac{1}{s^2}.$$

Therefore, it follows that

$$\lim_{k\to\infty}\sum_{t=1}^{k}\frac{X_t}{M+t} = \hat{X}_t \sim \mathcal{N}\left(0, 1/\left(\frac{\pi^2}{6} - \sum_{s=1}^{M}\frac{1}{s^2}\right)\right),$$

and the desired result follows. □

Lemma 1 states that the expected aggregated log-likelihood ratio for the surrogate likelihoods described in (1) converge in distribution to a random variable with a Normal distribution. Moreover, Lemma 1 explicitly characterizes the mean and precision of such asymptotic random variable. Additionally, the mean of the limiting distribution, i.e., $\hat{m}_M$, grows linearly with the amount of prior data M. This corresponds to the behavior of classical Bayesian learning since $M \to \infty$ will imply perfect knowledge of the hypotheses set. Thus, the log-likelihood ratio will grow unbounded, implying that the beliefs on θ_2 will go to zero. In Lemma 1, the variance $\hat{\sigma}_M^2$ also grows linearly with M. The first quadratic term in $\hat{\sigma}_M^2$ is $O(1)$, the second factor is M^2, but the third factor is $O(1/M)$ by the fact that $\frac{\pi^2}{6} - \frac{1}{n} \leq \sum_{i=1}^{n}\frac{1}{k^2} \leq \frac{\pi^2}{6} - \frac{1}{n+1}$. Importantly, the standard deviation $\hat{\sigma}_M$ will grow as $O(1/\sqrt{M})$. Therefore, for any $\alpha \in \mathbb{R}$, $\lim_{M\to\infty} P(W_M \leq \alpha) = 0$. We are now ready to state the main result of this paper.

Theorem 2. *Let $X_k \sim P^* = \mathcal{N}(\cdot; 0, 1)$ for $k \geq 1$, $Y_s^\theta \sim P_\theta = \mathcal{N}(\cdot; \theta, 1)$ for $1 \leq s \leq M$ for $M < \infty$ and $\theta \in \Theta = \{\theta_1, \theta_2\}$. Moreover, assume $D_{KL}(P^* \| P_{\theta_1}) < D_{KL}(P^* \| P_{\theta_2})$ and uniform prior beliefs, i.e., $\mu_0(\theta) = 1/2$. Then, for any $k \geq 1$, the sequence of believes $\{\mu_k(\theta)\}$ generated by (3) have the following property: there exists a $\delta > 0$ such that $P(\lim_{k\to\infty}\log(\mu_k(\theta_2)/\mu_k(\theta_1)) > 0) > \delta$.*

Proof. It follows from [18, Theorem 1], that $\mu_k(\theta_1) \to$

0 almost surely, if and only if the random variable $\log \frac{\mu_k(\theta_2)}{\mu_k(\theta_1)} \to -\infty$ as $k \to \infty$. However, Lemma 1 states that for any finite $M < \infty$, there exists a strictly positive probability such that $\lim_{k\to\infty} \mathbb{E}\left[\log \frac{\mu_{k+1}(\theta_2)}{\mu_{k+1}(\theta_1)} \mid \mathcal{F}_k\right] < \infty$. Thus, with a non-zero probability $\mu_k(\theta_2) > 0$. Similarly, with a non-zero probability $\mu_k(\theta_2) > \mu_k(\theta_1)$.

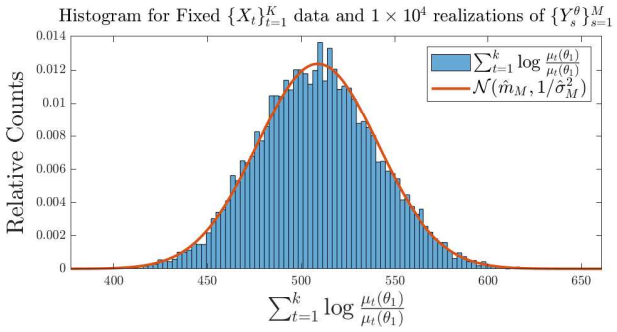
□

Theorem 2 implies that the estimator generated by the Bayesian posteriors with uncertain models of the form of posterior predictive distributions as surrogate likelihood functions are *inconsistent*. The beliefs generated by (3) will not concentrate around θ_1 with a strictly positive probability for any finite M.

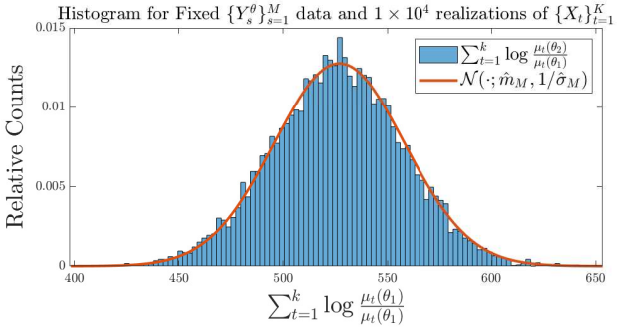
III. Numerical Analysis

This section shows the results of two independent numerical experiments that support the results provided in Lemma 1 and Theorem 2. We consider a sequence of $K = 1 \times 10^7$ observations, with $M = 1 \times 10^3$ prior data on two hypotheses. We set $X_k \sim P^* = \mathcal{N}(\cdot; 0, 1)$, $Y_s^{\theta_1} \sim P_{\theta_1} = \mathcal{N}(\cdot; 0, 1)$, and $Y_s^{\theta_2} \sim P_{\theta_2} = \mathcal{N}(\cdot; 1, 1)$.

Figure 1 shows the histogram of the cumulative log-likelihood ratio of the beliefs generated by (3). Specifically, in Fig. 1a, we fixed the set of observations as $\{X_t = x_t\}_{t=1}^K$, and run 1×10^{-4} Monte Carlo runs for the set of observations $\{Y_s^{\theta_1}, Y_s^{\theta_2}\}_{s=1}^M$. In Fig. 1b, we fixed set of prior data $\{Y_s = y_s\}_{s=1}^M$ and 1×10^{-4} Monte Carlo runs for the set of observations $\{X_t\}_{t=1}^K$. The obtained histogram matches the behavior of the cumulative log-likelihood ratio predicted by Lemma 1, which approximates a Normal distribution $\mathcal{N}(\cdot; \hat{m}_M, 1/\hat{\sigma}_M)$.



(a) Fixed set of observations $\{X_t = x_t\}_{t=1}^K$ and 1×10^{-4} Monte Carlo runs for the set of observations $\{Y_s^{\theta_1}, Y_s^{\theta_2}\}_{s=1}^M$.



(b) Fixed set of prior data $\{Y_s = y_s\}_{s=1}^M$ and 1×10^{-4} Monte Carlo runs for the set of observations $\{X_t\}_{t=1}^K$.

Fig. 1: Cumulative log-likelihood ratio of the beliefs generated by (3)

IV. Conclusions And Future Work

We considered the hypothesis testing problem with uncertain models, and each hypothesis is modeled as an unknown parameterized distribution. Unlike the traditional theory, which assumes the parameters are known, these environments lead to finite prior knowledge for each hypothesis. Previous work has found that using uncertain models, i.e., posterior predictive distributions, allows for an accurate estimate of the likelihood function and an unbiased estimate of the likelihood. However, we found that the generated estimator is inconsistent if uncertain models are used and the amount of prior knowledge is limited. Future work should study how to generalize to the problem of (i) multi-agent systems where agents' beliefs will be combined using non-Bayesian social learning fusion methods and (ii) scenarios where observations are drawn from the general exponential family of distributions.

References

[1] Arslan Munir, Alexander Aved, and Erik Blasch, "Situational awareness: Techniques, challenges, and prospects," *AI*, vol. 3, no. 1, pp. 55–77, 2022.

[2] Manzhu Yu, Myra Bambacus, Guido Cervone, Keith Clarke, Daniel Duffy, Qunying Huang, Jing Li, Wenwen Li, Zhenlong Li, Qian Liu, et al., "Spatiotemporal event detection: A review," *International Journal of Digital Earth*, vol. 13, no. 12, pp. 1339–1365, 2020.

[3] Fangzhou Wang, Pu Wang, Xin Zhang, Hongbin Li, and Braham Himed, "An overview of parametric modeling and methods for radar target detection with limited data," *IEEE Access*, vol. 9, pp. 60459–60469, 2021.

[4] Juliet Popper Shaffer, "Multiple hypothesis testing," *Annual review of psychology*, vol. 46, no. 1, pp. 561–584, 1995.

[5] Edward J Kelly, "An adaptive detection algorithm," *IEEE transactions on aerospace and electronic systems*, , no. 2, pp. 115–127, 1986.

[6] Ofer Zeitouni, Jacob Ziv, and Neri Merhav, "When is the generalized likelihood ratio test optimal?," *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1597–1602, 1992.

[7] Divyanshu Vats, Vishal Monga, Umamahesh Srinivas, and José MF Moura, "Scalable robust hypothesis tests using graphical models," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 3612–3615.

[8] Gökhan Gül and Abdelhak M Zoubir, "Robust hypothesis testing for modeling errors," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 5514–5518.

[9] Gökhan Gül and Abdelhak M Zoubir, "Minimax robust hypothesis testing," *IEEE Transactions on Information Theory*, vol. 63, no. 9, pp. 5572–5587, 2017.

[10] Jie Wang and Yao Xie, "A data-driven approach to robust hypothesis testing using sinkhorn uncertainty sets," *arXiv preprint arXiv:2202.04258*, 2022.

[11] Didier Dubois and Henri Prade, "Possibility theory, probability theory and multiple-valued logics: A clarification," *Annals of mathematics and Artificial Intelligence*, vol. 32, no. 1, pp. 35–66, 2001.

[12] Jean-Marc Bernard, "An introduction to the imprecise dirichlet model for multinomial data," *International Journal of Approximate Reasoning*, vol. 39, no. 2-3, pp. 123–150, 2005.

[13] Glenn Shafer, *A mathematical theory of evidence*, vol. 42, Princeton university press, 1976.

[14] James Z Hare, Cesar A Uribe, Lance Kaplan, and Ali Jadbabaie, "Non-

bayesian social learning with uncertain models," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4178–4193, 2020.
[15] James Z Hare, César A Uribe, Lance Kaplan, and Ali Jadbabaie, "A general framework for distributed inference with uncertain models," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 7, pp. 392–405, 2021.
[16] Aad W Van der Vaart, *Asymptotic statistics*, vol. 3, Cambridge university press, 2000.
[17] Ildar Abdulovich Ibragimov and Rafail Zalmanovich Has' Minskii, *Statistical estimation: asymptotic theory*, vol. 16, Springer Science & Business Media, 2013.
[18] A. Nedić, A. Olshevsky, and C. Uribe, "Fast Convergence Rates for Distributed Non-Bayesian Learning," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5538–5553, 2017.